UNIVERSIDAD DE GRANADA

Departamento de Didáctica de la Matemática

SIGNIFICADO DE LOS INTERVALOS DE CONFIANZA PARA LOS ESTUDIANTES DE INGENIERÍA EN MÉXICO

TESIS DOCTORAL

Presentada por: Eusebio Olivo Suárez

Directora: Dra. Carmen Batanero Bernabeu

Granada, 2008

Editor: Editorial de la Universidad de Granada
Autor: Eusebio Olivo Suárez
D.L.: Gr. 2595-2008
ISBN: 978-84-691-7889-8

SIGNIFICADO DE LOS INTERVALOS DE CONFIANZA PARA LOS ESTUDIANTES DE INGENIERÍA EN MÉXICO

Tesis Doctoral

MEMORIA realizada bajo la dirección dela
Dra. D. Carmen Batanero Bernabeu, que presenta
D. Eusebio Olivo Suárez
para optar al grado de Doctor

Fdo: Eusebio Olivo Suárez   V° B° Carmen Batanero Bernabeu

La investigación presentada en esta tesis doctoral ha sido cofinanciada por el Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Monterrey y la Fundación Carolina, entidad dependiente del Ministerio de Asuntos Exteriores, España.

Agradecimientos

En primer lugar quiero agradecer a mi directora de tesis, la Dra. Carmen Batanero por su sabiduría, paciencia y fortaleza inquebrantable que en todo momento me impulsaron y marcaron el rumbo en la realización de este trabajo de investigación.

A mi esposa Rebeca y a mis hijos por su amor, comprensión y estímulo.

A la Dra. Carmen Díaz Batanero, por su asesoramiento en la construcción del cuestionario y su ayuda en el análisis de los datos.

A mis compañeros del programa de doctorado del Departamento de Didáctica de la Matemática de la Universidad de Granada por su apoyo en los seminarios de investigación.

Al Ing. Tomás Sánchez, Oscar Villarreal, Germán Rodríguez y Juan Antonio López que generosamente colaboraron con nuestro estudio en la aplicación del cuestionario a sus alumnos en el semestre agosto-diciembre del 2007.

A mis amigos: Blanca Ruiz, Armando Albert y Olivia Carrillo por sus valiosos comentarios.

Al grupo de investigadores que colaboraron en el panel de expertos, por su seriedad, profesionalismo y valiosas sugerencias que contribuyeron a darle mayor calidad y solidez al trabajo de investigación.

El trabajo se ha realizado en el marco del proyecto SEJ2007-60110/EDUC, financiado por el MEC-FEDER.

INDICE

INTRODUCCIÓN

En una investigación didáctica es importante analizar en profundidad el significado de un objeto matemático en la institución de referencia, identificar sus elementos característicos y valorar cuáles de ellos y con qué intensidad y secuenciación deben introducirse en un proceso didáctico, cuando se pretende respetar en la enseñanza las características esenciales de dicho objeto matemático. Este análisis es particularmente necesario en el caso de la inferencia estadística, donde tanto las investigaciones psicológicas como las didácticas han llamado la atención sobre los errores frecuentes en su uso e interpretación, especialmente en relación al contraste de hipótesis (Morrison y Henkel, 1970; Vallecillos, 1994, 1999; Harlow, Mulaik y Steiger, 1997; Batanero, 2000; Díaz y de la Fuente, 2004).

Nuestro trabajo de investigación viene motivado por la anterior problemática, persigue, como objetivo central, *generar información sobre el significado personal que los estudiantes de las carreras de ingeniería asignan a los intervalos de confianza*, como un primer paso para explicar los posibles errores y dificultades y, posteriormente diseñar propuestas didácticas que permitan superarlos.

Más específicamente, nuestra investigación se orienta a la construcción de un instrumento de evaluación riguroso de dicho significado personal, apoyándonos en el enfoque ontosemiótico de la cognición matemática (EOS) propuesto por Godino y colaboradores (Godino y Batanero, 1994; Godino, 2002; Godino, 2003, Godino, Contreras y Font, 2006, Godino, Batanero y Font, 2007). La elección del tema se basa en que los intervalos de confianza constituye uno de los métodos sugeridos por diferentes asociaciones profesionales, como la American Psychological Association (APA) o la American Educational Research Association (AERA), para complementar los contrastes de hipótesis y mejorar de este modo los errores denunciados en la práctica de la inferencia estadística (Cumming, Williams y Fidler, 2004). Por otro lado, la investigación didáctica centrada en el intervalo de confianza es todavía muy incipiente.

Fundamentados en el marco teórico citado, el estudio histórico y de las investigaciones previas, realizaremos un análisis de contenido detallado de una muestra de libros de textos orientados a la enseñanza de estadística en ingeniería para llevar a cabo una definición semántica rigurosa de los *elementos de significado* que componen la variable objeto de medición. Una vez puesto a punto y validado el cuestionario, los resultados de aplicarlo a una muestra de 252 estudiantes nos permitirán analizar los conflictos semióticos que surgen en la comprensión y aplicación de este objeto matemático.

La Memoria se organiza en siete capítulos iniciándose en el Capítulo 1 con una contextualización de los estudios de estadística y su importancia para la formación de los ingenieros y los objetivos e hipótesis de la investigación. Se incluye en este capítulo un primer análisis del objeto "intervalo de confianza" cuyo significado institucional de referencia quedará detallado en los capítulos sucesivos.

En el segundo Capítulo presentamos los fundamentos del estudio, comenzando por un estudio histórico de la evolución de las ideas relacionadas con la estimación, que pasa por la estimación puntual y la estimación bayesiana mediante intervalos de credibilidad, culminando con la aparición de la estimación mediante intervalo de confianza. Seguidamente describimos brevemente las nociones del marco teórico del "enfoque ontosemiótico" que proponen Godino y colaboradores y que usaremos a lo largo de nuestro trabajo. Presentamos también en este capítulo una revisión bibliográfica de las más importantes investigaciones que se centran alrededor de la enseñanza y aprendizaje de los intervalos de confianza o de otros temas vinculados

En el Capítulo 3 recogemos la parte empírica orientada a describir el significado institucional de referencia del intervalo de confianza en nuestro trabajo, que determinaremos a partir del análisis de una muestra de 11 libros de texto recomendados en las asignaturas de estadística para ingenieros en el Sistema Tecnológico de Monterrey. Una vez descrita la metodología, se identifican los diferentes elementos de significado del concepto, analizando su presencia en los textos analizados con vista a tomar posteriormente los elementos más frecuentes como base de la construcción del cuestionario.

En el Capítulo 4 describimos el proceso de construcción del cuestionario de evaluación que es el principal producto de nuestro trabajo. Siguiendo la metodología sugerida por Díaz (2007), comenzamos por la definición semántica de la variable objeto de medición, a partir del análisis de contenido llevado a cabo en el Capítulo 3.

Elaborada la tabla de especificaciones del cuestionario, se describe la recopilación, traducción y depuración de un banco inicial de ítems, y la selección final de unidades de contenido e ítems para cada una de ellas a partir de juicio de expertos y pruebas empíricas de los ítems con muestras piloto de estudiantes.

El Capítulo 5 recoge los indicios de la fiabilidad y validez de nuestro instrumento de medición para comprobar que era útil para los objetivos pretendidos, siempre en la hipótesis de que el proceso de validación ha de ser continuo y no se agota en las pruebas iniciales (Meliá, 2001). Las evidencias de validez se recogen por tres métodos diferentes: validez de contenido (justificada mediante análisis teórico de los ítems y juicio de expertos), validez discriminativa (mediante análisis de diferencia de ejecución en los ítems en los grupos de puntuación superior e inferior) y validez de constructo (analizando los factores determinados en el análisis factorial y comparando con nuestras expectativas). La aproximación a la fiabilidad se realiza desde la teoría clásica de los tests, calculando el coeficiente Alfa (Martínez Arias, 1995; Barbero, 2003; Muñiz, 2004), basados en los resultados del análisis factorial (coeficiente theta; Carmines y Zeller, 1979; Barbero, 2003); además se hace una aproximación al problema de la fiabilidad desde la teoría de la generalizabilidad (López Féal, 1986; Feldt y Brennan, 1991), obteniéndose valoraciones de generalizabilidad a otros ítems y otros estudiantes.

En el Capítulo 6 se realiza un estudio detallado de evaluación a través de las respuestas al cuestionario de una muestra de 252 estudiantes de ingeniería, después de haber estudiado el tema. El análisis de frecuencia de respuestas a ítems de opciones múltiples y su comparación con los resultados obtenidos en investigaciones previas se complementa con el análisis semiótico de los ítems de respuesta abierta. Dicho análisis exhibe, por un lado, la complejidad de las respuestas correctas y los distintos objetos matemáticos que el estudiante ha de asociar entre sí por medio de funciones semióticas y por otro permite identificar conflictos semióticos no descritos en las investigaciones previas.

En el Capítulo 7, correspondiente a las conclusiones, revisamos nuestros objetivos e hipótesis iniciales y además señalamos líneas de continuación de este trabajo. Los anexos incluyen datos sobre los resultados en la prueba piloto, cuestionario para recogida de datos de expertos y el cuestionario definitivo.

En resumen nuestro trabajo analiza, en primer lugar la complejidad del significado institucional de referencia del objeto “intervalo de confianza”, ampliando en este análisis la tradicional división entre comprensión conceptual y procedimental del tema.

Aportamos también un cuestionario construido de forma objetiva y rigurosa que permite evaluar el significado personal de los estudiantes sobre dicho objeto matemático y que complementa otros cuestionarios previos, en particular el de Behar (2001), algunos de cuyos ítems se han incluido en nuestro estudio. Describimos asimismo el significado personal que adquieren los alumnos en una institución específica sobre los intervalos de confianza, comparándolo con el significado institucional de referencia y describiendo algunos nuevos conflictos semióticos. Otra aportación es el estado de la cuestión preparado. Toda esta información recogida puede ser de utilidad a otros investigadores que se interesen por el tema y permite definir líneas de investigación para continuar el camino emprendido.

CAPÍTULO 1
CONTEXTUALIZACIÓN DEL PROBLEMA DE INVESTIGACIÓN

1.1. INTRODUCCIÓN

Nuestro trabajo de investigación se centra en la comprensión que adquieren los alumnos sobre los intervalos de confianza, tema que es parte fundamental del material impartido en los cursos de probabilidad y estadística en todas las carreras universitarias. En particular nuestra investigación se focaliza en en las carreras de ingeniería y nos interesamos especialmente por las ofrecidas en el Sistema Multicampus del Tecnológico de Monterrey, aunque los resultados esperamos que puedan extenderse a otros centros universitarios. El interés por este tema proviene de nuestra propia experiencia como profesor de estadística en ingeniería, tanto por las dificultades de aprendizaje que hemos encontrado por parte de los alumnos, como por la búsqueda de elementos que puedan mejorar nuestra enseñanza, y en consecuencia el aprendizaje de nuestros alumnos.

Nuestra investigación continúa otras realizadas en el Grupo de Educación Estadística de la Universidad de Granada enseñanza a nivel universitario (Estepa, 1993; Vallecillos, 1994; Sánchez-Cobo, 1999; Tauber, 2001; Alvarado, 2004, 2007). Para contextualizarla, empezamos en este capítulo dando una perspectiva internacional de las carreras de Ingeniería desde la óptica de las acreditaciones, centrándonos luego en la perspectiva en México y Sistema Tecnológico de Monterrey y analizando la importancia de la estadística en la formación del ingeniero.

En segundo lugar, incluímos un análisis inicial de la configuración matemática conocida como “intervalo de confianza”, mostrando los objetos matemáticos que lo componen o se relacionan con los mismos y describiendo el procedimiento general de construcción. La finalidad es mostrar la complejidad sistémica del tema que servirá posteriormente para explicar posibles dificultades de los alumnos. Seguidamente analizaremos la importancia del objeto de aprendizaje intervalos de confianza en la estadística y discutiremos los retos que en la dimensión didáctica enfrenta su enseñanza. Finalizamos este capítulo describiendo los objetivos de la investigación y las hipótesis iniciales que la guían.

1.2. PERSPECTIVA DE LAS CARRERAS DE INGENIERÍA.

1.2.1. INTRODUCCIÓN

Según la Acreditation Board for Engineering Technology (ABET, 1997) la Ingeniería es la profesión en la cual el conocimiento de las ciencias naturales y matemáticas, obtenido por estudio, experiencia y práctica, es aplicado con criterio al desarrollo de formas de emplear, económicamente, los materiales y fuerzas de la naturaleza para el beneficio de la humanidad.

Los objetivos de la carrera de ingeniería, como de cualquier otra son consistentes con la visión de la misma desde la institución y son periódicamente evaluados por la institución o por organismos externos. Esta evaluación, realizada en forma continua, generalmente se traduce en políticas universitarias y programas de acción encaminados a mejorar la calidad de la enseñanza, en términos de su contribución al desarrollo personal de los alumnos y al bienestar y desarrollo sostenible de la comunidad a la que impacta la institución universitaria.

1.2.2. PANORAMA INTERNACIONAL

Las universidades son conscientes que la acreditación por un organismo de prestigio contribuye a asegurar que sus graduados están preparados adecuadamente para la práctica de la ingeniería, estimula la mejora de la educación y fomenta nuevos e inovativos enfoques de la misma. De ahí el interés por lograr la acreditación de sus programas educativos.

En particular las carreras de ingeniería establecen como prioridad que organismos como la Acreditation Board for Engineering Technology (ABET) certifiquen la calidad de sus programas. Dicha institución había acreditado 1613 programas de ingeniería en 328 instituciones y 714 programas de ingeniería con formación tecnológica en 238 instituciones hasta 2001 (ABET, 2001) y actualmente son 2700 los programas acreditados. La acreditación que proporciona dicho organismo establece dos esquemas distintos que prevalecen en las carreras de ingeniería:

1. Carreras con extensa base científica (engineering programs), cuyos programas en su mayoría tienen una duración entre 4 y 5 años, que están orientadas al diseño, gestión y producción.
2. Carreras con base científica limitada (engineering technology programs), con una buena dósis de formación tecnológica, cuyos programas tienen una duración entre 3 y 4 años, que van orientadas básicamente a la supervisión y producción.

En el primero de los esquemas el curriculo matemático incluye cálculo de una y varias variables, ecuaciones diferenciales, probabilidad y estadística y además el graduado de esta carrera debe demostrar suficiencia en física basada en cálculo entre otras habilidades[1] El segundo esquema está más orientado a las aplicaciones tecnológicas, supervisión y control. Ejemplos de estos esquemas son el Chartered Engineers (orientación científica) y los Incorporated Engineers (orientación tecnológica) en Inglaterra; los grados de Bachelor of Science in Engineering y Bachelor of Engineering Technology en los Estados Unidos de Norteamérica.

Merece mención aparte el SECAI (1998) -Sistema de Evaluación de la Calidad de las enseñanzas de Ingeniería- que integra un conjunto de conceptos, metodología e instrumentos que tienen por objeto el análisis sistemático de los factores asociados a la calidad de las enseñanzas de ingeniería y su evaluación. SECAI fue desarrollado en el marco Columbus, creado por la Conferencia de Rectores, Cancilleres y Vicecancilleres de Universidades Europeas (CRE) para promover la colaboración entre estas instituciones y las universidades latinoamericanas. Dicho programa constituyó un grupo de expertos en evaluación de la enseñanza de la ingeniería procedentes de las siguientes instituciones: Escuela Politécnica Nacional de Ecuador; Imperial College of Science, Technologie and Medicine del Reino Unido; Instituto Tecnológico y de Estudios Superiores de Monterrey, México; Politécnico di Torino, Italia; Pontificia Universidad Javeriana de Bogotá, Colombia; Technische Hochschule Darmstadt, Alemania; Universidad Católica de Valparaiso, Chile;Universidad de Río de Janeiro, Brasil; Universidad Nacional del Mar de Plata, Argentina y Universidad Politécnica de

[1] ABET (2005). 2006-2007 Criteria for Accrediting Engineers Programs (pp. 6-11), On line: www.abet.org/forms.shtml

Madrid, España. Este grupo de expertos desarrolló un sistema completo de evaluación de la calidad de la enseñanza aplicable a las carreras de ingeniería y fácilmente adaptable a otras carreras universitarias. Otros organismos que realizan también estas funciones son el Consejo de Acreditación de la Enseñanza de la Ingeniería (CACEI, 2007) y el Canadian Council of Profesional Engineers (CEAB).

1.2.3. ESTUDIOS DE INGENIERIA EN MÉXICO Y EL ITESM.

La práctica de la ingeniería en México requiere de especialistas en obra civil, química, sistemas computacionales, electrónica y comunicaciones, etc. que estén mejor preparados y que puedan afrontar los rápidos cambios que se están produciendo en sus campos de especialidad. La preparación en su campo es muy importante para incrementar la productividad industrial, la eficiencia, la calidad y con todo ello mejorar los estándares de vida de la población, especialmente teniendo en cuenta los cambios generados en la mayoría de sus actividades diarias, como el uso de computadoras en oficinas, control automático en la industria, telecomunicaciones, etc.

Un número importante de las universidades que en México ofrecen estas carreras de ingeniería, como la Universidad Nacional Autónoma de México (UNAM), Universidad de las Américas (en Puebla), Universidad de Guadalajara y en particular el Sistema Tecnológico de Monterrey están fuertemente preocupados porque sus programas educativos estén acreditados.

El Tecnológico de Monterrey, tiene actualmente 33 campus en México y una Universidad Virtual con presencia en México y otros países. Está acreditado por la Comisión de Universidades y Escuelas del Sur de Estados Unidos (SACS), para otorgar títulos profesionales y grados académicos de maestría y doctorado y por la Federación de Instituciones Mexicanas Particulares de Educación Superior (FIMPES). Los programas de ingeniería están acreditados a nivel nacional por el Consejo de Acreditación de la Enseñanza de la Ingeniería (CACEI) y en el nivel internacional por: el Acreditation Board for Engineering and Technology (ABET). En un documento publicado en 2005 establecen los objetivos que deberán normar las actividades de toda la comunidad en los próximos 10 años (ITESM, 2005), indicando que:

> *"En el año 2015 el Tecnológico de Monterrey será la institución educativa más reconocida de América Latina por el liderazgo de sus egresados en los sectores privado, público social; y por la investigación y desarrollo tecnológico que*

realiza para impulsar la economía basada en el conocimiento" (p.7).

También se señala que:

"*Es misión del Tecnológico de Monterrey formar personas íntegras, éticas, con una visión humanística y competitivas internacionalmente en su campo profesional, que al mismo tiempo sean ciudadanos comprometidos con el desarrollo económico, político, social y cultural de su comunidad y con el uso sostenible de los recursos naturales"* (p. 9).

El documento propone también las estrategias que habrán de seguirse para lograr estos fines, entre las que se encuentran: a) asegurar la calidad académica y enriquecer el modelo educativo y b) impulsar y enfocar la investigación y el posgrado hacia el cumplimiento de la misión.

El aseguramiento de la calidad académica se inició 10 años atrás con una acción que contempló el proyecto educativo más importante que jamás se halla emprendido en el Tecnológico de Monterrey: el rediseño del proceso de enseñanza y aprendizaje para llegar a un modelo educativo universitario con características propias. Dicho modelo educativo, centrado en el aprendizaje, está en consonancia con las demandas actuales de ingenieros y en general de profesionistas que requieren mucho más que conocimientos sobre una determinada disciplina.

Este modelo educativo centrado en el alumno requiere definitivamente del análisis didáctico de los objetos de aprendizaje que están contenidos en los programas analíticos de los diferentes cursos que conforman el curriculum de las distintas carreras de ingeniería. Nuestra investigación, se orienta hacia dicho análisis didáctico, sobre el tema particular de los intervalos de confianza, para colaborar al intento de contar con el modelo educativo ideal de aprendizaje de dicho objeto.

1.2.4. IMPORTANCIA DE LA ESTADÍSTICA EN LA FORMACIÓN DE LOS INGENIEROS.

La estadística es un área de conocimiento de fundamental importancia en toda situación del campo de la ingeniería que requiera el análisis de datos para la toma de decisiones informadas en presencia de incertidumbre y variación. Contribuye a facilitar el análisis estadístico de información que se requiera, entre otros, para el diseño de plantas de control de contaminación, diseño estructural de edificios comerciales e industriales y diseño de sistemas de transporte (Vardeman, 2002).

La estadística facilita la obtención de información adecuada y actualizada que propicia el análisis del ambiente empresarial y la generación, el mantenimiento y la administración de información que benefician el desarrollo de competencias (Vera, Olivo, Alvarado y Batanero, 2007). En un enfoque hacia la producción o la administración, permite apoyar los procesos productivos con síntesis y análisis de información. A través de la adquisición de los conocimientos estadísticos, el ingeniero está en condiciones de ser un científico aplicado a la solución de problemas prácticos y dar respuesta a los desafíos técnicos del sector productivo, investigar y desarrollar procesos industriales. Esto se consigue especialmente cuando los cursos incluyen la capacidad de comunicación estadística en sus objetivos (Mac Gillivray, 2002a).

La estadística favorece la adquisición de conocimientos para la optimización de inventarios de materiales, partes y productos, así como el diseño e implantación de sistemas de control estadístico de procesos que permite a las compañías mejorar la calidad de sus productos y con ello aspirar a competir en un mercado cada vez más globalizado, especialmente cuando el estudiante trabaja a partir de proyectos de ingeniería (Mac Gillivray, 2002b). Los métodos estadísticos para mejorar la calidad incluyen: cartas de control, índices de capacidad del proceso, diagramas de Pareto, diagramas causa-efecto, cartas de control multivariadas, regresión lineal, diseño de experimentos, métodos Taguchi, etc. Esos métodos ayudan a identificar variaciones inusuales y las causas de esas variaciones, sean procesos de manufactura o en general procesos de negocios. El uso de estos métodos estadísticos impactan finalmente en incrementos en la productividad (Ryan, 1989).

El campo de la ingeniería está envuelto en un entorno cada vez más informatizado, las revistas y publicaciones en este campo están cada día más repletas de información estadística y el hecho de que la planeación y la interpretación de una investigación en ingeniería se basan en metodologías estadísticas, hace de la estadística, la materia imprescindible en el currículo del ingeniero.

1.3. UNA PRIMERA APROXIMACIÓN AL INTERVALO DE CONFIANZA

1.3.1. INTRODUCCIÓN

El *intervalo de confianza* es uno de los procedimientos generales de inferencia estadística que puede aplicarse a diversos problemas, como estudio de una o varias muestras, análisis de correlación o regresión, etc. En su definición intervienen una multitud de objetos matemáticos. En el contexto de estimar un *parámetro poblacional*, un intervalo de confianza es un *rango de valores* (calculado a partir de los *datos* de una *muestra*) en el cual podría encontrarse el *verdadero valor* del parámetro, junto con un *coeficiente de confianza* que indica el *porcentaje de muestras* tomadas en las mismas condiciones, en las cuales el intervalo cubriría el verdadero valor del parámetro. Como procedimiento general, describe una regla general de construcción de dicho rango de valores a partir de un *estadístico* calculado en los datos de la muestra, para el parámetro correspondiente. La idea general de intervalo de confianza se particulariza dependiendo del parámetro a estimar (media, proporción, varianza, etc.) y según las condiciones (tipo de distribución, qué se conoce de la misma, etc.).

La presentación anterior nos sugiere considerar el intervalo de confianza como una configuración matemática compleja, que engloba diversos problemas, conceptos (definiciones), propiedades, lenguaje, procedimientos y argumentos que el estudiante ha de comprender de antemano para abordar su estudio, lo que le dota de complejidad semiótica. En el siguiente apartado analizamos algunas definiciones actuales del “intervalo de confianza”, mostrando algunas ejemplificaciones de su uso y formulaciones simbólicas y elaborando un mapa conceptual de los conceptos que lo componen, para mostrar esta complejidad.

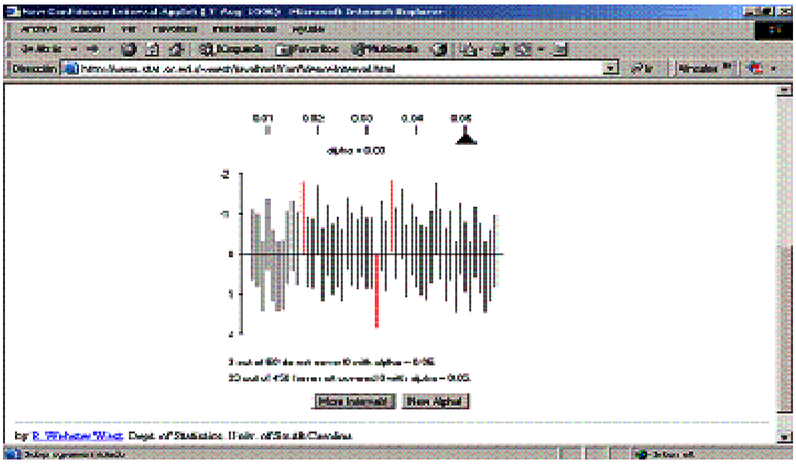
1.3.2. LA ESTRUCTURA CONCEPTUAL ASOCIADA AL INTERVALO DE CONFIANZA

El intervalo de confianza se introduce incluso en la educación secundaria. En M.E.C. (2004), en el tema de *estimación*, se indica que, en una *población* con una *variable aleatoria* de interés cuya *distribución* es de una familia conocida y queda determinada por *parámetros* desconocidos, podemos *estimar* uno de estos parámetros a partir de una *muestra aleatoria* de suficiente tamaño. Un e*stimador* es un resumen estadístico cuyo valor puede calcularse a partir de los datos de una *variable estadística* en la muestra y que proporciona información sobre el valor del parámetro. Por ejemplo, la media muestral es un estimador de la media poblacional, la proporción observada en la muestra es un estimador de la proporción en la población.

Una estimación es *puntual* cuando se obtiene un sólo valor para el parámetro. Los estimadores más probables en este caso son los estadísticos obtenidos en la muestra, aunque es necesario cuantificar el riesgo que se asume al considerarlos. La distribución muestral indica la distribución de los valores que tomará el estimador al seleccionar todas las muestras distintas de la población, es decir, al considerar el estimador como variable aleatoria. Las dos medidas fundamentales de esta distribución son su *media* que indica el valor promedio del estimador y la *desviación típica del estimador*, también denominada *error típico de estimación,* que indica la desviación promedio que podemos esperar entre el estimador y el valor del parámetro.

Más útil que la estimación puntual es la *estimación por intervalos* en la que calculamos dos valores entre los que se encontrará el parámetro, con un *nivel de confianza* fijado de antemano. Llamamos *intervalo de confianza* al intervalo que con un cierto nivel de confianza, contiene al parámetro que se está estimando.

Figura 1.1. Simulación de intervalos de confianza



El *nivel de confianza* es la *probabilidad* de que un intervalo calculado a partir de una muestra aleatoria contenga al verdadero valor del parámetro. Se suele representar por 1-α y habitualmente se da en porcentaje. La interpretación correcta de este nivel de confianza es que, si repitiésemos el proceso de construcción de intervalos con el mismo nivel y en muchas muestras de la misma población, el (1-α)% de los intervalos así construidos contendría al verdadero valor del parámetro. Por tanto, no se refiere a la muestra particular y no podemos saber cuál es la probabilidad de que el valor del parámetro se encuentre en el intervalo dado. En la Figura 1.1 se representan los datos de una simulación de muestras de la misma población con nivel de confianza dado (Simulador tomado de webstat http://www.stat.sc.edu/~west/javahtml/ConfidenceInterval.html). Observamos que algunos intervalos no cubren el valor verdadero del parámetro.

La definición intuitiva anterior se suele presentar en los libros de texto universitarios con mayor formalidad. Por ejemplo, para Walpole, Myers y Myers (1999), una estimación por intervalo de un parámetro poblacional θ es un intervalo de la forma $\hat{\theta}_L < \theta < \hat{\theta}_U$, donde $\hat{\theta}_L$ *y* $\hat{\theta}_U$ dependen del valor del estadístico $\hat{\Theta}$ para una muestra particular y también de la distribución de muestreo de $\hat{\Theta}$.

Como muestras distintas por lo general darán valores diferentes de $\hat{\Theta}$ y, por tanto, valores diferentes de $\hat{\theta}_L$ y $\hat{\theta}_U$, estos puntos extremos del intervalo son valores de las variables aleatorias correspondientes $\hat{\Theta}_L$ y $\hat{\Theta}_U$. De la distribución muestral de $\hat{\Theta}$ seremos capaces de determinar $\hat{\theta}_L$ y $\hat{\theta}_U$ de modo que $p(\hat{\Theta}_L < \theta < \hat{\Theta}_U)$ sea igual a algún valor fraccional positivo que queremos especificar. Si por ejemplo, encontramos $\hat{\theta}_L$ y $\hat{\theta}_U$, tales que $p(\hat{\Theta}_L < \theta < \hat{\Theta}_U) = 1-\alpha$ para 0< α <1, tenemos entonces una probabilidad de 1-α de seleccionar una variable aleatoria que produzca un intervalo que contenga a θ. El intervalo $\hat{\theta}_L < \theta < \hat{\theta}_U$ que se calcula a partir de la muestra seleccionada, se llama entonces intervalo de confianza de *(*1-*α)*100%, la fracción 1-*α* se denomina coeficiente de confianza o grado de confianza, y los puntos extremos $\hat{\theta}_L$ y $\hat{\theta}_U$, se llaman límites de confianza inferior y superior respectivamente.

La definición anterior se suele concretizar para parámetros particulares. Por ejemplo, Miller, Freund y Johnson (1992) introducen la construcción del intervalo de confianza para la media μ de una población, en caso de muestras grandes y σ conocida de la manera siguiente:

Si $\bar{x}$ es la media de una muestra aleatoria de tamaño *n* de una población con σ conocida, el intervalo de confianza de *(*1-*α)* 100%, para μ es,

$$\bar{x} - Z_{\alpha/2}\,\sigma\big/\sqrt{n} < \mu < \bar{x} + Z_{\alpha/2}\,\sigma\big/\sqrt{n}$$

donde $Z_{\alpha/2}$ es el valor de *Z* a la derecha del cuál se tiene un área de *α/2*.

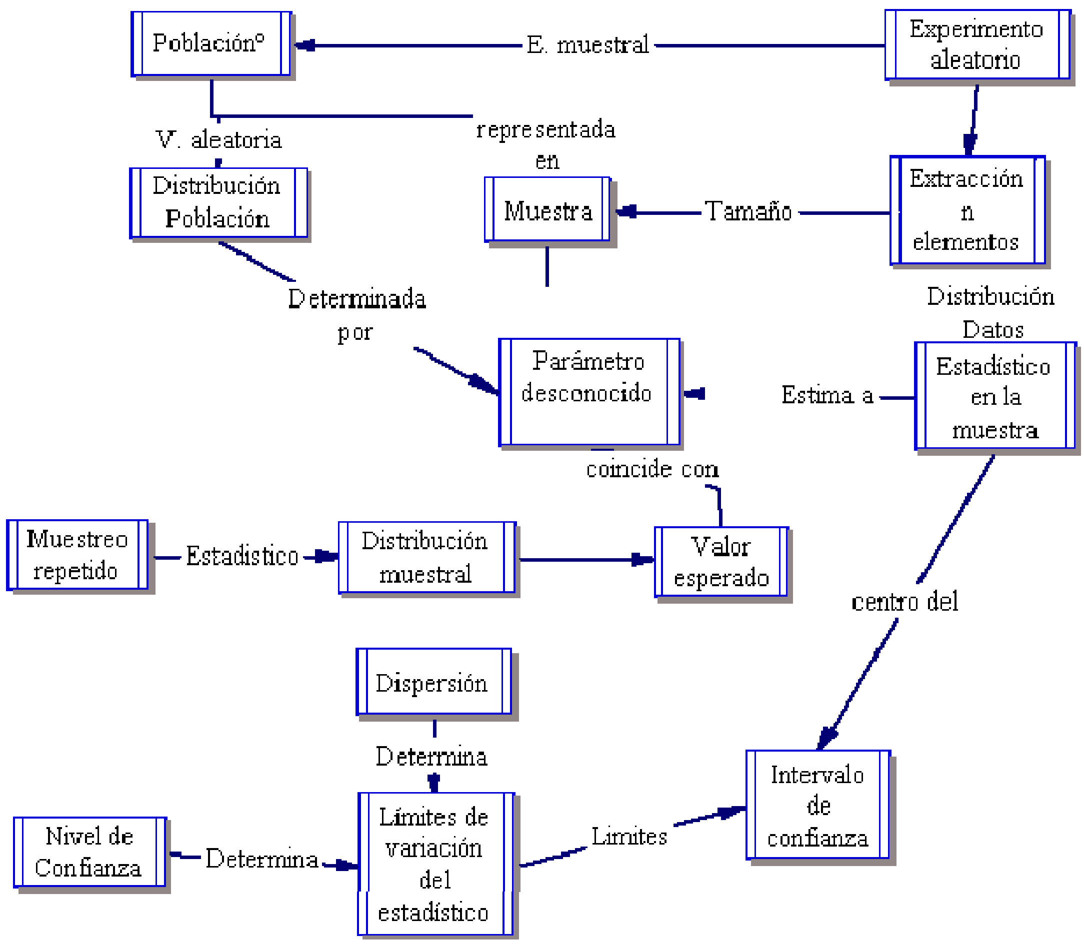
1.3.3. COMPONENTES DE UN INTERVALO DE CONFIANZA

De acuerdo a las definiciones y procedimientos anteriores, los siguientes objetos matemáticos intervienen en la definición del intervalo de confianza:

- *Población*: Conjunto de personas, mediciones u objetos con características bien definidas, que serán el grupo de interés de nuestra investigación.
- *Variable aleatoria.* Variable que estudiamos en la población y que tiene una *distribución de probabilidad*, que depende de uno o varios parámetros.
- *Parámetro:* Propiedad numérica de una población, constante, pero desconocido, que determina la distribución de la variable en la población.
- *Muestra*: Subconjunto de la población.
- *Tamaño de la muestra*: número de sujetos de que esta compuesta la muestra que estamos analizando. Se representa por (n).
- *Variable estadística*: conjunto de valores de la variable aleatoria en la muestra; le corresponde una *distribución de frecuencias*.
- *Estadístico:* Propiedad numérica de la distribución de frecuencias en la muestra. Valor conocido, pero varía de una muestra a otra, es decir, es una variable aleatoria. El estadístico es un *estimador* del parámetro.
- *Distribución muestral:* distribución de los valores que tomará el estadístico al seleccionar todas las posibles distintas muestras de la población. Es decir, distribución de probabilidad del estadístico, considerado como variable aleatoria. Depende también del parámetro desconocido.
- *Valor esperado:* es el promedio de todos los valores de una variable aleatoria. Se puede referir a la población (media de la población) o a la distribución muestral (media del estadístico).
- *Dispersión:* Medida que indica en promedio la variabilidad de los datos, respecto a la media.
- *Nivel de confianza*: valor de *probabilidad* que fijamos para construir el intervalo. Indica la frecuencia de muestras que darán un intervalo conteniendo el valor del parámetro.
- *Intervalo de confianza*: intervalo aleatorio que con un cierto nivel de confianza, contiene al parámetro que se está estimando.
- *Límites de variación*: determinan el error en la estimación.

Estos objetos y sus relaciones se resumen en el mapa conceptual representado en la figura 1.2. Observamos la complejidad del objeto intervalo de confianza, donde los mismos objetos asociados aparecen a diferentes niveles. Por ejemplo, la idea de distribución aparece relacionada con la población (distribución de la variable aleatoria), con la muestra (distribución estadística de datos) y con el muestreo (distribución muestral del estadístico). Esto supone una gran complejidad semiótica, pues cuando nos referimos a uno de estos niveles el estudiante podría interpretar otro.

Figura 1.2. Objetos matemáticos ligados al intervalo de confianza



1.3.4. PROCEDIMIENTO DE CONSTRUCCIÓN DE INTERVALOS DE CONFIANZA

Al concretizar el intervalo de confianza para un parámetro particular se determina un procedimiento de obtención del mismo. Como ejemplo, vamos a ilustrar el procedimiento de obtención de un intervalo de confianza, considerando una población normal con varianza desconocida, siendo el parámetro a estimar su media μ siguiendo a López Sánchez y cols. (2004). Consideramos una muestra aleatoria $X_1, X_2, ..., X_n$ de tamaño *n* extraída de la población *X*. El estadístico que emplearemos, relacionado con el parámetro μ, será:

$$\frac{\bar{x}-\mu}{s/\sqrt{n}},$$

siendo $\bar{x}$ la media de la muestra y *s* su desviación típica. Este estadístico sigue una distribución muestral *T* de Student con *(n-1)* grados de libertad. El nivel de confianza 1-α, establecido a priori por el experimentador (los usuales son 0.95, 0.90 y 0.99). Dada la distribución del estadístico y el nivel de confianza, se tiene la siguiente igualdad probabilística:

$$P(-t_{\alpha/2} < \frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{\alpha/2}) = 1-\alpha$$

siendo $t_{\alpha/2}$ el valor crítico en la *T* de Student. La expresión anterior es equivalente a:

$$p(\bar{x} - t_{\alpha/2}\, s\big/\sqrt{n} < \mu < \bar{x} + t_{\alpha/2}\, s\big/\sqrt{n}) = 1-\alpha$$

que hace referencia a que con una probabilidad 1-α el intervalo aleatorio

$$\bar{x} - t_{\alpha/2}\, s\big/\sqrt{n},\ \bar{x} + t_{\alpha/2}\, s\big/\sqrt{n}$$

contendrá el valor medio μ.

El intervalo es aleatorio ya que sus extremos se determinan a partir de los estimadores media muestral y desviación típica muestral, que son variables aleatorias. La probabilidad a que se refiere dicho intervalo aleatorio, puede interpretarse de manera informal pero quizás más clara: Si consideramos todas las muestras distintas de tamaño *n* que puedan ser extraídas de la población *X*, y con las observaciones de cada una construimos los correspondientes intervalos, según la estructura anterior, el 1-α de estos intervalos contendrán el parámetro μ.

Por tanto, si extraemos una muestra de tamaño *n* y con los datos u observaciones, $x_1, x_2, ..., x_n$, calculamos los extremos del intervalo, dispondremos del concreto intervalo de confianza para el parámetro μ

$$(\bar{x} - t_{\alpha/2}\, s\big/\sqrt{n}, \bar{x} + t_{\alpha/2}\, s\big/\sqrt{n})$$

que, en función de la interpretación informal anterior, contendrá dicho parámetro con una confianza 1-α.

Una observación es que el nivel de confianza establece en alguna medida la longitud del correspondiente intervalo de confianza. Aumentando el nivel de confianza (mayor certeza), aumenta la longitud (menor precisión).

Intervalo de confianza y tamaño de la muestra

La amplitud del intervalo de confianza depende del valor de:

$$E = Z_{\alpha/2}\, \sigma \big/ \sqrt{n}$$

Con un nivel de confianza del 1-α % admitimos que la diferencia entre la estimación para la media a partir de la muestra y su valor real es menor que E, que llamaremos error máximo admisible.

El tamaño de la muestra depende del nivel de confianza que se desee para los resultados y de la amplitud del intervalo de confianza, es decir del error máximo que se esté dispuesto a admitir. Fijados estos, 1-α y E, podemos calcular el tamaño mínimo de la muestra que emplearemos

$$n = (Z_{\alpha/2}\, \sigma \big/ E)^2$$

Uso de intervalos de confianza para verificar hipótesis

Como se indica en Escuela de Medicina (1996), los intervalos de confianza permiten verificar hipótesis planteadas respecto a parámetros poblacionales. Por ejemplo, supongamos que se plantea la hipótesis de que el promedio de peso de nacimiento de cierta población es igual a la media nacional de 3250 gramos. Al tomar una muestra de 30 recién nacidos de la población en estudio, se obtuvo:

$$\bar{x} = 2930,\ s = 450,\ n = 30$$

Al construir un intervalo de 95% de confianza para la media poblacional, se obtiene:

$$2930 - (1.96)(450)\big/\sqrt{30} \le \mu \le 2930 + (1.96)(450)\big/\sqrt{30}$$

Luego, el peso de nacimiento varía entre 2769 y 3091 gramos, con una confianza de 95%. Como el intervalo no incluye el valor μ = 3250 gramos planteado en la hipótesis, entonces esta es rechazada con confianza 95% (o un valor p menor a 0.05).

Como resumen del análisis, podemos ver que en el intervalo de confianza se recogen muchos contenidos anteriormente estudiados por el alumno y, a su vez, es base para el desarrollo posterior de gran parte de temas. Su utilidad práctica queda de manifiesto, tanto para estimación, como para el contraste de hipótesis.

En el capítulo 3 completaremos este estudio utilizando las herramientas del "enfoque ontosemiótico" de los conocimientos matemáticos desarrollado por Godino y colaboradores, del que presentaremos una síntesis en la sección 2.3.

1.4 RELEVANCIA DEL OBJETO DE ESTUDIO

1.4.1. RELEVANCIA DE LOS INTERVALOS DE CONFIANZA EN LA FORMACIÓN DEL INGENIERO

Casi cualquier investigación empírica, entre ellas las que ha de llevar a cabo el ingeniero en su vida profesional requiere análisis de datos y de la realización de inferencias a partir de los resultados de dicho análisis. Uno de los métodos implicados en la inferencia estadística es la estimación, la cual trata con la predicción de los valores de parámetros poblacionales específicos, tales como la media, varianza, proporción, coeficientes de regresión y correlación poblacionales, así como de la comparación de los valores de un parámetro en dos o más poblaciones. Estos procesos implican calcular a partir de los datos de una muestra algún estadístico que será utilizado como estimador del parámetro poblacional específico (estimación puntual).

Una forma de completar la estimación puntual del valor de un parámetro poblacional es utilizar un intervalo de confianza que proporciona unos límites aleatorios de variación del parámetro, así como una confianza en la estimación. Un intervalo de confianza para un parámetro, digamos la media poblacional μ se calcula a partir de un estadístico muestral (media) más o menos una medida del error del muestreo (el cual es el error de una muestra aleatoria), multiplicado por un valor crítico que se obtiene de la distribución de muestreo del estadístico. Todo este proceso le permite al estudiante de ingeniería aprender a razonar de manera lógica. Su pensamiento analítico se ve fortalecido, por ejemplo, al hacer comparación de parámetros poblacionales a través de la interpretación de intervalos de confianza de los parámetros correspondientes. Los escenarios donde aplica estos procesos y los procesos mismos le permiten reflexionar acerca del comportamiento de muchos fenómenos en donde está presente la incertidumbre y la variación.

1.4.2. RELEVANCIA DE LOS INTERVALOS DE CONFIANZA EN LA ESTADÍSTICA.

La relevancia que tienen los intervalos de confianza en estadística, es destacada por el hecho de que están siendo cada vez más apreciados para realizar inferencias estadísticas en comparación a los contrastes de hipótesis. Muchos autores indican que los contrastes de hipótesis deben complementarse con la estimación del efecto, proporcionando intervalos de confianza para los mismos, especialmente si se llega a un resultado no significativo (Thompson, 1996, 2002; Pascual, García y Frías, 2000; Vacha-Haase, 2001).

Autores como Clark (2004) consideran que una de las ventajas de los intervalos de confianza es que revelan al investigador el margen de error de sus resultados y la magnitud del efecto que observa. Cumming y Finch (2001) señalan que los intervalos de confianza por su enfoque en el efecto del tamaño de la muestra, tienen el potencial de facilitar el pensamiento meta-analítico. Su escaso uso actual se debe al hecho de que usualmente son demasiado amplios, pero se pueden mejorar aumentando la fiabilidad de los procedimientos y el tamaño de las muestras (Cohen, 1994). Podemos apuntar también una declaración dada por la Asociación Americana de Psicología acerca de la importancia de los intervalos de confianza,

> "*Debido a que los intervalos de confianza combinan información sobre localización y precisión y a menudo pueden ser directamente usados para inferir niveles de significancia, ellos son, en general, la mejor estrategia de reporte. El uso de intervalos de confianza es entonces fuertemente recomendado*" (APA, 2001, p. 22).

Cumming y Finch (2001) señalan que los intervalos de confianza, por su enfoque en el efecto del tamaño de la muestra, tienen el potencial de facilitar el pensamiento meta-analítico, atributo que es esencial en la diseminación de los datos estadísticos y en la elección de las representaciones que mejor comuniquen los resultados de una investigación. Según Wolfe y Cumming (2004) los intervalos de confianza son un medio efectivo de cuantificar la incertidumbre inherente en los resultados de los estudios**.**

Por ejemplo, para el caso de las ciencias médicas y sociales, Clark (2004) sugiere que los intervalos de confianza aventajan a las pruebas de significación como instrumento analítico para muchos tipos de investigación, entre ellos los estudios observacionales y experimentales, avisando que la interpretación de un resultado solamente a la luz de un *valor p* es una práctica mecánica e irreflexiva cuya persistencia es difícil de comprender. Esta autora considera que los intervalos de confianza, revelan al investigador el margen de error de sus resultados y la magnitud del efecto que observa, fomentan la actividad analítica imprescindible para la evolución del conocimiento científico. En el mismo sentido se expresa Davies (1998) quien sugiere que los intervalos de confianza proporcionan la misma información que un *valor p*, pero más que esto, esos intervalos muestran los efectos más grandes y más pequeños que son dados probablemente por los datos observados. Por supuesto, tal como hemos indicado en el análisis conceptual, estas afirmaciones deben ser precisadas, puesto que en inferencia clásica no podemos dar estas acotaciones. En todo caso, estas afirmaciones indican lo extendido de los errores en torno a los intervalos de confianza.

1.4.3. RETOS DIDÁCTICOS QUE ENFRENTA LA ENSEÑANZA DE LOS INTERVALOS DE CONFIANZA.

La lógica de la estadística inferencial es difícil de comprender y en consecuencia su aplicación e interpretación no son siempre atinados. Las críticas que han surgido alrededor de la inferencia estadística han generado una controversia que ha ido en aumento en los últimos años en algunas instituciones profesionales (Vallecillos, 1994, 1995, 1996; Thompson, 1996, 2002; Batanero, 2000b; Díaz y de la Fuente, 2004, Batanero, Díaz y Wilhelmi, en prensa) que sugieren importantes cambios en sus políticas editoriales. La American Psychological Association, por ejemplo, constituyó una comisión (Wilkinson, 1999) donde recomienda publicar los valores-*p* exactos, las estimaciones de los efectos y los intervalos de confianza cuando se usen los contrastes de hipótesis. En la American Education Research, Thompson (1996) se sugiere un uso más adecuado del lenguaje estadístico en la elaboración de los informes de investigación, resaltando la interpretación del tamaño de los efectos y evaluando la replicabilidad de los resultados.

Otros problemas que han sido denunciados tienen que ver con los factores psicológicos que están presentes en la interpretación incorrecta de los resultados de la inferencia estadística. Por ejemplo, Falk y Greenbaum (1995) sugieren la existencia de mecanismos psicológicos fuertemente arraigados que llevan a las personas a creer que eliminan el azar y minimizan su incertidumbre cuando obtienen un resultado significativo en los contrastes de hipótesis. Existe un abuso del razonamiento lógico a la inferencia estadística, y prueba de ello es la confianza excesiva que los investigadores experimentan en los contrastes estadísticos debido a la sofisticación de los términos y fórmulas matemáticas, que contribuyen a fomentar la sensación de que la significación estadística es una garantía de objetividad (Batanero y Díaz, 2005).

Hay que resaltar que los intervalos de confianza son intervalos en los que el verdadero valor del parámetro se encuentra en un porcentaje dado de muestras, aunque no aseguran en que intervalo estará el parámetro en nuestro experimento particular. Por tanto su función es la de complementar los contrastes de hipótesis y están sujetos a las mismas controversias e interpretaciones incorrectas que aquellos. Puesto que es extraño tratar los datos como aleatorios una vez recogidos hay peligro de olvidar la interpretación frecuencial a los intervalos de confianza y dar una interpretación bayesiana, que es más natural. En este sentido, los intervalos de confianza no son una solución al contraste, ya que, aunque dan una idea del signo del efecto, están sujetos a los mismos errores de interpretación que los contrastes y los investigadores pueden interpretar los coeficientes de confianza (que son probabilidades iniciales) como probabilidades finales (Díaz, 2007).

La investigación didáctica muestra la existencia de concepciones erróneas ampliamente extendidas en el trabajo cotidiano entre los usuarios de la inferencia estadística. En el caso de los intervalos de confianza algunos autores (Behar, 2001; Cumming, Williams y Fidler, 2004; Fidler y Cumming, 2005) están avisando que los investigadores y estudiantes, entre ellos los de ingeniería, tienen errores conceptuales y no interpretan los conceptos apropiadamente. Muchos de estos errores pueden ser debidos a fallos en la comprensión de otros conceptos previos, como el de distribución muestral (Vallecillos, 1995; Batanero, Tauber y Sánchez, 2004).

Asimismo, Kahneman, Slovic y Tversky (1982) indican que las personas que siguen la heurística de representatividad, consistente en prescindir del tamaño de la muestra, y de la variabilidad del muestreo, tienen confianza indebida en las pequeñas muestras. Este error puede tener importantes consecuencias de cara a la investigación experimental, ya que estas personas subestiman la amplitud de sus intervalos de confianza y tienen unas expectativas injustificadas en la replicabilidad de experimentos realizados con pequeñas muestras.

Como indica Batanero (2004), la población de estudiantes que acude a los cursos de estadística es cada vez mayor y más variada e, incluso como en el caso de los ingenieros, donde el alumno tiene una sólida base en cálculo y álgebra, la limitación temporal de los cursos hace que tengamos que enseñar inferencia avanzada sin usar las matemáticas avanzadas. Ello plantea el reto de repensar qué es en realidad el pensamiento matemático avanzado y cómo transmitirlo a diversos tipos de estudiantes.

Por otro lado, la estimación por intervalos, igual que los contrastes estadísticos solo son parte de un proceso más general de inferencia científica. Sin embargo, frecuentemente encontramos que la estadística se enseña aisladamente, sin conectarla con un marco más general de metodología de investigación y diseño experimental (Batanero, 2005). Entonces el reto que surge, concretamente con los intervalos de confianza, es que su enseñanza deberá ser organizada de tal manera que el estudiante reconozca algunos elementos de significado, tales como los diferentes campos de problemas, sus propiedades o el lenguaje y relacione cada uno de ellos con otros objetos de aprendizaje, como por ejemplo las distribuciones muestrales, y que además adquiera la lógica general que subyace en la construcción e interpretación de los intervalos.

1.5 OBJETIVOS DE LA INVESTIGACIÓN

Esta investigación, en que nos planteamos elaborar un instrumento que nos permita realizar una evaluación de la comprensión de los intervalos de confianza por estudiantes de las carreras de ingeniería, nos permitirá identificar durante este proceso algunas dificultades y errores de aprendizaje. El diseño de este instrumento estará basado en el estudio que hemos realizado sobre el análisis de la enseñanza actual de los intervalos de confianza a partir de la determinación de sus diversos significados históricos y sus elementos; estudio que presentamos en los capítulos segundo y tercero. Asimismo se aplicará un enfoque metodológico que nos permita asegurar la validez y fiabilidad del instrumento. En concreto, nuestro objetivo central lo podemos expresar en la forma siguiente.

Objetivo general: *Generar información que contribuya a explicar el significado personal que los estudiantes de las carreras de ingeniería asignan a los intervalos de confianza.* La importancia de este objetivo es tanto a nivel de diagnóstico, como para planificar las posibles acciones correctivas.

Este objetivo central se descompone en los siguientes objetivos específicos, los cuales describimos a continuación, siguiendo las mismas líneas de trabajo de Alvarado (2004) sobre el Teorema Central del Límite y de Díaz (2004, 2007) sobre construcción de cuestionarios (razonamiento condicional).

Objetivo 1. Realizar un análisis del intervalo de confianza desde el punto de vista de su evolución histórica, identificando información que nos proporcione pistas para la reconstrucción de sus diferentes significados a lo largo de la historia. Este análisis se presenta en el Capítulo 2.

Objetivo 2: Realizar una síntesis de las investigaciones previas relacionadas con la comprensión de los intervalos de confianza y seleccionar ítems que puedan usarse en la construcción de un cuestionario.

Esta síntesis que presentaremos en el capítulo segundo, nos permitirá afianzar nuestra familiaridad con el tema. Además podremos identificar los errores importantes descritos en investigaciones previas, así como algunas carencias, lo cual nos pondrá en posición de confirmar la necesidad de elaborar un instrumento de evaluación de la comprensión.

Objetivo 3: Realizar un análisis de los elementos del significado de los intervalos de confianza que nos fundamente la definición semántica de la variable "comprensión de los intervalos de confianza" y delimitar las principales áreas de contenido.

Para alcanzar este objetivo, y utilizando el enfoque ontosemiótico de la cognición matemática (EOS) propuesto por Godino y colaboradores, realizaremos en el capítulo tercero un análisis y clasificación de los *elementos de significado* del objeto matemático intervalo de confianza en una muestra de libros de texto recomendados en la asignatura de estadística para ingenieros en el Sistema Tecnológico de Monterrey. Este estudio, junto con el análisis histórico, permitirá describir el significado de referencia del intervalo de confianza en nuestra investigación. A partir de aquí se concretarán las áreas de contenido que serán objeto de evaluación.

Objetivo 4: Construir un instrumento piloto que permita evaluar la comprensión de los intervalos de confianza, siguiendo unos criterios metodológicos rigurosos en la selección de los ítems que lo compondrían.

La lectura de diversos libros de psicometría nos orientará en las fases iniciales de selección y depuración de ítems. Prepararemos la tabla de especificaciones, seleccionaremos inicialmente un número adecuado de ítems (algunos en varias versiones), tomados de investigaciones previas o libros de texto, de modo que podamos disponer de dos o tres ítems diferentes para cada uno de los contenidos previstos.

Un panel de expertos colaborará en la selección de los ítems y contenidos definitivos, cuya dificultad y legibilidad será también probada en distintas muestras de estudiantes de ingeniería. Describiremos en el capítulo cuarto la metodología que seguiremos en este proceso y en la selección de los ítems para configurar un cuestionario piloto.

Objetivo 5: Realizar un estudio de evaluación con una amplia muestra de estudiantes de ingeniería, para analizar tanto las tendencias como la variabilidad del significado personal de los estudiantes

El interés de este objetivo es describir las semejanzas y diferencias entre el significado de referencia y el significado personal adquirido por los estudiantes de la muestra. Se pondrá especial énfasis en identificar los conflictos semióticos y dificultades de los estudiantes. Los resultados serían aplicables a otros estudiantes de ingeniería en contextos similares al tomado en esta investigación.

1.6. HIPÓTESIS INICIALES

Puesto que apenas hay investigaciones didácticas sobre el tema de los intervalos de confianza, nuestro estudio es de tipo descriptivo y exploratorio. Las hipótesis que formulamos a continuación, deben entenderse como conjeturas o expectativas de lo que esperamos encontrar. Nuestra investigación no se orienta al estudio de hipótesis, en el sentido estadístico del término.

Hipótesis 1: *Existen diversidad de significados en los libros de texto de estadística para ingenieros en cuanto a los campos de problemas, a las representaciones, propiedades, formulaciones, argumentos y algoritmos de cálculo de los intervalos de confianza. Algunos de estos significados podrían ser incompletos o sesgados.*

Algunos autores han denunciado la existencia de errores conceptuales en la enseñanza del contraste de hipótesis en los libros de texto (Falk y Greembaum, 1995) y en otros casos se mezclan conceptos de la metodología de Fisher y Neyman – Pearson, e incluso se da una interpretación bayesiana al nivel de significación (Gigerenzer, 1993). Pensamos que estos errores también podrían aparecer al interpretar el concepto de coeficiente de confianza y pueden contribuir a los errores de comprensión denunciados por Cumming y Fidler (2005). También hemos constatado personalmente que algunos libros dan fórmulas incorrectas para el cálculo del error típico de la media, lo que puede afectar en el cálculo del intervalo de confianza.

Hipótesis 2: *Los textos analizados dan más importancia a los enunciados y propiedades del concepto, que a sus aplicaciones. Además, rara vez presentan el concepto a través de situaciones problemáticas.*

Esta hipótesis se fundamenta en los resultados encontrados a este respecto en el análisis de otros conceptos estadísticos en libros de texto universitarios (Alvarado, 2004), donde se ha encontrado un énfasis excesivo en los elementos conceptuales. Por otro lado nuestra experiencia como profesor nos ha llevado a observar que, en general, los libros introducen primeramente la definición del concepto y luego las aplicaciones, mientras que, didácticamente, sería preferible al contrario.

Hipótesis 3. *Los campos de problemas propuestos en los textos no coinciden con los encontrados en el análisis histórico.*

Esta hipótesis es más bien una conjetura; nos basamos por un lado en los resultados de Alvarado (2004) para el teorema central del límite y en el estudio de Cobo (2003) para el concepto de media, en que se presenta este problema.

Hipótesis 4. *Los autores de los textos, recién empiezan a darle importancia al uso del ordenador como herramienta didáctica en la comprensión de este concepto.*

De nuevo nos basamos en los trabajos previos citados que indican que la propuesta de actividades basadas en el ordenador era muy escasa hace unos años, pero que progresivamente se van incluyendo. De este modo se enriquece el uso de lenguajes gráficos junto a los simbólicos y algebraicos.

Hipótesis 5. *Existe una amplia variedad de posibles conflictos semióticos y dificultades procedimentales de los estudiantes en relación a los intervalos de confianza.*

Esta hipótesis se fundamenta en nuestra experiencia como profesor donde frecuentemente hemos observado confusión en las Distribuciones Muestrales o en la determinación de los valores críticos, así como en los resultados descritos en los antecedentes previos.

Con la comprobación de estas hipótesis continuaremos la investigación didáctica iniciada en Olivo (2006), analizando en profundidad el significado del objeto en la institución de referencia, identificando los elementos característicos y valorando cuáles de ellos y con qué intensidad son comprendidos por los estudiantes.

CAPÍTULO 2.
FUNDAMENTOS TEÓRICOS

2.1. INTRODUCCIÓN

Este capítulo está dedicado a presentar los fundamentos teóricos de nuestro trabajo de investigación. Comenzamos analizando la evolución histórica del intervalo de confianza, desde las primitivas ideas de estimación, pasando por el estudio de las probabilidades de causas por Bayes y el método fiducial de Fisher, hasta llegar a la formulación actual debida a Neymann. El estudio de este desarrollo nos aportará información para la reconstrucción posterior de los significados institucionales del intervalo de confianza según nuestro marco teórico, en el que diferenciaremos los elementos que los constituyen. Servirá también para explicar, en parte, las dificultades que tienen los alumnos en el aprendizaje de dicho objeto.

Seguidamente presentamos brevemente algunos elementos relevantes para nuestro trabajo del enfoque ontosemiótico para la cognición e instrucción matemática, propuesto por Godino y colaboradores y en el que apoyaremos nuestro trabajo. Específicamente, nos centraremos en los componentes epistémico y cognitivo de dicho marco teórico, es decir la teoría de los significados institucionales y personales de los objetos matemáticos (Godino y Batanero, 1994, 1998) y teoría de las funciones semióticas (Godino, 2002; Godino, Batanero y Roa, 2005). Utilizaremos también los criterios de idoneidad de un cuestionario de evaluación adaptados por Batanero y Díaz (2005) de los propuestos por Godino, Contreras y Font (2006) para los procesos de estudio.

Y finalmente presentamos el estado de la cuestión de las investigaciones previas, relacionadas con la comprensión del intervalo de confianza y los objetos en los que se apoyan, principalmente las distribuciones muestrales y distribución normal.

2.2. DESARROLLO HISTÓRICO

2.2.1. INTRODUCCIÓN

El origen histórico de los intervalos de confianza y la teoría de estimación estadística se remonta a muchos años atrás y ha tomado distintas formas y matices, como exponemos en Olivo y Batanero (2006) y Olivo, Batanero y Ortiz (2006).

Según Rao (1992, p.36), a principios del siglo XX, antes de 1912, la estimación estadística se basaba fundamentalmente en el método de mínimos cuadrados, debido a Gauss y el método de desviación mínima absoluta debida a Laplace, que eran usados generalmente para estimar parámetros en un modelo lineal. Desde el punto de vista de la filosofía de la ciencia, el problema de la estimación se relaciona con la inferencia inductiva, es decir aquella forma de razonamiento según la cual la verdad de las premisas no comporta necesariamente la de la conclusión. En términos estadísticos se relaciona con las argumentaciones desde la muestra hacia la población de la cual, se extrajo dicha muestra (Vallecillos, 1994).

Matemáticamente, el problema de la estimación, podemos plantearlo de distintas maneras. Una de ellas es la que se inicia con el deseo de estudiar un fenómeno aleatorio que viene caracterizado por una distribución de probabilidad, que depende de uno o varios parámetros. Pero no es posible recolectar los datos de toda la población, sino que hemos de conformarnos con una muestra aleatoria, de la misma población, en donde se calcula un estadístico. El problema es dar un valor aproximado del parámetro (o parámetros) a partir de los datos observados del estadístico (o estadísticos).

Dicho de otra manera: Sean $X_1, X_2,..., X_n$ un sistema de *n* variables aleatorias, cuyos valores particulares pueden ser obtenidos a través de muestras aleatorias. Como la ley de probabilidad de estas variables $P(x_1, x_2,..., x_n | \theta_1, \theta_2,..., \theta_k)$ para $x_1, x_2,..., x_n$ valores de las *n* variables respectivamente, dependen de *k* parámetros desconocidos, es necesario estimar estos parámetros haciendo uso en este proceso de los valores observados $x'_1, x'_2,..., x'_n$ del sistema de *n* variables aleatorias.

En lo que sigue describimos las soluciones aportadas históricamente a este problema, la primera de las cuales fue aportada por Bayes en su famoso teorema, aunque no se publica hasta después de su muerte por Laplace.

Laplace rescata en 1818 el teorema de Bayes y desarrolla él mismo algunas de sus consecuencias, a la vez que establece la primera formulación del problema de *estimación puntual.* Tanto Laplace como, posteriormente (hacia 1887) Gauss contemplan un valor desconocido del parámetro θ y un cierto número de sus mediciones x_i todas sujetas a un error aleatorio. También los dos contemplan la formulación de lo que se llama la "función pérdida" $L(\hat{\theta},\theta)$. Esta función representa el error que el estadístico asumirá por adoptar a $\hat{\theta}$ como estimador de θ. Laplace usó el valor absoluto de la diferencia $L_L(\hat{\theta},\theta)=\left|\hat{\theta}-\theta\right|$, mientras que Gauss prefirió su cuadrado $L_G(\hat{\theta},\theta)=(\hat{\theta}-\theta)^2$. De ahí resultó la teoría de mínimos cuadrados. (Neyman, 1976).

Posteriormente, Karl Pearson introdujo el método de momentos para estimar los parámetros de distribuciones (Rao, 1992). Otras claves históricas importantes, que en los siguientes apartados abordaremos, tienen que ver con la doctrina de la probabilidad fiducial de Fisher, quién estudió extensamente el problema de la estimación estadística e introdujo conceptos tales como: *consistencia, eficiencia y suficiencia* de un estimador y la teoría de los intervalos de confianza de Neyman.

2.2.2. EL TEOREMA DE BAYES

El Teorema de Bayes constituye el primer esfuerzo de solución del problema de la estimación, aunque considerando que los parámetros $\theta_1,\theta_2,...,\theta_k$ son ellos mismos variables aleatorias (Rivadulla 1991). El teorema de Bayes proporciona un método de cálculo de las posibilidades a posteriori de que un efecto ocurrido sea debido a una causa dada, tomando como punto de partida las probabilidades a priori, de las posibles causas y luego las probabilidades de que cada una de las causas produzca el efecto dado. Thomas Bayes, en un ensayo póstumo, publicado en 1763 propone una respuesta al citado problema, mediante su famoso Teorema, que en notación integral tiene el siguiente desarrollo.

Si en la serie de experimentos el suceso ha aparecido *p* veces y *q* veces el suceso contrario, la posibilidad *w* de que la probabilidad buscada para el suceso se encuentre entre *x* y *X* viene dada por la siguiente expresión:

$$W=\frac{\int_x^X x^P(1-x)^q\,dx}{\int_0^1 x^P(1-x)^q\,dx}$$

Puesto que Bayes considera el parámetro como variable aleatoria (al igual que hace hoy día la escuela Bayesiana), al dar la probabilidad de que el parámetro se encuentre entre ciertos valores, en realidad está dando un avance hacia la construcción de intervalos de credibilidad, por lo que estos serían anteriores a los intervalos de confianza.

Aportación Bayesiana de Laplace

El resultado anterior se vincula a la teoría de la probabilidad de las causas y de los acontecimientos futuros con base en sucesos observados y es recogido por Laplace. Hald (1998, pp. 23-4) encuentra un análisis Bayesiano en "Mémoire sur la Probabilité des Causes par les événements", publicada por Laplace en 1774, donde se ocupa de la inferencia, con la finalidad de determinar una probabilidad desconocida en una distribución binomial. En concreto trata de una urna que contiene un número dado de boletos blancos y boletos negros en una relación desconocida; si uno extrae un boleto y encuentra que es blanco, se desea determinar la probabilidad de que la razón de boletos blancos a negros es de p a q. El evento es conocido y la causa es desconocida.

Considerando a *F* como el evento que ha ocurrido y $\theta_1, \theta_2, ..., \theta_n$ las *n* causas, a priori igualmente probables, el Principio axiomático que establece Laplace para resolver esta situación es el siguiente:

$$\frac{p(\theta i / F)}{p(\theta j / F)} = \frac{p(F / \theta i)}{p(F / \theta j)}$$

y

$$P(\theta j / F) = \frac{p(F / \theta j)}{\sum_{i=1}^{n} p(F / \theta i)}$$

La idea que surge de este Principio es que si un evento puede ser producido por un número *n* de diferentes causas, el cociente de las probabilidades de esas causas dados los eventos es lo mismo que el cociente de las probabilidades de los eventos dadas las causas, y la probabilidad de la existencia de cada una de ellas es igual a la probabilidad del evento dada esa causa, dividida por la suma de todas las probabilidades de los eventos dadas cada una de esas causas. Este principio enunciado por Laplace en 1774 ahora se reconoce como equivalente al teorema de Bayes, es decir, la solución aportada por Laplace al problema de estimación es una solución bayesiana (Rivadulla, 1991).

Asimismo, en la sección V del “Mémoire sur la Probabilité des Causes par les événements”, Laplace se enfoca en la estimación del valor medio a partir de tres observaciones. Incentivado por la nota publicada en un artículo de J. Bernoulli III que indicaba que el problema de estimación de la media era de considerable interés para los astrónomos, Laplace realiza su tratamiento del problema estudiando la distribución del error y encuentra una expresión general para la *distribución posterior* de la media. La importancia que guarda el artículo citado de Laplace radica en que sienta las bases de la teoría de la decisión moderna y de la inferencia bayesiana (Stigler 1986).

De hecho, el teorema de Bayes puede ser orientado a calcular la probabilidad a posteriori $P(\theta_1, \theta_2, ..., \theta_n \mid x'_1, x'_2, ..., x'_n)$ de los parámetros indicados con base en los valores indicados $x'_1, x'_2, ..., x'_n$ en todas aquellas situaciones en que es conocida la probabilidad a priori $P(\theta_1, \theta_2, ..., \theta_k)$. El Teorema de Bayes, en consecuencia, permite también calcular la probabilidad de que un parámetro determinado θ_i caiga dentro de un intervalo determinado; siendo considerado el valor $\hat{\theta}_i$ el estimador de θ_i por ser el más probable de θ_i. La probabilidad:

$P(\hat{\theta}_i - \gamma < \theta_i < \hat{\theta}_i + \gamma \mid x'_1, x'_2, ..., x'_n)$ corresponderá a una medida de que tan buena en exactitud es la estimación $\hat{\theta}_i$, donde γ es un valor fijo positivo. Algunos autores creen que la aplicación del Teorema de Bayes se debilita porque la probabilidad a priori $P(\theta_1, \theta_2, ..., \theta_k)$ casi siempre es desconocida, pero otros sugieren usar en este caso distribuciones uniformes o no informativas (Lecoutre, 1999, Díaz, 2007).

2.2.3. FISHER Y LA TEORÍA FIDUCIAL

Para tratar de salvar las dificultades que surgen al intentar usar el Teorema de Bayes cuando no existe información a priori sobre los parámetros, Fisher (1930) introduce la idea de *intervalos fiduciarios.* El argumento fiducial representa una nueva manera de razonar de las observaciones a los valores de sus parámetros bajo discusión.

Fisher, que siempre se preocupó por mantener una visión frecuencial de la probabilidad desarrolló el método llamado "fiducial" (Rouanet, 1998) basado en la función de verosimilitud. La probabilidad fiducial trata de expresar la frecuencia con que el valor verdadero de un parámetro toma un valor determinado; a partir de los datos observados, por ejemplo que la probabilidad de que cierto parámetro poblacional sea menor que un valor dado es del 5%. Más concretamente, para el caso de una población normal con σ conocida, digamos que para el valor dado $\mu_1 = \overline{X} - 1.64 \frac{\sigma}{\sqrt{n}}$. Lo que el enunciado de probabilidad fiducial afirma es que P ($\mu < \mu_1$) = 0.05 lo cual se interpreta como que en el 5 % de todas las posibles muestras, μ no alcanzarán el valor obtenido en los datos, esto es μ_1.

Si $\hat{\theta}$ es un estadístico y P la probabilidad de que $\hat{\theta}$ sea menor que un valor específico, entonces tenemos una relación de la forma:

$$P = F(\hat{\theta}, \theta)$$

Si ahora damos a P cualquier valor particular tal como 0.95, tenemos una relación entre el estadístico $\hat{\theta}$ y el parámetro θ, tal que $\hat{\theta}$ es el valor en el 95 % percentil correspondiente a un θ dado, y esta relación implica que en el 5% de las muestras $\hat{\theta}$ excederá al percentil 95% correspondiente al valor real de θ de la población de la cual fue extraída, Fisher (1930) llama a esa relación el "5% fiducial del valor de θ" correspondiente a un valor dado de $\hat{\theta}$. Si $\hat{\theta}$ se incrementa con θ para todos los valores posibles, debemos expresar la relación diciendo que el verdadero valor de θ será menor que el 5% fiducial del valor correspondiente al valor observado de $\hat{\theta}$ en exactamente 5 intentos de 100" (Fisher, 1930).

Este método trata de evitar las probabilidades a priori de las hipótesis (como la estadística frecuencial), pero produce probabilidades a posteriori de las hipótesis, dados los datos (como la inferencia bayesiana). Rouanet (1998) indica que en algunos casos las distribuciones fiduciales de Fisher coinciden con las distribuciones bayesianas a posteriori, de modo que se podría considerar que Fisher fue un bayesiano, sin saberlo.

El procedimiento de máxima verosimilitud de Fisher

Para Fisher uno de los objetivos de la investigación empírica debe ser la búsqueda de la máxima verosimilitud. Esta idea da lugar a los métodos de estimación de máxima verosimilitud. (Rivadulla, 1991). En sus inicios, Fisher (1922) introdujo la verosimilitud como una "cantidad para designar el estado de nuestra información con respecto a los parámetros de la población hipotética" y más concretamente como una "medida de nuestro orden de preferencias entre las diferentes poblaciones posibles" (mencionado en Rao, 1992).

El principio de estimación de máxima verosimilitud nos conduce a escoger como el estimador del parámetro poblacional, para un conjunto de datos, ese valor para el cual el conjunto de datos observados habrían de ser los más probables de ocurrir. En este procedimiento se denomina a la función de máxima verosimilitud de los parámetros θ_i a la medida:

$$L = constante * P(x'_1, x'_2, ..., x'_n \mid \theta_1, \theta_2, ..., \theta_k)$$

de nuestra confianza en los valores correspondientes de θ.

Los valores $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k$ para los que L es máxima, recogen nuestra mayor confianza. $\hat{\theta}_i$, que es una función de $x'_1, x'_2, ..., x'_n$ constituye la estimación de máxima verosimilitud de θ_i el cual, si satisface ciertas condiciones de regularidad (Wasan, 1970, Bickel y Doksum, 1977) tiene las siguientes propiedades: existe y es único, estimador consistente de θ_i y es eficiente asintóticamente.

Algunas críticas alrededor de esta teoría son apuntadas por Rao (1992) quién afirma que no resulta claro como la medida de la verosimilitud de un valor dado del parámetro puede ser de ayuda en guiar investigaciones, además de que la función de verosimilitud no siempre puede ser definida y de que hay algunas otras dificultades con el manejo de los estimadores de máxima verosimilitud cuando están presentes parámetros que causan ruido en la estimación de los parámetros de interés (Cox, 1978). Pearson (1966) tampoco estuvo de acuerdo con las nociones teóricas de Fisher.

2.2.4. EL PROGRAMA NEYMAN- PEARSON Y LOS INTERVALOS DE CONFIANZA

Jerzy Neyman (1894-1981) inicia la teoría moderna de intervalos de confianza e introduce el término[1] en 1934 con su trabajo, "On the two different aspects of the representative method"[2], arrancando con ello la teoría estadística de la estimación del programa Neyman- Pearson. Sin embargo su trabajo más importante y más difundido sobre el problema de la estimación estadística es "Outline of a theory of statistical estimation based on the classical theory of probability " de 1937.

Neyman (1941) intenta probar, aunque pasando por algunos momentos de serias dudas, que entre la teoría fiducial y la teoría de los intervalos de confianza no había ninguna relación. Dudas provocadas, entre otras cosas, por la identidad numérica de los

[1] El término intervalos de confianza es una traducción del término polaco original ‘prezedzial ufnosci’.

[2] Aparecido en el volumen 97 del *Journal of the Royal Statistical Society*, en 1934**,** 558-625.

límites fiduciales de Fisher con los límites de sus intervalos de confianza (Rivadulla, 1991).

El problema de la estimación, para Neyman puede descomponerse en dos. Uno práctico, tiene que ver con la aritmetización que se necesita realizar con los datos obtenidos durante el muestreo para alcanzar un resultado, llamado *estimación* (del parámetro) que se acerque al valor estimado, es decir, con el procedimiento de cálculo en la estimación. El otro subproblema tiene que ver con la precisión matemática de ciertas nociones, que se presentan en forma vaga, en el trabajo práctico.

Neyman define la estimación en términos de probabilidad que para él constituye el vínculo matemático más cercano al término“presumiblemente”. La validez de la ley de los grandes números, es un supuesto importante en su teoría de la estimación estadística, que tiene una base frecuencial, pues asume que siempre que repitamos un experimento aleatorio con una probabilidad constante *P* de cierto resultado *A*, a la larga, la frecuencia relativa de la aparición de *A* se aproximará a *P*.

En el procedimiento usado para dar una estimación $\hat{\theta}$ del parámetro θ, Neyman considera que la estimación $\hat{\theta}$ nunca será exactamente igual al parámetro θ, por lo cual resulta necesaria una medida de la exactitud de la estimación $\hat{\theta}$. Para resolver esta cuestión sugiere calcular la desviación $S_{\hat{\theta}}$ con respecto a $\hat{\theta}$ y luego expresar el resultado de los cálculos en la forma $\hat{\theta} \pm S_{\hat{\theta}}$. Es decir, Neyman, sustituye una sola estimación, por dos estimaciones:

$$LI = \hat{\theta} - k_1 S_{\hat{\theta}} \text{ y } LS = \hat{\theta} + k_1 S_{\hat{\theta}}$$

donde: *LI* es la estimación inferior, *LS* es la estimación superior, $k_1 \; y \; k_2$ son constantes que indican los límites dentro de los cuáles, se presume, cae el verdadero valor del parámetro θ (Rivadulla, 1991).

Neyman (1934) sugiere que la solución del problema de estimación consiste en determinar ciertos intervalos, que denomina *intervalos de confianza,* en los cuales debemos asumir están contenidos los valores de los parámetros; la probabilidad de un error será igual o menor que 1-ε donde ε es cualquier número 0< ε <1, escogido anticipadamente. El número ε lo llama coeficiente de confianza.Los actuales procedimientos de la estimación estadística acostumbran designar como 1-α el coeficiente de confianza representado como ε por Neyman quién sugiere para la estimación del parámetro poblacional θ, se realicen los siguientes pasos (Mayo, 1981):

1. Elegir un valor del coeficiente de confianza, por ejemplo 1-α = 0.95,
2. El experimentador o estadístico debe llevar a cabo el experimento aleatorio y observar los valores obtenidos en la muestra $x_1, x_2, ..., x_n$ de las variables aleatorias,
3. Enseguida debe hacer uso de estos valores con el propósito de obtener los valores correspondientes *LI* y *LS* de las estimaciones inferior y superior respectivamente,
4. Debe afirmar que $LI < \Theta < LS$, donde Θ designa el valor verdadero del parámetro θ.

Neyman piensa que al aplicar los pasos a), b), c) y d) para la estimación del valor del parámetro θ, el experimentador o estadístico tendrá razón en dicha estimación, en *aproximadamente un porcentaje de los casos* a la larga. Neyman presenta de esta manera la teoría de la estimación estadística por medio de sus intervalos de confianza, buscando la máxima exactitud posible del resultado al determinar las estimaciones inferior (*LI*) y superior (*LS*). Con ello muestra de paso, que el problema de la estimación puede ser resuelto con base en la teoría frecuentista de probabilidades y sin requerir algún conocimiento de probabilidades a priori.

Neyman consideraba, según Mayo (1981), que era esencial que se entendiera que la probabilidad 1- α se refiere a los valores de la estimación inferior y estimación superior, pero que de ninguna manera representa la probabilidad de que el verdadero valor del parámetro θ esté situado dentro de tales límites.

Además puntualiza Mayo (p.273) “todo lo que se puede concluir es que los intervalos de confianza NP no proporcionan medidas de precisión final[3], sino más bien medidas de precisión inicial”[4].

2.2.5. CONCLUSIONES DEL ESTUDIO HISTÓRICO

Como conclusión del estudio histórico señalamos algunos hallazgos interesantes como lo son los siguientes:

El procedimiento bayesiano constituyó un esfuerzo interesante como solución al problema de la estimación, limitado por la necesidad de conocimiento de la probabilidad a priori de las causas, aunque salvable bajo ciertas circunstancias

[3] La probabilidad de que el parámetro caiga dentro de un intervalo específico es referido como una medida de precisión final (Savage, 1962)

[4] La precisión inicial es la probabilidad -antes de observar los datos-de que el intervalo de confianza que generarán los datos, contendrá el verdadero valor del parámetro (Mayo, 1981).

(distribuciones uniformes o no informativas). Es además la primera solución aportada al problema de estimación por intervalos, casi 200 años antes de que se creasen los intervalos de confianza frecuenciales.

Fisher con sus constructos de verosimilitud y probabilidad fiducial trata de sustituir la probabilidad subjetiva, aportada por el método bayesiano, por una medida de creencia racional, basada únicamente en los datos observados, pero continúa argumentando sobre una población tomando como punto de partida los datos recolectados en el muestreo. Por ello en algunos casos llega a las mismas soluciones que los métodos bayesianos.

Las teorías de la probabilidad fiducial de Fisher permiten calcular valores de verosimilitud sobre los parámetros, pero no proporcionan una distribución de probabilidad acerca de parámetros desconocidos, por lo que no resultaron exitosos. Aún así, hoy día la idea de verosimilitud y máxima verosimilitud es una de las principales teorías en estimación. Por otro lado los métodos bayesianos no informativos se apoyan en el método fiducial de Fisher (Lecoutre, 1999).

La historia nos revela que el nivel de confianza $(1 - \alpha)100\%$ significa para la teoría Neyman- Pearson que a la larga el $(1 - \alpha)100\%$ de los intervalos incluyen el valor verdadero del parámetro θ que se desea estimar y que la teoría Bayesiana lo considera como la probabilidad de que el verdadero valor del parámetro θ está contenido dentro de los límites de confianza. Esta segunda interpretación, que es la que intuitivamente da (erróneamente) la mayoría de los investigadores, es históricamente anterior; por tanto podría ser más intuitivo el método Bayesiano que el frecuencial.

El objeto intervalos de confianza emerge ante la resistencia de muchos investigadores a aceptar las probabilidades subjetivas propugnadas por el procedimiento bayesiano. Sin embargo, el razonamiento utilizado en su construcción es poco natural; incluso el sentido frecuencial de la probabilidad puede ser menos intuitivo en inferencia que el sentido subjetivo.

En resumen esta sección nos permitirá analizar la interpretación dada a los intervalos en los libros de texto (Capítulo 3) e interpretar los errores y conflictos semióticos en el aprendizaje de nuestros alumnos del objeto matemático intervalos de confianza (Capítulo 6).

2.3. MARCO TEÓRICO

2.3.1. INTRODUCCIÓN

En esta tesis aplicaremos algunas de las nociones que componen el modelo teórico propuesto por Godino y colaboradores (Godino y Batanero, 1994; Godino, 2002; Godino, Batanero y Roa, 2005; Godino Batanero y Font, 2007) quienes lo denominan "enfoque ontosemiótico" (EOS) de la cognición matemática. Este enfoque teórico proporciona una perspectiva pragmática – antropológica sobre el conocimiento matemático que puede complementar y enriquecer el análisis que hemos hecho en la sección 1.3.

El EOS propone tres dimensiones en el análisis de la enseñanza y el aprendizaje de las matemáticas: epistemológica, cognitiva e instruccional. Cada una de ellas se aborda con herramientas agrupadas en tres modelos teóricos: teoría de los significados institucionales y personales de los objetos matemáticos (Godino y Batanero, 1994, 1998); teoría de las funciones semióticas (Godino, 2002; Godino, Batanero y Roa, 2005) y teoría de las configuraciones didácticas (Godino, Contreras y Font, 2006). En esta Memoria nos ocuparemos principalmente de las dimensiones epistemológica y cognitiva, que describimos con más detalle en lo que sigue.

El punto de partida del EOS es una ontología de objetos matemáticos que considera un triple aspecto de la actividad matemática: como actividad de resolución de problemas socialmente compartida, como lenguaje simbólico y como sistema conceptual lógicamente organizado. Tomando como noción primitiva la de situación-problemática, se definen los conceptos teóricos de práctica, objeto (personal e institucional) y significado, con el fin de hacer patente la personal e institucional del conocimiento matemático.

2.3.2. SISTEMAS DE PRÁCTICAS OPERATIVAS Y DISCURSIVAS LIGADAS A CAMPOS O TIPOS DE PROBLEMAS

Los autores denominan *práctica matemática* a toda acción o expresión (verbal, gráfica, etc.) realizada por alguien para resolver problemas matemáticos, comunicar a otros la solución obtenida, validarla o generalizarla (Godino y Batanero, 1994, p. 334). Por ejemplo, al resolver problemas de toma de decisión, se realizan prácticas de análisis de datos tomados en muestras de las poblaciones, estudio de sus estadísticos y construcción de un intervalo de valores en torno a los mismos, en los cuáles se piensa podría estar un parámetro de una población. Las prácticas pueden ser específicas de una persona o compartidas en el seno de una institución. Una institución está constituida por las personas involucradas en una misma clase de situaciones problemáticas; lo que conlleva la realización de unas prácticas sociales que son generalmente condicionadas por los instrumentos disponibles en la misma, sus reglas y modos de funcionamiento. En el caso que nos ocupa podemos identificar diferentes instituciones, como la de estadísticos teóricos, investigadores que usan la inferencia o instituciones de enseñanza. En cada una de ellas las prácticas sociales son diferentes al variar los instrumentos disponibles y reglas de funcionamiento.

En el estudio de las matemáticas, más que una práctica particular ante un problema concreto, interesa considerar los sistemas de prácticas de las personas en su actuación ante tipos de situaciones problemáticas. A la pregunta, sobre qué es el objeto matemático[5] intervalo de confianza y qué significa esta expresión se propone como respuesta, el sistema de prácticas que realiza una persona (significado personal), o compartidas en el seno de una institución (significado institucional) para resolver un tipo de situaciones-problemas en los cuales se requiere estimar uno o varios parámetros en una población, especificando el margen de error, por medio de una probabilidad.

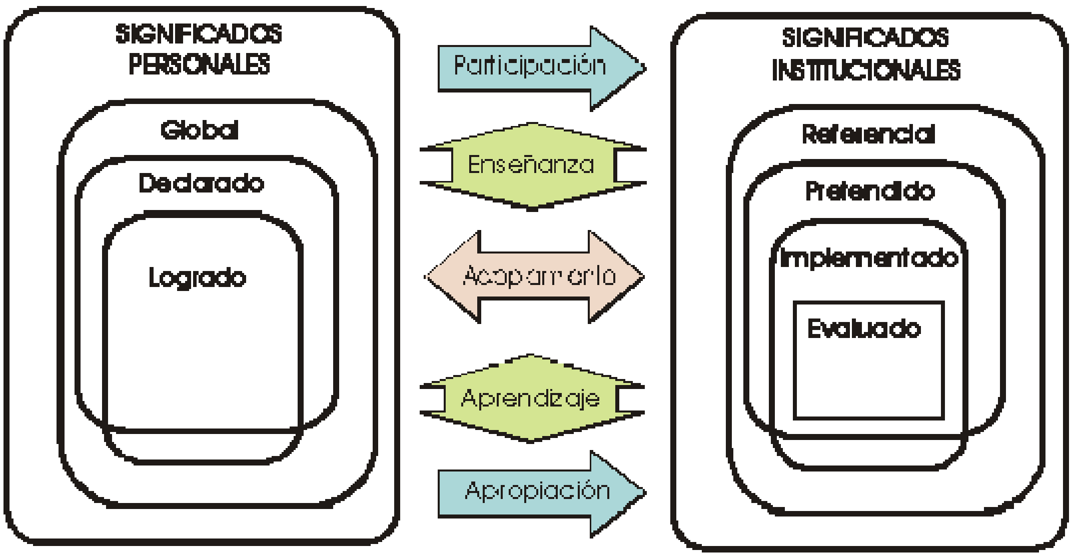
La relatividad socioepistémica y cognitiva de los significados, entendidos como sistemas de prácticas, y su uso en el análisis didáctico lleva a introducir la tipología básica de significados que se resume en la figura 2.1 (Godino, 2003, p. 141). En la parte central de la figura 2.1 se indican las relaciones dialécticas entre enseñanza y aprendizaje, que supone el acoplamiento progresivo entre los significados personales e institucionales. La enseñanza también supone la participación del estudiante en la comunidad de prácticas que soporta los significados institucionales, y el aprendizaje, en última instancia, supone la apropiación por el estudiante de dichos significados.

En nuestro trabajo nos interesamos tanto por los significados institucionales como personales. Respecto a los significados institucionales describiremos el significado de referencia, a partir del análisis de libros de texto y, a partir de él definiremos el significado evaluado al analizar el contenido del cuestionario. Respecto a los significados personales, nos interesamos por el significado global, pero tendremos que limitarnos al declarado en las respuestas al cuestionario y dentro de éste diferenciaremos

[5] Inicialmente los autores usaron la expresión ‘objeto matemático’ como sinónimo de ‘concepto matemático’. Más adelante extienden su uso y engloba proposiciones, argumentos, procedimientos, etc. que intervienen de algún modo en la actividad matemática.

en logrado (de acuerdo con el significado institucional) del no logrado.

Figura 2.1. Tipos de significados institucionales y personales



2.3.3. OBJETOS INTERVINIENTES Y EMERGENTES DE LOS SISTEMAS DE PRÁCTICAS

Godino (2003) indica que en las prácticas matemáticas intervienen objetos matemáticos que evocamos y que son representados en forma textual, oral o gráfica. También de los sistemas de prácticas matemáticas emergen nuevos objetos que dan cuenta de su organización y estructura. Los objetos emergentes son "objetos institucionales" si los sistemas de prácticas son compartidos en el seno de una institución, y "personales"si corresponden a una persona[6].

Godino, Batanero y Font (2006) proponen la siguiente tipología de objetos matemáticos primarios que analizaremos en nuestro trabajo y que a su vez se organizan en entidades más complejas como sistemas conceptuales o teorías:

- *Situaciones-problemas* (aplicaciones extra-matemáticas, ejercicios, ...)
- *Lenguajes* (términos, expresiones, notaciones, gráficos, ...) en sus diversos registros (escrito, oral, gestual, ...)
- *Conceptos- definición* (introducidos mediante definiciones o descripciones, por ejemplo: recta, punto, número, media, función, ...)

[6] En la idea de "objetos personales" los autores incluyen las concepciones, esquemas, representaciones internas, etc.

- *Proposiciones* (enunciados sobre conceptos, ...)
- *Procedimientos* (algoritmos, operaciones, técnicas de cálculo, ...)
- *Argumentos* (enunciados usados para validar o explicar las proposiciones y procedimientos, deductivos o de otro tipo, ...).

Estos seis tipos de entidades primarias amplían la tradicional distinción entre conceptos y procedimientos así como los componentes del "triángulo epistemológico" (signo/símbolo, objeto/contexto de referencia, concepto). Las situaciones-problemas son el origen de la actividad; el lenguaje representa las restantes entidades y sirve de instrumento para la acción; los argumentos justifican los procedimientos y proposiciones que relacionan los conceptos entre sí.

El modelo descrito es recursivo pues, dependiendo del tipo de análisis, cada unidad primaria puede estar compuesta por entidades de los restantes tipos (un problema, por ejemplo, puede poner en juego conceptos, proposiciones, procedimientos, etc.).

Configuraciones de objetos

Godino, Batanero y Font (2006) indican que los seis tipos de entidades primarias descritos anteriormente están relacionados entre sí formando *configuraciones*, es decir redes de objetos intervinientes y emergentes de los sistemas de prácticas. Los autores clasifican estas configuraciones en *epistémicas* (cuando se trata de objetos institucionales) o *cognitivas* (si se refieren a objetos personales). Describiremos algunas de ellas relacionadas con los intervalos de confianza en nuestro trabajo.

La constitución de estos objetos y configuraciones, tanto en su faceta personal como institucional, tiene lugar a lo largo del tiempo mediante procesos matemáticos, que los autores interpretan como "secuencias de prácticas" dando criterios para categorizarlos. La formación de objetos lingüísticos, problemas, definiciones, proposiciones, procedimientos y argumentos tiene lugar mediante los respectivos procesos matemáticos primarios, de comunicación, problematización, definición, enunciación, elaboración de procedimientos y argumentación.

2.3.4.3. RELACIONES ENTRE OBJETOS: FUNCIÓN SEMIÓTICA

Otro punto de interés en nuestro trabajo es la idea de función semiótica que los autores introducen a partir de trabajos previos de Hjelmslev (1971) (noción de función de signo) y Eco (1979) (función semiótica). Para Godino (2002) una función semiótica es unas correspondencia entre un antecedente (expresión, representante) y un consecuente (contenido, representado), establecida por un sujeto (persona o institución) de acuerdo con un cierto criterio o código de correspondencia. Por ejemplo la expresión "nivel de confianza" remite a un valor de probabilidad, pero interpretado como un porcentaje de intervalos construidos a partir de una misma muestra que cubren el verdadero valor del parámetro. En este ejemplo vemos que el código "nivel de confianza" se ha establecido mediante un convenio e informa a los sujetos sobre qué objetos hay que poner en correspondencia con el mismo.

Las relaciones de dependencia entre expresión y contenido pueden ser de tipo representacional (un objeto se pone en lugar de otro para un cierto propósito, como en el ejemplo dado), instrumental (un objeto usa a otro u otros como instrumento, por ejemplo cuando operamos con símbolos), y estructural (dos o más objetos componen un sistema del que emergen nuevos objetos, como el caso de los extremos del intervalo que conjuntamente lo definen). De esta manera, las funciones semióticas y la ontología matemática asociada, tienen en cuenta la naturaleza esencialmente relacional de las matemáticas y generalizan la noción de representación, que no queda asumido en exclusividad por el lenguaje. Los autores indican que cualquier tipo de objeto (situaciones-problemas, conceptos, proposiciones, procedimientos y argumentos), pueden ser expresión o contenido de las funciones semióticas.

En algunas ocasiones se produce un desajuste entre el significado asociado a una expresión por dos personas y hablamos de "conflicto semiótico". Un ejemplo es cuando se da una interpretación bayesiana al nivel de confianza, que es una probabilidad condicional, intercambiando sus términos. En nuestro trabajo se tratará de caracterizar algunos conflictos semióticos de los estudiantes en relación al objeto "intervalo de confianza" y se llevará a cabo un análisis semiótico de respuestas a los ítems abiertos (capítulo 7) para mostrar la complejidad del significado personal de los estudiantes en relación al objeto "intervalo de confianza".

2.3.5. EVALUACIÓN DE LA COMPRENSIÓN

El objetivo principal de nuestro trabajo es construir un cuestionario para evaluar el significado personal que los estudiantes asignan al intervalo de confianza y determinar la parte del mismo que está de acuerdo al significado institucional. Godino (1996) sugiere que para analizar la comprensión es preciso responder a dos preguntas:

qué comprender, y *cómo* lograr la comprensión. Su modelo de la comprensión tiene dos ejes: uno *descriptivo*, que indica los aspectos o componentes de los objetos a comprender, y otro *procesual* que indica las fases o niveles necesarios en el logro de la "buena" comprensión. Esta investigación se centrará principalmente en el eje descriptivo.

Godino (1996) también recuerda que la comprensión personal es inobservable (sería un constructo, en términos psicológicos) mientras que el significado, como conjunto de prácticas, es por lo menos potencialmente observable. El autor concibe la evaluación como el estudio de la correspondencia entre los significados personales e institucionales. La evaluación de la comprensión de un sujeto tiene que ser relativizada a los contextos institucionales. Desde el punto de vista de la institución un sujeto "comprende" el significado de un objeto si dicho sujeto es capaz de realizar correctamente las distintas prácticas que configuran el significado de dicho objeto institucional.

Cuando se quiere caracterizar el significado de un objeto en una institución o para una persona, las prácticas observables son los indicadores empíricos que nos permiten esta caracterización. Por lo tanto, en esta investigación y siguiendo el marco teórico, se infiere la comprensión personal de los alumnos participantes sobre el intervalo de confianza mediante el análisis de las respuestas (entendidas como prácticas), realizadas por la persona en la resolución de tareas problemáticas o ítems de evaluación propuestas.

Para poder comprender la diferencia entre los significados que son objetos del estudio y lo que realmente es posible determinar en la investigación, Batanero y Díaz (2005) recuerdan los diferentes procesos de inferencia llevado a cabo en un proceso de evaluación que condicionan la generalizabilidad de las inferencias. Nosotros estamos interesados en determinar el significado institucional del intervalo de confianza en la enseñanza a ingenieros. Pero no es posible acceder a todas las clases que se dan en esta institución (que además podrían potencialmente variar de un curso a otro, incluso para un mismo profesor). Una forma de acercarse al significado de interés es el análisis de libros de texto recomendados, que no puede ser completo, sino se lleva a cabo un muestreo de libros. Con ello construimos el significado de referencia del proceso de evaluación, en cuanto en base a él se organiza la construcción del cuestionario e interpretación de las respuestas de los alumnos.

Una vez fijado el significado de referencia hay infinitos posibles instrumentos de evaluación. Nosotros trataremos de construir un instrumento lo más objetivo posible, y para ello recopilaremos diferentes ítems, tomados de investigaciones previas en las que han sido evaluados y han dado resultados probados, para cada una de las unidades de contenido. Para evitar el introducir elementos subjetivos en la elección de los ítems o tareas recurrimos a juicio de expertos tratando de crear un *significado compartido* de lo que sería un instrumento de evaluación. Las pruebas de los ítems con estudiantes tratan de asegurar la legibilidad y la idoneidad cognitiva de los mismos. El cuestionario resultante define un significado nuevo, sería el significado evaluado, diferente de los anteriores.

Finalmente el instrumento se prueba con una muestra de estudiantes. La respuesta que cada uno de ellos proporciona a cada ítem no es la única que puede dar. Dependiendo del interés, cansancio, concentración y otros factores, sus respuestas reflejan una parte de lo que el estudiante realmente conoce. Sería el *significado declarado* que es el finalmente accesible al investigador.

2.3.6. IDONEIDAD DE UN CUESTIONARIO DE EVALUACIÓN

Además de estudiar con detalle la comprensión adquirida por los estudiantes sobre los diferentes elementos de significado incluidos en el cuestionario, este trabajo se interesa por analizar hasta qué punto el cuestionario construido ha sido eficaz. Para ello se analizarán los criterios de idoneidad de un cuestionario de evaluación indicados por Batanero y Díaz (2005b) quienes adaptan al caso de un cuestionario, algunos de los propuestos por Godino, Contreras y Font (2006) para los procesos de estudio.

El interés de nuestro estudio no se limita al significado declarado sino estamos interesados en el *significado personal global* de los alumnos respecto a las tareas propuestas (las respuestas dadas se suponen una muestra representativa de las que darían los mismos estudiantes en la misma prueba en otras ocasiones). Batanero y Díaz (2005b) indican que si las tareas del cuestionario son suficientemente representativas (para evaluar las unidades de contenido definidas), podemos hacer una inferencia sobre lo que cada alumno de la muestra sería capaz de hacer y decir en otras tareas relacionadas con el objeto.

Es decir, si las unidades del contenido están bien definidas y representan el objeto intervalo de confianza podríamos acercarnos al significado personal de los alumnos de la muestra sobre el mismo. Asimismo, podemos obtener conocimiento generalizable sobre las dificultades y capacidades de otros estudiantes similares a los de la muestra. Las autoras indican que la posibilidad de generalizar en cada uno de los pasos descritos depende de la representatividad y la variabilidad de la muestra elegida en cada uno de los procesos de muestreo.

En este sentido, aplican el concepto *idoneidad* y sus tipos (Godino, 2003; Godino, Contreras y Font, en revisión) al caso de la evaluación:

- La *dificultad* de un ítem o tarea daría una medida de su *idoneidad cognitiva*; es decir del grado de representatividad de los significados evaluados respecto a los significados personales.
- La *discriminación* de un ítem valoraría su *idoneidad evaluadora*, un ítem puede ser adecuado cognitivamente, pero no diferenciar (por ser demasiado fácil) los alumnos que tienen un mayor o menor conocimiento del concepto. Esta idoneidad podría ser un componente de la idoneidad *instruccional*, en cuanto uno de los objetivos de la instrucción es la función evaluadora.
- La *validez de contenido* de un cuestionario indicaría una *idoneidad epistémica*, o grado de representatividad del instrumento en cuanto al significado objeto de evaluación.
- La *fiabilidad* o *generalizabilidad a otros ítems* daría una medida de la estabilidad de la respuesta, es decir sería otro componente de la *idoneidad evaluadora*
- La *validez externa* y *generalizabilidad a otros estudiantes,* sugeriría una *idoneidad generalizadora o externa* en cuanto los resultados se generalizarían a otros estudiantes.

2.4. ESTADO DE LA CUESTIÓN

2.4.1. INTRODUCCIÓN

En este apartado presentamos investigaciones que se centran alrededor de la enseñanza y aprendizaje de los intervalos de confianza o de otros temas vinculados. Resumiremos los trabajos que de algún modo se relacionan con la comprensión de los intervalos de confianza y otros donde, con ayuda del ordenador, se hacen esfuerzos por facilitar la enseñanza de los intervalos de confianza o bien la comprensión de algún otro tema relacionado con ellos. También presentamos otros trabajos enfocados en temas muy relacionados como son la comprensión del teorema central del límite, de la distribución normal y de las distribuciones muestrales, elementos que tienen un peso importante en los intervalos de confianza.

2.4.2. COMPRENSIÓN DEL INTERVALO DE CONFIANZA

Proporcionar intervalos de confianza es un método que está siendo cada vez más apreciado para realizar inferencias estadísticas y al que muchos investigadores le añaden ventajas pedagógicas en la enseñanza, en comparación a las hipótesis nulas de las pruebas de significación. Sin embargo su interpretación no está exenta de problemas como mostramos a continuación.

Errores de interpretación de los intervalos de confianza por los investigadores

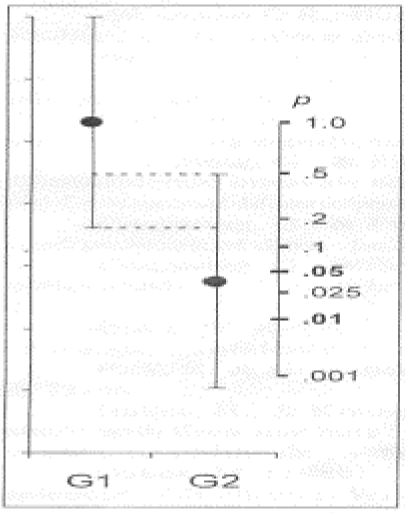
Las dificultades en la enseñanza de las pruebas de hipótesis son ampliamente conocidas y trabajos como los de Vallecillos (1994; 1998, 1999) y Falk y Greenbaum (1995) y otros recogidos en Harlow, Mulaik y Steiger (1997) y Batanero (2000) han demostrado que, aunque se enseñe a los estudiantes las posibles fallas de las pruebas de hipótesis, las interpretaciones de los estudiantes sobre los resultados de la investigación no mejoran. Los intervalos de confianza, entre otros beneficios, pueden aclarar concepciones erróneas asociadas a la significación estadística.

Cumming, William y Fidler (2004) investigan sistemáticamente los errores de los investigadores, la mayoría de los cuáles se interpretan en nuestro marco teórico como conflictos semióticos, pues se trata de errores de interpretación. En correos electrónicos enviados a autores de artículos publicados en revistas internacionales los invitaron a visitar un sitio de la red donde se presentaban diagramas de una media con sus intervalos de confianza al 95% y se les preguntó que indicaran donde pensaban que era más probable que cayera la media si el experimento se replicaba un número de veces. La mayoría mantenían la concepción equivocada de una alta probabilidad (sobre 95%) de la replicación, esto es que la media caiga de nuevo en el intervalo de confianza original, cuando de hecho esta probabilidad en promedio es únicamente alrededor del 83% (Estes, 1997).

En un segundo estudio con apoyo de la Internet, Belia, Fidler y Cumming (2005) invitaron a investigadores a que visitaran un sitio de la red donde se presentaba la figura 2. Se les pidió que pincharan para mover la media G2 hasta que ellos juzgaran que las dos medias eran significativamente diferentes en la prueba *t*, de dos colas para α =.05 (la configuración en la figura 2.2 es la respuesta correcta, ya que el *valor p* es cerca de

.05). Las respuestas fueron muy diversas y solo unos cuantos sujetos se acercaron a la respuesta correcta. El 34% de los que respondieron ajustaron las medias de modo que los intervalos de confianza se alinearan en el punto final de una recta con el punto final de la otra, para evitar el traslape. Esta es una creencia errónea muy común en medicina (Schenker y Gentleman, 2001), de que los intervalos de confianza al 95% de medias independientes son significativamente diferentes cuando los dos intervalos se tocan justo extremo con extremo.

Figura 2.2. Medias de dos grupos independientes, con IC al 95%



Un grupo de los que contestaron observó la misma figura, pero con las dos medias etiquetadas como pretest y postest, y describió los resultados como puntajes para un grupo simple. Pero para los datos apareados, no es posible realizar la tarea en la forma supuesta por los participantes, porque el intervalo de confianza de la diferencia de medias no es igual a la diferencia de dos intervalos de confianza por separado. Una gran mayoría de los que respondieron resolvió el problema como si se tratase de dos medias independientes.

En relación a la interpretación de intervalos de confianza para dos medias independientes, una solución que se sugiere para ayudar a la interpretación es a través de las “reglas a ojo”. Cumming y Finch (2005) dan la siguiente regla, para una diferencia de medias estadísticamente significativa:

> “*Se puede comparar dos medias independientes a nivel $p \leq .05$, cuando el traslape de los intervalos de confianza al 95% no es más que alrededor de la mitad del*

margen promedio del error, esto es cuando la proporción del traslape es alrededor de .5 o menos.

Para $p \leq .01$ cuando los dos intervalos de confianza no se traslapan, esto es cuando la proporción de traslape es alrededor de 0 o hay un espacio vacío positivo.

Esas relaciones son lo suficientemente exactas cuando ambas muestras son al menos de tamaño 10, y cuando los márgenes de error no difieran en más de un factor de 2. Esto es $\frac{w_1}{w_2} \leq 2$, donde w_i es el valor sumado y restado del punto estimado en el grupo i, en la construcción del intervalo de confianza" (pg. 176).

Comprensión por estudiantes

Se conoce muy poco acerca de la comprensión de los estudiantes de los intervalos de confianza, sin embargo en los últimos años se han realizado importantes avances.

Fidler y Cumming (2005) proveen evidencia de cómo los intervalos de confianza dan sentido a la ausencia de significación estadística. En su estudio piden a estudiantes de ciencias del ambiente (último año de estudios y postgrado) que interpreten resultados de valores *p* e intervalos de confianza a partir de publicaciones en revistas científicas. Se les preguntó a los estudiantes si los resultados proveían evidencia fuerte, moderada o ambigua a favor o en contra de la hipótesis nula. Cuando se da el *valor p*, el 44% de los estudiantes (24 de 55) malinterpretan los resultados como evidencia a favor de la hipótesis nula. Sin embargo menos de la mitad (18%, 10 de 55) lo malinterpretan cuando se les presentaron los resultados como intervalos de confianza. También hubo evidencia de un efecto de aprendizaje: los estudiantes que estudiaron primero el caso de intervalos de confianza y luego vieron la presentación como *valor p*, dieron la respuesta correcta a la pregunta sobre pruebas de hipótesis con más frecuencia que los estudiantes que vieron los *valores p* primero.

Los autores del estudio también avisan que pueden existir ideas erróneas acerca de los intervalos de confianza. En una encuesta a 180 estudiantes de psicología, muchos estudiantes sólo consideraban a los intervalos de confianza como estadísticos descriptivos, ignorando su naturaleza inferencial. Por ejemplo, un 38% dijo que los intervalos de confianza proveían "valores plausibles para la media de la muestra". El 19% pensaba que era el "rango" o "el rango truncado de los valores individuales".

Por otra parte, algunos estudiantes tienen ideas equivocadas sobre cómo los distintos aspectos de los intervalos de confianza se relacionan entre sí. Por ejemplo, 20% de los estudiantes dijeron que el ancho del intervalo de confianza se incrementaría si se aumentara el tamaño de la muestra, un 29% dijo que el ancho del intervalo de confianza no se vería afectado y el 36% de los estudiantes no sabía si existía alguna relación. Solamente un 16% pudo contestar correctamente esta pregunta elemental. Esto después de haber tomado un curso de un año de introducción en estadística con un fuerte enfoque en intervalos de confianza

Behar (2001), señala como hipótesis plausible, que el tener la heurística de representatividad en la asignación de probabilidades, no parece tener fuerte asociación con la comprensión de conceptos sobre intervalos de confianza y sobre contrastes de hipótesis. Siguiendo a Kahneman, Slovic y Tvesky (1982), Behar (2001) aplica un test a un grupo de expertos en estadística, y un grupo de estudiantes universitarios que habían tomado cursos de estadística equivalentes a 3 horas / semana, para valorar conceptos sobre intervalos de confianza y comprensión de conceptos sobre hipótesis estadísticas. En concreto, esta investigación explora tres aristas esenciales:

- La validación de los planteamientos de Kahneman, Slovic y Tvesky (1982), sobre algunos juicios bajo incertidumbre. Comparando la población de expertos con la de no expertos.
- Valoración de las concepciones relacionadas con la estimación por intervalos y contraste de hipótesis estadísticas, enfocados más específicamente hacia el concepto del *valor p*, los factores que lo afectan, su significado e interpretación en un contexto de aplicación.
- Estudio de las características de la batería de ítems incluida en el instrumento.

Siguiendo la orientación de Garfield, delMas y Chance (1999) y complementada con preguntas de otros autores del Departamento de Matemáticas y Estadística de la Simon Frase University, Behar(2001) elabora un instrumento en lo que respecta a la evaluación conceptual del aprendizaje sobre intervalos de confianza. El instrumento de 71 preguntas fue contestado por 47 expertos y 297 estudiantes de pregrado. Las preguntas de la 1 a la 9 corresponden a juicios sobre incertidumbre, de la 10 a la 33 valoran la comprensión de los conceptos sobre intervalos de confianza y de la 34 a la 71 comprensión de los conceptos de contrastes de hipótesis. De las 24 preguntas relacionadas con la comprensión de los conceptos sobre intervalos de confianza, seis preguntas son respondidas por menos del 50% de los expertos y el número de preguntas comparable con los no expertos es 11. De las respuestas a estas 24 preguntas se obtuvieron las siguientes conclusiones:

1. Se detecta una debilidad conceptual en la comprensión de la manera como se relacionan los distintos factores asociados con un intervalo de confianza, en especial el ancho del intervalo y el nivel de confianza, que parece no poder separarse. Es decir, se nota una tendencia a negar la posibilidad de disminuir la confianza, como si esto fuera inconsistente.
2. En lo más básico, en el significado esencial de un intervalo de confianza, se aprecia bastante debilidad, al parecer, porque se piensa que los valores que constituyen un intervalo de confianza, se refiere a posibles valores que toma la variable aleatoria que define la población, o a valores del estadístico que se usa como estimador, lo cual podría surgir del propio procedimiento de construcción del intervalo, que se deduce de probabilidades asociadas con dicho estadístico.
3. Una buena proporción de los participantes de ambos grupos, expertos y no expertos, parecen no asociar la confianza a un mecanismo aleatorio generador de intervalos, a partir de muestras aleatorias, ni asociar el nivel de confianza, con la frecuencia relativa a la larga, de que los intervalos generados por tal mecanismo aleatorio atrapen al verdadero parámetro de la población.
4. La utilidad de los intervalos de confianza para tomar decisiones sobre hipótesis, parece no estar comprendida por una buena parte de los participantes en la investigación. Una posible razón parece estar relacionada con no considerar los valores del intervalo como un conjunto de valores plausibles del parámetro.

2.4.3. ENSEÑANZA DE INTERVALOS DE CONFIANZA CON RECURSO DIDÁCTICO DEL ORDENADOR.

Hemos encontrado también algunas investigaciones sobre enseñanza basadas en simulación que, según Godino (1995), permite al alumno realizar exploración, mejorar su intuición probabilística y descubrir conceptos y principios que sin la ayuda de este instrumento serían mucho más abstractos.

Investigaciones de delMas, Garfield y Chance

Un trabajo importante es el de delMas, Garfield y Chance (2004), quienes tuvieron un proyecto para desarrollar y diseminar materiales que ayuden a los estudiantes en la comprensión de los conceptos centrales de la inferencia estadística. Han diseñado módulos instruccionales que usan software para simulación de intervalos de confianza, distribuciones muestrales y test de hipótesis. Basan su investigación en un modelo de cambio conceptual donde los estudiantes primero hacen predicciones y luego las prueban usando simulación con el software. Los módulos en Internet incluyen instrumentos donde evalúan el conocimiento previo requerido e instrumentos para evaluar la comprensión de los conceptos estadísticos, que se encuentran inmersos en la estadística inferencial.

En la lista de herramientas para los intervalos de confianza que incluyen delMas, Garfield y Chance en su proyecto están las siguientes:

1. Lo que deben comprender los estudiantes acerca de este tópico:

- Un intervalo de confianza para la media poblacional es la estimación de un intervalo de un parámetro poblacional desconocido (la media), basado en una muestra aleatoria de la población.
- Un intervalo de confianza para la media poblacional es un conjunto de valores plausibles del parámetro (μ) que pudiera haber generado el dato observado como un posible resultado.
- Un intervalo de confianza para la media poblacional consiste en un estadístico muestral más o menos una medida del error del muestreo (el cual es un error del muestreo aleatorio), cuando tenemos normalidad en la distribución muestral.
- El nivel de confianza nos dice la probabilidad de que el método construya un intervalo que incluya el parámetro desconocido. La probabilidad se relaciona al método (datos, intervalo), no al parámetro.
- Un incremento en el tamaño de la muestra conduce a decrecer el ancho del intervalo: grandes muestras tienen anchos de intervalos más estrechos que muestras más pequeñas (todos los otros elementos permaneciendo constantes).
- Altos niveles de confianza tienen intervalos más anchos que niveles de confianza más bajos (todos los otros elementos permaneciendo constantes).
- Intervalos estrechos y altos niveles de confianza son deseables, pero esas dos cosas se afectan una a la otra.
- Si muchas muestras aleatorias son recolectadas independientemente de una población y el 95% de los intervalos de confianza son construidos para cada

muestra, pudiéramos esperar que alrededor de un 5% de los intervalos no incluyen la media poblacional (el parámetro poblacional). Este 95% se refiere al proceso de tomar muestras repetidas y construir intervalos de confianza para cada muestra.

- Un intervalo de confianza sugiere que valores de parámetro son razonables dados los datos y todos los valores en el intervalo son igualmente plausibles como valores de μ que pudieran haber producido la media muestral observada.
- Después de que usted calcula un intervalo de confianza, el parámetro puede estar incluido o no, pero usted no lo sabe.
- Es deseable tener un ancho de intervalo angosto (para una estimación más precisa) con un alto nivel de confianza. Un ancho de intervalo angosto no es suficiente (si tiene un bajo nivel de confianza).

2. Lo que deben ser capaces de hacer los estudiantes con este conocimiento:

- Conocer como hacer un intervalo de confianza más ancho o más angosto (que factores pueden ser cambiados).
- Conocer como calcular un intervalo de confianza para una media dado un dato muestral.
- Conocer como interpretar un intervalo de confianza, hacer una inferencia apropiada (en contexto) y ser capaz de hacer una declaración correcta de probabilidad como una interpretación de una confianza.

3. Algunas concepciones erróneas comunes que los estudiantes no debieran tener:

- Hay un 95% de probabilidad de que el intervalo de confianza incluya la media muestral.
- Hay un 95% de probabilidad de que la media poblacional estará entre los dos valores (límite superior y límite inferior).
- 95% de los datos están incluidos en el intervalo.
- Un intervalo ancho significa menos confianza.
- Un intervalo de confianza más angosto es siempre mejor (a pesar del nivel de confianza).

4. Conocimientos previos requeridos que los estudiantes deben tener antes de usar las herramientas:

- Ideas de población, muestra, parámetro, estadístico.
- Desviación estándar de la media muestral, que significa (sigma dividido por la raíz cuadrada del tamaño de la muestra), como encontrarlo, que es una medida de la variabilidad de la media muestral, que dice acerca de la variabilidad de la media muestral.
- Variabilidad del muestreo/ error del muestreo, error estándar (s dividido por raíz cuadrada de n), media, desviación estándar, el concepto de probabilidad, Distribución z, Distribución muestral.
- Las estimaciones provienen de muestras, el verdadero valor puede únicamente ser obtenido conociendo toda la población. Esas estimaciones son cercanas al parámetro poblacional, pero no exactas.

El programa SIM Sampling fue el Software que originalmente fue desarrollado para investigar el impacto en los estudiantes sobre la comprensión de las distribuciones muestrales (delMas, 2001). Los autores usaron un Pretest para acercarse al conocimiento previo requerido antes de usar la herramienta, el cuál sirve como una medida de diagnóstico.

La actividad aborda el problema de un estudiante de economía que se graduará en unos cuantos meses, quién está interesado en estimar el sueldo que puede esperar cuando obtenga su grado, utilizando para ello solamente la información de que dispone de 16 compañías de los Estados Unidos de Norteamérica. La pregunta que plantea la actividad es si con esta información el estudiante puede estimar el salario promedio que se ofrece a los recién graduados en Economía.

A través de una lista de 82 preguntas organizadas en ocho partes y con la ayuda del programa SIM Sampling se va conduciendo al estudiante en el aprendizaje de los intervalos de confianza. En la organización de estas ocho partes se empieza con el caso simple de la población, de la cual se toma la muestra, sigue una distribución normal. En la parte I se simula la selección de una muestra de tamaño 16 de una población normal con media μ y σ conocida. Con los valores de la muestra el estudiante calcula el límite superior y el límite inferior y el software calcula también esos límites. Estos resultados son registrados en una tabla en donde además se registran: $\bar{x}$, s, el ancho del intervalo y el margen de error. En la parte II se toma una segunda muestra del mismo tamaño con la cual hace los mismos cálculos y luego aparecen una serie de preguntas comparativas con los resultados de la primera muestra y así continúa la actividad tomando muestras más pequeñas y luego muestras más grandes, tomando muestras de una población que no sigue una distribución normal, etc. El Post test incluye algunas cuestiones del pretest, algunas cuestiones relacionadas directamente con el software, y algunos problemas de aplicación.

En delMas, Garfield y Chance (1998) y Chance, Garfield y delMas (2004) se analiza el efecto de la simulación por ordenador para facilitar la comprensión conceptual de las distribuciones muestrales. En el primer estudio se plantea un modelo de evaluación basado en diversas recomendaciones de la literatura (Nickerson, 1995; Snir, Smith y Grosslight, 1995) para simulaciones mejoradas conceptualmente. Este modelo de evaluación, que fue diseñado con preguntas con mediciones en forma gráfica, está constituido por los siguientes elementos:

1. Un pretest que evalúa el conocimiento previo requerido y las intuiciones que pueden afectar las interacciones de los estudiantes con la actividad planeada.
2. Un listado de objetivos de evaluación final que especifican el aprendizaje deseado y que son usados para desarrollar la actividad de aprendizaje.
3. Preguntas de auto-evaluación involucradas en las actividades de aprendizaje, mediante las cuales los estudiantes realizan y evalúan sus predicciones, para promover los cambios conceptuales.
4. Un post test de resultados deseados que evalúan los tipos correctos de razonamiento conceptual o los errores.
5. Un post test que consta de preguntas que podrían ser incluidas en un examen final del curso para evaluar la retención a largo plazo.

En su investigación trabajan con el micro mundo Sampling distributions, diseñado para trabajar distribuciones muestrales y una actividad basada en el modelo de evaluación que previamente se describió, con el fin de evaluar la efectividad de las actividades planteadas por los investigadores. Las cuestiones nucleares de la investigación de Garfield, Chance y delMas son: identificar los problemas de aprendizaje, diseñar una técnica instruccional para mejorarlas, determinar si es o no efectiva y obtener conclusiones válidas para continuar la investigación. La experiencia se llevó a cabo en tres fases, en cada una de las cuáles se iban modificando las actividades o el software para tratar de mejorar la comprensión de los estudiantes sobre las distribuciones muestrales. En la primera fase se instruyó sobre como crear una distribución muestral por medio del software, de este modo realizaron varias simulaciones para distintos tamaños de muestra y luego se les pidió estudiar las siguientes cuestiones:

1. Relación entre tamaño de muestra y dispersión de la distribución muestral.
2. Determinar el tamaño de muestra en el cual empieza la estabilización de los estadísticos de la distribución.
3. Analizar si los estadísticos obtenidos son buenos estimadores de los parámetros de la población.
4. Estudiar el comportamiento de los estadísticos de la distribución muestral, revisando si hay concordancia con lo anticipado por el teorema central del límite.

Estas actividades son realizadas con poblaciones normales, así como poblaciones con formas asimétricas, uniformes, etc. Las preguntas y actividades logran que los estudiantes atiendan diversos puntos del teorema central del límite y evalúen sus hipótesis acerca del comportamiento de las distribuciones muestrales.

En la segunda fase o versión de la experiencia los investigadores hicieron mejoras al software y con ello los estudiantes realizaron sus propias predicciones y luego las evaluaron usando el nuevo software. En esta segunda versión se enfatizó mayormente en las comparaciones de la forma y la dispersión que en los parámetros y estadísticos y fue necesario que los estudiantes realizaran una comparación de sus predicciones en el pretest en forma directa con las distribuciones muestrales producidas a través del software.

En la tercera y última versión se incluye un programa que permite visualizar la distribución y los estadísticos para cada una de las muestras. Durante la prueba piloto aplicada de esta tercera versión, en el pretest, se incluyeron ítems relacionados con los conceptos que los investigadores consideraban que debían conocerse previos al manejo de las distribuciones muestrales, conceptos tales como: muestra, población, distribución, variabilidad y muestreo. delMas, Garfield y Chance (1998) señalan en sus conclusiones que la presentación de información en forma simple, sin significados ocultos, no conduce necesariamente a la comprensión sólida de las distribuciones muestrales.

Otras investigaciones

En los últimos años se han desarrollado numerosos programas computacionales instruccionales para ayudar a los estudiantes a desarrollar su aprendizaje sobre las distribuciones muestrales y otros conceptos estadísticos: HyperStat (Lane, 2001), Visual Statistics (Doane, Tracy y Mathieson, 2001), StatPlay (Cumming y Thomason 1998) y ActivStats (Velleman, 2003), no obstante es muy poco lo que se ha publicado que evalúe la efectividad de las actividades de simulación sobre la mejora en el razonamiento de los estudiantes sobre intervalos de confianza. Cumming (2002) ha desarrollado un software exploratorio para intervalos de confianza (ESCI). Con este software, que corre en Microsoft Excel, se pueden realizar simulaciones gráficas, provee interactivamente demostraciones y herramientas para la comprensión de los intervalos de confianza. También proporciona apoyo para la exploración de otros conceptos relacionados como pruebas de significación estadística, muestreo, etc.

Hin y Ping (2002**)** usan la simulación para intentar enseñar el intervalo de confianza para el parámetro pendiente e intervalo de predicción para una observación futura en el modelo de regresión lineal simple. Los autores señalan que se puede comprender fácilmente el significado de los estimados β_0, β_1 y $\hat{\sigma}^2$, pero no se puede comprender tan fácilmente las propiedades del intervalo de confianza y del intervalo de predicción, por el hecho de que para expresar dichas propiedades, se usa la palabra “probabilidad”. La manera en que los autores explican la “probabilidad” es mediante el uso de una simulación ilustrada paso a paso en programas hechos en JAVA para ayudar a los estudiantes a entender las propiedades del intervalo de confianza y del intervalo de predicción.

El programa en JAVA genera el vector de observación $(y_1, y_2, ..., y_n)$ repetidamente, y encuentra el intervalo de confianza que le corresponde a cada vector de observación generado. Mientras que los vectores de observación son generados, se construye el histograma para la *iésima* observación y_i para cada $i = 1,2,...,n$, se construyen los intervalos de confianza, y se actualiza la proporción de intervalos de confianza que cubrirán el valor real del parámetro. Los estudiantes se dan cuenta que la proporción tiende al valor deseado.

El intervalo de predicción también es enseñado usando simulación. Los estudiantes observarán que el valor medio de la proporción de las futuras observaciones que caen dentro del intervalo de predicción tiende al valor deseado. Durante el proceso de simulación el estudiante puede observar la aleatoriedad de los centros, anchos de los intervalos de predicción y el cambio gradual de la probabilidad hacia el valor objetivo de 0.95.

Significado de los intervalos de confianza

Terán (2006) investiga el significado de los intervalos de confianza para estudiantes del primer año en un curso de estadística en la Universidad Nacional de Rosario. Basándose en el marco teórico del "Modelo de Cognición Matemática" de Godino (2003), su propósito es detectar como se presentan los diferentes elementos de significado que surgen de los diálogos de dos estudiantes en su interacción con el profesor y el ordenador.

En una sesión de laboratorio de 2 horas, relacionado con el tópico de intervalos de confianza, Terán (2006) propone a los alumnos resolver la práctica propuesta en el libro de "Estadística Aplicada a la Administración y Economía" de L. Kazmier, utilizando software de elaboración propia, basado en hojas de trabajo Excel. La autora analiza el diálogo entre los dos estudiantes frente al ordenador, mientras ellos resuelven el problema.

Las nociones de objeto extensivo, intensivo; ostensivo, no ostensivo son usadas por Terán (2006), cuando los dos alumnos interpretan el problema, introducen datos y obtienen el primer resultado y además los elementos intensivos y de validación cuando los alumnos obtienen conclusiones a partir de los resultados. Terán (2006) considera que la posibilidad de interacción con la computadora incrementa la relevancia del aprendizaje, pues los estudiantes exploran y experimentan diferentes situaciones de acción, formulación y validación. Los estudiantes justifican la praxis explorando, empezando desde el problema mismo, con variaciones de los elementos que participan en la construcción de los intervalos de confianza.

2.4.4. COMPRENSIÓN DE LAS DISTRIBUCIONES MUESTRALES.

Puesto que las distribuciones muestrales son objetos de aprendizaje que deben ser considerados en la construcción de los intervalos de confianza, su comprensión es una componente esencial para nuestro estudio.

Comprensión del efecto del tamaño de muestra sobre la variabilidad de la distribución muestral

Well, Pollatsek y Boyce (1990) investigaron la comprensión de las distribuciones muestrales para la media, con estudiantes de psicología sin instrucción, para identificar las razones que conducen a utilizar la heurística de la representatividad. En el primer experimento de una serie de cuatro experimentos, Well, Pollatsek y Boyce comparan el rendimiento de un mismo grupo de sujetos en relación con dos versiones de enunciados de problemas, cada uno de las cuales presentaba dos variantes: uno relacionado con el grado de exactitud con el que la población es representada por la distribución de las medias muestrales, y la otra, en la que se debía de realizar una estimación del porcentaje de que un determinado valor promedio sea mayor que $\overline{X}_o$ ($\overline{X}_o$ es un valor determinado para cada problema), en esta situación se usa la cola superior de la distribución muestral.

En el segundo experimento los estudiantes son expuestos a un mismo enunciado con tres versiones relacionadas con la estimación de porcentajes en una sola cola de la distribución muestral, en las dos colas y, en el intervalo central. Además de seleccionar la respuesta correcta, se pidió a los sujetos que justificaran su elección. El diseño y la construcción de estos problemas estuvieron orientados con base en las respuestas dadas por los alumnos en el primer experimento. Los resultados obtenidos para la estimación de porcentajes en el centro de la distribución fueron mejor que para las otras dos versiones, lo cuál condujo a los investigadores a intentar obtener más información al respecto. En consecuencia, quitaron la versión relacionada con la exactitud y agregaron una versión relacionada con la estimación de porcentajes en ambas colas de la distribución (tercer experimento).

En el último experimento se cambió la estrategia, dándoles a los sujetos una breve instrucción sobre el concepto de distribuciones muestrales y pidiéndoles luego que explicaran sus razonamientos. Se pasó un pretest, en el cual el sujeto recibía una hoja con el enunciado de uno de los problemas. Una vez que el sujeto lo resolvía, el entrevistador hacía una breve explicación en relación con la diferencia entre la distribución muestral y poblacional, y por último, el sujeto podía experimentar con el ordenador para luego expresar sus conclusiones.

Los sujetos que participaron en el estudio fueron 114 estudiantes en el primer experimento, 151 estudiantes en el segundo experimento, 120 estudiantes en el tercero y

27 sujetos en el último. La conclusiones fueron que los sujetos utilizan la información del tamaño muestral en forma más apropiada cuando se les pregunta sobre la exactitud de las medias muestrales o sobre la región central de la distribución muestral que cuando se les pide información sobre las colas de las distribuciones muestrales. Esto sugiere que la variable más importante que influye en el éxito de la tarea es la similitud entre la media muestral y poblacional. En general, los sujetos parecen comprender que los promedios de muestras más grandes se acercan más a la media de la población, pero no comprenden las implicaciones de esto sobre la variabilidad de la media muestral. Esto sigue aconteciendo aún después de haberlos instruido al respecto y aún después que los mismos sujetos han podido experimentar realizando simulaciones con el ordenador.

Otras investigaciones que han diseñado cursos para estudiar el aprendizaje activo de las distribuciones muestrales, basado en el uso del ordenador concluyen en general: 1) Los avances en el aprendizaje son modestos, no hay diferencias significativas en el aprendizaje de los estudiantes que han usado diferentes softwares, 2) La simulación en general, mejora la comprensión conceptual de los estudiantes, aunque también introduce la posibilidad de generar nuevas fuentes de concepciones erróneas (Hodgson, 1996), 3) El uso del ordenador agrega nueva información en el aprendizaje, y esto en consecuencia agrega un costo al estudiante.

El rol de las distribuciones muestrales en el contraste de hipótesis.

El análisis de los errores en la comprensión del contraste estadístico de hipótesis ha sido abordado por varios investigadores y se han centrado típicamente en las dificultades sobre la interpretación del nivel de significación. Estos trabajos, que se describen en Vallecillos (1994, 1996, 1999) y Batanero (2000), han dado explicación de dichos errores argumentando con razones del tipo psicológico.

Vallecillos (1994) realiza un estudio epistemológico y matemático del contraste estadístico de hipótesis, relacionándolo con la problemática de la inducción en las ciencias empíricas y analizando hasta que punto la inferencia estadística proporciona una solución a esta problemática. De este análisis teórico determina una serie de elementos de significado sobre el contraste estadístico, mostrando la complejidad conceptual del tema y la multitud de conceptos que el alumno debe integrar para lograr una comprensión completa del mismo. Uno de estos elementos es el de distribución muestral del estadístico.

Para realizar el estudio construye un cuestionario de evaluación del significado personal atribuido por estudiantes del curso introductorio de estadística, al finalizar el mismo. Después de varias muestra piloto se pasó el cuestionario a un grupo de 436 estudiantes universitarios de la Universidad de Granada en diversas especialidades. El cuestionario contiene ítems de opciones múltiples, ítems en que el alumno debe justificar su respuesta, y problemas abiertos. El análisis cualitativo y cuantitativo de las respuestas, incluyendo métodos multivariantes, como análisis factorial y análisis cluster, permitió determinar diferentes componentes en el significado que los alumnos atribuyen al contraste de hipótesis.

El trabajo se complementa con entrevistas clínicas a un grupo reducido de alumnos. Como consecuencia, se describen concepciones erróneas respecto al nivel de significación, hipótesis y la lógica global del proceso (Vallecillos, 1995). También se estudian las relaciones entre el conocimiento conceptual y conocimiento procedimental de los estudiantes. Los errores detectados aparecen en todas las muestras de estudiantes, incluidos alumnos que cursan la licenciatura de matemáticas.

Uno de los factores determinados por Vallecillos se refiere a las relaciones entre el parámetro y el estadístico de contraste. Sobre este factor se observa la falta de comprensión de la variabilidad del estadístico en el muestreo y la confusión entre estadístico y parámetro. Así mismo se observa una falta de comprensión del papel que juega la distribución del estadístico en la determinación del nivel de significación y las regiones crítica y de aceptación.

En conclusión, se observa en este grupo de investigaciones las siguientes dificultades de comprensión en las distribuciones muestrales que van a incidir consecuentemente con el tema de los intervalos de confianza, y que deberán ser consideradas a lo largo del trabajo:

1. Dificultad de comprensión en el efecto del tamaño de la muestra sobre las distribuciones asintóticas en el muestreo.
2. Confusión entre estadístico y parámetro, lo que nos hace suponer que los alumnos tienen serias dificultades en distinguir entre los datos empíricos y el modelo que se usa para describirlos.

2.4.5. COMPRENSIÓN DEL TEOREMA CENTRAL DEL LÍMITE

El teorema central del límite es uno de los teoremas estadísticos más importantes que influye, a través de las distribuciones muestrales, en el adecuado aprendizaje de los intervalos de confianza. En relación a su comprensión destacamos los trabajos de Méndez (1991), Alvarado y Batanero (2006, 2007), Alvarado (2007).

Méndez (1991) tiene como propósito, extraer datos fenomenológicos para representar las creencias que los sujetos tienen sobre los aspectos fundamentales del teorema central del límite y luego clasificar los errores más comunes. Para ello realiza un análisis de 10 libros de estadística básica e identifica cuatro propiedades básicas en la comprensión firme del teorema:

1. La media de la distribución muestral es igual a la media de la población, e igual a la media de una muestra cuando el tamaño de la muestra tiende a infinito.
2. La varianza de la distribución muestral es menor que la de la población (cuando $n>1$).
3. La forma de la distribución muestral tiende a ser acampanada a medida que se incrementa el tamaño muestral, y aproximadamente normal, independientemente de la forma de la distribución en la población.
4. La forma de la distribución muestral crece en altura y decrece en dispersión a medida que el tamaño muestral crece.

Estas cuatro propiedades son utilizadas como base de un modelo mental de expertos sobre el teorema central del límite con dos niveles que resultan del análisis realizado a dos grupos (expertos y noveles):

- Nivel definido por las habilidades y conocimientos requeridos para resolver los ejercicios que aparecen en los libros de texto;
- Nivel que representa aspectos de conocimientos que usualmente no están en los libros de texto

En la segunda fase del estudio se realizaron entrevistas clínicas a un grupo de alumnos que resolvieron los tests. Se tomaron tres grupos de sujetos según si eran principiantes, (dividiendo lo que tenían estudios previos de estadística y los que no), o si eran expertos (estudiantes de doctorado en estadística y economía). El objetivo general era observar como los sujetos comprenden o no el teorema central del límite, y en particular observar la comprensión de las cuatro propiedades descritas anteriormente.

Méndez (1991) encontró: a) los estudiantes que estaban realizando estudios de doctorado mostraron relativamente, una buena comprensión del teorema central del límite. Carecían de capacidad de expresión intuitiva y además sus explicaciones eran formales, pero no mostraban una comprensión profunda, b) ninguno de los grupos presentó demasiado conocimiento del uso del vocabulario técnico, y además, los estudiantes de doctorado presentaban una falta de conocimiento en la utilización de expresiones cualitativas, c) no se encontraron grandes diferencias en los otros dos grupos de estudiantes, quienes en su mayoría, usaban los datos disponibles sin tener en cuenta la población de la que provenían y sin considerar el tamaño de muestra.

Considerando estos resultados Méndez (1991) recomienda: a) en un curso introductorio de estadística se tenga en cuenta la naturaleza de los materiales de aprendizaje, los conceptos y los procedimientos que se quieran enseñar, b) la utilización de datos para ayudar a los alumnos a observar los aspectos principales del teorema central del límite, así como la utilización de material manipulable, como por ejemplo, simulación por medio del lanzamiento de dados, así como el ordenador para realizar simulaciones en las que el tamaño de muestra sea muy grande, c) crear conciencia en los profesores de los diversos niveles de comprensión que pueden adquirir sus alumnos y d) propiciar la comprensión intuitiva antes de conducir a los alumnos a un pensamiento más formal.

Otro trabajo que destaca como representativo de la comprensión del teorema central del límite es el realizado por Alvarado (2004) quien siguiendo la metodología y filosofía de análisis de otros trabajos desarrollados en nuestro grupo de investigación, estudia los elementos del significado del teorema central del límite en textos de estadística para ingenieros .

Trabajo de Alvarado

La parte central de este trabajo consistió en el análisis histórico de la evolución del teorema, para determinar el *significado global del TCL* y el análisis específico de los distintos elementos de significados presentes en una muestra de 16 textos, que son los más consultados por los estudiantes de ingeniería y que no requieren un nivel avanzado de conocimiento en matemáticas, para determinar el *significado de referencia*.

En este estudio de Alvarado (2004), en que se realiza un análisis conceptual-epistemológico del significado del teorema central del límite, los principales resultados obtenidos son:

- Análisis de la importancia del teorema central del límite en estadística y la problemática didáctica que implica su enseñanza en el salón de clase
- Seguimiento de la evolución histórica del teorema, identificando nueve campos de problemas, que han dado origen al desarrollo teórico del teorema.
- Selección de investigaciones relacionadas con la enseñanza y aprendizaje del teorema central del límite.

Respecto al significado del teorema central del límite en los libros universitarios los resultados que arroja el estudio son:

- En relación a los campos de problemas en donde surge el teorema, se encontraron situaciones problemas nuevos (en relación al estudio histórico) mientras en general, los textos carecen de la mayoría de los campos históricos del teorema central del límite, teniendo prioridad los campos de problemas indirectos. Uno de los campos de problemas es precisamente la construcción de intervalos de confianza.
- Se observa también variedad de ejercicios predominantemente del área de ingeniería industrial y ausencia de problemas con dispositivos de simulación manipulable y debilidad de rigor matemático en los ejercicios resueltos.

Sobre el lenguaje y procedimientos en que se trata el teorema en los libros de texto universitarios:

- La forma usual de expresión es simbólica, descuidando las representaciones gráficas y simulaciones, propio de la disciplina ingenieril.
- En la resolución de problemas, el cálculo algebraico es el más común y por otro lado los textos son cada vez más técnicos en ausencia de la fundamentación matemática en las propiedades.
- Son prácticamente inexistentes los problemas reales con salida de software estadístico.

Alvarado (2004) señala que la determinación del significado del teorema central del límite, permitirá mejorar el proceso de enseñanza en la comprensión del teorema, así como también sienta las bases para la construcción de los instrumentos de evaluación y la interpretación de las respuestas de los alumnos de ingeniería, identificando potenciales conflictos semióticos.

Recientemente, como resultado del análisis del significado institucional de referencia del teorema central del límite, Alvarado (2007) ha hecho todo un experimento de enseñanza seleccionando los elementos de significados más adecuados del teorema para la construcción de su propuesta didáctica, organizándolos en un proceso de estudio pretendido, que ponen en contexto los diferentes elementos de significado en ejemplos vinculados con la ingeniería. Su experiencia de enseñanza se llevará a cabo en el curso de estadística para ingenieros con alumnos de segundo y tercer año, incorporando recursos computacionales a la acción didáctica con el uso de la hoja Excel y el programa estadístico @risk para simular distribuciones de probabilidad en el estudio de algunas situaciones de riesgo en la ingeniería. El curso de estadística para ingenieros tiene como prerrequisito la materia de Probabilidades (que se cursa a inicios del segundo año), por tanto los alumnos potenciales tienen conocimientos de estadística descriptiva, cálculo de probabilidades, variables aleatorias, distribuciones de probabilidad clásicas, distribuciones bivariadas y de Excel.

Un producto de este experimento de enseñanza será la comparación de los elementos del significado institucional pretendidos con los significados implementados y adquiridos por los estudiantes, estableciendo el grado de correspondencia de apropiación de los elementos relevantes del teorema central de límite de acuerdo a los objetivos propuestos.

Otras investigaciones

Entre otros autores que han publicado al respecto encontramos a Sedlmeier (1999), quién diseñó un software basado en el modelo de la urna flexible para entrenar a los estudiantes en problemas de distribuciones muestrales. Sedlmeier encontró efectos significativos en el mediano y largo plazo de este entrenamiento, lo que no queda claro en estos estudios es si los estudiantes desarrollaron una comprensión abstracta de las distribuciones muestrales a través de esas actividades o aún dependen de la traducción al modelo de la urna.

2.4.6. COMPRENSIÓN DE LA DISTRIBUCIÓN NORMAL

Según Huck y cols. (1986) una de las concepciones erróneas ampliamente extendidas entre los estudiantes, son las referentes al rango de variación de las puntuaciones Z, cuando se calculan a partir de una muestra finita o de una distribución uniforme. En contradicción de la opinión de Huck, los trabajos de Hawkins, Jolliffe y Glickman (1992), avisan que la mayoría de los estudiantes no tienen dificultad en comprender este concepto ni en calcular las puntuaciones Z para un conjunto de datos particular. Para ellos, los errores más comunes en un curso introductorio de estadística provienen del rol dominante que tiene el teorema central del límite y, del hecho de que la estimación y el contraste de hipótesis están con frecuencia basados en la hipótesis de que se trabaja con una distribución normal.

Los mismos autores Hawkins, Jolliffe y Glickman (1992), señalan que se producen errores cuando los estudiantes usan la distribución normal como una aproximación de la distribución binomial. Los errores provienen del hecho de que, muchas veces, los estudiantes no ven la diferencia entre lo discreto y lo continuo, y en muchos casos aplican la corrección por continuidad de una forma mecánica pero sin entender el significado de dicha corrección y sin saber en qué circunstancias debería ser aplicada. Resaltan también que se debería poner sobre aviso a los estudiantes cuando utilizan paquetes de software estadístico, ya que éstos generalmente utilizan la corrección por continuidad y los alumnos no siempre prestan atención a estos detalles o no conocen las razones por las cuales se aplica dicha corrección.

El estudio más relacionado con el nuestro, ya que usa el mismo marco teórico es el de Tauber (2001), quien realiza las siguientes contribuciones:

- Describe el significado institucional del tema en los libros de textos universitarios, destacando el hecho de que la integración del ordenador al salón de clase impulsa la necesidad de plantear una revisión de textos (Batanero, Tauber y Meyer, 1999).
- Diseña una secuencia de enseñanza basada en el uso de ordenadores, sin confinarse a seleccionar los elementos del significado institucional de referencia, sino planteando también problemas más realistas, en donde se incorporan los recursos gráficos y de cálculo que permite el ordenador y se intercalan las actividades tradicionales, tales como la resolución de problemas con papel y lápiz y exposiciones del profesor (Batanero, Tauber y Sánchez, 2001).
- Analiza la diferencia entre la secuencia diseñada y el significado institucional de referencia. También analiza la secuencia de enseñanza, tal como fue observada, proporcionando con ello información de primera mano, que es valiosa para los profesores, pues les permite anticiparse a las dificultades y posibilidades que genera

la decisión de incorporar el ordenador al salón de clase, así como proporcionar información de la factibilidad del trabajo con este recurso tecnológico y a la vez didáctico.

- Describe dificultades y errores en un tema que es prácticamente terreno virgen. Además describe las diferencias en la evaluación de dos tipos de alumnos: con o sin conocimientos previos en estadística. Lo cual es sumamente importante en una formulación realista de objetivos de aprendizaje.

2.4.7. CONCLUSIONES DEL ESTADO DE LA CUESTIÓN

Las conclusiones de todos estos estudios es que los estudiantes aprenden mejor cuando las actividades son estructuradas para ayudarlos a evaluar las diferencias entre sus propias creencias acerca de los fenómenos aleatorios y los resultados empíricos. (delMas, Garfield y Chance, 1998). Las principales dificultades señaladas respecto a los intervalos de confianza en investigadores y estudiantes, así como los distintos instrumentos utilizados para explorarlas son las siguientes:

Cumming, William y Fidler (2004) encontraron concepciones equivocadas de investigadores acerca de donde era más probable que cayera la media si el experimento se replicaba un número de veces usando diagramas de una media con sus intervalos de confianza. Belia, Fidler y Cumming (2005) encontraron concepciones equivocadas de investigadores acerca de comparación de dos medias, presentados los intervalos en forma gráfica. Ambos estudios utilizaron tests aplicados por internet.

Utilizando tests presenciales Fidler y Cumming (2005) encuentran que muchos estudiantes consideraban a los intervalos de confianza como estadísticos descriptivos, ignorando su naturaleza inferencial.

Behar (2001) aplicando tests presenciales a expertos en estadística y estudiantes detecta una debilidad conceptual en la comprensión de la manera como se relacionan los distintos factores asociados con un intervalo de confianza, en especial el ancho del intervalo. Además encuentra que la utilidad de los intervalos de confianza para tomar decisiones sobre hipótesis, parece no estar comprendida por una buena porción de los participantes.

También se han descrito errores en conceptos relacionados con los intervalos, como por ejemplo: Well, Pollatsek y Boyce (1990) utilizando pre y post-tests, con el propósito de realizar comparaciones antes y después del desarrollo de la experiencia de aspectos relacionados con las distribuciones muestrales, encuentra que los estudiantes, en general, parecen comprender que los promedios de muestras más grandes se acercan más a la media de la población, pero no comprenden las implicaciones de esto sobre la variabilidad de la media muestral.

delMas, Garfield y Chance (1998, 1999, 2004) hacen análisis de pruebas de simulación de intervalos de confianza, distribuciones muestrales y test de hipótesis realizadas con apoyo del ordenador pero no describen específicamente errores en los estudiantes. Sin embargo señalan que la presentación de información en forma simple, sin significados ocultos, no conduce necesariamente a la comprensión sólida de las distribuciones muestrales.

Utilizando cuestionarios y además entrevistas clínicas Vallecillos (1994 1995 encuentra concepciones erróneas en estudiantes universitarios respecto al nivel de significación, hipótesis y la lógica global del proceso. Además falta de comprensión de la variabilidad del estadístico en el muestreo y la confusión entre estadístico y parámetro.

Realizando entrevistas clínicas y tests a estudiantes principiantes y con estudios de doctorado para evaluar la comprensión del teorema central del límite, Méndez (1991) encontró: a) los estudiantes que estaban realizando estudios de doctorado carecían de capacidad de expresión intuitiva y además sus explicaciones eran formales, pero no mostraban una comprensión profunda, b) ninguno de los grupos presentó demasiado conocimiento del uso del vocabulario técnico y c) ambos grupos usaban los datos disponibles sin tener en cuenta la población de la que provenían y sin considerar el tamaño de muestra.

Utilizando cuestionarios para evaluar la comprensión de elementos de significado de la distribución normal en estudiantes, Tauber (2001) encontró: a) dificultades para distinguir entre distribución teórica y empírica, b) no identifican la frecuencia en un histograma con la altura de las barras, c) no identifican el modelo normal con una ecuación, d) dificultades para realizar argumentaciones para justificar una afirmación, e) errores para decidir si las variables pueden aproximarse o no por medio de una distribución normal y f) dificultades para comprender la diferencia entre variable cualitativa y cuantitativa, discreta y continua. En tanto que utilizando pruebas de ensayo para ser resuelta con ayuda del ordenador, Tauber (2001) encontró que los estudiantes tuvieron problemas de interpretación del tipo de asimetría y determinación de los valores exactos de los valores atípicos.

El uso del ordenador permite rediseñar y con ello redimensionar el aprendizaje de los intervalos de confianza, pues le permite al alumno explorar en forma más activa los efectos que se producen cuando se cambian los elementos que intervienen en la construcción de los intervalos de confianza. Por el hecho de que el uso del ordenador, es por si mismo fuente de nuevos errores en el aprendizaje de los conceptos estadísticos, incluidos en ellos los intervalos de confianza, será necesario asegurarse de que los estudiantes tengan un mínimo de entrenamiento para utilizar el ordenador, y además que entienden las características de las actividades de simulación.

No obstante, de las facilidades que brindan los programas de simulación de proporcionar excelentes visualizaciones de los procesos abstractos de creación de las distribuciones muestrales, algunos estudiantes continúan con serias dificultades en la comprensión de los intervalos de confianza y del teorema central del límite, debido en parte porque algunos estudiantes no alcanzan a integrar conceptos tales como: variabilidad, distribución de probabilidad, distribución normal y muestreo, a pesar de que se intentan algunas medidas remediales.

Los estudios previos sobre los intervalos de confianza y las investigaciones relacionadas, en donde se resaltan algunos errores y se determinan diferentes componentes en el significado que los alumnos atribuyen al concepto estadístico en cuestión, serán el hilo conductor del desarrollo de nuestro futuro trabajo de investigación. Trabajo en donde nos enfocaremos, con la ayuda del análisis de textos, en el uso cuantitativo y cualitativo de los intervalos de confianza, en el dominio del lenguaje técnico, en la determinación del nivel de comprensión formal y de uso correcto de los intervalos de confianza.

CAPÍTULO 3
ANALISIS DE CONTENIDO DE LIBROS DE TEXTO

3.1. INTRODUCCIÓN

En la sección 1.3, hemos hecho una primera aproximación a la noción de intervalo de confianza centrando la atención en los aspectos conceptuales y procedimentales implicados. Desde el punto de vista del "enfoque ontosemiótico" (EOS) descrito en la sección 2.3, el análisis realizado en la sección 1.3 es todavía incompleto, ya que no tiene en cuenta las cuestiones del origen fenomenológico (las situaciones – problemas que motivan la creación de los objetos matemáticos), ni el papel de los diversos lenguajes (expresiones verbales, notaciones, etc.) mediante los cuales se describen, representan y comunican los objetos matemáticos. Según estos planteamientos, los significados de los intervalos de confianza son diversos según el contexto institucional en que se usan. Por ejemplo, será diferente en la comunidad de estadísticos profesionales y en la comunidad de ingenieros que los usan como parte de sus herramientas profesionales. Así mismo, habrá potencialmente diferencias en los significados desde el punto de vista personal (dimensión cognitiva de los significados).

En este capítulo analizaremos los significados de los intervalos de confianza en una muestra de 11 libros de texto de estadística usados en las carreras de ingeniería en los campus del Instituto Tecnológico de Monterrey (México). Con este estudio empírico mostraremos en qué aspectos se diferencian y coinciden los significados implementados en los diversos libros, lo que nos permitirá posteriormente establecer un "significado de referencia" de los intervalos de confianza en el marco institucional fijado. Este significado de referencia lo usaremos para elaborar un instrumento de evaluación de los significados personales de los estudiantes que reciben sus enseñanzas en el Tecnológico de Monterrey.

La noción de significado (sistémico – pragmático) del EOS se hace operativa mediante la categorización descrita en la sección 2.3 de los objetos emergentes de los sistemas de prácticas y su articulación en configuraciones (epistémicas, cuando se trata de significados institucionales; o cognitivas, si se trata de significados personales).

En consecuencia, el estudio empírico se centra en la identificación y clasificación de los tipos de objetos puestos en juego en los diversos textos. Una metodología similar ha sido aplicada en otras investigaciones realizadas en la línea de investigación sobre Educación Estadística de la Universidad de Granada (Ortiz, 1999; Sánchez- Cobo, 1999; Tauber, 2001; Cobo, 2003; y Alvarado, 2004, 2007).

En lo que sigue, describimos los objetivos y la metodología utilizada en el análisis. A continuación, se analizan los elementos de significado del intervalo de confianza, mostrando ejemplos que los aclaren. Por último, organizaremos la información en tablas para resumir los elementos de significado en los distintos textos analizados. Los resultados serán la base de la definición semántica de la variable objeto de evaluación para la construcción de nuestro cuestionario.

3.2. OBJETIVOS ESPECÍFICOS DEL ANÁLISIS

El objetivo 3 de nuestra investigación (capítulo 1) era realizar un análisis de los elementos del significado de los intervalos de confianza que nos fundamente la definición semántica de la variable “comprensión de los intervalos de confianza” y delimitar las principales áreas de contenido.

En nuestro caso, hemos fijado una institución formada por los *cursos de probabilidad y estadística para ingenieros.* Dentro de esta institución el significado de los intervalos de confianza viene definido por los *elementos situacionales del significado* que incluye los campos de problemas que abordaremos, las herramientas disponibles en su solución, que son los *elementos lingüísticos, conceptuales, proposicionales, procedimentales y argumentativos*.

La importancia de este objetivo se deduce del hecho de que la construcción del instrumento de evaluación ha de estar fundamentada en el análisis del significado del concepto en Matemáticas, que nos sirva de patrón de comparación en el estudio. El estudio de la evolución histórica nos puede proporcionar pistas de las dificultades en el desarrollo del concepto, que también pueden aparecer en los alumnos. El estudio de los textos, nos da criterio para su mejora, nos proporciona ejemplos de situaciones que podemos usar en el diseño de nuestro cuestionario y nos avisa de posibles sesgos de significado.

Objetivos específicos

Este objetivo general se concretará en otros más específicos a la luz del marco teórico adoptado, los cuales describiremos a continuación**,** en la misma línea del trabajo de Alvarado (2004) sobre el Teorema Central del Límite y son los siguientes:

O1. *Descripción del significado que se presenta de intervalo de confianza* en la institución: "Enseñanza de estadística a ingenieros en el Instituto Tecnológico y de Estudios Superiores de Monterrey". Dicho significado institucional, lo vamos a analizar a partir del conjunto de libros de texto que son utilizados en mi universidad, el Tecnológico de Monterrey.

Compararemos el significado presentado en los libros con el análisis conceptual e histórico, previamente realizado. Nuestra premisa es que el libro de texto refleja el significado que el autor considera adecuado para la institución de enseñanza particular a que el texto va dirigido. Si el autor tiene un gran prestigio en la estadística este libro es representativo de lo que se considera adecuado por los miembros de dicha institución.

O2. *Obtención de conclusiones preliminares* que nos permitan en una segunda fase seleccionar los elementos de significados más adecuados para la construcción de lo que sería nuestro instrumento de evaluación centrado en unos campos de problemas y contextos específicos, así como un conjunto de instrumentos idóneos para abordarlos.

3.3. MUESTRA DE TEXTOS DE LIBROS UNIVERSITARIOS SELECCIONADOS

Para delimitar la población posible de textos entre los cuales seleccionar la muestra, decidimos recurrir a los libros de texto recomendados en las asignaturas de estadística para ingenieros en el sistema del ITESM. Para elegir los libros que vamos a analizar, hemos realizado una encuesta a los profesores que imparten la asignatura de Probabilidad y Estadística para ingenieros en 29 campus del Sistema Tecnológico de Monterrey que ofrecen carreras profesionales, de un total de 33 campus.

En dicha encuesta les hemos pedido la lista de textos que recomiendan a sus alumnos y que usan para la preparación de la asignatura. De los 29 campus encuestados vía correos electrónicos, recibimos respuesta de 16. Otra fuente que utilizamos para obtener información fue a través de las oficinas de la Vicerrectoría Académica del ITESM, quienes nos proporcionaron una lista de los 6 libros más utilizados en el sistema. En esta muestra de libros, como se puede observar aparecen textos específicos de estadística para ingenieros, uno de estadística elemental y no aparece ninguno de estadística matemática. La lista de textos recomendados, que son los que analizaremos, se incluye a continuación. La muestra incluye un periodo de 13 años.

Tabla 3.1. Muestra de libros usada en el estudio

	Título	Autores	Editorial	Edición
A	Probabilidad y estadística.	Walpole, R.E. y Myers R.H.	Mc Graw Hill (México)	1992
B	Probabilidad y estadística para ingeniería	Scheaffer, R. y McClave, J.	Editorial Iberoamericana	1993
C	Probability and Statistics for Engineering and Scientists	Hayter, A.	International Thomson	1996
D	Probabilidad y estadística para ingenieros	Miller, I., Freund, J. y Johnson, R.	Prentice Hall (México)	1997
E	Probabilidad y estadística para ingeniería y ciencias	Mendenhall, W. y Sincich, T.	Prentice Hall	1997
F	Probabilidad y estadística para ingenieros.	Walpole, R., Myers, R. y Myers, S	Prentice Hall	1999
G	Probabilidad y estadística para ingeniería y ciencias	Velasco, G., y Wisniewski, P.	Thomson Learning	2001
H	Estadística elemental	Johnson, R. y Kuby, P.	Thomson	2004
I	Probabilidad y estadística (con aplicaciones para ingeniería y ciencias computacionales)	Milton, J. y Arnold, J,	McGraw Hill Interamer	2004
J	Probabilidad y estadística aplicadas a la ingeniería	Montgomery, D. y Runger G.	McGraw Hill	2004
K	Probabilidad y estadística para ingeniería y ciencias.	Devore, J.	Thomson	2005

3.4. METODOLOGÍA DE ANÁLISIS

El análisis de contenido es un tipo de medición que se basa en la idea de que las unidades del texto pueden clasificarse en un número reducido de categorías (Weber, 1985). Sirve para efectuar inferencias mediante la identificación sistemática y objetiva de las características específicas del texto (Ghiglione y Matalón, 1989). En este estudio, se han seguido las técnicas lógico-semánticas del llamado análisis de contenido temático, que son las más frecuentes y típicas. En ella se recurren a la lógica, y al conocimiento previo sobre el tema adquirido a través de la revisión bibliográfica y el estudio histórico previo, para resumir el contenido del texto, definir categorías y verificar la validez de las mismas. Sólo considera la presencia de términos o conceptos, independientemente de las relaciones entre ellos (Ghiglione y Matalón, 1989).

El marco teórico del “enfoque ontosemiótico” nos proporciona una categorización de entidades matemáticas que aportan orientación y concreción al análisis de contenido. El objetivo es la caracterización de los significados del intervalo de confianza en los diversos textos. Siendo un análisis de datos cualitativo, el proceso se realizó en varias etapas dividiendo el texto en unidades, que fueron clasificadas en categorías y a continuación ensambladas y relacionadas para conseguir un todo coherente (Goetz y Lecompte, 1988). El proceso se dividió en tres etapas: la reducción de datos, el análisis de datos y la obtención y verificación de conclusiones (Huberman y Miles, 1994):

- La primera operación fue la separación de segmentos o unidades de análisis, en varios niveles. Siguiendo este criterio, una vez seleccionados los libros, una primera lectura de los mismos permitió seleccionar los capítulos que contuviesen temas relacionados con el intervalo de confianza.
- Para cada uno de estos capítulos, se identificaron *unidades secundarias* de análisis que fueron los párrafos que hacía referencia implícita o explícita a algunos de los elementos de significado considerados en el marco teórico, examinando los datos para clasificarlos en una u otra categoría de contenido (Gil Flores, 1994).
- Se compararon los diversos párrafos clasificados en un mismo apartado para cada categoría, depurándose la clasificación. La categorización permitió clasificar conceptualmente las unidades que correspondían a un mismo tópico.
- Se elaboraron tablas recogiendo los elementos de significado (componentes de las configuraciones epistémicas mediante las que se hacen operativos los significados) incluidos en los diferentes textos. Esta transformación y ordenación de los datos permitió presentarlos en modo abarcable y operativo para la extracción de conclusiones.
- Se elaboró el informe de análisis incluyendo ejemplos de las diferentes categorías, elaboradas mediante un proceso interpretativo guiado por el análisis epistemológico de los conceptos intervinientes y por los resultados de las investigaciones previas.

A continuación se presenta el análisis y sus resultados.

3.5. CAMPOS DE PROBLEMAS

En primer lugar analizaremos los distintos campos de problemas cuya resolución hace surgir la idea de intervalos de confianza, y que en su conjunto definen el campo de problemas relativo al concepto. También tendremos en cuenta el estudio conceptual e histórico que hemos llevado a cabo anteriormente. Estos campos de problemas llevan a la emergencia de los objetos matemáticos, como entidades culturales socialmente compartidas; ya que los problemas matemáticos y sus soluciones son compartidos en el seno de instituciones o colectivos específicos implicados en el estudio de ciertas clases de problemas. Los campos de problemas encontrados en los textos analizados se describen a continuación (Olivo, Ruiz y Albert, 2006).

CP1: *Estimar un parámetro desconocido para una población*

Con frecuencia se desea estudiar un fenómeno aleatorio que viene caracterizado por una distribución de probabilidad, que depende de uno o varios parámetros, por ejemplo, la media. Pero no es posible recolectar los datos de toda la población, sino que hemos de contentarnos con una muestra aleatoria de la misma en donde se calcula un estadístico que sirve para estimar el parámetro. Una estimación por intervalos tiene la ventaja de dar un margen de error.

En este campo, encontrado en el estudio histórico del tema, encontramos muchos subcampos diferentes, según el parámetro a estimar (media, proporción, varianza, etc.). En los libros hemos encontrado algunos de ellos que a su vez pueden diferenciarse según las condiciones dadas. No solo da origen al intervalo de confianza, sino que aparecen una diversidad de objetos matemáticos, especialmente las distribuciones de probabilidad.

CP1.1: *Estimar la media de una población*

Debido a que la media muestral $\overline{X}$ es un estimador insesgado de μ y de mínima varianza, $\overline{X}$ se considera el mejor estimador de la media μ. En la construcción del intervalo de confianza para la media poblacional, es necesario tomar en cuenta los supuestos que le dan validez a los procedimientos. Hemos encontrado en los libros los siguientes subcampos.

CP1.1.1: *Media de una población normal o en una muestra grande con, σ conocida***.**

La media muestral $\overline{X}$ está distribuida en forma normal cuando la distribución poblacional es normal o si el tamaño de muestra es suficientemente grande. La mayoría de los autores de los libros de probabilidad y estadística que analizamos manejan como regla aproximada utilizar un mínimo de 30 observaciones para ser considerada muestra grande, mientras Devore (2005, pg. 292) considera más de 40 observaciones y Johnson y Kuby (2004, pg. 368) sugieren que las poblaciones no simétricas requieren tamaños de muestra mayores, con 50 como número suficiente. A continuación se ilustra el procedimiento de construcción de un intervalo de confianza en estos casos.

Se supone que las observaciones muestrales reales $x_1, x_2, \ldots, x_n$ son el resultado de una muestra aleatoria $X_1, X_2, \ldots, X_n$ de una distribución normal con media μ y desviación estándar σ. Entonces, sin tomar en cuenta el tamaño de muestra *n*, la media muestral $\overline{X}$ tiene una distribución normal con valor esperado μ y desviación estándar $\frac{\sigma}{\sqrt{n}}$. La estandarización de $\overline{X}$ restando primero su valor esperado y dividiendo después entre su desviación estándar, produce la variable: $z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ que tiene una distribución normal estándar. Debido a que el área bajo la curva normal estándar entre -1.96 y 1.96 es .95 $p(-1.96 < \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < 1.96) = .95$ Después de manipular las desigualdades se obtiene: $p(\overline{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96 \frac{\sigma}{\sqrt{n}}) = .95$ Devore (2005, pg. 283)

CP1.1.2: *Estimar la media de una población aproximadamente normal cuando, σ es desconocida.*

Debido a que σ es desconocida en muchas aplicaciones prácticas, el procedimiento para estimar la media de una población es similar al utilizado cuando se conoce la varianza excepto que σ se sustituye por *S, la estimación en la muestra* y la distribución normal estándar por la distribución *t* de Student. Un supuesto que debe cumplirse es que la población de la cual se obtuvo la muestra sea normal. Sin embargo, el estadístico *t* es tan robusto que algunos autores indican que en el caso de violación de esta condición de normalidad, lo único que se exige es que la distribución poblacional no sea muy asimétrica. Ilustramos con un ejemplo:

Si tenemos una muestra aleatoria de tamaño *n* a partir de una distribución normal, entonces la variable aleatoria $t = \frac{\overline{X} - \mu}{S / \sqrt{n}}$ tiene una distribución *t* de Student con *(n-1)* grados de libertad. Se puede asegurar que $P(-t_{\alpha/2} < t < t_{\alpha/2}) = 1 - \alpha$ Al sustituir por t y luego resolver para μ, obtenemos: $P(\overline{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}) = 1 - \alpha$ Walpole, Myers y Myers (1999, pg. 246)

CP1.1.3: *Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande* **(***$n > 30$***).**

Con σ desconocida para muestras grandes se procede a la aproximación de sustituir σ por la desviación estándar de la muestra s y se utiliza nuevamente la distribución normal estándar en la construcción del intervalo de confianza para la media, el Teorema Central del Límite valida estos procedimientos. Un ejemplo es el siguiente:

> Con referencia a los datos de emisión de óxidos de azufre de la página 8, para los que teníamos n=80, $\bar{x}$ = 18.85 y s^2 =30.77, elabore un intervalo de confianza de 99% para la verdadera emisión diaria promedio de óxidos de azufre de la planta. Miller, Freund y Johnson (1997, pg. 223)

CP1.1.4: *Estimar la media en una distribución exponencial.*

En el tratamiento de problemas de confiabilidad resulta de interés estimar la vida media de un componente. Una muestra de n componentes se selecciona al azar de un lote, se prueba en condiciones ambientales especificadas y se observan los tiempos a la falla de los componentes individuales. Si esperamos hasta que los últimos componentes fallen, se puede perder mucho tiempo y dinero.

Con el propósito de reducir costos una decisión que frecuentemente se toma es concluir la prueba después de un tiempo de prueba t específico. En esta decisión decimos que la prueba de vida se censura. Un segundo tipo de decisión de muestreo censurado consiste en concluir la prueba después de que falla una cantidad fija de componentes. Si la prueba se censura en un tiempo t, el tiempo de duración de la prueba es fijo, en tanto que cuando la prueba se censura después de observar un número fijo de r componentes que fallan, se tiene la certeza de que siempre se obtienen los valores de r tiempos de falla y la duración de la prueba de vida será variable e igual al tiempo t_r hasta la falla del r-ésimo componente. De estas decisiones de muestreo surgen dos apartados que a continuación analizamos:

CP1.1.4.1: *Estimar la media en una distribución exponencial, muestreo censurado respecto al número de componentes.*

Los estimadores de la vida media μ del componente son los mismos sin importar si la prueba de vida se censura o no. Cuando la decisión de muestreo es esperar hasta que fallen los *n* componentes, el estimador es el tiempo de falla medio de la muestra:

$$\hat{\mu} = \frac{\sum_{i=1}^{i=n} t_i}{n} = \bar{t}$$

Para hacer la estimación de la vida media μ del componente, el método se apoya en el hecho de que $2T_r/\mu$ es un valor de una variable aleatoria que tiene la distribución Ji cuadrada con 2*r* grados de libertad, donde T_r es la vida acumulada hasta *r* fallas. En Miller, Freund y Johnson (1997, pg. 561) se ilustra el procedimiento de obtención de un intervalo de confianza para la vida media.

Suponga que 50 unidades se someten a una prueba de ciclo de vida (sin reemplazo) y que la prueba va a ser truncada después de que *r* =10 unidades hayan fallado. Supondremos además que los tiempos de las primeras 10 fallas son 65,110, 380, 420, 505, 580, 650, 840, 910 y 950. Calcule un intervalo de confianza de .90 para μ.

CP1.1.4.2. *Estimar la media en una distribución exponencial, muestreo censurado respecto al tiempo.*

En el caso de un muestreo censurado con un tiempo de prueba fijo *T*, el estimador $\hat{\mu}$ está dado por:

$$\hat{\mu} = \frac{\sum_{i=1}^{r} t_i + (n-r)T}{r} = \frac{vida \quad total \quad observada}{r} \qquad para \qquad r \geq 1$$

La expresión del numerador corresponde a la duración total de vida registrada para los *n* componentes hasta que la prueba concluye en el tiempo *T*. En Mendenhall y Sincich (1997, pg. 1046) se ejemplifica la obtención de un intervalo de confianza para la vida media.

Suponga que el tiempo entre descomposturas de cierto tipo de motor para avión tiene una distribución de tiempo de falla exponencial. Diez motores se probaron hasta que seis de ellos se descompusieron. Los tiempos hasta descompostura fueron: 48, 35, 91, 62, 59 y 77 horas respectivamente. Establezca un intervalo de confianza de 95% para la confiabilidad del sistema a las 50 horas.

CP1.1.5. *Estimar la media de una población no normal, en una muestra pequeña (n < 30).*

Intervalos de confianza no paramétricos se construyen en situaciones donde el tamaño de la muestra es bastante pequeño (menos de 30 observaciones) y las observaciones de los datos no están normalmente distribuidos. Tal es el caso del intervalo de confianza para la media relacionado con la prueba de rangos con signo de

Wilcoxon. La hipótesis que se prueba es $H_0: \mu = \mu_0$ contra $H_a: \mu \neq \mu_0$, donde μ es la media de una distribución simétrica continua, los valores absolutos $|x_1 - \mu_0|,\ldots$ $|x_n - \mu_0|$ se ordenan de menor a mayor, y se asigna el rango 1 al valor menor y el rango n al valor mayor. Luego, a cada rango se le asigna el signo de su $x_i - \mu_0$, correspondiente, y el estadístico de prueba es la suma de los rangos con signo positivo. Para $x_1, x_2, ..., x_n$ fijas, el intervalo de rangos con signo de 100(1-α) % consistirá en las μ_0 para las que $H_0: \mu = \mu_0$ no se rechaza al nivel α. Para identificar el intervalo se expresa el estadístico de prueba S_+ de la siguiente forma: $S_+ =$ número de promedios por pares $(X_i + X_j)/2$ con $i \leq j$ que son $\geq \mu_0$.

Es decir, si se promedia cada x_j en la lista con cada x_i a su izquierda, incluso $(X_i + X_j)/2$ (que es solo x_j), y se cuenta el número de estos promedios que son $\geq \mu_0$, se obtiene S_+. En Devore (2005, pg. 685) se presenta un ejemplo de obtención de un intervalo de confianza para el valor medio de la tasa metabólica cerebral para monos Rhesus.

CP1.2. *Estimar la mediana de una población no normal, en una muestra pequeña (n < 30).*

Otro intervalo de confianza sin distribución que se obtiene es el caso del intervalo de confianza para la mediana construido a partir de la prueba del signo, que es un procedimiento de prueba no paramétrico que puede ser usado para evaluar la plausibilidad de la hipótesis nula.

$H_0: F(x_0) = p_0$ basada en un conjunto de observaciones $x_1, x_2, ... x_n$ independientes e idénticamente distribuidos de una distribución desconocida *F(x).* Esta distribución tiene la propiedad de que la probabilidad de que ocurra una observación menor o igual a x_0 es p_0.

Las siguientes observaciones son valores de la tasa metabólica cerebral para monos Rhesus: $x_1 = 4.51$, $x_2 = 4.59$, $x_3 = 4.90$, $x_4 = 4.93$, $x_5 = 6.80$, $x_6 = 5.08$, $x_7 = 5.67$. Los 28 promedios por pares son, en orden creciente,

4.51, 4.55, 4.59, 4.705 ,4.72, 4.745, 4.76, 4.795, 4.835, 4.90
4.915 ,4.93 ,4.99 ,5.005 ,5.08, 5.09, 5.13, 5.285, 5.30, 5.375
5.655, 5.67, 5.695,,5.85, 5.865, 5.94, 6.235, 6.80

Debido a la naturaleza discreta de la distribución de S_+, α=0.05 no se puede obtener de modo exacto. La región de rechazo $\{0,1,2,26,27,28\}$ tiene α=0.046, que está lo más cerca posible a 0.05, así que el nivel es aproximadamente 0.05. Por tanto, si el número de promedios por pares $\geq \mu_0$ está entre 3 y 25, inclusive, no se rechaza H_0. El intervalo de confianza de 95 % (aproximado) para μ es (4.59, 5.94).

El análisis trata con lo siguiente: si una colección es hecha de todos los valores de x_0 para los cuales la prueba de hipótesis de nivel α de la declaración de que $F(x_0) = p_0$ se acepta, entonces la colección forma un intervalo de confianza para el cuantil $F^{-1}(p_0)$ con un nivel de confianza igual a 1-α

El interés de este subcampo, para el caso concreto de la prueba del signo, es que la obtención del intervalo de confianza para la mediana de la distribución $F^{-1}(0.5)$, es generalmente de más uso por los analistas que el cálculo del valor *p* para la afirmación de que la mediana toma un valor específico. Este intervalo se puede concebir como un resumen razonable de valores para la mediana. En Hayter (1996, pg. 812) se presenta un ejemplo de obtención de un intervalo de confianza para la mediana.

La figura 15.11 presenta una salida de Minitab de un intervalo de confianza para la mediana del tiempo de ejecución. Este conjunto de datos permite al experimentador concluir que con cerca de un 95% de confianza la mediana del tiempo de ejecución está entre 8 y 13 segundos.

CP1.3. *Estimación de una proporción.*

Otras situaciones problemáticas que se ubican en este campo tienen que ver con la estimación del parámetro proporción que encuentran muchas aplicaciones en ingeniería, porque en muchos problemas interesa estudiar si un atributo determinado está presente o no entre los elementos de una población y en cada población se desconoce la proporción que tiene determinada cualidad.

En una muestra aleatoria de 85 rodamientos para el cigüeñal del motor de un automóvil, 10 tienen un acabado de la superficie con más asperezas que las que permiten las especificaciones. Por lo tanto, una estimación puntual de la proporción de rodamientos de la población que excede la especificación de aspereza $\hat{p} = \frac{x}{n} = \frac{10}{85} = 0.12$ es . Calcular el intervalo de confianza de dos colas de 95% .
Montgomery y Runger (2004, pg. 352)

CP1.4. *Estimación de una varianza.*

Además de la proporción, se presenta la estimación de la varianza poblacional que, aunque de menor interés que la media, en ocasiones resulta necesaria con el propósito de tener la fotografía completa del proceso.

Los datos siguientes acerca de la tensión disruptiva de circuitos forzados eléctricamente, se leyeron de una gráfica de probabilidad normal que apareció en el artículo “Damage of Flexible Printed Wiring Boards Associated with Lightning-Induced Voltaje Surges” (IEEE Transactions on Components, Hybrids,and Manuf.Tech., 1985, pp. 214-220). La linealidad de la gráfica respaldó la suposición de que la tensión disruptiva tiene una distribución aproximadamente normal. Sea σ^2 la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es s^2 =137324.3, n =17. Calcular el intervalo de confianza de 95% para σ.

Devore (2005, pg. 310)

CP1.5. *Estimación de ρ coeficiente de correlación poblacional.*

Frecuentemente en ingeniería surge el problema de relacionar dos o más variables para ver si están asociadas. El coeficiente de correlación ρ nos permite medir la fuerza de la intensidad de la relación entre las variables x y y en la población. La construcción de este intervalo de confianza toma como base la transformación Z de Fisher.

$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$, este estadístico es un valor de una variable aleatoria que tiene aproximadamente una distribución normal con media $\mu_z = \frac{1}{2}\ln\frac{1+\rho}{1-\rho}$ y varianza $\frac{1}{n-3}$

Una advertencia es que no hay que inferir una relación causal a partir de una alta correlación muestral y que un resultado de r=0 implica una falta de linealidad y no de asociación. Un ejemplo que ilustra la construcción del intervalo de confianza para este parámetro es el siguiente:

El coeficiente de correlación muestral entre el nitrógeno entrante y el porcentaje de nitrógeno removido fue r=.733, así se obtuvo υ=.935. Con n=20, un intervalo de 95% para μ_V es $(.935-1.96/\sqrt{17}\ ,\ .935+1.96/\sqrt{17}) = (.460,\ 1.41) = (c_1, c_2)$ .El intervalo de 95% para ρ es:

$$\left[\frac{e^{2(0.46)}-1}{e^{2(0.46)}+1},\frac{e^{2(1.41)}-1}{e^{2(1.41)}+1}\right] = (0.43{,}0.89)$$

Devore (2005, pg. 546)

CP1.6. *Estimación del valor de parámetros en un modelo de regresión lineal simple.*

En ingeniería ocurren muchos problemas cuya solución implica variables de las cuales se conoce que existen relaciones inherentes entre ellas. Disponer de un método que nos permita anticipar el comportamiento de una de las variables (y) como resultado del cambio de la otra variable (x) es uno de los beneficios de utilizar un modelo de regresión. Conviene resaltar, en términos generales, el valor exacto de y no es predecible; sin embargo si al evaluar el modelo de regresión nos encontramos que las predicciones son razonablemente acertadas, cuando te mueves en una vecindad de puntos muy cercana a los valores observados en la muestra, los acercamientos al valor exacto de *y* serán de mucha utilidad para las futuras decisiones.

La importancia de disponer de un modelo de regresión lineal simple es que nos permite investigar la relación que existe entre dos variables. Este modelo también nos permite la posibilidad de indagar los niveles de optimización de un proceso o simplemente los niveles en los cuales está bajo control un proceso. Dentro de este subcampo vamos a analizar dos apartados que corresponden a la estimación de los parámetros pendiente y la ordenada al origen y respuesta media.

CP.1.6.1 *Estimación del valor de la pendiente y la ordenada al origen.*

Obtener estimaciones de la pendiente β_1 y la intersección β_0 en la ecuación de regresión lineal simple es de suma importancia por la intensa aplicación del modelo para propósitos de predicción en ingeniería y en el campo de los negocios. La validez de las inferencias acerca de β_0 y β_1 dependerá de las distribuciones de muestreo de los estimadores, que a su vez dependen de la distribución de probabilidad del error aleatorio, ε. Tales condiciones específicas son: la media de la distribución de probabilidad de ε es 0, la varianza de la distribución de probabilidad es constante para todos los valores de la variable independiente *x*, la distribución de probabilidad de ε es normal y los errores asociados a dos observaciones distintas cualesquiera son independientes.

Aquí se continúa el análisis de los datos sobre la magnitud de la evaporación de solvente durante la pintura con pistola y la humedad relativa, mediante el cálculo de los intervalos de confianza de β_0 y β_1 .
Se necesitan las siguientes estadísticas resumen, ya calculadas:

$$s^2 = 0.79 \quad \Sigma x^2 = 76\ 308.53 \quad b_0 = 13.64$$

$$S_{xx} = 7150.05 \quad b_1 = -0.08 \quad n = 25$$

Construya un intervalo de confianza de 99% para la pendiente de la línea de regresión.

Milton y Arnold (2004, pg. 398)

CP1.6.2 *Estimación de la respuesta media.*

En este subcampo se construye un intervalo de confianza a partir de la respuesta media en un valor específico x_0. Este intervalo se construye alrededor de $E(Y|\ x_0)=$ μ_{Y/x_0}, al que se acostumbra a llamar intervalo de confianza alrededor de la recta de regresión. Cuando *x* no contribuye con información a la predicción de *y*, la respuesta media, que es la parte determinística del modelo, no cambia cuando *x* cambia. Un ejemplo se presenta a continuación:

Se construirá un intervalo de confianza de 95% alrededor de la respuesta media para los datos del ejemplo 10-1. El modelo ajustado es $\mu_{Y/x_0} = 74.20 + 14.97x_0$, y el intervalo de confianza de 95% para μ_{Y/x_0} se encuentra con la ecuación 10-33. Montgomery y Runger (2004, pg.456)

CP1.7. *Estimación del valor de parámetros en un modelo de regresión lineal múltiple.*

La importancia de la construcción de la ecuación de regresión lineal múltiple es que esta es una situación menos restrictiva, porque la respuesta que se observa en la variable dependiente tiene que ver con un fenómeno en el que interaccionan más de una variable independiente. Realizar estimaciones de los parámetros que intervienen en este modelo con una mayor cantidad de términos nos da la ventaja de que podemos usarlo en relaciones más retadoras y más complejas. En este subcampo se describen y analizan dos apartados que a continuación presentamos.

CP1.7.1. *Estimación del valor de los parámetros coeficientes de x_i en un modelo de regresión lineal múltiple.*

En la estimación de β_i, los coeficientes de x_i en la ecuación de regresión lineal múltiple:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon$$

los estimadores $\hat{\beta}_i$ que se utilizan tienen una distribución de muestreo normal, debido a que $\hat{\beta}_i$ es una función lineal de *n* variables aleatorias normalmente distribuidas, además $\hat{\beta}_i$ es un estimador insesgado de β_i y el error estándar de la distribución de muestreo $\hat{\beta}_i$ está dado por: $\sigma_{\hat{\beta}_i} = \sigma\sqrt{c_{ii}}$ (*i* = 0,1,2, …*k*) ,donde c_{ii} es un elemento del i-ésimo renglón y la i-ésima columna en la matriz:

$$(X'X)^{-1} = \begin{bmatrix} c_{00}\, c_{01} & \ldots c_{ok} \\ c_{10} \quad c_{11} & \ldots c_{1k} \\ . \quad\quad . & . \quad\quad . \\ c_{k0} c_{k1} & \ldots \quad c_{kk} \end{bmatrix}$$

y *X*, que se muestra enseguida, es una matriz cuyas columnas son los coeficientes de β_i, así por ejemplo la primera columna está formada por unos que son los coeficientes de β_0, la segunda columna los coeficientes de β_1, etc.

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots x_{1k} \\ 1 & x_{21} & x_{22} & \ldots x_{2k} \\ . & . & . & . \\ . & . & . & . \\ 1 & x_{n1} & x_{n2} & x_{nk} \end{bmatrix}$$

Un ejemplo que ilustra el uso de este tipo de inferencia práctica es el siguiente:

> A fin de predecir el rendimiento de combustible (millaje) de un automóvil con base en su peso y la temperatura en el momento de conducirlo, se desarrolló la adecuación (véase el ejemplo 12.2.3). Un intervalo de confianza de 95 % para β_0 , la intersección de este modelo, es:
>
> $$\hat{\beta}_0 \pm t_{\alpha/2} s\sqrt{c_{00}}$$
>
> Milton y Arnold (2004, pg. 470)

CP1.7.2. *Estimación de la respuesta media.*

La obtención del intervalo de confianza sobre la respuesta media $\mu_{Y/x_{10},x_{20},...,x_{k0},}$, proporciona una de las inferencias de mayor utilidad e interés que se pueden realizar respecto a la calidad de la respuesta pronosticada y_0 utilizando en su construcción el conjunto de condiciones sobre las *x*, representado por el vector:

$$x_0 = \lfloor 1, x_{10}, x_{20}, ..., x_{k0} \rfloor$$

Suponiendo que los errores son independientes y se distribuyen normalmente, entonces las β_i's son normales con las propiedades de $E(\hat{\beta}_i) = \beta_i$ y $\sigma_{\hat{\beta}_i} = \sigma\sqrt{c_{ii}}$. Entonces $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_{10} + \hat{\beta}_2 x_{20} + ... + \hat{\beta}_k x_{k0}$ también tiene distribución normal y es un estimador insesgado para la respuesta media. como se ilustra en el siguiente ejemplo:

> Con el uso de los datos del ejemplo 12.3 (los datos corresponden al porcentaje de supervivencia de cierto tipo de semen animal, después del almacenamiento, en varias combinaciones de concentraciones de tres materiales que se utilizan para aumentar su oportunidad de supervivencia) construya un intervalo de confianza de 95% para la respuesta media cuando x_1 = 3%, x_2 = 8% y x_3 =9%.
>
> Walpole, Myers y Myers (1999, pg. 421)

CP1.8. *Estimación de parámetros en un modelo de regresión polinomial.*

En muchas ocasiones ocurren fenómenos en la ingeniería que no son lineales por naturaleza, en tales situaciones para modelar la variable de respuesta se requiere la construcción de un modelo de regresión no lineal y múltiple. Esas situaciones donde se presentan relaciones no lineales generan funciones de regresión polinomiales de la forma: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + ... + \beta_K X^K + \varepsilon$ donde ε es una variable aleatoria distribuida normalmente con media cero y varianza σ^2.

Una vez construido el modelo que mejor ajusta los datos recolectados del experimento, el siguiente punto de interés es disponer de métodos para hacer inferencias acerca de los parámetros del modelo elegido. Distinguimos dos apartados en este subcampo que a continuación analizamos.

CP1.8.1. *Estimación del valor de los parámetros coeficientes de x^i en un modelo de regresión polinomial.*

Frecuentemente en los problemas de ingeniería ocurre que al construir la gráfica de dispersión de los datos se indica que la función de regresión verdadera $\mu_{Y.X} = \beta_0 + \beta_1 X + \beta_2 X^2 + ... + \beta_K X^K$ tiene uno o más picos o valles, lo cual es señal de que existe por lo menos un mínimo o un máximo. En esas situaciones, una función polinomial $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + ... + \beta_K X^K + \varepsilon$ nos daría un valioso acercamiento para la función de regresión verdadera.

Los estimadores: $\hat{\beta}_0, \hat{\beta}_1, ...\hat{\beta}_K, T = \dfrac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}$ son funciones lineales de Y_i de tal manera que cada $\hat{\beta}_i$ tienen una distribución normal y son estimadores insesgados de β_i Los procedimientos inferenciales para estimar los coeficiente de x^i en la función de regresión polinomial usan la variable estandarizada $T = \dfrac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}$ que tiene una distribución t basada en n – (k+1) grados de libertad. En Devore (2005, pg. 581), aparece la fórmula del intervalo de confianza de 100(1-α)% para β_i el coeficiente de x^i en la función de regresión polinomial, pero no se ilustra con ejemplo.

CP1.8.2 *Estimación del valor del parámetro $\mu_{Y.X}$ en un modelo de regresión polinomial.* La estimación puntual de la función de regresión verdadera es:

$$\hat{\mu}_{Y.X} = \hat{\beta}_0 + \hat{\beta}_1 x + ...\hat{\beta}_k x^k .$$

Matemáticamente la desviación estándar estimada para cualquier valor de x es una expresión muy complicada, por lo que en la práctica es obtenida a través del uso de software estadístico. Este resultado junto con la variable estandarizada *t*, son utilizados para la estimación del parámetro $\mu_{Y.X}$.

> En la figura 13.11 (b) se muestra que $\hat{\beta}_2$ = -4.5358 y $s_{\hat{\beta}_2} = 0.6744$ (de la columna Desv. Est. en la parte superior del resultado). La última línea del resultado en la figura 13.11 (b) contiene información de la estimación *y* predicción para *x* = 25:
>
> $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1(25) + \hat{\beta}_2(25)^2 = \text{ Ajuste} = 3431.8$$
>
> $$s_{\hat{Y}} = \text{ajuste de desv. est.} = 69.2$$
>
> Un intervalo de confianza de 95% para el rendimiento medio cuando la fecha de cosecha = 25 es
>
> Devore (2005, pg. 582)

CP2. *Comparar los valores de un parámetro en dos o más poblaciones.*

En ingeniería una situación que ocurre continuamente es la comparación de dos o más tratamientos, para ver si uno es superior al otro en la respuesta que se observa. En estos problemas, necesitamos que se cumpla el supuesto de que la diferencia de las medias muestrales $\overline{X}_1 - \overline{X}_2$ se distribuya en forma normal, por lo cual tenemos que considerar nuevamente las condiciones experimentales del proceso de recolección de los datos. El estimador $\overline{X}_1 - \overline{X}_2$ tiene la propiedad de que es un estimador insesgado de $\mu_1 - \mu_2$ y la varianza $V(\overline{X}_1 - \overline{X}_2)$ es la más pequeña de entre todos los estimadores insesgados, por lo cual es utilizado en la construcción del intervalo de las diferencias de medias poblacionales. Si los tamaños de las muestras son suficientemente grandes, el Teorema Central del Límite respalda la validez del intervalo, sin importar la suposición de normalidad en las poblaciones. En este campo distinguimos varios subcampos, que analizamos enseguida.

CP2.1. *Comparar las medias en dos poblaciones.*

CP2.1.1. *Comparar las medias en dos poblaciones conociendo σ_1^2 y σ_2^2, muestras independientes.*

Muchas de las decisiones en ingeniería tienen que ver con la reducción de la variabilidad y una política sistemática de cero defectos, así como también decisiones acerca de las llamadas características de calidad de diseño como lo son: el nivel de desempeño, duración, confiabilidad y funcionamiento. Dentro de este juego de decisiones, uno que se presenta con mucha frecuencia es el que tiene lugar en los procesos de fabricación en que interesa comparar la media de algunas características de calidad de diseño en dos proveedores del proceso.

Se realizaron pruebas de resistencia a la tensión en largueros de aluminio usados en la fabricación del ala de un avión de transporte comercial, de dos calidades diferentes. Por experiencias pasadas con el proceso de fabricación de largueros y el procedimiento de prueba, se suponen conocidas las desviaciones estándar de las resistencias a la tensión. En la tabla 9-1 se muestran los datos obtenido. Si μ_1 y μ_2 denotan las verdaderas medias de la resistencia a la tensión para las dos calidades de los largueros, entonces puede encontrarse un intervalo de confianza de 90% para la diferencia de la resistencia media $\mu_1 - \mu_2$ Montgomery y Runger (2004, pg. 384)

CP2.1.2. *Comparar las medias en dos poblaciones,* $\sigma_1^2 = \sigma_2^2$ *pero desconocidas, muestras independientes pequeñas.*

El gran interés de este subcampo reside en el hecho de que en muchos experimentos el investigador deberá de planear la estrategia de acuerdo a que en la práctica por lo menos un tamaño de muestra es pequeño y los valores de las varianzas poblacionales son desconocidos, cuando está comparando medias μ_1 *y* μ_2 de dos poblaciones. Utilizando información proveniente de dos muestras aleatorias independientes, el análisis del investigador empieza suponiendo que las muestras provienen de dos poblaciones normales con $\sigma_1 = \sigma_2$. El estadístico que se utiliza en la construcción del intervalo es *t* de Student, el cual no es excesivamente sensible a pequeñas diferencias entre las varianzas poblacionales. Por regla, si una varianza es de cuatro veces la otra, debemos de preocuparnos acerca de la validez de los procedimientos, por lo cual lo más indicado será correr la prueba F para confirmar la hipótesis de igualdad de varianzas, revisando previamente que existan las condiciones necesarias para aplicar la prueba como son el supuesto de normalidad, no usarse cuando difieran mucho los tamaños muestrales y ejecutar la prueba utilizando un nivel de significancia relativamente alto (son valores adecuados usar niveles α hasta de 0.2).

Tratamientos para la Acrofobia. La figura 9.25 muestra los puntajes de 15 pacientes bajo un tratamiento estándar y los puntajes de 15 pacientes bajo el nuevo tratamiento. Puntajes más altos corresponden a mayores mejoras en la condición del paciente. Son dos diferentes conjuntos de pacientes en las dos terapias, de modo que son muestras independientes. El puntaje promedio con el nuevo tratamiento es $\bar{y}$ =51.20, mientras que con el tratamiento estándar un puntaje promedio es $\bar{x}$ = 47.47. ¿Es esta diferencia estadísticamente significativa? Las desviaciones estándar muestrales son respectivamente 11.40 y 10.09. Los resultados estadísticos de las muestras son obtenidos corriendo el paquete estadístico Minitab.   Tratamiento estándar: 33, 54, 62, 46,52,42,34,51,26,68,47,40,46,51,60   Tratamiento nuevo: 65, 61, 37, 47,45,53,53,69,49,42,40,67,46,43,51 Hayter (1996, pg.460)

CP2.1.3. *Comparar las medias en dos poblaciones,* $\sigma_1^2 = \sigma_2^2$ *desconocidas, muestras independientes grandes.*

De la misma manera como se procedió en la estimación del valor del parámetro media en una población, en este subcampo con σ_1^2 y σ_2^2 desconocidas para muestras grandes se procede a la aproximación de sustituir σ_1 y σ_2 por las desviaciones estándar de las muestras s_1 y s_2 y se utiliza nuevamente la distribución normal estándar en la construcción del intervalo de confianza para la media, el Teorema Central del Límite valida estos procedimientos, sean poblaciones normales o no. Mostramos en lo que sigue un ejercicio.

Se comparan dos tipos de rosca de tornillo para ver su resistencia a la tensión. Se prueban 50 piezas de cada tipo de cuerda bajo condiciones similares. La marca A tuvo una resistencia promedio a la tensión de 78.3 kg. con una desviación estándar de 5.6 kg., mientras que la marca B tuvo una resistencia promedio a la tensión de 87.2 con una desviación estándar de 6.3 kg. Determine un intervalo de confianza de 95% para la diferencia de las dos medias poblacionales

Velasco y Wisniewski (2001, pg. 241)

CP2.1.4. *Comparar las medias en dos poblaciones,* $\sigma_1^2 \neq \sigma_2^2$ *y desconocidas, muestras independientes pequeñas.*

En este subcampo se consideran distribuciones aproximadamente normales con varianzas diferentes y desconocidas. El procedimiento Smith-Satterhwaite se usa para determinar los límites de confianza. La variable *t* utilizada en esta técnica es una variable aleatoria con aproximadamente la distribución *t* con γ grados de libertad estimados a partir de los valores observados de las varianzas muestrales s_1^2 y s_2^2.

$$\gamma = \frac{(s^2{}_1/n_1 + s^2{}_2/n_2)^2}{\left[(s^2{}_1/n_1)^2/(n_1-1)\right] + \left[(s^2{}_2/n_2)^2/(n_2-1)\right]}$$

El valor de γ que resulta en esta expresión no siempre es un entero. Cuando eso ocurre se hace un redondeo al entero inmediato inferior, con propósitos conservadores. A medida que aumenta el valor de γ relacionados con variables aleatorias *t*, las curvas acampanadas se vuelven más compactas. Lo cual significa que el punto $t_{.025}$ relacionado con la curva t_{10} es un poco mayor que el punto $t_{.025}$ relacionado con la curva t_{11}, con sus implicaciones en la construcción del intervalo. Un ejemplo que ilustra esta comparación, bajo estas condiciones, aparece en Walpole y Myers (1992, pg. 262):

En 1980, el Departamento de Zoología en el Virginia Polytechnic Institute and State University dirigió un estudio sobre la “Nutrient Retention and Macroinvertebrate Comunity Response to Sewage Stress in a Stream Ecosystem” para estimar la diferencia en la cantidad de ortofósforo químico medido en dos estaciones diferentes en el río James. El ortofósforo se mide en miligramos por litro. Se sacaron quince muestras de la estación 1 y 12 de la estación 2. Las 15 primeras tuvieron un contenido promedio de ortofósforo de 3.84 miligramos por litro y una desviación estándar de 3.97 miligramos por litro, mientras que en las 12 segundas, estos datos fueron de 1.49 miligramos por litro y 0.80 miligramos por litro, respectivamente. Encuentre un intervalo de confianza de 95% para la diferencia en los contenidos promedios reales de ortofósforo en las dos estaciones, asumiendo que las observaciones surgen de poblaciones normales con varianzas distintas.

CP2.1.5. *Comparar las medias en dos poblaciones,* $\sigma_1^2 = \sigma_2^2$ *pero desconocidas, muestras no son independientes y las varianzas de las dos poblaciones no son necesariamente iguales.*

La situación considerada en este subcampo se refiere a un caso experimental muy especial, llamado observaciones pareadas. Las condiciones de las dos poblaciones no se asignan en forma aleatoria a unidades experimentales. Cada unidad experimental homogénea recibe ambas condiciones poblacionales. Cada unidad experimental tiene un par de observaciones, una para cada población. Parear observaciones es una estrategia que tiene muchas aplicaciones. La importancia de esta estrategia es que se reduce la eficacia de la “varianza del error experimental” y por lo tanto se reduce el error estándar de la estimación puntual. Como vemos en el siguiente ejemplo.

Construya un intervalo de confianza del 95% para la diferencia media de los datos pareados relacionados con el desgaste de los neumáticos, según se reporta en la tabla 10.1. La información muestral es n=6 posiciones de datos pareados, $\bar{d}$ = 6.3 y s_d = 5.1 (calculada en el ejemplo 10.3). Suponga que la cantidad de desgaste posee una distribución aproximadamente normal para ambas marcas de neumático.

Johnson y Kuby (2004, pg. 414)

CP2.2. *Comparación de varianzas.*

Otras situaciones problemáticas que se ubican en este campo tienen que ver con la comparación de los valores de los parámetros: proporción y varianza en dos poblaciones. Muchos de los problemas de ingeniería tienen que ver con la comparación de los porcentajes o proporciones de dos grupos, en tanto que la importancia de la comparación de la varianza en dos poblaciones radica en que continuamente en ingeniería se requiere comparar la uniformidad de dos procesos. Presentamos el siguiente ejemplo para la comparación de varianzas:

> Una empresa ha estado experimentando con dos disposiciones físicas distintas de su línea de ensamble. Se ha determinado que ambas disposiciones producen aproximadamente el mismo número de unidades terminadas al día. A fin de obtener una disposición que permita un mayor control del proceso, usted sugiere que se adopte de manera permanente la disposición que exhiba la varianza más pequeña con el número de unidades terminadas producidas al día. Dos muestras aleatorias independientes producen los resultados que se muestran en la tabla 8.9. Establezca un intervalo de confianza de 95% para $\sigma_1^2 / \sigma_2{}^2$, la razón de las varianzas del número de unidades terminadas para las dos disposiciones de línea de ensamble. Con base en el resultado, ¿cuál de las dos disposiciones recomendaría usted?
>
> Mendenhall y Sincich (1997, pg. 403)

CP2.3.*Comparación de proporciones.*

En tanto que para la comparación de proporciones un ejemplo se reproduce en lo que sigue:

> El director de personal de una empresa grande desea comparar dos pruebas distintas de aptitud que se supone son equivalentes en cuanto a las aptitudes que miden. Tiene motivos para sospechar que el grado de la diferencia en las calificaciones no es el mismo para hombres que para mujeres. Por ello, organiza un experimento en el que se seleccionan cien hombres y cien mujeres, de aptitudes casi iguales, para tomar parte en la prueba. Cincuenta hombres hacen el examen I y otros 50, seleccionados en forma independiente, hacen el examen II. Igualmente, 50 mujeres hacen la prueba I y 50, seleccionadas en forma independiente, hacen la prueba II. Las proporciones, de cada uno de los grupos de 50, que reciben calificaciones aprobatorias se anotan en cada caso y los datos son los siguientes:

Prueba	I	II
Hombres	0.7	0.9
Mujeres	0.8	0.9

> Si p_{ij} representa la probabilidad verdadera de pasar en el renglón i y la columna j, el director desea estimar $(p_{12} - p_{11}) - (p_{22} - p_{21})$ para ver si el cambio en la probabilidad para los hombres es distinto del correspondiente cambio para las mujeres. Estimar esta cantidad en un intervalo de confianza del 95%.
>
> Scheafer y McClave (1993, pg. 275)

CP3. *Contrastes de hipótesis.*

CP3.1 *Contrastar una hipótesis sobre una sola media poblacional.*

La importancia de este campo de problemas radica en que los intervalos de confianza están en relación muy cercana con el procedimiento de contrastes de hipótesis. La estimación del intervalo de confianza implica calcular límites para los cuales se considera plausible que el parámetro media poblacional esté dentro de ellos. Por ejemplo la prueba de H_o: $\mu=\mu_o$ en contraste con H_1: $\mu\neq\mu_o$ en un nivel de significancia α es equivalente a calcular un intervalo de confianza del $100(1-\alpha)$ % de μ y si μ_o no está dentro del intervalo de confianza rechazar H_o.

La relación de equivalencia del intervalo de confianza con los contrastes de hipótesis se amplía a diferencias entre dos medias poblacionales, relaciones de varianzas, etc. Un ejemplo que ilustra esta equivalencia, es el siguiente:

> Considérese el ejemplo 7.2 que dice: Se calcula que la media de los promedios de los puntos de calidad de una muestra aleatoria de 36 alumnos universitarios de último año es 2.6. Encuentre un intervalo de confianza del 95% para la media del total de alumnos del último año. Asuma que la desviación estándar de la población es 0.3. El intervalo de confianza del 95% de la media lo dan los límites [2.5 y 2.7]. Así, con la misma información muestral, no se rechazará una hipótesis de dos colas de μ que involucra cualquier valor hipotético entre 2.5 y 2.7.
>
> Walpole y Myers (1992, pg. 322)

CP4: *Hallar un tamaño adecuado de muestra de modo que se obtenga un error dado de estimación con una cierta probabilidad.*

Se calcula un tamaño de muestra, que satisface ciertas condiciones de nivel de confianza, margen de error y dispersión de la variable relacionada con el parámetro que se desea estimar. Para ello, en este caso, se estandariza el error de estimación del parámetro media poblacional, $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$, a la normal tipificada. Obteniendo el valor de z por tablas, calculadora o el ordenador, y despejando el valor de n, es posible calcular el tamaño de muestra óptimo que satisface las condiciones establecidas.

El gran interés de esta aplicación, es que en muchas investigaciones se desea saber qué tan grande debe ser una muestra para asegurar que el error en la estimación de μ será menor que una cantidad específica e. Un ejemplo el siguiente:

> Encuentre el tamaño de muestra que se requiere para estimar la media de la población dentro de 1/5 de la desviación estándar con el 99 % de confianza.
>
> Johnson y Kuby (2004, pg. 318)

CP5. *Cálculo de límites de tolerancia.*

Frecuentemente el interés del estadístico o el científico por estimar parámetros pasa a un segundo plano, y en lugar de ello se interesa más en donde podrían caer las observaciones individuales o mediciones. Si un ingeniero fabrica partes y dispone de las especificaciones, formuladas por el cliente o el ingeniero de diseño, de determinada dimensión para la manufactura de la parte, el valor medio de estas partes es una cuestión secundaria, pues lo importante es intentar determinar los límites en los cuales se cubran los valores individuales de la población (es decir, los valores medidos de las dimensiones) tomando en cuenta el azar.

Para establecer un límite sobre valores sencillos en la población un método es determinar un intervalo de confianza sobre una proporción fija de las mediciones. Un límite que cubre el 90% de la mitad de la población de las observaciones es, $\mu \pm 1.64\sigma$ este intervalo recibe el nombre de intervalo de tolerancia. Sin embargo en la práctica se debe aplicar: $\bar{x} \pm ks$ donde s es la desviación estándar muestral y k se determina de tal forma que se pueda asegurar, con una confianza de 100(1-γ)%, que los valores dados contienen al menos la proporción de 1-α de las mediciones.

Una máquina produce piezas metálicas de forma cilíndrica. Se toma una muestra de tales piezas y se encuentra que sus diámetros son 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03 centímetros. Encuentre los límites de tolerancia del 99% que contenga el 95% de las piezas metálicas producidas por esta máquina, suponiendo una distribución aproximadamente normal.
Walpole y Myers (1992, pg. 254)

Como resumen, mostraremos en la Tabla 3.2 los campos de problemas encontrados en cada texto analizado. Mostramos también las distribuciones muestrales asociadas.

En la muestra de los libros analizados hemos encontrado cinco campos de problemas, dos de los cuales (CP1 y CP2) con sus 22 subcampos han dado origen históricamente, al desarrollo teórico del objeto estadístico intervalos de confianza. Desarrollo que en sus inicios con Laplace y Gauss se hicieron contemplaciones acerca de un valor desconocido del parámetro θ y que generalmente eran formuladas para estimar parámetros en un modelo lineal.

En la indagación de la respuesta a la pregunta de cuál es el valor del parámetro poblacional, observamos en los campos de problemas encontrados que en general se prefieren los estimadores insesgados y de mínima varianza en la construcción del intervalo de confianza.

En relación a este campo de estimar un parámetro desconocido se puede apreciar en la tabla que los subcampos que aparecen en todos los libros son el de estimación de media de una población o en una muestra grande, con σ conocida (CP1.1.1), estimación de media de una población aproximadamente normal cuando, σ es desconocida (CP1.1.2) y estimación de una proporción (CP1.3).

Tabla 3.2. Campos de problema en los libros de texto

		Libro										
Campo de problemas	Distribución muestral	A	B	C	D	E	F	G	H	I	J	K
CP1.1.1	Normal	*	*	*	*	*	*	*	*	*	*	*
CP1.1.2	*t* de Student	*	*	*	*	*	*	*	*	*	*	*
CP1.1.3	Normal	e	*		*	e	e					*
CP1.1.4.1	Exponencial				*	*						
CP1.1.4.2	Exponencial					*						
CP1.1.5	Libre de Distribución			*		e						*
CP1.2	Libre de Distribución											
CP1.3	Normal	*	*	*	*	*	*	*	*	*	*	*
CP1.4	Chi-cuadrada	*	*		*	*	*	*		*	*	*
CP1.5	Normal				*					*	*	*
CP1.6.1	*t* de Student	*	*	*		*	*			*	*	*
CP1.6.2	*t* de Student	*	*	*		*	*			*	*	*
CP1.7.1	*t* de Student			*		*				*	*	*
CP1.7.2	*t* de Student	*		*		*	*			*	*	*
CP1.8.1	*t* de Student											e
CP1.8.2	*t* de Student											*
CP2.1.1	Normal	*	*	*	*	*	*	*	*	*	*	*
CP2.1.2	*t* de Student	*	*	*	*	*	*	*	*	*	*	*
CP2.1.3	Normal		*		*	e		*				*
CP2.1.4	*t* de Student	*			e	e	*			*	e	
CP2.1.5	*t* de Student	*		*	*	*	*	e	*	*	*	*
CP2.2	F	*	*			*	*	e			*	*
CP2.3	Normal	*	*	*	*	*	*	*	*	*	*	*
CP3.1	Normal, *t*, F, χ^2	*	*	*	*	*	*			e	*	e
CP4	Normal	*	*	*	e	*	*		*	*	*	*
CP5	Normal y *t*	*			*		*					*

*: el texto contiene el campo de problemas. e: no hay ejemplo, aparece un ejercicio.

Con respecto al campo de *problemas comparación de los valores de un parámetro en dos o más poblaciones* los subcampos comunes a todos los libros son aquellos en donde se conocen σ_1^2 y σ_2^2, muestras independientes (CP2.1.1), $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas (CP2.1.2) y finalmente comparación de proporciones (CP2.3). Cada campo o tipo de problemas involucra en su resolución una red de objetos matemáticos, principalmente, procedimientos, conceptos, proposiciones y argumentaciones específicas, esto es, implica una configuración epistémica particular. No obstante, es posible identificar algunos elementos comunes, lo que permite agrupar dichas configuraciones en otras configuraciones más generales. En particular, para el caso del intervalo de confianza en este contexto formativo podemos agruparlas en tipos de configuraciones, según la distribución de probabilidad que intervenga. Las dos principales (porque incluye mayor número diferentes de campos de problemas) corresponden a la distribución normal o la de *t* de Student:

1. Las distribuciones muestrales que siguen o se aproximan mediante la distribución normal; en este segundo caso, justificado por el teorema central del límite (Alvarado, 2004). Son los campos de problemas CP1.1.1, CP1.1.3, CP1.3, CP1.5, CP2.1.1, CP2.1.3, CP2.3, CP3.1, CP4, CP5. Esta configuración epistémica, parte de estos campos de problemas y al buscar la distribución muestral llega a uno de los campos de problemas que da origen al Teorema Central del Límite: hallar la distribución de la suma o la media de un conjunto de variables aleatorias idénticamente distribuidas. Lleva a procedimientos específicos, como son la tipificación de variables y el uso de las tablas de la distribución normal *N (0,1).*
2. Las distribuciones muestrales que siguen o se aproximan mediante la distribución *t* de Student, son los campos: CP1.1.2, CP1.6.1, CP1.6.2, CP1.7.1, CP1.7.2, CP1.8.1, CP1.8.2, CP2.1.2, CP2.1.4, CP2.1.5, CP3.1, CP5. Esta configuración epistémica también conduce a los procedimientos específicos: tipificación de variables y el uso de las tablas de la distribución *t*.

Aparte de ellos surgen otros correspondientes a la distribución exponencial, F o que combinan varias distribuciones.

El texto que contiene la mayor parte de los campos y subcampos es Devore (2005) con 20 subcampos y dos ejercicios, luego le siguen Mendenhall (1997) 17 subcampos y cuatro ejercicios y Montgomery y Runger (2004) 16 subcampos y un ejercicio. En particular Devore y Montgomery y Runger son los dos libros más usados en todo el Sistema Tecnológico de Monterrey. Los subcampos estimación de la media en una distribución exponencial (CP1.1.4), estimación de parámetros en un modelo de regresión polinomial (CP1.7) y estimación de la media de una población no normal, en una muestra pequeña (CP1.1.5) prácticamente no aparecen por su grado de especialización, solo Mendenhall y Sincich (1997), Miller, Freund y Jonson (1997), Hayter (1996) y Devore (2005) los contienen parcialmente.

Los libros que menos campos cubren son Velasco y Wisniewski (2001) y Johnson y Kuby (2004), ambos con ocho campos. El primero es un libro específicamente elaborado para cubrir los objetivos de un curso introductorio de probabilidad y estadística que se imparte en uno de los campus del sistema (Ciudad de México) y el segundo es un libro que gusta a los profesores más por su diseño (explicaciones gráficas, instantáneas del mundo real, fomento de uso de software, materiales de apoyo para el profesor, etc.) que por la amplitud y variedad de sus contenidos, razón por la cual es utilizado más como libro de consulta en los cursos para ingenieros en los campus del sistema.

Uno de los tres nuevos campos encontrados, que guardan una relación indirecta (de equivalencia) con los intervalos de confianza como son los problemas de pruebas de hipótesis están presentes o bien se les menciona en nueve de los 11 libros analizados. Un segundo nuevo subcampo que tiene que ver con la determinación del tamaño de muestra (CP4) también se presenta con importante frecuencia, en nueve de los 11 libros. El que tiene solo regular presencia (cuatro de 11) en los libros es el tercer nuevo campo (CP5) que tiene que ver con el cálculo de límites de tolerancia. En síntesis, en el análisis que hemos realizado de los 11 libros de texto enfocado en este primer elemento de nuestro marco teórico, las palabras que describen el color y tonalidad de los campos encontrados son amplitud, variedad, supuestos y restricciones.

Un aspecto que representa un área de oportunidad para desarrollar materiales didácticos para los alumnos es que la mayoría de los libros analizados solo contienen un ejemplo ilustrativo del campo de problemas, ejemplos que en la mayoría de los casos son muy puntuales y son pocos los libros que contienen ejemplos bien estructurados que manejan información real. Un objetivo de nuestra investigación a largo plazo es generar secuencias de enseñanza que tome en cuenta los resultados de esta fase de la investigación. Diseñar situaciones que contengan escenarios con un contexto atractivo, con problemáticas reales y de actualidad, que fomenten el razonamiento, que estimulen la reflexión y que sean un verdadero reto para los estudiantes.

3.6. PROCEDIMIENTOS Y ALGORITMOS

Para resolver los problemas anteriores, se realizan diversas técnicas o procedimientos. En lo que sigue analizamos los procedimientos descritos en los textos para la resolución de las situaciones problemáticas relacionados con los intervalos de confianza. En los libros analizados hemos encontrado los siguientes:

PA1. *Determinar un modelo de distribución muestral del estadístico:* El interés de la construcción del intervalo es estimar el parámetro o parámetros que se consideran constantes en inferencia clásica. El primer paso será determinar el estadístico adecuado para estimarlo, usualmente su equivalente en la muestra (por ejemplo, la media en la población será estimada por la media muestral).

Puesto que el estadístico varía muestra a muestra es una variable aleatoria. El segundo paso será determinar la distribución muestral del estadístico, que, generalmente está centrada en el parámetro, es decir, su valor medio sería el valor desconocido del parámetro que queremos estimar.

Este procedimiento implica elegir una familia de funciones de densidad, que ha venido determinada por el teorema central del límite y otros teoremas matemáticos, Por ejemplo, para estimar la media y desviación típicas utilizamos la media muestral y raíz cuadradas de la cuasivarianza y la distribución de densidad muestral es normal, como aseguran los teoremas de límite.

Además, como vimos al analizar los campos de problemas hay que analizar las condiciones del problema. Por ejemplo, si el tamaño de muestra es grande, se podría elegir la distribución normal incluso si la población de partida no es normal. Si es pequeña, hay que ver si la distribución de la variable es conocida o no, si se toma la distribución normal, debemos saber si la varianza es conocida, etc. Con todo ello se elige la distribución muestral que servirá posteriormente para calcular el intervalo de confianza.

Hacemos notar que en realidad esta acción consiste en resolver un nuevo problema, pero al llegar al estudio de los intervalos de confianza, el alumno domina la técnica para resolverlo, por lo que ya no es un problema como tal (en relación al intervalo de confianza) sino una técnica. Cuando el alumno por primera vez se enfrentó al estudio de las distribuciones muestrales, sin embargo, este mismo elemento sería un campo de problemas (en relación a otro objeto matemático, las distribuciones muestrales).

Esta es una característica de los objetos matemáticos, que pueden jugar diferente papel a diversos niveles de análisis, pudiendo en unos casos jugar el papel de problemas, en otro de técnicas, etc. Esta misma característica es compartida por otras acciones que describimos a continuación.

PA2. *Determinar valores críticos en la distribución del estadístico:* Los límites del intervalo de confianza para un coeficiente de confianza α son los valores críticos correspondientes a $\alpha/2$ y (1- $\alpha/2$) en la distribución muestral de estadístico, como se muestra, en el siguiente ejemplo:

$$P(\overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}) = 1 - \alpha$$

Pero estos valores no vienen dados directamente en las tablas de la distribución. Según la distribución (normal, *t*, Chi Cuadrada) y según el tipo de tabla (cola superior, cola inferior) a veces hay que emplear simetría u otras propiedades o bien interpolar, depende si se tiene calculadora u ordenador. Todas estas técnicas son objeto de enseñanza. Montgomery y Runger (2004, pg. 162), plantean un ejemplo donde se calculan valores críticos usando la distribución normal estándar.

> Encuentre el valor de *z* tal que *P*(-*z* < *Z* < *z*)= 0.99. Debido a la simetría de la distribución normal, si el área de la región sombreada en la figura 5-14 (7) es igual a 0.99, entonces el área en cada cola de la distribución debe ser igual a 0.005. Por lo tanto, el valor de *z* corresponde a una probabilidad de 0.995 en la tabla II. La probabilidad más próxima en la tabla II es 0.99506, cuando *z* = 2.58.

PA3. *Resolver inecuaciones, para despejar los límites del intervalo*. A partir de expresiones tales como: $P(-t_{\alpha/2,n-1} \le \frac{\sqrt{n}(\overline{X}-\mu)}{S} \le t_{\alpha/2,n-1}) = 1-\alpha$ o bien $P\left[-t_{\alpha/2} < \frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{(1/n_1)+(1/n_2)}} < t_{\alpha/2} = 1-\alpha\right]$ se llevan a cabo una serie de procedimientos de reordenamiento de los elementos en los dos términos de la inecuación, con los cuales finalmente se obtienen los límites de variación del parámetro o parámetros despejados.

> El resultado al final de la Sección 7.3.3 establece que $\frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t_{n-1}$, y de modo que la definición de los puntos críticos de la distribución *t* asegura que $P(-t_{\alpha/2,n-1} \le \frac{\sqrt{n}(\overline{X}-\mu)}{S} \le t_{\alpha/2,n-1}) = 1-\alpha$. Sin embargo, la desigualdad $-t_{\alpha/2,n-1} \le \frac{\sqrt{n}(\overline{X}-\mu)}{S}$ puede ser rescrita $\mu \le \overline{X} + \frac{t_{\alpha/2,n-1}S}{\sqrt{n}}$ y la desigualdad $\frac{\sqrt{n}(\overline{X}-\mu)}{S} \le t_{\alpha/2,n-1}$ puede ser rescrita $\overline{X} - \frac{t_{\alpha/2,n-1}S}{\sqrt{n}} \le \mu$ de modo que $P(\overline{X} - \frac{t_{\alpha/2,n-1}S}{\sqrt{n}} \le \mu \le \overline{X} + \frac{t_{\alpha/2,n-1}S}{\sqrt{n}}) = 1-\alpha$
>
> Hayter (1996, pg. 375)

PA4. *Obtención de intervalos de confianza utilizando un programa de ordenador*. Para calcular intervalos de confianza con un ordenador, a veces el intervalo viene dado directamente (por lo que los alumnos no han de usar la tabla) y no requieren las operaciones o procedimientos anteriores P2 y P3. En otros casos, aunque se necesiten las tablas, en ellas el valor crítico viene dado directamente, no hay que resolver inecuaciones ni usar las propiedades de simetría. Por ello el cálculo se facilita notablemente, aunque el alumno ha de saber cuál es el programa a utilizar. Johnson y Kuby (2004), presenta ejemplos de los comandos que se utilizan para calcular intervalos de confianza de (1-α) para la media μ utilizando los paquetes Minitab, Microsoft Excel y la calculadora TI-83. Devore (2005), Johnson y Kuby (2004) y Mendenhall y Sincich (1997), presentan ejemplos y ejercicios propuestos de intervalos de confianza utilizando los paquetes estadísticos Minitab, SAS y Microsoft Excel.

En el siguiente resultado de **MINITAB** se resumen los cálculos:

Two sample **t** for sin fusion vs fundido

	N	Mean	StDev	SE Mean
sin fusión	10	2903	277	88
fundido	8	3108	206	73

95% CI for mu sin fusi – mu fundido : (-448;38)
T-Test mu sin fusi= mu fundido(vs not =): T = -1,80 P=0.092 DF=15

Devore (2005, pg. 376)

PA5. *Transformaciones geométricas de las variables*. En ocasiones se requieren, por ejemplo, para la construcción del intervalo de confianza del coeficiente de correlación toma como base la transformación *Z* de Fisher.

$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$, este estadístico es un valor de una variable aleatoria que tiene aproximadamente una distribución normal con media $\mu_z = \frac{1}{2}\ln\frac{1+\rho}{1-\rho}$ y varianza $\frac{1}{n-3}$. En consecuencia, los alumnos han de usar los conceptos de función, inversa de una función y conocer las funciones específicas en que se basan las transformaciones.

PA6. *Calculo matricial.* La obtención del intervalo de confianza requiere operaciones matriciales como la transpuesta de una matriz, multiplicación de dos o más matrices, la inversa del resultado de una multiplicación matricial, etc. son procedimientos que se presentan frecuentemente en la obtención de algunos intervalos de confianza, como por ejemplo en el intervalo para β_i, el i-ésimo parámetro del modelo lineal general $\hat{\beta}_i \pm t_{\alpha/2} S\sqrt{c_{ii}}$ el valor c_{ii} es obtenido a través de cálculos matriciales. Por ello se requieren conocimientos de álgebra matricial y su operatoria. Un ejemplo, que aparece en que ilustra el uso de estos procedimientos, es el siguiente:

A fin de predecir el rendimiento de combustible (millaje) de un automóvil con base en su peso y la temperatura en el momento de conducirlo, se desarrolló la adecuación de regresión (véase el ejemplo 12.2.3):

$$\mu_{Y/x_1,x_2} = 24.75 - 4.16x_1 - 0.014897x_2$$

En el ejemplo 12.3.3, se determinó que $s^2 = 0.0214$ y, por ende, $s = \sqrt{s^2}$ = 0.1463. La matriz de varianza-covarianza es:

$$(X'X)^{-1}\sigma^2 = \begin{bmatrix} 6.0707 & -3.0258 & -0.0171 \\ -3.0258 & 1.7385 & 0.0021 \\ -0.0171 & 0.0021 & 0.0002 \end{bmatrix}\sigma^2$$

Un intervalo de confianza de 95 % para β_0, la intersección de este modelo, es:

$$\hat{\beta}_0 \pm t_{\alpha/2} s\sqrt{c_{00}}$$

Puesto que *n* = 10 y *k* = 2, el número de grados de libertad relacionado con el punto $t_{0.025}$ es 10-2-1 =7. El intervalo de confianza está dado por: $24.75 \pm 2.365\,(0.1463)\sqrt{6.0707}$

Milton y Arnold (2004, pg. 470)

PA7. *Comprobar hipótesis sobre la población objetivo con ayuda del ordenador*. Un algoritmo indirectamente relacionado con los intervalos de confianza, tiene que ver con la comprobación de la hipótesis de normalidad de la población de partida. El supuesto de normalidad es una condición que debe cumplirse para la construcción de los intervalos de confianza de parámetros como la media poblacional, por esta razón a los valores obtenidos en las observaciones de la muestra se les aplica la prueba de normalidad usando algún software estadístico. Algunos autores de los libros que hemos analizado utilizan, para la comprobación de la normalidad, gráficas de tallo y hoja, gráficas de caja e histogramas que construyen usando software como el SAS. Otros autores usando un software como el Minitab la comprueban a través del uso del valor *p* que resulta cuando corren la prueba Anderson- Darling o bien haciendo una inspección visual de la gráfica de probabilidad normal que el mismo paquete computacional produce. Un ejemplo que ilustra la salida gráfica de probabilidad normal de los datos correspondientes al módulo de elasticidad de paneles de recubrimiento reforzado se presenta enseguida.

Como parte de un proyecto más grande para estudiar el comportamiento de paneles con recubrimiento esforzado, un componente estructural que se utiliza de manera extensa en Norteamérica, en el artículo “Time- Dependent Bending Properties of Lumber” (J. of Testing and Eval., 1996: 187-193) aparecen varias propiedades mecánicas de especimenes de madera de pino escocés. Considere las observaciones siguientes acerca del módulo de elasticidad (MPa) obtenido un minuto después de cargar en cierta configuración:

10490 16620 17300 15480 12970 17260 13400 13900
13630 13260 14370 11700 15470 17840 14070 14760

En la figura 7.8 se muestra una gráfica de probabilidad normal obtenida de MINITAB.

Devore (2005, pg. 303)

Tabla 3.3. Algoritmos y procedimientos en los libros de texto

Algoritmos y Procedimientos	Libro										
	A	B	C	D	E	F	G	H	I	J	K
PA1.Elegir un modelo de distribución muestral del estadístico	*	*	*	*	*	*	*	*	*	*	*
PA2. Determinar valores críticos en la distribución del estadístico	*	*	*	*	*	*	*	*	*	*	*
PA3. Resolver inecuaciones, para despejar los límites del intervalo	*	*	*	*	*	*	*	*	*	*	*
PA4. Obtención de intervalos de confianza utilizando un programa de ordenador	b	*	*	c	*	b		*	*	*	*
PA5 Transformaciones geométricas de las variables	*	c	*	*	*	*			*	*	*
PA6 Realización de cálculos matriciales	*		*		*	*			*	*	
PA7. Comprobar hipótesis sobre la población objetivo con ayuda del ordenador		c	b	b		*			*	*	*

b: Aparece solo durante el análisis en el modelo de regresión, c: solo se menciona, no hay ejemplos.

En la tabla 3.3 observamos que solamente el 54% de los libros que hemos analizado contiene el procedimiento “comprobar hipótesis sobre la población objetivo con ayuda del ordenador”. Los textos más recientes del 2004 y 2005 si lo contienen, a excepción de Johnson y Kuby (2004) del cual ya mencionamos, en los comentarios de la tabla anterior, algunas de sus características de diseño. De los libros de la década anterior, los únicos que sí lo contienen son Hayter (1996) y Miller, Freund y Johnson (1997). Uno de los errores frecuentemente denunciados sobre el uso incorrecto de la estadística en la investigación es no comprobar los supuestos requeridos para aplicarlos. Resulta por tanto preocupante que los libros de texto universitarios analizados en su mitad no incluyan esta práctica como objeto de estudio.

El uso de los cálculos matriciales es un procedimiento que se presenta con regular frecuencia (seis de 11). Falta en el texto editado por Walpole y Myers y el de Devore. Miller, Freund y Johnson (1997) maneja débilmente el uso de la notación matricial en general, argumentando que como no se utiliza en ningún capítulo del texto, excepto en la regresión lineal múltiple se puede omitir esta sección sin perder continuidad. En concordancia con esta justificación, no aborda la construcción de intervalos de confianza para los parámetros x_i en el modelo de regresión lineal múltiple.

Los procedimientos elegir un modelo de distribución muestral del estadístico, determinar valores críticos en la distribución del estadístico y resolver inecuaciones, para despejar los límites del intervalo (PA1, PA2 y PA3) se presentan en todos los libros analizados. Esto debía de esperarse, porque son el núcleo de procedimientos básicos en la construcción de los intervalos de confianza. Textos como Walpole y Myers (1992) y Velasco y Wisniewski (2001) son representantes de los libros en que apenas empezaba a centrarse la atención en el análisis de datos utilizando el ordenador. Milton y Arnold (2004) y Montgomery y Runger (2004) son los libros que contienen todos los algoritmos y procedimientos que hemos encontrado en los 11 libros de texto, el bloque de libros que contiene solamente tres o cuatro procedimientos está integrado por Miller, Freund y Johnson (1997), Velasco y Wisniewski (2001), Johnson y Kuby (2004) y Scheaffer y McClave (1993).

3.7. LENGUAJE

Los elementos del lenguaje, tales como términos, expresiones, notaciones, gráficos, etc. son necesarios para resolver los problemas matemáticos, para representar objetos abstractos, para generalizar su solución o para describirlos a otras personas. En los libros analizados hemos hallado lo siguiente:

Términos y expresiones

Encontramos una gran variedad entre los que citamos los siguientes: estimación, estimación puntual, estimación por intervalos, estimador, insesgado, de mínima varianza, estimador de máxima verosimilitud, estimador por los momentos, función de decisión, errores aleatorios, intervalo, intervalo de confianza, límites, coeficiente de confianza, muestra, muestra grande, muestra pequeña, muestreo, población, población finita, población infinita, población aproximadamente normal, error de estimación, media, valor esperado, dispersión, desviación típica muestral, desviación típica poblacional, varianza poblacional, varianza muestral, parámetro, estadístico, estadístico robusto, idénticamente distribuidas, no idénticamente distribuidas, simulación con uso del ordenador, demostración, confiabilidad, promedio, datos, simetría, asimétrica, tamaño de muestra, observaciones independientes, fórmula, método, β_i el coeficiente de x^i, coeficiente de correlación poblacional, coeficientes de regresión, intervalos de confianza no paramétricos, etc.

Asimismo encontramos *términos asociados a los conceptos relacionados*: distribución muestral, distribución normal, distribución Ji-cuadrada, distribución *t* de Student, distribución *f*, distribución de probabilidad, variables aleatorias discretas y continuas, propiedades estadísticas, curva normal estándar, curva normal, forma de campana, campana de Gauss, distribución Gaussiana, cálculo de probabilidades de variables y estadísticos, ley de los grandes números, distribuciones clásicas, distribución de diferencias de medias muestrales, función de una variable aleatoria, función de densidad de probabilidad, contraste de hipótesis. recta de mínimos cuadrados, gráfica de dispersión, modelo de regresión lineal simple, estimación de mínimos cuadrados del coeficiente β_i ,variable independiente, predictora o explicativa, variable dependiente o de respuesta, desviación aleatoria o término de error aleatorio, valores ajustados, coeficiente de determinación, cantidad de variabilidad en el estimador, correlación, función densidad de probabilidad normal bivariada, función de densidad de probabilidad conjunta, regresión no lineal múltiple, residuos estandarizados, varianza no constante, mínimos cuadrados ponderados, regresión polinomial, coeficiente de determinación múltiple, coeficiente de x^i , cálculo matricial, Libre de distribución, representación gráfica de datos, análisis, función, ecuaciones normales, pendiente, residuos, suma total de cuadrados, efecto de regresión, función lineal, modelo, observación con gran influencia, relación, matricial.

Hacemos notar que cada una de estas expresiones corresponde a conceptos o propiedades matemáticas que el estudiante debería haber aprendido antes de abarcar el tema, lo que muestra la complejidad del objeto de estudio.

Expresiones algebraicas

Se da la distribución del estadístico y el nivel de confianza:

$$P(\bar{x} - t_{\alpha/2}\, s \big/ \sqrt{n} < \mu < \bar{x} + t_{\alpha/2}\, s \big/ \sqrt{n}\,) = 1 - \alpha$$

$$(\,\hat{p} - p) \big/ \sqrt{p(1-p)/n}\ :$$

así como la fórmula de estandarización de la proporción muestral:

$$\hat{\beta}_i \pm t_{\alpha/2} S \sqrt{c_{ii}}\ \ :$$

y la fórmula del intervalo de confianza para β_i , el i-ésimo parámetro del modelo lineal general.

La primera expresión algebraica corresponde a una igualdad probabilística; hacemos notar la complejidad de las expresiones y los muchos conceptos involucrados en la misma, así como la ambigüedad, que puede llevar a confusión. Mediante la expresión se indica el rango de valores del estadístico (considerado éste como variable aleatoria) que está a una distancia dada del valor desconocido del parámetro con una cierta probabilidad. Sin embargo, una interpretación incorrecta es que el intervalo de los valores del parámetro a una distancia fija del estadístico (considerado éste constante) con una cierta probabilidad. Este tipo de expresiones cuando no se aclaran pueden contribuir a la difusión del error citado (Cumming y Fidler 2005 a y b).

Otras expresiones algebraicas se usan al describir los conceptos relacionados: Por ejemplo, es frecuente $n=\left[Z_{\alpha/2}\,\sigma/E\right]^2$: fórmula para calcular el tamaño de muestra, cuando el parámetro que se quiere estimar es una media poblacional, $y=\beta_0+\beta_1 x+\beta_2 x^2+\varepsilon$: modelo de regresión no lineal; $\left(X^{'}X\right)^{-1}\sigma^2$: matriz de varianza-covarianza que es utilizada en la obtención del intervalo de confianza para β_i en el modelo lineal general. En la sección sobre campos de problemas se presentan muchos otros ejemplos de expresiones algebraicas

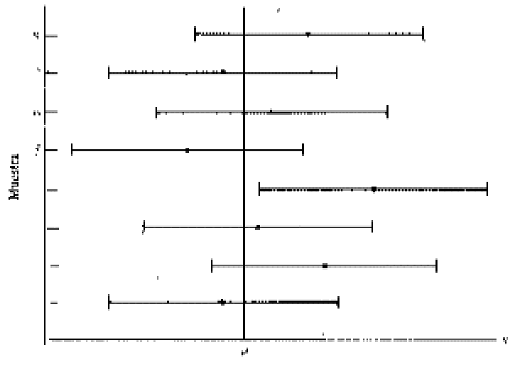
Representaciones gráficas

Los gráficos es uno de esos elementos que siempre están presentes en la Estadística. Se reconoce en los gráficos un medio elocuente de comunicar ideas, para refutar teorías, de representar relaciones entre varias variables, de usarlos para distinguir con ellos el foco de atención, en cómo se cuestionan los datos, en sus propósitos, en la precisión de la información cualitativa extraída de ellos y la realización ocasional de funciones instrumentales. En la muestra de libros revisados hemos encontrado los siguientes tipos:

- Al describir los intervalos de confianza: *diagramas de caja* para comparar las medias provenientes de muestras de dos poblaciones diferentes.
- *Gráfica de dispersión con intervalos de confianza,* esta gráfica se obtiene al ajustarse el modelo de regresión lineal simple y calcular los intervalos de confianza para un valor medio.
- *Gráficas de probabilidad normal, gráficas de tallo- hoja, histogramas y diagramas de caja* para verificar el supuesto de normalidad.
- *Figura con diferentes intervalos de estimación del parámetro* μ, construida para visualizar la interpretación clásica de los intervalos de confianza. En la figura 3.1 tomada del libro Walpole y Myers (1992, pg. 248) se exhibe este último tipo de gráfica que es una representación típica que se presenta en muchos de los libros que

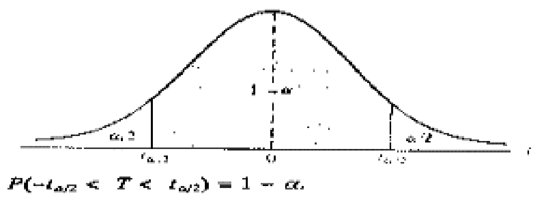
hemos analizado.

Figura 3.1. Diferentes estimaciones por intervalos de μ para diferentes muestras.



Para describir los conceptos relacionados hemos hallado: *gráfica de dispersión con la recta de mínimos cuadrados* en el modelo de regresión lineal simple, *gráficas de funciones intrínsecamente lineales*; funciones que relacionan a *y* con *x*, en donde por medio de una transformación las funciones se pueden expresar como $y' = \beta_0 + \beta_1 x'$; *diagramas residuales* se grafican los residuos estandarizados en función de *x;*, *distribución de* ε, distribución del error aleatorio en el modelo de regresión; *gráfica de fdp normal bivariada y gráficas de funciones de densidad;* para las distribuciones de probabilidad continuas Chi-cuadrada, *F*, *t* de Student y normal estándar , con el área sombreada mostrando el coeficiente de confianza. Enseguida se presenta la figura 3.2 exhibida en Walpole y Myers (1992, pg. 251).

Figura 3.2 Distribución t de Student mostrando el coeficiente de confianza 1-α



Representaciones elaboradas por el ordenador

Los paquetes estadísticos Minitab, SAS, el paquete Microsoft Excel, S-Plus y otros tantos que existen actualmente en el mercado, ofrecen una gran variedad de representaciones, tanto numéricas como gráficas, que se realizan en forma casi instantánea.

Merece mención aparte de representación, *la simulación*, la cual permite a través del ordenador estudiar las propiedades de un fenómeno aleatorio reemplazándolo por otro isomorfo. Por medio de dispositivos estadísticos como el modelo de urnas es factible simular experimentos probabilísticos, tales como el lanzamiento de un dado, de una moneda, etc. Podemos simular también la cobertura de intervalos de confianza, para ello se simula con el ordenador la obtención de una muestra de tamaño *n*, generada aleatoriamente proveniente de una población normal con parámetros μ y σ y luego con los valores de las observaciones muestrales se construye un intervalo de confianza de digamos 95%, intervalo que posteriormente es graficado. Este procedimiento es repetido k veces generando con ello k intervalos de diferentes longitudes. La representación gráfica de este proceso pone en evidencia que:

$$\frac{\textit{número de intervalos que cubren la media real}}{\textit{número de intervalos calculados}} \rightarrow 0.95$$

Otra situación de interés lo constituye el desarrollo de actividades de simulación en las distribuciones muestrales, las cuales han permitido mejoras en el razonamiento estadístico de los alumnos (delMas, Garfield y Chance, 1999). El papel que juega la simulación en la enseñanza de la estadística y la probabilidad ha sido resaltado entre otros por Biehler (1991, 1997), Estepa (1993), Batanero (2001) y Cumming (2002), avisando que la simulación permite poner en manos del estudiante un instrumento que hace posible la exploración y el descubrimiento de conceptos y principios que de otro modo serían mucho más abstractos. Sin embargo “guías prácticas” de la exploración del simulador, donde se diseña de manera parcial y difusa los significados pretendidos en el proceso de estudio provocará que los estudiantes tengan dificultades para observar el comportamiento del fenómeno, Godino y cols (2006).

Enseguida se presenta un ejemplo de la simulación de muestreo realizada por el ordenador y obtención del intervalo de confianza usando el paquete MINITAB que aparece en Johnson y Kuby (2004, pg. 363).

8.135 La siguiente salida de computadora presenta una muestra simulada de tamaño 25 generada aleatoriamente de una población normal con $\mu = 130$ y $\sigma = 10$. Luego se usó un comando para establecer un intervalo de confianza del 95% para μ.

```
116.187   119.832   121.782   122.320   141.436   129.197   119.172
120.713   135.765   131.153   122.307   126.155   137.545   141.154
123.405   143.331   121.767   109.742   140.524   150.600   121.655
127.992   136.434   139.768   125.594

   N        MEAN       STDEV       SE MEAN     95.0 PERCENT C.I.
   25       129.02      10.18        2.00       (125.10, 132.95)
```

a. Escriba el intervalo de confianza que se obtuvo.

Como resumen, mostraremos en la Tabla 3.4 el lenguaje encontrado en cada texto analizado. Observamos que las primeras dos categorías están presentes en todos los libros. Las representaciones gráficas, que para el análisis de los libros solamente hemos considerado, son las situaciones en las cuales se describe los intervalos de confianza, dejando de lado las situaciones para describir los conceptos relacionados, solamente Mendenhall y Sincich (1997) y Velasco y Wisniewski (2001) no las contienen.

Tabla 3.4 Lenguaje en los libros de texto

Lenguaje	Libro										
	A	B	C	D	E	F	G	H	I	J	K
Términos	*	*	*	*	*	*	*	*	*	*	*
Expresiones algebraicas	*	*	*	*	*	*	*	*	*	*	*
Representaciones Gráficas	*	*	*	*		*		*	*	*	*
Simulación		*	*	*	*			*			*

Las gráficas más utilizadas en los libros para verificar el supuesto de normalidad son las gráficas de probabilidad y la gráfica de caja y bigote y el diagrama tallo y hoja solo apareció en el libro de Milton y Arnold (2004). Con respecto a la simulación esta aparece en solo seis de los 11 libros (cuatro de ellos de la década pasada). Johnson y Kuby (2004) simula la obtención de una muestra por métodos manuales a través del uso de una tabla de números aleatorios para luego construir un intervalo de confianza con el uso de software estadístico. Scheaffer y McClave (1993). Uno de los libros de la anterior década contiene una simulación de intervalos de confianza generando al azar muestras aleatorias de tamaño *n* a partir de una población con distribución exponencial, con el uso del ordenador. Cuando encontramos presente la simulación, siempre fue para estimar una media poblacional, de los otros campos y subcampos no encontramos testimonios.

Hemos encontrado en los libros analizados una gran riqueza y diversidad de términos, la importancia que tiene el lenguaje en la comprensión de los intervalos de confianza se manifiesta por los múltiples elementos implicados en el tema. Hemos realizado la confrontación de los términos, las expresiones algebraicas, las representaciones gráficas y las representaciones ofrecidas por el ordenador, en particular, la simulación, en los diferentes textos, lo cual nos ha permitido una mejor evaluación de las fuerzas, debilidades y oportunidades de los textos en este elemento de significado.

3.8. DEFINICIONES DEL INTERVALO DE CONFIANZA

En la actividad matemática que realiza el sujeto, necesita evocar diferentes conceptos o nociones sobre los símbolos u objetos materiales con los que opera. Además en dicha actividad necesita evocar diferentes conceptos o nociones matemáticas que conoce con antelación y en los que se apoya para resolver el problema, mediante sus definiciones. Las que hemos encontrado son las siguientes:

D1. *Definición a partir de la fórmula de cálculo*

D1.1. *Definición a partir de la fórmula de cálculo, interpretando correctamente el significado.* En Miller, Freund y Johnson (1997, pg. 222) se presenta la siguiente definición del intervalo de confianza y luego unas cuantas líneas después aparece la interpretación, que corresponde a la clásica, la cual es una interpretación correcta.

> Cuando disponemos del valor observado *x*, obtenemos
>
> $$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
>
> De esta manera, cuando se ha obtenido una muestra y se ha calculado el intervalo de $\overline{x}$, podemos sostener con (1-α) 100% de confianza que el intervalo de $\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ contiene a μ. A un intervalo de esta clase suele designársele con el nombre de **intervalo de confianza** para μ con el grado de confianza.
> 1-α o (1-α) 100% y a sus puntos extremos **límites de confianza**. El grado de confianza (1-α) 100% significa que el método con el cual se obtuvo el intervalo funciona el (1-α)100% de las veces. En otras palabras, en aplicaciones repetidas de la fórmula del intervalo de confianza, puede esperarse que el 95% de los intervalos contenga las medias de las poblaciones respectivas (pg. 223).

D1.2. *Definición a partir de la fórmula de cálculo, interpretando en forma incorrecta o ambigua el significado***.**

En Walpole, Myers y Myers (1999, pg. 244) se presenta la siguiente definición del intervalo de confianza y luego unas cuantas líneas después aparece una interpretación. Hacemos notar que la interpretación dada es ambigua, puesto que a pesar de hablar de aplicaciones repetidas, no aclara por completo que porcentaje de los intervalos contendrían a μ para luego esto relacionarlo con el coeficiente de confianza. Esta ambigüedad puede llevar a una interpretación errónea, como que el intervalo da la probabilidad de que el parámetro esté comprendido en los extremos (considerados éstos fijos) con una cierta probabilidad. Es decir, dar una interpretación bayesiana como intervalo de credibilidad.

> Si $\bar{x}$ es la media de una muestra aleatoria de tamaño *n* de una población con varianza σ^2, conocida, un intervalo de confianza de (1-α)100%, para μ está dado por:
>
> $$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
>
> donde $z_{\alpha/2}$ es el valor *z* que deja un área de α/2 a la derecha.
>
> Explicación del autor: Muestras diferentes darán valores diferentes de $\bar{x}$ y por tanto producirán diferentes intervalos de estimación del parámetro μ como se muestra en la figura 9.3. Los puntos circulares al centro de cada intervalo indican la posición de la estimación puntual $\bar{x}$ para cada muestra aleatoria. Se ve que la mayor parte de los intervalos contienen a μ, pero no todos los casos. Entre más grande elegimos el valor $z_{\alpha/2}$ hacemos más anchos todos los intervalos y podemos tener una confianza en que la muestra particular que se seleccione producirá un intervalo que contenga el parámetro desconocido μ.

D2. *Definición como intervalo con extremos que dependen del valor de una variable aleatoria* (*estadístico en la muestra*). En esta definición se enfatiza el carácter aleatorio del intervalo en las diversas muestras.

> Una estimación por intervalo de un parámetro poblacional θ es un intervalo de la forma $\theta_L < \theta < \theta_U$ donde θ_L y θ_U , dependen del valor del estadístico Θ para una muestra particular y también de la distribución muestral de Θ. Ya que muestras distintas generalmente dan valores distintos de Θ y, por lo tanto, de θ_L y θ_U, estos puntos extremos del intervalo son los valores de las variables correspondientes Θ_L y Θ_U. A partir de la distribución muestral de Θ será posible determinar θ_L y θ_U, tales que la $P(\Theta_L < \theta < \Theta_U)$ sea igual que cualquier valor fraccional positivo que se desee especificar. Si, por ejemplo, se encuentran θ_L y θ_U tales que,
>
> $$P(\Theta_L < \theta < \Theta_U) = 1 - \alpha$$
>
> Para 0< α < 1, entonces se tiene una probabilidad de 1- α de seleccionar una muestra aleatoria que produzca un intervalo que contenga a θ. El intervalo de la forma $\theta_L < \theta < \theta_U$, que se calcula a partir de la muestra seleccionada, se denomina intervalo de confianza del (1-α)100%, la fracción 1-α se denomina coeficiente de confianza o grado de confianza, y los puntos extremos θ_L y θ_U se llaman límites de confianza inferior y superior respectivamente.
>
> Walpole y Myers (1992, pg. 245)

D3. *Definición de intervalo de confianza usando el método de inversión del estadístico de prueba*.

Sabemos que hay una fuerte relación entre pruebas de hipótesis e intervalos de confianza, por lo que resulta razonable que entre los métodos que identifican autores de textos de Estadística Matemática para encontrar intervalos de estimación está el de invertir un estadístico de prueba, en el cual se canaliza esa relación de equivalencia. Un ejemplo del uso de esta definición aparece durante la construcción de intervalos de confianza no paramétricos.

> Si la prueba de Wilcoxon de los rangos con signo el nivel α para $H_0 : \mu = \mu_0$ contra $H_a : \mu \neq \mu_0$ es rechazar H_0 si $s_+ \geq c$ o $s_+ \leq n(n+1)/2 - c$, entonces un intervalo de confianza de 100(1- α)% para μ es $(\overline{x}_{(n(n+1)/2-c+1)}, \overline{x}_{(c)})$ donde: $s_+ =$ número de promedios por pares $(x_i + x_j)/2$ con $i \leq j$ que son $\geq \mu_0$
>
> Devore (2005, pg. 686)

Como resumen, mostraremos en la Tabla 3.5 las definiciones encontradas en cada texto analizado.

Tabla 3.5. Definiciones en los libros de texto

definiciones	Libro										
	A	B	C	D	E	F	G	H	I	J	K
D1.1. Definición a partir de la fórmula de cálculo, interpretándose correctamente el significado.		*	*	*	*			*	*	*	*
D1.2. Definición a partir de la fórmula de cálculo, interpretándose el significado en forma ambigua (A) o incorrecta (I).	A					A					
D2. Definición como intervalo con extremos que dependen del valor de una variable aleatoria (estadístico en la muestra).	*	*				*	*		*	*	*
D3. Definición de intervalo de confianza usando el método de inversión del estadístico de prueba.			*								*

En la tabla 3.5 observamos que todos los libros, excepto Velasco y Wisniewski (2001) definen intervalo de confianza a partir de la fórmula de cálculo y en algún otro espacio, líneas atrás o líneas adelante, dan la interpretación. La mayoría de los libros dan la interpretación correcta del significado. Solamente los libros Walpole y Myers (1992) y Walpole, Myers y Myers (1999) dan una interpretación ambigua. La interpretación que presentan esos libros corresponde a la construcción del intervalo para la media poblacional, salvo un par de excepciones (Walpole y cols) que además interpretan para diferencias de medias poblacionales, luego una vez hecha esta interpretación las siguientes definiciones que presentan lo hacen dando solamente la fórmula de cálculo sin insistir más en su interpretación.

Así podemos resaltar que las definiciones que encontramos para la mayoría de los otros parámetros: diferencia de medias poblacionales, correlación poblacional, coeficientes de regresión β_i, etc. son a partir de la fórmula de cálculo, sin insistir más en la interpretación.

En relación a la definición como intervalo con extremos que dependen del valor de una variable aleatoria (estadístico en la muestra) existe una presencia en los libros que calificaríamos de regular, son siete de 11 los libros en que se presenta esta definición. La definición de intervalo de confianza usando el método de inversión de un estadístico de prueba, se encuentra únicamente en dos de los 11 libros, Hayter (1996) y Devore (2005). Los libros en que menos aparecen estas definiciones son Johnson y Kuby (2004), Mendenhall y Sincich (1997) y Miller, Freund y Johnson (1997). En tanto que el libro que contiene estas tres definiciones es Devore (2005).

3.9. PROPIEDADES Y PROPOSICIONES

Frecuentemente los libros usan diferentes definiciones de un mismo objeto matemático, cada una de las cuales enfatizan propiedades o atributos específicos de dicho objeto. Algunas de estas definiciones las relacionan con otros conceptos o propiedades que ya se conocen. En lo que sigue listamos las propiedades específicas de los intervalos encontradas en los textos. Estas propiedades pueden ser enunciadas como proposiciones que se demuestran, o también pueden quedar implícitas.

P1. *El ancho de los intervalos de confianza disminuye cuando el tamaño de la muestra aumenta***.** Al calcular los límites, de por ejemplo el intervalo de confianza para la media, el elemento que se le agrega a la media muestral, llamado error de la estimación, contiene en su denominador el factor tamaño de muestra, de ahí que al aumentar el tamaño de muestra, el efecto que produce es el de disminuir el ancho del intervalo.

P2. *El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta.* En la construcción de los intervalos quisiéramos tener aquel que tuviera la mayor confianza, sin embargo esta condición tiene su precio, al aumentar el nivel de confianza se tiene un intervalo más amplio, con lo cual se produce una mayor imprecisión en la estimación del parámetro. La ganancia en confiabilidad trae inherente una pérdida de precisión.

P3. *Significado del nivel de confianza (variación del intervalo en diferentes muestras).* La interpretación del nivel de confianza 100(1- α) % es que, si repitiésemos el proceso de construcción de intervalos con el mismo nivel y en muchas muestras de la misma población, tomadas todas las muestras en las mismas condiciones, el 100(1- α) % de los intervalos así construidos contendría al verdadero valor del parámetro (MEC, 2004).

Además de estas tres propiedades básicas, hemos encontrado otra serie de propiedades no relacionadas directamente con el intervalo, sino con la distribución muestral del estadístico y la distribución del error aleatorio del modelo de regresión.

P4. *Los estimadores de máxima verosimilitud tienen distribución asintótica normal.* El método de máxima verosimilitud que es una técnica aplicable a una gran variedad de problemas relacionados con los intervalos de confianza tiene las propiedades de ser asintóticamente insesgado, consistente e invariante. Cabe señalar que aún cuando el método de máxima verosimilitud es con frecuencia una técnica que prefieren los estadísticos matemáticos, en ocasiones surgen complicaciones en su uso. Por ejemplo no siempre es sencillo maximizar la función de verosimilitud porque la ecuación obtenida no es fácil de resolver o bien no es posible la utilización directa de los métodos de cálculo para determinar el máximo de $L(\theta)$.

P5. *La media de una muestra suficientemente grande sigue una distribución aproximadamente normal.* Este resultado es una consecuencia inmediata del Teorema Central del Límite, el teorema más importante de la probabilidad. Devore (2005, pg. 239) presenta el enunciado de este teorema donde se resalta esta propiedad.

Sea $X_1, X_2,..., X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces si n es suficientemente grande, $\overline{X}$ tiene aproximadamente una distribución normal con $\mu_{\overline{x}} = \mu$ y $\sigma_{\overline{x}}{}^2 = \sigma^2/n$ y T_0 tiene una distribución casi normal con $\mu_{T_0} = n\mu$, $\sigma^2{}_{T_0} = n\sigma^2$. Mientras más grande sea el valor de n, mejor es la aproximación.

P6. *La proporción muestral usada para estimar la proporción poblacional sigue una distribución normal, si n es grande.* Sea y_i el resultado de la i-ésima prueba de Bernoulli, donde

$$y_i = \begin{cases} 1 \ \ si \ \ el\, resultado\, es\, un\, éxito \\ 0 \ \ si\, el\, resultado\, es\, un \ \ fracaso \end{cases}$$

entonces el número de *y* éxitos en *n* pruebas es igual a la suma de *n* variables aleatorias de Bernoulli independientes:

$$\sum_{i=1}^{n} y_i$$

De ahí que, $\hat{p} = y/n$ es una media de muestra *y*, en consecuencia por el teorema central del límite $\hat{p}$ se distribuye aproximadamente normal.

P7. *La diferencia de medias muestrales en dos poblaciones sigue una distribución normal.* Si tomamos dos muestras aleatorias que se escogen en forma independiente de las poblaciones objetivo. Es decir la selección, en estas condiciones, de elementos para una muestra no afecta, ni es afectada por la selección de elementos para la otra muestra. Supongamos además que las poblaciones objetivo tienen medias μ_1 y μ_2 y varianzas $\sigma_1{}^2$ y $\sigma_2{}^2$ respectivamente, entonces la distribución muestral para la diferencia de medias $\overline{X}_1 - \overline{X}_2$ es aproximadamente normal, con media: $\mu_1 - \mu_2$ *y varianza* $\sigma_1{}^2/n_1 + \sigma_2{}^2/n_2$, Bain y Engelhardt (1989, pg. 356).

P8. *La diferencia de proporciones muestrales en dos poblaciones sigue una distribución normal*. El teorema central del límite asegura que si los tamaños de muestras n_1 y n_2 tienen el tamaño suficientemente grande, la distribución muestral de $(\hat{p}_1 - \hat{p}_2)$ es aproximadamente normal.

P9. *La distribución del estimador* $\hat{\beta}_i$ *del coeficiente de regresión* β_i *en el modelo de regresión es aproximadamente normal.* Según Hayter (1996, pg. 635) puesto que $\hat{\beta}_1$ el estimador de la pendiente en el modelo de regresión lineal simple, es una combinación lineal de variables aleatorias distribuidas normalmente Y_i, entonces $\hat{\beta}_1$ es también distribuida normalmente, de modo que:

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{XX}}\right)$$

donde: β_1 es la verdadera pendiente, σ^2 es la varianza del término del error ε en el modelo de regresión y $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$. En el modelo de regresión lineal múltiple el estimador $\hat{\beta}$ tiene una distribución normal multivariada con matriz de covarianza $\sigma^2(X'X)^{-1}$.

$$\hat{\beta} \sim N_{k+1}\left(\beta, \sigma^2(X'X)^{-1}\right)$$

P10. *La distribución del error aleatorio ε es aproximadamente normal.* La predicción de *y* para un valor dado de *x*, en un modelo probabilístico, es un proceso inferencial que requiere conocer las propiedades del error de la predicción si queremos hacer predicciones que sean de utilidad en el mundo real. En un modelo probabilístico como el modelo de regresión lineal simple $Y = \beta_0 + \beta_1 x + \varepsilon$, ε es denominado el error aleatorio del modelo, que captura la limitación de la inteligencia humana para construir un modelo exacto de la realidad.

Los errores asociados a cualquier dos observaciones distintas son independientes, ε cambia en cada observación, es una variable aleatoria con una distribución de probabilidad específica, que se asume ser normal. El supuesto de normalidad de la distribución del error con media 0 y varianza σ^2, se apoya en el hecho de la ocurrencia frecuente de la distribución normal en la naturaleza, Mendehall, Scheaffer y Wackerly (1986, pg. 460).

Finalmente hay propiedades que indican cuáles son los mejores estimadores de ciertos parámetros.

P11. *La media muestral, usada para estimar la media poblacional, es un estimador insesgado de mínima varianza.* Se puede demostrar que el valor esperado de la media muestral es la media poblacional, lo que justificaría la propiedad de estimador insesgado, Mendehall, Scheafefr y Wackerly (1986, pg. 213). También es posible demostrar que la varianza de la media muestral es la más pequeña en comparación a la de cualquier otro estimador de la media poblacional. Por ejemplo, en particular, se puede demostrar que para muestras aleatorias del mismo tamaño de poblaciones normales la varianza de la distribución de muestreo de la mediana es aproximadamente $(\pi/2)$ veces la varianza de la distribución de muestreo de la media muestral. Una demostración de la propiedad de mínima varianza aparece en Bain y Engelhardt (1990, pg. 280) para una muestra en una población exponencial.

P12. *La proporción muestral es un estimador insesgado de la proporción poblacional y de mínima varianza.* Para obtener el valor esperado de $\hat{p}$, es necesario considerarla como una función lineal de una sola variable aleatoria *y*: $\hat{p} = a_1 y_1 = (\frac{1}{n}) y$, aplicando luego el valor esperado se obtiene finalmente $E(\hat{p}) = p$ con lo que se demuestra la propiedad de insesgadez.

P13. *La diferencia de medias muestrales en dos poblaciones, usada para estimar* $\mu_1 - \mu_2$*, es un estimador insesgado de mínima varianza.* La propiedad de insesgadez y de mínima varianza se observa también en el estimador $\overline{X}_1 - \overline{X}_2$ en los campos de problemas relacionados con la comparación de dos o más tratamientos. Montgomery y Runger (2004) presentan la demostración (pg. 377), pero omiten la propiedad de mínima varianza.

P14. *La diferencia de proporciones muestrales en dos poblaciones usadas para estimar* $P_1 - P_2$ *es un estimador insesgado de mínima varianza*. Como $(\hat{p}_1 - \hat{p}_2)$ es una función lineal de las variables aleatorias binomiales y_1 y y_2, se puede demostrar que $E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$, y también el estimador $(\hat{p}_1 - \hat{p}_2)$ tiene mínima varianza. Mendenhall y Sincich (1997, pg. 391) presentan la demostración, en tanto que la propiedad de mínima varianza está fuera del alcance de los objetivos del texto.

P15. *La varianza muestral es un estimador insesgado de la varianza poblacional.* Es posible demostrar que la media de la distribución muestral de S^2 es igual a σ^2, es decir, $E(S^2) = \sigma^2$ cuando se define a $S^2 = \sum (x_i - \overline{x})^2 / n - 1$, con lo cual queda justificada esta propiedad, Walpole y Myers (1992, pg. 244).

P16. *El estimador* $\hat{\beta}_i$ *es un estimador insesgado del coeficiente de regresión* β_i

en la ecuación de regresión. Los estimadores de los mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$ para los parámetros en el modelo lineal simple $y = \beta_0 + \beta_1 x + \varepsilon$ son estimadores insesgados de los respectivos parámetros, Mendehall, Scheafer y Wackerly (1986, pg. 455). La propiedad también se presenta en los estimadores $\hat{\beta}_i$ de los coeficientes de x_i en la ecuación de regresión lineal general $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon$, es decir $E(\hat{\beta}_i) = \beta_i$ Mendehall, Scheaffer y Wackerly (1986, pg. 463).

Como resumen, mostraremos en la Tabla 3.6 las propiedades encontradas en cada texto analizado. En ella observamos que son cinco las propiedades que se presentan en todos los libros que analizamos (P5, P6, P7, P11 y P12), son propiedades que tienen que ver con la insesgadez y mínima varianza para la media muestral y la proporción muestral, además con la forma normal en que aproximadamente se distribuyen la media muestral, la proporción muestral y las diferencias de medias muestrales.

La propiedad que se presenta con menor frecuencia es la de los estimadores de máxima verosimilitud tienen distribución asintótica normal (P4) en seis de los 11 libros. Los libros en donde aparecen las 16 propiedades que hemos identificado son: Devore (2005), Montgomery y Runger (2004) y Walpole, Myers y Myers (1999). Los libros que menos propiedades presentan con un total de 10 son dos textos de la década pasada: Scheaffer y McClave (1997) y Miller, Freund y Johnson (1997). En relación a la propiedad, "la distribución del error aleatorio ε es aproximadamente normal", existen varias técnicas para probar su validez en el modelo de regresión, como por ejemplo el análisis de residuos, los ocho libros donde se encuentra esta propiedad (P10) discuten esta técnica.

La demostración de mínima varianza para los estimadores que identificamos en esta tabla está fuera del alcance de los textos analizados, inclusive en Velasco y Wisniewski (2001) en donde solo se hace mención de la desigualdad de la cota inferior de Cramer-Rao. Estas demostraciones son más bien típicas de los libros de Estadística Matemática, que están fuera del conjunto de libros que hemos analizado.

Tabla 3.6. Propiedades en los libros de texto

Propiedades	Libro										
	A	B	C	D	E	F	G	H	I	J	K
P1.El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra.	*		*	*		*	*	*	*	*	*
P2. El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta.	*		*			*	*	*	*	*	*
P3. Significado del nivel de confianza (variación del intervalo en diferentes muestras).	*	*	*	*	*	*		*	*	*	*
P4.Los estimadores de máxima verosimilitud tienen distribución asintótica normal.	*				*	*	*			*	*
P5. La media de una muestra suficientemente grande sigue una distribución aproximadamente normal.	*	*	*	*	*	*	*	*	*	*	*
P6. La proporción muestral usada para estimar la proporción poblacional sigue una distribución muestral, si n es grande.	*	*	*	*	*	*	*	*	*	*	*
P7. La diferencia de medias muestrales en dos poblaciones sigue una distribución normal.	*	*	*	*	*	*	*	*	*	*	*
P8. La diferencia de proporciones muestrales en dos poblaciones sigue una distribución normal.	*		*	*	*	*	*	*	*	*	*
P9. La distribución del estimador $\hat{\beta}_i$ del coeficiente de regresión β_i en la modelo de regresión es aproximadamente normal.	*	*	*		*	*			*	*	*
P10. La distribución del error aleatorio ε es aproximadamente normal.	*	*	*		*	*			*	*	*
P11. La media muestral, usada para estimar la media poblacional, es un estimador insesgado de mínima varianza.	*	*	*	*	*	*	*	*	*	*	*
P12. La proporción muestral es un estimador insesgado de la proporción poblacional y de mínima varianza.	*	*	*	*	*	*	*	*	*	*	*
P13. La diferencia de medias muestrales en dos poblaciones, usada para estimar $\mu_1 - \mu_2$, es un estimador insesgado de mínima varianza.		*			*	*	*	*	*	*	*
P14. La diferencia de proporciones muestrales en dos poblaciones usadas para estimar $P_1 - P_2$ es un estimador insesgado de mínima varianza.	*		*	*	*	*	*	*	*	*	*
P15. La varianza muestral es un estimador insesgado de la varianza poblacional.	*	*	*		*	*	*		*	*	*
P16. El estimador $\hat{\beta}_i$ es un estimador insesgado del coeficiente de regresión β_i en el modelo de regresión.	*	*	*	*	*	*			*	*	*

3.10. ARGUMENTOS

Los procedimientos, algoritmos, propiedades, definiciones y objetos matemáticos se enlazan entre sí mediante razonamientos o argumentos que son utilizados para comprobar las soluciones de los problemas, explicar y justificar la solución. Los tipos de argumentos ligados a los intervalos de confianza que aparecen en los libros de texto que hemos analizado son:

A1. *Mediante demostraciones formales*. Las demostraciones que aparecen hacen uso de definiciones, símbolos, propiedades de desigualdades, fórmulas de estandarización de la media muestral y algunas manipulaciones algebraicas. Mendenhall y Sincich (1997, pg. 353) presenta los siguientes argumentos para la obtención del intervalo de confianza (1-α)100% de muestra grande para un parámetro poblacional θ.

> Sea $\hat{\theta}$ un estadístico con una distribución de muestreo aproximadamente normal para muestras grandes con media $E(\hat{\theta}) = \theta$ y error estándar $\sigma_{\hat{\theta}}$. Entonces, $z = \dfrac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$ es una variable aleatoria normal estándar. Puesto que *z* también es una función de únicamente el estadístico de la muestra $\hat{\theta}$ y del parámetro θ, la utilizaremos como estadístico del pivote. A fin de obtener un intervalo de confianza para θ, primero necesitamos una expresión de probabilidad para el estadístico del pivote. Esto lo hacemos localizando los valores $z_{\alpha/2}$ y $-z_{\alpha/2}$ que ubican una probabilidad de $\alpha/2$ en cada cola de la distribución de *z* ; es decir, $P(z > z_{\alpha/2}) = \alpha/2$. Es evidente que $P(-z_{\alpha/2} \le z \le z_{\alpha/2}) = 1-\alpha$
> Si sustituimos la expresión para *z* en la expresión de probabilidad y aplicamos alguna de las operaciones algebraicas sencillas a la desigualdad, obtenemos
> $$P(-z_{\alpha/2} \le z \le z_{\alpha/2}) = P(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \le \theta \le \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha$$
> Por tanto, la probabilidad de que el intervalo formado por $LCI = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}$ a $LCS = \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}$ incluirá a θ es igual a (1-α). Las cantidades LCI y LCS se denominan límite de confianza inferior y límite de confianza superior, respectivamente para el intervalo de confianza. El coeficiente de confianza del intervalo será (1-α).

A2. *Simulación gráfica con el ordenador*. La tecnología electrónica ha invadido prácticamente todos los ámbitos de nuestras vidas, incluidos en éstos el campo de la Estadística. Los ordenadores, buenos para realizar operaciones que en ocasiones resultan tediosas, se magnifican cuando cuentan con paquetes estadísticos bajo los cuales se pueden correr simulaciones.

Un ejemplo proporcionado por Hayter (1996) exhibe una justificación del significado del término nivel de confianza a través de una simulación de 500 muestras de tamaño 30 de una población normal con media 10 y varianza 3; el intervalo de confianza de 95% es calculado para cada muestra y luego graficado para cada una de las 500 muestras. La ilustración gráfica que presenta el autor (pg. 383) exhibe algunos de los intervalos de confianza simulados.

A3. *Simulación gráfica sin usar el ordenador*. Un método alternativo con el cual se justifica el significado del término nivel de confianza a través de una simulación sin usar el ordenador es utilizando una tabla de dígitos aleatorios. Johnson y Kuby (2004, pg. 314) realizan una simulación de 15 muestras de tamaño 40, seleccionando para cada muestra 40 dígitos de una tabla de dígitos aleatorios (apéndice B), tablas de la RAND Corporation, que tienen un valor medio de 4.5 y una desviación estándar de 2.87. Calcula manualmente intervalos de confianza del 90%, con lo cual obtiene 15 intervalos de confianza, luego los grafica sin ayuda del ordenador.

A4. *Justificación mediante ejemplos o contraejemplos. (Justificación de casos especiales).* Un ejemplo de los razonamientos que presenta el Devore (2005, pg. 294) en la obtención de casos especiales de intervalos de confianza tomando como punto de partida un intervalo de confianza general basado en una muestra grande.

Suponga que $\hat{\theta}$ que satisface las propiedades siguientes: (1) tiene aproximadamente una distribución normal; (2) es (por lo menos de forma aproximada) insesgado, y (3) está disponible una expresión para $\sigma_{\hat{\theta}}$ la desviación estándar de $\hat{\theta}$. Por ejemplo, en el caso $\theta = \mu$, $\hat{\mu} = \overline{X}$ es un estimador insesgado cuya distribución es casi normal cuando *n* es grande y $\sigma_{\hat{\mu}} = \sigma_{\overline{X}} = \sigma/\sqrt{n}$. Al estandarizar $\hat{\theta}$ se obtiene la va $Z = \frac{(\hat{\theta}-\theta)}{\sigma_{\hat{\theta}}}$, que tiene aproximadamente una distribución normal estándar. Esto justifica el enunciado de probabilidad: $P(-Z_{\alpha/2} < \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} < Z_{\alpha/2}) \approx 1-\alpha$. Suponga, primero, que $\sigma_{\hat{\theta}}$ no requiere parámetros desconocidos (p. ej., σ conocida en el caso θ=μ). Entonces reemplazar cada < en (1) por = da como resultado $\theta = \hat{\theta} \pm Z_{\alpha/2}\sigma_{\hat{\theta}}$, así que los límites de confianza inferior y superior son $\hat{\theta} - Z_{\alpha/2}\sigma_{\hat{\theta}}$ y $\hat{\theta} + Z_{\alpha/2}\sigma_{\hat{\theta}}$ respectivamente. Ahora suponga que $\sigma_{\hat{\theta}}$ no tiene relación con θ, pero si con por lo menos otro parámetro desconocido (p. ej., $s/\sqrt{n}$ estima $\sigma/\sqrt{n}$). En condiciones generales (en esencia que $s_{\hat{\theta}}$ esté cerca de $\sigma_{\hat{\theta}}$ para la mayor parte de las muestras), un intervalo de confianza válido es $\hat{\theta} \pm z_{\alpha/2}(s_{\hat{\theta}})$. El intervalo $\overline{x} \pm z_{\alpha/2}\, s/\sqrt{n}$ es un ejemplo.

Por último, suponga que $\sigma_{\hat{\theta}}$ requiere el parámetro desconocido θ. Éste es el caso, por ejemplo, cuando θ=*p*, una proporción poblacional. Entonces puede ser difícil resolver $(\hat{\theta}-\theta)/\sigma_{\hat{\theta}} = z_{\alpha/2}$. Una solución aproximada se obtiene al sustituir θ en $\sigma_{\hat{\theta}}$ por su estimación $\hat{\theta}$. Esto da como resultado una desviación estándar estimada $s_{\hat{\theta}}$ y el intervalo correspondiente es de nuevo $\hat{\theta} \pm z_{\alpha/2}(s_{\hat{\theta}})$

A5. *Justificación mediante la relación de equivalencia con las pruebas de hipótesis.* Otro argumento que hemos encontrado es el razonamiento con el que se justifica la relación de equivalencia entre la prueba de hipótesis acerca de cualquier parámetro θ, y el intervalo de confianza para θ. Miller y cols. (1997, pg.244) lo presentan para el caso particular de $\theta = \mu$.

Consideremos el intervalo de confianza de 100(1-α)% para μ dado en la página 223:

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Este intervalo se relaciona estrechamente con un nivel α de prueba de $H_0 : \mu = \mu_0$ contra la alternativa de dos regiones $H_0 : \mu \neq \mu_0$. En términos de valores de $\bar{x}$ y s, esta prueba tiene una región crítica

$$\left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right| = |\,t\,| \geq t_{\alpha/2}$$

La región de aceptación de esta prueba se obtiene invirtiendo la desigualdad para obtener todos los valores de $\bar{x}$ y s que no conduzcan al rechazo de μ_0, es decir, la región de aceptación es $\left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right| < t_{\alpha/2}$ La región de aceptación también puede expresarse como: $\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu_0 < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$ donde los límites del intervalo son idénticos al intervalo de confianza anterior. Esto es, la hipótesis nula μ_0 no será rechazada en el nivel α si μ_0 se encuentra dentro del intervalo de confianza 100(1- α)% de μ.

En la tabla 3.7 presentamos los argumentos o justificaciones ligados directamente a los intervalos de confianza que aparecen en la muestra de libros que hemos revisado. En esta tabla observamos que las demostraciones formales, aunque aparecen en todos los libros, en tres de ellos: Miller, Freund y Johnson (1997), Johnson y Kuby (2004) y Velasco y Winiewski (2001) son técnicamente débiles en sus demostraciones, son mucho menos formales que el resto de los autores.

Tabla 3.7. Argumentos en los libros de texto

Justificaciones	Libro										
	A	B	C	D	E	F	G	H	I	J	K
AR1.Mediante demostraciones formales.	*	*	*	d	*	*	d	d	*	*	*
AR2.Simulación gráfica con el ordenador.		*	*	*	*			*			*
AR3.Simulación gráfica sin usar el ordenador.								*			
AR4.Justificación mediante ejemplos o contraejemplos. (Justificación de casos especiales).											*
AR5.Justificación mediante la relación de equivalencia con las pruebas de hipótesis.	*	*	*	*	*	*				*	

Nota: d = poco rigor en la demostración.

El argumento mediante simulación gráfica sin ayuda del ordenador sin muchos recursos contribuye a asimilar la idea intuitiva del resultado; esta justificación solamente aparece en Johnson y Kuby (2004). La justificación mediante la simulación gráfica usando el ordenador aparece apenas en el 54% de los libros; cuatro de ellos de la década pasada. La justificación mediante ejemplos o contraejemplos (Justificación de casos especiales) solamente la encontramos en Devore (2005).

En los años recientes se ha observado una tendencia de los especialistas en estadística a preferir la información disponible en un enunciado de intervalo de confianza en oposición a la información disponible en una prueba de hipótesis; en virtud de este hecho resulta importante que los libros de texto presenten la justificación de esa relación de equivalencia para que el alumno pueda reflexionar acerca de esta oportunidad que brindan los intervalos de confianza de hacer inferencia estadística evitando aquellos elementos que provocan conflictos al usarse pruebas de hipótesis, como por ejemplo el valor *p* (Davies, 1998). La justificación mediante la relación de equivalencia con las pruebas de hipótesis se presenta en siete de los 11 libros.

Finalmente podemos observar que de las 55 celdas (5 argumentos x 11 libros) que contiene nuestra tabla 3.7, solamente 26 celdas están ocupadas, por lo que nuestra tabla confirma lo que Ortiz (1998) señala que *se ha descuidado la actividad de formación de argumentos matemáticos porque posiblemente los alumnos no realizan muchas actividades que impliquen la producción de sus propios textos matemáticos*.

3.11. CONCLUSIONES DEL ANÁLISIS DE CONTENIDO

En este capítulo hemos realizado un análisis de los elementos del significado de los intervalos de confianza en una muestra de libros de texto dirigidos a ingenieros, siguiendo la metodología de otros estudios anteriores, como Alvarado (2007). Algunas conclusiones obtenidas del estudio son las siguientes:

Hemos encontrado una gran variedad de campos de problemas, que se han clasificado según se trata de la estimación de una sola variable, de la comparación de varias muestras, contrastes de hipótesis, cálculo de tamaño de muestra o límites de tolerancia. En cada caso se han diferenciado subcampos de problemas, en función de los parámetros a estimar o de las condiciones de partida. Muchos de estos subcampos comparten objetos matemáticos similares, por lo que podrían agruparse en configuraciones epistémicas relativas al intervalo de confianza.

Algunos de los algoritmos constituyen a su vez campos de problemas relativos a otros objetos matemáticos que el alumno debería tener apropiados. Se muestra así el carácter recursivo del modelo teórico utilizado en este trabajo.

Hemos encontrado una gran variedad de lenguaje en sus categorías de lenguaje verbal, simbólico y gráfico, así como la simulación asociada al uso de ordenador.

Asimismo la variedad de propiedades asociadas al intervalo u objetos relacionados, y los argumentos confieren a este objeto de una gran complejidad.

Hacemos también notar que hemos encontramos un tratamiento muy desigual del tema en los libros de texto, lo cual supone la diversidad de significados personales que podrían adquirir los estudiantes, según se utilizase uno u otro texto en la enseñanza.

Finalmente, el estudio permite cumplir uno de los objetivos del trabajo al fundamentar la definición semántica de la variable “comprensión de los intervalos de confianza”, mediante una definición precisa y sistemática de las unidades de contenido que se abordará en el capítulo 4 y que se basará en el análisis de contenido de los libros de texto presentados en este capítulo.

CAPITULO 4
CONSTRUCCIÓN DEL CUESTIONARIO

4.1.INTRODUCCIÓN

Describimos en este capítulo los pasos que se han seguido hasta llegar a la construcción del cuestionario y que son los siguientes: definir la variable a partir del análisis de contenido (descrito en el capítulo 3), elaborar una tabla de especificaciones de nuestro instrumento de medición; recopilar, traducir y depurar de un banco inicial de ítems; llevar a cabo una serie de pruebas piloto con los ítems y seleccionar los ítems que conformarán la versión definitiva del cuestionario que pretendemos elaborar, a través de dichos ensayos piloto, así como a partir de la valoración de los ítems mediante juicio de expertos.

Todo ello siguiendo el método de Díaz (Díaz y de la Fuente, 2005, 2006, 2007a y b; Díaz, 2007) en su construcción de un cuestionario de evaluación del razonamiento condicional dirigido a estudiantes de Psicología. Seguiremos los pasos de la autora, quien se basa en las recomendaciones metodológicas de APA, AERA y NCME (1999).

4.2.OBJETIVOS Y CLASIFICACIÓN DEL INSTRUMENTO

La necesidad de recoger información válida y fiable es el punto nuclear de la decisión de elaborar un instrumento efectivo de evaluación (Peterson, 2000) y el objetivo para el que será empleado el instrumento determina sus características. Como primer paso en el desarrollo de un instrumento efectivo de evaluación, hemos determinado su propósito y la naturaleza de las inferencias que vamos a extraer de los puntajes obtenidos por la muestra de sujetos a los que les aplicaremos el instrumento. Un acotamiento claro de los objetivos nos pondrá en condiciones de establecer un marco global para las especificaciones de la variable, así como para el desarrollo, prueba y revisión de los ítems (Sax, 1989).

En nuestro intento por evaluar la comprensión sobre un cierto objeto matemático de un grupo de alumnos, hemos de tener en cuenta que la comprensión es un *constructo inobservable* (León y Montero, 2002), por lo que al tratar de evaluarla, sus características deben ser inferidas a partir de las respuestas de los alumnos. Se trata de establecer un conjunto de reglas para asignar a los sujetos números que representan cantidades de atributos (Castro Posada, 2001). Estas reglas empleadas deben estar exentas de ambigüedad. Pueden basarse en la experiencia previa, surgir del sentido común o simplemente tener origen en corazonadas, pero el punto crucial es la manera en que los usuarios concuerdan de manera consistente en la medida (Nunnally y Bernstein, 1995). Por medio de las preguntas planteadas a los encuestados, pretendemos obtener una estimación de conocimientos y capacidades de los sujetos, que no son accesibles por simple observación o encuesta, ya que los mismos sujetos no son conscientes de ellas plenamente (Fox, 1981; Dane, 1990).

En nuestro trabajo, nos ubicamos en un dominio curricular, que se refiere a un conjunto de habilidades y conocimientos que se desarrollan como resultado de la instrucción reflexionada sobre un cierto contenido curricular, siguiendo la clasificación de Millman y Green (1989), quienes detallan los principales objetivos en la construcción de cuestionarios, diferenciando entre posibles escenarios de decisiones educativas y tipo de inferencias deseadas.

El objetivo principal de nuestro instrumento de evaluación es hacer diagnóstico sobre la comprensión que tienen del concepto intervalos de confianza, los alumnos de tercer semestre del curso de Probabilidad y Estadística en las carreras de ingeniería ofrecidas en el Sistema Tecnológico de Monterrey. Nuestro instrumento de medida podría también ser aplicado a otras poblaciones dentro del Sistema como por ejemplo alumnos que cursan Estadística para Ciencias Sociales o Bioestadística en las carreras de Ciencias de la Salud. O bien ampliarnos a otras comunidades estudiantiles en otras universidades.

Un segundo objetivo es diferenciar, a partir de las puntuaciones en el cuestionario, a los alumnos que tengan buena comprensión del concepto de intervalos de confianza de los que no lo tengan y proporcionar esta información a los distintos profesores para que puedan preparar algunas medidas remediales. Este cuestionario nos permitirá así tomar decisiones sobre posibles acciones educativas dirigidas a estudiantes particulares con referencia a los intervalos de confianza.

En estos dos primeros objetivos las decisiones que se tomarían, usando la información de las puntuaciones en el cuestionario, son relativas a nivel de instrucción.

En este tipo de decisiones es fundamental si un sujeto o grupo de sujetos domina o no ciertas destrezas predeterminadas (Martínez Arias, 1995). Durante este proceso, también tomaremos decisiones de diagnóstico, cuyo objetivo es identificar algunas dificultades y errores particulares en el rendimiento de un sujeto o grupo de sujetos. El análisis relativo a los significados personales de los estudiantes y de los conflictos semióticos que frenan su evolución, permitirá en el corto plazo, llevar a cabo una intervención educativa apropiada.

Clasificación del instrumento

Siguiendo a Díaz (2007), consideramos que nuestro cuestionario es un instrumento de medida *centrada en el sujeto* (Millman y Green, 1989), pues las puntuaciones en el cuestionario nos van a permitir ordenar a los estudiantes en el continuo "comprensión del intervalo de confianza", entendida la comprensión como acoplamiento del significado personal del estudiante con el significado institucional de referencia para su enseñanza.

Trabajaremos con un test referido a criterio, pues las interpretaciones que realicemos del cuestionario se refieren a un dominio específico, que hemos delimitado en el análisis de los libros de textos. Hacemos referencia a lo que los alumnos hacen o son capaces de hacer y sus conocimientos o errores sobre el tema (Sax, 1989).

De ahí que no se compararán los resultados con respecto a un grupo normativo. En el mejor de los casos, dentro del grupo de estudiantes se diferenciarán a los alumnos con más baja puntuación en el cuestionario para así poder diseñar en el corto plazo, una instrucción personalizada a estos alumnos. Debido a que el tiempo no determina el resultado, el *cuestionario es de potencia*, las diferencias en la puntuación entre los sujetos que respondan al cuestionario serán atribuibles a la calidad de su ejecución y conocimiento (Martínez Arias, 1995).

Desde la perspectiva de esta autora se trata de un test psicométrico porque evalúa las respuestas en forma cuantitativa y se refiere a un rasgo diferenciado del sujeto. Su finalidad es doble: investigación y diagnóstico. Se trata de un test de ejecución máxima (en cada ítem el sujeto pone en funcionamiento su capacidad) es de rendimiento académico y su construcción la basamos en la teoría clásica de los tests, que ve a la medición como la determinación de la cantidad o cuánto de un atributo está presente en un objeto a través de la suma, quizá ponderada, de respuestas a reactivos individuales (es decir, como una combinación lineal).

En relación a la decisión entre precisión en la puntuación y amplitud de contenido evaluado en el cuestionario, se requiere llegar a un compromiso, que está en función del tiempo disponible para que los estudiantes respondan. Mientras que un test de contenido muy uniforme se gana en fiabilidad, se pierde en validez o en cobertura de contenidos. En los tests referidos a criterio usualmente el contenido es más amplio y la fiabilidad podría ser menor. Esto está relacionado con la dimensionalidad del contenido del instrumento que se refiere a la homogeneidad o heterogeneidad teórica del contenido (Magnusson, 1990).

Podríamos considerar, en nuestro caso, el cuestionario como unidimensional, en el sentido que, derivaremos una puntuación total al mismo. Pero, es posible que, por la variedad de tipos de problemas y propiedades identificados en el estudio teórico, los resultados empíricos exhiban más de una dimensión y, en principio, nuestra prioridad es reflejar el contenido del dominio.

Todas estas consideraciones, afectarán a las decisiones que tomemos en relación al cuestionario.

4.3.DEFINICIÓN DE LA VARIABLE

El paso inicial que hemos dado para la construcción de nuestro instrumento es definir lo que vamos a medir. Como hemos indicado en este instrumento tratamos de medir el constructo "comprensión del concepto de intervalo de confianza". En este apartado describimos la metodología empleada para definir esta variable, que sigue la descrita en Díaz (2007).

4.3.1. FUNDAMENTOS DE LA DEFINICIÓN

Un constructo es un atributo para explicar un fenómeno (Wiersma, 1986). Es un atributo psicológico que caracteriza los comportamientos de los individuos y nos permite explicar patrones de comportamiento que sólo pueden ser observados indirectamente, están sujeto al cambio y sólo los comprendemos vagamente (Ghiselli, Campbell y Zedeck, 1981). Esto explica la dificultad de su evaluación que llevamos a cabo mediante alguna *variable* observable, por ejemplo, la puntuación en un cuestionario (Osterlind, 1989). La medición de constructos presenta las siguientes dificultades:

- No suele haber una única aproximación a la medida de un constructo.

- Las medidas psicológicas están basadas en muestras limitadas de conductas.
- La medida siempre tiene un error, ligado al muestreo de tareas, ocasiones o situaciones.
- Las escalas no tienen un origen ni unidades de medida bien definidos (Martínez Arias, 1995).

Según esta autora, la definición del constructo se describe a continuación y se realiza a dos niveles:

- *Definición semántica*: en términos de comportamientos observables o reglas de correspondencia entre el constructo y la conducta.
- *Definición sintáctica*: en términos de las relaciones lógicas o matemáticas del constructo con otros constructos o variables dentro de un marco teórico.

Nuestro paso inicial ha sido elaborar un procedimiento para identificar el constructo, mediante una definición semántica precisa, siguiendo el procedimiento de Díaz (2007). Al igual que en dicho trabajo, y para dotar de una mayor objetividad a la definición de nuestra variable, nos hemos basado en un análisis de contenido en una muestra de libros de texto, en nuestro caso los que utilizan los alumnos de Ingeniería en los campus del Sistema Tecnológico de Monterrey en la asignatura de probabilidad y estadística. De este modo conseguiremos una adecuada perspectiva sobre la enseñanza del tema, como sugiere Selander (1990).

4.3.2. RESULTADOS ANÁLISIS DE CONTENIDO

En el capítulo tres se describió con detalle el análisis de contenido del tema dedicado al intervalo de confianza, en una muestra de libros de texto. El análisis se apoyó en el marco teórico descrito en el capítulo dos y, específicamente, se determinaron los principales campos de problemas, lenguaje, procedimientos, definiciones, propiedades y argumentos presentados en este tema en los libros. Con todo ello determinamos el significado de referencia del objeto "intervalo de confianza" en esta investigación.

Los resultados se presentaron en forma de tablas y serán la base de la definición semántica de la variable "comprensión del intervalo de confianza" que se realiza en esta sección.

Fiabilidad y validez

Dos atributos esenciales son la validez y fiabilidad. Respecto a la validez en el análisis de contenido, Fox (1981) indica que para conseguirla, los temas y las categorías de análisis deben tener una relación directa con la finalidad para la que se han creado. Esto se observa claramente en nuestro caso, donde hemos recurrido a un juicio externo de expertos (que se describe en la sección 4.6.2) para valorar la pertinencia de las categorías elaboradas (especificaciones de contenidos) a la finalidad del instrumento, eliminándose aquellas categorías en las que no se alcanzó acuerdo en este proceso.

La fiabilidad del análisis de contenido depende del procedimiento de codificación y de la ambigüedad de las categorías (Ghiglione y Matalón, 1989). Metodológicamente se estima calculando el porcentaje de veces que dos codificaciones independientes coinciden cuando codifican el mismo material.

En nuestro caso, las tablas resúmenes de datos fueron revisadas dos veces por el autor del trabajo y por otro investigador, para comprobar la ausencia o presencia en los libros de texto de determinadas categorías, procediéndose a la corrección de los pocos desacuerdos con la codificación original. Pensamos que con este procedimiento puede garantizarse una adecuada fiabilidad en el proceso de codificación de los datos a partir de las categorías definidas.

4.4. TABLA DE ESPECIFICACIONES

Nuestro trabajo, está orientado a la definición semántica de nuestro constructo, que considera la especificación detallada de la variable de interés (en este caso la comprensión de los intervalos de confianza). Los tipos de prácticas operatorias y discursivas (Godino, 2003) asociadas al objeto “intervalo de confianza”, contempladas en nuestro marco teórico del enfoque ontosemiótico, son encapsuladas a través de las especificaciones del contenido. La diferenciación entre constructos y variables se recoge en la teoría de las funciones semióticas (TFS), puesto que se diferencia entre el dominio de las ideas u objetos abstractos (personales e institucionales) y el dominio de los significados o sistemas de prácticas de donde emergen tales objetos inobservables lo que permite plantear con claridad la dificultad del problema de la evaluación.

La evaluación debe comprender ítems que representen los objetivos claramente delimitados (Sax, 1989). Una forma efectiva de asegurar una adecuada representación de ítems es desarrollar una tabla de especificaciones (Tabla 4.1). Según Martínez Arias (1995) la tabla debe contener tres elementos básicos: contenido, procesos e importancia de estos procesos. En una situación de construcción de un test orientado a criterio, como es nuestro caso, esta tabla debe especificar qué contenidos principales del dominio se contemplan en el cuestionario. Las especificaciones del cuestionario nos ayudan a describir en forma completa y fácil la estructura y contenido del test, marcando con esto el primer paso en un enfoque sistemático, en que cada tarea debe ser secuencialmente y exitosamente completada antes de moverse al paso siguiente.

Esta tabla nos permite estructurar el constructo que queremos medir, indicando los comportamientos que lo evidencian, las mediciones en que se agrupan y la importancia específica relativa de cada uno de los comportamientos, que se traduce en el número de ítems utilizados para medir cada uno de los comportamientos.

Como resultado del análisis realizado de los libros recomendados en la asignatura de estadística para ingenieros en el Sistema Tecnológico de Monterrey y la revisión bibliográfica de las investigaciones que se centran alrededor de la enseñanza y aprendizaje de los intervalo de confianza, hemos identificado las siguientes conductas observables que utilizaremos como indicadores de la comprensión del concepto.

Elementos conceptuales y proposicionales de significado

1. Ser capaz de dar una definición correcta de intervalos de confianza. En esta definición hemos identificado tres diferentes categorías: a) Definición a partir de la fórmula de cálculo, interpretando correctamente el significado (aparece en ocho libros), b) Definición a partir de la fórmula de cálculo, interpretando en forma incorrecta o ambigua el significado (aparece en forma ambigua en dos libros), c) Definición como intervalo con extremos que dependen del valor de una variable aleatoria (aparece en siete libros) y d) Definición de intervalo de confianza usando el método de inversión del estadístico de prueba (aparece en dos de los libros consultados).
2. Reconocer que el ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra. Aparece explícitamente en nueve de los libros analizados.
3. Reconocer que el ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta. Aparece explícitamente en ocho de los libros analizados.
4. Reconocer que el ancho de los intervalos de confianza aumenta cuando aumenta la varianza. Aparece explícitamente en siete de los libros analizados.

5. Reconocer el significado del nivel de confianza (variación del intervalo en diferentes muestras) Aparece en casi todos los libros analizados.
6. Reconocer que la proporción muestral usada para estimar la proporción poblacional sigue una distribución normal, si *n* es grande. Aparece explícitamente en todos los libros analizados.
7. Reconocer que la media de una muestra suficientemente grande sigue una distribución aproximadamente normal. Aparece en todos los libros analizados.
8. La diferencia de medias muestrales en dos poblaciones sigue una distribución normal en muestras suficientemente grandes. Aparece en todos los libros analizados.
9. Reconocer que la media muestral es un estimador insesgado de mínima varianza. Aparece en todos los libros analizados.
10. Reconocer que la cuasivarianza muestral es un estimador insesgado de la varianza poblacional. Aparece prácticamente en todos los libros analizados.
11. La proporción muestral es un estimador insesgado de la proporción poblacional y de mínima varianza. Aparece en todos los libros analizados.
12. Reconocer que la diferencia de medias muestrales en dos poblaciones, usada para estimar $\mu_1 - \mu_2$ es un estimador insesgado de mínima varianza para la diferencia de medias en las poblaciones. Aparece en ocho de los libros analizados.

Elementos situacionales de significado

13. Estimar la media de una población normal o en una muestra grande con σ conocida Aparece en todos los libros analizados.
14. Estimar la media de una población aproximadamente normal cuando, σ es desconocida. Aparece en todos los libros analizados.
15. Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande. Aparece en tres de los libros analizados
16. Estimar una proporción. Aparece en todos los libros analizados.
17. Estimar una varianza. Aparece en nueve de los libros analizados.
18. Comparar las medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes. Aparece en todos los libros analizados.
19. Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas. Aparece en todos los libros analizados.
20. Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ desconocidas, muestras

independientes grandes. Aparece en cuatro de los libros analizados.

21. Resolver correctamente problemas de estimación intervalos de confianza para comparar dos varianzas poblacionales. Aparece en seis de los libros analizados.
22. Resolver correctamente problemas de contrastar una hipótesis sobre una sola media poblacional a través de la relación de equivalencia con el intervalo de confianza. Aparece en siete de los libros analizados.
23. Resolver correctamente problemas de calcular un tamaño adecuado de muestra, de modo que se obtenga un error dado de estimación con una cierta probabilidad. Aparece en prácticamente todos los libros analizados.

Elementos procedimentales de significado

24. Elegir un modelo de distribución muestral del estadístico. Aparece en todos los libros analizados.
25. Determinar valores críticos en la distribución del estadístico. Aparece en todos los libros analizados.
26. Obtención de intervalos de confianza utilizando un programa de ordenador Aparece en siete de los libros analizados.

Elementos linguísticos de significado

27. Expresiones. Aparece en todos los libros analizados.
28. Símbolos. Aparece en todos los libros analizados.
29. Gráficos. Aparece en prácticamente todos los libros analizados.

Elementos argumentativos de significado

30. Mediante demostraciones formales. Aparece en ocho de los libros analizados.
31. Simulación gráfica con el ordenador. Aparece en seis de los libros analizados.
32. Justificación mediante la relación de equivalencia con las pruebas de hipótesis. Aparece en siete de los libros analizados.

El análisis de contenido y la revisión bibliográfica de las investigaciones previas, los canalizamos en la elaboración de la tabla de especificaciones de nuestro cuestionario, que se presenta en la Tabla 4.1. Para los diferentes elementos de significado considerados por Godino (2003), analizaremos su comprensión y las relaciones que se establecen (configuraciones personales). Las situaciones problema, las propiedades, procedimientos y el lenguaje serán evaluados con ítems en formatos de opción múltiple, y algunos elementos de significado, como por ejemplo los argumentos, serán evaluados en los ítems de respuesta abierta, que analizaremos, para estudiar también, la comprensión de este tipo de elementos.

Tabla 4.1. Especificaciones del contenido del cuestionario

Tipo	Principales áreas de contenido
Defin.	1. Definición de intervalo de confianza.
Propiedades	2. El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra.
	3. El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta.
	4. El ancho de los intervalos de confianza aumenta cuando aumenta la varianza.
	5. Significado del nivel de confianza (variación del intervalo en diferentes muestras).
	6. La proporción muestral usada para estimar la proporción poblacional sigue una distribución normal, si n es grande.
	7. La media de una muestra suficientemente grande sigue una distribución aproximadamente normal.
	8. La diferencia de medias muestrales en dos poblaciones sigue una distribución normal en muestras suficientemente grandes.
	9. La media muestral es un estimador insesgado de mínima varianza.
	10. La cuasivarianza muestral es un estimador insesgado de la varianza poblacional.
	11. La proporción muestral es un estimador insesgado de la proporción poblacional y de mínima varianza.
	12. La diferencia de medias muestrales en dos poblaciones, usada para estimar $\mu_1 - \mu_2$, es un estimador insesgado de mínima varianza.
Campos de problemas	13. Estimar la media de una población normal o en una muestra grande con σ conocida.
	14. Estimar la media de una población aproximadamente normal cuando, σ es desconocida.
	15. Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande.
	16. Estimar una proporción.
	17. Estimar una varianza.
	18. Comparar las medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes.
	19.Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas.
	20.Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ desconocidas, muestras independientes grandes.
	21. Resolver correctamente problemas de estimación intervalos de confianza para comparar dos varianzas poblacionales.
	22. Resolver correctamente problemas de contrastar una hipótesis sobre una sola media poblacional a través de la relación de equivalencia con el intervalo de confianza.
	23. Resolver correctamente problemas de calcular un tamaño adecuado de muestra de modo que se obtenga un error dado de estimación con una cierta probabilidad.
Proced.	24. Elegir un modelo de distribución muestral del estadístico.
	25. Determinar valores críticos en la distribución del estadístico.
	26. Obtención de intervalos de confianza utilizando un programa de ordenador.
Leng-	27. Expresiones.
	28. Símbolos.
	29. Gráficos.
Arg.	30. Mediante demostraciones formales.
	31. Simulación gráfica con el ordenador.
	32. Justificación mediante la relación de equivalencia con las pruebas de hipótesis.

Para decidir cuales especificaciones del contenido eran relevantes para los propósitos del instrumento, así como para reforzar los resultados obtenidos del análisis de contenido de los libros de texto, recurrimos al juicio de expertos.

Siguiendo a Millman y Green (1989) y Díaz (2007) el objetivo que perseguimos al acudir al juicio de expertos fue establecer un consenso de opiniones entre ellos sobre cómo cada ítem particular se ajusta adecuadamente para evaluar el contenido específico para el cuál ha sido diseñado y que sirviesen como base para elegir los ítems definitivos.

En esta tabla podemos apreciar que han resultado 32 especificaciones o conductas observables. Sin embargo algunas unidades de contenido están incluidas en otras, con lo cual se pueden suprimir para simplificar el proceso de elaboración del cuestionario. Por ejemplo, los argumentos, lenguajes y procedimientos se pueden evaluar en las respuestas del alumno al resolver los reactivos de preguntas abiertas. Además, los contenidos seis a 12 aunque son prerrequisitos, no están estrictamente relacionados con el intervalo de confianza y están incluidos en los contenidos 13 a 23. Por ejemplo, para resolver un problema del contenido 23 (estimar una media) el alumno ha de usar el contenido siete y nueve. Asimismo los contenidos 22 y 23 no son estrictamente relacionados al tema del intervalo de confianza.

Tabla 4.2. Especificaciones del contenido del cuestionario

Tipo	Principales áreas de contenido
Def.	1. Definición de intervalo de confianza.
Propiedades	2. El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra.
	3. El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta.
	4. El ancho de los intervalos de confianza aumenta cuando aumenta la varianza.
	5. Significado del nivel de confianza (variación del intervalo en diferentes muestras).
Campos de problemas	6. Estimar la media de una población normal o en una muestra grande con σ conocida.
	7. Estimar la media de una población aproximadamente normal cuando, σ es desconocida.
	8. Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande.
	9. Estimar una proporción.
	10. Estimar una varianza.
	11. Comparar las medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes.
	12. Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas.
	13. Comparar las medias en dos poblaciones, con varianzas desconocidas, muestras independientes grandes.
	14. Comparar dos varianzas poblacionales.
Proc.	15. Elegir un modelo de distribución muestral del estadístico.
	16. Determinar valores críticos en la distribución del estadístico.
	17. Interpretar intervalos de confianza obtenidos de un programa de ordenador.
	18. Interpretar gráficos de intervalos de confianza.

Por todo ello, se decidió reducir los contenidos a los más esenciales y evitar los que estaban incluidos en los de tipo procedimental (pues quedarían evaluados en los ítems de respuesta abierta). Todo ello con la finalidad de que el cuestionario no sea excesivamente largo, con lo cual la nueva tabla que considera estas inclusiones sería la Tabla 4.2. Esta es la tabla que se toma como definición semántica del constructo, aunque después de realizar el juicio de expertos pudiera cambiar.

4.5. ELABORACIÓN DE UN BANCO DE ÍTEMS

Según Osterlind (1989), un ítem de un cuestionario es una unidad de medida que consta de un estímulo y una forma prescriptiva de respuesta. Se utiliza para evaluar la capacidad del examinado en un cierto constructo proporcionando datos cuantificables sobre la persona que lo completa. En la selección de ítems hemos tratado de apegarnos a los criterios de calidad habituales sugeridos por el autor:

- Congruencia entre el ítem particular y el objetivo del cuestionario, en un alto grado, puesto que es determinante para poder lograr la validez del cuestionario y acoplar con nitidez el ítem con un objetivo claramente definido. Estos dos criterios se han garantizado mediante la técnica del juicio de expertos (que describiremos en otro apartado).
- Reducir al máximo posible, la contribución de cada ítem al error de la medida de puntuación en el test. Para lograrlo revisaremos algunos ítems que resulten inadecuados en las pruebas piloto (una acción pudiera ser, pasar de opciones múltiples a respuestas abiertas) y no tomar en cuenta aquellos ítems que tengan una excesiva dificultad. También estudiaremos la contribución de cada ítem al error de medida, como parte del estudio de la confiabilidad del instrumento.
- Consistencia entre los fines del cuestionario y el formato de los ítems. Para ello hemos considerado dos tipos de formato: de opciones múltiples y respuesta abierta; los primeros son adecuado para evaluar los conocimientos conceptuales y los segundos para evaluar conocimientos procedimentales y argumentos.
- Nuestros ítems, además, satisfacen los criterios de homogeneidad (Magnusson, 1990), determinados a través de la independencia local (la respuesta a un ítem, no tiene ninguna relación con la respuesta a otro ítem), unidimensionalidad (todos los ítems se refieren al constructo intervalo de confianza).

Número de ítems

El número total de ítems contemplados, intenta cubrir en forma adecuada el contenido y asegurar una fiabilidad satisfactoria (Millán y Green, 1989), considerando la restricción de la longitud total posible del test para que los estudiantes tengan tiempo suficiente para responderlo con comodidad. En la decisión de cuantos ítems elegir, debemos tomar en cuenta la amplitud del contenido (López Feal, 1986), la estructura homogénea o heterogénea del contenido, su división o no en componentes, la forma de aplicación, características de los sujetos, etc.

Hemos considerado dar un peso igual a los diferentes contenidos y elaboraremos un cuestionario con un ítem por contenido. El propósito es disponer de un cuestionario de 18 ítems que pudiese ser respondido en un tiempo aproximado de hora y media. Para la prueba piloto, el número de ítems habrá de ser mayor que para la prueba definitiva, para que después de todas las pruebas, terminásemos con un número de ítems que nos permitiera completar el cuestionario.

Tipos de ítems considerados

Para la redacción del grupo inicial de ítems consideramos tanto los objetivos del proyecto de investigación, naturaleza de la información que va a ser recolectada, y la población a la que aplicaremos el cuestionario, así como las limitaciones de tiempo disponible (Peterson, 2000). En lo que sigue describimos estos requisitos y la manera en que los hemos tomado en cuenta (Millman y Green, 1989):

Características de la población a la que se va a administrar el cuestionario: para cuestionarios de habilidad y ejecución, las características más relevantes a tener en cuenta en la población son sus capacidades cognitivas y conocimientos. En nuestro estudio la población de interés son estudiantes inscritos en la asignatura de estadística para ingenieros. La formación previa de estos alumnos es ligeramente diversa. E intervalo de confianza se ha estudiado previamente durante el curso en que se pasa el cuestionario, dedicando al tema cinco sesiones. Otro supuesto es que el cuestionario final será utilizado al terminar de estudiar el tema de intervalos de confianza por los alumnos de esta población.

Tipo de administración: el cuestionario será administrado de forma grupal, con el fin de hacer eficiente el proceso. Para lograr esta eficiencia el cuestionario deberá diseñarse con instrucciones muy claras y fáciles de seguir. Sobre todo en un ambiente en donde la interacción entre el administrador del cuestionario y las personas que responden a él es menor que en administraciones individuales.

Limitaciones temporales en la administración: al construir un cuestionario se debe balancear entre un número suficiente de ítems (y de dificultad razonable) para el objetivo del cuestionario y el tiempo disponible (considerando además el factor fatiga de las personas que responden el cuestionario). En nuestro caso particular, no consideramos oportuno excedernos de una hora de administración del cuestionario. Decidimos que el cuestionario contuviera un número moderado de ítems y proporcionamos un tiempo de hora y media para evitar que algunos reactivos queden sin contestar por falta de tiempo.

Formato de los ítems

Hemos utilizado el formato de opciones múltiples y respuesta abierta en nuestro cuestionario. Como señalan Nunnally y Bernstein (1995, pg. 331) *"la naturaleza del material que será evaluado obviamente desempeña un papel importante en la elección"* Por otro lado Millman y Green (1989, pg. 143) indican:

> "*Algunos instrumentos pueden contener varios formatos diferentes*". "*Las decisiones sobre el tipo de ítem a usar se deben apoyar en la conceptualización de los componentes específicos del dominio de rendimiento o habilidad, tal como se refleja en las especificaciones del contenido. Estas especificaciones podrían sugerir más de un tipo particular de ítem*".

Cada cuestionario se confecciona de acuerdo a distintas necesidades. Mientras unos cuestionarios contienen sólo ítems abiertos o cerrados, otros cuestionarios contienen ítems de los dos tipos (Hernández, Fernández y Baptista, 1998). Entonces al construir un cuestionario se debe determinar:

a) Los elementos estructurales de los ítems que son necesarios para hacer que surjan las habilidades cognitivas y procesos identificados en la tabla de especificaciones del test. La preocupación principal es identificar los ítems que mejor generarán las muestras apropiadas de los componentes especificados en el contenido del dominio.
b) Qué tipo de ítems incluyen estos elementos estructurales.
c) Seleccionar, entre los distintos tipos de ítems, el formato preferido por ajuste, precisión y economía a la población. Aún en los casos en donde al emparejar el

formato del ítem con el contenido del cuestionario pueda haber conflictos con otros aspectos, tales como facilidad de codificación o fiabilidad, no se debe sacrificar la validez a ninguno de estos requisitos (Millman y Green, 1989).

Debido a que nuestro test es de ejecución óptima (cuestionario de rendimiento), consideramos adecuado incluir una parte de los ítems en un formato de opción múltiple. En los ítems en un formato de opción múltiple se proporciona la respuesta correcta y se pide al examinado elegir una de las opciones. Este tipo de ítems tiene tres elementos: 1) un tronco (introducción), 2) una alternativa u opción correcta y 3) varias alternativas u opciones incorrectas a las que se les denomina en forma variada "distractores" (Cohen y Swerdlik, 2001).

Esto nos facilita asignar puntuaciones a los sujetos e identificar patrones de respuesta erróneos. Además esta elección se justifica porque tomaremos los ítems de investigaciones previas donde se ha logrado obtener una "fotografía" detallada de los razonamientos de los estudiantes. Esto nos coloca en posición de predecir las dificultades y errores más frecuentes y tenerlas en cuenta en los distractores que diseñaremos en los nuevos ítems.

Otras ventajas son que estas preguntas requieren menos esfuerzo físico y mental, existe un menor "sesgo de campo" que aparece por la manera en que los entrevistadores registran las respuestas (Peterson, 2000). Además existe la familiaridad con este formato de ítem que ya tienen los estudiantes que responderán a nuestro cuestionario, ya que muchos de los exámenes que resuelven nuestros alumnos de ingeniería los realizan en el formato de opción múltiple. También este tipo de ítems se puede adaptar a una variedad de objetivos en educación, favorecen una interpretación precisa de la respuesta que puede constituir evidencia para la validez del contenido del cuestionario. Su formato es flexible y fácil de usar (Osterlind, 1989).

Díaz (2007) indica que en un ítem de opción múltiple, el número de respuestas usualmente varía entre tres y cinco, aunque no hay consenso sobre el número de alternativas que se pudiera considerar óptimo. Un número pequeño de alternativas, favorece el error aleatorio, aunque otros autores señalan que es preferible un mayor número de ítems con menos alternativas cuando se desea aumentar la fiabilidad (Martinez Arias, 1995). Lo que es inobjetable es que no se debe forzar una alternativa más, cuando esto no es natural.

Nunnally y Bernstein (1995) señalan los siguientes principios para el caso de ítems de opción múltiple:

- Asegurarse de que el tronco formula con claridad el problema.
- Incluir solo lo que es necesario en el tronco.
- Asegurarse que los distractores (alternativas incorrectas) son plausibles.
- Usar con moderación las expresiones “ninguna de las anteriores” o “todas las anteriores”, si es que se usan. En el caso de “todas las anteriores”, un estudiante que está bastante seguro de que dos alternativas son correctas puede elegir “todas las anteriores” sin saber si las alternativas adicionales son correctas también.
- Hacer cada alternativa aproximadamente de igual longitud y construcción gramatical paralela.
- Aleatorizar la ubicación de la alternativa correcta.
- Hacer cada alternativa aproximadamente de igual longitud y construcción gramatical paralela.
- Formular las alternativas incorrectas de modo que detecten errores comunes que los estudiantes puedan cometer.
- En general, tratar de eliminar cualquier factor que haga que la alternativa correcta se destaque para una persona no informada. Lo ideal es hacer que las alternativas le parezcan igualmente atractivas a dichas personas.

Nuestros ítems contienen entre tres y cuatro alternativas y cada alternativa fue diseñada teniendo en cuenta estos principios y recomendaciones. Incluiremos también algunos ítems abiertos donde la respuesta es suministrada por el examinado en oposición a sólo seleccionarla. Más específicamente los ítems de respuestas abiertas que incluiremos son problemas simples sobre intervalos de confianza. Estos ítems además de brindarnos la solución final, nos permitirán acceder a los algoritmos y procedimientos empleados en su solución, posibles expresiones algebraicas, representaciones gráficas y argumentos.

Podremos también asignar una puntuación basada en qué tan completa y precisa ha sido la respuesta. En muchas situaciones de evaluación en las que se desea tener información del conocimiento parcial de los examinados es deseable asignar un puntaje parcial. La ventaja más importante es que puede contribuir a aumentar la fiabilidad de un instrumento de longitud restringida, que es nuestro caso (Millman y Green, 1989).

Nivel de dificultad

La decisión acerca del nivel de dificultad de los ítems, es otra cuestión que debemos considerar en la planificación del cuestionario. Puesto que nuestro cuestionario es de referencia a un dominio, no es indispensable especificar por anticipado el nivel de dificultad de los ítems, ya que el contenido está especificado muy concretamente por el dominio. Nuestro interés está, por lo tanto, en incluir ítems concretos, incluso aunque estos ítems sean difíciles, puesto que nuestra intención es evaluar la dificultad real de conceptos seleccionados en forma deliberada. En estas condiciones, la dificultad del ítem tiene que ver más con la capacidad del examinado, que una determinada característica del ítem (Millman y Green, 1989).

Selección inicial de ítems

Hemos realizado una revisión de diversas investigaciones acerca de intervalos de confianza y hemos recolectado los ítems que se han utilizado para conformar nuestro banco de ítems, que posteriormente serán canalizados en la construcción de nuestro cuestionario. Esta estrategia de adaptar cuestionarios a poblaciones diferentes de aquellas en que inicialmente se utilizaron los ítems, es una práctica muy común (López Feal, 1986). Hemos analizado cada uno de esos ítems, para tener en cuenta los contenidos evaluados, comparándolos con nuestra tabla de especificaciones y luego eligiendo, cuando ha sido posible, los ítems que también sirven para nuestros propósitos, integrando de esta manera nuestro conjunto inicial.

El conjunto de ítems seleccionado ha sido comparado con la tabla de especificaciones del contenido del cuestionario, completando el banco de ítems de manera progresiva, según encontrábamos otras investigaciones cuya tarea pudiéramos reutilizar, tal cual, o con algunas modificaciones de formato o redacción; esta operación la realizamos en forma repetida.

Para la elaboración de distractores en los ítems de opciones múltiples, hemos tomado en cuenta los razonamientos equivocados en que frecuentemente incurren los examinados. Nos hemos apoyado en los ejercicios propuestos en los libros de texto analizados en el capítulo tres, cuando no hemos encontrado ítems o tareas adecuadas para alguna de nuestras especificaciones. Unos cuantos ítems han sido de elaboración propia cuando no hemos encontrado ítems adecuados con los dos procedimientos anteriores. En los ítems abiertos que elaboramos, algunos principios adaptados de Thorndike (1989) que hemos tomado en cuenta son:

- Escribir los ítems de la manera más simple y más directa posible.
- Variar la complejidad y dificultad de los ítems. Este principio es útil para mejorar la capacidad para discriminar todos los niveles de conocimiento. La idea de colocar algunos ítems más fáciles al principio permite a los estudiantes “instalarse” en el examen y reducir sus ansiedades.
- Evitar los ítems redactados de manera negativa tanto como sea posible. En caso de usarlo, se debe subrayar la palabra “no” para mayor claridad en dichos ítems.
- Nunca usar dobles negativos.

El objetivo principal de todo este proceso iterativo fue representar adecuadamente las especificaciones del contenido.

El banco de ítems inicial, obtenido con el proceso que hemos descrito previamente, se ha modificado, en algunos casos, atendiendo nuevamente las recomendaciones de Nunnally y Bernstein (1995) para el caso de ítems de opción múltiple y a los principios adaptados de Thorndike (1989) para los reactivos abiertos.

Finalmente obtuvimos un conjunto de 40 ítems, que cubren la tabla de especificaciones del contenido, eligiendo al menos dos ítems para cada contenido. En el Anexo A2 se presentan estos ítems, juntos con los resultados obtenidos en la prueba inicial de evaluación de los mismos.

4.6. SELECCIÓN DE ÍTEMS DEL CUESTIONARIO PILOTO

Concluida la planificación del cuestionario, seguidamente realizamos una selección de los ítems que conformarían el cuestionario piloto, siguiendo para ello los dos procedimientos utilizados en Díaz (2007) y recomendados en Osterlind (1989):

- La valoración empírica de los ítems, administrándolos a una muestra de estudiantes y estudiando una serie de indicadores estadísticos de los mismos.
- El análisis *a partir de un juicio*, pidiendo a una serie de expertos que realicen la valoración de los ítems particulares, de acuerdo con algunos criterios, que se les detalla.

Discribiremos ahora la metodología que aplicamos para cada uno de estos tipos de valoraciones así como los resultados obtenidos.

4.6.1 ENSAYOS PILOTO CON LOS ÍTEMS

En un primer nivel de análisis, la administración del ensayo piloto de los ítems se llevó a cabo en condiciones lo más semejantes posibles a las que se utilizarán en la forma final del cuestionario, con la intención de conseguir información objetiva, no ambigua y específica tanto de los ítems, como de las dificultades, legibilidad y ajuste al tiempo previsto.

El comportamiento de los alumnos, mientras respondían los ítems, fue monitoreado. Se respondió a sus dudas acerca de la redacción, tomando nota de los términos demasiado abstractos, demasiado técnicos o pobremente redactados, con lo cual se extrajeron indicadores de confusión de algunos de los ítems, para posteriormente mejorar su redacción y acercarse hacia una mejor comprensión literal y pragmática del significado de los ítems (Peterson, 2000).

En un segundo nivel de análisis todos los ítems fueron sometidos a prueba, calculando sus índices de dificultad, y obteniendo frecuencias de respuestas para los diferentes distractores. Los 40 ítems fueron divididos en cuatro cuestionarios, con el propósito de no hacer tan desgastante la sesión para el grupo de alumnos que respondieron los cuestionarios. En el Anexo 2, se incluyen los resultados de manera detallada.

4.6.1.1. SUJETOS

Los estudiantes inscritos en el curso de Probabilidad y Estadística para ingenieros llegan con diversas competencias previas. Muchos de ellos recibieron bases de probabilidad y de estadística en sus preparatorias y una porción reducida llega sin conocimiento alguno acerca de esta área. Su nivel de comprensión lectora, en general, es muy bueno. Tan pronto los alumnos de ingeniería del curso de Probabilidad y Estadística estudien el tema de intervalos de confianza, el cuestionario sería aplicado a estos estudiantes.

Los grupos utilizados en el estudio, aunque intencionales, pueden ser considerados muestras representativas del grupo objeto de estudio en la investigación final (Ghiglione y Matalón, 1991). Participaron en las pruebas piloto 90 alumnos del tercer semestre de las distintas carreras de ingeniería impartidas en el Campus Monterrey (Ingeniero Químico Administrador, Ingeniero en Biotecnología, Ingeniero en

Mecatrónica, Ingeniero en Sistemas Computacionales, etc.), que estaban llevando el curso de Probabilidad y Estadística. La calificación media de estos alumnos que participaron en el estudio, en su semestre anterior, fue de 82 (sobre 100). El total de alumnos inscritos en este curso de Probabilidad y Estadística fue de 805.

Los 90 alumnos fueron divididos en dos grupos de 48 y 42 alumnos respectivamente. Cada grupo contestó dos cuestionarios repartidos en dos sesiones de clase (un cuestionario por sesión, por grupo). Se aplicaron dos cuestionarios en una sesión y las otras dos versiones en otra sesión, de tal manera que al acomodarlos en forma alternada, en la sala donde se probaron estos cuestionarios, se controló la posibilidad de que se copiasen, logrando de paso probar simultáneamente la mitad de los ítems.

En las sesiones de trabajo se animó a los alumnos a que respondiesen con el máximo cuidado y la mayor seriedad. Su respuesta fue satisfactoria, mostrando en todo momento entusiasmo e interés por colaborar con el estudio.

4.6.1.2. MATERIAL

En la redacción del conjunto inicial de ítems se consideraron los objetivos del proyecto de investigación, la naturaleza de la información que va a ser recolectada la población de interés a la que se administraría el cuestionario, así como las restricciones de tiempo disponible (Peterson, 2000). Los dos últimos factores ya fueron discutidos ampliamente en el desarrollo de los apartados 4.5 y 4.6.1.1.

Fue importante realizar dos tareas que nos proveerían de la perspectiva que se necesita cuando se construye un cuestionario efectivo. La primera tarea consistió en la revisión y comprensión de los requerimientos de información del problema y las decisiones que nos guiaron a la necesidad del cuestionario. Fallar en esta tarea, provocaría que el esfuerzo realizado fuera una pérdida de tiempo para nuestra investigación y los resultados serían de muy poco valor. Una segunda tarea consistió en generar la lista de los posibles ítems de investigación que cubrieran adecuadamente el contenido y proveyeran de la información requerida. El número final preliminar de ítems en el cuestionario creemos que cumplió ese objetivo. Los cuatro cuestionarios utilizados en este ensayo piloto contenían 10 ítems cada uno, en total 40 ítems, que son un número mayor que el número contemplado para las pruebas definitivas. En su conjunto estos 40 ítems evaluaban las unidades de contenido de la definición semántica de la variable. En la tabla 4.3 se presentan los ítems elaborados para cada unidad de contenido.

Tabla 4.3. Contenidos e ítems pasados a prueba.

Contenido 1: Definición de intervalo de confianza.

Ítem 1. Explica con tus propias palabras la diferencia entre una estimación puntual y un intervalo de confianza.

Ítem 2. El intervalo de confianza del 50% para la media de una población μ es:

a. El rango dentro del cual caen el 50% de los valores de la media de la muestra $\overline{X}$.
b. Un intervalo más ancho que el intervalo de confianza del 95%.
c. Un intervalo de valores calculado a partir de los datos de la muestra. En el 50% de las muestras de una población, el intervalo calculado contiene a la media.
d. Dos veces más ancho que el intervalo de confianza del 100%.

Ítem 3. En un intervalo de confianza:

a. De una muestra a otra, el intervalo es constante.
b. Se especifica un rango de valores dentro de los cuales supuestamente cae el parámetro con seguridad.
c. **Indica un intervalo de posibles valores para el parámetro, y un porcentaje de intervalos que cubrirán, aproximadamente dicho valor, para el mismo tamaño de muestra.**
d. Siempre contienen el parámetro poblacional.

Contenido 2: El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra

Ítem 4. Dos muestras diferentes se toman de una población donde la media poblacional y la desviación estándar poblacional son desconocidas. La primera muestra tiene 25 datos, y la segunda muestra 64 datos. Se construye un intervalo de confianza de 95 % para cada muestra para estimar la media poblacional. ¿Que intervalo de confianza esperaría que tenga mayor precisión?

a. Espero que ambos intervalos de confianza tengan la misma precisión.
b. **Espero que el intervalo de confianza basado en una muestra de 64 datos sea más preciso**.
c. Espero que el intervalo de confianza basado en la muestra de 25 datos sea más preciso.
d. No puedo determinar cuál de los dos tendrá más precisión.

Ítem 5. Hemos calculado un intervalo de confianza al 90% basado en el valor medio $\overline{x}$ obtenido de una muestra de 10 casos. Si incrementamos el tamaño de la muestra a 1000, y calculamos un segundo intervalo al 90 % de confianza:

a. Debemos tener más confianza de que μ caerá en nuestro segundo intervalo.
b. **Sabemos que el segundo intervalo será 10 veces más angosto**.
c. El intervalo de confianza tiene el mismo ancho, pero su centro se desplaza.
d. El segundo intervalo de confianza es 100 veces más ancho que el primero.

Ítem 6. Comparado a los intervalos de confianza calculados en muestras de tamaño *n*=4, el ancho de los intervalos de confianza de la media de la población calculado en muestras de tamaño *n* = 50:

a. Variará más que los anchos de los intervalos para muestras de tamaño *n* =4.
b. **Variará un poco, pero no tanto como lo hicieron los anchos de los intervalos para muestras de tamaño *n* =4.**
c. Tomarán valores parecidos.

Contenido 3: El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta

Ítem 7. Si manteniendo todos los demás datos fijos el nivel de confianza se reduce (por ejemplo de 90% a 80%):

a. El intervalo de confianza no cambia.
b. El intervalo de confianza será más ancho.
c. El intervalo de confianza será más angosto.
d. El cambio en el intervalo de confianza no es predecible.

Ítem 8. Habrá más muestras donde la media de la población caiga en el intervalo de confianza con:

a. Un aumento en el tamaño de la población.
b. **Coeficientes de confianza más grandes (por ejemplo 95% en vez de 90%).**
c. Muestras pequeñas.
d. Más variabilidad en la población.

Ítem 9. En un intervalo de confianza, el ancho del intervalo puede ser reducido por:

a. Disminuyendo el tamaño de la muestra.
b. Bajando el nivel de confianza (por ejemplo de 0.99 a 0.90).
c. Aumentando la magnitud de $\sigma_{\bar{x}}$.
d. Un aumento en el tamaño de la población.

Contenido 4: El ancho de los intervalos de confianza aumenta cuando aumenta la varianza

Ítem 10. ¿Cuál de las siguientes afirmaciones es verdadera?:

a. Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza no cambia.
b. Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza disminuye.
c. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza disminuye.
d. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza no cambia.

Ítem 11. Explica cómo varía la anchura del intervalo de confianza de la media si, conservando el mismo tamaño de muestra y el mismo coeficiente de confianza tomamos una población con varianza doble.

Contenido 5: Significado del nivel de confianza (variación del intervalo en diferentes muestras)

Ítem 12. Un intervalo del 95% de confianza para la diferencia media en la producción de leche de una ganadería después de un tratamiento resultó ser (1.5, 3.5) Litros/vaca. ¿Cuál de las siguientes afirmaciones es verdadera?:

a. **No sabemos el verdadero aumento medio en la producción, pero en el 95% de intervalos calculados con el mismo tamaño de muestra y población se incluye el aumento medio en la producción.**
b. Debido a que el intervalo de confianza no contiene el cero, hay 95% de probabilidad de que tocar música no produce ningún efecto.
c. Debido a que el intervalo de confianza no contiene el cero, hay 95% de probabilidad que el verdadero aumento en la producción es 2.5 Litros/vaca.

Ítem 13. En un intervalo de confianza del 95% para la media:

a. Si se toman muchas muestras $\bar{x}$ caerá dentro del intervalo de confianza el 95% de las veces.
b. La probabilidad de que $\bar{x}$ caiga dentro de un intervalo de confianza calculado de una muestra es 0.95.
c. **Si se toman muchas muestras, el intervalo de confianza calculado contendrá a μ 95% de las veces**.

Ítem 14. Un estudiante toma una muestra de 16 compañías en Estados Unidos. El salario medio ofrecido por esas 16 compañías es de $30,600 dólares. El intervalo de confianza al 95% calculado para esta muestra va de un límite inferior de $20,500 a un límite superior de $40,500.
¿Cuál de las siguientes afirmaciones es verdadera?

a. El estudiante tiene una probabilidad de 95% de que su salario inicial estará entre $20,500 y $40,500.
b. Esto significa que el 95% de las compañías tienen un salario promedio entre $20,500 y $40,500.
c. Si tomamos otras 16 compañías diferentes, los límites del intervalo podrían variar, pero el ancho del intervalo sería aproximadamente 20,000 dólares.

Contenido 6: Estimar la media de una población normal o en una muestra grande con σ conocida

Ítem 15. La media muestral de 100 observaciones en una prueba de matemáticas es 75. Encuentre el intervalo de confianza al 95% para la media de la población, asumiendo que $\sigma = 7$

a. (61.28, 88.72) (no divide la desviación típica por 10).
b. **(73.63, 76.37).**
c. (68, 82) no multiplica por 1.96 ni divide por 10.
d. (74.3 , 75.7) no multiplica por 1.96.

Ítem 16. El propietario de una tienda desea estimar el número promedio de lápices vendidos por día. Una muestra aleatoria de 25 días es seleccionado de una población normal y el valor de la media muestral es 100. La desviación estándar de la población es $\sigma = 15$. El límite superior para un intervalo de confianza al 95% es:

a. 104.92 usa 1.64.
b. 103.00 no multiplica por 1.96.
c. 106.18 usa la T, 24 grados de libertad.
d. 105.88.

Contenido 7: Estimar la media de una población aproximadamente normal cuando, σ es desconocida

Ítem 17. Construya un intervalo de confianza al 95% para la media de una población normal de desviación típica **σ** desconocida, si en una muestra de tamaño 10 la media de la muestra es $\overline{x}$ =25 y la estimación de la desviación típica en la muestra es *s* = 6:

a. **20.71 – 29.29.**
b. 21.28 - 28.72 usa la normal en vez de la *t*.
c. 21.44 – 38.56 no divide por *n*.
d. 23.11 - 26.89 no multiplica por el valor de *t*.

Ítem 18. Un fabricante asegura que sus garrafones, contienen un litro de cloro puro. Al tomar una muestra de 16 garrafones se determinó que en promedio contenían .94 litros de cloro puro, con desviación estándar de la muestra de .097. Construir un intervalo de confianza al 95 %, para el verdadero contenido promedio de litros de cloro puro. No se conoce la desviación típica de la población

Contenido 8: Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande

Ítem 19. Se han obtenido los siguientes datos de emisión diaria de óxidos de azufre, para una muestra de tamaño n=100: media: $\overline{x}$ = 18 y cuasivarianza s^2 =36. Elabore un intervalo de confianza de 95% para la verdadera emisión diaria promedio de óxidos de azufre:

a. (17.016, 18.984) usar 1.64.
b. (16.824, 19.176).
c. (6.24, 29.76) no dividir entre *n*.
d. (8.16, 27.84) no dividir entre *n* y usar 1.64.

Ítem 20. Un estudiante de economía toma una muestra de 36 compañías a través de los Estados Unidos. Imagine que el salario medio ofrecido por esas 36 compañías es de $30000 dólares con una desviación estándar de $20000. El intervalo de confianza al 95% para el verdadero salario medio es:

a. (0, 69200) no dividir por *n*.
b. (10000, 50000) sumar/restar la desviación típica.
c. (23466, 36533).
d. (26667, 33333) no multiplicar por 1.96.

Contenido 9: Estimar una proporción

Ítem 21. En una muestra aleatoria de 100 rodamientos, 10 tienen un acabado de especificaciones defectuoso. Calcular el intervalo de confianza de 95% para la proporción verdadera de rodamientos defectuosos.

Ítem 22. En un estudio con 240 jóvenes estadounidenses cuyas edades van de 16 a 19 años, seleccionados al azar, 36 presentaron problemas graves de sobrepeso. El intervalo de confianza de 99% para la verdadera proporción *p* de jóvenes de esta población con problemas graves de sobrepeso es:

a. (.105, .195) se multiplicó por 1.96.
b. **(.091, .209).**
c. (0, .849) se multiplicó por 1.96 y no se dividió por *n.*
d. (.097, .203) se multiplicó por 2.33.

Contenido 10: Estimar una varianza

Ítem 23. Sea σ la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es =13700, n=16. Calcular el intervalo de confianza de 95% para σ.

Ítem 24. La cantidad de dióxido de carbono (CO) liquido presente en un proceso inclusión geológico en cinco días distintos en una roca cristalizada tuvo una varianza muestral igual a 80. e hace una estimación de la precisión de la técnica LRM estableciendo un intervalo de confianza de 99% para la variación en las mediciones de concentración de CO. El intervalo es:

a. (19.10, 776.69) usar 5 grados de libertad.
b. (1723.2, 123671.2) elevar al cuadrado 80.
c. (21.54, 1545.89).
d. (1528, 62135.2) elevar al cuadrado 80 y usar 5 grados de libertad.

Contenido 11: Comparar las medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes

Ítem 25. La siguiente tabla contiene un resumen de información sobre la resistencia a la compresión de cubos (N/mm^2) para especímenes de concreto:

	Tamaño muestral	Media Muestral
Tipo 1	1.68	26.99
Tipo 2	2.72	37.56

Suponga que las desviaciones estándar poblacionales de ambos grupos son $\sigma_1 = 4.89$ y $\sigma_2 = 6.43$ respectivamente. Calcule un intervalo de confianza de 99% para hallar la diferencia entre el verdadero promedio de resistencia en el Tipo1 y el verdadero promedio de resistencia en el Tipo 2:

a. (-13.02, -8.12).
b. (-12.437, -8.70) usar 1.96.
c. (-31.32, 10.18) no dividir entre *n* y *m.*
d. (-18.64, -2.5) no dividir entre n y m y no multiplicar por 2.57.

Ítem 26. La siguiente tabla contiene un resumen de información sobre la duración de baterías AA alcalinas marca Duracell y marca Eveready Energizer:

	Tamaño muestral	Media muestral
Duracell	100	4.1 hrs
Eveready Energizer	100	4.5 hrs

Suponga que las desviaciones estándar poblacionales de ambos grupos son $\sigma_1 = 1.8$ y $\sigma_2 = 2.0$ respectivamente. Calcule un intervalo de confianza de 95% para hallar la diferencia entre el verdadero promedio de duración para Duracell y el verdadero promedio de duración de Eveready Energizer.

Contenido 12: Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas

Ítem 27. En un experimento para estudiar los efectos de aplicación de fertilizante de cal y urea sobre la retención de dimetoato (un insecticida) por suelo arcilloso, se observaron los siguientes porcentajes de recuperación de dimetoato:

Suelo tratado con cal: n = 5, $\bar{x} = 26.58$ y s = 2.43.

Suelo tratado con urea: m = 5, $\bar{x} = 40.24$ y s = 2.93.

$s_p = 2.69$ (desviación típica conjunta).

Calcule un intervalo de confianza de 95% para la diferencia entre los dos porcentajes.

a. (-17.42,-9.9) usar 10 grados de libertad.
b. (-19.86, -7.45) no usar m y n.
c. (-16.75, -10.5) usar 8 grados de libertad y $\alpha = .05$
d. (-17.58,-9.74).

Ítem 28. Se compararon dos soluciones de grabado diferentes, usando dos muestras aleatorias de tamaño 10. Los resultados de la rapidez de grabado fueron:

Solución 1: $\bar{x} = 9.97$ y s = .422.

Solución 2: $\bar{x} = 10.4$ y s = .073.

$s_p = 0.34$ (desviación típica conjunta).

Calcule un intervalo de confianza de 90% para la diferencia de las medias de la rapidez de grabado.

Contenido 13: Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ desconocidas, muestras independientes grandes

Ítem 29. La tabla siguiente resume algunos datos de un experimento realizado para estudiar varias características de tornillos de anclaje:

		Resistencia al corte	
Diámetro de tornillo	Tamaño muestra	Media muestral	Desviación estándar
3/8	100	4.25	1.3
½	100	7.25	1.7

Calcule e interprete un intervalo de confianza de 95% para la diferencia del verdadero promedio de resistencias al corte:

a. (-3.41, -2.58): podemos tener una alta confianza de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.
b. (-3.41, -2.58): podemos tener una confianza moderada de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.
c. (-3.35, -2.65): podemos tener una confianza moderada de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.
d. (-3.35, -2.65): podemos tener una alta confianza de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.

Ítem 30. Se investiga el diámetro de varillas de acero fabricadas en dos máquinas

	tamaño muestral	media muestral	varianza muestral
Máquina 1	36	8.73	0.36
Máquina 2	40	8.08	0.40

El intervalo de confianza de 95% para la diferencia de las medias del diámetro de las varillas es:

a. (-1.05, 2.35) no dividir entre *n* y *m*.
b. (.319, .881) usar 1.64.
c. (.373, .927).
d. (-.776, 2.076) no dividir entre *n* y *m* y usar 1.64.

Contenido 14: Comparar dos varianzas poblacionales

Ítem 31. Una compañía quiere seleccionar el proceso de pulido que presente la variabilidad menor. Una muestra aleatoria de n_1=16 piezas del primer proceso da como resultado una desviación estándar muestral s_1= 5 micropulgadas, y una muestra aleatoria de n_2=11 piezas del segundo proceso da como resultado una desviación estándar muestral s_2= 4 micropulgadas. Establezca un intervalo de confianza de 90% para σ_1^2 / σ_2^2, suponiendo que los dos procesos son independientes y que la aspereza superficial tiene una distribución normal. ¿Cuál de los dos procesos recomendaría usted?

a. Como todos los valores están dentro del intervalo (.54, 3.96) recomendaría el proceso 1.
b. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría el proceso 2. (se cambian los grados de libertad del numerador y denominador).
c. Como el cociente de valores está dentro del intervalo (.54, 3.96) recomendaría cualquiera de los dos.
d. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría cualquiera de los dos. (se cambian los grados de libertad del numerador y denominador).

Ítem 32. Una empresa ha estado experimentando con dos disposiciones físicas distintas de su línea de ensamble. Dos muestras aleatorias independientes producen los resultados que se muestran en la siguiente tabla. Establezca un intervalo de confianza de 95% para σ_1^2 / σ_2^2, la razón de las varianzas del número de unidades terminadas para las dos disposiciones de línea de ensamble. Con base en el resultado, ¿Cuál de las dos disposiciones recomendaría usted?

Línea de ensamble 1	Línea de ensamble 2
n_1 = 21 días	n_2 = 25 días
s_1^2 = 1,432	s_2^2 = 2,864

Contenido 15: Elegir un modelo de distribución muestral del estadístico

Ítem 33. La distribución muestral utilizada en la construcción de intervalos de confianza para la varianza en muestras pequeñas es:

a. Distribución t de Student.
b. Distribución Ji-cuadrada.
c. Distribución Normal.
d. Distribución F.

Ítem 34. Cómo serán afectados los intervalos de confianza usando una desviación estándar muestral (*s*) y un valor *t* en vez de la desviación estándar poblacional (σ) y un valor *z?:*

a. Los centros de los intervalos de confianza serán diferentes e igual de amplios.
b. **Los intervalos tendrán mayor amplitud y el mismo centro.**
c. Los intervalos tendrán menor amplitud y el mismo centro.
d. Los intervalos tendrán mayor amplitud y diferente centro.

Contenido 16: Determinar valores críticos en la distribución del estadístico

Ítem 35. Al calcular un nivel de confianza para un 90% para un grupo de puntuaciones distribuido normalmente, usted pudiera usar un valor *z* de:

a. 1.96.
b. 1.65.
c. 0.90 interpretar el 90% como el valor de *z*.
d. 1.29 acumulada hasta .90.

Ítem 36. Si el nivel de confianza es 0.95, para un intervalo de confianza para la media poblacional con desviación estándar poblacional desconocida y para un grupo de puntajes distribuido normalmente de tamaño $n = 20$, los valores críticos pudieran ser:

a. -1.65 y 1.65 uso de normal estándar.
b. -1.96 y 1.96 uso de normal estándar.
c. -2.093 y 2.093.
d. -2.085 y 2.085 usar 20 grados de libertad.

Contenido 17: Interpretación de intervalos de confianza utilizando un programa de ordenador

Ítem 37. La siguiente salida de computadora presenta una muestra simulada de una población normal con μ =130 y σ =10. Luego se usó un comando para establecer un intervalo de confianza del 95% para μ.

118.690	144.226	138.827	125.934	136.198	148.731	133.394
134.184	127.260	145.345	121.966	125.435	141.236	150.021
126.895	118.499	137.225	136.781	119.526	125.628	134.865
116.416	134.312	138.053	140.828			

Mean	Median	TrMean	StDev	95% Confidence Interval for Mu
132.82	134.31	132.78	9.74	(128.798 , 136.840)

Escriba el intervalo de confianza que se obtuvo e interprete el resultado.

Ítem 38. La siguiente salida de computadora presenta dos muestras simuladas de dos poblaciones normales. La población 1 con μ=90 y σ =10 y la población 2 con μ =92 y σ=10

Muestra 1:

83.3195	87.6793	86.7831	95.0518	92.9781	86.6457
85.1305	97.5013	83.1112	82.2751	82.7831	90.2786
89.5876	71.2591	82.0282	90.6264		

Muestra 2:

82.312	95.098	92.598	85.959	91.319	108.130
90.392	90.074	78.789	100.923	85.601	89.861
78.685	100.354	81.267	101.432		

	N	Media	D. Típica	Error típico
C1	16	86.69	6.25	1.6
C2	16	90.80	8.70	2.2

95% intervalo de confianza para mC1 - mC2: (-9.6; 1.4)

Escriba el intervalo de 95% de confianza que se obtuvo para la diferencia de medias e interprete el resultado.

Contenido 18: Interpretar Gráficos de Intervalos de confianza

Ítem 39. Considere el gráfico siguiente del rendimiento medio de cebada en 1980,1984 y 1988 junto con un intervalo de 95% de confianza

```
Intervalos del 95% para la media

Año N Media StDev       ----+---------+---------+---------+--
1980 6 184.00 2.61       (----*---)
1984 5 212.40 14.36                          (-----*----)
1988 5 182.40 1.82     (-----*----)
                        ----+---------+---------+---------+--
Desv. Típica conjunta =.19 180       195       210       225
```

¿Cuál de las siguientes afirmaciones es verdadera?:

a. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay buena evidencia que las medias de las muestras difieran.
b. La estimación de la media de la población en 1980 es menos precisa que en 1988.
c. Puesto que los intervalos de confianza para 1980 y 1984 no se solapan, hay poca evidencia que las medias de las poblaciones respectivas difieran.
d. **Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay poca evidencia que las medias de las poblaciones difieran.**

Ítem 40. Considere el gráfico siguiente del rendimiento medio de cebada en 1990,1994 y 1998 junto con un intervalo de 95% de confianza

```
Individual 95% CIs For Mean

Año N Media StDev           --+---------+---------+---------+--
1990 36 238.83 26.57                           (-----*----)
1994 75 186.60 2.51         (-----*-----)
1998 67 186.71 2.63           (----*----)
                            --+---------+---------+---------+----
Desv. Típica conjunta=15.49 175      200       225       250
```

¿Cuál de las siguientes afirmaciones es verdadera?:

a. Puesto que los intervalos de confianza para 1994 y 1998 tienen considerable solape, hay poca evidencia que las medias de las poblaciones difieran.
b. La estimación de la media de la población en 1998 es menos precisa que en 1994.
c. Puesto que los intervalos de confianza para 1990 y 1994 no se solapan, no hay evidencia que las medias de las poblaciones respectivas difieran.
d. Puesto que los intervalos de confianza para 1994 y 1998 tienen considerable solape hay buena evidencia que las medias de las muestras difieran.

4.6.1.3. MÉTODO

Los datos que obtuvimos al administrar los cuestionarios en el piloteo, fueron codificados y luego analizados a través del uso del paquete estadístico ITEMAN (Item and Test Analysis Program). El análisis se realizó a partir índices de dificultad y discriminación derivados de los porcentajes de respuestas en los diferentes distractores y de la distribución de frecuencias, sin incluir estudios de puntuaciones globales o análisis de fiabilidad.

El índice de dificultad del reactivo es calculado dividiendo el número de alumnos que contestaron correctamente entre el número de personas que tomó la prueba luego este resultado multiplicado por 100. A medida que este índice se aproxima a 1.00 o a 0.00, menos información diferencial aporta sobre los examinados. En tanto que el índice de discriminación provee información acerca de la validez del reactivo. Es decir, indica si el reactivo mide lo que deseaba medir. Este índice separa la muestra en grupos homogéneos y nos señala las diferencias en ejecución entre los examinados del grupo superior y los del grupo inferior. Una forma de computar este índice es la siguiente:

$$ID = \frac{\begin{bmatrix}\text{número de examinados del grupo} \\ \text{superior que contestó correctamente}\end{bmatrix}}{\begin{bmatrix}\text{número de examinados del} \\ \text{grupo superior}\end{bmatrix}} - \frac{\begin{bmatrix}\text{número de examinados del grupo} \\ \text{inferior que contestó correctamente}\end{bmatrix}}{\begin{bmatrix}\text{número de examinados del} \\ \text{grupo inferior}\end{bmatrix}}$$

Un poder discriminativo de +.30 o más fue el criterio que seguimos para decidir que el ítem era bueno (López e Hinojosa, 2001; Hinojosa y Góngora, 2006). El propósito de recolectar estos datos fue tener información que complementara a la recogida a través del juicio de expertos, que nos ayudara en la decisión de los ítems que serían seleccionados, de tal manera que logremos cubrir el contenido, tener una dificultad variada y que podamos detectar sesgos y errores descritos en la investigación. El número de alumnos que completó cada ítem puede variar debido a que los alumnos del segundo grupo fue ligeramente menor. Es por esta razón que en los anexos presentamos el análisis de los ítems aisladamente.

En los cuestionarios hemos intercalado ítems de opción múltiple y de respuesta abierta. En los de respuesta abierta dimos créditos parciales al momento de calificarlos. Hemos graduado la puntuación asignada, dependiendo del grado de corrección en las respuestas de los examinados. En esas tareas asignamos una puntuación menor a la solución que era incompleta o correcta a medias. En el anexo 2, se describen los criterios que seguimos en la asignación de los puntajes para los casos que se nos presentaron.

Al ser el índice de dificultad la proporción de examinados que contestaron correctamente el ítem, pudimos calcular a partir de este estadístico, intervalos de confianza para la proporción verdadera que contestaría correctamente el ítem en la población de la cual se obtuvo la muestra, utilizando la aproximación normal, puesto que nuestras muestras eran de más de 30 elementos. En los ítems de respuestas abiertas, para efectos de la obtención del índice de dificultad, se consideraron respuestas correctas, además de los ítems de respuesta totalmente correcta, aquellos ítems en los que hubiera prevalecido un razonamiento correcto, con identificación correcta de los datos y solamente errores leves de cálculo.

4.6.1.4. RESULTADOS

Presentamos un análisis detallado en el anexo 2 de los resultados obtenidos en cada ítem. En la tabla 4.4 mostramos un resúmen de los porcentajes de respuestas correctas y resultados en los distractores. Podemos destacar que en algunos ítems uno de los distractores predominó sobre el resto, siendo elegido por más de un 45% de los alumnos (algunos de los ítems presentaron distractores que no fueron elegidos), por lo que algunos de esos ítems serían descartados o se necesitaría revisar nuevamente la redacción de las opciones.

En general los ítems tuvieron una dificultad aceptable, aunque algunos resultaron demasiado difíciles, como el ítem 14 que evalúa la comprensión del significado del nivel de confianza (variación del intervalo en diferentes muestras). Otros fueron excesivamente fáciles como el ítem 1 que evalúa la comprensión de la definición de intervalo de confianza.

Puesto que el índice de dificultad es una estimación puntual, hemos añadido el cálculo de intervalos de confianza, para dar una medida de cómo el índice podría extenderse a la población de estudiantes.

En la tabla 4.5 se presenta la estimación clásica con los índices de dificultad. El cálculo de los intervalos de confianza se hace mediante la aproximación normal. Los resultados se obtuvieron mediante el paquete estadístico Minitab. Observamos que los intervalos de confianza son relativamente amplios, debido a que los tamaños de muestra son moderados. No obstante, se conserva la dificultad relativa de los ítems.

Tabla 4.4. Porcentaje de respuestas correctas y observaciones para los distractores

Ítem	N	Índice de dificultad	Observaciones
1	48	.89	
2	42	.76	
3	48	.75	
4	48	.25	Domina el distractor c (.63)
5	42	.19	
6	42	.57	
7	48	.70	
8	42	.57	
9	48	.20	
10	48	.29	distractor d (.33)
11	42	.42	
12	48	.64	
13	42	.76	
14	42	.14	Domina el distractor b (.52)
15	48	.68	
16	42	.47	distractor a no fue elegido
17	48	.58	
18	42	.40	
19	48	.67	
20	42	.83	
21	44	.73	
22	47	.72	
23	48	.41	
24	42	.40	distractor b (.36)
25	44	.57	distractor d no fue elegido
26	47	.81	
27	44	.45	
28	47	.68	
29	44	.59	
30	47	.85	distractor a y d no fueron elegidos
31	44	.80	
32	47	.64	distractor a y d no fueron elegidos
33	48	.57	
34	42	.21	domina el distractor c (.71) y el distractor d no fue elegido
35	48	.75	Distractor d no fue elegido
36	42	.59	
37	48	.72	
38	42	.76	
39	48	.68	
40	42	.78	

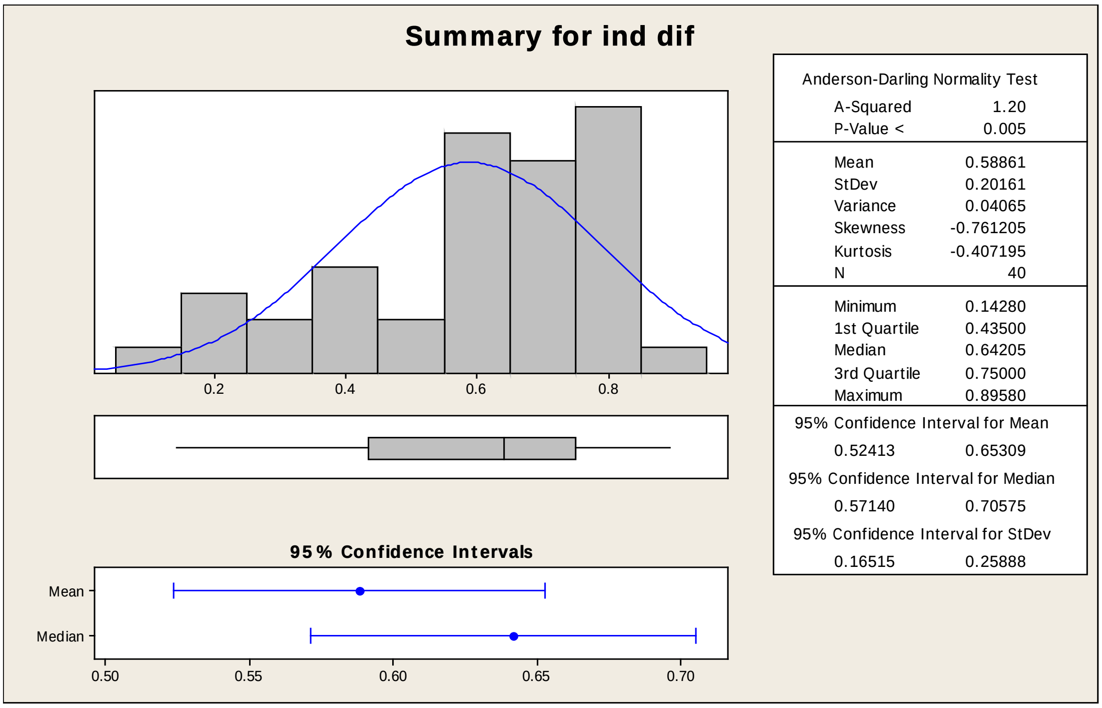
Tabla 4.5. Estimación de los índices de dificultad

Item	n	Índice dificultad	Límite inferior	Límite superior
I1	48	0.895833	(0.773422;	0.965302)
I2	42	0.761905	(0.605498;	0.879484)
I3	48	0.750000	(0.604041;	0.863628)
I4	48	0.250000	(0.136372;	0.395959)
I5	42	0.190476	(0.086006;	0.341184)
I6	42	0.571429	(0.409611;	0.722793)
I7	48	0.708333	(0.559365;	0.830469)
I8	42	0.571429	(0.409611;	0.722793)
I9	48	0.208333	(0.104691;	0.349910)
I10	48	0.291667	(0.169531;	0.440635)
I11	42	0.428571	(0.277207;	0.590389)
I12	48	0.645833	(0.494568;	0.778394)
I13	42	0.761905	(0.605498;	0.879484)
I14	42	0.142857	(0.054284;	0.285394)
I15	48	0.687500	(0.537486;	0.813404)
I16	42	0.476190	(0.320041;	0.635822)
I17	48	0.583333	(0.432132;	0.723873)
I18	42	0.404762	(0.256291;	0.567179)
I19	48	0.687500	(0.537486;	0.813404)
I20	42	0.833333	(0.686359;	0.930259)
I21	44	0.727300	(0.590300;	0.864200)
I22	47	0.702100	(0.566400;	0.837900)
I23	48	0.416667	(0.276127;	0.567868)
I24	42	0.404762	(0.256291;	0.567179)
I25	44	0.568200	(0.415800;	0.720500)
I26	47	0.787200	(0.665800;	0.908700)
I27	44	0.454500	(0.301400;	0.607700)
I28	47	0.680900	(0.542500;	0.819200)
I29	44	0.590900	(0.439700;	0.742100)
I30	47	0.829800	(0.718200;	0.941300)
I31	44	0.795500	(0.671400;	0.919500)
I32	47	0.638300	(0.495700;	0.780900)
I33	47	0.574468	(0.421785;	0.717421)
I34	42	0,214286	(0.102960;	0.368116)
I35	48	0.750000	(0.604041;	0.863628)
I36	42	0.595238	(0.432821;	0.743709)
I37	48	0.729167	(0.581544;	0.847218)
I38	42	0.761905	(0.605498;	0.879484)
I39	48	0.687500	(0.537486;	0.813404)
I40	42	0.785714	(0.631884;	0.897040)

Para verificar el supuesto de normalidad en la estimación del valor medio de la distribución de índices de dificultad, se hicieron varias pruebas. En la figura 4.1 se presenta un resumen gráfico de los índices de dificultad obtenidos en las pruebas empíricas de ítems. Ahí podemos apreciar una distribución que podemos considerar que tiene una simetría aceptable, es decir nuestra distribución es aproximadamente normal con un valor medio cercano al .58. El 50% de los valores centrales tienen un valor entre

.43 y .75. También se presenta en este resumen estadísticos descriptivos de los índices de dificultad tales como coeficientes de asimetría y curtosis que al ser valores cercanos al cero son otra evidencia de normalidad. Los valores de medidas de posición de media y mediana en una distribución normal son iguales y en nuestro caso estos valores son ligeramente parecidos.

Fig. 4.1 Resumen gráfico de los índices de dificultad



Finalmente hemos realizado pruebas de normalidad utilizando el programa Statgraphics y los resultados se muestran a continuación en la Tabla 4.6. Estos resultados reflejan valores aceptables de curtosis y simetría y el valor *p* en la prueba de bondad de ajuste es muy cercano a 0.05, lo que significa que no hay una discrepancia en demasía para concluir normalidad de la distribución, lo que nos autorizará a realizar contrastes de diferencias de medias en condiciones de normalidad supuesta.

Tabla 4.6. Pruebas de normalidad para índices de dificultad y valores p

Pruebas de normalidad	p-value
Computed Chi-Square goodness-of-fit statistic = 23.75	0.0491087
Shapiro-Wilks W statistic = 0.910618	0.00396225
Z score for skewness = 1.41696	0.156495
Z score for kurtosis = -0.47044	0.638038

4.6.2 VALORACIÓN DE ÍTEMS MEDIANTE JUICIO DE EXPERTOS

Según Losada y López Feal (2003) uno de los criterios fundamentales usados en la construcción de instrumentos, es la selección de ítems a partir del juicio de expertos. En los estudios cualitativos y exploratorios las muestras de expertos son frecuentes para generar el recurso intelectual en el diseño de un cuestionario (Hernández, Fernández y Baptista, 1998). Esto asegura una selección de ítems que representan manifestaciones concretas del constructo, relevantes para el uso que se dará a las puntuaciones y representativos del dominio de ítems de interés. Lo que se pretende es la validación del contenido del cuestionario y usar la información recogida de los expertos para seleccionar los ítems que finalmente constituyan el cuestionario. Millman y Green (1989) y Thorndike (1989) indican que algunos criterios que los expertos podrían valorar son los siguientes:

- Comunicabilidad: redacción correcta, claridad y consistencia. El director del trabajo, así como dos compañeros del Departamento de Matemáticas que son profesores de estadística leyeron las versiones previas. Además también nos ayudaron los estudiantes que han participado en las pruebas, a los que se pidió nos señalaran las posibles dificultades de comprensión.
- Nivel adecuado de dificultad. Nuestro cuestionario incluye distintos niveles de dificultad, inferidos de los datos en las pruebas piloto de ítems.
- La adecuación a las especificaciones del test. El grado en que el contenido del instrumento se adecua a las especificaciones del test.

Preparada la tabla de especificaciones del cuestionario y colectados los ítems que según nuestra perspectiva cubrían los objetivos propuestos, la fase que le siguió fue la de valoración por medio de expertos externos que valoraron todos los criterios descritos.

4.6.2.1. SUJETOS

Millman y Green (1989) señalan que el "experto" lo define el propósito del instrumento y que el grupo elegido de expertos ha de representar una diversidad relevante de capacidades y puntos de vista. En nuestro trabajo, los expertos que participaron fueron seleccionados con base en su conocimiento experto, por ser profesores de estadística con amplia experiencia docente y más específicamente de la enseñanza de los intervalos de confianza, así como un buen grado de experiencia de investigación sobre la didáctica de la inferencia estadística. Sus características se definen en la tabla 4.7.

Tabla 4.7. Expertos participantes en la valoración del contenido

Experto	Formación y Experiencia
1	Profesora de Estadística y Educación Matemática en la Universidad de Santa Fe, Argentina. Realizó una tesis doctoral sobre didáctica de la estadística
2	Profesora de Estadística en la Universidad Nacional de México. Doctora en Matemáticas. Imparte clase de estadística a ingenieros. Investiga en didáctica de la matemática
3	Profesora de Estadística en la Universidad de Cantabria. Imparte estadística en empresariales y sociología. Investiga en didáctica de la estadística. Doctora en Estadística
4	Profesora de Estadística en la Universidad Nacional de Rio Cuarto en Argentina. Master en Estadística. Realiza una tesis en didáctica de la estadística.
5	Matemático. Master en Estadística. Master en Física. Profesor en Estadística en el Instituto Tecnológico y de Estudios Superiores de Monterrey, México. Catedrático universitario durante 21 años.
6	Profesor de estadística en la Universidad de Cali, Colombia. Realizó una tesis doctoral sobre la enseñanza de los intervalos de confianza. Construyó un cuestionario de evaluación sobre intervalos de confianza y contraste de hipótesis.
7	Profesora de Estadística en la Universidad Politécnica de Barcelona. Doctora en Estadística. Investiga en Didáctica de la Estadística
8	Matemática. Doctora en Estadística. Profesora en Estadística en el Instituto Tecnológico y de Estudios Superiores de Monterrey, México. Catedrática universitaria durante 15 años.
9	Profesora de Métodos de investigación en la Universidad de Huelva. Doctora en Metodología de las Ciencias del Comportamiento. Imparte clases de estadística en Psicología. Construyó un cuestionario de evaluación como parte de su tesis doctoral.

La localización de nuestro grupo de expertos se inicia a partir del estado de la cuestión, en donde identificamos algunos investigadores que han publicado en los últimos 10 años en revistas o congresos internacionales de prestigio o tesis doctorales vinculadas estrechamente con la enseñanza de los intervalos de confianza. Nuestra lista inicial contemplaba 20 investigadores; 5 estadounidenses, 7 españoles y 8 de otros países (Argentina, Chile, Colombia y México). Sin embargo nuestra lista la redujimos a 15 debido a que nuestro cuestionario va dirigido a estudiantes de habla hispana. A esos 15 investigadores se les podía localizar por correo electrónico, accediendo a estas direcciones por internet, por medio de sus departamentos o bien a través de las listas de miembros de Sociedades Internacionales.

Finalmente decidimos invitar a participar a un total de 12 expertos, de los cuales 9 aceptaron. Se eligieron de modo que se cubriesen diversos países (España, México, Colombia, Chile y Argentina) y tuvieran en cuenta la mayoría de investigadores de América (cubrieron 4 universidades de América y otras cuatro españolas).

Nuestra muestra de expertos incluye matemáticos y estadísticos, profesores de matemáticas, didáctica de la matemática y estadística. Uno de ellos es investigador en estadística y el resto son investigadores en didáctica de la probabilidad y estadística. La edad comprende un rango desde 27 a 58 años, con experiencia investigadora de al menos 5 años en didáctica de la estadística.

4.6.2.2. MÉTODO

Terminada la tabla de especificaciones y seleccionado el conjunto inicial de ítems, preparamos un cuestionario que se recoge en el Anexo 1 para ser completado por el grupo de expertos, con la finalidad de aportar evidencias de validez de contenido (Martínez Arias, 1995, p. 337). Nuestro objetivo era entonces doble (Díaz, 2007):

- Llegar a establecer un consenso sobre la tabla de especificaciones del instrumento, determinando cuales especificaciones del contenido eran relevantes para los propósitos del instrumento. Esto permitiría reforzar los resultados obtenidos del análisis de contenidos de los libros de texto.
- Establecer también un consenso de opiniones de los expertos acerca de cómo cada ítem particular se ajusta de manera adecuada para evaluar el contenido específico para el cuál ha sido diseñado y que sirva como base para seleccionar los ítems definitivos.

Antes de enviarles el cuestionario a los expertos, primero les enviamos una carta de presentación, explicando los propósitos de la investigación y solicitándoles su ayuda. A continuación se incluían unas instrucciones para completar el cuestionario. Luego, se describían cada uno de los contenidos, seguidos de los ítems que servían para evaluarlos y una tabla de datos que el experto debía llenar, tomando en cuenta dos aspectos (ver ejemplo, en figura 4.3), ambos de acuerdo a una escala tipo Likert. Proporcionamos a los expertos, para realizar esta tarea, la tabla de especificaciones del contenido, pidiéndoles que para cada contenido propuesto evaluasen su relevancia para la comprensión de los intervalos de confianza, de acuerdo a una escala de cinco puntos, en la que los valores extremos y el intermedio toman el siguiente significado:

1. *Nada:* bajo grado de relevancia o relevancia incierta.
3. *Regular:* moderado grado de relevancia.
5. *Mucho:* alto grado de relevancia.

La decisión de usar una escala de valoración (1-5) se tomó atendiendo a las sugerencias de Osterlind (1989), quién opina que se debe pedir algo más que el acuerdo/desacuerdo a los jueces, y recomienda usar solamente las categorías anteriores, ya que resulta de muy poca utilidad práctica cuando se usa una discriminación mayor y ello puede complicar de manera innecesaria el proceso de alcanzar el acuerdo.

El método seguido por los expertos para la valoración de los ítems de nuestro cuestionario, cae dentro de un esquema conocido como panel “ciego”; los jueces no tienen un encuentro cara a cara, ni conocen la identidad de los otros jueces, por lo tanto se logra que no haya contaminaciones o sesgos por persuasión o por los efectos de prestigio personal, rango o carisma (Osterlind, 1989, pg. 259)

El punto inicial fue la valoración de la importancia de la unidad de contenido para la comprensión del intervalo de confianza. Para ejecutar esta tarea se proporcionó a los expertos, al final del cuestionario, la tabla de especificaciones de contenido. También se les pidió que agregaran otros posibles contenidos que no hubiéramos contemplado en la tabla de especificaciones, si así lo consideraban pertinente.

Luego se pidió a los expertos que expresaran su grado de acuerdo respecto al emparejamiento de cada uno de los ítems con su contenido principal, utilizando la misma escala (1 a 5). Se pidió a los jueces que evaluaran el emparejamiento hecho previamente entre ítems y especificaciones de contenido, según lo sugiere Osterlind (1989). Con ello se indicaba a los jueces nuestra forma de pensar acerca de cuáles ítems evaluaban cada contenido específico y el rol de los jueces consistió en confirmar o rechazar esta opinión.

Destacamos el gran apoyo que recibimos del grupo de expertos al evaluar el conjunto completo de ítems y objetivos, frente a la otra alternativa de dividir el banco inicial de ítems en varios grupos y entonces requerir un mayor número de expertos.

Figura 4.2. Ejemplo del contenido del cuestionario de expertos

CUESTIONARIO PARA RECOGIDA DE DATOS DE EXPERTOS

A continuación presentamos una lista de contenidos que consideramos relevantes para evaluar la comprensión del tema de intervalos de confianza. Presentamos también, junto a cada contenido, una lista de ítems de evaluación. El cuestionario está dirigido a estudiantes de ingeniería en México. Se usa la notación habitual en este país. Requerimos su colaboración para evaluar:

- El grado en que el contenido propuesto es relevante para la comprensión del intervalo de confianza en un curso de estadística dirigido a ingenieros.
- El grado en que cada ítem es adecuado para evaluar la comprensión del contenido específico propuesto en esta especialidad.

Le agradeceríamos marque para cada uno de estos dos aspectos su opinión en la escala 1 a 5, donde 1 indica: nada relevante y 5 muy relevante.

Contenido 1: Definición de intervalo de confianza.

Ítem 1. Explica con tus propias palabras la diferencia entre una estimación puntual y un intervalo de confianza

Ítem 2. El intervalo de confianza del 50% para la media de una población μ es:
a. El rango dentro del cual caen el 50% de los valores de la media de la muestra $\overline{X}$.
b. Un intervalo más ancho que el intervalo de confianza del 95%.
c. Un intervalo de valores calculado a partir de los datos de la muestra. En el 50% de las muestras de una población, el intervalo calculado contiene a la media.
d. Dos veces más ancho que el intervalo de confianza del 100%.

Ítem 3. En un intervalo de confianza:
a. De una muestra a otra, el intervalo es constante.
b. Se especifica un rango de valores dentro de los cuales supuestamente cae el parámetro con seguridad.
c. **Indica un intervalo de posibles valores para el parámetro, y un porcentaje de intervalos que cubrirán, aproximadamente dicho valor, para el mismo tamaño de muestra.**
d. Siempre contienen el parámetro poblacional.

	1: Nada	2	3	4	5: Mucho
El contenido “Definición de intervalo de confianza” es relevante					
El ítem 1 es adecuado para este contenido					
El ítem 2 es adecuado para este contenido					
El ítem 3 es adecuado para este contenido					

4.6.2.3. RESULTADO DEL JUICIO DE EXPERTOS

El grado de emparejamiento que todos los jueces dan al dominio de contenido(o al ítem), lo resumimos globalmente, a través de la obtención de índices cuantitativos como la media o la mediana, que es uno de los procedimientos habituales para resumir las decisiones de los jueces (Martínez Arias, 1995, p.338).

En este estudio se han calculado ambos, pero se ha usado la media como criterio para la selección, ya que este índice toma en cuenta los valores extremos, por lo que cualquier falta de coincidencia importante es tomada en cuenta. Además se ha calculado la desviación típica, como índice adicional para las valoraciones en que se presente igualdad de medias. De esta manera la selección de los contenidos, se rigió bajo un criterio más exigente.

En la tabla 4.8 se presentan los índices cuantitativos media, mediana y desviación típica de las puntuaciones que el grupo de expertos asignaron a cada uno de los contenidos de la tabla de especificaciones del cuestionario. Todos los contenidos que recibieron una alta valoración (por encima de 4 puntos) fueron aceptados.

A excepción de los contenidos 15 y 16, los cuales recibieron una calificación mediana de 4, el resto de los contenidos recibió una alta valoración. En esos contenidos 15 y 16, algunos desacuerdos que se presentaron tienen que ver con la consideración por parte de dos expertos de que el contenido era poco adecuado, que esos contenidos están implícitos en otros contenidos como el 6, 7, 8, etc. En consecuencia decidimos mantener ambos contenidos porque tenían una calificación media aceptable de 4.11 y 4.22 respectivamente en la valoración hecha por los expertos.

Por otra parte, algunos contenidos que los jueces consideraron pertinentes agregar, fueron los siguientes:

- *Interpretar el significado de intervalo de confianza según el contexto* (experto 4). Este contenido está incluido en los ítems 12 y 14 del contenido 5.
- *Estimar la media de la diferencia con dos muestras dependientes o pareadas (experto 4).* Este contenido no está contemplado en el programa analítico del curso, por lo cual no se incluyó en el cuestionario final.
- *Calcular intervalos de confianza en cartas de control* (experto 2). Igual que la recomendación anterior, este contenido no está contemplado en el programa analítico del curso.

Por lo tanto se decidió conservar todos los contenidos y no agregar más, debido a que los recomendados por los jueces se descartaron por las razones ya señaladas.

Tabla 4.8. Frecuencia de asignación de acuerdo por los expertos (1-5) a cada contenido, media, mediana y desviación típica

Tipo	contenido	1	2	3	4	5	$\overline{x}$	med	s
Def.	1. Definición de intervalo de confianza					9	5	5	0
Propiedades	2. El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra					9	5	5	0
	3. El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta					9	5	5	0
	4. El ancho de los intervalos de confianza aumenta cuando aumenta la varianza				1	8	4.88	5	0.33
	5. Significado del nivel de confianza (variación del intervalo en diferentes muestras)				1	8	4.88	5	0.33
Campos de problemas	6. Estimar la media de una población normal o en una muestra grande con σ conocida				3	6	4.66	5	0.50
	7. Estimar la media de una población aproximadamente normal cuando, σ es desconocida				1	8	4.88	5	0.33
	8. Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande				1	8	4.88	5	0.33
	9. Estimar una proporción				1	8	4.88	5	0.33
	10. Estimar una varianza.				2	7	4.77	5	0.44
	11. Comparar las medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes				3	6	4.44	5	0.72
	12. Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas.			1	3	5	4.44	4.5	0.72
	13. Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ desconocidas, muestras independientes grandes.			1	3	5	4.44	5	0.72
	14. Comparar dos varianzas poblacionales				2	7	4.77	5	0.44
	15. Elegir un modelo de distribución muestral del estadístico			3	2	4	4.11	4	0.92
	16. Determinar valores críticos en la distribución del estadístico			1	5	3	4.22	4	0.66
Proc.	17. Interpretar intervalos de confianza obtenidos de un programa de ordenador			1		8	4.77	5	0.66
	18. Interpretar gráficos de intervalos de confianza					9	5	5	0

En la tabla 4.9 se exhiben los resultados del grado de acuerdo que los expertos han sentenciado entre cada uno de los ítems y su contenido principal. En estos resultados podemos apreciar, en general, ítems bien valorados, con puntuaciones de al menos 4 puntos. Los ítems que tienen o bien están ligeramente alrededor de ese promedio son: 6, 12, 14, 27, 28, 29, 33, 34, 35 y 36. En algunos de esos casos los expertos consideraban que era necesario mejorar la redacción (I6 e I34). En el grupo de ítems en que la puntuación media se encuentra por debajo de 4 puntos, están ítems que los expertos juzgaban debieran evaluarse dentro de un problema de ingeniería al calcular el intervalo de confianza (I35 e I36), es decir donde sugieren cambiar el contexto.

Tabla 4.9. Frecuencia de asignación de acuerdo (1-5) por los expertos a cada ítem respecto a su contenido principal, media, mediana y desviación típica

ítem	1	2	3	4	5	Media	Mediana	Desviación típica
I1				2	7	4.778	5	0.441
I2				3	6	4.667	5	0.500
I3			1	3	5	4.444	5	0.726
I4		1		1	7	4.556	5	1.014
I5				5	4	4.444	4	0.527
I6			2	5	2	4.000	4	0.707
I7				1	8	4.889	5	0.333
I8			2	3	4	4.222	4	0.833
I9				2	7	4.778	5	0.441
I10				1	8	4.889	5	0.333
I11				1	8	4.889	5	0.333
I12		1	1	4	3	4.000	4	1.000
I13		1	1	2	5	4.222	5	1.093
I14		1	4	3	1	3.444	3	0.882
I15			1	5	3	4.222	4	0.667
I16				7	2	4.333	4	0.500
I17				5	4	4.556	5	0.527
I18				3	6	4.667	5	0.500
I19			2	3	4	4.222	4	0.833
I20				2	7	4.778	5	0.441
I21				5	4	4.444	4	0.527
I22			1	1	7	4.667	5	0.707
I23				5	4	4.444	4	0.527
I24			2	2	5	4.333	5	0.866
I25		1	1	2	5	4.222	5	1.093
I26			1	2	6	4.556	5	0.726
I27		2	2	1	4	3.778	4	1.302
I28		1	1	3	4	4.111	4	1.054
I29	1	1		1	6	4.111	5	1.537
I30			2	3	4	4.222	4	0.833
I31		1		3	5	4.333	5	1.000
I32				4	5	4.556	5	0.527
I33			3	3	3	4.000	4	0.866
I34		2	2	3	2	3.556	4	1.130
I35		2	1	5	1	3.556	4	1.014
I36		1	3	2	3	3.778	4	1.093
I37			1	4	4	4.333	4	0.707
I38			1	4	4	4.333	4	0.707
I39			1	2	6	4.556	5	0.726
I40			1	2	6	4.556	5	0.726

El hecho de contar con al menos dos ítems nos permitió elegir el mejor valorado por los jueces. En varios de los ítems, atendiendo las recomendaciones de los expertos se mejoró la redacción o se amplió el contexto para brindar una mayor claridad o relacionar más el ítem con la especialidad cursada por los estudiantes.

4.7. SELECCIÓN DE ÍTEMS PARA EL CUESTIONARIO PILOTO

En esta sección describimos las acciones que tomamos para seleccionar los ítems que compondrían el cuestionario piloto, una vez terminadas las fases anteriores:

- Como primer paso se desecharon aquellas especificaciones de contenido en las que el grado de acuerdo sobre su relevancia no fuese suficientemente elevado y consensuado. El grado de acuerdo se valoró mediante la puntuación media dada al contenido por los nueve expertos (a mayor valor medio mayor relevancia) y el grado de acuerdo por la desviación típica (a menor valor, mayor acuerdo).
- Para aquellos contenidos con alto grado de acuerdo respecto a su relevancia se examinaron los ítems.
- Se desecharon los ítems que no fuesen altamente valorados por los expertos en cuanto a su emparejamiento con el contenido (puntuación menor de 4 en las escalas 1-5).
- De entre aquellos ítems bien valorados por los expertos, en cuanto a su emparejamiento con el contenido (puntuación superior a 4), se eligió el que hubiese presentado menor índice de dificultad en las pruebas piloto de ítems y buen poder discriminativo.

El resto de los ítems formaron parte de una reserva, para el caso de que los resultados del cuestionario piloto hagan necesarios aumentar el número de ítems para un cierto contenido o sustituir algunos de los ítems previamente seleccionados. Destacamos que cada ítem permite evaluar la comprensión de otros contenidos secundarios además del contenido principal. Un análisis completo de los ítems finalmente seleccionados, y del contenido abarcado por el cuestionario piloto se incluye en el capítulo 6.

Posterior a la selección de los ítems, con base en su contenido principal, conduciremos un análisis teórico del contenido evaluado, el cuál reportaremos en otro apartado. Analizamos seguidamente, siguiendo la misma metodología de Díaz (2007) para cada uno de los contenidos, los resultados de los ítems diseñados para evaluarlo, así como los criterios aplicados para seleccionar los ítems que conformaran nuestro instrumento piloto.

Contenido 1: Definición de intervalo de confianza.

Este contenido fue valorado muy positivamente ($\bar{x} = 5, s = 0$), (ver tabla 4.7). Incluye los ítems 1, 2 y 3. Se decidió conservar para el cuestionario el ítem 2. Puesto que el ítem 1 resultó muy fácil para los alumnos y en el ítem 3 una de las opciones fue elegida por un solo alumno, se decidió desechar los ítems 1 y 3.

Tabla 4.10. Resultados en los ítems relacionados con el contenido 1

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I1				2	7	4.778	0.441	.895	48	
I2				3	6	4.667	0.500	.76	42	
I3			1	3	5	4.444	0.726	.75	48	La opción d (.02) fue escogida por un solo alumno

Contenido 2: El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra.

Este contenido fue valorado muy positivamente ($\bar{x} = 5, s = 0$), (véase tabla 4.7). Escogemos el ítem 6. Descartamos el ítem 4 porque uno de los distractores atrae a más del 60 % de los alumnos que lo contestaron. Otro inconveniente en relación al ítem 4, manifestado por uno de los expertos, es el uso del término precisión. *La palabra "precisión" en el lenguaje cotidiano es virtuosa es decir es deseable tener alta precisión. Considero que por sentido común podría responderse acertadamente este ítem, sin que la persona lo esté asociando con las ideas de estimación por intervalos. .*" (experto 6). También descartamos el ítem 5 porque resultó muy difícil para los alumnos.

Tabla 4.11. Resultados en los ítems relacionados con el contenido 2

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I4		1		1	7	4.556	1.014	.25	48	Domina el distractor c(.63)
I5				5	4	4.444	0.527	.19	42	
I6			2	5	2	4.000	0.707	.57	42	

Contenido 3: El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta.

Este contenido fue valorado muy positivamente ($\bar{x} = 5, s = 0$). (vease tabla 4.7). Incluye los ítems: 7.8 y 9. Escogemos el ítem 7. Descartamos el ítem 8 porque presenta un poder discriminativo bajo y el ítem 9 porque resultó difícil para los alumnos.

Tabla 4.12. Resultados en los ítems relacionados con el contenido 3

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I7				1	8	4.889	0.333	.70	48	
I8			2	3	4	4.222	0.833	.57	42	índice de discriminación (-.15)
I9				2	7	4.778	0.441	.20	48	

Contenido 4: El ancho de los intervalos de confianza aumenta cuando aumenta la varianza.

Este contenido con una valoración media de los expertos de $\bar{x} = 4.88$ y $s = 0.33$ incluye los ítems 10 y 11. Los expertos valoraron exactamente igual ambos ítems. Elegimos el ítem 11 que tiene mejor índice de dificultad y un buen poder discriminativo.

Tabla 4.13. Resultados en los ítems relacionados con el contenido 4

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I10				1	8	4.889	.333	.29	48	Distractor d (.33)
I11				1	8	4.889	.333	.42	42	Índice de discriminación (.88)

Contenido 5: Significado del nivel de confianza (variación del intervalo en diferentes muestras).

Este contenido fue valorado con una media $\bar{x} = 4.88$ y $s = 0.33$ incluye los ítems: 12, 13 y 14. De los ítems 12 y 13 nos quedamos con el ítem 13 porque fue bien valorado por los expertos y tiene un adecuado poder de discriminación. El ítem 12, no obstante que tiene buenos indicadores en el índice de dificultad y valoración media de los expertos, se desecha por tener un pobre poder discriminativo. El ítem 14 fue mal valorado por los expertos, por lo cual lo descartamos. Otro asunto que llamó la atención del ítem 14 fue en relación al contexto. Uno de los expertos señaló…En *el ítem 14 me parece interesante el planteo pero dado que muchas investigaciones aconsejan el trabajo con datos más reales y, considerando que son alumnos mexicanos, ¿no sería mejor cambiar el contexto y ubicarlo en empresas y dinero mexicano?* (experto 1).

Tabla 4.14. Resultados en los ítems relacionados con el contenido 5

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I12		1	1	4	3	4.000	1.000	.64	48	Índice de discriminación (.01)
I13		1	1	2	5	4.222	1.093	.76	42	Índice de discriminación (.63)
I14		1	4	3	1	3.444	.882	.14	42	Domina el distractor b (.52)

Contenido 6: Estimar la media de una población normal o en una muestra grande con σ conocida.

En este contenido se incluyen los ítems 15 y 16 (valoración por los expertos con una media $\bar{x} = 4.66$ y $s = 0.50$). Escogemos el ítem 15 porque fue bien valorado por los expertos y tiene buenos índices de dificultad y de discriminación. El ítem 16 lo descartamos porque una de las opciones no fue elegida.

Tabla 4.15. Resultados en los ítems relacionados con el contenido 6

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I15			1	5	3	4.222	.667	.68	48	Índice de discriminación (.88)
I16				7	2	4.333	.500	.47	42	Distractor a) no fue elegido

Contenido 7: Estimar la media de una población aproximadamente normal cuando, σ es desconocida.

Este contenido fue valorado positivamente ($\bar{x} = 4.88, s = 0.33$) . Incluye los ítems 17 y 18. Los dos ítems fueron bien valorados por los expertos, tienen buenos resultados en la pasación a los alumnos y tienen buen poder discriminativo, por lo que decidimos conservarlos. Nos quedaremos con el ítem 18 y también pediremos al alumno que interprete el resultado atendiendo a la sugerencia de uno de los expertos: "*Soy partidaria de hacer hincapié en la interpretación de los intervalos más que en su cálculo*" (experto 4).

Tabla 4.16. Resultados en los ítems relacionados con el contenido 7

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I17				5	4	4.556	.527	.58	48	Índice de discriminación (.53)
I18				3	6	4.667	.500	.40	42	Índice de discriminación (.82)

Contenido 8: Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande.

En este contenido (valoración de los expertos $\bar{x} = 4.88, s = 0.33$) se incluyen los ítems 19 y 20. Decidimos conservar el 19 porque tiene mejor poder discriminativo y su índice de dificultad están dentro de lo que se considera un rango aceptable (Osterlind, 1989, p. 269).

Tabla 4.17. Resultados en los ítems relacionados con el contenido 8

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I19			2	3	4	4.222	.833	.67	48	Índice de discriminación (.63)
I20				2	7	4.778	.441	.83	42	Índice de discriminación (.38)

Contenido 9: Estimar una proporción

Fue valorado positivamente ($\bar{x} = 4.88, s = 0.33$) Decidimos conservar el 21, aunque ambos fueron bien valorados por los expertos, tienen buen poder discriminativo y aceptables índices de facilidad (Losada y López Feal, 2003). La razón de escoger el 21 es tener un ítem abierto para el cálculo del intervalo de confianza de la proporción.

Tabla 4.18. Resultados en los ítems relacionados con el contenido 9

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I21				5	4	4.444	.527	.73	44	Índice de discriminación (.57)
I22			1	1	7	4.667	.707	.72	47	Índice de discriminación (.46)

Contenido 10: Estimar una varianza

En este contenido (valoración de los expertos $\bar{x} = 4.77, s = 0.44$) se incluyen los ítems 23 y 24. Los dos ítems fueron bien valorados por los expertos, tienen buen poder discriminativo y aceptables índices de dificultad (Losada y López Feal, 2003, p.167). Decidimos conservar el 23, mejorando la redacción.

Tabla 4.19. Resultados en los ítems relacionados con el contenido 10

Ítem	Valoración expertos							Índice Dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I23				5	4	4.444	.527	.41	48	
I24			2	2	5	4.333	.866	.40	42	Distractor b (.36)

Contenido 11: Comparar las medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes

Este contenido fue valorado con una media $\bar{x} = 4.44$ y s = 0.72. Incluye los ítems 25 y 26. Aunque tiene un pobre poder discriminativo decidimos conservar el ítem 25 porque tiene buen índice de dificultad y fue bien valorado por los expertos. Además de mejorar la redacción, para aumentar la discriminación pediremos al alumno que interprete el resultado en este ítem 25.

Tabla 4.20. Resultados en los ítems relacionados con el contenido 11

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I25		1	1	2	5	4.222	1.093	.45	44	índice de discriminación (.05)
I26			1	2	6	4.556	0.726	.81	47	índice de discriminación (.11)

Contenido 12: Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas

Este contenido fue valorado positivamente ($\bar{x} = 4.44, s = 0.72$). Incluye los ítems 27 y 28. Decidimos conservar el ítem 28 porque tiene buen índice de dificultad y un buen poder discriminativo. Pediremos al alumno que además interprete el resultado en el ítem 28. Descartamos el ítem 27 porque fue valorado por los expertos con menos de 4 puntos en promedio.

Tabla 4.21. Resultados en los ítems relacionados con el contenido 12

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I27		2	2	1	4	3.778	1.302	.45	44	índice de discriminación (.55)
I28		1	1	3	4	4.111	1.054	.68	47	índice de discriminación (.41)

Contenido 13: Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ desconocidas, muestras independientes grandes.

Este contenido con una valoración media de los expertos de $\bar{x} = 4.44$ y $s = 0.72$ incluye los ítems 29 y 30. Conservamos el ítem 29 porque el ítem 30 resultó muy fácil para la muestra de alumnos. En relación al ítem 29 una cuestión que tenemos que revisar en la redacción es el uso de palabras como "alta confianza ". Los expertos 6 y 8 lo señalaron*…"alta confianza" y "confianza moderada" son calificativos subjetivos.*

Tabla 4.22. Resultados en los ítems relacionados con el contenido 13

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I29	1	1		1	6	4.111	1.537	.59	44	Índice de discriminación (.64)
I30			2	3	4	4.222	0.833	.85	47	Índice de discriminación (.31)

Contenido 14: Comparar dos varianzas poblacionales.

Los expertos valoran este contenido con una media de $\bar{x} = 4.77$ y $s = 0.44$. Incluye los ítems 31 y 32. Decidimos conservar solamente el ítem 31 porque el ítem 32, aunque tiene un mejor índice de dificultad, tiene un pobre poder discriminativo.

Tabla 4.23. Resultados en los ítems relacionados con el contenido 14

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I31		1		3	5	4.333	1.000	.80	44	Índice de discriminación (.21)
I32				4	5	4.556	0.527	.64	47	Índice de discriminación (-.03)

Contenido 15: Elegir un modelo de distribución muestral del estadístico.

La puntuación media de la valoración de los expertos de este contenido fue de $\bar{x} = 4.11$ y $s = 0.92$. Incluye los ítems 33 y 34. Escogemos el ítem 33 para la prueba piloto. Puesto que el ítem 34 tiene un grado de dificultad importante y un pobre poder discriminativo, lo descartamos. Además de que presenta algunos otros inconvenientes: domina el distractor c, el distractor d no fue elegido y fue mal valorado por los expertos.

Tabla 4.24. Resultados en los ítems relacionados con el contenido 15

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I33			3	3	3	4.000	0.866	.57	48	
I34		2	2	3	2	3.556	1.130	.21	42	Domina el distractor c(.71) y el distractor d no fue elegido

Contenido 16: Determinar valores críticos en la distribución del estadístico.

Este contenido recibió una valoración media de $\bar{x} = 4.22,$ y $s = 0.66$. Incluye los ítems 35 y 36. El ítem 36 tiene un índice de dificultad aceptable, pero debido a que una de las opciones no fue elegida por los alumnos, lo descartamos. Además de que algunos expertos recomendaban no tomarlo en cuenta reflejado claramente en la valoración media de menos de 4 puntos. Decidimos dejar el ítem 35 por la buena valoración del contenido y porque pensamos que está en la frontera de índices de dificultad aceptables (Losada y López Feal, 2003, p.167).

Tabla 4.25. Resultados en los ítems relacionados con el contenido 16

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I35		2	1	5	1	3.556	1.014	.75	48	
I36		1	3	2	3	3.778	1.093	.59	42	Distractor d no fue elegido

Contenido 17: Interpretación de intervalos de confianza utilizando un programa de ordenador

La puntuación media de este contenido valorado por los expertos fue $\bar{x} = 4.77$ y $s = 0.66$. Incluye los ítems 37 y 38. Ambos ítems fueron bien valorados y tienen un índice de dificultad aceptable. Se decide conservar el 38 porque tiene mejor índice de discriminación (Hinojosa y Góngora, 2006).

Tabla 4.26. Resultados en los ítems relacionados con el contenido 17

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I37			1	4	4	4.333	.707	.72	48	Índice discriminación (.27)
I38			1	4	4	4.333	.707	.76	42	Índice discriminación (.63)

Contenido 18: Interpretar gráficos de intervalos de confianza

Este contenido fue valorado muy positivamente $(\bar{x} = 5, s = 0)$. Incluye los ítems 39 y 40. Se decidió conservar para la prueba piloto el ítem 39. El ítem 40 aunque fue bien valorado por los expertos y tiene aceptable índice de dificultad (Osterlind, 1989) tiene un pobre poder discriminativo por lo que se decidió desecharlo.

Tabla 4.27. Resultados en los ítems relacionados con el contenido 18

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I39			1	2	6	4.556	.726	.68	48	Índice de discriminación(.75)
I40			1	2	6	4.556	.726	.78	42	Índ disc (-.35)

4.8 CONCLUSIONES SOBRE LA CONSTRUCCIÓN DEL CUESTIONARIO

En el estudio de los ítems, el juicio de expertos, por medio de un procedimiento estructurado, ha permitido emparejar los ítems con el dominio de conductas (habilidades y conocimientos) que queremos observar. Este emparejamiento ha determinado que el contenido de nuestro futuro instrumento de medición sea relevante y representativo de un universo de observaciones admisibles ligadas a la comprensión de intervalos de confianza.

Finalizado este estudio de los ítems, con el doble procedimiento de juicio de expertos y de prueba empírica, nos encontramos con un cuestionario validado que consta de un total de 18 ítems que cubren las unidades de contenido fijadas en la definición semántica de la variable. Este cuestionario será ensayado con una nueva muestra de estudiantes de tamaño suficiente para llevar a cabo la validación completa del cuestionario (estudios de validez y fiabilidad). Asimismo, en el capítulo 6 llevaremos a cabo un estudio de evaluación (sobre la misma muestra usada en la validación) que permitirá una primera aproximación a los errores y dificultades de los estudiantes sobre el intervalo de confianza.

En el ensayo piloto que hemos realizado, los resultados dan cuenta que nuestros ítems tienen una dificultad e índices de discriminación aceptables. Atribuimos la obtención de estos indicadores satisfactorios a la fuerte motivación por dar su mejor esfuerzo de los estudiantes que participaron en la muestra. Motivación originada por el interés y entusiasmo de los profesores que participaron activamente en el proceso de enseñanza de nuestro objeto de estudio.

Esos resultados del ensayo piloto nos revelan además las fuerzas y debilidades de los ítems probados y nos han proporcionado un criterio racional de decisión en la elección de los que conformarán el cuestionario. Asimismo nos proporcionan una primera evidencia de validez de contenido que se comentarán con mayor detalle en el capítulo 5.

CAPITULO 5
VALIDACIÓN DEL CUESTIONARIO

5.1. INTRODUCCIÓN

Una vez elegidos los ítems que se incluirían en el instrumento, se llevó a cabo una prueba del mismo, con objeto de evaluar sus características psicométricas y comprobar que era útil para los objetivos para los que fue diseñado. Al mismo tiempo se utilizó la misma muestra para realizar un estudio de evaluación que se describe en el capítulo 6 y que permitirá describir las dificultades y conflictos semióticos más frecuentes entre nuestros estudiantes.

Puesto que el fin principal de la investigación es llevar a cabo una evaluación de los conocimientos de los estudiantes a partir de las respuestas al cuestionario se está sujeto a errores sistemáticos (sesgos) y aleatorios, como en todo proceso de inferencia (Carmines y Zeller, 1979; Thorndike, 1989). Los análisis de validez y fiabilidad, incluidos en este capítulo atienden a estas dos dimensiones.

En primer lugar se lleva a cabo un estudio de validación del instrumento, considerando la validez como un constructo unitario (Meliá, 2001). En consecuencia las pruebas de validación son parte de un proceso continuo de acumulación de evidencias de validez que ha de continuar en el futuro. Más en concreto, las evidencias de validez se recogen por tres métodos diferentes: validez de contenido, validez discriminante de los ítems y validez de constructo, siguiendo la metodología descrita en Díaz (2007). También se analiza el control del error aleatorio, desde una doble vertiente:

- Desde la teoría clásica de los tests (Muñiz, 1994, Martínez Arias, 1995), se estimará el coeficiente de fiabilidad de consistencia interna Alfa así como el coeficiente θ basados en los resultados del análisis factorial y que es adecuado cuando el constructo tiene más de una dimensión. Como veremos en el análisis factorial, se confirma nuestra hipótesis de que el constructo "comprensión de intervalos de confianza" es multifactorial, debido a la variedad de unidades de contenido descritas en la definición semántica de la variable (Capítulo 4).

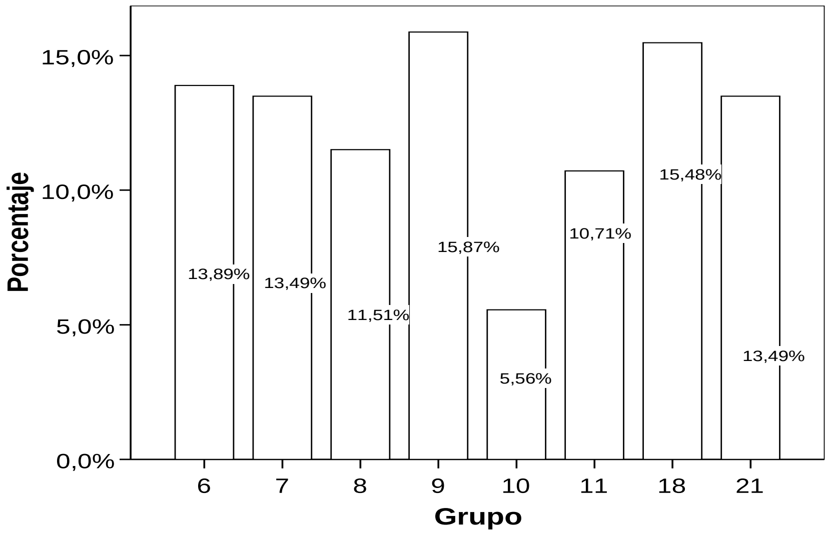
- Además se hace una aproximación al problema de fiabilidad desde la teoría de la generalizabilidad (López Feal, 1986; Feldt y Brennan, 1991; Martínez Arias, 1995), calculando un coeficiente de generalizabilidad, tanto para ítems como para sujetos, atendiendo, de este modo, a estas dos posibles fuentes de error en el muestreo.

5.2. MUESTRAS PARTICIPANTES

En el semestre agosto-diciembre del 2007 fue aplicado nuestro cuestionario a estudiantes de ingeniería del curso de Probabilidad y Estadística en el Campus Monterrey del Sistema Tecnológico de Monterrey, que recién habían estudiado el tema de intervalos de confianza. La muestra fue del tipo no probabilística, aunque intencional, puede ser considerada representativa del grupo objeto de estudio en la investigación (Ghiglione y Matalón, 1991). Los resultados se han presentado en Olivo y Batanero (2007a y b, en prensa). Participaron en las pruebas 252 alumnos del tercer semestre de las distintas carreras de ingeniería impartidas en el Campus Monterrey (Ingeniero Industrial y de Sistemas, Ingeniero en Biotecnología, Ingeniero en Mecatrónica, Ingeniero en Sistemas Electrónicos, etc.), que estaban llevando el curso de Probabilidad y Estadística. Los alumnos tuvieron que responder un cuestionario de 18 ítems; 12 ítems de opción múltiple y seis de respuesta abierta. La calificación media de estos alumnos que participaron en el estudio, en su semestre anterior, fue de 84 (sobre 100). El total de alumnos inscritos en este curso fue de 815.

De los 252 alumnos, 128 alumnos contestaron el cuestionario A y 124 alumnos el cuestionario B. No había diferencia en el contenido de los cuestionarios, solamente el orden de los ítems estaba invertido. Se aplicaron los dos cuestionarios simultáneamente a través de varias sesiones de hora y media, de tal manera que al distribuir los cuestionarios en forma alternada, en la sala donde se probaron estos cuestionarios, se controló la posibilidad de que se copiasen. La muestra de 252 alumnos provino de ocho grupos del curso de Probabilidad y Estadística, cuyos porcentajes de contribución al total de la muestra se aprecia en la Figura 6.1. Esta muestra total se consiguió después de varias semanas de trabajo, durante las cuales se estuvieron aplicando los cuestionarios a grupos que variaban entre 15 y 40 alumnos por sesión de hora y media.

Figura 6.1. Gráfico de barras para los grupos.



En las sesiones de trabajo, los profesores que colaboraron con el estudio animaron a los alumnos a que contestaran el cuestionario de una manera responsable, con el máximo cuidado y la mayor seriedad. La respuesta de los alumnos fue satisfactoria, mostrando una buena disposición por colaborar con el estudio.

5.3. MATERIAL

El material utilizado es el cuestionario en su versión final, que se compone de los ítems seleccionados mediante pruebas de ítems y juicio de expertos, con el proceso descrito en el Capítulo 4. De este cuestionario se elaboraron dos versiones (A y B) que difieren sólo en el orden de presentación de los ítems, con el fin de evitar que los alumnos copien la respuesta y que los ítems finales tengan mayor proporción de no-respuesta debido al posible cansancio. Los cuestionarios son exactamente iguales, salvo el orden invertido de presentación de los ítems y se incluyen en el Anexo 3. En la Tabla 5.1 se presenta el cuestionario en versión A.

Tabla 5.1. Cuestionario sobre intervalos de confianza

Ítem 1. El intervalo de confianza del 50% para la media de una población μ es:

a. El rango dentro del cual caen el 50% de los valores de la media de la muestra $\bar{x}$.
b. Un intervalo más ancho que el intervalo de confianza del 95%.
c. **Un intervalo de valores calculado a partir de los datos de la muestra. En el 50% de las muestras de una población, el intervalo calculado contiene a la media de la población.**
d. Dos veces más ancho que el intervalo de confianza del 100%.

Ítem 2. Comparado a los intervalos de confianza calculados en muestras de tamaño n=4 en una población normal, el ancho de los intervalos de confianza de la media de la población calculado en muestras de tamaño n = 50:

a. Variará más que los anchos de los intervalos para muestras de tamaño n =4.
b. **Variará, pero no tanto como lo hicieron los anchos de los intervalos para muestras de tamaño *n* =4.**
c. Tomarán valores parecidos.

Ítem 3. Si, manteniendo todos los demás datos fijos, el nivel de confianza se reduce (por ejemplo de 90% a 80%):

a. El intervalo de confianza no cambia.
b. El intervalo de confianza será más ancho.
c. **El intervalo de confianza será más angosto.**
d. El cambio en el intervalo de confianza no es predecible.

Ítem 4. Explica cómo varía la anchura del intervalo de confianza de la media si, conservando el mismo tamaño de muestra y el mismo coeficiente de confianza, tomamos una población con varianza cuatro veces mayor.

Ítem 5. En un intervalo de confianza del 95% para la media:

a. Si se toman muchas muestras y con cada una se construye el intervalo, la media muestral $\overline{x}$ caerá dentro del intervalo de confianza el 95% de las veces.
b. La probabilidad de que $\overline{x}$ caiga dentro de un intervalo de confianza calculado de una muestra específica es 0.95.
c. **Si se toman muchas muestras de igual tamaño, el 95% de los intervalos de confianza calculado contendrían a μ.**

Ítem 6. La media muestral de 100 observaciones en una prueba de matemáticas es 75. Encuentre el intervalo de confianza al 95% para la media de la población, asumiendo que σ =7:

a. (61.28, 88.72).
b. **(73.63, 76.37).**
c. (68, 82).
d. (74.3, 75.7).

Ítem 7. Un fabricante asegura que sus garrafones, contienen un litro de cloro puro. Al tomar una muestra de 16 garrafones se determinó que en promedio contenían 0.94 litros de cloro puro, con desviación estándar de la muestra de 0.097. Construir un intervalo de confianza al 95 %, para el verdadero contenido promedio de litros de cloro puro. No se conoce la desviación típica de la población. (La distribución del contenido de cloro por botella puede considerarse normal).

Ítem 8. Se han obtenido los siguientes datos de emisión diaria de óxidos de azufre, para una muestra de tamaño n=100, media: $\overline{x}$ = 18 y varianza muestral s^2=36. Elabore un intervalo de confianza de 95% para la verdadera emisión diaria promedio de óxidos de azufre:

a. (17.016, 18.984).
b. **(16.824, 19.176).**
c. (6.24, 29.76).
d. (8.16, 27.84).

Ítem 9. En una muestra aleatoria de 100 rodamientos, 10 tienen un acabado de especificaciones defectuoso. Calcular el intervalo de confianza de 95% para la proporción verdadera de rodamientos defectuosos.

Ítem 10. Sea σ^2 la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es s^2=13700, n=16. Calcular el intervalo de confianza de 95% para σ. Considere la muestra tomada de una población Normal

Ítem 11. La siguiente tabla contiene un resumen de información sobre la resistencia a la compresión de cubos (N/ mm^2) para especimenes de concreto :

	Tamaño muestral	Media muestral
Tipo1	68	26.99
Tipo2	74	37.56

Suponga que las desviaciones estándar poblacionales de ambos grupos son σ_1=4.89 y σ_2=6.43 respectivamente. Calcule un intervalo de confianza de 95% para hallar la diferencia entre el verdadero promedio de resistencia en el Tipo1 y el verdadero promedio de resistencia en el Tipo 2:

a. (-13.02, -8.12) usar 2,57.
b. **(-12.437, -8.70).**
c. (-31.32, 10.18) no dividir entre n y m.
d. (-18.64, -2.5) no dividir entre n y m y no multiplicar por 2.57.

Ítem 12. Se compararon dos soluciones de grabado diferentes, usando dos muestras aleatorias de tamaño 10 cada una. Las poblaciones son normales y tienen la misma varianza. Los resultados de la rapidez de grabado fueron:
Solución 1: $\bar{x}$ = 9.97 y s = .422
Solución 2: $\bar{x}$ =10.4 y s = .073
s_p=0,34 (desviación típica conjunta).
Calcule un intervalo de confianza de 90% para la diferencia de las medias de la rapidez de grabado.

Ítem 13. La tabla siguiente resume algunos datos de un experimento realizado para estudiar varias características de tornillos de anclaje:

		Resistencia al corte	
Diámetro de tornillo	Tamaño muestra	Media muestral	Desviación estándar
3/8	100	4.25	1.3
½	100	7.25	1.7

Calcule e interprete un intervalo de confianza de 95% para la diferencia del verdadero promedio de resistencias al corte:

a. **(-3.41, -2.58): en el 95% de las muestras del mismo tamaño en esta población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.**
b. (-3.41, -2.58): la probabilidad de que la verdadera diferencia promedio de la resistencia al corte cae en el intervalo (-3.41, -2.58) es 0.95.
c. (-3.35, -2.65): en el 95% de las muestras del mismo tamaño en esta población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.
d. (-3.35, -2.65): la probabilidad de que la verdadera diferencia promedio de la resistencia al corte cae en el intervalo (-3.35, -2.65) es 0.95.

Ítem 14. Una compañía quiere seleccionar el proceso de pulido que presente la variabilidad menor. Una muestra aleatoria de n_1=16 piezas del primer proceso da como resultado una desviación estándar muestral s_1=5 micropulgadas, y una muestra aleatoria de n_2=11 piezas del segundo proceso da como resultado una desviación estándar muestral s_2=4 micropulgadas. Establezca un intervalo de confianza de 90% para σ_1^2/σ_2^2 , suponiendo que los dos procesos son independientes y que la aspereza superficial tiene una distribución normal. ¿Cuál de los dos procesos recomendaría usted?

a. Como todos los valores están dentro del intervalo (.54, 3.96) recomendaría el proceso 1.
b. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría el proceso 2 (se cambian los grados de libertad del numerador y denominador).
c. **Como el cociente de valores está dentro del intervalo (.54, 3.96) recomendaría cualquiera de los dos.**
d. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría cualquiera de los dos (se cambian los grados de libertad del numerador y denominador).

Ítem 15. La distribución muestral utilizada en la construcción de intervalos de confianza para la varianza en muestras pequeñas, tomadas de una población normal, es:

a. Distribución *t* de Student.
b. Distribución Ji-cuadrada.
c. Distribución Normal.
d. Distribución *F*.

Ítem 16. El nivel de confianza es 0.95, para un intervalo de confianza para la media poblacional con desviación estándar poblacional desconocida para un grupo de puntajes distribuido normalmente de tamaño $n = 20$. Los valores críticos han de ser:

a) -1.65 y 1.65 uso de normal estándar.
b) -1.96 y 1.96 uso de normal estándar.
c) -2.093 y 2.093 uso de distribución *t* con 19 grados de libertad.
d) -2.085 y 2.085 uso de distribución *t* con 20 grados de libertad.

Ítem 17 La siguiente salida de computadora presenta dos muestras simuladas de dos poblaciones normales. La población 1 con $\mu = 90$ y $\sigma = 10$ y la población 2 con $\mu = 92$ y $\sigma = 10$

Muestra 1:
83.3195 87.6793 86.7831 95.0518 92.9781 86.6457 85.1305 97.5013 83.1112 82.2751 82.7831 90.2786 89.5876 71.2591 82.0282 90.6264

Muestra 2:
82.312 95.098 92.598 85.959 91.319 108.130 90.392 90.074 78.789 100.923 85.601 89.861 78.685 100.354 81.267 101.432

	N	Media	D. Típica	Error típico
C1	16	86.69	6.25	1.6
C2	16	90.80	8.70	2.2

95% IC para mC1 - mC2: (-9.6; 1.4)

Escriba el intervalo de confianza que se obtuvo para la diferencia de medias e interprete el resultado.

Ítem 18. Considere el gráfico siguiente del rendimiento medio de cebada en 1980,1984 y 1988 junto con un intervalo de 95% de confianza respectivos:

```
Intervalos del 95% para la media

Año N Media StDev        ----+---------+---------+---------+--
1980 6 184.00 2.61        (----*---)
1984 5 212.40 14.36                          (-----*----)
1988 5 182.40 1.82      (-----*----)
                         ----+---------+---------+---------+--
Desv. Típica conjunta =.19 180       195       210       225
```

¿Cuál de las siguientes afirmaciones es verdadera?:

a. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay buena evidencia que las medias de las muestras difieran.
b. La estimación de la media de la población en 1980 es menos precisa que en 1988.
c. Puesto que los intervalos de confianza para 1980 y 1984 no se solapan, hay poca evidencia que las medias de las poblaciones respectivas difieran.
d. **Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay poca evidencia que las medias de las poblaciones difieran.**

5.4. ESTUDIOS DE VALIDACIÓN

5.4.1. VALIDEZ DE CONTENIDO

En primer lugar se inició el estudio de la validez de contenido para asegurar que el instrumento recoge una muestra representativa de los contenidos que se pretenden evaluar (Messick, 1998) y que fueron especificados mediante la tabla de contenidos (Sección 4.4). La validez de contenido es justificada por el grado en que el contenido del instrumento refleja en forma representativa una muestra de un universo más amplio de ítems posibles que constituyen el contenido (Martínez Arias, 1995). En nuestro caso serían todos los posibles ítems sobre comprensión de intervalos de confianza.

En la construcción del cuestionario se proporcionaron tres tipos de evidencia de validez de contenido, siguiendo la metodología de Díaz (2007):

1. La definición semántica de la variable se basó en un estudio objetivo y exhaustivo de libros de texto usados en ingeniería, a partir del cual se construyeron las especificaciones del contenido. Esto contribuiría a recoger evidencias de *validez curricular*, es decir, el grado en que el cuestionario es relevante para los objetivos del currículo de la formación de ingenieros (Martínez Arias, 1995).
2. Además se hizo una planificación cuidadosa de los ítems que se incluirán en el cuestionario, que fueron tomados de la bibliografía previa o los libros de texto y se realizó un estudio de cómo estos ítems contribuyen a evaluar el constructo subyacente. De este modo llegamos a la tabla de especificaciones del contenido (Sección 4.4).
3. También se usó un juicio en el que participaron 9 expertos en el tema para analizar la congruencia entre los ítems y las especificaciones (Osterlind, 1989). Para ello se emparejan los ítems con objetivos y se da una valoración numérica. Los resultados se resumen y presentan en el Capítulo 4 y muestran una alta valoración y grado de acuerdo en los contenidos finalmente seleccionados para construir el instrumento. También se han proporcionado evidencias de la relevancia y representatividad de los ítems mediante la alta valoración y grado de acuerdo que los jueces han concedido a la idoneidad de los ítems finalmente seleccionados para evaluar los contenidos especificados.

Análisis a priori del contenido del cuestionario

Otro tipo de evidencia de la validez de contenido del cuestionario la proporciona el análisis de las posibles respuestas correctas e incorrectas a los diferentes ítems que componen el cuestionario, y la identificación de los conceptos y propiedades necesarios para su resolución, así como los posibles errores que llevarían a una solución errónea y comprobando que todo ello cubre la definición semántica de la variable.

A continuación se lleva a cabo un análisis teórico de las respuestas esperadas a cada uno de los ítems que constituyen el cuestionario. Se usa el orden en que aparecen en el primer cuestionario, ya que el segundo contiene los mismos ítems, aunque colocados en orden inverso.

Ítem 1. El intervalo de confianza del 50% para la media de una población μ es:

a. El rango dentro del cual caen el 50% de los valores de la media de la muestra $\bar{x}$.
b. Un intervalo más ancho que el intervalo de confianza del 95%.
c. **Un intervalo de valores calculado a partir de los datos de la muestra. En el 50% de las muestras de una población, el intervalo calculado contiene a la media de la población.**
d. Dos veces más ancho que el intervalo de confianza del 100%.

Este ítem se ha tomado de Cruise, Dudley y Thayer (1984), adaptándolo para nuestro trabajo y traduciéndolo. El contenido primario evaluado es el *1 Definición de intervalo de confianza*. La opción correcta es la c) que da la interpretación correcta del intervalo (tanto por ciento de muestras en la misma población e igual tamaño cuyo intervalo calculado cubre al parámetro). Requiere también como conocimientos secundarios el significado del *coeficiente de confianza (variación del intervalo en diferentes muestras).* Las opciones restantes evalúan los siguientes posibles conflictos semióticos:

- Opción a: Pensar que el intervalo se refiere a la media de la muestra y no de la población, es decir confusión entre *estadístico* y *parámetro*. Esta opción es errónea, puesto que la media muestral siempre cae dentro del intervalo de confianza, no el 50% de las veces.
- Opciones b y d: Conflicto en la comprensión del efecto del coeficiente de confianza sobre el tamaño del intervalo, puesto que el intervalo aumenta con la confianza (sería un contenido secundario de este ítem).

Ítem 2. Comparado a los intervalos de confianza calculados en muestras de tamaño n=4 en una población normal, el ancho de los intervalos de confianza de la media de la población calculado en muestras de tamaño n = 50:

a. Variará más que los anchos de los intervalos para muestras de tamaño n =4.
b. **Variará, pero no tanto como lo hicieron los anchos de los intervalos para muestras de tamaño n=4.**
c. Tomarán valores parecidos.

Este ítem es una adaptación de otro elaborado por Garfield, delMas y Chance (2004). El contenido primario evaluado es el *2 El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra*. La opción correcta es la b) que da la interpretación correcta del efecto que se produce sobre el ancho del intervalo al disminuir o aumentar el tamaño de la muestra. Ello es debido a que el intervalo se obtiene sumando y restando a la media muestral el producto del valor crítico por el error estándar de la media, que es igual a la desviación típica de la población, dividida por la raíz cuadrada del tamaño de la muestra. Por tanto, el error estándar será menor a mayor tamaño de muestra, y en consecuencia la anchura del intervalo. Entonces un contenido secundario para este ítem será la *estimación de la media de una población normal, tanto en caso de desviación típica conocida o desconocida*. Las opciones restantes evalúan los siguientes conflictos semióticos:

- Opción a: Pensar justamente lo contrario al efecto verdadero. Conflicto causado por el olvido de la fórmula del error estándar de la media.
- Opción c: Detecta completa ignorancia del tamaño muestral sobre la variabilidad de la distribución muestral y puede ser un conflicto debido a la heurística de representatividad descrita por Kahneman, Slovic y Tversky (1982).

Ítem 3. Si, manteniendo todos los demás datos fijos, el nivel de confianza se reduce (por ejemplo de 90% a 80%):

a. El intervalo de confianza no cambia.
b. El intervalo de confianza será más ancho.
c. **El intervalo de confianza será más angosto.**
d. El cambio en el intervalo de confianza no es predecible.

Este ítem se ha tomado de Cruise, Dudley y Thayer (1984), adaptándolo para nuestros fines. El contenido primario evaluado es el *3 El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta* La opción correcta es la c) que da la interpretación correcta del efecto que se produce al reducir el nivel de confianza. Corresponde a la propiedad descrita por delMas, Garfield y Chance (2004): Un incremento en el tamaño de la muestra conduce a decrecer el ancho del intervalo, por lo que grandes muestras tienen anchos de intervalos más estrechos que muestras más pequeñas (todos los otros elementos permaneciendo constantes). Las opciones restantes evalúan los siguientes conflictos:

- Opción a: Suponer que no hay relación entre nivel de confianza y ancho del intervalo. Este conflicto puede estar asociado a la creencia en la replicación de resultados descrita por Estes (1997).
- Opción b: Conflicto consistente en creer en el efecto opuesto, un intervalo ancho significa menos confianza, conflicto descrito por delMas, Garfield y Chance (2004) y que supone confundir confianza y precisión.
- Opción d: El estudiante exhibe una clara ausencia de conocimiento sobre este contenido, no siendo capaz de relacionar el ancho del intervalo con el coeficiente de confianza.

Ítem 4. Explica cómo varía la anchura del intervalo de confianza de la media si, conservando el mismo tamaño de muestra y el mismo coeficiente de confianza tomamos una población con varianza cuatro veces mayor.

Este ítem se ha tomado de Cruise, Dudley y Thayer (1984). El contenido primario evaluado es el *4 El ancho de los intervalos de confianza aumenta cuando aumenta la varianza.* La respuesta correcta debe suponer que la anchura del intervalo de confianza aumenta el doble. Ello es debido a que el intervalo se obtiene sumando y restando a la media muestral el producto del valor crítico por el error estándar de la media, que es igual a la desviación típica de la población, dividida por la raíz cuadrada del tamaño de la muestra. Como la desviación típica es la raíz cuadrada de la varianza, al aumentar la varianza cuatro veces, la fórmula de cálculo del intervalo se mantiene, excepto que la desviación típica va multiplicada por dos. Por tanto, la anchura del intervalo será doble. En consecuencia un contenido secundario para este ítem será la *estimación de la media de una población normal, tanto en caso de desviación típica conocida o desconocida*.

Algunas respuestas erróneas pueden incluir un factor raíz cuadrada de 2 entre los elementos del aumento, al producirse un conflicto consistente en confundir la varianza con la desviación típica en la fórmula del error típico o bien no indicar exactamente la proporción en que aumenta el intervalo.

Ítem 5. En un intervalo de confianza del 95% para la media:

a. Si se toman muchas muestras y con cada una se construye el intervalo, la media muestral $\bar{x}$ caerá dentro del intervalo de confianza el 95% de las veces.

b. La probabilidad de que $\bar{x}$ caiga dentro de un intervalo de confianza calculado de una muestra específica es 0.95.

c. **Si se toman muchas muestras de igual tamaño, el 95% de los intervalos de confianza calculado contendrían a μ.**

Este ítem se ha tomado de Cruise, Dudley y Thayer (1984). El contenido primario evaluado es el *5. Significado del nivel de confianza (variación del intervalo en diferentes muestras).* La opción correcta es la c) que da la interpretación correcta del intervalo (tanto por ciento de muestras en la misma población e igual tamaño cuyo intervalo calculado cubre al parámetro). Además expresa que el intervalo se refiere a la media de la población. Las opciones restantes evalúan los siguientes conflictos:

- Opción a: Pensar que el intervalo se refiere a la media de la muestra y no de la población, es decir confusión entre *estadístico* y *parámetro*. Esta opción es errónea, puesto que la media muestral siempre cae dentro del intervalo de confianza, no el 95% de las veces.
- Opción b: Pensar que el intervalo da el rango de valores en que cae la media de la muestra, es decir confusión entre *estadístico* y *parámetro*. Sería además dar la interpretación bayesiana del intervalo de confianza (Lecoutre, 1999).

Ítem 6. La media muestral de 100 observaciones en una prueba de matemáticas es 75. Encuentre el intervalo de confianza al 95% para la media de la población, asumiendo que σ =7.

a. (61.28, 88.72)
b. **(73.63, 76.37)**
c. (68, 82)
d. (74.3, 75.7)

Este ítem se ha tomado de Cruise, Dudley y Thayer (1984). El contenido primario evaluado es el *6 Estimar la media de una población normal o en una muestra grande con σ conocida*. Como contenido secundarios contiene: *definición del intervalo de confianza* y *elegir un modelo de distribución muestral del estadístico.* La opción correcta es la b) en que se calcula el intervalo sumando y restando a la media el error típico luego multiplicado por el valor crítico. Las opciones restantes evalúan los siguientes conflictos:

- Opción a: Conflicto consistente en confundir la fórmula de cálculo del error estándar, no dividiendo la desviación típica por el tamaño de la muestra. Esta confusión ha sido descrita por Alvarado (2007) en su estudio sobre el teorema central del límite.
- Opción c: Conflicto consistente en calcular el intervalo simplemente sumando y restando a la media la desviación típica poblacional.
- Opción d: Conflicto consistente en calcular el intervalo simplemente sumando y restando a la media el error típico sin multiplicar por el valor crítico.

Ítem 7. Un fabricante asegura que sus garrafones, contienen un litro de cloro puro. Al tomar una muestra de 16 garrafones se determinó que en promedio contenían 0.94 litros de cloro puro, con desviación estándar de la muestra de 0.097. Construir un intervalo de confianza al 95 %, para el verdadero contenido promedio de litros de cloro puro. No se conoce la desviación típica de la población. (La distribución del contenido de cloro por botella puede considerarse normal).

Este ítem se ha tomado de Cruise, Dudley y Thayer (1984). El contenido primario evaluado es el *7 Estimar la media de una población aproximadamente normal cuando, σ es desconocida*. Como contenido secundarios contiene: *definición del intervalo de confianza* y *elegir un modelo de distribución muestral del estadístico* La solución correcta esperada es la que incluye:

1. La ecuación correcta del intervalo de confianza para la media, de acuerdo a la información del ítem.
2. La aplicación la distribución *t* para obtener correctamente los puntos críticos del intervalo, a partir de 15 grados de libertad y nivel de confianza correspondiente, puesto que no se conoce la desviación típica de la población y el tamaño de la muestra es 16.
3. La obtención correcta del intervalo de confianza.

Ítem 8. Se han obtenido los siguientes datos de emisión diaria de óxidos de azufre, para una muestra de tamaño *n*=100, media: $\overline{x}$ = 18 y varianza muestral s^2=36. Elabore un intervalo de confianza de 95% para la verdadera emisión diaria promedio de óxidos de azufre.

a. (17.016, 18.984)
b. **(16.824, 19.176)**
c. (6.24, 29.76)
d. (8.16, 27.84)

Este ítem se ha tomado de Miller, Freund y Johnson (1997, pg. 223). El contenido primario evaluado es el *8 Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande*. Como contenido secundarios contiene: *definición del intervalo de confianza* y *elegir un modelo de distribución muestral del estadístico*. La opción correcta es la b) que captura la obtención correcta del intervalo de confianza para la media en muestras grandes. Las opciones restantes evalúan los siguientes conflictos:

- Opción a: Error consistente en usar 1.64 como punto crítico en vez de 1.96.Se utiliza incorrectamente la distribución normal estándar para obtener el punto crítico.
- Opción c: El conflicto consiste en no dividir la desviación estándar muestral entre *n*.
- Opción d: No se divide la desviación estándar muestral entre *n* y se usa 1.64.

Ítem 9. En una muestra aleatoria de 100 rodamientos, 10 tienen un acabado de especificaciones defectuoso. Calcular el intervalo de confianza de 95% para la proporción verdadera de rodamientos defectuosos.

Este ítem se ha tomado de Montgomery y Runger (2004, pg. 352). El contenido primario evaluado es el *9 Estimar una proporción*. Como contenido secundarios contiene: *definición del intervalo de confianza* y *elegir un modelo de distribución muestral del estadístico.* La solución correcta que el estudiante debe escribir incluye:

1. La ecuación correcta del intervalo de confianza para la proporción para muestras grandes.
2. La aplicación la distribución normal estándar para obtener correctamente los puntos críticos del intervalo, a partir del nivel de confianza correspondiente.
3. La obtención correcta del intervalo de confianza a partir de esa información.

Ítem 10. Sea σ^2 la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es s^2=13700, *n*=16. Calcular el intervalo de confianza de 95% para *σ*.

Este ítem se ha tomado de Devore (2005, pg. 310). El contenido primario evaluado es el *10 Estimar una varianza*. Como contenido secundarios contiene: *definición del intervalo de confianza* y *elegir un modelo de distribución muestral del estadístico.* La solución correcta que el estudiante debe escribir incluye:

1. La ecuación correcta del intervalo de confianza para la varianza para muestras pequeñas.
2. La aplicación la distribución Chi cuadrado para obtener correctamente los puntos críticos del intervalo, a partir del nivel de confianza correspondiente.
3. La obtención correcta del intervalo de confianza a partir de esa información.

Ítem 11. La siguiente tabla contiene un resumen de información sobre la resistencia a la compresión de cubos (N/ mm^2) para especimenes de concreto:

	Tamaño muestral	Media muestral
Tipo1	68	26.99
Tipo2	74	37.56

Suponga que las desviaciones estándar poblacionales de ambos grupos son σ_1=4.89 y σ_2=6,43 respectivamente. Calcule un intervalo de confianza de 95% para hallar la diferencia entre el verdadero promedio de resistencia en el Tipo1 y el verdadero promedio de resistencia en el Tipo 2:

e. (-13.02, -8.12) usar 2,57.
f. **(-12.437, -8.70) usar 1.96.**
g. (-31.32, 10.18) no dividir entre *n* y *m*.
h. (-18.64, -2.5) no dividir entre *n* y *m* y no multiplicar por 1.96.

Este ítem se ha tomado de Devore (2005, pg. 369). El contenido primario evaluado es el *11 Comparar las medias en dos* poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes. Como contenido secundarios contiene: definición del intervalo de confianza y elegir un modelo de distribución muestral del estadístico. La opción correcta b) requiere que el alumno:

1. Escriba la ecuación correcta del intervalo de confianza para la diferencia de medias con σ_1^2 *y* σ_2^2 *conocidas.*
2. Aplique la distribución normal estándar para obtener correctamente los puntos críticos utilizando el nivel de confianza correspondiente. Área a la derecha del punto crítico de .025.
3. Sustituir los valores y obtener correctamente el intervalo de confianza.

Las opciones restantes evalúan los siguientes errores:

- Opción a: Error consistente en usar el valor crítico 2.57 en vez de 1.96.
- Opción c: El error consiste en no dividir entre n y m.
- Opción d: Los errores consisten en no dividir entre n y m y no multiplicar por 1.96.

Ítem 12. Se compararon dos soluciones de grabado diferentes, usando dos muestras aleatorias de tamaño 10. Las poblaciones tienen la misma varianza. Los resultados de la rapidez de grabado fueron:
Solución 1: $\bar{x} = 9.97$ y $s = .422$
Solución 2: $\bar{x} = 10.4$ y $s = .073$
$s_p = 0.34$ (desviación típica conjunta).
Calcule un intervalo de confianza de 90% para la diferencia de las medias de la rapidez de grabado.

Este ítem se ha tomado de Montgomery y Runger (2004, pg. 40). El contenido primario evaluado es el *12 Comparar las medias en dos poblaciones,* $\sigma_1^2 = \sigma_2^2$ *pero desconocidas, muestras independientes pequeñas*. La solución correcta de este ítem requiere que el alumno:

1. Escribir la ecuación correcta del intervalo de confianza para la diferencia de medias con σ_1^2 *y* σ_2^2 desconocidas pero iguales y muestras pequeñas
2. Aplique la distribución *t* para obtener correctamente los puntos críticos utilizando los grados de libertad y el nivel de confianza correspondiente.
3. Sustituir los valores y obtener correctamente el intervalo de confianza.

Ítem 13. La tabla siguiente resume algunos datos de un experimento realizado para estudiar varias características de tornillos de anclaje:

		Resistencia al corte	
Diámetro de tornillo	Tamaño muestra	Media muestral	Desviación estándar
3/8	100	4.25	1.3
½	100	7.25	1.7

Calcule e interprete un intervalo de confianza de 95% para la diferencia del verdadero promedio de resistencias al corte.

a. **(-3.41, -2.58): en el 95% de la muestras del mismo tamaño en esta población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.**
b. (-3.41, -2.58): la probabilidad de que la verdadera diferencia promedio de la resistencia al corte cae en el intervalo (-3.41, -2.58) es 0.95.
c. (-3.35, -2.65): en el 95% de la muestras del mismo tamaño en esta población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.
d. (-3.35, -2.65): la probabilidad de que la verdadera diferencia promedio de la resistencia al corte cae en el intervalo (-3.35, -2.65) es 0.95.

Este ítem es una adaptación a un problema de Devore (1998, 8b, pg. 342). El contenido primario evaluado es el *13 Comparar las medias en dos poblaciones,* $\sigma_1^2 = \sigma_2^2$ *desconocidas, muestras independientes grandes*. La opción correcta es la a) estimación correcta del intervalo y además da la interpretación correcta del intervalo (tanto por ciento de muestras en la misma población e igual tamaño cuyo intervalo calculado cubre al parámetro Las opciones restantes evalúan los siguientes errores:

- Opción b: Se obtiene una estimación correcta del intervalo pero proporciona una interpretación incorrecta del intervalo.
- Opción c: Aunque da la interpretación correcta del intervalo (tanto por ciento de muestras en la misma población e igual tamaño cuyo intervalo calculado cubre al parámetro), se obtiene una estimación incorrecta del intervalo y el error consiste en la obtención de los puntos críticos, se usa 1.64 en vez de 1.96.
- Opción d: Se obtiene una estimación incorrecta del intervalo y además se proporciona una interpretación incorrecta del intervalo.

Ítem 14. Una compañía quiere seleccionar el proceso de pulido que presente la variabilidad menor. Una muestra aleatoria de $n_1 = 16$ piezas del primer proceso da como resultado una desviación estándar muestral $s_1 = 5$ micropulgadas, y una muestra aleatoria de $n_2 = 11$ piezas del segundo proceso da como resultado una desviación estándar muestral $s_2 = 4$ micropulgadas. Establezca un intervalo de confianza de 90% para σ_1^2 / σ_2^2, suponiendo que los dos procesos son independientes y que la aspereza superficial tiene una distribución normal. ¿Cuál de los dos procesos recomendaría usted?

a. Como todos los valores están dentro del intervalo (.54, 3.96) recomendaría el proceso 1.
b. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría el proceso 2 (se cambian los grados de libertad del numerador y denominador).
c. **Como el cociente de valores está dentro del intervalo (.54, 3.96) recomendaría cualquiera de los dos**.
d. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría cualquiera de los dos (se cambian los grados de libertad del numerador y denominador).

Este ítem se ha tomado de Montgomery y Runger (2004, pg. 416) y las opciones son de elaboración propia. El contenido primario evaluado es el *14 Comparar dos varianzas poblacionales*. La opción correcta es la c) que obtiene correctamente el intervalo de confianza del cociente de varianzas poblacionales y además da la interpretación correcta del intervalo. Al variar el intervalo en valores menores de 1 y mayores de 1 cualquiera de los dos procesos se puede recomendar. Las opciones restantes evalúan los siguientes errores:

- Opción a: Se obtiene una estimación correcta del intervalo pero proporciona una interpretación incorrecta del intervalo. Al recomendar solamente el proceso 1 se demuestra ausencia de comprensión del significado del conjunto de valores del intervalo.
- Opciones b: y d: El error consiste en cambiar los grados de libertad en el numerador y denominador, obteniéndose valores críticos incorrectos de F. También se da una interpretación incorrecta del intervalo aunque ese intervalo hubiese resultado correcto.

Ítem 15. La distribución muestral utilizada en la construcción de intervalos de confianza para la varianza en muestras pequeñas es:
a. Distribución *t* de Student.
b. Distribución Ji-cuadrada.
c. Distribución Normal.
d. Distribución *F*.

Este ítem es de elaboración propia. El contenido primario evaluado es el *15 Elegir un modelo de distribución muestral del estadístico*. La opción correcta es la b) que da la distribución muestral correcta para la construcción del intervalo. Las opciones restantes evalúan los siguientes errores:

- Opción a: El error consiste en suponer una distribución que se aplica también en condiciones de muestras pequeñas.
- Opción c: El error consiste en utilizar una distribución que en algunos casos (muestras grandes) se utiliza en la construcción de intervalos de confianza para la varianza.
- Opción d: El error consiste en utilizar una distribución que solamente se aplica para la construcción de intervalos de confianza para cociente de varianzas poblacionales.

Ítem 16. El nivel de confianza es 0.95, para un intervalo de confianza para la media poblacional con desviación estándar poblacional desconocida para un grupo de puntajes distribuido normalmente de tamaño $n = 20$. Los valores críticos han de ser:
a. -1.65 y 1.65 uso de normal estándar
b. -1.96 y 1.96 uso de normal estándar
c. **-2.093 y 2.093 uso de distribución *t* con 19 grados de libertad**
d. -2.085 y 2.085 uso de distribución *t* con 20 grados de libertad

Este ítem es una adaptación a ítem de Cruise, Dudley y Thayer (1984). El contenido primario evaluado es el *16 Determinar valores críticos en la distribución del estadístico*. La opción correcta es la c) que obtiene los puntos críticos correctos del intervalo de confianza para la media a partir del uso de la distribución *t*, para esas condiciones particulares. Las opciones restantes evalúan los siguientes errores:

- Opción a: El error consiste en suponer una distribución que se aplica en condiciones de muestras grandes y además se tiene error en el uso del valor del nivel de confianza.
- Opción b: El error consiste en suponer una distribución que se aplica en condiciones de muestras grandes aunque en esta solución se maneja adecuadamente el valor del nivel de confianza.

- Opción d: Se usa la distribución correcta pero se utiliza incorrectamente el valor de los grados de libertad de la distribución.

Ítem 17. La siguiente salida de computadora presenta dos muestras simuladas de dos poblaciones normales. La población 1 con μ=90, σ=10 y la población 2 con μ=92, σ=10.
Muestra 1:
83.3195 87.6793 86.7831 95.0518 92.9781 86.6457 85.1305 97.5013 83.1112 82.2751 82.7831 90.2786 89.5876 71.2591 82.0282 90.6264

Muestra 2:
82.312 95.098 92.598 85.959 91.319 108.130 90.392 90.074 78.789 100.923 85.601 89.861 78.685 100.354 81.267 101.432

	N	Media	D. Típica	Error típico
C1	16	86.69	6.25	1.6
C2	16	90.80	8.70	2.2

95% IC para mC1 - mC2: (-9.6; 1.4)

Escriba el intervalo de confianza que se obtuvo para la diferencia de medias e interprete el resultado.

Este ítem es una adaptación usando Minitab de problema de Johnson y Kuby (2004, pg.363). El contenido primario evaluado es el *17 Interpretar intervalos de confianza obtenidos de un programa de ordenador*. La solución que se espera del estudiante incluye:

1. Escribir la ecuación correcta del intervalo de confianza para la diferencia de medias a partir de la salida del ordenador.
2. La interpretación correcta del intervalo (tanto por ciento de muestreos repetidos en las mismas condiciones cuyo intervalo calculado cubre al parámetro diferencia de medias poblacionales o bien, puesto que el intervalo pasa por el cero, las medias poblacionales pueden ser iguales).

Ítem 18. Considere el gráfico siguiente del rendimiento medio de cebada en 1980,1984 y 1988 junto con un intervalo de 95% de confianza respectivos

```
Intervalos del 95% para la media

Año N Media StDev       ----+---------+---------+---------+--
1980 6 184.00 2.61       (----*---)
1984 5 212.40 14.36                          (-----*----)
1988 5 182.40 1.82     (-----*----)
                        ----+---------+---------+---------+--
Desv. Típica conjunta =.19 180       195       210       225
```

¿Cuál de las siguientes afirmaciones es verdadera?:

a. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay buena evidencia que la medias de las muestras difieran.
b. La estimación de la media de la población en 1980 es menos precisa que en 1988.

c. Puesto que los intervalos de confianza para 1980 y 1984 no se solapan, hay poca evidencia que las medias de las poblaciones respectivas difieran.
d. **Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay poca evidencia que las medias de las poblaciones difieran.**

Este ítem es tomado de Behar (2001). El contenido primario evaluado es el *18 Interpretar gráficos de intervalos de confianza*. La opción correcta es la d) que da la interpretación correcta al concluir que al producirse solapes de los intervalos, esto implica que las medias de las poblaciones no difieren. Las opciones restantes evalúan los siguientes errores:

Tabla 5.2. Contenido cubierto por los ítems del cuestionario

Tipo	Principales áreas de contenido	Principal	Secundaria
Def.	Definición de intervalo de confianza.	1	6 a 18
Propiedades	El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra.	2	
	El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta.	3	1
	El ancho de los intervalos de confianza aumenta cuando aumenta la varianza.	4	
	Significado del nivel de confianza (variación del intervalo en diferentes muestras).	5	1
Campos de problemas	Estimar la media de una población normal o en una muestra grande con σ conocida.	6	2, 4
	Estimar la media de una población aproximadamente normal cuando, σ es desconocida.	7	2, 4
	Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande.	8	2, 4
	Estimar una proporción.	9	
	Estimar una varianza.	10	
	Comparar las medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes.	11	
	Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas.	12	
	Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ desconocidas, muestras independientes grandes.	13	
	Comparar dos varianzas poblacionales.	14	
Proc.	Elegir un modelo de distribución muestral del estadístico.	15	6 a 18
	Determinar valores críticos en la distribución del estadístico.	16	
	Interpretar intervalos de confianza obtenidos de un programa de ordenador	17	
	Interpretar gráficos de intervalos de confianza.	18	

- Opción a: El error consiste en suponer todo lo contrario al producirse solapes de los intervalos
- Opción b: El error consiste en una interpretación inadecuada de la variabilidad para esos dos intervalos

- Opción c: El error está en concluir que no hay solape en esos dos intervalos. Además aunque no hubiera solape la conclusión que de ahí se desprende es también incorrecta.

Como resumen del análisis presentamos en la tabla 5.2 el contenido primario y secundario cubierto por los ítems finalmente seleccionados para formar parte del cuestionario. Podemos comprobar en esta tabla que todos los contenidos de la tabla de especificaciones están cubiertos al menos una vez, lo que asegura la validez de contenido del cuestionario.

5.4.2. VALIDEZ DISCRIMINANTE DE LOS ÍTEMS

Para estudiar la validez discriminante de los ítems se usó la puntuación total en la prueba, dividiendo a los estudiantes en dos grupos (grupos superior e inferior) de acuerdo con dicha puntuación. Se espera que los ítems discriminen a los estudiantes con puntuaciones altas o bajas en el total, es decir, los estudiantes con puntuación total alta tengan mayor facilidad para contestar cada uno de los ítems. En lo que sigue se describe la muestra, procedimiento, análisis y resultados.

Sujetos

Se dividió la muestra total descrita en la sección 5.2. en tres partes, en función del percentil (percentil 33 y percentil 66) en la puntuación total, es decir, se formaron tres grupos del mismo tamaño compuestos por los sujetos con puntuación baja, mediana y alta en la prueba. Las puntuaciones de corte de los grupos fueron 11 y 15 de una puntuación máxima de 22. Por tanto, para hacer la comparación entre grupos se han escogido los sujetos con puntuación igual o menor a 11 y sujetos con puntuación igual o superior a 15. El criterio de comparar los grupos superior e inferior en una prueba es ampliamente seguido, por ejemplo en Ebel y Friskies (1986).

A continuación presentamos la diferencia de puntuación total media en los dos grupos de estudiantes: estadísticos descriptivos, intervalos de confianza y credibilidad y contraste *t* de diferencia de medias independientes.

En primer lugar se analiza el número total de ítems completamente correctos por alumno en el cuestionario en los dos grupos. Como se observa en la tabla 5.3, el grupo de alumnos con instrucción tuvo en promedio 8 ítems correctos más y un valor similar del error típico en los dos grupos.

Tabla 5.3. Número de ítems totalmente correctos por grupo

Grupo	*N*	Media	Desviación típ.	Error típ. de la media
Superior	101	16.98	2.110	0.224
Inferior	89	8.40	2.168	0.216

La igualdad de varianzas se confirma en la prueba Levene, que no nos permite rechazar la hipótesis de igualdad de varianzas. La prueba *t* de diferencia de medias, realizada para el supuesto de varianzas iguales, así como con el intervalo de confianza de la diferencia de medias en los dos grupos permite rechazar la hipótesis de igualdad de puntuaciones medias e indica por tanto que el criterio es adecuado para estudiar la discriminación de los ítems.

Tabla 5.4. Prueba de muestras independientes

Prueba de Levene		Prueba t para la igualdad de medias						
F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típico	Intervalo de confianza (95%) diferencia	
							Inferior	Superior
.017	.896	-27.564	188	.000	-8.581	.311	-9.196	-7.972

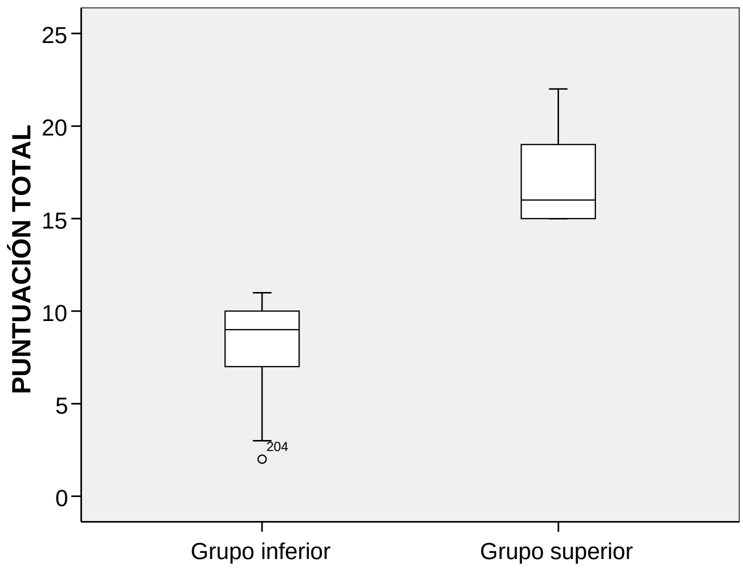
Se completa el estudio de la puntuación total con los estadísticos descriptivos por grupo (Tabla 5.5) y la representación gráfica de cajas (Figura 5.1) que muestra los valores claramente superiores de mediana y cuartiles en el segundo grupo.

Tabla 5.5. Estadísticos descriptivos de la puntuación total por grupos

	Superior (*n*=101)		Inferior (*n*=89)	
	Estadístico	Error típ.	Estadístico	Error típ.
Media	16.98	.224	8.40	.216
Límite inferior intervalo confianza 95%	7.97		7.97	
Límite superior Intervalo confianza 95%	8.82		8.82	
Límite inferior intervalo credibilidad 95%				
Límite superior Intervalo credibilidad 95%				
Media recortada al 5%	16.85		8.55	
Mediana	16.00		9.00	
Varianza	4.454		4.702	
Desv. típ.	2.110		2.168	
Mínimo	15		2	
Máximo	22		11	
Rango	7		9	
Amplitud intercuartil	4		3	
Asimetría	.765	.255	-.825	.240
Curtosis	-.719	.506	.218	.476

A partir de los resultados mostrados en la Tabla 5.5 y en la Figura 5.1 podemos inferir que la diferencia se produce no sólo en el ámbito de promedios, sino que una distribución supera a la otra en todos sus percentiles.

Figura 5.1. Gráfico de la caja para la puntuación total



Discriminación entre grupos

El estudio de la discriminación de los ítems se llevó a cabo mediante un análisis discriminante, tomando como variable dependiente el criterio (grupo de pertenencia) y como conjunto de variables independientes las puntuaciones tipificadas a los ítems, para dar igual peso a todos ellos. Este programa proporciona simultáneamente las comparaciones de las puntuaciones medias en cada ítem en el total de la prueba. En la Tabla 5.6 se presentan las medias y desviaciones típicas de los ítems (índices de dificultad) en los dos grupos y las pruebas *F* de diferencia de medias. Observamos que se obtiene una diferencia estadísticamente significativa en todos los ítems, excepto el primero (cuyo contenido primario evaluado es "*Definición de intervalo de confianza*"). La diferencia es siempre a favor del grupo superior, lo que indica una adecuada capacidad de discriminación de los ítems y del cuestionario en su conjunto.

Los ítems que mejor discriminan los dos grupos fueron los siguientes: i9, i10 e i17 cuyos contenidos son respectivamente: estimar una proporción, estimar una varianza e interpretar intervalos de confianza obtenidos de un programa de ordenador; todos ellos de resolución de problemas con respuesta abierta. Excepto el ítem 1, el resto de los ítems discriminan bien.

Tabla 5.6. Estadísticos de grupo

	Media	Desv. típ.	Media	Desv. típ.	Lambda de Wilks	F	Sig.
	Grupo inferior (n=89)		Grupo superior (n=101)				
i1	.50	.502	.57	.497	.995	.877	.350
i2	.42	.495	.67	.471	.933	13.462	.000
i3	.56	.498	.80	.404	.938	12.363	.001
i4	.81	.902	1.33	.750	.913	17.938	.000
i5	.27	.445	.52	.503	.934	13.184	.000
i6	.58	.495	.99	.106	.767	57.048	.000
I7	.66	.765	1.57	.638	.707	78.025	.000
I8	.57	.497	.89	.318	.878	26.021	.000
I9	.10	.387	1.30	.858	.538	161.630	.000
I10	.42	.711	1.63	.664	.562	146.642	.000
I11	.62	.487	.90	.303	.899	21.174	.000
I12	.91	.763	1.52	.624	.842	35.337	.000
I13	.16	.367	.46	.501	.892	22.840	.000
i14	.34	.475	.65	.479	.901	20.648	.000
i15	.35	.478	.72	.452	.862	30.223	.000
i16	.33	.471	.63	.486	.909	18 .933	.000
i17	.27	.527	.98	.866	.797	47.859	.000
i18	.52	.502	.85	.355	.876	26.559	.000

5.4.3. VALIDEZ DE CONSTRUCTO

Teniendo en cuenta que un constructo es una conceptualización teórica sobre un aspecto medible del comportamiento, la validez de constructo trata de evaluar hasta qué punto una prueba mide los constructos sobre los que se sustenta. Se trata de comprobar si el instrumento mide el rasgo o concepto teórico o si se cumplen las hipótesis sobre la estructura del constructo (Martínez Arias, 1995).

Entre las técnicas posibles para analizar la validez de constructo la más usada es el análisis factorial o el análisis de componentes principales. Al igual que Batanero y Díaz (2007), en nuestro caso, se usó el análisis factorial exploratorio del conjunto de respuestas a los ítems del cuestionario para examinar la estructura factorial de las puntuaciones al cuestionario y detectar las fuentes de variación en las medidas observadas. Se trataba de confirmar, por un lado, la existencia de un constructo subyacente que agrupase a la mayor parte de los ítems y que explicase el razonamiento sobre el intervalo de confianza. Al mismo tiempo se esperaba un constructo multidimensional en que algunos de los factores explicasen los conflictos denunciados en la investigación previa y evaluados en el cuestionario

La extracción de factores se llevó a cabo mediante el método de componentes principales; con objeto de obtener factores estadísticamente independientes y de máxima variabilidad al tiempo que no se deforma la estructura de los datos. Este método parte de una estimación inicial más alta de las comunalidades (Martínez Arias, 1995). Como método de rotación se usó la Varimax, método ortogonal, que conserva las comunalidades y la suma de porcentajes de varianza explicados por los factores (Afifi y Clark, 1990). Está orientado a maximizar la varianza de los factores. Antes de aplicar el método, se comprobaron los supuestos de aplicación (más de 10 casos por variable, unidad experimental, factorizabilidad de la matriz de correlaciones, normalidad, linearidad y ausencia de multicolinealidad).

En la tabla 5.7 se presentan las comunalidades obtenidas que oscilan entre .459 (ítem 2, efecto tamaño muestra) y .764 (ítem 1, definición) lo que indica la variabilidad de cada ítem que es explicada por el conjunto de factores retenidos (Afifi y Clark, 1990). Esto indica que cada ítem tiene una parte específica fuerte, en especial alguno de ellos. Los ítems más específicos son: el ítem 2 que evalúa la comprensión del efecto del tamaño de la muestra sobre el ancho del intervalo, debido a que su proporción común a todos los ítems es la más baja con un valor de .459 y el que le sigue con un valor de.512 en la proporción común es el ítem 9, en donde se pide al alumno construir un intervalo de confianza para estimar una proporción poblacional. De los tres parámetros: media poblacional, varianza poblacional y proporción poblacional, este último es el que menos aparece en el conjunto de ítems.

Tabla 5.7. Comunalidades

	Inicial	Extracción
i1.Definición	1.000	.764
i2.Efecto tamaño muestra	1.000	.459
i3. Efecto nivel confianza	1.000	.646
i4. Efecto varianza	1.000	.668
i5. Variación en diferentes muestras	1.000	.688
i6. Estimar media σ conocida	1.000	.589
i7. Estimar media σ desconocida	1.000	.600
i8 Estimar media σ desconocida, n grande	1.000	.722
i9. Estimar proporción	1.000	.519
i10. Estimar varianza	1.000	.594
i11. Comparar medias, varianzas conocidas	1.000	.659
i12. Comparar medias, v. desc. n pequeña	1.000	.650
i13 Comparar medias, v. desc. n grande	1.000	.667
i14. Comparar varianzas	1.000	.525
i15. Elegir distribución muestral	1.000	.751
i16. Determinar valor crítico	1.000	.676
i17. Interpretar salidas ordenador	1.000	.549
i18. Intepretar gráficos intervalos	1.000	.563

Método de extracción: Análisis de Componentes principales.

En tanto que los ítems menos específicos son: el ítem 1 con un valor de .764 en la proporción común, evalúa la comprensión de la definición del intervalo de confianza y el ítem 15, con un valor de .751 en la proporción común con el resto de los ítems, evalúa el conocimiento de la distribución muestral necesaria para calcular un intervalo de confianza para la varianza en muestras pequeñas. La comprensión de la definición del intervalo de confianza y el uso adecuado de las distribuciones muestrales son determinantes en el proceso de construcción de intervalos de confianza de ahí su baja especificidad.

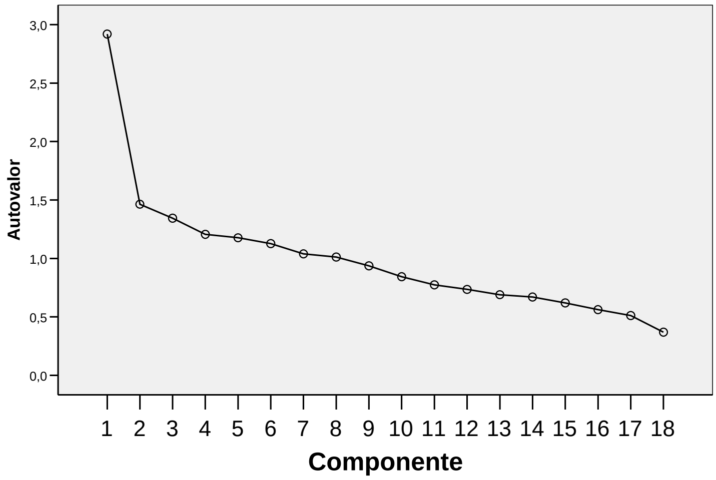
Se usaron múltiples métodos para determinar cuántos factores extraer. La extracción inicial obtuvo 8 factores con autovalor mayor que 1, que explicaron el 62.7 % de la varianza total (Tabla 5.8). También se sigue la sugerencia de Peña (2002) de que el número máximo de factores a extraer ha de ser menor a la mitad del número inicial de variables menos 1.

Tabla 5.8 Varianza total explicada por cada uno de los factores extraídos

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2.919	16.219	16.219	2.919	16.219	16.219	2.225	12.364	12.364
2	1.464	8.132	24.350	1.464	8.132	24.350	1.418	7.878	20.242
3	1.344	7.466	31.816	1.344	7.466	31.816	1.373	7.625	27.868
4	1.207	6.704	38.520	1.207	6.704	38.520	1.357	7.540	35.408
5	1.177	6.539	45.059	1.177	6.539	45.059	1.279	7.106	42.514
6	1.126	6.258	51.317	1.126	6.258	51.317	1.278	7.102	49,617
7	1.039	5.772	57.089	1.039	5.772	57.089	1.219	6.770	56.387
8	1.012	5.620	62.708	1.012	5.620	62.708	1.138	6.322	62.708
9	.937	5.204	67.912						
10	.844	4.688	72.601						
11	.774	4.301	76.902						
12	.735	4.084	80.986						
13	.690	3.833	84.820						
14	.670	3.723	88.542						
15	.620	3.443	91.986						
16	.562	3.121	95.107						
17	.512	2,842	97.949						
18	,369	2.051	100.000						

El primer factor explica un 16 % de la varianza, mientras que los siguientes entre el ocho y el cinco por ciento, cada uno, lo que indica la importancia relativa del primer factor. Este porcentaje proporciona una evidencia de validez del constructo, en cuanto que no hay acuerdo sobre el porcentaje de varianza mínima que debe explicar el primer factor pero sí que este porcentaje debe ser claramente superior al explicado por los restantes (y en nuestro caso es el doble). Además el resto de los factores explican cada uno aproximadamente la misma varianza.

Figura 5.2. Gráfico de sedimentación



En el gráfico de sedimentación (scree plot) (Figura 5.2) se observa de nuevo que la mayor varianza es debida al primer factor (que interpretaríamos como razonamiento sobre intervalo de confianza).

Tabla 5.9. Matriz no rotada de componentes

	Componente							
	1	2	3	4	5	6	7	8
i1	-.053	.346	-.069	.517	.175	-.040	.033	.580
i2	.234	-.103	.200	.144	-.568	.077	-.064	.024
i3	.140	.463	.301	-.326	-.408	.170	.129	.054
i4	.176	-.216	.413	.299	.369	-.230	.337	-.165
i5	.152	.391	.485	.173	.190	-.214	-.385	.129
i6	.546	-.399	.131	.134	.182	.222	-.090	-.078
i7	.598	-.247	-.183	-.086	-.168	-.290	.105	.130
i8	.397	-.380	-.034	.270	.000	.335	-.452	-.170
i9	.624	.116	-.093	.095	-.145	-.238	.055	.133
i10	.668	.148	-.309	.021	-.090	-.127	-.021	.072
i11	.366	.094	-.035	-.130	.322	.566	.105	.250
i12	.410	-.274	.148	-.286	.230	-.074	.419	.262
i13	.277	.422	.325	-.188	.222	.118	.129	-.438
i14	.347	.035	.127	.185	-.296	.423	.286	.068
i15	.313	.461	-.334	.432	.069	.110	.117	-.335
i16	.335	.164	-.222	-.468	.291	.083	-.390	.157
i17	.516	.127	-.263	-.099	.021	-.320	.009	-.289
i18	.404	-.036	.522	-.041	-.108	-.178	-.277	.071

Método de extracción: Análisis de componentes principales a 8 componentes extraídos

Por otro lado el punto de inflexión se produce en el sexto factor, lo que sería también un criterio para decidir el número de factores a retener, pues a partir de él los autovalores de los factores siete y ocho apenas cambian de magnitud (Martínez Arias, 1995). Por otro lado los factores cuyo autovalor es menor que uno suelen descartarse.

Incluimos en la tabla 5.9 la matriz no rotada de componentes, donde las variables se presentan ordenadas según la importancia relativa de su contribución al primer factor. Incluso antes de la rotación se observa que la mayoría de los ítems contribuyen con correlaciones positivas al primer factor y un número apreciable de ellos tiene un peso importante en el mismo, lo cual contribuye una nueva evidencia de existencia del constructo subyacente (Díaz, 2007).

Con objeto de obtener una estructura más simple se realizó una rotación Varimax, que maximiza la varianza de los coeficientes que definen los efectos de cada factor sobre las variables observadas (Peña, 2002). Tras la rotación (ver tabla 5.10) la estructura se clarifica, ya que las variables con correlaciones negativas prácticamente desaparecen o los valores de la correlación son muy pequeños (por debajo de 0.3), mientras que las correlaciones positivas fuertes en cada factor se mantienen o incluso crecen.

Tabla 5.10. Matriz de componentes rotados

	Componente							
	1	2	3	4	5	6	7	8
i1.Definición	-.005	-.125	.085	.133	-.013	.127	.024	.840
i2.Efecto tamaño muestra	.143	.165	-.208	.109	.580	-.104	-.084	-.037
i3. Efecto nivel confianza	.031	-.427	.197	.269	.506	.052	-.216	-.214
i4. Efecto varianza	.019	.095	-.031	.151	-.059	.051	.793	.000
i5. Variación en diferentes muestras	-.019	-.013	-.007	.794	-.071	.104	.043	.196
i6. Estimar media σ conocida	.211	.607	.264	.075	.110	-.037	.276	-.105
i7. Estimar media σ desconocida	.719	.124	.014	-.069	.101	-.203	.098	-.043
i8 Estimar media σ desconocida, n grande	.079	.825	.048	.065	.116	.065	-.085	-.062
i9. Estimar proporción	.661	.032	.045	.155	.177	,084	.052	.118
i10. Estimar varianza	.707	.105	.150	.045	.080	.187	-.113	.067
i11. Comparar medias, varianzas conocidas	.027	.151	.780	-.034	.101	.086	-.012	.093
i12. Comparar medias, v. desc. n pequeña	.347	-.081	.437	-.059	.023	-.356	.442	-.081
i13 Comparar medias, v. desc. n grande	-.009	-.166	,297	.344	.043	.457	.199	-.426
i14. Comparar varianzas	.083	.106	.245	-.087	.637	.140	.100	.064
i15. Elegir distribución muestral	.242	.064	.025	-.077	.057	.811	.006	.143
i16. Determinar valor crítico	.282	.086	.486	.212	-.347	-.056	-.410	-.128
i17. Interpretar salidas ordenador	.623	.014	-.034	.044	-.147	.278	.017	-.241
i18. Intepretar gráficos intervalos	.210	.163	-.018	.615	.203	-.214	.112	-.122

Método de extracción: Análisis de componentes principales.
Método de rotación: Normalización Varimax con Kaiser. La rotación ha convergido en 9 iteraciones.

Para facilitar la interpretación, se repiten los resultados en la Tabla 5.11, suprimiendo las correlaciones menores a 0.3 con lo que se ve más fácilmente la estructura. Para elegir la matriz final se siguieron los criterios (Martínez Arias, 1995):

- Tomar sólo factores que sean interpretables para el investigador (Afifi y Clark, 1990).
- Cada fila de la matriz rotada tenga al menos un cero; es decir para cada variable debe haber al menos un factor que no contribuya a su varianza.
- Para cada factor, habrá un conjunto de variables cuyas saturaciones se aproximen a cero (Afifi y Clark, 1990).
- Si hay cuatro o más factores comunes, hay un conjunto de variables que tienen saturaciones próximas a cero en las dos columnas (Martínez Arias, 1995).
- Cada factor debe tener peso importante de al menos dos variables, pues de otro modo sería un factor específico (Afifi y Clark, 1990). Los factores se interpretan a continuación.

Tabla 5.11. Matriz de componentes rotados simplificada

	Componente							
	1	2	3	4	5	6	7	8
i1.Definición								.840
i2.Efecto tamaño muestra					.580			
i3. Efecto nivel confianza		-.427			.506			
i4. Efecto varianza							.793	
i5. Variación en diferentes muestras				.794				
i6. Estimar media σ conocida		.607						
i7. Estimar media σ desconocida	.719							
i8 Estimar media σ desconocida, n grande		.825						
i9. Estimar proporción	.661							
i10. Estimar varianza	.707							
i11. Comparar medias, varianzas conocidas			.780					
i12. Comparar medias, v. desc. n pequeña	.347		.437			-.356	.442	
i13 Comparar medias, v. desc. n grande				.344		.457		-.426
i14. Comparar varianzas					.637			
i15. Elegir distribución muestral						.811		
i16. Determinar valor crítico			.486		-.347		-.410	
i17. Interpretar salidas ordenador	.623							
i18. Intepretar gráficos intervalos				.615				

Los factores separan claramente diferentes elementos de significado del intervalo de confianza, puesto que los primeros factores parecen ligados a la construcción de intervalos (campos de problemas y procedimientos) mientras que los últimos se ligan a las definiciones y propiedades.

Primer Factor: Agrupa con altas puntuaciones los ítems abiertos que piden construir intervalos para la media con desviación típica desconocida (r=0.719), proporción (r=0.661), varianza desconocida (r=0.707), comparar medias con varianzas desconocidas, *n* pequeñas (r=.347); todos ellos requieren recordar y discriminar diferentes distribuciones muestrales: *t* de Student, *F*, y normal. Asimismo incide en este factor la interpretación de salidas con ordenador (r=0.623), posiblemente porque el problema propuesto se refiere también a la comparación de dos medias.

Factor segundo: Construcción de intervalos que involucran la distribución normal. Agrupa los problemas de estimación de la media con varianza conocida y varianza desconocida pero muestras grandes; los dos requieren de la distribución normal. Aunque se trate de construcción de intervalos, son problemas de opción múltiple, por lo que sólo podemos tener en cuenta la respuesta totalmente correcta o incorrecta, mientras que en los ítems agrupados en el primer factor tenemos en cuenta las parcialmente correctas. También el estudio de la relación del coeficiente con el ancho del intervalo, que aparece con signo contrario, lo que nos indica que algunos estudiantes que construyen bien los intervalos pedidos no comprenden la definición de coeficiente de confianza.

Factor tercero: Construcción de intervalos que involucran la distribución t de Student. Agrupa los problemas de comparación de medias con varianza desconocida, *n* pequeña (r=0.437), y determinación de valor crítico para un intervalo con desviación estándar desconocida (r=0.486); los dos problemas requieren de la distribución *t* de Student. También se incluye el estudio de la comparación de medias, varianzas conocidas (r=0.780).

Factor cuarto: Significado del nivel de confianza. Agrupa los problemas Significado del nivel de confianza (r=0.794) e interpretación de gráficos de intervalos (r=0.615); en ambos ítems la comprensión del significado del nivel de confianza es primordial. Además se incluye en este factor la comparación de medias con varianza desconocida, *n* grande (r=0.344) en donde se incluye la pregunta interpretar el resultado. El ítem es de opción múltiple y el alumno para contestar correctamente, el significado de nivel de confianza, debe de comprenderlo.

Factor quinto: Habilidad analítica e interpretativa. Agrupa los problemas donde se exhibe una relación entre efecto tamaño muestra (r=0.580), efecto nivel de confianza(r=0.506) y comparar varianzas (r=.637). Podemos reconocer a este factor como la habilidad para analizar e interpretar el efecto que se produce en el ancho del intervalo cuando cambia el tamaño de la muestra, disminuye o aumenta el nivel de confianza, o bien analizar el intervalo obtenido en el cociente de varianzas pasando o no por el uno, para finalmente hacer la interpretación correcta. Aparece en este factor un problema de determinar valor crítico con un valor negativo (r=-.347) lo que indica que algunos estudiantes que pudieran realizar un buen análisis e interpretación no determinan correctamente el valor crítico en la construcción del intervalo.

Factor sexto: Agrupa los problemas de comparar medias con varianzas desconocidas, *n* pequeñas (r=-.356), comparar medias con varianzas desconocidas, *n* grande (r=.457); y elegir distribución muestral (r=.811). Igual que en el factor primero, todos ellos requieren recordar y discriminar diferentes distribuciones muestrales: *T* de Student, normal y Chi-cuadrado. El valor negativo (r=-.356) en el problema de comparar medias con varianzas desconocidas, *n* pequeñas, nos indica que la distribución que más conflictos causa en algunos estudiantes es la *t* de Student. Este factor aunque de naturaleza similar al primer factor, solo explica el 6.25% de la varianza, en tanto que el factor primero explica el 16.22% de la varianza.

El resto de los factores sugieren que los diferentes sesgos que afectan el razonamiento de los intervalos de confianza aparecen solo parcialmente relacionados con la ejecución matemática; algunos de los sesgos en estos factores están ligados a las definiciones y propiedades.

En consecuencia se confirma la hipótesis previa de existencia de un constructo subyacente (definido por el peso comparativamente alto del primer factor antes de la rotación) que se subdivide en diferentes componentes, todos ellos relacionados con el razonamiento sobre intervalos de confianza.

5.5. ESTUDIOS DE FIABILIDAD

5.5.1. FIABILIDAD DE CONSISTENCIA INTERNA

Un ítem contribuye a la fiabilidad del test cuando mide la misma clase de puntaje verdadero que los otros ítems del test; esto significa que contribuye a la fiabilidad si el componente verdadero del ítem está determinado por el mismo factor que determina la magnitud de los componentes verdaderos, medidos por los otros ítems (Magnusson, 1990).

El *coeficiente de fiabilidad* es un indicador de la fiabilidad teórica de las puntuaciones observadas, en el sentido de proporcionar un valor numérico que indica el *grado de confianza* que podíamos tener en dichas puntuaciones como estimadores de las puntuaciones verdaderas de los sujetos. Entre los diversos procedimientos para el cálculo del estimador del coeficiente de fiabilidad se tomó el coeficiente Alfa de Cronbach (López Féal, 1986; Meliá, 2001) por varias propiedades:

- Es la cota inferior de la que se obtendría por el método de la prueba repetida y estima la *fiabilidad en el acto* (Martínez Arias, 1995; Meliá, 2001).
- Refleja el grado en el que covarían los ítems que constituyen el test y es un valor conservador.
- Es el valor medio de todos los que se obtendrían con el método de las dos mitades si se utilizasen todas las combinaciones de ítems.

Se calculó este coeficiente a partir de la muestra de estudiantes (*n =252*) mediante el programa correspondiente del paquete estadístico SPSS, analizando los estadísticos de cada ítem si se suprime del instrumento y estudiando el efecto sobre el coeficiente de ir suprimiendo sucesivamente los ítems que presentan peores resultados.

Tabla 5.12. Estadísticos de fiabilidad de la prueba (*n*=252)

Alfa de Cronbach	Alfa de Cronbach basada en los elementos tipificados	N de elementos
.654	.644	18

El valor obtenido de Alfa=0.654 para el coeficiente de Cronbach (Tabla 5.12), corresponde a una correlación entre la puntuación observada y la puntuación verdadera de 0.809 e indica una consistencia bastante alta (Meliá, 2001). En todo caso, Santisteban (1990) indica, como limite general aceptable en los cuestionarios de evaluación 0.50. Es importante recordar que en la valoración de este coeficiente interviene la complejidad con que se ha definido el constructo y en nuestro caso es altamente complejo.

La correlación corregida ítem- test (Tabla 5.13) ha oscilado desde valores .099 a .488. Es un nuevo indicador de la validez discriminante de los ítems que en general es buena. Como vemos en esta tabla, el ítem 1 presenta una correlación negativa de .018

con respecto al valor total de la prueba y se refiere a sesgos en la definición sobre intervalo de confianza. Quiere decirse que este ítem no discrimina bien y su eliminación elevaría un poco la fiabilidad, dando uno de los dos valores más altos del total de la tabla (.668).

Tabla 5.13. Estadísticos total-elemento de la prueba (*n*=252)

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
i1	12.04	17.401	-.018	.668
i2	12.04	16.735	.144	.652
i3	11.94	16.953	.099	.656
i4	11.48	16.099	.109	.668
i5	12.23	16.831	.128	.654
i6	11.81	16.165	.378	.632
i7	11.46	14.552	.391	.619
i8	11.87	16.579	.220	.645
i9	11.90	13.835	.451	.607
i10	11.60	13.620	.488	.600
i11	11.85	16.519	.244	.643
i12	11.40	15.460	.269	.639
i13	12.30	16.634	.196	.647
i14	12.07	16.390	.230	.644
i15	12.07	16.398	.228	.644
i16	12.13	16.510	.200	.647
i17	12.04	14.942	.358	.625
i18	11.90	16.249	.298	.638

La correlación media entre ítems (Tabla 5.14) es aceptable, por lo que se decidió conservarlos a todos debido a nuestro interés en la medición de todas las unidades de contenido sobre comprensión de intervalos de confianza en los estudiantes de ingeniería.

Tabla 5.14. Estadísticos de resumen de los elementos

	Media	Mínimo	Máximo	Rango	Máximo/ mínimo	Varianza
Medias de los elementos	.700	.294	1.194	.901	4.068	.069
Varianzas de los elementos	.373	.167	.779	.612	4.670	.051
Covarianzas inter-elementos	.035	-.032	.337	.369	-10.524	.003
Correlaciones inter-elementos	.091	-.089	.434	.523	-4.874	.009

La prueba T-cuadrado de Hotelling genera un contraste multivariado sobre la hipótesis nula de que todos los elementos de la escala tienen la misma media. Como era de esperar el resultado es estadísticamente significativo (Tabla 5.15), pues las medias son diferentes, ya que se obtuvo una gama variada de índices de dificultad.

Tabla 5.15. Prueba T cuadrado de Hotelling

T-cuadrado de Hotelling	F	gl1	gl2	Sig.
669.472	36.870	17	235	.000

El coeficiente de correlación intraclase (Tabla 5.16) proporcionó un alto valor promedio, indicador de la fiabilidad de una sola medida y fue significativo tanto para los valores promedios como para las medidas individuales, aunque para éstas el valor fue menor.

Tabla 5.16. Coeficiente de correlación intraclase

	Correlación intraclase	Intervalo de confianza 95%		Prueba F con valor verdadero 0			
		Límite inferior	Límite superior	Valor	gl1	gl2	Sig.
Medidas individuales	.095(b)	.074	.122	2.894	251.0	4267	.000
Medidas promedio	.654(c)	.589	.714	2.894	251.0	4267	.000

Coeficiente basado en el análisis factorial

También se calculó el coeficiente Theta de Carmines basado en el análisis factorial que hemos llevado a cabo anteriormente.

$$\theta = \frac{n}{n-1}\left(1-\frac{1}{\lambda_1}\right)=0.696$$

En la expresión anterior donde *n* es el número de ítems y λ el primer autovalor en el análisis factorial. El valor obtenido próximo a 0.7 es suficientemente alto para llevar a cabo un estudio de evaluación (Santisteban, 1990). Como vemos sube respecto al coeficiente Alfa.

5.5.2. GENERALIZABILIDAD

Se utilizó la teoría de la generalizabilidad que usa el análisis de varianza para estudiar diferentes fuentes de error en un proceso de medida y extiende el concepto de fiabilidad (Feldt y Brennan, 1991). Siguiendo a López Feal (1986, p. 512) se consideró que "*una faceta puede ser un conjunto de ítems, un conjunto de métodos particulares de recogida de datos, un conjunto de sistemas escolares, un conjunto de evaluadores, etc.*" Cuando el objeto de estudio son las personas, la fiabilidad del estudio se maximiza cuando la varianza entre sujetos aumenta y es grande respecto a la de los ítems, que son homogéneos entre sí. Pero, si el objeto de estudio es medir el grado en que los diferentes objetivos educativos se alcanzan, la varianza de los ítems será grande respecto a la de los sujetos y se podría obtener una generalizabilidad grande –referida a los ítems como objeto de estudio (López- Feal, 1986).

Este es precisamente el caso de esta investigación puesto que para alcanzar una validez alta, se incluyó un conjunto de ítems muy variados en cuanto a los componentes de comprensión del concepto que evalúan. Esto explica la variabilidad de los ítems. Por ello se decidió aplicar la teoría de la generalizabilidad, considerando las facetas de ítems y sujetos, siguiendo la misma aproximación de otras investigaciones sobre comprensión de conceptos matemáticos avanzados (ej. Navarro- Pelayo, 1994, Díaz, 2007).

En el estudio de generalizabilidad la puntuación media X_{pi} de una persona *p* bajo la condición *i* se descompone en la forma siguiente (Martínez Arias, 1995):

(1) $$X_{pi} = \mu + (\mu_p - \mu) + (\mu_i - \mu) + (X_{pi} - \mu_p + \mu_i - \mu)$$

siendo μ la media de todos los sujetos en la prueba; $\alpha_p = (\mu_p - \mu)$ el efecto debido a la persona, $\beta_i = (\mu_i - \mu)$ el efecto de la condición *i* (en nuestro caso del ítem *i*) y $(\alpha\beta)_e = (X_{pi} - \mu_p + \mu_i - \mu)$ la interacción entre persona- condición. Se estimaron las diferentes varianzas a partir de la tabla de análisis de varianza (Tabla 5.17) y el modelo de estimación de Dunn y Clark (1987) para el análisis de varianza de medidas repetidas:

Tabla 5.17. ANOVA con prueba de Friedman

		Suma de cuadrados	gl	Media cuadrática	Chi-cuadrado de Friedman	Sig.
Inter-personas		245.040	251	0.976		
Intra-personas	Inter-elementos	294.619	17	17.331	727.882	.000
	Residual	1439.381	4267	0.337		
	Total	1734.000	4284	0.405		
Total		1979.040	4535	0.436		

Media global = .70
a Coeficiente de concordancia W de Kendall = .149.

- Varianza debida a los sujetos σ^2_s o varianza de diferenciación de todas las personas del universo de personas entre sí.
- Varianza debida a los ítems σ^2_i refleja los errores asociados con los diferentes niveles de dificultad de los ítems.
- Varianza residual σ^2_r o del error aleatorio de medida, que en este tipo de diseño está

confundido con la interacción ítem- persona.

De esta tabla se obtuvo los cuadrados medios entre sujetos, entre los diferentes ítems y residual, así como sus grados de libertad.

CM_r = es un estimador de $s^2_{r.}$

CM_p = es un estimador de $bs^2_p + s^2_r$ siendo *b* el número de ítems.

CM_i = es un estimador de $as^2_i + s^2_r$ siendo *a* el número de sujetos.

Tabla 5.18. Estimaciones de los componentes de la varianza

Fuente de variación	Componentes varianza	% Varianza total
Sujetos	0.03758824	8.50
Ítems	0.06770518	15.31
Residual	0.33700000	76.19
Total	0.44229341	100

Despejando en las anteriores ecuaciones se obtuvo las estimaciones de componentes de varianza presentadas en la Tabla 5.18. La variabilidad total explicada por los ítems fue el doble que la explicada por los sujetos, lo que confirma el hecho de que el coeficiente de fiabilidad *a* de Cronbach aunque suficiente para nuestro análisis, no se haya aproximado más a la unidad, debido a la heterogeneidad entre los diversos ítems.

El coeficiente de generalizabilidad se define con el cociente (2),

(2)
$$G = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2}$$

es decir. como cociente entre la varianza de las puntuaciones verdaderas de la prueba y la suma de la varianza verdadera más la varianza debida al error aleatorio. La varianza de error depende de cuál es el universo de puntuaciones verdaderas considerado (Thorndike, 1989). En nuestro trabajo se han considerado dos casos diferentes.

Generalizabilidad a otros ítems

Se fijó, en primer lugar, como faceta las personas, tomando como universo de generalización el conjunto de todos los ítems posibles asociados a la variable medida. El universo de generalización sería el universo de tests de la misma longitud con ítems muy similares a los dados y elegidos aleatoriamente. La fórmula (2) del coeficiente de generalizabilidad se transforma en este caso en la expresión (3).

(3)
$$G_i = \frac{\sigma_s^2}{\sigma_s^2 + \frac{\sigma_r^2}{22}} = 0.659$$

Se obtuvo un valor próximo al del coeficiente Alfa, lo cual es de esperar, puesto que el coeficiente de generalizabilidad a otros ítems coincide con la fiabilidad de consistencia interna, salvo errores de redondeo.

Generalización a un universo de alumnos

En segundo lugar se consideró la población de alumnos como universo de generalización y la faceta ítems como fija. Es decir se trata de generalizar los resultados a otros alumnos cuando se conserva exactamente el mismo cuestionario. Esta elección no es usual, pero la teoría de la generalizabilidad nos lo permite: "*Desde una perspectiva general, cualquier grupo de condiciones (personas, ítems, codificadores, ocasiones, clases, etc.) puede ser el objeto de medida*" (Brennann, 1983, p. 31).

La fórmula (2) del coeficiente de generalizabilidad se transforma en este caso en la expresión (4):

(4)
$$G_i = \frac{\sigma_i^2}{\sigma_i^2 + \frac{\sigma_r^2}{37}} = 0.987$$

Para este coeficiente se obtuvo un valor muy cercano a la unidad. Ello quiere decir que, cuando se considera los alumnos como universo posible de generalización, fijando los ítems de la prueba, tenemos muy altas probabilidades de generalización de los resultados. Para ello los alumnos han de tener las mismas características sociológicas y educativas que aquellos a los que se pasó la prueba.

5.6. CONCLUSIONES

En este capítulo en que hemos realizado el análisis de validez de contenido, validez discriminante de los ítems y validez de constructo; así como fiabilidad de nuestro cuestionario, destacamos las siguientes conclusiones:

- El análisis realizado al número total de ítems correctamente resueltos por cada uno de los grupos (grupo superior e inferior) exhibió valores con significancia estadística para la diferencia de medias, medianas y cuantiles, mostrando un desplazamiento de toda la distribución (gráfico de cajas), pero mostrando valores similares en los estadísticos de dispersión.
- El análisis discriminante mostró que todos los ítems resultaron discriminativos, excepto el ítem1 relacionado con sesgos en la definición del intervalo de confianza. Esto nos permite especular que dicho sesgo constituye uno o varios factores diferenciados en el estudio de la validez factorial.
- El análisis factorial nos ha permitido identificar 8 factores que explicaron en su conjunto 62.71% de la varianza y con autovalores mayores a 1. Además con relación a los factores resaltamos lo siguiente:
  a) El primer factor tuvo un peso mayor que los restantes, en forma considerable, aunque explica solamente el 16% de la varianza, lo que nos sugiere el carácter no unidimensional de nuestro instrumento de medición.
  b) Hemos interpretado los primeros seis factores como componentes específicos del razonamiento matemático en la comprensión de los intervalos de confianza; los factores primero, segundo, tercero y sexto comparten la habilidad, por parte del alumno, de discriminar diferentes distribuciones muestrales: *t* de Student, Normal, *F* y Chi-cuadrado. Las dificultades para distinguir los modelos que se usan para describir los datos empíricos han sido reportadas en los trabajos de Vallecillos (1994, 1996, 1999) y Batanero (2000).
  c) Los factores cuarto y quinto tienen que ver con la interpretación del efecto que se produce en el ancho del intervalo cuando disminuye o aumenta el nivel de confianza o cuando cambia el tamaño de la muestra. Las dificultades conceptuales e interpretativas de los estudiantes con relación a los intervalos de confianza pueden estar relacionadas con el hecho de que muchos estudiantes visualizan los intervalos de confianza como estadísticos descriptivos, ignorando su naturaleza inferencial (Cumming y Fidler, 2005).
  d) El resto de los factores corresponden, parcialmente, cada uno a los sesgos descritos en razonamiento sobre intervalos de confianza, que aparecen independientes del razonamiento lógico matemático.

- Del análisis factorial en que se muestra que los seis primeros factores dan cuenta de la dimensión lógico-matemática del constructo “razonamiento en intervalos de confianza” y el resto de los factores los sesgos de razonamiento que tienen base psicológica, según se planteó hipotéticamente en la definición semántica de la variable, podemos concluir que estos resultados aportan evidencias de validez de constructo.
- El análisis factorial muestra la existencia de componentes diferenciados, todos ellos relacionados con el razonamiento de los intervalos de confianza, lo cual está en comunión con nuestro marco teórico.
- Los coeficientes de fiabilidad y generalizabilidad que intentan dar una medida objetiva de la estabilidad de las puntuaciones obtenidas frente a variaciones aleatorias, nos permiten concluir, según los índices obtenidos, que nuestro cuestionario lo podríamos calificar como moderadamente fiable, en cuanto a generalizabilidad a otros ítems; sin embargo es altamente generalizable a otros alumnos.

CAPITULO 6
ESTUDIO DE EVALUACIÓN

6.1. INTRODUCCIÓN

Una vez finalizadas las pruebas de validez y fiabilidad del cuestionario, se completó el estudio de los datos de la muestra descrita en el capítulo 5, con objeto de obtener información sobre los errores y dificultades específicos de los estudiantes que la componen.

En este capítulo analizamos estos datos desde diferentes puntos de vista. La primera sección se dedica al análisis detallado de respuestas en ítems de opciones múltiples y su comparación con los resultados obtenidos en investigaciones previas.

Seguidamente realizamos un análisis semiótico de los ítems con respuesta abierta para mostrar, por un lado, la complejidad de las respuestas correctas y los diferentes objetos matemáticos que el estudiante ha de relacionar entre sí por medio de funciones semióticas. Asimismo mostramos los principales conflictos semióticos que dan lugar a respuestas parcialmente correctas o incorrectas. El análisis se completa con el estudio de las frecuencias de respuestas en ítems abiertos.

El estudio culmina con el análisis de la dificultad y discriminación de los ítems, puntuación global, estudio de número total de respuestas correctas y conflictos semióticos por estudiantes en ítems de respuesta abierta e interrelación entre respuestas a diferentes ítems.

6.2. ANÁLISIS DETALLADO DE ÍTEMS DE OPCIONES MÚLTIPLES

A continuación se analizan los resultados en los ítems de opciones múltiples y se comparan con los obtenidos en las pruebas piloto y en las investigaciones previas que trataron el mismo o similar ítem.

Ítem 1. El intervalo de confianza del 50% para la media de una población μ es:
a. El rango dentro del cuál caen el 50% de los valores de la media de la muestra $\bar{x}$.
b. Un intervalo más ancho que el intervalo de confianza del 95%.
c. **Un intervalo de valores calculado a partir de los datos de la muestra. En el 50% de las muestras de una población, el intervalo calculado contiene a la media de la población.**
d. Dos veces más ancho que el intervalo de confianza del 100%.

Tabla 6.1. Frecuencias y porcentajes de respuestas a las opciones del ítem 1.

Opciones	Frecuencia	Porcentaje
a	65	25.8
b	40	15.9
c	**140**	**55.6**
d	7	2.8
Total	252	100.0

En el ítem 1 tomado de Cruise, Dudley y Thayer (1984), se evalúa la comprensión de la definición del intervalo de confianza y su variación cuando se calculan diferentes intervalos tomando muestras de la misma población. La respuesta mayoritaria ha sido la correcta con un 55.6% de frecuencia, mientras que en la prueba piloto del ítem la respuesta mayoritaria fue la correcta con un 76% de frecuencia. Un 25.8% de los estudiantes piensa que el intervalo se construye para estimar una media muestral, es decir presenta una confusión entre estadístico y parámetro, que también aparece en las investigaciones de Vallecillos (1994, 1995). En investigaciones hechas por delMas, Garfield, Ooms y Chance (2007, pg. 40) encontraron un 55.8% de estudiantes que también mostraron ese error en la interpretación. Los estudiantes que contestan b) o d), que son el 18.7%, no tienen claro el efecto del nivel de confianza sobre la anchura del intervalo, suponiendo que la anchura aumenta al disminuir el coeficiente.

Ítem 2. Comparado a los intervalos de confianza calculados en muestras de tamaño n=4 en una población normal, el ancho de los intervalos de confianza de la media de la población calculado en muestras de tamaño n = 50:
a. Variará más que los anchos de los intervalos para muestras de tamaño n =4.
b. **Variará, pero no tanto como lo hicieron los anchos de los intervalos para muestras de tamaño *n*=4.**
c. Tomarán valores parecidos.

Este ítem, que se adaptó de otro elaborado por Garfield, delMas y Chance (2004), estudia la variación del ancho de los intervalos de confianza con el tamaño de la muestra para el caso particular de estimación de la media de una población normal, tanto en caso de desviación típica conocida o desconocida. Los resultados se recogen en la tabla 6.2. La opción más frecuente ha sido la b (correcta) y el porcentaje de respuestas correctas ha sido de 55.2% por lo que los alumnos han comprendido bien esta propiedad. Los resultados son muy similares a los de la muestra piloto, donde la respuesta mayoritaria fue la correcta con un 57% de frecuencia. Al comparar con las investigaciones de Fidler y Cumming (2005) observamos que en aquel estudio el 20% de los estudiantes dijeron que la variación del ancho de los intervalos de confianza se incrementaría si se aumentara el tamaño de la muestra y en nuestro estudio esta respuesta resultó el 33%; un 29% dijo el ancho que no se vería afectado y en nuestro estudio resultó un 11.1%. En el estudio de Fidler y Cumming (2005) solamente un 16% pudo contestar correctamente. Es preocupante, no obstante que más de un 40% de estudiantes en nuestro estudio no comprende esta propiedad, relacionada con la menor variabilidad de la distribución muestral en muestras grandes. Pensamos que este error puede estar relacionado con la heurística de la representatividad (Kahneman, Slovic y Tversky, 1982) por la que las personas no son conscientes del efecto del tamaño muestral sobre la variabilidad de la distribución del estadístico.

Tabla 6.2. Frecuencias y porcentajes de respuestas a las opciones del ítem 2

Opciones	Frecuencia	Porcentaje
a	84	33.3
b	**139**	**55.2**
c	28	11.1
Blanco	1	.4
Total	252	100.0

Ítem 3. Si, manteniendo todos los demás datos fijos, el nivel de confianza se reduce (por ejemplo de 90% a 80%):

a. El intervalo de confianza no cambia.
b. El intervalo de confianza será más ancho.
c. **El intervalo de confianza será más angosto.**
d. El cambio en el intervalo de confianza no es predecible.

Este ítem, que se adaptó de Cruise, Dudley y Thayer (1984), evalúa la variación *del ancho de los intervalos de confianza cuando el nivel de confianza aumenta,* es decir, se trata de analizar si los estudiantes relacionan el significado del nivel de confianza con el cálculo del valor crítico y cómo este afecta al intervalo obtenido. Los resultados se recogen en la tabla 6.3. La opción más frecuente ha sido la c (correcta) y el porcentaje de respuestas correctas ha sido de 65%, por lo que los alumnos parecen comprender bien la propiedad. En la muestra piloto la respuesta mayoritaria fue la correcta, con un 70% de frecuencia, por lo que los porcentajes son muy similares en las dos muestras. Al comparar con las investigaciones de Behar (2001) en su formato de verdadero y falso, observamos que el porcentaje de respuestas correctas en el grupo de expertos en estadística fue de 63.8% y en el grupo de estudiantes universitarios que habían tomado cursos de estadística (no expertos) fue de 52.2%. Los resultados de nuestra muestra son muy similares a los resultados del grupo de expertos. Los principales errores encontrados en nuestro estudio son: El 26% de los estudiantes piensan que una reducción en el nivel de confianza produciría un intervalo de confianza más ancho, error comentado anteriormente; el 6 % de los estudiantes respondió que el cambio en el intervalo de confianza no era predecible y el 2.8% creía que el ancho del intervalo de confianza no cambiaba con la reducción en el nivel de confianza.

Tabla 6.3. Frecuencias y porcentajes de respuestas a las opciones del ítem 3

Opciones	Frecuencia	Porcentaje
a	7	2.8
b	66	26.2
c	**164**	**65.1**
d	15	6.0
Total	252	100.0

Ítem 5. En un intervalo de confianza del 95% para la media:

a. Si se toman muchas muestras y con cada una se construye el intervalo, la media muestral $\bar{x}$ caerá dentro del intervalo de confianza el 95% de las veces.

b. La probabilidad de que $\bar{x}$ caiga dentro de un intervalo de confianza calculado de una muestra específica es 0.95.

c. **Si se toman muchas muestras de igual tamaño, el 95% de los intervalos de confianza calculado contendrían a μ.**

Este ítem que se adaptó de Cruise, Dudley y Thayer (1984), evalúa *el significado del nivel de confianza* (variación del intervalo en diferentes muestras). Los resultados se exhiben en la tabla 6.4. La opción más frecuente ha sido la correcta c y el porcentaje de respuestas correctas fue de 36.5%, por lo que los alumnos parecen no comprender bien la propiedad, porque se reparten casi por igual las respuestas en las tres opciones. Por el contrario, en la muestra piloto la respuesta mayoritaria fue la correcta con un 76% de frecuencia. Al comparar con las investigaciones de Behar (2001, pg. 220) en un ítem similar, observamos que el porcentaje de respuestas correctas en el grupo de expertos en estadística fue de 48.9% y en el grupo de estudiantes universitarios que habían tomado cursos de estadística (no expertos) fue de 35.7%. En este ítem existe coincidencia de resultados para los grupos de estudiantes en ambos estudios. Un 29% de los estudiantes piensan que el intervalo se refiere no a la media poblacional sino a la media muestral; error que fue observado para el caso del contraste de hipótesis en la investigación de Vallecillos (1994). Un 34% de estudiantes hacen una interpretación bayesiana del intervalo de confianza, reproduciendo otro error muy frecuente en la interpretación del contraste de hipótesis que ha sido denunciado, entre otros por Batanero (2000b) y Batanero y Díaz (2005).

Tabla 6.4. Frecuencias y porcentajes de respuestas a las opciones del ítem 5

Opciones	Frecuencia	Porcentaje
a	73	29.0
b	86	34.1
c	**92**	**36.5**
Blanco	1	.4
Total	252	100.0

Ítem 6. La media muestral de 100 observaciones en una prueba de matemáticas es 75. Encuentre el intervalo de confianza al 95% para la media de la población, asumiendo que σ=7:

a. (61.28, 88.72)
b. **(73.63, 76.37)**
c. (68, 82)
d. (74.3, 75.7)

Este ítem evalúa *estimar la media de una población normal o en una muestra grande con σ conocida*. Los resultados se recogen en la tabla 6.5. La opción más frecuente ha sido la b, que da el resultado de aplicar correctamente el procedimiento de cálculo y el porcentaje de respuestas correctas es 79% por lo que los alumnos parecen comprender bien dicho procedimiento, en tanto que en la muestra piloto la respuesta mayoritaria fue la correcta con un 68%. No tenemos comparación con investigaciones previas porque este ítem se ha tomado de Cruise, Dudley y Thayer (1984), del cual no se tienen investigaciones. En relación a la prueba piloto, los nuevos resultados exhiben una mejor comprensión de la definición, propiedades y los procedimientos requeridos en la solución del problema, incluidos en ellos la elección del modelo de distribución muestral del estadístico.

En la construcción del intervalo de confianza para la media poblacional, de una población normal o en una muestra grande con σ conocida, que se obtiene sumando y restando, a la media muestral, el error estándar (la desviación típica poblacional dividida por la raíz cuadrada del tamaño de muestra) y multiplicado por el valor crítico de la distribución normal estándar. Encontramos que el 8.3% de los estudiantes en la obtención del intervalo de confianza olvida el valor crítico de la distribución normal estándar que multiplica al error estándar (opción d). El 6.7% de los estudiantes en la obtención del intervalo de confianza olvida la raíz cuadrada de la muestra que divide a la desviación típica poblacional y el 4.4% de los estudiantes comete ambos errores en la obtención del intervalo de confianza.

Tabla 6.5. Frecuencias y porcentajes de respuestas a las opciones del ítem 6

Opciones	Frecuencia	Porcentaje
a	17	6.7
b	**199**	**79.0**
c	11	4.4
d	21	8.3
Blanco	4	1.6
Total	252	100.0

Ítem 8. Se han obtenido los siguientes datos de emisión diaria de óxidos de azufre, para una muestra de tamaño *n*=100, media: $\overline{x}$ = 18 y varianza muestral s^2=36. Elabore un intervalo de confianza de 95% para la verdadera emisión diaria promedio de óxidos de azufre:

a. (17.016, 18.984)
b. **(16.824, 19.176)**
c. (6.24, 29.76)
d. (8.16, 27.84)

Este ítem evalúa la capacidad para *estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande.* En la tabla 6.6 se recogen los resultados. La opción más frecuente ha sido la b (la correcta) y el porcentaje de respuestas correctas es 73% por lo que los alumnos parecen comprender bien el procedimiento. Al comparar con la muestra piloto la respuesta mayoritaria fue la correcta con una frecuencia del 68%. No tenemos comparación con investigaciones previas porque este ítem es un problema tomado del libro de texto de Miller, Freund y Johnson (1997, pg. 223) del cual no se tienen estudios, ni investigaciones de un ítem similar. En relación a la muestra piloto, los resultados de la nueva muestra son mejores.

El intervalo de confianza para la media poblacional a partir de datos experimentales σ desconocida, muestra grande, se obtiene sumando y restando, a la media muestral, el error estándar (en este caso la desviación típica muestral dividida por la raíz cuadrada del tamaño de muestra) multiplicado por el valor crítico de la distribución normal estándar. El 11.5% de los estudiantes obtiene incorrectamente el intervalo de confianza porque obtienen incorrectamente el valor crítico en la normal estándar usando 1.64 en vez de 1.96. Este error lo cometen porque en la tabla de la normal estándar buscan para un área a la derecha de la curva de 5%, cuando lo correcto es 2.5% para el nivel de confianza de 95%. Al ser intervalos bilaterales el 5% restante al 95% debe ser dividido por dos y esto no lo toman en cuenta. Por otro lado, el 6% de los estudiantes no dividen entre el tamaño de la muestra en la obtención del intervalo, es decir, confunden la desviación típica de la población con la desviación típica del estadístico (error de muestreo) y el 5.2% de los estudiantes incurre en ambos errores. Un 4.4% de estudiantes dejaron en blanco el ítem.

Tabla 6.6. Frecuencias y porcentajes de respuestas a las opciones del ítem 8

Opciones	Frecuencia	Porcentaje
a	29	11.5
b	**184**	**73.0**
c	15	6.0
d	13	5.2
Blanco	11	4.4
Total	252	100.0

Ítem 11. La siguiente tabla contiene un resumen de información sobre la resistencia a la compresión de cubos (N/ mm^2) para especimenes de concreto:

	Tamaño muestral	Media muestral
Tipo1	68	26.99
Tipo2	74	37.56

Suponga que las desviaciones estándar poblacionales de ambos grupos son σ_1=4.89 y σ_2=6,43 respectivamente. Calcule un intervalo de confianza de 95% para hallar la diferencia entre el verdadero promedio de resistencia en el Tipo1 y el verdadero promedio de resistencia en el Tipo 2:

a. (-13.02, -8.12)
b. **(-12.437, -8.70)**
c. (-31.32, 10.18)
d. (-18.64, -2.5)

Este ítem evalúa la capacidad de *estimar la diferencia de las medias en dos poblaciones, conociendo* σ_1^2 y σ_2^2, *muestras independientes.* Los resultados se recogen en la tabla 6.7. La opción más frecuente ha sido la b (la correcta) y el porcentaje de respuestas correctas es 74.6% por lo que los alumnos parecen comprender bien el procedimiento, mientras que en la muestra piloto la respuesta mayoritaria fue la correcta con un 57%. No tenemos comparación con investigaciones previas porque este ítem es un problema tomado del libro de texto de Devore (2005, pg. 371).

El intervalo de confianza para la diferencia de medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes, se obtiene usando la siguiente expresión: $\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}$. El 7.5% de los estudiantes obtiene incorrectamente el intervalo de confianza porque no dividió entre los tamaños de muestra del Tipo 1 (n_1) y Tipo 2 (n_2), el 7.1% de los estudiantes usó incorrectamente el valor crítico de 2.58 para $z_{\alpha/2}$, en vez de 1.96, obtenido en la tabla de la distribución normal estándar. Usar el valor crítico de 2.58 corresponde a un intervalo con un nivel de confianza del 99%. Los alumnos que pudimos observar sus procedimientos y tuvieron este error, suponemos que estaban pensando en 5% y por distracción leen un área a la derecha de la distribución normal de .005 y hubo otros alumnos que escogieron esta opción, pero no mostraron procedimiento alguno. Un 3.6% de los estudiantes cometieron ambos errores, no dividen por los tamaños de muestra n_1 y n_2 y usan el valor crítico de 2.58. Un 7.2% de los estudiantes no contestaron este ítem. En la muestra piloto (n =44) no hubo respuestas en blanco en este ítem.

Tabla 6.7. Frecuencias y porcentajes de respuestas a las opciones del ítem 11

Opciones	Frecuencia	Porcentaje
a	18	7.1
b	**188**	**74.6**
c	19	7.5
d	9	3.6
Blanco	18	7.2
Total	252	100.0

Ítem 13. La tabla siguiente resume algunos datos de un experimento realizado para estudiar varias características de tornillos de anclaje:

		Resistencia al corte	
Diámetro de tornillo	Tamaño muestra	Media muestral	Desviación estándar
3/8	100	4.25	1.3
½	100	7.25	1.7

Calcule e interprete un intervalo de confianza de 95% para la diferencia del verdadero promedio de resistencias al corte.

a. **(-3.41, -2.58): en el 95% de la muestras del mismo tamaño en esta población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.**
b. (-3.41, -2.58): la probabilidad de que la verdadera diferencia promedio de la resistencia al corte cae en el intervalo (-3.41, -2.58) es 0.95.
c. (-3.35, -2.65): en el 95% de la muestras del mismo tamaño en esta población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.
d. (-3.35, -2.65): la probabilidad de que la verdadera diferencia promedio de la resistencia al corte cae en el intervalo (-3.35, -2.65) es 0.95.

Este ítem que se adaptó de Devore (1998, 8b, pg. 342) evalúa la capacidad para *obtener la diferencia de las medias en dos poblaciones,* $\sigma_1^2 = \sigma_2^2$ *desconocidas, muestras independientes grandes y luego interpretarla.* Los resultados se recogen en la tabla 6.8. La opción más frecuente ha sido la b (incorrecta) con un porcentaje de respuestas incorrectas de 32.1% por lo que los alumnos no parecen comprender bien el procedimiento o bien dan una interpretación incorrecta. Es el hecho que la interpretación bayesiana pareciera ser más intuitiva y pudiera estar más difundida en las explicaciones que dan los profesores que la interpretación clásica. En la muestra piloto la respuesta mayoritaria fue la correcta con un 59%**.** En ítem estudiado por Behar (2001, pg. 223), en donde los estudiantes interpretan un intervalo de confianza para diferencia de medias, encontró que un 34.3% de los estudiantes lo responden incorrectamente, resultado muy parecido al de nuestro estudio. En el muestreo piloto la opción b tuvo un porcentaje de respuestas incorrectas de 30%, resultado muy parecido al obtenido en el nuevo muestreo.

El 32.1% de los estudiantes obtiene correctamente el intervalo de confianza pero hacen la interpretación bayesiana (opción b), este porcentaje es mayor que el de la opción correcta de 29.4% (opción a). El 19.0% interpreta correctamente el intervalo pero se equivocan en la obtención del valor crítico $z_{\alpha/2}$ (usan 1.64, que correspondería a un nivel de confianza de 90%, en vez de 1.96) y el 16.7% interpreta incorrectamente el intervalo y además se equivocan en la obtención del valor crítico $z_{\alpha/2}$ (usan también 1.64 en vez de 1.96). Un 2.8% de los estudiantes dejaron en blanco la respuesta.

Tabla 6.8. Frecuencias y porcentajes de respuestas a las opciones del ítem 13

Opciones	Frecuencia	Porcentaje
a	**74**	**29.4**
b	81	32.1
c	48	19.0
d	42	16.7
Blanco	7	2.8
Total	252	100.0

Ítem 14. Una compañía quiere seleccionar el proceso de pulido que presente la variabilidad menor. Una muestra aleatoria de n_1 =16 piezas del primer proceso da como resultado una desviación estándar muestral s_1=5 micro pulgadas, y una muestra aleatoria de n_2 =11 piezas del segundo proceso da como resultado una desviación estándar muestral s_2=4 micro pulgadas. Establezca un intervalo de confianza de 90% para, suponiendo que los dos procesos son independientes y que la aspereza superficial tiene una distribución normal. ¿Cuál de los dos procesos recomendaría usted?

a. Como todos los valores están dentro del intervalo (.54, 3.96) recomendaría el proceso 1.
b. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría el proceso 2 (se cambian los grados de libertad del numerador y denominador).
c. **Como el cociente de valores está dentro del intervalo (.54, 3.96) recomendaría cualquiera de los dos**.
d. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría cualquiera de los dos (se cambian los grados de libertad del numerador y denominador).

Este ítem evalúa si los estudiantes son capaces de *comparar dos varianzas poblacionales*. Este ítem se ha tomado de Montgomery y Runger (2004, pg.416) y las opciones, que son de elaboración propia, además de evidenciar como interpreta el intervalo de confianza el estudiante, proporcionan un aporte para valorar el razonamiento estadístico del estudiante (delMas, 2002).

Los resultados se recogen en la tabla 6.9. La opción más frecuente ha sido la c y el porcentaje de respuestas correctas 52.8% por lo que los alumnos parecen comprender bien el procedimiento, mientras que en la muestra piloto la respuesta mayoritaria fue la correcta con un 80%. No encontramos resultados de investigaciones de ítems similares.

En el intervalo de confianza para comparar dos varianzas, calculado usando la siguiente expresión:

$$\frac{S_1^2}{S_2^2}\left(\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(\nu_2,\nu_1)$$

Un 18.7% de los estudiantes obtiene incorrectamente el intervalo porque intercambia indebidamente los grados de libertad de numerador con los del denominador en sus lecturas de las tablas, para obtener los valores críticos $f_{\alpha/2}$ en la distribución *F.* Un 13.9% de los estudiantes obtiene correctamente el intervalo de confianza, pero se equivocan en la interpretación, porque al pasar el intervalo de valores decimales menores que 1 a valores mayores que 1, significa que cualesquiera de los dos procesos es menos variable que el otro, es decir se pueden recomendar cualquiera de los dos procesos. Un 12.3% de los estudiantes hacen una interpretación correcta pero obtienen incorrectamente el intervalo de confianza porque intercambia indebidamente los grados de libertad de numerador con los del denominador en sus lecturas de las tablas, para obtener los valores críticos en la distribución *F*. Hubo un 2.4% de respuestas en blanco.

Tabla 6.9. Frecuencias y porcentajes de respuestas a las opciones del ítem 14

Opciones	Frecuencia	Porcentaje
a	35	13.9
b	47	18.7
c	**133**	**52.8**
d	31	12.3
Blanco	6	2.4
Total	252	100.0

Ítem 15. La distribución muestral utilizada en la construcción de intervalos de confianza para la varianza en muestras pequeñas es:
a. Distribución *t* de Student.
b. Distribución Ji-cuadrada.
c. Distribución Normal.
d. Distribución *F*.

Este ítem evalúa si los estudiantes saben *elegir un modelo de distribución muestral del estadístico.* Los resultados se recogen en la tabla 6.10. La opción más frecuente ha sido la b y el porcentaje de respuestas correctas es 52.8% por lo que los alumnos parecen comprender bien la propiedad de discriminar las diferentes distribuciones muestrales, mientras en la muestra piloto la respuesta mayoritaria fue la correcta con un resultado casi idéntico de 56%. Este es un ítem de elaboración propia, y una comparación muy cercana es con investigaciones de delMas, Garfield, Ooms y Chance (2007, pg. 40), en ítem que evalúa la habilidad para seleccionar una distribución de muestreo apropiada para una población y un tamaño de muestra, observamos que el porcentaje de respuestas correctas es de 44.2%, siendo un porcentaje menor que el de nuestro estudio.

Un 31% de los estudiantes escogieron incorrectamente la Distribución *t* de Student, una posible explicación es que los alumnos tienen fuertemente asociados Distribución *t* con muestras pequeñas. Un 9.9% de los estudiantes seleccionaron incorrectamente la Distribución Normal y un 6% de los estudiantes seleccionaron de forma incorrecta la Distribución *F*. Estos dos últimos porcentajes suponemos que en su mayoría corresponden a respuestas de estudiantes que contestaron al azar y unos cuantos de los que escogieron la Distribución *F*, es porque supusieron alguna relación entre varianza y esa distribución.

Tabla 6.10. Frecuencias y porcentajes de respuestas a las opciones del ítem 15

Opciones	Frecuencia	Porcentaje
a	78	31.0
b	**133**	**52.8**
c	25	9.9
d	15	6.0
Blanco	1	.4
Total	252	100.0

Ítem 16. El nivel de confianza es 0.95, para un intervalo de confianza para la media poblacional con desviación estándar poblacional desconocida para un grupo de puntajes distribuido normalmente de tamaño $n = 20$. Los valores críticos han de ser:
a. -1.65 y 1.65
b. -1.96 y 1.96
c. -2.093 y 2.093
d. -2.085 y 2.085

Este ítem evalúa la capacidad de *determinar valores críticos en la distribución del estadístico*. Los resultados se recogen en la tabla 6.11. La opción más frecuente ha sido la c y el porcentaje de respuestas correctas 46.8% por lo que los alumnos parecen manejar bien el cálculo de valores críticos. En la muestra piloto la respuesta mayoritaria fue la correcta con un 59% y no tenemos comparación con investigaciones previas, porque este ítem es una adaptación a uno elaborado por Cruise, Dudley y Thayer (1984) quien no da datos de su dificultad. En comparación con la prueba piloto, el porcentaje de respuestas correctas disminuyó en el nuevo muestreo.

Un 41.3% de los estudiantes, porcentaje muy cercano al de la respuesta correcta, determinó el valor crítico en forma incorrecta al hacerlo a partir de la distribución normal estándar, es decir, aunque manejan el cálculo de valores críticos confunden los casos en que se debe usar la distribución *t* o la normal para la estimación de la media. El 5.2% de los estudiantes respondió incorrectamente al usar grados de libertad incorrectos en la distribución *t* y el 4.8% de los estudiantes se equivoca al calcular el valor critico correspondiente a $\alpha/2$ en las tablas de la normal estándar. Un 2% de los alumnos dejaron en blanco la respuesta.

Tabla 6.11. Frecuencias y porcentajes de respuestas a las opciones del ítem 16

Opciones	Frecuencia	Porcentaje
a	12	4.8
b	104	41.3
c	**118**	**46.8**
d	13	5.2
Blanco	5	2.0
Total	252	100.0

El siguiente ítem tomado de Behar (2001) evalúa la *interpretación de gráficos de intervalos de confianza,* un tema que algunos autores consideran difícil (Schenker y Gentleman, 2001; Cumming, William y Fidler, 2004; Belia, Fidler y Cumming, 2005). Los resultados se recogen en la tabla 6.12. La opción más frecuente ha sido la d y el porcentaje de respuestas correctas 69.4% por lo que los alumnos interpretan correctamente en su mayoría estos gráficos, aunque el porcentaje baja con relación a la muestra piloto, donde la respuesta mayoritaria fue la correcta con un 78%.

Ítem 18. Considere el gráfico siguiente del rendimiento medio de cebada en 1980,1984 y 1988 junto con un intervalo de 95% de confianza respectivos

```
Intervalos del 95% para la media

Año N Media StDev       ----+---------+---------+---------+--
1980 6 184.00 2.61        (----*---)
1984 5 212.40 14.36                           (-----*----)
1988 5 182.40 1.82     (-----*----)
                        ----+---------+---------+---------+--
Desv. Típica conjunta =.19 180      195       210       225
```

¿Cuál de las siguientes afirmaciones es verdadera?:

a. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay buena evidencia que las medias de las muestras difieran.
b. La estimación de la media de la población en 1980 es menos precisa que en 1988.
c. Puesto que los intervalos de confianza para 1980 y 1984 no se solapan, hay poca evidencia que las medias de las poblaciones respectivas difieran.
d. **Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay poca evidencia que las medias de las poblaciones difieran.**

Al comparar con las investigaciones de Behar (2001) en su formato de verdadero y falso, observamos que el porcentaje de respuestas correctas en el grupo de expertos en estadística fue de 61.7% y en el grupo de estudiantes universitarios que habían tomado cursos de estadística (no expertos) fue de 55.2%. Los resultados del grupo de expertos son cercanos a los de nuestro estudio.

Un 12.7% de los estudiantes responden incorrectamente, al interpretar que hay poca evidencia que las medias de las poblaciones respectivas difieran cuando no se solapan los intervalos de confianza. El 9.1% de los estudiantes interpretan incorrectamente a pesar de que el lenguaje utilizado es más claro que en la opción c y el 6.7% de los estudiantes interpretan incorrectamente al confundir la expresión “menos precisa” con menos variable. Un 2% de los estudiantes dejaron en blanco la respuesta**.**

Tabla 6.12. Frecuencias y porcentajes de respuestas a las opciones del ítem 18

Opciones	Frecuencia	Porcentaje
a	23	9.1
b	17	6.7
c	32	12.7
d	**175**	**69.4**
Blanco	5	2.0
Total	252	100.0

Como resumen del análisis de los ítems de opciones múltiples, podemos observar que los ítems fueron, en general bien respondidos por los estudiantes, siendo los más fáciles los ítems 6, 11 y 8 que evalúan los campos de problemas: a) *Estimar la media de una población normal o en una muestra grande con σ conocida*, b) *Comparar las medias en dos* poblaciones, conociendo σ_1^2 y, *muestras independientes* y c) *Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande,* respectivamente.

Los más difíciles fueron los ítems 13, 5 y 16 que evalúan: a) el campo de problemas*: Comparar las medias en dos poblaciones,* $\sigma_1^2 = \sigma_2^2$ *desconocidas, muestras independientes grandes,* b) la propiedad*: Significado del nivel de confianza (variación del intervalo en diferentes muestras)* y c) los procedimientos*: Determinar valores críticos en la distribución del estadístico*.

Respecto a los principales errores detectados, destacamos los siguientes:

- Un 25.8% de estudiantes confunde estadístico y parámetro, pensando que el intervalo se construye para estimar la media muestral (Ítem 1). Vallecillos y Batanero (1997) en su estudio con relación a la comprensión de los contrastes de hipótesis estadísticas, reportan también esta confusión. Avisan que no se trata de un problema de uso incorrecto de la notación adecuada para cada concepto, sino al hecho de no tomar en consideración las distintas medias y distribuciones implicadas, en concreto la distribución muestral del estadístico. Este error también ha sido reportado por Behar (2001) y delMas, Garfield, Ooms y Chance (2007).
- Un 33.3% de los estudiantes no comprenden como varía el ancho del intervalo al disminuir el tamaño de la muestra (Ítem 2).
- Un 26.2% de los estudiantes no comprenden como varía el ancho del intervalo al reducirse el nivel de confianza (Ítem 3). Estos resultados de los ítems 2 y 3 confirman los obtenidos por Fidler y Cumming (2005) y Behar (2001).
- Solo un 36.5% comprende que el intervalo de confianza representa el porcentaje de intervalos de muestras, tomadas todas bajo las mismas condiciones, dentro de los cuales estará contenido el verdadero valor del parámetro (Ítem 5). Resultado que corrobora los obtenidos por Behar (2001).
- El 32.1% de los estudiantes obtiene correctamente el intervalo de confianza para diferencia de medias poblacionales, pero cometen error al hacer la interpretación

bayesiana (Ítem 13). Aunque Behar (2001) también estudia la interpretación de intervalos de confianza para diferencia de medias, el ítem que él evalúa, plantea como proposición correcta, una diseñada bajo el enfoque de interpretación bayesiana… "*Nosotros no sabemos el verdadero aumento medio en la producción, pero estamos 95% seguros que el aumento medio en la producción ha quedado atrapado por este intervalo*" Behar (2001, pg. 223). De ahí que el resultado de nuestro estudio sea un nuevo aporte a la investigación empírica de las dificultades en la comprensión de los intervalos de confianza.

- Un 31% de los estudiantes escogieron incorrectamente la Distribución *t* de Student en la construcción de intervalos de confianza para la varianza en muestras pequeñas (Ítem 15), siendo la distribución Chi cuadrada la apropiada para esas condiciones.
- Un porcentaje alto (41.3%) de los estudiantes, determinó un valor crítico en forma incorrecta al hacerlo a partir de la distribución normal estándar, cuando la distribución requerida era la Distribución *t* de Student (Ítem 16). Las dificultades encontradas a través de las respuestas a los ítems 15 y 16 han sido estudiadas en un contexto más general por Schuyten (1991), quién indica que la utilización simultánea de conceptos con diferentes niveles de concreción supone una dificultad para los estudiantes. Por ejemplo distribución muestral para la media, media de la muestra y la de la población tienen diferentes niveles de abstracción que confunden y complican el trabajo de los estudiantes. Nuestros resultados confirman los señalamientos de Schuyten (1991).
- El 12.7% de los estudiantes interpretan incorrectamente los intervalos de confianza a partir de un gráfico, confundidos posiblemente por el lenguaje utilizado en el distractor c (Ítem 18). Behar (2001), en su formato de falso o verdadero, no incluye en su conjunto de ítems alguno parecido al distractor c (si están incluidos en el cuestionario de Behar cada uno de los otros distractores del ítem 18). La redacción de este distractor c incluye varias expresiones de carácter negativo que nos hacen pensar que debiéramos modificar este ítem en futuras ediciones del cuestionario.
- El 18.7% de los estudiantes obtienen incorrectamente el intervalo de confianza, para la comparación de dos varianzas poblacionales, porque intercambia indebidamente los grados de libertad de numerador con los del denominador, en sus lecturas de las tablas, para obtener los valores críticos en la distribución *F* (ítem 14). Una posible explicación de este error es que no llegan a interiorizar la diferencia entre los valores

críticos $f_{\alpha/2}(\nu_1,\nu_2)$ y $f_{\alpha/2}(\nu_2,\nu_1)$. No hemos encontrado este error en otros trabajos, por lo cual este resultado también es una nueva aportación a la investigación del tema.

- Otros errores que podemos resaltar y que tienen que ver con la obtención incorrecta del intervalo de confianza son: a) los estudiantes no dividen entre los tamaños de muestra en la construcción del intervalo para la diferencia de medias poblacionales, σ_1^2, σ_2^2 conocidas, muestras independientes grandes (7.5%, Ítem 11) y b) olvidan en la fórmula del intervalo de confianza, para la media poblacional, σ conocida, el valor crítico de la distribución normal estándar (8.3%, Ítem 6). Estos dos tipos de errores, que no hemos encontrado reportados en otras investigaciones, los pudiéramos explicar desde una simple falta de memorización de las fórmulas correspondientes, hasta una sugerente falta de conectividad entre símbolos y entidades conceptuales.

6.3. ANÁLISIS SEMIÓTICO DE ÍTEMS ABIERTOS

6.3.1. INTRODUCCIÓN

Para completar el estudio de ítems de opción múltiple, en lo que sigue realizamos el estudio semiótico de las respuestas de los estudiantes de la muestra en los ítems abiertos, con el fin de iniciar la identificación de conflictos semióticos en el aprendizaje de los intervalos de confianza. Seguimos el método usado en investigaciones previas, por ejemplo, Cobo (2003), Godino, Batanero y Roa (2005).

La importancia del estudio de conflictos semióticos la resalta Godino (2002, 2003), quien indica que en las prácticas matemáticas intervienen objetos ostensivos (símbolos, gráficos, etc.) y no ostensivos (que evocamos al hacer matemáticas) y que son representados en forma textual, oral, gráfica o incluso mediante gestos. El autor señala que en el trabajo matemático los símbolos (significantes) remiten a entidades conceptuales (significados) y un punto crucial de la enseñanza es lograr que los alumnos dominen la semántica (además de la sintaxis) de estos símbolos. Indica que la investigación en didáctica de las matemáticas ha mostrado la importancia que tienen las representaciones en la enseñanza y el aprendizaje, pero una cuestión todavía no suficientemente analizada es la variedad de objetos que desempeñan el papel de representación y de los objetos representados (Godino, Contreras y Font, 2006).

Precisamente el interés de Godino en sus trabajos es analizar esta cuestión. El autor asume la noción de función semiótica como una "correspondencia entre conjuntos", que pone en juego tres componentes:

- Un plano de expresión (objeto inicial, considerado frecuentemente como el signo);
- Un plano de contenido (objeto final, considerado como el significado del signo, esto es, lo representado, lo que se quiere decir, a lo que se refiere un interlocutor);
- Un criterio o regla de correspondencia, esto es un código interpretativo que relaciona los planos de expresión y contenido.

El modelo descrito por Godino (2002, 2003) destaca la diversidad de objetos puestos en juego en la actividad matemática, para referirse a los elementos de significado que hemos descrito en el marco teórico, cada uno de los cuales puede aparecer como parte de la función semiótica tanto en el plano de la expresión como en el del contenido. También indica que en ocasiones el significado que el profesor o investigador quiere atribuir a una expresión no es interpretado correctamente por el alumno y se produce el *conflicto semiótico.* En estos casos, el error se produce no por una falta de conocimientos, sino por no relacionar adecuadamente los dos términos de una función semiótica, como haremos ver en los ejemplos que analizamos en esta sección.

La finalidad de la investigación didáctica es encontrar dispositivos “idóneos” para la enseñanza y el aprendizaje de objetos matemáticos. Por ello un objetivo importante es describir y valorar la pertinencia de la enseñanza, incluidos los instrumentos de evaluación y determinar criterios para mejorarlos (Wilhelmi, Godino y Lacasta, 2007). La idoneidad semiótica de un proceso instruccional debe atender a evitar conflictos semióticos y la idoneidad didáctica debe procurar que estos conflictos afloren y el profesor sea capaz de resolverlos (Godino, Contreras y Font, 2006). Un primer paso para lograrlo será identificar los conflictos que cabe esperar ante determinadas tareas en una proporción importante de los estudiantes.

Para cada uno de los ítems abiertos del cuestionario, analizamos, a continuación, los tipos encontrados de respuesta correcta, parcialmente correcta e incorrecta. Para cada una de las categorías de respuestas incorrectas encontradas, reproducimos un ejemplo, para el cual se realiza un análisis semiótico, dividiéndolo en unidades de análisis y destacando para cada unidad las principales funciones semióticas establecidas por el alumno así como sus conflictos semióticos. Presentamos finalmente la tabla de frecuencias de respuestas diferentes en el ítem.

6.3.2. ANÁLISIS DEL ÍTEM 4. EFECTO DE LA VARIANZA SOBRE LA AMPLITUD DEL INTERVALO

Ítem 4. Explica cómo varía la anchura del intervalo de confianza de la media si, conservando el mismo tamaño de muestra y el mismo coeficiente de confianza tomamos una población con varianza cuatro veces mayor.

En este ítem deseábamos evaluar la capacidad del alumno para reconocer el efecto de la varianza sobre la anchura del intervalo. La respuesta correcta esperada debe suponer que la anchura del intervalo de confianza tiene un aumento en un factor de 2. Un ejemplo se reproduce a continuación y se analiza en la Tabla 6.13. Observamos la complejidad de la respuesta en donde se han de conjugar conceptos, procedimientos, representaciones, propiedades y argumentos; además se manifiestan las facetas ejemplar- tipo en relación con varios objetos matemáticos y en general la faceta ostensivo-no ostensivo en todos ellos. Además de esta solución correcta, a la que daremos código C10 hemos encontrado una variante de respuesta correcta, un ejemplo de la cual se analiza en la tabla 6.14.

> *"La fórmula del intervalo de confianza de la media es, siendo $Z_{\alpha/2}$ el valor crítico correspondiente al coeficiente de confianza 1-α, σ la desviación típica de la población y n el tamaño de la muestra. Como la varianza es el cuadrado de la desviación típica, si la aumentamos 4 veces, la desviación típica aumenta dos veces. La fórmula se transforma en $\bar{x} \pm Z_{\alpha/2} \times \frac{2\sigma}{\sqrt{n}}$ Por tanto el intervalo de confianza tiene una anchura doble. Si no se conoce la desviación típica de la población se usa la de la muestra y la distribución T en su lugar"* (Alumno HG).

Tabla 6.13. Análisis de la solución correcta al ítem 4

Unidad	Contenido	Funciones semióticas
U1	La fórmula del intervalo de confianza de la media es $\bar{x} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$, siendo $Z_{\alpha/2}$ el valor crítico correspondiente al coeficiente de confianza 1-α, σ la desviación típica de la población y n el tamaño de la muestra.	*Fórmula*: se refiere a la expresión $\bar{x} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$ y también al procedimiento de cálculo del intervalo (procedimiento y lenguaje). *Intervalo*: Concepto; en este caso se refiere a un ejemplar particular (intervalo de confianza) y al tipo general, es decir se manifiesta la dimensión ejemplar-tipo. *Media*: Evoca la media de la población (ejemplar) y al concepto general de media (tipo). $\bar{x}$ hace referencia a la media de la muestra, que es un estadístico.

		El alumno tiene que diferenciar entre ella y la media de la población μ que es un parámetro (conceptos) ± hace referencia a que el intervalo se construye sumando y restando una misma cantidad (operación), así como a los conceptos suma y diferencia. $Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$ indica la cantidad a sumar, que es a su vez un producto (operación). $Z_{\alpha/2}$ se refiere al valor crítico (concepto) en la distribución normal *N (0,1)* (concepto); *1-α* se refiere al nivel de confianza (concepto) α/2 a la división del complemento del nivel de confianza por dos (procedimiento), puesto que el nivel de confianza tiene que contemplar la probabilidad por encima y debajo de la media muestral (propiedad). $\frac{\sigma}{\sqrt{n}}$ indica la división (procedimiento) de la desviación típica (concepto) por la raíz (procedimiento) del tamaño de la muestra (concepto) o también al error típico (concepto y procedimiento de cálculo).
U2	Como la varianza es el cuadrado de la desviación típica, si la aumentamos 4 veces, la desviación típica aumenta en dos veces.	*Varianza, cuadrado, desviación típica*: el alumno nombra diferentes conceptos usándolos como ejemplar (en el caso particular) y general (tipo). *Aumentamos 4 veces*: se describe una operación. *La varianza es el cuadrado de la desviación típica*: (el estudiante explicita una propiedad). *Si la aumentamos 4 veces, la desviación típica aumenta en dos veces*: razonamiento deductivo, donde se pone en juego la función raíz cuadrada (concepto).
U3	La fórmula se transforma en $\bar{x} \pm Z_{\alpha/2} \times \frac{2\sigma}{\sqrt{n}}$. Por tanto el intervalo de confianza tiene una anchura doble.	*Transforma*: expresa una operación (procedimiento) $\bar{x} \pm Z_{\alpha/2} \times \frac{2\sigma}{\sqrt{n}}$ hace referencia al procedimiento de cálculo *Intervalo de confianza, anchura, doble* : el alumno evoca diferentes conceptos *El intervalo de confianza tendrá una anchura doble*: razonamiento deductivo
U4	Si no se conoce la desviación típica de la población se usa la de la muestra y la distribución T en su lugar	*La distribución t en su lugar:* El alumno discrimina la distribución muestral a emplear en caso de desviación típica conocida o desconocida (propiedad) y hace referencia al concepto de distribución (caso general) y un ejemplo particular (distribución t).

C11: El alumno indica que, al aumentar la varianza, la anchura del intervalo aumenta, aunque no indica la proporción ni argumenta su respuesta mediante una cadena de razonamientos similar a la anteriormente descrita. Explícitamente, al menos, el número de funciones semióticas establecidas por el estudiante, es mucho menor que en el caso anterior, y hemos de suponer que el estudiante relaciona algunos de los objetos anteriores, pero no es capaz de dar una justificación, aunque no se producen explícitamente conflictos semióticos. Un ejemplo se incluye y analiza a continuación.

> *"La anchura del intervalo de confianza aumenta ya que la varianza afecta directamente según la fórmula"* (Alumno TL).

Tabla 6.14. Análisis de la solución correcta al ítem 4

Unidad	Contenido	Funciones semióticas
U1	La anchura del intervalo de confianza aumenta ya que la varianza afecta directamente según la fórmula.	*Intervalo de confianza, anchura, aumenta, varianza*: conceptos que se usan como ejemplares y tipo del concepto general correspondiente. *Fórmula:* referencia al procedimiento de cálculo; se deduce que el alumno recuerda el procedimiento de construcción pues justifica la variación por medio de la fórmula. *La anchura del intervalo de confianza aumenta ya que la varianza afecta directamente según la fórmula*: razonamiento deductivo.

Hemos encontrado también la siguiente respuesta parcialmente correcta:

C21: El alumno indica que la anchura del intervalo varía en la misma proporción que la varianza (cuatro veces mayor). Comienza correctamente la construcción del intervalo. Se produce un *conflicto* relativo a una proposición (el intervalo tiene anchura proporcional a la varianza) y otro relativo al procedimiento de cálculo (usar la varianza y no la desviación típica en la construcción del intervalo). Un ejemplo (Alumno AG) se reproduce a continuación y se analiza en la Tabla 6.15.

"La anchura del intervalo de confianza se hace más grande
$\bar{x} - \frac{Z_{\alpha/2}(4\sigma^2)}{\sqrt{n}} \quad \mu \quad < \bar{x} + \frac{Z_{\alpha/2}(4\sigma^2)}{\sqrt{n}}$ ".

Tabla 6.15. Análisis de la solución parcialmente correcta al ítem 4

Unidad	Contenido	Funciones semióticas
U1	La anchura del intervalo de confianza se hace más grande.	*Intervalo de confianza, anchura, más grande*: Conceptos que se usan como ejemplares y tipo del concepto general correspondiente. La anchura del intervalo de confianza se hace más grande: razonamiento deductivo.
U2	$\bar{x} - \frac{Z_{\alpha/2}(4\sigma^2)}{\sqrt{n}} \quad \mu \quad < \bar{x} + \frac{Z_{\alpha/2}(4\sigma^2)}{\sqrt{n}}$	La expresión μ se refiere a la media poblacional. El alumno tiene que distinguir la media de la población de la media muestral $\bar{x}$, esto es, entre parámetro y estadístico (conceptos) y saber que el intervalo se refiere a la de la población (propiedad). $\frac{Z_{\alpha/2}(4\sigma^2)}{\sqrt{n}}$ hace referencia a la cantidad a sumar (o restar), que es a su vez un producto (operación). $Z_{\alpha/2}$:se refiere al valor crítico (concepto) en la distribución normal *N (0,1)* (concepto); *1-*α se refiere al nivel de confianza (concepto) α/2 a la división del complemento del nivel de confianza por dos (procedimiento), puesto que el nivel de confianza tiene que contemplar la probabilidad por encima y debajo de la media muestral (propiedad). El alumno olvida escribir el símbolo <, en el límite inferior del intervalo.

		$\bar{x} - \frac{Z_{\alpha/2}(4\sigma^2)}{\sqrt{n}}$ y $\bar{x} + \frac{Z_{\alpha/2}(4\sigma^2)}{\sqrt{n}}$ hace referencia al nuevo procedimiento de cálculo y a los límites del intervalo (conceptos, procedimientos y propiedades). El alumno tiene un *conflicto* en el manejo de la proporción cuatro veces mayor. Tiene otro *conflicto* al usar la varianza y no la desviación típica en la construcción del intervalo.

C31: El estudiante responde que se produce una disminución del ancho del intervalo sin emitir alguna justificación. Suponemos que el alumno difícilmente puede establecer relaciones entre los objetos puestos en juego. Se reproduce un ejemplo y se analiza en la Tabla 6.16.

"La anchura del intervalo de confianza se ve hace más pequeña" (Alumno FQ).

Tabla 6.16. Análisis de la solución incorrecta al ítem 4

Unidad	Contenido	Funciones semióticas
U1	La anchura del intervalo de confianza se ve hace más pequeña.	*Intervalo de confianza, anchura, más pequeña:* el estudiante evoca diferentes conceptos. *La anchura del intervalo de confianza se ve hace más pequeña*: razonamiento deductivo. *Conflicto* relativo a las propiedades relaciona incorrectamente el efecto de la varianza sobre el ancho de intervalo (propiedad).

C32: Otros errores donde el estudiante da una respuesta no relacionada con el problema.

En la Tabla 6.17 presentamos los resultados del ítem 4 (efecto de la varianza sobre el ancho del intervalo), que podemos considerarlo de dificultad moderada al resolverlo correctamente un 42.9 % .

Tabla 6.17. Frecuencias (y porcentajes) de respuestas en el ítem 4 (n=252)

	Frecuencia	Porcentaje	Porcentaje acumulado
Respuestas correctas C11	108	42.9	42.9
Respuestas parcialmente correctas C21	66	26.2	69.1
Respuestas incorrectas C31	65	25.8	94.9
Respuestas incorrectas C32	3	1.1	96.0
Blanco	10	4.0	100.0
Total	252	100.0	

Un 26.2% resolvió parcialmente correcto, mostrando que la dificultad más importante en este problema tiene que ver con un *conflicto* relativo a una proposición (el

intervalo tiene anchura proporcional a la varianza) y otro relativo al procedimiento de cálculo (usar la varianza y no la desviación típica en la construcción del intervalo).

Otro 26% tiene el conflicto también relativo a una proposición (el intervalo tiene anchura proporcional a la varianza), aunque, no podemos plantear posibles explicaciones por la falta de justificaciones a la respuesta. Lo más que podemos hacer es suponer que esos conflictos no explicitados lo limitan para establecer relaciones con los objetos involucrados, además de imposibilitarlo para argumentar. En ítem similar estudiado por Behar (2001, pg. 221) se encontró que un 40.7% de los estudiantes no comprendieron el efecto del aumento en la desviación estándar de la población sobre el ancho del intervalo y corrobora los resultados obtenidos por este autor.

6.3.3. ANÁLISIS DEL ÍTEM 7. CONSTRUCCIÓN DEL INTERVALO DE CONFIANZA PARA LA MEDIA

Ítem 7. Un fabricante asegura que sus garrafones, contienen un litro de cloro puro. Al tomar una muestra de 16 garrafones se determinó que en promedio contenían 0.94 litros de cloro puro, con desviación estándar de la muestra de 0.097. Construir un intervalo de confianza al 95 %, para el verdadero contenido promedio de litros de cloro puro. No se conoce la desviación típica de la población (La distribución del contenido de cloro por botella puede considerarse normal).

La respuesta correcta esperada es que el alumno construya un intervalo de confianza correcto, utilizando la distribución *t* con 15 *g.l.*, puesto que no se conoce la desviación típica de la población y el tamaño de la muestra es 16. Reproducimos un ejemplo, que se analiza en la Tabla 6.18.

"Como no se conoce la desviación típica de la población, para calcular el intervalo de confianza tenemos que usar la distribución t. Los grados de libertad son n-1=15. La desviación estándar de la muestra es S=0.097. El intervalo de confianza viene dado por la siguiente fórmula: $\bar{x} \pm t_{\alpha/2} \times \frac{s}{\sqrt{n}}$*, siendo* $t_{\alpha/2}$ *el valor crítico correspondiente al coeficiente de confianza 1-*α*, s la desviación típica de la muestra y n el tamaño de la muestra. El valor de la media muestral es 0.94 y tamaño de la muestra es 16, por lo que buscando en tablas en 15 grados de libertad y* $\alpha/2$*= 0.025 obtengo el valor crítico de T, con lo cual tengo todos los datos para sustituirlos en la fórmula:*

$$0.94-(2.13)\frac{0.097}{\sqrt{16}} < \mu < 0.94+(2.13)\frac{0.097}{\sqrt{16}}$$

Realizando las operaciones se llega al intervalo 0.8884<μ*<0.9916"* (Alumno BM).

Tabla 6.18 Análisis de la solución correcta al ítem 7

Unidad	Contenido	Funciones semióticas
U1	Como no se conoce la desviación típica de la población, para calcular el intervalo de confianza tenemos que usar la distribución *t*. Los grados de libertad son *n*-1=15. La desviación estándar de la muestra es *S*=0.097.	*Desviación típica de la población, intervalo de confianza, distribución, grados de libertad, desviación estándar de la muestra*: lenguaje y conceptos que se usan como ejemplar y tipo. *Tenemos que usar la distribución t*: el alumno identifica correctamente la distribución muestral (concepto usado como ejemplar) y distribución t (tipo particular) y los grados de libertad (propiedad).
U2	El intervalo de confianza viene dado por la siguiente fórmula: $\bar{x} \pm t_{\alpha/2} \times \frac{s}{\sqrt{n}}$, siendo $t_{\alpha/2}$ el valor crítico correspondiente al coeficiente de confianza 1-α, s la desviación típica de la muestra y n el tamaño de la muestra.	*Fórmula*: se refiere a la expresión: $\bar{x} \pm t_{\alpha/2} \times \frac{s}{\sqrt{n}}$ (lenguaje) y también al procedimiento de cálculo del intervalo. $\bar{x}$: hace referencia a la media de la muestra, que es un estadístico. El alumno tiene que diferenciar entre ella y la media de la población μ que es un parámetro (conceptos) $\pm$: hace referencia a que el intervalo se construye sumando y restando una misma cantidad (operación). $t_{\alpha/2} \times \frac{s}{\sqrt{n}}$: indica la cantidad a sumar, que es a su vez un producto (operación). $t_{\alpha/2}$ se refiere al valor crítico (concepto) en la distribución *T* de Student (concepto); *1-* α se refiere al nivel de confianza (concepto), $\alpha/2$ al área que queda a la derecha del valor de *t* (concepto), puesto que el nivel de confianza tiene que contemplar la probabilidad por encima y debajo de la media muestral (propiedad).
U3	El valor de la media muestral es 0.94 y tamaño de la muestra es 16, por lo que buscando en tablas en 15 grados de libertad y $\alpha/2$= 0.025 obtengo el valor crítico de *t*, con lo cual tengo todos los datos para sustituirlos en la fórmula. $0.94-(2.13)\frac{0.097}{\sqrt{16}} < \mu < 0.94+(2.13)\frac{0.097}{\sqrt{16}}$	*Media muestral, tamaño de la muestra, grados de libertad* : el alumno nombra diferentes conceptos. *Tablas* :lenguaje y conjunto de valores en ellas reflejados *Valor crítico de t*: concepto. *Por lo que buscando en tablas en 15 grados de libertad y $\alpha/2$= 0.025 obtengo el valor crítico de t* (procedimiento). *Con lo cual tengo todos los datos para sustituirlos en la fórmula*: (procedimiento). $0.94-(2.13)\frac{0.097}{\sqrt{16}} < \mu < 0.94+(2.13)\frac{0.097}{\sqrt{16}}$ (lenguaje y operaciones).
U4	Realizando las operaciones se llega al intervalo $0.8884<\mu<0.9916$.	El alumno ha de llevar a cabo varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), para llegar finalmente al intervalo pedido. $0.8884<\mu<0.9916$ (lenguaje).

Además de esta solución correcta, a la que daremos código C10, hemos encontrado la siguiente variante de respuesta, C11, en la que el alumno escribe correctamente la fórmula, al realizar el proceso obtiene correctamente el valor crítico de $t_{\alpha/2}$, pero al realizar las operaciones algebraicas tiene un conflicto relativo al procedimiento en el manejo de inecuaciones en el último paso; entonces el intervalo obtenido es incorrecto. Se reproduce un ejemplo que se analiza en la tabla 6.19.

"$n= 16, s = 0.097, 1\text{-}\alpha = 0.95, \bar{x} = 0.94$

$$\bar{x} \pm t_{\alpha/2} \times \frac{s}{\sqrt{n}} = 0.94 \pm (2.131)\frac{0.097}{\sqrt{16}} = 0.94 \pm 0.0516$$

$-0.0515 \leq \mu \leq 0.9916$" (Alumno MR).

Tabla 6.19. Análisis de la solución correcta al ítem 7

Unidad	Contenido	Funciones semióticas
U1	n = 16, s = 0.097 1-α = 0.95, $\bar{x}$ = 0.94	El alumno escribe una representación simbólica de los datos del problema que remite a varios conceptos (tamaño de muestra, desviación estándar muestral, nivel de confianza y media muestral) en el caso general, así como a los valores particulares en el problema.
U2	$\bar{x} \pm t_{\alpha/2} \times \frac{s}{\sqrt{n}}$	Escribe correctamente la fórmula del intervalo de confianza para la media (procedimiento). Recuerda la distribución muestral de la media para el caso de muestras pequeñas, σ desconocida (propiedad).
U2	$0.94 \pm (2.131)\frac{0.097}{\sqrt{16}}$	$= 0.94 \pm (2.131)\frac{0.097}{\sqrt{16}}$ lenguaje que se refiere al intervalo (concepto) y su cálculo (procedimiento). Calcula el valor crítico en la tabla de la *t* de student (procedimiento) y usa las tablas (lenguaje y procedimiento) correctamente, usando los grados de libertad correctos (propiedad).
U3	= 0.94 ± 0.0516	= 0.94 ± 0.0516 (lenguaje) que se refiere a los límites del intervalo (concepto) y su determinación (procedimiento) lleva a cabo correctamente varias operaciones algebraicas (procedimientos).
U4	-0.0515 ≤ μ ≤ 0.9916	-0.0515 ≤ μ ≤ 0.9916 (lenguaje); referencia al intervalo (concepto). No llega al intervalo correcto por tener un *conflicto* relativo al procedimiento con las operaciones algebraicas de manejo de inecuaciones.

Sólo hemos encontrado un tipo de respuesta parcialmente correcta C21: Calcula el valor crítico en la tabla de la *t* de Student y uso de las tablas correctamente, pero luego aparece un conflicto relacionado con el procedimiento de cálculo de los grados de libertad de la distribución *t*, como se exhibe en el caso siguiente, que se analiza en la Tabla 6.20, donde además hay una imprecisión de la notación.

"$\bar{x} \pm t \times \frac{s}{\sqrt{n}}$; $0.94 \pm (2.11)\frac{0.097}{\sqrt{16}}$; $0.089 < \mu < 0.991$" (Alumno HE).

Tabla 6.20. Análisis de la solución parcialmente correcta al ítem 7

Unidad	Contenido	Funciones semióticas
U1	$\bar{x} \pm t \times \frac{s}{\sqrt{n}}$	El alumno escribe la fórmula del intervalo de confianza para la media en forma parcialmente correcta. El alumno recuerda la distribución muestral de la media para el caso de muestras pequeñas, σ desconocida (propiedad). No escribe el subíndice α/2 en *t*.

U2	$0.94 \pm (2.11)\frac{0.097}{\sqrt{16}}$	Calcula el valor crítico en la tabla de la *t* de Student (procedimiento) y usa las tablas (lenguaje) correctamente, pero toma los grados de libertad incorrectos. $0.94 \pm (2.11)\frac{0.097}{\sqrt{16}}$ (lenguaje).
U3	$0.089 < \mu < 0.991$	Luego lleva a cabo correctamente varias operaciones algebraicas (procedimientos), pero no llega al intervalo correcto por las razones previamente expuestas. $0.089 < \mu < 0.991$ (lenguaje).

Respuestas incorrectas:

C31: El alumno tiene un conflicto relacionado con el cálculo de valores críticos de $t_{\alpha/2}$. Calcula correctamente los grados de libertad pero comete error en el uso de las tablas al tomar 5% a cada lado de las colas de la distribución *t* de Student, en vez de 2.5%, como se aprecia en el caso siguiente, que se analiza en la Tabla 6.21:

"*n = 16,* $\bar{x}$ *= .94. s = .097,* $.94 \pm 1.75\frac{(.097)}{\sqrt{16}}$ *, .8975<* μ*< .9824*" (Alumno LO).

Tabla 6.21. Análisis de la solución incorrecta al ítem 7

Unidad	Contenido	Funciones semióticas
U1	$n = 16$, $s = 0.097$ $\bar{x} = .94$	El alumno representa simbólicamente los datos del problema (tamaño de la muestra, media muestral y desviación estándar muestral).
U2	$.94 \pm 1.75\frac{(.097)}{\sqrt{16}}$	El alumno tiene un *conflicto* con la fórmula del intervalo de confianza para la media. Inicia sustituyendo los valores en un intervalo que correspondería al de la media poblacional. Calcula el valor crítico en la tabla *t* de Student (procedimiento) y hay *un conflicto* en el uso de las tablas (lenguaje) en forma incorrecta al tomar un valor de área de .05 en vez de .025 para 15 grados de libertad.
U4	$.8975 < \mu < .9824$	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto). $.8975 < \mu < .9824$ (lenguaje). Pero, por la razón señalada previamente, el resultado es incorrecto.

C32: El alumno tiene un *conflicto relativo a los conceptos* al escribir el intervalo de confianza para la media poblacional, escribe en el intervalo media muestral en lugar de la media poblacional y además realiza cálculo incorrecto de valores críticos de, al tomar 5% a cada lado de las colas de la distribución *t*, en lugar de 2.5%; como se observa en el siguiente caso que analizamos en la Tabla 6.22.

"*n = 16, s = 0.097 ,* $\bar{x}$ *= 0.94 ,* $.94 \pm 1.75\frac{(.097)}{\sqrt{16}}$ *, .8975 <* $\bar{x}$ *< .9824*" (Alumno AC).

Tabla 6.22. Análisis de la solución incorrecta al ítem 7

Unidad	Contenido	Funciones semióticas
U1	$n = 16$, $s = 0.097$ $\bar{x} = 0.94$	El alumno representa simbólicamente los datos del problema (tamaño de la muestra, media muestral y desviación estándar muestral).
U2	$.94 \pm 1.75 \frac{(.097)}{\sqrt{16}}$	El alumno tiene un *conflicto* con la fórmula del intervalo de confianza para la media, (olvida escribirla). Inicia sustituyendo valores en un intervalo que correspondería al de la media poblacional. Calcula el valor crítico en la tabla *t* de Student (procedimiento) y uso de las tablas (lenguaje) incorrectamente al tomar un valor de área de .05 en vez de .025 para 15 grados de libertad.
U3	$.8975 < \bar{x} < .9824$	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto). Pero al escribir la expresión del intervalo, tiene un c*onflicto* y escribe media muestral en lugar de media poblacional. $.8975 < \bar{x} < .9824$ (lenguaje).

C33: El alumno escribe la fórmula del intervalo de confianza para la media utilizando el valor crítico de $Z_{\alpha/2}$ en vez de. Usa la distribución normal para obtener los valores críticos, en vez de la distribución *t*, y además se presenta un conflicto conceptual al confundir la desviación estándar de la muestra con la desviación estándar de la población, como se exhibe en el caso que presentamos a continuación y que luego se analiza en la Tabla 6.23. Tanto en este caso como el anterior se produce una confusión entre ejemplar (media) y tipo (media de la muestra, media de la población); error que ha sido descrito entre otros por Vallecillos (1994).

"*n= 16, s= 0.097,* $\bar{x}$ *= 0.94, 1-α = 0.95, α/2 = .025,*

$$\bar{x} \pm \frac{(Z_{\alpha/2})\sigma}{\sqrt{n}}\ ,\ .94 \pm \frac{(Z_{\alpha/2}).097}{\sqrt{16}}\ ,\ .94 \pm \frac{(1.96).097}{\sqrt{16}}$$

.94 ± .047, (.893, .987)" (Alumno LK).

Tabla 6.23. Análisis de la solución incorrecta al ítem 7

Unidad	Contenido	Funciones semióticas
U1	$n = 16$, $s = 0.097$ $\bar{x} = 0.94$, $1\text{-}\alpha = 0.95$	El alumno representa simbólicamente los datos del problema (tamaño de la muestra, media muestral, nivel de confianza). Hace error de notación con la desviación estándar muestral. representándola como si fuera la desviación estándar poblacional.
U2	$\bar{x} \pm \frac{(Z_{\alpha/2})\sigma}{\sqrt{n}}$	El alumno escribe la fórmula del intervalo de confianza para la media utilizando el valor crítico de $Z_{\alpha/2}$ en vez de $t_{\alpha/2}$, como resultado de su confusión de la desviación estándar de la muestra con la desviación estándar de la población. Ello condiciona toda la solución.
U2	$1\text{-}\alpha = 0.95$, $\alpha/2 = .025$	El alumno calcula $\alpha/2$ haciendo la división del complemento del nivel de confianza por dos (procedimiento).

U3	$.94 \pm (1.96(\frac{.097}{\sqrt{16}}))$	Calcula el valor crítico en la tabla de la normal estándar (procedimiento) y uso de las tablas (lenguaje) correctamente.
U4	.94± .047	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto). .94± .047 (lenguaje)
U5	(.893, .987)	Para llegar finalmente al intervalo pedido (.893, .987) (lenguaje) Pero, por la razón señalada previamente el resultado es incorrecto.

C34: El alumno se equivoca, al escribir en su solución la fórmula del intervalo de confianza para la varianza. Se presenta un conflicto relativo a los conceptos pues los mezcla en la fórmula. El estudiante no puede distinguir entre desviación estándar de la muestra y la desviación estándar poblacional; asocia el parámetro σ con la distribución chi -cuadrado y esto lo transfiere a la fórmula del intervalo de confianza para la media, como se aprecia en el siguiente caso, cuyo análisis se presenta en la Tabla 6.24.

"*n=16, σ=0.097,* $\bar{x}$ *=0.94, 1-α=0.95, α/2=.025,* $\chi^2_{.025}$ *=27.488,* $\chi^2_{.975}$ *=6.262,* $\frac{15(.097)^2}{27.488} < \mu < \frac{15(.097)^2}{6.262}$ *, .005<* μ*<.023"* (Alumno HG).

Tabla 6.24. Análisis de la solución incorrecta al ítem 7

Unidad	Contenido	Funciones semióticas
U1	n= 16, $\bar{x}$ = 0.94, 1-α = 0.95 σ = 0.097 $\chi^2_{.025}$ =27.488 $\chi^2_{.975}$ = 6.262	El alumno representa simbólicamente los datos del problema (tamaño de la muestra, media muestral, nivel de confianza; conceptos). Conflicto consistente en confundir la desviación estándar de la muestra con la desviación estándar poblacional (conceptos). Confunde por ello la distribución muestral de la media para el caso de muestras pequeñas, σ desconocida (propiedad).
U2	$\frac{15(.097)^2}{27.488} < \mu < \frac{15(.097)^2}{6.262}$	El alumno escribe en su solución la fórmula del intervalo de confianza para la varianza (procedimiento). El valor que él había tomado como desviación estándar poblacional ahora lo usa como desviación estándar de la muestra (conflicto). Calcula correctamente los valores críticos en la tabla del Chi cuadrado (procedimiento) y uso de las tablas (lenguaje).
U3	.005< μ<.023	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto) para llegar finalmente al intervalo pedido. .005< μ<.023 (lenguaje). Pero, por la razón señalada previamente el resultado es incorrecto.

C35: El alumno escribe en forma incompleta un intervalo que corresponde al de la media poblacional y además se produce un conflicto relativo al procedimiento de cálculo al sustituir en la posición del valor crítico de $z_{\alpha/2}$ por el valor de probabilidad de

95%, como se observa en el caso que presentamos a continuación y que luego analizamos en la Tabla 6.25.

“*n = 16, x = 0.94, 1-α = 0.95, s = 0.097*

$0.95 \rightarrow \frac{(0.097)}{\sqrt{16}} < \mu < 0.94 + \frac{(0.95)(0.097)}{\sqrt{16}}$ ” (Alumno SL).

Tabla 6.25. Análisis de la solución incorrecta al ítem 7

Unidad	Contenido	Funciones semióticas
U1	*n*= 16, *x* = 0.94, *1-α* = 0.95 *s*= 0.097	El alumno, en la identificación de los datos, escribe el símbolo de media muestral en forma incorrecta. Usa varios conceptos generales y los valores particulares en el problema.
U2	$0.95 \rightarrow \frac{(0.097)}{\sqrt{16}} < \mu < 0.94 + \frac{(0.95)(0.097)}{\sqrt{16}}$	*El* alumno no escribe en su solución la fórmula del intervalo de confianza para la media. Escribe en forma incompleta un intervalo que corresponde al de la media. Tiene un conflicto al sustituir el valor de probabilidad de 95% en la posición correspondiente al valor del punto crítico. Y ya no continúa el proceso.

En la Tabla 6.26 presentamos los resultados del ítem 7 (construcción del intervalo de confianza para la media). Este ítem lo contestaron correctamente un 39.6% y una tercera parte de los estudiantes contestó en forma parcialmente correcta, por lo cual lo consideramos un ítem de dificultad moderada para los estudiantes.

Tabla 6.26. Frecuencias (y porcentajes) de respuestas en el ítem 7 (n=252)

	Frecuencia	Porcentaje	Porcentaje acumulado
Respuestas correctas C10	89	35.3	35.3
Respuestas correctas C11	11	4.4	39.6
Respuestas parcialmente correctas C21	87	34.5	74.2
Respuestas incorrectas C31	8	3.2	77.3
Respuestas incorrectas C32	3	1.2	78.5
Respuestas incorrectas C33	29	11.5	90.0
Respuestas incorrectas C34	7	2.8	92.8
Respuestas incorrectas C35	2	.8	93.6
Blanco	16	6.3	100.0
Total	252	100.0	

El 34.5% de los estudiantes que respondió en forma parcialmente correcta tuvo conflicto en los procedimientos de cálculo, en la obtención de los grados de libertad de la distribución *t*, tomando *n* en lugar de *n-1*, confusión simplemente memorística, que no tiene mucha importancia en cuanto los programas estadísticos hoy día dan el cálculo automático. Un 19.5% lo resolvió incorrectamente. El 3.2% tuvo conflicto en la obtención de los valores críticos a partir de las tablas, al tomar 5 % a cada lado de las

colas de la distribución *t*, en vez de 2.5% lo que es una dificultad no ligada al concepto de intervalo de confianza sino la falta de comprensión de la idea de valor crítico.

Un 1.2% escribe en el intervalo media muestral, en vez de media poblacional además del conflicto relativo a sus procedimientos en la obtención de los valores críticos de, al tomar 5 % a cada lado de las colas de la distribución *t*, en vez de 2.5%. Un 11.5% tiene el conflicto relativo a las entidades conceptuales al confundir la desviación estándar de la muestra con la desviación estándar poblacional, por lo que utilizan la distribución normal para calcular los valores críticos en vez de utilizar la distribución *t.* Un 2.8% hace error al escribir la fórmula del intervalo de confianza para la varianza, pero el parámetro que escribe al centro del intervalo es el de la media poblacional, con lo cual se exhibe el conflicto del alumno de su incapacidad de relacionar adecuadamente los parámetros media y varianza con sus límites de confianza correspondientes. Finalmente el .8% (dos alumnos) tienen un conflicto relativo al procedimiento de cálculo del valor crítico, escriben el valor de probabilidad de 0.95 en lugar del valor crítico de *t*. Llama la atención que un 6.3% fueron respuestas en blanco.

6.3.4. ANÁLISIS DEL ÍTEM 9. CONSTRUCCIÓN DEL INTERVALO DE CONFIANZA PARA LA PROPORCIÓN

Ítem 9. En una muestra aleatoria de 100 rodamientos, 10 tienen un acabado de especificaciones defectuoso. Calcular el intervalo de confianza de 95% para la proporción verdadera de rodamientos defectuosos.

La solución correcta que el estudiante debe escribir incluye: la ecuación correcta del intervalo de confianza para la proporción poblacional para muestras grandes, la aplicación de la distribución normal estándar para obtener correctamente los puntos críticos del intervalo, a partir del nivel de confianza correspondiente y la obtención correcta del intervalo de confianza utilizando esa información. Reproducimos un ejemplo, que se analiza en la Tabla 6.27. Esperamos el siguiente razonamiento de los estudiantes (C10):

> *"Para calcular el intervalo de confianza para una proporción poblacional con muestras grandes tenemos que usar la distribución normal estándar. La proporción de casos favorables la calculamos dividiendo 10 éxitos entre el tamaño de la muestra 100. El intervalo de confianza viene dado por la siguiente fórmula:* $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(\hat{q})}{n}}$, *siendo* $Z_{\alpha/2}$ *el valor crítico correspondiente al coeficiente de confianza 1-*α*,* $\hat{p}$ *la proporción*

favorable de la muestra, $\hat{q}$ la proporción desfavorable y n el tamaño de la muestra. El valor de la proporción favorable de la muestra es 0.10 y valor de la proporción desfavorable de la muestra es 0.90. Buscando en tablas en la normal estándar, para $\alpha/2 = 0.025$ obtengo el valor crítico de, con lo cual tengo todos los datos para sustituirlos en la fórmula,, obteniendo finalmente el intervalo de confianza para la proporción poblacional 0.0412 < p < 0.1588" (Alumno JN).

Tabla 6.27. Análisis de la solución correcta al ítem 9

Unidad	Contenido	Funciones semióticas
U1	Para calcular el intervalo de confianza para una proporción poblacional con muestras grandes tenemos que usar la distribución normal estándar. La proporción de casos favorables la calculamos dividiendo 10 éxitos por el tamaño de la muestra 100.	*Proporción poblacional, intervalo de confianza, muestras grandes*: el alumno nombra diferentes conceptos. *Tenemos que usar la distribución normal estándar:* El alumno discrimina la distribución muestral a emplear. *Proporción de casos favorables*: el alumno evoca un concepto. *Dividiendo*: describe una operación, *éxitos, tamaño de la muestra*: conceptos.
U2	El intervalo de confianza viene dado por la siguiente fórmula: $\hat{p} \pm Z_{\alpha/2} \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$, siendo $Z_{\alpha/2}$ el valor crítico correspondiente al coeficiente de confianza 1-α, $\hat{p}$ la proporción favorable de la muestra, $\hat{q}$ la proporción desfavorable y n el tamaño de la muestra	*Fórmula*: se refiere a la expresión: $\hat{p} \pm Z_{\alpha/2} \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$ y también al procedimiento de cálculo del intervalo (procedimiento y lenguaje). $\hat{p}$ hace referencia a la proporción muestral, que es un estadístico. El alumno tiene que diferenciar entre ella y la proporción de la población p que es un parámetro (conceptos). $\pm$ hace referencia a que el intervalo se construye sumando y restando una misma cantidad (operación) $Z_{\alpha/2} \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$ hace referencia a la cantidad a sumar, que es a su vez un producto (operación). $Z_{\alpha/2}$ indica al valor crítico (concepto) en la distribución normal estándar (concepto); 1-α se refiere al nivel de confianza (concepto), α/2 al área que queda a la derecha del valor de $Z_{\alpha/2}$ (concepto), puesto que el nivel de confianza tiene que contemplar la probabilidad por encima y debajo de la proporción muestral (propiedad).
U3	El valor de la proporción favorable de la muestra es 0.10 y valor de la proporción desfavorable de la muestra es 0.90, por lo que buscando en tablas en la normal estándar, para *$\alpha/2 = 0.025$* obtengo el valor crítico de $Z_{\alpha/2}$, con lo cual tengo todos los datos para sustituirlos en la fórmula. $0.1 \pm (1.96)(\sqrt{\frac{(0.1)(0.9)}{100}})$	*Proporción favorable de la muestra, proporción desfavorable* (conceptos), *tablas*(lenguaje) *valor crítico de $Z_{\alpha/2}$* (concepto). *Por lo que buscando en tablas en la normal estándar, para $\alpha/2 = 0.025$ obtengo el valor crítico de $Z_{\alpha/2}$*: (procedimiento). *Con lo cual tengo todos los datos para sustituirlos en la fórmula*: (procedimiento). $0.1 \pm (1.96)(\sqrt{\frac{(0.1)(0.9)}{100}})$: (lenguaje).
U4	obteniendo finalmente el intervalo de confianza para la proporción poblacional *0 .0412< p < 0.1588*	*Obteniendo finalmente el intervalo de confianza para la proporción poblacional*: (concepto y procedimiento). 0 .0412< *p* < 0.1588 (lenguaje).

A esta solución correcta le hemos dado código C10, y además hemos encontrado esta otra forma de respuesta correcta:

C11: En la fórmula no escribe $\alpha/2$ como subíndice de *Z*, $\hat{p} \pm Z \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$ pero al realizar el proceso obtiene correctamente el valor crítico de $z_{\alpha/2}$ como se observa en el siguiente caso. La tabla 6.28 muestra su análisis (Alumno JRC).

“ $\hat{p}$ = 10/100 = 0.1, $\hat{q}$ = 1-0.1 = 0.9, $\hat{p} \pm Z \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$

$0.1-(1.96)(\sqrt{\frac{(0.1)(0.9)}{100}}) < p < 0.1+(1.96)(\sqrt{\frac{(0.1)(0.9)}{100}})$

0.1-(1.96)(0.03) < p < 0.1+ (1.96)(0.03), 0 .0412< p < 0.1588”.

Tabla 6.28. Análisis a otra solución correcta al ítem 9

Unidad	Contenido	Funciones semióticas
U	$\hat{p} \pm Z \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$	El alumno escribe la fórmula del intervalo de confianza para la proporción poblacional incorrectamente. Hace error de notación al no escribir $\alpha/2$ como subíndice. Aunque ello no afecta la solución al continuar en forma correcta el resto de los procedimientos.
U2	$\hat{p}$ = 10/100 = 0.1, $\hat{q}$ = 1-0.1 = 0.9	El alumno realiza en forma correcta las operaciones necesarias para obtener los valores de $\hat{p}$ y $\hat{q}$.
U3	$0.1+(1.96)(\sqrt{\frac{(0.1)(0.9)}{100}})$	Obtiene el valor crítico en la tabla de la normal estándar (procedimiento) y uso de las tablas (lenguaje) correctamente.
U4	0 .0412< p < 0.1588	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), para llegar finalmente al intervalo pedido. 0 .0412< p < 0.1588 (lenguaje).

Respuestas parcialmente correctas:

C21: Luego de escribir la fórmula del intervalo de confianza para la proporción poblacional correctamente, se presenta un *conflicto relativo al procedimiento de cálculo de los valores críticos* de, al tomar 5% a cada lado de las colas de la distribución normal, en vez de 2.5%, como se observa en el caso que presentamos a continuación y que se analiza en la Tabla 6.29.

“ $\hat{p}$ = 10/ 100= .10, $\hat{q}$ = 1-.10= .9, $0.1 \pm (1.64)(\sqrt{\frac{(0.1)(0.9)}{100}})$

0 .0412< p < 0.1588” (Alumno JS).

Tabla 6.29. Análisis a una solución parcialmente correcta al ítem 9

Unidad	Contenido	Funciones semióticas
U1	$\hat{p} \pm Z_{\alpha/2} \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$	El alumno escribe la fórmula del intervalo de confianza para la proporción poblacional correctamente. Los símbolos remiten a la proporción poblacional y su complementario, tamaño de muestra y coeficiente de confianza (conceptos) y las operaciones con ellos (suma, raíz, multiplicación).
U2	$\hat{p}$ = 10/ 100= .10 $\hat{q}$ = 1-.10= .90	El alumno realiza en forma correcta las operaciones necesarias para obtener los valores de $\hat{p}$ y; se refiere a las estimaciones de la proporción poblacional y su complementario (conceptos) y los valores obtenidos en las mismas (datos).
U3	$0.1 \pm (1.64)(\sqrt{\frac{(0.1)(0.9)}{100}})$	Obtiene el valor crítico en la tabla de la normal estándar y lleva a cabo varios cálculos (procedimientos) y uso de las tablas (lenguaje) incorrectamente. El alumno tiene un conflicto al tomar 5% a cada lado de las colas.
U4	0 .0508 < p < 0.1492	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), pero no obtiene el intervalo correcto por las razones previamente expuestas. 0 .0508 < p < 0.1492 (lenguaje).

Respuestas incorrectas:

C31: *Error al escribir la fórmula del intervalo de confianza para la proporción poblacional.* El alumno escribe incorrectamente la fracción muestral de éxitos y la fracción muestral de fracasos en el radicando. Luego se presenta un conflicto relacionado con la obtención de los valores de $\hat{p}$ y. La presencia de este conflicto y del error en la fórmula se observa en el siguiente caso. La Tabla 6.30 muestra su análisis.

"$\hat{p} + \hat{q} = 1$, $\hat{p} = 1 - \hat{q}$, $\hat{p} = 1\text{-}10$, $\hat{p} - Z_{\alpha/2} \times \sqrt{\frac{p(q)}{n}} < p < \hat{p} + Z_{\alpha/2} \times \sqrt{\frac{p(q)}{n}}$

$9 - (1.96)(\sqrt{\frac{(9)(10)}{100}}) < p < 9 + (1.96)(\sqrt{\frac{(9)(10)}{100}})$, *7.1406< p < 10.8594"* (Alumno LL).

Tabla 6.30. Análisis de la solución incorrecta al ítem 9

Unidad	Contenido	Funciones semióticas
U1	$\hat{p} - Z_{\alpha/2} \times \sqrt{\frac{p(q)}{n}} < p < \hat{p} + Z_{\alpha/2} \times \sqrt{\frac{p(q)}{n}}$	El alumno escribe la fórmula del intervalo de confianza para la proporción poblacional incorrectamente. Hace referencia a la proporción, su complementaria, tamaño de muestra, valor crítico y nivel de confianza (conceptos) y varias operaciones entre ellos (procedimientos). Hay un *conflicto* al escribir los símbolos para la fracción muestral de éxitos y la fracción muestral de fracasos en el radicando. Aunque ello pudiera no haber afectado toda la solución de haber obtenido correctamente dichas fracciones muestrales.
U2	$\hat{p} + \hat{q} = 1$, $\hat{p} = 1 - \hat{q}$, $\hat{p} = 1\text{-}10$	El alumno tiene un conflicto sobre el significado de $\hat{p}$ y $\hat{q}$ por lo que hace errores en la obtención de sus valores.
U2	$9 + (1.96)(\sqrt{\frac{(9)(10)}{100}})$	Calcula el valor crítico en la tabla de la normal estándar (procedimiento) y uso de las tablas (lenguaje)

		correctamente, pero tiene un conflicto en la obtención del valor de fracción muestral de éxitos y la fracción muestral de fracasos.
U5	7.1406< p< 10.8594	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), para llegar finalmente al intervalo pedido. 7.1406< p < 10.8594 (lenguaje). Pero, por las razones señaladas el resultado es incorrecto.

C32: El alumno escribe correctamente la fórmula del intervalo de confianza para la proporción poblacional. Luego se presenta un conflicto relativo a los procedimientos de cálculo. Error en la obtención de los valores de $\hat{p}$ y $\hat{q}$ y además el alumno hace error en la obtención de los valores críticos de $Z_{\alpha/2}$ al tomar 5% a cada lado de las colas de la distribución normal estándar, en lugar de 2.5%. La Tabla 6.31 muestra el análisis de un caso en donde podemos observar este conflicto.

"$\hat{p} \pm Z_{\alpha/2} \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$, $\hat{p}$ = *10*, $\hat{q}$ = *90*, $10 \pm (1.64)(\sqrt{\frac{(10)(90)}{100}})$
5.08< p < 14.92" (Alumno AB).

Tabla 6.31. Análisis a una solución incorrecta al ítem 9

Unidad	Contenido	Funciones semióticas
U1	$\hat{p} \pm Z_{\alpha/2} \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$	El alumno escribe la fórmula del intervalo de confianza para la proporción poblacional correctamente. Hace referencia a la proporción, su complementaria, tamaño de muestra, valor crítico y nivel de confianza (conceptos) y varias operaciones entre ellos (procedimientos).
U2	$\hat{p} = 10$, $\hat{q} = 90$	El alumno tiene un conflicto y obtiene en forma incorrecta los valores de $\hat{p}$ y $\hat{q}$.
U3	$10 \pm (1.64)(\sqrt{\frac{(10)(90)}{100}})$	Obtiene el valor crítico en la tabla de la normal estándar (procedimiento) y hay un conflicto en el uso de las tablas (lenguaje) al tomar 5% a cada lado de las colas.
U4	5.08< p < 14.92	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), pero no obtiene el intervalo correcto por las razones previamente expuestas. 5.08< p < 14.92 (lenguaje).

C33: Otros. En este código encasillamos una respuesta de un alumno que escribe una expresión en el intervalo de confianza para la proporción poblacional que es muy distinta a los límites $\hat{p} \pm Z_{\alpha/2} \times \sqrt{\frac{\hat{p}(\hat{q})}{n}}$, se parecía un poco a un intervalo de confianza para la varianza poblacional. También incluimos en esta codificación la respuesta de otro alumno que escribe como intervalo 10.

Tabla 6.32. Frecuencias (y porcentajes) de respuestas en el ítem 9 (n=252)

	Frecuencia	Porcentaje	Porcentaje acumulado
Respuestas correctas C10	64	25.4	25.4
Respuestas correctas C11	7	2.8	28.2
Respuestas parcialmente correctas C21	33	13.1	41.3
Respuestas incorrectas C31	91	36.1	77.4
Respuestas incorrectas C32	30	11.9	89.3
Respuestas incorrectas C33	2	.8	90.1
Blanco	25	9.9	100.0
Total	252	100.0	

En la Tabla 6.32 presentamos los resultados del ítem 9 (construcción del intervalo de confianza para la proporción). Este ítem resultó difícil para los estudiantes, ya que el porcentaje de respuestas incorrectas es de 48.8 %, por lo que más de la mitad de los alumnos tuvieron dificultades con este ítem. El 13.1% de los que contestaron correctamente en forma parcial tuvieron conflicto en el procedimiento de cálculo, en la obtención de los valores críticos $z_{\alpha/2}$ al tomar 5% a cada lado de las colas de la distribución normal, en lugar de 2.5%, por lo que el concepto de coeficiente de confianza o el uso de las tablas en la determinación de valores críticos no son bien comprendidos.

El 36.1% presentó como conflicto principal uno relacionado también con el procedimiento de cálculo (la obtención de los valores de $\hat{p}$ y) y además tiene conflictos en la aprehensión de la fórmula, escriben incorrectamente la fracción muestral de éxitos y la fracción muestral de fracasos en el radicando. El 11.9% tiene conflictos relativos a la proposición fracción muestral de éxitos y fracción muestral de fracasos, obtienen incorrectamente los valores de $\hat{p}$ y; además otro de los conflictos tiene que ver con el procedimiento de la obtención de los valores críticos $z_{\alpha/2}$ al tomar $\alpha/2 =.05$ a cada lado de las colas de la distribución normal, en lugar de .025. Resaltamos que un porcentaje importante (9.9%) dejó la respuesta en blanco.

6.3.5. ANÁLISIS DEL ITEM 10. CONSTRUCCIÓN DEL INTERVALO DE CONFIANZA PARA LA VARIANZA

Ítem 10. Sea σ^2 la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es s^2=13700, *n*=16. Calcular el intervalo de confianza de 95% para *σ*.

En este ítem, tomado de Devore (2005, pg. 310), se pide calcular un intervalo de confianza para la desviación estándar poblacional, conocida la varianza muestral y el tamaño de muestra. Puesto que el tamaño de la muestra es pequeño se debe usar la distribución Chi cuadrado. Se reproduce un ejemplo que se analiza en la tabla 6.33. El razonamiento correcto que esperamos es el siguiente:

> *"Sea σ^2 la varianza. Sabemos que se verifica $\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(1-\alpha/2)}}$, entonces $\frac{(16-1)13700}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(16-1)13700}{\chi^2_{(1-\alpha/2)}}$, es decir, $\frac{(16-1)13700}{27.488} < \sigma^2 < \frac{(16-1)13700}{6.262}$, sustituyendo en las inecuaciones se llega al intervalo 7975.9895< σ^2 <32816.991, luego obteniendo raíz cuadrada a ambos lados de la desigualdad se obtiene el intervalo de confianza para la desviación estándar poblacional, 89.308< σ <181.154"* (Alumno AC).

Tabla 6.33. Análisis de la solución correcta al ítem 10

Unidad	Contenido	Funciones semióticas
U1	Sea σ^2la varianza	*Sea σ^2* la *varianza:* El alumno evoca la varianza de la población (ejemplar) y el concepto general de varianza (tipo). El alumno tiene que distinguir la varianza de la población y la varianza de la muestra (conceptos) y saber que el intervalo se refiere a la de la población (propiedad).
U2	Sabemos que se verifica $\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(1-\alpha/2)}}$,	$\chi^2_{\alpha/2}$, $\chi^2_{(1-\alpha/2)}$ se refieren a la distribución Chi cuadrada (concepto). El alumno ha de recordar la distribución muestral de la varianza para el caso de muestras pequeñas (propiedad). Además se refiere a los valores críticos de $\chi^2_{\alpha/2}$ y $\chi^2_{(1-\alpha/2)}$ correspondientes al nivel de confianza 95% (conceptos y propiedades). Por otro lado debe recordar que las probabilidades necesarias para calcular el intervalo son precisamente estas, siendo 1-α el coeficiente de confianza (propiedad).
U3	entonces $\frac{(16-1)13700}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(16-1)13700}{\chi^2_{(1-\alpha/2)}}$	El alumno sustituye los datos del problema en las inecuaciones anteriores (procedimiento).
U4	es decir, $\frac{(16-1)13700}{27.488} < \sigma^2 < \frac{(16-1)13700}{6.262}$,	Ha de calcular los valores críticos en la tabla del Chi cuadrado (procedimiento) y uso de las tablas (lenguaje).
U5	sustituyendo en las	El alumno ha de llevar a cabo varias operaciones algebraicas

	inecuaciones se llega al intervalo *$7975.9895 < \sigma^2 < 32816.991$*	(procedimientos) con inecuaciones (concepto), para llegar finalmente al intervalo pedido. A lo largo de toda la solución ha de usar lenguaje simbólico, verbal y numérico.
	luego obteniendo raíz cuadrada a ambos lados de la desigualdad se obtiene el intervalo de confianza para la desviación estándar poblacional $89.308 < \sigma < 181.154$	*Raíz cuadrada:* (describe una operación). *Desviación estándar poblacional*: (el alumno evoca un concepto). *Obteniendo raíz cuadrada a ambos lados de la desigualdad se obtiene el intervalo de confianza para la desviación estándar poblacional*: razonamiento deductivo donde se pone en juego la función raíz cuadrada de la varianza. $89.308 < \sigma < 181.154$ (lenguaje).

A esta solución correcta le hemos asignado el código C10, además de esta solución hemos encontrado las siguientes variantes de respuestas:

C11: El alumno olvida escribir la expresión del intervalo de confianza para la varianza, pero luego continúa correctamente el resto del proceso hasta llegar a obtener finalmente el intervalo para la desviación estándar, como se puede observar en el caso que se presenta enseguida y que se analiza en la Tabla 6.34.

"*$1-\alpha = 0.95$, $\alpha = 0.05$, $\alpha/2 = 0.025$ $\chi^2_{.025,15} = 27.488$, $\chi^2_{.975,15} = 6.262$*
$\frac{(16-1)13700}{27.488} < \sigma^2 < \frac{(16-1)13700}{6.262}$, *$7475.99 < \sigma^2 < 32817$, $86.46 < \sigma < 181.15$*" (Alumno RF).

Tabla 6.34. Análisis a otra solución correcta del ítem 10

Unidad	Contenido	Funciones semióticas
U1	$\frac{(16-1)13700}{27.488} < \sigma^2 < \frac{(16-1)13700}{6.262}$	El alumno ha olvidado escribir la fórmula del intervalo de confianza para la varianza. Sin embargo ello no condiciona toda la solución, pues luego continúa todo el proceso en forma correcta.
U2	$1-\alpha = 0.95$, $\alpha = 0.05$, $\alpha/2 = 0.025$	El alumno obtiene el valor de $\alpha/2$, que será posteriormente utilizado para obtener los puntos críticos.
U2	$\chi^2_{.025,15} = 27.448$ $\chi^2_{.975,15} = 6.262$	Calcula los valores críticos en la tabla del Chi cuadrado (procedimiento) y uso de las tablas (lenguaje) correctamente.
U3	$\frac{(16-1)13700}{27.488} < \sigma^2 < \frac{(16-1)13700}{6.262}$ *$7475.99 < \sigma^2 < 32817$*	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), obteniendo correctamente el intervalo de confianza para la varianza. *$7475.99 < \sigma^2 < 32817$* (lenguaje).
U4	$86.46 < \sigma < 181.15$	Obtiene raíz cuadrada en los extremos del intervalo para llegar finalmente al intervalo pedido. $86.46 < \sigma < 181.15$ (lenguaje).

Respuestas parcialmente correctas

C21: El alumno inicia correctamente con una representación gráfica de la distribución adecuada para este intervalo, dividiendo además la probabilidad total en tres partes, pero al avanzar surge un conflicto relacionado con los procedimientos al calcular incorrectamente los grados de libertad de la distribución Chi cuadrado. Los grados de libertad los obtiene sin restar 1 al valor del tamaño de la muestra, como en el caso siguiente cuyo análisis se presenta en la Tabla 6.35, donde además hay una confusión de la notación: (Alumno EG).

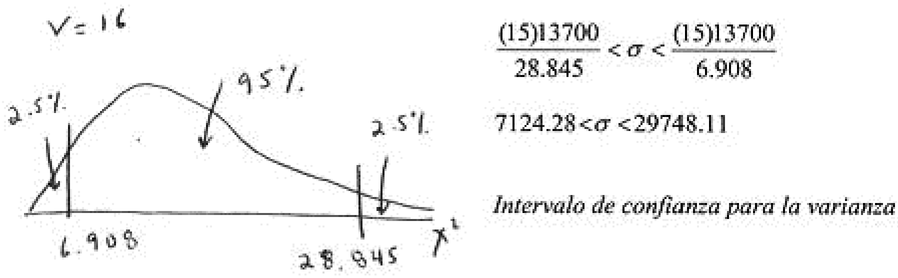
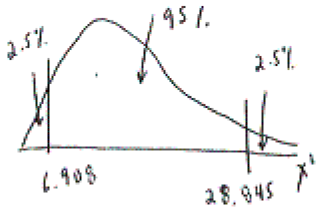


Tabla 6.35. Análisis de la solución parcialmente correcta al ítem 10

Unidad	Contenido	Funciones semióticas
U1		El alumno introduce una representación gráfica de la distribución Chi cuadrado que ha identificado correctamente. Divide la probabilidad total bajo la curva (1) en tres partes, dejando el coeficiente de confianza (95% en el centro) y el nivel 5% dividido en dos partes a cada lado (procedimiento y propiedades). Sin embargo tiene un conflicto relativo al procedimiento al obtener los grados de libertad, tomando 16 en vez de 15. Esto condiciona toda la solución, pues los límites obtenidos serán incorrectos.
U2	$\frac{(15)13700}{28.845} < \sigma < \frac{(15)13700}{6.908}$,	Calcula los valores críticos en la tabla del Chi cuadrado (procedimiento) y uso de las tablas (lenguaje) correctamente, pero al ser los grados de libertad incorrectos el valor obtenido es incorrecto. El alumno ha hecho correctamente todos los pasos descritos en la categoría de solución correcta, aunque no los explicita. Tiene también un conflicto de notación, al usar la terminología de la varianza con la correspondiente a la desviación típica.
U5	7124 < σ < 29748.11 Intervalo de confianza para la varianza	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), para llegar finalmente al intervalo pedido. $7124 < \sigma < 29748.11$ (lenguaje). Pero, por la razón señalada el resultado es incorrecto *Intervalo de confianza para la varianza:* el alumno resalta su confusión en la notación.

C22: El estudiante escribe el intervalo $\frac{(n-1)S^2}{\chi^2_?} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_?}$ olvidando las expresiones en los denominadores del intervalo, luego continúa en forma correcta el resto del procedimiento. No llega al paso final de obtener el intervalo para la desviación estándar poblacional. La Tabla 6.36 muestra el análisis de un caso en donde podemos observar este error.

"$\frac{(n-1)S^2}{\chi^2_?} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_?}$, $\frac{(16-1)13700}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(16-1)13700}{\chi^2_{\alpha/2}}$,

$\frac{(16-1)13700}{6.262} < \sigma^2 < \frac{(16-1)13700}{24.736}$" (Alumno DL).

Tabla 6.36. Análisis de la solución parcialmente correcta al ítem 10

Unidad	Contenido	Funciones semióticas
U1	$\frac{(n-1)S^2}{\chi^2_?} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_?}$	El alumno hace referencia al tamaño de muestra, varianza y distribución Chi cuadrado (conceptos) y escribe una fórmula (procedimiento). Ha olvidado, momentáneamente, el orden en que deben aparecer los puntos críticos de la distribución Chi cuadrado que ha identificado correctamente. Sin embargo ello no condiciona toda la solución, pues luego escribe los valores críticos obtenidos en el orden correcto.
U2	$\frac{(16-1)13700}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(16-1)13700}{\chi^2_{\alpha/2}}$	El alumno tiene un *conflicto* relativo al procedimiento, en las expresiones de los denominadores de la fórmula.
U3	$\frac{(16-1)13700}{6.262} < \sigma^2 < \frac{(16-1)13700}{24.736}$	Calcula los valores críticos en la tabla del Chi cuadrado (procedimiento) y uso de las tablas (lenguaje) en forma parcialmente correcta. Tiene un *conflicto* procedimental al leer en 13 grados de libertad en vez de 15 para uno de los puntos críticos.
U4	$\frac{(16-1)13700}{6.262} < \sigma^2 < \frac{(16-1)13700}{24.736}$	El alumno, siendo consistente con lo que escribió en el segundo paso, cambia el orden de los extremos, escribiendo la fracción de mayor valor a la izquierda. Deja incompleto el proceso que debiera terminar con la obtención del intervalo de confianza para la desviación estándar poblacional.

Respuestas incorrectas

C31: El alumno recuerda levemente un límite de la fórmula del intervalo de confianza para la varianza. Pone en juego distintas funciones semióticas antes de presentarse un conflicto con la obtención de los grados de libertad de la distribución Chi- cuadrado. Se analiza en la Tabla 6.37 un caso que exhibe este error.

“$\sigma^2 = \dfrac{(n-1)s^2}{\chi_{1-\alpha/2,\nu}} = \dfrac{(15)(13700)}{\chi_{1-\alpha/2,\nu}}$, *(1-α)% = 95%; α = 1-95%; α = 5%*

$\dfrac{(15)13700}{\chi^2_{.025,16}} = \dfrac{205500}{28.845} = 7124.284$” (Alumno MV).

Tabla 6.37. Análisis de la solución incorrecta al ítem 10

Unidad	Contenido	Funciones semióticas
U1	$\sigma^2 = \dfrac{(n-1)s^2}{\chi_{1-\alpha/2,\nu}} = \dfrac{(15)(13700)}{\chi_{1-\alpha/2,\nu}}$	Hace referencia al tamaño de muestra, varianza y distribución Chi cuadrado, grados de libertad (conceptos) y escribe una fórmula (procedimiento). El alumno tiene un *conflicto* al recordar solamente un límite de la fórmula del intervalo de confianza para la varianza.
U2	(1-α)% = 95% ; α = 1-95%; α = 5%	Sabe como identificar el valor, que al dividir por 2, usará posteriormente para obtener el valor crítico de la distribución chi cuadrada.
U3	$\dfrac{(15)13700}{\chi^2_{.025,16}} = \dfrac{205500}{28.845} = 7124.284$	Calcula el valor crítico en la tabla del Chi- cuadrado (procedimiento) y uso de las tablas (lenguaje) correctamente, pero tiene un *conflicto* sobre los grados de libertad. Luego lleva a cabo correctamente varias operaciones algebraicas (procedimientos), pero eso es todo lo que escribe en su solución.

C32: Otros. En este grupo están las respuestas de los alumnos que escriben una fórmula para el intervalo de confianza para la varianza que se parece más bien a la fórmula del intervalo para la media poblacional. También se incluyen las respuestas de los que escriben una fórmula que se acerca a la de la varianza, porque escriben $Z_{\alpha/2}$ en vez de escribir, escriben $\dfrac{(n-1)Z_{\alpha/2}}{\chi^2_{\alpha/2}} < \sigma^2 < \dfrac{(n-1)Z_{\alpha/2}}{\chi^2_{(1-\alpha/2)}}$.

Tabla 6.38. Frecuencias (y porcentajes) de respuestas en el ítem 10 (n=252)

	Frecuencia	Porcentaje	Porcentaje acumulado
Respuestas correctas C10	94	37.3	37.3
Respuestas correctas C11	3	1.2	38.5
Respuestas parcialmente correctas C21	42	16.7	55.2
Respuestas parcialmente correctas C22	15	6.0	61.2
Respuestas incorrectas C31	70	27.8	89.0
Respuestas incorrectas C32	10	4.0	93.0
Blanco	18	7.1	100.0
Total	252	100.0	

En la Tabla 6.38 presentamos los resultados del ítem 10 (calcular un intervalo de confianza para la desviación estándar poblacional). Este ítem resultó difícil para los estudiantes, solamente 38.5% respondieron en forma correcta. Un 16.7% respondieron en forma parcialmente correcta, su conflicto es de procedimiento, ya que calculan incorrectamente los grados de libertad de la distribución Chi-cuadrado. Podríamos suponer que el exceso de símbolos utilizados en la obtención de intervalos de confianza, provoca que se pierda el alumno y algo tan sencillo para los estadísticos como es el concepto de grados de libertad de la distribución no alcanza a ser retenido. En la sexta parte de los estudiantes que participaron en el estudio, esa desatención se magnifica con otro error de distracción que sufre el alumno al escribir desviación estándar, en vez de varianza. Sea desatención o no, lo que podríamos sospechar es de un vacío de significados para el estudiante de los objetos ostensivos intervinientes en estos procesos.

Otro 6% que cae en esta categoría de responder en forma parcialmente correcta, corresponde a los alumnos que tienen el conflicto procedimental de olvidar parte de la fórmula del intervalo de confianza para la varianza. Un 31.8% responden en forma incorrecta, que sumado al porcentaje de alumnos que dejaron en blanco la respuesta (7.1%), igualan al porcentaje de los que respondieron en forma correcta En este grupo que contestan incorrectamente el mayor porcentaje (27.8%) corresponden al conflicto procedimental relacionado con los límites del intervalo de confianza para la varianza, los alumnos recuerdan solamente un límite del intervalo. La familiaridad que alcanzaron con las fórmulas del intervalo de confianza para la media poblacional, no logran transferirla a este intervalo.

6.3.6. ANÁLISIS DEL ITEM 12. INTERVALO PARA LA DIFERENCIA DE MEDIAS

Ítem 12. Se compararon dos soluciones de grabado diferentes, usando dos muestras aleatorias de tamaño 10 cada una. Las poblaciones tienen la misma varianza. Los resultados de la rapidez de grabado fueron:
Solución 1: $\bar{x} = 9.97$ y $s = .422$
Solución 2: $\bar{x} = 10.4$ y $s = .073$
$s_p = 0.34$ (desviación típica conjunta)
Calcule un intervalo de confianza de 90% para la diferencia de las medias de la rapidez de grabado.

La respuesta correcta esperada de este ítem requiere que el alumno escriba la ecuación correcta del intervalo de confianza para la diferencia de medias con σ_1^2 *y* σ_2^2 desconocidas pero iguales y muestras pequeñas. Se espera que aplique la distribución *t* para obtener correctamente los valores críticos utilizando los grados de libertad y el nivel de confianza correspondiente. Debe sustituir los valores y obtener correctamente el intervalo de confianza. Un ejemplo se reproduce a continuación y se analiza en la Tabla 6.39. Esperamos el siguiente razonamiento de los estudiantes:

> *"Como desconocemos las varianzas de las poblaciones y las muestras son pequeñas, para calcular el intervalo de confianza para la diferencia de medias poblacionales, tenemos que usar la distribución t. Las desviaciones estándar muestrales de ambos grupos son s = .422 y s = .073 respectivamente. El intervalo de confianza viene dado por la siguiente fórmula:* $\overline{x_1} - \overline{x_2} \pm t_{\alpha/2} \times (S_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ *, siendo* $t_{\alpha/2}$ *el valor crítico correspondiente al coeficiente de confianza 1-α,* S_p *la desviación estándar conjunta y* n_1 *y* n_2 *los tamaños de las muestras. El valor de las medias muestrales son 9.97 y 10.4 respectivamente, los tamaños de las muestras son 10 para cada una, por lo que buscando en tablas de la distribución t de student en α/2= 0.025 obtengo el valor crítico de* $t_{\alpha/2}$ *que es de 1.734, con lo cual tengo todos los datos para sustituirlos en la fórmula. El intervalo para la diferencia de medias es de: -.6937 <* $\mu_1 - \mu_2$ *< -.1663"* (Alumno FG).

Tabla 6.39. Análisis de la solución correcta al ítem 12

Unidad	Contenido	Funciones semióticas
U1	Como desconocemos las varianzas de las poblaciones y las muestras son pequeñas, para calcular el intervalo de confianza para la diferencia de medias poblacionales, tenemos que usar la distribución t. Las desviaciones estándar muestrales de ambos grupos son *s* = .422 y *s* = .073 respectivamente.	*Varianzas de las poblaciones, intervalo de confianza*: (el alumno nombra diferentes conceptos). *Diferencia* (describe una operación), *medias poblacionales, distribución t* (conceptos). *Tenemos que usar la distribución t:* El alumno discrimina la distribución muestral a emplear en caso de desviación típica conocida o desconocida (propiedad). *Desviaciones estándar muestrales*: (concepto).
U2	El intervalo de confianza viene dado por la siguiente fórmula: $\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \times (S_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, siendo $t_{\alpha/2}$ el valor crítico correspondiente al coeficiente de confianza 1-α, S_p la desviación estándar conjunta y n_1 y n_2 los tamaños de las muestras.	*Fórmula*: se refiere a la expresión: $\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \times (S_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ y también al procedimiento de cálculo del intervalo *(*procedimiento y lenguaje). $\bar{x}_1 - \bar{x}_2$ hace referencia a la diferencia de medias muestrales, que son estadísticos. El alumno tiene que diferenciar entre ellas y las medias de las poblaciones μ_1 *y* μ_2 que son parámetros (conceptos). $\pm$ hace referencia a que el intervalo se construye sumando y restando una misma cantidad (operación) $t_{\alpha/2} \times (S_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$: hace referencia a la cantidad a sumar, que es a su vez un producto (operación). $t_{\alpha/2}$ se refiere al valor crítico (concepto) en la distribución *t de Student (*concepto*); 1-α* se refiere al nivel de confianza (concepto), α/2 al área que queda a la derecha del valor de *t* (concepto) puesto que el nivel de confianza tiene que contemplar la probabilidad por encima y debajo de la diferencia de medias muestrales (propiedad).
U3	El valor de las medias muestrales son 9.97 y 10.4 respectivamente, los tamaños de las muestras son	*Medias muestrales, tamaños de las muestras, distribución t, valor crítico de t* (conceptos), *Tablas* (lenguaje).

	10 para cada una, por lo que buscando en tablas de la distribución *t de student* en $\alpha/2$= *0.025* obtengo el valor crítico de $t_{\alpha/2}$ que es de 1.734, con lo cual tengo todos los datos para sustituirlos en la fórmula. El intervalo para la diferencia de medias es de : -.6937 < μ_1 - μ_2 < -.1663	*Por lo que buscando en tablas de la distribución t de Student en* $\alpha/2$*= 0.025 obtengo el valor crítico de* $t_{\alpha/2}$ *que es de 1.734*: (procedimiento y propiedad). *Con lo cual tengo todos los datos para sustituirlos en la fórmula*: (procedimiento). -.6937 < μ_1 - μ_2 < -.1663 (lenguaje).

Adicionalmente a esta solución correcta, a la que asignaremos código C10, hemos encontrado las siguientes formas diferentes de respuestas:

C11: El alumno escribe incorrectamente los símbolos $x_1 - x_2$ en lugar de escribir los símbolos de las medias muestrales correspondientes en la fórmula del intervalo de confianza para la diferencia de medias poblacionales, pero el resto del proceso lo ejecuta correctamente. A continuación presentamos un caso que luego es analizado en la tabla 6.40.

"x_1 =*9.97,* n_1 = *10,* s_1 =*.422,* x_2 =*10.4,* n_2 =*10,* s_2 =*.073,* v= n_1 + n_2 -*2=18*

$$x_1 - x_2 - t * s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < x_1 - x_2 + t * s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(9.97-10.4)-1.734*.34\sqrt{\frac{1}{10}+\frac{1}{10}} < \mu_1 - \mu_2 < (9.97-10.4)+1.734*.34\sqrt{\frac{1}{10}+\frac{1}{10}}$$

-*.6937 <* μ_1 - μ_2 *< -.1663"* (Alumno MS).

Tabla 6.40. Análisis a otra solución correcta al ítem 12

Unidad	Contenido	Funciones semióticas
U1	x_1=9.97, n_1= 10, s_1=.422, x_2=10.4, n_2 =*10,* s_2 =*.073,* v= n_1 + n_2 -*2=18* $\ldots < \mu_1 - \mu_2 < x_1 - x_2 + t * s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	Escribe los datos del problema (tamaño de la muestra, desviación estándar de la muestra, grados de libertad) haciendo referencia a todos estos conceptos. Hace error al escribir el símbolo de las medias muestrales. Escribe incorrectamente la fórmula del intervalo de confianza (procedimiento) para la diferencia de medias poblacionales, fallando en los símbolos de las medias muestrales, pero ello no condiciona la solución correcta del problema.
U2	$< \mu_1 - \mu_2 < (9.9-10.4)+1.73*.34\sqrt{\frac{1}{10}+\frac{1}{10}}$	Calcula los valores críticos en la tabla *t* (procedimiento) y uso de las tablas (lenguaje) correctamente.
U3	-.6937 < μ_1 - μ_2 < -.1663	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), para llegar finalmente al intervalo pedido. -.6937 < μ_1 - μ_2 < -.1663 (lenguaje).

Respuestas parcialmente correctas:

C21: El alumno escribe correctamente la fórmula del intervalo de confianza para la diferencia de medias poblacionales σ_1, σ_2 desconocidas pero iguales y muestras pequeñas. Durante su proceso de solución se le presenta un conflicto con el cálculo de los grados de libertad de la distribución *t*, para el caso de dos muestras que fueron tomadas de poblaciones normales. El alumno obtiene los grados de libertad sumando solamente los tamaños de muestra n_1 y n_2 sin restarle 2 como en el caso siguiente que se analiza en la tabla 6.41.

"$\bar{x}_1 = 9.97, n_1 = 10,\ s_1 = .422,\ \bar{x}_2 = 10.4, n_2 = 10,\ s_2 = .073$

$\overline{x_1} - \overline{x_2} \pm t_{\alpha/2} \times (S_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, $(-.43) \pm 1.812*.34\ \sqrt{\frac{1}{10} + \frac{1}{10}}$

$-.706 < \mu_1 - \mu_2 < -.154$" (Alumno YW).

Tabla 6.41. Análisis de la solución parcialmente correcta al ítem 12

Unidad	Contenido	Funciones semióticas
U1	$\bar{x}_1 = 9.97, n_1 = 10,\ s_1 = .422,$ $\bar{x}_2 = 10.4, n_2 = 10, s_2 = .073$	Escribe los datos del problema haciendo referencia al tamaño de la muestra, media muestral, desviación estándar de la muestra (conceptos).
U2	$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \times (S_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	Escribe correctamente la fórmula del intervalo de confianza para la diferencia de medias poblacionales (procedimiento).
U3	$(-.43) \pm 1.812*.34$ $*\sqrt{\frac{1}{10} + \frac{1}{10}}$	Calcula los valores críticos en la tabla *t* (procedimiento) y uso de las tablas (lenguaje) correctamente, pero al ser los grados de libertad incorrectos el valor obtenido es incorrecto.
U4	$-.706 < \mu_1 - \mu_2 < -.154$	El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), para llegar finalmente al intervalo pedido. Pero, por la razón señalada el resultado es incorrecto. $-.706 < \mu_1 - \mu_2 < -.154$ (lenguaje).

Respuestas incorrectas:

C31: El estudiante olvida la fórmula del intervalo de confianza para la diferencia de medias poblacionales con $\sigma_1^2\ y\ \sigma_2^2$ desconocidas pero iguales y muestras pequeñas. El principal conflicto procedimental que se presenta es el uso de la distribución de muestreo correcta. El estudiante utiliza la distribución normal estándar en lugar de la distribución *t*. Se analiza en la tabla 6.42 un caso que muestra el error descrito.

"$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$, 9.97- 10.4$\pm$ 1.96$\sqrt{\frac{1}{10}+\frac{1}{10}}$, -1.3 < $\mu_1 - \mu_2$ < .44"

(Alumno AW).

Tabla 6.42. Análisis de la solución incorrecta al ítem 12

Unidad	Contenido	Funciones semióticas
U1	$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	El alumno escribe en forma incorrecta la fórmula del intervalo de confianza para la diferencia de medias poblacionales con σ_1^2 *y* σ_2^2 desconocidas pero iguales y muestras pequeñas. Escribe una fórmula parecida a la que se utiliza para obtener el intervalo de confianza para la diferencia de medias poblacionales, conocidas σ_1^2 *y*, se presenta un conflicto, ya que las posiciones que debieran ocupar las varianzas poblacionales son sustituidas por el valor de 1.
U2	9.97-10.4$\pm$ 1.96$\sqrt{\frac{1}{10}+\frac{1}{10}}$	Calcula los valores críticos en la curva normal estándar (procedimiento) y uso de las tablas (lenguaje) correctamente, pero usa una distribución muestral incorrecta.
U3	-1.3 < $\mu_1 - \mu_2$ < .44	El alumno lleva a cabo varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), para llegar finalmente a un intervalo. Pero, por las razones señaladas el resultado es incorrecto -1.3 < $\mu_1 - \mu_2$ < .44 (lenguaje).

Tabla 6.43. Frecuencias (y porcentajes) de respuestas en el ítem 12 (n=252)

	Frecuencia	Porcentaje	Porcentaje acumulado
Respuestas correctas C10	93	36.9	36.9
Respuestas correctas C11	5	2.0	38.9
Respuestas parcialmente correctas C21	105	41.7	80.6
Respuestas incorrectas C31	33	13.1	91.7
Blanco	16	6.3	100.0
Total	252	100.0	

En la Tabla 6.43 presentamos los resultados del ítem 12 (intervalos de confianza para diferencia de medias poblacionales, $\sigma_1 = \sigma_2$ pero desconocidas, muestras independientes pequeñas). Este ítem resultó sencillo para los alumnos, puesto que contestó correctamente un 80.6% de los estudiantes (considerando tanto parcial como totalmente correctas).

Un 41.7% de los estudiantes que respondieron en forma parcialmente correcta. Su conflicto tiene que ver con el procedimiento de cálculo, obtienen incorrectamente los grados de libertad para la distribución *t* para el caso de dos muestras. Hacen error al sumar solamente los tamaños de muestra n_1 y n_2 sin restarle 2. Un 13.1% de los estudiantes tienen un conflicto relativo a los procedimientos; el intervalo de confianza que escriben es el correspondiente a la diferencia de medias poblacionales con σ_1 y σ_2 conocidas. Llama también la atención el 6.3% de respuestas en blanco, a pesar de ser un ítem sencillo.

6.3.7. ANÁLISIS DEL ITEM 17. INTEPRETACIÓN DE RESULTADOS DE ORDENADOR

Ítem 17. La siguiente salida de computadora presenta dos muestras simuladas de dos poblaciones normales. La población 1 con μ=*90* y σ =10 y la población 2 con μ =*92* y σ =10:

Muestra 1:
83.3195 87.6793 86.7831 95.0518 92.9781 86.6457 85.1305 97.5013 83.1112 82.2751 82.7831 90.2786 89.5876 71.2591 82.0282 90.6264

Muestra 2:
82.312 95.098 92.598 85.959 91.319 108.130 90.392 90.074 78.789 100.923 85.601 89.861 78.685 100.354 81.267 101.432

	N	Media	D. Típica	Error típico
C1	16	86.69	6.25	1.6
C2	16	90.80	8.70	2.2

95% IC para mC1 - mC2: (-9.6; 1.4)

Escriba el intervalo de confianza que se obtuvo para la diferencia de medias e interprete el resultado.

La respuesta correcta esperada del estudiante incluye: a) escribir la ecuación correcta del intervalo de confianza para la diferencia de medias a partir de la salida del ordenador y b) la interpretación correcta del intervalo que puede ser dada en términos de un tanto por ciento de muestreos repetidos en las mismas condiciones cuyo intervalo calculado cubre al parámetro diferencia de medias poblacionales o bien a través de la comparación de las medias en dos poblaciones de la forma, a partir de observar el comportamiento del intervalo de confianza.

Si el intervalo va de valores negativos a valores positivos, es decir pasa por el cero, el alumno deberá concluir que las medias de las poblaciones pueden ser iguales. Si el intervalo está formado solamente por valores positivos entonces deberá concluir que la media poblacional 1 es superior a la media poblacional 2. Si el intervalo está formado solamente por valores negativos, entonces deberá concluir que la media poblacional 2 es superior a la media poblacional 1. Un ejemplo se reproduce a continuación y se analiza en la Tabla 6.44. Esperamos el siguiente razonamiento de los estudiantes:

"La última línea de la salida del ordenador en que se lee: 95% IC para mC1-mC2:(-9.6; 1.4) nos indica que los extremos del intervalo de confianza para la diferencia de medias poblacionales son -9.6 y 1.4 a un nivel de confianza de 95%. Simbólicamente podemos escribirlo: -9.6< $\mu_1 - \mu_2$ < 1.4. Interpretando que en el 95% de muestreos ejecutados en estas mismas condiciones estará contenido el verdadero valor de la diferencia de medias para estas poblaciones. O bien que no es posible suponer que uno de los grupos tenga una media poblacional superior al otro, debido a que el intervalo incluye valores tanto positivos como negativos" (Alumno RV).

Tabla 6.44. Análisis de la solución correcta al ítem 17

Unidad	Contenido	Funciones semióticas
U1	La última línea de la salida del ordenador en que se lee: 95% IC para mC1-mC2*:(-9.6; 1.4)* Indica que los extremos del intervalo de confianza para la diferencia de medias poblacionales son -9.6 y 1.4 a un nivel de confianza de 95%.	*La última línea de la salida del ordenador en que se lee:* 95% *IC para mC1-mC2*: (-9.6; 1.4): concepto y lenguaje. *Extremos del intervalo de confianza, medias poblacionales, nivel de confianza:* el alumno nombra distintos conceptos. *Diferencia:* describe una operación.
U2	Simbólicamente podemos escribirlo*:* -9.6< $\mu_1 - \mu_2$ < 1.4 Interpretando que en el 95% de muestreos ejecutados en estas mismas condiciones estará contenido el verdadero valor de la diferencia de medias para estas poblaciones.	*Simbólicamente podemos escribirlo* -9.6< $\mu_1 - \mu_2$ < 1.4: (lenguaje). *En el 95% de muestreos ejecutados en estas mismas condiciones estará contenido el verdadero valor de la diferencia de medias para estas poblaciones*: razonamiento deductivo para explicar el significado del nivel de confianza.
U3	O bien que no es posible suponer que uno de los grupos tenga una media poblacional superior al otro, debido a que el intervalo incluye valores tanto positivos como negativos.	*Media poblacional*, *valores positivos*, *valores negativos*: el estudiante evoca diferentes conceptos. *No es posible suponer que uno de los grupos tenga una media superior al otro, debido a que el intervalo incluye valores tanto positivos como negativos*: razonamiento deductivo.

A esta solución correcta le daremos código C10. Otras variantes de respuesta que hemos encontrado son:

C11: El alumno escribe correctamente el intervalo de confianza para la diferencia de medias y su interpretación, aunque no es completa, se toma como aceptable, como se puede apreciar en el siguiente caso que analizamos en la tabla 6.45.

" -9.6< $\mu_1 - \mu_2$ < 1.4 Hay evidencia de que las medias puedan ser iguales" (Alumno ES).

Tabla 6.45. Análisis de la solución correcta al ítem 17

Unidad	Contenido	Funciones semióticas
U1	-9.6< $\mu_1 - \mu_2$ < 1.4	μ_1, μ_2 : son las medias de las poblaciones 1 y 2 respectivamente El alumno escribe correctamente el intervalo de confianza para la diferencia de medias. -9.6< $\mu_1 - \mu_2$ < 1.4 (lenguaje).
U2	Hay evidencia de que las medias pueden ser iguales	*Hay evidencia de que las medias pueden ser iguales:* razonamiento deductivo.

Respuestas parcialmente correctas

C21: El estudiante escribe correctamente el intervalo de confianza para la diferencia de medias, pero da una interpretación incorrecta. Su respuesta, aunque breve, nos permite suponer que tiene un conflicto conceptual relativo a la distinción entre media y una medida de variabilidad. En el caso siguiente se puede observar este conflicto y su análisis se presenta en la tabla 6.46.

"*-9.6< $\mu_1 - \mu_2$ < 1.4 C1 es menos variable*" (Alumno JLO).

Tabla 6.46. Análisis de la solución parcialmente correcta al ítem 17

Unidad	Contenido	Funciones semióticas
U1	-9.6< $\mu_1 - \mu_2$ < 1.4	μ_1, μ_2 : son las medias de las poblaciones 1 y 2 respectivamente El alumno escribe correctamente el intervalo de confianza para la diferencia de medias. 9.6< $\mu_1 - \mu_2$ < 1.4 (lenguaje).
U2	C1 es menos variable	*C1 es menos variable*: razonamiento que pone de manifiesto sus conflicto conceptual relacionado con la distinción entre media y una medida de dispersión.

Respuestas incorrectas:

C31: El alumno no escribe el intervalo de confianza y además se presenta un conflicto relativo al procedimiento. El estudiante evoca algunos conceptos, pero su interpretación se llena de errores por el uso de una terminología inexacta que no llega a especificar los parámetros que son estimados por el intervalo en cuestión, como se muestra en el siguiente caso cuyo análisis se presenta en la tabla 6.47.

"Existe un 95% de confianza de que la salida de computadoras de 2 poblaciones varíe entre -9.6 y 1.4. Tomando en cuenta que la C1 es más confiable" (Alumno NM).

Tabla 6.47. Análisis de la solución incorrecta al ítem 17

Unidad	Contenido	Funciones semióticas
U1	Existe un 95% de confianza de que la salida de computadoras de 2 poblaciones varíe entre -9.6 y 1.4.	El alumno no especifica que parámetros reporta la salida de computadora. *Existe un 95% de confianza de que la salida de computadoras de 2 poblaciones varíe entre -9.6 y 1.4:* razonamiento con descripción incompleta e imprecisa.
U2	Tomando en cuenta que la C1 es más confiable.	El alumno hace error en su conclusión manejando términos inadecuados. *Tomando en cuenta que la C1 es más confiable:* razonamiento deductivo incorrecto.

C32: El alumno escribe el intervalo, copiando la misma salida de computadora y además tiene un conflicto procedimental al dar una interpretación con errores por el uso de términos incorrectamente articulados, que no describen ni la operación ni los parámetros que intervienen en el intervalo de confianza. En el siguiente caso se aprecia este conflicto que luego se analiza en la tabla 6.48.

> *"(-9.6; 1.4), (mc1; mc2). El intervalo de confianza obtenido con un 95 % de confianza al tener a muestras simuladas, salidas de una compu en poblaciones normales es entre -9.6 y 1.4"* (Alumno JMC).

Tabla 6.48. Análisis de la solución incorrecta al ítem 17

Unidad	Contenido	Funciones semióticas
U1	(-9.6; 1.4) (mc1; mc2)	El alumno no escribe el intervalo en la forma esperada. Hace error al escribir en los límites de su intervalo los símbolos de lo que son para él las medias poblacionales 1 y 2.
U2	El intervalo de confianza obtenido con un 95 % de confianza al tener a muestras simuladas, salidas de una compu en poblaciones normales es entre -9.6 y 1.4.	En su interpretación el alumno en ningún momento señala que el intervalo es para una diferencia de medias poblacionales. *Intervalo de confianza, 95 % de confianza, muestras simuladas, poblaciones normales:* el alumno evoca distintos conceptos. *El intervalo de confianza obtenido con un 95 % de confianza al tener a muestras simuladas, salidas de una compu en poblaciones normales es entre -9.6 y 1.4:* razonamiento con una interpretación incompleta.

C33: Otros. En este código están las respuestas de dos alumnos que escribieron un intervalo de confianza para una diferencia de medias muestrales, en lugar de medias poblacionales y que luego no dan alguna interpretación. En estas respuestas surge nuevamente la confusión entre estadísticos y parámetros.

Tabla 6.49. Frecuencias (y porcentajes) de respuestas en el ítem 17 (n=252)

	Frecuencia	Porcentaje	Porcentaje acumulado
Respuestas correctas C10	39	15.5	15.5
Respuestas parcialmente correctas C21	61	24.2	39.7
Respuestas incorrectas C31	101	40.1	79.8
Respuestas incorrectas C32	34	13.5	92.3
Respuestas incorrectas C33	2	.8	94.1
Blanco	15	6.0	100.0
Total	252	100.0	

En la Tabla 6.49 presentamos los resultados del ítem 17 (interpretación de intervalo de confianza para diferencia de medias obtenida a partir del ordenador). Observamos que este ítem no resultó sencillo para los alumnos, puesto que, un 54.4% de los alumnos que interpretaron en forma incorrecta el intervalo, un 40.1% no escriben el intervalo de confianza y además dan una interpretación incorrecta con el uso de términos completamente ajenos al lenguaje requerido, otro porcentaje (13.5%) escribe el intervalo copiando la misma salida de computadora y dan una interpretación incorrecta con el uso de una terminología que no tiene relación alguna con los parámetros en cuestión.

De los que respondieron en forma parcialmente correcta, 24.2% escriben correctamente el intervalo de confianza para la diferencia de medias, pero dan una interpretación incorrecta cuando señalan que la población 1 es la menos variable. Además observamos un número alto de alumnos con respuestas en blanco para este ítem. Una posible explicación es que algunos grupos de alumnos que participaron en la muestra no recibieron instrucciones de cómo hacer una interpretación a partir de una salida de ordenador, que adicionada a la dificultad, per se, de interpretar el intervalo de confianza para diferencia de medias, da por resultado un alto porcentaje de respuestas incorrectas.

6.4. ESTUDIO DE LA PUNTUACIÓN GLOBAL

Para analizar el número de respuestas correctas que cada estudiante ha obtenido, les hemos asignado el valor de 1, por lo que al sumarlas podríamos encontrar una variación entre 0 y 22 respuestas correctas. Tal como se observa en la Tabla 6.50, en donde podemos apreciar que el número de respuestas correctas osciló entre 3 y 21 para el cuestionario A (se alcanzó el máximo teórico y hubo un alumno con solo tres respuestas correctas).

Para el cuestionario B la puntuación osciló entre 2 y 22 (se alcanzó el máximo teórico para este cuestionario y hubo un alumno con solo dos respuestas correctas). El número medio de respuestas correctas de los estudiantes es de 12.43 sobre 21 para el cuestionario A y de 12.77 sobre 22 para el cuestionario B, lo que nos da resultados aceptables, que se acercan a lo esperado. La mediana de 13 para el cuestionario A y de 14 para el cuestionario B, indican que más de la mitad de los estudiantes responden más de la mitad de las preguntas. La variabilidad en el número de respuestas en los dos cuestionarios es muy parecida. Por todo ello podemos considerar equivalentes los dos cuestionarios.

Tabla 6.50. Estadísticos Descriptivos

cuestionario			Estadístico	Error típ.
A	Media		12.43	.363
	Intervalo de confianza para la media al 95%	Límite inferior	11.71	
		Límite superior	13.15	
	Media recortada al 5%		12.42	
	Mediana		13.00	
	Varianza		16.877	
	Desv. típ.		4.108	
	Mínimo		3	
	Máximo		21	
	Rango		18	
	Amplitud intercuartil		7	
	Asimetría		.020	.214
	Curtosis		-.580	.425
B	Media		12.77	.385
	Intervalo de confianza para la media al 95%	Límite inferior	12.00	
		Límite superior	13.53	
	Media recortada al 5%		12.79	
	Mediana		14.00	
	Varianza		18.376	
	Desv. típ.		4.287	
	Mínimo		2	
	Máximo		22	
	Rango		20	
	Amplitud intercuartil		5	
	Asimetría		-.122	.217
	Curtosis		-.322	.431

Los coeficientes de asimetría y curtosis (Tabla 6.50) son aceptables para asumir la normalidad de las puntuaciones, lo que también se observa en la Tabla 6.51, que permite aceptar esta normalidad.

Tabla 6.51. Prueba de Kolmogorov-Smirnov para una muestra (n=252)

Parámetros normales(a,b)	Media	12.60
	Desviación típica	4.192
Diferencias más extremas	Absoluta	.084
	Positiva	.066
	Negativa	-.084
Z de Kolmogorov-Smirnov		1.327
Sig. asintót. (bilateral)		.059

Figura 6.1. Histograma de la puntuación total

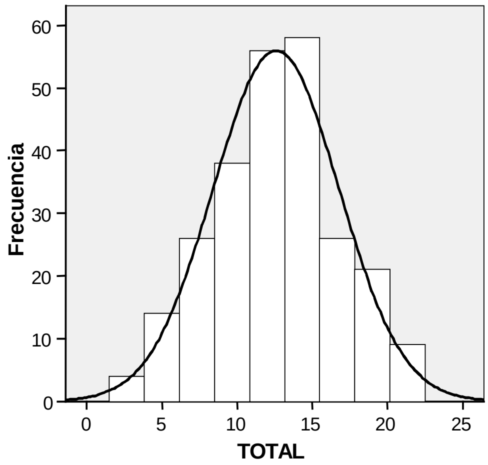
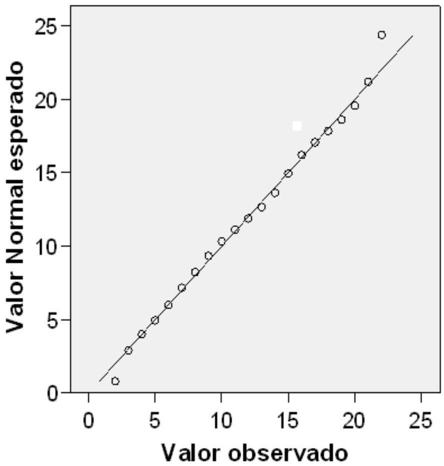


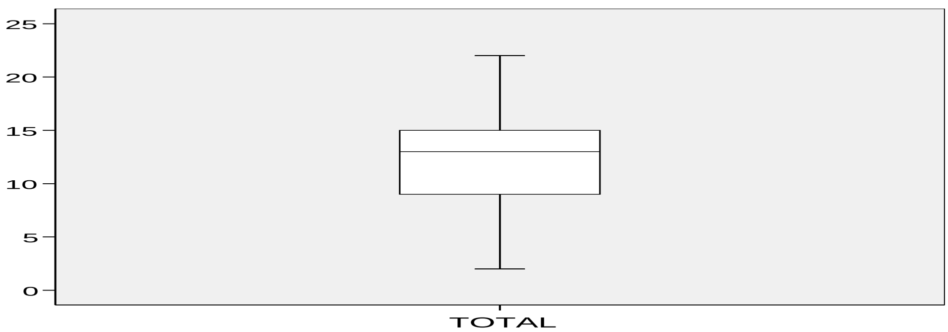
Figura 6.2. Gráfico Q-Q Normal de Total



Otras evidencias se presentan con la distribución en forma de campana de las puntuaciones (Figura 6.1) y la normalidad aproximada de la distribución de la puntuación total (Figura 6.2).

En la figura 6.3 observamos los cuartiles superior e inferior de 9 y 15, que indican que el conjunto central (50% central) responde entre 9 y 15 preguntas correctamente y el máximo se ubica en 22 preguntas contestadas correctamente.

Figura 6.3. Gráfico de la caja para la puntuación total



6.5. CONCLUSIONES DEL ESTUDIO DE EVALUACIÓN

En este capítulo en que hemos realizado un estudio de evaluación de la comprensión que demostraron los estudiantes de ingeniería de los intervalos de confianza, destacamos que dos terceras partes de los ítems del cuestionario fueron resueltas por más de la mitad de los alumnos. Resultados que apuntan a que la mayoría de los estudiantes parecen comprender este concepto, aunque la comprensión es sólo aparente pues se han presentado errores procedimentales y numerosos conflictos semióticos, algunos de los cuales no han sido descritos en investigaciones previas relacionadas con el tema.

Un primer conflicto es que los estudiantes no comprenden cabalmente que los intervalos de confianza representan el porcentaje de intervalos de muestras, tomadas todas bajo las mismas condiciones, dentro de los cuales estará contenido el verdadero valor del parámetro, es decir, dan una interpretación bayesiana, suponiendo los extremos del intervalo fijos y no aleatorios. Mientras en uno de los ítems el porcentaje de los que dan respuestas correctas sobre este punto es de 55.6% (ítem 1), en otro es solamente el 36.5% (ítem 5). Aunque este tipo de interpretación bayesiana se ha descrito con relación al contraste de hipótesis, consideramos que nuestro trabajo aporta la detección de este conflicto para el caso del intervalo de confianza.

En la comparación de dos medias poblacionales, solo un 29% de los estudiantes comprendieron que los intervalos de confianza no representan el porcentaje de valores de las diferencias de medias poblacionales entre los límites de confianza (ítem 13). Este resultado, ligado a los anteriores, reproduce lo descrito en las investigaciones de Behar (2001), quien indica que una buena proporción de los participantes en su investigación parecen no asociar la confianza a un mecanismo aleatorio generador de intervalos, a partir de muestras aleatorias, ni asociar una probabilidad, como el nivel de confianza, con la frecuencia relativa a la larga, de que los intervalos generados por tal mecanismo aleatorio atrapen al verdadero parámetro de la población. Los resultados nos avisan que esta propiedad tampoco es bien comprendida en nuestro estudio.

Un porcentaje importante de estudiantes (25.8%) tiene un conflicto al indicar en forma equivocada que el nivel de confianza representa el porcentaje de valores de la media de la muestra entre los límites de confianza (ítem 1; ítem 4). Es decir hay una confusión entre ejemplar (media de la muestra, media de la población) y tipo (media) que no se diferencian. Este tipo de confusión fue descrito entre otros por Schuyten, 1991 y Vallecillos, 1994).

Una tercera parte de los alumnos mostraron conflictos de comprensión de la manera como se relaciona el tamaño de la muestra con el ancho del intervalo (ítem 2). Resultado que nos revela un aspecto estructural de los significados personales de los estudiantes (Budé, 2006). Nuestro estudio corrobora los resultados obtenidos por Fidler y Cumming (2005) y Behar(2001).

Un poco más de la mitad de los alumnos tuvieron dificultades procedimentales en la construcción de intervalos de confianza para una proporción poblacional y para la varianza poblacional (ítems 9 y 10 respectivamente), bajo los cuáles subyacen conflictos de tipo conceptual. En nuestra búsqueda de investigaciones relacionadas encontramos un trabajo de Newcombe (1998) con relación a los intervalos de confianza para una proporción, pero no explora acerca de dificultades de comprensión en los estudiantes. Respecto al parámetro varianza tampoco hemos encontrado estudios relacionados, por lo cual nuestro resultado es una aportación a la investigación empírica de las dificultades en la comprensión de los intervalos de confianza.

Un resultado inesperado ocurrió con la interpretación de los intervalos de confianza, para diferencia de medias poblacionales, a partir de una salida de ordenador (ítem 17); 54.4% de los estudiantes tuvo dificultades en su interpretación. La comparación más cercana de este resultado es con el reportado por Belia, Fidler y Cumming (2005). En su estudio por internet, dirigido a autores de artículos publicados en revistas internacionales, un 34% interpretan equivocadamente la comparación en forma gráfica de medias de dos grupos independientes. Por tanto, al no tener una comparación directa y con trabajos relacionados con grupos de estudiantes, este resultado es también una aportación a la investigación del tema.

Finalmente podemos señalar que apenas la mitad de los estudiantes demostraron habilidad para seleccionar una distribución de muestreo apropiada y luego uso adecuado de las tablas en la construcción del intervalo de confianza (52.8% en el ítem 15 y 46.8% en el ítem 16). Sin embargo aparecen problemas en el cálculo de valores críticos en algunos casos, en que el 5% se coloca a ambos lados de la distribución, mostrando un conflicto en la lectura de tablas y cálculo de valores críticos. Otros conflictos de menor importancia han sido el olvido de los grados de libertad en las distribuciones T y Chi cuadrado o de algunos de los factores en las fórmulas de cálculo de los intervalos de confianza.

En resumen, nuestro estudio de evaluación confirma los resultados obtenidos en otros trabajos con relación a la comprensión de las distribuciones muestrales (Schuyten, 1991; Vallecillos y Batanero, 1997; delMas, Garfield, Ooms y Chance, 2007). Al mismo tiempo los amplia detectando nuevos tipos de conflictos como la interpretación bayesiana del significado del coeficiente de confianza, las dificultades con la obtención de valores críticos y de interpretación de salidas de ordenador. También ponemos de manifiesto, mediante el análisis semiótico realizado la complejidad de los problemas, aparentemente simples y la multitud de objetos matemáticos que el estudiante ha de poner en relación.

Con todo ello mostramos la idoneidad del instrumento de evaluación obtenido, que era uno de los fines de la investigación y sobre la que comentaremos con más detalle en el capítulo de conclusiones. Al mismo tiempo la identificación de conflictos y dificultades es el primer paso en el diseño de futuras propuestas didácticas orientadas a su superación.

CAPÍTULO 7
CONCLUSIONES

1.1. INTRODUCCIÓN

Como se indicó en la introducción, en este trabajo hemos analizado el significado personal que adquieren los alumnos sobre los intervalos de confianza, tema que es parte fundamental del material impartido en los cursos de probabilidad y estadística en todas las carreras universitarias. Para finalizar esta memoria, en este capítulo presentamos las principales conclusiones obtenidas de nuestro estudio, comenzando por las relacionadas con los objetivos y las hipótesis formuladas en el Capítulo 1 y estudiando la idoneidad del instrumento de evaluación construido.

Seguidamente, analizamos las aportaciones más relevantes del estudio, y realizamos una valoración global de sus alcances y limitaciones, respecto a las posibilidades de ser generalizado a otras muestras de textos y de estudiantes, dentro y fuera del Sistema Tecnológico de Monterrey o a otros cursos de estadística.

Describimos, finalmente, algunas cuestiones sobre las que otros investigadores podrían tomar como punto de partida y dar continuidad a esta línea de investigación iniciada. Incluimos también algunas recomendaciones, que se deducen de nuestras conclusiones, y que pueden contribuir a mejorar el proceso enseñanza- aprendizaje del tema.

7.2. CONCLUSIONES RESPECTO A LOS OBJETIVOS DE LA INVESTIGACIÓN

En el capítulo 1 señalamos que el objetivo central de nuestro trabajo era *generar información que contribuya a explicar el significado personal que los estudiantes de las carreras de ingeniería asignan a los intervalos de confianza.*

Creemos que nuestro trabajo cumple con este objetivo a pesar de tener un tamaño de muestra limitado y estar la muestra muy focalizada, en relación con la población general de estudiantes de ingeniería. La información recogida a través del cuestionario nos ha permitido confirmar los resultados de investigaciones anteriores sobre el tema, así como describir la existencia de algunas nuevas dificultades y conflictos semióticos no descritos en los trabajos previos, que afectan no sólo a la definición de los intervalos de confianza, sino a cada uno de los elementos que lo constituyen.

Utilizando nuestro marco teórico y la técnica de análisis semiótico, hemos analizado con detalle la complejidad de los problemas abiertos propuestos, mostrando la diversidad de elementos en el significado del intervalo de confianza que han de usar y relacionar los estudiantes para resolverlos. Esta variedad y complejidad de elementos explica en parte los conflictos semióticos que surgen en la comprensión y aplicación de este objeto matemático en la resolución de los problemas.

Destacamos los siguientes *conflictos semióticos* observados en los alumnos de las carreras de ingeniería, respecto a los diferentes elementos del significado del intervalo de confianza

a) Respecto a la definición, propiedades y relaciones con otros conceptos:

- Conflicto respecto a la definición del coeficiente de confianza (ítem 5), ya que sólo una tercera parte de los estudiantes de la muestra comprende que el coeficiente da el porcentaje de intervalos de muestras, tomadas todas bajo las mismas condiciones, dentro de los cuales estará contenido el verdadero valor del parámetro, confirmándose los resultados obtenidos por Behar (2001).
- Una tercera parte de los estudiantes, aunque obtiene correctamente el intervalo de confianza para diferencia de medias poblacionales, hacen una interpretación bayesiana del mismo suponiendo que el coeficiente de confianza es la probabilidad a posteriori de obtener el parámetro dentro del intervalo, una vez recogida la muestra (Ítem 13). Esta interpretación, que es la que intuitivamente da (erróneamente) la mayoría de los investigadores, también es dada para intervalos de confianza para una media poblacional (36.5%, ítem 5). La detección de este conflicto, en estudiantes es una de las aportaciones de nuestro trabajo.
- Algunos estudiantes confunden estadístico y parámetro, pensando que el intervalo se construye para estimar la media muestral (Ítem 1). Esta dificultad fue reportada

por Vallecillos y Batanero (1997) en su estudio con relación a la comprensión de los contrastes de hipótesis estadísticas, así como en los trabajos de Behar (2001) y delMas, Garfield, Ooms y Chance (2007). Nuestra explicación es que en inferencia el concepto media lo usamos como ejemplar (media de la muestra, media de la población, media de la distribución muestral) y tipo (media, en general). Los estudiantes no diferencian estos dos planos en la faceta ejemplar-tipo, por lo que confunden los diferentes ejemplares de media que usamos en inferencia.

- Los estudiantes no relacionan los distintos factores asociados con el intervalo de confianza, en especial el ancho del intervalo y el nivel de confianza (Ítem 2, Ítem 3), resultados que confirma los resultados de Fidler y Cumming (2005) y Behar (2001). Pensamos que estos errores pudieran estar relacionadas con el hecho de que muchos estudiantes visualizan los intervalos de confianza como estadísticos descriptivos, ignorando su naturaleza inferencial (Cumming y Fidler, 2005).

b) Respecto a los campos de problemas y procedimientos, se han identificado los siguientes conflictos:

- Confusión de las distribuciones de muestreo apropiadas para construir un intervalo de confianza (ítem 15), que influirá en la confusión entre campos de problemas y la correspondiente aplicación de procedimientos incorrectos de cálculo de intervalos.
- Los alumnos muestran conflictos en el manejo de los parámetros de las distribuciones muestrales en el cálculo de intervalos. El 18.7% de los estudiantes obtienen incorrectamente el intervalo de confianza, para la comparación de dos varianzas poblacionales, porque intercambia los grados de libertad de numerador con los del denominador en la obtención de los valores críticos cuando usan la distribución F (ítem 14).
- No hay suficiente competencia en la determinación del valor crítico a partir de una tabla de la distribución, para usarlo en la construcción del intervalo de confianza (ítem 16). Por ejemplo, los alumnos toman un área de cola doble de la debida, lo que resulta en un intervalo de menor confianza que la requerida. Nuestros resultados en este sentido coinciden con los señalamientos de Schuyten (1991).
- Otros conflictos se relacionan con el manejo de inecuaciones para calcular el intervalo de confianza para la media (4.4%, ítem7) y al calcular el intervalo de

confianza para la varianza, cuando cambian el orden de los extremos, escribiendo la fracción de mayor valor en el lado izquierdo (6%, ítem 10).

c) Respecto al lenguaje y argumentación:

- Se presentan conflictos en la interpretación de los intervalos de confianza para diferencia de medias poblacionales, a partir de una salida del ordenador (ítem 17; 50% de estudiantes). Esto indica la necesidad de instruir a todos los alumnos sobre cómo interpretar una salida del ordenador, aunque en la muestra piloto (n=48) solamente un 27% de los estudiantes presentó esta dificultad.
- Aunque no es muy generalizado, hay estudiantes que no interpretan correctamente los intervalos de confianza a partir de un gráfico (Ítem 18).
- Los estudiantes confunden varianza poblacional con desviación típica poblacional. Una sexta parte de los estudiantes en el ítem 10 (calcular un intervalo de confianza para la desviación estándar poblacional) escriben desviación estándar en vez de varianza. Confunden desviación típica muestral con desviación típica poblacional, como por ejemplo en el proceso de obtención del intervalo de confianza para la media (11.5%, ítem7) usan la distribución normal para obtener los valores críticos, en vez de la distribución *T*. Un 41.3% de los estudiantes no discrimina los casos de desviación estándar poblacional conocida y desconocida al determinar un valor crítico haciéndolo en todos los casos a partir de la distribución normal en vez de la distribución *T* (ítem 16).
- Con relación al lenguaje simbólico, un 16.7% de los estudiantes escribe el símbolo de desviación estándar poblacional en lugar del símbolo de la varianza poblacional cuando calcula el intervalo de confianza para la varianza (ítem 10). Otras dificultades, consisten en escribir el símbolo de la media muestral en vez del símbolo de la media poblacional (ítem 7), no escribir α/2 como subíndice de Z (C11, ítem 9), olvidar escribir los subíndices para chi- cuadrado en el intervalo de confianza para la varianza (C22, ítem 10), en el intervalo de confianza para diferencia de dos medias poblacionales (ítem12) escribe $Z_{\alpha/2}$ en vez de $t_{\alpha/2}$ y escribe un intervalo de confianza para una diferencia de medias muestrales, en lugar de medias poblacionales (C33, ítem 17).

Además de este objetivo central nos planteamos otros objetivos específicos, siguiendo las mismas líneas de trabajo de Alvarado (2004, 2007) sobre el Teorema Central del Límite y de Díaz (2004, 2007) sobre construcción de cuestionarios (razonamiento condicional).

Objetivo 1. Realizar un análisis del intervalo de confianza desde el punto de vista de su evolución histórica, identificando información que nos proporcione pistas para la reconstrucción de sus diferentes significados a lo largo de la historia.

Este objetivo se ha alcanzado razonablemente al presentar el análisis del desarrollo histórico en el Capítulo 2. En el estudio de dicha evolución, hemos comenzando por las primitivas ideas de estimación y la insuficiencia de la estimación puntual, que lleva a la idea de dar intervalos para acotar la variación del parámetro en una población. Este primitivo campo de problemas se diversifica posteriormente en una gran, multiplicidad, dependiendo del parámetro a estimar y sus diversas condiciones.

Por otro lado, no se produce una única solución al problema, sino se crean dos escuelas de inferencia bayesiana y clásica que no sólo proporcionan diferentes procedimientos sino dan diverso significado al mismo problema de estimación, el concepto de parámetro y la solución encontrada (intervalo de credibilidad o de confianza) e incluso a la misma idea de probabilidad. Mientras que los intervalos de confianza son conceptualizados como proporción de repeticiones de este intervalo que incluye el verdadero valor del parámetro en el largo plazo los intervalos de credibilidad contemplan la proporción de valores posibles del parámetro que son capturados por este intervalo. Otra alternativa la proporciona la teoría fiducial de Fisher, que permite calcular valores de verosimilitud sobre los parámetros, pero no resulta exitosa, porque no proporciona una distribución acerca de los parámetros desconocidos. A pesar de ello verosimilitud, y máxima verosimilitud son ideas importantes en la teoría de estimación, donde la controversia sobre la mejor metodología continúa en la estadística, su enseñanza y aplicaciones.

Otras conclusiones son que el procedimiento bayesiano representó el primer aporte a la solución al problema de la estimación (200 años anterior a los intervalos de confianza), pero fue abandonado en su día por la necesidad de conocimiento de la probabilidad a priori de las causas. Hoy día esta dificultad está superada, además de ser rescatable bajo ciertas circunstancias (distribuciones uniformes o no informativas). En realidad los intervalos de confianza emergen ante la resistencia de muchos investigadores a aceptar las probabilidades subjetivas propugnadas por el procedimiento bayesiano.

Estos resultados nos proporcionan luz sobre uno de los conflictos detectados en los estudiantes en nuestra investigación y que consiste en dar una interpretación bayesiana al intervalo y coeficiente de confianza, interpretación que se presenta en un 36.5% (ítem5) de la muestra. Posiblemente y como indica Lecoutre (1999) la inferencia bayesiana (por medio de intervalos de credibilidad) podría ser más intuitiva a los estudiantes. El análisis del aprendizaje real, iniciado por Díaz (2007) en cursos de psicología parece apoyar esta hipótesis y podría ser un tema para continuar la investigación futura.

Objetivo 2: Realizar una síntesis de las investigaciones previas relacionadas con la comprensión de los intervalos de confianza y seleccionar ítems que puedan usarse en la construcción de nuestro cuestionario.

La síntesis que hemos realizado del estado de la cuestión y que se presenta en el capítulo 2 nos sirvió para identificar las principales dificultades respecto a intervalos de confianza en estudiantes e investigadores, así como errores importantes reportados en los estudios previos e investigaciones relacionadas. Con todo ello, pudimos fundamentar el análisis de nuestros resultados. Además de familiarizarnos con el campo, esta síntesis nos permitió hacer una primera selección de los ítems que conformarían el banco inicial que utilizamos en la construcción del instrumento de medición. Pensamos que también proporciona un punto de partida a investigadores que se interesen por el tema.

Objetivo 3: Realizar un análisis de los elementos del significado de los intervalos de confianza que nos fundamente la definición semántica de la variable "comprensión de los intervalos de confianza" y delimitar las principales áreas de contenido.

Este análisis y clasificación de los elementos de significado de los intervalos de confianza realizados en una muestra de libros de texto dirigidos a ingenieros, lo hemos presentado en el capítulo tercero. En este estudio hemos encontrado una gran variedad de campos de problemas, diferenciados en función de los parámetros a estimar o de las condiciones de partida. Muchos de estos subcampos diferenciados comparten objetos similares, que podrían agruparse en configuraciones epistémicas relativas al intervalo de confianza. También hemos encontrado una gran variedad de lenguaje (verbal, simbólico y gráfico), de argumentos y propiedades asociadas al intervalo u objetos relacionados como las distribuciones de muestreo. De este análisis concluimos que la enseñanza del tópico debiera ser organizada de tal manera que el estudiante aprenda a reconocer los diferentes campos de problemas y relacione cada uno de ellos a las distribuciones de muestreo y procedimientos computacionales apropiados, así como también adquiera la lógica general subyacente detrás de la construcción e interpretación de todos esos intervalos.

Este análisis de los elementos de significado de los intervalos de confianza, que se apoya en el enfoque ontosemiótico de la cognición matemática (EOS) propuesto por Godino y colaboradores, junto con el análisis en el plano histórico de los intervalos de confianza, nos ha permitido fundamentar la definición semántica de la variable "comprensión de los intervalos de confianza", mediante una definición precisa y sistemática de las unidades de contenido que serían luego objeto de evaluación. Puede ser también la base de diseño de futuras experiencias de enseñanza.

Objetivo 4: Construir un instrumento piloto que permita evaluar la comprensión de los intervalos de confianza, siguiendo unos criterios metodológicos rigurosos en la selección de los ítems que lo compondrían.

En el capítulo cuarto describimos paso a paso el proceso que seguimos en la construcción del instrumento piloto de medición. Este proceso sistemático, que se inició con la definición de la variable objeto de medición, termina con la construcción de un cuestionario validado que demostró tener un contenido relevante, con ítems que tienen una dificultad e índices de discriminación aceptables, y además son representativos de un universo de observaciones admisibles ligadas a la comprensión de los intervalos de confianza. El conjunto de respuestas a los ítems dio cuenta de la validez de constructo a través del análisis factorial. Los índices obtenidos de fiabilidad y generalizabilidad, como medidas objetivas de la estabilidad de las puntuaciones, nos permitieron concluir que nuestro cuestionario era moderadamente fiable en cuanto a generalizabilidad a otros ítems, pero altamente generalizable a otros alumnos semejantes a los que participan en el estudio.

Objetivo 5: Realizar un estudio de evaluación con una amplia muestra de estudiantes de ingeniería, para analizar tanto las tendencias como la variabilidad del significado personal de los estudiantes

En el capítulo seis, en nuestro estudio de evaluación, hemos presentado información sobre los errores y dificultades específicos de una muestra amplia de estudiantes de ingeniería. Entre los resultados obtenidos señalamos que más de la mitad de los alumnos han respondido correcta o casi correctamente los dos tercios de preguntas planteadas Sin embargo, al analizar los distractores elegidos en los ítems de opción múltiple o hacer un análisis semiótico de las respuestas abiertas vemos que en muchos de los elementos de significado asociados al intervalo de confianza aún prevalecen dificultades y conflictos que podrían superarse con mayor tiempo dedicado al tema y la implementación de futuras propuestas didácticas diseñadas a partir de este estudio.

Nuestro estudio pone de manifiesto las tendencias generales de la comprensión (dadas por las respuestas mayoritarias en cada uno de los ítems); el análisis semiótico detallado y las categorías encontradas de respuestas correctas, parcialmente correctas e incorrectas en los ítems abiertos muestra la variabilidad del significado personal logrado por los estudiantes de la muestra sobre el intervalo de confianza.

7.3. CONCLUSIONES RESPECTO A LAS HIPÓTESIS INICIALES

Además de los objetivos iniciales, planteamos también unas hipótesis aunque con final exploratoria, pues parte de estas hipótesis eran novedosas al no haber encontrado estudios previos sobre algunos de los puntos abordados. A continuación presentamos de nuevo las hipótesis y las conclusiones obtenidas sobre las mismas.

Hipótesis 1: *Existen diversidad de significados en los libros de texto de estadística para ingenieros en cuanto a los campos de problemas, a las representaciones, propiedades, formulaciones, argumentos y algoritmos de cálculo de los intervalos de confianza. Algunos de estos significados podrían ser incompletos o sesgados.*

Esta hipótesis se basaba en los resultados de otros estudios sobre libros de texto universitarios de estadística (Tauber, 2001; Alvarado, 2004, Alvarado y Batanero, 2004, 2005) y se ha confirmado solo parcialmente.

Respecto a la definición, la mayoría de los libros analizados (10/11) definen intervalo de confianza a partir de la fórmula de cálculo (en forma procedimental) y solo ocho de los 11 presentan la interpretación correcta del significado, es decir, la interpretación frecuencial correcta. Un texto no interpreta la definición y en los dos libros restantes aparece una interpretación ambigua que puede llevar a la idea de intervalo de credibilidad (que sería el punto de vista bayesiano y no es apropiada en la inferencia frecuencial).

Hipótesis 2: *Los textos analizados dan más importancia a los enunciados y propiedades del concepto, que a sus aplicaciones. Además, rara vez presentan el concepto a través de situaciones problemáticas.*

Esta hipótesis se apoya en los resultados de análisis de texto de Tauber (2001) y nuestra propia experiencia docente. En primer lugar hemos observado que en nueve de los 11 libros, las definiciones y proposiciones tienen prioridad sobre las aplicaciones, mostrando con ello un énfasis excesivo sobre la teoría. Por tanto esta parte de la hipótesis se confirma.

Por el contrario, respecto a la segunda parte se contradicen nuestras expectativas, pues las definiciones presentadas se refieren a la construcción del intervalo para la media poblacional. Aunque se trata de un ejemplo restringido, la definición se hace a partir del ejemplo. Junto a esto, hemos encontrado también una definición interesante de intervalo de confianza usando el método de inversión del estadístico de prueba aunque sólo está presente en dos de los once libros analizados: Devore (2005) y Hayter (1996); el primero es uno de los libros de texto más reconocidos en los campus del Sistema Tecnológico de Monterrey, según los resultados de las encuestas que realizamos.

Hipótesis 3. *Los campos de problemas propuestos en los textos no coinciden con los encontrados en el análisis histórico.*

También esta hipótesis se contradice, al contrario de lo que ocurrió en la investigación de Alvarado (2004) quien encontró este resultado sobre el teorema central del límite. Con relación a los campos de problemas hemos encontrado cinco tipos de problemas, dos de los cuales: estimar un parámetro desconocido para una población y comparar los valores de un parámetro en dos o más poblaciones, con sus 22 subcampos han dado origen históricamente, al desarrollo teórico del objeto intervalos de confianza. Observamos que en los campos de problemas encontrados en general se prefieren los estimadores insesgados y de mínima varianza en la construcción del intervalo.

Encontramos también que un buen número de los subcampos corresponde a dos configuraciones epistémicas principales como son: a) Las distribuciones muestrales que siguen o se aproximan mediante la distribución normal, esta configuración parte de estos campos de problemas y al buscar la distribución muestral llega a uno de los campos de problemas que dan origen al TCL y b) Las distribuciones muestrales que siguen o se aproximan mediante la distribución *t* de Student.

Hipótesis 4. *Los autores de los textos, recién empiezan a darle importancia al uso del ordenador como herramienta didáctica en la comprensión de este concepto.*

Esta hipótesis se confirma. En relación a los algoritmos y procedimientos, una observación que es un tanto preocupante, es que los textos analizados en su mitad no incluyen como práctica de estudio "comprobar hipótesis sobre la población objetivo con ayuda del ordenador" cayendo de esta manera en uno de los errores más frecuentemente denunciados en el uso incorrecto de la estadística en la investigación, como lo es no comprobar los supuestos requeridos para aplicarlos. Con respecto a la simulación, usando o no el ordenador, ésta aparece en solo el 54% de los libros analizados, con lo que esta herramienta tecnológica es subutilizada como recurso de apoyo a la enseñanza de la estadística.

Añadidas a estas conclusiones, la investigación da cuenta de las múltiples dimensiones del significado de los intervalos de confianza. Por ejemplo se muestra una amplitud, variedad, supuestos y restricciones en los términos que mejor describen los campos encontrados en los 11 libros de texto analizados.

Hipótesis 5. *Existe una amplia variedad de posibles conflictos semióticos y dificultades procedimentales de los estudiantes en relación a los intervalos de confianza.*

Esta hipótesis se confirma, en nuestro estudio de evaluación de la comprensión que demostraron los estudiantes de ingeniería de los intervalos de confianza, hemos encontrado errores procedimentales muy diversos y numerosos conflictos semióticos. Además de reforzar las conclusiones de investigaciones previas como las de Schuyten (1991), Vallecillos y Batanero (1997), Behar (2001), Fidler y Cumming (2005) y delMas, Garfield, Ooms y Chance (2007), hemos encontrado nuevas dificultades y conflictos semióticos descritos con anterioridad.

7.4. IDONEIDAD DEL CUESTIONARIO DE EVALUACIÓN

Una contribución importante de nuestro trabajo es la construcción de un instrumento validado de evaluación de la comprensión del intervalo de confianza. A continuación utilizamos algunas ideas teóricas de Batanero y Díaz (2005b) para analizar la idoneidad didáctica de nuestro instrumento de evaluación.

Las autoras indican que el interés en la evaluación es el significado personal de los estudiantes, aunque, por medio del instrumento solo logramos acceder a su significado declarado (las respuestas dadas se suponen una muestra representativa de las que darían los mismos estudiantes en la misma prueba en otras ocasiones).

Sin embargo, si las tareas son suficientemente representativas (para evaluar las unidades de contenido definidas), podríamos hacer una inferencia sobre lo que cada alumno de la muestra sería capaz de hacer y decir en otras tareas relacionadas con el intervalo de confianza. Si las unidades del contenido están bien definidas y representan este objeto matemático, entonces podríamos acercarnos al significado personal de los alumnos.

Por tanto, la posibilidad de generalizar en cada uno de los pasos descritos depende de la representatividad y la variabilidad de la muestra de tareas y también de la muestra de estudiantes elegida en cada uno de los procesos de muestreo en la construcción de nuestro cuestionario. A lo largo de nuestro trabajo, nosotros hemos tratado de conseguir la mayor generalizabilidad posible, aunque reconocemos nuestras limitaciones.

Siguiendo a las autoras, a continuación aplicamos el concepto de *idoneidad* y sus tipos (Godino, 2003; Godino, Contreras y Font, 2006) al cuestionario construido:

- La *dificultad* de un ítem o tarea daría una medida de su *idoneidad cognitiva*; es decir del grado de representatividad de los significados evaluados respecto a los significados personales. En nuestro trabajo, hemos obtenidos ítems con un grado de dificultad que varía de 79% (ítem6) hasta 32% (ítem 13). En general los resultados reportaron índices de dificultad aceptables y con una distribución normal (Losada y López Feal, 2003). Es decir tenemos ítems aplicables a una variedad de conocimientos de los estudiantes (desde fáciles a difíciles) abundando los de dificultad intermedia.
- La *discriminación* d*e un ítem valoraría su idoneidad evaluadora,* un ítem puede ser adecuado cognitivamente, pero no diferenciar (por ser demasiado fácil) los

alumnos que tienen un mayor o menor conocimiento del concepto. Esta idoneidad podría ser un comp*onente de la* idoneidad *instruccional*, en cuanto uno de los objetivos de la instrucción es la función evaluadora. En nuestro cuestionario el estudio de discriminación se llevó a cabo mediante un análisis discriminante, programa en donde se comparan las puntuaciones medias en cada ítem en el total de la prueba. Los resultados obtenidos indican una adecuada capacidad de discriminación de los ítems y del cuestionario en su conjunto (tabla 5.6).

- La *validez de contenido de un cuestionari*o *indicaría una idoneid*ad *epistémica*, o grado de representatividad del instrumento en cuanto al significado objeto de evaluación. Nosotros hemos asegurado la validez de contenido, mediante el análisis de contenid*o detallado q*ue ha dado *la definició*n semántica de la variable, así como mediante el juicio de expertos en la selección y depuración de los ítems, por lo que consideramos que el cuestionario tiene aceptable idoneidad epistémica.
- La *fiabilidad* o *generalizabilidad a otros ítems* daría una medida de la estabilidad de la respuesta, es decir sería otro componente de la *idoneidad evaluadora.* El coeficiente de fiabilidad o *generalizabilidad a otros ítems* coincide con la fiabilidad de consistencia interna, en nuestro caso, el índice obtenido ($G_i = 0.799$) para *nuestro cuestionari*o nos permitió calificarlo como *fiable*, la fiabilidad no es mayor debido a la variedad de contenidos evaluados, lo que hace que el cuestionario sea multidimensional, como se puso de manifiesto en el análisis factorial.
- La *validez externa y generalizabilidad a otros estudiantes*, sugeriría una *idoneidad generalizadora o externa* en cuanto los resultados se generalizarían a otros estudiantes. En este sentido nuestro cuestionario es limitado pues las pruebas se han hecho solamente con estudiantes de ingeniería, aunque hay una alta generalizabilidad a otros estudiantes con las mismas características, pues el índice de generalizabilidad da valores cercanos a la unidad.

7.5. APORTACIONES Y LIMITACIONES DEL ESTUDIO

Pensamos que el trabajo realizado muestra la formación alcanzada en los estudios de doctorado y aporta contribuciones originales. Las principales aportaciones de nuestro trabajo las podemos resumir en los siguientes puntos:

- La principal aportación es la construcción de un cuestionario validado, de una alta consistencia interna, con alta idoneidad didáctica y que permite proporcionar información valiosa de lo que los estudiantes de ingeniería aprenden y comprenden acerca de los intervalos de confianza después de ser instruidos acerca de ese tema. El cuestionario evalúa el significado personal adquirido por los estudiantes sobre el objeto de estudio, atendiendo a las diferentes dimensiones en este significado. Por lo tanto es un instrumento que puede ayudar a promover investigación futura en la enseñanza y aprendizaje de los intervalos de confianza.

- Otra aportación es la síntesis del estudio histórico sobre el intervalo de confianza que, aunque no se basa en textos originales, resume diferentes análisis históricos y filosóficos de otros autores y los presenta en forma asequible para los profesores. Esta revisión nos ha permitido también esclarecer los presupuestos y significados a través del tiempo del intervalo de confianza.
- Hemos completado una descripción del significado de referencia del intervalo de confianza en los libros de texto para ingenieros. Aunque restringido dicho significado de referencia al Sistema Tecnológico de Monterrey, esta descripción nos ha permitido al descubrir la riqueza del término en sus múltiples campos de problemas y diversas representaciones, favorecer el proceso de contextualización y personalización que deben realizar los alumnos para apropiarse de este saber.
- Tenemos una descripción detallada del significado personal en los estudiantes de ingeniería de los intervalos de confianza una vez estudiado el tema. Descripción que analiza la comprensión de diferentes elementos de significado y que ha confirmado resultados obtenidos en otros trabajos, pero que también presenta resultados con nuevos tipos de conflictos detectados.
- Tenemos un estado de la cuestión preparado que nos ha servido para identificar las principales dificultades en la comprensión de los intervalos de confianza que puede ayudar a otros investigadores interesados en trabajar en este tema y a los profesores que quieran promover cambios en la enseñanza de nuestro objeto de estudio.

Limitaciones del estudio

Estamos conscientes de que, a pesar de la importancia de estas aportaciones, los resultados no pueden ser generalizados en el sentido de que todos los estudiantes son de la especialidad de ingeniería y cursan sus estudios en el mismo centro. En este sentido nuestra investigación no se diferencia de otras, pero creemos conveniente que el estudio sea hecho de nuevo con otros grupos más grandes que representen mayormente a la población de alumnos en otros centros y otras especialidades, intentando con ello profundizar en la exploración de las dificultades de la comprensión de los intervalos de confianza.

Reconocemos también el carácter limitado de la investigación, en cuanto a la muestra de libros de texto utilizada, que, aunque recoge los que principalmente se usan en el sistema educativo del Tecnológico de Monterrey, pudiera diferir en otro sistema de enseñanza de estadística a ingenieros o bien en otras especialidades. Frente a un estudio más superficial de un mayor número de documentos, hemos preferido un estudio analítico profundo de una muestra más restringida

Otra limitación inherente a cualquier proceso de evaluación de los significados personales de los estudiantes y que constituye una fuente potencial de error en los resultados del cuestionario, es la posibilidad de que agentes como motivación, cansancio u otros afecten las respuestas de los estudiantes, a pesar de que en nuestro estudio hemos tratado en todo momento de que los estudiantes colaborasen y tratasen de responder lo mejor que eran capaces.

7.6. LÍNEAS DE INVESTIGACIÓN ABIERTAS.

Los resultados presentados en nuestro tema de investigación sientan las bases para continuar con otras líneas de trabajo que conduzcan al investigador a plantear nuevas propuestas para el análisis del significado y comprensión de los intervalos de confianza para los estudiantes. A continuación señalamos algunos temas en que este trabajo puede ser extendido, con el propósito de orientar futuras investigaciones sobre el tópico en cuestión.

Estudio sobre el significado de los intervalos de confianza.

En nuestro trabajo describimos el significado que se presenta de intervalo de confianza en la institución: “Enseñanza de estadística a ingenieros” ”, sin embargo, puesto que el significado institucional puede variar de acuerdo al enfoque o tratamiento que se le da al tema, una extensión de nuestra investigación podría ser dirigida a analizar el significado de los intervalos de confianza a otras áreas de enseñanza como ciencias sociales o ciencias de la salud o a los cursos de formación de investigadores.

Mejora del cuestionario

Algunos resultados en la pasación del cuestionario a la muestra de 252 alumnos nos sugieren la necesidad de hacer cambios en algunos reactivos que van desde una modificación en la redacción de un distractor, la eliminación del ítem, hasta la generación de nuevos ítems. Todo ello pone en evidencia la pertinencia de mejorar nuestro cuestionario repitiendo el proceso metodológico que seguimos en la construcción de nuestro instrumento de medición. Sería también necesario completar el estudio de indicadores psicométricos con la estimación de la fiabilidad test-retest o aplicando la teoría de respuesta al ítem.

Analizar el proceso de aprendizaje y el significado personal individual

Las evaluaciones que hemos realizado con el instrumento de medición que hemos construido nos han proporcionado información sobre las tendencias y variabilidad en los errores y dificultades específicos en el global de una amplia muestra de estudiantes. Sin embargo es posible progresar en la información y que podamos tener una visión más completa si utilizamos otras técnicas cualitativas. Por ejemplo entrevistas clínicas, observación del trabajo que desarrollan en clase los alumnos, nos permitirían analizar casos específicos y determinar niveles o tipologías en la comprensión del intervalo de confianza.

El diseño, experimentación y evaluación de una propuesta didáctica de enseñanza.

Nuestros resultados podrían ser la base para la preparación de material didáctico con aprendizaje en contexto sobre los intervalos de confianza. Dicha propuesta podría estar apoyada con software de corte didáctico con laboratorios virtuales, como el Statiscope que permite simular datos, introducir datos manualmente o a través de internet y analizarlos, mostrando la generación de muestras, distribuciones muestrales y el cálculo de intervalos de confianza. Esa interacción con la computadora deberá ser acompañada de guías didácticas, en donde se diseñe de manera clara y directa los significados pretendidos en el proceso de estudio, que permitan al alumno observar el comportamiento del fenómeno.

Estudios anteriores sugieren que la computadora incrementa la relevancia del aprendizaje, pues los estudiantes exploran y experimentan el significado de los intervalos y el efecto del tamaño de muestra, varianza y coeficiente (Terán, 2006). Sería asimismo necesario relacionar este tema con el estudio de las distribuciones de probabilidad y hacer ver a los estudiantes el error involucrado en usar una distribución muestral errónea en el cálculo de los intervalos de confianza.

Estudio sobre el significado de los intervalos de confianza en el tránsito de bachillerato a la universidad.

La inferencia estadística se incluye ahora en los cursos de Bachillerato, donde los estudiantes no cuentan con las herramientas avanzadas de cálculo y de probabilidades para poder abordarlo. Otro sendero de investigación que sería importante analizar, es el significado de los intervalos de confianza que sería posible incorporar para los estudiantes en el nivel de bachillerato; construir propuestas didácticas adaptadas y experimentarlas para luego observar el progresivo acoplamiento de los significados personales que construyen los alumnos a los significados institucionales que pretendemos adquieran.

Las posibles líneas de investigación aquí planteadas brindan perspectivas interesantes para continuar el análisis relativo a las concepciones iniciales de los estudiantes, la enseñanza y aprendizaje del tema y los obstáculos que la condicionan. Todo ello permitirá, a la larga, formular acciones controladas que nos permitan caracterizar, explicar, diseñar, gestionar y evaluar este objeto de estudio, con un mayor grado de profesionalización en nuestro quehacer docente.

REFERENCIAS

ABET (1997). Accreditation Board for Engineering and Technology. Trabajo presentado en el *Thirteenth Annual Meeting of National Electrical Engineering Department Heads Association*. Orlando, FL.

ABET (2001). Programme self study report for civil engineering. *EC2000 Open Enrolment Engineering Faculty Workshop*. Polytechnic University of Puerto Rico. On Line: http://www.ce.washington.edu/about/abet/abet_2001.html.

ABET (2005). 2006-2007. *Criteria for accrediting engineers programs*. On line: www.abet.org/forms.shtml.

Afifi, A. y Clark, V. (1990). *Computer- aided multivariate analysis.* Nueva York: Van Nostrand Reinhold.

Alvarado, H. (2004). *Elementos del significado del teorema central del límite en textos de estadística para ingenieros.* Departamento de Didáctica de la Matemática. Universidad de Granada.

Alvarado, H. (2007). *Significados institucionales y personales del teorema central del límite en la enseñanza de estadística en ingeniería.* Tesis Doctoral. Universidad de Granada.

Alvarado, H. y Batanero, C. (2004). Elementos del significado del teorema central del límite. *Actas del VIII Simposio SEIEM, Sociedad Española de Investigación en Educación Matemática, Comunicaciones en los grupos de investigación.* CD-ROM. La Coruña: SEIEM.

Alvarado, H. y Batanero, C. (2005). Diferentes aproximaciones al teorema central del límite en la enseñanza a ingenieros. *Actas del IX Congreso de Metodología de las Ciencias Sociales y de la Salud*. CD-ROM.. Granada: Asociación Española de Metodología de las Ciencias del Comportamiento.

Alvarado, H. y Batanero, C. (2006). Designing a study process of the central limit theorem for engineers. En A. Rossman y B. Chance (Eds.). *Proceedings of the Seventh International Conference on Teaching Statistics.* CD-ROM Salvador de Bahia: International Association for Statistical Education.

Alvarado, H. y Batanero, C. (2007). Dificultades de comprensión de la aproximación normal a la distribución binomial. *Números, 67*. On line: http://www.sinewton.org/numeros/.

American Psychological Association, American Educational Research Association y National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Psychological Association (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

Bain, L. y Engelhardt, M. (1989). *Introduction to probability and mathematical statistics*. Boston: PWS-Kent.

Barbero, M. (2003). *Psicometría II. Métodos de elaboración de escalas*. Madrid: UNED.

Batanero, C. (2000). Controversies around significance tests. *Journal of Mathematics Thinking and Learning, 2*(1-2), 75-98.

Batanero, C. (2001). *Didáctica de la estadística.* Granada: Grupo de Investigación en Educación Estadística.

Batanero, C. (2004). Statistics education as a field for research and practice. Regular Lecture. *ICME-10*. Copenhague.

Batanero, C. (2005). Controversias sobre el uso de la inferencia en la investigación experimental. Conferencia plenaria. *Actas del XVI Simposio Nacional de Estadística de Colombia*. CD-ROM. Paipa, Boyacá, Colombia: Universidad Nacional de Colombia..

Batanero, C. y Díaz, C. (2005). Análisis del proceso de construcción de un cuestionario sobre probabilidad condicional. Reflexiones desde el marco de la TFS. En A. Contreras, L. Ordóñez y C. Batanero (Eds.). *Investigación en Didáctica de las Matemáticas. I Congreso Internacional sobre Aplicaciones y Desarrollos de la Teoría de las Funciones Semióticas* (pp. 13–36). Jaén: Universidad de Jaén.

Batanero, C. y Díaz, C. (2007). Análisis del proceso de construcción de un cuestionario sobre probabilidad condicional. *Educação e Pesquisa, 8*(2), 197-223

Batanero, C., Tauber, L. y Meyer, R. (1999). From data analysis to inference: A research project on the teaching of normal distributions. *Proceedings of the 52nd Session of the International Statistical Institute* (Tome LVIII, Book 1, pp. 57-58).

Helsinki: ISI

Batanero, C., Tauber, L. y Sánchez, V. (2001). Significado y comprensión de la distribución normal en un curso de análisis de datos. *Quadrante, 10*(1), 59-92

Batanero, C., Tauber, L. y Sánchez, V. (2004). Student's reasoning about the normal distribution. En D. Ben-Zvi y J. B. Garfield (Eds), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 257-276). Dordrecht: Kluwer.

Behar, R. (2001). *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística.* Tesis doctoral. Universidad Politécnica de Cataluña.

Belia, S., Fidler, F. y Cumming, G. (2005). Researchers misunderstand confidence intervals and standar error bars. *Psychological Methods, 4,* 389-396.

Bickel P. J. y Doksum K. A. (1977). *Mathematical statistics: Basic ideas and selected tópics.* Holden-Day, San Francisco.

Biehler, R. (1991). Computers in probability education. En R. Kapadia y M. Borovcnick (Eds.), *Chance encounters: Probability in education* (pp. 169-211). Dordrecht: Kluwer.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review,* 65(2), 167-190.

Brennan, R.L. (1983). *Elements of generalizability theory*: Iowa, Blackwell.

Budé, L. (2006). Assessing students' understanding of statistics. En A. Rossman y Beth Chance (Eds.). *Proceedings of the Seventh International Conference on Teaching of Statistics.* CD-ROM. Salvador (Bahía), Brasil: International Association for Statistical Education.

Carmines, E. G. y Zeller, R. A. (1979). *Reliability and validity assessment*. Londres: Sage.

CACEI (2007). *Normatividad*. México: Consejo de Acreditación de la Enseñanza de la Ingeniería. On line: http://itsta.edu.mx/cacei/Normatividad2007.html.

Castro-Posada, J. (2001). *Metodología de la investigación*. *Fundamentos.* Salamanca: Amaru.

Clark, M. L. (2004). Los valores p y los intervalos de confianza, ¿En qué confiar? *Revista Panamericana de Salud Pública, 15*(5), 295-296.

Cobo, B. (2003). *Significado de las medidas de posición central para los estudiantes de secundaria.* Tesis doctoral. Universidad de Granada.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist, 49*(12), 997 – 1003.

Cohen, R. J. y Swerdlik, M. E. (2001). *Pruebas y evaluación picológicas*. México: McGraw- Hill.

Colle, R. (2001) Análisis de contenido. El autor. On line: http://www.puc.cl/curso_dist/conocer/analcon/index.html.

Cox, D. R.(1978). Foundations of statistical inference: The case for eclecticism. *Australian Journal of Statistics, 20,* 43-59.

Cruise, R., Dudley, R. y Thayer, J (1984). *A resource guide for introductory statistics.* Nueva York: Kendall/Hunt.

Cumming, G. (2002). Live figures: Interactive diagrams for statistical understanding. En L. B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics.* CD-ROM. Ciudad del Cabo: International Association for Statistical Education.

Cumming, G. y Fidler, F. (2005). Interval estimates for statistical communication: problems and possible solutions. Trabajo presentado en la *IASE Satellite Conference on Communication of Statistics.* Sydney: International Association for Statistical Education.

Cumming, G. y Finch, S. (2001). A primer understanding, use, and calculation of confidence intervals that are based on central and non-central distributions. *Educational and Psychological Measurement, 61*, 530-572.

Cumming, G. y Finch, S. (2005). Inference by eye: Confidence intervals, and how to read pictures of data. *American Psychologist, 60,170-180.*

Cumming, G. y Thomason, N. (1998). StatPlay: Multimedia for statistical understanding. En L. Pereira-Mendoza, L. Kea, T. Kee, & W-K. Wong (Eds.) *Statistical education: Expanding the network. Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 947-952). Voorburg, The Netherlands: International Statistical Institute.

Cumming, G., Williams, J. y Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics*, 3, 299-311

Chance, B. L., delMas, R. C., y Garfield, J. B. (2004). Reasoning about sampling distributions. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295-324). Dordrecht, The Netherlands: Kluwer.

Dane, F. C. (1990). *Research methods*. Thompson. Pacific Grow. CA.

Davies, H. (1998). *What are confidence intervals?* On line: http://www.evidence-based-medicine.co.uk.

delMas, R. (2001). *What makes the standard deviation larger or smaller?* Statistics Teaching and Resource Library (STAR). On line: http://www.causeweb.org/repository/StarLibrary/activities/delmas2001/ .

delMas, R.(2002). Statistical literacy, reasoning, and learning: A commentary. *Journal of Statistics Education, 10* (3). On line: www.amstat.org/publications/jse/.

delMas, R. C., Garfield, J. B. y Chance, B. L. (1998). Exploring the role of computer simulations in developing understand of sampling distributions. Trabajo presentado en el *American Educational Research Association. Annual Meeting*. Montreal.

delMas, R. C., Garfield, J. B. y Chance, B. L. (1999). A model of classroom research in action: developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education, 7*(3). On line: http://www.amstat.org/publications/jse.

delMas, R. C., Garfield, J. B. y Chance, B. L. (2004). Using assessment to study the development of students' reasoning about sampling distributions. Trabajo presentado en el *American Educational Research Association. Annual Meeting*. California.

delMas, R.C., Garfield, J.B., Ooms, A. y Chance, B.L. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28-58. On line: www.stat.auckland.ac.nz/serj.

Devore, J. (1998). *Probabilidad y estadística para ingeniería y ciencias.* Cuarta edición. México: Thomson.

Devore, J. (2005). *Probabilidad y estadística para ingeniería y ciencias*. Sexta edición. México: Thomson.

Díaz, C. (2004). *Elaboración de un instrumento de evaluación del razonamiento condicional. Un estudio preliminar.* Memoria DEA. Universidad de Granada.

Díaz, C. (2007). *Viabilidad de la enseñanza de la inferencia bayesiana en el análisis de datos en psicología*. Tesis Doctoral. Universidad de Granada.

Díaz , C., Batanero, C. y Wilhelmi, M. R. (En prensa). Errores frecuentes en el análisis de datos en educación y psicología. *Publicaciones.*

Díaz, C. y de la Fuente, I. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento.* Volumen especial 2004, 161-167.

Díaz, C., y de la Fuente, E. I. (2005). Construcción de un cuestionario sobre comprensión de la probabilidad condicional. En J. Ortiz y A. Montenegro (Eds.), *Actas del XV Simposio de Estadística.* CD-ROM. Bogotá: Universidad Nacional de Colombia.

Díaz, C. y de la Fuente, E. I. (2006). Assessing psychology students' difficulties with conditional probability and Bayesian reasoning. En A. Rossman y B. Chance (Eds.), *Proceedings of Seventh International Conference on Teaching of Statistics* CD-ROM. Salvador de Bahia: International Association for Statistical Education.

Díaz, C., y de la Fuente, I. (2007a). Assessing students' difficulties with conditional probability and Bayesian reasoning. *International Electronic Journal of Mathematics Education, 2*(3), 129-148. On line: http://www.iejme.com/.

Díaz, C. y de la Fuente, I. (2007b). Validación de un cuestionario de razonamiento probabilístico condicional. *REMA*, *12*(1). On line: http://www.psico.uniovi.es/REMA/.

Doane, D. P., Tracy, K. y Mathieson, D. (2001). *Visual statistics :Student's book.* Boston: McGraw-Hill.

Dunn, O. J. y Clark, V.A. (1987). *Applied statistics: Analysis of variance and Regression*. Nueva York: John Wiley.

Ebel, R .L. y Frisbie, D.A. (1986). *Essentials of education measurement*. Englewood Cliffs, NJ: Prentice Hall.

Eco, U. (1979). *Tratado de semiótica general.* Barcelona: Lumen

Escuela de Medicina (1996). *Epidemiología descriptiva.* Pontificia Universidad Católica de Chile. On line: http://escuela.med.puc.cl/.

Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores.* Tesis Doctoral. Universidad de Granada.

Estes, W. K. (1997). On the communication of information by displays of standard errors and confidence intervals. *Psychonomic Bulletin and Review, 4,* 330-341.

Falk, R., y Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory and Psychology, 5* (1), 75-98.

Feldt, L. S. y Brennan, R. L. (1991). Reliability. En R. Linn (Ed.), *Educational measurement* (pp. 105-146). Nueva York, Mc- Millan.

Fidler, F. y Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *Proceedings of the 55th International Statistics Institute Session* CD-ROM. Sidney, Australia: International Statistical Institute.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A, 222*, 309-368.

Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society, 26*, 528-535.

Fox, D. J. (1981). *El proceso de investigación en la educación*. Pamplona: Eunsa.

Garfield, J. B., delMas, R. C., y Chance, B. L. (1999). The role of assessment in research on teaching and learning statistics. Trabajo presentado en el *American Educational Research Association. Annual Meeting*. Montreal.

Garfield, J. B., delMas, R. C., y Chance, B. L. (2004). *Tools for teaching and assessing statistical inference.* On line: http://www.gen.umn.edu/research/stat_tools.

Ghiglione, R. y Matalón, B. (1991). *Les enquêtes sociologiques. Théorie et practique.* París: Armand Colin.

Ghiselli, E., Campbell, J. y Zedeck, S (1981). *Measurement theory for the behavioral Sciences*. USA: W.H. Freeman and Company.

Gil Flores, J. (1994). *Análisis de datos cualitativos*. Barcelona. P.P.U.

Gingerenzer, G. (1993). The superego, the ego and the id in statistical reasoning. En G. Keren y C. Lewis (Eds.*), A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 311-339). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goetz, J. P. y Lecompte, M. D. (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid: Morata.

Godino, J. D. (1995). ¿Qué aportan los ordenadores a la enseñanza y aprendizaje de la estadística? *UNO, 5*, 45-55.

Godino, J. D. (1996). Mathematical concepts, their meanings and understanding. En L. Puig y A. Gutiérrez (Eds.), *Proceedings of the 20th PME Conference* (v.2, 417-424). Universidad de Valencia.

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathematiques, 22* (2 y 3), 237-284

Godino, J. D. (2003). Teoría de las funciones semióticas. Un enfoque ontológico-semiótico de la cognición e instrucción matemática. *Departamento de Didáctica de la Matemática. Universidad de Granada*. On line: URL: http://www.ugr.es/local/jgodino/.

Godino, J. D. y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathematiques*, *14*(3), 325-355.

Godino, J. D. y Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area of research in Mathematics Education. En A. Sierpinska y J. Kilpatrick (Eds.), *Mathematics education a research domain: A search for identity* (pp. 177-195). Dordrecht: Kluwer.

Godino, J. D. Batanero, C. y Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM. The International Journal on Mathematics Education, 39* (1-2), 127-135.

Godino, J. D., Batanero, C. y Roa, R. (2005). An onto-semiotic analysis of combinatorial problems and the solving processes by university students. *Educational Studies in Mathematics, 60*(1), 3-36.

Godino, J. D., Bencomo, D., Font, V. y Wilhelmi, M. R. (2006). Análisis y valoración de la idoneidad didáctica de procesos de estudio de las matemáticas. *Paradigma,* XXVII (2), 221-252.

Godino, J. D., Contreras A. y Font, V. (2006). Análisis de procesos de instrucción basado en el enfoque ontológico-semiótico de la cognición matemática. *Recherches en Didactiques des Mathematiques*, *26* (1), 39-88.

Hald, A. (1998). *A history of mathematical statistics from 1750 to 1930*. Nueva York: John Wiley.

Harlow, L., Mulaik, S. A. y Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.

Hawkins, A., Jolliffe, F. y Glickman, L. (1992). *Teaching statistical concepts*. Essex: Longman.

Hayter, A. (1996). *Probability and statistics for engineering and scientists*. Boston: PWS Publishing Company.

Hernández, R., Fernández, C. y Baptista, P. (1998). *Metodología de la investigación*. México: Mc Graw-Hill

Hin, P. y Ping, T. (2002). The teaching of confidence interval and prediction interval. En L. B. Phillips (Ed.), *Proceedings of the Seventh International Conference on Teaching Statistics.* CD-ROM. Ciudad del Cabo: International Association for Statistical Education.

Hinojosa, E y Góngora, J. (2006). *Taller de evaluación y Diseño de reactivos*. Monterrey: ITESM.

Hjelmslev, L. (1971). *Prolegómenos a una teoría del lenguaje*. Madrid: Gredos, Trabajo original publicado en 1943.

Hodgson, T. (1996). The influence of hands-on activities on student´s understanding of selected statistical concepts. En E. Jacubowski, D. Watkins & H. Biske (Eds.), *Proceedings of Eighteenth Annual Meeting of North American Chapter of the International Group for the Psychology of Mathematics* Education (pp. 241-246). Panamá City, Fl: PME-NA.

Huberman, A. M. y Miles, M. (1994). Data management and analysis methods. En N. K. Denzin y Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428 – 444). Londres: Sage Publications.

Huck, S., Cross, T. y Clark, S. (1986). Overcoming misconceptions about z-scores. *Teaching Statistics, 8,* 38-40.

ITESM (2005). *Visión y misión 2015 del Tecnológico de Monterrey*. Monterrey: ITESM.

Johnson, R. y Kuby, P. (2004). *Estadística elemental*. México: Thompson.

Kahneman, D., Slovic, P., y Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Nueva York: Cambridge University Press.

Lane, D.M. (2001). *HyperStat.* Online: http://www.ruf.rice.edu/~lane/rvls.html.

Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes? *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute* (Tome 58, Book 2) (pp. 205 – 208). Helsinki: International Statistical Institute.

León, O. G. y Montero, I. (2002). *Métodos de investigación en psicología y educación.* Madrid: McGraw-Hill.

López, B. e Hinojosa, E. (2001). *Evaluación del aprendizaje*. México: Trillas.

López Feal, R. (1986). *Construcción de instrumentos de medida en ciencias conductuales y sociales.* Barcelona: Alamex.

López Sánchez, J., Pérez de Vargas, A., Zamora Romero, J., Murciano Cespedosa, A., Alonso Fernández, J., Reviriego Eiros, M. y Lahoz Beltrá, R. (2004). *Aula virtual de Bioestadística.* On line: http://e-stadistica.bio.ucm.es/.

Losada, J. L. y López Feal, R. (2003). *Métodos de investigación en ciencias humanas y sociales*. Madrid: Thomson.

MacGillivray, H. L. (2002a). Technology, statistical thinking and engineering students. En B. Phillips (Ed), *Proceedings of the Sixth International Conference on Teaching Statistics*, Ciudad del Cabo: ISI. CD ROM.

MacGillivray, H. L. (2002b). Lessons from engineering student projects in statistics, *Proceedings of the Australasian Engineering Education Conference* (pp. 225-230). Australia: The Institute of Engineers.

Magnusson, D. (1990). *Teoría de los tests: Psicometría diferencial, psicología aplicada, orientación vocacional.* México: Editorial Trillas.

Martínez Arias, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos.* Madrid: Síntesis.

Mayo, D. G. (1981). In defense of the Neyman-Pearson theory of statistics. *Philosophy of Science 48*, 269-280.

M.E.C. (2004). *Una educación de calidad para todos y entre todos. Propuestas para el debate*. Madrid: Ministerio de Educación y Ciencia.

Meliá, J. L. (2001). *Teoría de la fiabilidad y la validez*. Valencia: Cristóbal Serrano.

Mendenhall, W., Scheaffer, R. y Wackerly, D. (1986). *Estadística matemática con aplicaciones.* Primera edición. México: Grupo Editorial Iberoamérica.

Mendenhall, W. y Sincich, T. (1997). *Probabilidad y estadística para ingeniería y ciencias*. Quinta edición. México: Prentice Hall.

Méndez, H. (1991). *Understanding the central limit theorem.* Tesis doctoral. Universidad de California. UMI 6369.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45,* 35-44.

Miller, I., Freund, J. y Johnson, R (1997). *Probabilidad y estadística para ingenieros*. Quinta edición. México: Prentice Hall.

Millman, J. y Green, J. (1989). The specification and development of test of achievement and ability. En R. L. Linn (Ed.), *Educational Measurement* (pp. 335 – 366). Londres*:* Macmillan.

Milton, J. y Arnold, J. (2004). *Probabilidad y estadística (con aplicaciones para ingeniería y ciencias computacionales)*. Cuarta edición. México: McGraw Hill.

Montgomery, D. y Runger, G. (2004). *Probabilidad y estadística aplicadas a la ingeniería*. Segunda edición. México: Limusa.

Muñiz, J. (1994). *Teoría clásica de los tests.* Madrid: Pirámide.

Navarro–Pelayo, V. (1994). *Estructura de los problemas combinatorios simples y del razonamiento combinatorio en alumnos de secundaria*. Tésis doctoral. Universidad de Granada.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, *17*, 857-872

Neyman, J. (1934). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosofical Transaction of the Royal Society of London, series A, Mathematical and Physical Sciences 236* (767), 33-380.

Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biométrica* 32, 2, 128-150.

Neyman, J. (1976). The emergence of mathematical statistics En D. B. Owen (Ed.), *On the history of statistics and probability* (pp. 149-189). Nueva York: Marcel Dekker.

Nickerson, R. S. (1995). Can technology help teach for understanding? En D. N. Perkins, J. L. Schwartz, M. M. West y M. S. Wiske (Eds.), *Software goes to school: Teaching for understanding with new technologies*. Nueva York: Oxford University Press.

Nunnally, J. y Bernstein, I. (1995). *Teoría psicométrica*. Tercera edición. México: McGraw-Hill.

Olivo, E. (2006). *Análisis de la presentación de intervalos de confianza en textos de estadística para ingenieros*. Trabajo de Investigación Tutelada. Universidad de Granada.

Olivo, E. y Batanero, C. (2006). Intervalos de confianza, desarrollo histórico e implicaciones didácticas. Comunicación presentada en la *X Escuela de Invierno en Matemática Educativa*. Tlaxcala, México.

Olivo, E. y Batanero, C. (2007a). Un acercamiento exploratorio a las dificultades de comprensión del intervalo de confianza. Comunicación presentada en la *XI Escuela de Invierno en Matemática Educativa*. Mérida (Yucatán), México.

Olivo, E. y Batanero, C. (2007b). Un estudio exploratorio de dificultades de comprensión del intervalo de confianza. *UNION*, 12. On line: http://www.sinewton.org/numeros/.

Olivo, E. y Batanero, C. (En prensa). Dificultades de comprensión del intervalo de confianza en estudiantes universitarios. *Educación Matemática.*

Olivo, E., Batanero, C. y Ortiz, J. J. (2006). Notas históricas sobre los intervalos de confianza e implicaciones didácticas. Comunicación presentada en el *IX Simposio de SEIEM. Grupos de Trabajo.* Granada.

Olivo, E., Ruiz, B. y Albert, A. (2006). Confidence intervals: fields of problems in engineering textbooks. En A. Rossman y Beth Chance (Eds.). *Proceedings of the Seventh International Conference on Teaching of Statistics.* Salvador (Bahía), Brasil: International Association for Statistical Education.

Ortiz, J. J. (1999). *Significado de los conceptos probabilísticos elementales en los textos de Bachillerato*. Tesis Doctoral. Universidad de Granada.

Osterlind, S. J. (1989). *Constructing test items*. Boston: Kluwer.

Pascual, J., García, J. F. y Frías, M. D. (2000). Significación estadística, importancia del efecto y replicabilidad de los datos. *Psicothema, 12* (2), 408-412.

Pearson, E. S. (1966). *Selected papers*. Cambridge University Press.

Peña, D. (2002). *Análisis de datos multivariables.* Madrid: McGraw-Hill.

Peterson, R. (2000). *Constructing effective questionnaires*. Thousand Oaks: Sage.

Rao, C. R. (1992). R. A. Fisher: The founder of modern statistics. *Statistical Science*, *7*, 1, 34-48.

Rivadulla, A. (1991). *Probabilidad e inferencia científica.* Barcelona: Anthropos.

Rouanet, H. (1998). Statistics for researchers. En H. Rouanet (Eds.), *New ways in statistical methodology* (pp. 1 – 28). Berna: Peter Lang.

Ryan, T. (1989). *Statistical methods for quality improvement*. Nueva York: Wiley.

Sánchez-Cobo, F. (1999). *Significado de la regresión y correlación para estudiantes universitarios*. Tesis Doctoral. Universidad de Granada.

Santisteban, C. (1990). *Psicometría aplicada.* Madrid: Norma.

Savage, L.J. (1962). *The foundations of statistical inference*. Londres: Methuen.

Sax, G. (1989). *Principles of educational and psychological measurement and evaluation.* Belmont, CA: Wadsworth.

Scheaffer, R. y McClave, J. (1993). *Probabilidad y estadística para ingeniería*. Tercera edición. México: Grupo Editorial Iberoamérica.

Schenker, N. y Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician, 55*, 182-186.

Schuyten, G. (1991). Statistical thinking in psychology and education. En D. Vere-Jones (Ed.). *Proceeding of the Third International Conference on Teaching Statistics* (pp. 486-490).Voorburg, The Netherlands: International Statistical Institute.

SECAI (1998). *Conceptos, alcance, metodología e instrumentos.* Madrid: Instituto de Ciencias de la Educación (ICE).

Sedlmeier, P. (1999). *Improving statistical reasoning. Theoretical models and practical implications.* Mahwah, NJ: Erlbaum.

Selander, S. (1990). Análisis del texto pedagógico. En J. García y M. Beas (Comp.), *Libro de texto y construcción de materiales curriculares,* (pp. 131 – 161). Granada: Proyecto Sur de Ediciones.

Snir, J., Smith, C. y Grosslight, L. (1995). Conceptually enhanced simulations: Acomputer tool for science teaching. En D. N. Perkins, J. L. Schwartz, M. M. Westy M. S. Wiske (Eds). *Software goes to school: Teaching for understanding with new technologies*. Nueva York: Oxford University Press.

Stigler, S. M. (1986). *The history of statistics the measurement of uncertainty before 1900*. Prensa de Belknap: Universidad de Harvard.

Tauber, L. (2001). *Significado y comprensión de la distribución normal a partir de actividades de análisis de datos.* Tesis Doctoral. Universidad de Sevilla.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*(2), 26 – 30.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 25 – 32.

Thorndike, R. L. (1989). *Psicometría aplicada.* Mexico: Limusa.

Terán, T. (2006). Elements of meaning and its role in the interaction with a computacional program. En A. Rossman y B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* CD-ROM. Salvador (Bahia): International Association for Statistical Education.

Vacha-Haase, T. (2001). Statistical significance should not be considered one of life’s guarantees: Effect sizes are needed. *Educational and Psychological Measurement, 61*, 219-224.

Vallecillos, A. (1994). *Estudio teórico - experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios.* Tesis doctoral Departamento de Didáctica de la Matemática. Universidad de Granada.

Vallecillos, A. (1995). Consideraciones epistemológicas sobre la inferencia estadística: implicaciones para la práctica docente. *UNO,* 5, 80-90.

Vallecillos, A. (1996). *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis*. Granada: Comares.

Vallecillos, A. (1998). Experimental study on the learning of the significance level concept. En L.Pereira-Mendoza, L. Seu Kea, T. Wee Kee & W. Wong (Eds.), *Proceedings of ICOTS 5* (pp. 1475-1476). Singapore: Nanyang Technological University.

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute* (Tome 58, Book 2, pp. 201-204). Helsinki, Finland.

Vallecillos, A. y Batanero, C. (1997). Conceptos activados en el contraste de hipótesis estadísticas y su comprensión por estudiantes universitarios. *Recherches en Didactique des Mathématiques, 17* (1), 29-48.

Vardeman, S. R. (2002). Providing "real" context in statistical quality control courses for engineers. En B. Phillips (Ed), *Proceedings of the Sixth International Conference on Teaching Statistics.* CD-ROM, Ciudad del Cabo: IASE.

Velasco, G y Wisniewski, P. (2001). *Probabilidad y estadística para ingeniería y ciencias*. México: Thompson .

Velleman, P. F.(2003). *ActivStats*. Data Description, Inc. Information. Online: http: www.datadesk.com.

Vera. O., Olivo, E., Alvarado, H. y Batanero, C. (2007). Estadística y competencias en la formación del ingeniero. En M. Molina, P. Pérez-Tyteca y M. A. Fresno (Eds.). *Jornadas de Investigación en el Aula de Matemáticas. Competencias matemáticas*. CD- ROM. Granada. Sociedad Thales y Departamento de Didáctica de las Matemáticas.

Walpole, R., Myers, R. y Myers, S. (1992). *Probabilidad y estadística*. Sexta edición. México: Prentice Hall, Pearson.

Walpole, R. E. y Myers, R. H. (1999). *Probabilidad y estadística*. Cuarta edición. México: McGraw Hill.

Wasan M.T. (1970). *Parametric estimation.* Nueva York: McGraw-Hill.

Weber, R. P. (1985). *Basic content analysis.* Londres: Sage.

Well, A. D., Pollastsek, A.y Boyce, S. J. (1990). Understanding the effects of the sample size on the variability of the means. *Organizational Behavior and Human Decision Processes, 47*, 289-312.

Wiersma, W. (1986). *Research methods in education.* Boston: Allyn y Bacon.

Wilhelmi, M. R., Godino, J. D. y Lacasta, E. (2007). Configuraciones epistémicas asociadas a la noción de igualdad de números reales. *Recherches en Didactique des Mathematiques 27* (1), 77 - 120.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.

Wolfe, R. y Cumming, G. (2004). Communicating the uncertainty in research findings: Confidence intervals. *Journal of Science and Medicine in Sport, 7*, 138-143.

ANEXOS

ANEXO 1.
CUESTIONARIO PARA RECOGIDA DE DATOS DE EXPERTOS

Enseguida presentamos una lista de contenidos que consideramos relevantes para evaluar la comprensión del tema de intervalos de confianza. Presentamos también, junto a cada contenido, una lista de ítems de evaluación.

Requerimos su colaboración para evaluar:

- El grado en que el contenido propuesto es relevante para la comprensión de la probabilidad condicional.
- El grado en que cada ítem es adecuado para evaluar la comprensión del contenido específico propuesto.

Le agradeceríamos marque para cada uno de estos dos aspectos su opinión en la escala 1 a 5, donde 1 indica: nada relevante y 5 muy relevante.

Contenido 1: Definición de intervalo de confianza.

Ítem 1. Explica con tus propias palabras la diferencia entre una estimación puntual y un intervalo de confianza

Ítem 2. El intervalo de confianza del 50% para la media de una población μ es:

a. El rango dentro del cuál caen el 50% de los valores de la media de la muestra $\overline{x}$.
b. Un intervalo más ancho que el intervalo de confianza del 95%.
c. **Un intervalo de valores calculado a partir de los datos de la muestra. En el 50% de las muestras de una población, el intervalo calculado contiene a la media.**
d. Dos veces más ancho que el intervalo de confianza del 100%.

Ítem 3. En un intervalo de confianza:

a. De una muestra a otra, el intervalo es constante.
b. Se especifica un rango de valores dentro de los cuales supuestamente cae el parámetro con seguridad.
c. **Indica un intervalo de posibles valores para el parámetro, y un porcentaje de intervalos que cubrirán, aproximadamente dicho valor, para el mismo tamaño de muestra.**
d. Siempre contienen el parámetro poblacional.

	1: Nada	2	3	4	5: Mucho
El contenido “Definición de intervalo de confianza” es relevante					
El ítem 1 es adecuado para este contenido					
El ítem 2 es adecuado para este contenido					
El ítem 3 es adecuado para este contenido					

Contenido 2: El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra

Ítem 4. Dos muestras diferentes se toman de una población donde la media poblacional y la desviación estándar poblacional son desconocidas. La primera muestra tiene 25 datos, y la segunda muestra 64 datos. Se construye un intervalo de confianza de 95 % para cada muestra para estimar la media poblacional. Que intervalo de confianza esperaría que tenga mayor precisión?

a. Espero que ambos intervalos de confianza tengan la misma precisión.
b. **Espero que el intervalo de confianza basado en una muestra de 64 datos sea más preciso.**
c. Espero que el intervalo de confianza basado en la muestra de 25 datos sea más preciso.
d. No puedo determinar cuál de los dos tendrá más precisión.

Ítem 5. Hemos calculado un intervalo de confianza al 90% basado en el valor medio $\bar{x}$ obtenido de una muestra de 10 casos. Si incrementamos el tamaño de la muestra a 1000, y calculamos un segundo intervalo al 90 % de confianza:

a. Debemos tener más confianza de que μ caerá en nuestro segundo intervalo.
b. **Sabemos que el segundo intervalo será 10 veces más angosto**.
c. El intervalo de confianza tiene el mismo ancho, pero su centro se desplaza.
d. El segundo intervalo de confianza es 100 veces más ancho que el primero.

Ítem 6. Comparado a los intervalos de confianza calculados en muestras de tamaño n=4, el ancho de los intervalos de confianza de la media de la población calculado en muestras de tamaño n = 50:

a. variará más que los anchos de los intervalos para muestras de tamaño n =4.
b. **variará un poco, pero no tanto como lo hicieron los anchos de los intervalos para muestras de tamaño *n*=4.**
c. Tomarán valores parecidos**.**

	1: Nada	2	3	4	5: Mucho
El contenido “El ancho de los intervalos de confianza disminuye cuando aumenta el tamaño de la muestra” es relevante					
El ítem 4 es adecuado para este contenido					
El ítem 5 es adecuado para este contenido					
El ítem 6 es adecuado para este contenido					

Contenido 3: El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta

Ítem 7. Si manteniendo todos los demás datos fijos el nivel de confianza se reduce (por ejemplo de 90% a 80%)

a. El intervalo de confianza no cambia.
b. El intervalo de confianza será más ancho.
c. **El intervalo de confianza será más angosto.**
d. El cambio en el intervalo de confianza no es predecible.

Ítem 8. Habrá más muestras donde la media de la población caiga en el intervalo de confianza con:

a. Un aumento en el tamaño de la población.

b. **Coeficientes de confianza más grandes** (por ejemplo 95% en vez de 90%).

c. Muestras pequeñas.

d. Más variabilidad en la población.

Ítem 9. En un intervalo de confianza, el ancho del intervalo puede ser reducido por:

a. Disminuyendo el tamaño de la muestra.

b. **Bajando el nivel de confianza (por ejemplo de 0.99 a 0.90).**

c. Aumentando la magnitud de $\sigma_{\bar{x}}$.

d. Un aumento en el tamaño de la población.

	1: Nada	2	3	4	5: Mucho
El contenido “El ancho de los intervalos de confianza aumenta cuando el nivel de confianza aumenta” es relevante					
El ítem 7 es adecuado para este contenido					
El ítem 8 es adecuado para este contenido					
El ítem 9 es adecuado para este contenido					

Contenido 4: El ancho de los intervalos de confianza aumenta cuando aumenta la varianza

Ítem 10. ¿Cuál de las siguientes afirmaciones es verdadera?

a. Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza no cambia.

b. **Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza disminuye.**

c. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza disminuye.

d. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza no cambia.

Ítem 11. Explica cómo varía la anchura del intervalo de confianza de la media si, conservando el mismo tamaño de muestra y el mismo coeficiente de confianza tomamos una población con varianza doble

	1: Nada	2	3	4	5: Mucho
El contenido “El ancho de los intervalos de confianza aumenta cuando aumenta la varianza” es relevante					
El ítem 10 es adecuado para este contenido					
El ítem 11 es adecuado para este contenido					

Contenido 5: Significado del nivel de confianza (variación del intervalo en diferentes muestras)

Ítem 12. Un intervalo del 95% de confianza para la diferencia media en la producción de leche de una ganadería después de un tratamiento resultó ser (1.5, 3.5) Litros/vaca. ¿Cuál de las siguientes afirmaciones es verdadera?

a. **No sabemos el verdadero aumento medio en la producción, pero en el 95% de intervalos calculados con el mismo tamaño de muestra y población se incluye el aumento medio en la producción.**
b. Debido a que el intervalo de confianza no contiene el cero, hay 95% de probabilidad de que tocar música no produce ningún efecto.
c. Debido a que el intervalo de confianza no contiene el cero, hay 95% de probabilidad que el verdadero aumento en la producción es 2.5 Litros/vaca.

Ítem 13. En un intervalo de confianza del 95% para la media:

a. Si se toman muchas muestras $\bar{x}$ caerá dentro del intervalo de confianza el 95% de las veces.
b. La probabilidad de que $\bar{x}$ caiga dentro de un intervalo de confianza calculado de una muestra es 0.95
c. **Si se toman muchas muestras, el intervalo de confianza calculado contendrá a μ 95% de las veces**.

Ítem 14. Un estudiante toma una muestra de 16 compañías en Estados Unidos: El salario medio ofrecido por esas 16 compañías es de $30,600 dólares. El intervalo de confianza al 95% calculado para esta muestra va de un límite inferior de $20,500 a un límite superior de $40,500

Cuál de las siguientes afirmaciones es verdadera

a. El estudiante tiene una probabilidad de 95% de que su salario inicial estará entre $20,500 y $40,500.
b. Esto significa que el 95% de las compañías tienen un salario promedio entre $20,500 y $40,500.
c. **Si tomamos otras 16 compañías diferentes, los límites del intervalo podrían variar, pero el ancho del intervalo sería aproximadamente 20,000 dólares.**

	1: Nada	2	3	4	5: Mucho
El contenido “Significado del nivel de confianza (variación del intervalo en diferentes muestras)” es relevante					
El ítem 12 es adecuado para este contenido					
El ítem 13 es adecuado para este contenido					
El ítem 14 es adecuado para este contenido					

Contenido 6: Estimar la media de una población normal o en una muestra grande con σ conocida

Ítem 15. La media muestral de 100 observaciones en una prueba de matemáticas es 75. Encuentre el intervalo de confianza al 95% para la media de la población, asumiendo que $\sigma = 7$

a. (61.28, 88.72) (no divide la desviación típica por 10).
b. **(73.63, 76.37)**
c. (68, 82) no multiplica por 1.96 ni divide por 10.
d. (74.3 , 75.7) no multiplica por 1.96.

Ítem 16. El propietario de una tienda desea estimar el número promedio de lápices vendidos por día. Una muestra aleatoria de 25 días es seleccionado de una población normal y el valor de la media muestral es 100. La desviación estándar de la población es $\sigma = 15$ Calcule el límite superior para un intervalo de confianza al 95%

a. 104.92 usa 1.64.

b. 103.00 no multiplica por 1.96.

c. 106.18 usa la *t*, 24 grados de libertad.

d. **105.88**

	1: Nada	2	3	4	5: Mucho
El contenido "Estimar la media de una población normal o en una muestra grande con σ conocida" es relevante					
El ítem 15 es adecuado para este contenido					
El ítem 16 es adecuado para este contenido					

Contenido 7: Estimar la media de una población aproximadamente normal cuando, σ es desconocida

Ítem 17. Construya un intervalo de confianza al 95% para la media de una población normal de desviación típica **σ** desconocida si en una muestra de tamaño 10, la media de la muestra es $\bar{x}$ =25 y la estimación de la desviación típica en la muestra es s = 6

a. **20.71 – 29.29**
b. 21.28 - 28.72 usa la normal en vez de la *t.*
c. 21.44 – 38.56 no divide por *n.*
d. 23.11 - 26.89 no multiplica por el valor de *t.*

Ítem 18. Un fabricante asegura que sus garrafones, contienen un litro de cloro puro. Al tomar una muestra de 16 garrafones se determinó que en promedio contenían .94 litros de cloro puro, con desviación estándar de la muestra de .097. Construir un intervalo de confianza al 95 %, para el verdadero contenido promedio de litros de cloro puro. No se conoce la desviación típica de la población.

	1: Nada	2	3	4	5: Mucho
El contenido "Estimar la media de una población aproximadamente normal cuando, σ es desconocida" es relevante					
El ítem 17 es adecuado para este contenido					
El ítem 18 es adecuado para este contenido					

Contenido 8: Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande

Ítem 19. Se han obtenido los siguientes datos de emisión diaria de óxidos de azufre, para una muestra de tamaño *n*=100, media: $\bar{x}$ = 18 y cuasivarianza s^2 =36. Elabore un intervalo de confianza de 95% para la verdadera emisión diaria promedio de óxidos de azufre.

a. (17.016, 18.984) usar 1.64.
b. **(16.824, 19.176)**
c. (6.24, 29.76) no dividir entre *n.*
d. (8.16, 27.84) no dividir entre *n* y usar 1.64.

Ítem 20. Un estudiante de economía toma una muestra de 36 compañías a través de los Estados Unidos. Imagine que el salario medio ofrecido por esas 36 compañías es de $30000 dólares con una desviación estándar de $20000. Obtener un intervalo de confianza al 95% para el verdadero salario medio.

a. (0, 69200) no dividir por n.
b. (10000, 50000) sumar/restar la desviación típica.
c. **(23466, 36533)**
d. (26667, 33333) no multiplicar por 1.96.

	1: Nada	2	3	4	5: Mucho
El contenido "Estimar la media de una población a partir de datos experimentales σ desconocida, muestra grande" es relevante					
El ítem 19 es adecuado para este contenido					
El ítem 20 es adecuado para este contenido					

Contenido 9: Estimar una proporción

Ítem 21. En una muestra aleatoria de 100 rodamientos, 10 tienen un acabado de especificaciones defectuoso. Calcular el intervalo de confianza de 95% para la proporción verdadera de rodamientos defectuosos.

Ítem 22. En un estudio con 240 jóvenes estadounidenses cuyas edades van de 16 a 19 años, seleccionados al azar, 36 presentaron problemas graves de sobrepeso. Obtenga un intervalo de confianza de 99% para la verdadera proporción *p* de jóvenes de esta población con problemas graves de sobrepeso.

a. (.105, .195) se multiplicó por 1.96.
b. **(.091, .209)**
c. (0, .849) se multiplicó por 1.96 y no se dividió por n.
d. (.097, .203) se multiplicó por 2.33.

	1: Nada	2	3	4	5: Mucho
El contenido "Estimar una proporción" es relevante					
El ítem 21 es adecuado para este contenido					
El ítem 22 es adecuado para este contenido					

Contenido 10: Estimar una varianza

Ítem 23. Sea σ^2 la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es s^2=13700, n=16. Calcular el intervalo de confianza de 95% para σ.

Ítem 24. La cantidad de dióxido de carbono (CO_2) liquido presente en un proceso inclusión geológico en cinco días distintos en una roca cristalizada tuvo una varianza muestral igual a 80 Haga una estimación de la precisión de la técnica LRM estableciendo un intervalo de confianza de 99% para la variación en las mediciones de concentración de CO_2

a) (19.10, 776.69) usar 5 grados de libertad.
b) (1723.2, 123671.2) elevar al cuadrado 80.
c) **(21.54, 1545.89)**
d) (1528, 62135.2) elevar al cuadrado 80 y usar 5 grados de libertad.

	1: Nada	2	3	4	5: Mucho
El contenido "Estimar una varianza" es relevante					
El ítem 23 es adecuado para este contenido					
El ítem 24 es adecuado para este contenido					

Contenido 11: Comparar las medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes

Ítem 25. La siguiente tabla contiene un resumen de información sobre la resistencia a la compresión de cubos (N/mm^2) para especímenes de concreto

	Tamaño muestral	Media Muestral
Tipo 1	1.68	26.99
Tipo 2	2.72	37.56

Suponga que las desviaciones estándar poblacionales de ambos grupos son $\sigma_1 = 4.89$ y $\sigma_2 = 6.43$ respectivamente.Calcule un intervalo de confianza de 99% para hallar la diferencia entre el verdadero promedio de resistencia en el Tipo1 y el verdadero promedio de resistencia en el Tipo 2:

a. **(-13.02, -8.12)**
b. (-12.437, -8.70) usar 1.96.
c. (-31.32, 10.18) no dividir entre n y m.
d. (-18.64, -2.5) no dividir entre n y m y no multiplicar por 2.57.

Ítem 26. La siguiente tabla contiene un resumen de información sobre la duración de baterías AA alcalinas marca Duracell y marca Eveready Energizer.

	Tamaño muestral	Media muestral
Duracell	100	4.1 hrs
Eveready Energizer	100	4.5 hrs

Suponga que las desviaciones estándar poblacionales de ambos grupos son $\sigma_1 = 1.8$ y $\sigma_2 = 2.0$ respectivamente..Calcule un intervalo de confianza de 95% para hallar la diferencia entre el verdadero promedio de duración para Duracell y el verdadero promedio de duración de Eveready Energizer.

	1: Nada	2	3	4	5: Mucho
El contenido "Comparar las medias en dos poblaciones, conociendo σ_1^2 y σ_2^2, muestras independientes" es relevante					
El ítem 25 es adecuado para este contenido					
El ítem 26 es adecuado para este contenido					

Contenido 12: Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas

Ítem 27. En un experimento para estudiar los efectos de aplicación de fertilizante de cal y urea sobre la retención de dimetoato (un insecticida) por suelo arcilloso, se observaron los siguientes porcentajes de recuperación de dimetoato:

Suelo tratado con cal: n = 5, $\bar{x} = 26.58$ y s = 2.43

Suelo tratado con urea: m = 5, $\bar{x} = 40.24$ y s = 2.93

$s_p = 2.69$

Ítem 28. Se compararon dos soluciones de grabado diferentes, usando dos muestras aleatorias de tamaño 10. Los resultados de la rapidez de grabado fueron:

Solución 1: $\bar{x} = 9.97$ y s = .422

Solución 2: $\bar{x} = 10.4$ y s = .073

$s_p = 0.34$

Calcule un intervalo de confianza de 90% para la diferencia de las medias de la rapidez de grabado.

	1: Nada	2	3	4	5: Mucho
El contenido "Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ pero desconocidas, muestras independientes pequeñas" es relevante					
El ítem 27 es adecuado para este contenido					
El ítem 28 es adecuado para este contenido					

Contenido 13: Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ desconocidas, muestras independientes grandes

Ítem 29. La tabla siguiente resume algunos datos de un experimento realizado para estudiar varias características de tornillos de anclaje

		Resistencia al corte	
Diámetro de tornillo	Tamaño muestra	Media muestral	Desviación estándar
3/8	100	4.25	1.3
½	100	7.25	1.7

Calcule e interprete un intervalo de confianza de 95% para la diferencia del verdadero promedio de resistencias al corte.

a. **(-3.41, -2.58) podemos tener una alta confianza de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.**
b. (-3.41, -2.58) podemos tener una confianza moderada de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.
c. (-3.35, -2.65) podemos tener una confianza moderada de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.

Ítem 30. Se investiga el diámetro de varillas de acero fabricadas en dos máquinas

	tamaño muestral	media muestral	varianza muestral
Máquina 1	36	8.73	0.36
Máquina 2	40	8.08	0.40

Construya un intervalo de confianza de 95% para la diferencia de las medias del diámetro de las varillas.

a. (-1.05, 2.35) no dividir entre *n* y *m.*
b. (.319, .881) usar 1.64.
c. **(.373, .927)**
d. (-.776, 2.076) no dividir entre *n* y *m* y usar 1.64.

	1: Nada	2	3	4	5: Mucho
El contenido “Comparar las medias en dos poblaciones, $\sigma_1^2 = \sigma_2^2$ desconocidas, muestras independientes grandes” es relevante					
El ítem 29 es adecuado para este contenido					
El ítem 30 es adecuado para este contenido					

Contenido 14: Comparar dos varianzas poblacionales

Ítem 31. Una compañía quiere seleccionar el proceso de pulido que presente la variabilidad menor. Una muestra aleatoria de $n_1 = 16$ piezas del primer proceso da como resultado una desviación estándar muestral $s_1 = 5$ micropulgadas, y una muestra aleatoria de $n_2 = 11$ piezas del segundo proceso da como resultado una desviación estándar muestral $s_2 = 4$ micropulgadas. Establezca un intervalo de confianza de 90% para σ_1^2 / σ_2^2 , Suponiendo que los dos procesos son independientes y que la aspereza superficial tiene una distribución normal, ¿cuál de los dos procesos recomendaría usted?

a. Como todos los valores están dentro del intervalo (.54, 3.96) recomendaría el proceso 1.
b. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría el proceso 2. (se cambian los grados de libertad del numerador y denominador)
c. **Como el cociente de valores está dentro del intervalo (.54, 3.96) recomendaría cualquiera de los dos**
d. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría cualquiera de los dos. (se cambian los grados de libertad del numerador y denominador)

Ítem 32. Una empresa ha estado experimentando con dos disposiciones físicas distintas de su línea de ensamble. Dos muestras aleatorias independientes producen los resultados que se muestran en la siguiente tabla. Establezca un intervalo de confianza de 95% para σ_1^2 / σ_2^2 , la razón de las varianzas del número de unidades terminadas para las dos disposiciones de línea de ensamble. Con base en el resultado, cuál de las dos disposiciones recomendaría usted?

Línea de ensamble 1	Línea de ensamble 2
n_1 = 21 días	n_2 = 25 días
s_1^2 = 1,432	s_2^2 = 2,864

	1: Nada	2	3	4	5: Mucho
El contenido “Comparar dos varianzas poblacionales” es relevante					
El ítem 31 es adecuado para este contenido					
El ítem 32 es adecuado para este contenido					

Contenido 15: Elegir un modelo de distribución muestral del estadístico

Ítem 33. La distribución Muestral utilizada en la construcción de intervalos de confianza para la varianza en muestras pequeñas es

a. Distribución *t* de Student.
b. **Distribución Ji-cuadrada.**
c. Distribución Normal.
d. Distribución *F*.

Ítem 34. Cómo serán afectados los intervalos de confianza usando una desviación estándar muestral (s) y un valor *t* en vez de la desviación estándar poblacional (σ) y un valor *z?*

a. Los centros de los intervalos de confianza serán diferentes e igual de amplios
b. **Los intervalos tendrán mayor amplitud y el mismo centro**
c. Los intervalos tendrán menor amplitud y el mismo centro
d. Los intervalos tendrán mayor amplitud y diferente centro

	1: Nada	2	3	4	5: Mucho
El contenido “Elegir un modelo de distribución muestral del estadístico”es relevante					
El ítem 33 es adecuado para este contenido					
El ítem 34 es adecuado para este contenido					

Contenido 16: Determinar valores críticos en la distribución del estadístico

Ítem 35. Al calcular un intervalo de confianza del 90% para la media, para un grupo de puntuaciones distribuido normalmente, usted pudiera usar un valor de *z* de:

a. 1.96
b. **1.65**
c. 0.90 interpretar el 90% como el valor de *z*.
d. 1.29 acumulada hasta .90.

Ítem 36. Si el nivel de confianza es 0.95 para un intervalo de confianza para la media poblacional con desviación estándar poblacional desconocida. Para un grupo de puntajes distribuido normalmente de tamaño $n = 20$, los valores críticos pudieran ser:

a. -1.65 y 1.65 uso de normal estándar.
b. -1.96 y 1.96 uso de normal estándar.
c. **-2.093 y 2.093**
d. -2.085 y 2.085 usar 20 grados de libertad.

	1: Nada	2	3	4	5: Mucho
El contenido “Determinar valores críticos en la distribución del estadístico” es relevante					
El ítem 35 es adecuado para este contenido					
El ítem 36 es adecuado para este contenido					

Contenido 17: Interpretación de intervalos de confianza utilizando un programa de ordenador

Ítem 37. La siguiente salida de computadora presenta una muestra simulada de una población normal con $\mu = 130$ y $\sigma = 10$. Luego se usó un comando para establecer un intervalo de confianza del 95% para μ.

118.690	144.226	138.827	125.934	136.198	148.731	133.394
134.184	127.260	145.345	121.966	125.435	141.236	150.021
126.895	118.499	137.225	136.781	119.526	125.628	134.865
116.416	134.312	138.053	140.828			

Mean	Median	TrMean	StDev	95% Confidence Interval for Mu	
132.82	134.31	132.78	9.74	128.798	136.840

Escriba el intervalo de confianza que se obtuvo e interprete el resultado

Ítem 38. La siguiente salida de computadora presenta dos muestras simuladas de dos poblaciones normales. La población 1 con $\mu = 90$ y $\sigma = 10$ y la población 2 con $\mu = 92$ y $\sigma = 10$

Muestra 1:

83.3195	87.6793	86.7831	95.0518	92.9781	86.6457
85.1305	97.5013	83.1112	82.2751	82.7831	90.2786
89.5876	71.2591	82.0282	90.6264		

Muestra 2 :

82.312	95.098	92.598	85.959	91.319	108.130
90.392	90.074	78.789	100.923	85.601	89.861
78.685	100.354	81.267	101.432		

	N	Media	D. Típica	Error típico
C1	16	86.69	6.25	1.6
C2	16	90.80	8.70	2.2

```
95%  IC para mC1 - mC2: (-9.6;  1.4)
```

Escriba el intervalo de confianza que se obtuvo para la diferencia de medias e interprete el resultado

	1: Nada	2	3	4	5: Mucho
El contenido “Interpretación de intervalos de confianza utilizando un programa de ordenador ” es relevante					
El ítem 37 es adecuado para este contenido					
El ítem 38 es adecuado para este contenido					

Contenido 18: Interpretar Gráficos de Intervalos de confianza

Ítem 39. Considere el gráfico siguiente del rendimiento medio de cebada en 1980,1984 y 1988 junto con un intervalo de 95% de confianza

```
Intervalos del 95% para la media

Año N Media StDev        ----+---------+---------+---------+--
1980 6 184.00 2.61         (----*---)
1984 5 212.40 14.36                            (-----*----)
1988 5 182.40 1.82       (-----*----)
                         ----+---------+---------+---------+--
Desv. Típica conjunta =.19 180       195       210       225
```

Cuál de las siguientes afirmaciones es verdadera

a. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay buena evidencia que la medias de las muestras difieran

b. La estimación de la media de la población en 1980 es menos precisa que en 1988

c. Puesto que los intervalos de confianza para 1980 y 1984 no se solapan, hay poca evidencia que las medias de las poblaciones respectivas difieran.

d. **Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay poca evidencia que las medias de las poblaciones difieran.**

Ítem 40. Considere el gráfico siguiente del rendimiento medio de cebada en 1990,1994 y 1998 junto con un intervalo de 95% de confianza

```
Individual 95% CIs For Mean

Año N Media StDev          --+---------+---------+---------+--
1990 36 238.83 26.57                          (-----*----)
1994 75 186.60 2.51        (-----*-----)
1998 67 186.71 2.63          (----*----)
                           --+---------+---------+---------+----
Desv. Típica conjunta=15.49 175     200       225       250
```

Cuál de las siguientes afirmaciones es verdadera

a. **Puesto que los intervalos de confianza para 1994 y 1998 tienen considerable solape, hay poca evidencia que la medias de las poblaciones difieran.**

b. La estimación de la media de la población en 1998 es menos precisa que en 1994

c. Puesto que los intervalos de confianza para 1990 y 1994 no se solapan, no hay evidencia que las medias de las poblaciones respectivas difieran.

d. Puesto que los intervalos de confianza para 1994 y 1998 tienen considerable solape hay buena evidencia que la medias de las muestras difieran.

	1: Nada	2	3	4	5: Mucho
El contenido "Interpretar Gráficos de Intervalos de confianza" es relevante					
El ítem 39 es adecuado para este contenido					
El ítem 40 es adecuado para este contenido					

ANEXO 2
RESULTADOS ENSAYO PRE-PILOTO

Ítem 1. Explica con tus propias palabras la diferencia entre una estimación puntual y un intervalo de confianza.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Define correctamente solo una de las estimaciones pedidas
2: Define correctamente las dos estimaciones pedidas.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	0	0.00		
1	5	0.11	.89	.38
2	43	0.89		

Ítem 2. El intervalo de confianza del 50% para la media de una población μ es:

a. El rango dentro del cuál caen el 50% de los valores de la media de la muestra $\overline{x}$.

b. Un intervalo más ancho que el intervalo de confianza del 95%.

c. **Un intervalo de valores calculado a partir de los datos de la muestra. En el 50% de las muestras de una población, el intervalo calculado contiene a la media.**

a) Dos veces más ancho que el intervalo de confianza del 100%.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	4	0.10		
B	3	0.07		
C	32	0.76	.76	.44
D	3	0.07		

Ítem 3. En un intervalo de confianza:

a. De una muestra a otra, el intervalo es constante.

b. Se especifica un rango de valores dentro de los cuales supuestamente cae el parámetro con seguridad.

c. **Indica un intervalo de posibles valores para el parámetro, y un porcentaje de intervalos que cubrirán, aproximadamente dicho valor, para el mismo tamaño de muestra.**

a) Siempre contienen el parámetro poblacional.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	6	0.12		
B	5	0.10		
C	36	0.75	.75	.81
D	1	0.02		

Ítem 4. Dos muestras diferentes se toman de una población donde la media poblacional y la desviación estándar poblacional son desconocidas. La primera muestra tiene 25 datos, y la segunda muestra 64 datos. Se construye un intervalo de confianza de 95 % para cada muestra para estimar la media poblacional. ¿Qué intervalo de confianza esperaría que tenga mayor precisión?

a. Espero que ambos intervalos de confianza tengan la misma precisión.

b. **Espero que el intervalo de confianza basado en una muestra de 64 datos sea más preciso**.

c. Espero que el intervalo de confianza basado en la muestra de 25 datos sea más preciso.

d. No puedo determinar cuál de los dos tendrá más precisión.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	2	0.04		
B	12	0.25		
C	31	0.65	.25	-.56
D	3	0.06		

Ítem 5. Hemos calculado un intervalo de confianza al 90% basado en el valor medio $\overline{x}$ obtenido de una muestra de 10 casos. Si incrementamos el tamaño de la muestra a 1000, y calculamos un segundo intervalo al 90 % de confianza:

a. Debemos tener más confianza de que μ caerá en nuestro segundo intervalo.

b. **Sabemos que el segundo intervalo será 10 veces más angosto**.

c. El intervalo de confianza tiene el mismo ancho, pero su centro se desplaza.

d. El segundo intervalo de confianza es 100 veces más ancho que el primero.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	16	0.38		
B	8	0.19		
C	6	0.14	.19	.12
D	12	0.29		

Ítem 6. Comparado a los intervalos de confianza calculados en muestras de tamaño n=4, el ancho de los intervalos de confianza de la media de la población calculado en muestras de tamaño n = 50:

a. Variará más que los anchos de los intervalos para muestras de tamaño n =4.

b. **Variará un poco, pero no tanto como lo hicieron los anchos de los intervalos para muestras de tamaño *n* =4.**

c. Tomarán valores parecidos.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	12	0.29		
B	24	0.57		
C	6	0.14	.57	.57

Ítem 7. Si manteniendo todos los demás datos fijos, el nivel de confianza se reduce (por ejemplo de 90% a 80%):

a. El intervalo de confianza no cambia.

b. El intervalo de confianza será más ancho.

c. **El intervalo de confianza será más angosto.**

d. El cambio en el intervalo de confianza no es predecible.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	9	0.19		
B	3	0.06		
C	34	0.70	.70	.88
D	2	0.04		

Ítem 8. Habrá más muestras donde la media de la población caiga en el intervalo de confianza con:

a. Un aumento en el tamaño de la población.

b. **Coeficientes de confianza más grandes** (por ejemplo 95% en vez de 90%)

c. Muestras pequeñas.

d. Más variabilidad en la población.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	6	0.14		
B	24	0.57		
C	7	0.17	.57	-.15
D	5	0.02		

Ítem 9. En un intervalo de confianza, el ancho del intervalo puede ser reducido por:

a) Disminuyendo el tamaño de la muestra.

b) Bajando el nivel de confianza (por ejemplo de 0.99 a 0.90)

c) Aumentando la magnitud de $\sigma_{\bar{x}}$.

d) Un aumento en el tamaño de la población.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	14	0.29		
B	10	0.20		
C	3	0.06	.20	.39
D	21	0.44		

Ítem 10. ¿Cuál de las siguientes afirmaciones es verdadera?:

a. Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza no cambia.

b. **Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza disminuye.**

c. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza disminuye.

d. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza no cambia.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	4	0.08		
B	14	0.29		
C	14	0.29	.29	-.31
D	16	0.33		

Ítem 11. Explica cómo varía la anchura del intervalo de confianza de la media si, conservando el mismo tamaño de muestra y el mismo coeficiente de confianza tomamos una población con varianza doble.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Explica de forma imprecisa, pero señala algunas ideas importantes.
2: Explica en forma precisa.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	17	0.40		
1	7	0.17	.42	.88
2	18	0.42		

Ítem 12. Un intervalo del 95% de confianza para la diferencia media en la producción de leche de una ganadería después de un tratamiento resultó ser (1.5, 3.5) Litros/vaca. ¿Cuál de las siguientes afirmaciones es verdadera?

a. **No sabemos el verdadero aumento medio en la producción, pero en el 95% de intervalos calculados con el mismo tamaño de muestra y población se incluye el aumento medio en la producción.**

b. Debido a que el intervalo de confianza no contiene el cero, hay 95% de probabilidad de que el tratamiento no produce ningún efecto.

c. Debido a que el intervalo de confianza no contiene el cero, hay 95% de probabilidad que el verdadero aumento en la producción es 2.5 Litros/vaca.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	31	0.64		
B	6	0.13	.64	.01
C	10	0.21		

Ítem 13. En un intervalo de confianza del 95% para la media μ:

a. Si se toman muchas muestras $\overline{x}$ caerá dentro del intervalo de confianza el 95% de las veces.

b. La probabilidad de que $\overline{x}$ caiga dentro de un intervalo de confianza calculado de una muestra es 0.95.

c. **Si se toman muchas muestras, el intervalo de confianza calculado contendrá a μ 95% de las veces**.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	3	0.07		
B	7	0.17	.76	.63
C	32	0.76		

Ítem 14. Un estudiante toma una muestra de 16 compañías en Estados Unidos. El salario medio ofrecido por esas 16 compañías es de $30,600 dólares. El intervalo de confianza al 95% calculado para esta muestra va de un límite inferior de $20,500 a un límite superior de $40,500. ¿Cuál de las siguientes afirmaciones es verdadera?

a. El estudiante tiene una probabilidad de 95% de que su salario inicial estará entre $20,500 y $40,500.

b. Esto significa que el 95% de las compañías tienen un salario promedio entre $20,500 y $40,500.

c. **Si tomamos otras 16 compañías diferentes, los límites del intervalo podrían variar, pero el ancho del intervalo sería aproximadamente 20,000 dólares.**

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	14	0.33		
B	22	0.52	.14	.18
C	6	0.14		

Ítem 15. La media muestral de 100 observaciones en una prueba de matemáticas es 75. Encuentre el intervalo de confianza al 95% para la media de la población, asumiendo que $\sigma = 7$

a. (61.28, 88.72)

b. **(73.63, 76.37)**

c. (68, 82)

d. (74.3 , 75.7)

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	9	0.19		
B	33	0.68		
C	2	0.04	.68	.88
D	4	0.08		

Ítem 16. El propietario de una tienda desea estimar el número promedio de lápices vendidos por día. Una muestra aleatoria de 25 días es seleccionado de una población normal y el valor de la media muestral es 100. La desviación estándar de la población es $\sigma = 15$. El límite superior para un intervalo de confianza al 95% es:

a. 104.92

b. 103.00

c. 106.18

d. **105.88**

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	0	0.00		
B	6	0.14		
C	16	0.38	.47	1.00
D	20	0.47		

Ítem 17 Construya un intervalo de confianza al 95% para la media de una población normal de desviación típica **σ** desconocida, si en una muestra de tamaño 10 la media de la muestra es $\bar{x}$ =25 y la estimación de la desviación típica en la muestra es $s = 6$.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Plantea el intervalo con la distribución muestral correcta.
2: Plantea el intervalo con la distribución muestral correcta y además encuentran el valor crítico

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	8	0.17		
B	12	0.25		
C	28	0.58	.58	.53

Ítem 18. Un fabricante asegura que sus garrafones, contienen un litro de cloro puro. Al tomar una muestra de 16 garrafones se determinó que en promedio contenían .94 litros de cloro puro, con desviación estándar de la muestra de .097. Construir un intervalo de confianza al 95 %, para el verdadero contenido promedio de litros de cloro puro. No se conoce la desviación típica de la población.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Plantea el intervalo con la distribución muestral correcta.
2: Plantea el intervalo con la distribución muestral correcta y además encuentran el valor crítico

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	6	0.14		
1	19	0.45		
2	17	0.40	.40	.82

Ítem 19. Se han obtenido los siguientes datos de emisión diaria de óxidos de azufre, para una muestra de tamaño *n*=100: media: $\bar{x}$ = 18 y cuasivarianza s^2 =36. Elabore un intervalo de confianza de 95% para la verdadera emisión diaria promedio de óxidos de azufre.

a. (17.016, 18.984)
b. **(16.824, 19.176)**
c. (6.24, 29.76)
d. (8.16, 27.84)

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	5	0.10		
B	33	0.68		
C	7	0.15	.68	.63
D	2	0.04		
Other		0.02		

Ítem 20. Un estudiante de economía toma una muestra de 36 compañías a través de los Estados Unidos. Imagine que el salario medio ofrecido por esas 36 compañías es de $30000 dólares con una desviación estándar de $20000. El intervalo de confianza al 95% para el verdadero salario medio es:

a. (0, 69200)
b. (10000, 50000)
c. **(23466, 36533)**
d. (26667, 33333)

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	3	0.07		
B	3	0.07		
C	35	0.83	.83	.38
D	1	0.02		

Ítem 21. En una muestra aleatoria de 100 rodamientos, 10 tienen un acabado de especificaciones defectuoso. Calcular el intervalo de confianza de 95% para la proporción verdadera de rodamientos defectuosos.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Plantea el intervalo con la distribución muestral correcta.
2: Plantea el intervalo con la distribución muestral correcta y además encuentran el valor crítico

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	2	0.05		
1	10	0.23		
2	32	0.73	.73	.57

Ítem 22. En un estudio con 240 jóvenes estadounidenses cuyas edades van de 16 a 19 años, seleccionados al azar, 36 presentaron problemas graves de sobrepeso. Obtenga un intervalo de confianza de 99% para la verdadera proporción *p* de jóvenes de esta población con problemas graves de sobrepeso.

a. (.105, .195) se multiplicó por 1.96
b. **(.091, .209)**
c. (0, .849) se multiplicó por 1.96 y no se dividió por *n*
d. (.097, .203) se multiplicó por 2.33

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	5	0.11		
B	34	0.72		
C	1	0.02	.72	.46
D	7	0.15		

Ítem 23. Sea σ^2 la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es s^2=13700, *n*=16. Calcular el intervalo de confianza de 95% para σ.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Plantea el intervalo con la distribución muestral correcta.
2: Plantea el intervalo con la distribución muestral correcta y además encuentran el valor crítico

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	8	0.16		
1	20	0.41	.41	.18
2	20	0.41		

Ítem 24. La cantidad de dióxido de carbono (CO_2) liquido presente en un proceso inclusión geológico en cinco días distintos en una roca cristalizada tuvo una varianza muestral igual a 80. Se hace una estimación de la precisión de la técnica LRM estableciendo un intervalo de confianza de 99% para la variación en las mediciones de concentración de CO_2. El intervalo es:

a. (19.10, 776.69)
b. (1723.2, 123671.2)
c. **(21.54, 1545.89)**
d. (1528, 62135.2)

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	10	0 24		
B	15	0.36		
C	17	0.40	.40	.82
D	0	0.00		

Ítem 25. La siguiente tabla contiene un resumen de información sobre la resistencia a la compresión de cubos (N/mm^2) para especimenes de concreto.

	Tamaño muestral	Media muestral
Tipo 1	68	26.99
Tipo 2	72	37.56

Suponga que las desviaciones estándar poblacionales de ambos grupos son $\sigma_1 = 4.89$ y $\sigma_2 = 6.43$ respectivamente.
Calcule un intervalo de confianza de 99% para hallar la diferencia entre el verdadero promedio de resistencia en el Tipo1 y el verdadero promedio de resistencia en el Tipo 2.

a. **(-13.02, -8.12)**
b. (-12.437, -8.70) usar 1.96
c. (-31.32, 10.18) no dividir entre *n* y *m*
d. (-18.64, -2.5) no dividir entre *n* y *m* y no multiplicar por 2.57

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	25	0.57		
B	18	0.41		
C	1	0.02	.57	.05
D	0	0.00		

Ítem 26. La siguiente tabla contiene un resumen de información sobre la duración de baterías AA alcalinas marca Duracell y marca Eveready Energizer.

	Tamaño muestral	Media muestral
Duracell	100	4.1 hrs
Eveready Energizer	100	4.5 hrs

Suponga que las desviaciones estándar poblacionales de ambos grupos son $\sigma_1 = 1.8$ y $\sigma_2 = 2.0$ respectivamente.
Calcule un intervalo de confianza de 95% para hallar la diferencia entre el verdadero promedio de duración para Duracell y el verdadero promedio de duración de Eveready Energizer.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Plantea el intervalo con la distribución muestral correcta.
2: Plantea el intervalo con la distribución muestral correcta y además encuentran el valor crítico

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	0	0.00		
1	9	0.19	.81	.11
2	38	0.81		

Ítem 27. En un experimento para estudiar los efectos de aplicación de fertilizante de cal y urea sobre la retención de dimetoato (un insecticida) por suelo arcilloso, se observaron los siguientes porcentajes de recuperación de dimetoato:
Suelo tratado con cal: n = 5, $\bar{x} = 26.58$ y s = 2.43
Suelo tratado con urea: m = 5, $\bar{x} = 40.24$ y s = 2.93
$s_p = 2.69$
Calcule un intervalo de confianza de 95% para la diferencia en el verdadero promedio de porcentaje de recuperación de dimetoato entre los dos tratamientos.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Plantea el intervalo con la distribución muestral correcta.
2: Plantea el intervalo con la distribución muestral correcta y además encuentran el valor crítico

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	1	0.02		
1	23	0.52	.45	.55
2	20	0.45		

Ítem 28. Se compararon dos soluciones de grabado diferentes, usando dos muestras aleatorias de tamaño 10. Los resultados de la rapidez de grabado fueron:

Solución 1: $\bar{x} = 9.97$ y s = .422

Solución 2: $\bar{x} = 10.4$ y s = .073

$s_p = 0.34$

Calcule un intervalo de confianza de 90% para la diferencia de las medias de la rapidez de grabado.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Define correctamente solo una de las estimaciones pedidas
2: Define correctamente las dos estimaciones pedidas.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	5	0.11		
1	10	0.21	.68	.41
2	32	0.68		

Ítem 29. La tabla siguiente resume algunos datos de un experimento realizado para estudiar varias características de tornillos de anclaje.

		Resistencia al corte	
Diámetro de tornillo (pulg)	Tamaño muestra	Media muestral	Desviación estándar muestral
3/8	100	4.25	1.3
½	100	7.25	1.7

Calcule e interprete un intervalo de confianza de 95% para la diferencia del verdadero promedio de resistencias al corte.

a. **(-3.41, -2.58) podemos tener una alta confianza de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.**
b. (-3.41, -2.58) podemos tener una confianza moderada de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.
c. (-3.35, -2.65) podemos tener una confianza moderada de que la verdadera diferencia de la resistencia al corte cae en el intervalo anterior.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	26	0.59		
B	13	0.30	.59	.64
C	5	0.11		

Ítem 30. Se investiga el diámetro de varillas de acero fabricadas en dos máquinas

	tamaño muestral	media muestral	varianza muestral
Máquina 1	36	8.73	0.36
Máquina 2	40	8.08	0.40

Construya un intervalo de confianza de 95% para la diferencia de las medias del diámetro de las varillas.

a. (-1.05, 2.35) no dividir entre *n* y *m*
b. (.319, .881) usar 1.64
c. **(.373, .927)**
d. (-.776, 2.076) no dividir entre *n* y *m* y usar 1.64

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	0	0,00		
B	7	0.15		
C	40	0.85	.85	.31
D	0	0.00		

Ítem 31. Una compañía quiere seleccionar el proceso de pulido que presente la variabilidad menor. Una muestra aleatoria de $n_1 = 16$ piezas del primer proceso da como resultado una desviación estándar muestral $s_1 = 5$ micropulgadas, y una muestra aleatoria de $n_2 = 11$ piezas del segundo proceso da como resultado una desviación estándar muestral $s_2 = 4$ micropulgadas. Establezca un intervalo de confianza de 90% para $\sigma_1{}^2 / \sigma_2{}^2$, Suponiendo que los dos procesos son independientes y que la aspereza superficial tiene una distribución normal, ¿cuál de los dos procesos recomendaría usted?

a. Como todos los valores están dentro del intervalo (.54, 3.96) recomendaría el proceso 1.
b. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría el proceso 2. (se cambian los grados de libertad del numerador y denominador).
c. **Como el cociente de valores está dentro del intervalo (.54, 3.96) recomendaría cualquiera de los dos.**
d. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría cualquiera de los dos (se cambian los grados de libertad del numerador y denominador).

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	4	0.09		
B	3	0.07		
C	35	0.80	.80	.21
D	2	0.05		

Ítem 32. Una empresa ha estado experimentando con dos disposiciones físicas distintas de su línea de ensamble. Dos muestras aleatorias independientes producen los resultados que se muestran en la siguiente tabla. Establezca un intervalo de confianza de 95% para $\sigma_1{}^2 / \sigma_2{}^2$, la razón de las varianzas del número de unidades terminadas para las dos disposiciones de línea de ensamble. Con base en el resultado, cuál de las dos disposiciones recomendaría usted?

Línea de ensamble 1	Línea de ensamble 2
n_1 = 21 días	n_2 = 25 días
s_1^2 = 1,432	s_2^2 = 2,864

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: Obtiene el intervalo incorrectamente y explica de forma imprecisa.
1: Obtiene el intervalo correctamente, pero explica de forma imprecisa.

2: Obtiene el intervalo correctamente y explica en forma precisa.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	0	0.00		
1	17	0.36		
2	30	0.64	.64	-.03

Ítem 33. La distribución muestral utilizada en la construcción de intervalos de confianza para la varianza en muestras pequeñas, tomadas de una población normal, es:

a. Distribución t de Student.

b. **Distribución Ji-cuadrada.**

c. Distribución Normal.

d. Distribución F.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	17	0.35		
B	27	0.56		
C	3	0.06	.56	.91

Ítem 34. Cómo serán afectados los intervalos de confianza usando una desviación estándar muestral (s) y un valor *t* en vez de la desviación estándar poblacional (σ) y un valor *z?*

a. Los centros de los intervalos de confianza serán diferentes e igual de amplios.

b. **Los intervalos tendrán mayor amplitud y el mismo centro.**

c. Los intervalos tendrán menor amplitud y el mismo centro.

d. Los intervalos tendrán mayor amplitud y diferente centro.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	3	0.07		
B	9	0.21		
C	30	0.71	.21	-.56
D	0	0.00		

Ítem 35. Al calcular un nivel de confianza para un 90% para la media, para un grupo de puntuaciones distribuido normalmente, usted pudiera usar un valor z de:

a. 1.96

b. **1.65**

c. 0.90

d. 1.29

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	5	0.10		
B	36	0.75		
C	4	0.08	.75	.63
D	3	0.06		

Ítem 36. Si el nivel de confianza es 0.95, para un intervalo de confianza para la media poblacional con desviación estándar poblacional desconocida y para un grupo de puntajes distribuido normalmente de tamaño *n = 20*, los valores críticos pudieran ser:

a. -1.65 y 1.65

b. -1.96 y 1.96

c. **-2.093 y 2.093**

d. -2.085 y 2.085

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	5	0.12		
B	12	0.29		
C	25	0.59	.59	1
D	0	0.00		

Ítem 37. La siguiente salida de computadora presenta una muestra simulada de una población normal con $\mu = 130$ y $\sigma = 10$. Luego se usó un comando para establecer un intervalo de confianza del 95% para μ.

118.690	144.226	138.827	125.934	136.198	148.731	133.394
134.184	127.260	145.345	121.966	125.435	141.236	150.021
126.895	118.499	137.225	136.781	119.526	125.628	134.865
116.416	134.312	138.053	140.828			

Mean	Median	TrMean	StDev	95% Confidence Interval for Mu	
132.82	134.31	132.78	9.74	128.798	136.840

Escriba el intervalo de confianza que se obtuvo e interprete el resultado.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Escribe solo correctamente el intervalo.
2: Escribe el intervalo e interpreta correctamente

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	0	0.00		
1	13	0.27	.72	.75
2	35	0.72		

Item 38. La siguiente salida de computadora presenta dos muestras simuladas de dos poblaciones normales. La población 1 con $\mu = 90$ y $\sigma = 10$ y la población 2 con $\mu = 92$ y $\sigma = 10$

Muestra 1:

83.3195	87.6793	86.7831	95.0518	92.9781	86.6457
85.1305	97.5013	83.1112	82.2751	82.7831	90.2786
89.5876	71.2591	82.0282	90.6264		

Muestra 2:

82.312	95.098	92.598	85.959	91.319	108.130
90.392	90.074	78.789	100.923	85.601	89.861
78.685	100.354	81.267	101.432		

	N	Media	D. Típica	Error típico
C1	16	86.69	6.25	1.6
C2	16	90.80	8.70	2.2

95% IC para mC1 - mC2: (-9.6; 1.4)

Escriba el intervalo de confianza que se obtuvo para la diferencia de medias e interprete el resultado.

Este ítem se ha puntuado en una escala 0-1-2 bajo el siguiente criterio:
0: No se responde o se responde en forma incorrecta.
1: Escribe solo correctamente el intervalo.
2: Escribe el intervalo e interpreta correctamente

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
0	0	0.00		
1	10	0.24	.76	.63
2	32	0.76		

Ítem 39. Considere el gráfico siguiente del rendimiento medio de cebada en 1980,1984 y 1988 junto con un intervalo de 95% de confianza

```
Intervalos del 95% para la media

Año         N     Media     StDev  ----+---------+---------+---------
+--
1980        6    184.00      2.61    (----*---)
1984        5    212.40     14.36                        (-----*----)
1988        5    182.40      1.82  (-----*----)
                                  ----+---------+---------+---------+--
Dev. Tipica conjunta =.19            180       195       210       225
```

¿Cuál de las siguientes afirmaciones es verdadera?:

a. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay buena evidencia que las medias de las muestras difieran.

b. La estimación de la media de la población en 1980 es menos precisa que en 1988.

c. Puesto que los intervalos de confianza para 1980 y 1984 no se solapan, hay poca evidencia que las medias de las poblaciones respectivas difieran.

d. **Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay poca evidencia que la medias de las poblaciones difieran.**

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	10	0.21		
B	2	0.04		
C	3	0.06	.68	.75
D	33	0,68		

Ítem 40. Considere el gráfico siguiente del rendimiento medio de cebada en 1990,1994 y 1998 junto con un intervalo de 95% de confianza

```
Individual 95% CIs For Mean

Año         N     Media     StDev --+---------+---------+---------+--
1990       36    238.83     26.57                        (-----*----)
1994       75    186.60      2.51   (-----*-----)
1998       67    186.71      2.63    (----*----)
                                --+---------+---------+---------+----
Dev. Tipica conjunta = 15.49       175       200       225       250
```

Cuál de las siguientes afirmaciones es verdadera

a. **Puesto que los intervalos de confianza para 1994 y 1998 tienen considerable solape, hay poca evidencia que la medias de las poblaciones difieran.**

b. La estimación de la media de la población en 1998 es menos precisa que en 1994

c. Puesto que los intervalos de confianza para 1990 y 1994 no se solapan, no hay evidencia que las medias de las poblaciones respectivas difieran.

d. Puesto que los intervalos de confianza para 1994 y 1998 tienen considerable solape hay buena evidencia que las medias de las muestras difieran.

Opciones	Frecuencia	Porcentaje	Índice de dificultad	Índice de discriminación
A	33	0.78		
B	2	0.05		
C	2	0.05	.78	-.35
D	5	0.12		

ANEXO 3. CUESTIONARIO A

Nombre:_________________________________Matrícula:___________Carrera:______

Género: H__M__

Primera vez que llevo el curso: si___ No____

Preparatoria de Procedencia: Sistema Tec ____ Otra ____

Calificación final en Matemáticas II ________ Nacionalidad ________________

Ítem 1. El intervalo de confianza del 50% para la media de una población μ es:

a. El rango dentro del cuál caen el 50% de los valores de la media de la muestra $\bar{x}$.

b. Un intervalo más ancho que el intervalo de confianza del 95%.

c. Un intervalo de valores calculado a partir de los datos de la muestra. En el 50% de las muestras de una población, el intervalo calculado contiene a la media de la población.

d. Dos veces más ancho que el intervalo de confianza del 100%.

Ítem 2. Comparado a los intervalos de confianza calculados en muestras de tamaño n=4 en una población normal, el ancho de los intervalos de confianza de la media de la población calculado en muestras de tamaño $n = 50$:

a. Variará más que los anchos de los intervalos para muestras de tamaño n =4

b. Variará, pero no tanto como lo hicieron los anchos de los intervalos para muestras de tamaño n =4

c. Tomarán valores parecidos

Ítem 3. Si, manteniendo todos los demás datos fijos, el nivel de confianza se reduce (por ejemplo de 90% a 80%):

a. El intervalo de confianza no cambia.

b. El intervalo de confianza será más ancho

c. El intervalo de confianza será más angosto

d. El cambio en el intervalo de confianza no es predecible

Ítem 4. Explica cómo varía la anchura del intervalo de confianza de la media si, conservando el mismo tamaño de muestra y el mismo coeficiente de confianza, tomamos una población con varianza cuatro veces mayor.

Ítem 5. En un intervalo de confianza del 95% para la media:

a. Si se toman muchas muestras y con cada una se construye el intervalo, la media muestral $\bar{x}$ caerá dentro del intervalo de confianza el 95% de las veces.

b. La probabilidad de que $\bar{x}$ caiga dentro de un intervalo de confianza calculado de una muestra específica es 0.95.

c. Si se toman muchas muestras de igual tamaño, el 95% de los intervalos de confianza calculado contendrían a μ.

Ítem 6. La media muestral de 100 observaciones en una prueba de matemáticas es 75. Encuentre el intervalo de confianza al 95% para la media de la población, asumiendo que σ=7.

a. (61.28, 88.72)

b. (73.63, 76.37)

c. (68, 82)

d. (74.3, 75.7)

Ítem 7. Un fabricante asegura que sus garrafones, contienen un litro de cloro puro. Al tomar una muestra de 16 garrafones se determinó que en promedio contenían 0.94 litros de cloro puro, con desviación estándar de la muestra de 0.097. Construir un intervalo de confianza al 95 %, para el verdadero contenido promedio de litros de cloro puro. No se conoce la desviación típica de la población. (La distribución del contenido de cloro por botella puede considerarse normal).

Ítem 8. Se han obtenido los siguientes datos de emisión diaria de óxidos de azufre, para una muestra de tamaño n=100, media: $\overline{x}$ = 18 y varianza muestral s^2=36. Elabore un intervalo de confianza de 95% para la verdadera emisión diaria promedio de óxidos de azufre.

a. (17.016, 18.984)

b. (16.824, 19.176)

c. (6.24, 19.76)

d. (8.16, 27.84)

Ítem 9. En una muestra aleatoria de 100 rodamientos, 10 tienen un acabado de especificaciones defectuoso. Calcular el intervalo de confianza de 95% para la proporción verdadera de rodamientos defectuosos.

Ítem 10. Sea σ^2 la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es s^2=13700, n=16. Calcular el intervalo de confianza de 95% para σ. Considere la muestra tomada de una población Normal.

Ítem 11. La siguiente tabla contiene un resumen de información sobre la resistencia a la compresión de cubos (N/mm^2) para especimenes de concreto :

	Tamaño muestral	Media muestral
Tipo 1	68	26.99
Tipo 2	72	37.56

Suponga que las desviaciones estándar poblacionales de los grupos son σ_1^2=4.89 y σ_2^2=6.43 respectivamente. Calcule un intervalo de confianza de 95% para la diferencia entre el verdadero promedio de resistencia en el Tipo1 y Tipo 2:

a. (-13.02, -8.12)

b. (-12.437, -8.70)

c. (-31.32, 10.18)

d. (-18.64, -2.5)

Ítem 12. Se compararon dos soluciones de grabado diferentes, usando dos muestras aleatorias de tamaño 10 cada una. Las poblaciones son normales y tienen la misma varianza. Los resultados de la rapidez de grabado fueron:

Solución 1: $\bar{x} = 9.97$ y s = .422   Solución 2: $\bar{x} = 10.4$ y s = .073

$s_p = 0.34$ (desviación típica conjunta). Calcule un intervalo de confianza de 90% para la diferencia de las medias de la rapidez de grabado

Ítem 13. La tabla siguiente resume algunos datos de un experimento realizado para estudiar varias características de tornillos de anclaje:

	Tamaño muestral	Media muestral
Duracell	100	4.1 hrs
Eveready Energizer	100	4.5 hrs

Calcule e interprete un intervalo de confianza de 95% para la diferencia del verdadero promedio de resistencias al corte.

a. (-3.41, -2.58): En el 95% de la muestras del mismo tamaño en esta población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.

b. (-3.41, -2.58): La probabilidad de que la verdadera diferencia promedio de la resistencia al corte cae en el intervalo (-3.41, -2.58) es 0.95.

c. (-3.35, -2.65): En el 95% de la muestras del mismo tamaño en esta población, el intervalo cubre la verdadera diferencia promedio de la resistencia al corte.

d. (-3.35, -2.65): La probabilidad de que la verdadera diferencia promedio de la resistencia al corte caiga en el intervalo (-3.35, -2.65) es 0.95.

Ítem 14. Una compañía quiere seleccionar el proceso de pulido que presente la variabilidad menor. Una muestra aleatoria de n_1=16 piezas del primer proceso da como resultado una desviación estándar muestral $s_1 = 5$ micropulgadas, y una muestra aleatoria de n_2=11 piezas del segundo proceso da como resultado una desviación estándar muestral $s_2 = 4$ micropulgadas. Establezca un intervalo de confianza de 90% para σ_1^2 / σ_2^2, suponiendo que los dos procesos son independientes y que la variable tiene una distribución normal. ¿Cuál de los dos procesos recomendaría?

a. Como todos los valores están dentro del intervalo (.54, 3.96) recomendaría el proceso 1.

b. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría el proceso 2

c. Como el cociente de valores está dentro del intervalo (.54, 3.96) recomendaría cualquiera de los dos.

d. Como todos los valores están dentro del intervalo (.61, 4.44) recomendaría cualquiera de los dos.

Ítem 15. La distribución muestral utilizada en la construcción de intervalos de confianza para la varianza en muestras pequeñas, tomadas de una población normal, es:

a. Distribución t de Student.

b. Distribución Ji-cuadrada.

c. Distribución Normal.

d. Distribución F.

Ítem 16. El nivel de confianza es 0.95, para un intervalo de confianza para la media poblacional con desviación estándar poblacional desconocida para un grupo de puntajes distribuido normalmente de tamaño $n = 20$. Los valores críticos han de ser:

a. -1.65 y 1.65

b. -1.96 y 1.96

c. -2.093 y 2.093

d. -2.085 y 2.085

Ítem 17 La siguiente salida de computadora presenta dos muestras simuladas de dos poblaciones normales. La población 1 con $\mu = 90$ y $\sigma = 10$ y la población 2 con $\mu = 92$ y $\sigma = 10$

Muestra 1: 83.31 87.67 86.78 95.05 92.97 86.64 85.13 97.50 83.11 82.27 82.78 90.27 89.58 71.25 82.02 90.62

Muestra 2: 82.31 95.09 92.59 85.95 91.31 108.13 90.39 90.07 78.78 100.92 85.60 89.86 78.68 100.35 81.26 101.43

	N	Media	D. Tipica	Error tipico
C1	16	86.69	6.25	1.6
C2	16	90.80	8.70	2.2

95% IC para mC1 - mC2: (-9.6; 1.4)

Escriba el intervalo de confianza para la diferencia de medias e interprete el resultado.

Ítem 18. Considere el gráfico siguiente del rendimiento medio de cebada en 1980,1984 y 1988 junto con un intervalo de 95% de confianza respectivos. ¿Cuál de las siguientes afirmaciones es verdadera?

```
Intervalos del 95% para la media

Año N Media StDev       ----+---------+---------+---------+--
1980 6 184.00 2.61        (----*---)
1984 5 212.40 14.36                           (-----*----)
1988 5 182.40 1.82      (-----*----)
                        ----+---------+---------+---------+--
Desv. Típica conjunta =.19 180       195       210       225
```

a. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay buena evidencia que las medias de las muestras difieran.

b. La estimación de la media de la población en 1980 es menos precisa que en 1988.

c. Puesto que los intervalos de confianza para 1980 y 1984 no se solapan, hay poca evidencia que las medias de las poblaciones respectivas difieran.

d. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay poca evidencia que las medias de las poblaciones difieran.