

Manual práctico de SPSS para el análisis estadístico de datos biológicos

*Guía de procedimientos estadísticos aplicados en SPSS para consolidar
la formación práctica de la asignatura de Bioestadística del Grado en Biología*



José Antonio Sáez Muñoz

*Departamento de Estadística e Investigación Operativa, Universidad de Granada,
Fuente Nueva s/n, 18071 Granada, España (e-mail: joseasaezm@ugr.es)*

Manual práctico de SPSS para el análisis estadístico de datos biológicos

Guía de procedimientos estadísticos aplicados en SPSS para consolidar la formación práctica de la asignatura de Bioestadística del Grado en Biología

José Antonio Sáez Muñoz

Departamento de Estadística e Investigación Operativa
Universidad de Granada

Material elaborado con fines educativos para la formación práctica de la asignatura de Bioestadística del Grado en Biología de la Universidad de Granada.

Créditos y derechos

Título: Manual práctico de SPSS para el análisis estadístico de datos biológicos

Autor: José Antonio Sáez Muñoz (joseasaezm@ugr.es)

Dpto. Estadística e Investigación Operativa, Universidad de Granada,
Fuente Nueva s/n, 18071 Granada, España.

Año: © 2026 José Antonio Sáez Muñoz

Versión: v2.0

Licencia: Publicado bajo la licencia *Creative Commons Attribution 4.0 International* (CC BY 4.0), que permite su uso, copia, distribución y adaptación siempre que se cite correctamente al autor. Más información en <https://creativecommons.org/licenses/by/4.0/>

Finalidad educativa: Este material docente ha sido elaborado con fines exclusivamente educativos para la formación práctica de la asignatura de Bioestadística del Grado en Biología de la Universidad de Granada.

Citación recomendada: Sáez, J. A. (2026). *Manual práctico de SPSS para el análisis estadístico de datos biológicos*. Universidad de Granada, Granada, España.

Prólogo

La enseñanza de la Bioestadística en los estudios de Biología requiere integrar, de forma equilibrada, los fundamentos teóricos con la aplicación rigurosa de las técnicas de análisis. La creciente disponibilidad de datos en las ciencias biológicas y afines ha reforzado la necesidad de que el alumnado desarrolle competencias prácticas que permitan abordar con solvencia los problemas cuantitativos propios de la disciplina. En este contexto, el uso de herramientas informáticas especializadas constituye un elemento fundamental del proceso formativo.

El presente manual se ha elaborado con la finalidad de proporcionar un recurso claro, estructurado y orientado a la práctica. Su propósito es facilitar el aprendizaje autónomo del manejo de SPSS para la realización de procedimientos estadísticos esenciales en el análisis de datos biológicos. El contenido complementa la formación teórica impartida en las clases magistrales mediante una guía detallada de operaciones, configuraciones y comandos que permiten trasladar los conceptos estadísticos a contextos reales de trabajo.

La organización del manual sigue la progresión lógica de los contenidos propios de la Bioestadística. Cada capítulo introduce los procedimientos necesarios para abordar tareas concretas, desde la gestión de datos hasta la realización de análisis descriptivos, modelos de regresión, cálculos de probabilidades, intervalos de confianza y contrastes de hipótesis. La presentación paso a paso y el empleo de ejemplos vinculados con problemas biológicos facilitan que el alumnado desarrolle una comprensión operativa de los métodos y consolide competencias aplicadas.

Este recurso ha sido concebido tanto para su utilización durante la formación práctica de la asignatura como para su consulta independiente. Su estructura favorece un aprendizaje guiado y riguroso, ofreciendo al alumnado una herramienta actualizada para el trabajo cotidiano con SPSS y al profesorado un apoyo sólido para el desarrollo de las actividades prácticas que forman parte integral de la enseñanza de la Bioestadística.

José Antonio Sáez Muñoz
Departamento de Estadística e Investigación Operativa
Universidad de Granada
2026

Índice

Prólogo	i
1 Introducción a SPSS	1
1.1 Interfaz	2
1.2 Entrada de datos	3
1.3 Guardado de archivos	5
1.4 Cálculo de variables derivadas	5
1.5 Recodificación de variables	6
1.6 Selección de casos	7
1.7 Segmentación de archivos	9
1.8 Fusión de archivos	10
1.9 Resumen de procedimientos	11
1.10 Ejercicios	12
2 Estadística descriptiva	13
2.1 Medidas descriptivas	14
2.2 Gráficos	15
2.3 Datos con frecuencias ponderadas	18
2.4 Resumen de procedimientos	20
2.5 Ejercicios	21
3 Regresión	23
3.1 Diagrama de dispersión	24
3.2 Cálculo de la recta de regresión	26
3.3 Predicción de nuevos valores	27
3.4 Medidas de relación y ajuste	28
3.5 Otros modelos de regresión	29
3.6 Resumen de procedimientos	30
3.7 Ejercicios	31
4 Modelos de probabilidad	33
4.1 Cálculo de probabilidades	34
4.2 Modelo binomial	36
4.3 Modelo de Poisson	37
4.4 Modelo normal	38
4.5 Percentiles	39
4.6 Resumen de procedimientos	40

4.7	Ejercicios	41
5	Intervalos de confianza	43
5.1	IC para la media poblacional (una muestra)	44
5.2	IC para la diferencia de medias (muestras relacionadas)	45
5.3	IC para la diferencia de medias (muestras independientes)	47
5.4	Resumen de procedimientos	51
5.5	Ejercicios	52
6	Contrastes de hipótesis sobre medias poblacionales	55
6.1	Formulación e interpretación	56
6.2	CH sobre la media poblacional (una muestra)	57
6.3	CH sobre la diferencia de medias (muestras relacionadas)	59
6.4	CH sobre la diferencia de medias (muestras independientes)	61
6.5	Resumen de procedimientos	64
6.6	Ejercicios	65
7	Otros contrastes de hipótesis	67
7.1	CH para la normalidad	68
7.2	CH para la pendiente en regresión lineal	70
7.3	CH para la proporción poblacional	71
7.4	Resumen de procedimientos	72
7.5	Ejercicios	73
	Bibliografía	75

Capítulo 1

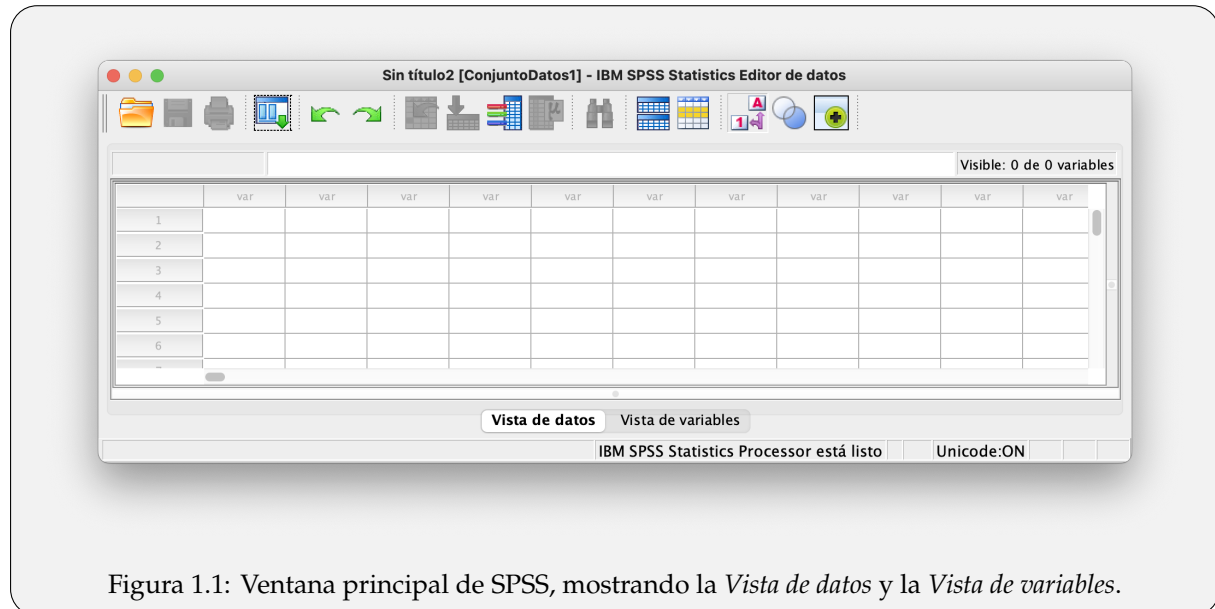
Introducción a SPSS



SPSS (*Statistical Package for the Social Sciences*) es un software estadístico ampliamente utilizado en diversas áreas de investigación, incluida la biología. Su facilidad de uso y sus potentes herramientas para la gestión y análisis de datos permiten manejar grandes volúmenes de información, realizar análisis descriptivos y estadísticos, y generar resultados que apoyan la toma de decisiones. En biología, SPSS se aplica para explorar patrones en datos experimentales, comparar grupos, modelar relaciones entre variables y evaluar tratamientos en áreas como genética, ecología o biomedicina. Este capítulo tiene como objetivo familiarizar al lector con las funciones básicas de SPSS, abarcando la exploración de la interfaz, la entrada y el guardado de datos, el cálculo de variables derivadas, la recodificación de variables, la selección y segmentación de casos, así como la fusión de conjuntos de datos, sentando las bases para análisis más complejos en capítulos posteriores.

1.1 Interfaz

Al abrir SPSS (versión 25, Figura 1.1), se presentan dos vistas principales: *Vista de datos* y *Vista de variables*.



Ambas vistas son fundamentales para trabajar con datos en SPSS, y cada una cumple un propósito específico que facilita el manejo y la organización de los datos:

1. **Vista de datos.** Esta vista permite visualizar y manipular directamente los datos introducidos. Los datos se organizan en filas y columnas, donde cada fila corresponde a un caso (un individuo o una unidad de observación) y cada columna corresponde a una variable (una característica medida o una variable que se estudia). La *Vista de datos* es muy similar a una hoja de cálculo de Excel, pero está optimizada para la gestión y análisis de datos estadísticos. En esta vista es posible introducir, editar o eliminar datos en las celdas correspondientes, visualizar las observaciones de cada variable y caso, y realizar operaciones básicas como copiar, pegar y ordenar los datos. Es importante tener en cuenta que los datos en esta vista están organizados en forma tabular, y es aquí donde se trabajan los valores numéricos o categóricos que se van a analizar.
2. **Vista de variables.** Esta vista permite ver y modificar las propiedades de las variables que componen los datos. En lugar de mostrar los datos, presenta información sobre cada variable, como su nombre, tipo y otras configuraciones relevantes para el análisis. Desde la *Vista de variables* se pueden definir nombres y tipos (numérico, cadena, fecha, etc.), asignar etiquetas descriptivas a variables y valores, y configurar valores perdidos o categorías específicas. Esta vista resulta esencial para organizar y preparar los datos antes de cualquier análisis estadístico, permitiendo adaptar las variables a las necesidades de la investigación.

Ambas vistas son complementarias y, al trabajar con ellas, se puede gestionar de manera eficiente tanto los datos como las características de las variables, garantizando que el análisis estadístico sea riguroso y pertinente al planteamiento del estudio.

1.2 Entrada de datos

Para introducir datos en SPSS, el proceso consta de dos pasos principales:

1. **Definir las variables (Vista de variables).** Antes de introducir los datos, es necesario definir las variables que se van a utilizar en el estudio. Para ello, se debe acceder a la *Vista de variables*, donde se debe rellenar información sobre cada una de ellas.

Imaginemos un estudio biológico sobre el crecimiento y desarrollo de individuos en una población determinada. A continuación se presenta un ejemplo con las variables **Nombre**, **Sexo**, **Peso** y **Altura**, junto con los campos más relevantes que deben completarse (Figura 1.2):

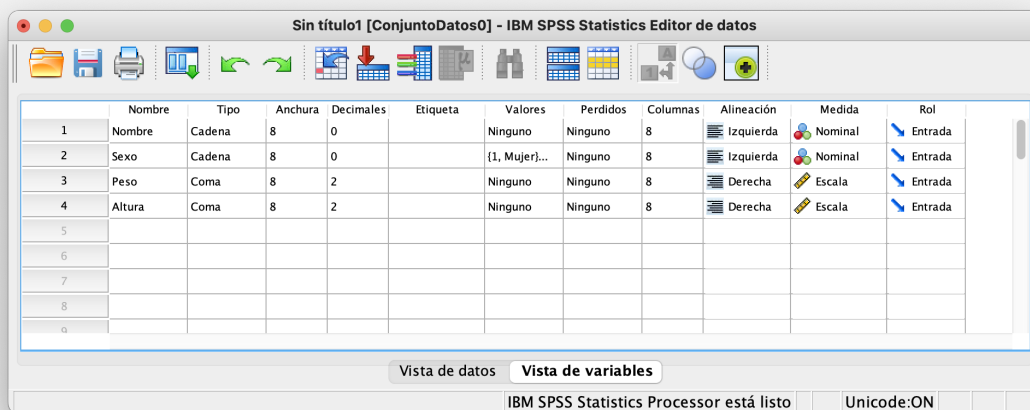


Figura 1.2: Vista de variables de SPSS.

- **Nombre.** El nombre de la variable debe ser único y descriptivo. En este caso, se utilizarán nombres como **Nombre**, **Sexo**, **Peso** y **Altura**. Es importante que el nombre de la variable sea una única palabra, sin espacios (por ejemplo, no usar **Peso corporal** o **Altura cm**). Además, se recomienda utilizar nombres cortos pero claros.
- **Tipo.** El tipo de la variable determina qué datos se almacenarán en esa columna. Para variables categóricas como **Sexo**, se puede seleccionar **Cadena** o, alternativamente, un tipo numérico y definir los valores y etiquetas correspondientes (por ejemplo, 1 = **Mujer**, 2 = **Hombre**). Para variables numéricas como **Peso** y **Altura**, se puede seleccionar un tipo numérico, como **Numérico** o **Coma**. Este último permite utilizar el punto (".") como separador decimal.
- **Decimales.** Este campo especifica el número de decimales que se utilizarán en una variable numérica. Por ejemplo, si se desea registrar el peso con dos decimales (como 65.50 kg), se debe configurar el número de decimales en 2. Para variables categóricas, este campo generalmente no se utiliza.
- **Valores.** Para las variables categóricas, como **Sexo**, es necesario especificar los valores posibles. Para ello, se debe acceder a **Valores**, donde se pueden introducir los valores y sus etiquetas asociadas. Por ejemplo, para **Sexo**, se puede asignar el valor 1 a **Mujer** y el valor 2 a **Hombre** (Figura 1.3).

Para asignar etiquetas a los valores de la variable, hacer clic en el botón Valores. En la ventana que aparece, introducir 1 en el campo Valor y Mujer en el campo Etiqueta, y hacer clic en Añadir. A continuación, introducir 2 en el campo Valor y Hombre en el campo Etiqueta, y hacer clic nuevamente en Añadir.

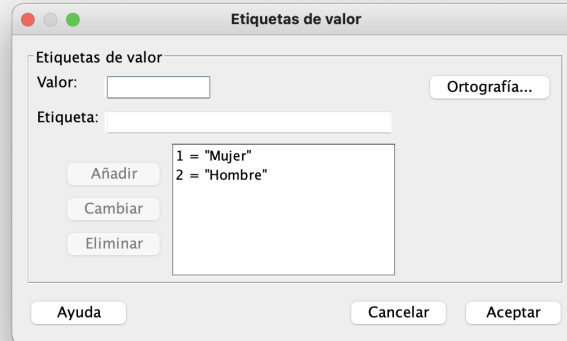


Figura 1.3: Configuración de valores para la variable categórica Sexo en SPSS.

- **Medida.** Este campo se utiliza para indicar el tipo de escala de la variable. Para variables categóricas, se puede seleccionar entre Nominal u Ordinal, dependiendo de la naturaleza de la variable. Por ejemplo, para Sexo, que es una variable categórica sin un orden específico, se selecciona Nominal. En cambio, para variables numéricas como Peso o Altura, que tienen una escala continua, se debe seleccionar Escala.

2. **Introducir los datos (Vista de datos).** Con las variables ya definidas, se puede pasar a la *Vista de datos*. En esta vista, cada fila corresponde a un caso o registro, y cada columna corresponde a una variable del estudio. En la *Vista de datos* se introducen los valores numéricos o categóricos para cada caso (Figura 1.4). Por ejemplo, para una persona con el nombre Ana, sexo Mujer, peso 60.21 kg y altura 1.60 m, se introducirán los siguientes datos en las celdas correspondientes: en la columna Nombre, Ana; en la columna Sexo, 1 (correspondiente a Mujer); en la columna Peso, 60.21; y en la columna Altura, 1.60. De este modo, se continuará introduciendo los datos para cada caso, completando la base de datos. Al pulsar el botón , se puede verificar la correspondencia entre los valores numéricos y sus etiquetas definidas.

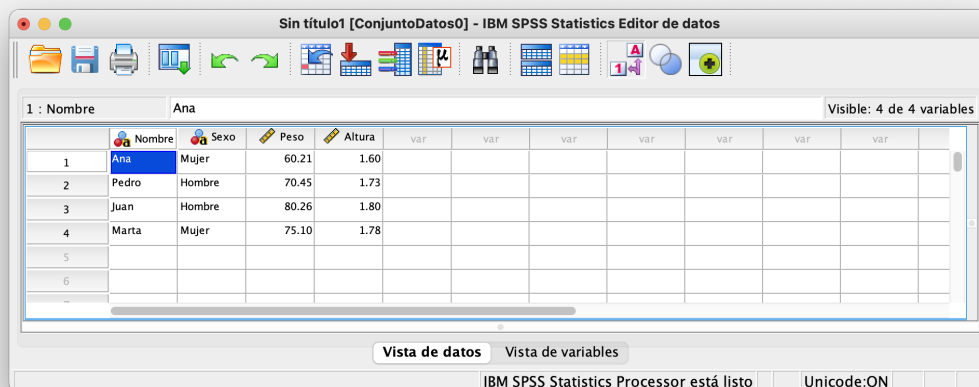


Figura 1.4: Vista de datos en SPSS, donde se introducen los datos para cada caso en las columnas correspondientes a las variables.

1.3 Guardado de archivos

Es fundamental guardar el archivo de datos para evitar la pérdida de información. Para ello, seleccionar en el menú principal de SPSS la opción:

Archivo > Guardar

En la ventana que aparece, elegir el formato de archivo deseado. El formato más común y recomendado para guardar los datos en SPSS es `.sav`, ya que se trata del formato nativo del programa y conserva toda la estructura y propiedades del archivo, incluyendo las definiciones de variables y los valores asignados. Al guardar el archivo, seleccionar la ubicación en el sistema donde se desea almacenarlo y asignarle un nombre adecuado que facilite su identificación. Es importante asegurarse de que el archivo se guarde correctamente para poder continuar con los análisis sin riesgo de perder los datos introducidos. A continuación, se muestra la ventana de *Guardar* de SPSS (Figura 1.5):

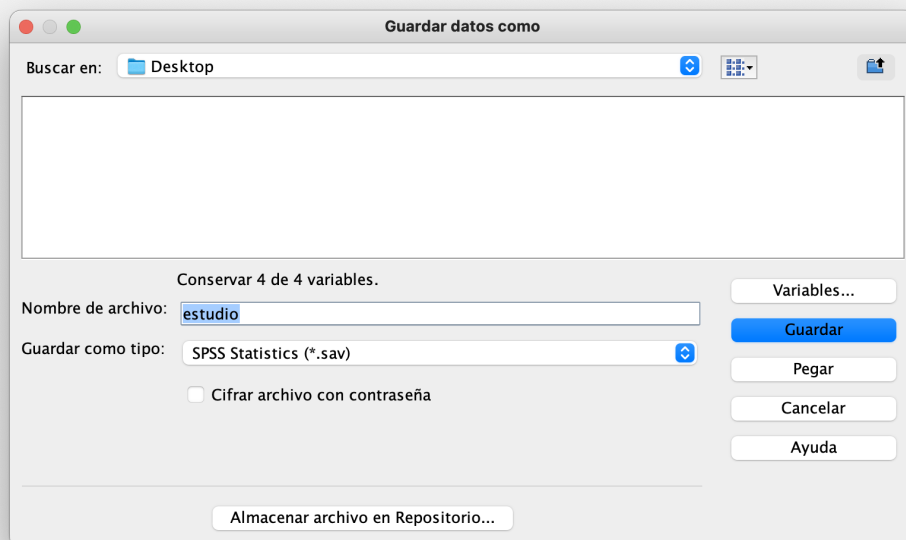


Figura 1.5: Ventana de *Guardar* en SPSS para guardar el archivo en formato `.sav`.

1.4 Cálculo de variables derivadas

En numerosas ocasiones, resulta necesario crear nuevas variables a partir de cálculos realizados con otras variables existentes. Para ello, utilizar la opción:

Transformar > Calcular variable

A continuación, se describen los pasos a seguir (Figura 1.6):

- En el campo *Variable objetivo*, escribir el nombre de la nueva variable. Por ejemplo, IMC.
- En el campo *Expresión numérica*, introducir la fórmula para calcular la nueva variable. Por ejemplo, para calcular el índice de masa corporal (IMC), se puede utilizar la expresión: Peso/Al^2

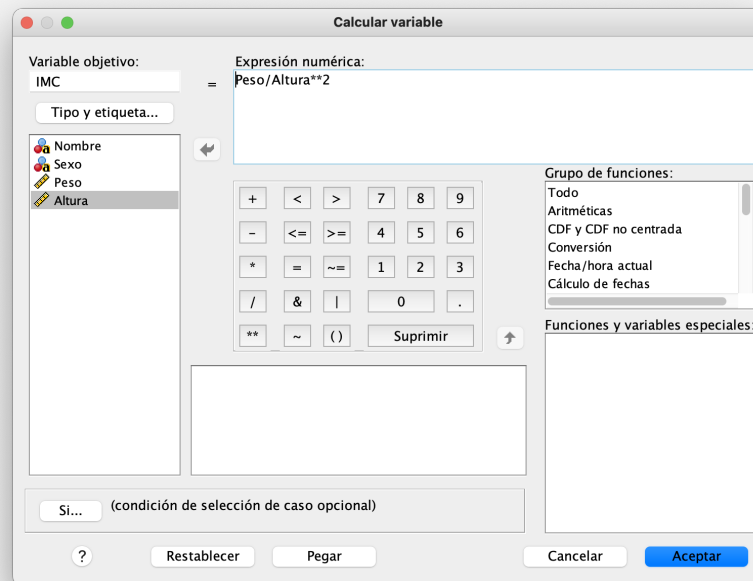


Figura 1.6: Ventana de SPSS para calcular una nueva variable a partir de otras existentes.

1.5 Recodificación de variables

La recodificación de variables permite transformar los valores de una variable según criterios previamente definidos. Para recodificar una variable en SPSS, se utiliza la opción:

Transformar > Recodificar en distintas variables

Esta opción crea una nueva variable a partir de la existente, sin modificar la original. Por ejemplo, para agrupar los valores de la variable *Altura* en categorías (Figura 1.7):

- Elegir la variable de origen, en este caso *Altura*, y moverla al cuadro central.
- Asignar un nombre a la nueva variable, por ejemplo *GrupoAltura*, y hacer clic en *Cambiar*.

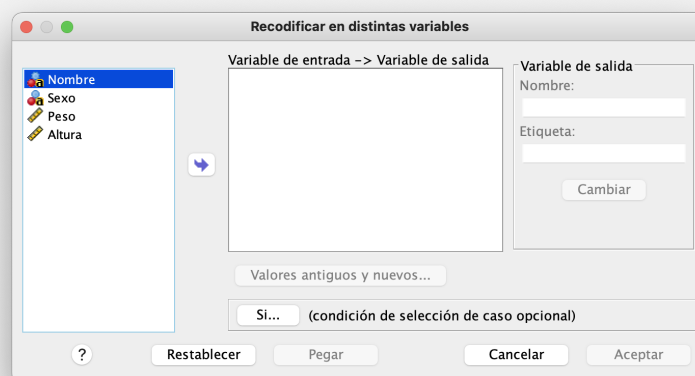


Figura 1.7: Ventana de recodificación de variables en SPSS.

A continuación, se deben definir las reglas que determinarán cómo se generará la nueva variable a partir de la original. Para ello, seleccionar la opción **Valores antiguos y nuevos** (Figura 1.8):

- Asignar los valores antiguos a sus nuevos valores. Por ejemplo, para agrupar alturas, se puede definir que los valores menores o iguales a 1.65 se asignen al valor 1, los valores en el intervalo (1.65, 1.75] se asignen al valor 2, y los valores mayores a 1.75 se asignen al valor 3.
- Tras introducir cada regla, hacer clic en **Añadir** y, finalmente, en **Aceptar**. Como resultado, se crea una nueva columna en el archivo de datos con los grupos definidos en la variable **GrupoAltura**.

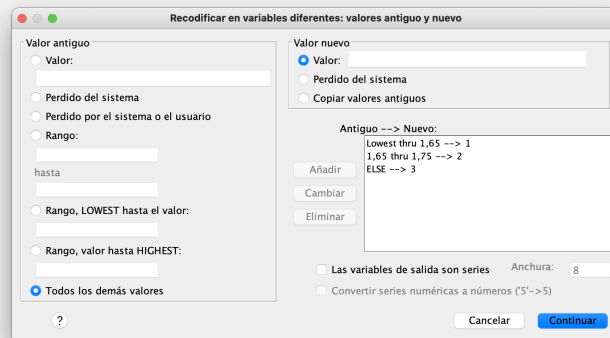


Figura 1.8: Definición de las reglas de recodificación en SPSS.

1.6 Selección de casos

SPSS permite seleccionar casos para realizar análisis sobre subconjuntos de datos. Esta función facilita la aplicación de criterios específicos, como seleccionar únicamente aquellos casos en los que el valor de una variable supere un determinado umbral. Para acceder a esta opción (Figura 1.9), seleccionar:

Datos > Seleccionar casos

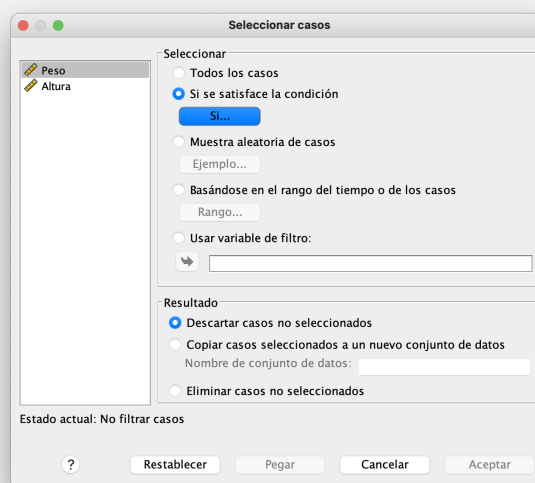


Figura 1.9: Ventana de selección de casos en SPSS.

En esta ventana, elegir la opción *Si se satisface la condición* (Si...), lo que abrirá una nueva ventana para especificar el criterio de selección. Por ejemplo, se puede definir una condición para seleccionar únicamente aquellos casos cuya variable *Altura* sea mayor a 1.7 (Figura 1.10):

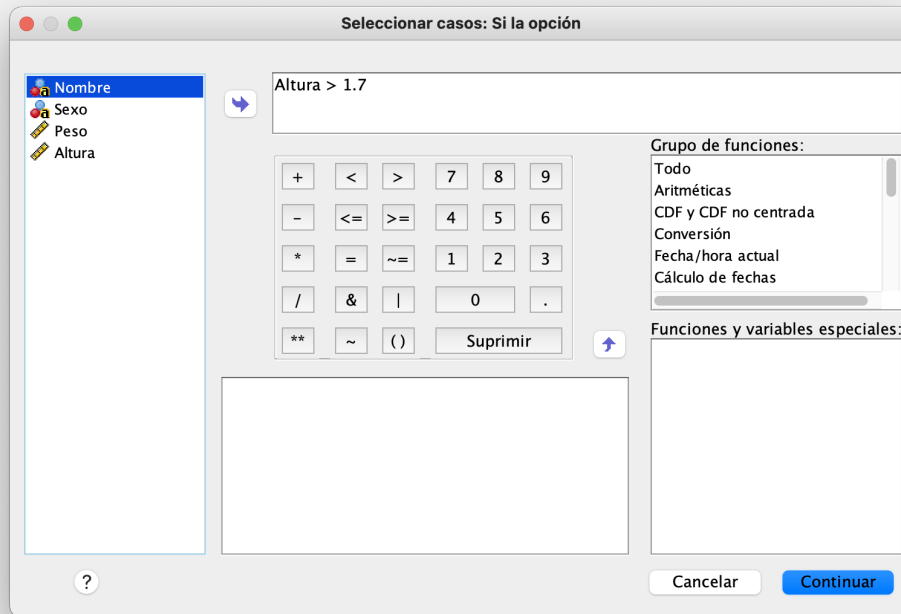


Figura 1.10: Definición de una condición para seleccionar casos en SPSS.

Una vez definida la condición (por ejemplo, *Altura > 1.7*), SPSS selecciona únicamente los casos que cumplen el criterio para los análisis posteriores (Figura 1.11). Para volver a trabajar con todos los casos, seleccionar nuevamente *Datos > Seleccionar casos* y elegir la opción de incluir todos los casos.

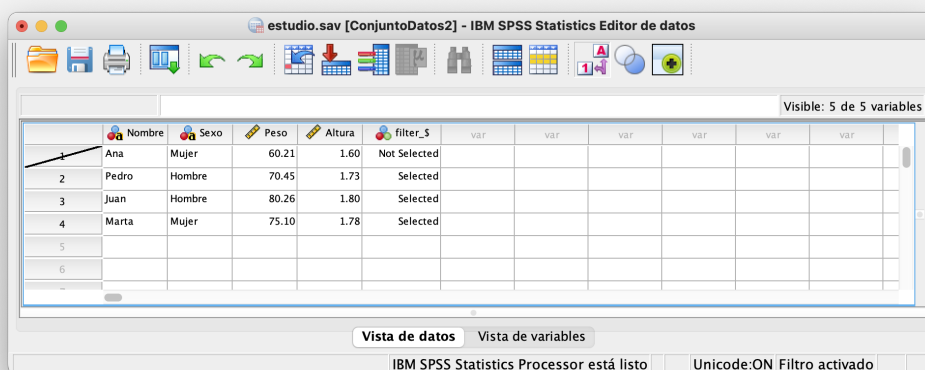


Figura 1.11: Base de datos con los casos seleccionados.

1.7 Segmentación de archivos

En determinadas situaciones, puede resultar útil dividir un archivo para realizar análisis separados por categorías, por ejemplo, según grupos de tratamiento. Esta opción se encuentra en:

Datos > Segmentar archivo

Al seleccionar esta opción, elegir la variable por la cual se desea segmentar el archivo (Figura 1.12).

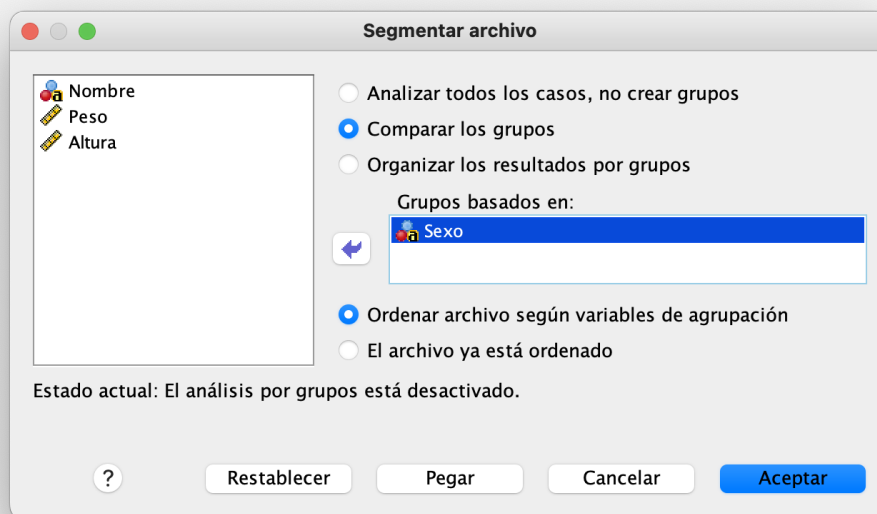


Figura 1.12: Ventana de segmentación de archivo en SPSS.

Existen varias alternativas al segmentar un archivo, según los objetivos del análisis:

- **Analizar todos los casos (no crear grupos):** Permite realizar un análisis sobre todos los casos de la base de datos sin dividirlos en grupos. Resulta útil cuando se desea examinar la base de datos completa sin diferenciar entre categorías.
- **Comparar los grupos:** Con esta opción, SPSS genera el análisis para todos los grupos dentro de una misma tabla o conjunto de resultados, facilitando la comparación directa. Es especialmente útil para observar diferencias entre categorías, como por ejemplo entre dos tratamientos en un estudio experimental.
- **Organizar los resultados por grupos:** Esta opción produce análisis separados para cada grupo definido por la variable seleccionada. SPSS genera tablas independientes para cada categoría, permitiendo examinar los resultados de manera más detallada y específica.

1.8 Fusión de archivos

Cuando sea necesario combinar datos procedentes de diferentes ficheros, se pueden utilizar una de las siguientes opciones:

Datos > Fusionar archivos > Añadir casos
Datos > Fusionar archivos > Añadir variables

Si se desea añadir nuevos casos (filas) al fichero actual, seleccionar *Añadir casos*. Esto permitirá incorporar más observaciones, como nuevos individuos o unidades de estudio, a la base de datos. En cambio, si se desea agregar nuevas variables (columnas) al fichero actual, seleccionar *Añadir variables*. Esta opción resulta útil cuando se dispone de un fichero con nuevas variables correspondientes a los mismos casos. Por ejemplo, al añadir nuevos casos, aparece la siguiente ventana (Figura 1.13):

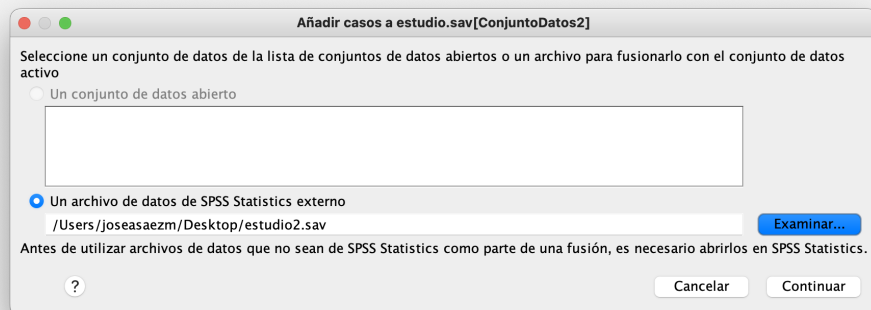


Figura 1.13: Ventana de SPSS para agregar nuevos casos al fichero.

Tras seleccionar el nuevo fichero, se mostrarán las variables que coinciden y las que no, como se aprecia en la Figura 1.14. Al realizar una fusión de ficheros, es fundamental comprobar que las variables y los casos coincidan correctamente entre los ficheros, con el fin de evitar errores en los análisis posteriores.

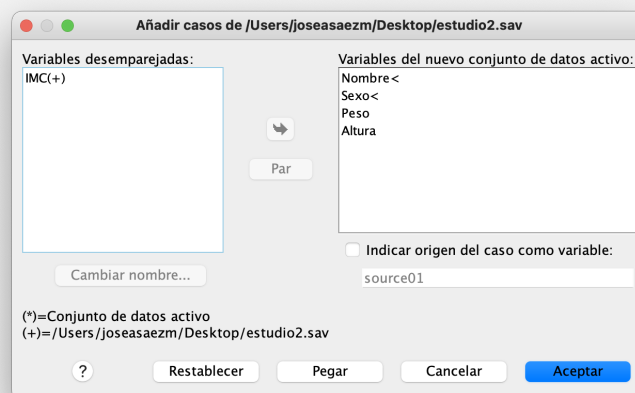


Figura 1.14: Ventana de SPSS que muestra las variables coincidentes y las que no al fusionar ficheros.

1.9 Resumen de procedimientos

Capítulo 1: Introducción a SPSS

1. **Introducción de datos: Vista de variables y Vista de datos.** En SPSS, la base de datos se construye a partir de dos vistas principales:

- *Vista de variables:* Para definir las características de las variables (nombre, tipo, etiquetas, etc.).
- *Vista de datos:* Para introducir y visualizar los valores de cada caso.

2. **Guardar un fichero.** Para guardar el archivo de datos en formato .sav:

Archivo > Guardar

3. **Calcular una nueva variable a partir de otras.** Para crear una variable calculada, como el *Índice de Masa Corporal* (IMC):

Transformar > Calcular variable

4. **Recodificar variables.** Para transformar los valores de una variable en categorías o grupos:

Transformar > Recodificar en distintas variables

5. **Seleccionar casos.** Para trabajar solo con un subconjunto de datos que cumplan ciertas condiciones:

Datos > Seleccionar casos

6. **Segmentar el archivo.** Para dividir el archivo y organizar los resultados por grupos basados en una variable categórica:

Datos > Segmentar archivo

7. **Fusionar archivos de datos.** Para combinar dos archivos añadiendo casos o variables:

Datos > Fusionar archivos > Añadir casos
Datos > Fusionar archivos > Añadir variables

1.10 Ejercicios

Ejercicio 1

Se desea analizar el comportamiento de una población de peces en un acuario experimental. Se han registrado datos de 10 individuos sobre sexo, longitud (cm), peso (g) y número de interacciones sociales observadas durante 1 hora. Se han obtenido los siguientes datos:

Sexo	Longitud	Peso	Interacciones
Hembra	5.2	10.1	12
Macho	6.0	12.3	15
Hembra	5.5	11.0	9
Macho	6.3	13.2	14
Hembra	5.0	9.8	11
Macho	6.1	12.8	16
Hembra	5.3	10.5	10
Macho	6.2	13.0	13
Hembra	5.1	10.0	12
Macho	6.4	13.5	17

A partir de estos datos, se pide realizar las siguientes tareas en SPSS:

- Definir las variables en la *Vista de variables*.
- Introducir los datos en la *Vista de datos*.
- Guardar el fichero en formato *.sav*.
- Calcular una nueva variable *IndiceCorporal* utilizando la fórmula:

$$\text{IndiceCorporal} = \frac{\text{Peso}}{\text{Longitud}^2}$$

- Recodificar la variable *Interacciones* en la nueva variable *NivelSocial* de acuerdo con los siguientes criterios:
 - $\text{Interacciones} \leq 11$: Nivel 1
 - $12 \leq \text{Interacciones} \leq 14$: Nivel 2
 - $\text{Interacciones} > 14$: Nivel 3
- Seleccionar los peces machos para analizar sus interacciones sociales.
- Segmentar el archivo por *NivelSocial* para comparar el comportamiento entre los diferentes niveles de interacción.

Capítulo 2

Estadística descriptiva



La *estadística descriptiva* es una rama fundamental de la estadística que permite resumir y presentar los datos de forma clara. En biología, ayuda a identificar patrones, resaltar tendencias y comprender las características principales de los datos. SPSS ofrece herramientas para calcular medidas descriptivas como media, mediana, moda, rango, desviación típica y percentiles, así como para generar representaciones gráficas como histogramas, diagramas de caja y gráficos de barras. También permite construir tablas de frecuencias para analizar la distribución de los datos. El objetivo de este capítulo es proporcionar una comprensión básica del uso de SPSS en estadística descriptiva, base esencial para la correcta interpretación de datos biológicos.

2.1 Medidas descriptivas

Las medidas descriptivas permiten caracterizar una variable y ofrecer una visión general de su distribución. Entre las más empleadas se encuentran la media, la mediana, la moda, la desviación típica y el rango, entre otras. Para ilustrar el cálculo de estas medidas, se utilizarán los datos introducidos en el capítulo anterior, correspondientes a un estudio biológico centrado en el desarrollo de individuos de una población determinada, que se presentan a continuación:

Nombre	Sexo	Peso (kg)	Altura (m)
Ana	Mujer	60.21	1.60
Pedro	Hombre	70.45	1.73
Juan	Hombre	80.26	1.80
Marta	Mujer	75.10	1.78

Para calcular las medidas descriptivas en SPSS, acceder al menú:

Analizar > Estadísticos descriptivos > Frecuencias

A continuación, seguir los pasos indicados:

1. Seleccionar las variables a analizar y moverlas al cuadro *Variables* (Figura 2.1).

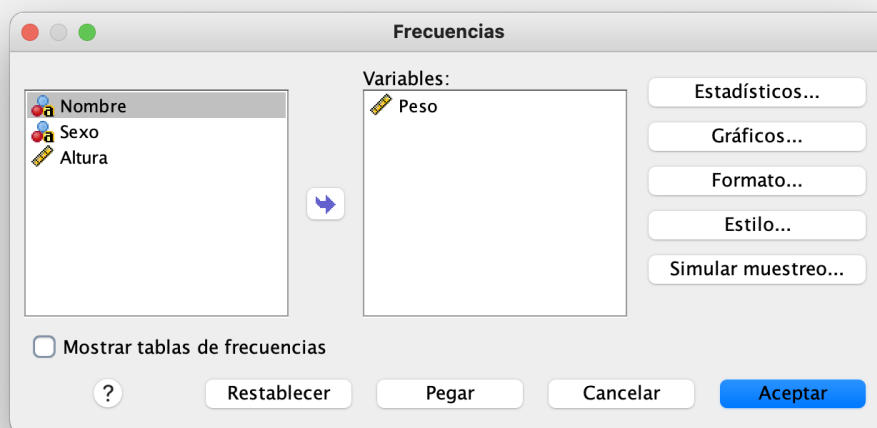


Figura 2.1: Ventana de selección de variables para el cálculo de medidas descriptivas.

2. Para obtener una tabla de frecuencias, se puede marcar la opción **Mostrar tabla de frecuencias**, que generará frecuencias absolutas, relativas y acumuladas para cada variable seleccionada.
3. Para calcular las medidas descriptivas, hacer clic en **Estadísticos**. En la ventana emergente (Figura 2.2), seleccionar las medidas de interés, como la media, la mediana, la desviación típica o los percentiles. En este ejemplo, se seleccionan la media, la desviación típica y el percentil 75.

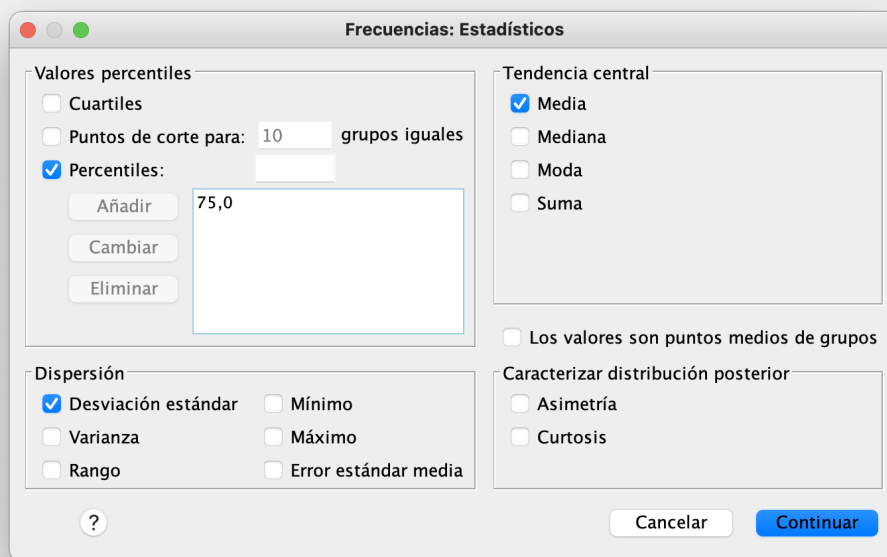


Figura 2.2: Selección de medidas descriptivas en SPSS.

Se generará una tabla titulada *Estadísticos* con las medidas seleccionadas para cada variable (Figura 2.3). Este procedimiento permite obtener de manera rápida las principales medidas descriptivas para caracterizar las variables y analizar su distribución.

Estadísticos		
Peso		
N	Válido	4
	Perdidos	0
Media		71.5050
Desv. Desviación		8.52964
Percentiles	75	78.9700

Figura 2.3: Tabla de resultados con las medidas descriptivas.

2.2 Gráficos

Los gráficos son herramientas esenciales para visualizar la distribución de los datos y facilitar su interpretación. SPSS ofrece diferentes tipos de representaciones que permiten mostrar la información de forma clara y estructurada. A continuación, se describen los procedimientos para generar algunos de los gráficos más utilizados:

- Gráficos de barras.
- Gráficos de sectores.
- Histogramas.
- Diagramas de cajas.

Gráfico de barras, gráfico de sectores e histograma

Estos gráficos pueden generarse a partir del mismo menú utilizado para obtener medidas descriptivas:

Analizar > Estadísticos descriptivos > Frecuencias

El procedimiento es el siguiente:

1. En la ventana que aparece (Figura 2.4), seleccionar las variables que se desean representar y moverlas al cuadro *Variables*.

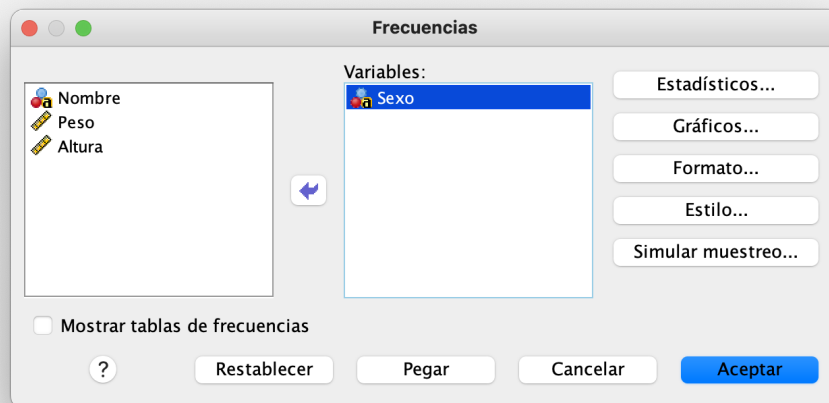


Figura 2.4: Ventana de selección de variables para generar gráficos.

2. Hacer clic en el botón *Gráficos*.
3. Elegir el tipo de gráfico: barras, sectores o histograma (Figura 2.5).

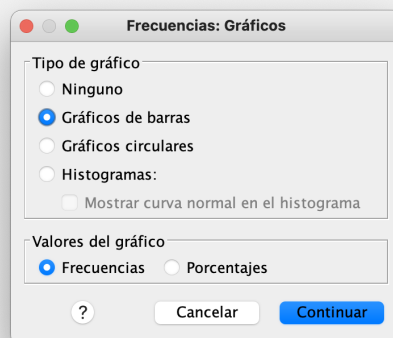


Figura 2.5: Opciones de gráficos disponibles en el cuadro *Frecuencias*.

4. Hacer clic en *Continuar* y, finalmente, en *Aceptar*.

SPSS generará el gráfico seleccionado, lo que permitirá visualizar la distribución de la variable de manera rápida y eficaz.

Diagrama de cajas

El diagrama de cajas es una herramienta útil para examinar la dispersión, la simetría y la presencia de valores atípicos en los datos. Para crear este gráfico en SPSS, se debe acceder al siguiente menú:

Gráficos > Cuadros de diálogo antiguos > Diagramas de cajas

El procedimiento es el siguiente:

1. En la ventana inicial (Figura 2.6), mantener seleccionada la opción **Simples** y elegir **Resúmenes para distintas variables** para distintas variables. Pulsar en **Definir**.

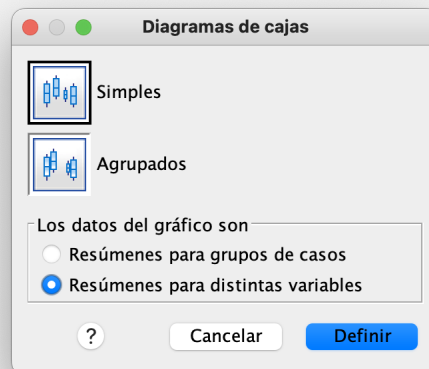


Figura 2.6: Opciones iniciales para generar un diagrama de cajas.

2. En la siguiente ventana (Figura 2.7), seleccionar la variable que se desea representar en el campo *Las cajas representan*.

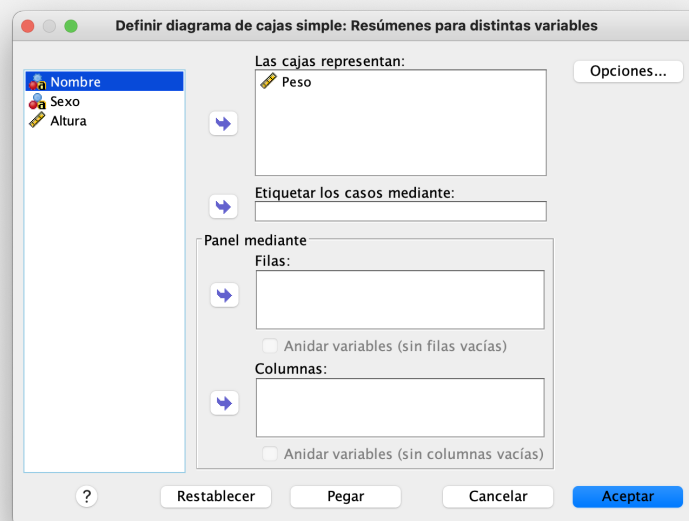


Figura 2.7: Selección de la variable para el diagrama de cajas.

3. Hacer clic en **Aceptar**. SPSS generará un diagrama de cajas (Figura 2.8) que mostrará la mediana, los cuartiles, la dispersión de los datos y la presencia de posibles valores atípicos.

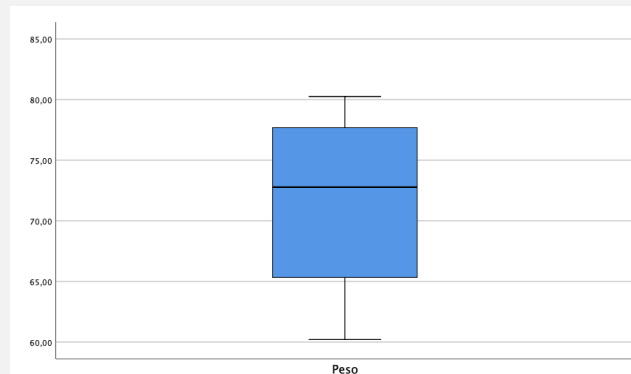


Figura 2.8: Diagrama de cajas generado en SPSS.

2.3 Datos con frecuencias ponderadas

Cuando los datos incluyen una columna que indica la frecuencia de cada caso, es necesario informar a SPSS para que cada fila se interprete como si apareciera tantas veces como señala la frecuencia correspondiente. De esta manera, los análisis reflejarán correctamente la distribución real de la variable.

Un ejemplo puede ser el registro del peso (en gramos) de larvas de una especie de pez. Aunque se observen muchos individuos, los datos pueden resumirse indicando la frecuencia de cada valor de peso:

Peso (g)	0.5	0.7	0.9	1.1
Frecuencia	6	4	3	2

Para indicar que se utilice la columna de frecuencias como ponderación, se debe acceder a:

Datos > Ponderar casos

En la ventana que aparece (Figura 2.9), se deben seguir los siguientes pasos:

1. Seleccionar la opción **Ponderar casos mediante**.
2. Elegir la variable que contiene la frecuencia de cada caso (por ejemplo, la columna **Frecuencia**) y moverla al cuadro *Variable de frecuencia*.

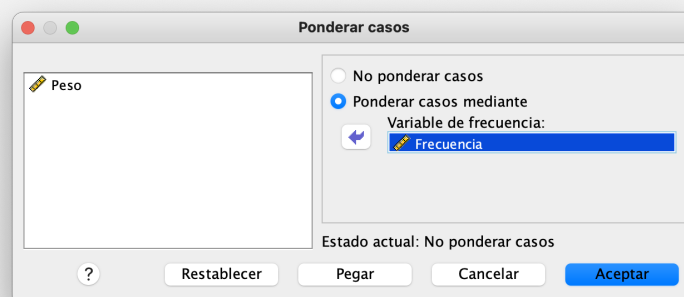


Figura 2.9: Ventana de SPSS para asignar la variable de ponderación.

3. Hacer clic en **Aceptar**. Tras completar este procedimiento, SPSS tratará cada caso con un peso igual al valor de la variable de frecuencia, de modo que los análisis se realizan como si cada observación apareciera ese número de veces.

2.4 Resumen de procedimientos

Capítulo 2: Estadística descriptiva

1. **Cálculo de medidas descriptivas.** Para obtener medidas como la media, mediana, moda, desviación típica y rango, entre otras:

Analizar > Estadísticos descriptivos > Frecuencias

En la ventana emergente, se seleccionan las variables de interés y se hace clic en Estadísticos para elegir las medidas deseadas. Al finalizar, se pulsa Aceptar. Opcionalmente, en la misma ventana, se puede marcar la casilla **Mostrar tablas de frecuencias** para generar una tabla con la distribución de los valores de la variable seleccionada.

2. **Creación de gráficos de barras, sectores e histogramas.** Estos gráficos se generan desde el mismo menú de frecuencias:

Analizar > Estadísticos descriptivos > Frecuencias

En la ventana que aparece, se selecciona el botón Gráficos y se elige el tipo de gráfico deseado (barras, sectores o histograma). Finalmente, se pulsa Aceptar.

3. **Creación de diagramas de caja.** Para visualizar la dispersión y valores atípicos en los datos:

Gráficos > Cuadros de diálogo antiguos > Diagramas de cajas

En la ventana emergente:

- Seleccionar la opción **Simple y Resúmenes** para distintas variables.
- Mover la variable de interés al cuadro *Las cajas representan*.
- Pulsar Aceptar.

4. **Ponderación de casos según frecuencias.** Cuando los datos incluyen la frecuencia de cada uno:

Datos > Ponderar casos

En la ventana que aparece:

- Seleccionar la opción **Ponderar casos** mediante.
- Mover la variable de frecuencias al cuadro *Variable de frecuencia*.
- Pulsar Aceptar.

2.5 Ejercicios

Ejercicio 1

Un biólogo registró el número de hojas presentes en diez plantas de una especie de hierba en un experimento de crecimiento. Los datos se presentan a continuación:

Número de hojas	5	3	20	3	8	2	7	10	3	3
------------------------	---	---	----	---	---	---	---	----	---	---

- Introducir los datos en SPSS.
- Calcular las medidas descriptivas: media, mediana y desviación típica.
- Determinar el valor de la variable que es mayor que el 80% de los datos.
- Determinar el valor de la variable que es menor que el 10% de los datos más grandes.
- Determinar entre qué valores está el 30% central de los datos.
- Generar el diagrama de cajas e indicar si se observa algún valor atípico.
- Indicar la forma de la distribución de los datos.

Ejercicio 2

Se midió la longitud (en cm) de los caparazones de larvas de cangrejo en un laboratorio. Como se observaron muchos individuos con longitudes repetidas, los datos se resumen en la siguiente tabla mediante frecuencias:

Longitud	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2
Frecuencia	4	5	7	6	8	6	6	5	4	3	2	1

- Introducir los datos en SPSS.
- Configurar SPSS para ponderar los casos según la columna de frecuencia.
- Calcular las medidas descriptivas: media, varianza y rango.
- Generar un diagrama de barras para visualizar la frecuencia de cada longitud.

Capítulo 3

Regresión



La *regresión* es una herramienta estadística fundamental para analizar la relación entre dos o más variables. En biología, resulta especialmente útil para estudiar cómo una variable dependiente (como peso, longitud, tasa de crecimiento o una medida ambiental) varía en función de una o más variables independientes. Este análisis permite identificar patrones, describir tendencias y formular predicciones basadas en los datos. En este capítulo se explica cómo utilizar SPSS para construir diagramas de dispersión que muestren la relación entre variables, calcular la recta de regresión y emplearla para predecir valores, así como obtener los coeficientes de correlación y de determinación, que cuantifican la fuerza de la asociación y el ajuste del modelo. También se aborda el uso de otros modelos de regresión cuando la relación entre las variables no puede representarse adecuadamente mediante un modelo lineal. El objetivo es que el estudiante pueda analizar relaciones entre variables, interpretar correctamente los resultados y aplicar modelos de regresión a situaciones reales en el ámbito biológico.

3.1 Diagrama de dispersión

Para ilustrar los procedimientos de regresión en SPSS se utilizará el siguiente conjunto de datos, que relaciona la cantidad de fertilizante aplicada a las plantas (X , en gramos) con la altura alcanzada (Y , en cm). El objetivo es analizar cómo la dosis de fertilizante influye en el crecimiento y ajustar un modelo de regresión que permita describir esta relación y formular predicciones a partir de los valores de la variable independiente. Los datos del experimento se muestran en la Tabla 3.1.

Fertilizante (X)	Altura (Y)
0.0	10.4
1.1	16.5
2.4	17.1
3.7	16.9
4.9	19.3
5.8	21.8

Tabla 3.1: Altura de las plantas en función de la cantidad de fertilizante aplicada.

Antes de iniciar el análisis de regresión es necesario introducir estos datos en SPSS siguiendo el procedimiento general descrito en el Capítulo 1.

En el contexto de la regresión, el **diagrama de dispersión** constituye el primer paso del análisis, ya que permite evaluar visualmente si existe una asociación entre las variables y si dicha relación presenta alguna tendencia. Este gráfico representa pares de valores (x, y) , donde x corresponde a la variable independiente e y a la variable dependiente. Su utilidad radica en facilitar la detección de patrones, tendencias y posibles relaciones entre las variables, lo que orienta la selección del modelo de regresión más adecuado. Para crear la nube de puntos en SPSS se debe seguir el siguiente procedimiento:

Gráficos > Cuadros de diálogo antiguos > Dispersión / Puntos

1. En la ventana inicial seleccionar **Dispersión simple** y pulsar **Definir** (Figura 3.1).

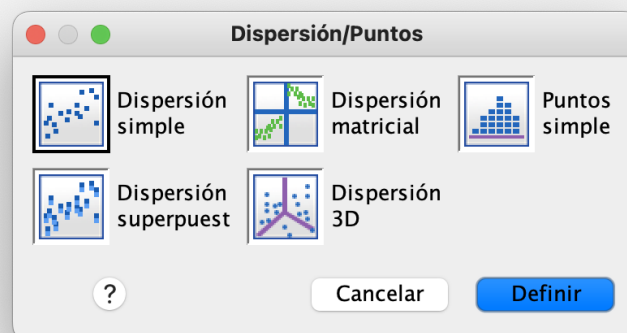


Figura 3.1: Selección del tipo de diagrama de dispersión en SPSS.

2. En la ventana emergente colocar la variable independiente en el *Eje X* y la variable dependiente en el *Eje Y* (Figura 3.2). Pulsar Aceptar para generar el gráfico.

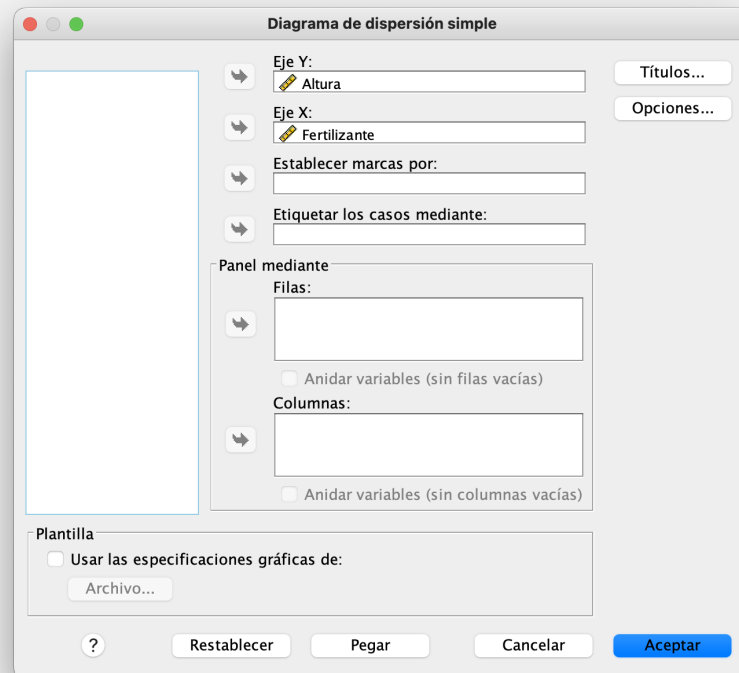


Figura 3.2: Asignación de variables a los ejes del diagrama de dispersión.

El gráfico obtenido (Figura 3.3) muestra los pares de datos distribuidos en el plano. Cuando los puntos se agrupan en torno a una línea ascendente, se puede inferir que existe una relación positiva o directa entre las variables: valores mayores de x se asocian a valores mayores de y . Una tendencia descendente indicaría una relación negativa o inversa.

El diagrama de dispersión también permite detectar valores atípicos (*outliers*) que pueden influir de forma notable en la estimación de un modelo de regresión. Por este motivo, siempre se recomienda examinar este gráfico antes de proceder al ajuste del modelo.

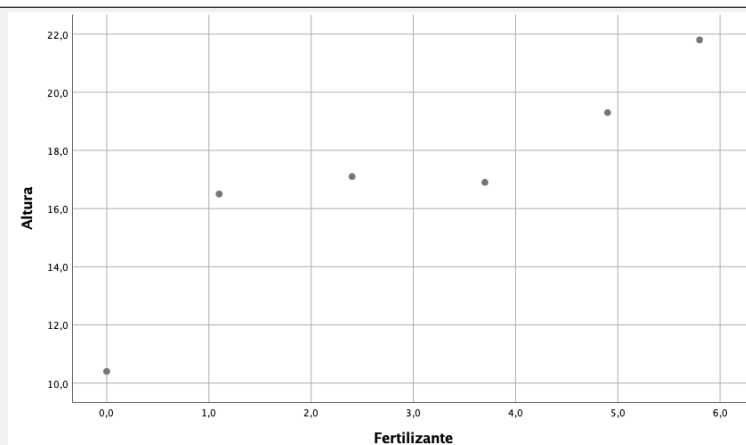


Figura 3.3: Diagrama de dispersión de altura de las plantas en función del fertilizante.

3.2 Cálculo de la recta de regresión

La **recta de regresión** permite expresar de forma matemática la relación entre una variable dependiente (Y) y una variable independiente (X) a partir de datos observados. Su ecuación general es:

$$\hat{y} = ax + b$$

donde:

- a es la **pendiente**, que indica el cambio esperado en $\hat{y}$ por cada unidad de cambio en x .
- b es el **intercepto** o término independiente, es decir, el valor de $\hat{y}$ cuando $x = 0$.

En un contexto biológico, este modelo permite estimar, por ejemplo, cómo una variable fisiológica responde a variaciones ambientales o cómo un parámetro de crecimiento depende de una medida corporal. En el ejemplo de este capítulo, se pretende predecir la **altura de las plantas** en función de la **cantidad de fertilizante** aplicada. La recta de regresión se ajusta a los datos observados minimizando la suma de los errores al cuadrado mediante el método de los *mínimos cuadrados*.

Para calcular la recta de regresión en SPSS, se debe seguir el siguiente procedimiento:

Analizar > Regresión > Lineales

1. En la ventana emergente (Figura 3.4):
 - Colocar la variable dependiente en el cuadro *Dependientes*.
 - Colocar la variable independiente en el cuadro *Independientes*.
2. Pulsar Aceptar para ejecutar el análisis.

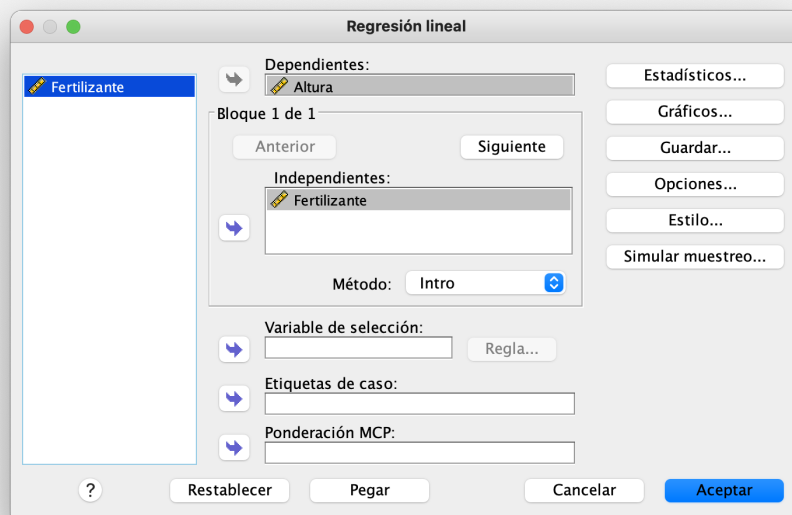


Figura 3.4: Configuración del análisis de regresión lineal en SPSS.

SPSS generará una tabla denominada *Coefficientes* (Figura 3.5), en la que se muestran los valores estimados de a (pendiente) y b (intercepto). En este ejemplo, dichos valores son $a = 1.543$ y $b = 12.397$. Sustituyéndolos en la ecuación general, se obtiene la recta de regresión ajustada: $\hat{y} = 1.543x + 12.397$

Coeficientes ^a					
Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	Sig.
		B	Desv. Error	Beta	
1	(Constante)	12,397	1,298		,001
	Fertilizante	1,543	,359	,906	,013

a. Variable dependiente: Altura

Figura 3.5: Tabla de coeficientes obtenida en el análisis de regresión lineal en SPSS.

3.3 Predicción de nuevos valores

Una vez obtenida la ecuación de la recta de regresión, es posible realizar predicciones del valor de la variable dependiente (Y) a partir de valores conocidos de la variable independiente (X). Este procedimiento constituye una de las aplicaciones más útiles del análisis de regresión, ya que permite estimar resultados esperados en nuevas observaciones basadas en la relación lineal identificada.

Para obtener la predicción, basta con sustituir el valor de x en la ecuación. Por ejemplo, si la ecuación es $\hat{y} = 1.543x + 12.397$ y queremos predecir $\hat{y}$ para $x = 4$: $\hat{y} = 1.543 \cdot 4 + 12.397 = 18.569$

Para realizar predicciones a partir de la recta de regresión en SPSS, se siguen los siguientes pasos:

1. Añadir, en la columna de la variable independiente (X), una fila con el valor para el cual se desea predecir Y . En nuestro caso, se introduce en la columna *Fertilizante* el valor 4 en la última fila.
2. Ir a *Transformar > Calcular variable*.
3. Introducir un nombre para la nueva variable que contendrá las predicciones, por ejemplo *Prediccion*.
4. En la casilla *Expresión numérica*, escribir la ecuación de la recta de regresión con la variable X (Figura 3.6). En nuestro caso: $1.543 * \text{Fertilizante} + 12.397$. Pulsar *Aceptar*.

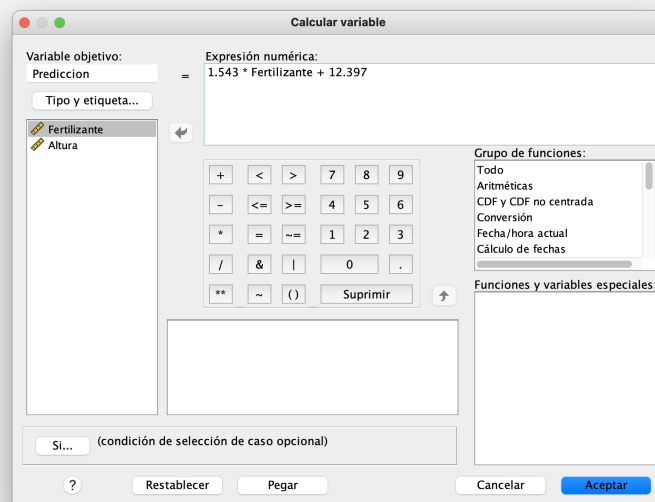


Figura 3.6: Ventana de cálculo de variable para crear predicciones.

3.4 Medidas de relación y ajuste

En el análisis de regresión, existen dos indicadores fundamentales que permiten cuantificar la relación entre las variables y evaluar la capacidad del modelo para explicar la variabilidad observada en los datos:

1. **Coefficiente de correlación (r):** El coeficiente de correlación de Pearson mide la fuerza y la dirección de la relación lineal entre dos variables cuantitativas. Su valor oscila entre -1 y $+1$:

$$-1 \leq r \leq +1$$

- Si r es positivo y cercano a $+1$, existe una relación lineal directa fuerte: a medida que X aumenta, Y también aumenta.
- Si r es negativo y cercano a -1 , la relación es inversa: al aumentar X , Y tiende a disminuir.
- Si r está próximo a 0 , no se observa una relación lineal apreciable entre las variables.

2. **Coefficiente de determinación (R^2):** El coeficiente de determinación indica la proporción de la variabilidad de la variable dependiente (Y) que puede explicarse mediante el modelo de regresión ajustado. Su valor varía entre 0 y 1 :

$$0 \leq R^2 \leq 1$$

- Un R^2 cercano a 1 indica que el modelo explica la mayor parte de la variación observada en Y .
- Un R^2 cercano a 0 indica que el modelo apenas explica la variabilidad de la variable dependiente.

Al realizar un análisis de regresión lineal en SPSS, los resultados incluyen una tabla titulada *Resumen del modelo*, en la que se muestran los valores de R y R^2 . En el caso de una regresión lineal simple, el valor de R coincide con el valor absoluto del coeficiente de correlación de Pearson ($|r|$). Para conocer la dirección de la relación entre las variables (ascendente o descendente), es necesario observar el signo de la pendiente de la recta de regresión.

En el ejemplo (Figura 3.7), se obtiene $R = 0.906$, lo que indica una relación lineal fuerte entre las variables, y $R^2 = 0.822$, lo que significa que el modelo explica el 82.2% de la variabilidad observada en la variable dependiente.

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,906 ^a	,822	,777	1.7936
a. Predictores: (Constante), Fertilizante				

Figura 3.7: Tabla *Resumen del modelo* en SPSS mostrando los valores de R y R^2 .

3.5 Otros modelos de regresión

En algunos casos, la relación entre dos variables no sigue un patrón lineal visible en un diagrama de dispersión. En esta sección se aborda un modelo frecuente para este tipo de situaciones: la **regresión cuadrática**. Este modelo permite representar relaciones de tipo parabólico, capturando curvaturas que no pueden describirse adecuadamente mediante una recta. La ecuación general es:

$$\hat{y} = ax^2 + bx + c,$$

donde:

- a es el **coeficiente cuadrático**, que determina la curvatura de la relación;
- b es el **coeficiente lineal**, que refleja la tendencia directa de X sobre Y ;
- c es el **intercepto**, el valor de $\hat{y}$ cuando $x = 0$.

Si $a > 0$, la curva tiene forma de U, indicando que $\hat{y}$ disminuye hasta cierto punto y luego aumenta. Si $a < 0$, la curva tiene forma de U invertida, mostrando un incremento inicial seguido de un descenso. Este tipo de modelo es frecuente en estudios biológicos en los que respuestas fisiológicas, ecológicas o de crecimiento no varían de manera estrictamente lineal con la variable independiente.

Para calcular la regresión cuadrática en SPSS, se siguen los pasos siguientes:

Analizar > Regresión > Estimación curvilínea

1. En la ventana emergente:

- Colocar la variable dependiente en el cuadro *Dependientes*.
- Colocar la variable independiente en el cuadro *Independiente*.

2. Seleccionar el modelo Cuadrático y, opcionalmente, Lineal para comparar ajustes. Pulsar Aceptar.

SPSS mostrará los coeficientes a , b y c en la tabla *Resumen de modelo y estimaciones de parámetro* (Figura 3.8). Con ellos, puede escribirse la ecuación ajustada del modelo:

$$\hat{y} = -0.129x^2 + 2.293x + 11.842$$

Resumen de modelo y estimaciones de parámetro								
Variable dependiente: Altura								
Resumen del modelo						Estimaciones de parámetro		
Ecuación	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2
Lineal	,822	18,431	1	4	,013	12,397	1,543	
Cuadrático	,837	7,677	2	3	,066	11,842	2,293	-,129
La variable independiente es Fertilizante.								

La variable independiente es Fertilizante.

Figura 3.8: Coeficientes de la regresión cuadrática en SPSS.

Este modelo permite describir relaciones curvilíneas con mayor precisión que una recta simple, siempre que la forma parabólica se ajuste adecuadamente a los datos observados. Además, en la misma tabla se puede comparar el ajuste de modelos distintos mediante R^2 .

3.6 Resumen de procedimientos

Capítulo 3: Regresión

1. **Creación de diagramas de dispersión.** Permite visualizar la relación entre dos variables y detectar patrones o valores atípicos (*outliers*):

Gráficos > Cuadros de diálogo antiguos > Dispersión / Puntos

En la ventana emergente, seleccionar *Dispersión simple*, pulsar *Definir*, asignar la variable independiente al eje X y la dependiente al eje Y, y pulsar *Aceptar*.

2. **Cálculo de la recta de regresión.** Ajusta un modelo lineal a los datos:

Analizar > Regresión > Lineales

Colocar la variable dependiente en *Dependientes* y la independiente en *Independientes*, y pulsar *Aceptar*.

3. **Realización de predicciones.** Estimar valores de la variable dependiente para nuevos valores de la variable independiente:

Transformar > Calcular variable

Introducir el nombre de la nueva variable (por ejemplo, *Prediccion*) y escribir la ecuación de la recta de regresión en *Expresión numérica*. Pulsar *Aceptar* para generar la columna de predicciones.

4. **Cálculo de los coeficientes de correlación y determinación.** Permite evaluar la calidad del ajuste del modelo:

Los valores de R y R^2 aparecen en la tabla "Resumen del modelo" al realizar la regresión lineal.

5. **Otros modelos de regresión (cuadrático).** Ajusta un modelo parabólico cuando la relación no es lineal:

Analizar > Regresión > Estimación curvilínea

Seleccionar la variable dependiente y la independiente, elegir el modelo *Cuadrático* y pulsar *Aceptar*. SPSS mostrará los coeficientes a , b y c .

3.7 Ejercicios

Ejercicio 1

En un estudio sobre plantas acuáticas se desea evaluar si la concentración de nitratos en el agua (mg/L) permite predecir la altura de las plantas (cm). Se sabe que los nitratos son un nutriente esencial que favorece el crecimiento, aunque cada planta puede responder de manera ligeramente distinta debido a la variabilidad biológica. Los datos obtenidos en el experimento se muestran a continuación:

Nitratos	7.4	7.5	8.5	9.0	9.0	11.0	13.0	14.0	14.5	16.0	17.0	18.0	20.0	23.0
Altura	30.0	25.0	31.5	27.5	39.5	38.0	43.0	49.0	55.0	48.5	51.0	64.5	63.0	68.0

- Construir un diagrama de dispersión de altura frente a concentración de nitratos.
- Ajustar la recta de regresión lineal para predecir la altura a partir de la concentración de nitratos.
- Calcular la altura prevista cuando la concentración de nitratos sea 12 mg/L.
- Evaluar la bondad del ajuste del modelo.
- Describir la relación lineal observada entre las variables.

Ejercicio 2

En un experimento de laboratorio se investigó cómo la intensidad de luz (lux) afecta al porcentaje de flores que abren en una especie de planta bajo condiciones controladas tras dos semanas. Se obtuvieron los siguientes datos:

Intensidad luz	50	100	150	200	250	300	350	400	450	500	550	600
Porcentaje flores	5	12	25	40	55	65	70	68	60	50	35	20

- Construir un diagrama de dispersión del porcentaje de flores abiertas frente a la intensidad de luz.
- Ajustar un modelo de regresión lineal para predecir el porcentaje de flores a partir de la intensidad de luz.
- Ajustar un modelo de regresión cuadrático para predecir el porcentaje de flores a partir de la intensidad de luz.
- Determinar qué modelo ofrece un mejor ajuste a los datos disponibles.

Capítulo 4

Modelos de probabilidad



Los *modelos de probabilidad* constituyen una herramienta fundamental en estadística para describir y cuantificar la incertidumbre asociada a los fenómenos observados. En biología y en otras ciencias permiten estimar la probabilidad de diversos eventos y analizar la variabilidad de los datos a partir de distribuciones teóricas. En este capítulo se estudiarán tres modelos ampliamente utilizados: la distribución binomial, la distribución de Poisson y la distribución normal. Se abordará el cálculo de probabilidades y la determinación de percentiles, útiles para situar un valor dentro de una distribución dada. El objetivo es proporcionar una comprensión sólida y aplicada de estos modelos en contextos biológicos, así como desarrollar la capacidad de interpretar correctamente los resultados.

4.1 Cálculo de probabilidades

El cálculo de probabilidades mediante modelos de probabilidad en SPSS (binomial, Poisson o normal) sigue un procedimiento general aplicable a distintas distribuciones. La idea consiste en expresar la probabilidad de interés mediante una función que SPSS pueda evaluar desde la ventana de cálculo de variables. Los pasos son los siguientes:

1. **Preparar la hoja de datos.** Introducir un valor cualquiera (por ejemplo, 1) en una celda de la hoja de datos (Figura 4.1). Esto es suficiente para poder acceder a las funciones de probabilidad y ejecutar los cálculos, ya que SPSS requiere al menos un caso en la hoja de datos.

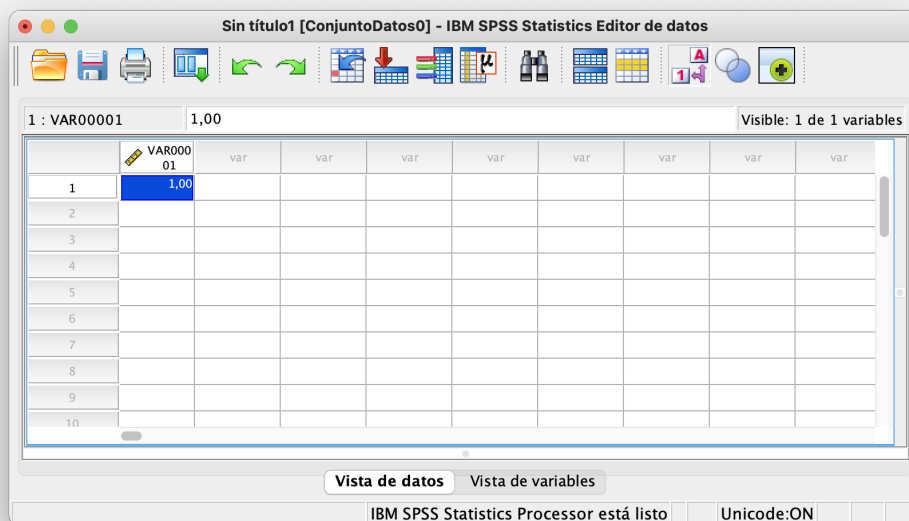


Figura 4.1: Ejemplo de valor en la hoja de datos para acceder al cálculo de probabilidades.

2. **Abrir la ventana de cálculo de variables.** Todas las probabilidades se obtienen desde el menú:

Transformar > Calcular variable

3. **Nombrar la variable de resultado.** En la ventana *Calcular variable*, introducir en el campo *Variable objetivo* el nombre de la variable que almacenará el valor de la probabilidad, por ejemplo Probabilidad.
4. **Seleccionar el tipo de probabilidad.** En el cuadro *Grupo de funciones*, SPSS permite calcular:
 - **Probabilidades puntuales** $P(X = a)$, mediante funciones del grupo PDF.
 - **Probabilidades acumuladas** $P(X \leq a)$, mediante funciones del grupo CDF.

En el caso de la distribución normal, al tratarse de una variable continua, las probabilidades puntuales son nulas y solo se utilizan probabilidades acumuladas.

5. **Seleccionar la distribución.** A continuación, se elige la función correspondiente en el cuadro *Funciones y variables especiales*.

Para probabilidades puntuales en distribuciones discretas (PDF):

- PDF.BINOM para la distribución binomial,
- PDF.POISSON para la distribución de Poisson.

Para probabilidades acumuladas (CDF):

- CDF.BINOM para la distribución binomial,
- CDF.POISSON para la distribución de Poisson,
- CDF.NORMAL para la distribución normal.

Por ejemplo, para obtener $P(X = 2)$ en una binomial con $n = 5$ y $p = 0.2$, se selecciona el grupo PDF en el cuadro *Grupo de funciones*, y después PDF.BINOM en *Funciones y variables especiales*.

6. **Introducir la expresión numérica.** En el cuadro *Expresión numérica*, escribir la función junto con sus parámetros correspondientes (Figura 4.2). De este modo, SPSS traduce la probabilidad deseada a una expresión evaluable.

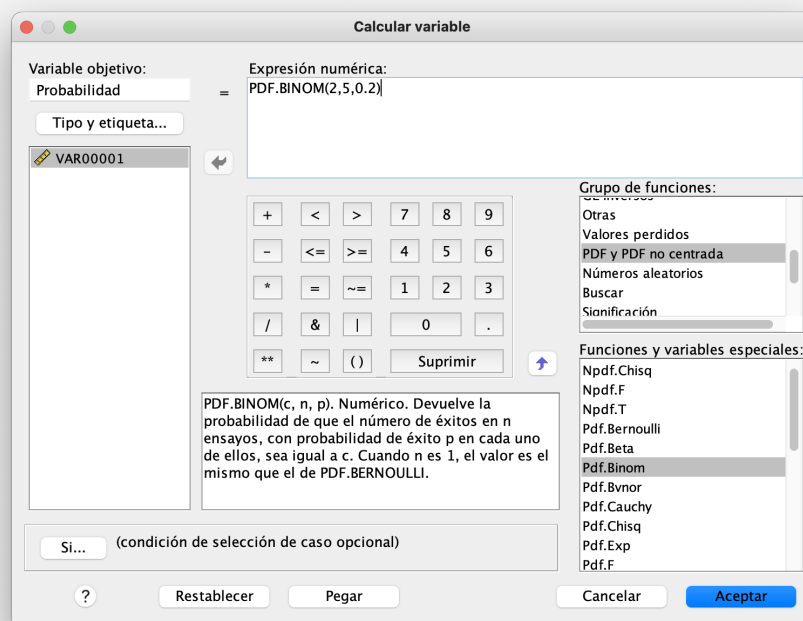


Figura 4.2: Ejemplo de expresión numérica para el cálculo de probabilidades.

7. **Calcular la probabilidad.** Pulsar Aceptar. SPSS generará una nueva columna en la *Vista de datos* con el valor de la probabilidad calculada. Al seleccionar la celda correspondiente, la barra de fórmulas mostrará el valor con mayor precisión decimal.

Este procedimiento permite obtener probabilidades puntuales y acumuladas para los distintos modelos de probabilidad disponibles en SPSS, siguiendo una pauta de trabajo uniforme y fácilmente aplicable a diferentes situaciones.

4.2 Modelo binomial

Para ilustrar el procedimiento de cálculo de probabilidades con el modelo binomial en SPSS, consideremos un estudio orientado a evaluar la eficacia de un tratamiento destinado a incrementar la germinación de semillas de una especie vegetal. Se tratan 15 semillas y se sabe que la probabilidad de germinación es $p = 0.3$. Con esta información se desea calcular las siguientes probabilidades:

- a) **Ninguna semilla germine.** La probabilidad buscada es puntual:

$$P(X = 0)$$

Como la expresión ya está formulada en términos de $P(X = a)$, puede introducirse directamente en SPSS mediante la función correspondiente en el cuadro *Expresión numérica*:

$$\text{PDF.BINOM}(0, 15, 0.3)$$

El valor obtenido es aproximadamente 0.00475.

- b) **Más de dos semillas germinen.** Para expresar esta probabilidad en una forma que SPSS pueda evaluar, se utiliza su equivalente acumulado:

$$P(X > 2) = 1 - P(X \leq 2)$$

La expresión adecuada en SPSS es:

$$1 - \text{CDF.BINOM}(2, 15, 0.3)$$

El resultado proporcionado es aproximadamente 0.8732.

- c) **Entre tres y cinco semillas germinen, ambos inclusive.** En este caso, la probabilidad puede expresarse mediante la diferencia de probabilidades acumuladas:

$$P(3 \leq X \leq 5) = P(X \leq 5) - P(X \leq 2)$$

La expresión correspondiente en SPSS es:

$$\text{CDF.BINOM}(5, 15, 0.3) - \text{CDF.BINOM}(2, 15, 0.3)$$

El valor obtenido es aproximadamente 0.5948.

Estos ejemplos muestran cómo traducir las probabilidades solicitadas a expresiones puntuales $P(X = a)$ o acumuladas $P(X \leq a)$, que pueden introducirse directamente en la ventana de cálculo de variables de SPSS.

4.3 Modelo de Poisson

Para ilustrar el cálculo de probabilidades con el modelo de Poisson, consideremos un estudio centrado en la llegada de granos de polen a una zona determinada. Se sabe que, en promedio, llegan $\lambda = 10$ granos por hora. Con esta información se desea calcular las siguientes probabilidades:

- a) **Que lleguen exactamente cinco granos de polen en una hora.** La probabilidad se formula como probabilidad puntual:

$$P(X = 5)$$

La expresión correspondiente en SPSS es:

$$\text{PDF.POISSON}(5, 10)$$

El valor obtenido es aproximadamente 0.0378.

- b) **Que lleguen menos de quince granos de polen en dos horas.** Al duplicar el intervalo temporal, el parámetro de la distribución pasa a ser $\lambda = 20$. La probabilidad buscada puede expresarse como acumulada:

$$P(X < 15) = P(X \leq 14)$$

En SPSS, la expresión adecuada es:

$$\text{CDF.POISSON}(14, 20)$$

El resultado aproximado es 0.1049.

- c) **Que lleguen más de cuatro y menos de ocho granos de polen en una hora.** Esta probabilidad puede expresarse mediante la diferencia de probabilidades acumuladas:

$$P(4 < X < 8) = P(X \leq 7) - P(X \leq 4)$$

La expresión correspondiente en SPSS es:

$$\text{CDF.POISSON}(7, 10) - \text{CDF.POISSON}(4, 10)$$

El valor aproximado obtenido es 0.191.

4.4 Modelo normal

Para ilustrar el cálculo de probabilidades con el modelo normal, consideremos un estudio sobre el contenido de clorofila en hojas de una planta. Se sabe que esta variable, medida en mg/100 g de hoja, sigue una distribución normal con media $\mu = 106$ y desviación típica $\sigma = 8$. Con estos parámetros se desea calcular las siguientes probabilidades:

- a) **Probabilidad de que el contenido de clorofila sea inferior a 120 mg/100 g de hoja.** La probabilidad se formula como probabilidad acumulada:

$$P(X < 120) = P(X \leq 120)$$

La expresión correspondiente en SPSS es:

$$\text{CDF.NORMAL}(120, 106, 8)$$

El valor obtenido es aproximadamente 0.9599.

- b) **Probabilidad de que el contenido de clorofila sea superior a 140 mg/100 g de hoja.** Esta probabilidad se calcula como el complemento de la acumulada:

$$P(X > 140) = 1 - P(X \leq 140)$$

En SPSS la expresión adecuada es:

$$1 - \text{CDF.NORMAL}(140, 106, 8)$$

El valor aproximado es 0.00011.

- c) **Porcentaje de plantas con contenido de clorofila entre 90 y 130 mg/100 g de hoja.** Esta probabilidad puede expresarse mediante la diferencia de probabilidades acumuladas:

$$P(90 \leq X \leq 130) = P(X \leq 130) - P(X \leq 90)$$

La expresión en SPSS es:

$$\text{CDF.NORMAL}(130, 106, 8) - \text{CDF.NORMAL}(90, 106, 8)$$

El resultado aproximado es 0.9759.

Este ejemplo muestra cómo calcular probabilidades en el modelo normal utilizando la misma estrategia general aplicada a las distribuciones binomial y de Poisson, basada en la formulación de probabilidades acumuladas compatibles con las funciones disponibles en SPSS.

4.5 Percentiles

Los percentiles permiten situar un valor dentro de una distribución y facilitan la comparación de posiciones relativas entre distintas poblaciones. El percentil k es el valor P_k que satisface $P(X \leq P_k) = k$, es decir, el punto de la distribución por debajo del cual se encuentra la proporción k de la población.

Para ilustrar su cálculo en un contexto biológico, consideremos un estudio sobre el peso de larvas de mariposa tras una semana de alimentación controlada. Se supone que el peso sigue una distribución normal con media $\mu = 2.5$ g y desviación típica $\sigma = 0.3$ g. El objetivo es determinar el peso que delimita el 25% inferior de las larvas, es decir, el percentil 25.

En SPSS, el cálculo de percentiles se realiza mediante el *Grupo de funciones* de GL inversos, seleccionando la función correspondiente a la distribución de interés. Para una distribución normal se utiliza `IDF.NORMAL`. Así, para obtener el percentil 25 se introduce:

```
IDF.NORMAL(0.25, 2.5, 0.3)
```

El valor devuelto por SPSS indica que aproximadamente el 25% de las larvas presentan un peso inferior a 2.3 g. Este procedimiento es aplicable a cualquier percentil de una distribución normal, modificando el primer argumento de la función `IDF.NORMAL`.

4.6 Resumen de procedimientos

Capítulo 4: Modelos de probabilidad

El cálculo de probabilidades en SPSS, ya sea mediante modelos binomiales, de Poisson o normales, sigue un procedimiento común y sistemático. A continuación se presenta el flujo de trabajo general:

1. **Preparar la hoja de datos.** Introducir un valor cualquiera en la *Vista de datos*. Esto permite activar el menú necesario para el cálculo de probabilidades.
2. **Abrir la ventana de cálculo de variables.** Todas las probabilidades, puntuales o acumuladas, se calculan desde:

Transformar > Calcular variable

3. **Nombrar la variable de resultado.** En el cuadro *Variable objetivo*, asignar un nombre identificativo a la nueva variable que contendrá los resultados, por ejemplo *Probabilidad*.
4. **Seleccionar el tipo de probabilidad.** En el cuadro *Grupo de funciones* se distinguen dos categorías fundamentales:
 - Probabilidad puntual (PDF), para calcular $P(X = a)$.
 - Probabilidad acumulada (CDF), para calcular $P(X \leq a)$.
5. **Seleccionar la distribución.** En el cuadro *Funciones y variables especiales* se elige la función que corresponde al modelo deseado:
 - PDF.BINOM / CDF.BINOM para la distribución binomial,
 - PDF.POISSON / CDF.POISSON para la distribución de Poisson,
 - CDF.NORMAL para la distribución normal.
6. **Introducir la expresión numérica.** En el cuadro *Expresión numérica* se escribe la función seleccionada con los parámetros correspondientes. Por ejemplo:
 - Binomial puntual: PDF.BINOM(valor, n, p)
 - Poisson acumulada: CDF.POISSON(valor, lambda)
 - Normal acumulada: CDF.NORMAL(valor, mu, sigma)
7. **Calcular la probabilidad.** Pulsar Aceptar. SPSS generará una nueva columna en la *Vista de datos* con los valores obtenidos. Para visualizar más decimales, basta con seleccionar una celda y consultar la barra de fórmulas.

Para **calcular percentiles**, debe utilizarse el grupo de funciones GL inversos. En el caso de la distribución normal, la función correspondiente es:

IDF.NORMAL(percentil, mu, sigma)

Modificando el primer parámetro se puede obtener cualquier percentil.

4.7 Ejercicios

Ejercicio 1

En un laboratorio de genética vegetal, se realiza una prueba para detectar un marcador genético específico en semillas. Cada prueba identifica correctamente el marcador con una probabilidad del 92%. Si se analizan 72 semillas en un experimento:

- a) Calcular la probabilidad de que 60 o menos semillas sean correctamente identificadas.
- b) Calcular la probabilidad de que menos de 60 semillas sean correctamente identificadas.
- c) Calcular la probabilidad de que exactamente 60 semillas sean correctamente identificadas.

Ejercicio 2

En un área de cultivo se ha observado que, en promedio, cada año caen 12 semillas de una planta determinada en la parcela. Suponiendo que el número de semillas que caen sigue una distribución de Poisson, calcular la probabilidad de que:

- a) Caigan exactamente 10 semillas en un año.
- b) Caigan 15 o más semillas durante un año.
- c) Caigan 10 o menos semillas en un periodo de seis meses.

Ejercicio 3

En un estudio sobre peces de un río, se ha medido la concentración de nitratos en su sangre. La concentración de nitratos sigue una distribución normal con media 0.25 partes por millón (ppm) y desviación típica 0.08 ppm.

- a) ¿Cuál es la probabilidad de que un pez presente una concentración de nitratos superior a 0.40 ppm?
- b) ¿Cuál es la probabilidad de que un pez tenga una concentración de nitratos entre 0.20 ppm y 0.50 ppm?
- c) ¿Cuál es el valor máximo de concentración de nitratos que caracteriza al 40% de los peces con menor concentración?

Capítulo 5

Intervalos de confianza



Los *intervalos de confianza* (IC) permiten estimar el rango de valores dentro del cual se sitúa un parámetro poblacional, como la media o la diferencia de medias. Frente a una estimación puntual, un IC expresa el grado de incertidumbre asociado al muestreo, lo que resulta especialmente relevante en estudios biológicos, donde la variabilidad es inherente a los sistemas vivos. En este capítulo se explica cómo construir intervalos de confianza en tres situaciones: para una sola media, para la diferencia de medias en muestras relacionadas y para la diferencia de medias en muestras independientes. Se presentan los supuestos de cada caso y las pautas para una interpretación correcta de los resultados.

5.1 IC para la media poblacional (una muestra)

Para ilustrar el cálculo de intervalos de confianza en SPSS, se utiliza un ejemplo basado en la concentración de clorofila (mg/g de tejido) en hojas de una especie vegetal cultivada bajo condiciones controladas. El objetivo es determinar un IC para la media poblacional de clorofila con un nivel de confianza del 99%. Se asume que la concentración de clorofila sigue una distribución aproximadamente normal. El IC se construye utilizando el estadístico t de Student con $n - 1$ grados de libertad, lo cual es adecuado cuando la desviación típica poblacional es desconocida. Se recogieron los siguientes valores:

Clorofila	2.8	3.0	2.6	3.1	2.9	2.7	3.2	2.8	3.0	2.9
-----------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Para calcular un intervalo de confianza para la media en SPSS, se debe acceder al menú:

Analizar > Comparar medias > Prueba T para una muestra

1. Seleccionar la variable Clorofila y moverla al cuadro *Variables de prueba* (Figura 5.1).

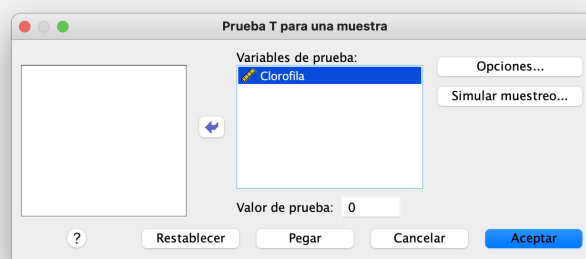


Figura 5.1: Selección de la variable para la prueba t de una muestra en SPSS.

2. En el apartado *Opciones*, establecer el *Porcentaje del intervalo de confianza* en 99% (Figura 5.2).

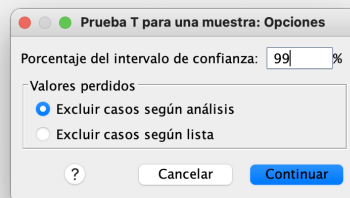


Figura 5.2: Configuración del nivel de confianza del 99% para el cálculo del intervalo en SPSS.

3. Pulsar *Aceptar*. SPSS mostrará el IC para la media poblacional en la parte derecha de la tabla *Prueba para una muestra* (Figura 5.3). En este ejemplo, la media de la muestra es 2.90 mg/g y el intervalo de confianza del 99% es [2.712, 3.088]. Este intervalo de confianza del 99% indica que, si se repitiera el muestreo muchas veces y se construyeran intervalos del mismo modo, aproximadamente el 99% de ellos contendrían la media poblacional (μ) de la concentración de clorofila.

Prueba para una muestra						
Valor de prueba = 0						
	t	gl	Sig. (bilateral)	Diferencia de medias	99% de intervalo de confianza de la diferencia	
					Inferior	Superior
Clorofila	50,229	9	,000	2.9000	2.712	3.088

Figura 5.3: Resultados del cálculo del intervalo de confianza para la media en SPSS.

5.2 IC para la diferencia de medias (muestras relacionadas)

Además de analizar una sola muestra, es común estimar el intervalo de confianza para la diferencia de medias poblacionales en mediciones relacionadas. Este intervalo permite determinar el rango en el que se encuentra el cambio promedio poblacional entre dos condiciones o momentos temporales.

Cuando los datos provienen de mediciones repetidas sobre los mismos individuos o de pares naturalmente relacionados, hablamos de **muestras relacionadas** o pareadas. En estos casos, el análisis se centra en la diferencia de cada par de observaciones, asumiendo que dichas diferencias siguen una distribución aproximadamente normal. El intervalo de confianza se calcula sobre estas diferencias utilizando la distribución t de Student con $n - 1$ grados de libertad. A modo de ejemplo, consideremos los datos de frecuencia cardíaca (latidos por minuto) de 10 voluntarios antes y después de una sesión de ejercicio moderado:

Antes del ejercicio	Después del ejercicio
70	85
65	78
72	88
68	80
75	90
70	82
66	79
69	84
71	86
67	81

El objetivo es calcular un intervalo de confianza al 99% para la diferencia de medias poblacionales ($\mu_d = \mu_{\text{Antes}} - \mu_{\text{Después}}$). Para ello, en SPSS se utiliza el siguiente menú:

Analizar > Comparar medias > Prueba T para muestras relacionadas

1. Seleccionar las variables correspondientes a la frecuencia cardíaca antes y después del ejercicio, y moverlas a los cuadros *Variable1* y *Variable2* (Figura 5.4).

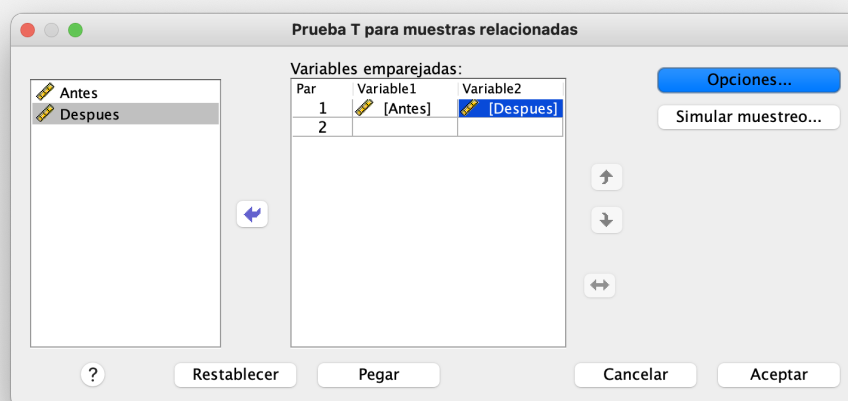


Figura 5.4: Selección de variables para la prueba t de muestras relacionadas en SPSS.

2. Pulsar el botón Opciones y establecer el *Porcentaje del intervalo de confianza* al 99% (Figura 5.5).

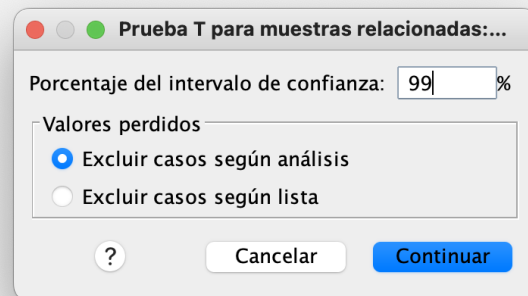


Figura 5.5: Configuración del nivel de confianza para la media de las diferencias.

3. Pulsar Continuar y luego Aceptar.

SPSS generará una tabla *Prueba de muestras emparejadas*, en la que se muestra el intervalo de confianza al 99% para la diferencia de medias poblacionales (Figura 5.6).

En este ejemplo, el intervalo de confianza al 99% para $\mu_d = \mu_{\text{Antes}} - \mu_{\text{Después}}$ es $[-15.453, -12.547]$. Esto indica que, si se repitiera el muestreo muchas veces y se construyeran intervalos del mismo modo, aproximadamente el 99% de ellos contendrían la diferencia de medias poblacionales. Dado que el intervalo no incluye el valor 0 y todos sus valores son negativos, existe evidencia estadística suficiente, con un nivel de confianza del 99%, de que la frecuencia cardíaca aumentó en promedio después de la sesión de ejercicio.

Prueba de muestras emparejadas									
Diferencias emparejadas									
		Media	Desv. Desviación	Desv. Error promedio	99% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
		Inferior	Superior						
Par 1	Antes – Despues	-14,000	1,414	,447	-15,453	-12,547	-31,305	9	,000

Figura 5.6: Intervalo de confianza al 99% para la media de las diferencias en muestras relacionadas.

5.3 IC para la diferencia de medias (muestras independientes)

Cuando los datos provienen de dos grupos distintos y no relacionados entre sí, se trata de **muestras independientes**. En estos casos, el análisis se centra en calcular un intervalo de confianza para la diferencia entre las medias poblacionales de ambos grupos.

Para ilustrar este procedimiento, consideremos la comparación de la concentración de glucosa en sangre entre dos grupos de ratones tras recibir dietas diferentes durante un mes: una dieta alta en carbohidratos (Grupo A) y una dieta baja en carbohidratos (Grupo B). El tamaño muestral es $n_A = 10$ para el Grupo A y $n_B = 8$ para el Grupo B. Se asume que la glucosa sigue una distribución aproximadamente normal en ambos grupos poblacionales y que las muestras son independientes. Los datos obtenidos (en mg/dL) son los siguientes:

Grupo A	Grupo B
120	110
125	108
118	115
130	112
127	109
122	114
124	111
129	113
121	-
128	-

Se desea calcular un intervalo de confianza al 90% para la diferencia de medias ($\mu_A - \mu_B$) con el fin de estimar la diferencia entre los niveles medios de glucosa en sangre entre ambas dietas.

Para llevar a cabo este análisis, es necesario organizar los datos en el formato adecuado: una columna que contenga todos los valores de ambas muestras y otra que indique a qué grupo pertenece cada dato. Para ello, se crean dos variables en la *Vista de variables* de SPSS: Datos y Grupo (Figura 5.7).

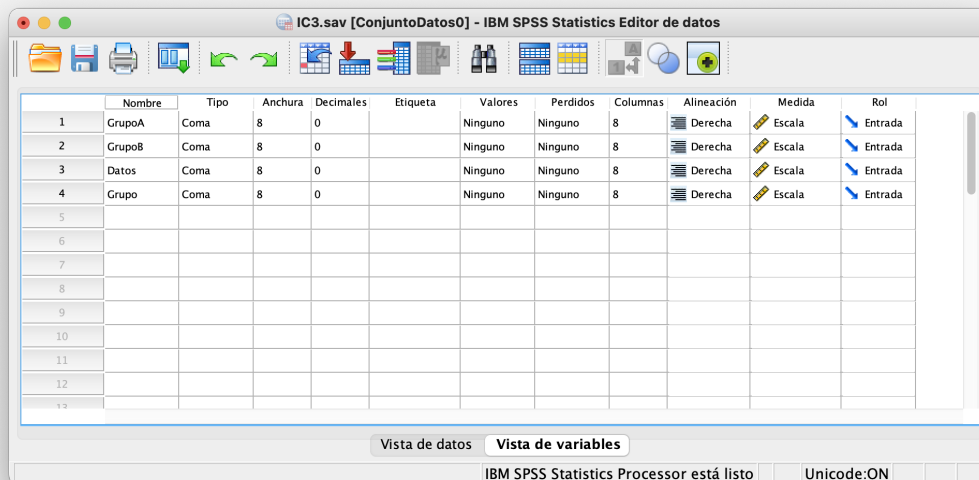


Figura 5.7: Creación de las variables Datos y Grupo en SPSS.

A continuación, rellenamos ambas columnas en la *Vista de datos* (Figura 5.8).

	GrupoA	GrupoB	Datos	Grupo
1	120	110	120	1
2	125	108	125	1
3	118	115	118	1
4	130	112	130	1
5	127	109	127	1
6	122	114	122	1
7	124	111	124	1
8	129	113	129	1
9	121	-	121	1
10	128	-	128	1
11	-	-	110	2
12	-	-	108	2
13	-	-	115	2
14	-	-	112	2
15	-	-	109	2
16	-	-	114	2
17	-	-	111	2
18	-	-	113	2
19				

Figura 5.8: Datos introducidos en SPSS para ambos grupos.

Con los datos ya organizados, podemos realizar el análisis mediante el menú:

Analizar > Comparar medias > Prueba T para muestras independientes

1. Seleccionar la variable Datos y moverla al cuadro *Variables de prueba*.
2. Mover la variable Grupo al cuadro *Variable de agrupación* (Figura 5.9).

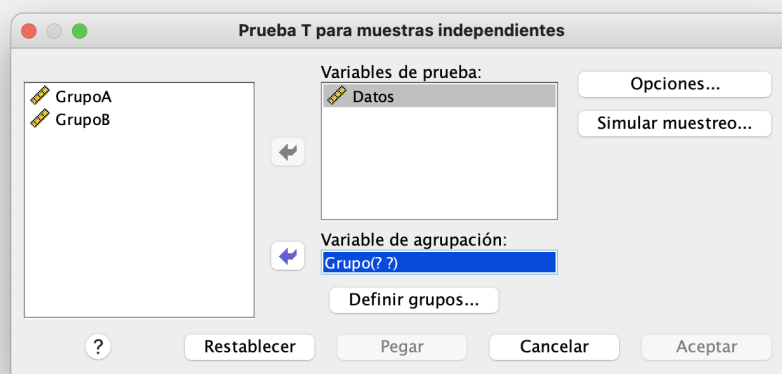


Figura 5.9: Asignación de variables en la prueba t para muestras independientes.

3. Hacer clic en **Definir grupos** y asignar los valores correspondientes a cada grupo: 1 para Grupo A y 2 para Grupo B (Figura 5.10). Pulsar **Continuar**.

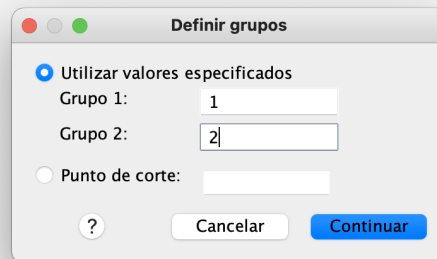


Figura 5.10: Definición de los grupos en SPSS.

4. Pulsar **Opciones** y especificar el *Porcentaje del intervalo de confianza* al 90% (Figura 5.11).

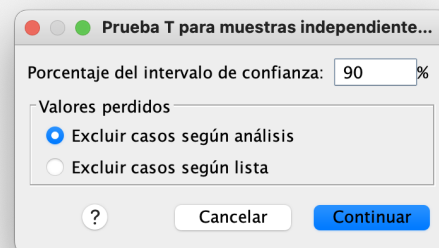


Figura 5.11: Opciones para establecer el nivel de confianza al 90%.

5. Pulsar **Continuar** y luego **Aceptar**.

SPSS generará la tabla *Prueba de muestras independientes* (Figura 5.12). En esta tabla aparecen dos resultados: uno bajo el supuesto de *varianzas iguales* y otro bajo el supuesto de *varianzas no iguales*. Para decidir cuál de ellos utilizar, se recurre a la **prueba de Levene** para igualdad de varianzas. Esta prueba indica si hay evidencia suficiente para rechazar la hipótesis de igualdad de varianzas poblacionales. Para ello, observamos el nivel de significación crítico o *p*-valor (etiquetado en SPSS como *Sig.*). Si este *p*-valor es menor que el nivel de significación establecido (usualmente $\alpha = 0.05$), se rechaza la hipótesis de igualdad de varianzas y se utiliza la fila correspondiente a *No se asumen varianzas iguales*. En caso contrario, no se rechaza la hipótesis y se utiliza la fila *Se asumen varianzas iguales*.

Prueba de muestras independientes									
		Prueba de Levene de igualdad de varianzas		prueba t para la igualdad de medias					90% de intervalo de confianza de la diferencia
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	Inferior Superior
Datos	Se asumen varianzas iguales	3,111	,097	7,843	16	,000	12.900	1.645	10.028 15.772
	No se asumen varianzas iguales			8,291	15,005	,000	12.900	1.556	10.172 15.628

Figura 5.12: Resultados de la prueba t para muestras independientes.

En nuestro ejemplo, $\text{Sig.} = 0.097$ y considerando $\alpha = 0.05$, no se rechaza la hipótesis de igualdad de varianzas. Por lo tanto, tomamos la fila correspondiente a *Se asumen varianzas iguales*. En la parte derecha de la tabla podemos observar el intervalo de confianza, que en este caso es $[10.028, 15.772]$.

Este intervalo indica que, si se repitiera el muestreo muchas veces y se construyeran intervalos del mismo modo, aproximadamente el 90% de ellos contendrían la verdadera diferencia de medias poblacionales ($\mu_A - \mu_B$). Existe evidencia estadística, con una confianza del 90%, de que el nivel medio de glucosa en el Grupo A es superior al del Grupo B, dado que el intervalo no incluye el cero y todos sus valores son positivos. Este procedimiento es fundamental para comparar tratamientos y tomar decisiones basadas en evidencia sobre los efectos de las dietas aplicadas.

5.4 Resumen de procedimientos

Capítulo 5: Intervalos de confianza

1. **Preparar la hoja de datos.** Asegurarse de que los datos están organizados correctamente:

- *Para una muestra:* una columna con los valores de la variable.
- *Para muestras relacionadas:* dos columnas, una para cada medición relacionada.
- *Para muestras independientes:* una columna con todos los valores y otra indicando el grupo.

2. **Calcular IC para una media.** Acceder al menú:

Analizar > Comparar medias > Prueba T para una muestra

Seleccionar la variable de interés, establecer el porcentaje del intervalo de confianza y pulsar Aceptar.

3. **Calcular IC para la diferencia de medias (muestras relacionadas).** Acceder al menú:

Analizar > Comparar medias > Prueba T para muestras relacionadas

Seleccionar las dos variables correspondientes a las mediciones relacionadas, elegir el nivel de confianza deseado y pulsar Aceptar.

4. **Calcular IC para la diferencia de medias (muestras independientes).** Acceder al menú:

Analizar > Comparar medias > Prueba T para muestras independientes

Mover la variable con los valores al cuadro *Variables de prueba* y la variable que indica el grupo al cuadro *Variable de agrupación*. Pulsar **Definir grupos**, asignar los códigos de los grupos, establecer el nivel de confianza y pulsar Aceptar.

5. **Interpretar los resultados.**

- En muestras independientes: identificar la fila correspondiente según la *Prueba de Levene*.
- En IC para la comparación de dos condiciones, examinar los límites del intervalo:
 - Si ambos son positivos, existe evidencia estadística de que la media de la primera variable o grupo es mayor que la de la segunda en la población.
 - Si ambos son negativos, existe evidencia estadística de que la media de la primera variable o grupo es menor que la de la segunda en la población.
 - Si el intervalo incluye el 0, no existe evidencia estadística suficiente para afirmar que las medias poblacionales sean distintas.

5.5 Ejercicios

Ejercicio 1

En un estudio sobre la calidad del ecosistema de un estanque, se midió la concentración de oxígeno disuelto (mg/L) en el agua en $n = 12$ puntos de muestreo seleccionados de forma aleatoria. Los valores obtenidos se presentan en la siguiente tabla:

Oxígeno	6.2	5.9	6.5	6.1	6.3	5.8	6.4	6.0	6.6	6.2	6.1	5.9
----------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Asumiendo que la concentración de oxígeno disuelto sigue una distribución normal, se solicita:

- Proporcionar una estimación puntual de la media poblacional.
- Construir un intervalo de confianza al 95% para la media poblacional de la concentración de oxígeno disuelto en el estanque.
- Interpretar los resultados obtenidos en el contexto del problema.

Ejercicio 2

Se realiza un experimento para evaluar el efecto de un suplemento vitamínico sobre la tasa de crecimiento de larvas de mosca. Para ello, se registró el peso (mg) de $n = 8$ larvas seleccionadas aleatoriamente, realizando una medición antes de iniciar el tratamiento y otra tras 5 días de administración del suplemento. Los datos obtenidos se muestran en la siguiente tabla:

Larva	Peso Antes	Peso Después
1	12.1	14.3
2	11.8	13.5
3	12.5	15.0
4	11.9	14.0
5	12.3	14.5
6	12.0	14.1
7	11.7	13.8
8	12.2	14.6

Asumiendo que la diferencia de peso ($d = \text{Después} - \text{Antes}$) sigue una distribución normal:

- Construir un intervalo de confianza al 99% para la media poblacional de las diferencias (μ_d).
- A partir del intervalo anterior, determinar si existe evidencia estadística suficiente para afirmar, con un nivel de confianza del 99%, que el suplemento produce un incremento significativo en el peso medio de las larvas.

Ejercicio 3

Se realizó un experimento para evaluar el efecto de la dieta sobre la concentración de glucosa en sangre (mg/dL) en ranas. Se asignaron aleatoriamente $n_A = 7$ ranas al Grupo A (dieta alta en carbohidratos) y $n_B = 7$ ranas al Grupo B (dieta baja en carbohidratos), y se registraron los valores de glucosa tras tres semanas de tratamiento. Los resultados se presentan en la siguiente tabla:

Grupo A (alta en carbohidratos)	Grupo B (baja en carbohidratos)
105	98
110	95
108	97
112	100
107	96
111	99
109	94

Asumiendo que las concentraciones de glucosa en ambos grupos siguen distribuciones normales y que las muestras son independientes:

- Construir un intervalo de confianza al 90% para la diferencia de medias poblacionales ($\mu_A - \mu_B$).
- A partir del intervalo obtenido, determinar si existe evidencia estadística suficiente para afirmar, con un nivel de confianza del 90%, que la dieta alta en carbohidratos produce un incremento significativo en la concentración media de glucosa respecto a la dieta baja en carbohidratos.

Capítulo 6

Contrastes de hipótesis sobre medias poblacionales



Los *contrastes de hipótesis* (CH) constituyen una herramienta esencial dentro del análisis estadístico inferencial. Su propósito es determinar, a partir de la información muestral, si existen evidencias suficientes para rechazar o no rechazar una afirmación sobre una población. En el ámbito de la biología, se emplean para valorar si un tratamiento tiene un efecto, si dos grupos muestran diferencias estadísticamente significativas en sus resultados o si la media poblacional se aparta de un valor teórico de referencia. En este capítulo se presentan los contrastes de hipótesis sobre medias poblacionales en tres situaciones: una muestra, dos muestras relacionadas y dos muestras independientes. Se expone el procedimiento general para formular e interpretar un contraste. Asimismo, se describe cómo llevar a cabo cada análisis en SPSS. El objetivo es que el estudiante adquiera la capacidad de aplicar correctamente los contrastes de hipótesis en estudios biológicos comprendiendo el desarrollo práctico de cada procedimiento.

6.1 Formulación e interpretación

Esta sección presenta el procedimiento general para formular e interpretar contrastes de hipótesis sobre medias poblacionales en el contexto de la prueba t . La Tabla 6.1 describe las posibles formas de establecer la hipótesis nula y la alternativa, así como los criterios de decisión en función del estadístico t , del p -valor o significación bilateral (p_{bi}), del p -valor unilateral ($p_{uni} = p_{bi}/2$) y del nivel de significación (α). Por claridad didáctica, se plantean las hipótesis comparando directamente los parámetros poblacionales (μ_1 y μ_2), lo que permite visualizar de forma intuitiva el sentido de la diferencia buscada¹.

Tipo de contraste	Bilateral	Unilateral izquierdo	Unilateral derecho
Hipótesis nula H_0	$\mu_1 = \mu_2$	$\mu_1 \geq \mu_2$	$\mu_1 \leq \mu_2$
Hipótesis alternativa H_1	$\mu_1 \neq \mu_2$	$\mu_1 < \mu_2$	$\mu_1 > \mu_2$
Regla de rechazo de H_0	Rechazar si $p_{bi} < \alpha$	Rechazar si $t < 0$ y $p_{uni} < \alpha$	Rechazar si $t > 0$ y $p_{uni} < \alpha$

Tabla 6.1: Tipos de hipótesis y criterios de decisión en la prueba t en SPSS.

En la Tabla 6.1, los parámetros μ_1 y μ_2 se adaptan al diseño de la investigación:

1. *CH para una muestra:* μ_1 es la media poblacional (μ) y μ_2 es el valor de prueba (μ_0).
2. *CH para muestras relacionadas:* μ_1 es la media del primer momento (usualmente *Antes*) y μ_2 es la media del segundo momento (*Después*).
3. *CH para muestras independientes:* μ_1 es la media del Grupo 1 y μ_2 es la media del Grupo 2.

Para una interpretación rigurosa del contraste, conviene seguir estas pautas:

- **Estructura de las hipótesis.** La hipótesis nula (H_0) siempre incluye la igualdad ($=, \geq, \leq$), mientras que la hipótesis alternativa (H_1) determina la dirección del contraste ($\neq, <, >$).
- **Interpretación del p -valor.**
 - *Contraste bilateral.* Se utiliza el p -valor bilateral (p_{bi}). Se rechaza H_0 cuando $p_{bi} < \alpha$.
 - *Contraste unilateral.* El rechazo de H_0 requiere que el signo del estadístico t sea coherente con la dirección marcada por H_1 y que el p -valor unilateral (p_{uni}) sea inferior a α . Si el signo de t es opuesto al esperado por H_1 , no procede el rechazo.
- **Nivel de significación.** El valor de α debe fijarse antes de ejecutar el contraste. Los niveles más habituales son 0.01 y 0.05, y en algunos contextos también se emplea 0.1.

¹En contrastes unilaterales, aunque H_0 abarca un rango de valores ($\geq$ o $\leq$), el cálculo del estadístico se realiza siempre asumiendo el caso límite de igualdad ($\mu_1 = \mu_2$)

6.2 CH sobre la media poblacional (una muestra)

En esta sección se describe cómo aplicar un contraste de hipótesis sobre la media poblacional y cómo interpretar los resultados generados por SPSS mediante la prueba t para una muestra.

Para ilustrarlo, consideremos un estudio biológico en el que se registra el tiempo (en minutos) que tardan determinadas larvas en entrar en estado de reposo tras la exposición a un estímulo ambiental. Los datos obtenidos en la muestra son los siguientes:

69.30, 56.30, 22.10, 47.60, 53.20, 48.10, 23.20, 23.80.

El objetivo es evaluar si el tiempo medio poblacional (μ) es significativamente superior a un valor de prueba de 20 minutos, con un nivel de significación $\alpha = 0.05$. El contraste se plantea como:

$$H_0 : \mu \leq 20$$

$$H_1 : \mu > 20$$

Se trata de un contraste unilateral derecho. El análisis permitirá establecer si los datos aportan evidencia suficiente para rechazar H_0 en favor de un tiempo superior a 20 minutos.

Para ejecutar este contraste en SPSS se debe acceder a:

Analizar > Comparar medias > Prueba T para una muestra

En la ventana correspondiente:

1. Trasladar la variable **Tiempo** al cuadro *Variables de prueba* (Figura 6.1).

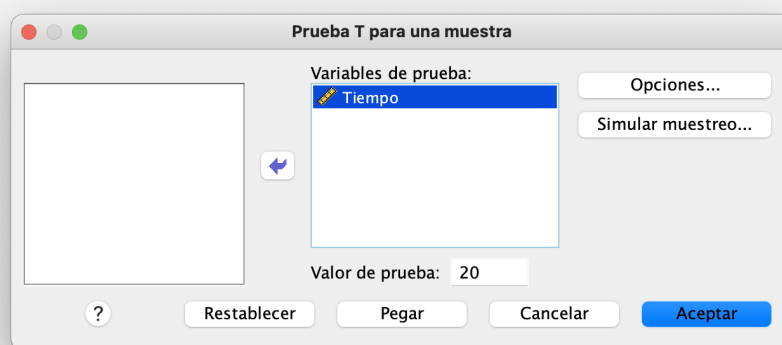


Figura 6.1: Ventana de selección de variables.

2. En el recuadro *Valor de prueba*, introducir el valor de referencia: 20.

3. Acceder a Opciones y fijar el nivel de confianza del intervalo en 95% (Figura 6.2).

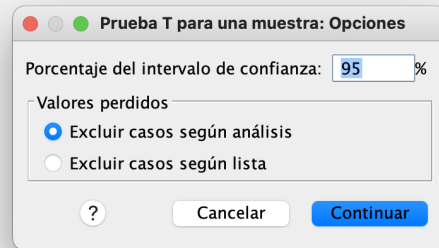


Figura 6.2: Configuración del nivel de confianza.

4. Pulsar Aceptar para obtener los resultados (Figura 6.3).

Prueba para una muestra						
Valor de prueba = 20						
	t	gl	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza de la diferencia	
					Inferior	Superior
Tiempo	3,647	7	,008	22.95000	8.0709	37.8291

Figura 6.3: Salida de SPSS para la prueba t para una muestra.

La salida de SPSS indica un estadístico $t = 3.647$. Al ser un valor positivo, existe concordancia con la dirección de la hipótesis alternativa ($H_1 : \mu > 20$).

La significación bilateral reportada es $p_{bi} = 0.008$. Para nuestro contraste unilateral, calculamos:

$$p_{uni} = \frac{p_{bi}}{2} = \frac{0.008}{2} = 0.004$$

Al cumplirse las dos condiciones establecidas en la regla de decisión para un contraste unilateral derecho ($t > 0$ y $p_{uni} < 0.05$), se concluye que existen evidencias suficientes para **rechazar la hipótesis nula** H_0 . Por tanto, se acepta, con un nivel de significación de 0.05, que el tiempo medio poblacional que tardan las larvas en entrar en reposo es superior a 20 minutos.

La tabla de resultados también muestra un intervalo de confianza al 95% para la diferencia entre la media poblacional y el valor de prueba (20). El hecho de que este intervalo esté compuesto íntegramente por valores positivos, y por tanto no incluya el valor 0, es coherente con el rechazo de la hipótesis nula.

6.3 CH sobre la diferencia de medias (muestras relacionadas)

Para ilustrar un contraste de hipótesis sobre la diferencia de medias en muestras relacionadas, consideremos un estudio sobre el efecto de un tratamiento de filtrado en la reducción de metales pesados en acuarios. Se midió la concentración de metales (en mg/L) en once acuarios antes y después de aplicar el tratamiento. Los datos obtenidos son los siguientes:

Concentración previa	198	210	194	220	138	220	219	161	210	313	226
Concentración posterior	65	77	94	73	37	131	77	24	99	321	57

Puesto que cada par de mediciones se realiza sobre el mismo sujeto, los datos son dependientes. El objetivo es determinar si el tratamiento reduce significativamente la concentración media de metales ($\alpha = 0.05$). El contraste se plantea como:

$$H_0 : \mu_1 \leq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

Este es un contraste unilateral derecho. Si el tratamiento es efectivo, la media inicial (μ_1) debe ser significativamente mayor que la posterior (μ_2).

Para ejecutar este contraste en SPSS se debe acceder a:

Analizar > Comparar medias > Prueba T para muestras relacionadas

1. Seleccionar las variables: primero *Antes* y luego *Después* (Figura 6.4). Mantener este orden asegura que SPSS calcule la diferencia como *Antes – Después*.

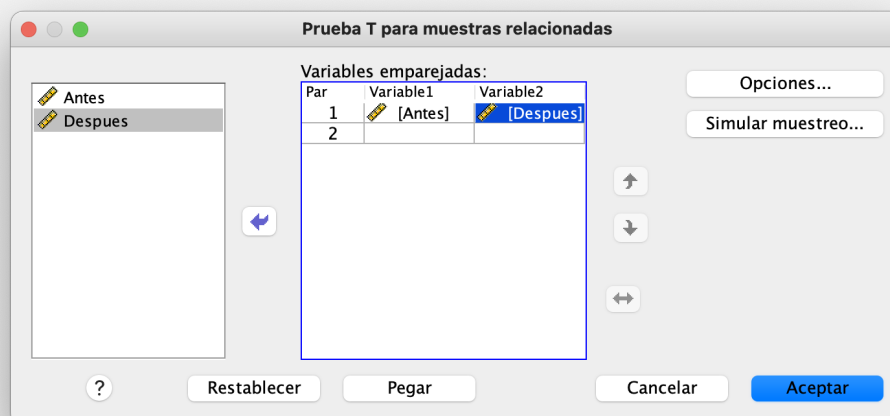


Figura 6.4: Selección de variables en SPSS.

2. Acceder a Opciones y fijar el nivel de confianza del intervalo en 95% (Figura 6.5).

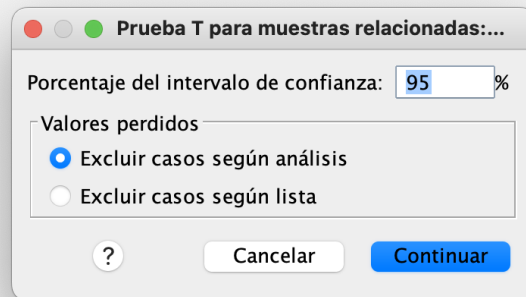


Figura 6.5: Configuración del intervalo de confianza en SPSS.

3. Pulsar Aceptar. SPSS mostrará los resultados de la prueba t (Figura 6.6).

Prueba de muestras emparejadas									
Diferencias emparejadas									
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
Par 1	Antes – Despues	114.000	46.938	14.152	Inferior	Superior	8,055	10	,000

Figura 6.6: Resultados de la prueba t para muestras relacionadas en SPSS.

Los resultados muestran un estadístico $t = 8.055$. Al ser un valor positivo, existe concordancia con la dirección de la hipótesis alternativa ($H_1 : \mu_1 > \mu_2$).

La significación bilateral reportada es $p_{bi} < 0.001$. Calculamos la significación unilateral:

$$p_{uni} = \frac{p_{bi}}{2} < 0.0005$$

Al cumplirse ambas condiciones ($t > 0$ y $p_{uni} < 0.05$), se **rechaza la hipótesis nula** H_0 . Existe evidencia estadísticamente significativa de que el nivel medio de metales pesados disminuye tras el tratamiento.

El intervalo de confianza del 95% para la diferencia es $[82.47, 145.53]$. El hecho de que sea íntegramente positivo (no incluye el 0) es coherente con el rechazo de H_0 y estima que la reducción media se sitúa entre dichos valores.

6.4 CH sobre la diferencia de medias (muestras independientes)

Retomando el ejemplo sobre el tiempo que tardan las larvas en entrar en estado de reposo tras un estímulo ambiental, se estudia ahora si existen diferencias estadísticamente significativas entre las medias poblacionales de dos grupos independientes de larvas.

Para ello se plantea un contraste de hipótesis ($\alpha = 0.05$) sobre la diferencia de medias en dos muestras independientes. Los datos disponibles (en minutos hasta el reposo) son los siguientes:

Grupo 1	69.30	56.30	22.10	47.60	53.20	48.10	23.20	23.80		
Grupo 2	18.60	25.10	26.40	14.90	29.80	28.40	12.10	10.20	11.60	12.80

El contraste bilateral es el siguiente:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

donde μ_1 y μ_2 representan las medias poblacionales del tiempo en los Grupos 1 y 2, respectivamente.

Para ejecutar el contraste, los datos deben organizarse en una columna para la variable cuantitativa (Datos) y otra para la variable de agrupación (Grupo), como se detalló en el Capítulo 5 (Figura 6.7).

	Grupo1	Grupo2	Datos	Grupo	var	var	var	var	var	var	var	var
1	69.30	18.60	69.30	Grupo1								
2	56.30	25.10	56.30	Grupo1								
3	22.10	26.40	22.10	Grupo1								
4	47.60	14.90	47.60	Grupo1								
5	53.20	29.80	53.20	Grupo1								
6	48.10	28.40	48.10	Grupo1								
7	23.20	12.10	23.20	Grupo1								
8	23.80	10.20	23.80	Grupo1								
9	.	11.60	18.60	Grupo2								
10	.	12.80	25.10	Grupo2								
11	.	.	26.40	Grupo2								
12	.	.	14.90	Grupo2								
13	.	.	29.80	Grupo2								
14	.	.	28.40	Grupo2								
15	.	.	12.10	Grupo2								
16	.	.	10.20	Grupo2								
17	.	.	11.60	Grupo2								
18	.	.	12.80	Grupo2								
19												

Figura 6.7: Entrada de datos en SPSS para el contraste de medias en muestras independientes.

El acceso al análisis se realiza a través de la ruta:

Analizar > Comparar medias > Prueba T para muestras independientes

En la ventana correspondiente:

1. Trasladar la variable Datos al cuadro *Variables de prueba* y la variable Grupo al cuadro *Variable de agrupación* (Figura 6.8).

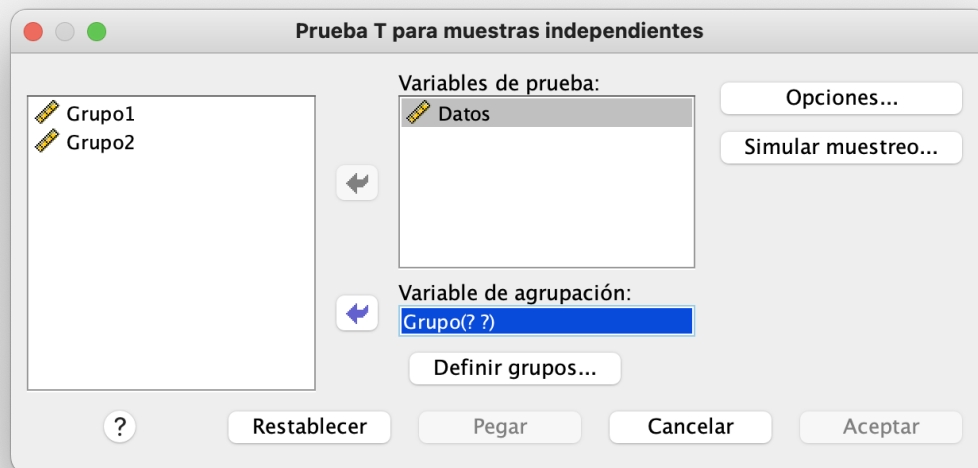


Figura 6.8: Asignación de variables en SPSS para la prueba t de muestras independientes.

2. Pulsar Definir grupos y especificar los códigos correspondientes (1 y 2), de modo que SPSS identifique las categorías a comparar (Figura 6.9).

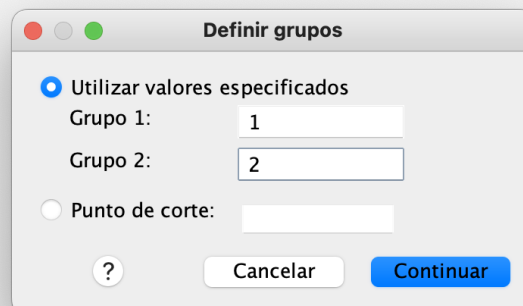


Figura 6.9: Definición de los códigos de grupo.

3. Fijar el nivel de confianza desde Opciones y pulsar Aceptar. SPSS generará la tabla *Prueba de muestras independientes* con los resultados del contraste (Figura 6.10).

Prueba de muestras independientes									
		Prueba de Levene de igualdad de varianzas		prueba t para la igualdad de medias					
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia
Datos	Se asumen varianzas iguales	9,369	,007	3,853	16	,001	23.96000	6.21805	10.77833 37.14167
	No se asumen varianzas iguales			3,552	9,089	,006	23.96000	6.74590	8.72254 39.19746

Figura 6.10: Resultados de la prueba t para muestras independientes en SPSS.

En la salida de SPSS aparecen dos filas de resultados. La elección de la fila adecuada depende del resultado de la prueba de Levene para la igualdad de varianzas:

- Si la significación de Levene es $p < \alpha$: Se rechaza la hipótesis de igualdad de varianzas. Se debe interpretar la fila *No se asumen varianzas iguales*.
- Si la significación de Levene es $p \geq \alpha$: No existen evidencias para rechazar la igualdad de varianzas. Se interpreta la fila *Se asumen varianzas iguales*.

En este caso, la prueba de Levene proporciona una significación de 0.007. Al ser inferior a $\alpha = 0.05$, se rechaza la igualdad de varianzas poblacionales y se interpretan los resultados de la fila inferior (*No se asumen varianzas iguales*). Para dicha fila, la prueba t proporciona una significación bilateral de:

$$p_{bi} = 0.006$$

Dado que $p_{bi} < 0.05$, **se rechaza la hipótesis nula H_0** . Se concluye que existe evidencia estadísticamente significativa de que el tiempo medio hasta el reposo difiere entre los dos grupos de larvas.

El intervalo de confianza del 95% para la diferencia de medias ($\mu_1 - \mu_2$) es aproximadamente [8.72, 39.20]. El hecho de que sea íntegramente positivo (no incluye el 0) es coherente con el rechazo de H_0 y estima que la verdadera diferencia de medias entre ambos grupos se sitúa entre dichos valores.

6.5 Resumen de procedimientos

Capítulo 6: Contrastes de hipótesis sobre medias poblacionales

1. **Prueba t para una muestra.** Para evaluar si la media poblacional difiere de un valor teórico de referencia:

Analizar > Comparar medias > Prueba T para una muestra

Introducir la variable en *Variables de prueba* y el valor de referencia en *Valor de prueba*. En *Opciones*, fijar el nivel de confianza para el IC.

2. **Prueba t para muestras relacionadas.** Para comparar las medias de dos mediciones dependientes (por ejemplo, en un diseño antes-después) sobre los mismos sujetos:

Analizar > Comparar medias > Prueba T para muestras relacionadas

Seleccionar el par de variables, fijar el nivel de confianza y ejecutar.

3. **Prueba t para muestras independientes.** Para comparar las medias de dos grupos físicamente distintos:

Analizar > Comparar medias > Prueba T para muestras independientes

- Mover la variable de datos a *Variables de prueba* y la variable de grupo a *Variable de agrupación*.
- Pulsar **Definir grupos** e introducir los códigos numéricos correspondientes.
- En *Opciones*, fijar el nivel de confianza para el IC.
- Ejecutar y seleccionar la fila de la tabla según el resultado de la prueba de Levene (igualdad de varianzas).

4. **Interpretación de resultados.** Criterios generales de decisión:

- *Contraste bilateral.* Se rechaza H_0 si $p_{bi} < \alpha$.
- *Contraste unilateral.* Se calcula la significación unilateral ($p_{uni} = p_{bi}/2$). El rechazo de H_0 requiere que el signo del estadístico t sea coherente con la dirección de H_1 y que $p_{uni} < \alpha$.

6.6 Ejercicios

Ejercicio 1

Un grupo de biólogos estudia la longitud de las hojas de una determinada especie vegetal en un bosque protegido. Estudios previos indican que la media poblacional de la longitud de las hojas en esta especie es de 50 mm.

Con el objetivo de evaluar si, en la población actual, la media poblacional (μ) de la longitud de las hojas difiere de dicho valor de referencia, se selecciona una muestra aleatoria simple de $n = 12$ hojas y se mide su longitud (en mm), obteniéndose los siguientes valores:

52.1, 47.8, 51.3, 49.5, 50.7, 48.9, 53.2, 51.0, 49.8, 50.5, 52.4, 48.6

Suponiendo que la distribución de la longitud de las hojas en la población es normal, determine si existe evidencia estadísticamente significativa para concluir que la media poblacional de la longitud de las hojas difiere del valor de referencia de 50 mm, utilizando un nivel de significación $\alpha = 0.05$.

Ejercicio 2

Se desea evaluar si un nuevo suplemento alimenticio produce un aumento en el crecimiento de polluelos de gallina durante el primer mes de vida. Para ello, se mide la altura (en cm) de cada polluelo al inicio y al final del periodo de estudio, considerándose las mediciones como observaciones pareadas correspondientes a los mismos individuos. Se selecciona una muestra de $n = 8$ polluelos, obteniéndose los siguientes valores:

Altura inicial	12.3	11.8	12.0	11.5	12.1	11.7	12.2	11.9
Altura final	13.1	12.6	12.9	12.2	12.8	12.3	12.9	12.5

Definiendo, para cada polluelo, la variable diferencia como la altura final menos la altura inicial, y suponiendo que la distribución de dichas diferencias en la población es aproximadamente normal, determine si existe evidencia estadísticamente significativa para concluir que el suplemento alimenticio produce un aumento en la altura media de los polluelos, utilizando un nivel de significación $\alpha = 0.01$.

Ejercicio 3

Se investiga si dos especies de peces que habitan en lagos distintos presentan diferencias en su actividad nocturna, medida en número de movimientos por hora. Se selecciona una muestra aleatoria de individuos de cada especie, y se registra la actividad de cada uno durante una hora:

Especie A	15	12	17	14	16	13	18
Especie B	20	22	19	21	23	20	24

Suponiendo que las distribuciones de actividad en cada especie son aproximadamente normales y que las muestras son independientes, determine si existe evidencia estadísticamente significativa de que la especie A tiene menor actividad nocturna que la especie B, utilizando un nivel de significación $\alpha = 0.1$.

Capítulo 7

Otros contrastes de hipótesis



En este capítulo se amplían los *contrastos de hipótesis* incorporando pruebas estadísticas de uso habitual que no se centran exclusivamente en la comparación de medias. En particular, se aborda el contraste de Kolmogorov-Smirnov de una muestra, utilizado para contrastar la adecuación de los datos a una distribución teórica (habitualmente la normal). También se analiza el contraste de la pendiente en el modelo de regresión lineal simple, que permite evaluar si la pendiente poblacional difiere de cero y, por tanto, si existe evidencia de una relación lineal estadísticamente significativa entre las variables. Finalmente, se estudia el contraste binomial, empleado para comprobar si la proporción poblacional asociada a una variable dicotómica coincide con un valor teórico preestablecido. El objetivo de este capítulo es proporcionar una guía práctica para la aplicación de estos procedimientos en SPSS, así como para la correcta interpretación de sus resultados.

7.1 CH para la normalidad

Antes de aplicar la mayoría de los métodos paramétricos, es necesario evaluar el supuesto de normalidad, es decir, comprobar si los datos proceden de una población cuya distribución pueda considerarse normal. Entre las herramientas disponibles en SPSS, el *contraste de Kolmogorov-Smirnov* (K-S) de una muestra constituye una opción habitual para contrastar la adecuación de una variable cuantitativa a una distribución teórica normal¹.

Para ilustrar su uso, consideremos un estudio en el que se desea evaluar la normalidad de los pesos (en gramos) de ratones adultos, con un nivel de significación $\alpha = 0.05$. Los datos registrados son:

22.5, 21.8, 23.1, 22.9, 23.5, 21.7, 22.3, 22.8, 23.0, 22.6, 23.2, 22.4

El contraste de hipótesis del test K-S se formula como:

- H_0 : La distribución poblacional de la variable es normal.
- H_1 : La distribución poblacional de la variable no es normal.

En SPSS, el procedimiento para realizar esta prueba estimando los parámetros a partir de la muestra (con corrección de Lilliefors) es el siguiente:

Analizar > Pruebas no paramétricas > Cuadros de diálogo antiguos > K-S de 1 muestra

En la ventana correspondiente, se debe mover la variable que se desea contrastar (en este caso, Pesos) al cuadro *Lista de variables de prueba* y pulsar **Aceptar** (Figura 7.1).

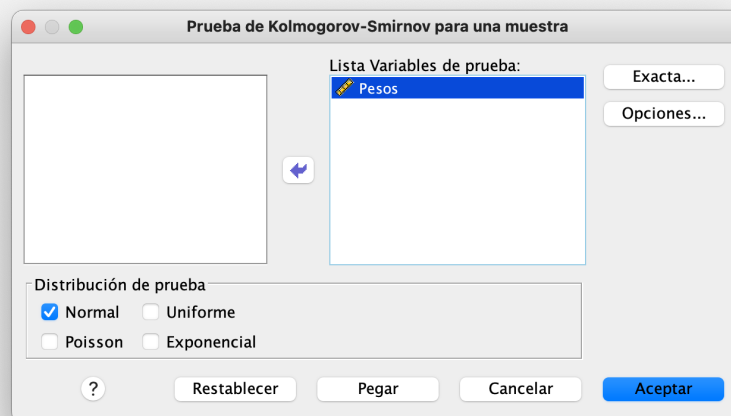
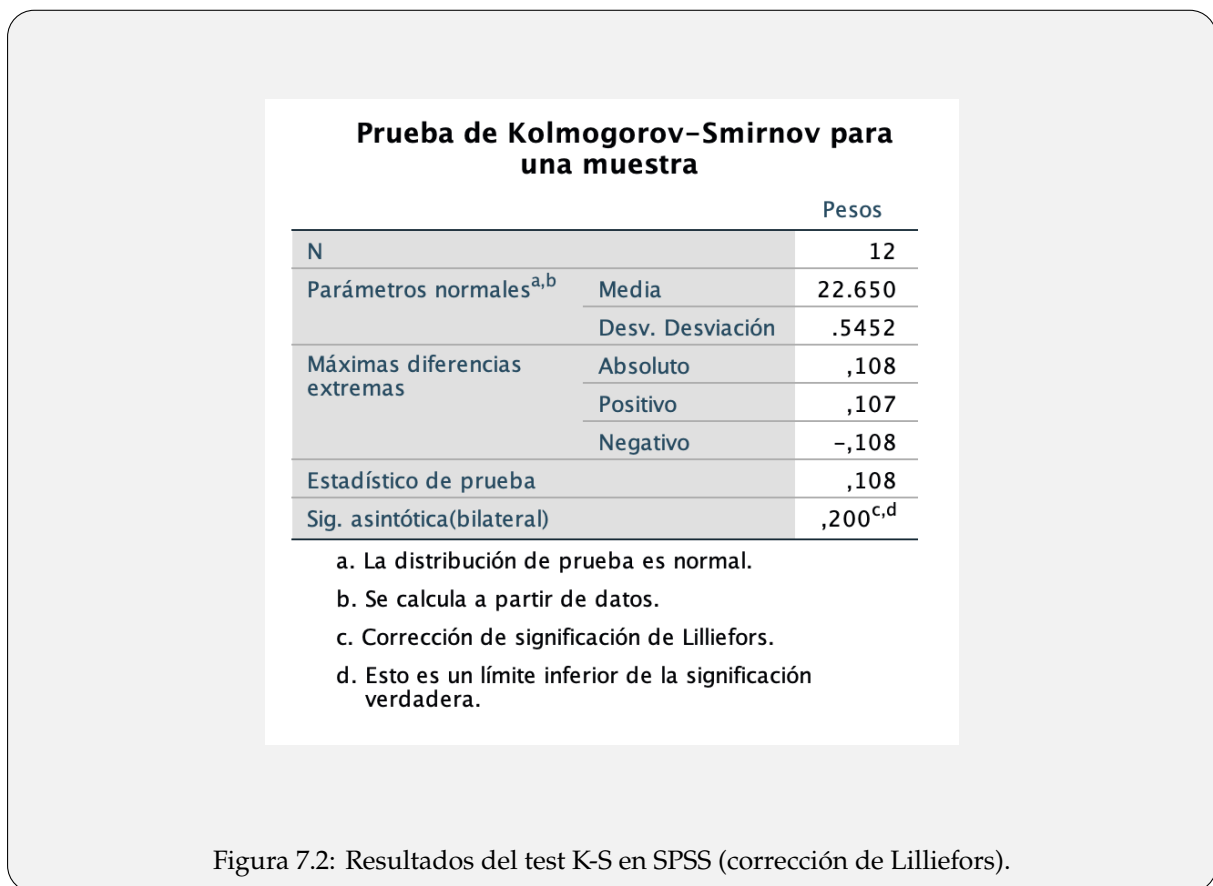


Figura 7.1: Cuadro de diálogo del test K-S en SPSS.

¹Especialmente en muestras pequeñas ($n \leq 50$) suele recomendarse el test de Shapiro-Wilk por su mayor potencia para detectar desviaciones de la normalidad, accesible desde *Analizar > Estadísticos descriptivos > Explorar > Gráficos > Gráficos de normalidad con pruebas*. Aquí se usa el test de K-S principalmente con fines ilustrativos para mostrar el procedimiento en SPSS.

El programa proporciona el valor del estadístico de Kolmogorov-Smirnov y el correspondiente p -valor (Figura 7.2). La decisión se toma según los criterios habituales:

- Si el p -valor (*Sig. asintótica bilateral*) es menor que el nivel de significación ($\alpha = 0.05$), se rechaza H_0 , lo que indica que los datos presentan desviaciones significativas respecto a la normalidad.
- Si el p -valor es mayor o igual que α , no se rechaza H_0 , lo que indica que los datos son compatibles con una distribución normal.



En este ejemplo, el contraste de Kolmogorov-Smirnov proporciona un p -valor mayor que 0.05. En consecuencia, no se dispone de evidencia estadística suficiente para rechazar la hipótesis nula, por lo que los datos resultan compatibles con una distribución normal.

Además del contraste K-S, es recomendable complementar la evaluación de la normalidad mediante herramientas gráficas como histogramas o gráficos de probabilidad normal. Estas representaciones permiten detectar desviaciones estructurales que pueden no reflejarse claramente en los contrastes numéricos, especialmente en muestras pequeñas. La combinación de pruebas estadísticas y exploración gráfica proporciona una base sólida para valorar el cumplimiento de los supuestos necesarios para el uso de métodos paramétricos.

7.2 CH para la pendiente en regresión lineal

En un análisis de regresión lineal, además de estimar la recta que relaciona dos variables, es fundamental evaluar si dicha relación es estadísticamente significativa. Con este fin se aplica un *contraste de hipótesis para la pendiente de regresión*, que permite contrastar si la pendiente poblacional difiere de cero y, por tanto, si existe evidencia de una relación lineal entre las variables estudiadas. Para ilustrar el procedimiento, consideremos un estudio que investiga la relación entre la temperatura de incubación (en °C) y el número de bacterias (en miles) presentes en 10 cultivos. Los datos registrados son los siguientes:

Temperatura	20.0	22.0	24.0	26.0	28.0	30.0	32.0	34.0	36.0	38.0
Número de bacterias	15.2	18.3	22.1	25.5	28.9	33.0	36.4	39.8	43.2	46.5

El contraste de interés es:

- $H_0 : \beta_1 = 0$ (la temperatura no presenta un efecto lineal sobre el número de bacterias),
- $H_1 : \beta_1 \neq 0$ (la temperatura presenta un efecto lineal sobre el número de bacterias),

donde β_1 es la pendiente poblacional de la recta de regresión. Para realizar este contraste, acceder a:

Analizar > Regresión > Lineales

A continuación, se introduce la variable dependiente (Bacterias) en el cuadro *Dependientes*, y la variable independiente (Temperatura) en *Independientes* (Figura 7.3). Finalmente, pulsar Aceptar.

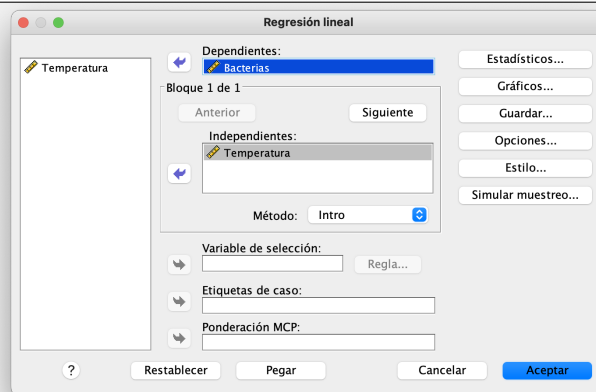


Figura 7.3: Selección de variables para el análisis de regresión en SPSS.

En el caso de la regresión lineal simple, el contraste global del modelo que aparece en la tabla ANOVA es equivalente al contraste de la pendiente². Esta tabla proporciona un *p*-valor (*Sig.*) inferior a 0.001, menor que el nivel de significación ($\alpha = 0.05$). Por tanto, se rechaza H_0 , lo que indica que existe evidencia estadísticamente significativa de una relación lineal entre la temperatura y el número de bacterias.

ANOVA ^a						
Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1023,969	1	1023,969	20466,971	,000 ^b
	Residuo	,400	8	,050		
	Total	1024,369	9			

a. Variable dependiente: Bacterias

b. Predictores: (Constante), Temperatura

Figura 7.4: Tabla ANOVA del modelo de regresión en SPSS.

²El *p*-valor también puede consultarse en la tabla de coeficientes asociado a la pendiente, proporcionando la misma información sobre la significación de la relación lineal.

7.3 CH para la proporción poblacional

En algunos estudios, se desea comprobar si la proporción de individuos que presentan una determinada característica coincide con un valor teórico o difiere significativamente de él. Este tipo de análisis se aplica cuando la variable de interés es cualitativa dicotómica, es decir, cuando solo puede adoptar dos valores posibles (por ejemplo, sí/no, positivo/negativo o normal/anormal).

Consideremos un estudio que evalúa la efectividad de un nuevo tratamiento para reducir la presencia de un parásito en un grupo de peces. Según estudios previos, aproximadamente el 70% de los peces tratados con el método estándar quedan libres del parásito. Se desea comprobar si el nuevo tratamiento aumenta la proporción de peces libres del parásito en la población. Para ello, se toma una muestra de 40 peces, de los cuales 32 quedaron libres del parásito tras el tratamiento. El CH se plantea como:

$H_0 : p \leq 0.70$ (el nuevo tratamiento no aumenta la proporción de peces libres de parásitos)

$H_1 : p > 0.70$ (el nuevo tratamiento aumenta la proporción de peces libres de parásitos)

donde p representa la proporción poblacional de peces libres de parásitos con el nuevo tratamiento. En SPSS, el contraste es unilateral y la dirección de H_1 depende de si la proporción observada es mayor o menor que la proporción de prueba.

Para analizar los datos en SPSS, la variable Libres debe ser numérica y codificarse, por ejemplo, como 1 = Libre y 2 = No libre. Dado que disponemos de frecuencias (32 libres y 8 no libres), es necesario ponderar los casos:

Datos > Ponderar casos

y seleccionar la variable que contiene las frecuencias (Figura 7.5).

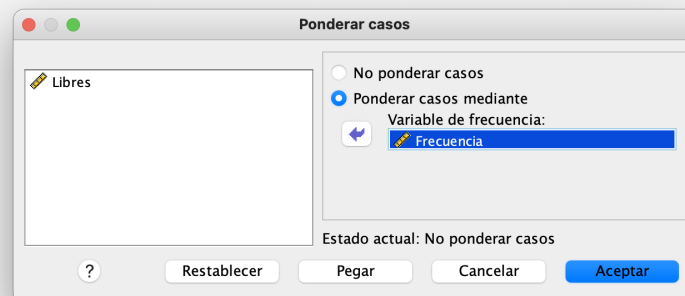


Figura 7.5: Ponderación de casos en SPSS.

A continuación, se realiza el contraste de hipótesis mediante el procedimiento binomial en SPSS:

Analizar > Pruebas no paramétricas > Cuadros de diálogo antiguos > Binomial

1. Seleccionar la variable Libres en el cuadro de variables.
2. Especificar la *Proporción de prueba* como 0.70 y pulsar **Aceptar** para ejecutar el análisis.

SPSS proporciona la proporción observada en la muestra ($32/40 = 0.80$) y el p -valor unilateral asociado al contraste binomial (0.111). Como $0.111 > 0.05$, no se rechaza la hipótesis nula H_0 . Por tanto, no existe evidencia estadística suficiente para afirmar que el nuevo tratamiento supere el 70% de peces libres de parásitos.

7.4 Resumen de procedimientos

Capítulo 7: Otros contrastes de hipótesis

1. **CH para la normalidad (Kolmogorov-Smirnov).** Para evaluar si los datos provienen de una distribución normal:

Analizar > Pruebas no paramétricas > Cuadros de diálogo antiguos > K-S de 1 muestra

Introducir la variable a contrastar en el cuadro *Lista de variables de prueba* y pulsar **Aceptar**. Interpretar el p -valor: si $p < \alpha$, se rechaza H_0 (los datos no son compatibles con la normalidad); si $p \geq \alpha$, no se rechaza H_0 .

2. **CH para la pendiente en regresión lineal.** Permite evaluar si existe una relación lineal significativa entre dos variables:

Analizar > Regresión > Lineales

Asignar la variable dependiente e independiente, pulsar **Aceptar** y consultar la tabla ANOVA. Si $p < \alpha$, se rechaza $H_0 : \beta_1 = 0$, indicando que existe evidencia estadísticamente significativa de una relación lineal entre las variables.

3. **CH para la proporción poblacional.** Permite comprobar si la proporción observada de una variable dicotómica difiere de un valor teórico:

Analizar > Pruebas no paramétricas > Cuadros de diálogo antiguos > Binomial

Si los datos se presentan en forma de frecuencias, es necesario ponderar los casos antes de realizar el contraste. Para hacer el contraste, seleccionar la variable dicotómica, indicar la proporción de prueba y pulsar **Aceptar**. Interpretar el p -valor: si $p < \alpha$, se rechaza H_0 .

7.5 Ejercicios

Ejercicio 1

Un biólogo mide el peso (en gramos) de 15 ratas adultas en un experimento de alimentación:

250, 245, 252, 248, 251, 247, 249, 253, 246, 250, 254, 248, 247, 251, 249

Utilizar los contrastes de normalidad en SPSS para comprobar, con un nivel de significación del 5%, si los pesos pueden considerarse normalmente distribuidos.

Ejercicio 2

Se desea estudiar la relación entre la concentración de oxígeno disuelto en agua (mg/L) y la tasa de crecimiento de algas (cm/semana) en 8 acuarios:

O ₂	4	5	6	7	8	9	10	11
Crecimiento	2.1	2.3	2.5	2.6	2.9	3.0	3.2	3.4

Realizar un contraste de hipótesis para la pendiente (β_1) mediante un análisis de regresión lineal simple en SPSS. Determinar, con un nivel de significación del 5%, si existe una relación lineal significativa entre ambas variables.

Ejercicio 3

En un estudio sobre la efectividad de un nuevo desinfectante, se observa que de 60 cultivos tratados, 45 quedan libres de contaminación bacteriana. Se sabe que el desinfectante estándar logra un 70% de cultivos libres.

Realizar un contraste binomial para la proporción poblacional p y comprobar, con un nivel de significación del 5%, si el nuevo desinfectante mejora la proporción de cultivos libres respecto al estándar.

Bibliografía

Angulo Díaz-Parreño, S., Cárdenas Rebollo, J. M., Romero Limón, A., & Sánchez Alberca, A. (2014). *Bioestadística aplicada con SPSS*. CEU Ediciones.

Canavos, G. C. (2003). *Probabilidad y Estadística: Aplicaciones y métodos*. McGraw-Hill / Interamericana de México.

Cobo, E., Muñoz, P., & González, J. A. (2007). *Bioestadística para no estadísticos: Bases para interpretar artículos científicos*. Elsevier-Masson.

Lara Porras, A. M. (2010). *Estadística para Biología y Ciencias Ambientales: Tratamiento informático mediante SPSS*. Proyecto Sur.

Milton, J. S. (2007). *Estadística para Biología y Ciencias de la Salud*. McGraw-Hill / Interamericana de España.

Ross, S. (2007). *Introducción a la Estadística*. Reverté.

Sánchez, M., Frutos, G., & Cuesta, P. L. (1996). *Estadística y matemáticas aplicadas*. Síntesis.

Santabárbara Serrano, J., Rubio Aranda, E., Feja Solana, C., & Martínez Terrer, T. (2015). *Manual de bioestadística aplicada con IBM SPSS*. Andavira.

