COMPARACION DE PROPORCIONES EN TABLAS 2x2

(una extensión particular a tablas rxs)

por

JUAN DE DIOS LUNA DEL CASTILLO

Tesis Doctoral realizada en el Departamento de Estadistica e Investigación Operativa, bajo la dirección del Prof. Dr. D. Antonio Martín Andrés.

Granada, 31 de Marzo de 1987

15/95

DILIGENCIA por la que se hace constar que la presente memoria fue defendida por D. Juan de Dios Luna del Castillo el día 24 de Abril de 1987 ante el tribunal compuesto por los profesores

Presidente: D. Ramon Gutierrez Jaimez
Vocales: D. Luis Parras Guijosa
D. Antonino García Rendon
D. Francisco Requena Guerrero
Secretario: D. Andrés González Carmona
obteniendo la calificación de

Apto cum laude por unanimidad.

R. 32.857

COMPARACION DE PROPORCIONES EN TABLAS 2x2

(una extensión particular a tablas rxs)

Memoria que para optar al grado de Doctor en Ciencias, Sección de Matemáticas, presenta el licenciado Juan de Dios Luna del Castillo.

Director de Tesis:

Prof. Dr. D.

Antonio Martín Andrés

Vº Bº

Ponente

Prof. Dr. D. Ramón Gutierrez Jaimez

Vº Bº

UNIVERSIDAD DE GRANADA. FACULTAD DE CIENCIAS

A Encarna,

tambien a Juan y a Pablito.

AGRADECIMIENTOS

El trabajo que aqui se presenta ha sido apoyado por muchas personas; las mas relevantes, a las que quiero expresar mi mas profundo agradecimiento son:

A mi amigo Antonio Martín Andrés, porque sin su trabajo no hubiera sido posible esta memoria.

A Ramón Gutierrez Jaimez y a mis compañeros del Departamento de Estadística e Investigación Operativa por su agradable acogida.

A Fatima, Mati, Maria, Trini y Victor, operadores del Servicio de Documentación del Hospital Clínico, por sus enormes facilidades para correr los programas de esta tesis; tambien gracias por su paciencia con la ejecución de tales programas.

A Puri que ha mecanografiado el texto, por su paciencia y rapidez.

A todos los que no figuran aqui, pero que se interesaron por este trabajo, tambien gracias.

INDICE

CAPITULO II

INTRODUCCION

En esta memoria nos ocuparemos de algunas situaciones que aparecen bajo el epigrafe general de tablas 2 x 2. De manera muy simple, una tabla 2 x 2 es una disposición de observaciones como la de la tabla 1.1.

Tabla 1.1

	B	$\bar{B}$	
A	x_1	y_1	n_1
$\bar{A}$	x_2	y_2	n_2
	a_1	a_2	n

Tal disposición aparece bajo tres situaciones distintas, que expondremos aqui, y que responden a tres esquemas de muestreo distintos; enumeremoslos a continuacionde manera detallada atendiendo fundamentalmente a tres cuestiones:

a) Situaciones reales en las que se presentan. b) Hipótesis que se contrasten con ellos y c) Inconvenientes para la determinación de la distribución de las variables implicadas en el problema.

Muestreo 1.- Si de una urna en la que hay n bolas - de las que n_1 son de tipo A y n_2 de tipo $\bar{A}$ ($n_1 + n_2 = n$) - se extraen sin reemplazamiento a_1 bolas, entonces los valores (x_1, x_2, y_1 , y_2) son las bolas de cada tipo, A o $\bar{A}$, que están fuera (B) ó dentro ($\bar{B}$) de la urna.

Como los marginales n_1, n_2, a_1 y a_2 están dados de antemano, la única variable independiente del problema es, por ejemplo, x_1.

Las hipótesis que se desea contrastar en este caso es $H_o^1 \equiv$ "Las a_1 bolas han sido extraidas al azar".

Este esquema de muestreo se corresponde con el celebre problema de Fisher (1942) de "la señorita y las tazas del te"; de una manera resumida tal problema consiste en determinar si una señorita dice la verdad cuando afirma que ella reconoce si en una taza de te con leche se ha puesto el te antes de la leche o viceversa; para ello se le presentan una serie de tazas y se le dice que en la mitad se puso antes la leche que en el te y viceversa. No obstante tal problema no es muy común en la práctica.

Sí aparece con mas asiduidad bajo la forma del test de la mediana de Mood y el test del rango intercuartilico (vease Bradley (1968)). Tal esquema de muestreo lleva a una determinación sencilla de la probabilidad de una tabla como la obtenida. En efecto, si la hipotesis nula es cierta, ("las a_1 bolas han sido extraidas al azar"), entonces la variable x_1 sigue una distribución hipergeometrica y la probabilidad de una configuración como la dada será:

$$P(x_1,x_2,y_1,y_2/n_1,n_2,a_1,a_2) = P(x_1/n_i,a_i) = \frac{\binom{n_1}{x_2}\binom{n_2}{x_2}}{\binom{n}{a_1}}$$

observese que en este caso particular la probabilidad de una tabla como la obtenida se puede calcular, pues todo lo que aparece en su formula es conocido.

Muestreo 2.- Si de dos poblaciones distintas, A y $\bar{A}$ se extraen dos muestras aleatorias e independientes de tamaños n_1 y n_2, respectivamente, y se anotan los números x_1 y x_2 de individuos de cada muestra que verifican una cierta propiedad dicotómica B, entonces los valores (x_1,x_2, y_1,y_2) son el número de individuos de cada muestra, A y $\bar{A}$, que verifican (B) ó no ($\bar{B}$) la caracteristica. Como los marginales n_1 y n_2 están dados de antemano, las únicas variables independientes del problema son, por ejemplo, la pareja (x_1,x_2).

La hipótesis nula que se desea contrastar en este caso es:

$H_o^2 \equiv$ "Las proporciones de individuos que presentan la caracteristica B en las dos poblaciones (p_1 y p_2) son iguales".

Este tipo de muestreo es muy común en la práctica y responde al problema general de si una determinada caracteristica se presenta por igual en dos poblaciones de las que se han extraido muestras independientes. En el caso del presente esquema la determinación de la propiedad de una tabla como la obtenida presenta un problema grave. Es claro que la probabilidad deseada no es mas que una convolución de los binomiales independientes, y que, si la hipotesis nula es cierta $(p_1=p_2=p)$, tal convolución toma la siguiente expresión:

$$P(x_1,x_2,y_1,y_2/n_1,n_2) = P(x_1,x_2/n_1,n_2) = \binom{n_1}{x_1}\binom{n_2}{x_2} p^{a_1}(1-p)^{n-a_1}$$

Aparece ahora de manera patente la dificultad a la que aludiamos antes: en la expresión anterior aparece el parametro "perturbador" (nuisance) p que es desconocido. Historicamente se han dado diferentes soluciones a tal problema, pero ello es algo que se verá con posterioridad.

Muestreo 3 .- Si de una población se extraen al azar n individuos, y en cada uno de ellos se anota si verifica ó no las características A y B, entonces los valores (x_1,x_2,y_1,y_2) son el número de individuos de la muestra que verifican ambas caracteristicas, la segunda pero no la primera, la primera pero no la segunda, y, finalmente, ninguna de las dos. Como ahora ningún marginal está dado de antemano, las variables independientes del problema son, por ejemplo, (x_1,x_2,y_1).

En esta ocasión la hipotesis nula es:

$H_o^3 \equiv$ "Las caracteristicas A y B son independientes".

La presente situación es muy frecuente en el campo aplicado bajo la pregunta de si dos características están relacionadas o no; entendiendo por relacionadas el que la probabilidad de la presentación de una modalidad de una de ellas influya en la probabilidad de presentación de una modalidad de la otra. Ahora, la distribución del vector (x_1, x_2,y_1,y_2) es una multinomial que, bajo la hipotesis nula, tendrá la expresión:

$$P(x_1,x_2,y_1,y_2/n) = P(x_1,x_2,y_1/n) = \frac{n!}{x_1!x_2!y_1!y_2!} \; p_a^{n_1}(1-p_a)^{n_2} \; p_b^{a_1}(1-p_b)^{a_2}$$

siendo p_a y p_b las probabilidades de presentar las caracteristicas A y B respectivamente. La situación aquí es análoga a la del muestreo 2 ya que ahora tampoco es posible el calculo directo de la probabilidad de una tabla como la presentada; aparecen aquí dos parámetros "perturbadores" (p_a y p_b) que impiden la solución directa de tal problema. A este tipo de muestreo también se ha dado una solución historica, que mas tarde comentaremos, aunque se ha dedicado a él bastante menos literatura.

Esta memoria está centrada fundamentalmente en el muestreo de tipo 2.

En efecto, el muestreo de tipo 1 tiene ya un test apropiado (el test exacto de Fisher) que solventa el problema de forma concluyente. No ocurre asi en el caso del muestreo de tipo 2, ya que las soluciones propuestas son en muchas ocasiones conservadores, y en cualquier caso no son óptimas para todos los casos. El problema con el muestreo de tipo 3 será, usando la "filosofia" aquí desarrollada, fruto de futuros estudios.

Quedando patenta ya que en el resto de la memoria nos ocuparemos del "test de homogeneidad de dos proporciones con muestras independientes" (y de una generalización de un caso particular de él), describamos suscintamente los abordajes que se han dado a tal problema. Recordemos que la situación era la aparición de un parametro perturbador en la verosimilitud de la tabla obtenida y la necesidad de estimarlo de manera que el test resultante sea lo "mejor" posible.

La eliminación de parametros perturbadores en un contraste de hipotesis ha generado varias soluciones diferentes (Basu, 1977) que se describen a continuación. La primera solución es la que podriamos llamar del metodo condicional y aparece bajo distintas situaciones (test de permutaciones, principio de condicionalidad, tests insesgados, etc...); este tipo de test se basa en la distribución condicional de algún estadístico observado tal que ella no dependa del parametro perturbador. Como veremos mas adelante, en nuestro caso el test condicionado resulta ser el test exacto de Fisher. La segunda alternativa es la impuesta por lo que llamaremos el método de estimación. Tal método, que es incondicionado, consiste en sustituir el parametro perturbador por un estimador. Evidentemente tales test están fundamentados en la teoria asintotica (es decir solo son aplicables para el caso de muestras suficientemente grandes) y son por tanto tests aproximados. En nuestro caso el más típico es el test de la χ^2, que tendremos la oportunidad de estudiar mas adelante. Por último existe otra manera de eliminar el parametro perturbador que es lo que podriamos llamar el método de maximización; consiste este método en sustituir el parámetro perturbador por el valor de él que hace máxima la función de potencia nula del test. En esta situación se obtiene un test incondicional que, aun siendo conservador, es el menos conservador de entre los que se pueden encontrar. El caso mas tipico en nuestro problema de homogeneidad de proporciones es el test propuesto por Barnard (1945 y 1947).

El estudio de los abordajes que acabamos de describir son el objetivo fundamental de esta memoria, que aparecerá dividida en tres capitulos claramente diferenciados. El primer capitulo, dedicado al problema del test de homogeneidad de proporciones con muestras independientes en tablas 2 x 2 para el caso de muestras pequeñas, cubrirá tanto el estudio del test de Fisher, como el de los diferentes tests no condicionados que existen. Un segundo capitulo en el que se estudiará el caso de las grandes muestras, es decir el caso de la sustitución del parametro perturbador por un estimador del mismo.

En el tercer capitulo se hará especial mención a un problema interesante que es el de la bioequivalencia en tablas 2 x 2; usando la teoria asintótica, se realizará una extensión de tal test al caso k x k , que como veremos alli es aplicable en otra serie de campos distintos al de la bioequivalencia.

CAPITULO I

TEST DE SIGNIFICACION EN TABLAS 2 x 2: PEQUEÑAS MUESTRAS

Con frecuencia, los investigadores que han de comparar dos proporciones independientes solo cuentan con muestras de pequeño tamaño que les impide confiar en los tests asintoticos del próximo capitulo. Ante tal situación, el investigador tiene a su disposición dos técnicas alternativas (test condicionados y test incondicionados) que son objeto del capítulo actual.

Todos los test condicionados están basados en la filosofia expuesta originalmente por Fisher (1935) y que, en correspondencia con su nombre, se basan en la distribución de x_1, condicionada por los marginales observados.

Todos los test incondicionados están basados en la filosofia expuesta originalmente por Barnard (1947) y que, en correspondencia con su nombre, se basan en la distribución conjunta de (x_1,x_2) sin condicionar en las a_i, obteniendose el tamaño del test mediante maximización de la region crítica en el valor p comun de las dos proporciones.

Los test condicionados admiten dos versiones (aleatorizados ó no aleatorizados) según que la frontera de la región crítica (RC) se haya construido ó no mediante algún mecanismo de sorteo (Tocher (1950)). Los test incondicionados son siempre no aleatorizados, aunque podria hacerse asi.

Finalmente, y para ciertas ocasiones en que un grupo de n individuos es repartido al azar en dos grupos que reciben cada uno un tratamiento, se tienen los test de aleatorizacion (que a su vez podrian ser condicionados o incondicionados).

En el capítulo actual se analizan en detalle todos los casos expuestos.

1.1. TESTS CONDICIONADOS NO ALEATORIZADOS

1.1.1. Test "exacto" de Fisher

1.1.1.1. Presentación de test

Recordemos la situación del problema. Se disponia de una tabla como la que figura a continuación con el número 1.1, y en ella los marginales n_1

	B	$\bar{B}$	
A	x_1	y_1	n_1
$\bar{A}$	x_2	y_2	n_2
	a_1	a_2	n

Tabla 1.1.

y n_2 eran fijos, es decir, se habian extraido dos muestras aleatorias e independientes de las poblaciones A y $\bar{A}$ respectivamente. La hipotesis nula en la que estabamos interesados era $H_o \equiv p_1 = p_2 (=p)$, siendo p_1 la proporción de individuos que presentan la caracteristica B en la población A y p_2 la proporción que la presentan en la población $\bar{A}$.

El problema es que la probabilidad de una tabla como la obtenida, supuesta cierta H_o, es:

$$P(x_1,x_2,y_1,y_2/n_1,n_2) = P(x_1,x_2/n_1,n_2) = \binom{n_1}{x_1}\binom{n_2}{x_2}p^{a_1}(1-p)^{n-a_1} \quad (1.1)$$

y tal probabilidad depende del parametro perturbador p. Fisher (1935) -y, casi simultáneamente, Irwing (1935) y Yates (1934)- para resolver el problema, propone la siguiente salida: para eliminar el parametro perturbador se debe comparar la muestra observada no con toda la población de muestras posibles, sino con una subpoblación de muestras relacionadas con la obtenida y tales que la distribución de un estadístico en ellas no dependa del parametro p. Esto se consigue en este caso sin mas que considerar los marginales n_1,n_2 y a_1,a_2 fijados de antemano; en ese caso la única variable que aparece en la tabla es la x_1 y estamos en la situación del tipo de Muestreo 1 del que hablabamos en la introducción; de ahi que ahora tengamos que

$$P(x_1/n_1,n_2,a_1) = \frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{n}{a_1}} \qquad r = \max(0,a_1-n_1)\leq x_1 \leq \min(a_1,n_1) \quad (1.2)$$

Es obvio que en la resolución del problema se juega pivotando sobre el estadístico a_1 que es lo que se llama un estadistico "auxiliar". Solo el condicionar sobre el a_1 es lo que permite la eliminación de p. Dado el importante papel que juega aqui el a_1 estudiemos a continuación mas detenidamente el concepto de auxiliariedad y su aplicación al caso que nos ocupa.

1.1.1.2. El concepto de auxiliariedad. Su aplicación al caso de homogeneidad de dos proporciones independientes.

Presentaremos en principio el concepto de auxiliariedad de manera rigurosa y despues veremos como disponer la verosimilitud que nos ocupa de forma que pueda contribuir el estadistico auxiliar al problema.

El concepto de auxiliariedad, que da lugar a los marginales fijos en las tablas 2 x 2, fue expuesto por Fisher (1956). Ha sido estudiado después por múltiples autores y puede formularse así:

Definición 1.1. Considerese el problema general de contraste de hipótesis $H_0 \equiv \theta_1 = \theta_1^0$, θ_2 (sin especificar) contra $H_1 \equiv \theta_1 \neq \theta_1^0$, θ_2 (sin especificar).

Tanto θ_1 como θ_2 pueden ser vectores o escalares. Se tiene un vector de observaciones X cuya verosimilitud se puede factorizar de esta manera:

$$L(\theta_1, \theta_2;X) = \exp\left[t_1\theta_1 + t_2\theta_2\right] j(\theta_1,\theta_2)\, K(X) \qquad (1.3)$$

donde K es solo función de X, j sólo de θ_1 y θ_2 y t_1 y t_2 son funciones de de X solamente. Bajo estas condiciones t_1 y t_2 son un conjunto minimal suficiente de estimadores de θ_1 y θ_2 respectivamente y, cuando uno hace inferencias sobre θ_1, t_2 es el llamado estadístico auxiliar.

Con esta definición se puede demostrar que la distribución de T_1 (uso la letra mayuscula para designar a la variable aleatoria) dado t_2 no es función de θ_2 (vease Lehmann (1959) pag. 52). De ahí que las inferencias acerca de θ_1 puedan hacerse a partir de la distribución de T_1 condicionado por t_2.

Definido ya el concepto de auxiliariedad, y vista la utilidad que presenta, estudiaremos como se puede aplicar en nuestro caso.

Como se sabe la verosimilitud de una tabla como la que tenemos se escribe

$$L(p_1,p_2;x_1,x_2) = \binom{n_1}{x_1}\binom{n_2}{x_2} p_1^{x_1}q_1^{n_1-x_1}\, p_2^{x_2}q_2^{n_2-x_2} \tag{1.4}$$

donde $q_i=1-p_i$, $i=1,2$. Tal verosimilitud puede ponerse de la siguiente manera

$$L(p_1,p_2;x_1,x_2)=\binom{n_1}{x_1}\binom{n_2}{x_2}q_1^{n_1}q_2^{n_2}\exp\left[x_2\left(\ln\frac{p_2}{q_2}-\ln\frac{p_1}{q_1}\right)+(x_1+x_2)\ln\frac{p_1}{q_1}\right] \tag{1.5}$$

Esta expresión está puesta simplemente para que se pueda apreciar que la verosimilitud que nos ocupa es factorizable según el formato de la expresión (1.3) y de ahí que la teoria allí expuesta sea aplicable. No obstante, para que sea más comodo el trabajo con la verosimilitud y sea más facil la obtención de las expresiones que nos hagan ver el estadistico auxiliar, usaremos una reparametrización de p_1 y p_2 a la forma logistica. LLamando

$$p_1=\frac{e^{\beta+\lambda/2}}{1+e^{\beta+\lambda/2}}, \qquad p_2=\frac{e^{\beta-\lambda/2}}{1+e^{\beta-\lambda/2}} \tag{1.6}$$

observerse que $e^{\lambda} = \dfrac{p_1q_2}{p_2q_1}$ o que $\lambda = \ln(p_1/q_1) - \ln(p_2/q_2)$ y que si $\lambda = 0$ entonces $p_1 = p_2$. De otro lado, se deduce facilmente que $\beta = \frac{1}{2}(\ln(p_1/q_1) + \ln(p_2/q_2))$. Por ello, λ esta relacionada con la diferencia de proporciones y β esta relacionada con la magnitud de las mismas.

Usando la reparametrización expuesta en (1.6), la expresión (1.4) queda

$$L(\lambda,\beta;x_1,x_2) = \frac{\binom{n_1}{x_1}\binom{n_2}{x_2}\, e^{(x_1-x_2)\lambda/2 + (x_1+x_2)\beta}}{(1+e^{\beta+\lambda/2})^{n_1}\,(1+e^{\beta-\lambda/2})^{n_2}} \tag{1.7}$$

De tal expresión es inmediato que el conjunto de estadísticos suficientes para λ y β son x_1-x_2 y x_1+x_2 respectivamente. Cuando se hagan inferencias sobre λ (que son las que interesan) entonces $x_1+x_2=a_1$ es el estadístico auxiliar para el parametro perturbador ,sin mas que aplicar lo dicho en la definición de auxiliariedad y parrafo posterior.

De la expresión (1.7) se obtiene, a partir de la distribución condicional de X_1-X_2 fijado a_1, que la distribución condicional para X_1 es

$$g(x_1/a_1;\lambda) = \frac{\binom{n_1}{x_1}\binom{n_2}{a_2-x_1} e^{\lambda x_1}}{\sum_{i=0}^{a_1}\binom{n_1}{i}\binom{n_2}{a_1-i} e^{\lambda i}} \tag{1.8}$$

y para $\lambda = 0$ que es el caso que interesa

$$g(x_1/a_1;\lambda=0) = \frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{n_1+n_2}{a_1}} \tag{1.9}$$

que es justamente la expresión en que está basado el test exacto de Fisher. Queda por tanto explicado el uso del test exacto de Fisher en el contraste de hipótesis de igualdad de dos proporciones independientes, uso que está apoyado por la propiedad de auxiliariedad del estadístico a_1. Además, como veremos más adelante, este test tiene una serie de propiedades muy interesantes y que discutiremos. De forma analoga a como se ha razonado aqui, se puede demostrar que el test exacto de Fisher es aplicable al tipo de Muestreo 3.

1.1.1.3. Revisión de los criterios clásicos de construcción de un test de dos colas

Planteemos primero el caso del test de una cola que nos ayudará a entender mejor el test de dos colas. En efecto, supongase que la alternativa es $H_1 \equiv p_1 < p_2$ y que deseamos realizar el test de Fisher a un error α. Todo consistirá en determinar un límite C tal que

$$\sum_{i=r}^{C} P(i) \leq \alpha \quad ,, \quad \sum_{i=r}^{C+1} P(i) > \alpha \tag{1.10}$$

con P(i) según la hipergeométrica (1.2). Si el x_1 observado es menor ó igual que C, la hipótesis nula habrá de rechazarse al error α. Similarmente para una alternativa como la $H_1' \equiv p_1 > p_2$.

Dado lo laborioso de los cálculos, se han dado numerosas tablas que los evitan (ver por ejemplo Pearson and Hartley (1966) y Finney et al (1963)). Todas ellas adolecen del defecto de su extensión, motivada fundamentalmente por la aparición de tablas repetidas. Tales repeticiones se deben a que se las presenta como tablas para comparar dos proporciones (cuando también valen para comprobar la independencia), ocasionando así una distinción ficticia entre filas y columnas; por otro lado, las tablas repiten regiones críticas para distintos valores de α, lo que ocasiona duplicaciones innecesarias; no sucedería así si, como es más razonable, se diera la probabilidad exacta de una tabla como la observada ó más extrema, es decir el valor de P(r) + P(r+1) + ... + P(x_1). Tales defectos fueron solventados por Neave (1982) que hizo una presentación más resumida de las tablas (pero de manejo mas complejo).

En cualquier caso, en los test de una cola no existe mayor problema. El problema fundamental surge en el caso de las dos colas pues en ella la determinación de cada una de las colas que formarian la región crítica está sujeta a debate por la multitud de criterios existentes, que enumeraremos, no habiendose dado uno que sea el óptimo para el problema que nos ocupa. Los criterios clásicos para la construcción del test de dos colas son:

<u>1^er criterio: Colas iguales</u>

Tal planteamiento, que defienden bastantes autores, esta basado en la idea, propia de las variables continuas,de que el error del test se va a repartir por igual en las dos colas de la RC. Bajo tal idea el experimentador se ve obligado a obtener dos numeros C y C' tales que:

$$\sum_{i=r}^{C} P(i) = \alpha_1 \leq \frac{\alpha}{2} \quad ,, \quad \sum_{i=r}^{C+1} P(i) > \frac{\alpha}{2} \tag{1.11}$$

$$\sum_{i=C'}^{s} P(i) = \alpha_2 \leq \frac{\alpha}{2} \quad ,, \quad \sum_{i=C'-1}^{s} P(i) > \frac{\alpha}{2} \tag{1.12}$$

con P(i) como en la (1.2) y α el error del test de dos colas. Obtenidos los valores C y C' el test es inmediato: rechazar H_o, en favor de $H_1 \equiv p_1 \neq p_2$, si $x_1 \leq C$ ó $x_1 \geq C'$. Sin embargo el error real de tipo I del test es $\alpha^* = (\alpha_1 + \alpha_2)$ menor o igual que el error objetivo α.

Armitage(1971,pgs.135-138) va mas alla y defiende que el error del test de dos colas es el doble del error del test de una cola. El criterio nos parece inadecuado por cuanto puede llegar a suceder que se asigne error a una cola que no existe.

2º Criterio: Tablas mas extremas

Otro criterio que suele aconsejarse (Armsen (1955) ó Cox y Hinkley (1974)) es el siguiente: Sea una Tabla T cualquiera y calculemos la probabilidad $\alpha_1(T)$ de obtener una tabla como T ó una más extrema en el mismo sentido; si $\alpha_2(T) < \alpha_1(T)$ es la probabilidad obtenida sumando una a una las probabilidades de tablas más extremas que la T pero en el otro sentido (las que se puedan hasta no superar α_1), entonces la tabla T pertenece a la región crítica del 100α% si $\alpha_1(T) + \alpha_2(T) \leq \alpha$. Provisionalmente podemos llamar a este criterio como el de "tablas más extremas".

3er Criterio: Tablas más improbables

Tal definición propuesta por Armsen (1955) y recogida también por Fleiss (1981) y que usan los paquetes BMDP y SPSS, consiste en ordenar todas las tablas posibles de mayor a menor probabilidad e ir añadiendo tablas a la región crítica hasta que la probabilidad acumulada sea tan próxima a α como se pueda (pero sin superarlo).

Estos criterios que acabamos de exponer son criticados en esta memoria con objeto de buscar una solución óptima a nuestro problema. Tal es el objeto de los subapartados siguientes.

1.1.1.4. Inconsistencia de los criterios clásicos de construcción del tes de dos colas (Aportación)

Estudiemos primero el criterio de "colas iguales". Recordemos que alli el error real del test era $\alpha^* = \alpha_1 + \alpha_2 \leq \alpha$.

El acercamiento a α no ha sido posible porque $\alpha/2 - \alpha_1 < P(C+1)$ y $\alpha/2 - \alpha_2 < P(C'-1)$, pero, en ocasiones, puede ocurrir que $\alpha^* - \alpha >$ $P(C+1)$ ó $P(C'-1)$, en cuyo caso la región crítica para un test de dos colas podría aumentar en uno de los dos puntos $(C+1)$ ó $(C'-1)$ con el consiguiente acercamiento de α^* a α y el consiguiente aumento de la potencia. Así pues tal criterio no es el idóneo.

Es claro que la definición de la región crítica de "tablas mas extremas" mejora la situación definida antes de "colas iguales", por cuanto todo punto de la región crítica de esta lo es de aquella, pero, sin embargo, tal definición no tiene aún las características de óptima.

La nueva definición de "tablas más improbables" suele dar regiones críticas con igual número de puntos que la de "tablas más extremas" (aunque no necesariamente con los mismos puntos), pero en ocasiones la primera contiene más puntos que la segunda, y, por tanto, da lugar a un test más potente. Esto se comprueba prácticamente con el ejemplo 1.1.

Ejemplo 1.1.

x_1	.	31
.	.	12
31	12	43

x_1	19	20	21 22 23 24	25	26	27	28	29	40	31
$p(x_1)$ en %	0,920	6,524	....	4,435	0,377	0,102	0,006	0	0	0

Tipo de test	Región crítica para el test de dos colas al error del 1%
"Tablas más extremas"	27, 28, 29, 30, 31
"Tablas más improbables"	26, 27, 28, 29, 30, 31

Sin embargo nadie ha sometido a comprobación cual pueda ser la mejor definición de región crítica de entre las muchas posibles.

En lo que sigue se dan varias definiciones de regiones críticas y se elige la definición óptima para comparar dos proporciones.

1.1.1.5. Elección de la región crítica optima para comparar dos proporciones mediante un test de dos colas (Aportación).

Las definiciones de regiones críticas a considerar son las siguientes:

Definición 1: Para un test al error objetivo α, construir la región crítica de colas simétricas según las (1.11) y (1.12) y añadir, si el error α lo permite, una de las dos fronteras (C+1) ó (C'-1), es decir añadir aquella frontera, si la hay, cuya probabilidad es $\leq \alpha-(\alpha_1+\alpha_2)$. Cuando las dos fronteras verifiquen tal condición entonces:

a) Incluir la frontera que más probabilidad tenga (es decir la que acerca más el error real α^* al error objetivo α).

b) Incluir la frontera que menos probabilidad tenga(es decir la más improbable bajo H_o).

c) Incluir la frontera que más iguale los errores de cola (es decir la que haga los α_i finales tales que $|\alpha_1-\alpha_2|$ sea mínima, ó, lo que es mismo, la que más simetrice las colas).

Si con cualquiera de los tres criterios se produce un empate, decidir por sorteo si se incluye (C+1) ó (C'-1).

Definición 2: Para un test al error objetivo α, ordenar los valores de x_1 de menor a mayor probabilidad e ir añadiendo a la región crítica los puntos más improbables hasta acercarse a α lo más posible (sin superarlo). En caso de empate, decidir por sorteo el punto a incluir.

Las Regiones Críticas anteriores se pueden llamar, de modo abreviado, como de "colas simétricas" y "de tablas mas improbables",respectivamente, y la lógica de las mismas es evidente. Pero, si se están comparando proporciones, parece tambien lógico que si una tabla con $|\hat{p}_1-\hat{p}_2|$ =10%, por ejemplo, sirve para rechazar H_o, una tabla con $|\hat{p}_1-\hat{p}_2|$ =15% tambien debería hacerlo. Esta es la base de la siguiente definición "de tablas más extremas":

Definición 3: Para un test al error objetivo α, ordenar los valores de x_1 de mayor a menor valor de $|\hat{p}_1-\hat{p}_2| = |x_1/n_1-x_2/n_2|$ e ir añadiendo puntos más extremos a la región crítica hasta que la suma de las probabilidades de ellos se acerque lo más posible a α (sin superarla). En caso de empate:

a) Añadir el punto de menor probabilidad

b) añadir el punto de mayor probabilidad

De persistir el empate, elegirlo al azar.

Como el criterio de valores crecientes de $|\hat{p}_1-\hat{p}_2|$ es equivalente al de valores crecientes de $|x_1-n_1a_1/n|$, y como las probabilidades de la hipergeométrica son monótonas crecientes en un tramo y decrecientes en el otro, queda asegurada la no existencia de huecos en la región crítica y la inclusión de las tablas más improbables como criterio subsidiario al de las más extremas.

Desde luego el criterio anterior "de tablas más extremas" se ha construido pensando en la medida de discrepancia $|\hat{p}_1-\hat{p}_2|$ ó valor absoluto de "la diferencia de Berkson". Tambien pueden plantearse criterios de tablas más extremas en otro sentidos:

Definición 4: Como en la definición 3, pero utilizando como medida de tabla extrema el "riesgo relativo" $\hat{p}_1/\hat{p}_2$.

Definición 5: Como en la definición 3, pero utilizando como medida de tabla extrema la "razón del producto cruzado" x_1y_2/x_2y_1.

Definición 6: Como en la definición 3, pero utilizando como medida de tabla extrema el "valor de χ^2_{exp}", es decir a $(x_1y_2-x_2y_2)^2n/n_1n_2a_1a_2$.

Las definiciones 3 y 6 son equivalentes (puesto que $|x_1y_2-x_2y_1|=|x_1n-n_1a_1|$), por lo que la segunda no será considerada mas. De hecho esta fue llamada "test χ^2 exacto" por Radlow y Alf (1975) en un contexto más general, probando que tenía características deseables frente a la definición 2. Vease Berry, K.J. (1985).

1.1.1.6. Comparación de los distintos criterios (Aportación)

La potencia de cualquier test de los señalados depende del número de puntos de la región crítica. Así, si suponemos que la distribución de p_i a lo largo de todas las experiencias de la vida de un experimentador es la uniforme en $[0.1]$, y si $\bigcup_i T_i$ son todas las tablas de una determinada región crítica para unos n_1 y n_2 dados, entonces la potencia será:

$$\beta = P(\text{decidir } H_o | \text{es cierta } H_1) = P(\bigcup_i T_i | H_1) = \sum_i P(T_i | H_1) =$$

$$\sum_i \int_0^1\int_0^1 P(p_1)P(p_2)\ P(T_i | p_1, p_2)\ dp_1\ dp_2 =$$

$$\binom{n_1}{x_1}\binom{n_2}{x_2} \int_0^1 p_1^{x_1} q_1^{n_1-x_1}\ dp_1 \int_0^1 p_2^{x_2} q^{n_2-x_2}\ dp_2 =$$

$$\sum_i \frac{1}{(n_1+1)\ (n_2+1)} = \frac{\text{nº de tablas de la región crítica}}{\text{nº total de tablas posibles}} \tag{1.13}$$

Es por ello que la definición clásica de "colas simétricas" se completó en la Definición 1 para que entrara el mayor número posible de puntos. Sin embargo, las tres versiones de la Definición 1 constan de igual número de puntos (aunque, circunstancialmente, no sean los mismos), por lo que habrá de considerarse algun criterio añadido que permita elegir alguna de las tres versiones como la óptima. De lo dicho queda claro que si se calcula la potencia de cada test para parejas dadas de (p_1, p_2), es decir:

$$\beta(p_1, p_2 | n_1, n_2) = \sum_{T_i} \binom{n_1}{x_2}\binom{n_2}{x_2} p_1^{x_1} q_1^{n_1-x_1} p_2^{x_2} q_2^{n_2-x_2} \tag{1.14}$$

en unas ocasiones será más potente un cierto método y, en otras, otro. Como quiera que estamos comparando proporciones, y que la mayoría de las veces p_1 no estará muy lejana de p_2 (que es justamente cuando la potencia es menor), añadamos, como criterio de test óptimo, la condición de elegir aquel que sea más potente en las cercanías de la hipótesis nula. Con tal fin, fijémonos en todas las parejas de tamaños muestrales $n_1 > n_2$ con $n_1 = 10(5)30$ (no considerando los casos $n_1=n_2$ pues en ellos, por simetria de la hipergeométrica, las tres regiones críticas de la definición 1 son iguales) y en todas las parejas de valores de $p_1 \neq p_2$, con $p_1 = 0,05(0,05)0,95$ (no considerando ahora el caso $p_1 = p_2$ pues en él sería cierta la hipótesis nula). Consideremos ahora la función:

$$\sum_{o < |p_1-p_2| \leq d} \beta(p_1,p_2|n_1,n_2) = \beta(d|n_1,n_2) \tag{1.15}$$

ó potencia de todas las alternativas que distan (en el sentido de $|p_1-p_2|$ menos ó igual que d (consideradas a saltos de 5% de probabilidad) con d = 0,10(0,20)0,70. Como las regiones críticas están construidas a un determinado error objetivo α , la función $\beta(d|n_1,n_2)$ dependerá tambien del α tomado; de ahí que podamos notarla por $\beta(d|n_1,n_2,\alpha)$ con α = 0.05,0,025 y 0,01. La tabulación de tales valores de β para cada uno de los tres criterios de la Definición 1, y la comparación de las mismas, permitirá determinar el criterio óptimo de región crítica, óptimo en el sentido de máxima potencia en las cercanias de H_o. La tabla 1.2 presenta tales valores, deduciéndose de ella que la definición óptima, en el sentido citado, es la 1.b ó 1.c (ambas son iguales).

Como aún quedan varias tablas del mismo tipo, y para no hacerlas demasiado extensas, en adelante únicamente se dan los valores

$$\beta(d|\alpha) = \sum_{n_1}\sum_{n_2} \beta(d|n_1\ n_2,\alpha) \tag{1.16}$$

si bien las conclusiones que aquí se den, han sido obtenidas con tablas similares a la Tabla 1.2. Así, la Tabla 1.3(a) constituye un buen resumen de la Tabla 1.2.

La Tabla 1.3(b) presenta los resultados de la comparación de las

TABLA 1.2

Valores de $\beta(d|n_1,n_2,\alpha)$ para las tres versiones de la Definicion 1 y para los valores reseñados de α, n_1,n_2,d.

α= .050		d=0.10			d=0.30			d=0.50			d=0.70		
n1	n2	1.a	1.b	1.c	1.a	1.b	1.c	1.a	1.b	1.c	1.a	1.b	1.c
**	**	****	****	****	****	****	****	****	****	****	****	****	****
10	5	.01	.01	.01	.06	.06	.06	.10	.10	.10	.15	.15	.15
15	10	.02	.02	.02	.10	.12	.12	.19	.22	.22	.25	.27	.27
15	5	.01	.02	.02	.06	.07	.07	.12	.13	.13	.17	.19	.19
20	15	.03	.03	.03	.17	.18	.18	.29	.30	.30	.33	.35	.35
20	10	.03	.03	.03	.14	.15	.15	.25	.26	.26	.30	.31	.31
20	5	.02	.02	.02	.08	.09	.09	.14	.16	.16	.20	.21	.21
25	20	.04	.04	.04	.24	.24	.24	.36	.36	.36	.39	.39	.39
25	15	.03	.04	.04	.19	.21	.21	.32	.34	.34	.37	.38	.38
25	10	.02	.03	.03	.14	.14	.14	.24	.25	.25	.30	.30	.30
25	5	.02	.02	.02	.08	.09	.09	.15	.17	.17	.21	.22	.22
30	25	.05	.05	.05	.28	.29	.29	.41	.41	.41	.43	.43	.43
30	20	.04	.05	.05	.26	.28	.28	.40	.41	.41	.42	.43	.43
30	15	.04	.04	.04	.21	.22	.22	.33	.34	.34	.37	.38	.38
30	10	.03	.03	.03	.16	.16	.16	.28	.28	.28	.33	.33	.33
30	5	.02	.02	.02	.08	.09	.09	.15	.17	.17	.20	.22	.22

α= .025		d=0.10			d=0.30			d=0.50			d=0.70		
n1	n2	1.a	1.b	1.c	1.a	1.b	1.c	1.a	1.b	1.c	1.a	1.b	1.c
**	**	****	****	****	****	****	****	****	****	****	****	****	****
10	5	.00	.00	.00	.01	.01	.01	.02	.02	.02	.03	.03	.03
15	10	.01	.01	.01	.04	.04	.04	.10	.10	.10	.16	.16	.16
15	5	.00	.00	.00	.01	.01	.01	.03	.03	.03	.05	.05	.05
20	15	.01	.01	.01	.09	.09	.09	.19	.19	.19	.25	.25	.25
20	10	.01	.01	.01	.04	.05	.05	.11	.11	.11	.15	.16	.16
20	5	.00	.00	.00	.01	.01	.01	.03	.03	.03	.05	.05	.05
25	20	.01	.02	.02	.13	.13	.13	.26	.26	.26	.31	.31	.31
25	15	.01	.01	.01	.07	.08	.08	.15	.16	.16	.19	.19	.19
25	10	.00	.00	.00	.04	.04	.04	.09	.09	.09	.12	.12	.12
25	5	.00	.00	.00	.01	.01	.01	.03	.03	.03	.05	.05	.05
30	25	.02	.02	.02	.16	.16	.16	.29	.29	.29	.33	.33	.33
30	20	.02	.02	.02	.17	.17	.17	.31	.32	.32	.36	.37	.37
30	15	.01	.01	.01	.09	.10	.10	.18	.19	.19	.22	.23	.23
30	10	.00	.01	.01	.04	.05	.05	.10	.11	.11	.15	.15	.15
30	5	.00	.00	.00	.01	.01	.01	.04	.04	.04	.06	.06	.06

α= .010		d=0.10			d=0.30			d=0.50			d=0.70		
n1	n2	1.a	1.b	1.c	1.a	1.b	1.c	1.a	1.b	1.c	1.a	1.b	1.c
**	**	****	****	****	****	****	****	****	****	****	****	****	****
10	5	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.01	.01
15	10	.00	.00	.00	.02	.02	.02	.06	.06	.06	.10	.10	.10
15	5	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.01	.01
20	15	.00	.00	.00	.05	.05	.05	.14	.14	.14	.21	.20	.20
20	10	.00	.00	.00	.02	.02	.02	.06	.07	.07	.12	.12	.12
20	5	.00	.00	.00	.00	.00	.00	.01	.01	.01	.02	.02	.02
25	20	.01	.01	.01	.09	.09	.09	.21	.22	.22	.28	.28	.28
25	15	.01	.01	.01	.06	.07	.07	.15	.17	.16	.21	.22	.22
25	10	.00	.00	.00	.02	.03	.03	.07	.08	.08	.11	.12	.12
25	5	.00	.00	.00	.00	.00	.00	.01	.01	.01	.02	.02	.02
30	25	.01	.01	.01	.12	.12	.12	.26	.26	.26	.32	.32	.32
30	20	.01	.01	.01	.09	.10	.10	.21	.22	.22	.26	.27	.27
30	15	.00	.00	.00	.06	.06	.06	.15	.15	.15	.20	.21	.21
30	10	.00	.00	.00	.02	.02	.02	.05	.06	.06	.09	.10	.10
30	5	.00	.00	.00	.00	.00	.00	.01	.01	.01	.02	.02	.02

TABLA 1.3

Valores de $\beta(d|\alpha)$ para las diferentes definiciones de region critica

(a)

Valores de d	d=0.10			d=0.30			d=0.50			d=0.70		
Definicion	1.a	1.b	1.c	1.a	1.b	1.c	1.a	1.b	1.c	1.a	1.b	1.c
β (d/ .050)	.41	.43	.43	2.23	2.36	2.36	3.73	3.90	3.90	4.43	4.57	4.57
β (d/ .025)	.11	.12	.12	.93	.96	.96	1.93	1.97	1.97	2.48	2.52	2.52
β (d/ .010)	.05	.05	.05	.54	.58	.58	1.39	1.45	1.45	1.98	2.03	2.02

(b)

Valores de d	d=0.10		d=0.30		d=0.50		d=0.70	
Definicion	3.a	3.b	3.a	3.b	3.a	3.b	3.a	3.b
β (d/ .050)	.50	.48	2.44	2.42	3.93	3.92	4.58	4.56
β (d/ .025)	.21	.15	1.08	1.01	2.04	1.99	2.56	2.52
β (d/ .010)	.12	.08	.69	.62	1.52	1.47	2.08	2.03

(c)

Valores de d	d=0.10			d=0.30			d=0.50			d=0.70		
Definicion	1.b	2	3.a	1.b	2	3.a	1.b	2	3.a	1.b	2	3.a
β (d/ .050)	.43	.43	.50	2.36	2.36	2.44	3.90	3.90	3.93	4.57	4.57	4.58
β (d/ .025)	.12	.12	.21	.96	.96	1.08	1.97	1.97	2.04	2.52	2.52	2.56
β (d/ .010)	.05	.05	.12	.58	.58	.69	1.45	1.45	1.52	2.03	2.03	2.08

(d)

Valores de d	d=0.10		d=0.30		d=0.50		d=0.70	
Definicion	4.a	4.b	4.a	4.b	4.a	4.b	4.a	4.b
β (d/ .050)	.36	.33	1.98	1.94	3.38	3.36	4.09	4.07
β (d/ .025)	.08	.07	.69	.65	1.45	1.42	1.92	1.86
β (d/ .010)	.03	.03	.39	.34	.98	.94	1.38	1.36

(e)

Valores de d	d=0.10		d=0.30		d=0.50		d=0.70	
Definicion	5.a	5.b	5.a	5.b	5.a	5.b	5.a	5.b
β (d/ .050)	.42	.38	2.14	2.12	3.54	3.53	4.18	4.17
β (d/ .025)	.14	.11	.90	.87	1.79	1.76	2.27	2.24
β (d/ .010)	.06	.04	.53	.49	1.30	1.25	1.75	1.74

(f)

Valores de d	d=0.10		d=0.30		d=0.50		d=0.70	
Definicion	3.a	5.a	3.a	5.a	3.a	5.a	3.a	5.a
β (d/ .050)	.50	.42	2.44	2.14	3.93	3.54	4.58	4.18
β (d/ .025)	.21	.14	1.08	.90	2.04	1.79	2.56	2.27
β (d/ .010)	.12	.06	.69	.53	1.52	1.30	2.09	1.75

dos versiones de la Definición 3, siendo preferible el criterio 3.a. En la siguiente -Tabla 1.3(c)- se comparan los criterios 1.b, 2 y 3.a, siendo seleccionado, con diferencia, el 3.a. De las dos definiciones 4 y 5 resulta elegida -ver tablas 1.3(d) y 1.3(e)- la Definición 5.a. Finalmente la tabla 1.3(f) compara los dos mejores criterios encontrados -el 5.a y el 3.a- siendo seleccionado el 3.a.

1.1.1.7. Tablas para el test exacto de Fisher según el criterio optimo (Aportación)

Se presentan a continuación las tablas construidas bajo el criterio optimo que acabamos de encontrar, haciendo referencia a las existentes y una pequeña crítica a ellas, razonando el formato optimo de tabla. Tambien se hace referencia a como se podrian emplear las tablas propuestas para el caso de Muestreo 3. Pero pasemos al análisis de las tablas existentes.

Cuando el test es de una cola, hay una única construcción razonable de región crítica, la cual ha sido tabulada extensamente por diversos autores (Pearson and Hartley (1966) y Finney et al (1963) por ejemplo). Tal construcción consiste en sumar las probabilidades de todas las tablas más extremas (en el sentido de H_1) hasta acercarse lo más posible (sin superarlo) al error α fijado.

No deja de ser curioso que casi todas las tablas de significación clásicas para tablas 2 x 2 se hayan construido pensando en un test de una cola, (solo Armsen (1955) y Neave (1982) presentan para 2 colas), cuando es casi infinitamente más frecuente un test de dos colas con una alternativa del tipo $H_1 \equiv p_1 \neq p_2$. El asunto no tendría mayor importancia si no fuera porque el criterio de aplicación para dos colas consiste en multiplar por dos el error nominal, lo que puede afectar sensiblemente a la potencia del test (como ya se ha visto). El error está seguramente inducido por un deseo de minimizar el espacio dedicado a las tablas y por el hecho de que, en grandes muestras, la distribución es prácticamente simétrica. Así pues, y como primera consecuencia, en toda tabla para test de significación en tablas 2 x 2 deben coexistir las tablas de una cola con la de las de dos colas.

Por otro lado, y esta es la segunda consecuencia, puesto que el investigador no tiene por qué estar sometido a los errores clásicos del 1%, 5% y 10%, las tablas deben especificar el error P exacto de significación (el mínimo error α al cual es rechazable la hipótesis nula). Tal proceder solventa, por ejemplo, el problema de aquellos investigadores que desean hacer comparaciones múltiples y que, al aplicar la regla de Bonferroni, se encuentran con un error α más pequeño que los usualmente tabulados.

Además, los criterios de construcción de tales tablas han dependido del gusto del autor más que de una investigación objetiva del método más adecuado (salvo alguna pequeña aportación del articulo citado de Armsen). Aqui se ha probado que, para comparar dos proporciones, el criterio "optimo" (en el sentido de que da regiones críticas con el máximo número posible de puntos y con la máxima potencia en las cercanias de H_o) consiste en ordenar las tablas de mayor a menor valor $|x_1/n_1 - x_2/n_2|$ e ir añadiendo a la región crítica las tablas más extremas hasta acercarse lo más posible (sin sobrepasarlo) al error objetivo α ; si al llegar a la tabla frontera hay dos tablas igual de extremas, añadir a la región crítica la que menor probabilidad tenga; si ambas tienen igual probabilidad, decidir por sorteo cual de ellas debe ser la incluida.

La filosofía general para test de dos colas es fácil de conjeturar: las tablas deben ordenarse de mayor a menor valor del parámetro cuyas discrepancias con su valor bajo H_o quieren detectarse.

Así, para problemas de asociación, habrá de considerarse como óptima -muy probablemente- la definición 5.a, y las "tablas más extremas" serán las que proporcionen un mayor valor de la razón del producto cruzado. De todos modos, puesto que el caso de dos proporciones ocupa una posición intermedia (en cuanto al número de marginales fijos) respecto de los otros dos casos, y puesto que la definición 5.a daba en aquel caso la potencia más cercana a la 3.a, parece razonablemente aceptable aplicar las mismas tablas en todos los casos (lo que evita la triplicación de las mismas a cambio, probablemente, de poco). Tal proceder no presenta problemas de aplicación por cuanto la ordenación de las tablas de mayor a menor valor de $|\hat{p}_1-\hat{p}_2|= \frac{n}{n_1 n_2} | x_1 - n_1 a_1/n|$ es idéntica, por la simetria de la expresión, si se consideran las $\hat{p}_i$ calculadas con respecto a las filas ó a las columnas. De igual modo, como $\hat{p}_1-\hat{p}_2=x_1/n_1 - x_2/n_2 = n(x_1-a_1n_1/n) /n_1n_2$ con la tabla ordenada por filas, y $\hat{p}_1-\hat{p}_2=x_1/a_1-y_1/a_1=n(x_1-a_1n_1/n) /a_1a_2$ con la tabla ordenada por columnas, ambas expresiones son positivas (ó negativas) en las mismas tablas, y el hecho de cambiar filas por columnas no altera la región crítica del test de una cola.

Como en variables discretas -tal es nuestro caso- el error de una cola no suele ser el de dos colas dividido por dos, las tablas de una cola deben coexistir junto a las de dos colas.

Finalmente y, por razones ya indicadas más arriba, las tablas no deben construirse -si se quiere tengan una validez universal- para errores α determinados de antemano (el 5% ó 1% usuales) sino que a cada tabla factible de ser significativa deberá indicársele su error

P mínimo de significación (lo que no hace ningun autor).

Somos conscientes de que tal proceder aumenta en gran medida el espacio requerido para las tablas, pero la búsqueda de la máxima potencia (especialmente con pequeños tamaños de muestra, como es el caso) justifica tal sacrificio. Además, una adecuada tabulación (que impida la repetición de tablas) puede paliar en algo el defecto.

A efectos prácticos, y si no se dispone de unas tablas como las señaladas, el modo de proceder para realizar un test de comparación de dos proporciones (similarmente en otro caso) es el siguiente: si experimentalmente la tabla obtenida es la tabla 1.1.:

1º) Si el test es de una cola, con $H_1 \equiv p_1 < p_2$, calcular primero $\hat{p}_1 = x_1/n_1$ y $\hat{p}_2 = x_2/n_2$. Si $\hat{p}_1 \geq \hat{p}_2$ aceptar H_o. Si $\hat{p}_1 < \hat{p}_2$, calcular $\sum_{i=r}^{x_1} P(i)=P$, con $r=\max(0; a_1-n_2)$ y $P(i)$ como en la (1.2). Similarmente para $H_o \equiv p_1 > p_2$.

2º) Si el test es de dos colas ($H_1 \equiv p_1 \neq p_2$) ordenar la tabla de modo que $\hat{p}_1 = x_1/n_1 < \hat{p}_2 = x_2/n_2$ (Si $\hat{p}_1 = \hat{p}_2$ aceptar H_o sin más) y calcular $P_1 = \sum_{i=r}^{x_1} P(i)$, con $r=\max(0;a_1-n_2)$ y $P(i)$ como en la (1.2). Determinar a continuación $C' = (2n_1a_1/n)-x_1$. Hacer $x'_1 = [C']^+$, el primer entero superior o igual que C', y calcular $P_2 = \sum_{i=x_1'}^{s} P(i)$, con $s = \min(a_1;n_1)$ y $P(i)$ como en la (1.2). Tal x'_1 daria lugar a la primera tabla que es tan extrema ó más, en el otro sentido, que la obtenida, es decir, la primera tabla que ocasiona $\hat{p}'_1 - \hat{p}'_2 \geq \hat{p}_2 - \hat{p}_1$.

Si C' no es un entero, $P=P_1+P_2$.

Si C' es un entero -en cuyo caso x_1 y x'_1 dan dos tablas igual de extremas- determinar $P(x_1)$ y $P(x'_1)$ mediante la (1.2). Si $P(x_1) < P(x'_1)$ entonces $P = P_1 + P_2 - P(x'_1)$. Si $P(x_1) = P(x'_1)$ sortear cual debe ser el elegido; si resulta ser el x_1, entonces de nuevo $P = P_1 + P_2 - P(x'_1)$; si resulta elegido el x'_1, entonces $P = P_1 + P_2$. Finalmente, si $P(x_1) > P(x'_1)$ tambien será $P = P_1 + P_2$.

Si C' es un entero y $P(x_1) = P(x'_1)$, entonces $n_1 = n_2 = n/2$, de modo que si se quiere eludir el sorteo (al igual que ocurría con el test de Tocher) entonces $P = P_1 + P_2 = 2P_1$, por la simetría de la hipergeométrica.

3º) En cualquier caso se rechaza H_o si $P \leq \alpha$, con α el error objetivo del test.

Siguiendo los criterios anteriores se ha elaborado un programa de lenguaje FORTRAN V que aparece en el Apendice 1.

También, y a través del mismo se han elaborado tablas de test hasta n= 30 que aparecen en el Apendice 2.

La filosofia de construcción de las tablas del Apendice 2 es la reseñada mas arriba. El formato de presentación de las mismas pretende ahorrar al máximo el espacio dedicado a ellas a la vez que les hace fácilmente manejables; asi, las cuatro tablas posibles que se obtendrian permutando filas entre si, columnas entre si, ó cambiando filas por columnas, de una determinada tabla (todas ellas equivalentes) aparecen en una sola ocasión.

La notación de las tablas del Apendice 2 es idéntica a la de la tabla 1.1 siendo el criterio de ordenación y busqueda el siguiente:

1º) Presentar la tabla problema de modo que a_1 sea el marginal mas pequeño y x_1 sea tal que $x_1/n_1 \leq x_2/n_2$ (es decir, el valor de menor frecuencia relativa entre los dos valores que suman a_1).

2º) Localizar en el Apéndice 2, y por este orden, la cuaterna (n; a_1; n_1; x_1).

3º) Si la cuaterna aparece, a su derecha figuran los valores P_1 y P_2. El valor P_1 es

$$P_1 = \sum_{i=r}^{x_1} P(i)$$

lo que constituye el error exacto α para un test de una cola cuya alternativa es $H_1 \equiv p_1 < p_2$ en el caso de comparar proporciones, ó $H_1 \equiv$ "asociación negativa" en el caso de tablas de contingencia. El valor P_2 es el error exacto P para el test de dos colas; cuando a tal valor le sigue el simbolo *, ello indica que el test de dos colas se ha realizado sorteando la frontera de la región de rechazo, de modo que si se desea eludir el sorteo habrá de hacerse $P_2 = 2P_1$.

4º) Si la cuaterna no aparece, ello puede deberse a que la tabla no es significativa ó muy significativa. La razón es que, para no hacer demasiado extensas las tablas, solo se han incorporado aquellas en las que $P_1 \leq 10\%$ y P_2 (ó $2P_1$ si hay *) $\geq 0,05\%$, con lo que se logra que aparezca cualquier tabla que sea significativa (para una ó dos colas) para valores de P entre el 0,05% y el 10%. La distinción entre un caso y otro es fácil: buscar en el Apéndice la terna (n; a_1; n_1) y anotar el menor (x_m) y el mayor (x_M) valor de x_1 que aparezca; entonces:

a) Si $x_1 < x_m$, la tabla es significativa para errores de una y dos colas inferiores a 0.05% y a los indicados para x_m.

b) Si $x_1 > x_M$, la tabla no es significativa pues sus errores P de una o dos colas son superiores al 10% y a los indicados para x_M.

Así, en el ejemplo 1.2, será n=22, a_1=6 y x_1=1 (ya que 1/11 < 5/11); con ello n_1=11 y las tablas del Apéndice indican que P_1 = 0,0743, P_2 = 0.0805* (con sorteo) y P_2 = 0.1486 (sin sorteo). Así, si en origen el problema era un test de una cola con una alternativa con la que

Ejemplo 1.2

6	10	16
5	$1=x_1$	$6=a_1$
11	$11=n_1$	22=n

Ejemplo 1.3

$1=x_1$	17	$18=n_1$
7	1	8
$8=a_1$	18	26=n

Ejemplo 1.4

$4=x_1$	14	$18=n_1$
4	4	8
$8=a_1$	18	26=n

Ejemplo 1.5

$0=x_1$	7	$7=n_1$
5	3	8
$5=a_1$	10	15=n

está de acuerdo lo experimental (1/6 < 10/16), el cambio de filas por columnas no altera este hecho (1/11 < 5/11) y el error de una cola es P_1 = 7.43%.

El ejemplo 1.3 no es localizable directamente en las tablas, pero como en ellas aparece la cuaterna n=26, a_1=8, n_1=18 y x_m=2, y x_1=1 < x_m=2, entonces P_1 y P_2 son menores que 0,05%.

El ejemplo 1.4 no es localizable directamente en las tablas, pero como en ellas aparece la cuaterna n=26, a_1=8, n_1=18 y x_M=3, y x_1=4 >

$x_M=3$, entonces P_1 y P_2 son superiores al 10%.

Finalmente, el ejemplo 1.5 presenta una tabla para la que se desea contrastar $H_o \equiv p_1 = p_2$ contra $H_1 \equiv p_1 > p_2$, pero como $\hat{p}_1 = 0/7 < \hat{p}_2 = 5/8$, H_o debe aceptarse sin más a pesar de que las tablas indiquen que $P_1 = 0,0187$ (pues tal error se refiere a la alternativa $H_1 \equiv p_1 < p_2$).

1.1.2. La vision de Pearson sobre el problema del test de homogeneidad de dos proporciones

Pearson (1947) realiza una serie de reflexiones acerca del contraste que nos ocupa. Las recogemos aqui pues entendemos que asi se ve de una manera mas precisa el problema que se plantea con estas tablas y se comprende mejor la solución propuesta por Tocher (1950) y que estudiaremos despues.

Recuerdese que la verosimilitud para nuestro problema era:

$$P(x_1,x_2/n_1,n_2) = P(x_1/n_1,p_1).\ P(x_2/n_2,p_2) \tag{1.18}$$

Si la hipótesis nula es cierta es claro que tal verosimilitud puede ponerse como

$$\begin{aligned} P(x_1,x_2/n_1,n_2) &= P(x_1/n_1,p).\ P(x_2/n_2,p) = \\ &= \frac{n_1!}{x_1!\ y_1!}\ p^{x_1}\ q^{y_1}\ \frac{n_2!}{x_2!\ y_2!}\ p^{x_2}\ q^{y_2} = \\ &= \frac{n!}{a_1!\ a_2!}\ p^{a_1}\ q^{a_2}\ \frac{n_1!\ n_2!\ a_1!\ a_2!}{n!\ x_1!\ x_2!\ y_1!\ y_2!} \end{aligned} \tag{1.19}$$

ahora bien en esa expresión se ve que

$$\frac{n!}{a_1!\ a_2!}\ p^{a_1}\ q^{a_2} = \binom{n}{a_1}\ p^{a_1}\ q^{n-a_1} = P(a_1/n,p) \tag{1.20}$$

que es la función de probabilidad de una Binomial de parametros n y p.

Por otro lado el resto de la expresión (1.19) se puede poner asi:

$$\frac{n_1!\ n_2!\ a_1!\ a_2!}{n!\ x_1!\ x_2!\ y_1!\ y_2!} = \frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{n}{a_1}} = P(x_1/n,n_1,a_1) \qquad (1.21)$$

que no es más que la función de probabilidad de una hipergeometrica de parametros n, n_1, a_1.

Por tanto se puede escribir

$$P(x_1,x_2|n_1,n_2) = P(a_1|n,p)\ .\ P(x_1|n,n_1a_1) \qquad (1.22)$$

Si ahora se plantea el test a un error α, la (1.22) permite resolverlo, en principio, facilmente: para cada a_1, hacer entrar en la region critica (RC) todas las parejas (x_1, a_1-x_1) mas improbables -en terminos de la (1.21)- hasta que se verifique

$$\sum_{x_1} P(x_1|n,n_1,a_1) = \alpha \qquad (1.23)$$

Si ello se hace para todo a_1, el tamaño del test se obtendrá sumando las probabilidades (1.22) en toda la RC:

$$\sum_{RC} P(x_1,x_2|n_1,n_2) = \sum_{a_1=0}^{n} P(a_1|n_1,p).\alpha = \alpha \qquad (1.24)$$

con lo que el test exacto de Fisher podria entenderse como uno incondicional cuyo tamaño es el error objetivo.

Desgraciadamente esto no es asi pues, por ser una variable discreta, la igualdad (1.23) no suele alcanzarse y, consiguientemente, tampoco la (1.24). En la práctica sucede que:

$$\sum_{x_1} P(x_1/n,n_1,a_1) = \alpha(a_1) \leq \alpha$$

$$\sum_{RC} P(x_1,x_2/n_1,n_2) = \sum_{a_1} P(a_1/n_1,p)\,\alpha(a_1) = \alpha^*(p) < \alpha \qquad (1.25)$$

en donde $\alpha - \alpha^*(p)$ depende de p y suele ser importante; de ahí el profundo conservadurismo del test exacto de Fisher aun cuando, por la filosofia anterior, pueda considerarse como un test incondicionado. De cualquier modo las tablas al uso no especifican el error $\alpha^*(p)$, sino el $\alpha(a_1)$, debiendo considerarlas como tests condicionados. Observese que la igualdad (1.23) no tendria dificultad en verificarse si la variable x_1 fuera continua.

1.1.3. Criticas al test "exacto" de Fisher

En los subapartados anteriores las criticas se han centrando exclusivamente sobre los criterios tradicionales de realizar un test de dos colas para comparar dos proporciones y sobre la forma clasica de presentar las tablas de significación, dejando en ambos casos incuestionable la utilidad del test de Fisher. Aqui entramos directamente a cuestionar la propia filosofia del test.

La adecuación del test de Fisher para comparar dos proporciones es uno de los temas de la Estadistica que roza el campo de la Filosofia, al menos en el modo en que defensores y atacantes hacen fortaleza de sus principios. Veanse los articulos de Berkson (1978), Kempthorne (1979), Yates (1984).

Barnard (1945), en su presentación menos formalizada de su criterio incondicionado, ya sostiene la primera discusión con Fisher (1945) acerca de cual test debe elegirse. Con posterioridad Barnard (1947) y Pearson (1947) vuelven a criticar como irreal el principio de condicionamiento que Fisher defiende por estar basado en un estadistico auxiliar. Garside and Mark (1967) apoya a Barnard y modifican algo su criterio. Berkson (1978) coincide con ellos al señalar que no es licito el condicionamiento en las a_1 pues, por ser valores aleatorios dependientes del resultado muestral, su eliminación constituye una perdida innecesaria de información. Cornfield (1966) señala diversas inconsistencias del test exacto de Fisher y, mas recientemente, Dupont (1986) presenta tablas en las que la aplicación del test de Fisher da lugar a graves irregularidades (ver 2.2.3).

El caso de Barnard es ciertamente curioso. Tras haber sido el precursor del metodo incondicionado, y haber discutido con Fisher, rapidamente se arrepiente (1949) y pasa mas adelante (1979, 1982) a una posición intermedia, algo oscura, en la que aboga por la realización del test de Fisher pero para valores de α variables, más ó menos altos según la sensibilidad de la experiencia.

La principal ventaja del test de Fisher es que se basa en un estadistico auxiliar, (vease para ello el articulo de Yates (1984)) pero en nuestra opinión tal ventaja es mas filosofica que real si ella no se trasplanta en una ganancia en potencia. Es mas, como demuestra Plackett (1977), los marginales no aportan información sobre los elementos de la tabla. La auxiliaridad permite a Tocher (1950) convertir el test de Fisher en un test UMPU, pero el propio test exacto de Fisher no goza de tal propiedad (propiedad discutible como se ve mas adelante). En principio, y de nuevo en nuestra opinion, los dos test son licitos, pues ambos se zafan del parametro perturbador por procedimientos tambien licitos, debiendo realizarse la elección en base a razones objetivas de potencia, no en base a razones filosóficas subjetivas.

Es un hecho cierto que el test de Fisher presenta graves irregularidades (ver Liddell(1978) y 2.2.4.), y eso es un dato en su contra. Tambien es un hecho cierto que el test de Fisher es ultraconservador, pues en tamaño suele estar entre ¼ y ½ del error nominal utilizado -Mc Donald (1977) y Liddell (1978)- y eso es otro dato en su contra; de hecho a muchos autores les parece presuntuoso llamar "exacto" a un test conservador. Finalmente, es tambien un hecho cierto que los test bajo la filosofia de Barnard son mas potentes que el de Fisher y, siendo mas reales por no condicionar en lo que inicialmente no estaba condicionado, deben ser preferidos. Historicamente ha sucedido que los test del tipo Barnard tienen tales dificultades de computo que no pudieron aplicarse. Pero hoy dia la situación es distinta.

En resumen, y siempre segun nuestra opinión, test exacto de Fisher debe aplicarse cuando en la realidad los marginales estan dados de antemano (tal sucede en el caso de la señorita y las tazas de te, el test de la mediana de Mood, el test del rango intercuartilico ó en el caso de test de aleatorización), lo que no es frecuente. Para comparar dos proporciones independientes, nos parecen preferibles -por razones de potencia- los test incondicionados que se veran. En todo caso, y para aquellos investigadores amantes del test de Fisher, con anterioridad se ha dado el criterio de elección de la RC que maximiza su potencia al comparar dos proporciones.

1.2. TEST CONDICIONADO ALEATORIZADO: TEST DE TOCHER-LEHMAN

Tocher (1950) propone un mecanismo de sorteo, para la elección de la región crítica, que permite obtener las igualdades (1.23) y (1.24), mecanismo al que llama una extensión de la teoria de Neyman--Pearson del contraste de hipótesis a variables discretas.

En efecto, la teoria de Neymam-Pearson del contraste de hipotesis considera a todos los test de igual tamaño y da "reglas" para de entre ellos seleccionar el "mejor" test. El problema es que para parametros de distribuciones discretas no existen, en general, regiones de probabilidad constante, no pudiendose aplicar aqui estrictamente tal teoria.

La alternativa que propone Tocher consiste en modificar el método de selección de la región del test. Lo usual, al definir la región crítica, es asignar un valor 1 o 0 a cada punto del espacio muestral si acaso este pertenece o no a la región crítica. Es decir $\omega = 1$ si el punto es de la region crítica y 0 si no lo es. Como tal ω puede ser mirado como una probabilidad, podrá considerarse $0 \leq \omega \leq 1$ y asignarle valores por un procedimiento aleatorio adecuado.

Es decir, si se tiene el contraste de hipotesis simple $H_o \equiv \theta = \theta_o$ y $H_1 \equiv \theta = \theta_1$, un procedimiento de test $\Omega(\omega_i)$ requiere dar las ω_i, $i = 1, 2 \ldots.$ que verifiquen:

$$0 \leq \omega_i \leq 1 \;,, \quad \sum_{1}^{\infty} \omega_i p_i(\theta_o) \leq \alpha \text{ fijado}, \quad \sum_{1}^{\infty} \omega_i p_i(\theta_1) = \beta \text{ (maximo)}$$

siendo $p_i(\theta)$ = Probabilidad (Suceso i/θ). (1.26)

Pues bien, Tocher prueba que la asignación optima, para que se verifiquen las (1.26), es la que tomando los puntos muestrales ordenados de mayor a menor valor de $\lambda_i = p_i(\theta_1)/p_i(\theta_o)$, asigna

$$\hat{\omega}_i = 1 \quad \text{si} \quad i = 1, 2, \ldots, S$$

$$\hat{\omega}_i = 0 \quad \text{si} \quad i = S + 2, \ldots.$$

$$\hat{\omega}_{S+1} = \frac{\alpha - \sum_{i=1}^{S} p_i(\theta_o)}{p_{S+1}(\theta_o)} \tag{1.27}$$

Siendo S el valor tal que:

$$\sum_{1}^{S} p_i(\theta_o) \leq \alpha < \sum_{1}^{S+1} p(\theta_o) \tag{1.28}$$

Con esto se consigue efectivamente un test de tamaño α.

Aplicandolo a nuestro caso particular, la forma de realizar el contraste seria:

1º) Fijar el tamaño del test, α.

2º) Para cada posible a_1, determinar el C tal que

$$\sum_{r}^{C} P(x_1/n,n_1,a_1) \leq \alpha < \sum_{r}^{C+1} P(x_1+1/n,n_1,a_1) \tag{1.29}$$

viniendo dada $P(x_1/n,n_1,a_1)$ por la hipergeometrica.

3º) Entrar los puntos 0,1,..,C en la región crítica y determinar si el punto C +1 entra en ella, mediante sorteo con probabilidad de entrar

$$\varrho = \frac{\alpha - \sum_{r}^{C} P(x_1/n,n_1 a_1)}{P(C+1/n,n_1,a_1)} \tag{1.30}$$

Bajo tales criterios, Tocher prueba que el test obtenido es UMPU. Igual conclusión obtiene Lehmann (1959) mediante un proceso de demostración mas elegante que incluye la presentación de la verosimilitud en el formato factorizado indicado en la (1.5) (ver sus pgs. 134-143).

El método de Tocher-Lehman, a pesar de parecer idóneo, ha sido sometido a numerosas criticas -Mc Donald (1977), Liddell (1978), Suissa and Shuster (1985), Mantel and Greenhouse (1968),...- todas ellas basadas en la irracionalidad de decidir por sorteo la posible significación de una tabla (lo que lleva a que, a igual α, investigadores distintos puedan tomar decisiones distintas en base a los mismos datos, perdiendose asi la objetividad cientifica). En la práctica es raro ver una publicación que tome decisiones en base a tal test, (Plackett (1964)) el cual es llamado por los últimos autores citados un procedimiento "repugnante".

Las opiniones anteriores, aun siendo bastante generalizadas, tienen un fuerte componente subjetivo. De mayor interés son las criticas objetivas de Liddell (1978) y Suissa and Shuster (1984). El primero presenta una serie de tablas experimentales en las que se pone de manifiesto lo absurdo de tomar decisiones con el test actual. Los segundos señalan que dicho test es UMPU, pero no UMP, probando que un test sesgado puede ser mas potente que uno insesgado en gran parte del espacio muestral; el argumento es de especial aplicación a las tablas 2 x 2 y valída las aproximaciones al problema que se ven a continuación, aproximaciones que, en nuestra opinión, son preferibles a cualquiera de las expuestas hasta ahora.

1.3. TESTS INCONDICIONADOS BAJO EL CRITERIO DEL MAXIMO

Los tests incondicionados contemplan como RC no los posibles valores de x_1 para el a_1 obtenido, sino los posibles valores de la pareja (x_1,x_2) para los tamaños de muestra dados en la experiencia. Así, el espacio muestral podria representarse como un rectángulo de (n_1+1) x (n_2+1) puntos como el de la Fig. 1.1 (construida para $n_1=6$ y $n_2=8$),

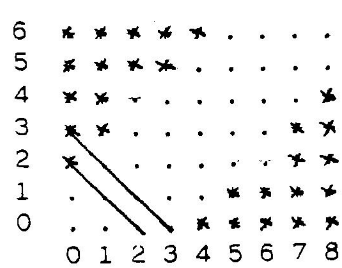

Fig. 1.1

y una determinada RC se representa poniendo una * en aquellos puntos del rectangulo que son de ella; los que no la tienen seran de la región de aceptación (RA). Así, una tabla como la (1;5) - $x_1=1$, $x_2=5$, $n_1=6$, $n_2=8$- daria lugar a un rechazo de H_o para el error que se haya tomado en la construcción de la RC.

Antes de usar una determinada RC es obvio preguntarse cual sera el tamaño del test basado en ella. Barnard (1947) sostiene que el proceder apropiado consiste en determinar

$$\alpha^*(p)=\sum_{RC}\binom{n_1}{x_1}\binom{n_2}{x_2}p^{a_1}q^{a_2} \tag{1.31}$$

pues tal seria el error de tipo I para el valor p del parametro perturbador. Como p es desconocido, un modo -conservador pero inevitable- de zafarse de él es calculando

$$\alpha^* = \underset{0<p<1}{\text{Max}}\ \alpha^*(p) \tag{1.32}$$

lo que constituirá el tamaño buscado del test. Tal criterio sera llamado <u>criterio del maximo</u>, y a él se dedican los apartados siguientes. Barnard tambien considera la posibilidad de obtener el error medio de la región crítica para una cierta distribución de p ó bien obtener $\alpha^*(p_o)$, con p_o el valor que se supone cierto para p, desechando ambos criterios por requerir suposiciones fuertes acerca de p.

Otro procedimiento incondicionado, no bajo el criterio del maximo, lo da la visión de Pearson del test de Fisher (ver 1.1.2). Ahora el error $\alpha^*(p)$ no es obtenido, sino que se anota solo su cota α. En definitiva ello es equivalente a hacer el test exacto de Fisher (de hecho, para una tabla dada, no es preciso obtener toda la RC, bastando con obtenerla para el valor constante a_1) y tal proceder ya fue descalificado en su momento por su conservadurismo. Es interesante observar que en la figura 1.1. las diagonales secundarias $x_1 + x_2 = a_1$ constituyen los espacios muestrales para el test de Fisher. En ella se ha representado, como ilustración, los casos $a_1 = 2$ y $a_1 = 3$.

Finalmente Liddell (1978), y posteriormente Burstein (1981), proponen otro metodo incondicional, tampoco bajo el criterio del maximo, bastante rustico. Su idea es obtener $\alpha^*(p)$ a partir de la (1.31) cambiando el parametro perturbador desconocido p por su estimador de maxima verosimilitud $\hat{p} = a_1/n$; es decir, para ellos el tamaño "estimado" del test seria $\alpha^*(\hat{p})$ para una determinada RC. En particular, para una tabla (x_1, x_2) dada, consideran formada la RC por todos los puntos cuya diferencia de proporciones, en valor absoluto, son superiores ó iguales a la observada, es decir:

$$RC = \left\{ (x_1', x_2') \mid \left| \frac{x_1'}{n_1} - \frac{x_2'}{n_2} \right| \geq \left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \right\} \tag{1.33}$$

con lo que el error P para la tabla problema seria estimado por:

$$P(\hat{p}) = \sum_{RC} \binom{n_1}{a_1} \binom{n_2}{a_2} \left(\frac{a_1}{n_1}\right)^{a_1} \left(\frac{a_2}{n_2}\right)^{a_2} \tag{1.34}$$

y de ahí el nombre de <u>test binomial</u> que le dan. Ambos autores comparan favorablemente su método con el de Fisher, y el segundo tambien con el de Barnard que se verá mas adelante. Sin embargo en tales comparaciones cometen un error bastante frecuente en este tipo de estudios: comparar punto a punto los errores P de dos métodos sin considerar si los mismos respetan el error α. Las potencias de dos test son solo comparables cuando ambos se realizan a igual error α; si uno de ellos se planifica para un α=5% y otro para un α=10%, sus resultados no son comparables.

El test binomial nos parece rechazable por dos razones:

1º) La información de la muestra es introducida en dos ocasiones: al determinar la RC y al obtener el tamaño del test.

2º) El tamaño estimado del test puede llegar a ser bastante irreal. Así, una tabla como la $x_1=3$, $x_2=11$, $n_1=5$, $n_2=11$ tiene un valor $P(\hat{p})=0.038$ según el test binomial, en tanto que $P(p=0.6)=0.079$, con lo que, para ese valor del parametro perturbador, el error α real es mas del doble del indicado por el test binomial. Tal situación es bastante frecuente y no circunstancial de una tabla. Notese además que tal solución es una de las rechazadas por Barnard si se hace $p_o=\hat{p}$. Yates (1984) ataca con algún otro razonamiento el test de Liddell sin citar lo expuesto aqui.

Como resumen, el test de Pearson es demasiado conservador y el test binomial es, en ocasiones, demasiado laxo. Es por ello que los unicos test incondicionados de interes son los obtenidos bajo el criterio del máximo, y a ellos nos dedicamos en lo que sigue. La única diferencia entre unas versiones y otras de tal criterio radica en el modo de obtener la RC, pues, obtenida ella, todos estan de acuerdo en aplicar la (1.32).

1.3.1. Test de Barnard

1.3.1.1. Presentación del test

Barnard (1947) considera que, puesto que se estan comparando proporciones, la RC deberá verificar ciertas condiciones de compatibilidad que le parecen razonables:

1º) Condicion de simetria (S)

Si se está contrastando $H_o \equiv p_1 = p_2$ conta $H_1 \equiv p_1 \neq p_2$, entonces tambien se está contrastando $H_o \equiv q_1 = q_2$ contra

$H_1 \equiv q_1 \neq q_2$, de modo que si el punto (x_1,x_2) es de la RC, tambien deberá serlo el punto $(n_1-x_1,\ n_2-x_2)$, ya que

$$\left|\frac{x_1}{n_1} - \frac{x_2}{n_2}\right| = \left|\frac{n_1-x_1}{n_1} - \frac{n_2-x_2}{n_2}\right| \qquad (1.35)$$

Ello da lugar a que la RC sea simetrica, en ese sentido, ya que basta con obtener la RC para $H_1 \equiv p_1 < p_2$, pues la otra se da por simetria (ver fig. 1.1)

2º) Condición de Convexidad (C)

Si en la RC esta el punto (x_1,x_2) - con $\frac{x_1}{n_1} < \frac{x_2}{n_2}$ - tambien han de estar los puntos que con igual abcisa estan por encima -es decir, los (x_1,x_2+k)- ó que con igual ordenada estan a la izquierda - es decir, los (x_1-k,x_2) - pues en ellos

$$\frac{x_2}{n_2} - \frac{x_1}{n_1} < \frac{x_2}{n_2} - \frac{x_1}{n_1} + \frac{k}{n_2} \qquad (1.36)$$

$(k > 0)$, asegurando tal condición que la RC construida no presenta huecos (ver de nuevo fig. 1.1).

Así pues, y si se va a hacer un test de una cola ($H_1 \equiv p_1 < p_2$) al error α, el investigador deberá comenzar por hacer entrar en la RC el punto del corner $(0,n_2)$ y a continuación obtener el valor α^*_1 para ella a partir de la (1.32), aludiendo el subindice 1 a que se trata del primer paso. Si $\alpha^*_1 < \alpha$, es factible que en la RC puedan entrar aun mas puntos, pero la dificultad está en saber cual es el siguiente punto a considerar dado que la regla de convexidad no establece un orden total entre ellos. Con tal fin Barnard propone un tercer criterio:

3º) Criterio del minimo (M)

Según C hay dos candidatos a entrar en la RC: los puntos

$(0,n_2-1)$ y $(1,n_2)$. Consideremos las RC ampliadas:

$$RC_1 = \left\{(0,n_2);\ (0,n_2-1)\right\}$$
$$RC_2 = \left\{(0,n_2);\ (1,n_2)\right\} \qquad (1.37)$$

y obtengamos los valores α^*_{11} y α^*_{12} que proporcionan a partir de la (1.32); la nueva RC será aquella que de un α^*_{1j} minimo. Si $\alpha^*_2=\mathrm{Min}\left\{\alpha^*_{11},\ \alpha^*_{12}\right\} < \alpha$, el proceso se repite en modo similar; si $\alpha^*_2=\alpha$, el proceso finaliza y tal RC es seleccionada; si $\alpha^*_2>\alpha$, el proceso finaliza y la RC es la del paso anterior (formada solo por el punto del corner). Si en el paso i, varios valores del conjunto $\left\{\alpha^*_{ij},\ j=1,\dots,r\right\}$ son coincidentes y todos ellos minimos, todos los puntos son entrados en la RC. Cuando el test es de dos colas ($H_1 \equiv p_1 \neq p_2$), las regiones criticas se formaron como en una cola mas los otros puntos que indica S; así, para las (1.37):

$$RC_1 = \left\{(0,n_2);(n_1,0);(0,n_2-1);(n_1,1)\right\}$$
$$RC_2 = \left\{(0,n_2);(n_1,0);(1,n_2);(n_1-1,0)\right\} \qquad (1.38)$$

pero el resto del proceder es igual.

Es de observar que el último criterio es llamado criterio del maximo por Barnard en alusión a la (1.32). En esta sección la (1.32) es el criterio que subyace a todos los métodos, y de ahí que el tercer criterio anterior se le haya llamado del mínimo, en alusión a que el punto a incluir en la RC se seleccione por ser el que minimiza los α^*.

El criterio de Barnard puede ser llamado abreviadamente criterio CSM aludiendo a las tres reglas que lo definen. Observese que los criterios C y S dominan sobre M, aplicandose el último solo para hacer entrar el punto ó puntos adecuados de la frontera, es decir, para decidir cuando C y S no son capaces de hacerlo. Barnard afirma que su criterio CSM es el que maximiza el número de puntos de la RC y, por tanto, el que proporciona la maxima potencia. En particular, y como defensa del mismo frente al criterio de Fisher, presentó los valores P de todos los puntos de varias tablas obtenidos por ambos criterios, valores que se reproducen-junto a otros que seran aludidos con posterioridad-en la Tabla 1.4. La comparación de tales valores de P es concluyente a favor del test CSM.

Notese finalmente que tal criterio anterior, al igual que los que siguen, proporcionan RC formados por los puntos mas distantes de la diagonal secundaria (ver Fig. 1.1), es decir por puntos cuya diferencia $|\hat{p}_1-\hat{p}_2|$ es grande, lo que entra dentro de la lógica.

Si $\alpha^*_1(p)$ y $\alpha^*_2(p)$ son los valores de la (1.31) en las RC para las alternativas $H_1 \equiv p_1 < p_2$ y $H'_1 \equiv p_1 > p_2$, entonces, por la condición S, es claro que $\alpha^*_1(p) = \alpha^*_2(1-p)$ y de ahí que $\alpha^*_1 = \alpha^*_2$ al aplicar la (1.32), es decir, los tamaños son el mismo en cualquiera de las colas. Sin embargo es claro que $\alpha^* = \underset{0<p<1}{\mathrm{Max}} \left\{ \alpha^*_1(p) + \alpha^*_1(1-p) \right\}$, el tamaño del test de dos colas, es siempre menor ó igual que el doble del tamaño del test de una cola. Esta caracteristica es general para todos los test de esta sección que respeten la condición de simetria.

1.3.1.1. Criticas y comentarios (Aportación)

A pesar de la evidente ventaja del test de Barnard (CSM), avalado además por Pearson (1947), las dificultades de conjunto del mismo han hecho que se olvidara su técnica durante un cuarto de siglo. Aun así, con el advenimiento de los computadores, tendremos oportunidad de ver como los autores que resucitan su filosofia se olvidan de su condición M, aun conservando el criterio del máximo, pues ella obliga a resolver un número excesivo de veces la (1.32), expresión que, en tiempo de computo , es precisamente la mas costosa. Por ello su criterio CSM no ha vuelto a ser puesto en práctica nunca.

Aun pareciendonos perfecto su criterio CSM, es factible hacerle alguna critica. Notese que las condiciones prioritarias S y C están presididas por la coherencia de no excluir de la RC puntos cuya $|\hat{p}_1-\hat{p}_2|$ sea superior a la de puntos ya incluidos. Sin embargo la condición M puede alterar el criterio. Así, en su ejemplo para $n_1=8$ y $n_2=6$, la primera RC contiene a los puntos corner (0,6) y (8,0), y, cuando aplica la condición M para optar entre incluir el (1,6) y su simetrico, ó el (0,5) y el suyo, concluye que debe incluirse el (0,5) y su simetrico. Cabe observar que el punto (0,5) incluido da una diferencia $|\hat{p}_1-\hat{p}_2|$ = 0,8333 inferior a la del punto (1,6) excluido que es de 0,8750, y ello es contrario a la filosofia que preside las reglas C y S.

Realmente lo que sucede es que Barnard se avala por dos criterios no necesariamente concordantes:

Criterio A: Ir añadiendo puntos a la RC de mayor a menor diferencia de proporciones muestrales $|\hat{p}_1-\hat{p}_2|$

Criterio B: Ir añadiendo a la RC los puntos que ocasionan un mínimo α^* en la (1.32): <u>Método ascendente</u>.

aunque este último admitiria otra versión no contemplada por Barnard y no necesariamente coincidente con la anterior:

Criterio C: Ir quitando de la RC los puntos que ocasionan un minimo α^* en la (1.32): <u>Método descendente</u>

Como el criterio B le llevaria a resolver un número excesivo de veces la (1.32), incluye las reglas S y C como una solución de compromiso entre los criterios A y B. Mas adelante se verá que sucede con el criterio A.

Finalmente, el método CSM presenta la desventaja de que la aplicación de la regla M puede dar lugar a incluir <u>simultaneamente</u> en la RC puntos que resultan empatados en su α^*. Como ello ocasiona un salto apreciable en el tamaño del test, nos parece mas adecuado buscar cualquier otra condición razonable, que añadida al criterio, permita deshacer los empates y obtener asi valores mínimos de P.

1.3.2. Test con fronteras de Fisher

1.3.2.1. <u>Test de Boschloo</u>

Boschloo (1970) es el primer autor que resucita el criterio del máximo de Barnard pero en un camino alternativo a su test CSM que requiere menos tiempo de conjunto. A fin de obviar el asunto de la obtención repetida del maximo de la (1.32) para decidir el orden en que los puntos van entrando en la RC, Boschloo renuncia a los criterios C y M, conserva el criterio S para los test de dos colas y añade un criterio ordenador que en realidad es un viejo conocido: el criterio del test "exacto" de Fisher.

La idea es que, puesto que el test de Fisher bajo la visión de Pearson es conservador, deberán obtenerse las RC (a_1) de una cola para un error γ de trabajo (mayor que el α propuesto) en cada uno de los posibles valores de a_1. Si RC = $\left\{RC(a_1),\ a_1=0,..,n_1+n_2\right\}$ es la RC asi obtenida, determinar para ella el tamaño $\alpha^*(\gamma)$ en base a la (1.32) e ir aumentando ó disminuyendo γ hasta obtener el mayor valor γ_M de él en el que ocurre $\alpha^*(\gamma_M)\leq\alpha$. Para test de dos colas la RC se amplia por la condición S y se procede de modo semejante.

Notese que en el test de Fisher es para $\gamma_M=\alpha$, en tanto que ahora γ_M puede llegar a ser el cuadruple de α. El incremento de puntos en la RC, y la consiguiente ganancia en potencia, son evidentes.

Finalmente, Boschloo presenta tablas de los diversos valores de γ_M que deben ser utilizados para test de 1 ó 2 colas, a diversos errores α, en una amplia gama de tablas (n_1,n_2).

1.3.2.2. Test de Mc Donald

Mc Donald et al (1977, 1981), sin conocer el test de Boschloo segun parece, proponen el mismo criterio de este y ofertan un programa en Fortran V que persigue, fijados n_1,n_2 y α, obtener la RC definitiva, el valor α^* exacto de ella y el valor de p en donde se alcanza tal maximo. Dado que su programa, ó partes de él, sera utilizado con frecuencia en esta memoria, y además es de las pocos que existen bajo el criterio del máximo, a continuación se le resume brevemente. Los pasos que realiza son:

1^er^ paso: Dados n_1,n_2 y α, obtener la RC de una cola por el test de Fisher en cada valor de a_1 y para un error de trabajo de γ. En la primera ocasión tomar $\gamma=3\alpha$ para que asi sobre RC y realizar, en adelante, un procedimiento descendente.

2º paso: Obtenida la RC(γ), calcular $\alpha^*(\gamma)$ por la (1.32). Si $\alpha^*(\gamma) > \alpha$, eliminar de la RC(γ) el punto ó puntos que tengan mayor valor de probabilidad segun la hipergeometrica y repetir este segundo paso hasta lograr una RC en la que $\alpha^* \leq \alpha$.

3er paso: Si el test es de dos colas, utilizar en el 1er paso el valor $\alpha/2$ y ampliar cada una de las RC del 2º paso por el criterio S antes de aplicar la (1.32)

1.3.2.3. Criticas y comentarios (Aportación)

Boschloo cometió dos errores. Publicó sus tablas indicando los valores de γ_M pero no los $\alpha^*(\gamma_M)$ obtenidos. Finalmente, no publico el programa base para realizar el test en cualquier situación. Además sus valores estaban basados en formulas aproximadas. De ahí que, injustamente, en la literatura es conocido el test descrito por test de Mc Donald.

Las tablas de Mc Donald tambien presentan defectos: no dan los valores P para cada punto (pero ese defecto es usual en la literatura) y solo da test de dos colas para errores del 10% y del 20%, que no son los frecuentes.

La base de ambos test es ir entrando (Boschloo) ó sacando (Mc Donald) puntos en la RC en orden a su valor de probabilidad segun la hipergeometrica, criterio de ordenación que sustituye a los C y M de Barnard. De modo explicito:

Criterio de la hipergeometrica (H): Los puntos van entrando en la RC de menor a mayor valor de probabilidad según la hipergeometrica con lo que los test actuales pueden ser llamados SH.

Mas aun, dada una cierta región crítica, los puntos a añadir (ó quitar) son siempre de la frontera externa (ó interna) de ella. Por otro lado, y si suponemos que estamos en la parte noroeste de la RC ($\hat{p}_1 < \hat{p}_2$), dado un punto (x_1, x_2) de ella se tiene que:

$$H(x_1,x_2) > H(x_1,x_2+1) \Longleftrightarrow \hat{p}_2-\hat{p}_1 > \frac{1}{n_2}(\hat{p}_1-1) \qquad (1.38)$$

con $H(x_1,x_2)$ la probabilidad hipergeometrica y $\hat{p}_i = x_i/n_i$. Como la segunda parte siempre es cierta en la RC correcta, tambien lo es la primera. Similarmente se prueba que $H(x_1,x_2) > H(x_1-1, x_2)$ siempre. De ambas se deduce que el criterio H respeta siempre el criterio C, lo que resulta conveniente para que las RC no presenten huecos. En definitiva $H \Rightarrow C$ y el test de Boschloo-Mc Donald puede ser identificado como el test CSH. Tal consideración, introducida en su programa, reducira obviamente el tiempo de computo.

La principal ventaja del método CSH sobre el CSM es su rapidez, pues sustituye el calculo de la (1.32) por el calculo de una probabilidad hipergeometrica, pero los autores no indican nada acerca de la ganancia ó pérdida de potencia que ello puede acarrear. Para investigar este asunto, se ha modificado apropiadamente el algoritmo de Mc Donald para obtener los valores P de todos los puntos relevantes de las tablas del articulo de Barnard (valor de P inferior al 15%). Los resultados se dan en la Tabla 1.4 y avalan las siguientes conclusiones:

1º) Los métodos CSM y CSH coinciden en la tabla 7 x 7 (salvo pequeñas diferencias debidas a los redondeos)

2º) El método CSM es preferido al CSH en las tablas 6 x 8 y 5 x 9 pues el numero de puntos en que P(CSM) < P(CSH) es mayor que el número de puntos en que ocurre lo contrario. Así, para la tabla 6 x 8, de los 10 puntos en que discrepan los valores P de ambos métodos, en 7 el del método CSM es menor que el del método CSH. Para la tabla 5 x 9, de 8 puntos desiguales, 5 son favorables al CSM.

La conclusión es general y se debe a que la condición M produce menos empates que la condición H y menores saltos en el valor de P. Sin embargo las diferencias no son tan manifiestas como para abandonar el criterio CSH dada su evidente ventaja de computo frente al CSM.

Notese que el problema surgido con Barnard respecto de la tabla 6 x 8 sigue ocurriendo ahora: el punto (0,5) entra antes que el punto (1,6).

Tabla 1.4

Valores de P (en %) para los puntos relevantes de las tablas n_1xn_2 que siguen, en los test de dos colas de Fisher, Barnard y McDonald

7 x 7				6 x 8				5 x 9			
Punto	Fisher	Barnard CSM	McDonald CSH	Punto	Fisher	Barnard CSM	McDonald CSH	Punto	Fisher	Barnard CSM	McDonald CSH
0;7	0.058	0.012	0.012	0;8	0.033	0.012	0.012	0;9	0.050	0.012	0.012
0;6	0.230	0.180	0.180	0;7	0.467	0.18	0.11	0;8	0.300	0.19	0.12
0;5	2.100	0.700	0.700	0;6	0.966	0.71	0.53	0;7	2.098	0.62	0.62
0;4	7.000	2.40	2.40	0;5	3.097	2.50	1.98	0;6	3.097	2.30	2.30
0;3	19.000	7.50	7.50	0;4	8.491	5.30	6.59	0;5	8.591	5.30	6.05
1;7	0.23	0.18	0.18	0;3	20.879	13.00	11.38	0;4	22.08	10.00	10.12
1;6	2.900	1.30	1.29	1;7	2.564	1.30	1.30	1;9	0.500	0.13	0.19
1;5	10.00	5.70	5.70	1;6	4.056	6.60	4.03	1;8	2.298	1.30	1.17
1;4	27.000	13.00	12.98	1;5	13.753	11.00	10.69	1;7	9.100	7.40	6.05
2;7	2.100	0.70	0.70	2;8	1.499	0.44	0.71	1;6	26.57	14.00	13.94
2;6	10.000	5.70	5.70	2;7	9.091	3.90	5.74	2;9	2.747	0.86	2.32
3;7	7.000	2.40	2.41	4;8	1.648	8.00	16.19	2;8	9.491	3.80	7.15
3;6	27.000	13.00						3;9	10.989	6.30	7.38

Con respecto al algoritmo de Mc Donald, pueden hacerse las siguientes criticas:

a) El algoritmo no prevee ninguna rutina de calculo que controle la frecuente aparición de "underflow" (frecuente por tratarse de potencias de cantidades que estan entre 0 y 1).

b) El algoritmo da el tamaño del test para una cierta RC, pero no el valor P de cada punto de ella, que es lo que usualmente resulta de interés.

c) El algoritmo, aun siendo rápido, no contiene formulas recurrentes que permitan obtener la probabilidad de un punto dado el anterior.

En el apendice 3 se da un programa realizado por nosotros que evita estos problemas. Tal programa que usa rutinas de paquetes estandares (IMSL, rutinas de Algorithm Section de Applied Statistics, etc...) solventa los problemas a) y b), resolviendo de manera parcial el c).

* * *

1.3.3. Test basado en intervalos de confianza: Test de Ballatori

1.3.3.1. Presentanción del test

Un test menos conocido es el de Ballatori (1982) que, aunque explicitamente no lo dice, está construido tambien bajo el principio del maximo.

Por seguir su notación, notemos con letras mayusculas a la variables aleatorias y con minusculas a la muestra obtenida. Dada una cierta tabla (x_1,x_2), su criterio de obtener el valor P para ella se basa en construir la RC que contiene a todos los puntos que proporcionan una diferencia de proporciones tan extrema ó mas, en igual sentido, que la obtenida; es decir:

$$RC = \left\{ T = X_1/n_1 - X_2/n_2 \geq t = x_1/n_1 - x_2/n_2 \right\} \qquad (1.39)$$

supuesto que $t > 0$ y que la alternativa es $H_1 \equiv p_1 > p_2$. Para dos colas bastará con poner valores absolutos. A continuación en lugar de resolver la (1.32) para tal RC, propone el siguiente método:

1º) Dar un intervalo de confianza para el p común (supuesta cierta H_0) a partir de a_1: sea este $(\bar{p}_1, \bar{p}_2)$, obtenido por el método clásico ó por el método bayesiano.

2º) Calcular $P\left\{RC|p\right\}$, a partir de la (1.31), para valores de $p = \bar{p}_1, \bar{p}_1+\delta, \ldots, \bar{p}_2$, con δ un valor pequeño (δ = 0.05 por ejemplo).

3º) Rechazar H_o si todas las probabilidades del 2º paso son menores ó iguales que α, con α el error de tipo I objetivo.

a cuyos efectos Ballatori (1983) ha propuesto el programa adecuado.

1.3.3.2. Criticas y comentarios (Aportación)

Notese que el criterio de Ballatori respeta las condiciones C y S, pero el orden de entrada de los puntos en la RC es de mayor a menor valor de su diferencia de proporciones, lo que podriamos llamar criterio D de la diferencia de proporciones. En tal sentido el test actual es un test CSD. El criterio D sustituye a los H y M de los métodos anteriores, estando por ver si tal situación mejora en algo a sus competidoras. Ballatori no habla de ello, pero esta memoria si contempla tal analisis con posterioridad.

Por otro lado, la obtención de la (1.32) es obviada por Ballatori mediante un proceso muy discutible:

1º) La información de la muestra, dada por (x_1,x_2), es introducida en dos ocasiones en el problema, no presentando el autor ningún análisis de la influencia de tal proceder ni del hecho de que se trabaja con dos errores simultáneamente: el del test y el del intervalo. Es obvio que ambas cuestiones influiran en el tamaño del test de algun modo.

2º) Cuando calcula un intervalo de confianza para p por métodos bayesianos, el autor utiliza una información "a priori" sobre p que luego no inserta en el test.

3º) Dado que su filosofia es del criterio del maximo, su test deberia compararlo con los de Barnard ó McDonald no con el de Fisher.

4º) Su criterio de decisión es defectuoso pues, bajo su filosofia, el test daria significativo si

$$\max_{p \in (\bar{p}_1, \bar{p}_2)} P\left\{RC \mid p\right\} = \alpha^* \leq \alpha \qquad (1.40)$$

y puede suceder que el máximo no se alcance en ningún punto de los de su malla.

Por todo ello su test nos parece absolutamente rechazable, si bien la idea que lo soporta será aprovechada mas adelante.

1.3.4. Test basado en el estadistico t_{exp} clasico: Test de Suissa

1.3.4.1. Presentación del test.

Suissa and Shuster (1985) proponen ir entrando puntos en la RC de mayor a menor valor del estadístico clásico para grandes muestras.

$$t_{exp} = \frac{(\hat{p}_2 - \hat{p}_1)}{\sqrt{\dfrac{\hat{p}_1\,\hat{q}_1 + \hat{p}_2\,\hat{q}_2}{n}}} \qquad (1.41)$$

supuesto que la alternativa es $H_1 \equiv p_1 < p_2$ y $n_1 = n_2$ (para el caso de dos colas, tomar valores absolutos). El hecho de que el test tradicional tenga por denominador

$$\sqrt{2\hat{p}\hat{q}/n} \qquad (1.42)$$

con $\hat{p}=a_1/n$ y $\hat{q}=a_2/n$, no afecta el orden de selección de los puntos pues cada estadistico es función monotona creciente del otro (ver articulo citado).

Aunque todos sus razonamientos los hace en base a tamaños muestrales iguales (pues su mayor objetivo es predecir tamaños adecuados de muestra), los autores sugieren el test tambien para $n_1 \neq n_2$ y ofrecen un programa al efecto.

En su articulo presentan tablas especificando, para diversos valores de n, los valores de t_{exp} para los que el tamaño del test es $\alpha^* \leq \alpha$, con α=5%, 2,5% y 1%, y los propios valores α^* alcanzados.

La mayor aportación al problema es que dan un método riguroso para acotar tanto como se desee el valor de α^*, método que, suplementado en un programa, dara mas precisión y rapidez que el método de maximización de McDonald (que es mas rústico). Un test en todo similar a este (está basado en el estadistico χ^2_{exp} con la corrección de Yates), presentaron Garside y Mack (1967, 1968).

1.3.4.2. Criticas y comentarios (Aportación)

Es inmediato comprobar que la (1.41) respeta la condición C. Asimismo puede verse que el valor de t_{exp} en el punto (x_1, x_2) difiere del valor de t_{exp} en el punto (x_1,x_2+1) en una cantidad cuyo signo es el de

$$- \left\{(x_2+x_1+1)\ (x_2-x_1)+2x_1(n-x_2)\right\} \qquad (1.43)$$

el cual, por ser negativo, asegura que la condición S tambien se respeta. Si finalmente a la condición de maximo valor de la t_{exp} la notamos por T, el test anterior podra denominarse CST.

Los autores validan su test por comparación con el de Fisher en una tabla 10 x 10. Tal proceder no es suficiente por cuanto ya se conoce que los test bajo el criterio del máximo son mas potentes que el de Fisher, de modo que la validación del mismo deberia haberse llevado a cabo por comparación con los criterios similares ya existentes (CSM y CSH fundamentalmente). Las tablas 1.5 y 1.6 rellenan el hueco.

La misma tabla 10 x 10 de Suissa, que utiliza para comparar con Fisher, es utilizada ahora para comparar su método con el de McDonald (Tabla 1.5). En ella aparecen los valores P, por los dos métodos, para todos los puntos relevantes. Notese que para aquellos puntos en que P(CST) $\leq$ 15%, sucede que P(CSH)< P(CST) en 15 ocasiones, y al reves solo en 3. En esta tabla el test de McDonald es netamente superior al de Suissa (se han hecho test de una cola pues tal era el ejemplo de Suissa).

La Tabla 1.6 presenta los valores α^* obtenidos por los dos test para un α =5% objetivo y diversos tamaños de muestra. Por paralelismo a las tablas de Suissa, el test es ahora de dos colas y para n_1=n_2. Conviene notar que la comparación de los

Tabla 1.5

Valores de P por el test de Suissa (CST) y de Mc Donald (CSH) -una cola- para los puntos relevantes de una tabla con $n_1=n_2=10$

Puntos	P(CST)	P(CSH)	Puntos	P(CST)	P(CSH)
0,10	$9x10^{-7}$	$9x10^{-7}$	1,4	0,0113	0,0064
0,9	$1x10^{-5}$	$2x10^{-5}$	1,3	0,1647	0,1317
0,8	0,0001	0,0001	1,2	0,3367	0,4119
0,7	0,0004	0,0002	2,8	0,0059	0,0039
0,6	0,0017	0,0012	2,7	0,0206	0,0114
0,5	0,0063	0,0059	2,6	0,0474	0,0312
0,4	0,0211	0,0207	2,5	0,1103	0,0755
0,3	0.0438	0,0576	2,4	0,2617	0,1648
0,2	0,1074	0,1316	2,3	0,3702	0,4119
0,1	0,2610	0,4119	3,7	0,0578	0,0447
1,9	0,0002	0,0001	3,6	0,1316	0,0987
1,8	0,0012	0,0004	3,5	0,2617	0,2003
1,7	0,0039	0,0039	3,4	0,3883	0,4119
1,6	0,0113	0,0064	2,6	0,2617	0,2251
1,5	0,0311	0,0211	2,5	0,4119	0,4119

Tabla 1.6

Valores del tamaño del test, α^*, para el test de Suissa (CST) y de Mc Donald (CSH) para diversas tablas con $n_1=n_2$ y para un test de dos colas al error objetivo $\alpha=0,05$. La columna Δ alude a la diferencia entre el número de puntos de las RC de CSH y CST

n_1	n_2	α^*(CST)	α^*(CSH)	Δ
10	10	0.0476	0.0422	0
15	15	0,0417	0,0433	0
20	20	0,0438	0,0464	+1
25	25	0,0458	0,0463	0
30	30	0,0467	0,0469	0
35	35	0,0476	0,0480	0

tamaños de los dos test no debe ser definitiva por cuanto pudiera ocurrir que $\alpha^*(CSH) < \alpha^*(CST)$ y, sin embargo, la RC del primero contengan mas puntos que la del segundo; en tal caso el primero seria mas potente. Como caso extremo piensese en una RC de un sólo punto en la que $\alpha^* = \alpha$; obviamente, y a pesar de alcanzarse el error objetivo α , tal RC no resulta de utilidad y tendrá una potencia despreciable. Sin embargo, el criterio es de utilidad si se contemplan varias tablas dado que las RC de ambos métodos respetan C y S y solo diferirán en algunos puntos de la frontera. Asi, aunque circunstancialmente en una tabla un test aparezca, indebidamente, como mejor que otro, ello será en general por haber permutado puntos, y así las otras tablas lo pondran de manifiesto. De todos modos, en nuestra opinión, ambos procedimientos de comparación deben simultanearse, y asi la Tabla 1.6 tambien incluye la diferencia entre los números de puntos de ambas RC.

La comparación es en general favorable al test CSH y el CST será abandonado en lo que sigue.

1.3.5. Propuestas de nuevos test incondicionados (Aportación)

Se presentan a continuación una serie de aportaciones de test incondicionados que se comparan con los ya existentes, haciendose al final una serie de disquisiones acerca de cual seria el test incondicionado mas potente.

1.3.5.1. Test basado en intervalos de confianza

El principal error de Ballatori (ver 1.3.3.) consistió en tomar la regla de los intervalos de confianza para resolver más fácilmente la (1.32). Aqui se va a ver que tal regla puede aplicarse como criterio de selección del orden de entrada de los puntos en la RC.

Mitra (1969) presentó un test parecido al que nosotros presentamos aqui pero con dos graves inconvenientes: a) Calcula los intervalos de confianza basandose en la aproximación normal; b) No determina el error real α, distinto del nominal de los intervalos de confianza. Por estas dos razones el test de Mitra nos parece desechable y no nos referiremos a él en adelante.

Consideremos de nuevo el caso $H_1 \equiv p_1 > p_2$. Un intervalo de confianza clásico para p_1, basado en su muestra, se obtiene determinando el valor p'_1 tal que $Pr\left\{X_1 \geq x_1 \mid p_1 = p'_1\right\} = \alpha$, con lo que $p_1 \geq p'_1$ con una confianza de $1-\alpha$. Con ello, a efectos del test, el resultado mas desfavorable será $p_1 = p'_1$ y la probabilidad de un resultado X_2 tan desfavorable ó mas que el obtenido es $Pr\left\{X_2 \leq x_2 \mid p_2 = p'_1\right\}$, cantidad que nos llevará a decidir H_1 si es menor ó igual que α.

Tal modo de proceder es equivalente a determinar intervalos de confianza de una cola para p_1 y p_2 en la forma:

$$p_1 \in (p'_1 \ ;1) \mid Pr\left\{X_1 \geq x_1 \mid p_1 = p'_1\right\} = \alpha \tag{1.44}$$

$$p_2 \in (0 \ ;p'_2) \mid Pr\left\{X_2 \leq x_2 \mid p_2 = p'_2\right\} = \alpha \tag{1.45}$$

y rechazar H_o cuando los intervalos sean disjuntos, es decir cuando $p'_1 > p'_2$. Graficamente, el procedimiento consiste en calcular intervalos de aceptación de una cola para X_1 y X_2 en cada posible valor de p. Así, dado un $X_1 = x_1$ será $p \geq p'_1$ (con una cierta confianza), y dado un $X_2 = x_2$ será $p \leq p'_2$ (con una cierta confianza), decidiéndose H_1 cuando $p'_1 > p'_2$ como pasa en la figura 1.2. En tal caso la región crítica será RC = $\bigcup\limits_{x_1} RC\,(x_1)$, con

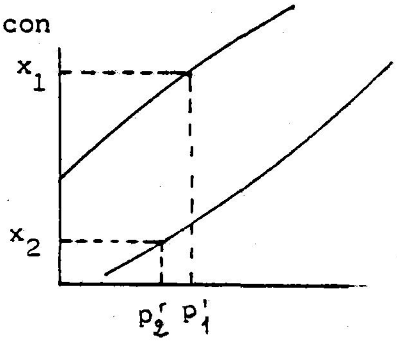

Fig. 1.2

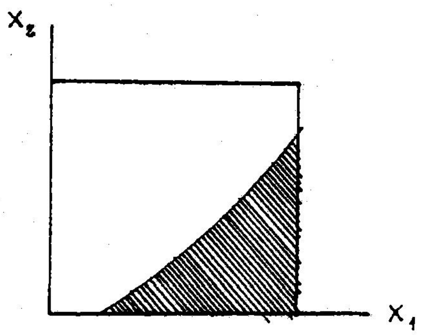

Fig. 1.3

$$RC(x_1) = \left\{ x_2 \mid p'_2\,(x_2) < p'_1\,(x_1) \right\}$$, y tendrá la forma de la figura 1.3.

Con la formulación descrita, el error real de tipo I del test (digamos α^*) no se puede obtener explicitamente, aunque si puede obtenerse una cota del mismo.

Así

$$\alpha^*(p) = Pr\left\{RC \mid H_o\right\} = \sum_{x_1} Pr\left\{X = x_1 \mid p\right\}.Pr\left\{RC(x_1) \mid p\right\}$$

ahora bien, si $x_2(p'_1)$ es el mas grande x_2 que proporciona un $p'_2 < p'_1$, entonces $RC(x_1) = \left\{x_2 \leq x_2(p'_1)\right\}$, y así:

$$\alpha^*(p) = \sum_{x_1} Pr\left\{X_1 = x_1 \mid p\right\} . Pr\left\{X_2 \leq x_2(p'_1) \mid p\right\}$$

Pero

$$Pr\left\{X_2 \leq x_2(p'_1) \mid p\right\} = Pr\left\{p < p'_1\right\}.Pr\left\{RC(x_1) \mid p < p'_1\right\} + P_r\left\{p \geq p_i\right\} \cdot Pr\left\{RC(x_1) \mid p \geq p'_i\right\} \leq \alpha + (1-\alpha)\alpha = 2\alpha - \alpha^2$$

y así $\alpha^* \leq (2\alpha - \alpha^2)\sum_{x_1} Pr\left\{X = x_1\right\} = 2\alpha - \alpha^2$, con lo cual $\alpha \leq 1 - \sqrt{1-\alpha^*}$

lo que indica que para hacer un test al error $\alpha^* = 0{,}05$, los intervalos deben construirse a un error $\alpha = 0{,}0253$ como máximo. Sin embargo tal criterio va a ser conservador y el α a tomar habrá de obtenerse por ensayo-error.

Obsérvese que, supuesto conocido α, el test pasa por determinar los valores p'_1 y p'_2 y compararlos. El modo más comodo de obtenerlos es a través de la relación entre la función de distribución de la Binomial y la distribución F de Snedecor. Así, como

$$Pr\left\{X \leq x\right\} = 1 - Pr\left\{F\left[2(x+1);\ 2(n-x)\right] \leq \frac{n-x}{x+1} \cdot \frac{p}{1-p}\right\}$$

con $X \sim B(n;p)$, entonces las (1.44) y (1.45) pueden resolverse explicitamente así:

$$p'_1 = \frac{x_1}{(n_1-x_1+1)F_1 + x_1} \qquad \text{con } F_1 = F_{1-\alpha}\left[2(n_1-x_1+1);\ 2x_1\right] \tag{1.46}$$

$$p'_2 = \frac{(x_2+1)\ F_2}{(n_2-x_2) + (x_2+1)\ F_2} \qquad \text{con } F_2 = F_{1-\alpha}\left[2(x_2+1);\ 2(n_2-x_2)\right] \tag{1.47}$$

de modo que se decidirá H_1 cuando $p'_1 \geq p'_2$, es decir cuando

$$\frac{x_1(n_2 - x_2)}{(x_2+1)\ (n_1-x_1+1)} \geq F_1 \cdot F_2 \tag{1.48}$$

a condición de que $x_1/n_1 > x_2/n_2$.

Como el valor de α no es conocido, habrá de adoptarse un criterio de aproximaciones sucesivas hasta lograr encontrar el valor que ocasiona un error real de test prefijado. Así, si se desea realizar un test de error α^+ de tipo I, los pasos a realizar para construir la región critica serán:

1º) Tomar un α inicial de trabajo ($\alpha = 2\alpha^+$ por ejemplo).

2º) Determinar $RC(\alpha, \alpha^+) = \left\{(x_1,x_2) \mid x_1/n_1 > x_2/n_2\right\}$ y verifican la (1.48)

3º) Calcular $\max\limits_{0<p<1} \left\{RC(\alpha,\alpha^+) \mid p\right\} = \alpha^*$

4º) Si $\alpha^* < \alpha^+$, aumentar α y repetir desde el paso 2º.

Si $\alpha^* > \alpha^+$, disminuir α y repetir desde el paso 2º.

Si $\alpha^* = \alpha^+$, el proceso finaliza y la última región crítica obtenida sirve para realizar el test al error α^+.

En la práctica, sin embargo, lo que interesa es decidir el nivel P de significación de una tabla determinada, y ello requiere menos tiempo de cómputo; así los pasos serían:

a) Dada una cierta pareja observada (x'_1, x'_2) verificando que $x'_1/n_1 > x'_2/n_2$, determinar el valor $\alpha(x'_1, x'_2)$ que verifica la (1.43) puesta como igualdad.

b) Determinar $RC(x'_1, x'_2) = \left\{(x_1, x_2) \mid x_1/n_1 > x_2/n_2\right\}$ y verifican la (1.48) para $\alpha = \alpha(x'_1, x'_2)$

c) Calcular $P(x'_1, x'_2) = \underset{0<p<1}{\text{Max}} \quad \Pr\left\{RC(x'_1, x'_2) \mid p\right\}$,

lo que constituye el error exacto P del test para la tabla propuesta.

Notese que el test actual tambien está basado en el criterio del maximo (ver los pasos 3º) y c) de los parrafos anteriores) y que la novedad del mismo es que entra los puntos en la RC en orden de menor a mayor valor del α que verifica la (1.48) puesta como igualdad. A tal criterio le llamaremos <u>criterio I de incompatibilidad de los intervalos de confianza</u>.

Por otro lado, si el punto (x_1, x_2) es de la RC -es decir verifica la (1.48)- el punto (x_1, x_2-1) tambien es de ella, con lo que se verifica la condición C. En efecto:

1) La fracción de la (1.48) para el punto (x_1, x_2) es siempre menor que tal fracción para el punto (x_1, x_2-1)

2) El valor de F_1 de la (1.48) es identico en ambos puntos.

El valor de F_2 en el primer punto es del tipo $F_2(r+2;s-2)$ y del tipo $F_2(r;s)$ en el segundo, y en una F de Snedecor es siempre $F_2(r+2;s-2) > F_2(r;s)$.

De ambas resulta que el 2º criterio de la (1.48) es siempre mayor en el primer punto que en el segundo.

3) De 1º) y 2º) se deduce que si (x_1,x_2) verifica la (1.48), (x_1,x_2-1) tambien la verifica.

Finalmente, y si se desea hacer un test de dos colas, notese que si (x_1,x_2) verifica la (1.48), entonces (n_1-x_1,n_2-x_2) tambien la verifica cambiando los subindices 1 por 2 en la (1.48), lo que constituye el test para la otra cola. Con ello tambien se verifica el criterio S y el test actual puede ser descrito como un test CSI.

Notese que el test actual está basado en un estadistico intimamente relacionado con la razón de odds clásica para tablas 2 x 2 (primera fracción de la (1.48)).

1.3.5.2. Test basado en la diferencia de proporciones

Recordemos que Ballatori (ver 1.3.3) utilizaba el criterio CSD para construir la RC, pero aplicaba mal el criterio del máximo. La propuesta actual es comprobar que sucede con el criterio CSD.

El análisis de la Tabla 1.7 permite concluir que el test actual es bastante mas insensible que el test CSI visto anteriormente, pues los tamaños del primero, para un test al 5%, son siempre inferiores (a veces apreciablemente) a los del segundo, produciendose las mayores diferencias en las tablas de igual tamaño.

Tabla 1.7

Tamaños α^* de los test CSD y CSI para diversos valores de tamaños de muestra y para un test de dos colas al error α=5%.

n_1	n_2	α^*(CSD)	α^*(CSI)	n_1	n_2	α^*(CSD)	α^*(CSI)
5	5	0,0117	0,0215	10	25	0,0467	0,0469
5	10	0,0236	0,0413	15	15	0,0411	0,0433
5	15	0,0245	0,0371	15	20	0,0444	0,0466
5	20	0,0322	0,0433	15	25	0,0453	0,0466
5	25	0,0380	0,0477	20	20	0,0385	0,0464
10	10	0,0296	0,0422	20	25	0,0485	0,0490
10	15	0,0400	0,0403	25	25	0,0328	0,0463
10	20	0.0425	0,0428				

La razón de la diferencia entre uno y otro metodo es que, estando ambos de acuerdo en las condiciones C y S, los puntos posibles a añadir a las fronteras resultan mas empatados con la regla D que con la regla I: la regla I da un orden parcial mas fino que la regla D y recuerdese que, cuando varios puntos resultan empatados por una regla, todos entran juntos en la RC. Una buena muestra de ello la dan las figuras 1.4: la RC del criterio CSI consta de mas puntos (merced a tener menos empates) que la del CSD.

Figura 1.4

Regiones críticas para un test de dos colas, al error $\alpha=5\%$, en una tabla con $n_1=n_2=10$ y con los métodos C.S.D. y C.S.I.

```
* * * * * * * . . . .            * * * * * * . . . . .
* * * * * . . . . . .            * * * * * . . . . . .
* * * * . . . . . . .      CSD   * * * * . . . . . . .
* * * . . . . . . . .            * * * . . . . . . . .
* * . . . . . . . . *            * * . . . . . . . . .
* . . . . . . . . . *            * . . . . . . . . . *
* . . . . . . . . * *            . . . . . . . . . * *
. . . . . . . . * * *            . . . . . . . . * * *
. . . . . . . * * * *  CSI       . . . . . . . * * * *
. . . . . . * * * * *            . . . . . . * * * * *
. . . . * * * * * * *            . . . . . * * * * * *
```

En resumen: el criterio C.S.D. es rechazable frente al criterio C.S.I.

1.3.5.3. Test basado en la maxima verosimilitud

Otra posible regla de entrada de puntos en la RC, siempre respetando C y S que parece es universal, seria con la condición V de máxima verosimilitud, para dar así un criterio CSV. Para un punto (x_1,x_2) dado, su verosimilitud seria

$$\binom{n_1}{x_1}\binom{n_2}{x_2} p^{a_1} q^{a_2} \tag{1.49}$$

la cual es máxima en el valor $\hat{p}=a_1/n$. La idea es ir entrando los puntos en la RC en orden de menor a mayor valor de su maxima verosimilitud

$$\binom{n_1}{x_1}\binom{n_2}{x_2} \hat{p}^{a_1} \hat{q}^{a_2} \tag{1.50}$$

es decir, de menor a mayor probabilidad maxima.

1.3.5.4. Comparación de los distintos métodos

De todos los métodos incondicionados vistos hasta ahora (excluido el CSM de Barnard), los unicos que no han sido descartados son los CSH (de McDonald), CSI (de intervalos de confianza) y CSV (de máxima verosimilitud). Para efectuar la comparación de ellos y poder elegir el mejor (el mas potente) se dan las tablas 1.7 y 1.8. De ellas se deduce que, aunque las diferencias no son excesivas, los test CSH y CSI son preferibles y prácticamente iguales (con ligera ventaja para el CSH).

Tabla 1.7

Valores de P, para los puntos relevantes de las tablas indicadas, con los métodos CSH (intervalos de confianza),CSI (Intervalos de Confianza) y CSV (maxima verosimilutd) para un test de dos colas.

7 x 7				6 x 8				5 x 9			
Punto	CSH	CSI	CSV	Punto	CSH	CSI	CSV	Punto	CSH	CSI	CSV
0,7	0.012	0.012	0.012	0,8	0.012	0.012	0.012	0,9	0.012	0.012	0.012
0,6	0.180	0.180	0.180	0,7	0.11	0.11	0.11	0,8	0.12	0.12	0.12
0,5	0.700	0.700	0.700	0,6	0.53	0.53	0.53	0,7	0.62	0.62	0.62
0,4	2.40	2.40	2.37	0,5	1.98	1.98	1.98	0,6	2.20	2.20	2.20
0,3	7.50	7.51	7.51	0,4	6.59	6.59	6.59	0,5	6.05	6.05	6.05
1,7	0.18	0.18	0.98	0,3	11.38	11.38	11.38	0,4	10.12	10.12	13.82
1,6	1.29	1.29	1.29	1,8	0.18	0.18	0.65	1,9	0.19	0.19	0.19
1,5	5.70	5.73	5.73	1,7	1.30	1.29	2.57	1,8	1.17	1.17	1.17
1,4	12.98	13.04	13.04	1,6	4.03	4.03	7.36	1,7	6.05	6.05	4.52
2,7	0.70	0.70	0.44	1,5	10.69	10.69	10.69	1,6	13.94	13.94	12.29
2,6	5.70	5.74	3.94	2,8	0.71	0.71	1.29	2,9	2.32	2.32	2.32
3,7	2.41	2.38	1.72	2,7	5.74	5.74	5.49	2,8	7.15	7.15	7.15
3,6	13.00	13.05	10.13	3,8	4.27	4.27	4.27	5,74	7.38	7.38	13.94
4,7	7.52	7.52	6.16								

Tabla 1.8

Tamaños α^* de los test de dos colas CSH, CSI y CSV, realizados a un $\alpha=5\%$, en diversas tablas. Las dos últimas columnas son las diferencias, Δ_I y Δ_V, entre el número de puntos de la RC del test CSH y de los otros dos test.

Tamaños de muestra	Tamaños α^* de los test			Incremento de puntos de los RC del test CSH vs. los otros dos	
(n_1,n_2)	CSH	CSI	CSV	Δ_I	Δ_V
5;5	0.0215	0.0215	0.0215	0	0
5;10	0.0413	0.0413	0.0474	0	- 1
5;15	0.0371	0.0370	0.0500	0	0
5;20	0.0433	0.0433	0.0433	0	0
5;25	0.0477	0.0477	0.0477	0	0
10;10	0.0422	0.0422	0.0422	0	0
10;15	0.0403	0.0403	0.0460	0	0
10;20	0.0428	0.0428	0.0427	0	0
10;25	0.0459	0.0469	0.0490	0	+ 3
15;15	0.0433	0.0433	0.0455	0	0
15;20	0.0466	0.0466	0.0465	0	0
15;25	0.0463	0.0466	0.0496	0	+ 2
20;20	0.0464	0.0464	0.0470	0	0
20;25	0.0490	0.0490	0.0490	0	0
25;25	0.0462	0.0462	0.0482	0	+ 4

1.4. TESTS DE ALEATORIZACION

Cuando los n individuos de una muestra se reparten al azar entre dos tratamientos produciendo estos en aquellos un efecto positivo (+) ó negativo (-), hay justificaciones teoricas -Lehmann (1959), pags 189-196- para emplear un test de aleatorización con el fin de decidir si ambos tratamientos son igualmente efectivos (H_o) ó no (H_1). Se entiende que la igual efectividad alude a la igualdad de los porcentajes de respuestas positivas con cada tratamiento ($H_o \equiv p_1=p_2(=p)$). En lo que sigue, por adecuarnos a la notación empleada en la literatura, la decisión se referira a una tabla como la 1.9 (que se explica por si sola) entendiendo que las letras mayusculas aluden a la tabla experimentalmente obtenida, en tanto que las mayusculas aludiran a los valores aleatorios que podrian haber ocurrido (compatibles siempre con los valores muestrales n_1 y n_2 que son, en todo caso, fijos).

Tabla 1.9

	B	$\bar{B}$	Total
A	a	c	n_1
$\bar{A}$	b	d	n_2
Total	m_1	m_2	n

Una vez más, el problema admite dos visiones: la condicionada y la incondicionada. Para cada una de ellas existe en la actualidad un test propuesto; ambos, y uno mas que se propondra en esta memoria, son el objeto de la sección actual,

1.4.1. Test condicionado: Test exacto de Fisher

Dado que la probabilidad de la hipergeometrica

$$P(a|n_1,n_2,m_1,m_2) = \frac{\binom{n_1}{a}\binom{n_2}{b}}{\binom{n}{m_1}} \qquad (1.51)$$

puede entenderse como la probabilidad de obtener "a" signos + en las n_1 primeras casillas y "b" signos + en las n_2 siguientes, cuando se colocan al azar m_1 signos + en las $n=n_1 + n_2$ casillas totales, entonces el test exacto de Fisher es tambien un test de aleatorización para el problema propuesto y su filosofia es plenamente adecuada.

Con ello, y para una alternativa de una cola ($H_1 \equiv$ el primer tratamiento produce menos signos + que el segundo), el valor P_F de dicho test para una tabla experimental como la obtenida (tabla 1.9) sera:

$$P_F = \sum_{A=r}^{a} \frac{\binom{n_1}{A}\binom{n_2}{m_1-A}}{\binom{n}{m_1}} \quad \text{con } r=\text{Max}\left\{0;\ m_1-n_2\right\} \qquad (1.52)$$

Para dos colas ver lo dicho en el apartado 1.1.1.

1.4.2. Tests incondicionados.

El test de Fisher actua determinando cual es la probabilidad de una tabla, tan extraña ó mas que la obtenida, de entre todas las que dan m_1 signos +. Los dos tests que siguen no efectuan tal condicionamiento.

1.4.2.1. Test de Ballatori.

El test "exacto" de Fisher es pues tambien un test de aleatorización condicional, por cuanto está construido bajo el supuesto de que necesariamente han de producirse m_1 signos + en total. Ballatori (1982) propone un test de aleatorización incondicional basado en la idea de que es posible cualquier número de signos +. Veamos el esquema de su razonamiento.

Si R es el conjunto de los modos posibles de asignar al tratamiento I n_1 de las n unidades experimentales, él constará de $\binom{n}{n_1}$ elementos equiprobables. Por cada uno de ellos hay un conjunto S de 2^n secuencias posibles de signos + y -.

Pero si la hipótesis nula (H_o) de igual eficacia de ambos tratamientos es cierta, entonces la respuesta observada en cada unidad experimental es independiente del tratamiento aplicado y, por el principio de indiferencia, cualquier elemento del producto RxS es equiprobable. Si ahora definimos $T=(B/n_2)-(A/n_1)$, entonces $E(T|H_o)=0$ y $E(T|H_1) > 0$, si la alternativa (H_1) es que el segundo tratamiento es mas eficaz que el primero, y la region crítica será:

$$RC_\alpha = \left\{T \geq t_\alpha \mid P(T \geq t_\alpha \mid H_o) \leq \alpha \right\} \qquad (1.53')$$

aunque en la práctica es más cómodo obtener el error P_B exacto de rechazo mediante la expresión:

$$P_B = P(T \geq t|H_o) \qquad (1.54)$$

con $t = (b/n_2)-(a/n_1)$ lo que experimentalmente ha ocurrido. El subindice B de P alude a que es el valor P del test de Ballatori.

Alternativamente, cuando la hipótesis alternativa (H'_1) sea que los dos tratamientos no son igual de efectivos entonces:

$$P'_B = P\ (|T| \geq |t|\ |H_o) \qquad (1.55)$$

será el error P'_B del test de dos colas. En consecuencia, y para el caso de H_1, el error P_B dado por la (1.54) se obtendrá dividiendo por 2^n el número de tablas del conjunto S que proporcionan un valor $T \geq t$; de ahí que:

$$P_B = \frac{1}{2^n} \sum_{A \in I_A} \sum_{B \in I_{B/A}} \binom{n_1}{A} \binom{n_2}{B} \tag{1.56}$$

con:

$$I_A = \left\{ A \mid 0 \leq A \leq a + \left[\frac{dn_1}{n_2}\right]^- \right\} \tag{1.57}$$

$$I_{B|A} = \left\{ B \mid b + \left[\frac{(A-a)n_2}{n_1}\right]^+ \leq B \leq n_2 \right\} \tag{1.58}$$

en donde el simbolo $[x]^-$ (ó $[x]^+$) alude al primer entero menor ó igual (ó mayor ó igual) que x.

1.4.2.2. Criticas al test de Ballatori (Aportación)

En el test descrito el conjunto R no juega ningun papel, por cuanto la situación probabilística es la misma para cada punto de R. De hecho la expresión (1.56) incluye como denominador 2^n (el número de puntos de S) puesto que el P_B así obtenido vale para todo punto de R y estos son equiprobables.

Por otro lado, y puesto que bajo H_o todos los elementos de RxS los supone equiprobables, todos los elementos de S habran de ser tambien equiprobables (dado que los R lo son y cada elemento de R da lugar a igual número (2^n) de elementos de S). Ahora bien, la suposición de que cualquier secuencia de n signos + y -es equiprobable (bajo H_o) es equivalente a suponer que la proporción de signos + es igual a la signos - e igual a ½, por lo que, siendo el signo independiente del tratamiento segun H_o, se desprende que se está comprobando $H_o \equiv p_1=p_2=p=½$, y el test planteado no es un test de aleatorización, sino uno paramétrico bien definido en el que se contrasta la hipótesis, bastante fuerte, de si las proporciones de signos + son idénticas con ambos tratamientos e iguales a su vez a ½. De hecho la expresión (1.56) puede ponerse en este otro formato:

$$P_B = \sum_{A \in I_A} \sum_{B \in I_{B/A}} \binom{n_1}{A} \binom{n_2}{B} \left(\frac{1}{2}\right)^{A+B} \left(\frac{1}{2}\right)^{n-A-B} \qquad (1.59)$$

lo que no es otra cosa que dar la probabilidad de que ocurra la region crítica cuando las variables A y B siguen distribuciones binomiales del tipo $B(n_1;\ p=½)$ y $B(n_2;\ p=½)$ respectivamente.

Es claro que, si acaso $p \neq ½$ - como sucederá en la mayoria de las ocasiones- el test de Ballatori dará un error de tipo I que puede sobrepasar el nominal (si p es suficientemente distinto de ½).

1.4.2.3. Propuesta de un nuevo test de aleatorización no condicionado. (Aportación)

Se ha visto que Ballatori utiliza el argumento de que cualquier secuencia de signos es equiprobable. Parece mas razonable suponer que, puesto que se desconoce la abundancia real de signos +, cualquier valor del numero M_1 de signos + es igualmente posible; con ello, y puesto que $M_1=0,1,\ldots,n$, se tendrá que:

$$P(M_1=m_1)=\frac{1}{n+1} \qquad (1.60)$$

Por otro lado, si es cierta la hipótesis nula (H_0) de que ambos tratamientos son igual de efectivos, cualquiera de las $\binom{n}{M_1}$ combinaciones posibles de M_1 signos + y $(n-M_1)$ signos - es igualmente posible, y asi cualquier pareja de muestras n_1 y n_2 elementos con A y B signos +, respectivamente, tendrá por probabilidad:

$$P(A,B|n_1,n_2)=P(A+B|n_1,n_2).\ P(A,B|n_1,n_2,A+B)=\frac{1}{n+1}\ \frac{\binom{n_1}{A}\binom{n_2}{B}}{\binom{n}{A+B}} \qquad (1.61)$$

y así el error P de la región crítica de Ballatori será:

$$P_A = \frac{1}{n+1} \sum_{A \in I_A} \sum_{B \in I_{B/A}} \frac{\binom{n_1}{A}\binom{n_2}{B}}{\binom{n}{A+B}} \qquad (1.62)$$

Con el subindice A indicando que P corresponde al test de aleatorización propuesto.

Veamos una forma alternativa de obtener P_A que es mas conveniente a efectos de cálculo. Dado un cierto valor de M_1, entonces las tablas deseadas habrán de verificar que $T \geq t$, es decir

$$\frac{B}{n_2} - \frac{A}{n_1} \geq t \tag{1.63}$$

por lo que, poniendo $B=M_1-A$ y despejando A queda:

$$\mathrm{Max}\left\{0;\ M_1-n_2\right\} \leq A \leq \left[\left(\frac{M_1}{n_2} - t\right)\frac{n_1 n_2}{n}\right]^- \tag{1.64}$$

debiéndose la desigualdad izquierda a que, por valer B como máximo n_2, A valdrá al menos M_1-n_2 (ó 0 si es negativo). Por otro lado el mínimo valor posible de M_1 se alcanza cuando $B=M_1$ y $A=0$, con lo que, por la (1.63), habrá de ser $M_1 \geq [tn_2]^+$; similarmente su máximo valor se alcanza cuando $B=n_2$ y $A=M_1-n_2$, con lo que, de nuevo por la (1.63), habrá de ser $M_1 \leq n - [tn_1]^+$. Resumiendo:

$$[tn_2]^+ \leq M_1 \leq n - [tn_1]^+ \tag{1.65}$$

Finalmente:

$$P_A = \frac{1}{n+1} \sum_{M_1} \sum_{A} \frac{\binom{n_1}{A}\binom{n_2}{B}}{\binom{n}{M_1}} \tag{1.66}$$

con los índices de las sumas tomando valores segun las (1.64) y (1.65). Si tal valor de P es notado por $P_A(a,b)$, aludiendo a que es el error P del test de aleatorización cuando se obtuvo la tabla indentificada por la pareja (a,b), entonces

$$P_A(a,b) = \frac{1}{n+1} \sum_{M_1} P_F\ (t^+|M_1) \tag{1.67}$$

Con $P_F\ (t^+|M_1)$ el error P del test de Fisher para una tabla cuyo total de signos + es M_1 y con un valor de (A,B) tal que su T sea lo más próximo posible, por arriba, a t (cuando $M_1=m_1$, $P_F(t|M_1)$ es justamente el error P del test exacto de Fisher clásico y dado por la (1.52)). Asi pues, el error P del método de aleatorización es un promedio de errores P del método exacto de Fisher aplicado a las diversas tablas que podrian haber sucedido y que son tan extremas ó mas que la obtenida.

Para el caso de un test de dos colas (H'_1), el error habrá de ser suma de dos cantidades. Si convenimos en ordenar la tabla experimental de modo que $t \geq 0$, entonces la (1.66), aplicada con los límites dados por las (1.64) y (1.65), da el error P_1 de una de las colas. Por simetría, el error P_2 de la otra cola se obtendrá tambien mediante la (1.66) cambiando $\sum_A$ por $\sum_B$ y aplicada con los límites:

$$\text{Max}\left\{0;\ M_1-n_1\right\} \leq B \leq \left[\left(\frac{M_1}{n_1} - t\right)\frac{n_1\, n_2}{n}\right]^- \qquad (1.68)$$

$$\left[tn_1\right]^+ \leq M_1 \leq n - \left[tn_2\right]^+ \qquad (1.69)$$

Ahora bien, a toda tabla $(M_1;A)$ de P_1 podemos asociarle otra tabla $(M'_1=n-M_1;B'=n_2-B)$ verificando que $T=-T'$, tabla que es una de las de P_2 pues el campo de variabilidad de B', teniendo en cuenta el de A, coincide con el reseñado en la (1.68); finalmente, como

$$\frac{\binom{n_1}{A}\binom{n_2}{B}}{\binom{n}{M_1}} = \frac{\binom{n_1}{n_1-A}\binom{n_2}{n_2-B}}{\binom{n}{n-M_1}} = \frac{\binom{n_1}{A'}\binom{n_2}{B'}}{\binom{n}{M'}} \qquad (1.70)$$

se desprende que $P_1=P_2$ y, por tanto, el error de dos colas será $2P_A$ (el doble del reseñado en la (1.66)).

1.4.2.4. El nuevo test como uno parametrico (Aportación)

Ahora se contrasta $H_o \equiv p_1 = p_2$ (=p) con p el valor común, desconocido, de ambos. Si al parámetro "perturbador" p se le supone la distribución uniforme,

$$P(M_1)=\int_0^1 P(p)P(M_1|p)dp=\int_0^1 \binom{n}{M_1} p^{M_1} (1-p)^{n-M_1} dp = \frac{1}{n+1} \qquad (1.71)$$

$$P(A,M_1)=P(M_1).P(A|M_1)=\frac{1}{n+1}\cdot\frac{\binom{n_1}{A}\binom{n_2}{B}}{\binom{n}{M_1}} \qquad (1.72)$$

que es la misma expresión que en (1.61).De ahi que el valor P_A de (1.66) permanezca.

Así pues el test de aleatorización propuesto es tambien un test parametrico obtenido pensando que a lo largo de las experiencias de la vida del investigador las diversas hipótesis nulas ciertas se corresponden con una proporción p comun que está distribuida según la uniforme. Las garantias acerca del error son pues a largo plazo, no para una experiencia particular. Alternativamente, el test puede entenderse como uno de tipo seudo-bayesiano en que el parámetro desconocido p se le supone la distribución uniforme.

1.4.3. Comparación de los test propuestos en pequeñas muestras (Aportación)

Con fines comparativos, tomemos la definición de sensibilidad de un test sugerida por Kempthorne (1967) en otro contexto y referida al efecto por Ballatori (1982): "Un test A es mas sensible que otro B, a igualdad de nivel de significación, si, para una tabla dada, la probabilidad de obtener una tabla no menos extrema que ella es menor para A que para B".

En base al criterio, Ballatori comprobó que, para pequeños tamaños de muestra, su test de aleatorización incondicionado era mas sensible que el condicionado de Fisher. Tal comprobación la hizo generando por ordenador todas las tablas posibles para diversas parejas de valores de n_1,n_2 , pero su método adolece de dos defectos. En primer lugar, sus resultados para pequeñas muestras no pueden considerarse como definitivos por cuanto el test para el que hace la comprobación es de una cola y, sin embargo, los tamaños de muestra que maneja son tales que $n_2 \geq n_1$, (es decir, la muestra de mayor tamaño asignada a la proporción mayor). En segundo lugar no hace comprobación alguna acerca de lo que ocurre en grandes muestras, pero eso se verá en el cap. II. Cualquier modo, y dado que que 1.4.2.2. se puso en duda la validez de su test, en lo que sigue no volveremos a aludirlo.

Veamos ahora la comparación propuesta de los dos test alternativos al problema actual: el test condicionado de Fisher y el test incondicionado actual (ambos de aleatorización). De modo general, la expresión (1.67) garantiza que, por ser P_A un promedio de valores de P_F, dado un valor de t, para unos valores de m_1 será $P_A < P_F$ y para otros ocurrirá lo contrario. En la tabla 1.10 se recogen diversas parejas en las que se han determinado todas las tablas (compatibles con $t \geq 0$) que son significativas al 5% de error, bien con el test de aleatorización incondicionado actual, bien con el condicinado de Fisher.

Para cada una de dichas tablas se han obtenido los errores P por ambos métodos $-P_A$ y P_F respectivamente que, por la condición de selección de tablas, serán tales que $P_A \leq 0{,}05$ ó $P_F \leq 0{,}05$-, anotándose en la Tabla 1.10 el número de veces en que $P_A < P_F$ (y el test incondicionado es mas sensible que el condicionado) ó $P_F < P_A$ (y el test condicionado es mas sensible que el incondicionado). A tal efecto ha de tenerse en cuenta que el test planteado es de una cola, y que las P_A y P_F se obtienen mediante las (1.66) y (1.52) respectivamente .

La observación de la Tabla 1.10 permite afirmar que el test propuesto es mas sensible que el de Fisher, siendo su ventaja mas apreciable con pequeñas muestras.

Tabla 1.10

Número de tablas en que el test incondicionado es mas sensible que el condicionado (S_A), ó al revés (S_F), para un test de una cola ($H_1 \equiv$ el tratamiento I da menos resultados + que el II) al error del 5% y para diversas parejas de n_1 y n_2.

n_1	n_2	S_A	S_F	n_1	n_2	S_A	S_F
5	5	6	0	20	5	26	2
	10	12	0		10	60	4
	15	22	0		15	95	5
	20	27	1		20	124	15
	25	36	2		25	157	21
	30	42	2		30	181	30
10	5	12	0	25	5	32	2
	10	28	0		10	76	5
	15	48	0		15	123	8
	20	60	4		20	161	17
	25	75	5		25	177	41
	30	90	10		30	240	35
15	5	18	0	30	5	41	3
	10	48	0		10	90	10
	15	60	8		15	136	21
	20	99	4		20	183	28
	25	123	8		25	237	38
	30	137	20		30	251	76

1.5. CONCLUSIONES (APORTACION)

De lo visto en este capitulo puede concluirse que:

1º) Ante una tabla 2x2 con marginales dados, aplicar el test exacto de Fisher. La versión optima del mismo no ha sido estudiada en la literatura ni en esta memoria. Presumiblemente será la de "mayor a menor probabilidad".

2º) Ante una tabla 2x2 para comparar dos proporciones:

a) Si uno es partidario de los métodos condicionados, aplicar el test exacto de Fisher en su version optima (definición 3.1 dada en 1.1.1.5)

b) Si uno es partidario de los métodos incondicionados, aplicar el test de Boschloo-McDonald visto en 1.3.2 (el test optimo de Barnard es impracticable por problemas de computo).

3º) Ante una tabla 2x2 de un problema de asociación:

a) Si uno es partidario de los métodos condicionados, aplicar el test exacto de Fisher en su versión óptima que, presumiblemente, sera la que proporciona la definición 5 dada en 11.1.5.

b) Si uno es partidario de los metodos incondicionados, no se ha desarrollado un método equivalente al de Boschloo-McDonald, pero todos los autores coinciden en que siempre será conveniente aplicar este al de Fisher.

4º) Ante una tabla 2x2 que surge de un problema de aleatorización,

a) Si uno es partidario de los métodos condicionados, aplicar el test exacto de Fisher

b) Si uno es partidario de los métodos incondicionados, aplicar el test propuesto en el subapartado 1.4.2.3.

5º) En todo caso, si uno es partidario de los test aleatorizados, utilizar el test de Tocher, por ser el UMPU.

CAPITULO II

TEST DE SIGNIFICACION EN TABLAS 2x2: GRANDES MUESTRAS

Todos los test descritos en el cap. I tienen la desventaja de requerir un gran número de calculos (a veces excesivos incluso para un ordenador), lo que puede obviarse si, como se hizo con el test de Fisher, se proporciona una tablas apropiadas. Sin embargo, y dado que las tablas han de ser necesariamente limitadas, a partir de ciertos valores del tamaño de muestra habran de buscarse métodos de aproximación que, basados en distribuciones conocidas, permitan obtener el valor P aproximado de una tabla dada ó determinar la RC oportuna de un modo sencillo. A tales métodos está dedicado el actual capitulo.

Si opiniones contrapuestas hubo con los métodos con pequeñas muestras, mas las hay ahora. En primer lugar existen varios estadisticos apropiados (χ^2, de maxima verosimilitud, etc.); en segundo lugar las correcciones por continuidad (c.p.c.) a realizar en los mismos (especialmente en χ^2) son origen de continuas controversias.

Por uniformidad de formato, en todo lo que sigue, salvo indicación expresa de lo contrario, se asumen test de dos colas y la notación de la Tabla 1.1.

2.1. METODOS DE APROXIMACION

2.1.1. En el caso de marginales dados

En este apartado se ve la aproximación asintotica al método "exacto" de Fisher para tablas con marginales n_i y a_i dados de antemano o que, no siendolos, se les considera como tal por los partidarios de los "métodos condicionados".

Dado que la v.a. hipergeometrica x_1 tiene por media $\mu=a_1n_1/n$ y por varianza $\sigma^2=a_1a_2n_1n_2/n^2(n-1)$, y dado que es conocido que para grandes muestras la hipergeometrica se aproxima a una normal, entonces $(x_1-\mu):\sigma$ seguirá aproximadamente una normal típica y su cuadrado una χ^2 teórica con 1 g.l. Así

$$\chi_u^2 = \frac{(n-1)\,\{x_1y_2-x_2y_1\}^2}{a_1a_2n_1n_2} \qquad (2.1)$$

será el estadistico alternativo al test "exacto" de Fisher para el caso de grandes muestras. Tal estadistico data de antiguo -recogiendolo así, por ejemplo, Pearson (1947)- pero ha sido tradicionalmente reconvertido en la llamada χ^2- no corregida.

$$\chi^2 = \frac{\{x_1y_2-x_2y_1\}^2}{a_1a_2n_1n_2}\cdot n \qquad (2.2)$$

dado que en grandes muestras es $(n-1)\simeq n$. χ_u^2 ha vuelto a considerarlo recientemente Upton (1982) y de ahí el subíndice u.

2.1.2. En el caso de dos proporciones (1 marginal dado)

En este apartado se ve la aproximación asintotica al caso de la comparación de dos proporciones independientes, caso en que el marginal n_i está dado de antemano y que en la literatura es denominado a veces "ensayos comparativos".

2.1.2.1. Métodos z

Dado que las variables $x_i \sim B(n_i,p_i)$, $i=1,2$, se pueden aproximar, en grandes muestras, a una normal, entonces, bajo $H_o \equiv p_1=p_2$ (=p), la v.a. $\hat{p}_1-\hat{p}_2 = x_1/n_1 - x_2/n_2$ es asintoticamente normal con media cero y varianza $p(1-p)(1/n_1 + 1/n_2)$. Un estimador insesgado de misma varianza de $p(1-p)$ es $\hat{p}\hat{q}n/(n-1)$- Brownlee (1967)- con lo cual el estadistico apropiado, a comparar con una normal tipica, es

$$Z_u = \frac{|\hat{p}_1-\hat{p}_2|}{\sqrt{\hat{p}\hat{q}\,\frac{n}{n-1}\,(\frac{1}{n_1}+\frac{1}{n_2})}} \tag{2.3}$$

en donde $Z_u^2 = \chi_u^2$. Usualmente el factor $n/(n-1)$ se iguala a 1 -Brownlee (1967) asi lo aconseja en base a investigaciones empiricas- y así

$$Z_1 = \frac{|\hat{p}_1-\hat{p}_2|}{\sqrt{\hat{p}\;\hat{q}\;(\frac{1}{n_1}+\frac{1}{n_2})}} \tag{2.4}$$

con $Z_1^2 = \chi^2$

Goodman (1964) propone utilizar un estadistico Z_2 que dé resultados compatibles con los intervalos de confianza para p_1-p_2; así, basandose en que x_i se aproxima a una normal de media $\mu_i=np_i$ y varianza $\sigma^2_i=np_i(1-p_i)$, y estimando individualmente con cada muestra el valor de $p_i(1-p_i)$, se obtiene el estadistico

$$Z_2 = \frac{|\hat{p}_1-\hat{p}_2|}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}} \tag{2.5}$$

con $\hat{p}_i=x_i/n_i$ y $\hat{q}_i=1-\hat{p}_i$. Su cuadrado $Z_2^2 = G^2$ seria un competidor del método χ^2 clásico.

Sathe (1982) propone utilizar:

$$Z_3 = \frac{|\hat{p}_1-\hat{p}_2|}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_2}+\frac{\hat{p}_2\hat{q}_2}{n_1}}} \tag{2.6}$$

que tambien es asintoticamente normal.

Aunque no han sido propuestos en la literatura, es obvio que, por iguales razones a las reseñadas para la (2.3), los estadisticos Z_2 y

Z_3 pueden ser transformados en los Z_2' y Z_3' cambiando n_i por n_i-1.

2.1.2.2. Método logit

Si la (1.4) se reparametriza en el formato de la (1.7), un estimador de máxima verosimilitud para $e^{\lambda} = \psi$ - la razón de odds ó razón del producto cruzado- es $\hat{\psi} = x_1 y_2 / x_2 y_1$, el cual para $0 < p_i < 1$ y $n \to \infty$ es asintoticamente insesgado de ψ y de varianza asintotica.

$$\psi^2 \left(\frac{1}{n_1 p_1 q_1} + \frac{1}{n_2 p_2 q_2}\right)$$

Cuando es cierta H_0, utilizando la escala logaritmica y la corrección de Haldane-Anscombe,

$$\hat{\lambda} = \ln \frac{(x_1+0,5)(y_2+0,5)}{(x_2+0,5)(y_1+0,5)}$$

$$\hat{V}(\hat{\lambda}) = \frac{1}{x_1+0,5} + \frac{1}{x_2+0,5} + \frac{1}{y_1+0,5} + \frac{1}{y_2+0,5}$$

con lo que el estadistico propuesto será:

$$L^2 = \frac{\ln \{(x_1+0,5)(y_2+0,5)/(x_2+0,5)(y_1+0,5)}{(x_1+0,5)^{-1} + (x_2+0,5)^{-1} + (y_1+0,5)^{-1} + (y_2+0,5^{-1})} \qquad (2.7)$$

cuyas referencias pueden verse en Gart (1971)

2.1.2.3. Método del arco seno

Basandose en que para una binomial B(n;p) el estadistico arc. sen $\sqrt{\hat{p}}$, con $\hat{p}=x/n$, converge en distribución a una normal de media arc. sen $\sqrt{p}$ y varianza 1/4n -ver por ejemplo Rao (1970)- es inmediato que el estadis tico.

$$Z_o = \frac{|\text{arc.sen}\sqrt{\hat{p}_1} - \text{arc. sen}\sqrt{\hat{p}_2}\,|}{\frac{1}{2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad (2.8)$$

comparando con una normal típica, sirve para realizar el test. Sin embargo Z_o no ha sido utilizado en la práctica a efectos de test (por haberse comprobado es menos eficiente que los tests clasicos) pero si a efectos de la determinación del tamaño de muestra (dado que, por no contener parametros desconocidos, es especialmente idóneo). Por ello no volverá a ser considerado.

2.1.2.4. Método de chi-cuadrado

Dada la función de densidad p-dimensional $f=(x_1 \ldots, x_p; \theta_1, \ldots, \theta_k)$ con $(x_1, \ldots, x_p)$ la v.a. y $(\theta_1, \ldots, \theta_k)$ el vector parametrico, y dada una muestra aleatoria de ella f_i, $i=1,\ldots,n$, si $L=\prod_i f_i$ es la verosimilitud y $\phi_i = \partial \ln L / \partial \theta_i$ las marcas eficientes, entonces $E(\phi_i)=0$ y Cov $(\phi_i, \phi_j) = \alpha_{ij}$ es la matriz de información con inversa α^{ij}. Es conocido -Rao (1970)- que en tal caso si las θ_j son desconocidas y estan sometidas a s restricciones independientes:

$$\sum_i \sum_j \alpha^{ij}(\hat{\theta})\, \phi_i(\hat{\theta})\, \phi_j(\hat{\theta}) \xrightarrow{d} \chi^2(s) \qquad (2.9)$$

con $\hat{\theta}_i$ los estimadores de maxima verosimilitud de las θ_i. Tal expresión, en el caso de multinomiales, se reduce a

$$\sum_i \frac{(n_i - n\hat{\pi}_i)^2}{n\hat{\pi}_i}$$

para una multinomial $M(n; \pi_1, \ldots, \pi_k)$ y con una Σ más para el caso de mas multinomiales.

En nuestro problema contamos con dos binomiales cuyo único parametro, bajo H_o, está estimado por $\hat{p}=a_1/n$ ($\hat{q}=a_2/n$), de modo que sustituyendo en la (2.10) y operando se obtiene la misma expresión (2.2).

2.1.2.5. Método de la razón de verosimilitudes

Obteniendo la verosimilitud de las dos binomiales en cuestion, el maximo de las mismas en el espacio parametrico general y bajo H_o, dividiendo ambos maximos, calculando su logaritmo neperiano y multiplicando por -2 se obtiene el estadistico:

$$Y^2 = 2\left\{ x_1 \ln \frac{x_1\, n}{a_1\, n_1} + x_2 \ln \frac{x_2\, n}{a_1 n_2} + y_1 \ln \frac{y_1\, n}{a_2 n_1} + y_2 \ln \frac{y_2\, n}{a_2\, n_2} \right\} \tag{2.11}$$

a comparar con una χ^2 con 1 g.l.

Williams (1976) sugiere una corrección multiplicativa del mismo en la forma siguiente:

$$Y^2_w = \left\{ 1 + (n^2-a_1a_2)(n^2-n_1n_2)/ 6a_1a_2n_1n_2n \right\}^{-1} . Y^2 \tag{2.12}$$

2.1.2.6. Método de los intervalos de confianza (Aportación)

Mitra (1969) propone obtener intervalos de confianza aproximados para las proporciones desconocidas p_i en base a la aproximación normal, los cuales se obtienen resolviendo

$$p_s - \hat{p}_1 = k \sqrt{p_s(1-p_s)/n_1}$$

$$\hat{p}_2 - p_I = k \sqrt{p_I(1-p_I)/n_2}$$

supuesto que $\hat{p}_1 < \hat{p}_2$, con k en la normal típica. Si $p_I > p_s$ se decide H_1.

Tal procedimiento es el analogo en grandes muestras del propuesto en esta memoria para pequeñas muestras (ver 1.3.5.1). Sin embargo Mitra no establece control alguno para el error del test, pues si k se corresponde con un error del 5% para los intervalos de confianza, el error del test no es del 5%. La dificultad de controlar el error real con grandes muestras hace inviable, en nuestra opinión, la aplicación del método. De todos modos no deja de ser interesante observar que la expresión (1.48) contiene un término muy parecido a la razón de odds y que su logaritmo dé un estadistico muy parecido al (2.7). Como se dijo en 1.3.5.1. un modo conservador de realizar el test seria utilizando un error maximo de trabajo de $1-\sqrt{1-\alpha}$.

2.1.3. En el caso de asociación (ningún marginal dado)

Ahora, y como es tradicional, conviene utilizar la notación usual en tablas rxk : n_{ij} para las observaciones en la casilla (i,j) ; $n_{i.}$ para el total de la fila i, $n_{.j}$ para el total de la columna j; y n.. para el gran total. Ahora los datos responden a una multinomial $M(n..;\pi_{ij})$ y H_o es que $\pi_{ij}=\pi_{i.}\,\pi_{.j}$, con $\pi_{i.}=\sum_j \pi_{ij}$ y $\pi_{.j}=\sum_i \pi_{ij}$

2.1.3.1. Método de chi-cuadrado

Teniendo en cuenta que la verosimilitud es, salvo constante, $L=\prod \pi_{ij}^{n_{ij}}$, bajo H_o sera $L=\prod(\pi_{i.}\,\pi_{.j})^{n_{ij}}$ y el estimador de maxima verosimilitud es $\hat{\pi}_{i.}=n_{i.}/n_{..}$ y $\hat{\pi}_{.j}=n_{.j}/n..$; con ello, y sustituyendo en la (2.10), vuelve a obtenerse el estadistico χ^2 de la (2.2)

2.1.3.2. Método de la maxima verosimilitud

La verosimilitud L anterior alcanza el máximo en $\hat{\pi}_{ij}=n_{ij}/n..$, de modo que $\mathcal{L}=n_{..}^{-n..}\prod n_{ij}^{n_{ij}}$. Por otro lado se vio arriba que L_o alcanza al maximo en las $\hat{\pi}_{i.}$ y $\hat{\pi}_{.j}$ indicadas antes, de modo que $\mathcal{L}_o=\prod(\hat{\pi}_{i.}\,\hat{\pi}_{.j})^{n_{ij}}$. Si $\lambda=\mathcal{L}_o/\mathcal{L}$ y se obtiene $-2\ln\lambda$, obtendremos el estadistico de Wilks:

$$Y^2=2\sum_i\sum_j n_{ij}\ln(n_{ij}n_{..}/n_{i.}n_{.j}) \qquad (2.12)$$

que, siguiendo una χ^2 con 1 g.l., es el mismo visto en la (2.11) para el caso de dos proporciones.

2.1.3.3. Método de Freeman-Tukey

Si $e_{ij} = n_{i.}n_{.j}/n..$, Freeman and Tukey (1950) sugieren el estadistico

$$T^2 = \sum_i \sum_j \left\{ n_{ij}^{1/2} + (n_{ij}+1)^{1/2} - (4e_{ij}+1)^{1/2} \right\}^2 \tag{2.13}$$

que asintoticamente sigue tambien una χ^2 con 1 g.l.

2.1.4. En el caso de los test de aleatorización (Aportación)

En este apartado la notación y conceptos son los vertidos en la sección 1.4.

La segunda parte de la función de probabilidad dada por la (1.61) se corresponde con la de una hipergeométrica, la cual se aproxima en grandes muestras a una $N(\mu = M_1 n_1/n; \sigma^2 = M_1 M_2 n_1 n_2/n^2(n\ 1))$ con ello la segunda sumatoria de la (1.66) será, aproximadamente, y empleando la clásica corrección por continuidad de Yates (ver sección 2.2).

$$\sum_A \frac{\binom{n_1}{A}\binom{n_2}{B}}{\binom{n}{M_1}} = \emptyset\left\{\frac{\left[\left(\frac{M_1}{n_2} - t\right)\frac{n_1 n_2}{n}\right]^{-} - \mu + 0,5}{\sigma}\right\} = \emptyset\left\{\frac{-t}{\sqrt{\frac{M_1 M_2}{n_1 n_2 (n-1)}}}\right\} \tag{2.13}$$

con $\emptyset(x)$ la función de distribunción de la normal típica y la segunda igualdad aproximada obtenida del hecho de que

$$\left[\left(\frac{M_1}{n_2} - t\right)\frac{n_1 n_2}{n}\right]^{-} - \mu + 0,5 = \left(\frac{M_1}{n_2} - t\right)\frac{n_1 n_2}{n} - \mu \tag{2.14}$$

Finalmente, la (1.66) será, para grandes muestras,

$$P_A = \frac{1}{n+1} \sum_{M_1 - [tn_2]^+}^{n - [tn_1]^+} \emptyset\left\{ - \frac{t}{\sqrt{\frac{M_1 M_2}{n_1 n_2 (n-1)}}} \right\} \tag{2.15}$$

Más generalmente, y puesto que las variables A y B siguen distribuciones binomiales $B(n_1;p)$ y $B(n_2;p)$, respectivamente, bajo H_o, distribuciones que se aproximan a distribuciones normales del tipo $N(\mu=n_i p;\ \sigma^2=n_i pq)$, entonces la variable T se distribuye como una $N(\mu=0;\ \sigma^2=p(1-p)n/n_1n_2)$, y así:

$$P_A=P\{T\geq t\}=\int_0^1 P\{T\geq t|p\}dp = 1-\int_0^1 \emptyset\{\frac{t}{\sqrt{p(1-p)n/n_1n_2}}\}dp \qquad (2.16)$$

ó, alternativamente, y teniendo en cuenta la aproximación de Pólya (1945) $\emptyset(x) \simeq \frac{1}{2}[1+\{1-\exp(-2x^2/\pi)\}^{1/2}]$,

$$P_A \simeq \frac{1}{2} - \frac{1}{2}\int_0^1 [1-\exp\{-\frac{2}{\pi}\frac{t^2n_1n_2}{p(1-p)n}\}]^{1/2} \qquad (2.17)$$

Para encontrar una formula explícita para P_A, consideremos que la (2.16) es tambien

$$P_A=\int_0^1\int_A^\infty (2\pi)^{-1/2}\exp(-x^2/2)dxdp \qquad (2.17)$$

con $A=t/\{p(1-p)n/n_1n_2\}^{1/2}$, por lo que permutando los limites de integración, y realizando la primera integral, queda que

$$P_A=\int_{\sqrt{B}}^\infty (2\pi)^{-1/2}\exp((-x^2/2)\ \{1-B/x^2\}^{1/2}dx \qquad (2.18)$$

con $B=4t^2n_1n_2/n$. Ahora bien, si llamamos por $f(x)=\{1-B/x^2\}^{1/2}$, entonces la (2.18) no es otra cosa que $\{1-\emptyset(\sqrt{B})\}\cdot E\{f(x)|x\geq\sqrt{B}\}$, con el segundo miembro la media de la función f(x) de la variable aleatoria x truncada en $x \geq \sqrt{B}$, variable que tiene por media (μ) y varianza (σ^2) -ver pag. 81 de Johnson y Kotz (1970)-:

$$\mu = [\ (2\pi)^{1/2}\exp(B/2)\ \{1-\emptyset(\sqrt{B})\}\ ]^{-1} \tag{2.19}$$

$$\sigma^2 = 1 + \mu\sqrt{B} - \mu^2 \tag{2.20}$$

Pero es conocido que $E\left\{ f(x)\ \middle|\ x \geq \sqrt{B} \right\} \simeq f(\mu) + \sigma^2\ f''(\mu)/2$, con lo que, como $f''(x) = -B(3x^2-2B)/x^3(x^2-B)^{3/2}$, se tendría finalmente que:

$$P_A \approx P'_A = \{1-\emptyset(\sqrt{B})\}\left[\ (1-B/\mu^2)^{1/2} - \frac{B\{1+\mu\sqrt{B}-\mu^2\}\{3\mu^2-2B\}}{2\mu^3(\mu^2-B)^{3/2}}\ \right] \tag{2.21}$$

si bien, en primera aproximación, y haciendo $E\left\{ f(x) \middle| x \geq \sqrt{B} \right\} \simeq f(\mu)$, se tendrá la expresión mas simple:

$$P_A \approx P''_A = \{1 - \emptyset(\sqrt{B})\}\{1-B/\mu^2\}^{1/2} \tag{2.22}$$

Como, por ser $\mu > B$, la fracción que resta en la (2.21) es siempre positiva, el valor P_A'' es siempre mayor que el P'_A, no siendo necesario descender a él si el primero ya dio significativo.

Respecto del test de Ballatori, el valor de significación de una tabla dada es, en grandes muestras,

$$P_B \simeq 1 - \emptyset\ (\sqrt{B}\ ) \tag{2.23}$$

dado que tal test es en realidad uno para $H_o : p_1 = p_2 = ½$, con lo que su valor de Z, según lo dicho al inicio de 2.1.2.1., es

$$Z = \frac{t}{\sqrt{\frac{1}{4}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \sqrt{B}$$

Finalmente, respecto del test de Fisher, ya se ha visto que el estadistico apropiado es el χ^2 dado por la (2.2) ó, equivalente, su raiz cuadrada dada por el Z_1 de la (2.4) en valor absoluto. Con ello, y en el nuevo formato,

$$Z_1^2 = t^2\, n_1 n_2/n\hat{p}\hat{q} \qquad (2.24)$$

$$P_F \simeq 1 - \emptyset\,(Z_1) \qquad (2.25)$$

2.1.5. Comparaciones teóricas de los metodos expuestos

La notación de esta sección es la correspondiente a la del apartado a que se aluda.

2.1.5.1. En el caso de los test Z.

Eberhardt (1977) realiza la comparación teórica de los test Z_1 y Z_2. Si notamos por D_i al término dentro de la raiz cuadrada de Z_i en las (2.4) a la (2.6), entonces el test mas potente será aquel que proporcione un valor mas grande para Z_i, es decir, el de menor valor de D_i. Basandonse en que

$$\begin{cases} \text{Si } f(.) \text{ es concava y } 0 \leq \gamma \leq 1: \\ f(\gamma a + (1-\gamma)\, b) \geq \gamma f\,(a) + (1-\gamma) f(b) \end{cases} \qquad (2.26)$$

es inmediato que para $\gamma = n_1/n = ½$, $a=\hat{p}_1$, $b=\hat{p}_2$ y $f(x)=x(1-x)$, es $D_2 \leq D_1$, de modo que Z_2 es siempre mas potente que Z_1 para iguales tamaños de muestra.

En el caso de tamaños de muestra distintos se prueba que si $n_1/n \to \lambda$, $0 \leq \lambda \leq 1$, $0 < p_i < 1$, Z_2 es mas potente si y solo si

$$(p_1-p_2)\left\{ k(\lambda)p_1 - k(1-\lambda)p_2 - \frac{k(\lambda) - k(1-\lambda)}{2} \right\} > 0 \qquad (2.27)$$

con $K(\lambda) = 1-\lambda-\lambda^2$, lo que conlleva que para mas de la mitad del espacio parametrico (p_1, p_2) sea preferible tambien Z_2.

Por otro lado Sathe (1982) realiza un estudio similar para comparar Z_3 con Z_1 y Z_2. Utilizando la misma función f(.) de entonces para $a=\hat{p}_1$, $b=\hat{p}_2$ y $\gamma=n_1/n$, se tiene que $D_1 \geq D_3$ y asi Z_3 es siempre mas potente que Z_1.

En el caso de que $n_1=n_2$, es claro que $Z_2=Z_3$ y la comparación no tiene sentido. Cuando $n_1 \neq n_2$ y $n_1/n \to \lambda$, Z_3 es mas potente que Z_2 cuando

(1) Si $\lambda < ½$: $p_1 > p_2$, $p_1+p_2 < 1$ ó $p_1 < p_2$, $p_1+p_2 > 1$

(2) Si $\lambda > ½$: $p_1 > p_2$, $p_1+p_2 > 1$ ó $p_1 < p_2$, $p_1+p_2 < 1$

con lo que la mitad de las veces es mejor Z_2 y la mitad Z_3.

2.1.5.2. El caso de los test de aleatorización (Aportación)

En 2.1.4. se dieron dos aproximaciones del valor de P_A(P_A' y P_A'') así como el valor de P_B, ambas para el caso de grandes muestras. La tabla 2.1 presenta los valores de ambas aproximaciones para diversos valores de B en el rango en que ocurren las significaciones usuales. Puede observarse que desde B=2,8 en adelante las diferencias no son excesivas. En la misma tabla se dan los valores de significación P_B del test de Ballatori, pudiendo verse que, en efecto, $P_B > P_A'$ y P_A''.

En realidad esto es algo general para todo valor de B puesto que, por ser siempre $x > \sqrt{B}$ en la (2.18), entonces $P_A < P_B$ siempre; la tabla sirve para probar que tales diferencias pueden ser importantes. Asi pues el test de aleatorización propuesto en esta memoria es siempre mas sensible-en el sentido utilizado en la sección 1.4- que el test de Ballatori.

Tabla 2.1

Valores de P, en %, para la primera (P'_A) y segunda (P''_A) aproximación del test de aleatorizacion actual y del test de Ballatori (P_B) en el caso de grandes muestras y para diversos valores del estadístico B de medida de discrepancia.

B	P'_A	P''_A	P_B	B	P'_A	P''_A	P_B
2,0	4,5	5,1	7,9	3,6	1,4	1,6	2,9
2,2	3,8	4,4	6,9	3,8	1,2	1,4	2,6
2,4	3,3	3,8	6,1	4,0	1,1	1,2	2,3
2,6	2,8	3,3	5,3	4,2	0,9	1,1	2,0
2,8	2,4	2,8	4,7	4,4	0,8	0,9	1,8
3,0	2,1	2,4	4,2	4,6	0,7	0,8	1,6
3,2	1,8	2,1	3,7	4,8	0,6	0,7	1,4
3,4	1,6	1,8	3,3	5,0	0,5	0,6	1,3

Como ejemplo, una tabla con $n_1=20$, $n_2=30$, a=5 y b=15, da unos valores t=1/4, B=3 y $\mu=2,14$. Con ello $P'_A=0,021$ y $P''_A=0,024$, que no son muy diferentes; de hecho, la expresión exacta (1.66) proporciona un $P_A=0,024$, y la aproximación funciona bien incluso para tan bajos tamaños de muestra. Sin embargo el test de Ballatori da un valor P_B=0,0420, en tanto que el de Fisher proporciona un P_F=0,0696, ambos menos significativos que el test actual. Finalmente, el test clasico de comparación de dos proporciones de una t_{exp}=1,768 sin corrección por continuidad, y ${}_c t_{exp}$=1,473 con ella, lo que proporciona unos valores de P de 0,039 y 0,070 respectivamente, ambos superiores tambien al valor de P_A.

Ballatori en su articulo no hace comparaciones de su test con el de Fisher en el caso de grandes muestras. Como $1/4 \geq \hat{p}\hat{q}$ siempre, es claro que $B \leq Z^2_1$ -ver (2.24)- y así $P_F \leq P_B$ siempre. El test de Fisher es siempre mas sensible que el test de Ballatori en grandes muestras.

2.2. CORRECCION POR CONTINUIDAD

Cuando se aproxima una variable aleatoria de tipo discreto por una de tipo continuo la aproximación mejora si se hace un artificio denominado corrección por continuidad. A tal artificio está dedicada en exclusiva esta sección.

2.2.1. Generalidades de la c.p.c.

Plackett (1964) hace notar que si $\mathfrak{X}$ es el número de exitos en la B(n;p), entonces

$$P(\mathfrak{X} \leq r) \simeq \Phi\left(\frac{r-np}{\sqrt{npq}}\right) + \frac{1}{2\sqrt{npq}}\, \phi\left(\frac{r-np}{\sqrt{npq}}\right) \qquad (2.28)$$

donde $\phi(\cdot)$ denota la función de densidad de la normal típica y $\Phi(.)$ la de distribución de la misma, expresión que historicamente ha sido expresada como

$$P(\mathfrak{X} \leq r) = \Phi\left(\frac{r+\frac{1}{2}-np}{\sqrt{npq}}\right) \qquad (2.29)$$

version usada comunmente por los estadisticos desde (al menos) 1921 (Pearson (1947)).

La adición de 1/2 al valor de r en el argumento de $\Phi(.)$ fue descrita por Yates (1934) como una corrección por continuidad, nomenclatura sugerida por el hecho de que se usa una función continua para aproximar a una función escalonada.

El mismo criterio se ha seguido historicamente para toda situación similar, con la única variación de que la c.p.c. habrá de ser la mitad del salto de la variable.

Usualmente la correccion por continuidad se introduce con argumentos prácticos y gráficos; es tradicional la justificación de Feller (1973) que se ilustra en la Figura 2.1 y que viene a considerar la

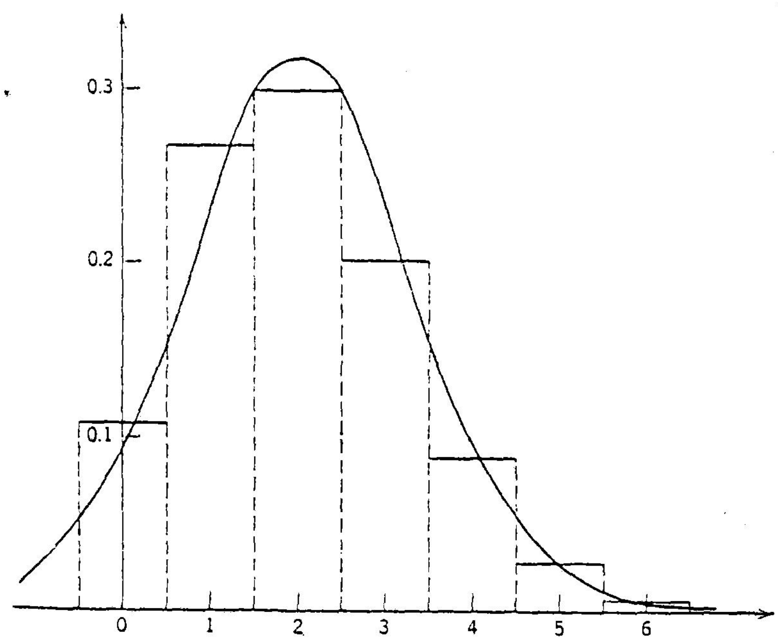

Figura 2.1

variable discreta como un redondeo de la continua a la que se aproxima.

El único argumento analítico paralelo al gráfico anterior es el de Cox (1970), que prueba que la mejor aproximación de una variable aleatoria discreta D por una variable continua C se obtiene poniendo

$$P(D=d+kh) \simeq \int_{d+(k-c_k)h}^{d+(k+c_k)h} f(c)dc \qquad (2.30)$$

en donde D toma valores d, d+h, d+2h,.... y $c_k \simeq ½$ en el caso de la normal y $c_k=½$ en cualquier otro caso si se desea hacer el promedio de error igual a 0.

2.2.2. Pruebas teóricas de la c.p.c. mas conveniente en tablas 2x2

2.2.2.1. Resultados conocidos

Dado que la distribución hipergeometrica tambien se aproxima en grandes muestras a una normal, Plackett (1964) justifica la historica introducción de la misma c.p.c. para tal variable. Como la hipergeometrica es la distribución base del test exacto de Fisher, Yates (1934) introdujo en dicho test la corrección y obtuvo que el estadistico χ^2, referido en la (2.2), habrá de ponerse en esta forma corregida por continuidad:

$$\chi^2_Y = \frac{\left[|x_1y_2-x_2y_1| - ½n\right]^2}{a_1a_2n_1n_2} \cdot n \qquad (2.31)$$

cantidad que se obtiene corrigiendo por continuidad la v.a. x_1 en la expresión $(x_1-\mu):\sigma$ citada al comienzo de 2.1.1..

El estadistico χ^2_Y puesto en formato Z se convierte en

$$Z_y = \frac{|\hat{p}_1-\hat{p}_2| - ½(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{\hat{p}.\hat{q}\,(\frac{1}{n_1} + \frac{1}{n_2})}} \qquad (2.32)$$

propuesto, entre otros, por Brownlee (1967) y que viene a ser el estadistico Z_1 de la (2.4) con c.p.c.

Dado que el test de Fisher es aplicado, por los partidarios del método condicionado, a cualquier tipo de tabla 2x2, el estadistico χ^2_Y

ha sido aplicado historicamente no solo al caso de los marginales dados sino tambien a los casos de comparación de dos proporciones y de asociación.

Plackett (1964) obtiene la diferencia entre P($\chi \leq x$) y $\Phi(x)$ en el caso de comparación de dos proporciones, probando que la misma puede ser positiva o negativa. Asimismo obtiene que P($\chi_y^2 \leq x)-\Phi(x)$ es siempre negativo (en ocasiones apreciablemente). De todo ello concluye que para comparar dos proporciones la c.p.c. del 0,5 da demasiado pocas significaciones y, consiguientemente, la misma no debe hacerse.

Pirie y Handam (1972) aplican el criterio de Cox (1970) al caso de la comparación de dos proporciones independientes y de asociación en tablas 2x2. De un modo general si P_D es una distribución discreta que es función de θ y hay que contrastar $H_o \equiv \theta = 0$ vs $H_1 \equiv \theta \neq 0$, sea U un estadístico insesgado y suficiente de θ, o de un múltiplo de él, con valores saltando de h en h unidades. Si P_D se aproxima a una normal:

$$P_D(-r < U < r) \simeq P_N\left\{-r + \frac{h}{2} \leq U \leq r - \frac{h}{2}\right\} \qquad (2.33)$$

con P_N la probabilidad bajo la normal. Si ademas $\hat{\sigma}^2$ es un estimador consistente de la varianza de U entonces

$$P_D\ (-r < U < r) \simeq P_N\left\{\frac{|U|}{\hat{\sigma}} \leq \frac{r - \frac{h}{2}}{\hat{\sigma}}\right\} \qquad (2.34)$$

con $U/\hat{\sigma}$ asintoticamente N(0,1). Los mismos autores señalan que la dificultad de la aproximación radica en que puede no existir el estimador consistente y que el resultado solo es válido en el límite; a estas razones le achacan las continuas discusiones al respecto de la c.p.c. y la tendencia a usar métodos condicionados.

En el caso de una tabla 2x2 de asociación, y utilizando la notación habitual en ellas, es $H_o \equiv p_{11}p_{22}=p_{12}p_{21}$ y asi $\theta = p_{11}p_{22}-p_{12}p_{21}$ da lugar a un estimador insesgado proporcional a $\mathcal{U}=n_{11}n_{22}-n_{12}n_{21}$ con una varianza estimada por maxima verosimilitud de $\hat{\sigma}^2 = (n..-1)n_{1.}n_{.1}n_{2.}n_{.2}/n_{..}^2$. Como el valor de $\mathcal{U}$ salta de 1 en 1 en la mayoria de las situaciones, la c.p.c. apropiada para el mismo es ½ y, haciendo $n_{..}-1 \simeq n..$, se obtiene (volviendo a la notación de la tabla 1.1):

$$ {}_o\chi^2_{PH} = \frac{\left\{|x_1y_2-x_2y_1| - \frac{1}{2}\right\}^2}{a_1a_2\,n_1n_2} \cdot n \tag{2.35} $$

con el subindice cero aludiendo a que hay cero marginales fijados. Como la nueva c.p.c. difiere de la de Yates en que el termino (½)n que resta en esta resulta cambiando por (½), es claro que el método de Yates es mas conservador que el de PH y este mas que el incorregido: $\chi^2_y < {}_o\chi^2_{PH} < \chi^2$.

En el caso de dos proporciones es $H_o \equiv p_1=p_2$, $\theta = p_1-p_2$ y $\mathcal{U}=\hat{p}_1-\hat{p}_2$. Para lograr que los saltos de $\mathcal{U}$ sean constantes, PH abordaron el problema bajo el supuesto de que $n_{1.}=n_{2.}=n/2$, con lo que $\mathcal{U}=x_1-x_2$, $\hat{\sigma}^2=a_1a_2/(n-1)$ y finalmente

$$ {}_1\chi^2_{PH} = \frac{\left\{|x_1y_2-x_2y_1| - \frac{1}{4}n\right\}^2}{a_1a_2n_1n_2}\, n \tag{2.36} $$

si se hace $(n-1) \simeq n$ y dado que en tal caso $\mathcal{U}$ salta de n/2 en n/2. El estadistico ${}_1\chi^2_{PH}$ (con el subindice 1 aludiendo a que hay 1 marginal dado) puesto en formato z se convierte en

$$ z_{PH} = \frac{|\hat{p}_1-\hat{p}_2| - \frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{2.37} $$

Ambos estadisticos son validos solo para el caso de que $n_1=n_2$, pero algunos autores (Liddell (1978)) proponen utilizarlos en cualquier situación. De nuevo sucede que $\chi^2_y < {}_1\chi^2_{PH} < \chi^2$. Como resumen, la c.p.c. en una tabla 2x2 no debe ser única, utilizandose las correcciones (-n/2), (-n/4) y (-½) para los casos, respectivamente, de dos, uno o cero marginales fijos.

Finalmente, Schouten et al (1980) afirman que en el caso de dos proporciones es inmediato que el salto en el estadistico $x_1y_2-x_2y_1$ que figura en el numerador de la χ^2 clasica no corregida es un número h tal que:

$$1 \leq MCD(n_1,n_2) \leq h \leq Min(n_1,n_2) \leq n/4 \tag{2.38}$$

(con M.C.D. el maximo comun divisor) proponiendo la c.p.c., conservadora, de $Min(n_1,n_2)/2$. Con ello

$$\chi^2_S = \frac{\left\{ |x_1y_2-x_2y_1| - \frac{1}{2} Min(n_1,n_2) \right\}^2}{a_1a_2n_1n_2} n \tag{2.39}$$

(si bien su oferta es con (n-1) en lugar de n) ó, puesto en formato Z :

$$Z_s = \frac{|\hat{p}_1-\hat{p}_2| - \frac{1}{2} Min(n_1,n_2)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \tag{2.40}$$

siendo evidente que cuando $n_1=n_2$ seran $\chi^2_s={}_1\chi^2_{PH}$ y $Z_s=Z_{PH}$

2.2.2.2. Aportación

Upton (1982) cita una comunicación privada de Cook (1981) en la que le sugiere el uso de la c.p.c. de $MCD(n_1,n_2)/2$ en el caso de comparación de dos proporciones. Por otro lado la expresión (2.38) que Schouten et al. proponen sin demostración es, en nuestra opinión, falsa. Para ver esto consideremos que $U=x_1y_2-x_2y_1$ y U_o es el valor experimental de una tabla dada (con $\hat{p}_1 > \hat{p}_2$). Los siguientes puntos menos extremos seran los de su frontera, puntos obtenidos restando 1 a x_1, sumando 1 a x_2 ó sumando 1 a x_1 y x_2; cada uno de ellos da un valor de $\mathcal{U}$ tal que $\mathcal{U} - \mathcal{U}_o$ es igual a $-n_2$, $-n_1$ y n_2-n_1 respectivamente, de donde la c.p.c. habra de ser $\frac{1}{2}\text{Min}\{n_1; n_2; n_1-n_2\}$, considerando el termino n_1-n_2 solo si la proporción muestral mas grande se corresponde con el tamaño muestral mas grande. Esto es parcialmente compatible con la afirmación de Schouten.

Sin embargo no hay porqué limitarse a la frontera del punto (x_1, x_2) pues, como ya se vio en los tests incondicionados, los puntos más o menos extremos a uno dado pueden caer en cualquier posición del espacio muestral. Bajo tal criterio, el siguiente punto menos extremo al (x_1, x_2) dado -siempre con $\hat{p}_1$ supuesto mayor que $\hat{p}_2$- sera aquel (x_1+r, x_2+S) que de un valor $\mathcal{U} < \mathcal{U}_o$ y lo mas cercano posible a $\mathcal{U}_o$, con r y s dos número enteros por determinar. Como $\mathcal{U} - \mathcal{U}_o = (x_1+r)(y_2-s)-(x_2+S)(y_1-r) - x_1y_2 + x_2y_1 = rn_2 - sn_1$, se trata de encontrar la solución (r,s) de la ecuación diofantica $rn_2-sn_1=y$ que da lugar a un $y<0$ y lo mas pequeño posible; es decir, la c.p.c. habrá de ser $\frac{1}{2}$ Min $\{rn_2-sn_1 < 0 \mid r, s \in \mathbb{Z}\}$. Como para que tal ecuación diofantica tenga solución hace falta que su termino independiente "y" sea un multiplo del $\text{MCD}(n_1, n_2)$, es evidente que el mas pequeño "y" que verifica la condición es $y = -\text{MCD}(n_1,n_2)$ y asi siempre existira una pareja (r,s) que haga $rn_2-sn_1 = -\text{MCD}(n_1n_2)$ y otra $(-r,-s)$ que hace $-rn_2 + sn_1 = \text{MCD}(n_1,n_2)$. Con ello el estadistico $\mathcal{U}$ salta de h en h, con $h = \text{MCD}(n_1,n_2)$ y la c.p.c. habrá de ser de $(\frac{1}{2})h$, es decir:

$$\chi_c^2 = \frac{\left\{|x_1y_2-x_2y_1| - \frac{1}{2}\,\text{MCD}(n_1,n_2)\right\}^2}{a_1a_2n_1n_2}\, n \tag{2.41}$$

$$Z_c = \frac{|\hat{p}_1-\hat{p}_2| - \frac{1}{2}\,MCD(n_1,n_2)}{\sqrt{\hat{p}\hat{q}\ \left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \tag{2.42}$$

Notese que todos los criterios teóricos de obtención de una c.p.c. están basados en el salto del estadístico $\mathcal{U} = x_1y_2-x_2y_1$. En un test de dos colas ello implica asumir que los puntos van entrando en la RC de mayor a menor valor de $|\mathcal{U}| = |x_1y_2-x_2y_1| = |\hat{p}_1-\hat{p}_2|\ n_1n_2$, es decir, de mayor a menor valor de $|\hat{p}_1-\hat{p}_2|$, y ya vimos que tal criterio de ordenación no era precisamente el óptimo (ver 1.3.5.2.). Pareceria mas conveniente obtener el valor de χ^2 en el punto (x_1,x_2) observado y su inmediato inferior χ^2_- de cualquier otro punto: con ello

$$\chi^2_{KS} = \frac{\chi^2 + \chi^2_-}{2} \tag{2.43}$$

-propuesto por Kendall and Stuart (1967)- seria la mejor aproximación al test incondicionado de Suissa (ver 1.3.4.), no al optimo de los incondicionados. Tal método presenta la desventaja de que habra de recurrirse al ordenador para determinar el punto $(x'_1,\ x'_2)$ que produce una χ^2_- menor que χ^2 y lo mas cercana a ella, y es claro que los métodos aproximados pretender obviar los calculos complejos y obtener los resultados "a mano ". Como por otro lado una prueba teorica, similar a la ofrecida arriba, acerca de cual puede ser χ^2_- es imposible (ó al menos no se ha dado, por lo compleja), los estadisticos se han visto abocados a adoptar el criterio de ordenación en base a $|\hat{p}_1-\hat{p}_2|$ -aunque explicitamente nadie lo reconozca- y a perder asi algo de potencia. Muy probablemente, la mitad de las discusiones habidas acerca de la c.p.c. idónea se deban a ello (ver 2.2.3).

2.2.3. Pruebas practica de la c.p.c. mas conveniente. Evolución histórica del problema.

Pearson (1947) es el primero en plantearse si el estadistico χ^2_Y es el adecuado para las tablas 2x2 con no todos los marginales fijos. Para el caso de dos proporciones, y a través de dos tablas con (n_1=18, n_2=12) y n_1=n_2=10, obtiene las RC , al 5% y 1%, con los criterios χ^2 y χ^2_Y y el tamaño de cada uno de los tests en una amplia gama de valores de p (bajo H_o). La comparación de los tamaños obtenidos (α*) con el error objetivo α permite concluir:

a) En el caso de χ^2, α * es casi siempre menor que α , pero en ocasiones lo desborda.

b) En el caso de χ^2_Y, α* es siempre menor que α y, muy frecuentemente, bastante menor.

De ahi concluye que, en el caso de dos proporciones, debe utilizarse el estadistico χ^2 (pues a lo largo de la vida del experimentador los valores $\alpha^* > \alpha$ se compensan con los $\alpha^* < \alpha$) y no el χ^2_Y -indicado para tablas 2x2 con marginales dados- pues resulta demasiado conservador. El problema es especialmente acusado en tablas asimetricas ($n_1 \neq n_2$) y en cualquier tabla del caso de asociación.

Grizzle (1967) simula los casos de dos proporciones y asociación obteniendo el porcentaje de veces que se decide H_1, siendo cierta H_o, con los test χ^2 y χ^2_Y, concluyendo que

a) El test χ^2_Y es demasiado conservador

b) El test χ^2 es preferible, si bien es algo conservador en el caso de muestras pequeñas

lo que es concordante con lo afirmado por Pearson

Mantel and Greenhouse (1968) mantienen que en efecto la χ^2 sigue mejor una χ^2 teorica con 1 g.l., pero eso es distinto a que se haga la c.p.c. por el paso de una discreta a una continua, abogando por el uso de la χ^2_Y por tratarse de la mejor aproximación a la hipergeometrica que es la base del test condicionado de Fisher que, a su vez, es segun ello el adecuado para toda tabla 2x2 en base al principio de "ancillarity". Gart (1971) apoya el argumento.

Rao (1970) cita una posible corrección, debida a Dandekar (sin ningun tipo de justificación teorica o práctica), que consistiria en obtenida la χ^2 para una tabla, χ^2_0, obtener las χ^2 para la tabla en la que la casilla menor se haya hecho inferior en una unidad (χ^2_{-1}) y para la tabla en la que la casilla menor se haya hecho superior en una unidad (χ^2_1) y utilizar la χ^2 corregida como aquella

$$\chi^2_D = \chi^2_0 - \frac{\chi^2_0 - \chi^2_{-1}}{\chi^2_1 - \chi^2_{-1}} (\chi^2_1 - \chi^2_0)$$

tal corrección no aparece referenciada en ningun otro sitio,

Garside (1971, 1972), para el caso de proporciones, indica que si los puntos han de entrar en la RC según su valor de χ^2, entonces para cada a_1 puede obtenerse el ultimo valor de x_2 que entraria en ella despejando $x_2(a_1)$ de la igualdad $\chi^2 = \chi^2_{1-\alpha}$:

$$x_2(a_1) = \left\{ n_2a_1 - \sqrt{n_1n_2a_1a_2\chi^2_{1-\alpha}/n} \right\}/n \qquad (2.44)$$

Si $[x_2(a_1)]$ es la parte entera de $x_2(a_1)$ y $\lambda(a_1) = x_2(a_1) - [x_2(a_1)]$, el autor propone entrar los puntos en la RC de menor a mayor valor de $\lambda(a_1)$, aplicando el criterio del maximo de Barnard para obtener el tamaño final de la RC. Al último valor $\lambda(a_1)$ le denomina λ_G y asi el nuevo estadístico será

$$\chi^2_G = \frac{\left\{ |x_1y_2 - x_2y_1| - \lambda_G n \right\}^2}{a_1a_2n_1n_2} \qquad (2.45)$$

proporcionando tablas de λ_G para diversos valores de n_i y α . Siempre sucede que $\chi^2_Y < \chi^2_G < \chi^2$, y, por simulación, comprueba que la χ^2_G nunca desborda al α objetivo y suele dar valores mas cercanos a él que las otras aproximaciones.

Conover (1974) recoge la sugerencia de Kendall and Stuart (1967) de utilizar el estadistico (2.43) en toda tabla 2x2, comprobando que para el caso de marginales fijos $\chi^2_{ks} \simeq \chi^2_Y$ en el caso de $n_1=n_2$ ó $a_1=a_2$, pero que en otro caso χ^2_{ks} se aproxima mas a los valores de la hipergeometrica. Para el caso de dos proporciones reconoce que χ^2_{ks} es dicicil de obtener, y por tanto impracticable, si bien es aproximadamente igual al χ^2 (aunque no ofrece pruebas de ello). Para el caso de asociación no se decide y solo afirma que χ^2 y χ^2_Y son cosas muy distintas.

Starmer et al (1974) contestan afirmando que debe utilizarse el estadistico χ^2 por ser el que mas se aproxima a la solución UMPU de Tocher. Mantel (1974) contesta a Conover indicando que el test bajo criterio χ^2_Y se aplica mal, pues para dos colas el error no es el nominal de la tabla χ^2 sino

$$P_M = \frac{P(\chi^2_Y) + P(\chi'^2_Y)}{2} \tag{2.46}$$

con $P(\chi^2_Y)$ y $P(\chi'^2_Y)$ los errores P nominales de la tabla experimental y la tabla que proporciona, por la otra cola, un valor $\chi'^2_Y \geq \chi^2_Y$ y lo mas proximo posible a él. Con tal criterio el test obtenido prueba que es aproximadamente igual al de Fisher (que piensa es el apropiado para toda tabla 2x2).

Garside and Mack (1976) indican que ninguno de los tests propuestos, salvo el de Tocher, logra alcanzar el error objetivo α, pareciendole que el mejor método será aquel que mas se acerque a α (aunque, señalan, algunos autores dicen que sin sobrepasarlo). A continuación comparan por simulacion los métodos de χ^2, Fisher, χ^2_Y, χ^2_G y Boschloo (1970) para el caso de dos proporciones, concluyendo:

a) Boschloo y χ^2_G son aproximadamente iguales en cuanto al tamaño de ambos tests y los que mas se aproximan al error objetivo α.

b) Fisher y χ^2_Y son aproximadamente iguales, pero demasiado conservadores, aproximandose a α con tablas de 500 datos menos que los dos anteriores con tablas de 40 datos.

c) El estadístico χ^2 desborda con frecuencia el error objetivo α, especialmente con tamaños muestrales distintos ó α pequeños, llegando incluso a dar tamaños de hasta 6α.

d) Los métodos de Boschloo y χ^2_G controlan las fuentes de error de continuidad y desviación de la normalidad, aproximandose mas a α que los otros.

Notese que esta es la primera ocasión en que no solo se intenta corregir el error que se produce al aproximar una discreta a una continua, sino tambien el error producido por el hecho de que la variable es solo aproximadamente normal, no "normal" de pleno derecho.

Haber (1980) se muestra partidario del metodo condicionado y compara varios métodos de aproximación al test de Fisher. En primer lugar, y puesto que es consciente de que el test "exacto" de Fisher puede realizarse bajo diversos puntos de vista, acoge el de Radlow and Alf (1975) explicitado en otro contexto: para una tabla dada, calcular la probabilidad (hipergeometrica) de todas las tablas tan extrañas ó mas (en χ^2) que la obtenida. A la probabilidad de todas las tablas que, para los marginales obtenidos, tienen un valor de χ^2 tan grande ó mas que el dado, se le anota como el valor P(Fisher) para esa tabla. Tal valor exacto es confrontado con los métodos aproximados clásicos de χ^2 y χ^2_Y y con otros tres métodos muy relacionados entre si, que tienen como base la tabla obtenida y la primera menos extrema que ella (en términos de χ^2) por el otro lado: i) el método $\chi^2_{ks} = (\chi^2 + \chi^2_-)/2$ citado en (2.43); ii) el método del P_M de Mantel citado en la (2.46); y iii) un nuevo método Z_H propuesto por él mismo, con

$$Z_H = \frac{\chi_+ \chi_-}{2} \qquad (2.47)$$

a enfrentar a una normal tipica y tal que, obviamente, $Z^2_H = \chi^2_H < \chi^2_{ks}$. Tras generar diversas tablas y obtener los valores de P por todos los metodos, concluye -basandose fundamentalmente en el cociente P(metodo ?)/P(Fisher)- que:

1) χ^2 debe descartarse por ser demasiado liberal (da demasiadas significaciones), llegando a ocurrir que $P(\chi^2) < P(\text{Fisher})/20$
2) χ^2_Y debe descartarse por ser demasiado conservador (da pocas significaciones), llegando a ocurrir que $P(\chi^2_Y) > 4P(\text{Fisher})$.
3) Los métodos χ^2_{ks}, Z_H y P_M son prácticamente iguales.

Schouten (1980) comparan los métodos χ^2_Y, χ^2_u y χ^2_s, en el caso de dos proporciones, obteniendo:

1) χ^2_Y es demasiado conservador
2) χ^2_u da demasiadas significaciones
3) χ^2_s se comporta mas normalmente, si bien sobreestima el valor exacto de P en el caso de que la muestra mas pequeña sea inferior a tres veces la muestra mas grande ($1/3 \leq n_1/n_2 \leq 3$)

Burstein (1981) hace notar que, en el caso de dos proporciones, los métodos χ^2 y χ^2_Y no son satisfactorios, siendo precisa una c.p.c. intermedia entre 0.n y ½ n. Tras comparar los valores de P obtenidos en diversas tablas, con los métodos χ^2_Y, χ^2, $\chi^2_{0,1}$ y $\chi^2_{0.25}$ (aludiendo χ^2_h a que la c.p.c. es de h.n.) concluye:

a) Si $n_1=n_2$ el estadistico apropiado es $\chi^2_{0.25}$

b) Si $n_1 \neq n_2$ (incluso si $|n_1-n_2|=1$) el estadistico apropiado es $\chi^2_{0,10}$)

Notese que en el caso de $n_1=n_2$ Burstein obtiene, por via numerica, la misma c.p.c. obtenida, por via teórica, por Pirie and Hamdan (1972) y dada por la (2.36). El nuevo estadístico seria pues:

$$\left\{ \begin{array}{l} \chi^2_B = \dfrac{\left\{ |x_1y_2-x_2y_1| - h.n \right\}^2}{a_1a_2n_1n_2} n \\ \begin{cases} h = 1/4 \text{ Si } & n_1=n_2 \\ h = 0,10 \text{ Si } & n_1 \neq n_2 \end{cases} \end{array} \right. \tag{2.48}$$

Upton (1982) realiza una revisión de 22 tests para comparar dos proporciones entre los cuales se encuentran los métodos χ^2 objeto de esta sección. Para todos ellos obtiene la RC (al 5%)en una tabla 36x16 y anota el número de puntos de la misma, obteniendo que el test exacto de Fisher es el que da la RC mas pequeña y que los números, M, de puntos de los distintos métodos de χ^2 se ordenan así:

$$M(\chi^2_Y) \leq M({}_1\chi^2_{PH}) \leq M(\chi^2_G) \leq M(\chi^2_u) \leq M({}_o\chi^2_{PH}) \leq M(\chi^2)$$

A continuación hace un estudio de las diferencias $A = \alpha^* - \alpha$ entre el tamaño de cada test y el error α de tipo I objetivo, a lo que llama <u>precision</u> del test. Es obvio que A depende de n_1, n_2 y p, de modo que para n_1, n_2 y α dados ocurre necesariamente que $-\alpha \leq A(p) \leq 1-\alpha$. Como H_o se acepta siempre cuando $x_1 = x_2 = 0$ (ó $y_1 = y_2 = 0$), y eso ocurre seguro cuando p=0 (ó 1), entonces $A(0)=A(1)=-\alpha$ en todos los tests y asi todos ellos son conservadores en ese caso. Parece claro que el test optimo sea aquel que hace A(p)=0 en $\delta \leq p \leq 1-\delta$ y con δ tan pequeño como se pueda. Como tal test no existe, Upton da unos ejemplos graficos de como evoluciona A(p) en los distintos test en una tabla de 48x36, obteniendo que:

1) χ^2 da valores de A(p) excesivamente grandes.

2) χ^2 y G^2 (ver 2.1.2.1) dan valores A(p) en el entorno de 0 cuando $0,2 < p < 0,8$.

Como es tentativo basar las conclusiones en una única tabla, el criterio de Cochran (1952) de que un test actuará razonablemente bien si $|A(p)| \leq 0,01$ para el caso de $\alpha = 0,05$, si bien otros autores sostienen que debe exigirse que $A(p) \leq 0$, $\forall_p$ prefiriendo asi que A(p) sea negativo a que sea positivo. Ante lo dificil de buscar una regla optima facil de evaluar en un gran número de tablas, adopta el criterio de considerar que P está distribuido como $\pi(p)=1$ ó $\pi(p)=p(1-p)$ para asi poder obtener los errores α^* promedios y realizar comparaciones. Tras indicar que rigidamente las comparaciones entre tests no son posibles a menos que tengan igual valor de α^*, ó que sus RC esten unas incluidas en otras, establece, a pesar de todo, las siguientes conclusiones basadas en las $\pi(p)$ anteriores:

1) Fisher y χ^2_Y son prácticamente idénticos y conservadores (α^* es como maximo de un 4% cuando α=5%)

2) χ^2_G es siempre conservador

3) ${}_1\chi^2_{PH}$ es generalmente conservador

4) G^2 es muy liberal en tablas con n_i muy distintos ó pequeños (puede llegar a dar $\alpha^* = 0,22$ cuando α =5%)

5) χ^2 da valores $E(\alpha^*) \simeq \alpha$

6) χ^2_c es demasiado conservador

7) χ^2, ${}_0\chi^2_{PH}$ y χ^2_u están bastante de acuerdo entre si

8) χ^2_s ocupa una posición intermedia entre ${}_1\chi^2_{PH}$ y χ^2_G

Finalmente recomienda usar χ^2_Y en el caso de marginales fijos, pero no en el caso de dos proporciones. En este último indica que el test a elegir debiera ser el mas preciso, poderoso y simple, con tal de que $\alpha^* > \alpha$ ocurra muy improbablemente; como todas las condiciones son dificiles de verificar simultaneamente, opta por el test χ^2_u si nos decidimos por no considerar la última.

Sachs (1986) tambien concluye el metodo χ^2_u es el que da una probabilidad de error promedio que esta mas cercana al nivel de significación α .

Tabla 2.2

Valores de P para las tablas y tests indicados.

x_1	y_1	n_1
x_2	y_2	n_2

(a)

12	108	120
24	96	120

(b)

10	90	100
20	80	100

(c)

10	91	101
20	80	100

	(a)	(b)	(c)
$\lvert\hat{p}_1-\hat{p}_2\rvert$	10%	10%	10,1%
TESTS DE DUPONT:			
1) Test exacto de Fisher			
a) P de dos colas (tablas mas improbables)	0,0457	0,0734	0,0497
b) Doble del P de una cola	0,0457	0,0734	0,0689
2) Tests del tipo χ^2:			
a) χ^2	0,0301	0,0477	0,0445
b) χ^2_Y	0,0468	0,0747	0,0701
c) χ^2_H	0,0468	0,0747	0,0478
d) χ^2_{KS}	0,0459	0,0729	0,0477
e) P_M	0,0468	0,0747	0,0509
TESTS ALTERNATIVOS:			
1) Tests incondicionados: McDonald	—	0,0596	0,0502
2) Tests del tipo χ^2:			
a) χ^2_B de Burstein	—	0,0599	0,0489
b) χ^2_C de Cook-Memoria	—	0,0599	0,0466

Dupont (1986) presenta tres tablas 2x2 que son las indicadas en la Tabla 2.2. Sobre ellas argumenta que si bien todos los metodos incluidos en la misma estan bastantes de acuerdo en el valor P cuando $n_1=n_2$ (salvo el χ^2 que es sustancialmente menor), se produce el hecho sorprendente de que una tabla como la 2.2(c), cuya diferencia de proporciones es de 10,1%, tenga valores de P sustancialmente distintos de los de la tabla 2.2(b), cuya diferencia de proporciones es de un 10%, con la mayoria de los métodos. De ella, y de un conjunto de tablas "similares" como las (b) y (c), concluye:

1º) χ^2_Y es el estadistico adecuado por ser coherente ante tablas "similares"

2º) El criterio de doblar el P del test exacto de Fisher de una cola es preferible al criterio de realizar tal test como de dos colas y obtener la probabilidad hipergeometrica de una tabla tan improbable ó mas que la obtenida.

2.2.4. Condiciones de validez de los test χ^2 con c.p.c.

Las condiciones de validez de los tests aproximados no han sido objeto de estudios sistematicos, teniendose algunos resultados solo en el caso de los test χ^2 con c.p.c.

Con la notación propia de las tablas de asociación, es conocido que a las cantidades n_{ij} se les suele llamar <u>cantidades observadas</u>, en tanto que a las cantidades $E_{ij} = E(n_{ij}) = n_{i.}n_{.j}/n..$ se les suele llamar <u>cantidades esperadas</u>.

Fisher (1941) aconseja utilizar tests exactos cuando $n_1+n_2 < 40$ y alguna E_{ij} es menor que 5.

Brownlee (1967) aconseja utilizar los tests χ^2 siempre que todas las E_{ij} sean superiores a 3,5.

Grizzle (1967) sugiere utilizar los métodos χ^2 solo si las E_{ij}

son todas las mayores que 5.

Haber (1980) indican que los criterios P_M, χ^2_{ks} y χ^2_H son apropiados si todas la E_{ij} son superiores a n/10.

2.3. COMPARACIONES PRACTICAS DE LOS TEST APROXIMADOS: EVOLUCION HISTORICA

Pearson (1947) señala que no hay diferencias apreciables entre χ^2 y χ^2_u.

Grizzle (1967) indica que el test Y^2 es demasiado liberal cuando los tamaños muestrales son pequeños, comportandose casi igual que el test χ^2 en otro caso.

Eberhardt and Fligner (1977) comparan los tests Z_1 y Z_2 observando que aunque ambos tienen un tamaño α^*_i que converge a α cuando $n_i \to \infty$, en la práctica α^*_2 desborda a α mas que α^*_1 (ver las tablas del articulo citado) en muestras moderadas. Para muestras grandes afirma preferir el Z_2 pues el exceso de $\alpha^*_2 - \alpha$ se ve compensado por una ganancia de potencia.

Larntz (1978) compara χ^2, Y^2 y T^2, obteniendo que este último es demasiado liberal frente a los otros dos.

Upton (1982), bajo los criterios señalados en 2.2.3, establece las siguientes conclusiones aparte de las alli señaladas:

1) L^2 es demasiado conservador y no es util con n pequeño

2) T^2 tiene amplias fluctuaciones en el valor de α^*, aunque por termino medio es cercano a α.

3) Y^2 es demasiado liberal en un amplio rango de valores de p.

4) Y^2_w es de los tests mas aceptables según el criterio de Cochran.

2.4. CRITICAS Y COMENTARIOS (APORTACION)

2.4.1. Sobre los principios

La mitad de las discusiones existentes acerca del test aproximado mas apropiado estan basadas en el mismo problema que en los tests del Capitulo I: depende del posicionamiento de cada investigador ante la cuestión de si debe utilizarse siempre el test condicionado de Fisher ó si (para proporciones y asociación) deben elegirse los incondicionados. Bajo ese punto de vista pierde sentido cualquier comparación de tests asintoticos con el test exacto de Fisher.

A estos efectos es interesante observar como el mismo Yates (1984), en uno de sus muchos argumentos descalificadores de los métodos incondicionados, llega a decir que todo el problema radica en que los estadisticos estan obsesionados con obtener significaciones. Resulta sorprendente tal afirmación. Dado que los tests estadísticos sirven para validar hipotesis alternativas a errores α controlados, pero no para validar hipótesis nulas, un estadistico que se precie debe perseguir obtener tests que den las mayores significaciones posibles, es decir, lo mas potentes posibles.

2.4.2. Sobre la comparación de metodos incomparables

Un buen reflejo del confusionismo anterior es el intento de poner en un plano de igualdad los métodos incondicionados exactos y los métodos aproximados.

Asi Garside (1971, 1972) propone un test bajo el formato χ^2 -ver χ^2_G de la expresión (2.45)- con una c.p.c. λ_G.n que depende de n_1, n_2 y α y que le obliga a dar una tabla para su obtención. Es claro que el formato χ^2 es engañoso pues en realidad no se trata de un test aproximado, sino de uno incondicionado, obteniendose los λ_G precisamente bajo el criterio de Barnard. Como las tablas de λ_G son necesariamente limitadas, es claro que el método solo puede entenderse como una alternativa a los métodos incondicionados de Boschloo, McDonald, etc. pero no a los aproximados (χ^2, Y^2, etc.), siendo su método, en esencia, una versión del de Suissa (1985), criterio que ya se vio no era el idóneo. Posteriormente -Garside and Mack (1976)- llega al máximo del confusionismo cuando pone en pie la igualdad, a efectos comparativos, un método condicionado (Fisher), dos incondicionados (Boschloo y Garside) y dos aproximados (χ^2 y χ^2_Y). Igual error comete Upton (1982).

2.4.3. Sobre las conclusiones infundadas.

Algunos autores intentan obtener conclusiones muy generales con pruebas numericas en tablas muy limitadas. Así Conover (1974), Schouten et al (1980) y Upton (1982), obtienen conclusiones tentativas por estar basadas en n_i pequeños ó moderados. Nos parece evidente que hoy dia, con los ordenadores, tal intento pierde validez pues para $n_i \leq 100$ los tests incondicionados son accesibles bien mediante tablas (χ^2_G por ejemplo) ó mediante ordenador (McDonald por ejemplo).

Schouten et al (1980) se limitan a comparar, para el caso de dos proporciones, su χ^2_s con las χ^2 y χ^2_Y, siendo asi que, por ser χ^2_s una cantidad intermedia entre ambas, es claro que no poseera algunos de los defectos que hacen dudar de la validez de las últimas. Se hecha en falta una comparación con los métodos que son sus competidores naturales: ${}_1\chi^2_{PH}$ y χ^2_c.

Burstein (1981), aun basandose en n_i de una amplia gama, comete el error de deducir cual es la c.p.c. apropiada para dos proporciones en base a la comparación de resultados con el test incondicionado binomial (propuesto tambien por Liddell), y ya vimos que tal test -ver introducción a 1.3- era rechazable. Asimismo no compara resultados con las c.p.c. alternativas y distintas de la χ^2_Y.

Dupont (1984) es un caso especial. Es razonable que se sorprenda de que dos tablas tan similares como las 2.2.(b) y (c) den valores de

P tan distintos mediante el test usual de Fisher de dos colas, pero no lo es que, a partir de ello, deduzca que el criterio adecuado es el de doblar el P de una cola. En primer lugar ya se vió en el Cap. I que dicho criterio no es el idóneo. En segundo lugar el problema radica en que para dos proporciones los métodos idóneos son los incondicionados, no los exactos; así los valores de P mediante el método de McDonald -ver tabla 2.2- no son tan sorprendentemente desiguales como los del test de Fisher tradicional. En tercer lugar ya vimos en el Cap. I. que cuando $n_1=n_2$ los criterios de construcción de la RC par los métodos incondicionados suelen dar lugar a mas empates que cuando $n_1 \neq n_2$; de ahí la preocupante diferencia entre los valores de P: con $n_1=n_2$ los puntos van entrando en la RC en grupos numerosos y el P da saltos apreciables.

Además es totalmente rechazable aceptar una versión del test exacto y una del test aproximado solo en base a que son concordantes pues puede que ambas sean incorrectas.

Resulta sorprendente que habiendo surgido el test de Yates como una aproximación a las versiones del test exacto que Dupont desprecia, él le conceda un lugar preponderante a lo que solo era una aproximación del método despreciado. Aceptado esto, el test de Armitage (doblar el P de una cola) tiene ventaja respecto de los demas no aproximados por tener errores de cola simetricos al igual que la aproximación χ^2_Y.

Sin embargo dicho test puede dar lugar a asignar una probabilidad de error (P/2) basada en una cola que <u>no exsite</u>. Desde luego no parece algo apropiado.

Finalmente, los métodos aproximados χ^2_B, χ^2_s y χ^2_c dan valores de P como los indicados en la Tabla 2.2, con lo que χ^2_B es bastante conforme con el incondicionado exacto de MacDonald.

2.4.4. Sobre el error α promedio

Otra fuente de error en las discusiones es la disparidad de criterios a la hora de elegir el método ó c.p.c. adecuada. Si los incondicionados -ver cap. I- estaban sometidos a la fuerte restricción de que su tamaño α^* fuera tal que $\alpha^* < \alpha$, no vemos por que se ha de ser mas laxo con los aproximados, ó, al menos, habrá de imponerse la condición de Cochran.

Asi Plackett (1964) rechaza, para el caso de dos proporciones, el test χ^2_Y por ser demasiado conservador (y tiene razón), aceptando el test χ^2 porque unas veces da un $\alpha^* < \alpha$ y otras $\alpha^* > \alpha$, arguyendo que unos se compensan con otros. Esto no es licito bajo los esquemas usuales pues el experimentador desea tener una garantia de <u>al menos</u> α en su problema, no en el promedio de sus problemas. Su conclusión deberia haber sido que hace falta una c.p.c. intermedia entre o.n y (½)n.

Igual argumento emplea Upton (1982) para seleccionar el test idóneo, aderezado esta vez con una distribución "a priori" de p. Si ese es su parecer, entendemos deberia utilizar algún test bayesiado.

Un problema distinto es el de Eberhardt (1977) que no se preocupa de que $\alpha^* > \alpha$ tanto con Z_1 como con Z_2. El caso de Sathe es aun peor pues ni siquiera se preocupa de determinar el valor α^*_3 de su test Z_3, aunque, dado su formato parecido al Z_2 y su mayor potencia que el Z_1, es presumible que su tamaño α^*_3 supere un exceso a α. Por las afirmaciones y datos de los autores parece claro que el test Z_3 es preferible al clásico con muy grandes muestras, pero, en tal caso, la propia magnitud de los n_i hace que todas las cantidades Z_i sean muy parecidas y la ganancia en potencia prácticamente nula. Así:

1) Si n_1=1000, x_1=400, n_2=1000, x_2=450, se obtiene que $Z_1=Z_2$ = Z_3= 2,26.

2) Si n_1=1000, x_1=400, n_2=2000, x_2=900, se obtiene que $Z_1=Z_3$= 2,61 y Z_2= 2,62, que son prácticamente iguales.

Asimismo, todas las c.p.c. pequeñas pueden hacerse bien con el factor "n", bien con el factor "n-1" (provenientes de la varianza de la hipergeometrica ó de la estimación de la varianza p.q)

2.4.5. **Sobre las comparaciones ausentes**

Además de faltar -por razones obvias- las comparaciones entre las c.p.c. no propuestas, faltan tambien comparaciones entre las c.p.c. mas plausibles en el caso de comparación de dos proporciones. Tanto Burstein como Schouten se dedican a validar sus χ^2_B y χ^2_S frente a las χ^2 y χ^2_Y clásicas, no habiendo comparación entre ellas y las χ^2_c. Unicamente Upton realiza comparaciones entre ellas (dentro de un contexto mas general), pero, por lo dicho hasta ahora, y por lo del proximo apartado, su método de comparación no parece satisfactorio.

Una comparación parcial, facilmente accesible, seria la de los métodos χ^2_B y χ^2_c bajo el criterio de Burstein.

Dado que $\chi^2_B=\chi^2_c$ cuando $n_1=n_2$, la Tabla 2.3 solo recoge una selección de las tablas de Burstein (1981) con $n_1\neq n_2$, incluyendose en ella los valores de P mediante los tests binomial y χ^2_B según los datos de dicho articulo, y los de χ^2_c segun nuestros calculos. Puede observarse que el test χ^2_c es muy concordante con el binomial, siendo asi preferible bajo el criterio de Burstein. Tal resultado no es concluyente pues, como se dijo, el test binomial es rechazable. De cualquier modo χ^2_c es, desde el punto de vista teórico, el idóneo.

Tabla 2.3

Valores de P para las tablas y métodos indicados

n_1,n_2	x_1,x_2	Binomial	χ^2_B	χ^2_c
100,50	18,0	0.0021	0.0017	0.0019
	18,2	0.041	0.038	0.041
	32,4	0.0015	0.0014	0.0015
	28,7	0.065	0.061	0.065
	64,18	0.0013	0.0013	0.0013
	60,22	0.072	0.069	0.065
100,70	14,0	0.0014	0.0013	0.0012
	12,2	0.036	0.038	0.034
	29,6	0.0013	0.0014	0.0012
	25,10	0.092	0.097	0.091
	62,26	0.0014	0.0016	0.0015
	58,30	0.053	0.056	0.053
100,90	10,0	0.0027	0.0026	0.0022
	8,2	0.085	0.086	0.078
	29,10	0.0024	0.0026	0.0024
	25,13	0.072	0.075	0.074
	61,34	0.0014	0.0015	0.0014
	57,38	0.043	0.045	0.043
100,99	10,0	0.0017	0.0016	0.0013
	8,2	0.063	0.062	0.054
	29,11	0.0016	0.0019	0.0017
	25,15	0.085	0.090	0.083
	61,39	0.0024	0.0025	0.0023
	57,43	0.057	0.059	0.056

Tabla 2.3 (Cont.)

n_1, n_2	x_1, x_2	Binomial	χ^2_B	χ^2_c
1000,500	22,0	0.0014	0.00098	0.0011
	8,0	0.059	0.054	0.060
	50,8	0.0016	0.0014	0.0015
	50,14	0.052	0.050	0.052
	100,25	0.0011	0.0010	0.0011
	100,35	0.060	0.058	0.060
	200,66	0.0013	0.0012	0.0013
	200,80	0.064	0.063	0.064
	500,205	0.0011	0.0010	0.0011
	500,225	0.071	0.069	0.071
1000,900	12,0	0.0012	0.0012	0.0010
	4,0	0.063	0.072	0.061
	20,3	0.0010	0.0011	0.0009
	20,8	0.047	0.049	0.049
	50,20	0.0014	0.0014	0.0014
	50,30	0.072	0.075	0.071
	100,54	0.0015	0.0015	0.0014
	100,67	0.050	0.051	0.050
	200,130	0.0014	0.0015	0.0014
	200,150	0.062	0.063	0.062
	500,385	0.0016	0.0017	0.0016
	500,410	0.053	0.054	0.053
1000,999	12,0	0.00079	0.00064	0.00052
	4,0	0.064	0.057	0.045
	20,4	0.0013	0.0012	0.0010
	20,10	0.073	0.072	0.066
	50,22	0.00088	0.00086	0.00076
	50,33	0.061	0.060	0.057
	100,60	0.0011	0.0011	0.0010
	100,75	0.050	0.051	0.049
	200,145	0.0012	0.0012	0.0012
	200,170	0.087	0.088	0.086
	500,430	0.0018	0.0019	0.0018
	500,460	0.075	0.078	0.077

2.4.6. **Sobre los criterios de comparación adecuados**

La metodologia seguida para seleccionar el mejor método asintotico está enfocado, en nuestra opinión, históricamente mal.

Si como es tradicional (y hay razones experimentales para hacerlo así) uno desea utilizar un test asintótico del tipo χ^2, lo primero es clarificar la posición del investigador acerca del condicionamiento.

1º) Si uno es partidario del condicionamiento, entonces la comparación se centra en ver qué metodo asintotico es mas conforme con el test de Fisher. Como en el capitulo I ya se observó que dicho test admite varias versiones de construcción de la RC segun el numero de marginales dados, la comparación habria de ser entre los distintos metodos aproximados (χ^2_{ks}, χ^2_H y P_M) y el método exacto "ad hoc" (por ejemplo, la definición 3.a de 1.1.1 para el caso de dos proporciones)

2º) Si uno es partidario de los test incondicionados, entonces la comparación se centra en ver qué método asintotico es mas conforme con el test incondicionado apropiado al problema (de dos proporciones ó de asociación). Así, en el caso de dos proporciones, habria de compararse el método de McDonald ó el de Intervalos de confianza con los métodos ${}_1\chi^2_{PH}$, χ^2_B, χ^2_s y χ^2_c. Sin embargo tales comparaciones pierden algo de sentido si se observa que en realidad tales métodos de chi-cuadrado son aproximaciones del método incondicionado de Suissa and Shuster (1985) -entrar los puntos en la RC de mayor a menor valor de su χ^2- que no era el optimo en tal situación. Así pues la proxima cuestión es decidir si las comparaciones van a hacerse con el óptimo ó con el metodo χ^2 incondicionado.

Una vez tomada la decisión de los tests a comparar, las comparaciones no deben hacerse en base a los tamaños de los tests para

RC'S al 5% o 1% sino en base a los valores de P en cada punto de cada tabla, imponiendo algunos topes a las diferencias P*-P entre los valores de P de la aproximación y las reales del método exacto utilizado.

De cualquier modo parece claro que la c.p.c. va a depender de n_1, n_2, x_1 y x_2, mientras que su utilidad practica requiere sea un número fácil de obtener. De ahi que, muy probablemente, unas c.p.c. sean mas apropiadas en unos casos y otras en otros, debiendo imponer condiciones no muy rigidas a las diferencias P*-P para que se pueda seleccionar, fuera de toda duda, el test aproximado optimo.

Pero todo esto será motivo de un trabajo futuro

2.5. CONCLUSIONES (APORTACION)

Para realizar un test aproximado en una tabla 2x2:

1ª) Si todos los marginales están fijos - ó, no estandolos, uno es partidario del método condicionado-, realizar el test χ^2_Y, ó, para mas precisión, realizar cualquiera de los tests χ^2_{ks}, χ^2_H ó P_M.

2ª) Si solo un marginal está fijo (comparación de dos proporciones) las investigaciones realizadas hasta la fecha no permiten decantarse por un test determinado. Razones teóricas y prácticas apuntan que el test idóneo debe ser uno de los χ^2_B, ${}_1\chi^2_{PH}$, χ^2_s ó χ^2_c, pero hasta la fecha no hay una comparación sólida entre ellos. Bajo un punto de vista puramente teórico, el test χ^2_c es el idóneo.

3ª) Si ningun marginal esta fijo (asociación) las investigaciones realizadas son menos amplias que las dedicadas al caso anterior, pero razones teóricas, y algunas prácticas, sugieren que el test adecuado es el ${}_o\chi^2_{PH}$ y ello practicamente si alternativa.

CAPITULO III

BIOEQUIVALENCIA CON PROPORCIONES

Y

TEST DE ELECCION MULTIPLE

Hasta ahora las tablas 2 x 2 se han venido analizando fundamentalmente bajo la perspectiva de que están basadas en dos binomiales independientes de tamaños n_1 y n_2 y parametros desconocidos p_1 y p_2 sobre los cuales nos preguntamos si son iguales ó no ($H_o \equiv p_1=p_2$ vs. $H_1 \equiv p_1 \neq p_2$ si es un test de dos colas). Tal proceder está basado en el hecho de que asumimos que la hipótesis a demostrar es que "las proporciones son distintas", y de ahí que ella sea colocada como H_1. Bajo tal esquema el experimentador se ve obligado a menudo a aceptar H_o pues no hay evidencias de significación, y sin embargo puede ser bastante factible que sea falsa. De hecho, es facil obtener un resultado no significativo, aun cuando exista una diferencia real entre los tratamientos, simplemente tomando tamaños de muestra inadecuados ó utilizando metodos de analisis poco sensibles. Alternativamente, tamaños muestrales grandes pueden dar lugar a un rechazo correcto de H_o basado en una diferencia p_1-p_2 estadisticamente significativa pero experimentalmente sin importancia.

En muchas aplicaciones, el proposito de la experiencia es establecer, fuera de toda duda, la equivalencia de dos tratamientos. Tales ejemplos surgen en los estudios de biodisponibilidad de drogas -Westlake (1974)- en los que uno desea determinar si una nueva formulación de una cierta droga es bioequivalente (produce prácticamente los mismos efectos) que la formulación estandar. Otro ejemplo -objeto de este capitulo- surge de los ensayos clínicos en que un cierto procedimiento nuevo de tratamiento de una enfermedad se compara con la práctica usual no con el fin de mejorar su porcentaje de curaciones, sino con el fin de probar que ellos son prácticamente iguales a un costo (de cualquier tipo) menor. Con tal planteamiento el test clasico no es válido por no servir para probar la igualdad, siendo tradicional usar criterios de intervalos para p_1-p_2 y decidir segun su rango. Tal aproximación es la intentada por Westlake para el caso de comparar medias, pero se ha demostrado poco potente frente a un criterio que está basado en test de hipotesis (Anderson and Hanck, 1983).

Alternativamente, y para el caso de dos proporciones, Dunnett and Gent (1977) proponen plantear un test de hipótesis en que la hipótesis nula es $H_o \equiv |p_1-p_2| = \Delta$; con $\Delta \neq 0$ una cantidad que el investigador considera despreciable y que, de ocurrir, le lleva a aceptar la equivalencia práctica de ambos tratamientos ($H_1 \equiv |p_1-p_2| < \Delta$).

3.1. BIOEQUIVALENCIA EN TABLAS 2 x 2

Dada una tabla 2 x 2 clásica como la Tabla 3.1, notemos por p_1 y p_2 a las proporciones reales y por $p_1-p_2=\delta$. El objetivo es contrastar $H_o \equiv \delta = \Delta$ vs. $H_1 \equiv \delta < \Delta$, con Δ dada de antemano (para un test de dos colas poner $|\delta|$) y para el caso de grandes muestras.

Tabla 3.1

	SI	NO	
Tratamiento 1	x_1	y_1	n_1
Tratamiento 2	x_2	y_2	n_2
	a_1	a_2	n

3.1.1. Solución por el método de chi-cuadrado

Aceptando que las frecuencias esperadas, bajo H_o, deben ser concordantes con los marginales realmente obtenidos, entonces habra de ocurrir que $\widetilde{p}_1=\widetilde{p}_2+\Delta$ y $n_1\widetilde{p}_1+n_2\widetilde{p}_2=x_1+x_2$, con lo que

$$\widetilde{p}_1 = \frac{a_1 + n_2\Delta}{n} \qquad \widetilde{p}_2 = \frac{a_1 - n_2\Delta}{n} \tag{3.1}$$

y las cantidades esperadas, bajo H_o, seran $\hat{x}_1=n_1\widetilde{p}_1$, $\hat{x}_2=n_2\widetilde{p}_2$, etc. Con ello el estadistico oportuno es

$$\chi^2_{exp} = (x_1 - \hat{x}_1)^2\left\{\frac{1}{\hat{x}_1}+\frac{1}{\hat{x}_2}+\frac{1}{\hat{y}_1}+\frac{1}{\hat{y}_2}\right\} \tag{3.2}$$

ó con corrección por continuidad

$$_c\chi^2_{exp}= \left\{|x_1-\hat{x}_1| - 0,5\right\}^2\left\{\frac{1}{\hat{x}_1} + \frac{1}{\hat{x}_2} + \frac{1}{\hat{y}_1} + \frac{1}{\hat{y}_2}\right\} \tag{3.3}$$

cantidades a comparar con una distribución χ^2 con 1 g.l., dividiendo el error P final por 2.

Para test de dos colas no es aconsejable multiplicar por 2 el error de una cola (es decir aceptar el error nominal de las tablas usuales). El procedimiento -Mantel (1974)- consiste ahora en determinar todas las frecuencias observadas que, con iguales marginales, dan desviaciones, en las cantidades esperadas, opuestas a la experimental y con verosimilitudes iguales ó menores que la de la tabla experimental. La suma de los dos errores P seria el error de dos colas. Como ello suele requerir un redondeo del valor x_1' previsto para la otra cola, la elección de si se redondea por abajo (x_1^-) ó por arriba (x_2^+) deberá hacerse determinando la razón de verosimiltud de cada uno de ellos al valor observado y eligiendo aquel que produzca un valor mas cercano, pero menor, a uno.

3.1.2. Solución mediante la aproximación normal

Es conocido que cuando se contrasta $H_o \equiv \delta = o$, la aproximación normal es equivalente al test χ^2. Cuando $\Delta \neq 0$ esto no es asi como se ve enseguida. Ahora, y basandose en que para grandes muestras $\hat{p}_1-\hat{p}_2$ se distribuye aproximadamente como una normal de media Δ y varianza $(p_1q_1/n_1) + (p_2q_2/n_2)$, con $q_i=1-p_i$ como siempre, la cantidad experimental oportuna será

$$t_{exp} = \frac{(\hat{p}_1-\hat{p}_2-\Delta) + \frac{1}{2}(n_1^{-1}+n_2^{-1})}{\sqrt{\dfrac{\tilde{p}_1\tilde{q}_1}{n_1} + \dfrac{\tilde{p}_2\tilde{q}_2}{n_2}}} \tag{3.4}$$

en donde $\hat{p}_i=x_i/n_i \hat{q}_i=1-\hat{p}_i$, $\tilde{q}_i=1-\tilde{p}_i$ y las $\tilde{p}_i$ obtenidas: i) como en las (3.1); ii) $\tilde{p}_i=\hat{p}_i$. Ello da lugar a dos valores ${}_1t_{exp}$ y ${}_2t_{exp}$, respectivamente, a comparar con una normal tipica.

3.1.3. Solución mediante el método de Gart

Otro modo de realizar el test anterior seria dar intervalos de confianza para δ, para distintas confianzas, y determinar aquella -(1-P) -que da un intervalo que por primera vez no incluye a cualquier valor $\delta \geq \Delta$. Así se obtiene el error P de la tabla utilizada.

La dificultad de tal proceder es que no existen limites de confianza exactos para δ por no haber ningun estadístico suficiente para δ (al depender todos de parametros perturbadores).

Un modo de obviar la dificultad consiste en replantear el problema en términos de parametros para los que si exista un estadístico suficiente. Tal parametro es la <u>razon de odds</u> $\psi=p_1q_2/p_2q_1$. Es conocido -Fisher (1935)- que la distribución de probabilidad condicional es

$$g(x_1|a_1,\psi) = \frac{\binom{n_1}{x_1}\binom{n_2}{a_1-x_1}\psi^{x_1}}{\sum_i \binom{n_1}{i}\binom{n_2}{a_1-i}\psi^{i}} \tag{3.5}$$

$$i_1=\max(0;a_1-n_2) \leq i \leq \min(n_1;a_1)=i_2$$

expresión que se reduce a la hipergeometrica del test exacto de Fisher cuando $\psi=1$ (ó $\delta=0$). Con ello -Cornfield (1956)- podemos determinar limites de confianza exactos para ψ en el modo usual: para dos colas

$$\sum_{i=i_1}^{x_1} g(i|a_1,\psi_I) = \alpha/2 = \sum_{i=x_1}^{i_2} g(i|a_1,\psi_S) \tag{3.6}$$

dará el intervalo $\psi_I \leq \psi \leq \psi_S$ de confianza $\geq 1-\alpha$.

La utilización directa del criterio anterior requiere formular el problema en términos de ψ en lugar de δ, pero en los problemas que nos ocupa las discrepancias entre p_1 y p_2 se mide de modo natural en términos de p_1-p_2 ó p_1/p_2, no de ψ . Una aproximación debida a Gart (1971) consiste en considerar $x^*_1=E(x_1|a_1,\psi)$ que esta relacionado con ψ en grandes muestras, mediante:

$$\psi = \frac{x_1^*(n_2+x_1^*-a_1)}{(n_1-x_1^*)(a_1-x_1^*)} \tag{3.7}$$

-Cornfield(1956)- con lo que resolviendo la ecuación en x^*_1 para valores $\psi=\psi_I,\psi_S$, se obtendrá el intervalo $x^*_{1I} \leq x^*_1 \leq x^*_{1S}$, y, consecuentemente, el intervalo para δ será:

$$\delta_I = \frac{x^*_{1I}}{n_1} - \frac{a_1-x^*_{1I}}{n_2} \;,,\; \delta_S = \frac{x^*_{1S}}{n_1} - \frac{a_2-x^*_{1S}}{n_2} \tag{3.8}$$

Sin embargo esta solución clásica, adopatada por diversos autores, no conduce a intervalos de confianza para $\delta=p_1-p_2$ pues δ no es una función monotona 1 a 1 de ψ.

La metodologia exacta consiste en despejar x^*_1 de $\Delta=x^*_1/n_1 - (a_1-x^*_1)/n_2$ y sustituirla en la (3.7), lo que da:

$$\hat{\psi} = \frac{(a_1+n_2\Delta)(n_1+n_2-a_1+n_1\Delta)}{(a_1-n_1\Delta)(n_1+n_2-a_1-n_2\Delta)} \tag{3.9}$$

obteniendo a continuación el error P de una cola:

$$P = \sum_{i=i_1}^{x_1} g(i|a_1,\hat{\psi}) \qquad \text{para la cola inferior}$$

$$P = \sum_{i=x_1}^{i_2} g(i|a_1,\hat{\psi}) \qquad \text{para la cola superior} \qquad (3.10)$$

procediendo, para el caso de dos colas, en un modo similar al descrito en 3.1.1.

3.1.4. Comparación numerica de las tres soluciones

Dunnett and Gent comparan las tres aproximaciones

(i) Test χ^2 con corrección por continuidad

(ii) Test de la aproximación normal con corrección por continuidad

(iii) Test basado en el método de Gart de la distribución condicional exacta.

en donde utilizan las correcciones por continuidad pues, experimentalmente, asi logran un mayor acuerdo con el método (iii). Tras la obtención de los valores P de diversas tablas concluye que los tres procedimientos dan, aproximadamente, los mismos resultados, pero el test χ^2 suele estar mas cercano al del método de Gart. El test que menos concuerda con el resto es el (ii) basado en los estimadores individuales $\hat{p}_i$.

La comprobación la hace solo para test de 1 cola

3.2. TEST DE ELECCION MULTIPLE EN TABLAS 2 x 2 (<u>APORTACION</u>)

3.2.1. Relación con los test de bioequivalencia

Supongamos que se tienen objetos de dos tipos (R y $\bar{R}$) y un experimentador que intenta reconocerlos. Un modo de estudiar su capacidad en eloficio consiste en presentarle un conjunto de n objetos en total -de los que n_1 son R y n_2 son $\bar{R}$- y solicitar que los identifique. Si notamos por A el hecho de que identifique a un objeto como de tipo R, y por $\bar{A}$ a lo propio con $\bar{R}$, el resultado de la experiencia puede resumirse en una tabla 2 x 2 como la Tabla 3.2 (la notación es distinta a la utilizada hasta ahora para poder generalizarla, posteriormente, con facilidad). Asi hay x_{11} objetos de tipo R correctamente

Tabla 3.2

Clasificación real y del experimentador en una prueba de clasificación de objetos de dos tipos

		Clasificacion del experimentador A	$\bar{A}$	Totales
Clasificación real	R	x_{11}	x_{12}	n_1
	$\bar{R}$	x_{21}	x_{22}	n_2
Totales		a_1	a_2	n

clasificados por el experimentador, etc.

El planteamiento del problema recuerda fuertemente al problema de Fisher (1942) de "la señorita y las tazas de te" aludido en la introducción, solo que entonces no solo se fijaron los dos marginales n_i como iguales (n_1=n_2) para optimizar la eficacia de la prueba, sino que, por dar tal información a la señorita , tambien quedaron fijados los marginales a_i (a_1=a_2=n_1=n_2) y el problema se convierte asi en una hipergeometrica. Para tal situación Gridgeman (1959 a) y b)) dio un modelo similar al descrito. Tales modelos condicionados no seran considerados aqui porque, ademas de no verse la ventaja de dar al experimentador la información del número de objetos de cada tipo (n_1 y n_2), se hace dificil pensar que, aun dandoselos, el mismo tenga tiempo para equilibrar sus respuestas al azar hasta conseguir que a_1=n_1 y a_2=n_2, especialmente si la prueba consta de un gran número de n objetos (como se supone en lo que sigue para que las aproximaciones sea validas) y el tiempo para las respuestas es limitado (como es lo usual). La dificultad es especialmente comprensible cuando el número de tipos de objetos es superior a 2 (como se estudiará en la proxima sección).

El planteamiento de la Tabla 3.2 es perfectamente asimilable a un examen de respuesta múltiple en el que se le proponen al alumno n preguntas en total, n_1 de ellas con respuesta real SI y n_2 de ellas con respuesta real NO; si el alumno no conoce tales valores, sus respuestas pueden agruparse como en la Tabla 3.3 (que es una particularización de la Tabla 3.2). En adelante nos referiremos a esta terminologia por ser mas facil y comprensible; así, ahora, x_{11} es

Tabla 3.3

Respuestas reales y del alumno a un examen de tipo test con respuestas dicotomicas del tipo SI-NO y con n preguntas en total

		Respuesta del alumno		
		SI $\equiv$ A	NO $\equiv \bar{A}$	Totales
Respuesta real	SI $\equiv$ R	x_{11}	x_{12}	n_1
	NO $\equiv \bar{R}$	x_{21}	x_{22}	n_2
	Totales	a_1	a_2	n

el número de respuestas SI del alumno ante preguntas cuya respuesta real es tambien SI, etc.

Si el alumno conoce el 100-Δ% de la asignatura (de las preguntas posibles) -ó, en la terminologia de la Tabla 3.2, si el experimentador conocer el 100.Δ% de los objetos de cualquier tipo- un posible test a plantear seria $H_o \equiv \Delta = \Delta_o$ contra $H_1 \equiv \Delta < \Delta_o$, con Δ_o un valor dado de antemano (Δ_o=0,5, por ejemplo, para adjudicar el aprobado). Si fuera cierto que $\Delta = \Delta_o$, y suponiendo que un alumno enfrentado a una pregunta de las que conoce la responde bien, en tanto que enfrentado a una de las que no conoce responde al azar A con probabilidad π y $\bar{A}$ con prob. $1-\pi$ (con π desconocida), entonces:

$$\begin{aligned} p_1 &= P(A|R) = \Delta_o + (1-\Delta_o)\pi \\ p_2 &= P(A|\bar{R}) = (1-\Delta_o)\pi \end{aligned} \tag{3.11}$$

(mas detalles en 3.3.1). Como $p_1 - p_2 = \Delta_o$, es claro que para contrastar las hipotesis anteriores habrá de utilizarse el test de bioequivalencia dado en la sección 3.1. Mas adelante se formalizará el problema en este caso y en el general.

3.2.2. Critica al criterio clasico de corrección de pruebas de elección múltiple

En la práctica es universal considerar que un examen como el de la Tabla 3.3 debe evaluarse, para decidir el grado de conocimientos del alumno, en función del número total $x=x_{11}+x_{22}$ de respuestas acertadas. El objeto de esta sección es probar que tal criterio es usualmente erroneo, validando asi la construcción del modelo alternativo que se explicitará en la proxima. Los argumentos que aqui se presentan son válidos en tablas generales kxk pero su comprensión es mas simple en una tabla 2 x 2; de ahí que los incluyamos ahora.

También para simplificar, supongamos que se va a determinar si el alumno sabe algo ó no (Δ_o=0). Llamando de nuevo por $p_1=P(A|R)$ y $p_2=P(A|\bar{R})$ a las probabilidades de que un alumno responda SI a una pregunta cuando su respuesta real es SI ó NO, respectivamente, entonces la comprobación de si un alumno sabe algo ó no (que es equivalente a preguntarse si las respuestas del alumno guardan alguna relación ó no con las reales) puede resolverse mediante el contraste de

$$H_o \equiv p_1 = p_2 \quad \text{vs.} \quad H_1 \equiv p_1 \neq p_2 \tag{3.12}$$

por la (3.11) ó tambien porque es claro que, si no sabe nada, respondera SI por igual ante preguntas cuya respuesta real es SI como ante las que son NO. Tenemos pues un test de dos colas de comparación de dos proporciones basado en los datos de la Tabla 3.3. Cuando n es grande (como se supondrá en lo que resta de memoria) el problema se solventa con el test clásico basado en la aproximación normal; bastará comparar la cantidad

$$t_{exp} = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}\,\hat{q}\,(\frac{1}{n_1} + \frac{1}{n_2})}} \tag{3.13}$$

con una N(0,1) -ver cap. II- con $\hat{p}_i = x_{i1}/n_i$, $\hat{p} = a_1/n$, $q = 1-\hat{p}$; la formula con corrección por continuidad es la usual.

Alternativamente, si el número total de aciertos fuera la variable adecuada para el test anterior, parecerá que el test oportuno para comprobar si un alumno sabe algo ó no deberá ser de

$$H'_o \equiv P(\text{acertar}) = 1/2 \qquad H'_1 \equiv P(\text{acertar}) \neq 1/2 \tag{3.14}$$

pues, si el alumno no sabe nada, acertará al azar la mitad de las preguntas. Tal afirmación es falsa como se ve enseguida. En general:

$$\begin{aligned} P(\text{acertar}) &= P(R \cap A) + P(\bar{R} \cap \bar{A}) = P(R)P(A|R) + P(\bar{R})P(\bar{A}|\bar{R}) = \\ &= \frac{n_1}{n} p_1 + \frac{n_2}{n}(1-p_2) = \frac{n_1}{n} p_1 - \frac{n_2}{n} p_2 + \frac{n_2}{n} \end{aligned} \tag{3.15}$$

y, si el alumno no sabe nada (H_o), será $p_1 = p_2 = p$, con lo que:

$$P(\text{acertar}) = \frac{n_1 - n_2}{n}\, p + \frac{n_2}{n} \qquad (3.16)$$

cantidad que vale ½ solo cuando $n_1=n_2=n/2$ ó $p=½$. Así pues un alumno que no sabe nada acertará la mitad de las preguntas, en largas series, solo si la mitad de ellas tienen SI por respuesta real ó si el alumno responde al azar SI ó NO por igual.

Como ilustración del peligro del falso razonamiento anterior, sea un examen de $n=100$ preguntas de las que $n_1=80$ tienen SI por respuesta real y $n_2=20$ son NO. Si un alumno responde al azar SI en el 90% de los casos ($p=0{,}90$), entonces, en promedio, $x_{11}=72$ y $x_{21}= 18$, con lo que en total acierta 74 preguntas (el 74% de ellas), y, posiblemente, se accederá a admitir que el alumno sabe algo (de hecho el test para una proporción -para H'_o- da una $t_{exp}=4{,}8$ que es fuertemente significativa). Sin embargo el test real (3.12) aplicado a tales datos da una $t_{exp}=0$, con lo que se acepta $H_o \equiv p_1=p_2$ y el alumno no sabe nada.

Para evitar tal error, y puesto que no podemos saber de antemano el porcentaje de respuestas SI que piensa dar un alumno que no sepa nada, podemos preveer que $n_1=n_2=n/2$, con lo que:

$$P(\text{acertar}) = \frac{p_1-p_2}{2} + \frac{1}{2} \qquad (3.17)$$

y así será $p_1=p_2=p(H_o)$ si y solo si $P(\text{acertar})=½$ (H'_o). Pareceria pues que lo adecuado es un test para la proporción de aciertos (en base a los datos $n=$ "número de preguntas realizadas" y $x=x_{11}+x_{22}=$ "número de respuestas adecuadas"). Sin embargo esto tampoco es asi por cuanto, bajo H_o, las variables x_{11} y x_{22} siguen, respectivamente, distribuciones binomiales $B(n/2;p)$ y $B(n/2;q)$ independientes, con lo que x es la suma de dos binomiales independientes de tamaños iguales pero proporciones distintas (aunque complementarias), la cual se convertirá en una binomial ordinaria del tipo $B(n;½)$ solo cuando $p=½$. Como p es desconocido y, presumiblemente, varia de un alumno a otro, el test indicado para una proporción no es el teoricamente adecuado.

Abandonando el falso criterio de la proporción de aciertos, intentemos utilizar en otro modo, pero con igual fin, la variable número de aciertos. El caso general de n_1, n_2 y p cualesquiera, las variables x_{11} y x_{21} siguen binomiales independientes $B(n_1;p)$ y $B(n_2;q)$ si H_o es cierta, y x será la convolucion de ambas. Parece claro que tal variable permitira rechazar H_o cuando tome valores pequeños (pocos aciertos) ó grandes (muchos aciertos). Como además tan extraño es un valor pequeño x_o de x como su homónimo grande $n-x_o$ (con $0 \leq x_o \leq n$), entonces, si x_o es el valor experimental y es pequeño, el error P de la tabla será función de p.

$$P(p)=P(x \leq x_o | p) + P(x \geq n-x_o | p) \tag{3.18}$$

y, puesto que el p común es desconocido, habra de calcularse

$$P = \max_{0<p<1} P(p) \tag{3.19}$$

segun el espiritu de Barnard (ver cap. I). Sin embargo esto tampoco es un buen criterio pues es inmediato comprobar que tales tablas son las que producen diferencias $|\hat{p}_1-\hat{p}_2|$ tan extremas ó mas que las observadas, y algunas mas, lo que indica que para esa misma tabla se obtiene un error P mayor con el criterio de (3.19) que con el clasico de comparación de dos proporciones. De hecho, y para grandes muestras, la variable x sigue, bajo H_o, una $N(n_1p + n_2q;\ npq)$, con lo que el test actual se basara en la cantidad

$$t_{exp} = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \cdot 2\sqrt{\frac{n_1 n_2}{n^2}} \tag{3.20}$$

que coincide con la (3.13) solo cuando $n_1=n_2=n/2$.

Finalmente, y puesto que $\hat{p}_1-\hat{p}_2=x_{11}/n_1 + x_{22}/n_2-1$, puede concluirse que el mejor modo de abordar el problema de si un alumno sabe algo ó no, no es a través del número total de respuestas acertadas, sino mediante la suma de las proporciones de respuestas acertadas de cada tipo. Cuando $n_1=n_2=n/2$, es indiferente utilizar una variable u otra, pero no basta con ninguna de ellas pues la (3.13) requiere tambien el conocimiento de a_1 (el número total de respuestas SI del alumno).

Los razonamientos anteriores son válidos en general para examenes con cualquier número (dos o más) de alternativas, así como para determinar no solo la independencia sino tambien si un alumno debe ó no aprobar etc. Lo que sigue constituye una formalización del problema para determinar el criterio generico adecuado.

3.3. TEST DE ELECCION MULTIPLE EN TABLAS k x k (BIOEQUIVALENCIA EN TABLAS k x k)

3.3.1. El modelo (Aportación)

Veamos de extender el modelo expuesto en el apartado 3.2.2. para una tabla 2 x 2 al caso de una tabla k x k.

Consideremos un examen de tipo test en el que cada pregunta tiene k respuestas a elegir, una sola de las cuales es la verdadera. Si R_i denota el suceso de que la respuesta real a una pregunta es la número i, y A_j al suceso de que la respuesta del alumno a una pregunta es la numero j, con $i,j=1,2,\ldots,k$, entonces el examen de un determinado alumno da lugar a x_{ij} respuestas en que el alumno decide A_j cuando lo real es R_i. Si ademas n_i es el número de preguntas que son R_i, y a_j las que son A_j, entonces $n_i=\sum_j x_{ij}$, $a_j=\sum_i x_{ij}$ y $n=\sum_i\sum_j x_{ij}=\sum_j a_j=\sum_i n_i$ sera el número total de preguntas realizadas. La Tabla 3.4 (a) constituye un buen resumen de estos datos y es una generalización de la Tabla 3.3.

Supongamos ahora que un alumno se sabe el 100Δ % de la asignatura, y que, cuando no sabe una pregunta, responde al azar A_j con probabilidad π_j ($\sum \pi_j=1$) independientemente de cual sea la respuesta real R_i. En tal caso, si C es el suceso de que el alumno conozca la respuesta correcta a una pregunta, se tiene:

$$\left\{\begin{aligned} &P(A_i|R_i) = P(C)\,P(A_i|R_i\cap C) + P(\bar{C})\,P(A_i|R_i\cap\bar{C}) = \\ &\qquad = \Delta + (1-\Delta)\pi_i = p_{ii} \\ &P(A_j|R_i) = (1-\Delta)\pi_j = p_{ij} \qquad (\text{si } i \neq j) \end{aligned}\right. \tag{3.20}$$

puesto que $P(C)=\Delta$, $P(\bar{C})=1-\Delta$, $P(A_i|R_i\cap C)=1$, $P(A_j|R_j\cap\bar{C}) = P(A_j|R_i\cap\bar{C}) = \pi_j$ y $P(A_j|R_i\cap C)=0$.

En tal expresión p_{ij} es la proporción de respuestas A_j ante preguntas cuya respuesta real es R_i.

Tal como se ha enfocado el problema es claro que Δ habra de verificar que $0\leq\Delta\leq 1$. ¿Tendrá sentido plantear el modelo con un Δ negativo?. Si convenimos en que un $\Delta=-0,5$ indica que el alumno se ha apren-

Tabla 3.4

(a)

Respuestas de un alumno a un examen

Real \ Alumno	A_1	A_2	......	A_K	Total
R_1	x_{11}	x_{12}	......	x_{1K}	n_1
R_2	x_{21}	x_{22}	......	x_{2K}	n_2
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
R_K	x_{K1}	x_{K2}	......	x_{KK}	n_K
Total	a_1	a_2	. . .	a_K	n

(b)

Proporciones teoricas de respuestas de um alumno que conoce el 100Δ% de la asignatura.

Real \ Alumno	A_1	A_2	......	A_K	Total
R_1	$p_1 + \Delta$	p_2	......	p_K	1
R_2	p_1	$p_2 + \Delta$	......	p_K	1
.	.	.	......	.	.
.	.	.	......	.	.
.	.	.	......	.	.
R_K	p_1	p_2	......	$p_K + \Delta$	1
	$p_i = (1 - \Delta)\pi_i$,, $\Sigma\pi_i = 1$,, $\Delta \geqslant 0$				

(c)

Proporciones teoricas de respuestas de un alumno que conoce al reves el 100x($-\Delta$)% de la asignatura

Real \ Alumno	A_1	A_2	......	A_K	Total
R_1	p_1	$p_2 - \frac{\Delta}{K-1}$	......	$p_K - \frac{\Delta}{K-1}$	1
R_2	$p_1 - \frac{\Delta}{K-1}$	p_2	......	$p_K - \frac{\Delta}{K-1}$	1
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
R_K	$p_1 - \frac{\Delta}{K-1}$	$p_2 - \frac{\Delta}{K-1}$	......	p_K	1
$p_i = (1 + \Delta)\ \pi_i$,, $\Sigma\pi_i = 1$,, $\Delta \leqslant 0$					

dido la mitad de la asignatura al reves, entonces las (3.20) para un alumno que conoce negativamente una fracción-Δde la asignatura, seran ahora:

$$\begin{cases} P(A_i|R_i) = p_{ii} = (1+\Delta)\,\pi_i \\ P(A_j|R_i) = p_{ij} = -\Delta \cdot \dfrac{1}{K-1} + (1+\Delta)\,\pi_j & (\text{si } i \neq j) \end{cases} \qquad (3.21)$$

si se supone que las respuestas opuestas a las correctas (que son las que cree conocer el alumno con probabilidad $-\Delta$) se encuentran repartidas por igual en las (k-1) respuestas falsas. Ambos modelos pueden especificarse en esta formula única

$$\begin{aligned} p_{ij} &= (1-\Delta)\,\pi_j + \delta_{ij}\,\Delta & \text{si} \quad \Delta \geq 0 \\ &= (1+\Delta)\,\pi_j - (1-\delta_{ij})\,\frac{\Delta}{K-1} & \text{si} \quad \Delta \leq 0 \end{aligned} \qquad (3.22)$$

con δ_{ij} las de Kronecker. Si ademas convenimos en que:

$$\begin{aligned} p_i &= (1-\Delta)\,\pi_i & \text{si} \quad \Delta \geq 0 \\ &= (1+\Delta)\,\pi_i & \text{si} \quad \Delta \leq 0 \end{aligned} \qquad (3.23)$$

entonces

$$\begin{aligned} p_{ij} &= p_j + \delta_{ij}\Delta & \text{si} \quad \Delta \geq 0 \\ &= p_j - (1-\delta_{ij})\,\frac{\Delta}{K-1} & \text{si} \quad \Delta \leq 0 \end{aligned} \qquad (3.24)$$

sera el modelo para un alumno que conozca la fraccion de la asignatura. Notese que los p_i y p_{ij} dependen del valor Δ considerado, por lo que, en ocasiones, expresaremos esta hecho especificando $p_i(\Delta)$ y $p_{ij}(\Delta)$. Los dos modelos descritos se reseñan mas ampliamente en las Tablas 3.4(b),(c).

El modelo descrito está basado en algunas hipótesis que conviene especificar:

HI: Las n preguntas del examen son una pequeña fracción del total de preguntas que podrian hacerse y estan repartidas por igual a lo largo de toda la asignatura. Asimismo las respuestas correctas a cada pregunta están repartidas al azar entre las k respuestas alternativas.

HII: El alumno ó conoce bien el $100\Delta\%$ de la asignatura ($\Delta \geqslant 0$) ó conoce al revés el $100(-\Delta)\%$ de ella ($\Delta < 0$).

H III: Un alumno con un $\Delta \geq 0$, enfrentado a una pregunta de las que conoce, la responde bien.

Un alumno con un $\Delta < 0$, enfrentado a una pregunta de las que cree conocer, responde cualquiera de las respuestas falsas, cada una de ellas con probabilidad $1/(k-1)$.

H IV: Cuando una pregunta no es conocida, el alumno responde A_j con probabilidad π_j desconocida pero constante en toda la prueba ($\sum \pi_j = 1$).

La condición H IV es la mas dificil de verificar por cuanto un alumno que no conoce la respuesta correcta a una pregunta, usualmente tendra duda de elegir alguna de entre 2 ó 3 alternativas, no de entre las k. Sea cual sea la situación, mas adelante se dará un test de comprobación del modelo que permitira salir de dudas. De todos modos, la precaución de tomar valores de k iguales a 2 ó 3 (que como veremos son optimos en otro sentido) hace desaparecer en todo ó en parte la dificultad anterior, especialmente si el examen esta cuidadosamente confeccionado.

La condición H II dificilmente se verificará de modo puro, pues es factible que el alumno conozca una parte, Δ_1, de la asignatura bien, y otra parte, Δ_2, al reves ($\Delta_1 + \Delta_2 \leq 1$); sin embargo usualmente el valor de Δ_2 sera despreciable frente a Δ_1 (en los alumnos que tienen posibilidades de aprobar) y las inferencias que veremos no se veran afectadas grandemente por ello. En realidad, la situación de un Δ negativo es poco frecuente (en comparación con la de un Δ positivo) y se le ha introducido exclusivamente para encontrarle una interpretación a cualquier valor de Δ, lo que sera de utilidad solo en los test de independencia que se veran.

A pesar del lenguaje utilizado hasta ahora, es evidente que el modelo descrito no es solo de aplicación a las pruebas de elección múltiple (examenes de tipo test), sino que tambien podrán emplearse en toda situación en que un experimentador deba clasificar n objetos (de los que se conoce su clasificación real) en k clases y sea razonable suponer que:

a) El experimentador es capaz de reconocer a una fracción Δ de los objetos, fracción que es independiente de la clasificación real del objeto.

b) Ante un objeto que no reconoce, el individuo lo clasifica al azar con probabilidades π_i constantes en toda la prueba.

Por ejemplo, y volviendo al clásico problema de Fisher de "la señorita y las tazas de te", si a ella se le dieran a juzgar tres grupos de tazas (uno con la leche echada antes que el té, otro con el té antes que la leche y, finalmente, un tercero con la leche y el té echados a la vez), sus clasificaciones y las reales darian lugar a una tabla 3 x 3 del tipo de la Tabla 3.4 (a), con Δ representando a la fracción de tazas de cualquier tipo que es capaz de reconocer.

3.3.2. Estimación de los parametros (Aportación)

Supuesto que el modelo incluye k parametros desconocidos (Δ y las k-1 primeras π_i) de las que dependen k distribuciones multinomiales independientes (las filas de la tabla 3.4 (a)), la estimación de los mismos podria intentarse por maxima verosimilitud. Sin embargo ello da lugar a ecuaciones de las que no se puede despejar explicitamente los parametros a estimar; de ahí que intentemos otra solución.

Una estimación de las p_i -y por tanto de las π_i- pasa por encontrar aquellos valores $\hat{p}_i$ que den lugar a unas frecuencias esperadas de columnas iguales a las observadas; con ello, como:

$$E(a_j|\Delta) = \sum_i n_i p_{ij} = \sum_i n_i \{p_j + \delta_{ij}\Delta\} = p_j n + n_j \Delta \quad \text{si } \Delta \geq 0$$
$$= \sum_i n_i \{p_j - (1-\delta_{ij})\frac{\Delta}{k-1}\} = p_j n - \frac{\Delta}{k-1}(n - n_j) \quad \text{si } \Delta < 0 \qquad (3.25)$$

haciendo $E(a_j|\Delta)=a_j$ y despejando p_j se obtiene:

$$\hat{p}_i = \frac{a_i - n_i\Delta}{n} \quad \text{si } \Delta \geq 0$$
$$= \frac{a_i + \frac{\Delta}{k-1}(n-n_i)}{n} \quad \text{si } \Delta < 0 \qquad (3.26)$$

y tambien:

$$p_{ij} = \frac{a_j + (\delta_{ij} n - n_j)\Delta}{n} \quad \text{si } \Delta \geq 0$$
$$= \frac{a_j + \frac{\Delta}{k-1}(\delta_{ij} n - n_j)}{n} \quad \text{si } \Delta \leq 0 \qquad (3.27)$$

cantidades todas ellas que, por depender de Δ, seran aludidas en ocasiones como $\hat{p}_i(\Delta)$ y $\hat{p}_{ij}(\Delta)$.

Para estimar Δ, aceptemos el principio de que la suma de las proporciones de aciertos observadas y esperadas coincidan. Asi, como

$$\Sigma p_{ii} = \Sigma p_i + K\Delta = 1 - \Delta + K\Delta \qquad \Delta \geq 0$$
$$= \Sigma p_i = 1 - \Delta \qquad \Delta < 0 \qquad (3.28)$$

entonces igualandolas a la suma de las proporciones x_{ii}/n_i observadas:

$$\hat{\Delta} = \frac{\Sigma x_{ii}/n_i - 1}{K - 1} \qquad \Delta \geq 0$$
$$= \Sigma \frac{x_{ii}}{n_i} - 1 \qquad \Delta < 0 \qquad (3.29)$$

Notese que para estimar Δ puntualmente basta conocer la suma de las proporciones de aciertos. En el ejemplo de la Tabla 3.5 es $\hat{\Delta}$= 13/40

3.3.3. Comprobación del modelo (Aportación)

Antes de abordar cualquier otro problema, convendria poner a prueba, en cada caso particular, la validez del modelo propuesto en las Tablas 3.4 (b),(c). Con tal fin habrán de calcularse las cantidades esperadas bajo cada modelo:

$$E_{ij}(\Delta) = \frac{a_j + (\delta_{ij} n - n_j)\Delta}{n} \cdot n_i \qquad \Delta \geqslant 0 \qquad (3.30)$$
$$= \frac{a_j + \frac{\Delta}{K-1}(\delta_{ij} n - n_j)}{n} \cdot n_i \qquad \Delta < 0$$

con lo que el test de bondad de ajuste clasico resolvera el problema. Así, el test consistira en comparar la cantidad

$$\chi^2_{exp} = \sum_i \sum_j \frac{\{x_{ij} - E_{ij}(\hat{\Delta})\}^2}{E_{ij}(\hat{\Delta})} \qquad (3.31)$$

con una χ^2 teorica con k(k-2) grados de libertad (g.l.), puesto que hasy k^2 casillas, k relaciones entre las p_{ij} ($\sum_j p_{ij}=1, \forall i$) y k parametros estimados (Δ y las k-1 primeras p_i), con lo que k^2-k-k = k(k-2). En la (3.31) las $E_{ij}(\hat{\Delta})$ aluden a las $E_{ij}(\Delta)$ de la (3.30) cambiando Δ por el $\hat{\Delta}$ de la (3.29). Sin embargo como $E_{ij}(\hat{\Delta}/\Delta>0)=E_{ij}(\hat{\Delta}/\Delta<0)$, la expresión (3.31) no varia, sea cual sea el modelo utilizado, y la formula es única. Así, si la entendemos escrita para $\Delta>0$, la (3.31) puede ponerse en esta otra forma alternativa mas simple:

$$\chi^2_{exp} = n \sum_j \frac{(\sum_i x_{ij}^2/n_i)(a_j + (n-n_j)\hat{\Delta}) - (n\hat{\Delta}\, x_{jj}^2/n_j)}{(a_j - n_j\hat{\Delta})(a_j + (n-n_j)\hat{\Delta})} - n \qquad (3.32)$$

con $\hat{\Delta}$ dada para $\Delta > 0$. El test es válido puesto que los estimadores propuestos son insesgados, asintoticamente normales y asintoticamente eficientes

Así, para los datos de la tabla 3.5, se tiene que $\chi^2_{exp}=3,39$, cantidad que constituye una medida del desajuste del modelo y que es significativa, con 3 g.l., solo para $P > 20\%$. No hay evidencias de fallo del modelo.

Observese que con k=2 no es posible comprobar el modelo pues no hay g.l. suficientes. En realidad lo que ocurre es que en tal circunstancia el modelo es siempre valido por haber una relación biunivoca entre los dos parametros del modelo (π y Δ) y los dos parametros (p_{11} y p_{21}) sin especificar modelo. Puede comprobarse que, en tal caso, ocurre que $x_{ij} = E_{ij}(\hat{\Delta})$, con lo que $\chi^2_{exp}=0$ siempre.

La comprobación del modelo en los alumnos con un valor grande de $|\Delta|$ se convierte en tentativa dado que, en tal caso, las cantidades esperadas de la diagonal principal (si $\Delta < 0$) ó del resto de las casillas (si $\Delta > 0$) seran demasiado bajas.

Tabla 3.5

Resultados de un alumno en un examen de 120 preguntas con k=3 respuestas alternativas

Alumno / Real	A_1	A_2	A_3	Total
R_1	$x_{11} = 22$	$x_{12} = 10$	$x_{13} = 8$	$n_1 = 40$
R_2	$x_{21} = 10$	$x_{22} = 26$	$x_{23} = 4$	$n_2 = 40$
R_3	$x_{31} = 14$	$x_{32} = 8$	$x_{33} = 18$	$n_3 = 40$
Total	$a_1 = 46$	$a_2 = 44$	$a_3 = 30$	$n = 120$

3.3.4. Estimación por intervalo del grado de conocimientos del alumno (Aportación)

El estimador puntual del grado de conocimientos Δ del alumno viene dado por las (3.29) eligiendose una u otra segun que $\sum x_{ii}/n_i$ sea mayor ó menor que la unidad. Ahora bien, cualquiera que sea el caso, x_{ii} sigue una binomial $B(n_i;p_{ii})$ que, para grandes muestras, se aproxima a una normal $N(\mu_i = n_i p_{ii};\ \sigma^2_i = n_i p_{ii}(1-p_{ii}))$; con ello $\hat{\Delta}$ se distribuirá aproximadamente como una normal $N(\Delta;\sigma^2(\Delta))$, con:

$$\sigma^2(\Delta) = \frac{1}{(K-1)^2}\cdot\sum_i \frac{p_{ii}(1-p_{ii})}{n_i} = \frac{1}{(K-1)^2}\cdot\sum_i \frac{(p_i+\Delta)(1-p_i-\Delta)}{n_i} \qquad \Delta \geqslant 0$$

$$= \sum_i \frac{p_{ii}(1-p_{ii})}{n_i} = \sum_i \frac{p_i(1-p_i)}{n_i} \qquad \Delta < 0 \qquad (3.33)$$

con p_i las propias de cada caso y dadas por la (3.23). Por otro lado, como $\sigma^2(\Delta)$ depende de las π_i que suelen ser desconocidas, y ademas varian con el alumno, una estimación $\hat{\sigma}^2(\Delta)$ apropiada puede obtenerse sustituyendo las $\hat{p}_i$ la (3.33) por las $\hat{p}_i$ de la (3.26); con ello:

$$\hat{\sigma}^2(\Delta) = \frac{1}{n^2(K-1)^2}\cdot\sum_i \{a_i+(n-n_i)\Delta\}\cdot\{n-a_i-(n-n_i)\Delta\}/n_i =$$

$$= \frac{1}{n^2(K-1)^2}\{n(1-2\Delta)\sum a_i/n_i - \sum a_i^2/n_i + n^2\Delta(1-\Delta)\sum 1/n_i +$$

$$+\Delta n(2+2\Delta K-K-\Delta) \qquad \Delta \geqslant 0 \qquad (3.34)$$

$$= \frac{1}{n^2}\sum\left\{a_i+\frac{\Delta}{K-1}(n-n_i)\right\}\left\{n-a_i-\frac{\Delta}{K-1}(n-n_i)\right\}/n_i \qquad \Delta < 0$$

Con ello, un intervalo al $100(1-\alpha)\%$ de confianza para Δ sera

$$\Delta \in \hat{\Delta} \pm t_{\alpha/2}\,\hat{\sigma}(\Delta) \qquad (3.35)$$

si se le desea de dos colas. Lo propio para una cola. Cuando k=2 convendria incluir la corrección por continuidad clásica. Como $\hat{\sigma}(\Delta)$ depende de Δ, un modo de observarlo seria sustituir en ella $\hat{\Delta}$ en lugar de Δ, con lo que:

$$\Delta \in \hat{\Delta} \pm t_{\alpha/2}\, \hat{\sigma}(\hat{\Delta}) \tag{3.36}$$

otra solución, mas exacta pero tambien mas laboriosa, consitiria en igualar Δ a los extremos del intervalo (3.35) y despejar Δ de las igualdades asi obtenidas.

Asi, para los datos de la Tabla 3.5, el primer intervalo, al 95% de confianza, daria $\Delta \in (0{,}192;\ 0{,}458)$, en tanto que el segundo indicaria que $\Delta \in (0{,}192;\ 0{,}453)$, lo que es prácticamente igual. Como en este caso $\hat{\Delta} = 13/40 > 0$, los intervalos se han constuido pensando que $\Delta > 0$; si acaso el extremo inferior hubiese dado negativo, el mismo se puede volver a determinar, como intervalo de una cola, en el supuesto de que $\Delta < 0$. Es inmediato que multiplicando por (k-1) los intervalos descritos para $\Delta > 0$ se obtienen los de $\Delta < 0$.

3.3.5. Test para asignar una calificación. (Aportación)

Supongamos que un profesor desea decidir que alumnos conocen al menos la fracción Δ_0 de la asignatura para asi asignarles una calificación (por ejemplo $\Delta_0 = 0{,}5$ indicaria el aprobado). Con tal fin habra de realizar un test de hipotesis para

$$H_0 \equiv \Delta = \Delta_0 \qquad\qquad H_1 \equiv \Delta > \Delta_0 \tag{3.37}$$

pues se decide a aprobar (por ejemplo) a un alumno solo cuando hay garantias de que $\Delta > 0{,}5$. Un profesor mas generoso plantearia el test

$$H_0 \equiv \Delta = \Delta_0 \qquad\qquad H_1 \equiv \Delta < \Delta_0 \tag{3.38}$$

pues quiere suspender a un alumno solo cuando hay evidencias ciertas de que debe suspender. Pero este no suele ser el caso (de ser asi, las soluciones que siguen son facilmente acomodables).

Para efectuar el test, como $\hat{\Delta}$ sigue aproximadamente una N(Δ; $\hat{\sigma}^2(\Delta)$), entonces la cantidad experimental:

$$t_{exp} = (\hat{\Delta} - \Delta_0)/\hat{\sigma}(\Delta_0) \qquad (3.39)$$

con $\hat{\sigma}(\Delta_o)$ obtenida de la (3.34), habrá de compararse, del modo usual con una t_α de la N(0;1) para obtener un test al error α (con t_α el valor de la normal típica que deja a su derecha un area de α; de igual modo en adelante). En particular, cuando todos los tamaños muestrales sean iguales ($n_1=n_2...=n_k=\bar{n}$), la (3.39) se simplifica en esta otra:

$$t_{exp} = K \frac{\Sigma x_{ii} - \bar{n}\{1 + \Delta_o (K-1)\}}{\sqrt{-\frac{\Sigma a_i^2}{\bar{n}} + K\bar{n}\left\{K + (K-1)(K-2)\Delta_o - (K-1)^2 \Delta_o^2\right\}}} \quad \Delta_o \geq 0 \qquad (3.40)$$

Alternativamente, el mismo argumento lleva a esta otra cantidad experimental:

$$t'_{exp} = (\hat{\Delta} - \Delta_o)/\hat{\sigma}(\hat{\Delta}) \qquad (3.41)$$

pues ambas, bajo H_o, son asintoticamente N(0;1). En este caso, la versión simplificada para cuando las n_i son iguales a $\bar{n}$, es la (3.40) cambiando Δ_o por $\hat{\Delta}$ en su denominador.

Aun puede encontrarse una version χ^2 del test. Al tratarse de un test de hipotesis compuesta (pues las π_i son desconocidas), entonces (ver Rao (1973), pag 395) la cantidad apropiada es:

$$\chi^2_{exp} = \sum_i \sum_j \frac{\{E_{ij}(\hat{\Delta}) - E_{ij}(\Delta_o)\}^2}{E_{ij}(\Delta_o)} \qquad (3.42)$$

ó, sustituyendo la (3.30) y operando:

$$\chi^2_{exp} = (\hat{\Delta}-\Delta_o)^2 \sum_i \frac{a_i n_i (n-n_i)}{\{a_i + (n-n_i)\Delta_o\}\cdot\{a_i - n_i\Delta_o\}} \qquad \Delta_o \geq 0 \qquad (3.43)$$

ó cuando las n_i son todas iguales a $\bar{n}$

$$\chi^2_{exp} = (\hat{\Delta}-\Delta_o)^2 \bar{n}^2 (K-1) \sum \frac{a_i}{\{a_i + (K-1)\bar{n}\Delta_o\}\{a_i - \bar{n}\Delta_o\}} \qquad (3.44)$$

cantidades que, extrayendo la raiz cuadrada, pueden ponerse en formato t como $t''_{exp} = \sqrt{\chi^2_{exp}}$, cantidad a comparar tambien con una t_α de la N(0;1) pues la χ^2_{exp} deberia compararse con una $\chi^2_{2\alpha}$ si $\hat{\Delta} > \Delta_o$, con tal valor el que deja a su derecha un area de 2α en una χ^2 con 1 g.l.

Obviamente, para contrastar las (3.38) las t_{exp} anteriores habran de compararse con una $-t_\alpha$ de la N(0;1).

Las expresiones (3.39), (3.41) y (3.42) son válidas para cualquier valor de Δ (positivo ó negativo), pero puesto que la situación $\Delta_o \geq 0$ sera la mas frecuente (practicamente la unica de interes), se han dado formulas particulares para tal caso, y en adelante, salvo indicación expresa de lo contrario nos referiremos a ella. De todos modos las formulas simplificadas anteriores tambien valen para el modelo $\Delta < 0$ con solo cambiar Δ_o por $\Delta_o/(k-1)$.

Así, de nuevo para el ejemplo de la Tabla 3.5, si se quiere contrastar el aprobado ($\Delta_o = ½$), se obtiene que t_{exp}=-2,73, t'_{exp}=- 2.59 y t''_{exp}=-2,89. Todas ellas, muy parecidas, llevan a que el alumno debe suspender segun (3.37)- pues $\hat{\Delta}$ =13/40 < ½- y segun (3.38)- pues dichas t_{exp} son significativas para un error P ≤ 1%.

Cuando k=2 ocurre que $\hat{\Delta} = (x_{11}/n_1) + (x_{22}/n_2) - 1 = (x_{11}/n_1) - (x_{21}/$

n_2) y la suma de las proporciones de aciertos es lo mismo que la diferencia de las proporciones de respuestas A_1 a preguntas que son R_1 ó R_2. Con ello la (3.39) pasa a ser:

$$ {}_c t_{exp} = n \frac{\left(\frac{x_{11}}{n_1} - \frac{x_{21}}{n_2} - \Delta_o\right) - \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{\frac{(a_1 + n_2\Delta_o)(a_2 - n_2\Delta_o)}{n_1} + \frac{(a_2 + n_1\Delta_o)(a_1 - n_1\Delta_o)}{n_2}}} \qquad (3.45) $$

y la (3.40)

$$ {}_c t_{exp} = \frac{x_{11} - x_{21} - \bar{n}\Delta_o - 1}{\sqrt{\frac{a_1 a_2 - \bar{n}^2 \Delta_o^2}{2\bar{n}}}} \qquad (3.46) $$

Es inmediato comprobar que $\hat{p}_{ij}(\hat{\Delta}) = \tilde{p}_{ij} = x_{ij}/n_i$, de modo que la (3.41) es ahora:

$$ {}_c t'_{exp} = \frac{(\tilde{p}_{11} - \tilde{p}_{21} - \Delta_o) - \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{\frac{\tilde{p}_{11}\tilde{p}_{12}}{n_1} + \frac{\tilde{p}_{21}\tilde{p}_{22}}{n_2}}} \qquad (3.47) $$

y con $\tilde{p}_{ij}$ los estimadores frecuentistas (individuales) de las p_{ij}, y, cuando todos los n_i son iguales a $\bar{n}$:

$$ {}_c t'_{exp} = \frac{x_{11} - x_{21} - \bar{n}\Delta_o - 1}{\sqrt{\frac{x_{11}x_{12} + x_{22}x_{21}}{\bar{n}}}} \qquad (3.48) $$

Finalmente, como ahora $E_{ij}(\hat{\Delta})=x_{ij}$, por ser $\hat{p}_{ij}(\hat{\Delta})=x_{ij}/n_i$ y como todas las $|E_{ij}(\hat{\Delta}) - E_{ij}(\Delta_o)|$ son ahora iguales, la (3.42) se convierte en:

$$ {}_c\chi^2_{exp} = (x_{11} - E_{11}(\Delta_o) - 0{,}5)^2 \left(\frac{1}{E_{11}(\Delta_o)} + \frac{1}{n_1 - E_{11}(\Delta_o)} + \frac{1}{a_1 - E_{11}(\Delta_o)} + \frac{1}{n_2 - a_1 + E_{11}(\Delta_o)}\right) $$

(3.49)

y la (3.43):

$$ {}_{c}\chi^2_{exp}=\left\{\hat{\Delta}-\Delta_0-½\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right\}^2 n_1n_2 \quad \frac{a_1}{(a_1+n_2\Delta_0)(a_1-n_1\Delta_0)}+\frac{a_2}{(a_2+n_1\Delta_0)(a_2-n_2\Delta_0)} $$

(3.50)

Observese que en el caso de k=2 el modelo incluye solo dos proporciones desconocidas, $p_{11}=p_1+\Delta$ y $p_{21}=p_1$, con lo que el test (3.37) se convierte en un contraste acerca de si la diferencia de dos proporciones $(p_{11}-p_{21})$ es Δ_o ó mas. Tal test fué ya tratado en la sección 3.1 y en el apartado 3.2.1. y se corresponde con el de Dunnett and Gent descrito en la primera. Asi la expresión (3.3) es la misma (3. 49) de ahora, en tanto que las dos versiones de la (3.4) son las (3. 45) y (3.47) de ahora (las correcciones por continuidad difieren en el signo pues las expresiones de entonces estaban pensadas para la alternativa $H_1 \equiv \delta < \Delta$ y las de ahora para la alternativa $H_1 \equiv \Delta > \Delta_o$). En este sentido podemos afirmar que el test actual para tablas k x k es una generalización del test de bioequivalencia de Dunnett and Gent.

En todos los casos las correcciones por continuidad habran de cambiarse de signo cuando la hipótesis a contrastar sean las (3.38).

De nuevo se ve que el estadistico adecuado para calificar a un alumno no es el número de preguntas acertadas, sino la suma $(\sum x_{ii}/n_i)$ de las proporciones de aciertos de cada tipo, suma en la que están basados los test y el estimador puntual $\hat{\Delta}$ del parametro de interes. Los dos criterios son equivalentes, a efectos de estimación puntual, cuando las n_i son todas iguales, pero no a efectos de test ó intervalo pues el mismo requiere tambien el conocimiento de las a_i (números totales de respuestas de cada tipo que da el alumno). En ultimo extremo, y dada la distinta varianza de $\hat{\Delta}$ para cada alumno, las calificaciones de suspenso, aprobado, etc. deberan darse en función de los distintos valores de t_{exp}, no de $\hat{\Delta}$.

3.3.6. Test de independencia (Aportación)

El caso de la independencia es algo especial y merece comentario aparte. Ahora contrastar la independencia parece mas aconsejable hacerlo mediante un test de dos colas

$$H_o \equiv \Delta = 0 \text{ (independientes) } ,, \ H_1 \equiv \Delta \neq 0 \text{ (dependientes)} \tag{3.51}$$

para asi determinar si el alumno sabe algo ó nada, aunque no hay inconveniente en hacerlo de una cola.

Es inmediato comprobar que tanto la (3.39) como la (3.41) toman igual valor en el caso de $\Delta \geq 0$ como en el de $\Delta < 0$ de ahi que los correspondientes tests para la (3.46) puedan ser escritos de modo unico asi:

$$t_{exp} = n\,(K-1)\,\frac{|\,\hat{\Delta}\,|}{\sqrt{\Sigma \dfrac{a_i\,(n - a_i)}{n_i}}} \tag{3.52}$$

$$t'_{exp} = n\,(K-1)\,\frac{|\,\hat{\Delta}\,|}{\sqrt{\Sigma\,\{a_i + (n-n_i)\hat{\Delta}\}\,.\,\{n-a_i-(n-n_i)\hat{\Delta}\}\;/n_i}} \tag{3.53}$$

Para el caso χ^2 ya se indico antes que la $E_{ij}(\Delta)$ son idénticas en los casos $\Delta > 0$ y $\Delta < 0$; como ademas igual ocurre con $E_{ij}(\Delta_o=0)=n_i a_j/$ $/n$, entonces la nueva cantidad experimental sera $t''_{exp} = \sqrt{\chi^2_{exp}}$, con:

$$\chi^2_{exp} = \hat{\Delta}^2 \Sigma\,\frac{n_i\,(n-n_i)}{a_i} \tag{3.54}$$

En todos los casos se sobreentiende $\hat{\Delta}$ obtenida para cuando $\Delta > 0$.

Así volviendo de nuevo al ejemplo de la Tabla 3.5, los test para la (3.51) dan lugar a las siguientes cantidades experimentales: t_{exp}=5,08 t'_{exp}=4,81 y $t''_{exp}=\sqrt{26.30}$ = 5,13, todas ellas significativas y similares. El alumno sabe algo, la cantidad teorica es ahora una $t_{\alpha/2}$ de la N(0;1).

Observese como en todos los test, un valor $\hat{\Delta}>0$ (obtenido bajo $\Delta>0$) es siempre mas significativo que un $\Delta<0$ (obtenido bajo $\Delta<0$) igual que el anterior en valor absoluto.

En el caso particular de k=2 las cantidades (3.45) (3.47) y (3.49), al hacer en ellas Δ_o=0, dan lugar a los test clásicos de comparación de dos proporciones independientes (ver cap. II).

Es obvio que, sea Δ_o=0 ó $\Delta_o\neq$0, los tres test t propuestos no son equivalentes, coincidiendo solamente los test t y t'' en el caso particular de k=2 y Δ_o=0.

Por otro lado, el test para las (3.51) es un test de independencia de las dos clasificaciones, test que seria un competidor del test χ^2 clasico de homogeneidad de varias proporciones. De verificarse las condiciones del modelo, es claro que el test actual sera preferible al test clásico de homogeneidad por cuanto aquel considera el hecho de que la no homogeneidad solo puede ocurrir de un modo (que las proporciones de la diagonal principal superen en una misma cantidad a todas las de su columna), y este la supone en muchos modos (cualquier situación en que las proporciones por columnas no sean todas iguales). Cuando k=2 ambos métodos coinciden pues ahora solo hay dos proporciones por columna y solo un modo de que estas no sean iguales: que una de ellas supere a la otra en Δ .

3.3.7 Discusión y comentarios (Aportacion)

En un examen de tipo test el alumno debe elegir una entre k respuestas alternativas, solo una de las cuales es verdadera. El criterio usual para calificar tales examenes se basa en la variable x= "numero de respuestas acertadas" (Σx_{ii} en la notación de la Tabla 3.4 (a)), variable que solo es adecuada cuando hay igual número de preguntas para cada respuesta real, pero que ni aun asi debiera ser suficiente pues los totales a_i de respuestas de cada tipo que da el alumno tambien intervienen en el test. En general las variables a tener en cuenta son la "suma de las proporciones de aciertos de cada tipo" junto con las citadas a_i.

En el texto se dan tres test alternativos para determinar si una calificación debe darse ó no a un alumno. Si, por alguna razón, el número de aprobados esta fijado de antemano, la variable "suma de las proporciones de aciertos", ó, equivalentemente, el estadistico $\hat{\Delta}$ dado por la (3.29), no debe servir como criterio de ordenación, pues, segun sean las π_i de cada alumno, mas ó menos variable es $\hat{\Delta}$; si Δ_o (usualmente Δ_o=0,5) señala el aprobado, la variable que serviria para ordenar a los alumnos de mejor a peor seria el valor de la t_{exp} del test utilizado.

Cuando los tamaños n_i son todos iguales, las variables "suma del numero de aciertos" y "suma de las proporciones de aciertos" se convierten en equivalentes. De hecho la forma de los test se simplifican bastante, como muestran las (3.40) y (3.44). Si en la (3.42) ponemos Δ_o=0,5 (lo usual para el aprobado), la t_{exp} comienza a ser positiva cuando $x = \sum x_{ii} > \bar{n}(k+1)/2$, y tal limite es el que se utiliza usualmente para aprobar a un alumno (si el alumno conoce la mitad de la asignatura, contestara bien, en promedio, a la mitad de las preguntas $-\bar{n}\ k/2-$ por sus conocimientos, y a la fracción k del resto- $\bar{n}$ k/2k- por azar, lo que suma $\bar{n}(k+1)/2$) sin tener en cuenta el denominador de la t_{exp}. Tal decisión está basada en la suposición de que todas las π_i son iguales a $1/k$, lo que no tiene por qué suceder en todos los alumnos; si ello fuera asi, la varianza (3.33) seria $\sigma^2=(k+1)/4k(k-1)\bar{n}$, por lo que ordenar a los alumnos en razon a su valor $t_{exp}=(\hat{\Delta}-½)/\sigma$ es equivalente a ordenarlos en razón a $\hat{\Delta}$ ó x. Supongamos que aceptamos los valores $\Delta_o=½$, $\pi_i=1/k$ y $n_i=\bar{n}$. En tal caso la cantidad de test es la anterior que, puesta explicitamente, será:

$$t_{exp} = \frac{x - \bar{n}(K+1)/2}{\sqrt{\dfrac{\bar{n}\ (K+1)\ (K-1)}{4K}}} \tag{3.55}$$

ó, si suponemos ademas que k=5 y $\bar{n}$=20, sera $t_{exp}=(x-60)/\sqrt{24}$. Ahora bien, un profesor estricto que, utilizando el test pata las (3.36), deseara aprobar indebidamente como maximo a un α =5% de los alumnos, aprobara a todos aquellos que verifiquen $x \geq 69$; un profesor bondadoso que, utilizando el test para los (3.38), deseara suspender indebidamente como maximo a α =1% de los alumnos, aprobara a todos aquellos que verifiquen $x \geq 49$. Lo usual es un criterio similar al primero. En las condiciones anteriores suele emplearse en la practica el criterio de aprobar cuando $x \geq 65$, lo que equivale a contrastar las (3.37) y aprobar indebidamente a, aproximadamente, un maximo de un 15% de los alumnos que conocen menos de la mitad de la asignatura. Sin embargo todos estos razonamientos estan construidos pensando que todos los alumnos adoptan el criterio de $\pi_i=1/k$, criterio que seguramente no sera cierto en general. Por derivación de la (3.55), para n_i iguales, es inmediato probar que dicha varianza es maxima cuando $\pi_i=1/k(\forall i)$; con ello está claro que la t_{exp} calculada para ese alumno "ideal" es siempre menor o igual que la calculada en cada alumno "real", alumno que podria resultar aprobado si se le tiene en cuenta su criterio particular de elección de las π_i. Una vez mas se concluye que, aun con tamaños muestrales iguales, es inadecuada la asignación de una cota unica para el "numero x de respuestas acertadas" con el fin de aprobar o no a un alumno. Desde luego, cuando ademas los tamaños muestrales son distintos, mas inadecuado es el limite unico. Finalmente, puesto que la derivación de la varianza respecto de las n_i señala que se alcanza el minimo valor de la misma cuando $n_1=n_2=...=n_k$ (como era de esperar), los examenes mas discriminatorios son aquellos en que los tamaños muestrales son iguales (ó al menos aproximadamente iguales, para evitar dar ese dato al alumno).

Cuando no se deseen test, sino solo estimar el grado de conocimientos del alumno, se disponen de dos procedimientos simples alternativos reseñados en el apartado 3.3.2.

Finalmente, la excesiva amplitud de los intervalos para Δ (puesta de manifiesto con el ejemplo de la Tabla 3.5) calculados con tamaños muestrales, n, de los usados normalmente en la practica, pone de manifiesto las pocas garantias que ofrece un examen de tipo test para asignar calificaciones. Mas adelante se insistira en el tema bajo otro punto de vista.

3.3.8. Aproximaciones clasicas al problema

La literatura Psicometrica esta llena de alusiones al problema de los tests de elección múltiple, pero las mismas se limitan a plantear modelos y dar estimaciones puntuales del grado de conocimientos del alumno (Δ en nuestra notacion), grado de dificultad de las preguntas, etc., no planteando problemas de inferencia (intervalos y test) tal y como aquí se ha hecho.

El criterio de evaluar un examen en base al número total de preguntas acertadas es el clasico en tal literatura, siendo preconizado por la mayoria de los modelos de rasgo latente (Lord and Movick (1968)). En nuestro caso Δ seria el rasgo latente.

El mas simple de los modelos psicométricos es el llamado FS-1-C por Hutchinson (1982) dentro del grupo de los modelos de la teoria de estados finitos, y se corresponde con el modelo actual si obligamos al alumno a que responda a todas las preguntas. La diferencia esencial entre ambos es que en nuestro caso las π_i pueden ser cualesquiera, en tanto que en el suyo se supone que π_i=1/k siempre. Creemos que nuestra hipótesis es mas realista por cuanto cada alumno tendra una tendencia particular a responder al azar unas letras ó números de alternativas mas que otros; de hecho el alumno no va al examen con unas tablas de números de azar bajo el brazo. Si acaso un determinado alumno es capaz de imitar bien al azar (π_i=1/k), el modelo lo detectara en sus estimaciones y no habra problema.

Tales diferencias en los criterios de asignación de las π_i son las que ocasionan los diferentes estimadores de ambas teorias (estimando Δ en base a la suma de las proporciones de aciertos -teoria actual- ó en base a la suma total de aciertos -teoria clasica-). El peligro de adoptar la teoria clasica ya ha sido reseñado repetidamente. Asi, bajo la teoria clasica,Lord y Novick(1968) dan el estimador usual:

$$\tilde{\Delta} = \frac{kx - n}{n(k-1)} \tag{3.56}$$

que prueban coincide con el de maxima verosimilitud, dado que el numero total de aciertos es x; el estimador actual, dado por la (3.29), coincide con aquel solo en el caso de n_i iguales y $\Delta \geq 0$. Para el ejemplo numerico citado en 3.2.2., seran $\tilde{\Delta} = (2x74-100)/100x1 = 0,48$ y $\hat{\Delta} = \{(72/80) + (2/20) - 1\}/(2-1) = 0$, lo que pone de manifiesto lo inadecuado del modelo clasico (pues el alumno realmente no sabia nada).

3.4. OPTIMIZACION EN LA EVALUACION DE TEST DE ELECCION MULTIPLE

En la evaluación de un test de elección múltiple, bajo el modelo aquí descrito, es inmediato que surjan preguntas del estilo de ¿cual de los tres test descritos debe aplicarse?, ¿de que número de preguntas debe constar el examen para una potencia dada?, ¿cuantas respuestas alternativas conviene que tenga cada pregunta?, etc. Esta sección está dedicada a ellas.

A tal efecto, conviene repetir la (3.33) en un modo mas explicito:

$$\sigma^2(\Delta) = \frac{1}{(k-1)^2}\sum\frac{\{\Delta+(1-\Delta)\pi_i\}\{1-\Delta-(1-\Delta)\pi_i\}}{n_i} \quad \text{si } \Delta \geq 0 \qquad (3.57)$$

$$= \sum\frac{(1+\Delta)\pi_i\{1-(1+\Delta)\pi_i\}}{n_i} \quad \Delta < 0$$

Por otro lado, y puesto que los test a plantear se referiran ordinariamente a valores de Δ_o mayores ó iguales que cero, las discusiones que siguen se refieren solo a tal situación.

Finalmente, conviene recordar que se van a manejar los tres criterios de test

$$t_{exp} = \frac{\hat{\Delta}-\Delta_o}{\hat{\sigma}(\Delta_o)} \;,, \quad t'_{exp} = \frac{\hat{\Delta}-\Delta_o}{\hat{\sigma}(\hat{\Delta})} \;,, \quad t''_{exp} = \frac{\hat{\Delta}-\Delta_o}{\hat{\theta}^{-1}(\Delta_o)} \qquad (3.53)'$$

con $\hat{\sigma}^2(\Delta)$ la (3.34) para el caso $\Delta \geq 0$ y

$$\hat{\theta}^2(\Delta)= \frac{a_i n_i (n-n_i)}{\{a_i+(n-n_i)\Delta\}\{a_i-n_i\Delta\}} \tag{3.59}$$

3.4.1. Potencias de los tests y comparación de los mismos (Aportación)

El objetivo de este apartado es calcular las potencias de los tres test propuestos para asi seleccionar el optimo. Con tal fin convengamos -como es usual- que las cantidades estimadas $\hat{\sigma}^2(\Delta_o)$, $\hat{\sigma}^2(\hat{\Delta})$ y $\hat{\theta}^2(\Delta_o)$ que aparecen en los tres tests toman por valor medio, para una alternativa Δ_1, lo que resulte de sustituir en ella cada a_i y cada $\hat{\Delta}$ por sus valores medios. Como en base al modelo:

$$\begin{aligned} E(\hat{\Delta}|\Delta_1) &= \Delta_1 && \text{si } \Delta_1 \geq 0 \\ &= \frac{\Delta_1}{k-1} && \text{si } \Delta_1 < 0 \end{aligned} \tag{3.60}$$

$$\begin{aligned} E(a_i|\Delta_1) &= n(1-\Delta_1)\pi_i + n_i\Delta_1 && \text{si } \Delta_1 \geq 0 \\ &= n(1+\Delta_1)\pi_i - \frac{\Delta_1}{k-1}(n-n_i) && \text{si } \Delta_1 < 0 \end{aligned} \tag{3.61}$$

entonces:

$$E\{\hat{\sigma}^2(\Delta_o)|\Delta_1\} \simeq \sigma^2(\Delta_o|\Delta_1)=$$

$$=\sum_i \frac{\{n_i(\Delta_1-\Delta_o)+n\Delta_o+n(1-\Delta_1)\pi_i\}\{n(1-\Delta_o)-n_i(\Delta_1-\Delta_o)-n(1-\Delta_1)\pi_i\}}{n^2(k-1)^2 n_i} \quad \text{si } \Delta_1 \geq 0$$

$$=\sum_i \frac{\{n(1+\Delta_1)(k-1)\pi_i+((k-1)\Delta_o-\Delta_1)(n-n_i)\}\{n(k-1)-n(k-1)(1+\Delta_1)\pi_i-((k-1)\Delta_o-\Delta_1)(n-n_i)\}}{n^2(k-1)^4 n_i} \quad \text{si } \Delta_1 < 0 \tag{3.62}$$

$$\begin{aligned} E\{\hat{\sigma}^2(\hat{\Delta})|\Delta_1\} &\simeq \sigma^2(\hat{\Delta}|\Delta_1) = \sigma^2(\Delta_1) && \text{si } \Delta_1 \geq 0 \\ &= \sigma^2(\Delta_1)/(k-1)^2 && \text{si } \Delta_1 < 0 \end{aligned} \tag{3.63}$$

$$E\{\hat{\theta}^2(\Delta_o)|\Delta_1\} \simeq \theta^2(\Delta_o|\Delta_1) =$$

$$=\sum \frac{n_i(n-n_i)\{n_i\Delta_1+n(1-\Delta_1)\pi_i\}}{\{n_i(\Delta_1-\Delta_o)+n\Delta_o+n(1-\Delta_1)\pi_i\}\{n_i(\Delta_1-\Delta_o)+n(1-\Delta_1)\pi_i\}} \quad \text{si } \Delta_1 \geq 0$$

(3.64)

$$=\sum \frac{(k-1)n_i(n-n_i)\{n(k-1)(1+\Delta_1)\pi_i - \Delta_1(n-n_i)\}}{\{n(k-1)(1+\Delta_1)\pi_i+(n-n_i)((k-1)\Delta_o-\Delta_1)\}\{n(k-1)(1+\Delta_1)\pi_i-\Delta_1(n-n_i)-(k-1)\Delta_o n_i\}}$$

$$\text{si } \Delta_1 < 0$$

Si el test va a ser de una cola a la derecha($H_1 \equiv \Delta > \Delta_0$), la potencia del test t para la alternativa $\Delta_1 \geq 0$ sera:

$$P(\Delta_1) = P\left\{t_{exp} = \frac{\hat{\Delta}-\Delta_0}{\hat{\sigma}(\Delta_0)} > t_\alpha \mid \Delta = \Delta_1\right\} \simeq P\left\{\frac{\hat{\Delta}-\Delta_0}{\sigma(\Delta_0|\Delta_1)} > t_\alpha \mid \Delta = \Delta_1\right\} =$$

$$= P\left\{\frac{\hat{\Delta}-\Delta_1}{\sigma(\Delta_1)} > \frac{\Delta_0-\Delta_1}{\sigma(\Delta_1)} + t_\alpha\frac{\sigma(\Delta_0|\Delta_1)}{\sigma(\Delta_1)}\right\} =$$

$$= \phi\left\{\frac{\Delta_1-\Delta_0}{\sigma(\Delta_1)} - t_\alpha\frac{\sigma(\Delta_0|\Delta_1)}{\sigma(\Delta_1)}\right\} \quad (3.65)$$

con $\emptyset(z)$ la función de distribución de la normal típica y t_α tal que $\emptyset(t_\alpha)=1-\alpha$. De igual modo con los otros dos test.

Cuando la hipotesis alternativa es $H_1 \equiv \Delta < \Delta_o$, la potencia para alternativas $\Delta_1 \geq 0$ es la de antes cambiando $\Delta_1-\Delta_o$ por $\Delta_o-\Delta_1$. Con ello, si P_2 es la potencia para $H_1 \equiv \Delta > \Delta_o$ y P_1 lo es para $H_1 \equiv \Delta < \Delta_o$, entonces para alternativas $\Delta_1 \geq 0$, sera:

$$P_i(\Delta_1) = \emptyset\left\{(-1)^i\frac{\Delta_1-\Delta_0}{\sigma(\Delta_1)} - t_\alpha\frac{\psi(\Delta_o|\Delta_1)}{\sigma(\Delta_1)}\right\} \quad (3.66)$$

surgiendo las potencias $P_i(\Delta_1)$, $P'_1(\Delta_1)$ y $P''_i(\Delta_1)$, de los test t, t' y t'', sustituyendo $\psi(\Delta_o/\Delta_1)$ por $\sigma(\Delta_o/\Delta_1)$, $\sigma(\hat{\Delta}/\Delta_1)=\sigma(\Delta_1)$ y $\theta^{-1}(\Delta_o/\Delta_1)$ respectivamente.

Cuando $\Delta_1 < 0$ (lo que solo puede ocurrir en la alternativa $H_1 \equiv \Delta < \Delta_o$) las potencias P_1 se ven alteradas asi:

$$P_1(\Delta_1) = \emptyset\left\{-\frac{\Delta_1-\Delta_0(k-1)}{\sigma(\Delta_1)} - t_\alpha(k-1)\frac{\psi(\Delta_o|\Delta_1)}{\sigma(\Delta_1)}\right\} \tag{3.67}$$

Finalmente, para un test de dos colas al error 2α, las potencias seran $P=P_1+P_2$ en cada caso, en donde $P_2(\Delta_1)$, cuando $\Delta_1 < 0$, viene dada por la (3.67) calculando el primer signo - por +.

En particular, cuando $\Delta_1=\Delta_o \geq 0$ se tiene que $\sigma^2(\Delta_o|\Delta_o)=\sigma^2(\hat{\Delta}|\Delta_o)=\sigma^2(\Delta_o)=\sigma^2(\Delta_1)$ por lo que $P_i(\Delta_o)=P_i(\Delta_o)=\alpha$ y los tamaños de los test t y t' son el α nominal. No ocurre asi con el test t'', por lo que convendria calcular su error real para ver su discrepancia con el nominal; ahora:

$$P_i''(\Delta_o) = \emptyset\left\{-\frac{t_\alpha}{\theta(\Delta_0|\Delta_0)\,\sigma(\Delta_0)}\right\} \tag{3.68}$$

$$\theta^2(\Delta_0|\Delta_0) = \frac{1}{n^2}\sum\frac{n_i(n-n_i)\left\{n_i\Delta_0+n(1-\Delta_0)\pi_i\right\}}{\left\{\Delta_0+(1-\Delta_0)\pi_i\right\}(1-\Delta_0)\pi_i} \tag{3.69}$$

El caso particular $\Delta_o=0$ ($H_o \equiv \Delta=0$ ó caso de independencia) simplifica algo las formulas. Ahora:

$$\sigma^2(0|\Delta_1) = \frac{\left\{n_i\Delta_1+n(1-\Delta_1)\pi_i\right\}\left\{n-n_i\Delta_1-n(1-\Delta_1)\pi_i\right\}}{n_i n^2(k-1)^2} \quad \text{si } \Delta_1\geq 0$$

$$= \sum\frac{\left\{n(1+\Delta_1)(k-1)\pi_i-\Delta_1(n-n_i)\right\}\left\{n(k-1)-n(k-1)(1+\Delta_1)\pi_i+\Delta_1(n-n_i)\right\}}{n_i n^2(k-1)^4} \quad \text{si } \Delta_1<0 \tag{3.70}$$

$$\theta^2(0|\Delta_1) = \sum\frac{n_i(n-n_i)}{n_i\Delta_1+n(1-\Delta_1)\pi_i} \quad (\text{si } \Delta_1\geq 0)$$

$$= \sum\frac{n_i(n-n_i)(k-1)}{n(k-1)(1+\Delta_1)\pi_i-\Delta_1(n-n_i)} \quad (\text{si } \Delta_1<0) \tag{3.71}$$

y para calcular el tamaño del test:

$$\sigma^2(0)=\sigma^2(0|0)=\frac{1}{(k-1)^2}\sum\frac{\pi_i(1-\pi_i)}{n_i} \qquad (3.72)$$

$$\theta^2(0|0)=\sum\frac{n_i(n-n_i)}{n\pi_i} \qquad (3.73)$$

A efectos comparativos, y puesto que los numeradores de las tres cantidades t_{exp} son el mismo, alcanzará mayor potencia aquel test que tenga un menor valor medio del denominador, es decir aquel test -t, t' ó t''- que alcance un menor valor de $\psi(\Delta_o|\Delta_1)$.

Con el fin de seleccionar el test mas potente, se han calculado las potencias y tamaños de los tres test para valores k=2, 3 y 5, α= 5% y 1%, n=90 y 150, $\pi_i=2ik^{-1}(k+1)^{-1}$, k^{-1} y $2(k+1-i)k^{-1}(k+1)^{-1}$, y n_i= nk^{-1} y $2ink^{-1}(k+1)^{-1}$, que son representativos de un amplio espectro, y para alternativas $\Delta_1<\Delta_o$ (cola izquierda), $\Delta_1>\Delta_o$ (cola derecha) y $\Delta_1\neq\Delta_o$ (dos colas). Las Tablas 3.6 a la 3.11 son una selección representativa de las 18 tablas citadas, pues el comportamiento de los test es similar en el resto. La potencia que figura para el test de dos colas es la potencia media en $-\Delta_1$ y $+\Delta_1$, pues los resultados comparativos de potencia varian segun cual sea la alternativa considerada.

El análisis de las Tablas 3.6 a la 3.11 permite concluir que los tamaños de los test t y t' son siempre el error nominal α (como ya sabiamos) y que el test t'' es tambien de tamaño α solo cuando k=2 y Δ_o=0, siendo moderadamente distinto de α cuando k=2 y $\Delta_o>0$, y pudiendo llegar a ser extremadamente distinto de α en los demás casos. En todo caso el tamaño de t'' es α si las π_i y las n_i son iguales. En consecuencia el test t'' debe descartarse siempre, salvo cuando k=2 y Δ_o=0, por ser un test sesgado. Finalmente, la comparación de las potencias de t y t' (y t'' en el caso citado) permiten concluir que el test mas potente es, en cada caso el indicado en la Tabla 3.12. Ella implica que, en el caso de la bioequivalencia en tablas 2x2, debe preferirse el test t para colas a la izquierda(que son las usuales).

TABLA 3.6

POTENCIAS DE LOS DIFERENTES TESTS: t(primer resultado), t'(segundo resultado) y t''(tercer resultado). $\alpha = 0,05$ k = 2 COLA IZQUIERDA.

		n =90					n =150				
	n_i =	45;45		30;60			75;75		50;100		
	Δ_1	(1)	(2)	(1)	(2)	(3)	(1)	(2)	(1)	(2)	(3)
0	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0	-.1	.257	.242	.231	.226	.247	.358	.336	.323	.311	.340
	-.1	.260	.245	.245	.228	.238	.362	.339	.339	.314	.330
	-.1	.257	.242	.231	.226	.247	.358	.336	.323	.311	.340
0	-.2	.630	.602	.576	.560	.598	.820	.794	.775	.752	.785
	-.2	.643	.615	.615	.572	.585	.830	.804	.804	.762	.775
	-.2	.630	.602	.576	.560	.598	.820	.794	.775	.752	.785
3	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.052	.050	.052	.050	.052	.052	.050	.052	.050	.052
3	.2	.277	.263	.244	.248	.280	.383	.363	.336	.339	.381
	.2	.262	.249	.237	.232	.249	.365	.346	.329	.320	.346
	.2	.287	.263	.250	.248	.291	.394	.363	.343	.339	.394
3	.1	.666	.629	.593	.593	.669	.844	.812	.782	.776	.840
	.1	.639	.603	.587	.561	.603	.826	.793	.777	.751	.793
	.1	.680	.629	.601	.594	.688	.853	.812	.788	.777	.852
5	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.055	.050	.054	.050	.054	.055	.050	.054	.050	.054
5	.4	.312	.302	.274	.286	.322	.428	.414	.374	.389	.436
	.4	.280	.271	.249	.252	.271	.391	.379	.346	.350	.379
	.4	.335	.302	.286	.286	.351	.453	.414	.389	.390	.467
	.3	.715	.690	.639	.659	.732	.878	.859	.817	.830	.885
	.3	.658	.635	.594	.591	.635	.842	.822	.784	.781	.822
	.3	.743	.690	.653	.661	.775	.895	.859	.827	.831	.909
	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.057	.050	.056	.050	.056	.057	.050	.056	.050	.056
	.6	.397	.389	.348	.372	.420	.536	.525	.470	.500	.557
	.6	.329	.323	.289	.299	.323	.463	.455	.405	.420	.455
	.6	.440	.389	.368	.374	.482	.579	.525	.491	.502	.618
	.5	.813	.798	.742	.776	.842	.938	.929	.892	.913	.950
	.5	.722	.707	.649	.663	.707	.892	.882	.835	.847	.882
	.5	.855	.798	.758	.782	.910	.956	.929	.901	.916	.976

(1) $\pi_i = 2i / K(K+1)$ (2) $\pi_i = 1/K$ (3) $\pi_i = 2(K+1-i)/K(K+1)$

TABLA 3.7

POTENCIAS DE LOS DIFERENTES TESTS: t(primer resultado),t'(segundo resultado) y t''(tercer resultado). α = 0,05 k = 2 COLA DERECHA.

		n=90					n=150				
n_i=		45;45		30;60			75;75		50;100		
λ_0	λ_1	(1)	(2)	(1)	(2)	(3)	(1)	(2)	(1)	(2)	(3)
.0	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
.0	.1	.257	.242	.247	.226	.231	.358	.336	.340	.311	.323
	.1	.260	.245	.238	.228	.245	.362	.339	.330	.314	.339
	.1	.257	.242	.247	.226	.231	.358	.336	.340	.311	.323
.0	.2	.630	.602	.598	.560	.576	.820	.794	.785	.752	.775
	.2	.643	.615	.585	.572	.615	.830	.804	.775	.762	.804
	.2	.630	.602	.598	.560	.576	.820	.794	.785	.752	.775
.3	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.052	.050	.052	.050	.052	.052	.050	.052	.050	.052
.3	.4	.256	.249	.235	.229	.236	.365	.354	.329	.324	.338
	.4	.280	.271	.249	.252	.271	.391	.379	.346	.350	.379
	.4	.262	.249	.241	.229	.239	.371	.354	.337	.324	.342
.3	.5	.661	.648	.601	.597	.618	.856	.845	.802	.801	.826
	.5	.722	.707	.649	.663	.707	.892	.882	.835	.847	.882
	.5	.665	.648	.610	.597	.620	.859	.845	.807	.802	.827
.5	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.055	.050	.054	.050	.054	.055	.050	.054	.050	.054
.5	.6	.280	.276	.249	.250	.259	.408	.401	.359	.364	.381
	.6	.329	.323	.289	.299	.323	.463	.455	.405	.420	.455
	.6	.291	.276	.263	.251	.264	.420	.401	.375	.364	.387
[illegible]	.7	.752	.746	.676	.687	.712	.928	.924	.877	.888	.909
	.7	.850	.844	.781	.805	.844	.966	.963	.930	.944	.963
	.7	.758	.746	.693	.689	.712	.931	.924	.887	.889	.909
[illegible]	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.057	.050	.056	.050	.056	.057	.050	.056	.050	.056
[illegible]	.8	.356	.353	.303	.310	.322	.537	.533	.461	.476	.500
	.8	.479	.475	.415	.439	.475	.659	.654	.580	.610	.654
	.8	.368	.353	.328	.313	.325	.550	.533	.489	.479	.503
[illegible]	.9	.952	.951	.895	.910	.927	.998	.998	.992	.994	.997
	.9	.997	.997	.988	.993	.997	1.000	1.000	1.000	1.000	1.000
	.9	.953	.951	.911	.916	.929	.998	.998	.994	.995	.997

(1) $\pi_i = 2i / K(K+1)$ (2) $\pi = 1/K$ (3) [illegible]

TABLA 3.8

POTENCIAS DE LOS DIFERENTES TESTS: t(primer resultado), t'(segundo resultado) y t''(tercer resultado). $\alpha = 0{,}05$ k = 2 DOS COLAS.

		n =90					n =150				
n_1=		45;45		30;60			75;75		50;100		
Δ_0	Δ_1	(1)	(2)	(1)	(2)	(3)	(1)	(2)	(1)	(2)	(3)
.0	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
.0	.1	.167	.156	.154	.144	.154	.249	.231	.227	.210	.227
	.1	.170	.159	.156	.146	.157	.253	.234	.230	.213	.230
	.1	.167	.156	.154	.144	.154	.249	.231	.227	.210	.227
.0	.2	.504	.475	.459	.433	.460	.724	.691	.674	.641	.674
	.2	.521	.491	.476	.447	.476	.738	.705	.688	.654	.688
	.2	.504	.475	.459	.433	.460	.724	.691	.674	.641	.674
.3	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.053	.050	.053	.050	.053	.053	.050	.053	.050	.053
.3	.4	.175	.167	.154	.154	.169	.262	.249	.227	.226	.251
	.4	.179	.171	.158	.157	.171	.266	.253	.232	.230	.253
	.4	.182	.167	.160	.154	.176	.271	.249	.235	.226	.259
.3	.5	.539	.512	.468	.467	.519	.762	.734	.687	.684	.740
	.5	.562	.535	.495	.489	.535	.777	.750	.709	.701	.750
	.5	.551	.512	.478	.468	.532	.771	.734	.695	.685	.750
.5	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.057	.050	.056	.050	.056	.057	.050	.056	.050	.056
.5	.6	.198	.192	.171	.176	.195	.300	.291	.255	.264	.293
	.6	.205	.200	.178	.183	.200	.310	.301	.265	.273	.301
	.6	.214	.192	.183	.176	.212	.320	.291	.271	.264	.313
.5	.7	.615	.597	.529	.545	.601	.835	.819	.757	.773	.825
	.7	.651	.634	.573	.587	.634	.844	.829	.780	.788	.829
	.7	.638	.597	.549	.548	.631	.851	.819	.772	.775	.844
.7	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.060	.050	.059	.050	.059	.060	.050	.059	.050	.059
.7	.8	.262	.257	.220	.233	.260	.407	.400	.340	.361	.400
	.8	.291	.286	.246	.260	.286	.440	.433	.372	.394	.433
	.8	.292	.257	.242	.236	.297	.440	.400	.367	.363	.439
.7	.9	.805	.795	.711	.742	.799	.946	.940	.903	.922	.953
	.9	.800	.791	.751	.763	.791	.911	.904	.872	.880	.904
	.9	.837	.795	.738	.754	.850	.962	.940	.913	.926	.975

(1) $\pi_i = 2i / K(K+1)$ (2) $\pi_i = 1/K$ (3) $\pi_i = 2(K+1-i)/K(K+1)$

TABLA 3.9

POTENCIAS DE LOS DIFERENTES TESTS: t(primer resultado), t'(segundo resultado) y t''(tercer resultado). $\alpha = 0,05$ k = 5 COLA IZQUIERDA.

		n = 90					n = 150					
	n_i =	18;18;18;18;18		6;12;18;24;30			30;30;30;30;30		10;20;30;40;50			
Δ_0	Δ_1	(1)	(2)	(1)	(2)	(3)	(1)	(2)	(1)	(2)	(3)	
.0	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	
	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	
	.0	.074	.050	.050	.074	.145	.074	.050	.050	.074	.145	
.0	-.1	.113	.111	.103	.098	.096	.144	.141	.131	.121	.117	
	-.1	.127	.125	.125	.111	.105	.161	.157	.157	.136	.128	
	-.1	.146	.111	.104	.140	.228	.182	.141	.133	.169	.264	
.0	-.2	.229	.224	.196	.179	.173	.332	.323	.290	.254	.241	
	-.2	.279	.271	.271	.223	.204	.390	.379	.379	.307	.278	
	-.2	.266	.224	.204	.246	.343	.376	.323	.300	.334	.434	
.3	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	
	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	
	.3	.063	.050	.066	.079	.125	.063	.050	.066	.079	.125	
.3	.2	.460	.452	.371	.362	.368	.642	.633	.532	.516	.521	
	.2	.483	.475	.407	.384	.379	.664	.654	.570	.539	.532	
	.2	.534	.452	.429	.466	.655	.709	.633	.591	.621	.785	
.3	.1	.947	.941	.885	.857	.855	.996	.995	.983	.972	.970	
	.1	.964	.959	.931	.893	.876	.997	.997	.992	.981	.976	
	.1	.975	.941	.914	.916	1.000	.998	.995	.989	.987	1.000	
.5	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	
	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	
	.5	.068	.050	.073	.083	.134	.068	.050	.073	.083	.134	
.5	.4	.467	.463	.372	.378	.392	.641	.636	.521	.527	.543	
	.4	.454	.450	.367	.364	.369	.629	.624	.516	.512	.519	
	.4	.576	.463	.445	.489	.756	.738	.636	.595	.637	.859	
.5	.3	.928	.924	.841	.841	.854	.991	.990	.962	.961	.965	
	.3	.925	.921	.846	.833	.834	.990	.989	.964	.958	.958	
	.3	.982	.924	.883	.907	1.000	.999	.990	.975	.981	1.000	
.7	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	
	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	
	.7	.073	.050	.077	.086	.145	.073	.050	.077	.086	.145	
.7	.6	.551	.548	.443	.459	.482	.727	.724	.600	.618	.643	
	.6	.495	.493	.392	.399	.412	.678	.676	.550	.559	.575	
	.6	.732	.548	.523	.582	1.000	.863	.724	.676	.730	1.000	
.7	.5	.954	.952	.881	.892	.909	.995	.995	.974	.978	.983	
	.5	.929	.927	.838	.842	.853	.991	.991	.960	.962	.967	
	.5	1.000	.952	.915	.946	.917	1.000	.995	.984	.991	.985	

(1) $\pi_i = 2i / K(K+1)$ (2) $\pi_i = 1/K$ (3) $\pi_i = 2(K+1-i)/K(K+1)$

TABLA 3.10

POTENCIAS DE LOS DIFERENTES TESTS: t(primer resultado), t'(segundo resultado) y t''(tercer resultado). α = 0,05 k = 5 COLA DERECHA.

		n=90					n=150				
	n_i=	18;18;18;18;18		6;12;18;24;30			30;30;30;30;30		10;20;30;40;50		
Δ_0	Δ_1	(1)	(2)	(1)	(2)	(3)	(1)	(2)	(1)	(2)	(3)
.0	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.074	.050	.050	.074	.145	.074	.050	.050	.074	.145
.0	.1	.604	.589	.583	.498	.459	.778	.763	.747	.661	.620
	.1	.530	.518	.469	.420	.404	.718	.704	.647	.587	.565
	.1	.653	.589	.583	.558	.630	.815	.763	.747	.714	.771
.0	.2	.967	.963	.945	.913	.895	.997	.997	.992	.984	.979
	.2	.940	.935	.880	.855	.848	.994	.993	.977	.967	.964
	.2	.973	.963	.945	.931	.944	.998	.997	.992	.988	.991
.3	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.063	.050	.066	.079	.125	.063	.050	.066	.079	.125
.3	.4	.458	.454	.380	.369	.366	.633	.628	.530	.517	.516
	.4	.454	.450	.367	.364	.369	.629	.624	.516	.512	.519
	.4	.487	.454	.434	.453	.516	.660	.628	.584	.602	.664
.3	.5	.926	.924	.843	.838	.840	.991	.990	.962	.961	.962
	.5	.929	.927	.838	.842	.853	.991	.991	.960	.962	.967
	.5	.932	.924	.875	.882	.906	.992	.990	.972	.975	.982
.5	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.068	.050	.073	.083	.134	.068	.050	.073	.083	.134
.5	.6	.462	.460	.366	.367	.371	.648	.645	.523	.525	.534
	.6	.495	.493	.392	.399	.412	.678	.676	.550	.559	.575
	.6	.497	.460	.442	.462	.529	.679	.645	.600	.622	.687
.5	.7	.953	.952	.868	.873	.882	.997	.997	.976	.979	.982
	.7	.972	.972	.908	.918	.930	.998	.998	.986	.989	.991
	.7	.957	.952	.908	.918	.936	.997	.997	.986	.988	.993
.7	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.073	.050	.077	.086	.145	.073	.050	.077	.086	.145
.7	.8	.562	.561	.428	.436	.446	.776	.775	.626	.638	.653
	.8	.666	.665	.532	.549	.570	.849	.848	.720	.738	.760
	.8	.593	.561	.529	.548	.607	.799	.775	.717	.737	.789
.7	.9	.999	.999	.981	.984	.988	1.000	1.000	1.000	1.000	1.000
	.9	1.000	1.000	.999	.999	1.000	1.000	1.000	1.000	1.000	1.000
	.9	.999	.999	.992	.994	.996	1.000	1.000	1.000	1.000	1.000

(1) $\pi_i = 2i / K(K+1)$ (2) $\pi = 1 / K$

TABLA 3.11

POTENCIAS DE LOS DIFERENTES TESTS: t(primer resultado), t'(segundo resultado) y t''(tercer resultado). α = 0,05 k = 5 DOS COLAS.

		n =90					n =150				
	n_i =	18;18;18;18;18		6;12;18;24;30			30;30;30;30;30		10;20;30;40;50		
Δ_0	Δ_1	(1)	(2)	(1)	(2)	(3)	(1)	(2)	(1)	(2)	(3)
.0	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.0	.085	.050	.050	.085	.208	.085	.050	.050	.085	.208
.0	.1	.280	.272	.269	.224	.204	.386	.376	.366	.313	.289
	.1	.243	.236	.213	.187	.178	.352	.343	.311	.273	.260
	.1	.326	.272	.270	.279	.380	.431	.376	.367	.370	.467
.0	.2	.541	.536	.513	.485	.469	.606	.602	.583	.564	.555
	.2	.539	.532	.492	.457	.446	.631	.625	.610	.572	.558
	.2	.564	.536	.516	.529	.594	.629	.602	.587	.607	.667
.3	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.3	.068	.050	.073	.093	.170	.068	.050	.073	.093	.170
.3	.4	.336	.331	.262	.254	.256	.513	.506	.404	.391	.393
	.4	.347	.341	.274	.263	.263	.524	.516	.418	.401	.401
	.4	.395	.331	.321	.355	.508	.574	.506	.472	.504	.651
.3	.5	.884	.877	.777	.756	.757	.984	.982	.944	.933	.933
	.5	.903	.897	.814	.789	.784	.987	.986	.953	.945	.943
	.5	.917	.877	.827	.840	.927	.989	.982	.961	.963	.983
.5	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.5	.075	.050	.083	.099	.187	.075	.050	.083	.099	.187
.5	.6	.341	.339	.257	.260	.268	.520	.516	.396	.400	.412
	.6	.353	.350	.268	.270	.277	.532	.528	.408	.411	.423
	.6	.425	.339	.336	.372	.583	.605	.516	.486	.525	.714
.5	.7	.890	.887	.764	.768	.784	.985	.984	.938	.939	.946
	.7	.908	.904	.803	.801	.810	.987	.986	.950	.948	.951
	.7	.943	.887	.832	.858	.944	.995	.984	.962	.971	.992
.7	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
	.7	.083	.050	.089	.104	.208	.083	.050	.089	.104	.208
.7	.8	.426	.425	.312	.323	.339	.638	.636	.484	.500	.521
	.8	.458	.457	.343	.355	.370	.661	.659	.515	.531	.552
	.8	.560	.425	.414	.459	.745	.752	.636	.591	.638	.846
.7	.9	.957	.956	.880	.892	.909	.995	.995	.975	.979	.984
	.9	.938	.936	.873	.876	.883	.990	.990	.962	.964	.968
	.9	.997	.956	.924	.951	.929	1.000	.995	.985	.993	.986

(1) $\pi_i = 2i / K(K+1)$ (2) $\pi_i = 1/K$ (3) $\pi_i = 2(K+1-i)/K(K+1)$

Tabla 3.12

En el interior de la tabla figura el test mas potente, de entre los ofrecidos en la expresión (3.58), para contrastar $H_o \equiv \Delta = \Delta_o$ en las condiciones indicadas en los margenes

	$H_1 \equiv \Delta < \Delta_o$		$H_1 \equiv \Delta > \Delta_o^c$		$H_1\ \Delta \neq \Delta_o$	
	$k=2$	$k>2$	$k=2$	$k>2$	$k=2$	$k>2$
$\Delta_o = 0$	t'	t'	t'	t	t'	t'
$\Delta_o > 0$	t	t	t'	t'	t'	t'

3.4.2. Tamaños de muestra

3.4.2.1. En intervalos de confianza (Aportación)

Si se conociera la $\sigma^2(\Delta)$ dada por la (3.33), un intervalo de confianza para Δ se dijo que era, en grandes muestras, $\Delta \in \hat{\Delta} \pm t_{\alpha/2}\ \sigma(\Delta)$. Supongamos que el objetivo es obtener el valor $\bar{n}=n_1=\ldots=n_k$ tal que $|\hat{\Delta}-\Delta| \leq \delta$ con una cierta confianza. En tal caso bastaria con resolver en $\bar{n}$ la igualdad:

$$t_{\alpha/2}\ \sigma(\Delta) = \delta \qquad (3.74)$$

Como la afirmación debe valernos para cualquier alumno, es obligatorio sustituir $\sigma(\Delta)$ por el maximo valor que pueda tomar. Derivando $\sigma^2(\Delta)$ en las π_i es inmediato probar que el maximo se alcanza cuando $\pi_i=1/k$, $\forall_i$, con lo que

$$\sigma^2(\Delta) = \frac{1}{\bar{n}k(k-1)}(1-\Delta)\left\{1+(k-1)\Delta\right\} \qquad (3.75)$$

(en el caso de $\Delta>0$ que es el de mayor interés: similarmente cuando $\Delta<0$) y asi, sustituyendo en la (3.74) y despejando $\bar{n}$:

$$\bar{n} = \frac{t^2_{\alpha/2}}{\delta^2 k(k-1)} (1-\Delta)\left\{1+(k-1)\Delta\right\} \tag{3.76}$$

expresión que, de no haber ninguna información acerca de Δ, habra de maximizarse en Δ. Por derivación es inmediato ver que el maximo se alcanza en

$$\Delta = \frac{1}{2}\cdot\frac{k-2}{k-1} \tag{3.77}$$

y asi, sustituyendo en la (3.76), queda:

$$\bar{n} = \frac{t^2_{\alpha/2}\cdot k}{4\,\delta^2 (k-1)^2} \tag{3.78}$$

En particular, para $\alpha=5\%$, $k=5$ y $\delta=0,1$, se tiene que $\bar{n} \simeq 30$ y el número total de preguntas del examen habran de ser de $n=150$ (si no hay información acerca de Δ).

3.4.2.2. En test de una cola (Aportación)

Supongamos se va a asignar una calificación como, por ejemplo, el aprobado ($\Delta_o=0,5$ digamos). Un profesor exigente contrastaria $H_o \equiv \Delta = \Delta_o$ (suspenso) conta $H_1 \equiv \Delta > \Delta_o$ (aprobado); un profesor bondadoso contrastaria $H_o \equiv \Delta = \Delta_o$ (aprobado) contra $H_1 \equiv \Delta < \Delta_o$ (suspenso); el primero busca razones suficientes para aprobar, ó, en otro caso, suspende; el segundo busca razones suficientes para suspender ó, en otro caso, aprueba. En realidad ninguna de las dos situaciones es satisfactoria por cuanto, si el tamaño de muestra -el número total de preguntas- no es lo suficientemente grande, habra demasiados alumnos que resulten suspendidos (con el primer test) ó aprobados (con el segundo) indebidamente. El modo de controlar ambos errores pasa por plantearse el tamaño adecuado de muestra para que, al realizar un test al error

α de tipo I, se obtenga un error β (dado) de tipo II, es decir, una potencia $1-\beta$. Con tal fin reescribamos las distintas varianzas promedio anteriores para valores iguales de n_i $(n_1=n_2=\ldots=n_k=\bar{n})$:

$$\sigma^2(\Delta_0|\Delta_1) = \frac{\Sigma\left\{\Delta_1+(k-1)\Delta_0+k(1-\Delta_1)\pi_i\right\}\left\{k-\Delta_1-(k-1)\Delta_0-k(1-\Delta_1)\pi_i\right\}}{\bar{n}\,k^2\,(k-1)^2} \quad \text{si } \Delta_1 \geq 0$$

$$= \sigma^2(\Delta_0|\,|\Delta_1|) \quad \text{si } \Delta_1 < 0 \qquad (3.79)$$

$$\sigma^2(\hat{\Delta}|\Delta_1) = \frac{1}{\bar{n}(k-1)^2}\Sigma\left\{\Delta_1+(1-\Delta_1)\pi_i\right\}\left\{1-\Delta_1-(1-\Delta_1)\pi_i\right\} = \sigma^2(\Delta_1) \quad \text{si } \Delta_1 \geq 0$$

$$= \frac{1}{\bar{n}(k-1)^2}\Sigma\left\{(1+\Delta_1)\pi_i\right\}\left\{1-(1+\Delta_1)\pi_i\right\} = \frac{\sigma^2(\Delta_1)}{(k-1)^2} \quad \text{si } \Delta_1 < 0 \qquad (3.80)$$

$$\theta^2(\Delta_0|\Delta_1) = \bar{n}(k-1)\Sigma\frac{\Delta_1 + k(1-\Delta_1)\pi_i}{\left\{\Delta_1+(k-1)\Delta_0+k(1-\Delta_1)\pi_i\right\}\left\{\Delta_1-\Delta_0+k(1-\Delta_1)\pi_i\right\}} \quad \text{si } \Delta_1 \geq 0 \qquad (3.81)$$

$$= \theta^2(\Delta_0|\,|\Delta_1|) \quad \text{si } \Delta_1 < 0$$

A efectos de calculo del tamaño de muestra $\bar{n}$ adecuado, las anteriores varianzas no tienen utilidad inmediata por cuanto ellas dependen de las π_i que, no solo son desconocidas, sino que varian de alumno a alumno. Para solventar el problema planteemos el tamaño adecuado de muestra para el caso mas desfavorable de selección de las π_i, es decir para aquella selección de las π_i que haga maximas las varianzas anteriores. Pero la derivación de tales varianzas respecto de π_h da lugar a dos terminos identicos -uno para π_h y otro para π_k- pero de signo contrario; de ahí que, en todos los casos, el maximo se da para el caso de $\pi_1=\pi_2=\ldots=\pi_k=1/k$, y, sustituyendo tales valores en las (3.79) a (3.81), se tiene que:

$$\sigma^2(\Delta_0|\Delta_1) = \frac{(1-\Delta_0)\left\{1+(k-1)\Delta_0\right\}}{\bar{n}k(k-1)} \qquad (3.82)$$

$$\sigma^2(\hat{\Delta}|\Delta_1) = \frac{(1-\Delta_1)\{1+(k-1)\Delta_1\}}{\bar{n}\,k\,(k-1)} = \sigma^2(\Delta_1) \qquad (\text{si } \Delta_1 \geq 0)$$

$$= \frac{(1+\Delta_1)\{k-1-\Delta_1\}}{\bar{n}\,k\,(k-1)^2} = \frac{\sigma^2(\Delta_1)}{(k-1)^2} \qquad (\text{si } \Delta_1 < 0) \qquad (3.83)$$

$$\vartheta^2(\Delta_0|\Delta_1) = \frac{\bar{n}\,k\,(k-1)}{(1+\Delta_0)\{(k-1)\,\Delta_0+1\}} \qquad (3.84)$$

luego $\sigma^2(\Delta_0|\Delta_1) = \vartheta^{-2}(\Delta_0|\Delta_1)$ siempre y, por tanto, las potencias de los test t y t'' para los casos de n_i y π_i iguales, deben coincidir. Tambien puede comprobarse que el tamaño del test es α en los tres casos. Todo esto es conforme con lo observado en las Tablas de potencia anteriores.

Con todo ello, el tamaño $\bar{n}$ adecuado de muestra se obtendra, para los test de una cola al error α, igualando las (3.66) y (3.67) a $(1-\beta)$- es decir, igualando a t_β la expresion tras el simbolo Ø- y despejando de ellas $\bar{n}$. Así, tras operar, tendremos:

$$\bar{n} = \frac{1}{k(k-1)}\left\{\frac{t_\alpha\sqrt{h(\Delta_0)}+t_\beta\sqrt{h(\Delta_1)}}{\Delta_1-\Delta_0}\right\}^2 \qquad \text{si } \Delta_1 \geq 0 \text{ y test t o t''} \qquad (3.85)$$

$$\bar{n} = \frac{h(\Delta_1)}{k(k-1)}\left\{\frac{t_\alpha+t_\beta}{\Delta_1-\Delta_0}\right\}^2 \qquad \text{si } \Delta_1 \geq 0 \text{ y test t'} \qquad (3.86)$$

$$\bar{n} = \frac{1}{k(k-1)}\left\{\frac{t_\alpha\sqrt{h(\Delta_0)}+t_\beta\sqrt{h(\Delta_1')}}{\Delta_1'-\Delta_0}\right\}^2 \qquad \text{si } \Delta_1 < 0 \text{ y test t o t''} \qquad (3.87)$$

$$\bar{n} = \frac{h(\Delta_1')}{k(k-1)}\left\{\frac{t_\alpha+t_\beta}{\Delta_1'-\Delta_0}\right\}^2 \qquad \text{si } \Delta_1 < 0 \text{ y test t'} \qquad (3.88)$$

en donde:

$$h(\Delta) = (1-\Delta)\{1+(k-1)\Delta\} \qquad (3.89)$$

$$\Delta'_1 = \Delta_1/(k-1) \qquad (3.90)$$

pudiendo observarse que las formulas del caso $\Delta_1 < 0$ son las del caso $\Delta_1 \geq 0$ cambiando Δ_1 por $\Delta_1/(k-1)$.

Por otro lado, y para el caso $\Delta_1 \geq 0$, sera $\bar{n}(t)=\bar{n}(t'') < \bar{n}(t')$ $\Longleftrightarrow h(\Delta_0) < h(\Delta_1) \Longleftrightarrow (\Delta_0-\Delta_1)\left\{(k-2)-(k-1)(\Delta_0+\Delta_1)\right\} < 0$ de modo que:

$$\bar{n}(t) = \bar{n}(t'') < \bar{n}(t') \Longleftrightarrow \begin{cases} \Delta_0 < \Delta_1 < \frac{k-2}{k-1} - \Delta_0 \\ \text{o} \\ \frac{k-2}{k-1} - \Delta_0 < \Delta_1 < \Delta_0 \end{cases} \quad (\text{si } \Delta_1 \geq 0) \qquad (3.91)$$

y similarmente

$$\bar{n}(t) = \bar{n}(t'') < \bar{n}(t') \Longleftrightarrow \begin{cases} (k-1)\Delta_0 < \Delta_1 < (k-2)-(k-1)\Delta_0 \\ \text{o} \\ (k-2)-(k-1)\Delta_0 < \Delta_1 < (k-1)\Delta_0 \end{cases} \quad (\text{si } \Delta_1 < 0) \qquad (3.92)$$

o, puesto que $\Delta_1 < 0$ y $\Delta_0 \geq 0$,

$$\bar{n}(t) = \bar{n}(t'') < \bar{n}(t') \Longleftrightarrow (k-2) - (k-1)\Delta_0 < \Delta_1 \quad (\text{si } \Delta_1 < 0) \qquad (3.93)$$

El análisis de las anteriores expresiones permite concluir que para el caso de n_i y π_i constantes, y para los valores usuales de k(desde 2 hasta 5)

1) Para asignar las calificaciones ordinarias (Δ_0=0,5; 0,7; 0,9), el test t' es el mas potente para la alternativa $H_1 \equiv \Delta > \Delta_0$, en tanto que los t y t'' lo son para $H_1 \equiv \Delta < \Delta_0$.

2) Para decidir la independencia (Δ_0=0):

a) Si k=2: siempre es preferible el test t'

b) Si k>2: es preferible el test t' para la alternativa $H_0 \equiv \Delta < 0$

lo que tambien concuerda con lo observado en las tablas de potencia.

Así, si un profesor riguroso va a contrastar el aprobado (H_o Δ=0,5 contra H_1: Δ>0,5) a un error α=5% y desea suspender indebidamente a solo un β=1% de los alumnos con conocimientos Δ_1=0,6, entonces empleara el test t' y el tamaño adecuado de muestra será, según la (3.86) $\bar{n} \simeq 108$ si el examen consta, como es lo usual, de k=5 alternativas. Con ello, el examen habra de constar de $n=5\bar{n}=540$ preguntas para contar con las garantias anteriores.

Si se bajan apreciablemente las condiciones, de modo que α=10% y β=5%, se obtiene $\bar{n}$=59 y n=295; siguen siendo precisas demasiadas preguntas.

3.4.2.3. <u>En test de dos colas</u> (<u>Aportación</u>)

Cuando se contrastan las calificaciones usuales, las alternativas de interes seran siempre positivas ($\Delta_1 \geq 0$), de modo que las (3.85) (3.86) valen de modo aproximado en los test de dos colas si se cambia α por $\alpha/2$. Sin embargo al ser $H_1 \equiv \Delta \neq \Delta_o$ la potencia ($1-\beta$) se solicitara tanto para valores Δ_1 a la derecha como a la izquierda de Δ_o, siendo usualmente del tipo $\Delta_1 = \Delta_o \pm \delta$, con $\delta>0$; ello implica que para $\Delta_1 = \Delta_o + \delta$ se utilizara el test t' y su tamaño de muestra vendra dado por la (3.86), en tanto que para $\Delta_1 = \Delta_o - \delta$ se utilizaran los tests t ó t'' y su tamaño de muestra vendra dado por la (3.85). El maximo de ambos seria el tamaño $\bar{n}$ buscado para la versión mas potente del test de dos colas: la mezcla de t' y los t y t''. Sin embargo ese no es el camino usual; lo comun es elegir el test t' ó los t ó t''. En ambos casos el tamaño de muestra vendria dado por

$$\bar{n} = \mathrm{Max}\left\{ \bar{n}(\Delta_o + \delta) \; ; \; \bar{n}(\Delta_o - \delta) \right\} \qquad (3.94)$$

con $\bar{n}(\Delta_o \pm \delta)$ calculado por la (3.85) para los test t ó t'' y por la (3.86) para el test t'. Tambien en ambos casos el $\bar{n}$ maximo es aquel que haga maxima la cantidad $h(\Delta_1)$, pero $h(\Delta_o+\delta) \geq h(\Delta_o-\delta) \Leftrightarrow \Delta_o \leq (k-2)/2(k-1)$, valor que siempre esta por debajo del 0,5 usual minimo. En consecuencia, para los test de dos colas habrá de tomarse siempre $h(\Delta_o-\delta)$ a efectos del tamaño de muestra y, en consecuencia, este sera el dado por las (3.85) y (3.86) (cambiando α por $\alpha/2$) en $\Delta_1 < \Delta_o$. De aqui se desprende que la potencia es mayor siempre en $\Delta_o + \delta$ que en $\Delta_o - \delta$, como ya se observo de las tablas de potencia.

Abordemos ahora el caso de la independencia (Δ_o=0) en que tanto una alternativa negativa como un test de dos colas, surgen de modo natural. Ahora, si se desean detectar alternativas $\pm\Delta_1$ (con $\Delta_1>0$) con potencia $(1-\beta)$, los tamaños muestrales adecuados se obtienen aproximadamente, cambiando α por $\alpha/2$ en las expresiones (3.85) a (3.88) y eligiendo el maximo de entre las (3.85) y (3.87) -en el caso de los test t y t''- ó de entre las (3.86) y (3.88)- en el caso del test t'-. Como quiera que $\bar{n}(+\Delta_1) \leq \bar{n}(-\Delta_1)$ en ambos casos, alcanzandose el igual cuando k=2, entonces las formulas a aplicar son las (3.87) y (3.88) cambiando α por $\alpha/2$ y haciendo Δ_o=0, es decir:

$$\bar{n} = \frac{1}{k}\left\{\frac{t_{\alpha/2}\sqrt{k-1} + t_\beta\sqrt{(1-\Delta_1)(k-1+\Delta_1)}}{\Delta_1}\right\}^2 \qquad (3.95)$$

si $\Delta_1 > 0$ y tests t o t''

$$\bar{n} = \frac{(1-\Delta_1)(k-1+\Delta_1)}{k}\left\{\frac{t_{\alpha/2}+t_\beta}{\Delta_1}\right\}^2 \qquad \text{si } \Delta_1>0 \text{ y test t'} \quad (3.96)$$

resultado que es coherente con el hecho de que los test detectan mas facilmente una alternativa $+\Delta_1$ que una $-\Delta_1$ (con $\Delta_1 > 0$). Observese que de nuevo estos resultados reiteran lo observado en las tablas de potencia de que $P(+\Delta) \geq P(-\Delta)$, alcanzandose el igual cuando k=2.

El planteamiento anterior es el adecuado cuando interesa detectar por igual un valor $+\Delta_1$, como uno $-\Delta_1$. Si a los valores negativos de Δ se les concede menos importancia, de modo que la potencia indicada se desea para alternativas $+\Delta_1$ y $-(k-1)\Delta_1$ con $\Delta_1 > 0$, entonces, sustituyendo Δ_1 por $-(k-1)\Delta_1$ en las (3.87) y (3.89)- caso $\Delta_0 = 0$- se obtienen las (3.85) y (3.86) respectivamente, con lo que la formula del tamaño de muestra para una u otra cola son coincidentes.

Respecto de las potencias del test de dos colas cabe afirmar lo ya señalado en el apartado 3.4.2.2. unos tests seran preferibles para una alternativa y otros para otra, existiendo unanimidad solo cuando k=2 en que el test t' es preferible para contrastar la independencia.

3.4.2.4. Caso particular de k=2 (Aportación)

Cuando k=2 las (3.85) y (3.86) coinciden, respectivamente, con las (3.87) y (3.88), con lo que:

$$\bar{n} = \frac{1}{2}\left\{\frac{t_\alpha\sqrt{1-\Delta_0^2} + t_\beta\sqrt{1-\Delta_1^2}}{\Delta_1 - \Delta_0}\right\} \qquad \text{si tests t o t''} \qquad (3.97)$$

$$\bar{n} = \frac{1-\Delta_1^2}{2}\left\{\frac{t_\alpha + t_\beta}{\Delta_1 - \Delta_0}\right\}^2 \qquad \text{si test t'} \qquad (3.98)$$

serán los tamaños de muestra para test de una cola, y esto con independencia de que Δ_1 sea positivo ó negativo. De nuevo se reiteran las conclusiones para la potencia hechas en el apartado 3.4.1. Para dos colas basta cambiar α por $\alpha/2$; en el caso de la independencia ($\Delta_o=0$) ya se ha dicho es preferible el test t', de modo que convendra utilizar la (3.98) con $\Delta_o=0$.

Así para contrastar $H_o \equiv \Delta = 0,5$ contra $H_1 \equiv \Delta > 0,5$ a unos errores $\alpha=5\%$ y $\beta=1\%$ para la alternativa $\Delta_1=0,6$, el test t' (que es el mas potente en ese caso) requiere un tamaño de muestra $\bar{n} \simeq 505$, por lo que el examen habra de constar de n=1010 preguntas en total.

3.4.2.5. Aproximaciones clásicas al problema

Ya se ha indicado antes que el planteamiento habitual en psicometria no suele incluir problemas de inferencia. Una ocasión en que si lo hacen es cuando tratan de determinar el número total de preguntas de que debe constar la prueba, pero ello bajo criterios muy particulares. Wilcox (1980) da una excelente revisión sobre el tema.

De un modo general la idea consiste en asumir un cierto valor critico Δ_o para el índice de conocimientos del alumno y decidir si el Δ real de un alumno es menor ó no que él en base a los resultados de su examen. Con tal fin el investigador ha de seleccionar un determinado estadistico ς relacionado con Δ (usualmente una funcion del numero total de preguntas acertadas) y determinar un valor ς_o ("marca de paso") de modo que si ς es la "marca del alumno"

$$\begin{array}{lll} \text{Si} & \varsigma < \varsigma_o & \text{decidir que } \Delta < \Delta_o \\ \text{Si} & \varsigma \geq \varsigma_o & \text{decidir que } \Delta \geq \Delta_o \end{array} \qquad (3.99)$$

El objetivo es determinar el número n de preguntas a realizar para que el criterio de decisión dado por la (3.99) tenga un determinado porcentaje de aciertos. Con tal fin es usual admitir que cuando $\Delta_o - \delta < \Delta < \Delta_o + \delta$, con δ fijado por el investigador, cualquier decisión es aceptable como correcta, y, en tal caso, se desea un n tal que la (3.99) de un porcentaje de aciertos mayor ó igual que $(1-\beta)$, con $(1-\beta)$ dado de antemano y tal que $0,5<1-\beta<1$ (se pone $1-\beta>0,5$ pues en otro caso un proceso de decisión al azar, sin necesidad de las observaciones del individuo, garantizan un $(1-\beta)=0,5$).

Con tal fin la literatura psicometrica utiliza diversos modelos, el mas conocido de los cuales consiste en tomar $\varrho=\tilde{\Delta}$ de la (3.56). En este formato Wilcox da la solución siguiente al problema.

Supongamos que $\Delta_o=0,8$ y que adoptamos el criterio de decidir que $\Delta<\Delta_o$ si $\hat{\Delta}<0,8$, y que $\Delta\geq\Delta_o$ si $\hat{\Delta}\geq 0,8$. Sea x_o el mas pequeño entero en que $\tilde{\Delta}\geq 0,8$ y sean $\Delta_1=\Delta_o-\delta$ y $\Delta_2=\Delta_o+\delta$, con δ dada de antemano e interpretada como mas arriba ($\delta=0,1$ por ejemplo). En tal caso, si $\pi=P\{$acertar una pregunta | el alumno no la conoce$\}$, seran:

$$p_1 = P\left\{A_i|R_i,\Delta_1\right\} = \Delta_1+\pi(1-\Delta_1) \qquad (3.100)$$
$$p_2 = P\left\{A_i|R_i,\Delta_2\right\} = \Delta_2+\pi(1-\Delta_2)$$

con lo que si p(DC) es la probabilidad de una decisión correcta:

$$P\left\{DC|\Delta_1\right\} = \sum_{x=0}^{x_o-1}\binom{n}{x}p_1^x(1-p_1)^{n-x} \geq 1-\beta$$

$$(3.101)$$

$$P\left\{DC|\Delta_2\right\} = \sum_{x=x_o}^{n}\binom{n}{x}p_2^x(1-p_2)^{n-x} \geq 1-\beta$$

de modo que el probema consistira en determinar el mínimo valor de n que verifique que $p(DC)\geq 1-\beta$, es decir que verifique simultaneamente las (3.101).Para el caso de k=4, $\beta=0,10$ y $\pi=1/4$ la solución es n=50.

3.4.2.6. Solución al problema clásico según la metodologia actual (Apor tación)

El planteamiento de Wilcox es susceptible de ser abordado por dos de las metodologias ya vistas para el modelo actual: las del tamaño de muestra para intervalos de confianza y para test de hipótesis.

En el segundo caso, como $\hat{\Delta} < \Delta_o \Leftrightarrow t_{exp} < 0$ con independencia de que utilicemos los test t, t' ó t'', el objetivo sera determinar el tamaño de muestra, $n=k.\bar{n}$, tal que los test:

$$
\begin{array}{lll}
H_o^1 \equiv \Delta = \Delta_o & \text{vs.} & H_1^1 \equiv \Delta > \Delta_o \\
H_o^2 \equiv \Delta = \Delta_o & \text{vs.} & H_1^2 \equiv \Delta < \Delta_o
\end{array}
\qquad (3.102)
$$

realizados a un error $\alpha = 0{,}50$ (para que asi $t_\cdot = 0$) tengan una potencia $1-\beta$ de detectar las alternativas Δ_2 y Δ_1 respectivamente. Para $\Delta_i > 0$, que es lo usual, las formulas de tamaño de muestra para los tests de una cola t, t' ó t'' son la misma e igual a (para la alternativa Δ):

$$
n(\Delta, \Delta_o) = \frac{h(\Delta)}{k-1} \cdot \frac{t^2_\beta}{(\Delta - \Delta_o)^2} \qquad (3.103)
$$

según las (3.85) y (3.86), con $h(\Delta)=(1-\Delta)\left\{1+\Delta(k+1)\right\}$ segun la (3.89). Como la condición de potencia deben verificarla los dos test (3.102) para las alternativas Δ_2 y Δ_1, entonces el tamaño ha de ser:

$$
n(\Delta_o) = \mathrm{Max}\left\{ n(\Delta_1, \Delta_o) \ ; \ n(\Delta_2, \Delta_o) \right\} \qquad (3.104)
$$

Alternativamente, el objetivo tambien será determinar el tamaño de muestra, $n=k\bar{n}$, tal que $\hat{\Delta}$ difiera del verdadero en una cantidad inferior que δ con confianza $1-2\beta$, con lo que la expresión (3.76), cambiando $\alpha/2$ por β, da la solución al problema. En ella Δ es sustituido por Δ_1 ó Δ_2 como información sobre los puntos a controlar. El tamaño resultante es el mismo dado en (3.104)

Las expresiones anteriores de tamaños de muestra fueron obtenidas originalmente para el caso de $n_1=n_2=\ldots=n_k=\bar{n}$ y en la situación mas desfavorable (varianza maxima) de que $\pi_i=1/k(\forall i)$. En tal circunstancia es $x_{ii}\sim B(n_i,p_{ii})$, con $p_{ii}=\Delta+(1-\Delta)(1/k)=p$ constante (ver 3.3.4), y así $x=\sum x_{ii}\sim B(n;p)$. Como en estas condiciones $\tilde{\Delta}=\hat{\Delta}$ y p vale p_1 y p_2 en los valores $\Delta=\Delta_1$ y $\Delta=\Delta_2$, entonces la (3.104) está solucionando las mismas ecuaciones clasicas (3.101) pero a traves de la aproximación normal a una binomial aproximación que será válida cuando la pareja de frecuencias esperadas responsables del maximo (np_1 y nq_1 ó np_2 y nq_2) sean mayores que 5 (lo que siempre ocurre para valores de Δ no demasiado extremos).

Asi, para el ejemplo citado en el subapartado anterior, seran $n(\Delta_1,\Delta_o)=51$ y $n(\Delta_2,\Delta_o)=21$, con lo que $n(\Delta_o)=51$, valor muy similar al 50 obtenido entonces por el metodo exacto (la aproximación es válida pues $np_1=88,75$ y $n(1-p_1)=11,475$ son ambas mayores que 5).

Sin embargo, y dado que $n(\Delta_1,\Delta_o)$ y $n(\Delta_2,\Delta_o)$ son tan distintos, el modo de proceder anterior (el clasico) lleva a un desaprovechamiento injustificado de los datos. El modo de evitarlo consistira en efectuar los test (3.102) para un Δ'_o tal que $n(\Delta_1,\Delta'_o)=n(\Delta_2,\Delta'_o)$, con $\Delta_1<\Delta'_o<\Delta_2$. Resolviendo la ecuación resultante en Δ'_o se obtiene

$$\Delta'_o=\frac{\Delta_1+H\,\Delta_2}{1+H}\qquad H=\sqrt{\frac{h(\Delta_1)}{h(\Delta_2)}}\tag{3.105}$$

con lo que la nueva regla de decision sera:

$$\begin{array}{l}\text{Si } \hat{\Delta}<\Delta'_o \Longrightarrow \text{decidir que } \Delta<\Delta_o\\ \text{Si } \hat{\Delta}\geq\Delta'_o \Longrightarrow \text{decidir que } \Delta\geq\Delta_o\end{array}\tag{3.106}$$

En particular, y de nuevo para el ejemplo citado, sera Δ'_o=0,8226 y n=36, tamaño apreciablemente inferior al 51 obtenido mas arriba.

Así, con menos muestra, pero cambiando el punto critico de decisión de Δ_o a Δ'_o, se consiguen las mismas garantias.

Sin embargo, una vez realizado el examen, y si los n_i no son exactamente iguales, la cantidad a comparar con Δ_o ó Δ'_o no debe ser $\tilde{\Delta}$, sino $\hat{\Delta}$, por las razones apuntadas en otras ocasiones.

3.4.3. Numero optimo de alternativas

3.4.3.1. <u>Solución mediante el modelo actual</u> (<u>Aportación</u>)

En el apartado anterior se han dado formulas para el tamaño $\bar{n}$ adecuado de muestra para el caso de contar con una alternativa Δ_j, de interes y supuesto k conocido y fijo. Dentro del esquema anterior parece conveniente plantearse cual es el valor óptimo de k, entendiendo el optimo en algun sentido que esta por especificar. La discusión que sigue está dedicada salvo indicación contraria, al caso de test de una cola para asignar calificaciones.

Una primera definición consistiria en entender por optimo a aquel valor de k que hace minimo el número total de preguntas a realizar. Como dicho número es $n=k.\bar{n}$, con $\bar{n}$ dada por las (3.85), (3.86) según el caso, y puesto que n decrece con k, es claro algo obvio: cuanto mas aumente k menos preguntas sera preciso realizar. Por otro lado, como n no puede llegar a ser menor que k (pues cuanto menos es $\bar{n}$=1), está claro que el minimo n posible es n=k, ó, equivalentemente, $\bar{n}$=1. Así, igualando las (3.90) y (3.86) a la unidad, se obtendra el valor deseado de k; por ejemplo, si Δ_o=0,5, Δ_1=0,6, α=5% y β=1%, la (3.81), que es la idónea para alternativas a la derecha, indica que k=381, y el examen deberá constar de 381 preguntas con 381 respuestas alternativas cada una.

Parece evidente que tal planteamiento de optimo no es el adecuado por cuanto, entre otras razones, redactar 381 alternativas para cada pregunta no es algo asequible. Mas razonable es obtener el minimo no del número de preguntas, sino del espacio total dedicado al examen. Como cada pregunta consta de k respuestas posibles, el conjunto del examen contendra $kn=k^2\bar{n}$ "frases-respuesta"; si ademas cada enunciado de pregunta lo consideramos equivalente a una "frase-respuesta", en total el examen constara de $N=(k+1)(k\bar{n})$ - "frases-respuesta", numero que habra de minimizarse con respecto a k, pero:

$$N = \left\{ \frac{t_\alpha \sqrt{g(\Delta_0)} + t_\beta \sqrt{g(\Delta_1)}}{\Delta_1 - \Delta_0} \right\}^2 \quad \text{si test t o t''} \qquad (3.107)$$

$$N = g(\Delta_1) \left\{ \frac{t_\alpha + t_\beta}{\Delta_1 - \Delta_0} \right\}^2 \quad \text{si test t'} \qquad (3.108)$$

con

$$g(\Delta) = \frac{(k+1)(\Delta k+1-\Delta)}{k-1} (1-\Delta) \qquad (3.109)$$

por lo que el signo de dN/dk, al derivar la (3. 107), es el mismo que el de la expresión

$$t_\alpha \varphi(\Delta_0) + t_\beta \varphi(\Delta_1) \qquad (3.110)$$

con

$$\varphi(\Delta) = \frac{\Delta(1-\Delta)}{\sqrt{(1-\Delta)(\Delta k+1-\Delta)}} (k-1+\sqrt{\frac{2}{\Delta}})(k-1-\sqrt{\frac{2}{\Delta}}) \qquad (3.111)$$

• Como $1 < 1+\sqrt{2/\Delta_0} < 1+\sqrt{2/\Delta_1}$, pues los tests t ó t'' se usaran cuando

$\Delta_1 < \Delta_o$, y como $\varphi(\Delta_o)$ y $\varphi(\Delta_1)$ son siempre negativas entre k=1 y k=1 $+\sqrt{2/\Delta_o}$ y positivas a partir de k=1 $+\sqrt{2/\Delta_1}$, entonces el valor k en el que la (3.107) se anula (y que hara minimo a N) será tal que:

$$1+\sqrt{\frac{2}{\Delta_o}} < k < 1+\sqrt{\frac{2}{\Delta_1}} \qquad (3.112)$$

por ejemplo, si Δ_o=0,50, Δ_1=0,40, α=1% y β=5%, sera $3 < k < 3{,}24$, y el valor optimo k_o de k, elegido de entre las k=3 y k=4, es k_o=3. De hecho tal optimo permanece para cualquier otro valor razonable de Δ_1 y para cualquier otra pareja de valores Δ_o y Δ_1 que pretendan dar el notable o el sobresaliente (por ejemplo Δ_o=0,70 y Δ_1=0,60, ó Δ_o=0,85 y Δ_1=0,75). Conviene observar tambien que el valor optimo k_o no es solo practicamente independiente de Δ_o y Δ_1, sino que lo es casi completamente de α y β como muestra la (3.112)(la completa independencia no se da pues la (3. 112) solo es una acotación y el verdadero valor, que debe ser entero, habra de localizarse a traves de la (3.107) que si depende de α y β).

Con el ejemplo citado antes sera k_o=3, $\bar{n}$=272 y n=816.

Para alternativas a la derecha ($\Delta_1 > \Delta_o$) el test a emplear sera el t' y dN/dk habra de calcularse de la (3.108); pero:

$$\frac{dN}{dk} = \left(\frac{t_\alpha + t_\beta}{\Delta_1 - \Delta_o}\right)^2 \left(\Delta_1 - \frac{2}{(k-1)^2}\right)(1-\Delta_1) \qquad (3.113)$$

con lo que el minimo se alcanza en

$$k = 1+\sqrt{2/\Delta_1} \qquad (3.114)$$

Bajo los supuestos ordinarios de Δ_1=0,6, 0,8, 0,95, los valores de k estan siempre comprendidos entre k=2 y k=3, pero por la (3.108) $N(k=2) > N(k=3) \Leftrightarrow \Delta_1 < 1$, por lo que k_o=3. En realidad el valor k_o=3 es optimo siempre que $\Delta_1 > 0{,}5$, pues, segun la (3.108), $2 < k < 3 \Leftrightarrow \frac{1}{2} < \Delta_1 < 1$. Ahora el optimo es por tanto independiente de α, β, Δ_o y Δ_1 a la hora de asignar las calificaciones de aprobado o superioes.

Realmente la funcion de N es algo opinable. Si cada enunciado de pregunta se considera equivalente a dos "frases-respuesta", entonces $N=(k+2)(k\bar{n})$ y las expresiones (3.107) a la (3.114) permanecen si se cambia el (k+1) de la (3.109) por (k+2) y los valores 2 de las (3.111) a la (3.114) por el valor 3. De nuevo puede comprobarse que el k optimo es siempre k_o=3, aunque a veces, para discriminar el aprobado, el valor k=4 es solo un poco mejor en el caso de la (3.107).

Sin embargo el caso de k=2 merece un comentario especial. Un examen en que las alternativas son solo dos puede confeccionarse sin "frases-respuesta", de modo que el alumno se limita a contestar SI o NO a la afirmacion del enunciado. En tal situación, y suponiendo que cada "enunciado-afirmación" equivales a dos "frases-respuesta" de las consideradas arriba, sera $N=4\bar{n}$, con $\bar{n}$ dado por las (3.97) y (3.98). Cuando k=3, N vendrá dado por las (3.107) y (3.108) si se supone que cada enunciado equivale a solo una "frase-respuesta". Como $N(k=2) < N(k=3)$, $\forall A$, entonces el examen de respuesta binaria es preferible a cualquier otro sistema por ser mas sensible a igualdad de espacio. En la discusión anterior no ha intervenido la posible mayor ó menor dificultad de confeccionar un examen de uno u otro tipo; tal aspecto es subjetivo y depende del tema objeto de examen.

En el caso de la independencia ($H_o \equiv \Delta=0$) y bajos los razonamientos anteriores, es N=(k+1)kn, con n dada por las (3.95) ó (3.96). Como, sea cual sea el caso, ocurre que dN/dk > 0, siempre, entonces el optimo es k_o=2 en el caso de test de dos colas, si bien en el contexto de los examenes tal comprobación no tiene interés alguno.

3.4.3.2. Soluciones clásicas

Lord (1977) cita cuatro métodos de aproximación al problema aqui planteado, asi como otras aproximaciones empiricas. Todas ellas son conformes con las conclusiones obtenidas anteriormente: k=2 ó 3 es el optimo, aunque en ciertas circunstancias puede serlo k=4. En definitiva, diversos métodos de aproximación al problema, basados en modelos muy distintos, dan soluciones similares.

Lord, en su cuarto método (basado en el modelo de la curva caracteristica del item), establece una conclusión añadida que es compatible con las conclusiones de 3.4.3.1.; el decrecimiento de k ocasiona un incremento en la eficacia del test para los alumnos buenos, y la decrece en los malos, de modo que el establecimiento del valor óptimo de k es relativo al fin perseguido.

APENDICE I

ESTRUCTURA DEL PROGRAMA

Procedemos ahora a describir la rutina FORTAN V que aparece en el APENDICE I y que ejecuta el algoritmo anteriormente presentado :

1. Llamada a la subrutina.-

   SUBROUTINE TEXFIS(IT, P, IA, IR, IER)

2. Descripción de los argumentos.-

   IT Matriz entera 3x3 en la que aparece la tabla para la que se desea calcular las probabilidades de error de una y dos colas. La tabla debe aparecer en las dos primeras filas y dos primeras columnas, dejado la fila 3 y la columna 3 para los marginales que calcula la rutina. En salida figura en IT la tabla mas extrema(si existe) en el sentido opuesto a la que originalmente se puso.

   P. Matriz real de longitud 2. En salida figuran las probabilidades de una y dos colas en P(1) y P(2) respectivamente.

   IA. Variable entera que aparecerá llena con dos asteriscos si al realizar el sorteo es elegido x_1, y, por tanto, P(2) <2.P(1). Si no ocurre esto aparecerá llena con dos blancos.

   IR. Matriz entera de dimensión 3 en la que el usuario de la rutina deberá introducir tres valores enteros (preferiblemente elegidos al azar) comprendidos entre 0 y 30323; dichos valores se usaran como entrada a la rutina RANDOM de Wichmann y Hill (1982); si se usa de manera continuada la rutina TEXFIS basta con introducirlos una vez, ya que la rutina RANDOM puede usar sus resultados como sucesivas entradas.

   IER. Variable entera donde apareceran los diferentes codigos de error.

   IER=1. Alguno de los marginales es menor o igual que cero. Los valores que aparecen a continuación provienen de la función CHYPER, Lund, R.E.(1980), y son los valores que puede tomar en ella IFAULT aumentados en una unidad.

IER=2. Si alguno de los marginales está fuera del rango de la hipergeometrica (tal ocurriria si en la rutina TEXFIS se encontraran negativos en la tabla IT).

IER=3. Si x_1 estuviera fuera de su limite inferior o por encima de su limite superior.

IER=0. Si no hay ningun error.

3. Comentarios adicionales al programa.

En la presente rutina, para el calculo de las probabilidades de una hipergeometrica (tanto de valores aislados como los acumulados) se ha usado la función CHYPER de Lund, R.E. (1980), tal función, dependiendo de uno de sus argumentos, calcula la probabilidad de que la variable tome un valor ó la de que tome un valor menor ó igual que uno dado. Tal función, CHYPER, como la mayoria de las funciones que calculan una probabilidad de manera recurrente, pueden presentar el problema del overflow (underflow). Ese problema ha sido puesto de manifiesto por diferentes autores, pero especialmente por Lillestol, J.(1982)con referencia a CHYPER. La solución al problema proviene de realizar un cambio de escala en los valores originales cuando estos vayan a alcanzar en el producto, un valor por encima del máximo permitido, volviendo a la escala original cuando los cocientes correspondientes se hayan hecho. Un procedimiento análogo, pero a la inversa, debe realizarse para evitar el "underflow". Teniendo en cuenta el uso que se hace de la rutina en este trabajo, los errores a los que nos acabamos de referir rara vez ocurren, por lo que no nos detendremos mas en ello. Por supuesto que en el caso de que el ordenador de que se disponga avise del "overflow", ó del "underflow", el problema está solventado inmediatamente. La versión de CHYPER usada es la de doble precisión.

En el caso de que no se disponga de la función CHYPER, para un usuario con conocimientos de FORTRAN no es dificil calcular la probabilidad puntual de una $H(n, n_1, a_1)$, ya que puede usar la formula de recurrencia $P(x_1+1)= P(x_1).T(x_1)$, donde $T(x_1)= (n_1-x_1)(a_1-x_1)/ [(x_1+1)(n-n_1-a_1-x_1+1)]$ y donde, evidentemente, $\max(0; a_1-n_2) \leq x_1 \leq \min(a_1; n_1)$. Una vez mas el usuario deberia controlar el "overflow" y / ó el"underflow".

Como en la rutina que presentamos se puede llevar a cabo un sorteo , hemos tenido que incluir en ella a un programa que genere numeros seudoaleatorios. Se ha usado la función RANDOM de Wichnann, B.A. and Hill, I.D. (1982) debido a su potencia (tiene un ciclo de repetición de 2.78×10^{13} sobrante para el uso que de ella se hace). Hemos implementado la versión de tal función en que los números iniciales se dan como argumentos de la función, es decir la llamada seria RA=RANDOM (IR),(siendo IR una matriz de dimensión 3); tal uso puede hacer un poco mas lenta de ejecución la rutina (pues la introducción de los valores a traves de un COMMON hace que la ejecución gane en velocidad) pero el número de sorteos es tan pequeño que esto no afecta al tiempo de ejecución de la rutina total. Asimismo, hemos usado la versión que supone un ordenador de 16 bits(enteros entre - 32768 y + 32767), lo que resulta suficiente para este caso.

En la rutina TEXFIS es necesario que el usuario introduzca tres números aleatorios (que iran en la matriz IR) cada vez que se la llame. Si se usa de manera continuada la rutina TEXFIS, los valores aleatorios resultantes de una llamada pueden ser usados como valores de entrada para la siguiente. Con respecto a los enteros aleatorios iniciales, lo mas común es proporcionar los que se consiguen como fruto de operaciones sencillas con la hora y la fecha que suelen incluir en la actualidad practicamente todos los ordenadores (los autores lo han hecho asi en la generación de tablas que han realizado y no han tenido ningún problema). En cualquier caso, el usuario puede mejorar la "calidad" de los numeros iniciales haciendo un numero (aleatorio tambien, si lo desea) de llamadas a la rutina TEXFIS con objeto de separarse suficientemente de los valores "aleatorios " de partida.

4. Tiempo de ejecución

La rutina tiene un tiempo de ejecución que crece aproximadamente de manera lineal con el valor de x_1.

5. Lenguaje y portabilidad del codigo fuente.

El programa ha sido realizado en FORTRAN V usando las características estandar de este lenguaje. La rutina corre en la actualidad en un ordenador ECLIPSE/140 (Data General) de 16 bits/palabra y 64K de memoria central para usuario, siendo el espacio requerido por la rutina mucho menor de las citadas 64K.

```
ÜDDHHS.ESTADISTICA:BEFISHPROP00.2

      SUBROUTINE TEXFIS(IT,P,IA,IR,IER)
C
C     ESTA RUTINA REALIZA EL CALCULO DE LAS PROBABILIDADES DE UNA
C     Y DOS COLAS POR EL METODO MAS POTENTE DEL TEST EXACTO DE
C     FISHER.
C
      INTEGER IT(3,3),IA,IER,IFA
      REAL P(2),C1,C2
      DOUBLE PRECISION PIGX(2),PMIX(2)
C
C     INICIALIZACION DE LOS PARAMETROS DE ERROR Y PROBABILIDADES.
C
      IA='  '
      IER=0
      IFA=0
      P(1)=1.0
      P(2)=1.0
      PMIX(2)=0.0D0
      PMIX(2)=0.0D0
C
C     CALCULO DE LOS MARGINALES DE LA TABLA
C
      IT(1,3)=IT(1,1)+IT(1,2)
      IT(2,3)=IT(2,1)+IT(2,2)
      IT(3,1)=IT(1,1)+IT(2,1)
      IT(3,2)=IT(1,2)+IT(2,2)
      IT(3,3)=IT(3,1)+IT(3,2)
C
C     SI LOS MARGINALES SON 0 O MENORES QUE 0, FINAL DE LA RUTINA
C     EN CONDICION DE ERROR.
C
      IF(IT(1,3).LE.0.OR.IT(2,3).LE.0)    GO TO 1000
      IF(IT(3,1).LE.0.OR.IT(3,2).LE.0)    GO TO 1000
C
      DO 100 IC=1,2
C
C     SELECCIONA X1 COMO AQUEL VALOR QUE DA UNA MENOR FRECUENCIA
C     REALTIVA.
C
         IP=1
         C1=FLOAT(IT(1,1)*IT(2,2))
         C2=FLOAT(IT(2,1)*IT(1,2))
         IF(C1-C2) 20,10000,10
10       IP=2
C
C     PREPARADOS YA LOS VALORES LLAMADA PARA EL CALCULO DE LAS
C     PROBABILIDADES DE LA HIPERGEOMETRICA.PRIMERO SE CALCULA
C     LA P(X=X1) Y DESPUES P(X<=X1).
C
20       MODO=1
         PIGX(IC)=CHYPER(MODO,IT(3,1),IT(IP,1),IT(3,3),IT(IP,3),IFA)
         MODO=2
         PMIX(IC)=CHYPER(MODO,IT(3,1),IT(IP,1),IT(3,3),IT(IP,3),IFA)
         P(IC)=PMIX(IC)
         IF(IFA.NE.0) GO TO 1010
C
C     PREPARA EL CALCULO DE LA PROBABILIDAD DE UNA TABLA TAN EXTREMA
C     O MAS QUE LA QUE SE TIENE, PERO EN EL SENTIDO OPUESTO A ESTA.
C
         IF(IP.EQ.[illegible]) GO TO [illegible]
```

```
HDD:US.ESTADISTICA:BEFISHPROP00.2

          IN=IP
          IP=3-IP
C
C       CALCULO DEL NUEVO X1 PARA LA TABLA MAS EXTREMA.
C
          CPRIMA=2.0*FLOAT(IT(IN,3)*IT(3,1))-FLOAT(IT(3,3)*IT(IN,1))
          IF(CPRIMA.LT.0.0) GO TO 10000
          CPRIMA=CPRIMA/FLOAT(IT(3,3))
C       TOMA X1 COMO EL ENTERO IGUAL O INMEDIATAMENTE SUPERIOR A CPRIMA
          IX1=CPRIMA
          IF(FLOAT(IX1).LT.CPRIMA) IX1=IX1+1
C       CALCULA LA NUEVA TABLA
          IT(IN,1)=IX1
          IT(IN,2)=IT(IN,3)-IX1
          IT(IP,1)=IT(3,1)-IT(IN,1)
          IT(IP,2)=IT(IP,3)-IT(IP,1)
C       SI NO ES POSIBLE LA NUEVA TABLA EXTREMA FIN DE LA RUITNA.
          IF(IT(IP,2).LT.0) GO TO 10000
100     CONTINUE
C
C       COLOCACION DE LOS CODIGOS DE ERROR.
C
1000    IER=1
        GO TO 10010
1010    IER=IFA+1
        GO TO 10010
C
C       CALCULO DE LA PROBABILIDAD DE DOS COLAS. CON SORTEO,SI HA LUGAR.
C
10000   P(2)=P(1)+PMIX(2)
C               SI CPRIMA NO ES ENTERO,FINAL.
        IF(FLOAT(IX1).NE.CPRIMA) GO TO 10010
C               SI P(X1))P(XPRIMA1),FINAL.
        IF(PIGX(1).GT.PIGX(2))     GO TO 10010
        P(2)=P(2)-PIGX(2)
C               SI P(X1)(P(XPRIMA1) DESCONTAR P(XPRIMA1) DE LA PROBA-
C               BILIDAD DE DOS COLAS E IRSE AL FINAL.
        IF(PIGX(1).LT.PIGX(2))     GO TO 10010
C               SI P(X1)=P(XPRIMA1) REALIZAR UN SORTEO PARA DESCONTAR
C               P(XPRIMA1) CON PROBABILIDAD 0.5.
        IA="**"
        RA=RANDOM(IR(1),IR(2),IR(3))
        IF(RA.LE.0.5) GO TO 10010
        P(2)=P(2)+PIGX(2)
        IA="  "
10010   RETURN
        END
```

APENDICE II

TABLAS PARA LA VERSION MAS POTENTE DEL TEST EXACTO DE FISHER

TABLAS DEL TEST EXACTO DE FISHER   PAG. 1

T O T A L(n)= 8

a1	n1	x1	P1	P2
2	6	0	.0357	.0357
3	4	0	.0714	.0714*
	5	0	.0179	.0179
4	4	0	.0143	.0143*

T O T A L(n)= 9

a1	n1	x1	P1	P2
2	6	0	.0833	.0833
	7	0	.0278	.0278
3	5	0	.0476	.0476
	6	0	.0119	.0119
4	4	0	.0397	.0476
	5	0	.0079	.0079

T O T A L(n)=10

a1	n1	x1	P1	P2
1	9	0	.1000	.1000
2	7	0	.0667	.0667
	8	0	.0222	.0222
3	5	0	.0833	.1667
	6	0	.0333	.0333
	7	0	.0083	.0083
4	4	0	.0714	.0762
	5	0	.0238	.0476
	6	0	.0048	.0048
5	5	0	.0040	.0040*

T O T A L(n)=11

a1	n1	x1	P1	P2
1	10	0	.0909	.0909
2	8	0	.0546	.0546
	9	0	.0182	.0182
3	6	0	.0606	.0606
	7	0	.0242	.0242
	8	0	.0061	.0061
4	5	0	.0455	.0606
	6	0	.0152	.0152
	7	0	.0030	.0030
		1	.0879	.0879
5	5	0	.0130	.0152
	6	0	.0022	.0022
		1	.0671	.0801

T O T A L(n)=12

a1	n1	x1	P1	P2
1	11	0	.0833	.0833
2	8	0	.0909	.0909
	9	0	.0455	.0455
	10	0	.0152	.0152

T O T A L(n)=12CONT.

a1	n1	x1	P1	P2
3	6	0	.0909	.1818
	7	0	.0455	.0455
	8	0	.0182	.0182
	9	0	.0045	.0045
4	5	0	.0707	.0808
	6	0	.0303	.0606
	7	0	.0101	.0101
	8	0	.0020	.0020
		1	.0667	.0667
5	5	0	.0265	.0278
	6	0	.0076	.0076*
	7	0	.0013	.0013
		1	.0455	.0720
6	6	0	.0011	.0022
		1	.0400	.0801

T O T A L(n)=13

a1	n1	x1	P1	P2
1	12	0	.0769	.0769
2	9	0	.0769	.0769
	10	0	.0385	.0385
	11	0	.0128	.0128
3	7	0	.0699	.0699
	8	0	.0350	.0350
	9	0	.0140	.0140
	10	0	.0035	.0035
4	5	0	.0979	.1049
	6	0	.0490	.0699
	7	0	.0210	.0210
	8	0	.0070	.0070
	9	0	.0014	.0014
		1	.0518	.0518
5	5	0	.0435	.0754
	6	0	.0163	.0210
	7	0	.0047	.0047
		1	.0863	.1026
	8	0	.0008	.0008
		1	.0319	.0319
6	6	0	.0041	.0047
		1	.0775	.1026
	7	0	.0006	.0006
		1	.0251	.0291

T O T A L(n)=14

a1	n1	x1	P1	P2
1	13	0	.0714	.0714
2	10	0	.0659	.0659
	11	0	.0330	.0330
	12	0	.0110	.0110
3	7	0	.0961	.0961*
	8	0	.0550	.0550
	9	0	.0275	.0275
	10	0	.0110	.0110

T O T A L(n)=14CONT.

a1	n1	x1	P1	P2
3	11	0	.0028	.0028
		1	.0934	.0934
4	6	0	.0699	.0849
	7	0	.0350	.0350*
	8	0	.0150	.0150
	9	0	.0050	.0050
		1	.0949	.0949
	10	0	.0010	.0010
		1	.0410	.0410
5	5	0	.0629	.0859
	6	0	.0280	.0310
	7	0	.0105	.0105*
	8	0	.0030	.0030
		1	.0629	.0909
	9	1	.0230	.0230
6	6	0	.0093	.0097
	7	0	.0023	.0023*
		1	.0513	.0536*
	8	1	.0163	.0256
7	7	0	.0003	.0003*
		1	.0146	.0149*

T O T A L(n)=15

a1	n1	x1	P1	P2
1	14	0	.0667	.0667
2	10	0	.0952	.0952
	11	0	.0571	.0571
	12	0	.0286	.0286
	13	0	.0095	.0095
3	8	0	.0769	.0769
	9	0	.0440	.0440
	10	0	.0220	.0220
	11	0	.0088	.0088
	12	0	.0022	.0022
		1	.0813	.0813
4	6	0	.0923	.1033
	7	0	.0513	.0769
	8	0	.0256	.0256
	9	0	.0110	.0110
	10	0	.0037	.0037
		1	.0769	.0769
	11	0	.0007	.0007
		1	.0330	.0330
5	5	0	.0839	.1009
	6	0	.0420	.0440
	7	0	.0187	.0256
	8	0	.0070	.0070
	9	0	.0020	.0020
		1	.0470	.0889
	10	1	.0170	.0170
6	6	0	.0168	.0278
	7	0	.0056	.0070
		1	.0839	.1189
	8	0	.0014	.0014
		1	.0350	.0406

T O T A L(n)=15CONT.

a1	n1	x1	P1	P2
6	9	1	.0110	.0110
7	7	0	.0012	.0014
		1	.0317	.0406
	8	1	.0089	.0101

T O T A L(n)=16

a1	n1	x1	P1	P2
1	15	0	.0625	.0625
2	11	0	.0833	.0833
	12	0	.0500	.0500
	13	0	.0250	.0250
	14	0	.0083	.0083
3	8	0	.1000	.1000*
	9	0	.0625	.0625
	10	0	.0357	.0357
	11	0	.0179	.0179
	12	0	.0071	.0071
	13	0	.0018	.0018
		1	.0714	.0714
4	7	0	.0692	.0885
	8	0	.0385	.0769
	9	0	.0192	.0192
	10	0	.0082	.0082
	11	0	.0028	.0028
		1	.0632	.0632
	12	0	.0006	.0006
		1	.0269	.0269
5	6	0	.0577	.0934
	7	0	.0289	.0337
	8	0	.0128	.0256
	9	0	.0048	.0048
		1	.0769	.1058
	10	0	.0014	.0014
		1	.0357	.0357
	11	1	.0128	.0128
6	6	0	.0262	.0338
	7	0	.0105	.0114
	8	0	.0035	.0035*
		1	.0594	.1189
	9	0	.0009	.0009
		1	.0245	.0350
	10	1	.0076	.0076
		2	.0919	.1181
7	7	0	.0032	.0032
		1	.0546	.0601
	8	0	.0007	.0007*
		1	.0203	.0210*
	9	1	.0056	.0087
		2	.0717	.1262
8	8	1	.0051	.0051*
		2	.0660	.1319

TABLAS DEL TEST EXACTO DE FISHER PAG. 2

T O T A L(n)=17

a1	n1	x1	P1	P2
**	**	**	******	******
1	16	0	.0588	.0588
2	12	0	.0735	.0735
	13	0	.0441	.0441
	14	0	.0221	.0221
	15	0	.0074	.0074
3	9	0	.0824	.0824
	10	0	.0515	.0515
	11	0	.0294	.0294
	12	0	.0147	.0147
	13	0	.0059	.0059
	14	0	.0015	.0015
		1	.0632	.0632
4	7	0	.0882	.1029
	8	0	.0529	.0824
	9	0	.0294	.0294
	10	0	.0147	.0147
	11	0	.0063	.0063
		1	.0987	.0987
	12	0	.0021	.0021
		1	.0525	.0525
	13	1	.0223	.0223
5	6	0	.0747	.1023
	7	0	.0407	.0441
	8	0	.0204	.0294
	9	0	.0091	.0091
	10	0	.0034	.0034
		1	.0600	.1007
	11	0	.0010	.0010
		1	.0276	.0276
	12	1	.0099	.0099
6	6	0	.0373	.0427
	7	0	.0170	.0345
	8	0	.0068	.0091
		1	.0882	.1312
	9	0	.0023	.0023
		1	.0430	.0498
	10	0	.0006	.0006
		1	.0175	.0175
	11	1	.0054	.0054
		2	.0721	.1094
7	7	0	.0062	.0098
		1	.0818	.1340
	8	0	.0019	.0023
		1	.0364	.0498
	9	1	.0134	.0152
	10	1	.0037	.0037
		2	.0522	.0584
8	8	1	.0122	.0152
	9	1	.0030	.0034
		2	.0445	.0567

****

T O T A L(n)=18

a1	n1	x1	P1	P2
**	**	**	******	******
1	17	0	.0556	.0556
2	12	0	.0980	.0980

T O T A L(n)=18CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
2	13	0	.0654	.0654
	14	0	.0392	.0392
	15	0	.0196	.0196
	16	0	.0065	.0065
3	10	0	.0686	.0686
	11	0	.0429	.0429
	12	0	.0245	.0245
	13	0	.0123	.0123
	14	0	.0049	.0049
	15	0	.0012	.0012
		1	.0564	.0564
4	8	0	.0686	.0915
	9	0	.0412	.0412*
	10	0	.0229	.0229
	11	0	.0114	.0114
	12	0	.0049	.0049
		1	.0833	.0833
	13	0	.0016	.0016
		1	.0441	.0441
	14	1	.0186	.0186
5	6	0	.0924	.1142
	7	0	.0539	.1013
	8	0	.0294	.0360
	9	0	.0147	.0294
	10	0	.0065	.0065
		1	.0882	.1177
	11	0	.0025	.0025
		1	.0474	.0474
	12	0	.0007	.0007
		1	.0217	.0217
	13	1	.0077	.0077
		2	.0987	.0987
6	6	0	.0498	.0537
	7	0	.0249	.0377
	8	0	.0113	.0128
	9	0	.0045	.0045*
		1	.0656	.1312
	10	0	.0015	.0015
		1	.0317	.0430
	11	1	.0128	.0128
	12	1	.0039	.0039
		2	.0573	.1070
7	7	0	.0104	.0128
	8	0	.0038	.0040
		1	.0566	.0656
	9	0	.0011	.0023
		1	.0249	.0498
	10	1	.0091	.0128
		2	.0882	.1448
	11	1	.0025	.0025
		2	.0387	.0491
8	8	0	.0010	.0011
		1	.0230	.0248
	9	1	.0076	.0078*
		2	.0767	.1534
	10	1	.0019	.0029
		2	.0307	.0536

T O T A L(n)=18CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
9	9	1	.0017	.0034
		2	.0283	.0567

****

T O T A L(n)=19

a1	n1	x1	P1	P2
**	**	**	******	******
1	18	0	.0526	.0526
2	13	0	.0877	.0877
	14	0	.0585	.0585
	15	0	.0351	.0351
	16	0	.0175	.0175
	17	0	.0058	.0058
3	10	0	.0867	.0867
	11	0	.0578	.0578
	12	0	.0361	.0361
	13	0	.0206	.0206
	14	0	.0103	.0103
	15	0	.0041	.0041
		1	.0970	.0970
	16	0	.0010	.0010
		1	.0506	.0506
4	8	0	.0851	.1032
	9	0	.0542	.0867
	10	0	.0325	.0325
	11	0	.0181	.0181
	12	0	.0090	.0090
	13	0	.0039	.0039
		1	.0709	.0709
	14	0	.0013	.0013
		1	.0374	.0374
	15	1	.0157	.0157
5	7	0	.0681	.1060
	8	0	.0397	.0446
	9	0	.0217	.0325
	10	0	.0108	.0108
	11	0	.0048	.0048
		1	.0710	.1108
	12	0	.0018	.0018
		1	.0379	.0379
	13	0	.0005	.0005
		1	.0173	.0173
	14	1	.0061	.0061
		2	.0844	.0844
6	6	0	.0633	.1093
	7	0	.0341	.0436
	8	0	.0170	.0181
	9	0	.0077	.0108
		1	.0913	.1409
	10	0	.0031	.0031
		1	.0495	.0573
	11	0	.0010	.0010
		1	.0237	.0408
	12	1	.0096	.0096
		2	.0947	.1287
	13	1	.0029	.0029
		2	.0460	.0460

T O T A L(n)=19CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
7	7	0	.0157	.0174
	8	0	.0065	.0128
		1	.0799	.1473
	9	0	.0024	.0031
		1	.0399	.0573
	10	0	.0007	.0007
		1	.0174	.0198
	11	1	.0063	.0063
		2	.0674	.0739
	12	1	.0017	.0017
		2	.0292	.0449
8	8	0	.0022	.0034
		1	.0371	.0587
	9	0	.0006	.0007
		1	.0149	.0198
	10	1	.0049	.0055
		2	.0549	.0698
	11	1	.0012	.0012
		2	.0216	.0237
9	9	1	.0045	.0055
		2	.0513	.0698
	10	1	.0010	.0011
		2	.0185	.0230

****

T O T A L(n)=20

a1	n1	x1	P1	P2
**	**	**	******	******
1	18	0	.1000	.1000
	19	0	.0500	.0500
2	14	0	.0789	.0789
	15	0	.0526	.0526
	16	0	.0316	.0316
	17	0	.0158	.0158
	18	0	.0053	.0053
3	11	0	.0737	.0737
	12	0	.0491	.0491
	13	0	.0307	.0307
	14	0	.0175	.0175
	15	0	.0088	.0088
	16	0	.0035	.0035
		1	.0877	.0877
	17	0	.0009	.0009
		1	.0456	.0456
4	9	0	.0681	.0941
	10	0	.0433	.0433*
	11	0	.0260	.0260
	12	0	.0145	.0145
	13	0	.0072	.0072
	14	0	.0031	.0031
		1	.0609	.0609
	15	0	.0010	.0010
		1	.0320	.0320
	16	1	.0134	.0134
5	7	0	.0830	.1137
	8	0	.0511	.0547
	9	0	.0298	.0379

TABLAS DEL TEST EXACTO DE FISHER PAG. 3

T O T A L(n)=20CONT.

al	nl	xl	P1	P2
**	**	**	******	******
5	10	0	.0163	.0163*
	11	0	.0081	.0081
		1	.0975	.1273
	12	0	.0036	.0036
		1	.0578	.1089
	13	0	.0014	.0014
		1	.0307	.0307
	14	1	.0139	.0139
	15	1	.0049	.0049
		2	.0726	.0726
6	6	0	.0775	.1149
	7	0	.0443	.0515
	8	0	.0238	.0419
	9	0	.0119	.0141
	10	0	.0054	.0054*
		1	.0704	.0759*
	11	0	.0022	.0022
		1	.0379	.0499
	12	0	.0007	.0007
		1	.0181	.0181
	13	1	.0072	.0072
		2	.0777	.1219
	14	1	.0022	.0022
		2	.0374	.0374
7	7	0	.0221	.0445
	8	0	.0102	.0147
	9	0	.0043	.0047
		1	.0579	.0703
	10	0	.0016	.0016*
		1	.0286	.0302*
	11	1	.0124	.0166
	12	1	.0044	.0044
		2	.0521	.0623
	13	1	.0012	.0012
		2	.0223	.0223
8	8	0	.0039	.0047
		1	.0542	.0697
	9	0	.0013	.0014
		1	.0249	.0281
	10	0	.0004	.0007
		1	.0099	.0102*
		2	.0849	.0948*
	11	1	.0032	.0045
		2	.0399	.0648
	12	1	.0008	.0008
		2	.0154	.0194
9	9	1	.0092	.0098
		2	.0799	.0923
	10	1	.0027	.0028*
		2	.0349	.0376*
	11	1	.0006	.0009
		2	.0124	.0216
		3	.0949	.1748
10	10	1	.0006	.0006*
		2	.0115	.0230
		3	.0895	.1010*

****

T O T A L(n)=21

al	nl	xl	P1	P2
**	**	**	******	******
1	19	0	.0952	.0952
	20	0	.0476	.0476
2	14	0	.1000	.1000
	15	0	.0714	.0714
	16	0	.0476	.0476
	17	0	.0286	.0286
	18	0	.0143	.0143
	19	0	.0048	.0048
3	11	0	.0902	.0902
	12	0	.0632	.0632
	13	0	.0421	.0421
	14	0	.0263	.0263
	15	0	.0150	.0150
	16	0	.0075	.0075
	17	0	.0030	.0030
		1	.0797	.0797
	18	0	.0007	.0007
		1	.0414	.0414
4	9	0	.0827	.1038
	10	0	.0551	.0902
	11	0	.0351	.0351
	12	0	.0211	.0211
	13	0	.0117	.0117
	14	0	.0058	.0058
		1	.0877	.0877
	15	0	.0025	.0025
		1	.0526	.0526
	16	0	.0008	.0008
		1	.0276	.0276
	17	1	.0115	.0115
5	7	0	.0984	.1235
	8	0	.0633	.1107
	9	0	.0389	.0451
	10	0	.0227	.0351
	11	0	.0124	.0124
	12	0	.0062	.0062
		1	.0805	.1194
	13	0	.0028	.0028
		1	.0475	.0475
	14	0	.0010	.0010
		1	.0251	.0251
	15	1	.0114	.0114
	16	1	.0040	.0040
		2	.0630	.0630
6	6	0	.0922	.1229
	7	0	.0553	.0609
	8	0	.0316	.0456
	9	0	.0170	.0186
	10	0	.0085	.0124
		1	.0936	.1486
	11	0	.0039	.0039
		1	.0550	.0635
	12	0	.0016	.0016
		1	.0294	.0464
	13	0	.0005	.0005
		1	.0139	.0139
	14	1	.0056	.0056

T O T A L(n)=21CONT.

al	nl	xl	P1	P2
**	**	**	******	******
6	14	2	.0642	.1196
	15	1	.0017	.0017
		2	.0307	.0307
7	7	0	.0295	.0468
	8	0	.0148	.0180
	9	0	.0068	.0071
		1	.0783	.0873*
	10	0	.0028	.0039
		1	.0426	.0635
	11	0	.0010	.0010
		1	.0209	.0237
	12	1	.0090	.0158
		2	.0805	.0873*
	13	1	.0032	.0032
		2	.0408	.0555
	14	1	.0009	.0009
		2	.0173	.0173
8	8	0	.0063	.0068
		1	.0738	.0850
	9	0	.0024	.0046
		1	.0375	.0669
	10	0	.0008	.0010
		1	.0170	.0237
	11	1	.0067	.0075
		2	.0635	.0805
	12	1	.0022	.0022
		2	.0294	.0318
	13	1	.0005	.0005
		2	.0113	.0176
		3	.0900	.1637
9	9	0	.0007	.0011
		1	.0159	.0244
	10	1	.0058	.0075
		2	.0563	.0805
	11	1	.0017	.0019
		2	.0242	.0300
	12	2	.0085	.0092
		3	.0713	.0872
10	10	1	.0016	.0019
		2	.0226	.0300
	11	2	.0073	.0089
		3	.0635	.0861

****

T O T A L(n)=22

al	nl	xl	P1	P2
**	**	**	******	******
1	20	0	.0909	.0909
	21	0	.0455	.0455
2	15	0	.0909	.0909
	16	0	.0649	.0649
	17	0	.0433	.0433
	18	0	.0260	.0260
	19	0	.0130	.0130
	20	0	.0043	.0043
3	12	0	.0779	.0779
	13	0	.0546	.0546

T O T A L(n)=22CONT.

al	nl	xl	P1	P2
**	**	**	******	******
3	14	0	.0364	.0364
	15	0	.0227	.0227
	16	0	.0130	.0130
	17	0	.0065	.0065
	18	0	.0026	.0026
		1	.0727	.0727
	19	0	.0007	.0007
		1	.0377	.0377
4	9	0	.0977	.1150
	10	0	.0677	.0964
	11	0	.0451	.0451*
	12	0	.0287	.0287
	13	0	.0172	.0172
	14	0	.0096	.0096
	15	0	.0048	.0048
		1	.0766	.0766
	16	0	.0021	.0021
		1	.0458	.0458
	17	0	.0007	.0007
		1	.0239	.0239
	18	1	.0100	.0100
5	8	0	.0760	.1154
	9	0	.0489	.0537
	10	0	.0301	.0396
	11	0	.0175	.0175*
	12	0	.0096	.0096
	13	0	.0048	.0048
		1	.0670	.1159
	14	0	.0021	.0021
		1	.0393	.0393
	15	0	.0008	.0008
		1	.0207	.0207
	16	1	.0093	.0093
	17	1	.0033	.0033
		2	.0549	.0549
6	7	0	.0671	.1207
	8	0	.0403	.0511
	9	0	.0230	.0461
	10	0	.0124	.0152
	11	0	.0062	.0062*
		1	.0743	.0805*
	12	0	.0028	.0028
		1	.0433	.0557
	13	0	.0011	.0011
		1	.0231	.0231
	14	1	.0109	.0109
		2	.0963	.1365
	15	1	.0043	.0043
		2	.0536	.0536
	16	1	.0013	.0013
		2	.0254	.0254
7	7	0	.0377	.0513
	8	0	.0201	.0225
	9	0	.0101	.0167
	10	0	.0046	.0054
		1	.0588	.0743
	11	0	.0019	.0039

TABLAS DEL TEST EXACTO DE FISHER PAG. 4

T O T A L(n)=22CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
7	11	1	.0317	.0635
	12	0	.0007	.0007
		1	.0155	.0201
	13	1	.0066	.0066
		2	.0642	.0743
	14	1	.0024	.0024
		2	.0322	.0524
	15	1	.0006	.0006
		2	.0136	.0136
8	8	0	.0094	.0177
		1	.0953	.1673
	9	0	.0040	.0055
		1	.0523	.0743
	10	0	.0016	.0017
		1	.0263	.0310
	11	0	.0005	.0010
		1	.0119	.0237
		2	.0913	.1032*
	12	1	.0046	.0062
		2	.0480	.0743
	13	1	.0015	.0015
		2	.0220	.0260
	14	2	.0083	.0083
		3	.0721	.0815
9	9	0	.0014	.0017
		1	.0247	.0306
	10	1	.0104	.0115
		2	.0820	.0991
	11	1	.0038	.0039*
		2	.0403	.0805
	12	1	.0011	.0016
		2	.0170	.0274
	13	2	.0059	.0073
		3	.0542	.0789
10	10	1	.0035	.0037
		2	.0380	.0427
	11	1	.0010	.0019
		2	.0150	.0159*
		3	.0992	.1984
	12	2	.0048	.0083
		3	.0456	.0836
11	11	2	.0045	.0046*
		3	.0431	.0861

****

T O T A L(n)=23

a1	n1	x1	P1	P2
**	**	**	******	******
1	21	0	.0870	.0870
	22	0	.0435	.0435
2	16	0	.0830	.0830
	17	0	.0593	.0593
	18	0	.0395	.0395
	19	0	.0237	.0237
	20	0	.0119	.0119
	21	0	.0040	.0040
3	12	0	.0932	.0932

T O T A L(n)=23CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
3	13	0	.0678	.0678
	14	0	.0474	.0474
	15	0	.0316	.0316
	16	0	.0198	.0198
	17	0	.0113	.0113
	18	0	.0056	.0056
	19	0	.0023	.0023
		1	.0666	.0666
	20	0	.0006	.0006
		1	.0344	.0344
4	10	0	.0807	.1045
	11	0	.0559	.0932
	12	0	.0373	.0373
	13	0	.0237	.0237
	14	0	.0142	.0142
	15	0	.0079	.0079
	16	0	.0040	.0040
		1	.0672	.0672
	17	0	.0017	.0017
		1	.0401	.0401
	18	0	.0006	.0006
		1	.0209	.0209
	19	1	.0087	.0087
5	8	0	.0892	.1221
	9	0	.0595	.1157
	10	0	.0383	.0457
	11	0	.0235	.0373
	12	0	.0137	.0137
	13	0	.0075	.0075
		1	.0886	.1269
	14	0	.0037	.0037
		1	.0562	.0562
	15	0	.0017	.0017
		1	.0329	.0329
	16	0	.0006	.0006
		1	.0173	.0173
	17	1	.0078	.0078
		2	.0886	.0886
	18	1	.0027	.0027
		2	.0482	.0482
6	7	0	.0793	.1243
	8	0	.0496	.0582
	9	0	.0298	.0481
	10	0	.0170	.0191
	11	0	.0091	.0137
		1	.0955	.1550
	12	0	.0046	.0046
		1	.0595	.0687
	13	0	.0021	.0021
		1	.0345	.0515
	14	0	.0008	.0008
		1	.0183	.0183
	15	1	.0086	.0086
		2	.0814	.1310
	16	1	.0034	.0034
		2	.0450	.0450
	17	1	.0010	.0010

T O T A L(n)=23CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
6	17	2	.0212	.0212
7	7	0	.0467	.0574
	8	0	.0263	.0520
	9	0	.0140	.0189
	10	0	.0070	.0075
		1	.0770	.0886
	11	0	.0032	.0046
		1	.0447	.0687
	12	0	.0014	.0014
		1	.0240	.0272
	13	1	.0116	.0186
		2	.0918	.1688
	14	1	.0049	.0049
		2	.0517	.0657
	15	1	.0018	.0018
		2	.0257	.0257
	16	2	.0107	.0107
		3	.0907	.1374
8	8	0	.0131	.0194
	9	0	.0061	.0072
		1	.0691	.0858
	10	0	.0026	.0059
		1	.0376	.0743
	11	0	.0010	.0014
		1	.0188	.0272
	12	1	.0084	.0094
		2	.0706	.0894
	13	1	.0033	.0033
		2	.0367	.0393
	14	1	.0011	.0011
		2	.0166	.0228
	15	2	.0062	.0062
		3	.0582	.0713
9	9	0	.0025	.0026
		1	.0355	.0397
	10	0	.0009	.0016
		1	.0166	.0288
	11	1	.0069	.0094
		2	.0602	.0894
	12	1	.0025	.0028
		2	.0291	.0361
	13	1	.0007	.0007
		2	.0122	.0131
		3	.0857	.1023
	14	2	.0042	.0066
		3	.0416	.0771
10	10	1	.0065	.0097
		2	.0571	.0903
	11	1	.0022	.0028
		2	.0260	.0361
	12	1	.0006	.0006
		2	.0101	.0123
		3	.0736	.0995
	13	2	.0032	.0034
		3	.0332	.0397
11	11	1	.0006	.0006
		2	.0095	.0123

T O T A L(n)=23CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
11	11	3	.0699	.0995
	12	2	.0028	.0033
		3	.0296	.0391

****

T O T A L(n)=24

a1	n1	x1	P1	P2
**	**	**	******	******
1	22	0	.0833	.0833
	23	0	.0417	.0417
2	17	0	.0761	.0761
	18	0	.0544	.0544
	19	0	.0362	.0362
	20	0	.0217	.0217
	21	0	.0109	.0109
	22	0	.0036	.0036
3	13	0	.0815	.0815
	14	0	.0593	.0593
	15	0	.0415	.0415
	16	0	.0277	.0277
	17	0	.0173	.0173
	18	0	.0099	.0099
	19	0	.0049	.0049
		1	.0988	.0988
	20	0	.0020	.0020
		1	.0613	.0613
	21	1	.0316	.0316
4	10	0	.0942*	.1140
	11	0	.0673	.0983
	12	0	.0466	.0466*
	13	0	.0311	.0311
	14	0	.0198	.0198
	15	0	.0119	.0119
	16	0	.0066	.0066
		1	.0909	.0909
	17	0	.0033	.0033
		1	.0593	.0593
	18	0	.0014	.0014
		1	.0353	.0353
	19	1	.0184	.0184
	20	1	.0076	.0076
5	9	0	.0706	.1181
	10	0	.0471	.0530
	11	0	.0303	.0412
	12	0	.0186	.0186*
	13	0	.0109	.0109
	14	0	.0059	.0059
		1	.0751	.1222
	15	0	.0030	.0030
		1	.0474	.0474
	16	0	.0013	.0013
		1	.0277	.0277
	17	1	.0145	.0145
	18	1	.0065	.0065
		2	.0785	.0785
	19	1	.0023	.0023
		2	.0425	.0425

TABLAS DEL TEST EXACTO DE FISHER PAG. 5

T O T A L(n)=24CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
6	7	0	.0920	.1300
	8	0	.0595	.0664
	9	0	.0372	.0519
	10	0	.0223	.0239
	11	0	.0128	.0162
	12	0	.0069	.0069*
		1	.0775	.1550
	13	0	.0034	.0034
		1	.0481	.0608
	14	0	.0016	.0016
		1	.0278	.0501
	15	0	.0006	.0006
		1	.0147	.0147
	16	1	.0069	.0069
		2	.0693	.1288
	17	1	.0027	.0027
		2	.0381	.0381
	18	1	.0008	.0008
		2	.0179	.0179
7	7	0	.0562	.0648
	8	0	.0331	.0538
	9	0	.0186	.0223
	10	0	.0099	.0188
		1	.0967	.1718
	11	0	.0050	.0059
		1	.0595	.0778
	12	0	.0023	.0023*
		1	.0343	.0687
	13	0	.0010	.0010
		1	.0183	.0233
	14	1	.0088	.0088
		2	.0751	.0850
	15	1	.0037	.0037
		2	.0420	.0606
	16	1	.0013	.0013
		2	.0207	.0207
	17	2	.0086	.0086
		3	.0774	.1336
8	8	0	.0175	.0222
	9	0	.0087	.0095
		1	.0875	.1782
	10	0	.0041	.0064
		1	.0508	.0791
	11	0	.0018	.0020
		1	.0274	.0335
	12	0	.0007	.0014
		1	.0136	.0143*
		2	.0965	.1101*
	13	1	.0061	.0078
		2	.0551	.0825
	14	1	.0024	.0024
		2	.0283	.0324
	15	1	.0007	.0007
		2	.0127	.0215
		3	.0907	.0994
	16	2	.0047	.0047
		3	.0474	.0649

T O T A L(n)=24CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
9	9	0	.0038	.0068
		1	.0481	.0803
	10	0	.0015	.0020
		1	.0245	.0333
	11	0	.0006	.0006
		1	.0114	.0131
		2	.0836	.1049
	12	1	.0047	.0094
		2	.0447	.0494*
	13	1	.0017	.0022
		2	.0214	.0327
	14	2	.0088	.0104
		3	.0673	.0918
	15	2	.0030	.0030
		3	.0322	.0361
10	10	0	.0005	.0006
		1	.0107	.0129
		2	.0796	.1041
	11	1	.0042	.0045
		2	.0403	.0472
	12	1	.0014	.0014*
		2	.0180	.0361
	13	1	.0004	.0005
		2	.0069	.0111
		3	.0551	.0953
	14	2	.0022	.0027
		3	.0244	.0352
11	11	1	.0013	.0014
		2	.0171	.0188
	12	1	.0003	.0006
		2	.0061	.0123
		3	.0498	.0995
	13	2	.0018	.0031
		3	.0207	.0377
12	12	2	.0017	.0017*
		3	.0196	.0391

****

T O T A L(n)=25

a1	n1	x1	P1	P2
**	**	**	******	******
1	23	0	.0800	.0800
	24	0	.0400	.0400
2	17	0	.0933	.0933
	18	0	.0700	.0700
	19	0	.0500	.0500
	20	0	.0333	.0333
	21	0	.0200	.0200
	22	0	.0100	.0100
	23	0	.0033	.0033
3	13	0	.0957	.0957
	14	0	.0717	.0717
	15	0	.0522	.0522
	16	0	.0365	.0365
	17	0	.0244	.0244
	18	0	.0152	.0152
	19	0	.0087	.0087

T O T A L(n)=25CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
3	20	0	.0043	.0043
		1	.0913	.0913
	21	0	.0017	.0017
		1	.0565	.0565
	22	1	.0291	.0291
4	11	0	.0791	.1052
	12	0	.0565	.0957
	13	0	.0391	.0391
	14	0	.0261	.0261
	15	0	.0166	.0166
	16	0	.0100	.0100
	17	0	.0055	.0055
		1	.0808	.0808
	18	0	.0028	.0028
		1	.0526	.0526
	19	0	.0012	.0012
		1	.0312	.0312
	20	1	.0162	.0162
	21	1	.0067	.0067
5	9	0	.0822	.1225
	10	0	.0565	.0613
	11	0	.0377	.0464
	12	0	.0242	.0391
	13	0	.0149	.0149
	14	0	.0087	.0087
		1	.0957	.1333
	15	0	.0047	.0047
		1	.0640	.1206
	16	0	.0024	.0024
		1	.0403	.0403
	17	0	.0011	.0011
		1	.0235	.0235
	18	1	.0123	.0123
	19	1	.0055	.0055
		2	.0699	.0699
	20	1	.0019	.0019
		2	.0377	.0377
6	8	0	.0699	.1292
	9	0	.0452	.0571
	10	0	.0283	.0508
	11	0	.0170	.0196
	12	0	.0097	.0149
		1	.0969	.1603
	13	0	.0052	.0052
		1	.0634	.0730
	14	0	.0026	.0026
		1	.0391	.0561
	15	0	.0012	.0012
		1	.0225	.0225
	16	1	.0119	.0119
		2	.0972	.1425
	17	1	.0055	.0055
		2	.0593	.0593
	18	1	.0022	.0022
		2	.0324	.0324
	19	1	.0007	.0007
		2	.0151	.0151

T O T A L(n)=25CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
7	7	0	.0662	.1326
	8	0	.0405	.0573
	9	0	.0238	.0267
	10	0	.0134	.0202
	11	0	.0071	.0078
		1	.0759	.0900
	12	0	.0036	.0052
		1	.0464	.0730
	13	0	.0017	.0017
		1	.0266	.0302
	14	0	.0007	.0007
		1	.0141	.0213
	15	1	.0068	.0068
		2	.0619	.0752
	16	1	.0029	.0029
		2	.0343	.0581
	17	1	.0010	.0010
		2	.0169	.0169
	18	2	.0069	.0069
		3	.0664	.0664
8	8	0	.0225	.0261
	9	0	.0119	.0218
	10	0	.0060	.0077
		1	.0655	.0875
	11	0	.0028	.0029
		1	.0377	.0421
	12	0	.0012	.0017
		1	.0202	.0302
	13	1	.0100	.0112
		2	.0766	.0968
	14	1	.0044	.0072
		2	.0433	.0810
	15	1	.0017	.0017
		2	.0221	.0280
	16	1	.0005	.0005
		2	.0099	.0099
		3	.0751	.0870
	17	2	.0037	.0037
		3	.0389	.0613
9	9	0	.0056	.0078
		1	.0623	.0875
	10	0	.0025	.0028
		1	.0340	.0405
	11	0	.0010	.0021
		1	.0172	.0330
	12	1	.0079	.0112
		2	.0634	.0968
	13	1	.0033	.0036
		2	.0335	.0414
	14	1	.0012	.0012
		2	.0159	.0168
		3	.0982	.1153
	15	2	.0065	.0089
		3	.0533	.0872
	16	2	.0022	.0022
		3	.0252	.0308
10	10	0	.0009	.0010

TABLAS DEL TEST EXACTO DE FISHER PAG. 6

T O T A L(n)=25CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
10	10	1	.0162	.0177
	11	0	.0003	.0006
		1	.0070	.0119
		2	.0576	.0992
	12	1	.0027	.0036
		2	.0287	.0414
	13	1	.0009	.0010
		2	.0127	.0154
		3	.0820	.1107
	14	2	.0048	.0051
		3	.0416	.0486
	15	2	.0015	.0024
		3	.0182	.0344
11	11	1	.0026	.0037
		2	.0273	.0419
	12	1	.0008	.0010
		2	.0114	.0154
		3	.0749	.1107
	13	2	.0040	.0048
		3	.0358	.0472
	14	2	.0012	.0012
		3	.0146	.0172
		4	.0887	.1160
12	12	2	.0038	.0048
		3	.0341	.0472
	13	2	.0010	.0012
		3	.0131	.0169
		4	.0812	.1152

****

T O T A L(n)=26

a1	n1	x1	P1	P2
**	**	**	******	******
1	24	0	.0769	.0769
	25	0	.0385	.0385
2	18	0	.0861	.0861
	19	0	.0646	.0646
	20	0	.0462	.0462
	21	0	.0308	.0308
	22	0	.0185	.0185
	23	0	.0092	.0092
	24	0	.0031	.0031
3	14	0	.0846	.0846
	15	0	.0635	.0635
	16	0	.0462	.0462
	17	0	.0323	.0323
	18	0	.0215	.0215
	19	0	.0135	.0135
	20	0	.0077	.0077
	21	0	.0039	.0039
		1	.0846	.0846
	22	0	.0015	.0015
		1	.0523	.0523
	23	1	.0269	.0269
4	11	0	.0913	.1134
	12	0	.0670	.1001
	13	0	.0478	.0478*

T O T A L(n)=26CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
4	14	0	.0331	.0331
	15	0	.0221	.0221
	16	0	.0141	.0141
	17	0	.0084	.0084
	18	0	.0047	.0047
		1	.0721	.0721
	19	0	.0023	.0023
		1	.0468	.0468
	20	0	.0010	.0010
		1	.0278	.0278
	21	1	.0144	.0144
	22	1	.0060	.0060
		2	.0987	.0987
5	9	0	.0941	.1286
	10	0	.0664	.1213
	11	0	.0457	.0527
	12	0	.0304	.0425
	13	0	.0196	.0196*
	14	0	.0120	.0120
	15	0	.0070	.0070
		1	.0823	.1279
	16	0	.0038	.0038
		1	.0549	.0549
	17	0	.0019	.0019
		1	.0345	.0345
	18	0	.0009	.0009
		1	.0200	.0200
	19	1	.0104	.0104
	20	1	.0047	.0047
		2	.0624	.0624
	21	1	.0016	.0016
		2	.0335	.0335
6	8	0	.0806	.1317
	9	0	.0538	.0634
	10	0	.0348	.0532
	11	0	.0217	.0238
	12	0	.0130	.0171
	13	0	.0074	.0149
		1	.0801	.0876*
	14	0	.0040	.0040
		1	.0522	.0652
	15	0	.0020	.0020
		1	.0321	.0539
	16	0	.0009	.0009
		1	.0184	.0184
	17	1	.0097	.0097
		2	.0841	.1379
	18	1	.0045	.0045
		2	.0510	.0510
	19	1	.0018	.0018
		2	.0278	.0278
	20	1	.0005	.0005
		2	.0129	.0129
7	7	0	.0766	.1338
	8	0	.0484	.0622
	9	0	.0296	.0578
	10	0	.0174	.0227

T O T A L(n)=26CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
7	11	0	.0098	.0208
		1	.0935	.1783
	12	0	.0052	.0064
		1	.0600	.0809
	13	0	.0026	.0052
		1	.0365	.0730
	14	0	.0012	.0012
		1	.0209	.0261
	15	0	.0005	.0005
		1	.0110	.0110
		2	.0848	.0946
	16	1	.0053	.0053
		2	.0513	.0687
	17	1	.0022	.0022
		2	.0283	.0283
	18	1	.0008	.0008
		2	.0138	.0138
	19	2	.0057	.0057
		3	.0572	.0572
8	8	0	.0280	.0601
	9	0	.0156	.0233
	10	0	.0082	.0095
		1	.0815	.0989
	11	0	.0041	.0074
		1	.0494	.0838
	12	0	.0019	.0022
		1	.0283	.0357
	13	0	.0008	.0008*
		1	.0151	.0302
	14	1	.0074	.0093
		2	.0612	.0895
	15	1	.0033	.0033
		2	.0343	.0384
	16	1	.0013	.0013
		2	.0174	.0256
	17	2	.0077	.0077
		3	.0626	.0781
	18	2	.0028	.0028
		3	.0321	.0321
9	9	0	.0078	.0094
		1	.0778	.0977
	10	0	.0037	.0085
		1	.0449	.0873
	11	0	.0016	.0024
		1	.0243	.0362
	12	0	.0006	.0007
		1	.0122	.0145
		2	.0847	.1100
	13	1	.0056	.0058*
		2	.0484	.0968
	14	1	.0023	.0029
		2	.0254	.0375
	15	1	.0008	.0008
		2	.0119	.0135
		3	.0792	.1034
	16	2	.0048	.0048
		3	.0425	.0461

T O T A L(n)=26CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
9	17	2	.0016	.0016
		3	.0199	.0277
10	10	0	.0015	.0026
		1	.0230	.0367
	11	0	.0006	.0007
		1	.0109	.0144
		2	.0776	.1093
	12	1	.0047	.0053
		2	.0420	.0511
	13	1	.0018	.0036
		2	.0207	.0225*
	14	1	.0006	.0008
		2	.0091	.0138
		3	.0634	.1054
	15	2	.0034	.0040
		3	.0317	.0426
	16	2	.0011	.0011
		3	.0137	.0152
		4	.0857	.1087
11	11	1	.0045	.0052
		2	.0401	.0506
	12	1	.0016	.0017
		2	.0187	.0214
	13	1	.0005	.0010
		2	.0077	.0082*
		3	.0554	.1107
	14	2	.0027	.0043
		3	.0260	.0447
	15	2	.0008	.0010
		3	.0105	.0149
		4	.0688	.1089
12	12	2	.0073	.0079
		3	.0529	.0618
	13	2	.0024	.0025*
		3	.0236	.0472
	14	2	.0006	.0011
		3	.0089	.0162
		4	.0602	.1131
13	13	2	.0006	.0006*
		3	.0085	.0169
		4	.0576	.0661*

****

T O T A L(n)=27

a1	n1	x1	P1	P2
**	**	**	******	******
1	25	0	.0741	.0741
	26	0	.0370	.0370
2	19	0	.0798	.0798
	20	0	.0598	.0598
	21	0	.0427	.0427
	22	0	.0285	.0285
	23	0	.0171	.0171
	24	0	.0085	.0085
	25	0	.0029	.0029
3	14	0	.0978	.0978
	15	0	.0752	.0752

TABLAS DEL TEST EXACTO DE FISHER PAG. 7

T O T A L(n)=27CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
3	16	0	.0564	.0564
	17	0	.0410	.0410
	18	0	.0287	.0287
	19	0	.0192	.0192
	20	0	.0120	.0120
	21	0	.0068	.0068
	22	0	.0034	.0034
		1	.0786	.0786
	23	0	.0014	.0014
		1	.0486	.0486
	24	1	.0250	.0250
4	12	0	.0778	.1060
	13	0	.0570	.0978
	14	0	.0407	.0407
	15	0	.0282	.0282
	16	0	.0188	.0188
	17	0	.0120	.0120
	18	0	.0072	.0072
		1	.0933	.0933
	19	0	.0040	.0040
		1	.0646	.0646
	20	0	.0020	.0020
		1	.0419	.0419
	21	0	.0009	.0009
		1	.0248	.0248
	22	1	.0128	.0128
	23	1	.0053	.0053
		2	.0918	.0918
5	10	0	.0767	.1240
	11	0	.0541	.0598
	12	0	.0372	.0470
	13	0	.0248	.0407
	14	0	.0159	.0159
	15	0	.0098	.0098
	16	0	.0057	.0057
		1	.0711	.1252
	17	0	.0031	.0031
		1	.0473	.0473
	18	0	.0016	.0016
		1	.0297	.0297
	19	0	.0007	.0007
		1	.0172	.0172
	20	1	.0089	.0089
		2	.0913	.0913
	21	1	.0040	.0040
		2	.0560	.0560
	22	1	.0014	.0014
		2	.0300	.0300
6	8	0	.0917	.1358
	9	0	.0627	.0707
	10	0	.0418	.0570
	11	0	.0271	.0536
	12	0	.0169	.0200
	13	0	.0101	.0159
		1	.0981	.1647
	14	0	.0058	.0058
		1	.0667	.0768

T O T A L(n)=27CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
6	15	0	.0031	.0031
		1	.0433	.0602
	16	0	.0016	.0016
		1	.0265	.0265
	17	0	.0007	.0007
		1	.0152	.0152
	18	1	.0080	.0080
		2	.0731	.1358
	19	1	.0037	.0037
		2	.0441	.0441
	20	1	.0014	.0014
		2	.0239	.0239
	21	2	.0111	.0111
7	7	0	.0873	.1369
	8	0	.0567	.0681
	9	0	.0358	.0593
	10	0	.0219	.0261
	11	0	.0129	.0216
	12	0	.0073	.0081
		1	.0749	.0914
	13	0	.0039	.0058
		1	.0478	.0768
	14	0	.0019	.0019
		1	.0290	.0329
	15	0	.0009	.0009
		1	.0165	.0238
	16	1	.0087	.0087
		2	.0711	.0840
	17	1	.0042	.0042
		2	.0428	.0647
	18	1	.0017	.0017
		2	.0235	.0235
	19	1	.0006	.0006
		2	.0114	.0114
		3	.0878	.1445
	20	2	.0047	.0047
		3	.0496	.0496
8	8	0	.0340	.0607
	9	0	.0197	.0258
	10	0	.0110	.0248
		1	.0986	.1895
	11	0	.0058	.0083
		1	.0625	.0899
	12	0	.0029	.0031
		1	.0377	.0433
	13	0	.0014	.0019
		1	.0215	.0329
	14	0	.0006	.0006
		1	.0114	.0128
		2	.0817	.1032
	15	1	.0056	.0085
		2	.0493	.0870
	16	1	.0025	.0025
		2	.0274	.0332
	17	1	.0009	.0009
		2	.0138	.0138
		3	.0910	.1019

T O T A L(n)=27CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
8	18	2	.0061	.0061
		3	.0524	.0721
	19	2	.0022	.0022
		3	.0267	.0267
9	9	0	.0104	.0116
		1	.0944	.1925
	10	0	.0052	.0088
		1	.0571	.0912
	11	0	.0024	.0030
		1	.0327	.0417
	12	0	.0011	.0011
		1	.0175	.0192
	13	0	.0004	.0006
		1	.0088	.0128
		2	.0659	.1032
	14	1	.0040	.0044
		2	.0373	.0461
	15	1	.0016	.0027
		2	.0194	.0369
	16	1	.0006	.0006
		2	.0090	.0115
		3	.0642	.0969
	17	2	.0037	.0037
		3	.0341	.0393
	18	2	.0012	.0012
		3	.0158	.0262
		4	.0981	.1085
10	10	0	.0023	.0031
		1	.0311	.0415
	11	0	.0010	.0011
		1	.0159	.0183
		2	.0998	.1241
	12	0	.0004	.0007
		1	.0075	.0140
		2	.0578	.1071
	13	1	.0032	.0044
		2	.0310	.0461
	14	1	.0012	.0013
		2	.0151	.0183
		3	.0891	.1201
	15	2	.0066	.0069
		3	.0493	.0568
	16	2	.0025	.0034
		3	.0244	.0402
	17	2	.0007	.0007
		3	.0104	.0127
		4	.0697	.1008
11	11	0	.0003	.0009
		1	.0071	.0147
		2	.0554	.1090
	12	1	.0029	.0047
		2	.0282	.0473
	13	1	.0010	.0013
		2	.0130	.0183
		3	.0789	.1201
	14	2	.0053	.0063
		3	.0412	.0542

T O T A L(n)=27CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
11	15	2	.0019	.0020
		3	.0191	.0220
	16	2	.0005	.0005
		3	.0076	.0079
		4	.0537	.0608
12	12	1	.0010	.0014
		2	.0124	.0185
		3	.0757	.1208
	13	2	.0048	.0063
		3	.0377	.0542
	14	2	.0016	.0018
		3	.0165	.0213
		4	.0906	.1284
	15	3	.0062	.0071
		4	.0450	.0574
13	13	2	.0015	.0018
		3	.0158	.0213
		4	.0871	.1284
	14	3	.0056	.0070
		4	.0412	.0570

****

T O T A L L(n)=28

a1	n1	x1	P1	P2
**	**	**	******	******
1	26	0	.0714	.0714
	27	0	.0357	.0357
2	19	0	.0952	.0952
	20	0	.0741	.0741
	21	0	.0556	.0556
	22	0	.0397	.0397
	23	0	.0265	.0265
	24	0	.0159	.0159
	25	0	.0079	.0079
	26	0	.0027	.0027
3	15	0	.0873	.0873
	16	0	.0672	.0672
	17	0	.0504	.0504
	18	0	.0366	.0366
	19	0	.0256	.0256
	20	0	.0171	.0171
	21	0	.0107	.0107
	22	0	.0061	.0061
	23	0	.0031	.0031
		1	.0733	.0733
	24	0	.0012	.0012
		1	.0452	.0452
	25	1	.0232	.0232
4	12	0	.0889	.1131
	13	0	.0667	.1016
	14	0	.0489	.0489*
	15	0	.0349	.0349
	16	0	.0242	.0242
	17	0	.0161	.0161
	18	0	.0103	.0103
	19	0	.0062	.0062
		1	.0841	.0841

TABLAS DEL TEST EXACTO DE FISHER PAG. 8

T O T A L(n)=28CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
4	20	0	.0034	.0034
		1	.0581	.0581
	21	0	.0017	.0017
		1	.0376	.0376
	22	0	.0007	.0007
		1	.0222	.0222
	23	1	.0115	.0115
	24	1	.0047	.0047
		2	.0856	.0856
5	10	0	.0872	.1282
	11	0	.0630	.1248
	12	0	.0444	.0525
	13	0	.0306	.0437
	14	0	.0204	.0407
	15	0	.0131	.0131
	16	0	.0081	.0081
		1	.0886	.1331
	17	0	.0047	.0047
		1	.0618	.0618
	18	0	.0026	.0026
		1	.0410	.0410
	19	0	.0013	.0013
		1	.0256	.0256
	20	0	.0006	.0006
		1	.0148	.0148
	21	1	.0077	.0077
		2	.0825	.0825
	22	1	.0034	.0034
		2	.0504	.0504
	23	1	.0012	.0012
		2	.0269	.0269
6	9	0	.0720	.1358
	10	0	.0493	.0619
	11	0	.0329	.0549
	12	0	.0213	.0237
	13	0	.0133	.0178
	14	0	.0080	.0159
		1	.0824	.0903*
	15	0	.0045	.0045
		1	.0558	.0691
	16	0	.0025	.0025
		1	.0361	.0573
	17	0	.0012	.0012
		1	.0221	.0221
	18	0	.0006	.0006
		1	.0126	.0126
		2	.0979	.1472
	19	1	.0066	.0066
		2	.0638	.0638
	20	1	.0031	.0031
		2	.0384	.0384
	21	1	.0012	.0012
		2	.0207	.0207
	22	2	.0096	.0096
		3	.0913	.0913
7	7	0	.0982	.1414
	8	0	.0655	.0749

T O T A L(n)=28CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
7	9	0	.0426	.0621
	10	0	.0269	.0302
	11	0	.0164	.0233
	12	0	.0097	.0103
		1	.0908	.1842
	13	0	.0054	.0069
		1	.0604	.0836
	14	0	.0029	.0058
		1	.0384	.0768
	15	0	.0015	.0015
		1	.0232	.0286
	16	0	.0007	.0007
		1	.0132	.0228
		2	.0934	.1031
	17	1	.0069	.0069
		2	.0600	.0764
	18	1	.0033	.0033
		2	.0359	.0627
	19	1	.0014	.0014
		2	.0196	.0196
	20	2	.0095	.0095
		3	.0769	.1423
	21	2	.0039	.0039
		3	.0432	.0432
8	8	0	.0405	.0628
	9	0	.0243	.0292
	10	0	.0141	.0251
	11	0	.0078	.0097
		1	.0767	.0987
	12	0	.0041	.0084
		1	.0483	.0882
	13	0	.0021	.0025
		1	.0290	.0377
	14	0	.0010	.0019
		1	.0164	.0329
	15	1	.0087	.0108
		2	.0667	.0957
	16	1	.0042	.0042
		2	.0399	.0441
	17	1	.0019	.0019
		2	.0221	.0299
	18	1	.0007	.0007
		2	.0111	.0111
		3	.0772	.0913
	19	2	.0048	.0048
		3	.0441	.0684
	20	2	.0018	.0018
		3	.0223	.0223
9	9	0	.0134	.0261
	10	0	.0070	.0098
		1	.0704	.0980
	11	0	.0035	.0039
		1	.0422	.0492
	12	0	.0017	.0028
		1	.0240	.0390
	13	0	.0007	.0008
		1	.0128	.0157

T O T A L(n)=28CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
9	13	2	.0855	.1145
	14	0	.0003	.0003*
		1	.0064	.0128
		2	.0516	.1032
	15	1	.0029	.0036
		2	.0290	.0418
	16	1	.0012	.0012
		2	.0149	.0166
		3	.0899	.1139
	17	2	.0069	.0104
		3	.0524	.0946
	18	2	.0028	.0028
		3	.0276	.0346
	19	2	.0009	.0009
		3	.0127	.0127
		4	.0834	.0968
10	10	0	.0033	.0039
		1	.0404	.0484
	11	0	.0015	.0033
		1	.0219	.0407
	12	0	.0006	.0009
		1	.0111	.0159
		2	.0758	.1144
	13	1	.0052	.0060
		2	.0434	.0546
	14	1	.0022	.0044
		2	.0230	.0461
	15	1	.0008	.0011
		2	.0111	.0163
		3	.0706	.1141
	16	2	.0048	.0054
		3	.0386	.0497
	17	2	.0018	.0018
		3	.0189	.0204
	18	2	.0005	.0005
		3	.0080	.0113
		4	.0570	.0974
11	11	0	.0006	.0009
		1	.0105	.0161
		2	.0728	.1150
	12	1	.0047	.0060
		2	.0398	.0540
	13	1	.0019	.0021
		2	.0201	.0238
	14	1	.0007	.0007*
		2	.0091	.0098*
		3	.0601	.1201
	15	2	.0037	.0056
		3	.0310	.0510
	16	2	.0013	.0015
		3	.0142	.0189
		4	.0813	.1212
	17	3	.0056	.0062
		4	.0422	.0527
12	12	1	.0018	.0021
		2	.0192	.0235
	13	1	.0006	.0006

T O T A L(n)=28CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
12	13	2	.0083	.0093
		3	.0554	.0671
	14	2	.0032	.0033*
		3	.0271	.0542
	15	2	.0010	.0016
		3	.0117	.0200
		4	.0695	.1248
	16	3	.0043	.0061
		4	.0339	.0531
13	13	2	.0030	.0032
		3	.0259	.0296
	14	2	.0009	.0018
		3	.0107	.0116*
		4	.0642	.0748*
	15	3	.0037	.0067
		4	.0298	.0557
14	14	3	.0035	.0070
		4	.0285	.0320*

****

T O T A L(n)=29

a1	n1	x1	P1	P2
**	**	**	******	******
1	27	0	.0690	.0690
	28	0	.0345	.0345
2	20	0	.0887	.0887
	21	0	.0690	.0690
	22	0	.0517	.0517
	23	0	.0370	.0370
	24	0	.0246	.0246
	25	0	.0148	.0148
	26	0	.0074	.0074
	27	0	.0025	.0025
3	15	0	.0996	.0996
	16	0	.0783	.0783
	17	0	.0602	.0602
	18	0	.0452	.0452
	19	0	.0328	.0328
	20	0	.0230	.0230
	21	0	.0153	.0153
	22	0	.0096	.0096
	23	0	.0055	.0055
		1	.0999	.0999
	24	0	.0027	.0027
		1	.0684	.0684
	25	0	.0011	.0011
		1	.0422	.0422
	26	1	.0216	.0216
4	13	0	.0766	.1067
	14	0	.0575	.0996
	15	0	.0422	.0422
	16	0	.0301	.0301
	17	0	.0208	.0208
	18	0	.0139	.0139
	19	0	.0088	.0088
	20	0	.0053	.0053
		1	.0760	.0760

TABLAS DEL TEST EXACTO DE FISHER PAG. 9

T O T A L(n)=29CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
4	21	0	.0030	.0030
		1	.0525	.0525
	22	0	.0015	.0015
		1	.0339	.0339
	23	0	.0006	.0006
		1	.0200	.0200
	24	1	.0103	.0103
	25	1	.0043	.0043
		2	.0800	.0800
5	10	0	.0979	.1336
	11	0	.0721	.1261
	12	0	.0521	.0588
	13	0	.0368	.0476
	14	0	.0253	.0422
	15	0	.0169	.0169
	16	0	.0108	.0108
	17	0	.0067	.0067
		1	.0775	.1296
	18	0	.0039	.0039
		1	.0539	.0539
	19	0	.0021	.0021
		1	.0357	.0357
	20	0	.0011	.0011
		1	.0223	.0223
	21	1	.0129	.0129
	22	1	.0067	.0067
		2	.0747	.0747
	23	1	.0030	.0030
		2	.0456	.0456
	24	1	.0010	.0010
		2	.0243	.0243
6	9	0	.0816	.1375
	10	0	.0571	.0676
	11	0	.0391	.0576
	12	0	.0261	.0563
	13	0	.0169	.0205
	14	0	.0105	.0169
		1	.0990	.1686
	15	0	.0063	.0063
		1	.0695	.0801
	16	0	.0036	.0036
		1	.0470	.0638
	17	0	.0020	.0020
		1	.0303	.0303
	18	0	.0010	.0010
		1	.0185	.0185
	19	1	.0105	.0105
		2	.0861	.1432
	20	1	.0055	.0055
		2	.0559	.0559
	21	1	.0025	.0025
		2	.0335	.0335
	22	1	.0010	.0010
		2	.0180	.0180
	23	2	.0083	.0083
		3	.0828	.0828
7	8	0	.0745	.1421

T O T A L(n)=29CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
7	9	0	.0497	.0661
	10	0	.0323	.0625
	11	0	.0204	.0259
	12	0	.0125	.0230
	13	0	.0073	.0084
		1	.0740	.0927
	14	0	.0041	.0063
		1	.0490	.0801
	15	0	.0022	.0022
		1	.0311	.0352
	16	0	.0011	.0011
		1	.0187	.0260
	17	0	.0005	.0005
		1	.0106	.0106
		2	.0796	.0920
	18	1	.0055	.0055
		2	.0508	.0712
	19	1	.0026	.0026
		2	.0302	.0302
	20	1	.0011	.0011
		2	.0164	.0164
	21	2	.0079	.0079
		3	.0676	.0676
	22	2	.0032	.0032
		3	.0377	.0377
8	8	0	.0474	.0662
	9	0	.0294	.0667
	10	0	.0176	.0265
	11	0	.0102	.0116
		1	.0918	.1097
	12	0	.0057	.0089
		1	.0600	.0926
	13	0	.0030	.0033
		1	.0377	.0443
	14	0	.0015	.0022
		1	.0225	.0352
	15	0	.0007	.0007
		1	.0127	.0142
		2	.0862	.1087
	16	1	.0067	.0097
		2	.0547	.0923
	17	1	.0033	.0033
		2	.0325	.0382
	18	1	.0014	.0014
		2	.0179	.0281
	19	1	.0005	.0005
		2	.0089	.0089
		3	.0658	.0834
	20	2	.0039	.0039
		3	.0374	.0374
	21	2	.0014	.0014
		3	.0188	.0188
9	9	0	.0168	.0270
	10	0	.0092	.0114
		1	.0847	.1072
	11	0	.0049	.0102
		1	.0529	.0959

T O T A L(n)=29CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
9	12	0	.0024	.0033
		1	.0316	.0432
	13	0	.0011	.0012
		1	.0179	.0200
	14	0	.0005	.0007
		1	.0095	.0142
		2	.0680	.1087
	15	1	.0047	.0052
		2	.0407	.0502
	16	1	.0021	.0033
		2	.0227	.0405
	17	1	.0009	.0009
		2	.0116	.0140
		3	.0744	.1059
	18	2	.0053	.0053
		3	.0430	.0478
	19	2	.0021	.0021
		3	.0225	.0317
	20	2	.0007	.0007
		3	.0103	.0103
		4	.0712	.0880
10	10	0	.0046	.0108
		1	.0507	.0976
	11	0	.0022	.0035
		1	.0289	.0436
	12	0	.0010	.0012
		1	.0155	.0191
		2	.0956	.1261
	13	0	.0004	.0010
		1	.0078	.0161
		2	.0579	.1142
	14	1	.0037	.0052
		2	.0329	.0502
	15	1	.0016	.0017
		2	.0173	.0209
		3	.0953	.1281
	16	1	.0006	.0006
		2	.0083	.0087
		3	.0563	.0641
	17	2	.0036	.0045
		3	.0304	.0460
	18	2	.0013	.0013
		3	.0148	.0169
		4	.0853	.1142
	19	3	.0062	.0062
		4	.0468	.0515
11	11	0	.0009	.0012
		1	.0148	.0190
		2	.0921	.1255
	12	1	.0071	.0080
		2	.0535	.0641
	13	1	.0031	.0058
		2	.0289	.0524
	14	1	.0013	.0017
		2	.0144	.0209
		3	.0821	.1281
	15	2	.0065	.0078

T O T A L(n)=29CONT.

a1	n1	x1	P1	P2
**	**	**	******	******
11	15	3	.0460	.0604
	16	2	.0026	.0027
		3	.0235	.0266
	17	2	.0009	.0013
		3	.0106	.0177
		4	.0651	.1186
	18	3	.0041	.0051
		4	.0333	.0482
12	12	1	.0030	.0060
		2	.0277	.0535
	13	1	.0011	.0018
		2	.0132	.0216
		3	.0762	.1298
	14	2	.0056	.0078
		3	.0408	.0604
	15	2	.0021	.0025
		3	.0197	.0253
		4	.0987	.1394
	16	2	.0007	.0007
		3	.0084	.0095
		4	.0535	.0667
	17	3	.0031	.0032
		4	.0258	.0287
13	13	2	.0054	.0079
		3	.0391	.0608
	14	2	.0019	.0025
		3	.0180	.0253
		4	.0919	.1394
	15	2	.0006	.0007
		3	.0073	.0092
		4	.0476	.0656
	16	3	.0025	.0029
		4	.0217	.0271
14	14	2	.0006	.0007
		3	.0070	.0092
		4	.0457	.0656
	15	3	.0023	.0028
		4	.0199	.0268
		5	.0974	.1431

****

T O T A L(n)=30

a1	n1	x1	P1	P2
**	**	**	******	******
1	27	0	.1000	.1000
	28	0	.0667	.0667
	29	0	.0333	.0333
2	21	0	.0828	.0828
	22	0	.0644	.0644
	23	0	.0483	.0483
	24	0	.0345	.0345
	25	0	.0230	.0230
	26	0	.0138	.0138
	27	0	.0069	.0069
	28	0	.0023	.0023
3	16	0	.0897	.0897
	17	0	.0704	.0704

TABLAS DEL TEST EXACTO DE FISHER PAG. 10

T O T A L(n)=30CONT.

al	n1	x1	P1	P2
3	18	0	.0542	.0542
	19	0	.0406	.0406
	20	0	.0296	.0296
	21	0	.0207	.0207
	22	0	.0138	.0138
	23	0	.0086	.0086
	24	0	.0049	.0049
		1	.0936	.0936
	25	0	.0025	.0025
		1	.0640	.0640
	26	0	.0010	.0010
		1	.0394	.0394
	27	1	.0202	.0202
4	13	0	.0868	.1129
	14	0	.0664	.1029
	15	0	.0498	.0996
	16	0	.0365	.0365
	17	0	.0261	.0261
	18	0	.0181	.0181
	19	0	.0120	.0120
	20	0	.0077	.0077
		1	.0952	.0952
	21	0	.0046	.0046
		1	.0690	.0690
	22	0	.0026	.0026
		1	.0475	.0475
	23	0	.0013	.0013
		1	.0307	.0307
	24	0	.0006	.0006
		1	.0181	.0181
	25	1	.0093	.0093
	26	1	.0038	.0038
		2	.0750	.0750
5	11	0	.0816	.1288
	12	0	.0601	.0657
	13	0	.0434	.0525
	14	0	.0307	.0447
	15	0	.0211	.0422
	16	0	.0141	.0141
	17	0	.0090	.0090
		1	.0943	.1378
	18	0	.0056	.0056
		1	.0681	.1282
	19	0	.0032	.0032
		1	.0472	.0472
	20	0	.0018	.0018
		1	.0312	.0312
	21	0	.0009	.0009
		1	.0195	.0195
	22	1	.0112	.0112
	23	1	.0058	.0058
		2	.0679	.0679
	24	1	.0026	.0026
		2	.0413	.0413
	25	1	.0009	.0009
		2	.0219	.0219
6	9	0	.0914	.1406

T O T A L(n)=30CONT.

al	n1	x1	P1	P2
6	10	0	.0653	.0741
	11	0	.0457	.0613
	12	0	.0313	.0568
	13	0	.0208	.0237
	14	0	.0135	.0185
	15	0	.0084	.0169
		1	.0843	.0927*
	16	0	.0051	.0051
		1	.0590	.0725
	17	0	.0029	.0029
		1	.0397	.0606
	18	0	.0016	.0016
		1	.0256	.0256
	19	0	.0008	.0008
		1	.0156	.0156
	20	1	.0088	.0088
		2	.0760	.1413
	21	1	.0046	.0046
		2	.0492	.0492
	22	1	.0021	.0021
		2	.0294	.0294
	23	1	.0008	.0008
		2	.0157	.0157
	24	2	.0072	.0072
		3	.0754	.0754
7	8	0	.0838	.1434
	9	0	.0571	.0710
	10	0	.0381	.0637
	11	0	.0248	.0292
	12	0	.0156	.0242
	13	0	.0096	.0104
		1	.0886	.1038
	14	0	.0056	.0073
		1	.0607	.0860
	15	0	.0032	.0032*
		1	.0400	.0801
	16	0	.0017	.0017
		1	.0253	.0309
	17	0	.0008	.0008
		1	.0152	.0247
	18	1	.0086	.0086
		2	.0681	.0837
	19	1	.0045	.0045
		2	.0433	.0680
	20	1	.0021	.0021
		2	.0256	.0256
	21	1	.0009	.0009
		2	.0139	.0139
		3	.0962	.1533
	22	2	.0067	.0067
		3	.0596	.0596
	23	2	.0027	.0027
		3	.0331	.0331
8	8	0	.0546	.0705
	9	0	.0348	.0665
	10	0	.0215	.0288
	11	0	.0129	.0275

T O T A L(n)=30CONT.

al	n1	x1	P1	P2
8	12	0	.0075	.0100
		1	.0727	.0994
	13	0	.0041	.0094
		1	.0474	.0924
	14	0	.0022	.0027
		1	.0296	.0395
	15	0	.0011	.0022
		1	.0176	.0352
	16	0	.0005	.0005
		1	.0099	.0121
		2	.0715	.1010
	17	1	.0052	.0052
		2	.0451	.0492
	18	1	.0025	.0025
		2	.0267	.0342
	19	1	.0011	.0011
		2	.0146	.0146
		3	.0911	.1040
	20	2	.0072	.0072
		3	.0563	.0778
	21	2	.0031	.0031
		3	.0318	.0318
	22	2	.0011	.0011
		3	.0159	.0159
9	9	0	.0205	.0289
	10	0	.0117	.0134
		1	.0998	.2035
	11	0	.0065	.0106
		1	.0646	.1000
	12	0	.0034	.0040
		1	.0401	.0492
	13	0	.0017	.0033
		1	.0238	.0417
	14	0	.0008	.0009
		1	.0134	.0169
		2	.0862	.1184
	15	0	.0004	.0007
		1	.0071	.0074*
		2	.0543	.1087
	16	1	.0035	.0043
		2	.0323	.0457
	17	1	.0016	.0016
		2	.0179	.0196
		3	.0994	.1232
	18	1	.0006	.0006
		2	.0091	.0125
		3	.0618	.1019
	19	2	.0042	.0042
		3	.0355	.0419
	20	2	.0017	.0017
		3	.0184	.0301
	21	2	.0005	.0005
		3	.0084	.0084
		4	.0611	.0816
10	10	0	.0062	.0110
		1	.0621	.1008
	11	0	.0031	.0041

T O T A L(n)=30CONT.

al	n1	x1	P1	P2
10	11	1	.0369	.0485
	12	0	.0015	.0016
		1	.0209	.0235
	13	0	.0007	.0011
		1	.0112	.0174
		2	.0743	.1194
	14	1	.0056	.0067
		2	.0446	.0577
	15	1	.0026	.0027*
		2	.0251	.0502
	16	1	.0011	.0014
		2	.0131	.0187
		3	.0771	.1216
	17	2	.0062	.0069
		3	.0451	.0562
	18	2	.0027	.0041
		3	.0242	.0450
	19	2	.0010	.0010
		3	.0116	.0147
		4	.0712	.1081
	20	3	.0048	.0048
		4	.0387	.0449
11	11	0	.0014	.0016
		1	.0200	.0231
	12	0	.0006	.0012
		1	.0102	.0182
		2	.0689	.1213
	13	1	.0049	.0067
		2	.0396	.0575
	14	1	.0021	.0024
		2	.0212	.0259
	15	1	.0009	.0017
		2	.0105	.0209
		3	.0641	.1281
	16	2	.0047	.0068
		3	.0355	.0567
	17	2	.0019	.0021
		3	.0179	.0228
		4	.0927	.1322
	18	2	.0006	.0006
		3	.0080	.0086
		4	.0524	.0626
	19	3	.0031	.0045
		4	.0265	.0465
12	12	1	.0046	.0068
		2	.0380	.0577
	13	1	.0019	.0024
		2	.0195	.0256
		3	.0999	.1414
	14	1	.0007	.0008
		2	.0091	.0106
		3	.0573	.0717
	15	2	.0039	.0041*
		3	.0302	.0341*
	16	2	.0015	.0022
		3	.0144	.0236
		4	.0776	.1349

TABLAS DEL TEST EXACTO DE FISHER PAG. 11

T O T A L(n)=30CONT.

a1	n1	x1	P1	P2
12	17	2	.0005	.0005
		3	.0061	.0080
		4	.0415	.0610
	18	3	.0022	.0024
		4	.0197	.0243
		5	.0982	.1362
13	13	1	.0007	.0008
		2	.0087	.0105
		3	.0552	.0711
	14	2	.0035	.0039
		3	.0279	.0329
	15	2	.0013	.0013*
		3	.0127	.0253
		4	.0697	.1394
	16	2	.0004	.0006
		3	.0051	.0086
		4	.0355	.0634
	17	3	.0017	.0024
		4	.0159	.0247
		5	.0824	.1376
14	14	2	.0012	.0013
		3	.0121	.0136
		4	.0673	.0813
	15	2	.0003	.0007
		3	.0046	.0092
		4	.0328	.0374*
	16	3	.0015	.0027
		4	.0140	.0261
		5	.0741	.1414
15	15	3	.0014	.0015*
		4	.0134	.0268
		5	.0716	.1431

****

T O T A L L(n)=31

a1	n1	x1	P1	P2
1	28	0	.0968	.0968
	29	0	.0645	.0645
	30	0	.0323	.0323
2	21	0	.0968	.0968
	22	0	.0774	.0774
	23	0	.0602	.0602
	24	0	.0452	.0452
	25	0	.0323	.0323
	26	0	.0215	.0215
	27	0	.0129	.0129
	28	0	.0065	.0065
	29	0	.0022	.0022
3	17	0	.0810	.0810
	18	0	.0636	.0636
	19	0	.0489	.0489
	20	0	.0367	.0367
	21	0	.0267	.0267
	22	0	.0187	.0187
	23	0	.0125	.0125
	24	0	.0078	.0078

T O T A L L(n)=31CONT.

a1	n1	x1	P1	P2
3	25	0	.0045	.0045
		1	.0879	.0879
	26	0	.0022	.0022
		1	.0601	.0601
	27	0	.0009	.0009
		1	.0369	.0369
	28	1	.0189	.0189
4	13	0	.0972	.1200
	14	0	.0756	.1075
	15	0	.0578	.1012
	16	0	.0434	.0434
	17	0	.0318	.0318
	18	0	.0227	.0227
	19	0	.0157	.0157
	20	0	.0105	.0105
	21	0	.0067	.0067
		1	.0868	.0868
	22	0	.0040	.0040
		1	.0627	.0627
	23	0	.0022	.0022
		1	.0432	.0432
	24	0	.0011	.0011
		1	.0278	.0278
	25	1	.0164	.0164
	26	1	.0084	.0084
	27	1	.0035	.0035
		2	.0704	.0704
5	11	0	.0913	.1328
	12	0	.0684	.1285
	13	0	.0504	.0580
	14	0	.0364	.0482
	15	0	.0257	.0434
	16	0	.0177	.0177
	17	0	.0118	.0118
	18	0	.0076	.0076
		1	.0833	.1338
	19	0	.0047	.0047
		1	.0600	.0600
	20	0	.0027	.0027
		1	.0416	.0416
	21	0	.0015	.0015
		1	.0274	.0274
	22	0	.0007	.0007
		1	.0171	.0171
	23	1	.0098	.0098
		2	.0932	.0932
	24	1	.0051	.0051
		2	.0619	.0619
	25	1	.0022	.0022
		2	.0376	.0376
	26	1	.0008	.0008
		2	.0199	.0199
6	10	0	.0737	.1411
	11	0	.0526	.0658
	12	0	.0369	.0585
	13	0	.0252	.0275
	14	0	.0168	.0209

T O T A L L(n)=31CONT.

a1	n1	x1	P1	P2
6	15	0	.0109	.0177
		1	.0999	.1719
	16	0	.0068	.0068
		1	.0721	.0829
	17	0	.0041	.0041
		1	.0503	.0671
	18	0	.0023	.0023
		1	.0338	.0590
	19	0	.0013	.0013
		1	.0217	.0217
	20	0	.0006	.0006
		1	.0132	.0132
		2	.0983	.1510
	21	1	.0075	.0075
		2	.0674	.0674
	22	1	.0039	.0039
		2	.0434	.0434
	23	1	.0018	.0018
		2	.0258	.0258
	24	1	.0007	.0007
		2	.0138	.0138
	25	2	.0063	.0063
		3	.0688	.0688
7	8	0	.0932	.1460
	9	0	.0649	.0766
	10	0	.0442	.0661
	11	0	.0295	.0665
	12	0	.0192	.0261
	13	0	.0121	.0245
	14	0	.0074	.0087
		1	.0733	.0940
	15	0	.0043	.0068
		1	.0500	.0829
	16	0	.0025	.0025
		1	.0329	.0373
	17	0	.0013	.0013
		1	.0207	.0281
	18	0	.0007	.0007
		1	.0124	.0124
		2	.0873	.0994
	19	1	.0070	.0070
		2	.0585	.0776
	20	1	.0036	.0036
		2	.0370	.0370
	21	1	.0017	.0017
		2	.0219	.0219
	22	1	.0007	.0007
		2	.0118	.0118
		3	.0856	.1504
	23	2	.0056	.0056
		3	.0528	.0528
	24	2	.0023	.0023
		3	.0292	.0292
8	8	0	.0622	.0756
	9	0	.0405	.0677
	10	0	.0258	.0317
	11	0	.0160	.0280

T O T A L L(n)=31CONT.

a1	n1	x1	P1	P2
8	12	0	.0096	.0116
		1	.0862	.1082
	13	0	.0056	.0096
		1	.0580	.0954
	14	0	.0031	.0035
		1	.0376	.0454
	15	0	.0016	.0025
		1	.0234	.0373
	16	0	.0008	.0008
		1	.0139	.0155
		2	.0900	.1134
	17	1	.0078	.0109
		2	.0596	.0971
	18	1	.0041	.0041
		2	.0374	.0429
	19	1	.0020	.0020
		2	.0220	.0316
	20	1	.0009	.0009
		2	.0120	.0120
		3	.0788	.0947
	21	2	.0059	.0059
		3	.0484	.0742
	22	2	.0026	.0026
		3	.0272	.0272
	23	2	.0009	.0009
		3	.0135	.0135
		4	.0921	.1542
9	9	0	.0247	.0315
	10	0	.0146	.0297
	11	0	.0083	.0116
		1	.0771	.1065
	12	0	.0046	.0118
		1	.0496	.1012
	13	0	.0024	.0036
		1	.0306	.0448
	14	0	.0012	.0013
		1	.0181	.0207
	15	0	.0006	.0008
		1	.0101	.0155
		2	.0697	.1134
	16	1	.0054	.0059
		2	.0437	.0538
	17	1	.0026	.0038
		2	.0258	.0439
	18	1	.0012	.0012
		2	.0142	.0166
		3	.0837	.1143
	19	2	.0072	.0072
		3	.0516	.0562
	20	2	.0033	.0033
		3	.0294	.0377
	21	2	.0013	.0013
		3	.0152	.0152
		4	.0900	.1045
	22	3	.0068	.0068
		4	.0526	.0772
10	10	0	.0080	.0118

APENDICE III

Rutinas para realizar el test incondicionado de Mc Donald

```
      DIMENSION ALFA(3)/0.05,0.025,0.01/,MM(101,2),FSUM(201),
     * FSUM2(201),LIN(60)
      OPEN 1,"BETCMDT01.PRU",ATT="A"
      OPEN 2,"BETCMDT02",ATT="A"
      OPEN 3,"BETCMDT03",ATT="A"
      DO 100 JA=1,1
      ALFA(JA)=ALFA(JA)/2.0
      DO 95 I=1,1
      N1=30
      DO 90 J=1,1
      N2=31
      CALL FGTIME(IH,IM,IS)
      TYPE N1,N2," INICIO ",IH,IM,IS
      CALL TABLE(N1,N2,ALFA(JA),MM,A1,PMAX,FSUM,A2,PMAX2,FSUM2,IER)
      ALFA(JA)=ALFA(JA)*2.0
      WRITE(JA,5000) N1,N2,ALFA(JA),A2,PMAX2
5000  FORMAT("<014>",3X,I3,1X,I3,2X,F5.3,5X,"A2=",F8.6,1X,"PMAX=",F8.6,/)
      PASO=1./FLOAT(N1+N2)
      P=0.0
      DO 50 IX=1,N1+N2+1
      IP=IX+IX-1
      IF(IP.GT.(N1+N2+1)) GO TO 50
      P=P+PASO
      IF(IX.EQ.1) P=0.0
      P2=P+PASO
      IX1=IX-1
      IF(IX.LE.(N1+1)) GO TO 40
      IF(IP.EQ.(N1+N2+1)) WRITE(JA,5005)
     * P,FSUM2(IP)
      IF(IP.NE.(N1+N2+1)) WRITE(JA,5005)
     * P,FSUM2(IP),P2,FSUM2(IP+1)
5005  FORMAT(22X,                 2(F8.6,1X,F8.6))
      GO TO 45
40                        WRITE(JA,5010) IX1,MM(IX,1),MM(IX,2),
     * P,FSUM2(IP),P2,FSUM2(IP+1)
5010  FORMAT(3X,I3,2X,I4,1X,I4,5X,2(F8.6,1X,F8.6))
45    P=P2
50    CONTINUE
      WRITE(JA,5000) N1,N2,ALFA(JA),A2,PMAX2
      ALFA(JA)=ALFA(JA)*0.5
      DO 80 II=1,N1+1
      DO 55 IL=1,120
      BYTE(LIN,IL)=46
      IF(IL.GT.N2+1) BYTE(LIN,IL)=0
55    CONTINUE
      IF(MM(II,1).EQ.999) GO TO 65
      DO 60 IL=MM(II,1)+1,N2+1
60    BYTE(LIN,N2+1-IL+1)=42
65    IF(MM(II,2).EQ.-999) GO TO 75
      DO 70 IL=1,MM(II,2)+1
70    BYTE(LIN,N2+1-IL+1)=42
75    WRITE(JA,5015) LIN(1)
5015  FORMAT(S120)
80    CONTINUE
90    CONTINUE
95    CONTINUE
100   CONTINUE
      CALL RESET
      STOP
      END
```

```
      SUBROUTINE TABLE(N1,N2,ALPHA,MM,A1,PMAX,FSUM,A2,PMAX2,FSUM2,IFAULT)
C
C     ALGORITMO AS 161 APPL. STATIS.(1981) VOL. 30 NO.2
C
      DIMENSION MM(101,2),FSUM(201),X(101,101),COM(101),
     * FSUM2(201),COMB(101),COMBO(201)
      IFAULT=0
      IF(N2.GT.100) IFAULT=IFAULT+1
      IF(N1.GT.N2) IFAULT=IFAULT+2
      IF(ALPHA.LE.0.0.OR.ALPHA.GE.0.3) IFAULT=IFAULT+4
      IF(IFAULT.NE.0) RETURN
      N11=N1+1
      N21=N2+1
      CALL COMBIN(N11,COM)
      CALL COMBIN(N21,COMB)
      CALL COMBIN(N1+N2+1,COMBO)
      DO 10 I=1,N11
      MM(I,1)=999
      MM(I,2)=-999
10    CONTINUE
      DO 15 I=1,N11
      DO 15 J=1,N21
      X(I,J)=-1.0
15    CONTINUE
C
C     SE HACE CRECER CHECK HASTA SIG, LA MENOR COTA SUPERIOR SOBRE LA
C     REGION CRITICA, ESTO ES .GE.ALPHA+0.0005
C
      DEL=0.1
      CHECK=3.0*ALPHA
20    IF(CHECK.GT.1.0) GO TO 30
      CALL UMPUAL(N1,N2,CHECK,MM,X,SIG,FSUM,PMAX,COM,COMB,COMBO)
CTYPE N1,N2,"  SIG=  ",SIG,"  ALFA= ",ALPHA
      IF(SIG.GE.ALPHA) GO TO 40
      CHECK=CHECK+DEL
      GO TO 20
30    CHECK=1.0
      CALL UMPUAL(N1,N2,CHECK,MM,X,SIG,FSUM,PMAX,COM,COMB,COMBO)
CTYPE N1,N2,"  SIG=  ",SIG,"  ALFA= ",ALPHA
40    IF(SIG.LE.ALPHA+0.0005) GO TO 50
CTYPE N1,N2,"  SIG=  ",SIG,"  ALFA= ",ALPHA
C
C     SE MODIFICA Y SE GUARDA SIG HASTA QUE SEA MENOR QUE ALPHA
C     MEDIANTE LA ELIMINACION REPETIDA DE LA REGION CRITICA
C     DEL PUNTO QUE TENGA MAYOR PROBABILIDAD
C
      J=1
      IF(MM(J,1).EQ.999) GO TO 50
      JOLD=MM(J,1)
      DO 45 I=1,N11
      IF(MM(I,1).EQ.999) GO TO 47
      INEW=MM(I,1)
      IF(X(I,INEW).LE.X(J,JOLD)) GO TO 45
      J=I
      JOLD=MM(J,1)
45    CONTINUE
C
C     QUITA TODOS LOS VALORES X QUE ESTAN CERCANOS AL MAXIMO VALOR
C
47    XVAL=X(J,JOLD)
      DO 49 I=1,N11
```

!UDD:US.ESTADISTICA:BETCP11

```
      IF(MM(I,1).EQ.999) GO TO 49
      INEW=MM(I,1)
      IF(ABS(X(I,INEW)-XVAL).GT.0.00005) GO TO 49
      MM(I,1)=MM(I,1)+1
      IF(MM(I,1).GT.N21) MM(I,1)=999
49    CONTINUE
      CALL TMAX(MM,N1,N2,SIG,PMAX,FSUM,COM,COMB)
CTYPE N1,N2,"  SIG= ",SIG,PMAX
      GO TO 40
50    A1=SIG
      DO 60 I=1,N11
      IF(MM(I,1).EQ.999) GO TO 60
      IREV=N11-I+1
      MM(IREV,2)=N21+1-MM(I,1)
60    CONTINUE
      CALL TMAX(MM,N1,N2,A2,PMAX2,FSUM2,COM,COMB)
CTYPE N1,N2,"  SIG= ",SIG,PMAX
      DO 70 J=1,N11
      IF(MM(J,1).NE.999) MM(J,1)=MM(J,1)-1
      IF(MM(J,2).NE.-999) MM(J,2)=MM(J,2)-1
70    CONTINUE
      RETURN
      END



      SUBROUTINE COMBIN(NP1,C)
C
C     ALGORITMO AS 161.1 APPL. STATIST. (1981) VOL. 30 NO. 2
C
C     CALCULA LAS COMBINACIONES DE N=NP1-1 ELEMENTOS
C     TOMADOS DE J EN J Y LOS GUARDA EN UNA MATRIZ C
C
C     C(J+1),J=0,1,2,...,N
C
      DIMENSION C(NP1)
      N=NP1-1
      C(1)=1.0
      C(2)=FLOAT(N)
      C(N)=FLOAT(N)
      C(N+1)=1.0
      IF(N.LE.3) RETURN
      M=N/2+1
      DO 10 J=3,M
      NMJP2=N-J+2
      C(J)=C(J-1)*(FLOAT(NMJP2)/FLOAT(J-1))
      C(NMJP2)=C(J)
10    CONTINUE
      RETURN
      END
```

```
      SUBROUTINE UMPUAL(N1,N2,CHECK,MM,X,SIG,FSUM,PMAX,COM,COMB,COMBO)
C
C     ALGORITMO AS 161.2 APPL. STATIST. (1981) VOL.30 NO.2.
C
      DIMENSION MM(101,2),COMBO(201),COMB(101),COM(101),FSUM(201)
            DIMENSION X(101,101)
      N11=N1+1
      N21=N2+1
      BETA0=CHECK
C
C     AÑADE PUNTOS A LA REGION CRITICA CON X1+X2 CONSTANTE
C     HASTA QUE LAS SUMAS DE LAS PROBABILIDADES (HIPERGEOMETRICA)
C     SEA TAN GRANDE COMO SEA POSIBLE SIN SOBREPASAR A CHECK
C
C     LA REGION CRITICA SE REPRESENTA POR MM Y HA SIDO INICIALIZADA
C     EN LA RUTINA TABLE.
C
C     AJUSTA LOS VALORES DE MM FILA A FILA (J1=1,N11).
      DO 100 J1=1,N11
C
C     SI NINGUNA ENTRADA EN LA FILA ESTA EN LA REGION CRITICA,
C     SUPONE QUE LA ES LA ULTIMA ENTRADA(N21).
C
      IF(MM(J1,1).EQ.999) MM(J1,1)=N21
C
C     INICIALIZA BETAU Y PONE J2 INCREMENTADO EN 1 PARA EL BUCLE
C
      J2=MM(J1,1)+1
      BETAU=-1.0
C
C     MIENTRAS ESTEMOS TODAVIA EN LA FILA Y LA POSICION ESTE
C     EN LA REGION CRITICA, CALCULA BETAU MEDIANTE LA HIPER-
C     GEOMETRICA SOBRE LA DIAGONAL. LA SUMA DE LAS PROBABILIDADES
C     SE ALMACENA EN X, QUE ES UNA MATRIZ DE DIMENSION N11*N21.
C     EL CALCULO SE SALTA SI YA SE HIZO ANTERIORMENTE.
C
10    IF(J2.LE.1.OR.BETAU.GT.BETA0) GO TO 40
      J2=J2-1
      IF(X(J1,J2).NE.-1.0) GO TO 30
      J1J2M1=J1+J2-1
      BETAU=COM(J1)*COMB(J2)/COMBO(J1J2M1)
      IF(J1.NE.1.AND.J2.NE.N21) BETAU=BETAU+X(J1-1,J2+1)
      X(J1,J2)=BETAU
      GO TO 10
C
C     TOMA BETAU DE LA MATRIZ X
C
30    BETAU=X(J1,J2)
      GO TO 10
C
C     SI SE PARA ES PORQUE J2=1, SI NO CONTINUA CON EL BUCLE
C     DE J1, PERO SALVANDO ANTES J2.
C
C
40    IF(BETAU.GT.BETA0        ) J2=J2+1

C
C     CHEQUEA SI HAY ALGUN PUNTO DE LA FILA EN LA REGION CRITICA
C
      IF(J2.GT.N21) J2=999
      MM(J1,1)=J2
```

```
100     CONTINUE
C
C       PARA LA REGION CRITICA DE TAMANO MENOR O IGUAL QUE CHECK
C       BAJO EL TEST NO-ALEATORIZADO UMPU, ENCUENTRA SIG TAL QUE
C       SEA EL MENOR DE LAS COTAS SUPERIORES SOBRE SU TAMANO BAJO
C       EL TEST INCONDICIONAL NO-ALEATORIZADO.
C
        CALL TMAX(MM,N1,N2,FMAX,PMAX,FSUM,COM,COMB)
        SIG=FMAX
        RETURN
        END
```

```
      SUBROUTINE TMAX(MM,N1,N2,FMAX,PMAX,FSUM,COM,COMB)
C
C     ALGORITMO 161.3 APPL. STATIST. (1981) VOL. 30.NO.2
C
      DIMENSION FSUM(201),MM(101,2),P(201),DFSUM(201),COM(101),
     * COMB(101)
      DATA EPS/0.00001/
C
C     EVALUA LA PROBABILIDAD,FSUM(P),DE LA REGION CRITICA BAJO LA
C     HIPOTESIS DE QUE P1=P2=P PARA P=0.0,...,1.0
C
C     PAUSE "P14"

      F=0.0
      KSTEP=N1+N2+1
      DEL =1.0/FLOAT(N1+N2)
      DO 10 K=1,KSTEP
      P(K)=F
      CALL SRCH(MM,F,N1,N2,FSUM(K),DFSUM(K),COM,COMB)
      TYPE "TMAX",N1,N2,K,FSUM(K),DFSUM(K)
      F=F+DEL
10    CONTINUE
      FMAX=0.0
      PMAX=0.0
      IF(MM(1,1).EQ.999.AND.MM(N1+1,2).EQ.-999) RETURN
C
C     APROXIMA LA MENOR COTA SUPERIOR,FMAX, POR LA PROBABILIDAD
C     DE LA REGION CRITICAEN EL INTERVALO DE 0.0 A 1.0
C
      J=1
      DO 20 I=2,KSTEP
      IF(FSUM(I).GT.FSUM(J)) J=I
20    CONTINUE
      FMAX=FSUM(J)
      PMAX=P(J)
C
C     SI FMAX ES UNO DE LOS EXTREMOS DE (0,1) VETE
C
C      TYPE "MAXIMO EN ",J,P(J)
      PJMID=P(J)
      SUMID=FSUM(J)
      DSUMID=DFSUM(J)
      IF(J.EQ.1) GO TO 60
      IF(J.EQ.KSTEP) GO TO 70
C
C     VETE SI LA FUNCION FSUM(P) ES LO SUFICIENTEMENTE UNIFORME
C     EN LAS CERCANIAS DE FMAX.
C
25    IF(DFSUM(J-1)*DFSUM(J+1).GT.0.0) GO TO 40
      SUMLO=FSUM(J-1)
      PJLO=P(J-1)
      SUMHI=FSUM(J+1)
      PJHI=P(J+1)
30    IF(ABS(SUMID-SUMLO).LE.EPS.AND.ABS(SUMID-SUMHI).LE.EPS)
     * GO TO 40
      IF(DSUMID) 34,40,32
32    PJLO=PJMID
      SUMLO=SUMID
      GO TO 38
34    PJHI=PJMID
      SUMHI=SUMID
```

```
38      PJMID=(PJLO+PJHI)*0.5
        CALL SRCH(MM,PJMID,N1,N2,SUMID,DSUMID,COM,COMB)
         TYPE PJMID,SUMID,DSUMID
        GO TO 30
40      FMAX=SUMID
        PMAX=PJMID
        RETURN
60      IF(FMAX.GE.0.9999) RETURN
        G=0.0001
        CALL SRCH(MM,G,N1,N2,RUM,DRUM,COM,COMB)
C
C       SI FMAX OCURRE EN EL 0, VETE.
C
        IF(RUM.LT.FMAX) RETURN
        J=2
        GO TO 25
70      IF(FMAX.GE.0.9999) RETURN
        G=0.9999
        CALL SRCH(MM,G,N1,N2,RUM,DRUM,COM,COMB)
C
C       SI FMAX OCURRE EN 1 VETE.
C
        IF(RUM.LT.FMAX) RETURN
        J=KSTEP-1
        GO TO 25
        END
```

```
      SUBROUTINE SRCH(MM,PP,N1,N2,SUMS,DSUMS,COM,COMB)
C
C     ALGORITMO AS 161.4 APPL. STATIST. (1981). VOL. 30 NO.2
C
C     ENCUENTRA EL TAMAÑO DE LA REGION CRITICA BAJO EL TEST INCON
C     DICIONAL NO ALEATORIZADO CON PROBABILIDAD P1=P2=PP
C
CDOUBLE PRECISION SUM,SUMM,DSUM,DSUMM,PPEXP,DPPEXP,DP1EXP,PP1EXP
      DIMENSION MM(101,2),COM(101),COMB(101)
      N11=N1+1
      N21=N2+1
      SUM=0.0
      SUMM=0.0
      DSUM=0.0
      DSUMM=0.0
      IF(PP.GE.0.99999) GO TO 30
      IF(PP.LE.0.00001) GO TO 40
C
C     BUSCA LA REGION CRITICA DE UNA COLA.
C
C     TYPE "SRCH",N1,N2,PP,SUMS,DSUMS
      PP1=1.0-PP
      N1N2M=N1+N2
      DO 20 I=1,N11
      IF(MM(I,1).EQ.999) GO TO 50
      NE=N21
      K11=I-1
      PR1=PP
      IF(I.NE.1) PR1=(-1)*PR1
      CALL MDBINT(PBIN,K11,N1,PR1,PA11,ICNT11)
      P11=PBIN
      PR1=1.0-PP
      IF(I.NE.1) PR1=(-1)*PR1
      CALL MDBINT(PBIN,K11,N1,PR1,PA12,ICNT12)
      P12=PBIN
      K12=N11-I
      IF(P11.EQ.0.0.AND.P12.EQ.0.0) GO TO 20
C
C
C     CALCULA LA PROBABILIDAD DE LOS PUNTOS DE LA REGION CRITICA
C     POR EL PRODUCTO DE DOS BINOMIALES AMBAS CON P1=P2=PP.SE
C     EVITA LA POSIBILIDAD DE UNDERFLOW CON LA RUTINA MDBINT QUE
C     ADEMAS NOS PERMITE USAR LOS LA PROBABILIDAD DE UN PUNTO PARA
C     CALCULAR LA PROBABILIDAD DEL SIGUIENTE. TAMBINE SE EVALUA LA
C     DERIVADA EN P1=P2=PP Y SU RESULTADO SE ACUMULA EN DSUM.
C
C
      DO 10 J=1,N21
      IF(P11.EQ.0.0) GO TO 4
      PR2=1.0-PP
      IF(J.NE.1) PR2=(-1.0)*PR2
      IF(MM(I,1).GT.NE) GO TO 20
      IF(MM(1,1).EQ.999) GO TO 4
      K2=J-1
      CALL MDBINT(PBIN,K2,N2,PR2,PA21,ICNT21)
      K2=N21-J
      N1N2M=N1+N2-K11-K2
      P2=PBIN
      IF(P2.EQ.0.0) GO TO 5
      POV=P11*P2
      CALL OVERFL(IOOD)
```

```
      IF(ICOD.NE.2) POV=0.0
      SUM=SUM+POV
      DSUMA=DSUM
      DSUM=DSUM+POV   *(((1.0-PP)*(K11+K2)-PP*N1N2M)/(PP*(1.0-PP)))
      CALL OVERFL(ICOD)
      IF(ICOD.NE.2) DSUM=DSUMA
4     IF(MM(N11,2).EQ.-999) GO TO 5
      IF(P12.EQ.0.0) GO TO 5
      K2=J-1
      PR2=PP
      IF(J.NE.1) PR2=(-1.0)*PR2
      CALL MDBINT(PBIN,K2,N2,PR2,PA22,ICNT22)
CTYPE N1,N2,PBIN,K2,'   DOS COLAS   '
      P2=PBIN
      IF(P2.EQ.0.0) GO TO 5
      N1N2M=N1+N2-K12-K2
      POV=P12*P2
      CALL OVERFL(ICOD)
      IF(ICOD.NE.2)POV=0.0
      SUMM=SUMM+POV
      DSUMMA=DSUMM
      DSUMM=DSUMM+POV   *(((1.0-PP)*(K12+K2)-PP*N1N2M)/(PP*(1.0-PP)))
      CALL OVERFL(ICOD)
      IF(ICOD.NE.2)DSUMM=DSUMMA
5     IF(J.EQ.1) PR2=(-1.0)*PP
      NE=NE-1
10    CONTINUE
20    CONTINUE
      GO TO 50
30    IF(MM(N11,1).NE.999) SUM=1.0
      SUMS=SUM+SUMM
      RETURN
40    IF(MM(1,1).EQ.1) SUM=1.0
      SUMS=SUM
      RETURN
50    SUMS=SUM+SUMM
      DSUMS=DSUM+DSUMM
      RETURN
      END
*************************************************************************
*************************************************************************
*     ESTAS ERAN LAS INSTRUCCIONES DE LA RUTINA ORIGINAL Y QUE SON
*     LAS QUE DEBEN IR ENTRE 23 Y 64. CON ESTAS INSTRUCCIONES HABIA
*     LA POSIBILIDAD DEL UNDERFLOW QUE HEMOS SOLVENTADO CON LAS
*     INSTRUCCIONES ALLI PUESTAS.
*************************************************************************
      DO 20 I=1,N11
      IF(MM(I,1).EQ.999) GO TO 50
      NE=N21
      DP1EXP=PP1**DFLOAT(I+N2)
      PPEXP=PP**DFLOAT(I+N2)
      DPPEXP=PP**DFLOAT(N1-I)
      PP1EXP=PP1**DFLOAT(N1-I)
C     TYPE DP1EXP,PPEXP,DPPEXP,PP1EXP
      DO 10 J=1,N21
      IF(MM(I,1).GT.NE) GO TO 20
      PPEXP=PPEXP/PP
      PP1EXP=PP1EXP*PP1
C
C     CALCULA LA PROBABILIDAD DE UN PUNTO EN LA REGION CRITICA
C     POR EL PRODUCTO DE DOS BINOMIALES INDEPENDIENTES Y LO SUMA
```

```
C       EN SUM. TAMBIEN EVALUA LA DEIVADA DE SUM EN P1=P2=PP.
C
        IF(MM(1,1).EQ.999) GO TO 4
        SUM=SUM+COM(I)*COMB(NE)*PPEXP*PP1EXP
        DSUM=DSUM+COM(I)*COMB(NE)*(FLOAT(I+N2-J)*PPEXP/PP*
     *  PP1EXP-FLOAT(N1+J-I)*PPEXP*PP1EXP/PP1)
C       TYPE "SUM/DSUM",SUM,DSUM
4       IF(MM(N11,2).EQ.-999) GO TO 5
        DPPEXP=DPPEXP*PP
        DP1EXP=DP1EXP/PP1
        SUMM=SUMM+COM(I)*COMB(NE)*DPPEXP*DP1EXP
        DSUMM=DSUMM+COM(I)*COMB(NE)*(FLOAT(N1+J-I)*DPPEXP/PP*
     *  DP1EXP-FLOAT(N2+I-J)*DPPEXP*DP1EXP/PP1)
C       TYPE "SUMM,DSUMM",SUMM,DSUMM
5       NE=NE-1
10      CONTINUE
20      CONTINUE
************************************************************************
************************************************************************
```

```
      SUBROUTINE MOBINT(PBIN,K,N,P,PA,ICNT)
      DATA EPS/.5398E-78/
      DATA ALEPS/-180.2182/
      PBIN=-99.99
      IF(K.GT.N.OR.K.LT.0) RETURN
      IF((ABS(P).GT.1.0).OR.(ABS(P).LT.0.0)) RETURN
      IF(ABS(P).GT.0.0) GO TO 10
      PBIN=0.0
      IF(K.EQ.0) PBIN=1.0
      RETURN
10    IF(ABS(P).LT.1.0) GO TO 20
      PBIN=0.0
      IF(K.EQ.N) PBIN=1.0
      RETURN
20    P1=ABS(P)
      Q1=1.0-P1
      K1=K
      XN=N
      XX=XN*P1
      IF(FLOAT(K).LE.XX) GO TO 25
      P1=Q1
      Q1=ABS(P)
      K1=N-K
      IF(P.LT.0.0) K1=K
25    ALQN=XN*ALOG(Q1)
      IF(P.LT.0.0) GO TO 27
      ICNT=ALQN/ALEPS
      ALQN=ALQN-ICNT*ALEPS
      PK=EXP(ALQN)
      GO TO 28
27    PK=PA
28    IF(K1.EQ.0) GO TO 35
      QP=P1/Q1
      XJ=0.0
      IF(P.LT.0.0) XJ=K1-1
      KI=1
      XN=XN+1.0
      IF(P.LT.0.0) KI=K1
      IF(FLOAT(K).GT.XX.AND.P.LT.0.0) QP=1.0/QP
      DO 30 J=KI,K1
      XJ=XJ+1.0
      PK=PK*(QP*(XN-XJ))
CTYPE PK,XJ,K1,QP,XN
      IF(PK.LT.XJ) GO TO 30
      PK=PK*EPS
      ICNT=ICNT-1
30    PK=PK/XJ
      IF(PK.LT.XJ) GO TO 35
      PK=PK*XJ
      PK=PK*EPS
      ICNT=ICNT-1
35    PA=PK
      PBIN=PK
      IF(ICNT.NE.0) PBIN=0.0
      RETURN
      END
```

REFERENCIAS

Anderson, S. and Hauck, W.W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. Communications in statistics, Theory and Methods 12(23), 2663-2692.

Armitage, P. (1971). Statistical Methods in Medial Research. Oxford, Blackwell Scientific Publications.

Armsen, P. (1955). Tables for significance tests of 2x2 contingency tables. Biometrika 42, 494-511.

Ballatori, E. (1982). Sui test statistici per il confronto tra due frequenze in tabelle 2x2. Metron XL, 3-4, 157-171.

Ballatori, E.; Cecconi, G. (1983). Calcolo di test statistici per i confronto tra due frequenze in tabelle 2x2. Rivista di Statistica Applicata, 16(2), 171-176.

Barnard, G.A. (1945). A new test for 2x2 tables. Nature 156, 177, 783--784.

Barnard, G.A. (1947). Significance tests for 2x2 tables. Biometrika 34, 123-138.

Barnard, G.A. (1949). Statistical Inference. J.R. Stat. Soc. B, 11, 115-139.

Barnard, G.A. (1979). In contradiction to J. Berkson's dispraisse:conditional tests can be more efficient. J. Stat. Planning Inf. 3, 181-187.

Barnard, G.A. (1982). Condicionality versus similarity in the analysis of 2x2 tables. Statistics and probability: Essays in Honor of C.R. Rao. North Holland Publishing Company, 59-65.

Basu, D. (1977). On the elimination of nuisance parameters. Journal of the American Statistical Association 75, 355

Basu, D. (1979). Discussion of Joseph Berkson's paper 'In dispraisse of the exact test' J. Stat. Planning Inf. 3, 189-192.

Berkson, J. (1978). In dispraisse of the exact test. J. Stat. Planning Inf. 2, 27-42.

Boschloo, R.D. (1970). Rassed conditional level of significance for the 2x2 Table when testing the equality of two probabilities. Statistica Neerlandica 24, 1, 1-35.

Bradley, J.V. (1968). Distribution-free statistical tests. Ed. Prentice Hall.

Brownlee K.A. (1967). Statistical theory and methodology in science and engineering. Wiley: New York

Burstein, H. (1981). Binomial 2x2 test for independent samples with independent proportions. Commun. Statist-Theoryand Methods. A10(1) 11-29.

Cohran, W.G. (1952). The χ^2 test of goodnes of fit. Ann. Math. Statist. 23, 315-345.

Conover, W.J. (1974) Some Reasons for not using the Yates continuity corrections on 2x2 contingency tables. Journal of the American Stat. Assoc., 69, 374-376.

Cook, I.T. (1981) On the continuity correction for bivariate discrete distributions. Private communication.

Cornfield, J. (1956). A statistical problem arising from retrospective studies. Proceeding of the Third Berkeley Symposium on Mathematical Statistics and Probability IV. J. Neyanan (ed), University of California Press, Berkeley, 135-148.

Cornfield, J. (1966). Sequential trials, sequential analysis and the likelihood principe. The American Statiscian 20, 18-23.

Cox, D.R.(1970). The continuity correction. Biometrika 57, 217-219.

Cox, D.R. and Hinkley, D.V. (1974). Theoretical statistics. Chapman and Hall, London.

Dunnett, C.W. and Gent, M. (1977). Significance testing to establish equivalence between treatments, with special reference to data in the form of 2x2 tables. Biometrics 33, 593-602.

Dupont (1986) Sensitivity of Fisher's exact test to minor perturbations in 2x2 contingency tables. Statistics in Medicine 5, 692-635.

Eberhardt, K.R. and Fligner, M.A. (1977). A comparison of two tests for equality of two proportions. The American Statistician, 31, 151-155.

Feller, W. (1973). Introduccion a la Teoria de Probabilidades y sus Aplicaciones Vol. I. Ed. Limusa Wiley.

Fisher, R.A. (1935). The logic of inductive inference. Journal of the Royal Statistical Society 98, 39-54.

Fisher, R.A. (1941). Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

Fisher (1942). The Design of Experiments. 3º Ed. Cap. 2. Oliver and Boyd.

Fisher, R.A. (1945). Nature pp 388.

Fisher, R.A. (1956). Statistical Methods and Scientific Inference. London, Oliver and Boyd.

Finney, D.I.; Latsda, R.; Bennet, B.B.Z.; Hsu, P. (1963). Tables for testing significance in an 2x2 contingency table. Ed. Cambridge University Press.

Fleiss, J.L. (1981). Statistical Methods for Rates and Proportions. (2nd Edn). N.Y., Ed. John Wiley and Sous.

Freeman, M.F. and Tukey, J.W. (1950). Transformations related to the angular and the square root. Ann. Math. Statist. 27, 607-611.

Garside, G.R. and Mack, C.L. (1967). Correct confidence limits for the 2x2 homogeneity contingency table with small frequencies. The New Journal of Statistics and Operational Research 3(2), 1-25.

Garside, G.R. and Mack, C. (1968). Further tables of confidence limits for the 2x2 homogeneity contingency table. New Journal of Statistical and Operational Research 4,3, 9-34.

Garside, G.R. (1971). An accurate correction for the χ^2-test in the homogeneity case of 2x2 contingency tables. New Journal of Statistics and Operational Research 7, 1-26.

Garside, G.R. and Mack, C. (1976). Actual Type 1 Error Probability for various test in the homogeneity case of the 2x2 Contingency Table. The American Statiscian, 30, 1, 18-20.

Gart, J.J. (1971). The comparison of proportions: A review of significance tests confidence intervals and adjustements for stratification. Review of the International Statistical Institute 39, 2, 148-169.

Goodman, L. (1964). Simultaneous Confidence Intervals fo Contrast Among multinomial Populations. Annals of Mathematical Statistics 35, 716-725.

Gridgeman, M.T. (1959a). The lady testing tea, and allied topics. JASA 54, 776-783.

Gridgeman, M.T. (1959b). Sensory item sorting, Biometrics 15, 298-306.

Grizzle, J.E. (1967). Cotinuity correction in the χ^2-test for 2x2 tables. The American Statiscian, 21(4), 28-32.

Haber, M. (1980). A comparison of some continuity corrections for the chi-square test on 2x2 tables. Journal of the American Stat. Assoc. 75, 510-515.

Hutchinson, T.P. (1982). Some theories of perfomance in multiple choice test and theri implications for variants of the task. British Journal of Mathematical and Statistical Psychology 35, 71-89.

Irwing, J.O. (1935). Test of significance for differences between percentages based on small numbers. Metron 12(2), 84-94.

Johnson, M.L. and Kotz, S. (1970) Distributions in statistics (Continuous Univariate Distributions-1) New-York: Wiley.

Kempthorne, O. (1967). The design and analysis of experiments New York: Wiley

Kempthorne, O. (1979). In dispraisse of the exact test: reactions J. Stat. Planning. Inf., 3, 199-213.

Kendall, M.G. and Stuart, A. (1967). The advanced Theory of Statistics. N.Y. Hafner Publishing Company.

Lehman, E.L. (1959). Testing for Statistical Hypotheses. N.Y. Ed. John Wiley and Sous.

Liddell, D. (1978). Practical tests of 2x2 contigency tables. The Statistician, 25(4), 295-304.

Lillestol, J. (1982). Remark on AS152. Cumulative Hypergeometric Probabilities. Applied Statist, 29, 221-223.

Lord, F.M. and Novick, M.R. (1968). Statistical Theories of Mental Test Scores. Menlo Park, C.A.: Addison-Wesley.

Lord, F.M. (1977). Optimal number of choices per item a comparison of four approaches. Journal of Educational Measurement. 14, 1, 33-38.

Lund, R.E. (1980). Cumulative hypergeometric probabilities. Applied Statistics 29, 221-223.

Mantel, M. and Greenhouse S.W. (1968). What is the continuity correction?. The American Statistician, 22(5), 27-30.

Mantel, M. (1974). Comment and a Suggestion. JASA, 69, 378-380.

McDonald, L.L.; Davis, B.M.; Milliken, G.A. (1975). A nonrandomized unconditional test for comparing two proportions. College of Commerce and Industry Research Paper 94, University of Wyoming, Laramie, Wyoming.

McDonald, L.L. et al. (1977). A nonrandomized unconditional test for comparing two proportions in a 2x2 contingency table. Technometrics 19, 145-150.

McDonald, L.L.; Davis, B.M.; Baner,, H.R. (III), Laby, B. (1981). Algorithm AS161: Critical Regions of an unconditional non-randomized test of homogeneity in 2x2 Contingency tables. Applied Statistics, 30, 2, 182-189.

Mitra, S. (1969). An alternative approach to the analysis of a fourfold table. The American Statistician, 23, 5, 22-24.

Neave, H.R. (1982). A new look at an old test. Bulletin of Applied Statistics 9(2) 165-178.

Plackett, R.L. (1964). The continuity correction in 2x2 tables. Biometrika, 51, 3 and 4, 327-337.

Plackett, R.L. (1977). The marginal totals of a 2x2 table. Biometrika 64, 1, 37-42.

Pearson, G.S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2x2 table. Biometrika 34, 139-167.

Pearson, G.S. and Hartley (1966). Biometrika Tables for Statisticians, Vol. 1 (3º Ed). Cambridge University Press.

Pirie, W.R. and Handam, M.A. (1972). Some revised continuity corrections for discrete distributions. Biometrics 28, 693-701.

Polya, G. (1945). Remark on computing the probability integral in one and two dimensions. Proceeding of the 1st Berkeley Symposium on Mathematical Statistics and Probability, 63-78.

Radlow, R. and Alf, E.F. Jr. (1975). An alternate multinomial Assessment of the Accuracy of the χ^2 test of Goodness of fit. Journal of the American Statistical Association, 70, 352, 811-813.

Rao, C.R. (1970). Advanced Statistical Methods in Biometric Research Hafner.

Rao, C.R. (1973). Linear Statistical Inference and Its Applications. 2nd. ed. John Wiley & Sons.

L. Sachs (1986). Alternatives to the chi-square test of homogeneity in 2x2 Tables and to Fisher's Exact Test. Biometrical Journal 28 (8), 975-979.

Sathe, Y.S. (1982). Another test for equality of two proportions. Commun. in Stat.-Simul. and Comp., 11, 3, 373-375.

Schouten, H.J.A. et al (1980). Comparing two independent binomial proportions by a modified chi square test. Biom Journal. 22, 3, 241-248.

Starmer, C.F. and Grizzle, J.E. (1974). Comment on : "Some reasons for not using the Yates continuity correction or 2x2 contingency tables". JASA 69, 376-378.

Suissa, S. and Shuster, S.J. (1984). Are uniformely most powerful unbiased tests really best?. The American Statistician, 38, 204-206.

Suissa, S. and Shuster, J.J. (1985). Exact unconditional samples sizes for the 2x2 binomial trial. Journal of the Royal Statistical Society. A. 148(4), 317-327.

Tocher, K.D. (1950). Extension of the Neyman-Pearson theory of tests of discontinous variables. Biometrika 37, 130-144.

Upton, G.J.G. (1982). A comparison of alternative tests for the 2x2 comparative trial. J.R. Stat. Soc. A, 145, 1, 86-105.

Westlake, W.J. (1974). The use of balanced incomplete block desings in comparative bioavailability trials. Biometrics 30, 319-327.

Wichnann, B.A. and Hill, I.D. (1982). An efficient and portable pseudo-random numbers generator. Applied Statistics 31, 188-190.

Wilcox, R.R. (1980). Determining the length of a criterion-referenced test. Applied Psychological Measurement 4, 4, 425-446.

Williams, D.A. (1976). Improved likelihood ratio tests for complete contingency tables. Biometrika 63, 33-37.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. J.R. Stat. Soc., Suppl. 1, 217-235.

Yates, F. (1984). Test of significance for 2x2 contingency tables. J.R. Statistic. Soc A, 147, 3, 426-463.