

03043

2ej

1

RECEIVED
LIBRARY
NOV 15 1987

**ANALISIS DE
TABLAS DE CONTINGENCIA 2X2**

APLICACIONES EN MEDICINA VETERINARIA Y ZOOTECNIA

MONOGRAFIA QUE PRESENTA EL

M.V.Z. RICARDO R. NAVARRO FIERRO

**PARA OBTENER EL DIPLOMA DE LA
ESPECIALIZACION EN ESTADISTICA APLICADA
QUE OTORGA LA UNIVERSIDAD NACIONAL AUTONOMA DE
MEXICO, A TRAVES DE LA UNIDAD ACADEMICA DE LOS CICLOS
PROFESIONAL Y DE POSGRADO DEL COLEGIO DE CIENCIAS Y
HUMANIDADES Y DEL INSTITUTO DE INVESTIGACIONES EN
MATEMATICAS APLICADAS Y EN SISTEMAS (IIMAS)**

**VEJIS CON
FALLA DE ORIGEN**

**CIUDAD UNIVERSITARIA, D. F.
NOVIEMBRE DE 1987.**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

TESIS CON FALLA DE ORIGEN

INDICE DE CONTENIDO

CONTENIDO

PAG.

I . LA TABLA DE CONTINGENCIA 2x2

1. INTRODUCCION	1
2. APLICACION DE LA TABLA 2x2	1
3. LAS VARIABLES QUE FORMAN LA TABLA 2x2	2
a) Dos valores exhaustivos	2
b) Valores mutuamente excluyentes	2
4. CUALIDADES DE LAS VARIABLES BINARIAS	3
a) La escala de medición más simple	3
b) La posesión de un orden intrínseco	5
5. NOMENCLATURA UTILIZADA EN LA TABLA 2x2	6
6. CALCULOS DE PROBABILIDAD EN UNA TABLA 2x2	7
a) Probabilidad de un evento	7
b) La probabilidad conjunta de dos eventos	8
c) Probabilidades condicionales	8
d) Independencia entre dos variables binarias	9

II . ESTUDIOS QUE ORIGINAN UNA TABLA 2x2

1. INTRODUCCION	11
2. DISTINCION ENTRE EXPERIMENTOS Y ESTUDIOS OBSERVACIONALES	11
3. TIPOS DE MUESTREO APLICABLES A LA TABLA 2x2	14
a) Muestreo multinomial	16
b) Muestreo comparativo	16
c) Muestreo experimental	17
4. MUESTRAS INDEPENDIENTES CONTRA DATOS APAREADOS	19
5. RELACION DE LA POBLACION CON LA MUESTRA	20
6. EJEMPLOS INTEGRADOS	21

III . MEDIDAS DESCRIPTIVAS PARA UNA TABLA 2x2

1. INTRODUCCION	24
2. ESTIMACION DE LAS MEDIDAS DESCRIPTIVAS	25
a) Estimación de punto o puntual	25
b) Estimación por intervalo	26
3. DESARROLLO DE UNA MEDIDA DESCRIPTIVA: LA PROPORCION	27
4. CLASIFICACION DE LAS MEDIDAS DESCRIPTIVAS	30
5. MEDIDAS DE ASOCIACION	31
a) Phi cuadrada (ϕ^2)	34
b) χ de Cramer	35
c) Proporción explicada de la varianza (P_{ex})	35
d) Phi (ϕ)	36
e) La tau b de Kendall (τ_b)	37
f) Coeficiente medio de contingencia (C_m)	38
6. MEDIDAS DE PREDICCIÓN	42
a) La delta de Somers (δ)	42
b) La gamma de Goodman y Kruskal (Γ)	46
c) La lambda de Goodman y Kruskal (λ)	51
d) Riesgo atribuible (R_a)	55
e) Potencia atribuible (P_a)	57
f) Proporción atribuible ($Prop_a$)	58

INDICE DE CONTENIDO

7.	MEDIDAS DE COMPARACION	59
	a) La razón de momios	59
	b) Riesgo relativo (Rr)	65
	c) Potencia relativa (Pr)	66
8.	MEDIDAS DE CONCORDANCIA	68
	a) Kappa de Cohen (K)	68
9.	VALORACION DE UNA PRUEBA DIAGNOSTICA	74
	a) Calidad general de una prueba	75
	b) Estimación de la sensibilidad y la especificidad	75
	c) Calidad de la predicción	80
IV .	PRUEBAS ESTADISTICAS PARA UNA TABLA 2x2	
1.	INTRODUCCION	85
2.	LOGICA Y FUNCIONAMIENTO DE LAS PRUEBAS ESTADISTICAS	85
	a) Tipos de fallas en las pruebas de hipótesis	87
	b) Indicadores de la calidad de una prueba	87
3.	DESARROLLO DE UNA PRUEBA ESTADISTICA	88
4.	APLICACION DE UNA PRUEBA ESTADISTICA	89
5.	PRUEBA DE JI CUADRADA (χ^2)	91
	a) Corrección de Yates	93
6.	PRUEBA EXACTA DE FISHER	96
7.	PRUEBA DE McNEMAR	101
8.	PRUEBA DE GART	104
	a) La prueba de Gart con muestras pequeñas	106
9.	PRUEBA PARA GRADO DE CONCORDANCIA	112
V .	EXTENSIONES DE LA TABLA 2x2	
1.	INTRODUCCION	115
2.	COMBINACION DE VARIAS TABLAS 2x2	115
	a) Método de Cochran	116
	b) Método de Mantel-Haenzel	121
3.	MAS DE DOS PORCENTAJES A COMPARAR: TABLA Cx2	125
	a) Variable nominal de C niveles	130
	comparar todos los pares de proporciones	130
	comparación con un testigo	134
	b) C grupos con un orden propio	136
	c) Variable cuantitativa en C puntos	140
VI .	DETERMINACION DEL TAMAÑO DE MUESTRA	
1.	INTRODUCCION	149
2.	RELACION DE LA POTENCIA CON EL TAMAÑO DE MUESTRA	149
3.	DIFERENCIA MINIMA POR DETECTAR	150
	a) Ventaja mínima sobre el estándar	151
	b) Razón de momios entre los grupos	151
4.	TAMAÑO DE LA MUESTRA PARA UN MUESTREO MULTINOMIAL	152
	a) Porcentajes medios	153
	b) Porcentajes pequeños	154
	c) Porcentajes grandes	155
5.	TAMAÑO DE LA MUESTRA PARA ESTUDIOS COMPARATIVOS	156
VII .	LITERATURA CITADA	161

CAPITULO I

LA TABLA DE CONTINGENCIA 2x2

1. INTRODUCCION

La tabla 2x2 es un medio para resumir y presentar los resultados de cierta clase de estudios, su forma es familiar a la mayoría de los profesionales, por lo que es ampliamente aceptada y cabalmente entendida.

El cuadro I.1 muestra el esquema general de la tabla 2x2, en el se puede constatar la sencillez de esta herramienta estadística: su lectura es simple y, lo más importante, es fácil de analizar e interpretar.

Cuadro I.1

Esquema general de la tabla de contingencia 2x2

		Variable X		
		0	1	Total
Variable Y	0	a	b	$r_{0.}$
	1	c	d	$r_{1.}$
Total		$c_{0.}$	$c_{1.}$	N

2. APLICACION DE LA TABLA 2x2

La tabla 2x2 es útil cuando se consideran dos variables en cada sujeto examinado, y cada una de ellas sólo tiene dos resultados posibles; más adelante se verá cómo muchas variables pueden expresarse en esta forma, lo que proporciona una amplia aplicación a la tabla 2x2.

Estas dicotomías son muy frecuentes en la vida real; "éxito o fracaso" y "día y noche" son ejemplos en los que sólo se consideran dos resultados, sin valores intermedios entre ellos. Estos casos son aún más frecuentes en problemas biológicos, donde clasificaciones como "vivo o muerto", y "hembra o macho" no dejan lugar a puntos medios; en general, abundan las situaciones en que se examina la "presencia o ausencia" de alguna característica.

Es común asociar un código binario a los resultados de estas variables. Lo más frecuente es utilizar las cifras "cero" y "uno" para designar a cada valor de la variable. El código binario 0 y 1 se aplicará a lo largo del texto.

3. LAS VARIABLES QUE FORMAN LA TABLA 2x2

Para lograr un uso correcto de la tabla 2x2, las variables que la forman deben poseer dos características: que estén constituidas por dos valores exhaustivos y mutuamente excluyentes. A continuación se explican cada una de ellas.

a) DOS VALORES EXHAUSTIVOS

La condición de dos valores exhaustivos indica que los dos únicos niveles de la variable cubren todos los resultados posibles y, por tanto, no existe otro resultado viable; por ejemplo, al clasificar una hembra en gestante o vacía, la hembra está o no en gestación, no hay más.

En algunos casos puede ocurrir que a pesar de que existen otros resultados posibles, éstos salen de la población cubierta por el estudio. Por ejemplo, si se clasifica a un grupo de vacas de las razas Pardo Suiza y Holstein es evidente que no se abarcan todas las razas posibles, pero la población de referencia para la investigación sólo incluye vacas de las dos razas citadas, y esta limitación deberá tenerse en cuenta al concluir sobre los resultados obtenidos.

b) VALORES MUTUAMENTE EXCLUYENTES

La condición de mutuamente excluyentes indica que cuando ocurre alguno de los resultados es imposible que suceda el otro. Así, no puede tenerse una hembra que al mismo tiempo esté y no en gestación.

También para esta segunda característica puede haber excepciones teóricamente viables. Al buscar información sobre alumnos de las facultades de veterinaria y de administración (se supone que se tiene interés en investigar sólo a la población de estas dos áreas), puede existir el caso, aunque es raro, de personas que cursen ambas carreras; por tanto el estudio y las conclusiones deberán limitarse a quienes siguen sólo una de estas dos profesiones.

Por supuesto, puede haber estudios en los que exista alguna sobreposición entre los resultados potenciales para alguna variable, con casos que combinen los dos niveles de la variable, pero quedan fuera de los propósitos de este libro.

4. CUALIDADES DE LAS VARIABLES BINARIAS

Las variables binarias tienen dos cualidades que les confieren ventajas interesantes: 1) cualquier medición puede reducirse a una escala dicotómica, y 2) pueden dar un orden propio a las variables nominales. Para apoyar la exposición, el cuadro I.2 contiene un resumen de las escalas de medición.

a) LA ESCALA DE MEDICION MAS SIMPLE

El hecho de que cualquier medida pueda simplificarse llevándola a una escala binaria es fácil de intuir. Por ejemplo, una variable cuantificada en escala nominal, como la raza de un borrego (Tabasco, Suffolk o Dorset), puede transformarse en dicotómica: "raza de lana o de pelo", o bien, si se atiende al número de crías por camada: "raza multipara o unipara", la clasificación dependerá del motivo de la investigación. Aun las variables numéricas, como el peso de una vaca, pueden reducirse a un juicio dicotómico: "hasta 350 kg o más de 350 kg"; otro ejemplo podría registrarse como "peso suficiente o insuficiente para salir al mercado".

Se debe hacer notar que reducir una variable cualquiera a una expresión dicotómica implica, de manera inevitable, una pérdida de información sobre el evento medido; a cambio, la reducción ofrece mayor simplicidad en el análisis e interpretación del fenómeno que se está estudiando. Es evidente que la conveniencia de reducir alguna variable a cierta expresión dicotómica dependerá de que la pérdida de información sea menos importante que la ventaja derivada de la facilidad de registro, de análisis o de interpretación.

Cuadro 1.2

Características de la escalas de medición*

ESCALA	CARACTERÍSTICAS
NOMINAL = y <>	-Es la escala de medición más simple. Asigna un nombre diferente a cada distinta manifestación de la variable, por lo que sus características son: - <u>IGUALDAD DENTRO DE GRUPOS</u> : todos los que tienen un mismo valor son iguales para esa variable. Los sujetos que reciben un mismo nombre para la característica de interés son iguales en cuanto a esa variable se refiere. - <u>DIFERENCIA ENTRE VALORES</u> : los casos clasificados en distintos grupos son diferentes entre sí. - <u>EJEMPLO</u> : al medir la especie, las vacas son iguales entre sí y diferentes de los cerdos, los que a su vez son todos cerdos y distintos de las vacas y las cabras.
BINARIA	-Variable nominal que comparte algunas cualidades de los ordinales.
ORDINAL =, <>, < y >	-Es el siguiente nivel en cuanto a la fuerza de la medición realizada. También se basa en asignar nombres; pero, además de las características anteriores, aquí se tiene un <u>ORDEN PROPIO</u> entre los valores de la variable, de modo que no sólo son distintos entre grupos, sino que se sabe cuáles son mejores o mayores que cuales otros - <u>EJEMPLO</u> : al medir el estado de carnes en bueno, regular y malo hay una relación ordinal en la que bueno es mejor que los otros y, a su vez, regular aventaja a malo.
INTERVALO =, <>, <, > y distancia	-Es la primera escala propiamente numérica; aunque en las dos primeras se pueden usar números, estos se aplican en un sentido estrictamente nominal. En esta escala, a las anteriores se agrega la noción de <u>DISTANCIA</u> de un punto a otro de la escala; tiene el concepto de magnitud de la diferencia de cada nivel con respecto a cualquier otro. Además de saber cuál es mejor, ahora se puede medir cuanto mejor. - <u>EJEMPLO</u> : al medir la temperatura, 15°C es menor que 22°C, exactamente 7°C menor.
RAZÓN O PROPORCIÓN =, <>, <, >, distancia, cero y razón matemática	-Aquí se añaden dos características importantes, la primera es que el cero ya no es arbitrario, sino que indica un <u>PUNTO DE PARTIDA REAL</u> ; por ejemplo, cero kg indica que no hay peso, y es el punto en que se basa la medición de este, sin importar que se usen kg, lb o ton. -Además, ahora <u>LA RAZÓN TIENE SENTIDO</u> : 6 kg es el doble de 3 kg y aunque se cambie la escala esta razón o proporción se mantiene. -Las características de esta escala permiten efectuar todas las operaciones numéricas con los resultados. Notese que en la escala de intervalo sólo se puede calcular la diferencia, ya que es la única apropiada a las características de esta escala. - <u>EJEMPLOS</u> : los ejemplos de variables medidas en una escala de razón son incontables, ya que incluyen todas las medidas de peso, volumen y distancia.
ABSOLUTA =, <>, <, >, distancia, cero, razón matemática y escala única	-Con frecuencia esta escala se considera dentro de la de razón, ya que la característica que se agrega no es de consecuencias importantes para la fuerza de la medición. La escala absoluta se refiere a los conteos de objetos, por lo que la <u>ESCALA ES ÚNICA</u> e invariante, ya que la medición se refiere al número de individuos, objetos o de sucesos, de manera que la escala de medición depende de lo que se cuenta y no puede cambiarse: si se cuentan algunos peces, la escala es peces y ninguna otra. - <u>EJEMPLOS</u> : Número de aves en una parvada, cantidad de lechones en la camada, conteo de parásitos en una muestra de heces, total de granjas en una región.

* Los símbolos debajo de los nombres de las escalas de medición resumen las características de cada una.

Por otro lado, es claro que la reducción aplicable en cada caso particular depende del interés que haya motivado el estudio y de las preguntas que se pretende responder con la información que se obtenga al terminar el trabajo.

b) LA POSESION DE UN ORDEN INTRINSECO

La segunda característica que favorece a las variables de dos niveles es que pueden dar un orden propio a las nominales, las cuales no poseen un orden intrínseco. Esto se debe a que cuando sólo hay dos valores posibles para la variable, puede juzgarse a cualquiera de los dos como mejor o mayor que el otro o simplemente asignarles los valores 0 y 1, sin importar a cuál se favorezca o se le dé el valor más grande, sin que esto afecte la estructura propia de los datos.

Se comprueba fácilmente que a una variable nominal con tres o más niveles no puede dársele un orden sin poner en juego un criterio subjetivo y poco sólido (o muy particular) para asignarles una posición relativa entre sí, ya que esta posición no forma parte de la organización de los datos, no puede ser única e inequívoca.

Usando de nuevo el ejemplo sobre las razas de borregos, si se insistiera en darles un orden, alguien podría ordenarlas de acuerdo al tamaño medio; otra persona podría usar la capacidad reproductiva como criterio de clasificación; algún investigador emplearía la producción de lana para acomodarlas, ya sea conforme la longitud o de acuerdo con la calidad de la lana producida. Así, habría tantos criterios para ordenar las razas como diferentes intereses pudieran existir sobre estos grupos de borregos, lo cual hace patente que no tienen un orden intrínseco y único entre sí.

Lo mismo sucede con todas las variables nominales, exceptuando las de dos niveles, ya que para éstas sólo hay una sucesión factible: un valor al lado de otro, sin importar cuál se coloca primero. Al registrar el peso de un grupo de lechones, la medición no se afecta si se considera al macho mejor que a la hembra o viceversa. También es indistinto asignar el código "0 = hembra" y "1 = macho" o decidirse por "1 = hembra" y "0 = macho". Es claro que la sucesión de los valores es única y que el orden relativo entre ambos no afecta la medición de la variable.

5. NOMENCLATURA UTILIZADA EN LA TABLA 2x2

Los valores que aparecen en toda tabla 2x2 son el resultado de un conteo, es decir, son el número de casos registrados. Según el código del cuadro I.1, el valor a es el número de casos en que se combina el nivel 0 para ambas variables, por lo que puede representarse como (0,0); b es la cantidad de individuos que tienen el valor 1 para la variable X (clasificación vertical) y 0 para la variable Y (renglones), los casos registrados en b pueden codificarse como (0,1); en forma similar c y d corresponden al conteo de sujetos del tipo (1,0) y (1,1), respectivamente.

Los valores r_0 , r_1 , c_0 y c_1 se conocen como totales marginales de renglón los dos primeros y de columna los dos últimos, en tanto que N es el total de casos contenidos en la tabla.

Cuadro 1.3

Ejemplo del uso de una tabla 2x2

	Estado reproductivo			
	Gestante	Vacía	Total	
Salud de la ubre	Sana	13	17	30
	Mastitis	7	43	50
Total	20	60	80	

Un ejemplo contribuirá a explicar la nomenclatura utilizada en la tabla 2x2. Considerando un total de 80 vacas ($N = 80$) clasificadas en sanas o con mastitis y en gestantes o vacías, como aparece en el cuadro I.3, r_1 indica el total de hembras que padecen mastitis ($r_1 = 50$) y r_0 incluye a las que no la padecen ($r_0 = 30$). En cuanto al estado reproductivo, los totales de columna expresan las frecuencias para gestantes ($c_0 = 20$) y vacías ($c_1 = 60$). El número de vacas sanas gestantes se indican en a ($a = 13$); b es el número de las que aparecieron sanas y vacías ($b = 17$); de la misma manera se representa el estado reproductivo de las hembras con mastitis: c y d reúnen a las gestantes ($c = 7$) y vacías ($d = 43$).

6. CALCULOS DE PROBABILIDAD EN UNA TABLA 2x2

La probabilidad de ocurrencia de un evento se expresa con un valor comprendido entre cero y uno. La probabilidad cero señala que el evento no sucederá, en tanto que el valor de uno se asocia con los resultados que con seguridad han de ocurrir. Los valores próximos a cero se aplican a resultados que difícilmente ocurren; a su vez, las probabilidades cercanas a uno se asignan a los eventos que suceden con mayor frecuencia.

a) PROBABILIDAD DE UN EVENTO

En el caso de la tabla 2x2, los conteos presentados pueden expresarse como fracciones o porcentajes del total de los datos o de un subconjunto de éstos. Cada una de estas fracciones se interpreta como una probabilidad; es decir, la probabilidad de que se encuentre un sujeto con determinada característica en una cierta población se indica con la fracción de los casos, en esa población, en que se presenta la característica.

Si se considera una población de 80 vacas (vease el cuadro I.3), se pueden calcular algunas proporciones; por ejemplo, aparecen $20/80 = 0.25$ vacas gestantes, por lo que la probabilidad, en ese conjunto, de que una hembra cualquiera, esté gestante es de 0.25.

Lo anterior muestra el concepto de probabilidad que se aplicará a lo largo del texto, entendiéndola como la frecuencia relativa con que ocurre un evento. Así, la probabilidad de encontrar una hembra sana es $30/80 = 0.375$, expresándola en forma de porcentaje es: $0.375 (100) = 37.5\%$, lo que indica que hay un 37.5% de vacas sanas en este conjunto de 80 hembras y, por tanto, la probabilidad de que una de ellas, elegida al azar, esté libre de mastitis es 0.375. Por lo anterior, se utilizarán como sinónimos de probabilidad términos como proporción, porcentaje y tasa.

En la tabla 2x2, a la probabilidad de un evento también se le llama probabilidad marginal, ya que su cálculo se basa en los totales marginales; basta comprobar que en el cuadro I.3 los valores del número de vacas gestantes -20- y el de hembras sanas -30- aparecen al margen, en los espacios correspondientes a las sumas parciales de cada variable.

En el cálculo de las probabilidades se pueden considerar diferentes conjuntos como base, de modo que resultan distintos tipos de probabilidad. En la tabla 2x2 se pueden obtener probabilidades de un evento, como las mostradas en los párrafos anteriores; probabilidades condicionales, cuando la base del cálculo es uno de los totales marginales; y probabilidades conjuntas de dos eventos, cuando se desea la probabilidad de que ocurran conjuntamente dos eventos determinados, como sería que una vaca este sana

y vacía o que tenga mastitis y este gestante. En las secciones siguientes se describen estos dos tipos de probabilidad.

b) LA PROBABILIDAD CONJUNTA DE DOS EVENTOS

Se mostró ya la manera de obtener la probabilidad de un evento, pero en la tabla 2x2 resulta de mucho interés calcular la probabilidad de que en un individuo ocurran dos eventos a la vez; por ejemplo, con base en la población presentada en el cuadro I.3, la probabilidad de encontrar una vaca gestante y de ubre sana. Este valor se conoce como probabilidad conjunta de dos eventos, ya que es la probabilidad de que se conjunten dos características distintas.

El cálculo de estas proporciones en la tabla 2x2 es sencillo, sólo requiere dividir el valor de a, b, c o d, según corresponda, entre el total de los casos, es decir, entre N. La probabilidad de una vaca gestante y de ubre sana en la población del ejemplo es $a/N = 13/80 = 0.1625$, expresado en porcentaje: $0.1625 (100) = 16.25\%$. La probabilidad de una hembra no gestante y con mastitis es $d/N = 43/80 = 0.538$.

c) PROBABILIDADES CONDICIONALES

Quando el porcentaje de casos que tiene cierta característica se calcula dentro de un subconjunto de los datos (r_2 , r_1 , c_2 o c_1) y no con respecto al total de ellos (N), el resultado recibe el nombre de probabilidad condicional, y representa la fracción de casos que poseen la característica de referencia dentro del subconjunto definido. Usando el cuadro I.3, la probabilidad de encontrar una vaca vacía entre las enfermas de mastitis, es de $d/r_1 = 43/50 = 0.86$ -en porcentaje es $0.86 (100) = 86\%$ -, esta es la probabilidad de que una vaca esté vacía tomando como referencia el grupo de vacas enfermas, lo cual señala que si se marca como condición que la hembra esté enferma, la probabilidad de encontrar una vacía es 0.86. Nótese que en el ejemplo la probabilidad marginal correspondiente es de 0.75, ya que 75% del total de 80 animales están vacíos.

Las probabilidades condicionales, junto con las de dos eventos son importantes para definir y entender las relaciones entre las variables que conforman la tabla 2x2. En la siguiente sección se aplican ambos conceptos para explicar la independencia entre dos variables binarias.

d) INDEPENDENCIA ENTRE DOS VARIABLES BINARIAS

Un concepto de gran importancia para los desarrollos posteriores es el de no asociación, también llamado independencia, entre las dos variables dicotómicas que forman la tabla 2x2. Si el valor que tiene un individuo para una de las variables no afecta las probabilidades para la otra, se dice que ambas son independientes, o bien, que no están asociadas.

En seguida se utilizarán los conceptos de probabilidad marginal y de probabilidad de dos eventos para describir la independencia entre dos variables dicotómicas. En este momento el propósito es conocer el concepto, por lo que se tratará con poblaciones completas. Las complicaciones inherentes a la estimación de valores en una muestra se tratarán en el Cap. III.

En esencia la independencia consiste en que lo sucedido en una de las variables no influye en el resultado de la otra; en términos de probabilidad, significa que el valor que tenga una de las variables no afecta las probabilidades en la otra.

En una tabla 2x2, si las dos variables consideradas son independientes, entonces las probabilidades condicionales serán iguales a la probabilidad marginal correspondiente; es decir, que si lo ocurrido en una variable no afecta lo que suceda en la otra, los probabilidades dentro de cada uno de los subconjuntos posibles deberán ser iguales. Por ejemplo, en la población de vacas del cuadro I.3 no hay independencia entre el estado reproductivo y la salud de la ubre, ya la probabilidad marginal de que una vaca esté gestante es de $20/80 = 0.25$, mientras que las probabilidades condicionales son: entre las de ubre sana de $13/30 = 0.433$, y entre las que tienen mastitis $7/50 = 0.14$; valores que son diferentes e indican que la probabilidad de encontrar una hembra gestante no es independiente de la salud de la glándula mamaria, lo que provocan que las probabilidades difieran de una columna a la otra, lo mismo que de un renglón a otro.

Considerando ahora una población de cerdas, (vease el cuadro I.4), si se calcula la probabilidad marginal de que una cerda haya abortado: $(20/160 = 0.125)$, resulta igual a la probabilidad condicionada a las cerdas con el síndrome del ojo azul $(5/40 = 0.125)$, y ambos cocientes son iguales a la probabilidad de aborto condicionada a las cerdas sanas: $(15/120 = 0.125)$. Entonces, es claro que los datos indican que la presencia de abortos es independiente de que la hembra tenga el síndrome mencionado.

La independencia también indica que la probabilidad conjunta de dos eventos específicos es el producto de las probabilidades marginales correspondientes, por ejemplo: $a/N = (r_0/N) (c_2/N)$, o bien $b/N = (r_1/N) (c_1/N)$. En el ejemplo de las cerdas, donde hay independencia, la probabilidad conjunta de aborto y ausencia del síndrome $(15/160 = 0.094)$ es igual al producto de las proporciones marginales respectivas: $(20/160) (120/160) = 0.094$; lo mismo que la probabilidad que una hembra tenga el síndrome y no aborte $(35/160 = 0.219)$ es igual al producto de las probabili-

ésta sobre la otra, llamada variable de respuesta. Por el contrario, en un estudio observacional solo se registran las condiciones en que ocurre el evento, para evaluar las relaciones entre las distintas variables registradas.

La variable cuyo efecto se analiza a través de un experimento se conoce como tratamiento, a diferencia de la que se evalúa mediante un estudio observacional, que se le designa atributo. Ambos son factores; en general, se llama factor a cualquier variable que pueda afectar la variable de respuesta.

Cuadro II.1

Clasificación de los estudios que originan una tabla 2x2

Sección	Clasificación				
II.2 *	Estudio de atributos		Estudio de tratamientos		
II.3 *	Muestreo multinomial	Muestreo comparativo	Muestreo experimental		
II.4 *	Muestras independ.	M. independientes	M. apareadas	M. independientes	M. apareadas
II.5 *	"	"	"	"	"
	Se obtiene una muestra de N individuos, que representa a la población de interés. El número de observaciones que conforman la muestra se determina antes de colectarla.	Se toma una muestra de F_0 casos con nivel 0 para X y otro con F_1 casos con la característica 1 en la variable Y . Dando un total de N unidades distintas: $N = F_0 + F_1$	Se examinan F_0 individuos con la característica 0 en la variable X y luego se les estudia cuando poseen el valor 1. Cada unidad se registra dos veces. $F_0 = F_1 = N/2$	Se eligen N unidades experimentales, a F_0 de ellas se les asigna el tratamiento 0 y a las restantes F_1 se les administra el otro. Como todo experimento implica la manipulación de los tratamientos. $N = F_0 + F_1$	A una muestra de F_0 sujetos se le da el tratamiento 0, luego se les somete al 1. En ambos casos se registra el resultado en Y . Cada sujeto origina dos observaciones. $F_0 = F_1 = N/2$
Ejemplo**	II.1	II.2	II.3	II.4	II.5

* Este número refiere a la sección del capítulo II en que se describe la clasificación.

** En todos los casos se puede dar la relación directa y la inversa entre la muestra y la población.

*** Hace referencia a los ejemplos presentados en el capítulo II.

Es importante considerar que el hacer referencia a tratamientos implica la manipulación del factor. Un ejemplo de factor que puede considerarse como tratamiento es el siguiente: un veterinario desea conocer la efectividad de cierta vacuna de reciente aparición en el mercado, en relación con la eficiencia de la que él aplica. Para lograrlo, administra al azar la nueva vacuna a la mitad de un grupo de perros y la otra al resto de ellos. De este modo, las vacunas serán los tratamientos y su comparación se hará bajo un estudio experimental. Si al concluir encuentra que el resultado fue distinto en ambos grupos, podrá inferir que se debe al efecto de la vacuna.

Para que los resultados de una investigación puedan apoyar la causalidad de un factor sobre determinada respuesta se requiere un estudio experimental basado en la aplicación de los tratamientos en forma aleatoria a los animales bajo estudio. La aleatorización al asignar los tratamientos en un experimento es lo que fundamenta la solidez de los resultados. Si la distribución de los tratamientos se hace realmente al azar, se espera que las diferentes características de los sujetos bajo estudio que puedan afectar los resultados de la investigación se presenten en forma similar en ambos grupos. Entonces, si hay diferencia en los resultados de ambos grupos, dado las muestras son similares en sus características y sólo difieren en el tratamiento que recibieron, es razonable suponer que la causa de la diferencia es, precisamente, el tratamiento considerado.

El estudio de un atributo nunca lleva a la conclusión de que el factor considerado es la causa de la respuesta; sólo se puede afirmar -si así lo indican los datos- que el factor examinado y la respuesta de interés están asociados. Esto podrá dar origen a un experimento con el fin de averiguar si la relación es o no causal.

El cuadro II.2 contiene una parte de los resultados obtenidos por Rivero (1984) que servirán para ilustrar lo explicado acerca de los atributos. Al encontrar una relación entre la talla del perro y la probabilidad de contraer dirofilariosis, Rivero especula que tal asociación se debe a que los animales de talla pequeña permanecen en interiores con mayor frecuencia que los más grandes, y por consecuencia están menos expuestos a los moscos que transmiten los parásitos causantes del problema.

Si la proposición es correcta, un factor que permanece fuera del alcance de la investigación: que el perro viva a la intemperie o en interior, es lo que predispone a la enfermedad, pero al asociarse con el atributo registrado, la talla del perro, hace parecer a éste como un factor que modifica la posibilidad de enfermar.

Incluso, existen estudios observacionales en los que no hay una separación clara y formal entre el atributo y la respuesta: desde diferentes puntos de vista cada una de las variables puede ser considerada como atributo o como respuesta de la otra. Todo esto lleva a concluir, aunque sea de manera intuitiva, que el análisis de atributos produce información menos sólida sobre su relación con la respuesta, que la obtenida en los experimentos.

Por ejemplo si se registra la presencia de enfermedades respiratorias en la marrana y de diarreas en la camada, habrá interés en saber si ambas se relacionan, y no en buscar causalidad de una sobre la otra; por tanto ninguna sería atributo o respuesta.

Cuadro II.2

Dirofilariasis de acuerdo a la talla del perro*

	T a l l a		Total
	Chica	Grande	
Dirofilariasis	Si	1 35	36
	No	19 19	38
	Total	20 54	74

* Modificado de Rivero (1984)

Muchos factores pueden estudiarse como atributos o como tratamientos, la distinción en un caso particular se desprenderá de la manera en que se desarrolle el trabajo, por tanto el informe de la investigación deberá señalar la clase de estudio utilizado, y la discusión y conclusiones habrán de hacerse conforme el tipo de investigación empleada.

3. TIPOS DE MUESTREO APLICABLES A LA TABLA 2X2

La forma en que se obtiene la muestra afecta de manera importante a la información contenida en ella; a la vez influye en el tipo y alcance de las conclusiones que pueden desprenderse del análisis. Por tanto, es importante conocer los diferentes procesos que pueden seguirse para obtener la muestra y cómo influyen los resultados del estudio.

En esta sección se describen los distintos tipos de muestreo que pueden emplearse en las investigaciones cuyos resultados pueden resumirse en tablas 2x2. También se analiza la información que puede obtenerse en cada caso a partir de la muestra.

a) MUESTREO MULTINOMIAL

La forma más simple de tomar una muestra para estudiar la relación entre dos variables dicotómicas es elegir aleatoriamente un número previamente estipulado de individuos de la población y luego clasificarlos de acuerdo con las dos variables de interés; esto se conoce como muestreo multinomial. Es importante que el total de casos que se incluirán en la muestra esté definido de antemano, de no ser así el comportamiento probabilístico de los resultados será diferente.

Como ejemplo se puede citar el siguiente estudio: para analizar la relación entre la madurez de la vaca y la mortalidad en los becerros, se toma una cantidad definida de los registros individuales, seleccionados al azar de los existentes en el rancho, y después se clasifican por el grado de madurez de la madre (joven o adulta) y la supervivencia del becerro hasta el destete. Debe notarse que primero se integra la muestra por analizar y luego se miden las dos variables binarias de interés (véase la Fig. 11.1a).

Un buen número de investigaciones siguen este esquema muestral, sobre todo de las que se realizan con base en registros de producción o a través de encuestas.

A partir de una muestra multinomial se puede estudiar la relación entre ambas variables, es decir la asociación de X con Y en la población de referencia.

Lo anterior implica que se pueden estimar todas las probabilidades marginales, lo mismo que las conjuntas de dos eventos, ya que el análisis de la asociación entre las variables requiere la comparación entre estos valores. El Cap. I contiene una explicación del concepto de asociación entre variables binarias y una descripción de la forma en que se manifiesta en la tabla 2x2.

b) MUESTREO COMPARATIVO

Cuando se planea realizar una comparación entre dos grupos de sujetos, se puede utilizar el muestreo comparativo, que consiste en seleccionar r_0 observaciones de un grupo y r_1 del otro; y después clasificar a cada sujeto conforme a la otra variable, denominada de respuesta.

Para ejemplificar el muestreo comparativo se empleará un brote de abortos en un rebaño. Si al analizar el problema surge la sospecha de que la enfermedad es más común en las borregas que reciben un cierto alimento concentrado comercial que en las hembras con distinta alimentación, se puede tomar una muestra aleatoria de r_0 animales que consumieron el alimento sospechoso y otra muestra de r_1 hembras que no lo comieron, para investigar a

continuación cuantas borregas de cada muestra abortaron. En este caso se está aplicando el muestreo comparativo para comparar el porcentaje de abortos en ambos grupos de hembras mediante un estudio retrospectivo (vease la Fig. 11.10.4).

Debe notarse que este esquema de muestreo sólo permite estimar las probabilidades condicionales, las probabilidades marginales reflejará el número de casos que se, si el investigador haya incluido en cada muestra; las proporciones marginales de X provienen directamente del número de casos que se foran cada muestra r_{0i} y r_{1i} , por lo que no reflejan las probabilidades marginales correspondientes. En tanto que las probabilidades marginales de la variable Y sólo serán alteradas si hay diferencias en los porcentajes de ambos grupos. Por ejemplo, si la tasa de abortos es mayor en las hembras que recibieron el alimento prohibido y durante el estudio se decide investigar una muestra tres veces mayor para este grupo, es decir tres borregas de la ración sospechosa por cada hembra de las que consiguieron otros alimentos, entonces el porcentaje global de abortos será sobreestimado porque se vería más influido por el grupo que tuvo una mayor frecuencia del problema.

Incluso considerar el mismo número de observaciones en cada muestra no garantiza una estimación válida de las probabilidades marginales de la variable de respuesta; en un caso donde la quinta parte de las borregas no recibió el alimento relacionado con el problema, se integran muestras del mismo tamaño para ambos grupos, se obtendrá una mayor influencia de la muestra (la mitad del valor global) que la importancia del grupo en el rebaño (la quinta parte). Resulta claro que al aplicar el muestreo comparativo, los valores de las probabilidades marginales tenderán a mostrar sesgos importantes, por lo que no se deben utilizar como estimadores de la tasa global de respuesta, si no que, manteniendo la concepción original de este esquema de muestreo, debe considerarse que cada muestra proviene de una población distinta y sólo deben estimarse los porcentajes de respuesta en cada una de ellas por separado. Si las variables son independientes, la estimación de probabilidades marginales para la variable Y es válida.

El muestreo comparativo permite contrastar las muestras con respecto a la proporción de individuos que poseen cierta característica. Conforme al ejemplo anterior, este esquema de muestreo permite estudiar si la tasa de abortos es igual en ambas poblaciones; o bien, si efectivamente difiere entre las hembras que recibieron distinta alimentación; pero no sirve para investigar si las variables son independientes.

c) MUESTREO EXPERIMENTAL

El esquema de muestreo que falta describir se utiliza en los estudios experimentales y consiste en una combinación de los anteriores. Se obtiene una muestra de N individuos y posteriormente se le asigna en forma aleatoria el tratamiento cero a r_0 de

ellos y el otro a los restantes r_1 casos. En la sección 11.2 se describen en detalle las características de un experimento.

Utilizando como ejemplo, de nuevo, el caso de los abortos en las borregas. Al sospechar la relación entre el aborto y el alimento concentrado podría seleccionarse una muestra de N hembras gestantes para administrar, aleatoriamente, el alimento problema a r_0 de ellas y el otro tipo de alimentación a las restantes r_1 . Así, se aplicaría un estudio experimental para analizar el problema descrito (véase la Fig. 11.1c).

Un muestreo experimental conduce a un estudio comparativo; sin embargo, existe una diferencia fundamental que es aplicar la aleatorización, es decir asignar por azar cada individuo a uno de los dos grupos, lo que permite la eliminación de sesgos en los resultados. Por tal motivo la información colectada es mucho más sólida en cuanto a la relación causa-efecto. En un muestreo comparativo del caso utilizado como ejemplo, podría ocurrir que por alguna razón técnica se diera el alimento problema a las borregas en peor estado físico y por este motivo, y no por el alimento, mostrarán una alta tasa de abortos. Tal tipo de dificultades no se presentan en un muestreo experimental, ya que al formar por azar cada grupo, se espera que las distintas características de los individuos que pudieran afectar la respuesta queden equitativamente distribuidas en ambas muestras, evitando sesgos por factores ajenos a la investigación.

Es evidente que la similitud entre muestras, resultante de aleatorizar, será mejor mientras más grande sea la muestra. En una muestra pequeña será difícil que se represente correctamente la población de referencia.

Obtener una muestra aleatoria o asignar al azar los tratamientos a los sujetos considerados en un experimento no se sólo hacerlo como parezca salir a la suerte, sino que debe aplicarse un mecanismo aleatorio. Si se quiere asignar dos tratamientos a un grupo de vacas y se decide aplicar un tratamiento a las primeras que entren al corral de manejo y el otro a las últimas, lo más posible es que un grupo incluya a las hembras de temperamento más tranquilo; si el estudio se enfoca, por ejemplo, al análisis de niveles hormonales o de eficacia reproductiva, es muy probable es que la manera de distribuir los tratamientos induzca algunas diferencia entre grupos y que estas se confundan con los resultados de los tratamientos.

Tampoco es necesario que el mecanismo de aleatorización sea complicado: pueden utilizarse trozos de papel en los que se ha escrito la identificación de cada uno de los sujetos considerados en el trabajo, revolverlos en un sombrero y separar a los que, por azar, recibirán cada tratamiento.

La falta de una verdadera aleatorización puede provocar que los datos generados del estudio carezcan de solidez. En una investigación sobre cojeras en pollos de engorda se analizaron, en varias parvadas, una serie de características en los pollos que cojearan y, para un estudio comparativo, se tomó una muestra de los pollos aparentemente sanos, pero en vez de una muestra alea-

toria se utilizaron los primeros pollos que se pudo capturar de entre los que no cojeaban. Antes de aceptar los resultados de tal estudio, cabría preguntarse si los animales incluidos como sanos eran los que empezaban ya con alguna molestia en las patas por lo que eran los más fáciles de atrapar, y que de no haberse retirado de la parvada muy pronto estarían cojeando. Esta duda invalida las conclusiones de tal investigación y hace obvia la importancia de emplear la aleatorización.

En un estudio experimental ni siquiera el tratar de "equilibrar" la asignación de tratamientos al gusto del investigador garantiza tan buenos resultados como una buena aleatorización. Al distribuir los tratamientos según algún criterio personal para equilibrar los grupos, es frecuente que se refleje la preferencia hacia uno de los tratamientos y que el grupo correspondiente tenga ventaja sobre el otro.

Debe notarse que el muestreo experimental permite inferir la causalidad del factor sobre la respuesta -en el ejemplo sería la causalidad del alimento sobre los abortos-, a diferencia del comparativo que sólo da base para investigar si la proporción de respuesta es similar en ambos grupos -en el ejemplo corresponde a verificar si la tasa de abortos es semejante entre los grupos de borregos que reciben una y otra ración-, sin indicar una relación de causa a efecto.

4. MUESTRAS INDEPENDIENTES CONTRA DATOS APAREADOS

Este criterio de clasificación se basa en la relación existente entre las dos muestras resultantes de los muestreos comparativos y experimentales. Se llaman muestras independientes a las que se integran con individuos que no tienen una relación especial entre sí; a su vez, se designan como muestras apareadas las que surgen cuando ambas muestras se forman con datos del mismo sujeto o de parejas de individuos que comparten una o varias características que pueden afectar los resultados del estudio.

Por supuesto, los individuos considerados en todas las muestras tienen en común las características generales que delimitan la cobertura de los resultados de la investigación, es decir las que definen la población de interés. Por ejemplo, al estudiar la fertilidad de cerdas jóvenes y adultas inseminadas artificialmente, todas las hembras comparten las cualidades que delimitan la población, como son especie, clima, tipo de alimentación sistema de crianza y las condiciones particulares de la granja, entre otras. Los resultados, en forma estricta, sólo pueden aplicarse a los animales que posean características similares.

Si la muestra de cerdas jóvenes se constituye con I_n puercas tomadas aleatoriamente y, por separado, se forma la muestra de

hembras adultas con n_1 animales elegidos al azar, se trata de un estudio con muestras independientes. Pero si se integra la muestra de cerdas jóvenes, registrando la respuesta a la inseminación (gestante o no) y cuando las mismas hembras sean adultas se toman como muestra de las adultas, entonces se trata de dos muestras relacionadas, ya que cada animal proporciona información a cada una de ellas.

No todas las muestras apareadas requieren que el mismo individuo se considere en ambos grupos, también pueden formarse parejas de sujetos semejantes en una o más características; éstas se utilizan como criterios de apareamiento porque se sabe o se cree que influyen en la variable de respuesta y se requiere evitar su efecto sobre el resultado de la investigación. Los factores utilizados como criterios de apareamiento no se estudian, sólo se controlan.

Un caso de se presentaría al formar parejas de cerdas jóvenes y adultas que hayan sido similares en peso al nacimiento, número de hermanos de camada, peso a los 100 días de edad y antecedentes reproductivos de su madre, para integrar las muestras de los dos grupos con un miembro de cada pareja. Aun cuando son dos animales distintos, cada pareja tiene importantes analogías en factores que podrían influir los resultados de la investigación.

Si bien las muestras apareadas son un método muy útil para evitar que factores extraños a los analizados modifiquen y confundan los resultados, es importante destacar que su uso implica una reducción de la potencia de las pruebas estadísticas; esto es, disminuye la probabilidad de que, en las muestras, se encuentre significativa una diferencia que existe entre las poblaciones de referencia (véase el Cap. IV.1. para una explicación del término potencia). La reducción de la potencia puede comprobarse en ejemplo de la inseminación artificial en cerdas: si un veterinario estudia dos muestras independientes, es decir un grupo de 20 hembras jóvenes y otro de 20 adultas constituidos por separado, cuenta con un total de 40 cerdas para su investigación. Por el contrario, si registra 20 puercas jóvenes y después examina la presencia o ausencia de gestación al inseminar las mismas hembras ya adultas, tiene dos muestras de 20 animales cada una para su investigación, pero en realidad sólo tiene un total de 20 hembras, por lo cual el tamaño efectivo de la muestra es la mitad que en el primer caso. En el Cap. VI.1 se explica la relación del tamaño de la muestra con la potencia de las pruebas estadísticas.

5. RELACION DE LA POBLACION CON LA MUESTRA

La relación de la población de interés con la muestra analizada es un punto de singular importancia por su impacto en las conclusiones de la investigación. Se distinguen dos tipos básicos en esta relación: la relación directa, que ocurre al definir la

población para luego extraer la muestra, y la relación inversa, que se presenta cuando la muestra existe o se agrupa por alguna circunstancia y después se especifica la población representada en ella, es decir la población a la cual pueden aplicarse las conclusiones logradas.

Un ejemplo de relación directa de la población con la muestra es el de un veterinario que está interesado en la frecuencia de cierta enfermedad en las granjas de la zona en que trabaja, y selecciona en forma aleatoria una muestra de las granjas que le interesan a partir de un registro de todas las existentes en el área. Por otro lado, usando el mismo ejemplo, la relación inversa se presenta cuando el médico estudia la enfermedad de interés en las granjas que atiende y posteriormente analiza y describe la muestra para indicar cuál es la población que se representa en ella; se pretende que las conclusiones de la investigación se puedan generalizar a esa población.

Es claro que la aplicación de los resultados de la investigación es más consistente y mejor fundamentada si la relación de la población con la muestra es directa que cuando es inversa. Esta última no invalida los resultados logrados pero sí obliga a tener precaución al extrapolar la información generada.

6. EJEMPLOS INTEGRADOS

Los puntos de vista expuestos en las cuatro secciones anteriores estudian separadamente los distintos aspectos de las investigaciones cuyos resultados se plasman en tablas 2x2. En esta sección se describen algunos casos que integran estos aspectos y que servirán para ejemplificar las combinaciones presentadas en el cuadro II.1

Ejemplo II.1: un caso de muestreo multinomial

Un médico veterinario desea comparar dos métodos para detectar cisticercos en canales de cerdo. Con tal propósito, acude a un rastro para examinar un número determinado de los canales de los puercos ahí sacrificados, y registrar en cada caso el resultado de la inspección con cada método (contaminado o libre de cisticercos). Al mismo tiempo anota las características generales del rastro y de los animales sacrificados.

Este caso presenta el estudio de atributos por medio de un muestreo multinomial. Se trata de una sola muestra, así que no existe el problema de clasificarla como apareada o independiente, ya que este criterio señala el nexo entre dos muestras. La rela-

ción entre la muestra y la población es inversa, es decir, la población a la que sean aplicables los resultados dependerá de la clase de cerdos que circunstancialmente constituyan la muestra; por tanto conviene contar con un registro detallado de las particularidades de esta, de modo que puedan describirse minuciosamente en el informe final.

Si el investigador hubiera definido una población porcina de interés, extraído una muestra de ella e inspeccionado las canales con los dos métodos a comparar, consignando en cada caso el resultado, entonces se tendría una relación directa de la muestra con la población, lo que evitaría dudas al utilizar las conclusiones del trabajo, ya que la población de referencia estaría bien definida, a la vez que se tendría más confianza en la representatividad de la muestra.

Ejemplo 11.2: muestreo comparativo

Al investigar acerca del trasplante de embriones en vacas lecheras, se desea saber si el éxito está influido por el historial reproductivo de la hembra nodriza. Con tal fin, dentro del hato de interés se toma una muestra de vacas con antecedentes de falla reproductiva y otra de hembras que nunca han presentado tales problemas. Posteriormente, todas ellas se utilizan como nodrizas de embriones transferidos y se determina si quedan o no gestantes.

Este ejemplo muestra una investigación sobre atributos, basada en un muestreo comparativo con muestras independientes. Es importante notar que la diferencia con el ejemplo anterior radica en que el número de vacas en cada tipo de historial reproductivo es decisión de quien diseña el estudio, con lo cual manipula este criterio de clasificación y sólo el éxito o fracaso del trasplante se produce en forma aleatoria. La relación entre el muestra y la población es directa: las dos muestras provienen de poblaciones previamente definidas, en este caso la de vacas con fallas reproductivas previas y la de aquellas con un historial reproductivo limpio. Estas dos poblaciones constituyen el hato de referencia.

Ejemplo 11.3: ejemplo de muestras apareadas

Para comparar la fertilidad de dos razas de cabras, en cierta zona, un investigador toma un grupo de animales de cada una en la región de interés y forma parejas de hembras con el mismo número de partos y no más de 5 kg de diferencia en peso. En seguida les da servicio con el macho, registrando en ambos casos si hubo preñez.

El caso descrito presenta un ejemplo del estudio de atributos mediante un muestreo comparativo con muestras apareadas. Es importante notar que aunque cada par de mediciones se obtiene de dos hembras distintas, ambas representan un tipo de cabra de cierta edad y peso.

Debido a que el investigador definió la población de interés y luego extrajo las muestras de cada raza, la relación entre las muestras y la población es directa.

Ejemplo II.4: estudio experimental

Como parte de una investigación se quiere conocer la eficacia de cierta vacuna para gallinas de postura. Con tal propósito, se reúne una muestra de esas aves y se divide aleatoriamente en dos grupos. Uno es tratado con la vacuna y el resto de las aves no reciben tratamiento. Se anota el número de gallinas enfermas de cada grupo durante los seis meses siguientes.

El párrafo anterior presenta el caso de un estudio experimental basado en muestras independientes, el cual se aplica para analizar tratamientos. El proceso parte de una sola muestra que luego se subdivide en dos y cada una de éstas recibe uno de los tratamientos, a cada sujeto experimental puede tocarle en suerte cualquiera de los dos. Frecuentemente este tipo de estudios se utiliza para comparar el resultado de la aplicación de un tratamiento con la ausencia de éste o la eficiencia de un nuevo proceso con el que se encuentra en uso.

Ejemplo II.5: experimento con muestras apareadas

Para estudiar las ventajas de un sustituto de leche para cerditos, sabiendo que hay muchas características peculiares a cada camada que pueden influir en el resultado, un grupo de investigadores toma un par de lechones de cada camada y asigna aleatoriamente uno de cada par a cada tratamiento: leche de cerda o sustituto de leche. Para cada par registran si se presenta diarrea durante la lactancia en uno (marcando cual de los dos fue), ninguno o ambos de los lechones.

Ahora se tiene un ejemplo de estudio experimental basado en muestras apareadas. Se ha utilizado nuevamente un caso en que cada pareja de datos se obtiene de dos individuos distintos, con objeto de recalcar este tipo de muestras apareadas; sin embargo, el esquema descrito es más usual en los estudios de antes y después, donde un individuo aporta datos antes del tratamiento y después de recibirlo, obteniéndose cada par de medidas a partir del mismo sujeto.

2. ESTIMACION DE LAS MEDIDAS DESCRIPTIVAS

Para calcular un parámetro es necesario disponer de los datos de toda la población de interés. En cualquier estudio cuyo objetivo tenga cierto grado de generalidad, el tamaño de la población hará prácticamente imposible medir a todos los individuos. Por tanto, se estudia una parte de la población, llamada muestra, que permita obtener un estimador del parámetro. Si la muestra cumple con los requisitos de calidad necesarios (véase la Sección II.3) puede esperarse que el valor del estimador y el del parámetro sean similares.

a) ESTIMACION DE PUNTO O PUNTUAL

Un estimador de punto es una estadística que, bajo ciertas condiciones, se presume tiene un valor cercano al del parámetro estimado. Por lo general, el estimador se calcula en la misma forma que el parámetro, pero en vez de usarse los datos de toda la población de referencia se utiliza la información obtenida de una muestra.

La similitud esperada entre el parámetro y su estimador es una función de la variabilidad de la estimación, la cual se evalúa con el error estándar del estimador, que indica la magnitud de la variabilidad por azar que puede ocurrir en el estimador.

El error estándar de un estimador es inversamente proporcional a su precisión, cuanto más pequeño es hay menores variaciones aleatorias en la estimación y, por tanto, se tiene un estimador más preciso.

Cuando no se analiza la población completa sino sólo una muestra, no basta con estimar el parámetro; es necesario evaluar la consistencia de la estimación, lo que puede hacerse, en parte, mediante los intervalos de confianza; pero, antes de cualquier generalización respecto al asunto investigado, se debe determinar la significancia de la estimación alcanzada, es decir, si los resultados obtenidos mediante la muestra verdaderamente apoyan tal o cual conclusión aplicable a la población de referencia. Para lograrlo deben utilizarse las pruebas de hipótesis descritas en el siguiente Cap. y, para algunos casos especiales, en el Cap. V.

b) ESTIMACION POR INTERVALO

A diferencia del estimador de punto, que es un sólo valor, la estimación por intervalo se constituye con dos límites, para los cuales se tiene cierto grado de confianza de que el parámetro esté contenido entre ambos; es decir, existe una probabilidad conocida de que el parámetro sea mayor que el límite inferior y menor que el superior. Los límites que cumplen con esa condición forman el intervalo de confianza.

La probabilidad de que un intervalo contenga al parámetro se conoce como confiabilidad del intervalo. El nivel de confianza para el intervalo es decisión de quien lo calcula. Los más comunes son los intervalos al 95% y al 99% de confianza.

Para una misma muestra, un intervalo al 95% de confianza es más estrecho que uno al 99%, pero menos confiable: la probabilidad de que contenga al parámetro es menor. El nivel de confianza para un intervalo se decide en función de la amplitud del mismo -buscando que sea lo más estrecho posible- y de la confianza que ofrece -tratando que sea máxima-. Como uno y otro criterio se contraponen, en cada caso debe decidirse a cuál dar más peso y con ello qué grado de confianza dar al estimador de intervalo.

Conceptualmente, el nivel de confianza equivale al complemento de la probabilidad de cometer un error tipo I al probar una hipótesis, es decir a $(1 - \alpha)$, por lo que la lectura de la introducción del capítulo IV apoyará la comprensión de lo expuesto.

La estimación por intervalo se basa en un estimador de punto del parámetro correspondiente y el error estándar del estimador, en combinación con un factor de confiabilidad.

El intervalo de confianza brinda información muy importante para lograr una buena interpretación de los resultados. La amplitud del intervalo depende de dos factores: primero, del nivel de confianza, como ya se explicó y, segundo, de la precisión del estimador de punto; si el estimador es muy preciso -el error estándar es pequeño- el intervalo de confianza será estrecho, pero si la estimación es vaga será amplio. Por ello un intervalo de confianza demasiado amplio señala la falta de precisión en el estimador del parámetro, indicando que los resultados no aportan información concluyente acerca del problema.

Cuando el intervalo de confianza es más amplio de lo aceptable, antes que hacer inferencias sobre el parámetro se debe pensar que la muestra no es lo bastante buena para brindar una base adecuada a la inferencia, por tanto conviene tomar una muestra mayor.

Aceptar el intervalo de confianza o considerarlo demasiado amplio es cuestión de criterio de quien lo interpreta, depende principalmente del problema que se estudia y de lo que se espera conocer acerca de él, por lo que no obedece a más reglas que la opinión del que realiza la investigación.

Aunque el aspecto descrito de la interpretación de los intervalos de confianza basta para justificarlos, no es el único uso que puede dárseles: el intervalo de confianza indica los niveles en que puede ocurrir el parámetro estimado, dando una idea del rango en que puede estar el valor real del parámetro.

Por ejemplo, si se estima que la proporción de animales no vacunados que puede enfermar es de 5%, un veterinario pensaría que no es conveniente vacunar; pero si se le informa que la proporción de animales que desarrollan la enfermedad varía de 0 a 15%, y él considera que un porcentaje de enfermos mayor al 10% justifica la protección, entonces concluirá que, en tanto no se tenga nueva información, será recomendable vacunar. Es posible que al mismo tiempo se dedique a estudiar mejor el problema a fin de determinar con mayor precisión el riesgo de no vacunar. Con sólo un estimador de punto no es posible lograr esta interpretación, por lo que no emplear el intervalo de confianza llevaría a perder parte importante del provecho que puede lograrse de los datos colectados.

Si bien la aplicación de los intervalos de confianza es de singular utilidad, no siempre pueden calcularse. Para las medidas presentadas se utilizan aproximaciones que permiten aplicar la distribución normal de probabilidades como base de la estimación por intervalo, para justificar una buena aproximación la muestra no debe ser muy pequeña. Aunque es difícil fijar tajantemente un punto mínimo para que la aproximación a la distribución normal sea válida, puede considerarse que sólo es útil en muestras con más de 20 observaciones.

Quando se trata de alguna estadística basada en proporciones y esta es menor que 30% o mayor que 70%, se requieren mas datos para lograr la aproximación deseada. Como una regla de uso práctico, se considera que si el producto de la fracción por el tamaño de la muestra es mayor de seis puede calcularse el intervalo; por ejemplo, en una muestra de 35 donde la proporción de interés es 18%, el producto es $35(0.18) = 6.3$, que permite la estimación por intervalo. Una muestra de 58 no sería suficiente para estimar un intervalo dirigido a una medida de 7%, porque $58(0.07) = 4.1$ es menor que seis. Cuando la fracción es mayor que 70%, el cálculo anterior debe hacerse con el complemento, es decir con [1-fracción]; para un caso de 94% y una muestra de 46 no habría suficiente información, porque $1-0.94 = 0.06$ y $46(0.06) = 2.8$ no alcanza el mínimo planteado.

3. DESARROLLO DE UNA MEDIDA DESCRIPTIVA: LA PROPORCION

La medida estadística más simple que puede asociarse a una variable binaria es la proporción. En esta sección se describe el proceso para obtener un estimador de punto y uno de intervalo para una proporción, lo que será útil en dos sentidos: a) se ex-

pondrá con más detalle lo analizado en las secciones anteriores acerca de la estimación y, b) se darán las bases para tratar las medidas descriptivas que son casos especiales de una proporción, como la prevalencia y la sensibilidad.

Una proporción expresa la frecuencia con que se presenta una cierta clase de datos dentro de un conjunto general. Si se designa con e los datos de interés y con N al total de observaciones, la proporción se calcula como:

$$e = e / N$$

Si se dispone de los datos de toda la población, e es resultado es un parámetro; pero, si sólo se cuenta con una muestra, el resultado será un estimador de la proporción.

Calcular la proporción poblacional evita todo el problema de la estimación: sabiendo que en un conjunto definido de datos la proporción es constante, esto es que si no varían los registros la proporción no cambia, entonces se tendría la información requerida sobre esa población.

Cuando sólo se dispone de una muestra el caso es muy distinto: se puede obtener un estimador de la proporción y no su valor verdadero en la población. Se espera que la proporción muestral se acerque al parámetro, esta cercanía depende de la cantidad de datos en la muestra y del procedimiento aplicado para reunirlos.

Para valorar mejor el resultado, se obtienen los intervalos de confianza que además de señalar un rango para el verdadero valor de la proporción -el parámetro-, también indican la precisión de su estimador. La forma general de un intervalo de confianza es:

Estimador \pm factor de confianza (error estándar del estimador)

El estimador de punto es el centro del intervalo y alrededor de él se extienden los límites; la distancia de los límites al centro depende del factor de confianza -que aumenta al disminuir α - y del error estándar del estimador, que será menor conforme la muestra sea mayor, y que también se reduce cuando la variabilidad en la muestra disminuye.

En el caso de una proporción, el estimador es precisamente la proporción en la muestra; el factor de confiabilidad proviene de la distribución normal: 1.960 para el 95% de confianza y 2.576 para el 99%. Para los intervalos de confianza de todas la estadísticas descritas en este capítulo se usarán estos factores de

confiabilidad. El error estándar del estimador de una proporción se calcula:

$$S(p) = \sqrt{\left(\frac{p(1-p)}{N}\right)}$$

De modo que el intervalo de confianza para una proporción es:

$$\text{Al 95\% de confianza: } p - 1.960 S(p); \quad p + 1.960 S(p)$$

$$\text{Al 99\% de confianza: } p - 2.576 S(p); \quad p + 2.576 S(p)$$

Para obtener el límite superior del intervalo de confianza, al estimador se le suma el resultado de multiplicar el factor de confiabilidad por el error estándar, y se le resta para calcular el límite inferior, constituyéndose así el espacio en el cual se tiene cierto grado de confianza de que contenga el parámetro representado por el estimador de punto que está al centro del intervalo.

EJEMPLO III.1: estimador de una proporción

Suponiendo que al inseminar un grupo de 500 borregas 149 quedaron gestantes. La proporción de ovejas gestantes es $149/500 = 0.298$, que corresponde a 29.8%. Considerando que las 500 hembras son una muestra de la población a la que puede aplicarse este método de inseminación, entonces conviene calcular un intervalo de confianza, al 95% se obtiene:

$$S(p) = \sqrt{\left(\frac{0.298(1-0.298)}{500}\right)}$$

$$p \pm 1.960 S(p)$$

$$0.298 \pm 1.960 (0.02045)$$

$$0.298 - 0.040 \text{ 08}; 0.298 + 0.040 \text{ 08}$$

Intervalo de confianza (95%): 0.257 92; 0.338 08

Por tanto, al inseminar borregos similares en las mismas condiciones en que se hizo el trabajo se tiene un 95% de confianza en obtener una fertilidad de 26% a 34%.

Si la confianza se aumenta al 99%, el factor de confiabilidad aplicable es 2.576 y el intervalo se amplía a 25-35% (0.245 32; 0.350 68); puede verse que estos límites son más amplios, y que se tiene mayor probabilidad de que el verdadero valor de la fertilidad se encuentre entre los límites.

EJEMPLO III.2: estimación de la mortalidad en una granja porcina

En otro caso puede estimarse la proporción de lechones muertos en lactancia. Para lograrlo, en una granja se tomó una muestra de 65 lechones, de los cuales murieron siete; la tasa de mortalidad es $7/65 = 0.107 \text{ 69}$. Con un 99% de confianza, la proporción de muertes estará entre:

$$0.107 \text{ 69} \pm 2.576 \text{ (0.038 45)}$$

$$0.008 \text{ 64}; 0.206 \text{ 74}$$

Lo cual indica que la mortalidad en esa granja, expresada en porcentaje y redondeando, debe fluctuar entre 1 y 21%.

4. CLASIFICACION DE LAS MEDIDAS DESCRIPTIVAS

No obstante que todas las medidas descriptivas aplicables a una tabla 2x2 son diferentes expresiones de la asociación entre las dos variables binarias que la forman, a fin de analizarlas mejor pueden dividirse en: las que sólo describen la asociación

entre las dos variables dicotómicas, aquellas que evalúan la capacidad descriptiva de una variable respecto a la otra y las que comparan los dos niveles de una variable (atributo o tratamiento) en cuanto al comportamiento de la otra (respuesta). Las primeras se englobarán como medidas de asociación, las segundas bajo el nombre de medidas de predicción y las últimas como medidas de comparación.

Aunque menos comunes que los anteriores, hay otros dos aspectos de la tabla 2x2 que resulta interesante evaluar: la concordancia entre dos criterios de clasificación y la validez de una prueba diagnóstica. Para presentar esos conceptos se formaron dos grupos con las medidas descriptivas destinadas a evaluar estos casos especiales de asociación, uno para las medidas de concordancia, que incluye las estadísticas que se enfocan a ese tipo singular de asociación, para lo cual evalúan la similitud de las decisiones tomadas al clasificar a un conjunto de sujetos conforme dos sistemas diferentes. El último grupo se constituye con las estadísticas utilizadas para valorar algunos aspectos de la validez de las pruebas diagnósticas, llamadas medidas de la calidad de una prueba diagnóstica, que llevan en sí conceptos singulares de asociación entre los resultados de la prueba en análisis y una clasificación utilizada como referencia; en este grupo es especialmente notorio que, a pesar que son estadísticas de uso muy frecuente en medicina, son útiles en otros campos.

El cuadro III.1 muestra la clasificación propuesta para las medidas estadísticas que se presentarán en este capítulo. A pesar de que las medidas pertenecientes a los grupos citados tienen diferentes objetivos, en la tabla 2x2 algunas de ellas coinciden; así se encontrarán medidas que difieren en tablas de contingencia mayores a 2x2, pero que en ésta arrojan el mismo resultado. Por tanto resulta difícil mantener todas las estadísticas dentro de un solo grupo de la clasificación expuesta. Por suerte este problema origina una ventaja: cuando dos estadísticas arrojan el mismo resultado, debe entenderse que en este valor se reúnen las cualidades de las dos medidas coincidentes.

5. MEDIDAS DE ASOCIACION

Las medidas de asociación tienen por objeto describir la intensidad con que se relacionan o asocian dos variables aleatorias; dicho de otro modo, aquellos individuos que tienen igual valor para una variable, qué tan parecidos son para la otra. En una tabla 2x2, la asociación es función de los porcentajes registrados en cada una de las cuatro celdas de la tabla.

Cuadro III.1

Clasificación de las medidas estadísticas de tablas 2x2*

Grupo en que se clasificó	Nombre y símbolo de la estadística
Medidas de asociación	Phi cuadrada (Φ^2) Proporción explicada de la varianza (P_{ev}) Phi (Φ) C de Cramer (C) Tau b de Kendall (τ_b) Coeficiente medio de contingencia (C_m)
Medidas de predicción	Delta de Somers (δ) Gamma (Γ) Lambda (λ) Riesgo atribuible (R_a) Potencia atribuible (P_a) Proporción atribuible ($Prop_a$)
Medidas de comparación	Razón de momios (M) Riesgo relativo (R_r) Potencia relativa (P_r)
Medidas de concordancia	Kappa de Cohen (K)
Evaluación diagnóstica	Sensibilidad (S) Especificidad (E) Sensibilidad diagnóstica (S_d) Especificidad diagnóstica (E_d)

Las pruebas estadísticas de hipótesis son el tema del siguiente capítulo. Primero se presentan las medidas estadísticas -en este Cap.- y luego se tratan los métodos enfocados a la inferencia de tales medidas. Sin embargo, hay varias estadísticas que se basan en el resultado de la prueba de hipótesis más conocida, la prueba de ji-cuadrada (χ^2) para independencia, por lo que aquí se adelantan algunos conceptos sobre ella.

La prueba de ji-cuadrada toma su nombre de la distribución ji-cuadrada de probabilidad en la cual se basa. Sirve para probar hipótesis en relación con la independencia entre dos variables nominales; proporcionando un criterio objetivo para verificar si las frecuencias observadas en cada categoría son compatibles con

la independencia entre las dos variables consideradas (en el Cap. I se explica el concepto de independencia). Como base de la prueba se obtienen los valores que resultarían de la independencia total, llamados frecuencias esperadas. Para el caso de la tabla 2x2 se obtienen mediante las siguientes expresiones:

$$\begin{aligned} \text{Esperado en la celda: } a &= \frac{r_0 \cdot c_0}{N} \\ b &= \frac{r_0 \cdot c_1}{N} \\ c &= \frac{r_1 \cdot c_0}{N} \\ d &= \frac{r_1 \cdot c_1}{N} \end{aligned}$$

La comparación de los valores esperados con sus correspondientes observados se efectúa mediante:

$$\chi^2_{\text{e}} = \sum_i \frac{(\text{OBSERVADO}_i - \text{ESPERADO}_i)^2}{\text{ESPERADO}_i}$$

La ecuación anterior indica que para cada celda se obtenga la diferencia entre la frecuencias observada en la muestra y la esperada conforme a la hipótesis de independencia, se eleve al cuadrado y se divida entre la esperada, por último se sumen los resultados de todas las celdas. Para el caso de la tabla 2x2, el cálculo se reduce a:

$$\chi^2_{\text{e}} = \frac{N (a \cdot d - b \cdot c)^2}{c_0 \cdot c_1 \cdot r_0 \cdot r_1}$$

El resultado de este cálculo se compara con la distribución de ji cuadrada con un grado de libertad, si el resultado es mayor que el valor de las tablas de χ^2 la desviación con respecto a la independencia es significativa, es decir que las variables no son independientes. Si el valor es menor se dice que los datos no indican una relación significativa entre las variables, o bien que no se encontró una desviación significativa respecto de la independencia entre ellas.

Sabiendo que el valor de χ^2 , resultante de aplicar la prueba χ^2 para independencia es una función del grado de asociación entre las dos variables, es natural pensar en emplearlo como una medida de la asociación presente en una tabla de contingencia. Sin embargo, aunque χ^2 es un excelente medio para decidir sobre la significancia de la relación entre las variables, no sirve por sí mismo como indicador del grado de asociación. El valor de χ^2 ,

En una tabla 2×2 , está relacionado con la proporción de cada celda y con el número total de casos estudiados; en tanto que una medida que refleje la fuerza de asociación entre ambas variables sólo debe ser función de las probabilidades contenidas en cada celda de la tabla 2×2 , por lo cual el valor de χ^2 no puede aplicarse con este propósito.

a) Phi Cuadrada (ϕ^2)

Karl Pearson (1900), al presentar el valor de χ^2 como estadística de prueba para la asociación entre dos variables cualitativas, propuso la estadística ϕ^2 como indicador del grado de asociación entre las variables que forman la tabla de contingencia. ϕ^2 proviene de estandarizar el valor de χ^2 según el número total de datos en la tabla, al efecto se usa χ^2 tal como se calculó en la sección anterior, sin aplicar la corrección de Yates (Véase esta corrección en la sección de χ^2 en el Cap. IV).

$$\phi^2 = \chi^2 / N$$

Aunque ϕ^2 puede utilizarse como medida de asociación, no siempre tiene una interpretación sencilla que permita comparar los valores de ϕ^2 calculados en tablas de distinto tamaño; incluso en tablas mayores a 2×2 , ϕ^2 puede tomar valores superiores a uno. La dificultad de interpretación y el problema de comparabilidad aparecen en todas las medidas descriptivas derivadas de χ^2 .

El problema planteado ha llevado a varios autores a afirmar que las estadísticas basadas en χ^2 son apenas más que simples índices que señalan cuál tabla tiene más asociación que otra, pero no dejan muy claro en qué sentido miden la asociación. Sin embargo, al restringir la discusión a la tabla 2×2 , varias medidas derivadas de χ^2 tienen cualidades interesantes que permiten una interpretación sin ambigüedades y la comparación objetiva entre resultados de varias tablas 2×2 , por lo que son de singular utilidad en esta tabla.

La correlación (r) -correlación lineal simple o coeficiente de correlación- es la estadística utilizada para medir la asociación entre dos variables aleatorias continuas, la frecuencia con que se aplica la ha llevado a ser un concepto que todos entienden. El cuadrado de la correlación, llamado coeficiente de determinación (r^2), también se emplea con frecuencia y se interpreta como la cantidad de varianza o variabilidad de una variable que se refleja en la otra, o bien como la proporción de varianza que es común a ambas variables. La correlación fluctúa entre menos uno y uno y el coeficiente de determinación va de cero a uno.

Es interesante advertir que θ^2 es igual al producto de la ξ de Somers (ξ se explica en la siguiente sección, que corresponde a las medidas de predicción), considerando la variable Y como respuesta, multiplicada por la ξ inversa -tomando X como respuesta-, con lo que el resultado tiene gran similitud con el del coeficiente de determinación obtenido a través del producto de la pendiente de la regresión de X en Y por la pendiente de Y en X .

$$\theta^2 = \tau_u^2 = s_{xy} s_{yx}$$

$$r^2 = b_{xy} b_{yx}$$

Así, la ξ de Somers es un concepto muy similar al de la pendiente en la regresión lineal simple, mientras que θ^2 y τ_u^2 son análogos al coeficiente de determinación y, por ende, θ y τ_u al de correlación. En las siguientes secciones se explican θ , τ_u y ξ .

b) ζ de Cramer

Se han desarrollado diversas modificaciones al valor de θ^2 ; Cramer propuso la llamada estadística ζ de Cramer, que se obtiene al ajustar θ^2 conforme el número de renglones y columnas en la tabla de contingencia. El ajuste consiste en dividir θ^2 entre el menor de los números de columnas y renglones menos uno, que en el caso de la tabla 2×2 es siempre dos menos uno, por lo cual, en este caso, θ^2 y la ζ de Cramer coinciden.

c) Proporción explicada de la varianza (P_{ev})

Una de las formas más significativas para evaluar la asociación entre dos variables es determinar cuánto de la varianza presente en una variable puede explicarse en función de la varianza en la otra, esto es que proporción de la varianza en la primera está explicada por la varianza de la segunda; el concepto es fácil de entender e interpretar por lo que se usa con frecuencia. Si se conoce el concepto de varianza, intuitivamente se da una interpretación correcta al de proporción explicada de la varianza; la varianza se explica en la mayoría de los libros de estadística general.

Gini (1912) desarrolló una expresión para evaluar la variabilidad presente en una tabla de contingencia; a su vez, Light y

Margolin (1971) demostraron que posee cualidades que la destacan como una buena medida de variancia, es decir, que logra una evaluación justa de la variabilidad. A partir de la medida de Gini puede aplicarse el concepto de proporción explicada de la variancia a las tablas de contingencia.

En cualquier tabla 2×2 , la proporción explicada de la variancia, P_{ev} , tiene el mismo valor que θ^2 , por lo cual no se requiere desglosar el cálculo (véase a Bishop, 1979, para una explicación de P_{ev} en tablas mayores a 2×2).

La estadística que señala la proporción explicada de la variancia entre dos variables cuantitativas es el coeficiente de determinación (r^2), por tanto se encuentra de nuevo que θ^2 representa en las tablas de contingencia 2×2 el mismo concepto que r^2 en las variables cuantitativas. De tal manera, θ^2 también puede interpretarse como la cantidad de variancia común a ambas variables o como la magnitud de la variancia en una variable que puede explicarse a través de la otra.

d) Phi (θ)

Conforme a lo explicado en la sección de θ^2 (III.5a), en las tablas 2×2 la raíz de θ^2 ($\theta = \sqrt{\theta^2}$) puede interpretarse de la misma manera que el valor absoluto del coeficiente de correlación. Por tanto, será cero cuando no exista asociación y uno cuando sea perfecta; los valores cercanos a uno indicarán una asociación fuerte, en tanto que los valores próximos a cero ocurrirán cuando la relación sea débil.

Berger (1961) describe un método que permite la comparación objetiva entre dos estadísticas θ obtenidas en estudios independientes. Por su parte, Fleiss (1973) destaca que no deben compararse las θ calculadas en investigaciones con diferentes esquemas de muestreo. El capítulo II describieron los esquemas de muestreo que pueden originar una tabla 2×2 .

Al analizar la aplicación de θ , Carrol (1961) encontró que si una o las dos variables dicotómicas son partición de una variable continua (la Sec. 1.4 contiene ejemplos de esas particiones), el valor de θ depende del punto de corte utilizado. De hecho, este problema aparece siempre que se utiliza una variable binaria generada al particionar un fenómeno continuo; es fácil ver que el punto de corte influye de modo determinante en la proporción incluida en cada uno de los dos grupos producidos y, por ende, en el resultado de los análisis del caso.

e) Tau b de Kendall (τ_b)

Una medida de asociación muy difundida es la τ_b de Kendall. En el caso de la tabla 2x2, los valores de ϕ y τ_b siempre son de la misma magnitud, la diferencia consiste en que τ_b es positiva cuando los sujetos con el valor considerado mejor en X tienden a mostrar el mejor dato en Y , y negativa cuando la tendencia es contraria: individuos con resultados buenos en X tienden a valores malos en Y (ϕ es igual al valor absoluto de τ_b : $\phi = |\tau_b|$). Para aprovechar sus cualidades, se requiere que τ_b se aplique sólo cuando las dos variables estudiadas sean verdaderamente ordinales.

En el capítulo I se demostró que todas las variables binarias pueden considerarse ordinales; aunque esto es cierto en general, para algunos casos resulta artificioso. Por ejemplo, si se examina la presencia de cierta enfermedad en dos regiones, puede argumentarse de modo natural que estar sano es mejor que tener enfermedad, pero es difícil afirmar que una región es indudablemente mejor que otra: puede serlo en algunos aspectos, mas no en todos. Con esto no se contradice lo demostrado en el capítulo I, en cuanto a que las variables binarias permiten un orden que no altera la medición del fenómeno, sólo se afirma que en algunos casos dicho orden no representa una cualidad de la variable en sí, sino una característica de la medición dicotómica.

Resulta entonces que existen variables dicotómicas que son de verdad ordinales, es decir que presentan un orden propio fuera de toda duda; al resto, aunque puede dársele un orden por ser dicotómicas, no les pertenece de modo intrínseco. Es en las primeras donde puede aplicarse τ_b con más provecho, en las segundas la indicada es ϕ .

Con esto se hace evidente que τ_b tiene la misma base conceptual que la correlación: los valores positivos se aplican cuando los datos en las dos variables se asocian directamente -tienden a ser parecidos- y los negativos se asignan a los casos en que la asociación es inversa, o sea que los sujetos con mejor resultado en una variable tienden a resultados bajos en la otra. De manera semejante a la correlación, cuanto más próxima a 1 o a -1 sea τ_b más fuerte será la asociación que expresa, al acercarse los números a cero señalarán una relación cada vez más débil. Para calcular τ_b se utiliza la fórmula que aparece a continuación:

$$\tau_b = (a d - b c) / \sqrt{(c_0 c_1 c_0 c_1)}$$

Observando que ϕ y τ_b son de igual magnitud en cualquier tabla 2x2, aunque τ_b puede ser de signo negativo, resulta que dos de las medidas de asociación más comunes para variables cualitativas toman el mismo valor en el caso de la tabla 2x2. Además,

ambas son iguales a la raíz cuadrada de ϕ^2 ($\sqrt{\phi^2}$) y la raíz de la C de Cramer (\sqrt{C}), por lo que puede aseverarse que todas estas medidas en esencia poseen la misma información; la única diferencia se encuentra entre las que presentan un concepto similar al del coeficiente de determinación, como se comentó en la sección de ϕ^2 , y las que son del tipo del coeficiente de correlación, como ϕ y T_c .

f) Coeficiente medio de contingencia (C_m)

El mismo Pearson propuso en 1904 otra medida de asociación basada en χ^2 , la cual se conoce como coeficiente medio de contingencia y se obtiene al calcular:

$$C_m = \sqrt{\left\{ \frac{\chi^2}{\chi^2 + N} \right\}}$$

Expresado en términos de ϕ^2 , el coeficiente medio de contingencia resulta:

$$C_m = \sqrt{\left\{ \frac{\phi^2}{\phi^2 + 1} \right\}}$$

El valor máximo de C_m depende del número de columnas y renglones que conforman la tabla de contingencia, por tal motivo no puede utilizarse para confrontar tablas de distintas dimensiones, pero si aplicarse cuando las tablas tienen igual tamaño. El hecho de que el máximo de C_m dependa de la magnitud de la tabla se refleja en que nunca rebasa 0.7071 en la tabla 2x2.

Aunque C_m puede aplicarse para comparar el grado de asociación entre dos tablas 2x2, ya que se cuenta con una amplia gama de estadísticas para medir la asociación, lo más indicado es elegir otra que sea más consistente en sus propiedades estadísticas. Al efecto vease el ejemplo III.3.

EJEMPLO III.3: Las medidas de asociación en tablas 2x2

Considerando la similitud entre las medidas de asociación descritas y la necesidad de compararlas, se presentan todas en un solo ejemplo. El cuadro III.2 contiene los datos que servirán de base para calcularlas y analizarlas. No se plantean condiciones específicas para este ejemplo a fin de poder aplicar todas las medidas y compararlas entre sí, aun las que están indicadas para distintas circunstancias.

Para mostrar el desarrollo de los cálculos necesarios, en seguida se muestra la obtención de las estadísticas de la tabla III.2d, para el resto de las tablas presentadas en el cuadro III.2 sólo se resumen los resultados en el cuadro III.3. Recuerdese que para estimar las medidas descriptivas basadas en χ^2 se emplea el valor de χ^2 sin la corrección de Yates, el cálculo adecuado aparece en el ejemplo.

$$\chi^2 = \frac{N (a d - b c)^2}{c_a c_b c_c c_d} = \frac{201 [(74)(98) - (15)(14)]^2}{(89)(112)(88)(113)} = 100.56$$

$$g^2 = \frac{\chi^2}{N} = G = P_{xy} = 100.56 / 201 = 0.5003$$

$$\theta = \sqrt{g^2} = |\tau_{bi}| = \sqrt{0.5003} = 0.7073$$

$$\tau_b = (a d - b c) / \sqrt{(c_a c_b c_c c_d)}$$

$$\tau_b = ((74)(98) - (15)(14)) / \sqrt{(89)(112)(88)(113)} = 0.7073$$

$$c_a = \sqrt{\left(\frac{\chi^2}{\chi^2 + N} \right)} = \sqrt{\left(\frac{100.56}{100.56 + 201} \right)} = 0.5775$$

La tabla III.2a señala la misma proporción en ambos sentidos; en los dos casos el resultado es el 64% del renglón, en tanto que de arriba hacia abajo cada dato es la mitad de la columna en que está. El valor de cero para χ^2 y para el resto de las medidas refleja tal situación (cuadro III.3) y permite verificar que se asigna cero para señalar la independencia o falta de

asociación entre las dos variables que forman la tabla 2x2 (en el Cap. I se describe el término independencia).

Cuadro III.2

Nueve tablas 2x2 para ejemplificar las medidas de asociación*

(a)	(b)	(c)																																													
<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>47 (64%)</td> <td>26</td> </tr> <tr> <th>-</th> <td>47 (64%)</td> <td>26</td> </tr> </tbody> </table>			Y				+	-	X	+	47 (64%)	26	-	47 (64%)	26	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>15 (100%)</td> <td>0</td> </tr> <tr> <th>-</th> <td>0 (0%)</td> <td>44</td> </tr> </tbody> </table>			Y				+	-	X	+	15 (100%)	0	-	0 (0%)	44	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>0 (0%)</td> <td>64</td> </tr> <tr> <th>-</th> <td>96 (100%)</td> <td>0</td> </tr> </tbody> </table>			Y				+	-	X	+	0 (0%)	64	-	96 (100%)	0
		Y																																													
		+	-																																												
X	+	47 (64%)	26																																												
	-	47 (64%)	26																																												
		Y																																													
		+	-																																												
X	+	15 (100%)	0																																												
	-	0 (0%)	44																																												
		Y																																													
		+	-																																												
X	+	0 (0%)	64																																												
	-	96 (100%)	0																																												
(d)	(e)	(f)																																													
<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>74 (83%)</td> <td>15</td> </tr> <tr> <th>-</th> <td>14 (13%)</td> <td>98</td> </tr> </tbody> </table>			Y				+	-	X	+	74 (83%)	15	-	14 (13%)	98	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>17 (23%)</td> <td>58</td> </tr> <tr> <th>-</th> <td>97 (81%)</td> <td>23</td> </tr> </tbody> </table>			Y				+	-	X	+	17 (23%)	58	-	97 (81%)	23	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>17 (63%)</td> <td>16</td> </tr> <tr> <th>-</th> <td>39 (91%)</td> <td>2</td> </tr> </tbody> </table>			Y				+	-	X	+	17 (63%)	16	-	39 (91%)	2
		Y																																													
		+	-																																												
X	+	74 (83%)	15																																												
	-	14 (13%)	98																																												
		Y																																													
		+	-																																												
X	+	17 (23%)	58																																												
	-	97 (81%)	23																																												
		Y																																													
		+	-																																												
X	+	17 (63%)	16																																												
	-	39 (91%)	2																																												
(g)	(h)	(i)																																													
<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>51 (41%)</td> <td>73</td> </tr> <tr> <th>-</th> <td>60 (73%)</td> <td>22</td> </tr> </tbody> </table>			Y				+	-	X	+	51 (41%)	73	-	60 (73%)	22	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>34 (29%)</td> <td>82</td> </tr> <tr> <th>-</th> <td>78 (54%)</td> <td>67</td> </tr> </tbody> </table>			Y				+	-	X	+	34 (29%)	82	-	78 (54%)	67	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>60 (74%)</td> <td>21</td> </tr> <tr> <th>-</th> <td>93 (70%)</td> <td>40</td> </tr> </tbody> </table>			Y				+	-	X	+	60 (74%)	21	-	93 (70%)	40
		Y																																													
		+	-																																												
X	+	51 (41%)	73																																												
	-	60 (73%)	22																																												
		Y																																													
		+	-																																												
X	+	34 (29%)	82																																												
	-	78 (54%)	67																																												
		Y																																													
		+	-																																												
X	+	60 (74%)	21																																												
	-	93 (70%)	40																																												

+ y - indican uno y otro valor de las variables X y Y que forman cada una de las tablas 2x2 presentadas.
 * Entre paréntesis se anota el porcentaje de renglón del primer dato, es decir (a / r₁) (100) y (c / r₂) (100).

Las tablas b y c descritas en el cuadro III.2 muestran las dos posibilidades de asociación absoluta: cuando sólo las frecuencias de una diagonal son distintas de cero. Para ambos casos las estadísticas toman su valor máximo, lo que hace patente la

Una limitación de C_m , que solo llega a 0.7071. Por su parte ϕ^2 , ϕ y τ_b si llegan a uno, que corresponde a lo que se acostumbra esperar como valor máximo de una estadística de asociación. En τ_b si se observa una diferencia entre ambas tablas: aunque en las dos τ_b es la misma magnitud, en la primera el valor es positivo y en la segunda negativo, señalando que en b los sujetos con mejor valor en X tienden a mostrar el mismo dato en Y -tendencia congruente-, en cambio en la tabla c se observa la tendencia inversa o incongruente: los que tienen el mejor resultado en X tienen el peor en Y .

Cuadro III.3

Medidas de asociación calculadas en nueve tablas 2x2

Tabla	Medida estadística				
	X^2	ϕ^2	ϕ	τ_b	C_m
d	0.00	0.0000	0.0000	0.0000	0.0000
e	59.00	1.0000	1.0000	1.0000	0.7071
f	160.00	1.0000	1.0000	-1.0000	0.7071
g	100.56	0.5003	0.7073	0.7073	0.5775
h	64.30	0.3298	0.5742	-0.5742	0.4980
i	18.89	0.2552	0.5052	-0.5052	0.4509
j	20.39	0.0990	0.3146	-0.3146	0.3001
k	15.77	0.0604	0.2458	-0.2458	0.2387
l	0.43	0.0020	0.0446	0.0446	0.0445

Nota: El valor de τ_b en que se basan las estadísticas, que no incluye la corrección de Yates.

Las tablas d, g y f del cuadro referido contienen casos de una asociación fuerte -d- y media -g y f-. En los resultados que del cuadro III.3 se observa que el coeficiente medio de contingencia (C_m) en los tres casos es menor que ϕ y que la diferencia es mayor conforme la asociación es más grande es, incluso en los resultados de las tres últimas tablas -g, h e i-, en las que la asociación es débil -g- o muy baja -h e i-, y llegan a ser prácticamente iguales. Puede concluirse que es preferirse ϕ cuando se quiere expresar la asociación presente en una tabla 2x2, inter-

pretándola en el mismo sentido que el valor absoluto del coeficiente de correlación, y que debe aplicarse ϕ^2 para indicar cuanto de la varianza existente en una de las variables de interés que está representada en la otra, entendiéndolo del mismo modo que el coeficiente de determinación.

En las tablas g , g , f , g , h e i , ϕ y τ_b son de signo contrario, destacando la cualidad de τ_b para distinguir entre la asociación directa y la negativa. Esta distinción sólo tiene sentido cuando las variables en examen son ordinales en sí mismas, por tal motivo sólo debe aplicarse τ_b para expresar la asociación cuando las dos variables que conforman la tabla 2×2 son claramente ordinales. Al presentar τ_b se explicaron las características de una variable dicotómica ordinal.

6. MEDIDAS DE PREDICCIÓN

El segundo grupo de estadísticas que se presentará tiene como propósito describir que tan útil es conocer una de las variables para predecir el resultado de la otra. En realidad es difícil estudiar el concepto mismo de asociación sin incluir el grado de habilidad de predicción; sin embargo, hay lugares en los que se puede tener una asociación significativa entre dos variables sin que haya una capacidad de predicción.

En múltiples circunstancias lo que se pretende es contar con una medición auxiliar para aproximarse al fenómeno de interés, por lo cual es importante valorar la capacidad predictiva de la variable auxiliar y no sólo la asociación entre ambas. En otras ocasiones es necesario averiguar si alguna manifestación que se mide en el presente puede emplearse para conocer con mayor precisión lo que ocurrirá en desempeños posteriores de un individuo o de un grupo. También en estos casos es útil aplicar las medidas estadísticas que evalúan la capacidad de predicción.

Existen varias estadísticas de predicción para la tabla 2×2 , aquí se describen las tres más útiles: la delta de Somers (δ) y las medidas lambda (λ) y gamma (Γ) de Goodman y Kruskal.

a) Delta de Somers (δ)

Desarrollada por Somers (1962), δ es una medida muy sencilla de interpretar. En el caso de la tabla 2×2 se reduce a la diferencia en la proporción de respuesta entre los dos niveles de la variable que se usará para predecir. Si se toman dos sujetos con diferentes niveles en una variable, δ indica la probabilidad

de que el individuo con mejor nivel en ésta también lo tenga en la otra variable.

Por ejemplo, si se toman dos embriones clasificados de acuerdo con su estado fisiológico, uno normal y otro no, y se quiere predecir la probabilidad de que se desarrollen al trasplantarlos, si $\xi = 0.3$ (véase el siguiente ejemplo), entonces el embrión clasificado como normal tendrá una probabilidad 30% mayor de implantarse correctamente que la de aquel que no parece estar en buen estado.

Ya que se refiere a "mejor y peor", para aplicar ξ en forma correcta se requiere que las variables sean propiamente ordinales. En la presentación de τ_b como medida de asociación se explican las características de una variable binaria estrictamente ordinal.

En forma análoga a lo que ocurre con la correlación, los valores posibles para ξ están entre -1 y 1. Cuando $\xi = 0$ la probabilidad de obtener cualquier resultado en la variable que se quiere predecir es igual para ambos sujetos. Los números negativos indican como más probable que quien obtuvo mejor resultado en la primera no lo tenga en la segunda variable; cuanto más cercano a -1, mayor será la probabilidad de que esto ocurra. Por su parte, los valores positivos se asocian a los casos en que la probabilidad de un buen resultado en la variable predicha favorece al que tiene ventaja en la variable base de la predicción. El cuadro III.4 muestra los datos organizados para valorar la capacidad de predicción.

Cuadro III.4

Esquema para valorar la capacidad predictiva de una variable ordinal sobre otra del mismo tipo

	Variable que se predice (Y)		Total	
	Mejor	Peor		
Variable en que se basa la predicción (X)	Mejor	a	b	C_0
	Peor	c	d	C_1
Total	C_0	C_1	N	

ξ es una medida asimétrica, es decir que indica la capacidad predictiva en un sentido específico, precisamente de una de las variables hacia la otra; así, en una tabla 2x2 pueden obtenerse

dos valores para S , uno de la variable X hacia la Y (S_{xy}) y otro para la capacidad de la segunda variable para predecir a la primera (S_{yx}). Para calcular S se aplican las siguientes fórmulas; la primera evalúa la capacidad predictiva de X hacia Y y la segunda lo hace en el sentido contrario, debe considerarse la primera como la expresión para calcular S .

$$S_{xy} = a / c_0 - c / c_1$$

$$S_{yx} = a / c_0 - b / c_1$$

La presentación del cálculo en los dos sentidos no indica que siempre sea útil valorar ambas capacidades predictivas; de hecho, en los casos en que se obtiene el mayor provecho es tan clara la distinción entre la variable base y la que se desea predecir que sólo se estima la capacidad predictiva en un sentido.

Para la estimación por intervalo de la S de Somers se usará el error estándar de la primera fórmula, S_{xy} , que corresponde a la manera más común de estimar S . El error estándar es:

$$S(S) = \sqrt{\frac{(a / c_0) (1 - [a / c_0])}{c_0} + \frac{(c / c_1) (1 - [c / c_1])}{c_1}}$$

El intervalo de confianza para la estimación de S , de acuerdo con el nivel de confianza que se quiera utilizar es:

Al 95% de confianza: $S - 1.960 S(S)$; $S + 1.960 S(S)$

Al 99% de confianza: $S - 2.576 S(S)$; $S + 2.576 S(S)$

EJEMPLO III.4: medición de la capacidad predictiva con S

En un estudio realizado con el fin de estimar la importancia de evaluar la apariencia morfológica en los embriones de bovino destinados al trasplante, se evaluó la calidad aparente de 48 embriones y se registró cuales continuaron la gestación dentro de la vaca nodriza. Los datos aparecen en el cuadro III.5.

Cuadro III.5

Relación entre la calidad del embrión y el éxito del trasplante*

		Resultado del trasplante		Total
		Implante	Falla	
Calidad aparente del embrión	Normal	14	16	30
	Mala	3	15	18
	Total	17	31	48

*Datos de Romo (1982)

$$S_{xy} = 14 / 30 - 3 / 18 = 0.3000$$

$$S_{yx} = 14 / 17 - 16 / 31 = 0.3074$$

Los resultados indican que determinando la calidad aparente de los embriones destinados al trasplante puede evitarse el 30% ($S = 0.3$) de los errores al predecir el resultado del trasplante. También puede interpretarse que los embriones de apariencia normal tienen una probabilidad de implantarse 30% mayor que los otros. En términos prácticos esto indica que la evaluación morfológica es útil para evitar el trasplante de aquellos embriones que no tengan apariencia normal, disminuyendo las fallas de implantación.

Aunque se calcularon los valores de S en ambos sentidos, el de la capacidad predictiva del resultado de la implantación sobre la calidad morfológica ($S = 0.3074$) es un mero ejercicio, ya que es evidente que nunca se intentará predecir en tal dirección. Con esto queda aún más claro lo explicado en el último párrafo de la presentación de S .

El intervalo de confianza al 95% para la estimación de S calculada en primer lugar se muestra a continuación. El error estándar del caso es:

$$S(\xi) = \sqrt{\left(\frac{(14 / 30) (1 - 14 / 30)}{30} + \frac{(3 / 18) (1 - 3 / 18)}{18} \right)}$$

$$S(\xi) = \sqrt{0.0160} = 0.1265$$

Intervalo de confianza al 95%: $0.3 \pm (1.960)(0.1265)$

Intervalo de confianza al 95%: 0.0520; 0.5480

Con lo cual se tiene un 95% de confianza de que un embrión clasificado como normal presenta una probabilidad de implantarse entre 5 y 55% mayor que la de uno considerado en proceso de degeneración.

b) Gamma de Goodman y Kruskal (Γ)

Cuando no se establece una distinción especial entre una variable y otra, por lo que no se considera variable explicativa a una y de respuesta a la otra, no tiene sentido aplicar una medida asimétrica como S , sino que se requiere una estadística simétrica que proporcione una evaluación general de la posibilidad de predecir en uno y otro sentido. Una de las medidas descriptivas más valiosas en este contexto es la gamma (Γ) desarrollada por Goodman y Kruskal (1954, 1959, 1972). Al igual que S , Γ se emplea cuando las dos variables en estudio son ordinales. Una ventaja interesante de Γ es que no le afectan los cambios en los totales marginales, siendo así una estadística muy estable.

La magnitud de Γ fluctúa entre -1 y 1 ; los valores positivos indican la probabilidad de que dos individuos que tuvieron diferentes resultados en una de las variables tengan rendimientos congruentes en la otra, en tanto que los negativos señalan la probabilidad de que los valores en la segunda variable sean incongruentes. Por congruentes se entiende que el individuo con un mejor resultado en la primera variable también lo tenga en la segunda, e incongruencia indica que el sujeto con mayor valor en la primera tenga peor registro en la variable que se considera en segundo término.

Un valor positivo de Γ se presenta, por ejemplo, cuando se espera que un individuo que apareció sano en un análisis sanitario también lo esté en un segundo escrutinio. Un ejemplo del valor negativo de Γ se da cuando los individuos resistentes a

cierta enfermedad tienden a ser de baja productividad y a su vez la mayor parte de los no resistentes logran una buena producción.

La correlación y Γ se interpretan de manera similar: los valores positivos se aplican si los resultados en las dos variables tienden a parecerse -congruentes- y los negativos cuando predomina la incongruencia; cuanto más cercanos a 1 o a -1, es más probable que ocurra la congruencia en los positivos o la incongruencia en los negativos. En la correlación, el cero se utiliza para señalar que no existe asociación, y en Γ indica que conocer una de las variables no mejora la capacidad de predecir la otra. Por tanto, $\Gamma = 1$ denota la seguridad de observar congruencia, porque al menos una de las celdillas b y c contiene frecuencia cero; $\Gamma = -1$ ocurre cuando a , d o ambas son iguales a cero, de modo que siempre sucederá la incongruencia. A continuación se indica la ecuación para calcular Γ :

$$\Gamma = (a \cdot d - b \cdot c) / (a \cdot d + b \cdot c)$$

El intervalo de confianza se obtiene, como en todos los casos, a partir del error estándar del estimador, que se calcula:

$$S(\Gamma) = \sqrt{0.5 (1 - \Gamma^2) \sqrt{(1/a + 1/b + 1/c + 1/d)}}$$

Intervalo de confianza al 95%: $\Gamma \pm 1.960 S(\Gamma)$

Intervalo de confianza al 99%: $\Gamma \pm 2.576 S(\Gamma)$

Considerando que Γ y τ_b son medidas simétricas y ambas se aplican cuando las dos variables analizadas son ordinales, parecería que ambos se reducen a un mismo concepto; sin embargo, existe una diferencia importante entre ellas: Γ alcanza un valor extremo (1 o -1) cuando existe una asociación completa, es decir que dos sujetos que difieren en una variable también serán distintos en la otra. Por su parte, τ_b sólo llega a un máximo cuando la asociación es absoluta, es decir que todos aquellos con determinado resultado en una variable tienen igual resultado en la otra. Si Γ es de 1 o -1, significa que al menos una variable permite predecir la otra sin error, mientras que un valor extremo de τ_b indica una asociación absoluta que permite predecir sin error en ambos sentidos.

La diferencia surge, precisamente, del distinto propósito de cada una: Γ procura la evaluación de la capacidad predictiva, y τ_b expresa la asociación desde un punto de vista general. Por lo

anterior Γ siempre es mayor que t_{α} , o por lo menos igual. En el ejemplo III.6 se muestra un caso que resalta la diferencia entre estas dos medidas estadísticas.

EJEMPLO III.5: aplicación de Γ para medir la capacidad predictiva

En una empresa avícola se quiere saber la proporción de gallinas que tienen coriza infecciosa, cuantas presentan bronquitis aviar y, al mismo tiempo, investigar la posible relación entre ambas afecciones. Resulta interesante la posibilidad de predecir la presencia de una enfermedad a partir de los datos de la otra y como no hay una que pueda considerarse respuesta de la otra, puede obtenerse provecho de la aplicación de Γ . Los datos del cuadro III.6 corresponden al caso.

Cuadro III.6

Frecuencia de bronquitis aviar y coriza infecciosa en gallinas

	Coriza infecciosa		Total
	Sí	No	
Bronquitis aviar	Sí	24	64
	No	485	496
Total	51	509	560

$$\Gamma = \frac{(40)(485) - (24)(11)}{(40)(485) + (24)(11)} = 0.9731$$

La estimación de Γ indica que hay una alta capacidad predictiva mutua entre las dos variables, de modo que puede emplearse la información de una para predecir la otra. Antes de emitir un comentario concluyente debe obtenerse el intervalo de confianza correspondiente. Primero se calcula el error estándar del estimador:

$$S(\Gamma) = \sqrt{0.5 (1 - 0.9731^2)} \sqrt{(1/40 + 1/24 + 1/11 + 1/485)}$$

$$S(\Gamma) = 0.1029$$

Intervalo de confianza al 99%: $0.9731 \pm (2.576) (0.1029)$

$$0.7081 ; 1.2382 \approx 0.7081 ; 1.0000$$

Como se ha hecho siempre que el intervalo de confianza sale del rango posible para el parámetro, el límite superior se truncó en uno. Puede afirmarse, con un 99% de confianza, que la habilidad predictiva entre una y otra enfermedad es de 71 a 100%, por lo que puede concluirse que, con una probabilidad de error tan baja como 0.01 —un error por cada 99 aciertos—, se espera que el valor de Γ se mantenga en niveles altos; por lo tanto los resultados acerca de una enfermedad tienen utilidad práctica para conocer sobre la otra afección. Con la información analizada en el ejemplo, se esperaría que una parvada con alguna de las dos enfermedades también presente la otra, ya que la mayor parte de las enfermas tiene las dos afecciones, de modo que una parvada que padece uno de los problemas estudiados debe examinarse en busca del otro.

EJEMPLO III.6: evaluación global de la capacidad predictiva con Γ

El cuadro III.7 contiene varios conjuntos de datos hipotéticos en los que se plantea el estudio de dos variables, considerando que el propósito de la investigación incluye evaluar el grado en que se predicen entre sí estas variables. En este caso no se describe un problema real para permitir la estimación de Γ y de τ_b sin provocar problemas de aplicabilidad de una u otra por el tipo de estudio; además, los distintos grupos de datos se formularon de modo que destaquen las cualidades de Γ y su comparación con τ_b . Ninguna de las dos variables se considera como explicativa o como respuesta de la otra, por tanto es un caso en el que están indicadas las estadísticas simétricas, como la Γ de Goodman y Kruskal y la τ_b de Kendall.

Al analizar los resultados de los distintos conjuntos contenidos en el cuadro III.7 se verifica lo expuesto acerca del valor relativo de Γ y τ_b , en cuanto a que Γ siempre es mayor o igual. Las tablas a, b, c y d presentan casos con un valor extremo para Γ , pero τ_b no lo tiene en las tablas c y d, haciendo patente la diferencia en el propósito de estas estadísticas. Ambas distinguen entre una relación directa y una inversa, por lo

cual tienen la misma magnitud en las dos primeras tablas, pero de signo positivo en **a** -relación congruente- y negativo en **b** para indicar la tendencia incongruente de esta tabla.

Cuadro III.7

Tablas 2x2 para mostrar el uso de la Γ de Goodman y Kruskal

(a)	(b)	(c)																																													
<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>34</td> <td>0</td> </tr> <tr> <th>-</th> <td>0</td> <td>21</td> </tr> </tbody> </table>			Y				+	-	X	+	34	0	-	0	21	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>0</td> <td>12</td> </tr> <tr> <th>-</th> <td>7</td> <td>0</td> </tr> </tbody> </table>			Y				+	-	X	+	0	12	-	7	0	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>27</td> <td>0</td> </tr> <tr> <th>-</th> <td>13</td> <td>46</td> </tr> </tbody> </table>			Y				+	-	X	+	27	0	-	13	46
		Y																																													
		+	-																																												
X	+	34	0																																												
	-	0	21																																												
		Y																																													
		+	-																																												
X	+	0	12																																												
	-	7	0																																												
		Y																																													
		+	-																																												
X	+	27	0																																												
	-	13	46																																												
(d)	(e)	(f)																																													
<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>0</td> <td>33</td> </tr> <tr> <th>-</th> <td>28</td> <td>50</td> </tr> </tbody> </table>			Y				+	-	X	+	0	33	-	28	50	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>15</td> <td>23</td> </tr> <tr> <th>-</th> <td>4</td> <td>72</td> </tr> </tbody> </table>			Y				+	-	X	+	15	23	-	4	72	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Y</th> </tr> <tr> <th colspan="2"></th> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <th rowspan="2">X</th> <th>+</th> <td>7</td> <td>58</td> </tr> <tr> <th>-</th> <td>91</td> <td>26</td> </tr> </tbody> </table>			Y				+	-	X	+	7	58	-	91	26
		Y																																													
		+	-																																												
X	+	0	33																																												
	-	28	50																																												
		Y																																													
		+	-																																												
X	+	15	23																																												
	-	4	72																																												
		Y																																													
		+	-																																												
X	+	7	58																																												
	-	91	26																																												

** Valores de Γ y τ_b para las tablas de la parte superior **

(a)	(b)	(c)
$\Gamma = 1.0000$	$\Gamma = -1.0000$	$\Gamma = 1.0000$
$\tau_b = 1.0000$	$\tau_b = -1.0000$	$\tau_b = 0.7254$
(d)	(e)	(f)
$\Gamma = -1.0000$	$\Gamma = 0.8430$	$\Gamma = -0.9333$
$\tau_b = -0.3778$	$\tau_b = 0.4327$	$\tau_b = -0.6441$

En las tablas **c** y **d** se presentan ejemplos de asociación completa, aunque no absoluta, por tanto Γ tiene su valor extremo pero no τ_b ; más aún, en **d** τ_b es relativamente pequeña. Las dos medidas descriptivas distinguen la relación directa de la inversa, por lo que tienen signo positivo en la tabla **c** y negativo en **d**. Las tablas **e** y **f** contienen casos de asociación intermedia, que tiene valores altos en Γ y medios para τ_b . En todo caso la

elección entre usar una u otra de las estadísticas comparadas en este ejemplo está determinada por las condiciones del problema y por el propósito de la investigación, pero no por la comparación de los valores alcanzados por una y otra estadística.

c) La lambda de Goodman y Kruskal (λ)

Si alguna de las variables analizadas no es propiamente ordinal, como se expuso en el apartado de τ_b , entonces no es posible aplicar λ ni Γ , pero puede utilizarse el concepto de proporción del error de predicción que es explicado por la variable empleada como base de la predicción. La evaluación numérica de la cantidad de error explicado constituye la λ de Goodman y Kruskal, la cual es la respuesta numérica a la pregunta: ¿que proporción de los errores que se cometerían al predecir sin información se eliminan al considerar la variable base de la predicción? A diferencia de Γ , λ sí resulta afectada por los totales marginales; esto debe tenerse en cuenta para decidir si puede aceptarse en un caso particular.

Existen dos versiones de λ , una que predice de una variable hacia la otra: λ_{12} , que es una medida asimétrica que actúa de manera similar a β ; aplicable cuando se define a una como variable de respuesta y a la otra como explicativa. La otra versión: λ -lambda simétrica-, analiza simultáneamente la capacidad de predicción en ambos sentidos, operando en forma parecida a las estadísticas simétricas como Γ y τ_b .

De acuerdo con los datos del cuadro III.8, para cualquier hembra se predeciría que es de buena habilidad productiva, ya que son las más frecuentes ($160 / 273 = 0.586$), así sólo se cometerían errores con todas las de baja capacidad de producción, que representan el 41.4% ($113 / 273 = 0.414$). Si al predecir se considera el lugar de origen de la cerda, para las del sur se esperaría un buen desempeño, al contrario de las del norte, por tanto se acertaría en 139 del sur y 42 del norte, errando sólo en el 33.7% de las hembras ($(71 + 21) / 273 = 0.337$). Si la predicción global conduce a 113 errores y la predicción tomando en cuenta el lugar de origen produce 92 errores ($71 + 21 = 92$), lo cual permite eliminar 21 ($113 - 92 = 21$) de los 113 errores originales, por lo que λ_{12} es 0.1858 ($21 / 113 = 0.1858$).

En el resultado anterior se puede constatar que λ se interpreta como la proporción de errores que se eliminan al incluir la variable base de la predicción. λ fluctúa de cero a uno, siendo cero cuando la capacidad predictiva es nula y uno si la predicción al considerar la variable predictiva está exenta de error; en el ejemplo, conocer el lugar de origen reduce en 18.58% los errores. La fórmula para obtener λ_{12} es:

$$f_2 = \frac{(\text{Número de errores sin } X) - (\text{Número de errores usando } X)}{(\text{Número de errores sin } X)}$$

Para uso práctico, el cálculo puede reducirse a:

$$f_2 = \frac{\text{Mín}(c_2, c_1) - \text{Mín}(a, b) - \text{Mín}(c, d)}{\text{Mín}(c_2, c_1)}$$

La instrucción "Mín" que aparece en la fórmula indica que se tome el menor de los dos valores que están dentro del paréntesis.

Cuadro III.8

Origen y habilidad productiva de un grupo de cerdas

	Habilidad productiva		Total
	Buena	Mala	
Lugar de origen de la hembra			
Sur	139	71	210
Norte	21	42	63
Total	160	113	273

La versión simétrica de f tiene un uso parecido al de f cuando se desea evaluar la capacidad mutua de predicción entre dos variables binarias y una o las dos no son propiamente ordinales, entonces debe calcularse la versión simétrica de f . La forma de obtenerla consiste en sumar los numeradores de f_1 en uno y otro sentido, sumar los denominadores y obtener la división de ellos. A continuación aparece la ecuación apropiada para estimar f :

$$f = \frac{\text{Mín}(c_2, c_1) - \text{Mín}(a, b) - \text{Mín}(c, d) + \text{Mín}(r_{21}, r_1) - \text{Mín}(a_1, c) - \text{Mín}(b_1, d)}{\text{Mín}(c_2, c_1) + \text{Mín}(r_{21}, r_1)}$$

Si se revisa la ecuación anterior se comprobará que es la suma de los numeradores y denominadores de f_a calculada en uno u otro sentido de la tabla 2x2. f toma un valor intermedio a las dos f_a .

EJEMPLO III.7: predicción entre variables sin orden propio

El desarrollo de este ejemplo se basa en los datos del cuadro III.B, que corresponden al origen de un grupo de cerdas y su habilidad reproductiva. La medición de f se muestra en seguida.

$$f_a = \frac{\text{Mín (160, 113)} - \text{Mín (139, 71)} - \text{Mín (21, 42)}}{\text{Mín (160, 113)}}$$

$$f_a = \frac{113 - 71 - 21}{113} = 0.1858$$

$$f = \frac{\text{Mín (160, 113)} - \text{Mín (139, 71)} - \text{Mín (21, 42)} + \text{Mín (210, 63)} - \text{Mín (139, 21)} - \text{Mín (71, 42)}}{\text{Mín (160, 113)} + \text{Mín (210, 63)}}$$

$$f = \frac{113 - 71 - 21 + 63 - 21 - 42}{113 + 63} = 0.1193$$

La capacidad de predicción del origen de la hembra sobre su habilidad productiva se estima en 19% ($f_a = 0.1858$), en tanto que el valor de f resultó cercano a 12% ($f = 0.1193$). Como se explicó, f está entre el valor de f_a en uno y otro sentido (para el ejemplo f_a obtenida de la habilidad reproductiva hacia el lugar de origen es cero).

EJEMPLO III.8: datos con distintos valores de λ

El cuadro III.9 contiene varios grupos de datos que representan los distintos casos que pueden ocurrir con λ , incluyendo una tabla en que la asociación es significativa y, sin embargo, no existe una capacidad de predicción entre las variables examinadas. El propósito de esos datos es mostrar los resultados que ocurren en λ_a ante distintos casos.

Cuadro III.9

Datos para ejemplificar la interpretación de λ

(a)	(b)	(c)																								
<table border="1"> <thead> <tr> <th colspan="2">\underline{Y}</th> </tr> <tr> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <td>+ 20</td> <td>2</td> </tr> <tr> <td>- 34</td> <td>28</td> </tr> </tbody> </table>	\underline{Y}		+	-	+ 20	2	- 34	28	<table border="1"> <thead> <tr> <th colspan="2">\underline{Y}</th> </tr> <tr> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <td>+ 2</td> <td>26</td> </tr> <tr> <td>- 23</td> <td>17</td> </tr> </tbody> </table>	\underline{Y}		+	-	+ 2	26	- 23	17	<table border="1"> <thead> <tr> <th colspan="2">\underline{Y}</th> </tr> <tr> <th>+</th> <th>-</th> </tr> </thead> <tbody> <tr> <td>+ 77</td> <td>89</td> </tr> <tr> <td>- 42</td> <td>28</td> </tr> </tbody> </table>	\underline{Y}		+	-	+ 77	89	- 42	28
\underline{Y}																										
+	-																									
+ 20	2																									
- 34	28																									
\underline{Y}																										
+	-																									
+ 2	26																									
- 23	17																									
\underline{Y}																										
+	-																									
+ 77	89																									
- 42	28																									
$\underline{X}^2 = 7.6979$ $\lambda_a = 0.0000$	$\underline{X}^2 = 15.8648$ $\lambda_a = 0.2400$	$\underline{X}^2 = 3.1265$ $\lambda_a = 0.1558$																								

*Como el propósito es mostrar si hay diferencias significativas, el valor de \underline{X}^2 incluye la corrección de Yates

La tabla a del cuadro III.9 contiene un conjunto de datos en los que existe una asociación significativa; para verificarlo basta ver el valor de \underline{X}^2 correspondiente (la prueba utilizada se explica en el Cap. IV), y sin embargo λ es cero, lo que indica una capacidad de predicción nula. Con esto se comprueba que una asociación significativa no garantiza la capacidad de predicción. De paso se muestra que ambas estadísticas evalúan un aspecto diferente de los datos.

Por su parte la tabla c presenta un caso en el que no hay asociación significativa y λ , aunque relativamente pequeña, es mayor que cero. En esos datos la variable \underline{X} permite reducir cerca del 16% ($\lambda = 0.1558$) de los errores que se cometerían al predecir sin usarla.

La tabla b presenta un caso intermedio, en el que existe una asociación significativa entre las dos variables ($\underline{X}^2 = 15.68$) y el valor de λ es un poco mayor que en la tabla c, que con-

tiene un caso en el que las dos estadísticas tienen valores medios.

d) Riesgo atribuible (R_a)

Una de las estadísticas predictivas más fáciles de interpretar es el riesgo atribuible (R_a), que indica la diferencia en la proporción de sujetos afectados entre los expuestos a un factor de riesgo, dicho de manera más exacta, indica la fracción de individuos cuya enfermedad es atribuible al factor analizado. El R_a es singularmente útil para valorar la importancia de cierto factor como contribuyente a la presencia de una enfermedad en estudio.

La estimación de R_a es sencilla: basta obtener la diferencia entre la proporción de individuos enfermos en el grupo que está expuesto al factor y la fracción de ellos en el grupo no expuesto.

$$R_a = \left(\frac{\text{\# de expuestos enfermos}}{\text{Total de expuestos}} \right) - \left(\frac{\text{\# de no expuestos enfermos}}{\text{Total de no expuestos}} \right)$$

Por la forma en que se calcula, R_a sólo es válido cuando se estima mediante estudios de causa a efecto; es decir, si primero se clasifica a los sujetos en expuestos y no expuestos y después se determina la presencia del problema. No debe estimarse R_a a partir de estudios de efecto-causa en los que se toma a un grupo de enfermos y otro de individuos que no sufren el problema, para determinar, con posterioridad, si estuvieron o no expuestos al factor.

Desde un punto de vista estadístico, R_a es equivalente a la S de Somers (véase la Sec. III.6.a) por lo cual la estimación de R_a , incluyendo los intervalos de confianza, se hace con los métodos ya explicados para S . Para obtener R_a y su error estándar se utilizan las siguientes fórmulas:

$$R_a = (a / r_0) - (c / r_1)$$

$$S(R_a) = \sqrt{\left(\frac{(a / r_0) (1 - [a / r_0])}{r_0} + \frac{(c / r_1) (1 - [c / r_1])}{r_1} \right)}$$

EJEMPLO III.9: uso del riesgo atribuible en un caso de lactancia

Para evaluar el riesgo de quitar las fuentes de calor para los lechones, en una granja se crió un lote de 100 lechones sin esas fuentes y otro con ellas. Se registró 15% de lechones con afecciones respiratorias entre los que no tenían calor artificial y sólo 7% con la fuente de calor durante la lactancia, por tanto el riesgo atribuible a la falta de una fuente de calor es de 8% ($0.15 - 0.07 = 0.08$), que debe entenderse como sigue: de los lechones que no tuvieron la fuente de calor, un 7% hubiese enfermado de cualquier manera, pero el 8% restante tuvo afecciones respiratorias a consecuencia de la falta del aditamento.

$$Ra = (15 / 100) - (7 / 100)$$

$$Ra = 0.15 - 0.07 = 0.08$$

$$S(Ra) = \sqrt{\left(\frac{(0.15)(1 - 0.15)}{15} + \frac{(0.07)(1 - 0.07)}{18} \right)}$$

$$S(Ra) = 0.1334$$

El intervalo al 95% de confianza: $0.08 \pm 1.960 (0.1334)$

$$0.08 - 1.960 (0.1334); 0.08 + 1.960 (0.1334)$$

$$-0.1815; 0.3415 = 0; 0.3415$$

Por lo que se espera que el riesgo atribuible a la falta de la fuente de calor esté entre 0 y 34%, es decir, que retirar las lámparas que proporcionan calor a los lechones incrementará la mortalidad entre 0 y 34%

e) Potencia atribuible (P_a)

Cuando el concepto de R_a se enfoca a la investigación de un tratamiento médico o de cualquier otro factor del que se espera un efecto favorable, manifestado a través de una disminución en la frecuencia del problema, se le llama potencia atribuible (P_a). La P_a se calcula a la inversa que el R_a , es decir, se obtiene mediante la diferencia de la fracción de individuos con el problema dentro de los no tratados menos la proporción correspondiente en los tratados.

$$P_a = (c / c_1) - (a / c_0)$$

Al igual que en todas las estadísticas anteriores, es conveniente obtener un intervalo de confianza antes de llegar a una conclusión definitiva, como la enunciada en el ejemplo anterior. De manera similar a lo explicado para R_a , para hacer la estimación de P_a se siguen los procedimientos descritos para \bar{x} (vease la Sección III.6.a). El error estándar del estimador de este parámetro es igual al presentado en la sección anterior para R_a , el cálculo aparece en el siguiente ejemplo.

EJEMPLO III.10: Aplicación de la potencia atribuible

Para estudiar el efecto de una monta con un macho criptorquideo previo al servicio fértil en cerdas, se dio monta a un grupo de 600 hembras y luego el servicio del seminal y a otro grupo de 500 sólo se le dio el servicio del seminal. Los resultados fueron: 93 y 88% de fertilidad, lo que implica 7 y 12% de cerdas vacías, respectivamente.

$$P_a = (60 / 500) - (42 / 600)$$

$$P_a = 0.12 - 0.07 = 0.05$$

Como $P_a = 0.05$, se estima que el procedimiento descrito incrementa la gestación en 5% en relación con el grupo testigo.

$$S(\underline{P}_a) = \sqrt{\left(\frac{(0.07)(1 - 0.07)}{15}\right) + \left(\frac{(0.12)(1 - 0.12)}{18}\right)}$$

$$S(\underline{P}_a) = 0.0179$$

El intervalo al 99% de confianza: $0.05 \pm 2.576 (0.0179)$

$$0.0034; 0.0966$$

La estimación por intervalo indica que se tiene un 99% de confianza en que el incremento en la tasa de gestación al aplicar el procedimiento descrito esté entre 0.34 y 9.66%

f) Proporción atribuible (Prop_a)

La proporción atribuible permite una evaluación similar al R_a, pero desde otro punto de vista sólo analiza los individuos enfermos, señalando que fracción lo está a consecuencia de su exposición al factor.

$$\underline{Prop}_a = \underline{R}_a / (\text{proporción de enfermos en los expuestos})$$

Cuando Prop_a se utiliza para valorar algún factor que tiene un efecto positivo, es decir cuando se aplica de modo similar a la P_a para evaluar un factor que disminuye la incidencia del problema, la Prop_a señala la fracción de individuos en los que se resolvió -o se evitó- el problema gracias al factor en análisis. Para tales casos la proporción atribuible puede expresarse como:

$$\underline{Prop}_a = \underline{P}_a / (\text{fracción de los testigos con el problema})$$

EJEMPLO III.11: cálculo de la proporción atribuible

Se estimará la proporción atribuible con los datos utilizados para los ejemplos de R_a y P_a . En el primer caso, se estimó que el riesgo de contraer la enfermedad respiratoria aumenta en 8% al retirar la fuente de calor de las lechonerías, para llegar a una incidencia de 15%, entonces la proporción atribuible es de 53%; es decir, que el 53% de los lechones enfermos lo están por haber sido retirado el calor artificial.

$$\text{Prop}_a = 0.08 / 0.15 = 0.5333$$

Para el ejemplo de P_a , la potencia atribuible al uso del macho criptorquideo se estimó en 5%, ya que al emplearlo se reducen las fallas reproductivas del 12 al 7%, por tanto la proporción atribuible al tratamiento descrito es de 42%; de donde se puede interpretar que al usar un macho criptorquideo antes de la monta fértil se evitan el 42% las fallas reproductivas.

$$\text{Prop}_a = 0.05 / 0.12 = 0.4167$$

7. MEDIDAS DE COMPARACION

Un grupo de estadísticas cuya importancia es evidente está constituido por las medidas que se enfocan a comparar la tasa de respuesta entre dos grupos; es fácil imaginar una situación en la cual se requiera comparar dos proporciones, así que seguramente es el grupo al que se le encontrarán más aplicaciones. La medida de comparación más importante para tablas 2x2 es la razón de momios, ya que es simple en su interpretación y tiene excelentes cualidades estadísticas. En este apartado se presentan medidas que son útiles en algunas situaciones específicas: el riesgo y la potencia atribuibles y la diferencia entre proporciones.

a) La razón de momios

Los momios son una manera de expresar las probabilidades, con la que se indica la posibilidad de que suceda un evento en relación con la probabilidad de otro. Por ejemplo, si la probabi-

lidad de que una cruda quede gestante es el cuádruple de la probabilidad de que permanezca vacía, el momio es de 4:1. Cuando sólo hay dos resultados posibles, en este caso 80 y 20%, el momio se calcula $0.80 / 0.20 = 4$ y $0.20 / 0.20 = 1$ para obtener el 4:1. El número que aparece como denominador es el que servirá como punto de referencia para expresar el momio, que aquí corresponde a la probabilidad de que no quede preñada.

Los momios se emplean con frecuencia para expresar las preferencias en apuestas, especialmente en el hipódromo. Otro uso común, en biología, son las expresiones de las leyes mendelianas; por ejemplo, se sabe que al aparear dos híbridos habrá una proporción de tres a uno de dominantes con respecto a los recesivos, esto se expresa con un momio de 3:1.

Cornfield (1951), propuso una medida para comparar los porcentajes en dos poblaciones, basada en calcular la razón entre el momio de respuesta en cada grupo, a la cual se le conoce como razón de momios. El desarrollo definitivo de esta medida como estadística de comparación se debe a Goodman y Kruskal (1954, 1959, 1963).

La razón de momios tiene una serie de características deseables que la hacen destacar como una estadística muy útil para tablas 2×2 ; entre otras, es válido obtenerla a partir de datos provenientes de muy distintos tipos de estudio. Si P_1 indica la proporción de respuesta en el primer grupo y P_2 en el otro, los momios respectivos son:

$$m_1 = P_1 / (1 - P_1)$$

$$m_2 = P_2 / (1 - P_2)$$

Y la razón de momios:

$$M = m_1 / m_2$$

Al revisar la expresión de la razón de momios (M), es fácil comprobar que $M = 1$ indica que ambos momios son iguales, esto es que la probabilidad de respuesta es igual en los dos grupos. Cuando M es mayor que uno, el momio del primer grupo es mayor, en cambio, los valores inferiores a uno corresponden a los casos en que el primer momio es inferior al del segundo grupo. La razón de momios puede tener cualquier valor.

Debido a que el momio es superior a uno cuando la respuesta es mayor en el primer grupo, y menor si la ventaja es del segundo

conjunto de datos, entonces puede variar de acuerdo con el orden de los renglones. Esto no representa una falla en la estadística, ya que su valor es simétrico, de tal modo que \bar{M} y $1 / \bar{M}$ expresan la misma comparación; si el orden de los renglones se intercambia el momio se modifica en esa relación: $1 / \bar{M}$. Calculando el logaritmo natural de la razón de momios, se genera una medida simétrica con respecto a cero que puede tener cualquier valor. Utilizando los logaritmos, la relación es $l_n(\bar{M}) = -l_n(1 / \bar{M})$.

Aplicando la nomenclatura definida para la tabla 2x2 y colocando los dos grupos a comparar en los renglones y la respuesta en las columnas, la razón de momios es:

$$\bar{m}_1 = \frac{a / r_0}{b / r_0} = \frac{a}{b}$$

$$\bar{m}_2 = \frac{c / r_1}{d / r_1} = \frac{c}{d}$$

$$\bar{M} = \frac{(a / b)}{(c / d)} = \frac{a d}{b c}$$

Por la expresión que aparece en la última fórmula, la razón de momios también recibe el nombre de razón de productos cruzados.

Cuando la respuesta de interés es rara, las tasas de respuesta en ambos grupos son cercanas a cero y, por tanto, el complemento de cada una es prácticamente uno; si $\bar{m}_1 \approx 0$, entonces $1 - \bar{m}_1 \approx 1$, lo mismo que si $\bar{m}_2 \approx 0$, resulta que $1 - \bar{m}_2 \approx 1$. En tal caso la razón de momios tiene un valor muy similar al del riesgo relativo, incluso algunos autores denominan riesgo relativo aproximado o riesgo relativo modificado a la razón de momios usada en este sentido.

Para calcular la variabilidad de estimación de la razón de momios se recurre a los logaritmos ya que al tomar el logaritmo de \bar{M} se logra una estadística simétrica con respecto a cero. El error estándar del logaritmo natural de la razón de momios es:

$$S(l_n \bar{M}) = \sqrt{S^2(l_n \bar{M})}$$

$$S^2(1_n \bar{M}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

A partir del error estándar se determina el intervalo de confianza del logaritmo de la razón de momios; después se calcula el antilogaritmo o exponencial de los límites de ese intervalo y el resultado es el intervalo de confianza para la razón de momios. Denotando con \underline{l} y \underline{g} los límites inferior y superior del intervalo de confianza para el logaritmo natural de la razón de momios, la estimación se hace mediante el siguiente procedimiento:

$$\underline{l} = 1_n \bar{M} - (Z S(1_n \bar{M}))$$

$$\underline{g} = 1_n \bar{M} + (Z S(1_n \bar{M}))$$

Z es 1.960 para el intervalo al 95% de confianza y 2.576 para el de 99%.

Intervalo de confianza para \bar{M} : \underline{g}^a ; $\underline{g}^b = 2.718 28^{\underline{l}}$; $2.718 28^{\underline{g}}$

Un problema importante en relación con la razón de momios y su error estándar es que, al existir alguna celda con frecuencia cero -en la razón de momios \underline{c} o \underline{d} y en la variancia cualquier celda- el cálculo conduce a una división entre cero, dando un resultado indefinido. Para resolver este problema se ha propuesto sumar 0.5 a cada frecuencia durante el cálculo de \bar{M} y de la variabilidad de estimación.

$$\bar{M} = \frac{(a + 0.5) (c + 0.5)}{(b + 0.5) (d + 0.5)}$$

$$S^2(1_n \bar{M}) = \frac{1}{(a + 0.5)} + \frac{1}{(b + 0.5)} + \frac{1}{(c + 0.5)} + \frac{1}{(d + 0.5)}$$

Las dos ecuaciones anteriores eliminan las dificultades inherentes a las celdas con frecuencia cero. Cuando se decide em-

plear estos últimos estimadores de \bar{Y} y su varianza, se siguen los mismos procedimientos de estimación ya explicados para la razón de monios; la única diferencia es la introducida por la modificación a las fórmulas del estimador \bar{Y} y de la varianza.

EJEMPLO III.12: cálculo de la razón de monios

Con el propósito de valorar el efecto producido por la monta de un macho infértil previa al servicio del semental sobre la tasa de gestación en cerdas, se tomaron 80 hembras. A 40 de ellas se les dio monta con un macho criollo-quideco y a continuación el servicio del semental; el resto se usó como lote testigo, por lo que sólo se le dio servicio. El cuadro III.10 contiene los datos del caso, considerando la fertilidad del servicio al parto.

Cuadro III.10

Fertilidad registrada en dos grupos de cerdas*

	Resultado		Total
	Parturidó	Vacía	
Monta previa al servicio fértil	Si 37	3	40
	No 3	5	40
Total	40	8	80

*Datos de Medina (1986)

La razón de monios estimada para este caso es de 1.7619, que indica un monio de gestación 76.19% mayor entre las cerdas tratadas que en el grupo testigo. Los datos incluyen un monio de gestación en las tratadas de 12.31 y un monio de 7.021 para el resto de las cerdas.

$$\bar{m}_a = \frac{1}{1} \cdot \frac{37}{3} = 12.33$$

$$\bar{m} = \frac{e}{d} = \frac{35}{5} = 7.00$$

$$\bar{M} = \frac{a d}{b e} = \frac{(37) (5)}{(3) (35)} = 1.7619$$

$$1_n \bar{M} = 1_n 1.7619 = 0.5664$$

$$S^2(1_n \bar{M}) = \frac{1}{37} + \frac{1}{3} + \frac{1}{35} + \frac{1}{5} = 0.5889$$

$$S(1_n \bar{M}) = \sqrt{S^2(1_n \bar{M})} = \sqrt{0.5889} = 0.7674$$

El intervalo al 95% de confianza para la razón de monios, conforme los resultados de la investigación, es de:

$$\underline{i} = 0.5664 - (1.960) (0.7674) = -0.9377$$

$$\underline{s} = 0.5664 + (1.960) (0.7674) = 2.0705$$

Intervalo de confianza para \bar{M} : $e^{-0.9377}$; $e^{2.0705} = 0.3915$; 7.9291

Finalmente, los datos del estudio indican que la razón de monios observada en la muestra es de 1.7619, y que se tiene un 95% de confianza en que el valor poblacional esté entre 0.39 y 7. Dado que una \bar{M} superior a 1 señala un monio mayor en el primer grupo y los valores menores a 1 indican una ventaja del segundo, esto significa que en la población de referencia bien puede ser que el monio esté a favor de uno -valores entre 0.39 y 1- o del otro grupo -valores de 1 a 7-. Si los datos provienen de una muestra lo bastante grande como para aportar información precisa, la conclusión debe ser que no hay ventaja de un grupo sobre otro, pero si la muestra es pobre, debe revisarse el problema con mayor número de datos. La amplitud del intervalo de confianza permite investigar si la estimación es o no precisa.

b) Riesgo relativo (R_r)

Para comparar la frecuencia con que se presenta un problema en dos grupos puede calcularse la razón entre ambas tasas; a tal relación se le conoce como riesgo relativo (R_r). En este sentido el riesgo relativo es semejante a la razón de momios, sólo que se basa en los porcentajes de respuesta y no en los momios

$$P_1 = a / c_0$$

$$P_2 = c / c_1$$

$$R_r = P_1 / P_2$$

El R_r es una de las estadísticas de comparación mejor comprendidas por los profesionales de la medicina, quienes la utilizan para comparar la frecuencia de una enfermedad en dos grupos, generalmente uno expuesto a un factor que se supone precursor del problema y el otro libre de tal factor.

Los valores de riesgo relativo muestran la frecuencia la afección en el grupo expuesto al factor considerado de riesgo en función de la tasa observada en el otro grupo. Si $R_r = 1$ significa que ambas tasas son iguales, valores entre cero y uno señalan que el problema ocurre con menor frecuencia en el primer grupo, y los R_r mayores que uno están asociados a los casos en que el problema se presenta más en el grupo expuesto. Un factor que en verdad represente riesgo, tendrá un resultado mayor a uno en R_r : cuanto más grande sea R_r mayor será el riesgo al estar expuesto a dicho factor.

Para evaluar el riesgo relativo se requiere tomar una muestra de los individuos que están expuestos al factor y otra de los no expuestos, para determinar posteriormente si presentan o no el problema, lo que exige un esquema del tipo causa-efecto. No es válido calcular el R_r a partir de los datos generados en una investigación de efecto a causa, en la que se registre un grupo de sujetos con el problema, otro sin él, y entonces se averigüe si están o estuvieron expuestos al factor de interés.

Si la tasa de respuesta es muy baja, se puede considerar que la razón de momios se aproxima al riesgo relativo y, debido a que la razón de momios sí puede calcularse en investigaciones de efecto a causa, representa una vía para estimar R_r en este tipo de estudios, siempre que la respuesta sea rara.

EJEMPLO III.13: Estimación del riesgo relativo

En un rancho se desea estimar el riesgo de fertilizar los potreros con urea, para lo cual se pastorea un grupo de novillos en un área fertilizada y otro lote se alienta en pastos sin fertilizar. Los datos se muestran en el cuadro III.11.

Cuadro III.11

Intoxicación en bovinos al aplicar urea al pasto

		Intoxicación		Total
		Sí	No	
Potrero en que se encontraban pastando	Fertilizado	7	33	40
	Sin fertilizar	9	45	54
	Total	16	78	94

$$R_r = (7 / 40) / (9 / 54) = 1.05,$$

De modo que el riesgo relativo de que el ganado padezca una intoxicación a consecuencia de fertilizar los potreros con urea se estima en 1.05 veces el riesgo presente sin la fertilización referida. Debido a que casi no aumenta el riesgo, la aplicación de este producto podría considerarse como una opción viable para promover el crecimiento de los pastos.

c) Potencia relativa (P_r)

Existe un procedimiento de valoración, similar al riesgo relativo, que se utiliza para estimar la efectividad de un factor aplicado para disminuir la frecuencia de algún problema; este procedimiento se denomina potencia relativa (P_r), su cálculo se efectúa mediante:

$$Pr = 1 - (P_2 / P_1)$$

$$P_1 = a / c_0$$

$$P_2 = e / c_1$$

En donde P_2 indica la frecuencia del problema entre los individuos no expuestos al factor en estudio y P_1 la tasa correspondiente en el grupo que sí lo estuvo. Pr puede tomar cualquier valor; los negativos se presentan cuando el factor en estudio no evita el problema, sino que lo promueve, y los valores positivos indican el grado en que el factor reduce la frecuencia del problema.

EJEMPLO III.14: uso e interpretación de la potencia relativa

A fin de buscar métodos para mejorar el resultado de la inseminación artificial en ganado cebú, se estudió el resultado obtenido al inseminar vacas con estro natural y estro inducido. Los datos se presentan en el cuadro III.12.

Cuadro III.12

Resultados al inseminar vacas con estro natural o inducido*

		Gestante		Total
		Sí	No	
Tipo de estro cuando se inseminó	Inducido	11	48	59
	Natural	17	40	57
	Total	28	88	116

*Datos de Landivar (1982)

$$Pr = 1 - (P_2 / P_1) = 1 - [(17 / 57) / (11 / 59)] = -0.5997$$

Los resultados del estudio indican que la inducción del estrógeno en las vacas cebú no ayudó a mejorar la eficacia reproductiva, sino que la disminuyó en casi 60%.

B. MEDIDAS DE CONCORDANCIA

Un aspecto valioso para interpretar los resultados en muchos estudios, que rara vez se evalúa, es la concordancia entre dos variables. Si se utilizan dos procedimientos para clasificar una serie de sujetos, ¿qué tanto coinciden ambas clasificaciones? Las medidas de concordancia responden a esta pregunta, por lo que resultan interesantes para comparar, por ejemplo, qué tan concordante es el diagnóstico de gestación hecho por dos veterinarios en un grupo de vacas, o que tanto concuerdan los resultados de dos pruebas de laboratorio o de dos métodos de inspección sanitaria; en fin, podrían citarse muchos casos en los que las medidas de concordancia proporcionan un medio importante para interpretar los resultados de una investigación.

Las medidas de concordancia son útiles para saber si dos clasificaciones tratan de igual manera los casos similares, que es una forma peculiar de asociación. Al comparar dos métodos de inspección sanitaria en rastros no sólo interesa averiguar si ambos detectan la misma proporción de casos, sino también si son los mismos. De ser así, pueden tomarse como métodos equivalentes y considerarlos opciones alternativas; pero si la concordancia es pobre deben aplicarse como métodos complementarios entre sí, ya que cada uno detecta diferentes casos.

Aunque la medida de concordancia más simple es la proporción de casos con igual clasificación en ambos procedimientos, su uso no es recomendable porque no refleja exclusivamente la concordancia, sino que está influida por los totales marginales.

a) Kappa de Cohen (K)

Cohen (1960) propuso una medida de concordancia llamada K de Cohen, basada en comparar la concordancia observada en un conjunto de datos con respecto a la que podría ocurrir por simple azar. Aunque la K de Cohen también es útil cuando la clasificación tiene más de dos niveles, aquí sólo se aplica al caso 2x2.

Si se denota la concordancia observada como C_{obs} y la esperada por azar como C_{azar} , la K de Cohen es:

$$K = \frac{C_{\text{obs}} - C_{\text{azar}}}{1 - C_{\text{azar}}}$$

Cuando K es cero indica que la concordancia observada coincide con la que ocurriría por simple azar, los valores positivos señalan mayor concordancia que la esperada por azar, siendo uno el valor asignado a la concordancia perfecta. Cuando K toma un valor negativo señala discordancia, llegando -en la tabla 2x2- hasta menos uno para señalar la discordancia total entre las dos clasificaciones. El cálculo de K en la tabla 2x2 es muy simple:

$$K = \frac{N(a+d) - B}{N^2 - B}$$

Donde: $u = C_o C_e$

$$t = C_x E_x$$

$$B = u + t$$

El error estándar de K es la raíz cuadrada de la varianza de estimación, de modo que se obtiene:

$$S(K) = \sqrt{\left(\frac{B(N+R/N) - u(C_o + C_e) - t(C_x + E_x)}{B^2 + N^2(N^2 - 2B)} \right)}$$

Con los datos anteriores es sencillo obtener un intervalo de confianza para el valor de K :

Al 95%: $K - 1.960 S(K)$; $K + 1.960 S(K)$

Al 99%: $K - 2.576 S(K)$; $K + 2.576 S(K)$

Para examinar en detalle la interpretación de K , se analizarán las tablas 2x2 contenidas en el cuadro III.13.

Cuadro III.13

Tablas 2x2 con distintos grados de concordancia*

(a)	Y		Total	(b)	Y		Total
	+	-			+	-	
X	0	50	50	X	0	75	75
	50	0	50		25	0	25
Total	50	50	100	Total	25	75	100

(c)	Y		Total	(d)	Y		Total
	+	-			+	-	
X	22	13	35	X	37	12	49
	12	18	30		9	42	51
Total	34	31	65	Total	46	54	100

Estadísticas de concordancia para las tablas de la parte superior

(a)

$K = -1$ (-1; - .80**)
Coincidencia = 0.0

(b)

$K = -0.6$ (-0.72; -0.48)
Coincidencia = 0.0

(c)

$K = 0.23$ (-0.02; 0.47)
Coincidencia = 0.62

(d)

$K = 0.58$ (0.38; 0.78)
Coincidencia = 0.79

* Entre paréntesis se muestra el intervalo de confianza para K .

** En a el límite inferior del intervalo se truncó en menos uno, ya que es el valor mínimo de K .

La tabla 2x2 que aparece en el cuadro III.13 a muestra la discordancia total entre las dos clasificaciones, $K = -1$; y aunque en III.13 b parece repetirse el mismo caso K sólo alcanza -0.6 , lo que puede interpretarse en el sentido de que ambas clasificaciones presentan tendencias que las llevan a discordar, una clasificación tiende al resultado positivo, mientras que la otra tiende a la clasificación negativa. De hecho, K sólo alcanza su valor mínimo - menos uno- cuando los totales marginales son iguales.

EJEMPLO III.15: estimación de la concordancia

Considerando que una de las dificultades para lograr éxito en programas reproductivos en ganado cebú radica en los errores de clasificación de las vacas, se decidió comparar el resultado de la evaluación reproductiva mediante la palpación rectal contra la cantidad de progesterona presente en la sangre al momento de palpar. Al efecto se consideró que niveles de progesterona mayores a 0.5 ng corresponden a un cuerpo lúteo activo.

Los datos del cuadro III.14, indican un 85% de coincidencia entre la palpación rectal y la medición de la progesterona circulante, que aparenta un alto grado de concordancia; pero al obtener K resulta que sólo es del 28% ($K = 0.28$), mucho más baja que la proporción de coincidencia. Puede verse que la tasa de coincidencia no es una buena medida de la concordancia, ya que está muy influenciada por los totales marginales; en el ejemplo, cerca del 90% de las vacas tienen ovarios activos, así que las dos clasificaciones no pueden menos que coincidir en muchos de los resultados que señalan actividad ovárica.

$$y = 42 (41) = 1\ 722$$

$$t = 6 (5) = 30$$

$$R = 1\ 722 + 30 = 1\ 752$$

$$K = \frac{47 (38 + 2) - 1\ 752}{47^2 - 1\ 752} = 0.280\ 09$$

$$S^2(K) = \frac{1752(47 + 1752/47) - 1722(42 + 41) - 30(5 + 6)}{1752^2 + 47^2(47) - (2)(1752)} = 0.02105$$

$$S(K) = \sqrt{0.02105} = 0.14509$$

Calculando el intervalo de confianza al 95% para el valor de K :

$$K - 1.960(0.14509); K + 1.960(0.14509)$$

$$-0.00438; 0.56438$$

Cuadro III.14

Evaluación reproductiva en vacas cebú mediante dos sistemas*

		Progesterona circulante		Total
		> 0.5 ng (Cuerpo lúteo)	< 0.5 ng (Sin c. lúteo)	
Palpación rectal	Con cuerpo lúteo	38	4	42
	Sin cuerpo lúteo	3	2	5
Total		41	6	47

* Datos de Rubio (1985)

El intervalo cubre un rango desde una ligera discordancia -señalada por el signo negativo- hasta un poco más de la mitad de concordancia. Estos límites tan amplios y el hecho de que el cero queda incluido en el intervalo hace pensar que la concordancia detectada en la muestra no es significativa. Para verificarlo debe aplicarse la prueba de hipótesis específica, como se explica en la sección IV.9.

Al confirmarse que no existe una concordancia significativa puede señalarse que efectivamente las fallas en palpación rectal

inciden en el resultado de esta clase de programas, conclusión que no podría lograrse a partir de otras estadísticas.

EJEMPLO III.14: concordancia entre dos métodos de inspección

Al valorar la confiabilidad de dos métodos de inspección para detectar cisticercos en canales de cerdo, se comparó la inspección basada en cortes en el brazuelo -método en uso- con respecto a la búsqueda de los cisticercos en el músculo macetero; los datos se muestran en el cuadro III.15.

Cuadro III.15

Concordancia entre dos métodos de inspección sanitaria*

	Brazuelo		Total
	+	-	
Macetero	+	2	15
	-	8538	8544
Total	19	8540	8559

* Datos de Vargas (1984)

En este ejemplo es aún más notorio que en el anterior el hecho de que el porcentaje de coincidencia entre las dos clasificaciones es poco útil, ya que al ser en extremo rara la infección por cisticercos, las dos se ven obligadas a coincidir, hasta en un 99.9%. De nuevo χ^2 brinda una descripción más útil:

$$\chi^2 = 15 (19) = 285$$

$$t = 8544 (8540) = 72965760$$

$$R = 285 + 72\ 965\ 760 = 72\ 966\ 045$$

$$K = \frac{8\ 559\ (13 + 8\ 538) - 72\ 965\ 760}{8\ 559^2 - 72\ 965\ 760} = 0.760076$$

K indica que la concordancia es del 76%, que puede considerarse un nivel razonablemente alto; sin embargo, conviene ver los límites del intervalo de confianza antes de emitir una opinión, trabajando al 99% de confianza:

$$S^2(K) = \frac{R(R\ 559 + R/8\ 559) - u(15 + 19) - t(8\ 544 + 8\ 540)}{R^2 + 8\ 559^2 (8\ 559^2 - 28)} = 0.00012$$

$$S(K) = \sqrt{0.00012} = 0.01073$$

$$0.76 - 2.576 (0.01073); \quad 0.76 + 2.576 (0.01073) \quad (2)(73)$$

$$0.73236; \quad 0.78764$$

Con un 99% de confianza, puede afirmarse que el valor poblacional de K fluctuará entre 73 y 79% en cerdos con condiciones similares de inspección, es decir, que ambos métodos coinciden en 73 a 79%, que es una buena concordancia, por lo que pueden considerarse como métodos alternativos.

9. VALIDACION DE UNA PRUEBA DIAGNOSTICA

Una de las inquietudes más comunes en medicina es la evaluación de las pruebas de laboratorio. Para el médico es indispensable conocer qué importancia debe dar al resultado de cada examen de laboratorio, misma que varía en función de la calidad de la prueba. Si se aplica un análisis que nunca falla, el médico basará sus conclusiones en el resultado del laboratorio; pero, si utiliza un procedimiento que presenta fallas frecuentes, sólo en

Plenará los resultados como una indicación más del estado del paciente y no como base de su diagnóstico final.

Es importante destacar que los conceptos contenidos en las siguientes secciones no sólo pueden aplicarse a pruebas de laboratorio enfocadas al diagnóstico clínico; de hecho, son útiles para valorar cualquier método cuyo objetivo sea conocer el estado que presenta un sujeto en cuanto a una característica específica. Es decir, sirven para evaluar cualquier procedimiento diagnóstico cuyos resultados sean dicotómicos -positivo o negativo-; lo mismo puede evaluarse un procedimiento para el diagnóstico de rinitis atrófica en cerdos que uno para predecir si una cerda será o no buena reproductora -diagnosticar su habilidad reproductiva- o un sistema para determinar si una vaca está en calor, lo importante es que todos son métodos que pretenden diagnosticar un hecho expresado con una dicotomía. Incluso podrían aplicarse a procedimientos destinados a detectar fallas administrativas, siempre y cuando lleguen a resultados dicotómicos: se concluye que el problema de interés está o no presente.

a) Calidad general de una prueba

Para valorar en forma global la eficacia de una prueba se emplean dos medidas específicas: la sensibilidad y la especificidad. La sensibilidad señala la proporción de enfermos que son detectados por la prueba; a su vez, la especificidad indica el porcentaje de sanos para los que la prueba arroja un resultado negativo. Estas dos medidas se refieren a las proporciones de resultados correctos, la primera es la tasa de aciertos entre los enfermos y la segunda entre los sanos. En ocasiones, los parámetros descritos se designan como sensibilidad y especificidad nosológicas.

En este y en los siguientes apartados la presentación se hará en el contexto de una prueba clínica, por lo que se usarán términos como "enfermo" y "sano"; esto no resta generalidad a los conceptos que se expongan, ni invalida la explicación de la sección anterior acerca de su aplicabilidad a cualquier prueba que tienda al diagnóstico.

b) Estimación de la sensibilidad y la especificidad

Para estimar los indicadores de la calidad general de una prueba diagnóstica se tiene el problema de conocer con exactitud cuáles casos presentan la enfermedad y cuáles no; casi nunca es posible determinarlo, así que se toma como punto de referencia algún procedimiento bastante preciso o, por lo menos, de calidad conocida.

Tal es el caso cuando un médico desea valorar cierto método de campo para la detección temprana de mastitis y compara los resultados con una prueba de laboratorio de calidad reconocida. A los estimadores así obtenidos se les conoce como sensibilidad y especificidad relativos, porque indican la calidad de la prueba de interés en relación con los resultados de otra y no en forma absoluta. La validez de la sensibilidad relativa (S_r) y de la especificidad relativa (E_r) dependen de la calidad de la prueba utilizada como punto de referencia.

Es importante estar conscientes de que, en sentido estricto, la estimación de la especificidad requiere que se aplique la prueba a individuos sanos, lo que es poco ético para aquellos exámenes clínicos que pueden provocar algún daño al paciente o poco adecuado económicamente cuando se trata de pruebas costosas, e incluso pueden presentarse ambos inconvenientes a la vez. Para evitarlo se ha propuesto que la valoración del procedimiento de diagnóstico se haga con sujetos sospechosos de padecer la enfermedad y no en una muestra aleatoria de la población de referencia.

Incluso es posible que una muestra de individuos sospechosos de la enfermedad represente mejor a la población susceptible de ser sometida al análisis; ya que, si se decide poner en práctica la prueba, se aplicará, precisamente, a los sujetos que tengan indicios de presentar el problema.

Cuando es factible conocer el estado real del sujeto se pueden estimar la sensibilidad y la especificidad reales, es decir, valorar la calidad verdadera del método y no sólo en relación con otra prueba. Esto sucede cuando se dispone de un procedimiento exento de error, como sería el caso al comparar un método diagnóstico con el resultado de la necropsia; no puede utilizarse la necropsia para una evaluación en vivo, pero sí para valorar la prueba de interés. También es posible obtener la calidad real de la prueba cuando el estado del sujeto se hace evidente con el tiempo; por ejemplo, al evaluar un método para diagnóstico temprano de gestación, basta aplicarlo y esperar la presentación del parto, lo que servirá para determinar su calidad. Lo mismo puede decirse de los procedimientos destinados a detectar ciertas enfermedades en sus fases iniciales.

Para estimar la sensibilidad y especificidad reales se sigue el mismo procedimiento que se describirá para los valores relativos, sólo que la prueba de referencia está libre de error y, en consecuencia, los estimadores de calidad son más confiables. El cambio se refleja en el hecho de que no se aplica el adjetivo de "relativos" a los estimadores resultantes, por lo que se elimina el subíndice "r", anotándose: sensibilidad, S y especificidad, E .

Aunque en una parte importante de los casos estos dos parámetros describen la calidad de la prueba como una cualidad intrínseca de la misma, en ocasiones los resultados dependen no sólo de la prueba, sino también de la población en que se hace el estudio. Cuando al análisis es sensible a otras enfermedades, similares a la de interés, como sucede en las "reacciones cruzadas"

de ciertas pruebas inmunológicas, la valoración final dependerá de la frecuencia de individuos que tiene esa otra, u otras, afecciones. Por tal motivo, siempre que exista la posibilidad de que la prueba responda a otras enfermedades, debe considerarse con cuidado la población en que se efectuó la estimación de los valores de calidad diagnóstica

El cuadro III.16 presenta el esquema de los datos utilizados en la estimación de sensibilidad y especificidad. Los renglones corresponden a los resultados de la prueba que se desea evaluar y las columnas a la de referencia. Las frecuencias anotadas en a y en d indican los aciertos, mientras que b contiene los falsos positivos y c los falsos negativos.

Cuadro III.16

Orden de los datos para estimar la sensibilidad y especificidad

	Prueba de referencia		Total	
	+	-		
Prueba en evaluación	+	<u>a</u>	<u>b</u>	<u>C₀</u>
	-	<u>c</u>	<u>d</u>	<u>C₁</u>
Total	<u>C₀</u>	<u>C₁</u>	<u>N</u>	

Considerando la organización del cuadro III.16, la sensibilidad y la especificidad se estiman como se indica en las siguientes ecuaciones. Si la prueba de referencia esta exenta de error, entonces los estimadores obtenidos corresponderán a la especificidad y sensibilidad reales (S y E), si no serán relativos (S_r y E_r).

$$\underline{S} = \text{Sensibilidad} = \frac{a}{C_0}$$

$$\underline{E} = \text{Especificidad} = \frac{d}{C_1}$$

Como ambos estimadores son un caso especial de la estimación de porcentajes en una población -sensibilidad: proporción de aciertos en una población de enfermos, y especificidad: tasa de aciertos entre los sanos-, los intervalos de confianza correspon-

dientes se calculan de acuerdo con los métodos descritos para una proporción:

$$\frac{a}{c_0} \pm \text{factor de confiabilidad} \sqrt{\left(\frac{(a/c_0) (c/c_0)}{c_0} \right)}$$

$$\frac{d}{c_1} \pm \text{factor de confiabilidad} \sqrt{\left(\frac{(d/c_1) (b/c_1)}{c_1} \right)}$$

El factor de confiabilidad adecuado para un intervalo de confianza al 95% es 1.960, para los intervalos al 99% de confianza deberá aplicarse un factor de 2.576.

EJEMPLD III.17: evaluación de la calidad diagnóstica

Con el propósito de valorar un procedimiento para el diagnóstico temprano de gestación en cerdas, basado en la medición de la resistencia eléctrica de las secreciones vaginales, se midió la resistencia en 40 hembras y se predijo su estado reproductivo; después se registraron los partos de ese grupo de cerdas. Los resultados aparecen en el cuadro III.17.

Cuadro III.17

Diagnóstico de preñez en cerdas a partir de la resistencia eléctrica*

	Ocurrió el parto		Total	
	Sí	No		
Diagnóstico basado en la resistencia eléctrica	Gest.	28	1	29
	Vacía	7	4	11
	Total	35	5	40

*Datos de López (1986)

Cap. III. Medidas Estadísticas

Los resultados indican que tanto la sensibilidad como la especificidad son del 80% ($28 / 35 = 0.8$; $4 / 5 = 0.8$). El hecho de que ambos estimadores tengan el mismo valor se debe a una coincidencia de este caso en especial. Calculando los intervalos de confianza al 95% para los parámetros considerados, se obtiene:

Sensibilidad: $28 / 35 \pm 1.960 \sqrt{(28 / 35) (7 / 35) / 35}$

$0.8 \pm 1.960 (0.0676)$

$0.6675 ; 0.9325$

ESTA TESIS
SALIR DE LA
NO DEBE
BIBLIOTECA

Especificidad: $4 / 5 \pm 1.96 \sqrt{(4 / 5) (1 / 5) / 5}$

$0.8 \pm 1.96 (0.1789)$

$0.4494 ; 1.1506$

La sensibilidad real tiene una probabilidad de 95% de caer entre 67 y 93% -redondeando los porcentajes-. La especificidad tiene un intervalo que va de 45 a 115%. El concepto de especificidad y su forma de cálculo la limitan a valores de cero a uno; por tanto el límite superior del intervalo se trunca en 1, anotándose el intervalo:

$0.4494 ; 1.0000$

En los datos del ejemplo, tanto la sensibilidad como la especificidad tienen el mismo valor estimado -80%, pero difieren sensiblemente en los intervalos de confianza. La causa es que el número de observaciones incluido en cada una es diferente: para la sensibilidad hay 35 datos y sólo cinco en especificidad, por lo que la estimación de la sensibilidad es más estable y, en consecuencia, el intervalo de confianza es más estrecho.

Cuando el intervalo de confianza es demasiado amplio, como sucede con el de la especificidad, en vez de inferir sobre el parámetro de interés, se debe pensar que la muestra es demasiado pequeña para aportar información concluyente; es decir, que se

sabe poco acerca del parámetro y por eso el intervalo cubre una amplia gama de valores. En estos casos conviene aumentar la cantidad de datos analizados antes de emitir una conclusión.

c) Calidad de la predicción

Para el profesional que utiliza una prueba diagnóstica es más interesante conocer el valor de la predicción basada en la prueba que la especificidad y sensibilidad de la misma. Los parámetros que indican la calidad de la predicción se conocen como sensibilidad y especificidad diagnósticas.

La sensibilidad diagnóstica (S_d) representa la tasa de aciertos cuando la prueba resulta positiva, por lo que también se le llama calidad de predicción de la prueba positiva. La proporción de predicciones correctas cuando la prueba es negativa se expresa mediante la especificidad diagnóstica (E_d) o calidad de predicción de la prueba negativa. En forma general, los parámetros que describen el valor de predicción se expresan con las siguientes relaciones:

$$S_d = \frac{\text{Número de resultados positivos correctos}}{\text{Número total de resultados positivos en la población}}$$

$$E_d = \frac{\text{Número de resultados negativos correctos}}{\text{Número total de resultados negativos en la población}}$$

Es importante notar que la calidad de predicción depende de dos factores: por un lado de la eficacia de la prueba misma, expresada mediante la sensibilidad y especificidad, y por otro, de la frecuencia de la enfermedad en la población donde se aplicará la prueba, o proporción de individuos que poseen la condición a diagnosticar. Mientras que la sensibilidad y especificidad son cualidades propias de la prueba, la sensibilidad y la especificidad diagnósticas son específicas de la población en que se calculan; esto aclara la distinta utilidad de los parámetros descritos: la sensibilidad y la especificidad indican la eficacia general de la prueba; a su vez, la sensibilidad y especificidad diagnósticas señalan su valor de uso en cierta población.

Existen dos procedimientos para estimar la sensibilidad y especificidad diagnósticas. El primero, conocido como método bayesiano, se basa en el teorema de Bayes y en una estimación a priori de la probabilidad de que un individuo sometido a la prueba tenga la condición que se desea diagnosticar. Por lo que

se requiere conocer de antemano el porcentaje de la población que presenta la enfermedad; Mendez et al. (1984) describen en detalle el fundamento de este método. A partir de la sensibilidad y especificidad ya estimadas y contando con una aproximación a priori de la probabilidad de que un individuo tenga la condición de interés (P_e), la estimación bayesiana de los parámetros de la calidad predictiva es:

$$S_d = \frac{S P_e}{S P_e + (1 - E) (1 - P_e)}$$

$$E_d = \frac{E (1 - P_e)}{E (1 - P_e) + (1 - S) P_e}$$

Existen algunas dificultades inherentes a la estimación basada en el teorema de Bayes. Es difícil obtener un buen estimador de P_e , ya que si no es una condición fácil de observar, lo que es obvio desde el momento mismo en que se está investigando una prueba que permita distinguir entre los que presentan el problema y los que no, será difícil que se cuente con un escrutinio completo que indique de modo preciso la tasa de presentación del problema. Tal vez se tengan datos de la proporción de individuos que sufren fases más avanzadas del padecimiento, cuando la condición se hace evidente, pero es claro que este dato no es un buen estimador del porcentaje que sufre las etapas iniciales.

Para subsanar estas deficiencias, puede emplearse el segundo procedimiento, conocido como estimación diagnóstica, que consiste en aplicarla a un grupo de individuos sospechosos de padecer el problema, esperar al desarrollo de etapas avanzadas de la enfermedad o hacer las pruebas necesarias para conocer con precisión el estado real de los pacientes y, por último, estimar directamente la tasa de aciertos entre los que tuvieron un resultado positivo (S_d), al igual que la proporción de individuos que, teniendo un resultado negativo, estaban libres del padecimiento (E_d). Colocando los datos como se indica en el cuadro III.14, las estimaciones de la calidad de predicción se hacen con las siguientes fórmulas:

Sensibilidad diagnóstica: $S_d = a / r_0$

Especificidad diagnóstica: $E_d = d / r_1$

De manera similar a lo que sucede con la especificidad y la sensibilidad, al trabajar con este tipo de estimación se puede calcular el intervalo de confianza con base en el procedimiento explicado para una proporción.

Al comparar las dos formas para estimar el valor de uso de una prueba diagnóstica se encontrará que producen resultados parecidos cuando la estimación utilizada en el método bayesiano es cercana a la frecuencia observada del problema en el procedimiento diagnóstico; de otra manera habrá diferencias, que serán mayores cuanto más distintas sean. Así, será preferible la estimación diagnóstica a la bayesiana. Siempre que sea posible debe utilizarse el método diagnóstico, dejando el procedimiento bayesiano como una opción para los problemas en que no sea factible la estimación directa.

EJEMPLO III.18: cálculo bayesiano del valor de predicción

Si en determinado rancho se considera que cerca del 15% de los animales están enfermos de brucelosis y se piensa aplicar una prueba con sensibilidad de 90% y especificidad de 95% para detectarla, el valor de predicción para ese hato es:

$$\underline{Sd} = \frac{(0.90) (0.15)}{(0.90) (0.15) + (1 - 0.95) (1 - 0.15)} = 0.7606$$

$$\underline{Ed} = \frac{(0.95) (1 - 0.15)}{(0.95) (1 - 0.15) + (1 - 0.90) (0.15)} = 0.9818$$

Estos valores indican que la probabilidad de que un individuo cualquiera esté enfermo es de 15% -la proporción de enfermos en ese hato-; si se aplica la prueba y el resultado es positivo, la probabilidad aumenta al 76% ($\underline{Sd} = 0.7606$). En cuanto a los que no tienen el problema, el porcentaje en el hato es de 85% ($1 - 0.15 = 0.85$), a su vez si la prueba es negativa, la probabilidad de que esté sano es de 98% ($\underline{Ed} = 0.9818$).

Estos estimadores también pueden interpretarse en el sentido de que la prueba positiva tiene una confiabilidad del 76%, o sea que el 76% de los que resultan positivos realmente están enfermos; y que la prueba negativa es 98% confiable, es decir, que el 98% de los que tienen un resultado negativo de verdad están libres del problema que se desea diagnosticar. Si en vez de 15%, el 50% de los animales tuviera la brucelosis, la calidad de la predicción resultaría:

$$\underline{Sd} = \frac{(0.90) (0.50)}{(0.90) (0.50) + (1 - 0.95) (1 - 0.50)} = 0.9474$$

$$\underline{Ed} = \frac{(0.95) (1 - 0.50)}{(0.95) (1 - 0.50) + (1 - 0.90) (0.50)} = 0.9048$$

En el segundo hato, la prueba positiva tiene una confiabilidad de casi el 95% -sensibilidad diagnóstica-, en tanto la especificidad diagnóstica resultó apenas arriba del 90%, por lo cual en este hato tiene más valor la prueba positiva que la negativa.

EJEMPLO III.19: comparación de los dos métodos analizados

Si el método examinado en el ejemplo de la sección anterior se aplica en una piara con 85% de fertilidad -85 de cada 100 cerdas en que se requiere la prueba están gestantes-, el cálculo bayesiano del valor de predicción es:

$$\underline{Sd} = \frac{0.8 (0.85)}{0.8 (0.85) + (1 - 0.8) (1-0.85)} = 0.9577$$

$$\underline{Ed} = \frac{0.8 (1 - 0.85)}{0.8 (1 - 0.85) + (1 - 0.8) (0.85)} = 0.4138$$

Lo que indica una alta confiabilidad para la prueba positiva, ya que la probabilidad de que una cerda con diagnóstico positivo esté gestante es de casi 96%. Por su parte, el diagnóstico negativo es de poca utilidad, debido a que sólo el 41% de las hembras con resultado negativo en la medición de resistencia eléctrica en verdad están vacías -casi 60% están preñadas-.

En conclusión, es una prueba válida sólo para separar las hembras gestantes -con diagnóstico positivo- de aquellas que posiblemente no lo estén -diagnóstico negativo-, para luego aplicar, en este último grupo, otra rutina que permita discernir entre las realmente vacías y las preñadas, es decir, que tenga mejor especificidad. Esta segunda prueba debe ser más cara o más complicada que el diagnóstico basado en la resistencia eléctrica,

de otro modo lo conveniente sería utilizar la prueba más precisa en todos los casos.

La poca confiabilidad de los resultados negativos de la prueba se debe a que más de la mitad de las veces son falsos negativos, de modo que en este caso no es la especificidad lo que induce a fallas, sino la falta de sensibilidad. Aun si la especificidad fuera de 100%, el valor de predicción de la prueba negativa sería bajo, menos del 47%. El cálculo de la especificidad diagnóstica si la especificidad del ejemplo fuera de 100% aparece en seguida:

$$\text{Si } \underline{E} = 1, \underline{Ed} = \frac{1 (1 - 0.85)}{1 (1 - 0.85) + (1 - 0.8) (0.85)} = 0.4688$$

Si se considera que la evaluación de la prueba se hizo en la piara en que se aplicará, es válido estimar el valor de predicción con el método diagnóstico. El estimador y los intervalos de confianza al 99% son:

$$\underline{Sd} = 28 / 29 = 0.9655$$

$$\underline{Ed} = 4 / 11 = 0.3636$$

$$\text{Sensibilidad: } 28 / 29 \pm 2.576 \sqrt{(28 / 29) (1 / 29) / 29}$$

$$0.9655 \pm 2.576 (0.0339)$$

$$0.8991 ; 1.0319 \approx 0.8991 ; 1.0000$$

$$\text{Especificidad: } 4 / 11 \pm 2.576 \sqrt{(4 / 11) (1 / 11) / 11}$$

$$0.3636 \pm 2.576 (0.1450)$$

$$-0.0100 ; 0.7372 \approx 0.0000 ; 0.7372$$

Como ya se explicó, los intervalos se truncaron al límite establecido para la variable analizada. Puede constatarse que la conclusión es parecida en ambos procedimientos; en este caso la semejanza se debe a que la fertilidad observada es parecida a la que se consideró en el método bayesiano.

CAPITULO IV

PRUEBAS ESTADISTICAS PARA UNA MUESTRA 2X2

1. INTRODUCCION

Con frecuencia el profesional de cualquier área debe enfrentarse a un nuevo método, por lo que se ve obligado a investigar si este brinda mejores resultados que el que se tiene en uso. Por tanto resolverá preguntas como ¿el nuevo medicamento aumenta la probabilidad de sanar?, ¿el nuevo método de inspección detecta un mayor porcentaje de los casos problemáticos? y otras similares. Al examinar estas preguntas y recordar que las medidas estadísticas describen la manera en que sucede un fenómeno, se deduce que la comparación entre dos métodos se reduce a contrastar los parámetros de ambos, mediante la comparación de los estimadores correspondientes. En esencia, este problema es igual al del investigador que quiere analizar un proceso como comparando el estimador del parámetro de interés con un valor previamente establecido, o bien, cuando requiere investigar dos procesos distintos y obtener una evaluación de los resultados que proporcionan.

Al comparar los estimadores obtenidos de las muestras de dos o más poblaciones, o los de una muestra con un valor fijo, el problema es decidir si la diferencia que se observa entre ellos (casi nunca son iguales) refleja alguna diferencia entre los parámetros de las poblaciones en estudio, o si sólo son pequeñas variaciones originadas por fluctuaciones aleatorias presentes en la muestra. Las pruebas estadísticas o de hipótesis proporcionan un medio para tomar tal decisión; este método se basa en criterios objetivos y bien definidos, lo cual permite realizar los conceptos de juicio sobre las diferencias observadas en los estimadores.

Las pruebas estadísticas se basan en un juicio de hipótesis: H_0 es la hipótesis nula y H_a la hipótesis alternativa. La hipótesis nula, H_0 , expresa la igualdad entre los parámetros que se quieren comparar, por su parte, la hipótesis alterna, H_a , señala la desigualdad entre los valores. Al aplicar la prueba de esta estadística a los datos de la muestra se decide por una u otra de las hipótesis y, por tanto, si los parámetros comparados son o no iguales.

2. LOGICA Y FUNCIONAMIENTO DE LAS PRUEBAS ESTADISTICAS

El esquema de funcionamiento de una prueba de hipótesis puede compararse con el de un examen de laboratorio en este último

timo, el animal está o no enfermo, lo cual es un hecho, y la prueba tiende al diagnóstico de este, por tanto, el resultado puede ser sano o enfermo; si el estado real del animal y el resultado del laboratorio coinciden el diagnóstico es correcto.

Sin embargo, es posible declarar enfermo a un animal que no lo está (falso positivo) o considerar sano a uno enfermo (falso negativo). Estos dos tipos de error también pueden ocurrir en las pruebas de hipótesis; cuando no hay diferencia entre los parámetros de interés pero el estudio de las muestras lleva a decidir que si la hay, el resultado es una falla llamada error tipo I, equivalente al falso positivo. Si la inspección de los estimadores conduce a una decisión de no diferencia o similitud y los parámetros en realidad si difieren, se presenta un error tipo II, que corresponde al falso negativo.

Representar los dos tipos de prueba con tablas 2x2 servirá para aclarar los conceptos anteriores; en el cuadro IV.1 se presenta la prueba de laboratorio y en el IV.2 la de hipótesis.

Cuadro IV.1

Esquema del funcionamiento de una prueba de laboratorio

		Estado real	
		Sano	Enfermo
Resultado de la prueba	Sano	Correcto (especificidad)	Error (falso negativo)
	Enfermo	Error (falso positivo)	Correcto (sensibilidad)
Total		p(sano)	p(enfermo)

El principio lógico de ambas es igual: se supone que el animal está sano (los parámetros se consideran iguales), a menos que haya suficiente evidencia en contra. Dado este principio, resulta natural que sean pruebas muy confiables cuando se decide por la diferencia de parámetros o por la enfermedad, pero la seguridad del resultado al declarar la igualdad de parámetros o la ausencia de enfermedad depende de la calidad de la muestra que se ha analizado en la prueba y del resto de la información que se tenga en relación al caso.

Cuadro IV.2

Esquema del funcionamiento de una prueba estadística.

		Relación real entre los parámetros	
		Iguals (H_0 verdadera)	Difieren (H_0 falsa)
Resultado de la prueba	Iguals (H_0)	Correcto (eficiencia)	Error (tipo II)
	Difieren (H_a)	Error (tipo I)	Correcto (potencia)
Total		$p(H_0 \text{ cierta})$	$p(H_0 \text{ falsa})$

a) TIPOS DE FALLAS EN LA PRUEBAS DE HIPOTESIS

Una prueba estadística permite controlar la probabilidad, conocida como alfa (α), de cometer un error tipo I, esta probabilidad es decisión del investigador. Lo más común es usar 1/20 (5%) o 1/100 (1%); por lo cual, se habla de $\alpha = 0.05$ y de $\alpha = 0.01$.

El problema surge al considerar la posibilidad del error tipo II, llamada beta (β). La magnitud de β depende del número de observaciones en la muestra y del nivel utilizado para α . Al complemento de β , $(1 - \beta)$, o probabilidad de no cometer un error tipo II se le conoce como potencia de la prueba. En el capítulo correspondiente a los métodos para determinar el tamaño de la muestra se analiza su relación con la potencia.

b) INDICADORES DE LA CALIDAD DE UNA PRUEBA

La sensibilidad de un examen de laboratorio es el porcentaje de enfermos con un diagnóstico correcto. La especificidad indica la proporción de sanos en que el resultado es acertado (cuadro IV.1). Si se continúa con la comparación de las pruebas de hipótesis con las de laboratorio, puede considerarse que en el

total de resultados que indican una diferencia y que son correctos, las fracciones $1-\alpha$ y $1-\beta$ corresponden a la especificidad y sensibilidad, respectivamente.

Por tanto, la especificidad y la sensibilidad de los exámenes de laboratorio equivalen, conceptualmente, a la eficiencia y la potencia de las pruebas estadísticas

En ambas existe una tercera medida indicativa de la calidad: la exactitud. El grado de exactitud de una prueba se define como el porcentaje de casos en que el resultado es correcto.

$$\text{Exactitud} = \frac{\text{Número de aciertos}}{\text{Número total de casos}}$$

Aunque la exactitud representa un concepto muy interesante de la calidad de una prueba, su evaluación no es sencilla. Se requiere conocer el porcentaje de sanos en la población para el examen de laboratorio, o la probabilidad de casos con H_0 verdadera para las pruebas de hipótesis. A continuación se muestra la ecuación apropiada para calcular la exactitud de una prueba de hipótesis y de una de laboratorio, en ese orden.

$$\text{Exactitud} = \frac{1 - \alpha}{p(H_0 \text{ cierta})} + \frac{1 - \beta}{p(H_0 \text{ falsa})}$$

$$\text{Exactitud} = \frac{\text{especificidad}}{p(\text{sano})} + \frac{\text{sensibilidad}}{p(\text{enfermo})}$$

3. DESARROLLO DE UNA PRUEBA ESTADÍSTICA

Al aplicar una prueba de hipótesis se calcula la probabilidad de obtener una diferencia igual o mayor que la observada entre lo estipulado en la hipótesis nula y los datos de la o las muestras, en caso de que H_0 fuese verdadera; esto es, se considera cierta la hipótesis nula y bajo estas condiciones se evalúa la probabilidad de que el resultado de una muestra difiera de la hipótesis nula tanto o más que lo detectado en la muestra.

Si esta probabilidad es pequeña (igual o menor que α) se afirma que la hipótesis nula es falsa. La afirmación se basa en que, si realmente se cumple la hipótesis nula, sería en extremo improbable obtener una diferencia como la detectada, por lo que la hipótesis nula debe ser falsa. Existe la posibilidad de equivocarse en tal afirmación, misma que está representada por α : la probabilidad de cometer un error tipo I.

Si la probabilidad de una diferencia tan extrema o más que la observada es mayor que α , se concluye que no existe evidencia suficientemente fehaciente para rechazar la hipótesis nula. Por lo tanto, se dice que la diferencia no es significativa, o que los datos son compatibles con la nulidad de efecto.

Cuando se rechaza la hipótesis nula se conoce la probabilidad de error (α), pero al no rechazarla por lo general se desconoce la probabilidad de cometer un error tipo II (β), por lo que es poco conveniente afirmar que H_0 es cierta. El peso de una conclusión en este sentido depende de la calidad de la muestra: si está constituida por un número suficiente de observaciones, seleccionadas con un método adecuado, entonces el no rechazo permite asegurar que en caso de existir alguna diferencia entre la hipótesis nula y el comportamiento de la población de referencia, lo más probable es que sea mínima y de poco interés práctico.

En síntesis, para desarrollar una prueba estadística se evalúa la probabilidad de que los datos y la hipótesis nula difieran tanto o más que lo detectado en la muestra. Si es baja se rechaza H_0 y se opta por la H_a , con una probabilidad α de error. Por el contrario, si la probabilidad es mayor que el valor definido para α se concluye que no hay evidencia para rechazar H_0 , pero no se afirma que sea verdadera.

4. APLICACION DE UNA PRUEBA ESTADÍSTICA

Para explicar la aplicación de una prueba de hipótesis se utilizará la prueba binomial. Este sencillo método sirve para comparar el porcentaje de respuesta en una muestra con respecto a un valor previamente estipulado.

En este caso la hipótesis nula señala que la tasa global de respuesta en la población representada en la muestra es el valor ya establecido; por supuesto, la hipótesis alterna indica la diferencia entre la proporción de respuesta en la población y el valor considerado como punto de referencia.

$H_0: p =$ valor ya determinado

$H_a: p \neq$ valor ya determinado

Si se quisiera verificar que el porcentaje de cerdas preñadas al utilizar un método de inseminación artificial es de 70%, tal como afirma la casa que lo comercializa, las hipótesis se representarían así:

$$H_0: P = 0.70$$

$$H_a: P < 0.70$$

(Recuérdese que las probabilidades se expresan con valores entre cero y uno)

Supóngase que se toma una muestra de ocho cerdas inseminadas, de las cuales cuatro resultaron gestantes, entonces $p = 4/8 = 0.50$, con lo cual se estima un fertilidad de 50%. A continuación se procede a evaluar la probabilidad de obtener una muestra con una diferencia de 20% o más ($70\% - 50\% = 20\%$), bajo el supuesto de que H_0 es cierta.

El cuadro IV.3 señala las probabilidades correspondientes a la distribución binomial con ocho casos y 70% de éxito: bin ($n = 8, p = 0.70$). La probabilidad de obtener cuatro éxitos o menos se calcula sumando las correspondientes a 4, 3, 2, 1 y 0 éxitos. Para llegar al resultado de la prueba también debe sumarse la probabilidad de ocho éxitos, ya que señala el 20% o más de diferencia en el otro sentido.

Cuadro IV.3

Probabilidades de la distribución binomial ($n = 8, p = 0.7$)

Porcentaje de éxitos	Número de éxitos	Probabilidad del suceso
100	8	0.057 65
88	7	0.197 65
75	6	0.296 48
63	5	0.254 12
50	4	0.136 14
38	3	0.046 67
25	2	0.010 00
13	1	0.001 22
0	0	0.000 07

La probabilidad de una discrepancia como la observada o mayor, en cualquier sentido, es de 0.251 75, por lo que se concluye que no hay evidencia suficiente para rechazar la H_0 . El resultado indica que en una población con $p = 0.70$ una de cada cuatro muestras de ocho datos ($0.251\ 75 \approx 1/4$) tendría una diferencia de 20% o más con respecto al 70% establecido en la hipótesis nula.

Si el interés de la prueba fuera sólo detectar la baja fertilidad, se plantea un juego de hipótesis direccionales:

$$H_0: p \geq 0.70$$

$$H_a: p < 0.70$$

Y al evaluar la probabilidad asociada a la diferencia observada sólo se consideran los valores en el sentido que señala H_a ; en este caso el resultado sería 0.194 10, que aun es demasiado grande para rechazar la hipótesis nula.

Si en vez de cuatro hubiera sólo dos hembras gestantes ($2/8 = 0.25$), la probabilidad asociada a una diferencia de 45% o más ($25\% - 70\% = 45\%$) es de sólo 0.011 29, es decir la suma de las correspondientes a 2, 1 y 0 éxitos, que implica una probabilidad pequeña de equivocarse al afirmar que H_0 es falsa, por lo cual en esta situación debería concluirse que el método ofrece resultados inferiores al 70% de fertilidad.

5. PRUEBA DE JI CUADRADA (χ^2)

Sin duda la prueba más conocida para el análisis de variables cualitativas es la de ji cuadrada (χ^2), que toma su nombre de la distribución ji cuadrada de probabilidad en la cual se basa. La prueba de χ^2 para independencia entre dos variables cualitativas fue desarrollada por Pearson (1900), y sirve precisamente para probar hipótesis en relación con la independencia entre dos variables nominales u ordinales; proporcionando un medio para verificar si las frecuencias observadas en cada categoría son compatibles con la independencia entre las dos variables consideradas (en el Cap. 1 se explica el concepto de independencia). Para lograrlo se calculan los valores que indicarían la independencia absoluta, llamados frecuencias esperadas. Las hipótesis consideradas en esta prueba son:

H₀: Las variables son independientes

H_a: Las variables están asociadas

La prueba de χ^2 solo puede aplicarse a estudios basados en muestras independientes (vease el Cap. II para aclarar este concepto) cuando todos los valores esperados son mayores de cinco. Los esperados se obtienen conforme la hipótesis de independencia, es decir son los que indican la independencia absoluta entre las dos variables. Para el caso de la tabla 2x2 se obtienen mediante las siguientes expresiones:

$$\begin{aligned} \text{Esperado en la celda: } a &= \frac{r_0 \cdot c_0}{N} \\ b &= \frac{r_0 \cdot c_1}{N} \\ c &= \frac{r_1 \cdot c_0}{N} \\ d &= \frac{r_1 \cdot c_1}{N} \end{aligned}$$

La prueba de χ^2 utiliza una aproximación a la distribución ji cuadrada para evaluar la probabilidad de una discrepancia igual o mayor que la existente entre los datos y la hipótesis nula; la exactitud de esta aproximación depende de que los valores esperados no sean muy pequeños y, en menor medida, de que el contraste entre ellos no sea demasiado grande. Al usar la prueba de χ^2 debe cuidarse que se cumplan estas condiciones.

La comparación de los valores esperados con sus correspondientes observados se efectúa mediante:

$$\chi^2_{\text{es}} = \sum \frac{(\text{OBSERVADO}_i - \text{ESPERADO}_i)^2}{\text{ESPERADO}_i}$$

La expresión indica que deben sumarse las diferencias entre las frecuencias observadas y las esperadas conforme a la hipótesis de independencia, elevadas al cuadrado y divididas entre la esperada. Para el caso de la tabla 2x2, el cálculo puede reducirse a:

$$\chi^2_{\text{es}} = \frac{N (a d - b c)^2}{r_0 r_1 c_0 c_1}$$

El resultado de este cálculo se compara con la distribución de χ^2 cuadrada con un grado de libertad, por lo que si es mayor de 6.63 ($\alpha = 0.01$) o de 3.84 ($\alpha = 0.05$) se considera que la desviación con respecto a la independencia es significativa, es decir que las variables no son independientes. Si el valor es menor que 3.84 se dice que los datos no indican una relación significativa entre las variables, o bien que no se encontró una desviación significativa respecto de la independencia entre ellas.

En datos de muestreos comparativos o experimentales, al probar las hipótesis relacionadas con la independencia entre la variable de respuesta y la explicativa, la prueba de χ^2 permite averiguar, de modo implícito, si el porcentaje de respuesta es igual en ambos grupos:

$$H_0: p_a = p_b$$

$$H_a: p_a \neq p_b$$

a) CORRECCIÓN DE YATES

A consecuencia de utilizar una distribución continua -la de χ^2 - para representar un fenómeno discreto -el número de casos en cada clasificación-, existe una pequeña falla en la aproximación. Yates (1934) propuso una corrección que ajusta a este hecho y que mejora la concordancia con la distribución de χ^2 .

Se ha hablado mucho sobre la conveniencia de usar la corrección de Yates, las conclusiones al respecto indican que debe aplicarse en todos los casos (Fleiss, 1973). El efecto de esta corrección es notorio en muestras pequeñas, pero es muy reducido cuando se tienen muchos datos. En seguida se muestra la expresión que permite calcular el valor de χ^2 para una tabla 2x2 con la corrección de Yates:

$$\chi^2_c = \frac{N (|a d - b c| - N / 2)^2}{r_0 r_1 c_0 c_1}$$

Las barras (|) indican que se debe hacer caso omiso del signo al obtener la diferencia señalada, dejando el resultado como positivo.

EJEMPLO IV.1: La prueba de χ^2

A fin de comparar la eficiencia de dos vacunas se formaron al azar dos lotes de gallinas de postura, cada lote recibió un tipo de vacuna. En un momento determinado se expusieron las aves a la enfermedad y se registró el número de gallinas enfermas en cada lote. El porcentaje de aves que resistieron el desafío señala la eficacia de la vacuna (80% vs 85%).

Cuadro IV.4

Resultado de la comparación entre dos vacunas para aves

		Vacuna A	Vacuna B	Total
Resultado del desafío	Enfermó	20	15	35
	No enfermó	80	85	165
	Total	100	100	200

$$\chi^2_{ca} = \frac{200 \left((20)(85) - (15)(80) \right)^2}{(100)(100)(35)(165)} = 0.5541$$

Como el valor de χ^2 calculada es menor que el de la distribución de χ^2 (3.84 para $\alpha = 0.05$), se concluye que los datos no indican una diferencia significativa en la eficiencia de ambas vacunas.

De una manera más formal se diría: no hubo diferencia significativa en la eficacia de ambas vacunas ($p > 0.05$). En el paréntesis se especifica que se usó 0.05 como nivel de significancia y que la eficacia estimada para las vacunas no fue lo bastante distinta como para rechazar la igualdad de efectos.

EJEMPLO IV. 2: otro caso de la prueba de χ^2

Para estudiar la ventaja de un tratamiento de acupuntura para vacas que no han quedado preñadas después de varios servicios, se dividió al azar en dos grupos a las vacas que mostraban problemas reproductivos; unas fueron sometidas a la acupuntura y en seguida se inseminaron, y otras sólo se inseminaron.

Cuadro IV.5

Efecto de la acupuntura sobre la concepción en vacas

		Acupuntura	Testigo	Total
Resultado de la inseminación	Gestantes	20	4	24
	No gestantes	17	27	44
	Total	37	31	68

Toscano de Vázquez et al. (1984)

$$\chi^2 = \frac{68 \left((20)(27) - (4)(17) \right)^2}{(37)(31)(24)(44)} = 10.7703$$

Para este ejemplo el valor calculado ($\chi^2 = 10.77$) es mayor que el correspondiente a $\alpha = 0.01$ en la distribución de χ^2 con un grado de libertad (6.63), por lo que se concluye que existe una diferencia altamente significativa en el porcentaje de vacas gestantes en ambos grupos: 54% contra 13% ($p < 0.01$). Anotar que se usó $\alpha = 0.01$ o señalar "p < 0.01" indica que se tiene una probabilidad de 1% o menor de cometer un error de tipo I, es decir de afirmar una diferencia inexistente.

Es importante destacar que la conclusión enunciada deriva de que la prueba detectó que la variable "aplicación o no de la acupuntura" está asociada con el resultado de la inseminación: gestante o vacía. Como el concepto de independencia entre dos variables binarias (véase el Cap. I) implica que cuando están asociadas se presentan diferentes probabilidades en una variable al cambiar de uno a otro nivel de la otra variable, entonces se deduce que los porcentajes de concepción difieren entre las vacas no tratadas y las sometidas a acupuntura.

6. PRUEBA EXACTA DE FISHER

Cuando al emplear la prueba de χ^2 se usan predeterminados valores, empíricos menores a cinco y cuando el total de casos es inferior a 60, se recomienda utilizar la prueba exacta de Fisher en vez de la de χ^2 . Sin embargo, existe una diferencia importante en el planteamiento de ambas pruebas: la de χ^2 evalúa la probabilidad de discrepancias en ambos sentidos, mientras que la prueba exacta de Fisher sólo considera diferencias en una dirección (véase la Sec IV.3). Para eliminar esta diferencia basta duplicar el resultado de la probabilidad exacta.

La prueba de Fisher se basa en el cálculo exacto de la probabilidad de encontrar diferencias iguales o mayores que las existentes entre los datos y la hipótesis de las nulas de independencia entre las variables. Aunque la evaluación de las probabilidades es simple, el procedimiento es largo y complicado porque requiere considerar la serie de tablas 2x2 que se forman al variar las diferencias mayores que las observadas.

Antes de exponer el cálculo de las probabilidades exactas, se introducirá la notación del factorial de un número entero, $n!$, que indica el factorial de n ; $n!$ significa el producto de todos los enteros menores o iguales a n , hasta uno:

$$n! = n (n-1) (n-2) (n-3) \dots (n-4) \dots (1) \quad (1)$$

$$\text{Por ejemplo: } 4! = 4 (3) (2) (1) = 24$$

$$7! = 7 (6) (5) (4) (3) (2) (1) = 5040$$

Por definición, el factorial de cero es uno ($0! = 1$).

La probabilidad de un conjunto específico de frecuencias en las celdas de una tabla 2x2, considerando que los totales marginales son valores fijos, se obtiene de la siguiente forma:

$$P = \frac{E_1! E_2! R_1! R_2!}{n! b_1! b_2! r_1! r_2!}$$

El procedimiento consiste en calcular esta probabilidad en la tabla que contiene los datos γ , también, en todas las tablas que impliquen una diferencia mayor que la observada; el resultado de la prueba es la suma de las probabilidades obtenidas. La probabilidad exacta se interpreta directamente, si es menor de 0.05 o de 0.01 se afirma que hay diferencia entre las proporciones en ambos grupos; la probabilidad de error tipo I es α calculada o la significancia observada. Feldman y Klinger (1963), idearon un método para abreviar el cálculo de esta prueba, en el ejemplo se describe ese método.

Para generar las tablas que indican una diferencia mayor que la existente en los datos, se toma el menor de a, b, c y d, se le resta uno y se ajustan los valores de las otras celdas a fin de conservar fijos los totales. Estos pasos se repiten hasta obtener una tabla cuya frecuencia menor sea cero.

EJEMPLO IV.3: Cálculo de la prueba exacta de Fisher

En un estudio acerca de la producción porcina en México, se analizaron granjas del bajo y del noroeste del país, y para cada una se averiguó si contaba con planes de producción bien establecidos. Los datos que aparecen en el cuadro IV.6 indican que 43% de las empresas en el bajo programan su producción y que en el noroeste el 71% lo hace.

Cuadro IV.6

Ubicación geográfica y planificación en 14 granjas

		Ubicación		Total
		Bajo	Noroeste	
Programas de producción	Planifican	3	5	8
	No planifican	4	2	6
Total		7	7	14

Si las dos variables son independientes, la probabilidad de que ocurra el arreglo de frecuencias que aparece en el cuadro IV.6 es:

$$P_2 = \frac{8! 6! 7! 7!}{3! 5! 4! 2! 14!} = 0.244 76$$

Se usa como subíndice de p la menor frecuencia que aparece en la tabla, en este caso el menor de 3, 5, 3 y 2, el cual es 2. Las tablas 2x2 que muestran diferencias mayores entre porcentajes de respuesta a la existente en los datos, se presentan en el cuadro IV.7.

Cuadro IV.7

Tablas 2x2 con mayor discrepancia respecto de la independencia que los datos del cuadro IV.6

(a)		Ubicación		Total
		Bajo	Noroeste	
Programas de producción	Planifican	2	6	8
	No planifican	5	1	6
	Total	7	7	14

(b)		Ubicación		Total
		Bajo	Noroeste	
Programas de producción	Planifican	1	7	8
	No planifican	6	0	6
	Total	7	7	14

$$P_1 = \frac{8! 6! 7! 7!}{2! 6! 5! 1! 14!} = 0.048 95$$

$$P_1 = \frac{(3) (2)}{(6) (5)} (0.244 76) = 0.048 95$$

$$P_0 = \frac{(2) (1)}{(7) (6)} (0.048 95) = 0.002 33$$

$$F = P_0 + P_1 + P_2 = 0.296 04$$

Puede corroborarse que este valor corresponde al resultado que se obtuvo con el método directo. La probabilidad exacta que se busca depende del planteamiento del problema; si la hipótesis original señala una dirección específica para la diferencia entre ambos grupos, entonces la probabilidad exacta de caer en un error del tipo I es el valor de F ya calculado. Este caso ocurre cuando la pregunta es ¿hay un mayor grado de planificación en el noroeste que en el bajo? El juego de hipótesis implicado es:

$$H_0: P_A \geq P_B$$

$$H_A: P_A < P_B$$

Como $F = 0.296 04$ es mayor que 0.05 se concluye que no hay una diferencia significativa en el porcentaje de planificación en ambos grupos.

Si, por el contrario, no se establece desde un principio que la diferencia entre los grupos ocurra de un modo específico, entonces la probabilidad de cometer un error tipo I será el doble de F. Así, al plantear preguntas como ¿difieren los empresas de ambas zonas en el grado de planeación de sus actividades? no se especifica ventaja de un grupo sobre el otro, por lo que las hipótesis estadísticas son:

$$H_0: P_A = P_B$$

$$H_A: P_A \neq P_B$$

Y la probabilidad exacta que resulta de la prueba de Fisher es el doble de F. Para este caso el ejemplo daría como resultado final: α observada = $2F = 2(0.296 04) = 0.592 08$ que por ser ma-

$$P_0 = \frac{8! \cdot 6! \cdot 7! \cdot 7!}{11! \cdot 7! \cdot 6! \cdot 0! \cdot 14!} = 0.00233$$

La probabilidad de cometer un error tipo I, al afirmar que los porcentajes son distintos en ambos grupos, se obtiene de la suma de las probabilidades asociadas a cada tabla:

$$F = \sum_k P_k$$

$$\text{Para este caso: } F = P_0 + P_1 + P_2 = 0.286 \ll 04$$

Feldman y Klinger (1963) desarrollaron un método abreviado para calcular F_k con base en una fórmula que se aplica de manera recurrente:

$$P_{k-1} = \frac{a_k \cdot d_k}{b_{k-1} \cdot c_{k-1}} P_k$$

Si la frecuencia más pequeña en la tabla 2x2 aparece en las casillas b o c , la expresión cambia a:

$$P_{k-1} = \frac{b_k \cdot c_k}{a_{k-1} \cdot d_{k-1}} P_k$$

Para realizar los cálculos conforme este método se obtiene la probabilidad en la tabla original, y para las subsiguientes se utiliza la fórmula mencionada; por ejemplo, usando los datos anteriores (cuadros IV.6 y IV.7):

$$P_2 = 0.24476$$

yendo una corrección por continuidad debida a Edwards (1948), se hace mediante:

$$M = \frac{(|b - c| - 1)^2}{b + c}$$

El valor de M se compara con la distribución de χ^2 con un grado de libertad, como se explicó en la prueba de χ^2 .

EJEMPLO IV.4: Aplicación de la prueba de McNemar

Para comparar la eficacia de dos desinfectantes de la ubre, se tomaron treinta vacas lecheras, y en cada ordeña se les aplicó un producto en dos de los cuartos de la ubre y, el otro en los restantes, habiendo definido con anterioridad que pezones recibirían cada tratamiento. Al cabo de un mes, se cultivó una alícuota de la leche producida en cada par de pezones y se registró si hubo o no crecimiento bacteriano. Los resultados aparecen en el cuadro IV.9.

Cuadro IV.9

Eficacia de dos desinfectantes aplicados en la ubre

		Cloruro de benzalconio al .05%		Total
		C/bacterias	S/bacterias	
Yodo	C/bacterias	6	1	7
	S/bacterias	8	15	23
Total		14	16	30

$$M = \frac{(|1 - 8| - 1)^2}{1 + 8} = 4$$

La eficacia del yodo ($7/30 = 0.23$) y la del cloruro de benzalconio ($14/30 = 0.47$) son significativamente distintas, ya que M ($M = 4$) es mayor que el valor de la distribución χ^2 con un grado de libertad para $\alpha = 0.05$ (3.84).

EJEMPLO IV.5: otro caso de la prueba de McNemar

Tal vez el uso más difundido para la prueba de McNemar sea en estudios del tipo antes y después, ya que son la situación más común de muestras apareadas. Por ejemplo, para evaluar el efecto de un tipo específico de piso sobre la frecuencia de problemas de patas en las hembras de una granja porcina, en un lote de sesenta cerdas se registró cuáles padecían dificultades de ese tipo, luego se cambió el piso de sus alojamientos y al cabo de cierto tiempo se inspeccionaron las condiciones de las patas de estas hembras. Los datos se listan en el cuadro IV.10.

Cuadro IV.10

Padecimientos de patas en cerdas antes y después de la colocación de un nuevo piso

		Antes		Total
		Sana	Enferma	
Después	Sana	40	11	51
	Enferma	5	4	9
Total		45	15	60

$$M = \frac{(11 - 5)^2}{11 + 5} = 1.5625$$

Como M ($M = 1.5625$) es menor que 3.84 (3.84 es el valor que corresponde a $\alpha = 0.05$ en la distribución de χ^2 con un grado de libertad) se concluye que no hay evidencia significativa para afirmar que el nuevo piso conduce a una distinta proporción de cerdas con problemas de patas ($p > 0.05$). La anotación $p > 0.05$ indica que debería aceptarse una probabilidad de error mayor de uno

en veinte (5%) para poder afirmar que el nuevo piso modifica la frecuencia de animales con el padecimiento.

8. PRUEBA DE GART

En muchas de las investigaciones comparativas basadas en muestras apareadas del tipo que se ha planteado en la prueba de McNemar y en la sección II.4, además de analizar la posible diferencia entre los tratamientos, surge el inters por averiguar si el orden en que se administran influye en su efecto; es decir, si el tratamiento A produce el mismo resultado cuando se aplica antes que el B que cuando se administra despues, o bien, si hay diferencia entre las secuencias A-B y B-A a causa de la secuencia de administración.

Gart (1966), desarrolló una prueba que permite determinar el efecto del tratamiento y el efecto de orden de administración. La prueba de Gart se basa en el planteamiento de McNemar de utilizar sólo los pares discordantes, pero emplea dos series de datos. Una en la que primero se aplicó el tratamiento A y luego el B, y otra en la que se administró primero B y despues A.

El cuadro IV.11 presenta el esquema de los datos para una prueba de Gart, de stos, sólo los pares discordantes son útiles para la prueba.

Cuadro IV.11

Esquema de los datos para la prueba de Gart

Orden de los tratamientos	Respuesta en:			
	A	B	Ambos	Ninguno
A, B	A+, B*	A-, B+	A+, B+	A-, B-
B, A	B-, A*	B+, A-	B+, A+	B-, A-

* Sólo las frecuencias de estas celdas se usan en la prueba.

La prueba de Gart evalúa en forma simultánea dos juegos de hipótesis; el primero, al igual que en las pruebas de Fisher y McNemar, se refiere a la igualdad (H_0) o diferencia (H_a) de la

tasa de respuesta a los tratamientos. El segundo juego de hipótesis se enfoca a la secuencia de administración:

Ho: el orden no afecta la probabilidad de respuesta

Ha: la probabilidad de respuesta depende del orden de aplicación

Así, cuando la prueba de Gart es significativa, indica que hay un efecto del orden, del tratamiento o de ambos, pero no especifica cuál; por tanto, debe investigarse cuál de los dos efectos está presente, o si son ambos a la vez. Para lograrlo se ordenan los datos como se indica en los cuadros IV.12 y IV.13 y se aplica en cada uno la prueba exacta de Fisher; el único cambio entre estos cuadros es que los valores b y d se han intercambiado. De acuerdo con el resultado de las dos pruebas se concluye si existen o no efectos de orden o de tratamiento.

Utilizando los datos que aparecen en el cuadro IV.12 (aunque pueden usarse los valores del IV.13 con el mismo resultado), la prueba de Gart se calcula:

$$G = \frac{(a - c)^2}{e_a} + \frac{(b - d)^2}{e_b}$$

Cuadro IV.12

Organización de los datos para probar el efecto de tratamiento

		Tratamiento aplicado en primer lugar		Total
		A	B	
Tratamiento en que se obtuvo respuesta	Primero	a (A+, B-)	b (B+, A-)	e _a
	Segundo	c (A-, B+)	d (B-, A+)	e _b
Total		e _a	e _b	N

El resultado se compara con la distribución χ^2 con dos grados de libertad, de modo que, cuando G es mayor que 9.21 ($\alpha = 0.01$) o que 5.99 ($\alpha = 0.05$), se concluye que los tratamientos, el orden de aplicación o ambos afectan la probabilidad de respuesta, y se procede a realizar las pruebas de Fisher.

Cuadro IV.13

Organización de los datos para probar el efecto de orden.

	Tratamiento aplicado en primer lugar		Total	
	A	B		
Tratamiento en que se obtuvo respuesta	A	a (A+, B-)	b (B-, A+)	r_{0a}
	B	c (A-, B+)	d (B+, A-)	r_{1a}
Total	c_{0a}	c_{1a}	N	

a) La prueba de Gart con muestras pequeñas

Si se quiere aplicar la prueba de Gart y el número de pares discordantes es pequeño, es decir cuando en una de las secuencias de aplicación (A-B, B-A o en ambas) haya menos de quince parejas discordantes, se efectúan directamente las dos pruebas exactas de Fisher.

EJEMPLO IV. la prueba de Gart con muestras grandes

Para analizar el efecto cicatrizante de la electroacupuntura y de la pomada de furacín se tomaron ochenta ratas; a cada una se le extirpó un círculo de piel del abdomen de 1 cm de diámetro. Se dividieron al azar en dos grupos de cuarenta ratas, tratando a unas con electroacupuntura y a las otras con furacín. A las tres semanas se registró si la herida había cicatrizado por completo o no. Un mes después se les practicó el otro tratamiento.

Los resultados aparecen en el cuadro IV.14, a su vez, los cuadros IV.15 y IV.16 muestran las frecuencias correspondientes a

los pares discordantes organizados para probar el efecto de tratamiento y orden, respectivamente.

Cuadro IV.14

Comparación del efecto de la electroacupuntura y de la pomada de furacín sobre la cicatrización

Orden de los tratamientos	Cicatrizó con:			
	Electroacup.	Furacín	Ambos	Ninguno
Electroac.-Furacín	14	3	10	13
Furacín-Electroac.	7	9	16	8

Cuadro IV.15

Datos ordenados para comparar el efecto de la acupuntura y el furacín

	Tratamiento aplicado en primer lugar		Total	
	Electroac.	Furacín		
Tratamiento con que cicatrizó la herida.	Primero	14	9	23
	Segundo	3	7	10
	Total	17	16	33

$$G = \frac{(14 - 3)^2}{17} + \frac{(9 - 7)^2}{16} = 7.37$$

Como G (G = 7.37) es mayor que 5.99, pero menor que 9.21, se concluye, con $\alpha = 0.05$, que existe algún efecto significativo, por lo que se procede a las pruebas exactas de Fisher. Para el cuadro IV.15, siguiendo el método de Feldman y Klinger (1963):

Cuadro IV.16

Datos organizados para evaluar el efecto de orden de administración

	Tratamiento aplicado en primer lugar		Total
	Electroac.	Furacín	
Tratamiento con el que cicatrizó la herida	Electroac.	14	21
	Furacín	3	12
	Total	17	33

$$f_3 = \frac{23!}{14! 9!} \frac{10!}{3! 7!} \frac{17!}{7! 10!} \frac{16!}{33!} = 0.084 04$$

$$f_2 = \frac{(9) (3)}{(15) (8)} P_3 = 0.018 91$$

$$f_1 = \frac{(8) (2)}{(16) (7)} P_2 = 0.002 70$$

$$f_0 = \frac{(7) (1)}{(17) (8)} P_1 = 0.000 14$$

$$f = P_0 + P_1 + P_2 + P_3 = 0.105 79$$

El resultado para el efecto de tratamiento ($F = 0.106$) señala que no hay suficiente evidencia para asegurar que los diferentes es el porcentaje de cicatrización. Al analizar el efecto de orden de administración de los tratamientos:

$$P_3 = \frac{21! 12! 17! 16!}{14! 7! 3! 9! 33!} = 0.021 92$$

$$P_2 = \frac{(7) (3)}{(15) (10)} P_2 = 0.003 07$$

$$P_1 = \frac{(6) (2)}{(16) (11)} P_1 = 0.000 21$$

$$P_0 = \frac{(5) (1)}{(17) (12)} P_0 = 0.000 01$$

$$F = P_0 + P_1 + P_2 + P_3 = 0.025 21$$

Ya que el valor de F es pequeño ($F = 0.025$), se concluye que existe un efecto importante del orden de aplicación; puede constatarse que en 70% (23/33) de los pares discordantes hubo respuesta al primer tratamiento, mientras que sólo 30% (10/33) cicatrizaron con el segundo.

Grizzle (1965) demostró que, al ocurrir este tipo de casos, es decir, cuando hay un efecto de orden y, en especial, cuando se influye de manera distinta a los tratamientos comparados, sólo los datos del primer periodo son válidos para estimar los porcentajes de respuesta. Para el ejemplo se tomarían sólo los datos del tratamiento que se aplicó en primer lugar, considerando entonces dos muestras independientes de cuarenta ratas cada una, con proporciones de cicatrización de 60% (24/40) para la acupuntura y 58% (23/33) para la pomada de furacín.

EJEMPLO IV.7: la prueba de Gart con muestras pequeñas

En una investigación sobre el bloqueo colinérgico y su relación con la manifestación del dolor, se comparó el efecto analgésico de la neomelubrina y de la atropina. A la mitad de un grupo de 32 ratones se les administró neomelubrina y se registró si respondían a un estímulo eléctrico de intensidad fija; después se les dio atropina y se repitió el examen del reflejo. La otra parte primero recibió la atropina y luego la neomelubrina. Los datos aparecen en el cuadro IV.17, en el IV.18 se ordenaron para

probar el efecto del tratamiento y en el IV.19 están organizados para evaluar el efecto de orden de aplicación.

Como el mínimo número de pares discordantes es diez -en ambas secuencias sólo hubo diez parejas discordantes-, se procede a las dos pruebas exactas de Fisher.

Cuadro IV.17

Acción analgésica de la neomelubrina y la atropina

Orden de los tratamientos	Hubo analgesia con:			
	Neomelubrina	Atropina	Ambas	Ninguna
Neomelub.-Atrop.	8	2	4	2
Atrop.-Neomelub.	7	3	5	1

Cuadro IV.18

Tabla 2x2 para comparar el efecto analgésico de dos fármacos

	Fármaco aplicado en primer lugar		Total	
	Neomelubrina	Atropina		
Fármaco con el que se obtuvo analgesia	Primero	8	3	11
	Segundo	2	7	9
	Total	10	10	20

A continuación se muestra la prueba exacta de Fisher enfocada al efecto de tratamiento, y después la que evalúa el orden de administración de los fármacos.

$$P_{12} = \frac{11! 9! 10! 10!}{8! 3! 2! 7! 20!} = 0.03215$$

$$P_1 = \frac{(3) (2)}{(9) (8)} P_2 = 0.00268$$

$$P_0 = \frac{(2) (1)}{(10) (9)} P_1 = 0.00006$$

$$F = P_0 + P_1 + P_2 = 0.03489$$

Cuadro IV.19

Tabla 2x2 para evaluar el efecto de orden de administración de atropina y neomelubrina

	Fármaco aplicado en primer lugar		Total
	Neomelubrina	Atropina	
Fármaco con el que se obtuvo analgesia	Neomelubrina	8	11
	Atropina	2	9
Total	10	10	20

Como F es menor que 0.05 (F = 0.03489), existe una diferencia significativa entre los fármacos en cuanto al porcentaje de animales que muestran analgesia.

Para analizar la significancia del efecto de orden de administración, se utilizan los datos organizados como se listan en el cuadro IV.19.

$$P_2 = \frac{15! 5! 10! 10!}{8! 7! 2! 3! 20!} = 0.34830$$

$$P_1 = \frac{(7)(2)}{(9)(4)} P_2 = 0.13545$$

$$P_0 = \frac{(4)(1)}{(10)(5)} P_3 = 0.01625$$

$$F = P_0 + P_1 + P_2 = 0.50000$$

Ya que el valor de F es grande ($F = 0.5$), no se puede afirmar que el orden de aplicación tenga influencia sobre la proporción de animales que responden a cada uno de los fármacos estudiados. Así, con estos datos se concluye que existe una diferencia en el poder analgésico de las drogas investigadas y que no depende del orden de aplicación.

Este ejemplo permite confirmar la importancia de proceder directamente a las dos pruebas exactas de Fisher cuando se tiene una muestra pequeña. Si se calcula la prueba de Gart ($G = 5.2$), el valor no es significativo, lo cual indica que no está presente ninguno de los dos efectos, conclusión a todas luces contraria a la obtenida al procesar los datos en forma correcta.

9. PRUEBA PARA GRADO DE CONCORDANCIA

Cuando se decide investigar acerca de la concordancia entre dos clasificaciones, como se planteó en la sección III.8, es importante determinar si el valor de Kappa estimado en la muestra indica una concordancia en la población. Es decir si los datos muestran una diferencia significativa con respecto a la concordancia que se esperaría por azar o bien si sólo se tiene una concordancia similar a la que ocurriría de manera aleatoria.

Para ejecutar la prueba se obtiene el estimador de la Kappa de Cohen y la varianza de estimación, conforme lo que se señaló en la sección III.8 y se calcula Cons_c , que será comparada con la distribución de χ^2 con un grado de libertad y el nivel de α correspondiente.

$$\text{Cons}_c = \frac{K^2}{v(K)}$$

Si Conc es significativa -mayor que 3.84 para $\alpha = 0.05$ o que 6.63 para $\alpha = 0.01$ - señala que la concordancia o la discordancia -según sea el signo de \underline{K} - se aleja significativamente de cero, lo cual indica que el acuerdo -o el desacuerdo- entre las dos formas de clasificar es mayor que lo que se esperaría por azar y que puede afirmarse, con una probabilidad α de error, que existe una concordancia real entre ambas clasificaciones.

EJEMPLO IV.8: prueba de concordancia

En el primer ejemplo de la sección III.7 sobre estimación de la concordancia se calculó un valor de 28% ($\underline{K} = 0.28$) y la varianza de este estimador fue 0.021 05 ($v(k) = 0.021\ 05$). Con esta información puede obtenerse directamente la prueba de concordancia:

$$\underline{Conc} = \frac{0.28^2}{0.021\ 05} = 3.72$$

Como Conc es menor que 3.84, la concordancia no es significativa. La concordancia observada en la prueba puede ser producto del azar y no necesariamente es una expresión de concordancia en la población. Este resultado está de acuerdo con el intervalo de confianza calculado para estos datos: -0.004; 0.564; al incluir el cero, el intervalo de confianza indica la posibilidad de que no haya concordancia entre ambas clasificaciones, lo mismo que señaló la prueba de hipótesis efectuada.

EJEMPLO IV.9: Una concordancia significativa

En el segundo ejemplo de la sección III.8 titulado "concordancia entre dos métodos de inspección" se encontró que la inspección sanitaria para detectar cisticercos basada en el brazuelo concuerda en 76% con la revisión del masetero ($\underline{K} = 0.76$); la varianza de \underline{K} fue 0.000 12 ($v(k) = 0.000\ 12$). Al obtener la estadística para probar la concordancia:

$$\underline{Conc} = \frac{0.76^2}{0.00012} = 4\ 813.33$$

Siendo Conc mayor que 6.63 la prueba indica una concordancia significativa ($p < 0.01$), por lo que puede afirmarse que los dos métodos de inspección son equivalentes. La fuerza de esta concordancia se estimó con un intervalo al 95% de confianza, y fue de 73 a 79% (0.732 36; 0.787 64). Así puede concluirse que los dos métodos de inspección tienen una concordancia significativa que varía entre 73 y 79%.

consideradas y el segundo método obtiene el promedio de la razón de momios, ponderando cada tabla de acuerdo con el tamaño de la muestra que la forma.

a) METODO DE COCHRAN

El método de Cochran (1954) es útil cuando se quiere detectar diferencias sistemáticas entre las proporciones a lo largo de las distintas tablas 2x2, sobre todo si las diferencias son pequeñas. En el ejemplo anterior, esto significa que el antibiótico en prueba produzca resultados similares en todas las regiones con respecto al producto testigo. Las hipótesis implicadas en este método son:

H₀: una o más de las diferencias ($p_1 - p_2$) no son distintas de cero o son de signos contrarios.

H_a: todas las diferencias ($p_1 - p_2$) son distintas de cero y del mismo sentido.

Por supuesto, lo anterior implica que el procedimiento no es aplicable cuando la relación entre las dos variables es notoriamente distinta en las tablas 2x2 consideradas en el estudio. De hecho, en ningún caso debe mezclarse información de tablas que muestren diferentes tendencias. En los datos del cuadro V.1 se observa que la relación entre los antibióticos es parecida en las tres regiones consideradas, de modo que es conveniente verificar si esta tendencia es significativa, es decir, si la información permite concluir que el producto B es superior al A como tratamiento en casos de mastitis.

La estadística C_n permite probar las hipótesis relacionadas con las posibles diferencias entre las proporciones de respuesta.

$$C_n = (\sum_i W_i e_i)^2 / \sum_i W_i$$

El subíndice i identifica los datos de cada una de las tablas y la sumatoria (Σ) señala la suma a lo largo de las distintas tablas en que se tiene interés. El valor de C_n se compara con la distribución de χ^2 con un grado de libertad: 3.84 para $\alpha = 0.05$ y 6.63 para $\alpha = 0.01$.

Cuadro V.1

Eficacia contra mastitis de dos antibióticos en tres regiones

Región	Antibiótico	Resultado		Total
		Curación	Ninguno	
La Laguna	A	95 (72%)	37	132
	B	102 (84%)	19	121
Querétaro	A	53 (76%)	17	70
	B	67 (84%)	13	80
Tizayuca	A	32 (62%)	20	52
	B	61 (77%)	18	79

Para resolver la ecuación de C_n se necesitan algunos cálculos previos, de cada tabla debe obtenerse:

n_1 : número de casos en el primer grupo.

n_2 : número de casos en el otro grupo.

p_1 y p_2 : Proporción de respuesta en cada grupo.

d : la diferencia entre proporciones de respuesta. $d = p_1 - p_2$

P : la proporción media. $P = (p_1 n_1 + p_2 n_2) / (n_1 + n_2)$

Q : el complemento de la proporción media. $Q = 1 - P$

$$W_i = P Q n_{1i} n_{2i} / (n_{1i} + n_{2i})$$

e_i : diferencia entre proporciones estandarizada. $e = d / (P Q)$

De acuerdo con el objetivo de este método se requiere, además, una prueba para verificar si todas las tablas son lo bastante parecidas como para pensar que son distintas manifestaciones de una misma tendencia general, llamada prueba para homogeneidad de tendencias; por supuesto que ésta sólo debe aplicarse si C_n es significativa, es decir cuando se detectan diferencias sistemáticas en las tablas. La prueba de homogeneidad de tendencias implica las siguientes hipótesis:

H_0 : las tablas presentan una tendencia homogénea.

H_a : existen diferentes tendencias en las tablas analizadas.

La prueba para homogeneidad de tendencias se calcula conforme se indica en seguida:

$$H = \sum_i W_i e_i^2 - C_n$$

El valor de H se compara con la distribución de χ^2 con tantos grados de libertad como el número de tablas evaluadas menos uno. Al final del texto aparece una tabla con los valores de la distribución de χ^2 para varios grados de libertad. Si H es menor que el valor de χ^2 cuadrada, se puede pensar que las distintas tablas manifiestan tendencias similares; por tanto, es válido calcular un solo estimador de la diferencia entre proporciones a lo largo de toda la información disponible:

$$\bar{d} = \sum_i d_i (n_{1i} + n_{2i}) / \sum_i (n_{1i} + n_{2i})$$

Pero, si el valor de H es significativo, es decir, mayor que el de χ^2 correspondiente a los grados de libertad y al nivel estipulado para α , entonces la conclusión es que la tendencia no es la misma en todas las tablas; por lo tanto, no debe

calcularse un valor único de toda la información, sino que debe manejarse cada tabla por separado tanto en el análisis como en el cálculo de la diferencia entre proporciones de respuesta.

EJEMPLO V.1: el método de Cochran

Este ejemplo se basa en los datos del cuadro V.1 que ya fue explicado. En el cuadro V.2 aparecen los cálculos preliminares a la obtención de C y de H.

Cuadro V.2

Cálculos preliminares para el método de Cochran

Tabla	n_1	n_2	p_1	p_2	d	P	Q	N	e
Laguna	132	121	.7197	.8430	-.1233	.7787	.2213	10.8790	-.7155
Queretaro	70	60	.7571	.8375	-.0804	.8000	.2000	5.9733	-.5025
Tizayuca	52	79	.6154	.7722	-.1568	.7099	.2901	6.4581	-.7614

$$C_n = \frac{(10.8790 (-.7155) + 5.9733 (-.5025) + 6.4581 (-.7614))^2}{10.8790 + 5.9733 + 6.4581} = 10.5779$$

$$H = 10.8790 (-.7155)^2 + 5.9733 (-.5025)^2 + 6.4581 (-.7614)^2 - 10.5779 = 0.2437$$

Como C_n es mayor que 6.63 ($C = 10.5779$, 6.63 es el valor crítico de χ^2 con $\alpha = 0.01$ y un grado de libertad), se rechaza la hipótesis nula y, por tanto, se infiere que el antibiótico B tiende a producir mayores porcentajes de vacas curadas que el A ($p < 0.01$). Por otro lado, ya que el valor de H es menor que el de la distribución de χ^2 con dos grados de libertad ($2 = 3 - 1$) y $\alpha = 0.05$ ($\chi^2 = 5.991$) se concluye que no hay evidencia para rechazar la hipótesis de homogeneidad de tendencias.

Conjuntando ambos resultados: el producto B fue mejor que el A y la ventaja de éste es similar en las tres poblaciones estudiadas. Por tanto, es conveniente obtener un estimador

general de la diferencia entre proporciones a lo largo de las distintas tablas. En este caso el cálculo es:

$$\bar{d} = \frac{-1.233 (132 + 121) + -.0804 (70 + 80) + -.1568 (52 + 79)}{132 + 121 + 70 + 80 + 52 + 79}$$

$$\bar{d} = -.1195$$

De modo que la diferencia entre ambos productos se estima en 11.95% , lo cual significa prácticamente un 12% de ventaja del producto B sobre el A en las tres zonas analizadas.

EXEMPLO V-2: otro caso del método de Cochran

En este ejemplo se presenta como la prueba de homogeneidad de tendencias impide mezclar estimadores de grupos con diferentes tendencias.

Dentro de una investigación sobre la eficiencia reproductiva y su relación con malformaciones del cérvix, se examinó el canal cervical en vacas Cebú (Bos indicus) y Pardo Suizo (Bos taurus) a distintas edades, los datos se resumen en el cuadro V.3. A su vez, el cuadro V.4 muestra los cálculos preliminares.

$$G^2 = \frac{(3.633 (.0131) + 2.7989 (.9184) + 4.7334 (1.8767))^2}{3.633 + 3.4284 + 4.7334} = 11.837$$

$$H = 3.633 (.0131)^2 + 2.7989 (.9184)^2 + 4.7334 (1.8767)^2 - 11.837 = 7.0885$$

En este ejemplo, el valor de G^2 es mayor que 3.84 , por lo cual puede afirmarse que existen diferencias en el porcentaje de vacas con malformaciones del cérvix entre ambas razas ($p < 0.01$). Sin embargo, el resultado de H indica que la tendencia no es homogénea. $H = 7.0885$ es significativo, ya que es mayor que 5.99 a $\alpha = 0.05$ con dos grados de libertad; por lo tanto, la diferencia en los porcentajes de malformaciones en las dos razas varía con la edad y debe utilizarse un estimador diferente para cada grupo de edad. El hecho de que la tendencia no sea homogénea hace que no se pueda conjuntar la información de las tablas 2x2, así que se emplean por separado los valores estimados

Cap. V Extensiones de la tabla 2x2

para d en cada una de las tablas (columna d del cuadro V.4); además, debe analizarse cada tabla por separado.

Cuadro V.3

Deformaciones del cervix por edades en dos razas de bovinos

Edad	Raza	Canal cervical		Total
		Deforme	Libre	
Jóvenes	Cebú	8 (24%)	25	33
	P. Suizo	12 (24%)	38	50
Adultas	Cebú	6 (17%)	30	36
	P. Suizo	9 (33%)	18	27
Viejas	Cebú	29 (48%)	32	61
	P. suizo	2 (6%)	31	33

Cuadro V.4

Datos para aplicar el método de Cochran al ejemplo V.2

Tabla	n_1	n_2	p_1	p_2	d	P	Q	w	e
Jovenes	33	50	.2424	.2400	.0024	.2410	.7590	3.6363	.0131
Adultas	36	27	.1667	.3333	.1666	.2381	.7619	2.7989	.9184
Viejas	61	33	.4754	.0606	.4148	.3298	.6702	4.7334	1.8767

b) METODO DE MANTEL-HAENZEL

En 1959, Mantel y Haenzel publicaron un método, ampliado por Mantel en 1963, muy similar al de Cochran (1963). Si el número

total de observaciones en cada una de las tablas ($n_1 + n_2$) es grande, la diferencia entre ambos métodos es inapreciable; en caso contrario, el cambio de uno a otro es notorio. La diferencia más importante entre los métodos es que el de Cochran se enfoca a la diferencia entre proporciones, mientras que el de Mantel-Haenzel analiza la razón de momios; de este modo, las hipótesis en prueba cambian a:

H_0 : no todas las razones de momios son distintas de uno o del mismo sentido.

H_a : todas las razones de momios tienen el mismo sentido con respecto al uno.

Durante el cálculo de C_n y de H (véase la sección anterior) sólo los valores W y e tienen cambios: se le resta uno al denominador de W y se multiplica e por una función del total de casos en la tabla. En las expresiones que aparecen a continuación es fácil verificar que el cambio en el resultado de la prueba es insignificante cuando la muestra es grande:

$$W = P Q n_1 n_2 / (n_1 + n_2 - 1)$$

$$e = \{ (p_1 - p_2) / (P Q) \} \{ (n_1 + n_2 - 1) / (n_1 + n_2) \}$$

Para elegir entre el método de Cochran y el de Mantel-Haenzel debe considerarse que el primero analiza las proporciones y el segundo los momios y la razón de momios, diferencia que es clara al ver que el estimador global producido por ambos es distinto. Entonces, la elección depende de cuál de los dos parámetros se quiera analizar.

Por su parte, Fleiss (1973), recomienda utilizar el procedimiento de Mantel-Haenzel si las muestras son pequeñas, es decir cuando la diferencia en los valores W y e de uno y otro métodos es notable.

Como ya se mencionó, el procedimiento de Mantel-Haenzel tiende al análisis de la razón de momios, por lo cual si C_n es significativa y H no lo es, debe estimarse la razón de momios global para todos los datos. Si se utiliza una cálculo auxiliar, las fórmulas son:

$$\bar{M} = \frac{\sum_i (f_i (P_{1i}) (1-P_{2i}))}{\sum_i (f_i (P_{1i}) (1-P_{2i}))}$$

El valor de f requerido en la fórmula anterior se calcula:

$$f = n_1 n_2 / (n_1 + n_2)$$

El cálculo anterior estima un promedio ponderado de la razón de momios, aunque se ha propuesto un estimador más sencillo, que tiende a un promedio no ponderado, el cual resulta de:

$$\bar{M} = \frac{\sum_i (a_i d_i / n_i)}{\sum_i (b_i c_i / n_i)}$$

Este último es adecuado sólo si las muestras de las tablas consideradas son semejantes; además, las complicaciones inherentes al primer estimador no son tantas como para optar por esta segunda forma, por lo que se recomienda trabajar con la primera ecuación.

EJEMPLO V.3: el método de Mantel-Haenzel

El cuadro V.5 contiene los cálculos preliminares para analizar los datos del cuadro V.1 con el método de Mantel-Haenzel. Aunque no se anotan los valores de n_1 y n_2 éstos son los mismos que en el cuadro V.2 y corresponden a los cálculos explicados en la sección anterior; r indica la razón de momios estimada en cada tabla.

$$\bar{M} = \frac{(10.9222 (-.7127) + 6.0134 (-.4992) + 6.5078 (-.7556))^2}{10.9222 + 6.0134 + 6.5078} = 10.5189$$

$$H = 10.9222(-.7127^2) + 6.0134 (-.4992^2) + 6.5078(-.7556^2) = 10.5189$$

$$H = 0.2430$$

Ya que C_n es significativa, indicando que la tendencia es sistemática y H no es significativo, señalando que las tendencias son homogéneas, las conclusiones son iguales a las enunciadas en el ejemplo V.1, es decir que la tendencia de los resultados es homogénea entre las distintas tablas 2x2 analizadas, por lo que conviene obtener un estimador global de la razón de momios.

$$\hat{M} = \frac{63.1304 (0.7197) (1 - 0.8430) + 37.3333 (0.7571) (1 - 0.8375) + 31.3586 (0.6154) (1 - 0.7722)}{63.1304 (0.8430) (1 - 0.7197) + 37.3333 (0.8375) (1 - 0.7571) + 31.3586 (0.7722) (1 - 0.6154)}$$

$$\hat{M} = 0.5066$$

El estimador de \hat{M} señala que el momio de que una vaca se cure con el antibiotico A es cercano a la mitad del momio correspondiente al producto B.

Cuadro V.5

Cálculos preliminares para el método de Mantel-Haenzel

Tabla	p_1	p_2	r	P	Q	W	e	f
Lag.	.7197	.8430	0.4783	.7787	.2213	10.9222	-.7127	63.1304
Gro.	.7571	.8375	0.6049	.8000	.2000	6.0134	-.4992	37.3333
Tiz.	.6154	.7722	0.4721	.7099	.2901	6.5078	-.7556	31.3588

EJEMPLO V.4: el método de Mantel-Haenzel con muestras chicas

Al examinar la fertilidad en dos razas de cerdos, se quiere investigar si el comportamiento de éstas es similar al compararlas bajo tres sistemas de producción porcina. A continuación se muestran los datos del caso (Cuadro V.6). Los cálculos preliminares aparecen en el cuadro V.7.

$$G_1 = \frac{(1.0807 (0.7931) + 0.9111 (-1.0354) + 1.2097 (0.7349))^2}{1.0807 + 0.9111 + 1.2097} = 0.2013$$

Como G_1 es menor que 3.84 (valor de χ^2 para $\alpha = 0.05$ con un grado de libertad), no se puede pensar que todas las tablas contengan razones de momios en el mismo sentido y distintas de uno; lo cual señala que no hay una tendencia sistemática en las distintas tablas. Por tal motivo ya no es necesario aplicar la prueba de homogeneidad de tendencias. Es claro que si la tendencia no es sistemática, mucho menos será homogénea.

Cuadro V.6

Fertilidad de dos razas de cerdos en tres sistemas de producción

Sistema de producción	Raza	Fertilidad		Total
		Gestante	Vacia	
Intensivo en jaula	Yorkshire	18 (90%)	2	20
	Hampshire	12 (80%)	3	15
Intensivo en corral	Yorkshire	15 (83%)	3	18
	Hampshire	16 (94%)	1	17
Semiextensivo	Yorkshire	11 (85%)	2	13
	Hampshire	10 (71%)	4	14

3. MAS DE DOS PORCENTAJES A COMPARAR: TABLA Cx2

Con frecuencia el interés acerca de un caso específico no se circunscribe a comparar porcentajes entre dos grupos, sino que se requiere incluir tres o más muestras en la prueba. En un estudio de este tipo, los resultados aparecen resumidos en una tabla Cx2. Una de las variables ya no es binaria, como todas las referidas anteriormente, ahora es nominal, ordinal e incluso de tipo cuantitativo (en el Capítulo 1 se describen los tipos de

variable), es la que distingue a los C grupos que se quiere comparar en cuanto a la tasa de respuesta en la variable de dos niveles.

Si bien todos los casos son semejantes, el tipo de variable que clasifica los C grupos incluye una forma especial de enfocar el análisis, así, no se comparan de la misma manera los porcentajes de respuesta entre los niveles de una variable nominal -como la raza en perros-, que entre los de una ordinal -como la edad: joven, adulto y viejo-, o en los distintos niveles de una variable cuantitativa -como la dosis de un fármaco: 150, 200, 400 y 1 600 mg/kg-. Los métodos que se describen más adelante permiten tratar cada caso en la forma adecuada.

Cuadro V.7

Cálculos previos a χ^2 y H para el ejemplo V.4

Tabla	p_1	p_2	r	P	Q	w	e	f
I.jau.	.9000	.8000	2.2500	.8571	.1429	1.0807	0.7931	8.5714
I.cor.	.8333	.9412	0.3125	.8857	.1143	0.9111	-1.0354	8.7429
Senie.	.8462	.7143	2.2000	.7778	.2222	1.2097	0.7349	6.7407

Para evaluar si los porcentajes difieren en los distintos grupos se aplica la prueba de χ^2 para independencia. Aunque es la misma prueba que se describió para la tabla 2x2 no puede utilizarse la fórmula reducida, por lo cual los cálculos varían en apariencia.

Si se utiliza la nomenclatura definida en el cuadro V.8, los valores esperados para χ^2 se obtienen con base en los totales marginales y en el número total de casos de la tabla Cx2:

$$E_{ij} = \frac{r_i \cdot t_j}{N}$$

El subíndice i señala a qué renglón pertenece el valor y el j indica la columna correspondiente.

Cuadro V.8

La tabla Cx2

		Variable de respuesta		
		Si	No	Total
Variable de clasificación	1	O_{11}	O_{12}	r_1
	2	O_{21}	O_{22}	r_2
	3	O_{31}	O_{32}	r_3

	C	O_{c1}	O_{c2}	r_c
Total		t_1	t_2	N

Los valores esperados de acuerdo con la hipótesis de igualdad de proporciones se comparan con las frecuencias observadas en cada grupo mediante la expresión expuesta a continuación:

$$\chi^2_c = \sum \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}}$$

Esta expresión indica que debe sumarse la comparación para los dos grupos de la variable binaria a lo largo de los C niveles de la otra variable. El resultado se confronta con la distribución de χ^2 para el valor de α seleccionado y con C-1 grados de libertad, esto significa que los grados de libertad son uno menos que el número de proporciones a comparar.

Si la prueba de χ^2 proporciona un valor significativo debe hacerse la partición de la tabla Cx2 para dilucidar entre cuáles de los porcentajes hay diferencia significativa. Aun cuando esta prueba no sea significativa puede convenir la partición, ya que es un método más potente para detectar diferencias específicas.

EJEMPLO V.5: la prueba de χ^2 en una tabla Cx2

Durante una investigación sobre las dificultades en el parto en vacas lecheras, se registraron los partos ocurridos en cuatro establos, en cada caso se anotó si hubo distocia; los datos aparecen en el cuadro V.9. Los valores esperados para cada una de las frecuencias observadas se resumen en el cuadro V.10.

Cuadro V.9

Frecuencia de distocias en cuatro establos lecheros

Establo	Parto		Total
	Distócico	Normal	
Uno	136 (6%)	1962	2098
Dos	185 (28%)	485	670
Tres	16 (9%)	165	181
Cuatro	21 (24%)	68	89
Total	358 (12%)	2680	3038

Datos de Bacha (1985)

Como ejemplo del cálculo de las frecuencias esperadas se muestra el desarrollo en cuanto a distocias en el primer establo y de partos normales en el tercero:

$$E_{11} = \frac{(2098)(358)}{3038} = 247.2298$$

$$E_{32} = \frac{(181)(2680)}{3038} = 159.6708$$

Cuadro V.10

Frecuencias esperadas conforme la hipótesis de independencia*

Establo	Parto	
	Distóxico	Normal
Uno	247.2298	1850.7702
Dos	78.9533	591.0467
Tres	21.3292	159.6708
Cuatro	10.4878	78.5122

*Esta hipótesis implica la igualdad entre establos para la proporción de distóxicos

Para obtener el resultado de la prueba sólo falta comparar las frecuencias observadas (cuadro V.9) con los valores esperados (cuadro V.10) mediante la fórmula ya descrita:

$$\chi^2_{pr} = \frac{(136 - 247.2298)^2}{247.2298} + \frac{(1962 - 1850.7702)^2}{1850.7702} +$$

$$\frac{(185 - 78.9533)^2}{78.9533} + \frac{(485 - 591.0467)^2}{591.0467} +$$

$$\frac{(16 - 21.3292)^2}{21.3292} + \frac{(165 - 159.6708)^2}{159.6708} +$$

$$\frac{(21 - 10.4878)^2}{10.4878} + \frac{(65 - 78.5122)^2}{78.5122} =$$

$$\chi^2_{pr} = 50.0428 + 6.6848 + 142.4374 + 91.0271 +$$

$$1.3315 + 0.1779 + 10.5367 + 1.4075 = 231.6457$$

El valor de la distribución χ^2 con tres grados de libertad y $\alpha = 0.01$ es 11.35, por lo que el número calculado (231.6457) es significativo. Por tanto, hay una diferencia significativa entre las proporciones de distocias en los ranchos considerados ($p < 0.01$). Con esta conclusión queda aún por resolver entre cuáles establos, en forma específica, se presentan diferencias significativas. En las siguientes secciones se presentan los métodos para contestar esta pregunta, mediante el análisis de la tabla $C \times 2$ de acuerdo con el tipo de variable que distingue los C grupos.

a) VARIABLE NOMINAL DE C NIVELES

El caso más general es cuando los grupos incluidos en la comparación de proporciones se diferencian con base en una variable nominal; por ejemplo, los porcentajes de enfermedad en varias regiones, las tasas de mortalidad en diez granjas avícolas o la proporción de respuesta a cuatro tratamientos distintos. La generalidad de los métodos aplicables a esta situación estriba en que también pueden aplicarse cuando la variable de C niveles es ordinal o cuantitativa; aunque la prueba obtenida es menos potente que las que se han desarrollado para cada caso específico.

Lancaster (1950) e Irwin (1949) demostraron que el valor de χ^2 calculado con el procedimiento de la sección anterior puede dividirse en tantos componentes independientes como grados de libertad tenga; sin embargo, este método no siempre permite todas las comparaciones que el investigador requiere. Para confrontar todas las proporciones se han propuesto varios sistemas, a continuación se describen dos, uno para comparar todos los pares de proporciones entre sí y otro para comparar la proporciones con una que sirve como testigo o punto de referencia.

COMPARAR TODOS LOS PARES DE PROPORCIONES

En muchas ocasiones se desea comparar todos los porcentajes, para esto es necesario formar las tablas 2×2 que permitan confrontar entre sí a todos los pares posibles de proporciones. El cuadro V.11 muestra las tablas 2×2 para comparar los porcentajes del ejemplo V.5 (cuadro V.9).

Puede verse que las tablas incluidas en el cuadro V.11 presentan todas las combinaciones posibles de las proporciones

entre los diferentes establos. Lo único que falta es aplicar las pruebas descritas en el capítulo IV para las tablas 2x2. En este caso, lo indicado es la prueba de χ^2 para independencia.

Sin embargo, efectuar una serie de pruebas presenta un problema: La probabilidad de error tipo I (α) se mantiene solo para una, y el método sugerido exige varias pruebas sobre los mismos datos. Si se realiza más de una prueba estadística, la probabilidad de rechazar alguna hipótesis nula verdadera aumenta más allá del valor nominal de α ; cuantas más se hagan, mayor será la probabilidad de cometer un error tipo I en alguna de ellas.

Cuadro V. 11

Tablas 2x2 para comparación de porcentajes uno a uno

Establo	Parto		Establo	Parto	
	Distóxico	Normal		Distóxico	Normal
(a) Uno	136	1962	(b) Uno	136	1962
Dos	185	485	Tres	16	165
(c) Uno	136	1962	(d) Dos	185	485
Cuatro	21	68	Tres	16	165
(e) Dos	185	485	(f) Tres	16	165
Cuatro	21	68	Cuatro	21	68

La probabilidad de error por prueba se indica en α , pero cuando se efectúan varias en una sola investigación también es importante la probabilidad de error por experimento, es decir, la probabilidad de que por lo menos una de las inferencias derivadas de un estudio específico esté equivocada.

El problema puede plantearse de esta manera: si durante un estudio las conclusiones se basan en una prueba con $\alpha = 0.05$ puede esperarse que haya error en uno de cada veinte casos ($1/20 = 0.05$), si la hipótesis nula es verdadera. Por otro lado, si se hacen 60 pruebas para un solo estudio y se usa $\alpha = 0.05$ es casi seguro que habrá alguna inferencia errónea; ya que aun cuando la

probabilidad de error en cada prueba sigue siendo uno en veinte (0.05) la tasa global de error llega casi a uno, lo que resulta inadmisibles desde cualquier punto de vista.

Se han propuesto varios métodos para controlar el error global cuando se hacen pruebas simultáneas; uno de los más prácticos es el descrito por Leach (1979). Con base en la demostración de que la probabilidad global de error nunca rebasa la suma de las probabilidades de error en cada prueba, para mantener la probabilidad de error global por debajo del valor que se quiere utilizar para α basta dividirla entre el número de comparaciones a realizar. De tal modo las confrontaciones individuales se basan en un valor de α ajustado a la cantidad de pruebas:

$$\alpha_a = \frac{\alpha}{K}$$

Donde K indica el número de comparaciones a realizar.

El procedimiento es bastante conservador y aunque resulta razonablemente preciso para un número reducido de comparaciones, no es muy acertado cuando se considera una gran cantidad de pruebas. Por conservador se entiende que exagera al controlar la probabilidad de error tipo I, manteniendo α por abajo del valor nominal que se está aplicando; sin embargo, el método es adecuado para la mayor parte de los casos que se presentan.

EJEMPLO V.6: comparaciones múltiples de proporciones

Para ejemplificar este método se continuará con el análisis iniciado en el ejemplo V.5. Para investigar entre que ranchos existen diferentes tasas de distocias, el primer paso es la formación de todas las tablas 2x2, lo cual se hizo en el cuadro V.11, a continuación debe calcularse el valor de α ajustado al número de comparaciones:

$$\alpha_a = \frac{\alpha}{K} = \frac{\alpha}{6} = \frac{0.05}{6} = 0.008$$

Como el valor ajustado es 0.008 y sabiendo que es un método conservador, puede utilizarse 0.01 al consultar las tablas de X². El resultado con un grado de libertad y $\alpha = 0.01$ es 6.63, contra

éste deberá compararse la χ^2 obtenida en cada prueba. El cuadro V.12 contiene los valores calculados para χ^2 en cada una de los pares a comparar, con anotación de la conclusión correspondiente.

Cuadro V.12

Resultados de comparar cada par de proporciones

Comparación	Porcentajes	$\chi^2 =$	Conclusión
Uno con Dos	6 y 28	219.09	Distintos
Uno con Tres	6 y 9	1.13	Similares
Uno con Cuatro	6 y 24	35.00	Distintos
Dos con Tres	28 y 9	26.80	Distintos
Dos con Cuatro	28 y 24	0.45	Similares
Tres con Cuatro	9 y 24	9.77	Distintos

*Para un valor global de $\alpha = 0.05$ se comparé con 6.63

En conclusión, existe diferencia significativa entre la tasa de distocias en los establos uno y tres con respecto a los ranchos dos y cuatro; a su vez, cada par de éstos tiene una proporción similar de partos difíciles. En forma análoga a las comparaciones múltiples de promedios, para expresar los resultados pueden ordenarse los porcentajes de acuerdo con su magnitud y subrayar juntos a los que son similares, señalando así los grupos que no difieren significativamente.

Establo	Uno	Tres	Cuatro	Dos
Porcentaje	<u>6</u>	<u>9</u>	<u>24</u>	<u>28</u>

Otra forma de hacerlo es marcar con una letra los porcentajes similares, de modo que las similitudes y diferencias pueden reconocerse verificando si dos estimadores en particular están marcadas o no con la misma letra.

Establo	Uno	Dos	Tres	Cuatro
Porcentaje	6 ^a	28 ^b	9 ^a	24 ^a

COMPARACION CON UN TESTIGO

En ocasiones no se requiere comparar entre sí todos los porcentajes, sino que se cuenta con algún grupo de referencia, conocido como grupo testigo; éste puede ser un producto de efecto ya conocido, una muestra sin ningún tratamiento, el proceso que se encuentra en uso o, de cualquier otra manera, un punto de referencia para evaluar el resto de los lotes en prueba. En estos casos, lo importante es contrastar cada una de las proporciones de respuesta con la del lote testigo.

Brunden (1975) propuso un sistema para hacer estas comparaciones: se forman todas las tablas 2x2 que combinan los datos de cada uno de los lotes en prueba con los del grupo testigo y cada una de las tablas se evalúa de acuerdo con lo descrito en el capítulo anterior. Otro cambio importante con respecto al método anterior consiste en que el ajuste de α es diferente. Brunden (1975) recomienda que el ajuste se haga con la siguiente ecuación:

$$\alpha = \frac{\alpha}{2(C - 1)}$$

Donde: C es el número de grupos que integran la tabla 2x2

El ajuste producido es muy similar al de la sección anterior cuando C, el número de grupos que integran la clasificación, es menor que cinco, pero difiere sensiblemente si hay seis o más grupos a comparar.

EJEMPLO V.7: Comparación de proporciones con un testigo

Durante un análisis de la productividad en cerdos, se requiere comparar tres razas con respecto a la raza Landrace que se tiene en una granja. Una parte de la evaluación se enfoca a la proporción de hembras jóvenes que alcanzan su primer parto, es decir, las que no se desechan a causa de fallas reproductivas o físicas. Por tanto no se busca contrastar los porcentajes de reemplazos útiles entre todas las razas, sino que sólo se requiere conocer cuál es su comportamiento con respecto a la que existe en la granja. Los datos aparecen en el cuadro V.13.

Cuadro V.13

Tasa de hembras de reemplazo útiles en cuatro razas de cerdos

Raza	Parieron	Desechadas	Total
Landrace	75 (96%)	3	78
Yorkshire	97 (90%)	11	108
Chester White	65 (81%)	15	80
Berkshire	38 (70%)	16	54
Total	275	45	320

El valor de la prueba de χ^2 es 20.3631; con tres grados de libertad y $\alpha = 0.01$, la tabla de χ^2 indica un valor de 11.35, por lo que el resultado señala que existe una diferencia significativa entre los porcentajes de hembras de reemplazo que llegan al primer parto en cada una de las razas consideradas.

Para hacer las comparaciones múltiples se requiere ajustar el valor de α al número de comparaciones del testigo -raza que ya se tiene- con los lotes en prueba -razas nuevas-, al efecto se aplica la fórmula descrita. Es conveniente usar como valor de α el mismo que se empleó en la prueba de χ^2 , y que en este ejemplo fue $\alpha = 0.01$.

$$\alpha_m = \frac{0.01}{2(4-1)} = 0.0017$$

La probabilidad de error tipo I en la tabla de la distribución χ^2 más próxima al valor ajustado de α es 0.001, por lo cual, para asegurar una tasa global de error menor a 0.01 se tomó el número correspondiente a un grado de libertad y $\alpha = 0.001$, que es 10.83.

El cuadro V.14 contiene las tablas 2x2 que permiten comparar la tasa de efectividad en el reemplazo de cada raza con respecto a la Landrace. Puede verse que sólo las hembras Berkshire mostraron un porcentaje distinto al grupo testigo. Si la comparación entre Chester White y Landrace se hubiera realizado sin el ajuste debido resultaría significativa, lo cual no es

correcto ya que la probabilidad global de error rebasaría el nivel nominal de α .

Cuadro V.14

Tablas 2x2 para comparar las proporciones con el testigo

	Raza	Parieron	Desechadas	X^2	Conclusión
(a)	York	97	11	1.78	Similares
	Landrace	75	3		
(b)	Chester White	65	15	7.28	Similares
	Landrace	75	3		
(c)	Berkshire	38	16	15.19	Diferentes
	Landrace	75	3		

* Para asegurar una probabilidad global de error de 0.01 se comparó con 10.83, valor correspondiente a $\alpha = 0.001$ y un grado de libertad.

b) C GRUPOS CON UN ORDEN PROPIO

Cuando la variable que distingue a los diferentes grupos es de tipo ordinal, es decir que indica un cambio de grado en alguna cualidad, y se desea probar si los porcentajes de respuesta están relacionados con ese orden propio de la variable de clasificación, entonces es útil aplicar la prueba de Jonckheere. Aunque ésta es más general, aquí sólo se presenta un caso particular: el que se utiliza para comparar varios grupos con respecto a una respuesta binomial.

Si la variable que distingue los lotes en prueba es ordinal pero no se quiere probar ninguna hipótesis acerca de la relación entre las tasas de respuesta y el orden de la agrupación, es más conveniente usar los métodos descritos en el apartado anterior (V.3,a) para una variable nominal de C niveles, ya que no implican una relación predicha entre el orden de la variable y la probabilidad de respuesta.

Como ejemplo del tipo de problemas que puede analizarse con la prueba de Jonckheere en la versión que se muestra aquí, puede

plantearse el siguiente caso: en un centro de trasplante de embriones se investiga si el grado de madurez del embrión está relacionado con la probabilidad de que se implante en la hembra nodriza. Para estudiarlo toman un grupo de embriones clasificados de acuerdo a su desarrollo -embrión joven, mórula, blástula y blastocistos- y los trasplantan, registrando después cuántos de cada grupo prosiguieron la gestación.

Es importante destacar que esta prueba sólo debe aplicarse si se quiere verificar la existencia de una relación entre el orden establecido en la variable de clasificación y la tasa de respuesta en cada grupo. Por ejemplo, si en un estudio epidemiológico sobre brucelosis se analiza un centro de cría de vacas lecheras, registrando la incidencia de la enfermedad en cada etapa del centro: becerras lactantes, desarrollo I, desarrollo II y vaquillas gestantes; pero el investigador no busca probar una relación de ese orden con la incidencia de brucelosis, aunque la clasificación se basa en una variable ordinal, resultará más adecuado aplicar los métodos descritos para variables nominales.

La prueba de Jonckheere se basa en el conteo de los casos que tienen una clasificación mayor en la variable ordinal y diferente en la variable binomial. Para cada celda de la tabla Cx2 se cuentan los casos de las celdas inferiores en la otra columna; el resultado se multiplica por la frecuencia de la celda. Las ecuaciones que se aplican son:

$$S = \sum_i f_{1i} (f_{i \text{ abajo der.}}) - \sum_i f_{i2} (f_{i \text{ abajo izq.}})$$

$$S_a = S \pm \frac{2 \sum_i N_i - r_1 - r_c}{2 (C - 1)}$$

El símbolo \pm en el ajuste de S indica que a los valores positivos de S debe restárseles el factor de ajuste, mientras que a los negativos debe sumárseles. De esta forma, en todos los casos el ajuste acerca el valor de S a cero. La varianza de S se calcula:

$$V = \frac{t_1 t_2 (N^3 - \sum_i r_i^3)}{3 N (N - 1)}$$

A partir de los valores anteriores se calcula la estadística de prueba:

$$J = \frac{S_{\alpha}^2}{V}$$

J se compara con la distribución de χ^2 con un grado de libertad, por lo que si es mayor que 3.84 puede concluirse, con una probabilidad de error tipo I de 0.05 ($\alpha = 0.05$), que existe una asociación significativa entre las tasas de respuesta y el orden de la variable de clasificación. Si J supera el valor de 6.63 esta afirmación sólo tiene 0.01 de posibilidad de error.

EJEMPLO V.8: la prueba de Jonckheere sobre la tabla 2x2

Con el propósito de estudiar la relación de la edad con la tasa de fracturas diagnosticadas en perros se analizaron los libros de un consultorio radiológico, registrando la cantidad de animales que presentaban fracturas en cada grupo de edad. Los datos se muestran en el cuadro V.15

Cuadro V.15

Frecuencia de fracturas en perros de distintas edades*

Edad	Fracturas	Otros	Total
Cachorro	70 (29%)	174	244
Joven	38 (26%)	107	145
Adulto	32 (12%)	244	276
Senil	10 (5%)	202	212
Total	150	727	877

* Datos de Correa (1985)

El cuadro V.16 contiene los casos que están abajo a la derecha (parte a) y abajo a la izquierda (parte b) para cada una

de las celdas de la tabla Cx2 que se está analizando. Puede constatarse en el primer caso que a todas las celdas del lado derecho les corresponde una cuenta de cero y lo mismo sucede en el otro para las celdas del lado izquierdo, a su vez, el último renglón no tiene frecuencias abajo de él y, por tal motivo, todas esas celdas se ignoran al calcular S.

Cuadro V.16

Sumas de frecuencias necesarias para calcular S

Edad	Sumas de frecuencias:			
	Abajo a la derecha		Abajo a la izquierda	
	Fractura	Otros	Fractura	Otros
Cachorro	553	0	0	80
Joven	446	0	0	42
Adulto	202	0	0	10
Senil	0	0	0	0
	(a)		(b)	

$$S = (553(70) + 446(38) + 202(32)) - (80(174) + 42(167) + 10(244))$$

$$S = 621\ 22 - 208\ 54 = 412\ 68$$

Como S es positiva el ajuste consiste en restarle el resultado del lado derecho de la fórmula; recuerdese que el ajuste se resta de los valores positivos de S y se suma a los negativos.

$$S_a = S - \left(\frac{2(877) - 244 - 212}{2(4 - 1)} \right) = 412\ 68 - 216 = 410\ 52$$

$$V = \frac{150(77) \left[\frac{150(77) - (24^2 + 14^2 + 27^2 + 21^2)}{150} \right]}{3(77)(77-1)} = 29.638150$$

$$J = \frac{S_{.j}^2}{V} = \frac{41.052^2}{29.638150} = \frac{1.685.267.000}{29.638150} = 56.86$$

El resultado de J es mayor que el valor de χ^2 con un grado de libertad aun para un valor de α tan pequeño como 0.001, por lo que se concluye que existe una relación significativa de la variable ordinal -edad- con la respuesta -proporción de fracturas-. Esto indica que la probabilidad de respuesta tiende a cambiar de un grupo a otro de acuerdo con el orden entre ellos; en este ejemplo la tendencia es a una disminución en la tasa de fracturas al aumentar la edad.

c) VARIABLE CUANTITATIVA EN C PUNTOS

Un caso bastante común es cuando se toman algunos niveles específicos de una variable cuantitativa y se analiza la tasa de respuesta en cada uno de ellos. Por ejemplo, cuando un fisiólogo necesita determinar la mortalidad causada por cierto producto, aplica algunas dosis específicas a lotes de animales experimentales y registra la proporción de muertes en cada uno.

En este tipo de problemas no sólo se busca inferir si el porcentaje de respuesta es distinto en las diferentes dosis, sino que se investiga si existe una relación del nivel de la variable cuantitativa -medido en forma numérica- y la proporción de respuesta.

En ocasiones no se utilizan puntos específicos de una variable cuantitativa, sino que se conjuntan las observaciones de un intervalo definido; en este caso el modelo puede utilizarse si se considera que en cada grupo todas las observaciones corresponden al punto medio del rango en que se encuentran.

Este sería el caso si un veterinario del hipódromo decide averiguar la relación del peso de los caballos de salto con la probabilidad de que sufran un accidente en la pista y agrupa los caballos de acuerdo con el peso y registra el número de accidentes en cada lote. Para analizar la proporción de accidentes puede asignar a todos los animales de un grupo el peso medio del rango que los incluye; en el ejemplo v.10 se muestra un problema semejante.

Para describir el cambio en la probabilidad de respuesta al modificar el nivel de la variable cuantitativa, se ha propuesto como modelo la ecuación de una recta:

$$p_i = \Gamma + \beta X_i$$

Donde X_i representa la variable cuantitativa y X_i señala el i -ésimo nivel de la variable; p_i es la probabilidad de la respuesta correspondiente a ese valor de la variable cuantitativa X_i . Γ y β son los parámetros que describen la forma en que se asocian los niveles de la variable cuantitativa y la medida de respuesta. β es la pendiente y puede interpretarse como el cambio en la probabilidad de respuesta por cada unidad de incremento en la variable X .

Este modelo es útil en muchos casos, pero como presenta algunos defectos importantes; el más claro de ellos es que puede indicar valores por debajo de cero o por encima de uno, lo que está fuera del intervalo de cero a uno en el que se define siempre la probabilidad; sin embargo, en muchos casos este procedimiento puede ser aceptado por la sencillez de aplicación y porque en un buen número de casos describe bien la relación entre ambas variables. Cuando la estimación obtenida mediante el modelo está fuera de los límites de la probabilidad, el valor se trunca en uno o en cero según sea el caso.

Cuando este modelo no pueda aplicarse, tiene una variedad de alternativas en otras técnicas enfocadas al análisis de variables de respuesta binarias, como los análisis logit y próbit. Siempre es posible encontrar una técnica adecuada para una investigación específica, pero no siempre una técnica sencilla de aplicar o de interpretar, de ahí que se planteen el que se describe en seguida.

El uso del modelo se basa en la estimación de los parámetros que lo forman. Para obtener estimadores de Γ y β se aplican las siguientes fórmulas:

$$\beta = \frac{F_{xy}}{F_x}$$

$$\Gamma = t_1 / N - \beta / N I$$

La varianza del estimador de β es:

$$V(E) = (1 / N) (F_v / F_w)$$

Los cálculos preliminares se obtienen como una relación de los valores asociados a los distintos niveles de X y las frecuencias en cada celda. En la variable binomial se utiliza el código (1,0) para indicar cada uno de los niveles, procurando asociar el uno a la respuesta de interés y el cero a la otra. En las fórmulas descritas a continuación se considera que la respuesta de interés se colocó en la columna derecha de la tabla 2x2.

$$W = \sum_i X_i r_i$$

$$F_{wy} = \sum_l D_{li} X_i - W (t_i) / N$$

$$F_w = \sum_i X_i^2 r_i - W^2 / N$$

$$F_v = t_1 - t_1^2 / N$$

Tan importante como obtener los estimadores para el modelo propuesto, es determinar si las estimaciones indican una asociación significativa entre la variable cuantitativa y el porcentaje de respuesta, para lo cual se obtiene una estadística de pruebas:

$$L = E^2 / V(E)$$

Si L es mayor que el valor de tablas en la distribución de χ^2 para un grado de libertad y el nivel de α estipulado, entonces existe una asociación lineal significativa y se concluye que la proporción de respuesta cambia en relación con los cambios en el nivel de la variable cuantitativa analizada.

Para estimar la importancia de las desviaciones con respecto a una asociación lineal se aplica otra prueba, la cual permite inferir si hay alguna tendencia distinta a la línea recta, como sería una relación curvilínea o con cambios en la tasa de

respuesta que varían al moverse en distintas zonas de la variable cualitativa.

La prueba consiste en restar L del valor total calculado para χ^2 utilizando el método descrito al inicio de la sección V.3:

$$D = \chi^2_c - L$$

El valor (D) se compara con la distribución de χ^2 con $C - 2$ grados de libertad, si es mayor que el de tablas se concluye que existe una desviación significativa con respecto a la tendencia lineal.

EJEMPLO V.9: respuesta binomial a un factor cuantitativo

Durante una investigación acerca de los efectos letales del cloranfenicol se escogieron seis dosis del producto. Se aplicó cada una a un lote de diez ratones, midiendo la mortalidad en cada grupo. Los datos se encuentran en el cuadro V.17.

El resultado de la prueba de χ^2 es de 36.3801 y es significativo para los cinco grados de libertad que corresponden al caso y $\alpha = 0.001$, indicando que la tasa de mortalidad varía con la dosis del cloranfenicol.

Los cálculos necesarios para la prueba de linealidad de la tendencia, lo mismo que para la prueba de desviaciones de la linealidad se muestran a continuación:

$$N = .66 (10) + 1.32 (10) + 1.65 (10) + 2 (10) + 2.64 (10) + 3 (10) = 112.7$$

$$F_{xy} = 0 (.66) + 1 (1.32) + 6 (1.65) + 8 (2) + 9 (2.64) + 10 (3) = 112.7 (34) / 60$$

$$F_{xy} = 17.1167$$

$$F_x = 10 (.66^2) + 10 (1.32^2) + 10 (1.65^2) + 10 (2^2) + 10 (2.64^2) + 10 (3^2) = 112.7 / 60$$

$$F_x = 37.0128$$

$$F_y = 34 - 34^2 / 60 = 14.7333$$

Cuadro V.17

Efecto letal del cloranfenicol en ratones*

Dosis (g / kg)	Muertos (Código 1)	Vivos (Código 0)	Total
0.66	0	10	10
1.32	1	9	10
1.65	6	4	10
2.00	8	2	10
2.64	9	1	10
3.00	10	0	10
Total	34	26	60

*Datos de González (1983)

Por lo tanto, los estimadores del modelo son:

$$\hat{\epsilon} = 17.1167 / 37.0128 = 0.4625$$

$$\hat{\Gamma} = 34 / 60 - 112.7 / 60 (0.4625) = -0.3025$$

$$V(\hat{\epsilon}) = 1 / 60 (14.7333 / 37.0128) = 0.006634$$

Antes de aplicar el modelo y de aceptarlo como útil; conviene hacer las pruebas sobre su validez:

Cuadro V.18

Tasas de respuestas observadas y estimadas mediante el modelo

Dosis (g / kg)	Mortalidad		Diferencia
	Observada	Estimada con el modelo	
0.66	0.0	$-.3025 + 0.4625 (0.66) = 0.0028$	0.28%
1.32	0.1	$-.3025 + 0.4625 (1.32) = 0.3080$	20.88%
1.65	0.6	$-.3025 + 0.4625 (1.65) = 0.4606$	-13.94%
2.00	0.8	$-.3025 + 0.4625 (2.00) = 0.6225$	-17.75%
2.64	0.9	$-.3025 + 0.4625 (2.64) = 0.9185$	1.85%
3.00	1.0	$-.3025 + 0.4625 (3.00) = 1.0000^*$	8.50%

*Este valor era 1.0850 y se truncó en 1 para no exceder los límites de la probabilidad, como se explicó antes.

$$L = 0.4625^2 / 0.006634 = 32.2439$$

Como L es mayor que 6.63, que es el valor de χ^2 con un grado de libertad y $\alpha = 0.01$, hay una relación lineal significativa entre la dosis de cloranfenicol y la mortalidad. Ahora sólo falta obtener D, que es la prueba para desviaciones con respecto a la linealidad:

$$D = 36.3801 - 32.2439 = 4.1360$$

El valor de la distribución de χ^2 con cuatro (6 - 2) grados de libertad y $\alpha = 0.05$ es 9.49, como D es menor, se concluye que no existen diferencias significativas con respecto a la linealidad. Sin embargo, parece haber una tendencia a estas desviaciones, en la dosis bajas el modelo sobreestima la mortalidad, después la subestima con los valores medios para volver a dar estimaciones excedidas en las mayores dosis analizadas. El cuadro V.18 muestra los valores estimados mediante el modelo.

Al parecer, aun cuando hay una aproximación a la recta, existe cierta tendencia hacia una relación curvilínea entre la dosis y la respuesta. Para interpretar correctamente estos casos

convienen preparar una gráfica como la V.1, donde pueden observarse las diferencias entre la línea estimada por el modelo y las tasas de respuesta observadas. Cuando se aprecia una tendencia clara a separarse de la relación lineal y D no es significativa, podría ser interesante incrementar el tamaño de muestra y repetir el experimento.

EJEMPLO V.10: variable cuantitativa dividida en intervalos

Como parte de una encuesta acerca de las condiciones de explotación de los bovinos en el trópico, se planteó investigar la relación entre el tamaño del hato y la desparasitación del ganado. Los ranchos se agruparon de acuerdo con el número de cabezas que tienen y en cada intervalo se contó cuántos aplican desparasitantes a sus animales. Los datos para cada grupo de ranchos aparecen en el cuadro V.19.

Cuadro V.19

Relación del tamaño del hato con la desparasitación del ganado*

Número de cabezas	Desparasita	No lo hace	Total
1-10	39	161	200
11-20	42	58	100
21-40	35	45	80
41-80	36	34	70
81-160	29	21	50
161-320	17	13	30
321-600	6	2	8
601-2000	2	2	4
Total	206	336	542

*Datos de Barajas (1985)

El valor de χ^2 es 54.0278, lo cual indica una diferencia significativa de la proporción de ganaderos que desparasita entre los diferentes grupos de ranchos. Nótese que el valor de tablas para $\alpha = 0.01$ y siete grados de libertad es 16.475 que es menor que el calculado.

Ahora se requiere analizar la relación entre el tamaño del hato y la probabilidad de que practiquen la desparasitación, para lograrlo se utiliza como nivel de X el valor medio de los dos límites de cada intervalo. Para el primer caso es $(1 + 10) / 2 = 5.5$, para el segundo $(11 + 20) / 2 = 15.5$ y así sucesivamente. Todos los valores medios se encuentran en la expresión de W .

$$W = 5.5 (200) + 15.5 (1100) + 30.5 (800) + 60.5 (700) + 120.5 (500) + 240.5 (300) + 460.5 (80) + 1300.5 (40) =$$

$$W = 314\ 51$$

$$F_{wy} = 5104.2989$$

$$F_w = 9\ 458\ 587.0850$$

$$F_y = 127.7048$$

$$\Gamma = 5\ 104.2989 / 9\ 458\ 587.0850 = 0.000\ 54$$

$$\epsilon = 206 / 542 - 31\ 451 / 542 (0.000\ 54) = 0.348\ 74$$

$$V (\Gamma) = (1 / 504) (127.7048 / 9\ 458\ 587.0850) = 0.000\ 000\ 027$$

La prueba de linealidad de tendencia.

$$L = 0.000\ 54^2 / 0.000\ 000\ 027 = 10.8$$

El resultado es significativo, por tanto existe una relación lineal entre la proporción de ganaderos que utilizan

parasitocidas y el número de cabezas que poseen. La prueba de desviaciones alrededor de la tendencia lineal es:

$$D = 54.0278 - 10.8 = 43.2278$$

Como χ^2 con 0.01 y cuatro grados de libertad es 18.467 también es significativa, lo cual indica que una relación rectilínea no es lo bastante buena como para describir la relación entre ambas variables por lo que se requiere un análisis detallado de esas diferencias; utilizando métodos como lógit y próbit puede cubrirse esta necesidad.

1. INTRODUCCION

Una de las preguntas iniciales en casi toda investigación es ¿cuántos casos se deben incluir en la muestra, para que sea confiable? En este capítulo se aborda el problema del tamaño de muestra.

Al utilizar una prueba estadística debe recordarse que el resultado puede conducir a dos tipos de error (descritos en el Cap. IV). La lógica en que se fundamentan las pruebas permite controlar la probabilidad del error tipo I (α : alfa), pero la probabilidad de cometer un error tipo II (β : beta) depende del valor de α , del número de observaciones en la muestra y de la magnitud de la diferencia entre los parámetros (diferencia entre las poblaciones o diferencia real). La dependencia entre α y β se expuso al explicar la lógica de las pruebas estadísticas y los dos tipos de error que pueden cometerse. Ahora se tratará la forma en que el tamaño de la muestra influye en β , a continuación se describen los métodos para determinar cuántas observaciones se deberán incluir en la muestra de interés.

2. RELACION DE LA POTENCIA CON EL TAMAÑO DE MUESTRA

La probabilidad de no cometer un error tipo II, o potencia de la prueba, cuyo valor es $(1-\beta)$ cuando en verdad los parámetros son iguales, está determinada en gran medida por una función del tamaño de la muestra estudiada; es decir, está relacionada con el número de observaciones que se tengan. Una muestra demasiado pequeña difícilmente proporcionará información confiable sobre la diferencia entre los parámetros de dos poblaciones; por ejemplo, si se tiene interés en comparar la tasa de fertilidad del ganado cebú con la de las razas europeas, y sólo se registran cuatro vacas de cada grupo, la cantidad de información disponible no permitiría obtener resultados concluyentes acerca del problema. Por esto se han generado métodos que permiten determinar el número mínimo de observaciones que deben constituir la muestra para obtener información suficiente y confiable.

Por otro lado, es importante destacar que una muestra demasiado grande permitiría declarar significativa casi cualquier diferencia, por pequeña y de poca repercusión práctica que ésta sea. Si en el estudio sobre fertilidad del ganado hindú comparado

con el europeo, se tomaran los datos de diez mil hembras de cada especie, se considerarían significativas discrepancias menores al 1% entre las tasas de fertilidad, aunque una diferencia tan reducida no es de interés en la práctica. Este problema es poco frecuente en la investigación pecuaria, ya que, por lo general, el alto costo de mantener a cada animal experimental obliga a reducir el tamaño de la muestra conforme a los recursos disponibles.

Incluso algunos autores afirman que el error tipo I casi nunca ocurre, ya que los parámetros de dos poblaciones casi siempre son distintos, aunque muchas veces la diferencia sea en extremo reducida. Se deduce entonces que el investigador debe poner mucha atención en obtener una muestra de tamaño suficiente para contar con información confiable, pero que no sea tan grande que encuentre significativa una discrepancia irrelevante.

En consecuencia, a lo largo de este capítulo se hace referencia a la "determinación del tamaño de muestra" y no al "tamaño mínimo necesario" de la misma. Los métodos expuestos permiten estimar el número mínimo de datos necesarios para no sufrir los problemas de una muestra pequeña, pero también indican el límite máximo con sentido práctico. La siguiente sección contiene dos métodos para determinar la diferencia de interés práctico que es deseable detectar en un estudio, y la Sección VI.3 presenta los procedimientos para determinar el tamaño de la muestra.

3. DIFERENCIA MINIMA POR DETECTAR

En la sección anterior se explicó la importancia de evitar el uso de muestras demasiado grandes que conduzcan a un resultado significativo para diferencias irrelevantes. Determinar el número apropiado de datos en la muestra exige conocer cuál es la diferencia más pequeña entre los parámetros que es de importancia práctica. En esta sección se analiza un método para calcular la "mínima diferencia importante" entre dos proporciones.

Al analizar algún fenómeno específico, por lo general se tiene una idea previa del resultado que cabe esperar, la cual surge de la experiencia del profesional dedicado al tema, de estadísticas anteriores, ya sean propias o de la literatura especializada, o bien de una muestra piloto. Dado el mínimo de información, el encargado del estudio podrá determinar el tamaño de la muestra. Cuando no se cuenta con ningún dato previo, difícilmente podrá determinarse el número de casos que habrá de contener la muestra, en esta situación será de gran utilidad tomar una muestra piloto y examinarla.

En algunos casos, cuando no se tiene información previa, es útil saber que, si los demás factores permanecen constantes, el tamaño de muestra para una variable binomial es máximo cuando la proporción de interés es de 50%, de modo que puede emplearse este

valor para calcular el número de observaciones que debe contener la muestra; es decir, que puede tomarse la muestra necesaria para analizar una proporción de la mitad y, en caso que se presentara un porcentaje menor, de todos modos la muestra sería suficiente. A continuación se detallan algunos procedimientos para estimar la proporción o la diferencia entre proporciones con que debe desarrollarse el cálculo del tamaño de la muestra para estudios basados en la tabla 2x2, y después se exponen los métodos para el cálculo mismo del tamaño de la muestra.

a) VENTAJA MINIMA SOBRE EL ESTANDAR

El punto de vista más utilizado cuando se compara la tasa de respuesta en un nuevo grupo con respecto a un estándar, consiste en fijar un mínimo de ventaja que el nuevo debe tener para considerarlo mejor. Si se busca la comparación entre un nuevo tratamiento y el que se tiene en uso, conociendo la tasa actual de respuestas favorables (P) y suponiendo que todos los que responden al tratamiento actual lo harán con el nuevo, entonces lo que se requiere es que el nuevo beneficie a una fracción suficiente (F) de los que no responden ($Q = 1 - P$) para justificar el cambio, de modo que alcance un porcentaje total de respuesta ($P + QF$).

Ejemplo VI.1: ventaja mínima sobre el estándar

Si un laboratorio fabrica una vacuna con 85% de eficacia y quiere probar una nueva vacuna, el responsable del estudio puede pensar que si ésta protege al menos a un tercio de los que no responden al producto en uso, entonces vale la pena cambiar a la producción de la nueva vacuna.

La vacuna actual falla en el 15% de las veces ($1 - 0.85 = 0.15$). Si la nueva es efectiva en la tercera parte de estos casos, $(15\%) (1/3) = 5\%$, entonces debe alcanzar por lo menos un 5% más de protección que la otra, con un total de $85\% + 5\% = 90\%$, por lo que sólo una tasa de protección del 90% justificaría la nueva vacuna.

b) RAZON DE MOMIOS ENTRE LOS GRUPOS

Otro método útil para planear valores esperados para la proporción de respuesta en los grupos a estudiar se basa en la razón de momios (ver el cap. III para una descripción de esta estadística). Cuando se requiere confirmar o refutar los resultados

de estudios previos y se cuenta con los datos necesarios para calcular la razón de mosios en esa investigación anterior, lo mejor será evaluar la proporción de respuesta en uno de los grupos con base en la razón de mosios del estudio anterior y en la información que se tenga sobre la otra muestra de interés.

$$P_2 = \frac{M P_1}{M P_1 + (1 - P_1)}$$

Donde M representa la razón de mosios y P_1 es la probabilidad de respuesta en uno de los grupos.

Teniendo la proporción de respuesta que se espera en cada grupo, se procede a calcular el tamaño de la muestra conforme los métodos expuestos en la siguiente sección.

Ejemplo VI.2: proporciones basadas en la razón de mosios

En un lote de gatos registrados en una clinica, se encontró que los animales adultos padecera cierta enfermedad con una frecuencia dos y media veces mayor que en los jóvenes ($M = 2.5$). Si se quiere verificar esta observación en otra población de gatos, sabiendo que el padecimiento se encuentra en el 15% de los adultos, puede determinarse la tasa de enfermos que cabria esperar entre los jóvenes:

$$P_2 = \frac{2.5 (0.15)}{2.5 (0.15) + 0.85} = 0.3061$$

Lo cual indica que, si los datos confirman la información anterior, la proporción de enfermos en el segundo grupo debe oscilar alrededor del 31%.

4. TAMARO DE LA MUESTRA PARA ESTUDIOS CON MUESTREO MULTINOMIAL

Al aplicar el muestreo multinomial (descrito en el Cap. II), se tendrá necesidad de saber el número de casos que deben formar

la muestra para representar adecuadamente a la población de referencia. Para tratar éste caso se determinará el tamaño de la muestra necesario para estimar las proporciones de una de las variables dicotómicas consideradas; al efecto, el cálculo se basará en las probabilidades marginales, consideradas aquí como estimaciones de la proporción de individuos que poseen uno u otro valor para cada una de las variables binarias que se están estudiando.

Lo primero que se requiere es señalar un valor esperado para las probabilidades marginales de una de las dos variables dicotómicas, es decir para la que se usará como base para determinar el tamaño de la muestra. De las variables consideradas, ésta deberá ser la que tenga un proporción más cercana al 50%, por que, como ya se explicó, mientras más cercano al 50% mayor será el tamaño de la muestra necesaria. El valor esperado puede surgir del conocimiento que se tenga del problema, de experiencia, de lo descrito en artículos relacionados, o bien de un estudio piloto; por ejemplo, alguien puede afirmar que la mortalidad durante la lactancia será de un 10%, o plantear: que de acuerdo con el tipo de alimento, las cualidades de las instalaciones y el programa sanitario aplicado, la tasa de diarreas en lechones debe ser de 20%.

El segundo punto de importancia es la precisión con que se quiere estimar el porcentaje referido. Es fácil ver que, para una evaluación muy vaga, como podría requerirse en una primera aproximación a un problema, basta un lote pequeño; pero lograr una estimación fina exige una muestra grande. Por ejemplo, para estimar con un error menor al 1% el porcentaje de vacas lecheras enfermas de mastitis en el Valle de México, se necesitaría una muestra numerosa: más de mil hembras; mientras que, permitiendo un error del 10% en la estimación, se tendría suficiente con unas cien vacas o incluso con menos. Lo común es fijar la precisión del estimador en 1, 5 o 10% (0.01, 0.05 o 0.1), ya que estimar una proporción con fluctuaciones mayores al 10% conduce a un resultado pobre en información sobre el problema.

a) PORCENTAJES MEDIOS

Si se conoce el valor predicho para el porcentaje y se tiene la precisión requerida para el estimador, sólo falta fijar el grado de confiabilidad con que se trabajará, la confiabilidad representa la probabilidad de que el error de estimación sea igual o menor que el límite señalado, por lo común se utiliza 95% o 99% de confiabilidad. Si el porcentaje predicho fluctúa entre 30% y 70%, el tamaño de la muestra se obtiene:

$$N = K P (1 - P) / e^2$$

Donde: P es la proporción predicha, e representa la precisión de estimador y K varía con la confiabilidad, K = 3.8416 para 95%; y K = 6.6358 para 99%

Ejemplo VI.3: muestra para un porcentaje medio

En un rancho ganadero se investiga la proporción de animales parasitados y su relación con el origen del ganado. Se espera que haya un 35% de animales afectados y se estima que 45% del total proviene de otros ranchos. Como 45% es el más cercano a 50%, se utiliza como base para determinar el tamaño de muestra, si se permite un error de estimación hasta del 10% -precisión del 0.1- con una confiabilidad de 95%, la muestra se calcula:

$$N = 3.8416 (0.45) (1 - 0.45) / 0.1^2 = 95$$

Lo cuál indica que si se examinan 95 bovinos y se califican de acuerdo con su origen y al diagnóstico parasitológico, se alcanzará una muestra de la calidad estipulada.

b) PORCENTAJES PEQUEÑOS

Cuando la proporción predicha se encuentra por debajo del 30%, la precisión del estimador que se busca se expresa en forma relativa, a través del coeficiente de variación. Este indica el error estándar del estimador como proporción del porcentaje estimado; por ejemplo, si se estima un porcentaje de 10% con un coeficiente de variación del 0.2, entonces se permite un error de estimación máximo de 2% (0.2 de 10%). Lo más común es fijar el coeficiente de variación permisible en 5%, 10% o 20% (0.05, 0.1 o 0.2). Para este tipo de casos, la muestra se determina mediante:

$$N = (1 - P) / (P V)$$

Donde: V es el coeficiente de variación expresado como fracción del estimador

Ejemplo VI.4: muestra para un porcentaje pequeño

Al analizar la proporción de vacas con mastitis y su relación con la desinfección del equipo de ordeño, se decide utilizar la variable de salud de la glándula mamaria como base para determinar el tamaño de la muestra. Se plantea que cabe esperar un 15% de vacas enfermas y que se quiere un coeficiente de variabilidad en el estimador de 5%:

$$N = (1 - 0.15) / (0.15 (0.05)) = 113$$

De este modo, con 113 observaciones puede determinarse la proporción de vacas con mastitis, logrando un coeficiente de variación máximo de 5% del estimador, esto significa un error estándar de hasta 5% del valor del estimador; si el estimador es de 20%, 5% de variación significa un error estándar de 1% ($15\% [20\%] = 1\%$). Si se optara por un coeficiente de variación del estimador de hasta 20%, la muestra debería incluir sólo 28 vacas:

$$N = (1 - 0.15) / (0.15 (0.2)) = 28$$

Por supuesto que, en esta segunda muestra, el resultado sería un estimador mucho menos preciso que en la primera.

c) PORCENTAJES GRANDES

Una situación muy similar a la anterior se presenta cuando la proporción que se espera encontrar está por arriba del 70%. La única diferencia es que se invierten la relaciones entre la proporción de interés y el resto, por lo cual la fórmula aplicable en este caso es:

$$N = P / ((1 - P) V)$$

Si se examina la expresión anterior, se encontrará que es la misma ecuación que en el caso anterior, pero los valores de P y de 1 - P se han intercambiado.

Ejemplo VI. 5: tamaño de la muestra para un porcentaje grande

Durante la evaluación de una empresa porcina se requiere estimar la tasa de fertilidad de las hembras del pie de cría. Se piensa encontrar alrededor del 80% de fertilidad al primer servicio. Buscando un coeficiente de variación no mayor del 5%, el tamaño de muestra se resuelve:

$$N = 0.8 / ((1 - 0.8) 0.05) = 80$$

Lo cuál indica la necesidad de integrar la muestra con los datos de fertilidad de 80 cerdas.

5. TAMAÑO DE LA MUESTRA PARA ESTUDIOS COMPARATIVOS

Si el propósito de un estudio es comparar las proporciones de dos grupos, es decir, cuando se aplican muestreos comparativos o experimentales, se debe determinar el tamaño de la muestra que permitirá contrastar los porcentajes de ambos grupos mediante una prueba de potencia adecuada.

Por tal motivo, para estos casos se plantea la determinación del número de casos que deben integrar la muestra con base en la potencia (en el Cap. IV se explica este concepto), de modo que permita una alta probabilidad de detectar como significativas las diferencias entre las muestras cuando los parámetros en verdad difieren tanto o más que la mínima diferencia de importancia práctica.

Para desarrollar los métodos que se describen a continuación debe especificarse el valor de α (véase el Cap. IV para una explicación de α) y la potencia con que se desea trabajar. Para definir α deben tomarse en cuenta las posibles consecuencias del resultado de la prueba: si encontrar que los porcentajes difieren significa que el tratamiento en uso será sustituido por el nuevo o que deberán invertirse mucho más recursos en la investigación al respecto, entonces cometer un error tipo I será grave y deberá usarse un valor pequeño para α , por ejemplo $\alpha = 0.01$.

Por el contrario, si las conclusiones serán utilizados como una parte más del conocimiento en relación con cierto fenómeno, no es tan importante el riesgo de cometer un error de tipo I y α puede elevarse a 5% e incluso a 10% (0.05, 0.1).

Además es necesario especificar la potencia de la prueba. Cohen (1969) propone un criterio útil para hacerlo: planteando que, en forma general, el error tipo I es cuatro veces más grave que el error tipo II, entonces β debe fijarse cerca de cuatro veces el valor de α y, por tanto, la potencia $(1 - \beta)$ debe fijarse

en $1 - \alpha$. Siguiendo este criterio, para $\alpha = 0.01$ se recomienda usar una potencia de 0.95; cuando $\alpha = 0.05$, la potencia debe establecerse en 0.8. Sin embargo, en ningún caso es conveniente que la potencia sea menor que el 70%, aun cuando se empleen valores de α de 10%.

A partir de los valores esperados para cada una de las proporciones a comparar, de la magnitud de α y de la potencia requerida puede definirse el tamaño de cada una de las dos muestras. Al efecto se necesitan los siguientes datos:

P_1 y P_2 : proporción de respuesta predicha para cada uno de los grupos interés.

K_{α} : constante que separa una fracción $\alpha/2$ en el extremo superior de la distribución normal estándar de probabilidades. El cuadro VI.1 muestra las constantes adecuadas para una amplia gama de valores de α .

K_{β} : constante para la potencia requerida, tomada del cuadro VI.1.

\bar{P} : proporción media de respuesta, $\bar{P} = (P_1 + P_2) / 2$

Fleiss (1972) demuestra que el número de casos necesario en cada una de las dos muestras se estima mediante:

$$N = \frac{(K_{\alpha} \sqrt{2 \bar{P} (1 - \bar{P})}) + K_{\beta} \sqrt{(P_1 (1 - P_1) + P_2 (1 - P_2))}}{(P_1 - P_2)^2}$$

El resultado indica cuántas observaciones deben integrar cada muestra, por lo que el total de casos en la investigación será el doble de lo que señala N ($N_1 + N_2$).

Si se prueban diferencias en un sentido específico, es decir cuando la prueba implica hipótesis unidireccionales, como en la prueba exacta de Fisher, debe tomarse la constante marcada con el doble del valor de α a utilizar. En la potencia no se hace ningún cambio; por ejemplo, si se aplica $\alpha = 0.01$ se buscan las constantes para $\alpha = 0.02$ y para una potencia de 95%.

En la expresión anterior no se considera la corrección por continuidad recomendada en muchas de las pruebas descritas. Kra-

mer y Greenhouse (1959) desarrollaron un método para ajustar el tamaño de la muestra cuando se ha de aplicar esta corrección:

$$N_{aj} = N / 4 (1 + 4[1 + 8 / (N |P_1 - P_2|)])^2$$

N_{aj} ajustada (N_{aj}) es el número de casos necesarios en cada muestra para los estudios que serán evaluados mediante una prueba con corrección por continuidad.

Al calcular el tamaño de la muestra, recuerdese que, manteniendo los demás factores constantes, al aumentar α el tamaño de la muestra será menor. Por otro lado, si solo se modifica la potencia, al incrementarla también habrá un aumento en el número de observaciones necesarias. Ahora bien, conservando fijos α y la potencia, cuanto mayor es la diferencia entre los porcentajes menor es la cantidad de datos necesaria para encontrar significativa esta diferencia.

Cuadro VI.1

Constantes para la determinación del tamaño de muestra

Alfa* (α)	K_{α}	Potencia**	K_{α}
0.001	3.296	0.995	2.576
0.002	3.090	0.99	2.326
0.005	2.813	0.98	2.055
0.01	2.576	0.95	1.645
0.02	2.326	0.90	1.282
0.05	1.960	0.80	0.842
0.1	1.645	0.75	0.674
0.2	1.282	0.70	0.524

*Para probar hipótesis unidireccionales se toma la constante correspondiente a 2α y la potencia del renglón α .

**La potencia escrita en el mismo renglón es el valor recomendado para cada nivel de α .

EJEMPLO VI.6: tamaño de la muestra para comparar porcentajes

Volviendo al ejemplo VI.1, se tiene una vacuna con 85% de eficacia y se quiere probar si la nueva vacuna alcanza el 90% de protección. Considerando que un error tipo I tendría consecuencias de importancia para el laboratorio, ya que invertiría en un cambio inadecuado, entonces se utiliza $\alpha = 0.01$ y, de acuerdo con lo recomendado en el cuadro VI.1, una potencia de 95%.

Nótese que la diferencia fundamental con los ejemplos VI.3, VI.4 y VI.5 es que aquí se aplicará un muestreo experimental -aunque podría ser comparativo-, que exige una muestra separada para cada grupo, mientras que en los ejemplos citados se trata de muestreos multinomiales, o sea una sola muestra que luego se clasifica de acuerdo con las dos variables binarias de interés.

$$\bar{P} = (0.85 + 0.9) / 2 = 0.875$$

$$N = \frac{(2.576)^2 \{2 (0.875) (1 - 0.875)\} + 1.645^2 \{(0.85 (0 - 0.85) + 0.9 (1 - 0.9))\}}{(0.85 - 0.9)^2}$$

$$N = 1\ 555$$

Si al analizar los resultados se utiliza una prueba con corrección por continuidad, el número de animales que debe someterse a cada vacuna aumenta a 1 634.

$$N_c = 1\ 555 / 4 \{1 + (1 + 8 / 4(1\ 555 \cdot 1.85 - 91))\}^2 = 1\ 634$$

Estos datos señalan que el estudio debe incluir un total de $2(1\ 555) = 3\ 110$ animales o $2(1\ 634) = 3\ 268$ con el ajuste, que es una muestra bastante numerosa, pero debe observarse que la investigación se enfoca a detectar una diferencia pequeña de -sólo 5% (90% - 85% = 5%)-, con base en una potencia grande y poca probabilidad de cometer un error tipo I. Si se redujera la potencia de la prueba a 75%, el resultado sería:

$$N = 606$$

$$N_{\alpha} = 684$$

Con lo cual se verifica que, a mayor potencia, la muestra necesaria se incrementa, en tanto que al aumentar o se disminuye el tamaño de la muestra. De manera semejante, si α se conserva en 1% y la potencia en 95%, pero se busca una comparación entre porcentajes esperados de 85% y 95%, sólo se necesitarían 317 casos -con el ajuste 356- lo que representa una reducción casi a la quinta parte al aumentar la diferencia buscada de 5% ($90\% - 85\% = 5\%$) a 10% ($95\% - 85\% = 10\%$).

LITERATURA CITADA

- Bacha, B.F. (1985): Frecuencia de distocias en ganado bovino especializado en producción de leche. Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Barajas, C.G. (1985): Estudio preliminar para un programa de mejoramiento genético para bovinos productores de carne en el trópico. Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Billewicz, W.Z. (1965): The efficiency of matched sample: An empirical investigation, Biometrics, 21: 623-644.
- Brunden, M.N. (1972): The analysis of non-independent 2 x 2 tables using rank sums, Biometrics, 24: 315-327.
- Cochran, W.G. (1969): Some methods of strengthening the common "Chi square" test, Biometrics, 10: 417-451.
- Cohen, J. (1960): A coefficient of agreement for nominal scales. Educ. Psychol. Measur., 20: 37-46.
- Cornfield, J. (1951): A method of estimating comparative rates from clinical data. J. Natl. Cancer Inst., 11: 1269-1275.
- Correa, M.C. (1985): Prevalencia de enfermedades diagnosticadas radiológicamente en perros y gatos. Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Chase, G.R. (1968): On the efficiency of matched pairs in Bernoulli trials, Biometrika, 55: 365-369.
- Edwards, A.L. (1948): Note on the "correction for continuity" in testing the significance of the difference between correlated proportions, Psychometrika, 13: 185-187.
- Feldman, S.E. and Klinger, E. (1963): Short-cut calculation of the Fisher "exact test", Psychometrika, 28: 289-291.
- Fleiss, J.L. (1973): Statistical methods for rates and proportions. Wiley, New York.
- Gart, J.J. (1969): An exact test for comparing matched proportions in crossover designs, Biometrika, 56: 75-80.
- Gonzalez, V.L. (1983): Efecto del halotano sobre la acción letal aguda del cloranfenicol. Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Goodman, L. A. and Kruskal, W. H. (1954): Measures of association for cross classification. J. Amer. Statis. Assoc., 49: 732-764.

- Goodman, L. A. and Kruskal, W. H. (1959): Measures of association for cross classification II: Further developments and references. J. Amer. Statist. Assoc., 54: 109-161.
- Goodman, L. A. and Kruskal, W. H. (1954): Measures of association for cross classification III: Approximate sampling theory. J. Amer. Statist. Assoc., 58: 310-364.
- Irwin, J.O. (1949): A note on the subdivision of chi-square into components. Biometrika, 36: 130-131.
- Kramer, M. and Greenhouse, S. W. (1959): Determination of sample size and selection of cases. In National Academy (NRC, 583), Psychopharmacology: problems in evaluation, Washington, D. C.
- Lancaster, H.O. (1949): The derivation and partition of chi-square in certain discrete distributions, Biometrika, 36: 117-129.
- Landivar, S.C. (1982) Fertilidad con inseminación artificial y sonda directa en ganado cebú a estrógeno saturable inducido con pgf2 α , Tesis de maestría, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Leach, C. (1979): Introduction to statistics, a nonparametric approach for the social sciences, Wiley, New York.
- Lopez, J.D. (1986): Diagnóstico temprano de gestación en cerdas mediante la resistencia eléctrica de las secreciones cervico-vaginales. Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Mantel, N. (1963): Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel method, J. Amer. Statist. Assoc., 58: 690-700.
- Mantel, N. and Haenszel, W. (1959): Statistical aspects of the analysis of data from retrospective studies of disease, J. Natl. Cancer Inst., 22: 719-744.
- McNemar, Q. (1947): Note on the sampling error of the difference between correlated proportions of percentages, Psychometrika, 12: 153-157.
- Medina, G.J. (1986): Efecto de la sonda con machos criptorquidos antes del servicio sobre la fertilidad y prolificidad en cerdas. Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Pearson, K. (1900): On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from the random sampling. Philos. Mag., 50: (5) 157-175.

- Rivero, M.P. (1984): Frecuencia estacional de Dirofilaria imitans en perros de la Ciudad de Chetumal, G.R., Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Romo, B.B.O. (1986): Evaluación de embriones bovinos para la técnica de trasplante de embriones. Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Rubio, G. I. (1985): Niveles séricos de progesterona en vacas Indobrasil. (Bos indicus) en el trópico. Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Somers, R. H. (1962): A new asymmetrical measure of association for ordinal variables Amer. Sociological Rev., 27: 799-811.
- Vargas, M. G. (1984): Distribución de cysticercus cellulosae en el cerdo y su importancia para la inspección sanitaria. Tesis profesional, Fac. Med. Vet. Zoot. U.N.A.M., México, D.F.
- Vázquez, F., Sumano, L.H. y Navarro Fierro, R. (1984): Efecto de la acupuntura sobre la concepción en vacas con más de cuatro repeticiones. Memoria de la Reunión de Investigación Pecuaría en México 1984, p 297. México, D.F.
- Yates, F. (1934): Contingency tables involving small number and the Chi square test, J. Roy. Statist. Soc. Suppl., 1: 217-291.
- Youkeles, L.H. (1963): Loss of power through ineffective pairing of observations in small two-treatment all-or none experiments, Biometrics, 19: 175-180.