Conceptos básicos sobre investigaciones para el conocimiento de las asociaciones

Carlos Ferrero

PROYECTO SOBRE

SISTEMAS DE INFORMACION ESTRATEGICA

PARA NIVELES MUNICIPALES

CONCEPTOS BASICOS SOBRE INVESTIGACIONES PARA EL CONOCIMIENTO DE LAS ASOCIACIONES

Dr. Carlos Ferrero

Instituto Nacional de
Estadística y Censos
Mayo, 1993

CONTENIDO

		Página
	CONCEPTOS BASICOS SOBRE INVESTIGACIONES PARA EL CONOCIMIENTO DE LAS ASOCIACIONES	1
1.	Determinación de Significado en Tables de 2x2 o 2xn	3
2.	Los Caracteres de la Asociación en una Tabla de 2x2	6
2.1.	La Signicación de la Asociación en una Tabla de 2x2	6
2.2.	La Magnitud de la Asociación	9
2.2.1.	Coeficiente Ý (Fi)	9
2.2.2.	Riesgo Relativo	9
2.2.3.	Relación de Contingencia u "Odds Ratio"	14
2.2.4.	Riesgo Diferencial	17
2.3.	Variabilidad de los Estimadores de Riesgo	18
3.	La Asociación en Tablas de 2xk	19
4.	La Asociación en Tablas de Contingencias de rxk	21
5.	La Asociación de Varias Tablas de 2x2	21
6.	Contraste Entre Grupos	26
6.1.	Análisis del Primer Estrato	26
6.2.	Análisis del Segundo Estrato	28
7.	El Caso General de Análisis de Asociaciones	32
8.	Procedimiento Abreviado de Cálculo	33
8.1.	Computaciones para el Primer Estrato	34
8.2.	Computaciones para el Segundo Estrato	35

CONCEPTOS BASICOS SOBRE INVESTIGACIONES PARA EL CONOCIMIENTO DE LAS ASOCIACIONES

La investigación estadística procura interpretar el comportamiento de los fenómenos biológicos y sociales a través del análisis descriptivo de sus características o de las inferencias que se realizan a partir de las muestras que de ellos se extraen. La inducción y la deducción son dos recursos lógicos normalmente usados.

Sin embargo toda la investigación estadística basada en métodos paramétricos da información sobre las probabilidades de los resultados de los tests realizados en cada situación, poniendo especial énfasis en aclarar que ellos son aplicables y extrapolables, con un determinado nivel de confianza, al o a los universos de los cuáles el o los estudios provienen; o a otros universos con caracteres semejantes a aquéllos en que se realizaron las pruebas.

Por ejemplo, cuando se estudia la correlación de los resultados obtenidos de la medición de dos eventos en cada uno de los individuos que pertenecen a una misma muestra, y cuando se asume que su distribución es normal (de la regresión de "y" en "x"), se afirma que con un determinado nivel de confianza, para cada nivel de variación de "x" corresponde una variación de "y". Pero cuidado que esto no indica que la variación de "x" es responsable por los cambios en "y".

Otro ejemplo, cuando se extraen dos muestras para determinar si hay discrepancias significativas entre ellas, a un determinado nivel de confianza, se dice que se rechaza o no la hipótesis nula que indica, en principio, que entre las distribuciones de ambos universos no hay diferencias. No se dice que los universos de las cuales ellas provienen son iguales o disímiles, sino que con ese nivel de confianza se rechazan o no las hipótesis originales.

La estadística paramétrica cuida mucho no establecer relaciones de causalidad, hace sólo hincapié en que para un determinado nivel de confianza, los resultados logrados pueden o no provenir del mismo o de diferentes universos.

Sin embargo el problema del investigador que procura detectar las relaciones entre los agentes etiológicos y los caracteres o hábitos de los integrantes de los grupos que están expuestos a ellos, es poder caracterizar al menos la asociación entre ambos hechos a la par que denotar la magnitud de la misma. Esto lo acerca mucho a poder lograr una identificación satisfactoria de las relaciones entre ambos factores que le de pie para otras deducciones y pruebas posteriores. Con el apoyo de las diferentes formas de la lógica se puede avanzar, a posteriori, a otras áreas semejantes o derivadas.

Como lo importante para el investigador es poder identificar asociaciones causales, resulta importante medir esa característica entre las variables y poder caracterizar la magnitud de la misma. Esto da mayor seguridad al investigador que arriesga establecer algunas relaciones de causalidad parcial o total. [1]/

Las características y magnitudes de las asociaciones se pueden estudiar a traves de dos tipos de diseños: 1) estudios de corte transversal y 2) estudios de cohorte. En el primero se estudia retrospectivamente a grupos de individuos expuestos o no a ciertos factores, y posteriormente se trata de detectar si tuvieron o no manifestaciones de ciertas enfermedades o condiciones sociales en estudio. En este primer caso lo único que se establece a priori es el tamaño de la muestra total. Las frecuencias parciales de cada celda, en las tablas de 2 x 2, son parte del total general citado, lo mismo que los subtotales marginales.

Ejemplos de las cuestiones que pueden formularse son de la siguiente naturaleza: 1) las personas de una comunidad en que una parte está expuesta a ciertas condiciones sociales o de salud que pueden condicionar la aparición de ciertos efectos negativos; y 2) los grupos humanos expuestos a ciertos riesgos laborales corren mayor peligro de contraer algunas afecciones debidas al tipo de sus actividades, que otros no expuestos o apropiadamente defendidos de esos factores.

Para responder a estos interrogantes, en este primer tipo de diseño, se selecciona una comunidad, una de cuyas partes estuvo expuesta al riesgo de las condiciones "negativas" y otra en que esa situación no se dio. En cada una de ellas se diferencia a los que estuvieron o no expuestos a las condiciones seleccionadas, y posteriormente se identifica a los que poseyeron (o adquirieron) la enfermedad durante un período relativamente breve, anterior a la investigación.

En el segundo se toman dos grupos, generalmente de tamaño semejante, y se constituyen dos cohortes (grupos de individuos que son seguidos a lo largo del tiempo), y se los observa y mide longitudinalmente, en forma prospectiva o retrospectiva, para contabilizar el número de personas que contraen o padecieron algún tipo de esas enfermedades profesionales; este diseño procura determinar si existen relaciones entre los riesgos laborales a que están expuestas las cohortes y la aparición de "casos" en cada una de ellas. En este caso es deseable que ambos grupos sean lo más semejantes posible y que sólo discrepen en la constatación de la condición que se estudia. Sólo así pueden ser aceptables las diferencias que se detecten en los grupos expuestos o no al riesgo.

Acá hay una determinación a priori del tamaño de cada uno de los dos grupos, de tal suerte que hay un "n_1" y un "n_2". En el caso anterior había un solo "n" o "$n_{..}$".

Como se dijo, en ambos tipos de estudio los datos se obtienen fijando a priori el(los) tamaño(s) de la(s) muestra(s). En el primero se establece un tamaño único de la misma y a partir de esta cantidad se identifican los expuestos y los que no lo

[1]/ Armitage, P.: "Statistical Methods in Medical Research"; 3a. impres..John Wiley and Son; 1974; Gran Bretaña; pag. 176 a 181.

fueron. Posteriormente se busca la cantidad de enfermos en ambos grupos y de ello se desprende las magnitudes de los que no llegan a adquirir la enfermedad.

Cuando las frecuencias absolutas de cada celda se transforman en probabilidades, éstas pueden ser solamente expresión de probabilidades parciales del total general. Lo mismo lo son las probabilidades provenientes de los totales marginales correspondientes. En este caso la medida del riesgo está dada por la **razón de contingencia** u **"odds ratio"** según se denomina en la literatura especializada de habla inglesa, y cuya estructura será detallada posteriormente.

Otra forma de constatar la existencia o no de riesgos es a través de las **tasas específicas** de cada grupo, esto es, las tasas diferenciales de enfermedad entre los que estuvieron o no expuestos. En este caso lo más importante es comparar ambas tasas específicas (los expuestos que enfermaron sobre el total de los expuestos y los no expuestos que enfermaron sobre el total de los no expuestos) y a partir de ésto inferir cuánto más severa es la situación de enfermedad según haya o no existido exposición.

Muchas veces se obtiene una razón entre ambas tasas y a esto se llama **"riesgo relativo"**. Si bien constituye una medida única que facilita las comparaciones entre diversas muestras, tiene el inconveniente que una razón de razones pierde capacidad de destacar las diferencias entre ambas tasas específicas. Berkson, citado por Fleiss, fué quién primero llamó la atención sobre estos hechos y lo particularizó con su crítica al valor de la razón de contingencia.[2]/ Estudiando el efecto del cigarrillo sobre el cáncer de pulmón y las enfermedades coronarias aclara con un ejemplo las formas erróneas de interpretación de esas razones. Se recomienda su lectura para entender las lógicas prevenciones que presenta tanto él como posteriormente Sheps, con el modelo de corrección que propone.

1. Determinación de Significado en Tablas de 2x2 o 2xn

Las tablas de "2 x 2" o de "2 x n" son frecuentemente usadas para determinar la existencia de asociaciones entre dos características estudiadas. Una tabla modelo puede ser la siguiente: los valores de las celdas son medidas de probabilidad que resultan del cociente entre el valor de la frecuencia absoluta registrada en cada una de ellas y el total general de la tabla. De este modo: p_{11} es igual a n_{11} / $n_{..}$ y asi sucesivamente. Ademas los totales marginales $p_{1.} = p_{11} + p_{12}$; $p_{.1} = p_{11} + p_{21}$; etc.

[2]/ Fleiss,J.L.: "Statistical Methods for Rates and Proportions"; 1a. ed.; John Wiley and Sons; New York, 1973; pag.60.

Tabla 1 (Modelo)
(Con Proporciones en las Celdas)

Factor A	Factor B		
	Presente	Ausente	Total
Presente	p_{11}	p_{12}	$p_{1.}$
Ausente	p_{21}	p_{22}	$p_{2.}$
Total	$p_{.1}$	$p_{.2}$	1,00

(1.1)

La tabla con las frecuencias absolutas, tiene la siguiente notación: n_{11} = "a", es igual al valor de la frecuencia absoluta de la celda con la probabilidad p_{11}; n_{12} = "b", es la frecuencia absoluta de la celda con la probabilidad p_{12}; n_{21} = "c", es la frecuencia absoluta de la celda con p_{21}; y finalmente n_{22} = "d", es la que tiene la probabilidad p_{22}. Los valores de los subtotales, por ejemplo $n_{1.} = n_{11} + n_{12}$ = "a+b"; y $n_{.1} = n_{11} + n_{21}$.= "a+c"; y asi sucesivamente.

Tabla 2 (Modelo)
(Con Valores Absolutos en las Celdas)
(Notación A)

Factor A	Factor B		
	Presente	Ausente	Total
Presente	n_{11}	n_{12}	$n_{1.}$
Ausente	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

(1.2)

Tabla 3 (Modelo)
(Con Valores Absolutos en las Celdas)
(Notación B)

Factor A	Factor B		
	Presente	Ausente	Total
Presente	a	b	a+b
Ausente	c	d	c+d
Total	a+c	b+d	N

(1.3)

Cuál es el significado usual de investigar asociaciones entre variables?. En el área epidemiológica, frecuentemente, lo que se desea conocer es si una condición que afecta a un grupo de personas es la responsable por la aparición de una enfermedad o de un problema social determinado. Si cada vez que se encuentra una enfermedad está también presente una condición en estudio, se puede inferir que al menos hay asociación entre ambos hechos.

Estadisticamente, **per** se no se puede establecer la existencia de una relación causal; sin embargo si se da el caso que un investigador administre una medicación a un grupo de enfermos y sistemáticamente aparezca en ellos una determinada reacción, es posible hablar aquí que esa droga "puede" tener **efectos causales** sobre la aparición de la reacción, pero no se puede decir que aquélla sea responsable por la aparición de ésta.

Como dice Armitage en la obra antes citada, la capacidad de realizar o no un experimento que oriente firmemente sobre las relaciones de causalidad es usualmente poco posible. Consecuentemente los epidemiólogos muchas veces deben estar satisfechos si pueden observar que la aparición de una enfermedad está **asociada** con la presencia de un hábito personal o una característica dada. Esto permite, después, tomar la decisión de hacer afirmaciones lógicas sobre **posibles nexos causales**.

Cuando en estadísticas se hacen estudios de **asociación** hay siempre una implicancia de comparación. Cuando se compara el número de casos de desnutrición que se presentan en familias de bajos ingresos, con los que aparecen en otras de ingreso alto, y se reconoce que las diferencias son evidentes, no se hace más que afirmar que hay asociación entre niveles de ingreso y desnutrición. (Maxwell: Analysing Quelitative Data, páginas 11 y 12). Los tests de significación estadística que se aplican cuando se comparan datos de este tipo, toman cuenta las discrepancias que se observan y que pueden ser debidas al azar, que se suman a las diferencias realmente debidas a las asociaciones. Esto permite "limpiar" las conclusiones.

Otro criterio importante a considerar en este tipo de análisis es el de **independencia**. Se considera que dos eventos son "independientes", esto es, que no hay asociación entre dos características, cuando la presencia de una de ellas en una no afecta las posibilidades de tener la otra. El concepto de independencia en las tablas de 2x2 se refiere a la condición de semejanza que muestra la proporción observada en cada celda (p_{12}) por ejemplo, con los valores de los productos de los totales marginales que a ella corresponden ($P_{1.}$ x $p_{.2}$) en este caso. Es por esto que se dice que dos criterios de clasificación son independientes cuando la proporción de individuos que adquieren una condición dada (frecuencia en la celda) muestra una gran semejanza con los valores esperados en cada una de las celdas cuando estos se calculan como los valores del universo correspondiente y se estima a través de los productos de los totales marginales correspondientes. **No se habla aca de independencia de las muestras sino de los criterios de clasificación.** (Para detalles sobre este tópico, ver Maxwell: páginas 14 y 15 y Fleiss:páginas 11, 16 y 17).

2. Los Caracteres de la Asociación en una Tabla de 2x2

2.1. La Significación Estadística de la Asociación

Usualmente es necesario conocer si la aparición de un proceso anómalo social o de salud, en presencia de un factor condicionante, tienen o no una asociación detectable. Para lograr ese conocimiento se requiere primero determinar el nivel de la significación estadística de esa asociación, cualquiera sea el nivel de la misma; y a posteriori cuantificar la magnitud de la relación entre el factor condicionante y la aparición del fenómeno estudiado a fin de poder emitir juicio válido sobre la cuantía de sus relaciones.

Es evidente que si la asociación es estadísticamente significativa y que su magnitud es alta, a medida que ésta aumenta, aumentan las posibilidades lógicas de una eventual relación causal entre ambos fenómenos, o que al menos se incrementa la detección de un correlato entre ambos.

La búsqueda posterior de nuevos conocimientos, o el descubrimiento de otras áreas de análisis, se derivan fehacientemente a partir de estas determinaciones. Siegel, en su libro sobre estadísticas no paramétricas, da lineamientos sobre como lograr esas aproximaciones válidas y sobre la forma de determinar la significación de la asociación a través del test de dos o más muestras con caracteres independientes. Las llama "tests de dos (o varias) muestras independientes", refiriéndose a la independencia de las características que se contrastan y no de las muestras en si mismas.

Cuando se trata de dos muestras independientes se usa la distribución de Ji Cuadrado y se pone a prueba la hipótesis de nulidad (H_0) a través de la obtención de un Ji cuadrado observado que resulta de la sumatoria de los cuadrados de las diferencias entre los valores observados en cada celda y los teóricos que cada una de ellas debería tener, divididos por los valores esperados respectivos.

Posteriormente se compara este valor con el valor del Ji cuadrado crítico que se obtiene de la tabla de la distribución de Ji cuadrado, con (c-1) x (f-1) grados de libertad, para un determinado nivel de α. "c" está dado por el número de las columnas de la tabla; y "f" por su número de filas. Usualmente el nivel de α es del 0,05 o en casos de gran rigurosidad, del 0,01.

La fórmula para lograr ese Ji cuadrado es la siguiente:

$$X^2 = \sum_{i=1}^{c} \sum_{j=1}^{f} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad (1.4)$$

En la cuál "O_{ij}" es el número de casos observados en cada celda de la tabla; mientras que "E_{ij}" es el número de casos esperados en cada una de ellas si no hubiera asociación entre las variables bajo las condiciones establecidas en H_0. La sumatoria doble indica que se sumen los cuadrados de las diferencias entre ambos valores de todas las celdas.

Este Ji observado se compara con el Ji crítico, obtenido como se mencionó anteriormente, y se determina en la tabla correspondiente si esa diferencia es o no significativa.

Un ejemplo. En un área geográfica se estudia la posible asociación entre la condición de ocupación del jefe del hogar y el nivel de desnutrición de los menores de cinco años de cada hogar. Se encontraron los siguientes valores:

Ocupación del Jefe del Hogar

	Ocupado	Descup.	Total
Nutrido	(16,67) 20	(8,33) 5	25
Desnutrido	(13,33) 10	(6,67) 10	20
Total	30	15	45

(1.5)

Con los valores esperados redondeados al entero más próximo se obtiene el siguiente Ji cuadrado observado:

$$X^2 obs = \frac{(20,00-16,67)^2}{16,67} + \frac{(5,00-8,33)^2}{8,33} + \frac{(10,00-6,67)^2}{6,67} +$$

$$+ \frac{(10,00-13,33)^2}{13,33} = 4.49 \approx 4.5$$

Como el $X^2_{crit.(1\ gl,\alpha = 0,05)} = 3,84$

Consecuentemente, se rechaza la hipótesis nula y se dice que entre ambas variables hay asociación.

Otros autores, Maxwell por ejemplo, [2]/ presentan una fórmula de cálculo para llegar a iguales resultados con cómputos más simples. Sin aplicar la corrección de continuidad de Yates, y partiendo de la notación de la Tabla 3 (Modelo),la fórmula de cálculo es la siguiente:

$$X^2_{obs} = \frac{N(|ad - bc|)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (1.6)$$

[2]/ Maxwell, A.E.: "Analysing Qualitative Data"; Methuen's Monographs; Londres, Gran Bretaña; 3a.reimpres.; 1967; págs. 11 a 20.

Calculando con los valores del ejemplo anterior:

$$X^2_{obs} = \frac{45(|\ 5 \times 10 - 10 \times 20|)^2}{15 \times 30 \times 20 \times 25} = \frac{1.012.500}{225.000} = 4.5$$

Este valor es igual al obtenido por el método original reproducido por Siegel. Trabajando con celdas que tienen frecuencias pequeñas (valores menores de 5) el cálculo del Ji cuadrado debe obtenerse aplicando la **Corrección de Continuidad de Yates**. Ella se logra aplicando la siguiente fórmula: [3]/

$$X^2_{obs} = \frac{N(|\ ad - bc|) - N/2)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (1.7)$$

Aplicada al ejemplo anterior, el resultado es el siguiente:

$$X^2_{obs} = \frac{45\ (|5x10 - 10x20| - 22.5)^2}{15x30x20x25} = \frac{731.531,25}{225.000,00} = 3.25$$

Este valor es menor que el obtenido con la fórmula sin corrección de continuidad y hace que no se rechace la hipótesis nula y no se pueda afirmar que hay una asociación significativa entre ambas variables.

Una última estimación del Ji Cuadrado es la dada por la corrección de Mantel-Haenzel, quién usa sólo el cuadrado de la diferencia entre el **valor observado** en n_{11} y el **valor esperado para esa misma celda.** Este resultado se divide por el cociente entre los subtotales marginales divididos por el cuadrado del total por el total menos uno.

La expresión algebraica es la siguiente:

$$X^2_{obs} = \frac{(n_{11} - [(n_{.1} \times n_{1.})]/\ n..)^2}{(n_{.1} \times n_{.2} \times n_{1.} \times n_{2.})\ /\ [(n..^2) \times (n..-1)]} \qquad (1.8)$$

Con los mismos datos anteriores, el cálculo da el siguiente valor:

$$X^2_{obs} = \frac{(5 - [(15 \times 25)\ /\ 45])^2}{(15 \times 30 \times 25 \times 20)\ /\ [(45)^2 \times 44)]} = \frac{11,11}{2,52} = 4.40$$

Este valor es ligeramente inferior al del Ji Cuadrado sin corrección.

---

[3]/ Maxwell A.E.; "Analysing Qualitative Data"; loc. cit.; pag. 20 a 23.

2.2. Magnitud de la Asociación

Hasta ahora se explicó cómo determinar la significación estadística de la asociación de dos o más variables, sin embargo no se dieron instrumentos sobre cómo determinar la magnitud o el grado de esa asociación. Para lograr esto se deben usar métodos como el "Coeficiente ϕ", el "Riesgo Relativo", y la "Razón de Contigencia" (Odds Ratio). Algunos de estos estadísticos permiten calcular estimadores de punto y de intervalo con mayor o menor nivel de eficiencia. Cuando se considere cada uno de ellos se analizará este tema.

2.2.1. Coeficiente ϕ (Fi)

Un procedimiento muy popular en las ciencias sociales es el "Coeficiente ϕ", debido a la gran simplicidad para su cálculo y a la fácil interpretación del mismo.

Tiene el problema que no hay un estimador apropiado para su error estandar, hecho que le impide determinar los límites de confianza de su estimación.[4]/ Su cálculo se realiza partiendo de la formula del Ji Cuadrado sin la corrección de Yates, y es el siguiente:

$$\phi = \sqrt{(X^2_u / n_{..})}$$

$$\text{donde:} \quad X^2_u = \frac{n_{..}(n_{11}\, n_{22} - n_{12}\, n_{21})^2}{n_{.1} \times n_{.2} \times n_{1.} \times n_{2.}} \qquad (1.9)$$

En el ejemplo de la expresión (1.5), el valor del X^2_{obs} = X^2_u = 4,5; de modo que el valor de $\phi = \sqrt{(4{,}5/ 45)} = 0.32$.

Como regla general, dice Fleiss[5]/, valores que se encuentren por debajo 0,30 o 0,35 deben indicar una asociación muy baja, mientras que cifras superiores indican lo contrario. La ventaja de ϕ es que se puede captar rápidamente ya que su interpretación es igual a la del coeficiente de correlación r, que varía entre 0 y 1. El cero denota la ausencia completa de asociación, mientras que el 1 indica una asociación total.

2.2.2. Riesgo Relativo

Otra forma de medir la magnitud de asociación es a través de la valoración del **riesgo de enfermar** a que están sujetos los individuos **que están expuestos** a una característica o hábito específico, con relación a los sujetos **que no están expuestos** a esos caracteres o hábitos y que sin embargo sí **enferman**. Esto se llama el **"Riesgo Relativo"** (RR).

En realidad se trata de un problema de **"probabilidad condicional"** en lugar de uno de **"probabilidad simple"** de

[4]/ Edwards A.L.: "Statistical Methods for the Behavioral Sciences"; Rinehart & Co.,Inc.; 3d printing; New York; 1956.

[5]/ Fleiss, J.L.: "Statistical Methods for Rates and Proportions"; loc. cit.; pag 42.

ocurrencia de un evento. Por ejemplo, una cosa es la probabilidad de tener un hijo de bajo peso al nacer, en una comunidad dada (probabilidad simple), y otra diferente es la de tener un hijo de bajo peso entre las madres menores de 20 años de esa misma comunidad (probabilidad condicional). [6]/

Esto es lo que en el área epidemiológica se trata como "tasa general" y como "tasa específica". En este caso la probabilidad de tener un hijo de bajo peso en la población total es igual a: (Ver tabla 1.3)

$$\text{(Tasa General): } \underline{\text{Tasa de bajo peso al nacer}} = \frac{\text{Nro. de casos en } (a+c)}{N}$$

Este valor es conocido como la tasa general de bajo peso al nacer en la población total (N). Por el contrario si se desea conocer la probabilidad de tener un hijo de bajo peso entre las madres menores de 20 años, (en el subuniverso "a+b" de esa población total) (Ver tabla 1.3).

$$\text{(Tasa específica): } \underline{\text{Tasa específica de bajo peso al nacer en madres} < 20 \text{ años}} = \frac{\text{Nro. de casos en "a"}}{a + b}$$

Lo que en estadísticas vitales se conoce como "tasa específica", en teoría de probabilidades se conoce como "probabilidad condicional":

P(B/A) = probabilidad de tener un hijo de bajo peso dado que proviene de una población de madres menores de 20 años de edad.(Probabilidad de B/A) [7]/

Que se entiende, entonces, por probabilidad condicional? Es la probabilidad que acaezca un evento en una población específica, dada la presencia o ausencia de una condición que puede modificar esas probabilidades. En el ejemplo presentado, las posibilidades de tener un hijo de bajo peso en madres menores de 20 años es mayor que las que tienen las madres de 20 a 40 años. De alguna manera la condición de tener cierta edad determina en este caso el nivel de probabilidades del evento.

La ecuación que define esta probabilidad es la siguiente:

$$P(B/A) = \frac{P(B \quad A)}{P(A)} \quad \text{(Si A no es igual a 0)} \qquad (1.10)$$

En esta expresión A es la totalidad del espacio de la submuestra de madres menores de 20 años. En el caso de una tabla de 2 x 2 esta submuestra es la totalidad de los casos correspondientes a A, estén o no en presencia de B (que tengan o no hijos de bajo peso al nacer: ("B + No B") probabilidad igual a $p_{11} + p_{12} = p_{1.}$ Por el contrario, su complemento es "No A", la

---

[6]/ Mosteller F., Rourke R., Thomas G.: "Probability and Statistics"; Addison-Wesley Publishing Co.; Massachusetts, USA; 1961; pags. 85 a 91.

[7]/ Fleiss, J.L.: loc cit.; sección 1.1.

probabilidad de tener o no hijos de bajo peso en madres de más de 20 años de edad, esto es $p_{21} + p_{22} = p_{2.}$

El espacio muestral total es igual a $p_{1.} + p_{2.}$ o a $p_{.1} + p_{.2}$ según se sumen los subtotales de las columnas o de las filas, respectivamente.

La probabilidad de tener un hijo de bajo peso al nacer, en madres menores de 20 años, (se lee como la probabilidad de tener B dado A), se expresa por el cociente entre los hijos de bajo al nacer, provenientes de madres menores de 20 años; (se lee "B" intersección "A"), sobre "A" (total de madres menores de 20 años). La notación de esa tabla de 2x2 es la que se muestra en (1.11). En (1.12) se ven valores numéricos que las ejemplifica.

Edad	Peso		
	B	No B	Total
A	p_{11}	p_{12}	$p_{11}+p_{12}$
No A	p_{21}	p_{22}	$p_{21}+p_{22}$
Total	$p_{11}+p_{21}$	$p_{12}+p_{22}$	1.00

(1.11)

En (1.11) se debe reconocer que:

$$p_{11} + p_{12} = p_{1.} = P(A) \quad ; \quad p_{21} + p_{22} = p_{2.} = P(\bar{A}) = P(\text{No } A)$$

$$p_{11} + p_{21} = p_{.1} = P(B) \quad ; \quad p_{12} + p_{22} = p_{.2} = P(\bar{B}) = P(\text{No } B)$$

De esto se puede deducir que:

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{p_{11}}{p_{11} + p_{12}} = \frac{p_{11}}{p_{1.}}$$

y análogamente:

$$P(B/\bar{A}) = \frac{P(\bar{B} \cap \bar{A})}{P(\bar{A})} = \frac{p_{21}}{p_{21} + p_{22}} = \frac{p_{21}}{p_{2.}}$$

En (1.12) los valores numéricos originales son:

Edad	Peso		
	-2500 gr	2500 o +	Total
- 20 años	40	50	90
20 o +	5	90	95
Total	45	140	185

(1.12)

Estas dan los siguientes valores porcentuales:

Edad	Peso		
	-2500 gr.	+2500 gr.	Total
- 20 años	0,2162	0,2703	0,4865
20 o +	0,0270	0,4875	0,5135
Total	0,2432	0,7568	1,0000

(1.13)

Esta asociación, el riesgo relativo, se mide dividiendo la proporción de los que enfermaron y que estuvieron expuestos a la condición (p_{11}) sobre el total de los que estuvieron expuestos a ese factor ($p_{1.}$); todo esto dividido por los que enfermaron y no estuvieron expuestos (p_{21}), sobre todos los no expuestos ($p_{2.}$). En términos de probabilidad condicional el riesgo relativo se expresa de la siguiente manera:

$$RR = \frac{P\ (B/A)}{P\ (B/\bar{A})} \qquad (1.14)$$

El riesgo relativo (RR) está dado por el cociente entre la P(B/A) y la P(B/A). Estas probabilidades se calculan de la siguiente manera:

$$P(B/A) = P_{11}\ /\ P_{1.}\ ; \qquad P(B/\bar{A}) = P_{21}\ /\ P_{2.}$$

La estimación del RR del universo se logra a través del cálculo del "rr" de la muestra, de la siguiente manera:

$$rr = \frac{p_{11}\ /\ p_{1.}}{p_{21}\ /\ p_{2.}} = \frac{p_{11} \times p_{2.}}{p_{21} \times p_{1.}}$$

Otro ejemplo para determinar el riesgo relativo con las proporciones de la tabla (1.13):

$$rr = \frac{0{,}2162 \times 0{,}5135}{0{,}0270 \times 0{,}4865} = \frac{0{,}1110}{0{,}0131} = 8{,}473 \simeq 8{,}47$$

El mismo valor, salvo por el redondeo de los decimales, se obtiene si se trabaja con las frecuencias absolutas de la tabla (1.12):

$$rr = \frac{40 \times 95}{5 \times 90} = \frac{3.800}{450} = 8{,}444 \approx 8{,}44$$

Cuál es el significado de este valor? Indica que la probabilidad de tener un nacido vivo de bajo peso, en presencia del factor edad materna menor de 20 años, es más de ocho veces mayor que la de tenerlo no estando expuesto al riesgo (que la madre tenga 20 o más años de edad).

Estos valores son importantes ya que en este ejemplo hay una asociación altamente significativa entre los riesgos de tener hijos de bajo peso al nacer entre madres de menos de 20 años de edad (X^2= 36,44 usando la corrección de Yates; o 38,54 sin ella). Esta cifra indica una diferencia realmente evidente entre los grupos de edades maternas como para orientar o inducir la generación de políticas y programas específicos.

Otro ejemplo para contrastar lo expuesto. Si no hubieran diferencias en el número de casos estando o no presente el factor interviniente, podría tenerse una tabla hipotética como la siguiente:

Edad	Peso		
	-2500 gr	+2500 gr	Total
- 20 años	10	20	30
20 o +	10	30	40
Total	20	50	70

(1.14)

Los valores porcentuales de esa tabla serían los siguientes:

Edad	Peso		
	-2500 gr	+2500 gr	Total
- 20 años	0,1429	0,2857	0,4286
20 o +	0,1428	0,4286	0,5714
Total	0,2857	0,7143	1,0000

(1.15)

El cálculo del riesgo relativo con esas proporciones de la tabla 1.15 es el siguiente:

$$rr = (0,1429 \times 0,5714) \,/\, (0,1428 \times 0,4286) =$$
$$= 0,0817 \,/\, 0,0612 = 1,3349 \approx 1,33$$

El mismo cálculo, pero considerando ahora los valores absolutos (tabla 1.14), es el siguiente:

$$rr = 10 \times 40) \,/\, (10 \times 30) = 400 \,/\, 300 = 1,3333 \approx 1,33$$

En este ejemplo, la interpretación de este resultado sería que no hay asociación entre el bajo peso al nacer y el factor analizado (edad de la madre), ya que en ambas situaciones, con o sin factor interviniente, el número de bajos pesos es casi el mismo. (Practicamente la relación es de 1 a 1 (1,3 a 1,0).

2.2.3. Razón de Contingencia u "Odds Ratio"

Hay otra forma de aproximarse al conocimiento de la asociación, y es a través del riesgo relativo aproximado o también llamado la razón de contingencia. En los textos en Inglés se la denomina "odds ratio"; que no es otra cosa que la razón de los productos cruzados. Esta expresión se refiere a la razón de la probabilidad de tener una enfermedad dada un cierta condición, y la probabilidad de no tenerla también en presencia de la misma condición.

"Odds" en inglés tiene las siguientes acepciones: "magnitud de la diferencia por la cual una característica excede o es menor que otra"; "cantidad en exceso o defecto"; "diferencia en favor de una entre dos circunstancias opuestas" etc. Consecuentemente, la razón de los productos cruzados es la que tiene en el numerador el producto de los valores de los casos que manifestaron el evento en presencia del factor condicionante, por el de no tenerlo en ausencia de ese factor; mientras que en el denominador se encuentran como multiplicando el no tener el evento en presencia del factor y como multiplicador el tenerlo en ausencia del factor.

Para ejemplificar este último párrafo, valga una aplicación simplificada de la tabla (1.3):

	B	$\bar{B}$	Total
A	a	b	a+b
$\bar{A}$	c	d	c+d
Total	a+c	b+d	N

(1.16)

$$\text{"or" o } rc = \frac{a \times d}{b \times c}$$

Esto quiere decir que la razón de consistencia (u "Odd Ratio") es igual al cociente de los productos de (a x d) / (b x c), que no es otra cosa que lo que se describió en el párrafo anterior.

Goodman y Kruskal, citados por Fleiss [8]/, presentan estas medidas de la magnitud de la asociación no basadas en la distribución de Ji Cuadrado. Sin embargo son tributarias, también, de las probabilidades condicionales.

$$\Omega_A = \frac{P(B/A)}{P(\bar{B}/A)}$$

Esta expresión mide la probabilidad de acaecimiento de un evento cuando el factor antecedente está presente.

[8]/ Goodman & Kruskal: citados por Fleiss, J.L.: op. cit./ pag.43.

Como una estimación de la P(B/A) está dada por la expresión:

$$P(B/A) = p_{11} \;/\; p_{1.} = 0{,}059 \;/\; 0{,}353 = 0{,}167$$

y la de
$$P(B/\bar{A}) = p_{12} \;/\; p_{1.} = 0{,}294 \;/\; 0{,}353 = 0{,}833$$

tenemos entonces $0{,}167 \;/\; 0{,}833 = 0{,}20$.

Consecuentemente la probabilidad de Ω_A se estima como:

$$RC_A = \frac{p_{11} \;/\; p_{1.}}{p_{12} \;/\; p_{1.}} = p_{11} \;/\; p_{12}$$

(La estimación de Ω_A es igual a la de RC_A, ambas iguales a 0,20). Usando los datos de la tabla (1.13) se tiene:

$$P(B/A) = 0{,}059 \;/\; 0{,}353 = 0{,}167 = 1/6.$$

Con los valores absolutos (1.12) el resultado es igual:

$$P(B/A) = 10 \;/\; 60 = 0{,}167 = 1/6.$$

Esto es, que por cada seis nacimientos, cinco lo serán de peso normal, mientras que uno será de bajo peso.

Cuando el factor condicionante A no está presente, la probabilidad de su complemento está estimada por:

$$\Omega_{\bar{A}} = \frac{P(B/\bar{A})}{P(\bar{B}/\bar{A})} \quad \text{y} \quad RC\bar{A} = \frac{p21 \;/\; p2.}{p22 \;/\; p2.} = p21 \;/\; p22$$

La medida de asociación más frecuentemente usada, basada en la razón de Ω_A y $\Omega_{\bar{A}}$, es su cociente:

$$w = \frac{\Omega_A}{\Omega_{\bar{A}}}$$

Que normalmente se estima por:

$$rc = \frac{O_A}{O_{\bar{A}}} = \frac{p_{11} \;/\; p_{12}}{p_{21} \;/\; p_{22}} = \frac{p_{11} \cdot p_{22}}{p_{12} \cdot p_{21}}$$

Valor que no es otra cosa que la "razón de los productos cruzados" o "razón de contingencia" (u "odds ratio"). La interpretación de estos valores es la siguiente: si la probabilidad de tener el evento B dado A es igual que la tener B dado $\bar{A}$, esto implica que $w = 1$ y que no hay asociación entre ambos (condicionante y evento). Si $w > 1$ implica que Ω_A es mayor que $\Omega_{\bar{A}}$ y se dice que si hay asociación y que ésta es mayor cuanto más se distancia de la unidad ($w = 4$ indica una asociación 4 veces mayor, o de 4 a 1). Si w < 1 se dice que también hay asociación pero que acá ésta es mayor con $\bar{A}$ y con $\bar{B}$ que con A y B. Esta situación es interesante ya que en esos casos las conclusiones tiene un sentido opuesto. En el ejemplo numérico que continúa y en su interpretación se aclara este hecho.[9]/

Sea el caso presentado en la tabla (1.13), entonces la razón de contingencia será igual a:

$$rc = \frac{0,2126 \times 0,4875}{0,2703 \times 0,0270} = \frac{0,1054}{0,0073} = 14,438 \approx 14,4$$

Con los valores absolutos el resultado es el mismo:

$$rc = \frac{40 \times 90}{5 \times 50} = \frac{3.600}{250} = 14,4$$

El análisis de esta situación indica que el factor en consideración favorece la aparición de la enfermedad en una relación de catorce a uno, con referencia a la aparición de la enfermedad sin que el factor esté presente. Si se da una situación diferente en que los valores sean los que aparecen en la tabla (1.17):

Vacunas	Enfermedad		
	Presente	Ausente	Total
Dadas	10	20	30
No dadas	40	10	50
Total	50	30	80

(1.17)

$$rc = \frac{10 \times 10}{20 \times 40} = \frac{100}{800} = 0.125 = 1 / 8$$

Acá se da el caso que la enfermedad está asociada con la vacunación, pero en un sentido inverso al caso anterior, ya que su ausencia no evita la aparición de la enfermedad. Por lo tanto la interpretación de este resultado es la siguiente: en una población sin vacunaciones las posibilidades de enfermar son ocho veces mayores que cuando la población está vacunada.

Este efecto protector se calcula de la siguiente manera:

$$1/rc = 1.000 / 0,125 = 8$$

[9]/ Fisler, J.L.; op. cit.; pgs. 44-45.

2.2.4. Riesgo Diferencial

El **riesgo diferencial** indica la magnitud de la diferencia entre el riesgo de enfermar estando expuesto con relación al riesgo similar pero no estando expuesto. Se calcula de la siguiente manera:

$$rd = (a/a+b) - (c/c+d)$$

Si no existiera ningún factor externo que pudiera explicar esa diferencia, ésta podría llamarse "riesgo atribuíble", como suelen hacerlo algunos epidemiólogos.También llegan a denominarla "fracción prevenible".

Pero estas denominaciones disminuyen las características de la investigación de ese riesgo ya que pueden ser muchos los factores externos que suelen ser responsables por esas diferencias, y no el hecho de estar o no expuesto al riesgo en consideración.[10]/ Si se usa el ejemplo de la tabla (1.12) los valores serían:

$$rd = (40/90) - (5/95) = 0,444 - 0,053 = 0,391$$

Esto puede implicar que en ausencia de otros factores explicativos, si la exposición fuera la única responsable por las diferencias, el tener más de 20 años de edad daría un 39% más de nacimientos de peso normal, que los hijos de las madres de menos de 20 años.

Si se consideran los valores de sus probabilidades, tabla (1.13), los resultados serían los siguientes:

$$rd = (0,2162/0,4865) - (0,0270/0,5153) = 0,444 - 0,053 = 0,391$$

Un caso diferente se plantea si se usan los datos de la tabla (1.17), en este caso es:

$$rd = (10/30) - (40/50) = 0,33 - 0,80 = - 0,47$$

Cuya interpretación sería la siguiente: la inmunización por medio de la vacunación da un 47 % menos de casos de enfermedad que los que se registran entre no vacunados.

Un comentario final. Esta no es una medida de riesgo que amerite mayores comentarios; son muy pocas las veces en que realmente puede asignarse valor a este indicador debido a la necesidad de eliminar todo otro factor interviniente que sea eventualmente responsable por esta diferencia atribuíble. Si bien este criterio vale para todas las medidas de comparación de riesgos, en este caso adquiere una relevancia particular.

[10]/ Andersson, N.: loc. cit., págs. 33 y 40/41.

2.3. Variabilidad de los estimadores de riesgo

Es importante calcular el error estandar de la razón de contingencia [e.e.(rc)], el que se realiza mediante las siguientes fórmulas:

$$e.e._{(rc)} = \frac{rc}{\sqrt{n_{..}}} \sqrt{\left(\frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}}\right)} \qquad (1.18)$$

$$e.e._{(rc)} = rc \sqrt{\left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)} \qquad (1.19)$$

Usando los datos de las tablas (1.12) y (1.13) los valores del error estandar son los siguientes:

$$e.e._{(rc)} = \frac{8,4}{\sqrt{185}} \sqrt{\left(\frac{1}{0,2162} + \frac{1}{0,2703} + \frac{1}{0,0270} + \frac{1}{0,4865}\right)} =$$

$$= 0,6176 \times 6,8860 = \mathbf{4,2528 \approx 4,25}$$

$$e.e._{(rc)} = 8,4 \sqrt{\left(\frac{1}{40} + \frac{1}{50} + \frac{1}{5} + \frac{1}{90}\right)} =$$

$$= 8,4 \sqrt{0,2565} = \mathbf{4.2546 \approx 4,25}$$

Esta es una estimación de buena aproximación al valor del error estandar verdadero de la razón de contingencia u odds ratio.

Hay otras maneras de estimar la variancia y el error estándar de la razón de contingencia (rc). Trabajando con el logaritmo natural (base e), la variancia del $\log_e$(rc) se obtiene de la siguiente manera:

$$Var(\log_e rc) = 1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22}$$

El valor de la variancia o del error estandar se logran sacando el anti-logaritmo natural del resultado. El error estandar es la raiz cuadrada de la varianza. [11]/

El intervalo de confianza para cualquier estimación de RC está dado por el valor estimado del RC correspondiente ± 1.96 x $e.e._{(rc)}$. Si bien ésta es una estimación en que se usa el valor de dos desvíos estándares de la distribución normal, los valores que se obtienen son aceptables para el cálculo del intervalo de confianza.

Otro ejemplo con este procedimiento de estimación se presenta a continuación.

[11]/ Armitage,P.: "Statistical Methods in Medical Research"; op cit.; pag.429.

Vacunas	Enfermedad		Total
	Presente	Ausente	
Dadas	15	20	35
No dadas	10	50	60
Total	25	70	95

(1.20)

rc = (15 x 50) / (10 x 20) = 3.75 ; Log$_e$ (rc) = 1.3218

Var.(log$_e$ rc) = 1/15 + 1/20 + 1/10 + 1/50 = 0.2367

e.e.(log$_e$ rc) = $\sqrt{0,2367}$ = 0,4865

L.C.(inf. y sup) de log$_e$ rc = 1,3218 ± (1,96 x 0,4865) =
= 0,3683 < log$_e$ rc < 2,2753

Sacando el antilog$_e$, el LC para RC es: **1,45 < RC < 9,73.**

3. La Asociación en Tablas de Contingencia de 2 x k

No siempre se trabaja en tablas de contingencia de 2 x 2 sin que algunas veces se requieren más celdas. Como ejemplo sea el caso siguiente: presencia de huevos de Taenia Saginata en heces de personas clasificadas por grupos de edad:

Edad	0-9	10-19	20-29	30-39	40-49	Total
Positivos	30	20	15	18	20	103
Negativos	40	35	38	76	70	259
Total	70	55	53	94	90	362

(1.21)

La denominación genérica de estos valores es:[12]/

Edad	1	2	3	i	k	Todos los grupos combinado
Positiv.	r_1	r_2	r_3	r_i	r_k	R
Negativ.	$n_1 - r_1$	$n_2 - r_2$	$n_3 - r_3$	$n_i - r_i$	$n_k - r_k$	N-R
Total	n_1	n_2	n_3	n_i	n_k	N
Prop.pos.	p_1	p_2	p_3	p_i	p_k	P = R/N

[12]/ Armitage,P.: loc. cit., pags. 207/211.

Para cada grupo se analizaron procedimientos para determinar la existencia o inexistencia de asociación entre la enfermedad o problema social y el factor de riesgo en cuestión; si se la detecta, se determina la magnitud de la asociación por medio de diferentes indicadores: coeficiente fi, razón de los productos cruzados o razón de contingencia ("odds ratio"), riesgo relativo, etc.

El paso siguiente de este proceso de análisis consiste en la búsqueda de evidencias de asociación cuando se trabaja con varias de esas tablas o estratos. Por ejemplo, si se trabaja con varios sitios centinela de un área geográfica dada, y si de todos ellos se obtienen valores sobre las asociaciones de dos variables, ambas iguales en cada uno de ellos, las preguntas que se deben formular deben dirigir el proceso de análisis hacia la determinación de: 1) el nivel de asociación de esas variables es consistente entre los sitios?; 2) si el nivel de asociación entre sitios es consistente, el grado de asociación común es significativo, desde un punto de vista estadístico?; y 3) asumiendo que éste es significativo, cuál es la mejor forma de estimar la magnitud de esa asociación y la medida de su variación?

Antes de presentar los procedimientos operacionales para el cálculo de la homogeneidad y la asociación conjunta entre los diferentes grupos, es importante hacer consideraciones sobre ciertos componentes de la teoría del Ji cuadrado. Se sabe que:

$$\text{El } X^2_{total} = X^2_{homog.} + X^2_{asoc.}$$

Consecuentemente:

$$X^2_{homog.} = X^2_{total} - X^2_{asoc.}$$

$$X^2_{asoc.} = X^2_{total} - X^2_{homog.}$$

En general puede decirse que:

$$X^2_{total} = \sum_{i=1}^{g} X^2_i = \sum_{i=1}^{g} w_i \, m_i^2 \qquad (1.24)$$

Donde m_i puede ser las (rc), los $\log_e$ (rc), las d_i

[diferencias estandarizadas: $(p_{i1} - p_{i2}) / (\bar{p}_i \, \bar{q}_i)$], etc. y w_i es la inversa del cuadrado del error estándar de m_i.

Para el caso en que $m_i = \log\ (rc_i)$:

$$w_i = \frac{1}{\{e.e.[\log_e(rc_i)]\}^2} \qquad (1.25)$$

El X^2_{total} tiene una distribución de X^2 con g grados de libertad ya que los g grupos son asumidos como independientes.

$$\text{El e.e.}[\log_e(rc_i)] = \sqrt{\left(\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}\right)} \qquad (1.26)$$

De esta manera, las fórmulas computacionales de los diferentes componentes del X^2 son las siguientes:

$$X^2_{total} = \sum_{i=1}^{g} w_i \, [\log_e(rc_i)]^2 \qquad (1.27)$$

$$X^2_{asoc.} = \frac{\left(\sum_{i=1}^{g} w_i \, [\log_e(rc_i)] \right)^2}{\sum_{i=1}^{g} w_i} \qquad (1.28)$$

$$X^2_{homog.} = \sum_{i=1}^{g} w_i \, [\log_e(rc_i)]^2 - \frac{\left[\sum_{i=1}^{g} w_i \log_e(rc_i) \right]^2}{\sum_{i=1}^{g} w_i} \qquad (1.29)$$

En estas ecuaciones el valor del X^2_{total} está dado por (1.27), mientras que el $X^2_{homog.}$ es el valor de (1.29), y finalmente el $X^2_{asoc.}$ es el de (1.28). Como se observa, no son mas que distintas expresiones de las operaciones algebraicas en relación con la igualdad:

$$X^2_{total} = X^2_{asoc.} + X^2_{homog.}$$

Cuando se tienen varios estratos y dentro de cada uno de ellos a varios grupos u otra unidad de muestreo, y cuando se realizan mediciones a todos los grupos de cada una de esas unidades; y cuando los datos vienen agrupados en cruces de tablas de contingencia del tipo de 2 x 2, entonces es indispensable conocer el estimador de la **"razón de contingencia conjunta"**, y además las **"razones de contingencia de los estratos"**.

Esto permite detectar la cuantía de las diferencias estadísticas entre los estratos, y además, entre los diferentes grupos dentro de los propios estratos, si se perciben discrepancias entre ellos.

La forma de estimar esa razón de contingencia conjunta se logra mediante la aplicación de la siguiente fórmula:

$$\overline{rcc} = \frac{\sum_{i=1}^{g} (a_i \, d_i \, / \, n_i)}{\sum_{i=1}^{g} (b_i \, c_i \, / \, n_i)} \qquad (1.30)$$

Los X^2 observados: total, de heterogeneidad y de asociación se computan de la manera siguiente:

# Orden	Sitio	Talla Baja — Estrato — Bajo (a_i)	Talla Baja — Estrato — Alto (b_i)	Talla Alta — Estrato — Bajo (c_i)	Talla Alta — Estrato — Alto (d_i)	($a_i d_i$)	($b_i c_i$)	rc_i
1	01	15	6	8	20	300	48	6,25
2	02	24	10	8	30	720	80	9,00
3	22	6	15	9	19	114	135	0,84
4	24	10	12	10	27	270	120	2,25
	Total	55	43	35	96	1.404	383	

(1.31)

rc_i	Lrc_i	A $1/a_i + 1/b_i + 1/c_i + 1/d_i$	$w_i = 1/A$	$w_i \times Lrc_i$	$w_i \times (Lrc_i)^2$	n_i	$\frac{a_i d_i}{n_i}$	$\frac{b_i c_i}{n_i}$
6,25	0,7959	0,4084	2,4486	1,9488	1,5511	49	6,12	0,98
9,00	0,9542	0,3000	3,3333	3,1806	3,0350	72	10,00	1,11
*0,84	-0,0735	0,3971	2,5183	**0,1851	0,0136	49	2,33	2,76
2,25	0,3522	0,3203	3,1221	1,0996	0,3873	59	4,58	2,03
			11,4223	6,0439	4.9870	229	23,03	6,88

* = 0,8444 ; ** = -0,1851 (1.32)

En las tablas presentadas se usa el log base 10 y se procede a su transformación en log base e en pasos sucesivos.

El estimador del $\overline{rcc}$, según Mantel Haenzel es:

$$\overline{rcc} = 23,03 \,/\, 6,88 = 3,3474 \approx 3,35 \qquad (1.33)$$

Es necesario destacar que el $\overline{rcc}$ según Mantel-Haenzel no tiene un estimador de sub variancia, mientras que el log. rc si la tiene. La media ponderada del log rc_i se obtiene mediante la siguiente fórmula:

$$\overline{L}\, rc_i = \frac{\sum_{i=1}^{g} w_i \log rc_i}{\sum_{i=1}^{g} w_i} = 6,0439 \,/\, 11,4223 = 0.5291 \qquad (1.34)$$

Su antilog da otra estimación de la rcc:

$$\text{Antilog } \overline{L}\, rc_i = \text{antilog } 0,5291 = 3,3814 \approx 3,38$$

Valor que es equivalente al presentado en (1.33).

Por su parte, la varianza de la media ponderada del log de las rc_i se logra mediante la fórmula:

$$\text{Var } \bar{L}\ rc_i = 1 / (2.3026)^2 \times \sum_{i=1}^{g} w_i = \qquad (1.35)$$
$$= 1 / (5,3020 \times 11,4223) = 0,0165$$

Si se desea un intervalo de confianza para el estimador de un 95 %, entonces:

$$0.5291 \pm (1.96) \sqrt{0,0165} = 0,5291 \pm 0.2518 \qquad (1.36)$$

Los valores sin corrección son: 0,2773 y 0,7809

Sus antilog dan: 1.894 y 6.038

De donde: $1.89 < \overline{rcc} < 6,04$

Ya se dijo anteriormente que el $X^2_{total} = \sum_{i=1}^{g} w_i \log rc_i^2$, de donde su estimación es la siguiente:

$$X^2_{total} = (2.3026)^2 \times 4,9870 = 26,4409 \approx 26,44 \qquad (1.37)$$

$$X^2_{crit.(gl=3, \alpha=0,05)} = 7.81$$

Este resultado es altamente significativo ($P < 0.001$). Esto es indicativo que las condiciones de exposición y los riesgos de enfermar de los cuatro estratos tomados en su conjunto muestran una fuerte asociación.

$$\text{El } X^2_{homog.} = \sum_{i=1}^{g} w_i (\log rc_i)^2 - [(\sum_{i=1}^{g} w_i \log rc_i)^2 / \sum w_i] \qquad (1.38)$$

En este caso, incorporando la corrección de log natural, se tiene la siguiente fórmula de cálculo:

$$X^2_{homog.} = (2,3026)^2 (4,9870 - [(6,0439)^2 / 11,4223]) =$$
$$= 5,3020 \ [4,9870 - (36,5287 / 11,4223)] =$$
$$= 5,3020 \ (4,9870 - 3,1980) = 5,3020 \times 1,7890 =$$
$$= 9.49$$

El $X^2_{crit.(gl=3,\ \alpha=0,05)} = 7,81$

Como el X^2_{homog} observado tiene una probabilidad que está entre $0,025 < P < 0.010$, podría afirmarse que hay heterogeneidad entre las rc_i de los estratos.

La estimación del $X^2_{asoc.}$ es igual al cociente de la rcc y su error estandar, todo elevado al cuadrado. La fórmula sería la siguiente:

$$X^2_{asoc.} = [\bar{L} / ee(\bar{L})]^2 \ ; \quad \text{donde:} \qquad (1.39)$$

$$\bar{L} = (\sum_{i=1}^{g} w_i \log rc_i) / (\sum_{i=1}^{g} w_i) \ ; \quad ee(\bar{L}) = 1 / (\sqrt{\sum_{i=1}^{g} w_i})$$

$$\bar{L} = 6{,}0439 / 11{,}4223 = 0{,}5291$$

$$ee\bar{L} = 1 / \sqrt{11.4223} = 1/ 3{,}3797 = 0{,}2959$$

$$X^2_{asoc.} = (0{,}5291 / 0{,}2959)^2 \times 2.3026^2 = $$
$$= 3{,}1973 \times 5{,}3020 = 16{,}9521 \approx 16{,}95$$

El $X_{crit.(gl=1,\alpha=0,05)} = 7{,}81$

Como la probabilidad del $X^2_{obs.}$ es < que 0,001 esto indica que si hay una fuerte asociación común entre las rc de los estratos . Dicho de otro modo, el valor hallado para el Ji Cuadrado de asociación indica que tomando el promedio de los cuatro grupos, existe una fuerte asociación entre el factor de riesgo y la ocurrencia de la condición estudiada.

6. Contrastes entre Grupos

Una primera forma de hacerlo, siguiendo los caracteres dados anteriormente, es aplicar el análisis de homogeneidad en cada uno de los estratos que constituyen la totalidad antes analizada. A los fines de este ejemplo, dada la estructura de los datos se diferencian dos estratos: el primero conteniendo a los grupos 1 y 2, y el otro con los 3 y 4.

En la primera de estas agrupaciones hay una asociación evidente entre la exposición al riesgo y la posibilidad de enfermarse; en la segunda, en cambio, ésta no es evidente. Consecuentemente la lógica indicaría que deberían esperarse resultados de un alto rcc en el primer grupo, mientras que el segundo éste debería ser bajo; el X^2_{total} del primero debería dar un valor elevado, mientras que el del segundo debería ser muy bajo; con relación al $X^2_{homog.}$ no se esperaría sea muy diferente entre ambos grupos; y por último, sí debería haber una asociación diferente ya que la rcc del primer grupo se espera sea alta mientras que la del segundo sea baja o casi inexistente.

6.1. Análisis del Primer Estrato

# orden	Sitio	Talla Baja Estrato Bajo (a_i)	Talla Baja Estrato Alto (b_i)	Talla Alta Estrato Bajo (c_i)	Talla Alta Estrato Alto (d_i)	(a_i d_i)	(b_i c_i)	rc_i
1	01	15	6	8	20	300	48	6,25
2	02	24	10	8	30	720	80	9,00
	Total	39	16	16	50	1,020	128	

(1.39)

rc_i	Lrc_i	A $1/a_i$ + $1/b_i$ + $1/c_i$ + $1/d_i$	$w_i = 1/A$	w_i x Lrc_i	w_i x Lrc_i^{2}	n_i	$a_i d_i$ n_i	$b_i c_i$ n_i
6,25	0,7959	0,4084	2,4486	1,9488	1,5511	49	6,12	0,98
9,00	0,9542	0,3000	3,3333	3,1807	3,0350	72	10.00	1,11
			5,7819	5.1295	4,5861	121	16,12	2,09

(1.40)

El estimador de la razón de contingencia conjunta de Mantel-Haenzel, en este grupo es el siguiente:

$$\overline{rcc} = 16,12 \;/\; 2,09 = 7.71 \qquad (1.41)$$

Una confirmación del valor del $\overline{rcc}$ encontrado por el método de Mantel Haenzel se logra de la manera siguiente:

Antilog $\overline{L}\ rc_i$ = antilog 0,8872 = 7,71

Para establecer el intervalo de variación esperado, con un 95% de confianza, hay que calcular la media ponderada del log rc_i, que según lo visto anteriormente, está dada por:

$\overline{L}\ rc_i = 5,1295 \;/\; 5,7819 = 0,8872$

Var $\overline{L}\ rc_i = 1 \;/\; [(2,3026)^2 \times 5,7819] = 1/30,6554 = 0,0326$

En consecuencia los límites del intervalo citado serían:

$$0,8872 \pm (1,96) \times \sqrt{0,0326} = 0,8872 \pm 0,3540 \qquad (1.42)$$

$$0,5332 < \overline{rcc} < 1,2412$$

Obteniendo sus antilogaritmos, el valor real será:

$$3,41 < \overline{rcc} < 17,42$$

Que es un valor semejante al estimador antes obtenido.

El cálculo del $X^2_{total} = (2,3026)^2 \times 4,5861 = 24,3155$
$\approx 24,32$ (1.43)

El $X^2_{crit.(gl=1,\alpha=0.05)} = 3,84$

Como la probabilidad del $X^2_{obs.}$ es < que 0,001, por lo tanto hay una diferencia notablemente significativa entre las posibilidades de enfermarse estando expuesto al riesgo en cuestión, y de no contraer la enfermedad, en ausencia de él, juntando todas las tablas de 2x2 del estrato.

Con relación al $X^2_{homog.}$ su expresión numérica es:

$X^2_{homog.} = (2,3026)^2 \times [4,5861 - (5,1295)^2 / 5,7819] =$
$= 5,3020 \times (4,5861 - 4,5507) =$
$= 5,3020 \times 0,0354 = 0.1877 \approx 0,19$ (1.44)

El $X^2_{crit.(gl=1, \alpha=0,05)} = 3,84$.

Por lo que se dice que no hay heterogeneidad entre las asociaciones de las tablas de 2x2 analizadas. Es lógico ya que la probabilidad del $X^2_{obs.}$ está entre: $0,500 < P < 0,250$

El $X^2_{asoc.}$, dado por el cociente entre $\bar{L}$ y el ee $\bar{L}$, es:

$$\bar{L} = 5,1295 / 5,7819 = 0,8872$$

$$ee\bar{L} = 1 / \sqrt{5,7819} = 1 / 2,4046 = 0,4159$$

Entonces, $X^2_{asoc.} = (0,8872 / (0,4159)^2 = 4,5506$

El antilog. es: $4,5506 \times 5.3020 = 24,1272 \approx 24,13$ (1.45)

Como el $X^2_{crit.(gl=1, \alpha=0,05)} = 3,84$ se dice que hay una diferencia significativa entre la posibilidad de contraer la enfermedad estando o no expuesto al factor de riesgo promedio de las tablas de 2x2 del estrato. En realidad la probabilidad de 24,13 es menor que 0,001. La diferencia demuestra una fuerte significación.

Por último: 24,13 + 0,19 = 24,32. O sea que numéricamente se corrobora que el $X^2_{total} = X^2_{asoc.} + X^2_{homog.}$

6.2. Análisis del Segundo Estrato

# Orden	Sitio	Talla Baja —Estrato— Bajo (a_i)	Talla Baja —Estrato— Alto (b_i)	Talla Alta —Estrato— Bajo (c_i)	Talla Alta —Estrato— Alto (d_i)	$(a_i d_i)$	$(b_i c_i)$	rc_i
3	22	6	15	9	19	114	135	0,84
4	24	10	12	10	27	270	120	2,25
	Total	16	27	19	46	384	255	

(1.46)

rc_i	Lrc_i	A $1/a_i + 1/b_i + 1/c_i + 1/d_i$	$w_i = 1/A$	$w_i \times Lrc_i$	$w_i \times Lrc_i^2$	n_i	$\frac{a_i d_i}{n_i}$	$\frac{b_i c_i}{n_i}$
0,84	-0,0757	0,3971	2,5183	-0,191*	0,0144	49	2,33	2,76
2,25	0,3522	0,3203	3,1221	1,0996	0,3873	59	4,58	2,03
			5,6404	0,9090	0,4017	108	6,91	4,79

* = -0,1906

(1.47)

El estimador de la razón de contingencia conjunta, según el método de Mantel Haenzel es:

$$\overline{rcc} = 6,91 / 4,79 = 1,4426 = 1,44 \qquad (1.48)$$

Una confirmación de estos cálculos se logra obteniendo el

antilog de $\bar{L}\ rc_i$ = anlilog 0,1612 = 1,45. Salvo redondeos este valor es similar al de la estimación por el método anterior de Mantel Haenzel.

El intervalo de variación esperado para el mismo, con un 95% de confianza, debe ser calculado estimando la media ponderada de los log rc_i y su variancia.

$$\bar{L}\ rc_i = 0.9090 \ / \ 5.6404 = 0,1612$$

$$\text{Var } \bar{L}\ rc_i = 1 \ / \ 5,3020 \times 5,6404 = 0,0334$$

En consecuencia, los límites de confianza están dados por:

$$0,1612 \pm (1,96) \times \sqrt{0,0334} = 0,1612 \pm 0,3584 =$$
$$-0,1972 \ \text{y} \ 0,5196 = 0,6350 \ \text{y} \ 3,3083 \approx$$

$$0,63 < \overline{rcc} < 3,31 \qquad (1.49)$$

$$\text{El } X^2_{total} = (2,3026)^2 \times 0,4017 = 2.1298 \approx 2,13 \qquad (1.50)$$

Como el valor del $X^2_{crit.(gl=1,\alpha=0.05)}$ = 3,84, se desprende que no hay una diferencia significativa entre las probabilidades de enfermarse o no estando expuestos al riesgo en cuestión, tomando el conjunto de los dos estratos. Esto coincide con la lógica de las observaciones.

El cálculo del $X^2_{homog.}$ es el siguiente:

$$X^2_{homog.} = (2.3026)^2 \times [\ 0,4017 - (0,9090)^2 / \ 5,6404] =$$
$$= 5.3020 \times 0,1465 = 1,3531 \approx 1.35 \qquad (1.51)$$

Como el valor del $X^2_{crit.(gl=1,\alpha=0,05}$ = 3,84 se desprende que no hay heterogeneidad entre las tablas de de 2x2 de este estrato. (Esto coincide con la simple observación de los valores de las a, b, c, y d).

Finalmente el cálculo del $X^2_{asoc.}$ está dado, como se dijo en (1.45), por el cociente entre la media ponderada del log rc_i y su error estándar:

$$\bar{L} = [(0,9090 \ / \ 5,6404)] = 0,1612$$

$$ee\bar{L} = 1 \ / \ \sqrt{5,6404} = 0,4211 \quad , \quad \text{de donde:}$$

$$X^2_{asoc.} = [5,3020 \times (0,1612 \ / \ 0,4211)^2 = 0,7770 \approx 0,78 \quad (1.52)$$

La probabilidad observada para el $X^2_{asoc.}$ estaría entre:

0,50 ... 0,30 (Confirmando: **2,13** = **1,35** + **0,78**.)

El $X^2_{crit.(gl=1,\alpha=0,05)}$ = 3,84, consecuentemente se afirma que no hay asociación entre las posibilidades de enfermarse y la exposición a los riesgos tomando el promedio de las tablas presentadas en este estamento.

Sintetizando estas observaciones se puede decir que el análisis de estos ejemplos conduce a las siguientes conclusiones:

Conjunto de Tablas:		Significancia estadistica:
$\overline{rcc}$	= 3,38	Alta asociación
X^2_{total}	= 26.44	**(Asociación altamente significat.)
$X^2_{homog.}$	= 9,49	* (Hay heterogeneidad en las asoc.)
$X^2_{asoc.}$	= 16,95	**(Asociación altamente significat.)

(1.53)

Tablas 1er Nivel:		Significancia Estadística:
$\overline{rcc}$	= 7,71	Muy alta asociación
X^2_{total}	= 24,32	**(Asociación altamente significat.)
$X^2_{homog.}$	= 0,19	NS(No hay heterogeneid.en las asoc.)
X^2_{asoc}	= 24,13	**(Asociación altamente significat.)

(1.54)

Tablas 2do. Nivel		Significancia estadística
$\overline{rcc}$	= 1,44	Muy baja asociación
X^2_{total}	= 2,13	NS(No hay asociación)
$X^2_{homog.}$	= 1.35	NS(No hay heterogeneid.en las asoc.)
$X^2_{asoc.}$	= 0,78	NS(No hay asociación)

(1.55)

Para interpretar estos resultados conviene recordar los significados de algunos conceptos. 1) Si el X^2 total es grande con una diferencia significativa con relación a su X^2 crítico, se puede concluir que hay asociación entre los factores de riesgo y la aparición de la condición que se está valorando, pudiendo esto suceder en todos o en algunos de los grupos estudiados. 2) El X^2 de homogeneidad indica la existencia de asociaciones diferentes dentro del conjunto de los grupos estudiados. 3) Por último, un X^2 de asociación significante (por encima de su valor crítico) indica que existe una asociación promedio considerable entre los grupos. Cuanto mayor es la asociación se puede inferir que existen grupos con alta relación entre los factores de riesgo y la presencia de los problemas que se están estudiando. Como el Ji Cuadrado de homogeneidad dio alto, es dable esperar que los grupos sean heterogéneos.

La tabla general habla de una razón de contingencia evidente (rcc = 3,38) y de un X^2 total altamente significativo con relación al conjunto de los estratos que se presentan (cada uno de ellos está representado por una tabla de 2x2). Esto dice que hay asociación en todos o por lo menos en algunos de esos grupos.

El X^2 de homogeneidad significante implica que hay grupos, dentro del conjunto analizado, que poseen niveles diferentes de asociación. El X^2 de asociación dice que ellas pueden ser muy evidentes en aquellos lugares en que ella pueda ser detectada.

Corresponde entonces hacer una partición de las tablas. Se identifican dos grupos diferentes. En el primer conjunto se detecta una fuerte razón de contingencia (odds ratio) entre los factores de riesgo y la presencia de los eventos estudiados (rcc

= 7,71); con un X^2 total muy significativo (24,32) ; un X^2 de homogeneidad sin diferencias evidenciables con sus valores críticos (0,19); y un X^2 de asociación (24,13) altamente discrepante.

En el segundo subgrupo la razón de contingencia (odds ratio) es casi igual a la unidad, o sea que las posibilidades de contraer una condición es independiente de la presencia del factor de riesgo en estudio. La $P(B/A)$ es apenas mayor que la $P(\bar{B}/A)$, por lo tanto se dice que los riesgos son semejantes tanto entre los expuestos como en los no expuestos a la condición en estudio.

El X^2 total es pequeño no indicando una asociación entre esos grupos estudiados. Además, con un X^2 de homogeneidad por debajo de su nivel crítico, no hay evidencias de que los grupos sean heterogéneos. Finalmente no hay en ellos asociación entre el factor de riesgo y la aparición del fenómeno en estudio.

Sumarizando los resultados del ejemplo, cualquiera sean las condiciones cuyo efecto se desea conocer (producción o no de una situación en cada uno de los grupos) ella existe con marcada evidencia entre los primeros dos subgrupos y no la hay en los segundos.

Otra forma de comentar esos resultados es la siguiente [14/]: "En el análisis de los datos en que se combinan los 4 estratos originales, se encontró una asociación global bien marcada entre el factor de riesgo y la presencia de la condición investigada.

Esa asociación "cruda" se parte en dos componentes: a) uno que mide la homogeneidad de los estratos que intervienen en el estudio; y b) otro que valora la asociación promedio de los cuatro estratos.

Los dos componentes mostraron una significación estadística suficiente como para considerar que hay una marcada asociación promedio y que existe un considerable grado de heterogeneidad entre esos estratos.

Sobre la base de la observación de los resultados de los cuatro estratos , se unieron los que mostraban una "odds ratio" o razón de contingencia alta, por una parte, y los que tenían valores bajos por otra (suponiendo que pertenecen a estratos diferentes). Esta es una decisión del que realiza la investigación y que con ella pretende distinguir aquéllos que tienen asociaciones evidentes, de los estratos que no la demuestran. A posteriori él interpretará las diferencias que encuentre.

En el estrato uno, donde la medida de asociación global dio significativa, sus componentes demostraron que ese valor dependía fundamentalmente del grado de asociación de los dos estratos (cuya semejanza derivó de la separación arbitraria de los grupos

---

[14/] Marchevsky, N.: Comunicación personal; Buenos Aires, 1992.

realizada por el propio investigador); y se evidenció, además, por su baja homogeneidad (Ji cuadrado de homogeneidad con bajo valor).

En el segundo subgrupo, en el cual la asociación global no es significativa, esos dos componentes tampoco son significativos, indicando que las asociaciones son de bajo nivel y que ambos sitios (de ese subgrupo) tienen baja asociación.

Como se dijo anteriormente esos resultados no son inesperados ya que los grupos se armaron a partir de la observación de sus resultados".

En estos estudios hay que tener en cuenta a los "distractores" que tienen una importancia muy evidente en las investigaciones epidemiológicas, pero que son mucho mas limitados en las investigaciones de los sitios centinela que se usan para dar apoyo a prioridades, planes o políticas y a sus estrategias.

7. El Caso General de Análisis de Asociaciones

El manejo de varias tablas de 2x2, en busca del conocimiento de las eventuales asociaciones entre los factores de exposición y la adquisición de ciertos caracteres (enfermedad, abandono escolar, desocupación, etc.), requiere ciertas deducciones de los valores que deben hallarse.

Usando la notación de Fleiss en la obra antes citada, se llama m_i a cualquier valor que mida alguna forma de asociación. Por ejemplo la diferencia entre dos proporciones ($p_1 - p_2$), el logaritmo de la razón de contingencia ($\log_e rc$), etc. Cualquiera sea m_i se llama w_i al peso que se asigna a cada valor de m_i.
Se define a $w_i = 1 / [ee(m_i)]^2$, (1.56)
que no es más que la recíproca del cuadrado del error estandar de m_i. Asumiendo que m_i es cero, esto es, que no hay asociación bajo la hipótesis nula, en el iésimo grupo:

$$X_i = m_i / ee(m_i) = m_i \times 1 / ee(m_i) = m_i \times \sqrt{w_i} \quad (1.57)$$

ya que: $w_i = 1 / ee\ m_i$

$$X^2 = (X)^2 = (\Sigma w_i\ m_i)^2 / (\sqrt{w_i})^2 = \Sigma w_i\ m_i \times \Sigma w_i\ m_i / \Sigma w_i$$

Multiplicando el numerador por $\Sigma w_i / \Sigma w_i$:

$$X^2 = \frac{(\Sigma w_i\ m_i) \quad (\Sigma w_i\ m_i) \times \Sigma w_i}{\Sigma w_i \qquad \Sigma w_i} = \bar{m}^2\ \Sigma w_i$$

ya que $(\Sigma w_i\ w_i\ ) / \Sigma \bar{w}_i) = \bar{m}$

Que no es más que la fórmula de cálculo del X^2_{total}.

Ya se dijo que un X^2_{total} significativo sólo indica que hay asociación en todos o, por lo menos, en algunos de los grupos. Hay que partirlo entonces en sus dos componentes: el $X^2_{asociación}$ y el $X^2_{homogeneidad}$; recordando lo que se dijo anteriormente, que el X^2 de homogeneidad habla del grado de igualdad o equivalencia entre las g medidas de asociación; mientras que el X^2 de asociación detecta el grado promedio de esa asociación. Esto es, si es alta o baja, o inexistente.

8. Procedimiento Abreviado de Cálculo.

Usando los logaritmos naturales se pueden calcular los valores del X^2 total, de asociación y de homogeneidad en una forma más simplificada que la general presentada en los puntos (1.33 a 1.39). Marchevsky, en la comunicación citada, propone:

#	Talla Baja — Estrato — Bajo	Talla Baja — Estrato — Alto	Talla Alta — Estrato — Bajo	Talla Alta — Estrato — Alto			
Orden	(a_i)	(b_i)	(c_i)	(d_i)	$(a_i d_i)$	$(b_i c_i)$	rc_i
1	15	6	8	20	300	48	6,2500
2	24	10	8	30	720	80	9,0000
3	6	15	9	19	114	135	0,8444
4	10	12	10	27	270	120	2,2500
	55	43	35	96	1404	383	

(1.58)

(m_i) $L.\ rc_i$	$(m_i)^2$ $L.\ (rc_i)^2$	(A) $1/ai + 1/b_i + 1/c_i + 1/d_i$	$w_i = 1/A$	$w_i m_i$ $w_i L.\ rc_i$	$w_i (m_i)^2$ $w_i L.\ rc_i^2$	$a_i d_i / n_i$	$b_i c_i / n_i$
1,8326	3,3584	0,4084	2,4486	4,4873	8,2234	6,12	0,98
2,1972	4,8277	0,3000	3,3333	7,3239	16,0922	10,00	1,11
-0,1744	0,0304	0,3971	2,5183	-0,4392	0,0766	2,33	2,76
0,8109	0,6576	0,3203	3,1221	2,5317	2,0531	4,58	2,03
			11,4223	13,9037	26,4453	23,03	6,88

(1.59)

$$\text{Log. } \overline{rcc} = \left(\sum_{i=1}^{g} w_i\, m_i\right) / \sum_{i=1}^{g} w_i = 13,9037 / 11,4223 = 1,2172 \qquad (1.60)$$

$$\text{Var Log. } \overline{rcc} = 1 / 11,4223 = 0,0875$$

$$\text{ee Log. } (\overline{rcc}) = \sqrt{0,0875} = 0,2958 \qquad (1.61)$$

La razón de contingencia conjunta ($\overline{rcc}$ u odds ratio) es igual al antilog. de 1,2172 = 3,35 . (ver 1.33)

El cálculo de su intervalo de confianza para un 95% es:

$$\overline{rcc} \pm 1.96\ (ee_{\overline{rcc}}) = 1,2172 \pm 1,96 \times 0,2958 =$$
$$= 1,2172 \pm 0,5800 = 0,6369 \text{ y } 1,7969$$

Sacando el antilogaritmo natural (base e):

$$1,89 < \overline{rcc} < 6,03$$

El $X^2_{total} = \sum_{i=1}^{g} w_i \, rcc_i^2 = 26,4453 \approx 26,45$ (Ver 1.53)

El $X^2_{homog.} = \sum_{i=1}^{g} w_i \, rcc_i^2 - \left(\sum_{i=1}^{g} w_i \, rcc_i\right)^2 / \sum_{i=1}^{g} w_i =$
$= 26,4453 - 13,9037 = 16,9242 \approx 9,52$ (Ver 1.53)

El $x^2_{asoc.} = (\overline{rcc} / ee_{\overline{rcc}})^2 = (1,2172 / 0,2958)^2 =$
$= 16,2397 \approx 16,24$ (Ver 1.53)

Como se puede observar los resultados son equivalentes usando este método abreviado en que se trabaja directamente con los $\log_e$.

8.1. Computaciones para el Primer Estrato

$L_e \, rc_i$	(A) $1/a_i + 1/b_i + 1/c_i + 1/d_i$	$w_i = 1/A$	$w_i \, rc_i$	$w_i \; rc_i^2$	$a_i \, d_i / n_i$	$b_i \, c_i / n_i$
1,8326	0,4084	2,4486	4,4873	8,2234	6,12	0,98
2,1972	0,3000	3,3333	7,3239	16,0922	10,00	1,11
		5,7819	11,8112	24,3156	16,12	2,09

$\overline{L_e \, rc_i} = 11,8112 / 5,7819 = 2,0428$

Antilog$_e$ $rc_i = 7,7119 \approx 7,71 = \overline{rcc}$ (Ver 1.54)

Var $\overline{L} \, rc_i = 1 / 5,7818 = 0,1739$

ee $\overline{L} \, rc_i = \sqrt{0,1739} = 0,4159$

Intervalo de confianza para Log$_e$ $\overline{rcc}$:

Log$_e$ $\overline{rcc} \pm 1,96$ ee Log$_e$ $\overline{rcc}$

$2,0428 \pm 0,8152 = 1,2276$ y $2,8580$

Sacando el antilog$_e$, nos da el intervalo de confianza

para $\overline{rcc}$: $3,41 < \overline{rcc} < 17,42$ (Ver 1.42)

$X^2_{total} = 24,3156 \approx 24,32$ (Ver 1.54)

$X^2_{homog.} = 24,3156 - (11,8112)^2 / 5,7819 =$
$= 24,3156 - 24,1278 = 0,1878 \approx 0,19$ (Ver 1.54)

$X^2_{asoc.} = (2{,}0428 \;/\; 0{,}4159)^2 = 24{,}1254 \approx$ **24,13** (Ver 1.54)

8.2. Computaciones para el Segundo Estrato

$L_e\, rc_i$	$L_e\, rc_i^{\,2}$	(A) $1/a_i$ + $1/b_i$ + $1/c_i$ + $1/d_i$	$w_i = 1/A$	$w_i\, \bar{L}_e\, rc_i$	$w_i\, L_e\, rc_i^{\,2}$	$\frac{a_i\, d_i}{n_i}$	$\frac{b_i\, c_i}{n_i}$
-0,1744	0,0304	0,3971	2,5183	-0,4392	0,0766	2,33	2,76
0,8109	0,6576	0,3203	3,1221	2,5317	2,0531	4,58	2,03
			5,6404	2,0925	2,1297	6,91	4,79

$\overline{L_e rc_i} = 2{,}0925 \;/\; 5{,}6404 = 0{,}3710$

$\text{Antilog}_e rc_i = 1{,}4492 \approx$ **1,45** (Ver 1.55)

$\text{Var Log}_e rc_i = 1 \;/\; 5{,}6404 = 0{,}1773$

$\text{ee. Log}_e rc_i = \sqrt{0{,}1773} = 0{,}4211$

El intervalo de confianza para $\text{Log}_e rc_i$ es:

$\overline{L_e rc_i} \pm 1{,}96\ (\text{ee. Log}_e rc_i)$

$0{,}3707 \pm 1{,}96 \times 0{,}4211 = -0{,}4547$ y $1{,}1961$

Obteniendo los antilogaritmos 0,6346 y 3,3072 de donde:

$0{,}63 < \overline{rcc} < 3{,}31$

$X^2_{total} = 2{,}1297 \approx$ **2,13** (Ver 1.55)

$X^2_{homog.} = 2{,}1297 - (2{,}0925)^2 \;/\; 5{,}6404 =$
$= 2{,}1297 - 0{,}7763 = 1{,}3534 \approx$ **1,35** (Ver 1.55)

$X^2_{asoc.} = (0{,}3710 \;/\; 0{,}4211)^2 = 0{,}7762 \approx$ **0,78** (Ver 1.55)