Problemas de la Práctica IX. Enunciados y soluciones.

Juan de Dios Luna del Castillo

UNIVERSIDAD
DE GRANADA

PROBLEMAS – PRÁCTICA IX. TEST CHI-CUADRADO Y TABLAS 2×2

IX. 1 Ante la sospecha de que el hábito de fumar en una embarazada pueda influir en el peso de su hijo al nacer, se tomaron dos muestras de madres fumadoras y no fumadoras clasificándose a sus hijos según su peso en tres categorías:

	Peso del Hijo		
	Menor del Percentil 10	Entre los percentiles10 y 90	Mayor del Percentil 90
Sí Fumadoras	117	529	19
No Fumadoras	124	1157	117

¿Qué conclusiones se pueden obtener de esta experiencia?

Solución con calculadora y tablas:

(Esta solución está para ejemplificar. Se recomienda al alumno que mire y estudie la solución con R)

Test de homogeneidad de dos muestras cualitativas (tabla 2×3) (**R.8.1.a**):

$H_0: p_{<p_{10}|F} = p_{<p_{10}|\bar{F}}$; $p_{p_{10}-p_{90}|F} = p_{p_{10}-p_{90}|\bar{F}}$; $p_{>p_{90}|F} = p_{>p_{90}|\bar{F}}$ ≡ Fumar NO influye en el Peso.

$H_1:$ Alguna de las igualdades no es cierta ≡ Fumar SÍ influye en el Peso.

O_{ij} / E_{ij}	$< p_{10}$	Entre p_{10} y p_{90}	$> p_{90}$	Totales
Sí Fumadoras	117 77,69	529 543,47	19 43,84	665 = fijo
No Fumadoras	124 163,31	1157 1.142,52	117 92,16	1.398 = fijo
Totales	241	1.686	136	2.063

$$E_{11} = \frac{F_1}{T}C_1 = \frac{665}{2.063}241 = 77,69;\ E_{12} = \frac{F_1}{T}C_2 = \frac{665}{2.063}1.686 = 543,47;\ E_{13} = \frac{665}{2.063}136 = 43,84$$

$E_{21} = C_1 - E_{11} = 241 - 77,69 = 163,31$; etc.

• Se verifican las <u>condiciones de validez</u>, pues todas las E_{ij} son > 5.

• $\chi^2_{exp} = \left(\frac{117^2}{77,69} + \frac{529^2}{543,47} + \ ..... \ + \frac{117^2}{92,16}\right) - 2.063 = 50,69$ *vs.* χ^2_α con (2–1)×(3–1) = 2 *gl* de la Tabla 9 ⇒ $P < 0,001$ ⇒ Altamente significativo ⇒ H_1: fumar influye en el peso.

• Calculando los % por filas (117×100/665 = 17,59 etc.):

%	$< p_{10}$	p_{10} y p_{90}	$> p_{90}$	Total
SÍ Fuma	17,59	79,55	2,86	100,0
NO Fuma	8,87	82,76	8,37	100,0

La causa de la significación es que nacen más hijos con peso bajo (y menos con peso alto) en las madres fumadoras, siendo prácticamente iguales las proporciones de niños con peso intermedio.

Solución con R:

Test de homogeneidad de dos muestras cualitativas (tabla 2×3) :

$H_0 : p_{<p_{10}|F} = p_{<p_{10}|\bar{F}}$; $p_{p_{10}-p_{90}|F} = p_{p_{10}-p_{90}|\bar{F}}$; $p_{>p_{90}|F} = p_{>p_{90}|\bar{F}}$ ≡ Fumar NO influye en el Peso.

H_1 **:** Alguna de las igualdades no es cierta ≡ Fumar SÍ influye en el Peso.

Los resultados aparecen en el siguiente recuadro.

```
> library(BioestadisticaR2)
>
> # Problema IX.1
> # Definir las columnas Fumadora_Si y Fumadora_NO
> Fumadora_SI=c(117,529,19)
> Fumadora_NO=c(24,1157,117)
> #Definir la tabla agregando Fumadora_SI y Fumadora_NO
> tabla=data.frame(Fumadora_SI, Fumadora_NO)
> # Calcular el estadístico chi-cuadrado
> tablarxc(frec=tabla, tablas="C")

Test Chi-cuadrado para tablas RxC
----------------------------------

# Frecuencias observadas
          Fumadora_SI   Fumadora_NO Total
  1               117            24   141
  2               529          1157  1686
  3                19           117   136
  Total           665          1298  1963


# Validez del test chi-cuadrado
  Frecuencia minima esperada = 46.07
  0 frecuencias esperadas son menores a 1 y 0 son menores a 5 (el 0% de la tabla)

# Test chi-cuadrado
  X2exp(2 gl) = 180.5261, p < 0.001

# Porcentajes por columnas
          Fumadora_SI   Fumadora_NO  Total
  1            0.1759        0.0185 0.0718
  2            0.7955        0.8914 0.8589
  3            0.0286        0.0901 0.0693
  Total        1.0000        1.0000 1.0000
```

Obsérvese que la tabla creada lo ha sido poniendo las muestras en las dos variables Fumadora_SI y Fumadora_NO y en cada una de ellas se han puesto las frecuencias correspondientes a cada una de las modalidades del peso del recién nacido. La función *tablarxc* saca en primer lugar la tabla de frecuencias observadas tal como se le ha introducido. Inmediatamente se nos proporciona la frecuencia esperada mínima que vale 46.07 y además la función nos da el número de frecuencias observadas por debajo de 1 y el número de frecuencias observadas por debajo de 5. En ambos casos no hay ninguna, lo que nos asegura que el test chi-cuadrado para esta tabla 2x3 es aplicable.

El valor del test resulta ser $\chi^2_{exp} = 180.53, 2gl, P < 0.001$. Siguiendo la Regla automática de como P<0.05(ya que es P<0.001), podemos rechazar la hipótesis nula pudiendo decir entonces que las probabilidades de las diferentes categorías del peso del recién nacido no son todas iguales.

Observando la tabla de porcentajes por columnas que hemos calculado podemos tener una idea de qué categorías proviene la significación: Obsérvese que en la categoría 1 del peso, el bajo peso, en las fumadoras hay un 17.59% de niños mientras que en las no fumadoras hay un 1.85%, por tanto en la modalidad del bajo peso hay una muy fuerte diferencia entre fumadoras y no fumadoras; en la modalidad del peso normal también existe una diferencia que es de alrededor del 10% pero en proporciones del 80% con lo cual vamos a considerar que esa diferencia no es muy grande; por último en la modalidad de peso alto, por encima del percentil 90, hay una discrepancia importante entre las fumadoras y las no fumadoras si bien es en el sentido contrario a la que había en el bajo peso. Aunque todo esto necesitaría de una prueba podemos decir que lo que indican los resultados es que en las madres fumadoras tienden a dar niños de bajo con más probabilidad que las no fumadoras y tienden a dar menos niños de peso elevado que las no fumadoras. Con respecto a los niños de peso normal no parece haber diferencias fuertes entre fumadoras y no fumadoras.

IX. 2 Se observó en 665 niños la huella plantar y la fórmula digital, a fin de estudiar la posible asociación entre ambas características del pie. Por su huella plantar se les clasificó en normales, cavos y planos. Respecto a la fórmula digital, en pie cuadrado, egipcio y griego. Los resultados fueron:

	Cuadrado	Egipcio	Griego
Normales	85	212	67
Cavos	37	93	19
Planos	32	69	51

¿Qué se puede afirmar acerca de la asociación de ambas características?

Solución con calculadora y tablas:

(Esta solución está para ejemplificar. Se recomienda al alumno que mire y estudie la solución con R,. además de leer ésta)

$$\begin{cases} H_0\text{: La Huella Plantar y la Fórmula Digital son independientes} \equiv \text{ NO están asociadas.} \\ \text{H}_1\text{: La Huella Plantar y la Fórmula Digital son dependientes} \equiv \text{ SÍ están asociadas.} \end{cases}$$

(1) Test de independencia para variables cualitativas (Tabla 3×3) (**R.8.1.b**). Esperadas:

- La fracción (364/665) se multiplica por 154, 374 o 137 para obtener la 1ª fila de E_{ij}.
- La fracción (149/665) se multiplica por 154, 374 o 137 para obtener la 2ª fila de E_{ij}.
- Para la 3ª fila: E_{31} = 154–84,29–34,51 = 35,20 y similarmente con el resto.

Fórmula Digital

	O_{ij} (E_{ij})	Cuadrado	Egipcio	Griego	Total
Huella Plantar	Normales	85 (84,29)	212 (204,72)	67 (74,99)	364
	Cavos	37 (34,51)	93 (83,80)	19 (30,70)	149
	Planos	32 (35,20)	69 (85,49)	51 (31,31)	152
	Total	154	374	137	665 = fijo

Todas las frecuencias esperadas son mayores que 5: se verifican las condiciones de validez.

(2) $\chi^2_{exp} = \left(\frac{85^2}{84,29} + \frac{212^2}{204,72} + \frac{67^2}{74,99} + \ldots + \frac{51^2}{31,31} \right) - 665 = 22,61$ *vs.* χ^2_α con (3–1)×(3–1) = 4 *gl* de la Tabla 9 ⇒ Altamente significativo ($P < 0,001$) ⇒ H_1: las dos características están asociadas.

(3) Tabla de % por filas para buscar las causas de la significación:

%	Cuadrado	Egipcio	Griego	Total
Normales	23,4	58,2	18,4	100%
Cavos	24,8	62,4	12,8	100%
Planos	21	45,4	33,6	100%

La significación se debe a que los pies griegos son más frecuentes en los individuos con los pies planos que en los normales o cavos.

Solución con R:

$$\begin{cases} H_0\text{: La Huella Plantar y la Fórmula Digital son independientes} \equiv \text{NO están asociadas.} \\ H_1\text{: La Huella Plantar y la Fórmula Digital son dependientes} \equiv \text{SÍ están asociadas.} \end{cases}$$

```
> library(BioestadisticaR2)
> # Problema IX.2
> # Definir las columnas fd_cuadrado, fd_egipcio y fd_griego
> fd_cuadrado=c(85,37,32)
> fd_egipcio=c(212,63,99)
> fd_griego=c(67,19,51)
>
> #Definir la tabla agregando Fumadora_SI y Fumadora_NO
> tabla_pies=data.frame(fd_cuadrado, fd_egipcio, fd_griego)
> # Calcular el estadístico chi-cuadrado
> tablarxc(frec=tabla_pies, tablas="F")

Test Chi-cuadrado para tablas RxC
----------------------------------

# Frecuencias observadas
          fd_cuadrado  fd_egipcio   fd_griego Total
  1                85         212          67   364
  2                37          63          19   119
  3                32          99          51   182
  Total           154         374         137   665


# Validez del test chi-cuadrado
  Frecuencia minima esperada = 24.52
  0 frecuencias esperadas son menores a 1 y 0 son menores a 5 (el 0% de la tabla)

# Test chi-cuadrado
  X2exp(4 gl) = 13.2405, p = 0.0102

# Porcentajes por filas
          fd_cuadrado  fd_egipcio   fd_griego  Total
  1            0.2335      0.5824      0.1841 1.0000
  2            0.3109      0.5294      0.1597 1.0000
  3            0.1758      0.5440      0.2802 1.0000
  Total        0.2316      0.5624      0.2060 1.0000
```

Obsérvese que la tabla creada generando tres variables (de 3 frecuencias en cada una) con la fórmula digital correspondiente: cuadrado, egipcio y griego. La tabla ha sido grabada como aparece en el enunciado del problema.

Inmediatamente se nos proporciona la frecuencia esperada mínima que vale 24.52y además la función nos da el número de frecuencias observadas por debajo de 1 y el número de frecuencias observadas por debajo de 5. En ambos casos no hay ninguna, lo que nos asegura que el test chi-cuadrado para esta tabla 3x3 es aplicable.

El valor del test resulta ser $\chi^2_{exp}=13,24,4gl, P=0.0102$. Siguiendo la Regla automática de decisión como P<0.05(ya que es P=0.0102), podemos rechazar la hipótesis nula y podemos concluir que la huella plantar y la fórmula digital no son independientes, es decir, están relacionados, o lo que es lo mismo, las personas con, al menos, una determinada huella plantar tienden a tener con más o con menos probabilidad, al menos, una determinada fórmula digital.

La observación de los porcentajes por filas nos indica por dónde iría esa significación: en efecto para el caso de la huella plantar 3, la de los pies Planos tendríamos un porcentaje de pies cuadrados inferior al de los pies Normales o Cavos, en las tres el porcentaje de pies Egipcios es muy parecido y los pies griegos se presentarían en mayor porcentaje que en los pies Planos que en los Normales o Cavos. Esto se debería verificar mediante los tersts correspondientes, pero siendo un procedimiento largo, lo dejaremos aquí.

IX. 3 Se desea estudiar la posible relación entre el hábito de fumar y la aparición de una cardiopatía coronaria. Se observaron 80 fumadores, de los que 35 sufrieron tal cardiopatía al cabo de un cierto período de tiempo; de los 70 no fumadores observados, aparecieron 16 cardiopatías. **a)** ¿Existe esa posible relación? **b)** Calcular las medidas de asociación que convienen al caso mediante estimación puntual y por intervalo.

Solución con calculadora y tablas:

(Esta solución está para ejemplificar. Se recomienda al alumno que mire y estudie la solución con R, además de leer ésta)

<u>Problema VI.3</u>

R.8.4.a): Estudio <u>prospectivo</u> (2 muestras: una de fumadores y otra de no fumadores).

(a) <u>Test previo a realizar</u> (**R.8.4.b.*ii***):

$$\left.\begin{array}{l} H_0: p_{Card|F} = p_{Card|\overline{F}} \\ H_1: p_{Card|F} \neq p_{Card|\overline{F}} \end{array}\right\} \begin{array}{l} \equiv \text{Fumar y padecer Cardiopatía son independientes.} \\ \equiv \text{Fumar y padecer Cardiopatía están relacionados.} \end{array}$$

La condición de validez se verifica pues: $(51\times 70)/150=23,8>7,7 \Rightarrow$

		Fumar		
		SÍ	NO	Total
Cardiopatía	SÍ	35	16	51
	NO	45	54	99
	Total	80 = fijo	70 = fijo	150

$\chi^2_{\exp} = \dfrac{\left(\left|35\times 54 - 16\times 45\right| - 1\right)^2}{51\times 99\times 80\times 70}\times 150 = 7,250$ *vs.* χ^2_α (1 *gl*) de Tabla 9 ⇒ $P < 0,01$ (muy significativo) ⇒ hay evidencias de que el hecho de fumar está asociado a la presencia de cardiopatía. Además, como $\hat{p}_{Car|F} = \dfrac{35}{80} = 0,4375 > \hat{p}_{Car|\overline{F}} = \dfrac{16}{70} = 0,2286$ ⇒ padecen más cardiopatías los fumadores (asociación positiva) ⇒ ser fumador es un factor de riesgo.

(b) Medidas de asociación lícitas (**Cuadro R.8.1**):

$\hat{R} = \dfrac{35\times 70}{16\times 80} = 1,91$ (Riesgo relativo) → Es casi el doble de probable padecer una cardiopatía cuando se es fumador que cuando no se es fumador.

$\hat{O} = \dfrac{35\times 54}{16\times 45} = 2,625$ (Razón del producto cruzado) → La fracción de personas con cardiopatía frente a los que no la tienen es 2,625 veces mayor para las personas que fuman que en las que no fuman.

Los intervalos de confianza para estas medidas de asociación se ven en el apartado de la solución con R.

Solución con R:

Estudio prospectivo, el contraste de hipótesis a realizar es:

$$\left.\begin{array}{l} H_0: p_{Card|F} = p_{Card|\overline{F}} \\ H_1: p_{Card|F} \neq p_{Card|\overline{F}} \end{array}\right\} \begin{array}{l} \equiv \text{Fumar y padecer Cardiopatía son independientes.} \\ \equiv \text{Fumar y padecer Cardiopatía están relacionados.} \end{array}$$

La llamada a la función tabla2x2 y los resultados que se obtienen figuran en el siguiente recuadro.

Obsérvese en `primer lugar que se ha dispuesto la tabla poniendo por filas la enfermedad, el problema, mientras que por columnas se ha puesto el factor de riesgo y que siempre la primera categoría ha sido o la presencia de la enfermedad o la presencia del factor de riesgo. Además, hemos añadido etiquetas a las categorías para que así se pueda leer mejor la tabla.

Para llevar a cabo el contraste de hipótesis se ha de verificar la condición de validez, E>7.7 para el estudio que tenemos. Efectivamente E=23.8>7.7, luego se puede emplear el estadístico chi-cuadrado. Si no se verificara la condición de validez se debería hacer el test empleando el test exacto de Fisher que aparece en los resultados del ejercicio. Pero esta situación no la veremos puesto que tal test no se ha explicado en las clases de teoría.

```
> tabla2x2(o=c(35,16,45,54), fcat=c("Cardiopatia SI","Cardiopatía NO"),
+          ccat=c("Fumador SI","Fumador NO"),estudio="P",
+          tablas=c("F,C"))

# Análisis de tablas 2x2
# --------------------------------

# Frecuencias observadas
                   Fumador SI  Fumador NO Total
  Cardiopatia SI          35          16    51
  Cardiopatía NO          45          54    99
  Total                   80          70   150
```

```
# Test Chi-cuadrado para un estudio prospectivo

  Frecuencia mínima esperada = 23.8  > 7.7

  X2 = 7.2498,   gl = 1,  p = 0.0071, (cpc = 1)
  Test exacto de Fisher: p = 0.0094

  --- Otros criterios X2:
  X2 = 7.2622,   gl = 1,  p = 0.0117, (sin cpc)
  X2 = 6.361,   gl = 1,  p = 0.0117, (cpc de Yates = 75)

# Medidas de asociación para un estudio prospectivo
  Las medidas de riesgo se calculan como riesgo de la categoría en
  la primera columna para la categoría en la primera fila

  Riesgo absoluto (diferencia de Berkson; método de Agresti-Caffo):
  d=0.2089; 95%-IC(d)=(0.0576, 0.3487)

  Riesgo relativo:
  Rr=1.9141; 95%-IC(Rr)=(1.1561, 3.0765)

  Razón del producto cruzado (odds ratio):
  OR=2.625; 95%-IC(OR)= (1.2743, 5.2116)
```

El resultado del test es: $\chi^2_{exp} = 13,24, 1gl, P = 0.0102$; con ese valor de P menor que el 5%, si se sigue la regla automática de decisión, podemos rechazar la hipótesis nula y por tanto podremos afirmar que el ser fumador (el hábito tabáquico) está asociado con el sufrimiento de una cardiopatía. Concretamente la estimación puntual de sufrir cardiopatía cuando se es fumador vale 35/80=0.4375 que es mayor que la estimación puntual de sufrir cardiopatías cuando no se fuma, 16/70=0.2286; con esos resultados no sólo podremos decir que las proporciones poblacionales son diferentes, sino que la probabilidad de sufrir cardiopatías cuando se fuma es superior a la probabilidad de sufrir una cardiopatía cuando no se fuma, o lo que es lo mismo, que el ser fumador incrementa el riesgo de sufrir cardiopatías.

Veremos a continuación diferentes medidas de ese incremento del riesgo que se pueden calcular en un estudio como el que nos ocupa.

En primer lugar en un estudio prospectivo se puede calcular el **Riesgo Absoluto(diferencia de Berkson)** que resulta:

```
  Riesgo absoluto (diferencia de Berkson; método de Agresti-Caffo):
  d=0.2089; 95%-IC(d)=(0.0576, 0.3487)
```

La estimación puntual del riesgo absoluto 0.2089 se interpreta como que el hábito tabáquico incrementa el riesgo de sufrir cardiopatía en 0.2089 o de una manera más coloquial, que el hábito tabáquico incrementa el riesgo de sufrir cardiopatías en un 20.89%.

El intervalo de confianza obtenido para la diferencia de Berkson se interpreta: el hábito tabáquico incrementa el riesgo de cardiopatía en una cantidad comprendida entre un 5.76% y un 34.87% y eso lo afirmo con una confianza del 95%. Obsérvese que el intervalo es coherente con el resultado del contraste de hipótesis ya que no contine al cero y el contraste de hipótesis dio significativo; esto no siempre será así porque el intervalo está derivado de manera diferente a como se ha hecho el contraste de hipótesis, de forma que cuando ocurra la discrepancia entre el intervalo y el test, el resultado de este debe prevalecer siendo el intervalo un añadido y teniendo realmente valor el extremo más alejado del cero del mismo.

Para el **Riesgo Relativo**, se obtienen los resultados:

```
Riesgo relativo:
  Rr=1.9141; 95%-IC(Rr)=(1.1561, 3.0765)
```

La interpretación de $\widehat{R} = 1.91$ es: la probabilidad de sufrir una cardiopatía cuando se fuma es 1.91 veces mayor que la probabilidad de sufrir una cardiopatía cuando no se fuma. Es una redundancia decir "…1.91 veces mayor…", puesto que, como 1.91 es mayor que 1, no puede ser otra cosa que mayor. No obstante se incurre con frecuencia en esa redundancia que por otro lado no es perniciosa.
Para el intervalo podemos decir que el riesgo de sufrir cardiopatía cuando se fuma, es entre 1.16 veces y 3.08 veces mayor que la probabilidad de sufrir cardiopatía cuando no se fuma, con una confianza del 95%.

Para la **Razón del Producto Cruzado**, se obtienen los resultados:

```
  Razón del producto cruzado (odds ratio):
 OR=2.625; 95%-IC(OR)= (1.2743, 5.2116)
```

La interpretación de $\widehat{O} = 2.63$ es: la fracción de personas que sufren cardiopatía frente a los que no sufren cardiopatía es 2.63 veces mayor entre los fumadores que entre los no fumadores.
Para el intervalo podemos decir que la fracción de personas que sufren cardiopatía frente a los que no la sufren es entre 1.27 y 5.21 veces superior entre los fumadores que entre los no fumadores, con una confianza del 95%.

Hemos de hacer una reflexión sobre el uso de la odds ratio en los estudios prospectivos. Si bien es lícito su calaulo en los estudios prospectivos no parece razonable usarla y mucho menos sin el riesgo relativo. La razón es que en los estudios prospectivos el parámetro epidemiológico de interés es el riesgo relativo o la diferencia de Berkson y la odds ratio en ese caso es una aproximación al riesgo relativo y esa aproximación no es buena sino que tiende a sobreestimar el riesgo relativo, como ocurre en nuestro caso que la $\widehat{O} = 2.63$ es mayor que el riego relativo $\widehat{R} = 1.91$. Por tanto aunque siempre se puede calcular la odds ratio no es aconsejable emplearla en el caso de estudios prospectivos si es que se puede estimar directamente el riesgo relativo.

IX. 4 Una edad elevada (mayor de 35 años) de la madre se considera un posible factor de riesgo para la aparición del síndrome de Down en los hijos. Para comprobarlo, se revisó una muestra aleatoria de 90 casos de tal síndrome y otra de 540 recién nacidos sanos. De entre los primeros, en 41 la edad de la madre resultó ser superior a 35 años, mientras que esto ocurría en 108 de los niños normales. **a)** ¿Se puede afirmar este factor de riesgo? **b)** Calcular las medidas de asociación que convienen al caso mediante estimación puntual, teniendo en cuenta que el síndrome es de escasa incidencia.

Solución con calculadora y tablas:

(Esta solución está para ejemplificar. Se recomienda al alumno que mire y estudie la solución con R,. además de leer ésta)

R.8.4.a): Estudio <u>retrospectivo</u> (2 muestras: una CON y otra SIN el síndrome de Down).

(a) <u>Test previo a realizar</u> (**R.8.4.b.*ii***):

$H_0 \equiv p_{>35}|_{DOWN} = p_{>35}|_{NODOWN}$ ≡ Síndrome de Down y Edad de la madre NO están asociados.

$H_1 \equiv p_{>35}|_{DOWN} \neq p_{>35}|_{NODOWN}$ ≡ Síndrome de Down y Edad de la madre SÍ están asociados.

Edad Madre (> 35 años)

Síndrome de Down		SÍ	NO	Total
	SÍ	41	49	90 = fijo
	NO	108	432	540 = fijo
	Total	149	481	630

La condición de validez se verifica pues: $(90 \times 149) / 630 = 21,28 > 14,9 \Rightarrow$

$\chi^2_{\exp} = \dfrac{\left(|41\times 432 - 49\times 108| - 1\right)^2}{90\times 540\times 149\times 481}\times 630 = 27,896$ *vs.* χ^2_{α} (1 *gl*) de Tabla 9 $\Rightarrow P < 0{,}001$ (altamente significativo) $\Rightarrow$ el síndrome está asociado a la edad.

(b) Medidas de asociación lícitas (**Cuadro R.8.1**):

- $\hat{O} = \dfrac{41\times 432}{49\times 108} = 3{,}35$ $\rightarrow$ La fracción de niños con Síndrome de Down frente a los que no lo tienen es 3,35 veces mayor cuando las madres los tuvieron con > 35 años que cuando los tuvieron con ≤ 35 años.

Para lo que sigue hay que tener en cuenta que el estudio es retrospectivo, pero de escasa incidencia (la prevalencia es baja).

- $\hat{R} \approx \hat{O} = 3{,}35$ $\rightarrow$ La probabilidad de nacer con Síndrome de Down es aproximadamente 3,35 veces mayor cuando las madres los tienen con > 35 años que cuando los tienen con ≤ 35 años.

Solución con R:

Estudio retrospectivo, dos muestras una con Síndrome de Down y otra sin Síndrome de Down.

Los resultados están en el siguiente recuadro:

```
> # Problema IX.4
> # Calcular la chi-cuadrado y las medidas de asociación.
> tabla2x2(o=c(41,49,108,432), fcat=c("Sindrome Down SI","Sindrome Down NO"),
+          ccat=c("Edad Madre >35 años SI","Edad Madre >35 años NO"),estudio="R"
,
+          tablas=c("F,C"))

# Análisis de tablas 2x2
# --------------------------------

# Frecuencias observadas
                     Edad Madre >35 años SI   Edad Madre >35 años NO Total
  Sindrome Down SI                       41                       49    90
  Sindrome Down NO                      108                      432   540
  Total                                 149                      481   630

# Test Chi-cuadrado para un estudio retrospectivo

  Frecuencia mínima esperada = 21.2857  > 14.9

  X2 = 27.8963,   gl = 1,  p < 0.001, (cpc = 1)
  Test exacto de Fisher: p < 0.001

  --- Otros criterios X2:
  X2 = 27.9008,   gl = 1,  p < 0.001, (sin cpc)
  X2 = 26.5035,   gl = 1,  p < 0.001, (cpc de Yates = 315)

# Medidas de asociación para un estudio retrospectivo

  Riesgo atribuible*:
  Ra=0.193; 95%-IC(Ra)= (0.105, 0.2722)
  * La estimación de Ra para estudios retrospectivos es una aproximación válida
si la prevalencia de la enfermedad es baja: P(E) < 10%

  Razón del producto cruzado (odds ratio):
  OR=3.3469; 95%-IC(OR)= (2.1032, 5.3103)
  * La estimación para OR sirve de aproximación al riesgo relativo siempre que l
a prevalencia de la enfermedad sea P(E) < 10%
```

El resultado del test es: $\chi^2_{exp} = 27.90, 1gl, P < 0.001$; con ese valor de P menor que el 5%, si se sigue la regla automática de decisión, podemos rechazar la hipótesis nula y por tanto podremos afirmar que la probabilidad de tener una madre mayor de 35 años es diferente en las personas con Síndrome de Down que en las personas sin Síndrome de Down. Concretamente la estimación puntual de la probabilidad de que la madre tenga más de 35 años si la persona tiene Síndrome de Down es 41/90=0.4456 y la probabilidad de que la madre tenga más de 35 años cuando la persona no tiene síndrome de Down es de 10/540=0.2000, por lo que mirando estas estimaciones puntuales podemos afirmar que entre las personas con Síndrome de Down es más frecuente que su madre renga más de 35 años que en las personas que no tienen Síndrome de Down. Por tanto el que la madre tenga más de 35 años al nacer el niño es un factor de riesgo para la presencia de síndrome de Down.

Veremos a continuación diferentes medidas de ese incremento del riesgo que se pueden calcular en un estudio como el que nos ocupa.

La única medida que podemos calcular, según lo explicado en teoría, es la razón del producto cruzado. El riego atribuible que proporciona el programa no se ha visto en teoría porque no ha sido explicado previamente.
La razón del producto cruzado es la única medida que se puede calcular en los estudios retrospectivos y con ello se estimará el riesgo relativo, pero en los estudios retrospectivos estimará de una manera sesgada al riesgo relativo que sólo será insesgada si la enfermedad es rara, tiene una incidencia menor del 10%, como es el caso del síndrome de Down según se nos indica en el ejercicio con la frase "…de escasa incidencia….".

Los resultados para la odds ratio son:
```
Razón del producto cruzado (odds ratio):
  OR=3.3469; 95%-IC(OR)= (2.1032, 5.3103)
  * La estimación para OR sirve de aproximación al riesgo relativo siempre que la prevale
ncia de la enfermedad sea P(E) < 10%
```

Siendo la enfermedad rara, incidencia inferior al 10% o prevalencia inferior al 10%, el valor de $\widehat{O}=3.35\cong\widehat{R}$ lo que nos permite afirmar que siendo la madre de una edad superior a 35 años, cuando tiene al niño, la probabilidad de que éste tenga síndrome de Down es 3.35 veces mayor que cuando la madre tiene 35 años o menos.

El intervalo se interpreta, como: teniendo la madre más de 35 años la probabilidad de que tenga un niño con Síndrome de Down es entre 2.10 veces superior y 5.31 veces superior a la probabilidad de que el niño tenga Síndrome de Down caundo la madre, en el momento del parto, no tiene más de 35 años, con una confianza del 95%.

IX. 5 Estudiando los factores de riesgo para el sobrepeso o la obesidad se tomó una muestra de 9206 niños de 6 años y se vio que 1288 eran obesos o tenían sobrepeso. Entre estos no habían recibido alimentación materna en su infancia 683. De entre los que no tenían sobrepeso u obesidad resulta que 3339 no habían tenido alimentación materna en la infancia. ¿Podemos decir que la alimentación materna está asociada a la obesidad o el sobrepeso en la edad infantil? De las medidas oportunas de la fuerza de asociación.

Solución con R:

Estamos en presencia de un estudio transversal porque se ha tomado una única muestra a cuyos individuos los hemos clasificado según que estén en sobrepeso u obesidad en su edad infantil y que tuvieran o no alimentación materna en su infancia. Es un transversal porque es una muestra que se clasifica según dos caracteres cualitativos.

H_0≡El sobrepeso-obesidad en la edad escolar es independiente de la alimentación materna.
H_1≡El sobrepeso-obesidad en la edad escolar no está relacionado con la alimentación materna.

De los datos que nos proporciona el enunciado hemos obtenido la siguiente tabla en la que por filas aparece el problema, la enfermedad, en nuestro caso sufrir sobrepeso u obesidad, y por columnas aparece el posible factor de riesgo la alimentación materna apareciendo como grupo de riesgo los que no hayan tenido en la infancia alimentación materna.

	Alimentación Materna NO	Alimentación Materna SI	Total
Sobrepeso u Obesidad	683	605	1288
Normopeso	3339	4579	7918
Total	4022	5184	**9206**

El único valor fijado de antemano, como en todos los estudios transversales es el número total de individuos de la muestra, 9206.

Veamos en el cuadro siguiente los resultados de la ejecución de la función tabla2x2 a este ejemplo.

```
> # Problema IX.5
> # Calcular la chi-cuadrado y las medidas de asociación.
> tabla2x2(o=c(683,605,3339,4579), fcat=c("Sobrepeso u obesidad SI","Sobrepeso u
obesidad NO"),
+          ccat=c("Lactancia Materna NO","Lactancia Materna SI"),estudio="T",
+          tablas=c("F,C"))

# Análisis de tablas 2x2
# --------------------------------

# Frecuencias observadas
                            Lactancia Materna NO   Lactancia Materna SI  Total
  Sobrepeso u obesidad SI                    683                    605   1288
  Sobrepeso u obesidad NO                   3339                   4579   7918
  Total                                     4022                   5184   9206

# Test Chi-cuadrado para un estudio transversal

  Frecuencia mínima esperada = 562.713  > 6.2

  X2 = 53.0898,    gl = 1,   p < 0.001, (cpc = 0.5)
  Test exacto de Fisher: p < 0.001

  --- Otros criterios X2:
  X2 = 53.0899,    gl = 1,   p < 0.001, (sin cpc)
  X2 = 52.6494,    gl = 1,   p < 0.001, (cpc de Yates = 4603)

# Estimación de la prevalencia en un estudio transversal
  Método de Wald ajustado:
  Prev=0.1401; 95%-IC(prev)=(0.133, 0.1472)

# Medidas de asociación para un estudio transversal
  Las medidas de riesgo se calculan como riesgo de la categoría en
  la primera columna para la categoría en la primera fila

  Riesgo absoluto (diferencia de Berkson; método de Agresti-Caffo):
  d=0.0531; 95%-IC(d)=(0.0386, 0.0677)

  Riesgo relativo:
  Rr=1.4551; 95%-IC(Rr)=(1.3147, 1.61)

  Riesgo atribuible:
  Ra=0.1658; 95%-IC(Ra)= (0.1197, 0.2096)

  Razón del producto cruzado (odds ratio):
  OR=1.5482; 95%-IC(OR)= (1.3755, 1.7421)
> 683/1208
[1] 0.1698
> 3339/7918
[1] 0.1167
```

Lo primero que hemos de hacer es verificar que la tabla que se ha introducido para su análisis se corresponde con la que nosotros tenemos; en este caso es así.
Verifiquemos que se puede aplicar el test chi-cuadrado para esta tabla según los resultados obtenidos:

Frecuencia mínima esperada = 562.713 > 6.2

Los resultados de la chi-cuadrado apropiada son: X2 = 53.0898, gl = 1, $p < 0.001$, (cpc = 0.5) lo que nos permite rechazar la hipótesis nula y decir que ambos caracteres están asociados. Calculando la estimación de la probabilidad de no habiendo tomado leche materna tener sobrepeso u obesidad, 683/4022=0.1698, y la estimación de la probabilidad de habiendo tomado leche materna sufrir sobrepeso u obesidad, 605/5184=0.1167, podemos afirmar que el no haber tomado leche materna aumenta el riesgo de haber sufrido sobrepeso u obesidad. Veamos ahora medidas de la fuerza de asociación entre ambos caracteres y sus interpretaciones.

Para un estudio transversal se puede estimar:

a) Riesgo absoluto (diferencia de Berkson; método de Agresti-Caffo):
d=0.0531; 95%-IC(d)=(0.0386, 0.0677)

Interpretación estimación puntual: El porcentaje de niños obesos o en sobrepeso que no recibieron lactancia materna es 5.31% superior al porcentaje de niños obesos o en sobrepeso que si recibieron lactancia materna.

Interpretación estimación por intervalo: El porcentaje de niños que no habiendo recibido lactancia materna son a la edad escolar obesos o tienen sobrepeso es entre un 3.86% y un 6.77% superior al porcentaje de niños que habiendo recibido lactancia materna están en obesidad o en sobre peso en edad escolar y esto lo afirmamos con una confianza del 95%.

b) Riesgo relativo:
Rr=1.4551; 95%-IC(Rr)=(1.3147, 1.61)

Interpretación estimación puntual: La probabilidad de estar en sobrepeso u obesidad en la edad escolar en niños que no recibieron lactancia materna es 1.46 veces mayor que la probabilidad de estar en sobrepeso u obesidad en la edad escolar cuando no se recibió lactancia materna.

Interpretación estimación por intervalo: El porcentaje de niños que no habiendo recibido lactancia materna son a la edad escolar obesos o tienen sobrepeso es entre un 1.31 y 1.61 veces superior al porcentaje de niños que habiendo recibido lactancia materna están en obesidad o en sobre peso en edad escolar y esto lo afirmamos con una confianza del 95%.

c) Razón del producto cruzado (odds ratio):
OR=1.5482; 95%-IC(OR)= (1.3755, 1.7421)

Interpretación estimación puntual: La fracción de niños en sobrepeso u obesidad frente a los que están en normopeso es 1.55 veces superior entre los que no recibieron la lactancia materna que entre los que la recibieron. En un estudio transversal la interpretación al revés también sería apropiada: La fracción de niños sin lactancia materna con respecto a los niños con lactancia materna es 1.55 veces superior en los que tienen sobrepeso u obesidad que entre los que no tienen normopeso.

Interpretación estimación por intervalo: La fracción de niños en sobrepeso u obesidad frente a los que están en normopeso es entre 1.38 y 1.74 veces superior entre los que no recibieron la lactancia materna que entre los que la recibieron con una confianza del 95%. Queda para el alumno la interpretación inversa.

En los estudios transversales la medida natural es la odds ratio y las otras dos medidas, aunque pueden calcularse no son las más apropiadas.