Cheat sheet Práctica 9: Análisis de Tablas de Contingencia y tablas 2 × 2

Juan Antonio Villatoro García – Grupo BioestadisticaR

UNIVERSIDAD DE GRANADA

Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo BioestadisticaR formado por:
Antonio Martín Andrés
Juan de Dios Luna del Castillo,
Pedro Femia Marzo,
Miguel Ángel Montero Alonso,
Christian José Acal González,
Pedro María Carmona Sáez,
Juan Manuel Melchor Rodríguez,
José Luis Romero Béjar,
Manuela Expósito Ruíz,
Juan Antonio Villatoro García.

Todos los integrantes del grupo han participado en todas las actividades, en su elección, construcción, correcciones o en su edición final, no obstante, en cada una de ellas, aparecerán uno o más nombres correspondientes a las personas que han tenido la máxima responsabilidad de su elaboración junto al grupo de BioestadisticaR.
Todos los materiales están protegidos por la Licencia Creative Commons CC BY-NC-ND que permite "descargar las obras y compartirlas con otras personas, siempre que se reconozca su autoría, pero no se pueden cambiar de ninguna manera ni se pueden utilizar comercialmente".

Práctica 9: : Análisis de Tablas de Contingencia y tablas 2 × 2

UNIVERSIDAD DE GRANADA

Homogeneidad de un carácter cualitativo entre muestras independientes

Función tablarxc() de la librería BioestadisticaR

Datos agregados

tablarxc(frecs, tablas)

frecs: *data.frame con los valores de la tabla de contingencia*
tablas: *indica si queremos representar alguna tabla más en el análisis:*

"E": Tabla de las frecuencias esperadas
"F": Tabla de los porcentajes por filas
"C": Tabla de los porcentajes por columnas

```
> hipertenso_SI=c(22,16,15,31,23,25)
> hipertenso_NO=c(78,84,85,69,77,75)
> tabla=data.frame(hipertenso_SI,hipertenso_NO)
> tablarxc(frecs=tabla, tablas="F")

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

# Frecuencias observadas
         hipertenso_SI   hipertenso_NO Total
  1                 22              78   100
  2                 16              84   100
  3                 15              85   100
  4                 31              69   100
  5                 23              77   100
  6                 25              75   100
  Total            132             468   600

# Validez del test chi-cuadrado
  Frecuencia minima esperada = 22
  0 frecuencias esperadas son menores a 1 y 0 son menores
 a 5 (el 0% de la tabla)

# Test chi-cuadrado
  X2exp(5 gl) = 10.2564, p = 0.0683

# Porcentajes por filas
         hipertenso_SI   hipertenso_NO Total
  1               0.22            0.78  1.00
  2               0.16            0.84  1.00
  3               0.15            0.85  1.00
  4               0.31            0.69  1.00
  5               0.23            0.77  1.00
  6               0.25            0.75  1.00
  Total           0.22            0.78  1.00
```

Distrito / Hipertensión	SI	NO	Totales
Distrito 1	$E_{11} = 22$	$E_{12} = 78$	$F_1 = 100$
Distrito 2	$E_{21} = 22$	$E_{22} = 78$	$F_2 = 100$
Distrito 3	$E_{31} = 22$	$E_{32} = 78$	$F_3 = 100$
Distrito 4	$E_{41} = 22$	$E_{42} = 78$	$F_4 = 100$
Distrito 5	$E_{51} = 22$	$E_{52} = 78$	$F_5 = 100$
Distrito 6	$E_{61} = 22$	$E_{62} = 78$	$F_6 = 100$
Totales	$C_1 = 132$	$C_2 = 468$	$T = 600$

Frecuencias observadas

Comprobación de la validez del test

P-valor del test

Porcentaje por filas

Datos en un data.frame

tablarxc(fvar, cvar, fcat, ccat, tablas)

fvar: *variable de las filas*
cvar: *variable de las columnas*
Fcat: *categorías de las filas*
ccat: *categorías de las columnas*
tablas: *indica si queremos representar alguna tabla más en el análisis:*

¿Hay homogeneidad en la actividad física entre los grupos de edad del fichero osteo.sav?

```
> tablarxc(fvar = osteo$grupo_edad, cvar = osteo$acfis,
+          fcat = levels(osteo$grupo_edad), ccat = levels(osteo$acfis),
+          tablas = "E")

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

# Frecuencias observadas
            Sí   No Total
  < 25      19   13    32
  25 - 33   13   19    32
  > 33      13   17    30
  Total     45   49    94

# Frecuencias esperadas
              Sí    No Total
  < 25    15.32 16.68 32.00
  25 - 33 15.32 16.68 32.00
  > 33    14.36 15.64 30.00
  Total   45.00 49.00 94.00

# Validez del test chi-cuadrado
  Frecuencia minima esperada = 14.36
  0 frecuencias esperadas son menores a 1 y 0 son menores a 5 (el 0% de
 la tabla)

# Test chi-cuadrado
  X2exp(2 gl) = 2.6179, p = 0.2701
```

Frecuencias observadas

Frecuencias esperadas

Comprobación de la validez del test

P-valor del test

© Juan Antonio Villatoro-García – Grupo BioestadisticaR

Práctica 9: : Análisis de Tablas de Contingencia y tablas 2 × 2

UNIVERSIDAD DE GRANADA

Independencia de caracteres cualitativos

Función tablarxc() de la librería BioestadisticaR (igual que para el caso de homogeneidad)

Datos agregados

tablarxc(frecs, tablas)

frecs: *data.frame con los valores de la tabla de contingencia*

tablas: *indica si queremos representar alguna tabla más en el análisis:*

"E": Tabla de las frecuencias esperadas

"F": Tabla de los porcentajes por filas

"C": Tabla de los porcentajes por columnas

```
> Palp_baja=c(20,60,45,20)
> Palp_media=c(5,20,15,15)
> Palp_alta=c(5,10,25,12)
> tabla=data.frame(Palp_baja,Palp_media,
+                  Palp_alta)
> tablarxc(frecs=tabla, tablas=c("F"))
```

Análisis / Palpación	Baja	Media	Alta	Totales
1	$E_{11} = 17.26$	$E_{12} = 6.55$	$E_{13} = 6.19$	$F_1 = 30$
2	$E_{21} = 51.78$	$E_{22} = 19.64$	$E_{23} = 18.57$	$F_2 = 90$
3	$E_{31} = 48.91$	$E_{32} = 18.55$	$E_{33} = 17.54$	$F_3 = 85$
4	$E_{41} = 27.05$	$E_{42} = 10.26$	$E_{43} = 9.70$	$F_4 = 47$
Totales	$C_1 = 145$	$C_2 = 55$	$C_3 = 52$	$T = 252$

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

# Frecuencias observadas
         Palp_baja   Palp_media   Palp_alta Total
  1             20            5           5    30
  2             60           20          10    90
  3             45           15          25    85
  4             20           15          12    47
  Total        145           55          52   252

# Validez del test chi-cuadrado
  Frecuencia minima esperada = 6.19
  0 frecuencias esperadas son menores a 1 y 0 son menores
 a 5 (el 0% de la tabla)

# Test chi-cuadrado
  X2exp(6 gl) = 15.033, p = 0.02

# Porcentajes por filas
         Palp_baja   Palp_media   Palp_alta  Total
  1         0.6667       0.1667      0.1667 1.0000
  2         0.6667       0.2222      0.1111 1.0000
  3         0.5294       0.1765      0.2941 1.0000
  4         0.4255       0.3191      0.2553 1.0000
  Total     0.5754       0.2183      0.2063 1.0000
```

Frecuencias observadas

Comprobación de la validez del test

P-valor del test

Porcentaje por filas

Datos provenientes de un fichero

tablarxc(fvar, cvar, fcat, ccat, tablas)

fvar: *variable de las filas*

cvar: *variable de las columnas*

fcat: *categorías de las filas*

ccat: *categorías de las columnas*

tablas: *indica si queremos representar alguna tabla más en el análisis:*

¿Hay independencia entre padecer retinopatía y el sexo en los datos de fichero osteo.sav?

```
> tablarxc(fvar=osteo$retin,cvar=osteo$sexo,
+          fcat=c("No","Leve","Grave"),
+          ccat = c("Hombre","Mujer"),tablas=c("F","C"))

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

# Frecuencias observadas
         Hombre   Mujer Total
  No         33      30    63
  Leve        8      10    18
  Grave       4       9    13
  Total      45      49    94

# Validez del test chi-cuadrado
  Frecuencia minima esperada = 6.22
  0 frecuencias esperadas son menores a 1 y 0 son menores
 a 5 (el 0% de la tabla)

# Test chi-cuadrado
  X2exp(2 gl) = 2.1218, p = 0.3461

# Porcentajes por filas
         Hombre   Mujer  Total
  No     0.5238  0.4762 1.0000
  Leve   0.4444  0.5556 1.0000
  Grave  0.3077  0.6923 1.0000
  Total  0.4787  0.5213 1.0000

# Porcentajes por columnas
         Hombre   Mujer  Total
  No     0.7333  0.6122 0.6702
  Leve   0.1778  0.2041 0.1915
  Grave  0.0889  0.1837 0.1383
  Total  1.0000  1.0000 1.0000
```

Frecuencias observadas

Comprobación de la validez del test

P-valor del test

Porcentaje por filas

Porcentaje por columnas

© Juan Antonio Villatoro-García – Grupo BioestadisticaR

Práctica 9: : Análisis de Tablas de Contingencia y tablas 2 × 2

UNIVERSIDAD DE GRANADA

Enfermedad vs factor de riesgo. Tablas 2 × 2. Medidas de asociación

Función tabla2x2() de la librería BioestadisticaR:

tabla2x2(o, fcat, ccat, estudio, tablas)

o: *valores de la tabla (orden (o11,o12,o21,o22)*
fcat: *categorías de las filas*
ccat: *categorías de las columnas*
estudio: *tipo de estudio:* ***"T":*** *Transversal /* ***"P":*** *Prospectivo" /* ***R":*** *Retrospectivo*
tablas: *indica si queremos representar alguna tabla más en el análisis.*

Trasversal

Enfermedad / Factor de Riesgo	FR	FR^c	Totales
E	$O_{11} = 47$	$O_{12} = 46$	$F_1 = 93$
E^c	$O_{21} = 15$	$O_{22} = 92$	$F_2 = 107$
Totales	$C_1 = 62$	$C_2 = 138$	$T = 200$

tabla2x2(o, fcat, ccat, estudio = "T" , tablas)

```
> tabla2x2(o=c(47,46,15,92), fcat=c("Niños con TDA-H","Niños sin TDA-H"),
+          ccat=c("Algún padre con TDA-H","Padres sin TDA-H"),estudio="T",)

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

# Frecuencias observadas
                    Algún padre con TDA-H   Padres sin TDA-H Total
  Niños con TDA-H                      47                46    93
  Niños sin TDA-H                      15                92   107
  Total                                62               138   200

# Test Chi-cuadrado para un estudio transversal

  Frecuencia mínima esperada = 28.83  > 3.9

  X2 = 31.0129,   gl = 1,  p < 0.001, (cpc = 0.5)
  Test exacto de Fisher: p < 0.001

  --- Otros criterios X2:
  X2 = 31.0215,   gl = 1,  p < 0.001, (sin cpc)
  X2 = 29.3377,   gl = 1,  p < 0.001, (cpc de Yates = 100)

# Estimación de la prevalencia en un estudio transversal
  Método de Wald ajustado:
  Prev=0.4657; 95%-IC(prev)=(0.3972, 0.5341)

# 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=0.4247; 95%-IC(d)=(0.2829, 0.5465)

  Riesgo relativo:
  Rr=2.2742; 95%-IC(Rr)=(1.7143, 2.9631)

  Riesgo atribuible:
  Ra=0.2832; 95%-IC(Ra)= (0.1666, 0.3834)

  Razón del producto cruzado (odds ratio):
  OR=6.2667; 95%-IC(OR)= (3.1103, 11.9483)
```

Comprobación de la validez del test

Resultados del test

Medidas de asociación

Prospectivo

Enfermedad / Factor de Riesgo	FR	FR^c	Totales
E	$O_{11} = 47$	$O_{12} = 46$	$F_1 = 93$
E^c	$O_{21} = 15$	$O_{22} = 92$	$F_2 = 107$
Totales	$C_1 = 62$	$C_2 = 138$	$T = 200$

tabla2x2(o, fcat, ccat, estudio = "P" , tablas)

```
> tabla2x2(o=c(47,46,15,92),
+          fcat=c("Niños con TDA-H","Niños sin TDA-H"),
+          ccat=c("Algún padre con TDA-H","Padres sin TDA-H"),
+          estudio="P")

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

# Frecuencias observadas
                    Algún padre con TDA-H   Padres sin TDA-H
  Niños con TDA-H                      47                46
  Niños sin TDA-H                      15                92
  Total                                62               138
                   Total
  Niños con TDA-H     93
  Niños sin TDA-H    107
  Total              200

# Test Chi-cuadrado para un estudio prospectivo

  Frecuencia mínima esperada = 28.83  > 7.7

  X2 = 31.0044,   gl = 1,  p < 0.001, (cpc = 1)
  Test exacto de Fisher: p < 0.001

  --- Otros criterios X2:
  X2 = 31.0215,   gl = 1,  p < 0.001, (sin cpc)
  X2 = 29.3377,   gl = 1,  p < 0.001, (cpc de Yates = 100)

# 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=0.4247; 95%-IC(d)=(0.2829, 0.5465)

  Riesgo relativo:
  Rr=2.2742; 95%-IC(Rr)=(1.7143, 2.9631)

  Razón del producto cruzado (odds ratio):
  OR=6.2667; 95%-IC(OR)= (3.1103, 11.9483)
```

Comprobación de la validez del test

Resultados del test

Medidas de asociación

Retrospectivo

Enfermedad / Factor de Riesgo	FR	FR^c	Totales
E	$O_{11} = 47$	$O_{12} = 46$	$F_1 = 93$
E^c	$O_{21} = 15$	$O_{22} = 92$	$F_2 = 107$
Totales	$C_1 = 62$	$C_2 = 138$	$T = 200$

tabla2x2(o, fcat, ccat, estudio = "R" , tablas)

```
> tabla2x2(o=c(47,46,15,92),
+          fcat=c("Niños con TDA-H","Niños sin TDA-H"),
+          ccat=c("Algún padre con TDA-H","Padres sin TDA-H"),
+          estudio="R")

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

# Frecuencias observadas
                    Algún padre con TDA-H   Padres sin TDA-H Total
  Niños con TDA-H                      47                46    93
  Niños sin TDA-H                      15                92   107
  Total                                62               138   200

# Test Chi-cuadrado para un estudio retrospectivo

  Frecuencia mínima esperada = 28.83  > 7.7

  X2 = 31.0044,   gl = 1,  p < 0.001, (cpc = 1)
  Test exacto de Fisher: p < 0.001

  --- Otros criterios X2:
  X2 = 31.0215,   gl = 1,  p < 0.001, (sin cpc)
  X2 = 29.3377,   gl = 1,  p < 0.001, (cpc de Yates = 100)

# Medidas de asociación para un estudio retrospectivo

  Riesgo atribuible*:
  Ra=0.6371; 95%-IC(Ra)= (0.4274, 0.77)
  * La estimación de Ra para estudios retrospectivos es una aproxim
ación válida si la prevalencia de la enfermedad es baja: P(E) < 10%

  Razón del producto cruzado (odds ratio):
  OR=6.2667; 95%-IC(OR)= (3.1103, 11.9483)
  * La estimación para OR sirve de aproximación al riesgo relativo
 siempre que la prevalencia de la enfermedad sea P(E) < 10%
```

Comprobación de la validez del test

Resultados del test

Riesgo atribuible

Razón de productos cruzados

© Juan Antonio Villatoro-García – Grupo BioestadisticaR