

Causalidad, Test McNemar, Coeficiente Kappa, Conceptos Sensibilidad y Especificidad

1. Investigue el término *prima facie* y su relación con la causalidad. Basado en el concepto anterior defina, en términos de probabilidades condicionales, cuándo un evento C causaría un evento E . Explique en sus propias palabras esta definición.

Respuesta Fuente 1 Fuente 2 Safety and Reliability. Theory and Applications

Una causa *prima facie* de un evento E , es un evento que precede temporalmente a E y que está positivamente relacionado con él. Lo anterior, se puede escribir matemáticamente de la siguiente forma:

1. Sea $r < s$ valores que denotan momentos en el tiempo, donde temporalmente r está previo a s . El evento C_r (que ocurre en el momento r) es una causa *prima facie* del evento E_s (que ocurre en el momento s) si:

$$P(E_s|C_r) > P(E_s)$$

Existen autores que además especifican formalmente que $P(C) > 0$. Parafraseando lo anterior, si C es un evento que ocurre antes de E y tal que, la probabilidad de ocurrencia del evento E aumenta si ésta se condiciona al evento C (es decir, dada la ocurrencia de C se observa un aumento en la probabilidad de ocurrencia de E), entonces se puede afirmar que C es una causa *prima facie*.

Por ejemplo, si tenemos información de pacientes con cáncer de pulmón y en su historial médico se conoce si el paciente fumaba o no fumaba antes de presentar el cáncer, y además observamos que la proporción de pacientes con cancer en fumadores es mayor que la proporción de pacientes con cáncer en no fumadores, podríamos entender el evento *fumar* como una causa *prima facie* del evento presentar cáncer de pulmón.

Es relevante notar que afirmar que C es una causa *prima facie* de E no significa realmente que C causa directamente E . La definición de causa *prima facie* está sujeta bajo dos condiciones que no asegurarían necesariamente la causalidad directa; de hecho, las causas *prima facie* pueden ser *espurias* o (exclusivo) *genuinas*.

2. Utilice la tabla 11.11 de Chernick and Friis (2003) y determine si fumar sería, a primera vista, una causa del cáncer de pulmón. ¿Cuál es el principal supuesto que se debe realizar para responder esta pregunta?

TABLE 11.11. Smoking and Lung Cancer Data for a Cohort Study

Smokers	Lung cancer		Total
	Present	Absent	
Yes	98	202	300
No	35	665	700

Respuesta

Sea E el evento presentar o padecer cáncer de pulmón y C el evento fumar/ser fumador. Se tiene que:

$$P(E) = \frac{98 + 35}{98 + 202 + 35 + 665} = \frac{133}{1000} = 0.133$$

$$P(E|C) = \frac{98}{202 + 98} = \frac{98}{300} = 0.326$$

Luego,

$$P(E|C) > P(E)$$

Para decir que C es una causa *prima facie* de E, se debe suponer que el evento medido C ocurrió antes de ocurrir E. Que la variable medida *Smokers=Yes* indica que el paciente fumó previo presentar cáncer de pulmón. Además, con la información que se posee y sólo afirmaría que C es una causa *prima facie* de E, no diría que C causa directamente E, pues podrían existir variables mediadoras o confundentes entre E y C que no han sido observadas, las cuales podrían indicar que C es una *prima facie espuria* de E.

3. Un grupo de especialistas analizó y discutió sobre las siguientes preguntas: 1) ¿Crees que una persona tiene derecho a terminar con su propia vida si tiene una enfermedad incurable? y 2) Cuando una persona tiene una enfermedad que no puede ser curada, ¿cree que a los médicos se les debe permitir terminar con la vida del paciente por algún medio indoloro si el paciente y su familia lo solicitan? Las opiniones fueron tabuladas en la siguiente tabla, en donde las preguntas 1 y 2 están referidas como “suicidio” y “dejar morir al paciente”.

	Dejar morir al paciente	
	Si	No
Suicidio	1097	90
No	203	435

Evalúe el test, exacto y asintótico, de McNemar e interprete. Escriba, en sus propias palabras, las hipótesis involucradas en el problema.

Respuesta Fuente 1

Lo que interesa comparar es si la muerte del paciente resulta ser más justificada o entendible para los especialistas en uno u otro caso (suicidio voluntario vs decisión médica). Por lo cual, la hipótesis nula del Test de McNemar sería que en ambos casos, la proporción de respuestas afirmativas (o respuestas no afirmativas, note que resulta equivalentes) para cualquiera de los casos sería la misma, es decir, entre el caso de suicidio voluntario y el caso de la decisión médica se observa una distribución igual de opiniones.

Se definen los datos en R:

```
(discusion<-matrix(c(1097,90,203,435),byrow=TRUE,ncol=2,nrow=2))
```

```

      [,1] [,2]
[1,] 1097  90
[2,]  203 435
colnames(discusion)<-c("Si","No")
> rownames(discusion)<-c("Si","No")
addmargins(discusion)
      Si  No  Sum
Si   1097  90 1187
No    203 435 638
Sum  1300 525 1825

round(addmargins(prop.table(discusion)),3)
      Si    No  Sum
Si  0.601 0.049 0.65
No  0.111 0.238 0.35
Sum 0.712 0.288 1.00

# Aproximado
mcnemar.test(discusion)

      McNemar's Chi-squared test with continuity correction

data:  discusion
McNemar's chi-squared = 42.812, df = 1, p-value = 6.025e-11

# Exacto
# install.packages("exact2x2")
library("exact2x2")
mcnemar.exact(discusion)

      Exact McNemar test (with central confidence intervals)

data:  discusion
b = 90, c = 203, p-value = 3.464e-11
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.3419718 0.5709854
sample estimates:
odds ratio
 0.4433498

```

Ambos tests, ya sea aproximado o exacto entregan como resultado que existirían diferencias significativas en la distribución de opiniones por caso (suicidio vs decisión médica). Es decir, realizar una u otra pregunta no genera las mismas opiniones, seguramente, porque hablar de *la persona se suicidó* genera un impacto distinto que hablar de *los médicos estuvieron de acuerdo con terminar con su vida*, quizás por temas religiosos o éticos.

4. **3.2)** In a study of intraobserver variability in assessing cervical smears, 3325 slides were screened for the presence or absence of abnormal squamous cells. Each slide was screened by a particular observer and then rescreened six months later by the same observer. The results are shown in Table E3.2. Is there a statistical relationship between first screening and second screening?

TABLE E3.2

First screening	Second screening		Total
	Present	Absent	
Present	1763	489	2252
Absent	403	670	1073
Total	2166	1159	3325

Respuesta

Se pretende estudiar si la detección de células escamosas anormales difiere en la primera evaluación y en la segunda. Se definen los datos en R:

```
(screening<-matrix(c(1763,489,403,670),byrow=TRUE,ncol=2,nrow=2))
      [,1] [,2]
[1,] 1763  489
[2,]  403  670

colnames(screening)<-c("Present","Absent")
rownames(screening)<-c("Present","Absent")

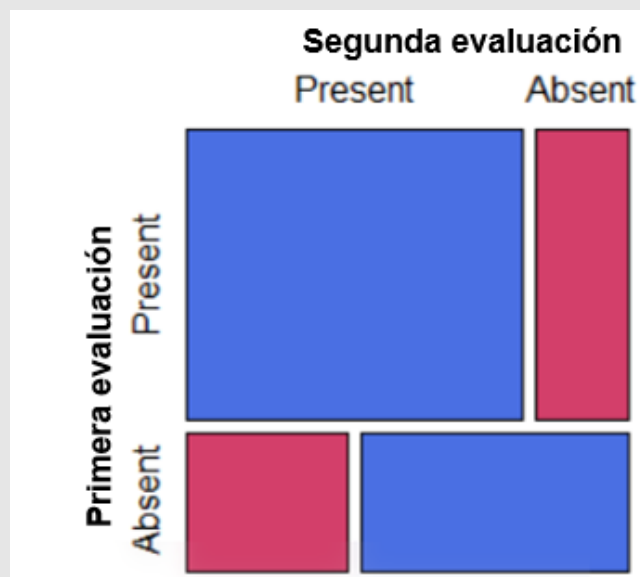
addmargins(screening)
      Present Absent Sum
Present   1763   489 2252
Absent     403   670 1073
Sum        2166  1159 3325

prop.table(screening,margin=1)

# install.packages("vcd")

library(vcd)

mosaic(screening,shade=TRUE)
```



Es posible observar que el diagnóstico *Present* es aquél predominante en ambas evaluaciones, ahora bien, es posible observar que en la segunda evaluación se tiene una menor proporción de evaluaciones *Absent* en comparación a la primera evaluación.

La pregunta de investigación es ¿son los resultados concordantes entre estas dos evaluaciones?. Para ésto se calculará el coeficiente κ , recordar que la hipótesis nula de este test es que $\kappa = 0$ es decir, los diagnósticos de ambas evaluaciones serían aleatorizados y por lo tanto, no concordantes ni discordantes, y la hipótesis alternativa es $\kappa \neq 0$.

```
Kappa.test(screening)
$Result

Estimate Cohen's kappa statistics and test the
null hypothesis that the extent of agreement is
same as random (kappa=0)

data:  screening
Z = 20.651, p-value < 2.2e-16
95 percent confidence interval:
 0.3651651 0.4326508
sample estimates:
[1] 0.3989079

$Judgement
[1] "Fair agreement"
```

Se obtiene $\hat{\kappa} = 0.398$ y utilizando un 95% de confianza el intervalo de confianza obtenido es de (0.365, 0.432), notar que el intervalo de confianza no contiene valores negativos ni tampoco al cero, por lo cual, podríamos decir que existe un nivel de acuerdo o de concordancia entre ambos métodos, si bien el nivel de acuerdo no es perfecto, los resultados de ambas evaluaciones no son opuestas y ambos resultados podrían complementarse.

3.3) From the intraobserver variability study above, find:

a) The probability that abnormal squamous cells were found to be absent in both screenings.

Respuesta

La probabilidad sería:

$$P_{\text{absent-absent}} = \frac{670}{3325} = 0.2015038$$

La probabilidad de que en ambos métodos no se detecten células escamosas anormales es de 0.2015.

b) The probability of an absence in the second screening given that abnormal cells were found in the first screening.

Respuesta

La probabilidad sería:

$$P_{\text{absent-present}} = \frac{489}{2252} = 0.2171403$$

La probabilidad de no detectar células escamosas anormales dado que se detectaron en la primera evaluación es de 0.217.

c) The probability of an abnormal presence in the second screening given that no abnormal cells were found in the first screening.

Respuesta

La probabilidad sería:

$$P_{\text{present-absent}} = \frac{403}{1073} = 0.3755825$$

La probabilidad de detectar células escamosas anormales dado que no se detectaron en la primera evaluación es de 0.376.

d) The probability that the screenings disagree.

Respuesta

La probabilidad sería:

$$P_{\text{discordancia}} = \frac{489 + 403}{3325} = 0.2682707$$

La probabilidad de que las evaluaciones entreguen resultados discordantes sería de 0.268.

5. In a case-control study, 317 patients suffering from endometrial carcinoma were matched individually with 317 other cancer patients in a hospital and the use of estrogen in the six months prior to diagnosis was determined (Table E6.2). Use McNemar's chi-square test to investigate the significance of the association between estrogen use and endometrial carcinoma; state your null and alternative hypotheses.

TABLE E6.2

Cases	Controls	
	Estrogen	No estrogen
Estrogen	39	113
No estrogen	15	150

Respuesta

La pregunta de investigación es si se ve una diferencia en la proporción de uso de estrógeno por grupo (grupo controles y grupo casos carcinoma endometrial). La hipótesis nula de interés sería:

$$P_{\text{Uso de estrógeno en Casos}} = P_{\text{Uso de estrógeno en Controles}}$$

Y la hipótesis alternativa es $P_{\text{Uso de estrógeno en Casos}} \neq P_{\text{Uso de estrógeno en Controles}}$.

Se definen los datos en R y se procede a realizar el test de McNemar:

```
(estrogen<-matrix(c(39,113,15,150),byrow=TRUE,ncol=2,nrow=2))
      [,1] [,2]
[1,]   39  113
[2,]   15  150
```

```
colnames(estrogen)<-c("Uso Estrogeno","No Uso Estrogeno")
rownames(estrogen)<-c("Uso Estrogeno","No Uso Estrogeno")
```

```
round(addmargins(prop.table(estrogen)),3)
      Uso Estrogeno No Uso Estrogeno Sum
Uso Estrogeno      0.123           0.356 0.479
No Uso Estrogeno    0.047           0.473 0.521
Sum                 0.170           0.830 1.000
```

```
mcnemar.test(estrogen) #Asintotico
```

McNemar's Chi-squared test with continuity

correction

```
data:  estrogen
McNemar's chi-squared = 73.508, df = 1, p-value <
2.2e-16
```

En el test asintótico se concluye utilizando un 95% de confianza que existe evidencia suficiente para rechazar la hipótesis nula de igualdad de proporciones. Es decir, la proporción de uso de estrógenos previo al diagnóstico no sería igual en ambos grupos de estudio, por lo que se podría asociar al carcinoma endometrial.

6. It has been noted that metal workers have an increased risk for cancer of the internal nose and paranasal sinuses, perhaps as a result of exposure to cutting oils. Therefore, a study was conducted to see whether this particular exposure also increases the risk for squamous cell carcinoma of the scrotum (see Example 6.2). Cases included all 45 squamous cell carcinomas of the scrotum diagnosed in Connecticut residents from 1955 to 1973, as obtained from the Connecticut Tumor Registry. Matched controls were selected for each case based on the age at death (within eight years), year of death (within three years), and number of jobs as obtained from combined death certificate and directory sources. An occupational indicator of metal worker (yes/no) was evaluated as the possible risk factor in this study. The results are shown in Table 6.2. Test to compare the cases versus the controls using the McNemar's chi-square test; state clearly your null and alternative hypotheses and choice of test size.

TABLE 6.2

Cases	Controls	
	Yes	No
Yes	2	26
No	5	12

Respuesta

La pregunta de investigación es si se ve una diferencia en la proporción de ocupación de tipo metalúrgica por grupo (grupo controles y grupo casos carcinoma endometrial). La hipótesis nula de interés sería:

$$P_{\text{Ocupación metalúrgica en Casos}} = P_{\text{Ocupación metalúrgica en Controles}}$$

Y la hipótesis alternativa sería su complementaria.

Se definen los datos en R y se procede a realizar el test de McNemar:

```
(metalurgia<-matrix(c(2,26,5,12),byrow=TRUE,ncol=2,nrow=2))
      [,1] [,2]
[1,]    2  26
[2,]    5  12

colnames(metalurgia)<-c("Ocupación Metalúrgica","No ocupación Metalúrgica")
rownames(metalurgia)<-c("Ocupación Metalúrgica","No ocupación Metalúrgica")

mcnemar.exact(metalurgia) #Test Exacto

Exact McNemar test (with central confidence
```

```

intervals)

data:  metalurgia
b = 26, c = 5, p-value = 0.0001922
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.96497 17.34044
sample estimates:
odds ratio
      5.2

```

Se realizó un test de McNemar exacto dado que el número de observaciones por celda no es muy alto. El valor-p obtenido al utilizar un 95% es de 0.0001, de lo cual, es posible concluir que se rechaza la hipótesis nula de igualdad de proporciones, es decir, la proporción de ocupación metalúrgica en casos y controles no es la misma, por lo que se podría asociar la ocupación metalúrgica al desarrollo de carcinoma de células escamosas del escroto.

7. Defina, en términos probabilísticos, sensibilidad y especificidad.

Respuesta

La sensibilidad de un test se define como la probabilidad de que un caso positivo se clasifique correctamente:

$$P(\text{Clasificado Positivo}|\text{Presenta enfermedad}) = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}}$$

La especificidad de un test se define como la probabilidad de que un caso negativo se clasifique correctamente:

$$P(\text{Clasificado Negativo}|\text{No presenta enfermedad}) = \frac{\text{Verdaderos negativos}}{\text{Verdaderos negativos} + \text{Falsos positivos}}$$

Recordemos que en clases definimos que los casos falsos positivos corresponden a errores de tipo I y los casos falsos negativos corresponden a errores de tipo II. De ahí es posible realizar la conexión con los errores de un test de hipótesis:

$$\alpha = \text{Tasa de falsos positivos} = \frac{FP}{FP + TN} = 1 - \text{Especificidad}$$

$$\beta = \text{Tasa de falsos negativos} = \frac{FN}{FN + TP} = 1 - \text{Sensibilidad} = \text{Poder del test}$$

De ahí que se dice que la sensibilidad corresponde al poder del test.

8. Defina falsos positivos, falsos negativos, prevalencia, predictividad positiva, predictividad negativa, incidencia.

Respuesta

Falsos Positivos Aquellos casos que fueron clasificados con la enfermedad, pero que en realidad no presentan la enfermedad.

Falsos Negativos Aquellos casos que fueron clasificados sin la enfermedad, pero que poseen la enfermedad. Estos son los casos en general, más graves, pues decirle a un

paciente que no presenta la enfermedad cuando en realidad, sí la posee, puede resultar perjudicial no sólo para él, sino que, para las personas que lo rodean.

Prevalencia Corresponde a la proporción de personas que presentan la enfermedad, es decir, qué tan probable es, en la muestra, encontrar un caso positivo. Este valor podría extrapolarse a una población de interés a través de la muestra.

Predictividad positiva Corresponde a la probabilidad de que un paciente que fue clasificado con la enfermedad, tenga efectivamente la enfermedad.

Predictividad negativa Corresponde a la probabilidad de que un paciente que fue clasificado sin la enfermedad, efectivamente no tenga la enfermedad.

Incidencia Las incidencias son los casos nuevos de contagio o adquisición de la enfermedad en un tiempo fijo. Y la incidencia mide la proporción de los nuevos contagios en el período de estudio en la población de interés.

9. ¿Por qué una curva ROC que se aprecia como la diagonal desde (0,0) a (1,1) implica un mal test?

Respuesta

El área de la diagonal desde (0,0) a (1,1) es de 0.5, lo cual indicaría que, la probabilidad de clasificar correctamente a dos individuos uno enfermo y otro no enfermo es del 50%, que básicamente nos dice que es similar a lanzar una moneda y si sale cara, clasificarlo como positivo o negativo. Además, una curva de este tipo indica que la sensibilidad crece a la misma velocidad que 1-especificidad, por lo que se le estaría entregando el mismo valor clasificar a un caso positivo que a un caso negativo, lo cual, en medicina, no es realmente el caso, pues clasificar correctamente a un caso positivo (enfermo) posee por lejos más valor dadas las consecuencias que podría generar no detectar a un paciente enfermo.