Ejercicios Inferencia Unidad 1. Probabilidad

Alex Sanchez y Santi Perez-Hoyos

2024-10-13

Table of contents

1 UNIDAD I. Ejercicios y cuestiones. Soluciones **1**
Problema 1 . . . 1
1.0.1 Solución . . . 1
Problema 2 . . . 2
1.0.1 Solución . . . 2
Problema 3 . . . 4
1.0.1 Solución . . . 4
1.1 Problema 4 . . . 5
1.1.1 Solución . . . 5
Problema 5 . . . 6
1.1.1 Solución . . . 6

1 UNIDAD I. Ejercicios y cuestiones. Soluciones

Problema 1

Sean A y B dos sucesos. Suponiendo que $P(A) = 0.3, P(B) = 0.6$, y $P(A \cap B) = 0.1$, calcula las siguientes probabilidades:

a) $P(A \cup B)$

b) $P(A^c)$

c) $P(Ac \cap B)$

d) $P(A \cap B^c)$

e) $P(A^c \cap B^c)$

1.0.1 Solución

a. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.6 - 0.1 = 0.8$
b. $P\left(A^c\right) = 1 - P(A) = 1 - 0.3 = 0.7$
c. $P\left(A^c \cap B\right) = P(B) - P(A \cap B) = 0.6 - 0.1 = 0.5$
d. $P\left(A \cap B^c\right) = P(A) - P(A \cap B) = 0.3 - 0.1 = 0.2$
e. $P\left(A^c \cap B^c\right) = 1 - P(A \cup B) = 1 - 0.8 = 0.2$

Problema 2

Una población está afectada por tres enfermedades diferentes A, B i C. La probabilidad de que una persona sufra A es 0.30 , la probabilidad de que sufra B es 0.20 y la probabilidad de que sufra C es 0.15 . La probabilidad de que una persona sufra A y B es 0.12 , la que sufra A y C es 0.09 y la que sufra B y C es 0.06 . La probabilidad de que una persona sufra las tres enfermedades es 0.03 . Se piden las probabilidades de que una persona escogida al azar:

1. padezca al menos una enfermedad
2. sólo sufra A
3. sufra B o C, pero no sufra A
4. sufra A o no sufra ni B ni C.

1.0.1 Solución

---

a) **¿Cuál es la probabilidad de que una persona padezca al menos una enfermedad?**

Queremos calcular la probabilidad de que una persona sufra al menos una de las tres enfermedades, es decir, $P(A \cup B \cup C)$.

Para calcular $P(A \cup B \cup C)$, usamos la regla de inclusión-exclusión:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Sustituyendo los valores dados en el enunciado:

$$P(A \cup B \cup C) = 0.30 + 0.20 + 0.15 - 0.12 - 0.09 - 0.06 + 0.03 = 0.41$$

Por lo tanto, la probabilidad de que una persona padezca al menos una enfermedad es **0.41**.

---

b) **¿Cuál es la probabilidad de que una persona solo sufra A?**

Para resolver esto, necesitamos calcular la probabilidad de que la persona sufra A, pero no B ni C, es decir, $P(A \cap B^c \cap C^c)$.

Podemos calcular $P(A \cap B^c \cap C^c)$ restando de $P(A)$ la probabilidad de que la persona sufra A junto con alguna de las otras dos enfermedades:

$$P(A \cap B^c \cap C^c) = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$$

Sustituyendo los valores:

$$P(A \cap B^c \cap C^c) = 0.30 - 0.12 - 0.09 + 0.03 = 0.12$$

Por lo tanto, la probabilidad de que una persona solo sufra A es **0.12**.

---

c) **¿Cuál es la probabilidad de que una persona sufra B o C, pero no sufra A?**

Aquí buscamos la probabilidad $P(A^c \cap (B \cup C))$, es decir, la probabilidad de que la persona no tenga A, pero tenga B o C.

Primero, calculamos $P(B \cup C)$ utilizando la regla de inclusión-exclusión:

$$P(B \cup C) = P(B) + P(C) - P(B \cap C)$$

Sustituyendo los valores:

$$P(B \cup C) = 0.20 + 0.15 - 0.06 = 0.29$$

Ahora, para calcular $P(A^c \cap (B \cup C))$, restamos de $P(B \cup C)$ la probabilidad de que la persona tenga A y alguna de las enfermedades B o C, es decir, $P(A \cap (B \cup C))$:

$$P(A \cap (B \cup C)) = P(A \cap B) + P(A \cap C) - P(A \cap B \cap C)$$

Sustituyendo los valores:

$$P(A \cap (B \cup C)) = 0.12 + 0.09 - 0.03 = 0.18$$

Finalmente, restamos:

$$P(A^c \cap (B \cup C)) = P(B \cup C) - P(A \cap (B \cup C)) = 0.29 - 0.18 = 0.11$$

Por lo tanto, la probabilidad de que una persona sufra B o C, pero no A, es **0.11**.

---

d) **¿Cuál es la probabilidad de que una persona sufra A o no sufra ni B ni C?**

Aquí buscamos la probabilidad $P(A \cup (B^c \cap C^c))$, es decir, que la persona sufra A o que no sufra ni B ni C.

Primero, calculamos $P(B^c \cap C^c)$, que es la probabilidad de que la persona no sufra ni B ni C. Esto es simplemente $1 - P(B \cup C)$, que ya calculamos previamente:

$$P(B^c \cap C^c) = 1 - P(B \cup C) = 1 - 0.29 = 0.71$$

Ahora, aplicamos la regla de la unión para calcular $P(A \cup (B^c \cap C^c))$:

$$P(A \cup (B^c \cap C^c)) = P(A) + P(B^c \cap C^c) - P(A \cap B^c \cap C^c)$$

Ya calculamos $P(B^c \cap C^c)$, y sabemos que $P(A \cap B^c \cap C^c)$ es la probabilidad de que una persona solo sufra A, que también calculamos previamente:

$$P(A \cap B^c \cap C^c) = 0.12$$

Sustituyendo los valores:

$$P(A \cup (B^c \cap C^c)) = 0.30 + 0.71 - 0.12 = 0.89$$

Por lo tanto, la probabilidad de que una persona sufra A o no sufra ni B ni C es **0.89**.

Resumiendo:

1. La probabilidad de que una persona padezca al menos una enfermedad es **0.41**.
2. La probabilidad de que una persona solo sufra A es **0.12**.
3. La probabilidad de que una persona sufra B o C, pero no A, es **0.11**.
4. La probabilidad de que una persona sufra A o no sufra ni B ni C es **0.89**.

Problema 3

Por los síntomas observados en un enfermo, y según la experiencia acumulada en un gran número de situaciones similares, se deduce que ha podido coger la enfermedad A con probabilidad 1/3, o la enfermedad B con probabilidad 2/3. Con el fin de precisar el diagnóstico, se hace un análisis clínico al enfermo con dos resultados posibles, positivo o negativo. Se sabe, también por experiencia, que en los pacientes que tienen la enfermedad En el análisis es positiva con probabilidad 0.99 , y en los que padecen la enfermedad B lo es con probabilidad 0.06

a) ¿Cuál es la probabilidad de que el análisis dé un resultado negativo?

b) Si el resultado ha sido positivo, ¿cuál es la probabilidad de que el paciente sufra la enfermedad A? ¿Y la probabilidad de que padezca la enfermedad B?

1.0.1 Solución

a.

$$P(Neg) = P(Neg|A) \cdot P(A) + P(Neg|B) \cdot P(B) = \\ = 0.01 \cdot 1/3 + 0.94 \cdot 2/3 = 0.63$$

b.

$$\mathrm{P(A}|Pos) = \frac{P(\text{ Pos }|A)P(A)}{P(\text{ Pos})} = 0.8919, \quad \text{para A,}$$
$$\mathrm{P(B}|Pos) = 1 - \mathrm{P(A}/Positiu) = 0.1081, \quad \text{para } B.$$

Las probabilidades las hemos calculado con R a partir de la información del enunciado:

```
pA<-1/3
pB<-2/3
ppA<-0.99
ppB<-0.06
pn<-(1-ppA)*pA+(1-ppB)*pB
pn
```

```
[1] 0.63
```

1.1 Problema 4

El embolismo pulmonar es una condición relativamente común que necesita hospitalización y que a menudo ocurre en pacientes hospitalizados. La presión arterial menor de 90 mm HG es uno de los criterios importantes para diagnosticar esta condición. Supongamos que la sensibilidad del test es del 95% y la especificidad del test es del 75% y la prevalencia es del 20%.

a) Calcula el valor predictivo positivo del test.

b) Calcula el valor predictivo negativo del test.

c) Responde a las preguntas anteriores si la prevalencia fuera del 80%.

1.1.1 Solución

a) Calcula el valor predictivo positivo del test

$$VP+ = P(\text{ Embolismo } / \text{ Test } +) = \frac{\text{Sens} \times \text{Prev}}{\text{Sens} \times \text{Prev } + (1 - \text{ Esp })(1 - \text{ Prev })}$$

```
sens<-0.95
esp<-0.75
prev<-0.20
vpp<-(sens*prev)/(sens*prev+(1-esp)*(1-prev))
vpp
```

```
[1] 0.4871795
```

b) Calcula el valor predictivo negativo del test

$$VP- = \frac{\text{Esp}(1 - \text{Prev})}{\text{Esp}(1 - \text{Prev}) + (1 - \text{ Sens }) \text{Prev}}$$

```
vpn<-(esp*(1-prev))/(esp*(1-prev)+(1-sens)*prev)
vpn
```

```
[1] 0.9836066
```

Como se observa al tratarse de una prueba muy sensible y poco específica hay pocos falsos negativos y cuando el test da negativo hay una probabilidad muy alta (0.98) de que el individuo sea sano. No así cuando da positivo. Sólo el 48% serán verdaderos enfermos.

c) Responde a las preguntas anteriores si la prevalencia fuera del 80%

```
prev<-0.80
vpp<-(sens*prev)/(sens*prev+(1-esp)*(1-prev))
vpp
```

```
[1] 0.9382716
```

```
vpn<-(esp*(1-prev))/(esp*(1-prev)+(1-sens)*prev)
vpn
```

```
[1] 0.7894737
```

Si la prevalencia es más alta, el VP- sigue siendo alto, aunque no tanto pero hemos aumentado el VP+ hasta el 93% y no habrá tantos falsos positivos. Lo que está claro es el VPN y el VPP dependen de la prevalencia de la enfermedad.

Problema 5

Un índice que evalúa el síndrome de la muerte súbita (SMS) tiene una sensibilidad del 68% y una especificidad del 82%. ¿Cuáles son los valores predictivos positivo y negativo del índice si se aplica a una población donde se producen un $0,21\%$ de muertes súbitas sobre el total de nacimientos?

1.1.1 Solución

La prevalencia del síndrome de la muerte súbita en la población es del 0.21%, es decir 0.0021.

Nos piden que calculemos respectivamente los valores predictivos positivo y negativo del test. Es decir, que tan bien funciona el test para detectar la enfermedad (SMS) cuando da un resultado positivo ($T+$) y para indicar su ausencia (SMS^c), mediante un resultado negativo ($T-$).

$$VP+ = P[SMS|T+], \qquad VP- = P[SMS^c|T-],$$

Puede hacerse el cálculo directamente a partir de las probabilidades condicionadas.

$$\begin{aligned} VP+ = P[SMS|T+] &= \frac{P[T+|SMS] \times P[SMS]}{P[T+]} = \\ &= \frac{P[T+|SMS] \times P[SMS]}{P[T+|SMS] \times P[SMS] + P[T+|SMS^c] \times P[SMS^c]} = \\ &= \frac{\text{Sensibilidad} \times \text{Prevalencia}}{\text{Sensibilidad} \times \text{Prevalencia} + \text{1-Especificidad} \times \text{1-Prevalencia}} \end{aligned}$$

De forma análoga:

$$\begin{aligned} VP- = P[SMS^c|T-] &= \frac{P[T-|SMS^c] \times P[SMS^c]}{P[T-]} = \\ &= \frac{P[T-|SMS^c] \times P[SMS^c]}{P[T-|SMS^c] \times P[SMS^c] + P[T-|SMS] \times P[SMS]} = \\ &= \frac{\text{Especificidad} \times \text{1-Prevalencia}}{\text{Especificidad} \times \text{1-Prevalencia} + \text{1-Sensibilidad} \times \text{Prevalencia}} \end{aligned}$$

Estos cálculos se reañlizan de forma imediata usando R:

```r
sensi <- 0.68
espec <- 0.82
prev <- 0.0021
vp.pos <- (sensi * prev )/ (sensi * prev + (1-espec)* (1-prev))
cat ("El valor predictivo positivo es: ", vp.pos)
```

```
El valor predictivo positivo es:  0.007887324
```

```r
vp.neg <- (espec * (1-prev) )/ (espec * (1-prev) + (1-sensi)* (prev))
cat ("El valor predictivo negativo es: ", vp.neg)
```

```
El valor predictivo negativo es:  0.9991794
```

Como en el caso anterior, podemos ver que. al ser la prevalencia muy baja, el valor predicpositivo del test también lo es puesto que un test + tan solo indica en un 0,79% de veces la presencia del síndrome, correctamente.