

Test del SIDA

J. Abellán

30/10/2019

Se trata de un problema clásico, sencillo pero importantísimo.

Enunciado

En cierta población existe una enfermedad grave con una incidencia muy baja y un proceso de incubación muy lento.

Se ha desarrollado un *test* para detectarla precozmente pero el *test* no es perfecto. Concretamente tiene una **sensibilidad** del 95% y una **especificidad** del 90%.

El problema es el siguiente:

- Se escoge un individuo al azar de la población.
- Se le hace el *test* y da **positivo** para la enfermedad.
- ¿Cuál es la probabilidad de que tenga realmente la enfermedad?

La solución

La solución nos la da el teorema de **Bayes**:

$$p(S|+I) = p(S|I) \times \frac{p(+|SI)}{p(+|I)}$$

donde:

1. I es la información general sobre el problema.
2. S significa que el individuo tiene el Sida.
3. + significa que ha dado positivo en el *test*.

Por supuesto, $p(S|I)$ es la probabilidad de que el individuo elegido tenga el Sida (sin tener ninguna otra información). Podemos asignar a esta probabilidad la incidencia de la enfermedad en dicha población.

$p(+|SI)$ es la **sensibilidad** del *test*: la probabilidad de dar positivo teniendo la enfermedad.

$p(+|I)$ es la probabilidad de dar positivo. De acuerdo con la fórmula de la probabilidad total:

$$p(+|I) = p(S|I) \times p(+|SI) + p(\bar{S}|I) \times p(+|\bar{S}I)$$

donde \bar{S} significa que el individuo no tiene el Sida.

```
# incidencia de la enfermedad
prior <- pSida <- 1/1000

#lógicamente, la probabilidad de no tener la enfermedad será
pNoSida <- 1 - pSida

#Sensibilidad = probabilidad de dar positivo teniendo la enfermedad
pPos.Sida <- vero <- .96

#Especificidad = probabilidad de dar negativo no teniendo la enfermedad
```

```

pNeg.NoSida <- .9

#por tanto, la probabilidad de falsos positivos será
pPos.NoSida <- 1 - pNeg.NoSida

#Probabilidad de dar positivo
pPos <- pSida * pPos.Sida + pNoSida * pPos.NoSida

#Bayes
pSida.Pos <- pSida * pPos.Sida / pPos

post <- round( pSida.Pos, 3 )
cbind( prior, vero, post)

```

```

##      prior vero post
## [1,] 0.001 0.96 0.01

```

Es decir, la probabilidad de que el individuo (recuerde: elegido al azar) tenga SIDA, habiendo dado positivo en el *test*, es aproximadamente del 1 % y no del 96 %.

Siguiente pregunta

¿Y si le hacemos de nuevo el *test* y da positivo? ¿Cuál será ahora la probabilidad de que tenga SIDA?

La respuesta es bien simple: volvemos a aplicar el teorema de **Bayes** pero ahora el *prior* es el *posterior*:

```

#probabilidad de tener sida ahora habiendo dado positivo
pSida <- pSida.Pos

#la probabilidad de no tener sida habiendo dado positivo será
pNoSida <- 1 - pSida

#Probabilidad de dar positivo
pPos <- pSida * pPos.Sida + pNoSida * pPos.NoSida

#Bayes
pSida.Pos <- pSida * pPos.Sida / pPos

post2 <- round( pSida.Pos, 3 )
cbind( prior, vero, post, vero, post2)

```

```

##      prior vero post vero post2
## [1,] 0.001 0.96 0.01 0.96 0.084

```

Y así sucesivamente.

En resumen

Una gráfica vale por mil palabras.

Representaremos $p(S|+I)$ en función de $p(S|I)$

```

#Incidencia de la enfermedad: de muy escasa a generalizada
pSida <- seq( 0, 1, len = 100 )

```

```

#lógicamente, la probabilidad de no tener la enfermedad será
pNoSida <- 1 - pSida

#Probabilidad de dar positivo
pPos <- pSida * pPos.Sida + pNoSida * pPos.NoSida

#Bayes
pSida.Pos <- pSida * pPos.Sida / pPos

plot( pSida, pSida.Pos,
      type = "l",
      xlab = "p(S / ahora)",
      ylab = "p(S / +)",
      main = paste("p(+ / S) = ", pPos.Sida, ", p(+ / nS) = ",pPos.NoSida ) )

abline( h = 0, v = 0 )
abline( 0, 1 )

#Supongamos 4 test positivos
prior <- 1/1000

for (i in 1:4){
  #Probabilidad de dar positivo
  pPos <- prior * pPos.Sida + ( 1 - prior ) * pPos.NoSida

  #Bayes
  posterior <- prior * vero / pPos

  segments( prior, prior, prior, posterior, col = 2 )
  segments( prior, posterior, posterior, posterior, col = 2 )

  prior <- posterior
}

segments( 0, posterior, posterior, posterior, col = 2 )

```

$$p(+ / S) = 0.96, p(+ / nS) = 0.1$$

