Probabilidad condicional

Antonio Falcó

Seminario 3

Cuestión

¿Cómo podemos calcular en la práctica $\mathbb{P}(\Omega_1 \cap E)$ o $\mathbb{P}(E \cap F)$?

Respuesta

- Si conocemos $\mathbb{P}(\Omega_1)$ o $\mathbb{P}(F)$,
- Si conocemos $\mathbb{P}(E|\Omega_1)$ o $\mathbb{P}(E|F)$, donde

$$\mathbb{P}(E|\Omega_1) := \frac{\text{número de individuos enfermos en la muestra}}{\text{número de individuos de la muestra}}$$

y

$$\mathbb{P}(E|F) := \frac{\text{número de individuos enfermos que tienen el factor de riesgo}}{\text{número de individuos que tienen el factor de riesgo}}$$

Entonces

$$\mathbb{P}(\Omega_1 \cap E) = \mathbb{P}(E|\Omega_1)\mathbb{P}(\Omega_1) \text{ y } \mathbb{P}(E \cap F) = \mathbb{P}(E|F)\mathbb{P}(F).$$

Cuestión

Tenemos una población 100000 individuos donde sabemos que 1 de cada 10000 padece una enfermedad E. Además, sabemos que 5 de cada 10 enfermos que llegaron a un hospital presentaron un factor de riesgo F. ¿Cuántos individuos estan enfermos y presentan el facfor de riesgo aproximadamente?

Respuesta

Si **conocemos** (información a priori) que una persona está enferma, entonces tiene una probabilidad de $\frac{5}{10} = \frac{1}{2}$ de tener el factor de riesgo, luego

$$\mathbb{P}(F|E) = \frac{1}{2} \text{ (con la notación } F|E \text{ indicamos que conocemos que está enfermo).}$$

Además, la probabilidad de padecer la enfermedad es $\mathbb{P}(E) = \frac{1}{10000}$, entonces

$$\mathbb{P}(E \cap F) = \mathbb{P}(F|E)\mathbb{P}(E) = \frac{1}{2} \times \frac{1}{10000} = \frac{1}{20000}.$$

Si hay 100000 individuos en la población tenemos que unos $\frac{1}{20000} \times 100000 = 5$ indivíduos estaran enfermos y presentaran el factor de riesgo aproximadamente.

Definición (Probabilidad Condicional)

Sean

- Ω una población y $\mathbb{P}$ una probabilidad definida sobre la población.
- A es un subconjunto de individuos de la población que comparten una propiedad $\mathcal{P}$, (información a priori).

Entonces dado cualquier otro subconjunto B de individuos la población que comparten una propiedad $\mathcal{Q}$, se define la probabilidad de que un individuo que cumple la propiedad $\mathcal{P}$, y en consecuencia está en A, sea susceptible de cumplir la propiedad $\mathcal{Q}$, es decir que esté en B, como

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\text{número de individuos de } A \text{ que están a su vez en } B}{\text{número de individuos de } A}.$$

$\mathbb{P}(B|A)$ es la **probabilidad de que ocurra B condicionada a que ocurra A.**

Independencia probabilista

Sean

- Ω una población y $\mathbb{P}$ una probabilidad definida sobre la población.
- Los individuos A cumplen la propiedad $\mathcal{P}$,
- Los individuos B cumplen la propiedad $\mathcal{Q}$.

Se dice que las propiedades $\mathcal{P}$ y $\mathcal{Q}$ relativas a los conjuntos A y B son independientes si

$$\mathbb{P}(B|A) = \mathbb{P}(B),$$

es equivalente a decir que

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Cuestión

En un grupo de 25 pacientes 5 presentan vómitos, 10 padecen migrañas, y 2 de los pacientes presentan a su vez vómitos y migrañas. ¿Podemos afirmar que presentar vómitos es independiente de padecer migrañas?

Respuesta

Conocemos que

$$\begin{array}{rcl} A & = & \text{ser susceptible de presentar vómitos y} \\ B & = & \text{ser susceptible de padecer migrañas} \end{array}$$

entonces de los datos facilitados se tiene que

$$\mathbb{P}(A) = \frac{5}{25} = \frac{1}{5}, \mathbb{P}(B) = \frac{10}{25} = \frac{2}{5} \text{ y } \mathbb{P}(A \cap B) = \frac{2}{25},$$

como

$$\mathbb{P}(A \cap B) = \frac{2}{25} = \frac{1}{5} \times \frac{2}{5} = \mathbb{P}(A)\,\mathbb{P}(B),$$

podemos afirmar que en nuestra población de pacientes presentar vómitos es independiente de padecer migrañas, es decir $\mathbb{P}(A|B) = \mathbb{P}(A)$ y $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Cuestión

En un grupo de 25 pacientes 6 presentan vómitos, 10 padecen migrañas, y 2 de los pacientes presentan a su vez vómitos y migrañas. ¿Podemos afirmar que presentar vómitos es independiente de padecer migrañas?

Respuesta

Conocemos que

$$\begin{array}{lcl} A & = & \text{ser susceptible de presentar vómitos y} \\ B & = & \text{ser susceptible de padecer migrañas} \end{array}$$

entonces de los datos facilitados se tiene que

$$\mathbb{P}(A) = \frac{6}{25}, \mathbb{P}(B) = \frac{10}{25} = \frac{2}{5} \text{ y } \mathbb{P}(A \cap B) = \frac{2}{25},$$

como

$$\mathbb{P}(A \cap B) = \frac{2}{25} \neq \frac{12}{125} = \frac{6}{25} \times \frac{2}{5} = \mathbb{P}(A)\,\mathbb{P}(B),$$

podemos afirmar que en nuestra población de pacientes presentar vómitos no es independiente de padecer migrañas.

Conclusión

La noción de independencia probabilista es una noción que afecta a la relación numérica entre dos propiedades que caracterizan a los individuos de una población, y en consecuencia **no explica una relación causa-efecto entre estas dos propiedades.**

Cuestión científica

¿Qué clase de relación explica la noción de probabilidad condicional?

Teorema de Bayes

Teorema

Sea Ω una población y $\mathbb{P}$ una probabilidad definida sobre ella. Si F es un conjunto de individuos expuestos a una propiedad y E es un conjunto de individuos que comparten una propiedad de interés de la que deseamos calcular su probabilidad, entonces

$$\mathbb{P}(E) = \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|\overline{F})\mathbb{P}(\overline{F}),$$

o bien

$$\mathbb{P}(E) = \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|\overline{F})(1 - \mathbb{P}(F)).$$

Demostración

Sabemos que

$$E = (E \cap F) \cup (E \cap \overline{F}) \text{ y ambos conjuntos no tienen individuos en común.}$$

Entonces

$$\mathbb{P}(E) = \mathbb{P}(E \cap F) + \mathbb{P}(E \cap \overline{F}) = \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|\overline{F})\mathbb{P}(\overline{F}).$$

Riesgos absolutos y relativos

Supongamos que estudiamos los factores de riesgo de una enfermedad (o los factores de riesgo de fallecer una vez hemos contraido la enfermedad). Denotaremos por E **el conjunto de individuos susceptibles de padecer una enfermedad** y por D **el conjunto de individuos susceptibles de fallecer**. Entonces

$$\mathbb{P}(E) = \text{ probabilidad de contraer la enfermedad}$$

y

$$\mathbb{P}(D) = \text{ probabilidad de fallecer }.$$

- A priori deseamos conocer si la probabilidad de fallecer es más alta dentro de las personas que han contraido la enfermedad y se encuentran en el conjunto de individuos E, que si no la tienen y se encuentran en el grupo $\overline{E}$.

Riesgos absolutos y relativos

- Como los indivíduos susceptibles de fallecer pueden tener o no la enfermedad, entonces tenemos que distinguir entre los susceptibles de fallecer que padecen la enfermedad que denotamos por $D|E$ y los susceptibles de fallecer que no tienen la enfermedad que denotamos por $D|\overline{E}$.
- En particular,

$$\mathbb{P}(D|E) = \text{ probabilidad de de fallecer una vez contraída la enfermedad}$$

Riesgos absolutos y relativos

Como los indivíduos susceptibles de fallecer pueden tener o no la enfermedad, entonces tenemos que distinguir entre los susceptibles de fallecer que padecen la enfermedad que denotamos por $D|E$ y los susceptibles de fallecer que no tienen la enfermedad que denotamos por $D|\overline{E}$. En particular,

$$\mathbb{P}(D|E) = \begin{array}{l}\text{probabilidad de de fallecer conociendo} \\ \text{que ha contraído la enfermedad}\end{array}$$

y

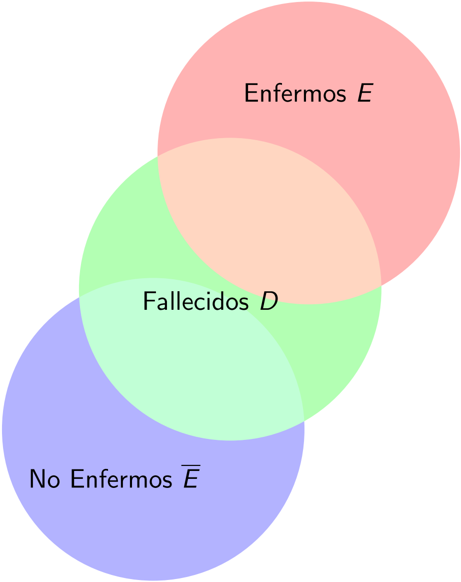
$$\mathbb{P}(D|\overline{E}) = \begin{array}{l}\text{probabilidad de de fallecer conociendo} \\ \text{que NO ha contraído la enfermedad.}\end{array}$$

Cuestión

¿Es metodológicamente correcto el comparar

$$\mathbb{P}(D|E) \text{ y } \mathbb{P}(D|\overline{E})$$

para determinar si el riesgo de fallecimiento dentro del grupo de personas que han contraído la enfermedad es superior al riesgo de fallecimiento dentro del grupo de personas que no han contraído la enfermedad?



Los fallecidos entre los enfermos $E \cap D$ no tienen indivíduos en común con los fallecidos entre los no enfermos $\overline{E} \cap D$: $(E \cap D) \cap (\overline{E} \cap D) = \emptyset$

Consecuencia

No se pueden comparar los fallecidos entre los enfermos y los fallecidos entre los no enfermos.

Ejemplo

Tenemos una población de 6 individuos donde hay 2 enfermos y 4 no enfermos, han fallecido 1 enfermo y 1 no enfermo. Entonces tenemos

$$\mathbb{P}(D|E) = \frac{1}{2} = 50\% \text{ y } \mathbb{P}(D|\overline{E}) = \frac{1}{4} = 25\%$$

que la probabilidad de fallecer estando enfermo es el doble de fallecer si no lo estás sin embargo

$$\mathbb{P}(E) = \frac{2}{6} = \frac{1}{3} \approx 33\% \text{ y } \mathbb{P}(\overline{E}) = \frac{4}{6} = \frac{2}{3} \approx 67\%,$$

la probabilidad de no estar enfermo es el doble que de estarlo. Si tenemos esto en cuenta

$$\mathbb{P}(D|E)\mathbb{P}(E) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6} \text{ y } \mathbb{P}(D|\overline{E})\mathbb{P}(\overline{E}) = \frac{1}{4} \cdot \frac{2}{3} = \frac{2}{12} = \frac{1}{6}.$$

Ejemplo

Tenemos una población de 6 individuos donde hay 2 enfermos y 4 no enfermos, han fallecido 1 enfermo y 1 no enfermo. Observemos que la probabilidad de fallecer en esta población es de

$$\mathbb{P}(D) = \frac{2}{6} = \frac{1}{3} = \frac{1}{6} + \frac{1}{6} = \mathbb{P}(D|E)\mathbb{P}(E) + \mathbb{P}(D|\overline{E})\mathbb{P}(\overline{E}),$$

que se reparte de igual manera entre los enfermos E que entre los no enfermos $\overline{E}$.

Definición (riesgo absoluto)

Llamamos riesgo absoluto a cualquiera de las probabilidades siguientes:

$$\mathbb{P}(D),\ \mathbb{P}(E),\ \mathbb{P}(D|E) \text{ y } \mathbb{P}(D|\overline{E}).$$

Definición (factor de riesgo)

Si F es el conjunto de individuos de una población susceptibles de tener un factor de riesgo para una enfermedad E entonces podemos definir los riesgos absolutos siguientes:

$$\mathbb{P}(E),\ \mathbb{P}(F),\ \mathbb{P}(E|F) \text{ y } \mathbb{P}(E|\overline{F}).$$

Ejemplo

E enfermos susceptibles de tener la gripe y F personas susceptibles de tener una temperatura o igual a 38°, o bien F personas susceptibles de estornudar de forma frecuente. Ambos casos pueden considerarse factores de riesgo para tener gripe.

Definición (Riesgo relativo)

Llamaremos **riesgo relativo** de contraer una determinada enfermedad E con respecto un factor de riesgo F al cociente:

$$RR = \frac{\mathbb{P}(E|F)}{\mathbb{P}(E|\overline{F})}.$$

Es la medida más empleada en epidemiología para medir con una única cifra la asociación entre la exposición a un factor F y una enfermedad E.

Discusión

- Si $RR >> 1$ entonces $\mathbb{P}(E|F) >> \mathbb{P}(E|\overline{F})$ entonces F se puede considerar un factor de riesgo de la enfermedad.
- Si $RR \leq 1$ entonces $\mathbb{P}(E|F) \leq \mathbb{P}(E|\overline{F})$ entonces F no se puede considerar un factor de riesgo de la enfermedad.

Definición (Riesgo atribuible)

El riesgo atribuible a un factor F es la proporción de casos que habríamos podido evitar si se hubiese evitado ese factor, ya que se le considera un factor en el desarrollo de la enfermedad E.

Ejemplo

Se conoce que el riesgo de cáncer de pulmón atribuible al factor del tabaco es del orden del 90% (esto quiere decir que si suprimiéramos de golpe el tabaco, alrededor del 10% de los casos de cáncer de pulmón seguirían ocurriendo).

Cálculo de riesgo atribuible (Proporción de casos evitables)

- Sea N el número total de individuos de una población Ω.
- Si se dan $N \times \mathbb{P}(E)$-casos de enfermos,
- no podemos evitar $N \times \mathbb{P}(E|\overline{F})$

La proporción máxima de casos **que podemos evitar** o riesgo atribuible RA será entonces

$$RA = \frac{\mathbb{P}(E) - \mathbb{P}(E|\overline{F})}{\mathbb{P}(E)}$$

Relación entre el riesgo relativo y el riesgo atribuible

Sea $\mathbb{P}(F)$ la proporción de sujetos **expuestos al factor de riesgo** de la población, entonces empleando el Teorema de Bayes podemos escribir:

$$\mathbb{P}(E) = \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|\overline{F})(1 - \mathbb{P}(F)),$$

recordemos que

$$RR = \frac{\mathbb{P}(E|F)}{\mathbb{P}(E|\overline{F})} \text{ y } RA = \frac{\mathbb{P}(E) - \mathbb{P}(E|\overline{F})}{\mathbb{P}(E)},$$

Entonces se puede deducir que

$$RA = \frac{\mathbb{P}(F)(RR - 1)}{\mathbb{P}(F)(RR - 1) + 1}.$$

Cuestión

En una población un individuo de cada 3 está expuesto a un factor F. Un individuo de cada diez desarrolla una enfermedad E. Dentro del conjunto de individuos que desarrollan la enfermedad E un 95% han estado expuestos al factor F. Calcular

1. Los riesgos absolutos y relativos,
2. la proporción de enfermos E dentro de conjunto de individuos expuestos la factor F que pueden ser atribuibles al factor F,
3. Si suprimiéramos el factor F, ¿cuál sería la proporción máxima de casos de enfermos que desaparecerían? y ¿bajo que hipótesis?

Respuesta

Los riesgos absolutos son

$$\mathbb{P}(F) = \frac{1}{3}, \quad \mathbb{P}(\overline{F}) = \frac{2}{3}, \quad \mathbb{P}(E) = \frac{1}{10}, \quad \mathbb{P}(\overline{E}) = \frac{9}{10}, \quad \mathbb{P}(F|E) = 0.95,$$

y $\mathbb{P}(\overline{F}|E) = 0.05$. Entonces

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(F|E)\mathbb{P}(E)}{\mathbb{P}(F)} = 0.285 \text{ y } \mathbb{P}(E|\overline{F}) = \frac{\mathbb{P}(\overline{F}|E)\mathbb{P}(E)}{\mathbb{P}(\overline{F})} = 0.0075$$

Respuesta

La proporción de enfermos E dentro de conjunto de individuos expuestos la factor F que pueden ser atribuibles al factor F,

$$\frac{\mathbb{P}(E|F) - \mathbb{P}(E|\overline{F})}{\mathbb{P}(E|F)} = \frac{0.285 - 0.0075}{0.285} = 0.9736842 = 97.3684211\%.$$