

Primeira Prova - FECD

April 2023

1. Suppose there is a rare disease that affects 1 in 1000 people, and there is a diagnostic test for the disease. The test results can be positive (indicating the presence of the disease), negative (indicating the absence of the disease), or inconclusive. The test is positive 99% of the time for people who have the disease. It is negative 95% of the time for people who do not have the disease, and inconclusive 3% of the time for each one of the two groups, for those who have the disease and for those who do not have the disease. If a person takes the test and gets a positive result, what is the probability that this person actually has the disease? (pode deixar as contas apenas indicadas).

Solução: Let A be the event that a person has the disease, B_1 be the event that the person tests positive for the disease, B_2 be the event that the person tests negative for the disease, and B_3 be the event that the person gets an inconclusive result. We want the probability that a person has the disease given that they get a positive result: $\mathbb{P}(A|B_1)$.

From the problem statement, we know that:

- $\mathbb{P}(A) = 0.001$
- $\mathbb{P}(B_1|A) = 0.99$
- $\mathbb{P}(B_2|A^c) = 0.95$
- $\mathbb{P}(B_3|A) = 0.03$
- $\mathbb{P}(B_3|A^c) = 0.03$

We use Bayes' rule to find $\mathbb{P}(A|B_1)$:

$$\mathbb{P}(A|B_1) = \frac{\mathbb{P}(B_1|A)\mathbb{P}(A)}{\mathbb{P}(B_1|A)\mathbb{P}(A) + \mathbb{P}(B_1|A^c)\mathbb{P}(A^c)} = \frac{0.99 * 0.001}{0.99 * 0.001 + \mathbb{P}(B_1|A^c)(1 - 0.001)}$$

We use that $\Omega = B_1 \cup B_2 \cup B_3$ and these three events are disjoint. Therefore,

$$1 = \mathbb{P}(\Omega|A^c) = \mathbb{P}(B_1 \cup B_2 \cup B_3|A^c) = \mathbb{P}(B_1|A^c) + \mathbb{P}(B_2|A^c) + \mathbb{P}(B_3|A^c)$$

and hence

$$\mathbb{P}(B_1|A^c) = 1 - \mathbb{P}(B_2|A^c) - \mathbb{P}(B_3|A^c) = 1 - 0.95 - 0.03 = 0.02$$

We can now conclude the calculation:

$$\mathbb{P}(A|B_1) = \frac{0.99 * 0.001}{0.99 * 0.001 + 0.02 * 0.999} = 0.047$$

Vocês notaram uma inconsistência nos dados. Devemos ter $1 = \mathbb{P}(B_1|A) + \mathbb{P}(B_2|A) + \mathbb{P}(B_3|A)$. Entretanto, a partir do que foi declarado no problem, temos

$$1 = \mathbb{P}(B_1|A) + \mathbb{P}(B_2|A) + \mathbb{P}(B_3|A) = 0.99 + \mathbb{P}(B_2|A) + 0.03 > 1.$$

Felizmente este erro não afeta o uso da regra de Bayes para obter a probabilidade solicitada.

-
2. Patients with Hodgkins lymphoma (a type of cancer of the lymphatic system) are classified according to their response to a certain treatment (3 possible answers) and their histological type (4 different types). The probability of falling into each of the 12 possible categories is given in the table below. Calculate the following probabilities: $\mathbb{P}(\text{Partial Resp})$, $\mathbb{P}(LP)$, $\mathbb{P}(LP|\text{Partial Resp})$, $\mathbb{P}(\text{Partial Resp}|LP)$ (pode deixar as contas apenas indicadas).

Histological type	Response		
	Positive	Partial	Unresponsive
LP	0.14	0.03	0.02
NS	0.13	0.03	0.02
MC	0.29	0.10	0.11
LD	0.03	0.02	0.08

Solução:

- $\mathbb{P}(\text{Partial Resp}) = \mathbb{P}(\text{Partial Resp} \cap LP) + \mathbb{P}(\text{Partial Resp} \cap NS) + \mathbb{P}(\text{Partial Resp} \cap MC) + \mathbb{P}(\text{Partial Resp} \cap LD) = 0.03 + 0.03 + 0.10 + 0.02 = 0.18.$
- $\mathbb{P}(LP) = 0.14 + 0.03 + 0.02 = 0.19$
- $\mathbb{P}(LP|\text{Partial Resp}) = \mathbb{P}(LP \cap \text{Partial Resp})/\mathbb{P}(\text{Partial Resp}) = 0.03/0.18 \approx 0.17.$
- $\mathbb{P}(\text{Partial Resp}|LP) = \mathbb{P}(LP \cap \text{Partial Resp})/\mathbb{P}(LP) = 0.03/0.19 \approx 0.16.$

Neste exemplo, tivemos $\mathbb{P}(\text{Partial Resp}|LP) \approx \mathbb{P}(LP|\text{Partial Resp})$ mas, em geral, as condicionais reversas diferem substancialmente.

3. Suppose we have a simple discrete random variable X that can take on the values 1, 2, 3, and 4 with the following probabilities: $\mathbb{P}(X = 1) = 0.4$, $\mathbb{P}(X = 2) = 0.3$, $\mathbb{P}(X = 3) = 0.2$, $\mathbb{P}(X = 4) = 0.1$. Calculate the expected value $\mathbb{E}(X)$ and the variance $\mathbb{V}(X)$. You don't have to finish the calculations, you can just leave them indicated.

Solução:

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^4 k\mathbb{P}(X = k) \\ &= 1 * \mathbb{P}(X = 1) + 2 * \mathbb{P}(X = 2) + 3 * \mathbb{P}(X = 3) + 4 * \mathbb{P}(X = 4) \\ &= 1 * 0.4 + 2 * 0.3 + 3 * 0.2 + 4 * 0.1 = 2\end{aligned}$$

The variance $\mathbb{V}(X)$ may be calculated in any of two ways:

$$\begin{aligned}\mathbb{V}(X) &= \sum_{k=1}^4 [k - \mathbb{E}(X)]^2 \mathbb{P}(X = k) \\ &= [1 - 2]^2 * 0.4 + [2 - 2]^2 * 0.3 + [3 - 2]^2 * 0.2 + [4 - 2]^2 * 0.1 \\ &= 1 * 0.4 + 0 * 0.3 + 1 * 0.2 + 4 * 0.1 = 1\end{aligned}$$

The other way is to first obtain

$$\mathbb{E}(X^2) = \sum_{k=1}^4 k^2 \mathbb{P}(X = k) = 1^2 * 0.4 + 2^2 * 0.3 + 3^2 * 0.2 + 4^2 * 0.1 = 5$$

and to calculate

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 5 - 2^2 = 1.$$

4. We have a simple continuous random variable X that can take on the values on the interval $[0, 1]$ with probability density function given by $f(x) = 2x$ in the interval $[0, 1]$ (and equal to zero elsewhere).

- Calculate $\mathbb{P}(X \in (0.0, 0.1))$.
- Obtain the cumulative probability density function $\mathbb{F}(x) = \mathbb{P}(X \leq x)$ for $x \in (-\infty, \infty)$.
- Also, obtain the expected value $\mathbb{E}(X)$.

Solução:

$$\mathbb{P}(X \in (0.0, 0.1)) = \int_0^{0.1} 2x \, dx = 2 \frac{x^2}{2} \Big|_{0.0}^{0.1} = 0.01$$

$$\mathbb{F}(x) = \begin{cases} 0, & \text{if } x < 0 \\ x^2, & \text{if } 0 < x < 1 \\ 1, & \text{if } x > 1 \end{cases}$$

$$\mathbb{E}(X) = \int_0^1 x \, 2x \, dx = \frac{2}{3} x^3 \Big|_0^1 = \frac{2}{3}$$

5. Um estudo observacional foi conduzido para investigar o impacto da atividade física na prevenção de doenças cardíacas (CHD). Um grande número de jovens saudáveis e sem CHD foi selecionado e acompanhado por muitos anos. O grau de atividade física de cada um deles foi monitorado, assim como o desenvolvimento de eventuais problemas cardíacos. Uma forte associação foi encontrada entre as duas variáveis: pessoas com menos atividade física eram mais propensas a desenvolver CHD. Por que essa associação não nos permite concluir decisivamente que a atividade física contribui para reduzir o risco de CHD?

Solução: Indivíduos que se exercitam mais provavelmente são sistematicamente diferentes daqueles que se exercitam menos, e os grupos provavelmente diferem em relação a outros fatores que afetam a probabilidade de desenvolvimento de CHD. As pessoas que se exercitam regularmente são mais propensas a se preocupar com a saúde e são mais propensas a comer uma dieta saudável, manter um peso saudável, verificar a pressão arterial regularmente e tomar vitaminas. Além disso, eles são menos propensos a fumar e menos propensos a ter diabetes ou hipertensão. Todas essas coisas os colocam em menor risco de CHD. Queremos medir o efeito direto da atividade física, independente de todos esses outros fatores de risco. Quando outros fatores de risco que afetam o resultado (neste caso, a ocorrência de CHD) são distribuídos de forma desigual entre nossos grupos de comparação, a medida de associação será tendenciosa como resultado da confusão.