Aula de Exercícios - Teorema de Bayes

Organização: Rafael Tovar *Digitação*: Guilherme Ludwig

Primeiro Exemplo - Estagiários

Três pessoas serão selecionadas aleatóriamente de um grupo de dez estagiários administrativos. Esses três formarão um comitê com três cargos diferentes: o primeiro será nomeado coordenador, o segundo fiscal e o terceiro secretário.

Metade do grupo são estudantes de último ano de graduação, sem nenhuma experiência dentro da empresa. Os outro cinco são estagiários há um semestre, e já concorrem por uma vaga efetiva na empresa.

Primeiro Exemplo - Estagiários

O espaço de configurações possíveis para a formação do comitê é:

$$H = \{NNN, NNA, NAN, ANN, NAA, ANA, AAN, AAA\}$$

Onde a ordem representa os cargos (coordenador, fiscal, secretário) e A indica um estagiário antigo, enquanto N um estagiário novo.

Defina o evento $A = \{$ O coordenador é um estagiário antigo $\}$, de modo que $A^c = \{$ O coordenador é um estagiário novo $\}$. Defina também os eventos B_0, B_1, B_2 e B_3, associados ao número de estagiários novos no comitê.

$$B_k = \{k \text{ estagiários novos no comitê}\}$$

Primeiro Exemplo - Estagiários

Para cada configuração, temos uma probabilidade associada:

Evento	Probabilidade
NNN	$5/10 * 4/9 * 3/8 = 3/36$
NNA	$5/10 * 4/9 * 5/8 = 5/36$
NAN	$5/10 * 5/9 * 4/8 = 5/36$
ANN	$5/10 * 5/9 * 4/8 = 5/36$
NAA	$5/10 * 5/9 * 4/8 = 5/36$
ANA	$5/10 * 5/9 * 4/8 = 5/36$
AAN	$5/10 * 4/9 * 5/8 = 5/36$
AAA	$5/10 * 4/9 * 3/8 = 3/36$

Primeiro Exemplo - Estagiários

Observando os pontos amostrais na tabela anterior (NNN, NNA, etc.), construimos uma tabela de distribuição conjunta de A e B.

	B_0	B_1	B_2	B_3	Total
A	$3/36$	$10/36$	$5/36$	0	$18/36$
A^c	0	$5/36$	$10/36$	$3/36$	$18/36$
Total	$3/36$	$15/36$	$15/36$	$3/36$	$18/36$

Primeiro Exemplo - Estagiários

(1) Qual é a probabilidade de que o comitê tenha pelo menos dois estagiários novos?
Resp.: O evento é $B_2 \cup B_3$. Note que é uma união disjunta, isto é, $B_2 \cap B_3 = \emptyset$. Então $P(B_2 \cup B_3) = P(B_2) + P(B_3) = 18/36$.

(2) Qual é a probabilidade de ter um coordenador antigo no comitê?
Resp.: $P(A^c) = 18/36 = 1/2$.

Primeiro Exemplo - Estagiários

(3) Qual é a probabilidade de ter dois estagiários novos no comitê, e um deles ser o coordenador?
Resp.: A conjunção “e” indica intersecção de eventos. No caso, $P(B_2 \cap A^c)$, que é a probabilidade conjunta, ou simplesmente $P(B_2 \cap A^c) = 10/36$.

(4) Qual é a probabilidade de ter pelo menos um estagiário novo no comitê e um deles ser o coordenador?
Resp.: $P(A^c \cap \{B_1 \cup B_2 \cup B_3\}) =$
$P(A^c \cap B_1) + P(A^c \cap B_2) + P(A^c \cap B_3) = 18/36$

Primeiro Exemplo - Estagiários

(5) Se sabemos que o coordenador é um estagiário antigo, qual é a probabilidade de que o comitê tenha dois estagiários novos?
Resp.: Queremos $P(B_2|A)$. Pela definição de probabilidade condicional, $P(B_2|A) = P(B_2 \cap A)/P(A) = \frac{5/36}{18/36} = 5/18$

(6) Se o comitê tem dois estagiários novos, qual é a probabilidade que o coordenador seja um estagiário antigo?
Resp.: Queremos agora $P(A|B_2)$. Novamente pela definição, $P(A|B_2) = P(B_2 \cap A)/P(B_2) = 5/15$

Primeiro Exemplo - Estagiários

(7) Se o comitê tem pelo menos dois estagiários novos, qual é a probabilidade de que o coordenador seja um estagiário antigo?
Resp.:
$P(A|\{B_2 \cup B_3\}) = \frac{P(A \cap \{B_2 \cup B_3\})}{P(B_2 \cup B_3)} = \frac{P(A \cap B_2) + P(A \cap B_3)}{P(B_2) + P(B_3)} = 5/18$

(8) Se o comitê tem pelo menos um estagiário novo, qual é a probabilidade de que o coordenador seja um esgatiário novo?
Resp.: De modo semelhante ao item anterior,
$P(A^c|\{B_1 \cup B_2 \cup B_3\}) = 18/33$

Primeiro Exemplo - Estagiários - Exercício

(9) São independentes os eventos A e B_i, $i = 0, 1, 2, 3$?
Lembre-se da definição de independência: os eventos G e H são independentes se, e somente se, $P(G \cap H) = P(G)P(H)$.

(10) Defina, a partir da tabela, dois eventos mutuamente exclusivos.

Segundo Exemplo - Controle de Qualidade

Uma companhia multinacional tem três fábricas que produzem o mesmo tipo de produto. A fábrica I é responsável por 30% do total produzido, a fábrica II produz 45% do total, e o restante vem da fábrica III. Cada uma das fábricas, no entanto, produz uma proporção de produtos que não atendem aos padrões estabelecidos pelas normas internacionais. Tais produtos são considerados "defeituosos" e correspondem a 1%, 2% e 1,5%, respectivamente, dos totais produzidos por fábrica.

No centro de distribuição, é feito o controle de qualidade da produção combinada das fábricas.

Segundo Exemplo - Controle de Qualidade

(1) Qual é a probabilidade de encontrar um produto defeituoso durante a inspeção de qualidade?
Resp.: Seja o evento $A = \{\text{Produto Defeituoso}\}$ e $F_i = \{\text{Produto da Fábrica } i\}$. Sabemos, pelo enunciado, que $P(F_1) = 0{,}3$, $P(F_2) = 0{,}45$ e $P(F_3) = 0{,}25$. Além disso, sabemos que $P(A|F_1) = 0{,}01$, $P(A|F_2) = 0{,}02$ e $P(A|F_3) = 0{,}015$.
Então, pela lei da probabilidade total,

$$P(A) = P(A|F_1)P(F_1) + P(A|F_2)P(F_2) + P(A|F_3)P(F_3) =$$

$$= 0{,}3 * 0{,}01 + 0{,}45 * 0{,}02 + 0{,}25 * 0{,}015 = 0{,}01575$$

Segundo Exemplo - Controle de Qualidade

(2) Se durante a inspeção, encontramos um produto defeituoso, qual é a probabilidade que ele tenha sido produzido na fábrica II?

Resp.: Aqui, aplicaremos o Teorema de Bayes usando o item anterior para encontrar $P(A)$:

$$P(F_2|A) = \frac{P(A|F_2)P(F_2)}{P(A)} = \frac{0{,}02 * 0{,}45}{0{,}01575} = 0{,}5714$$

Terceiro Exemplo - PSA com um padrão de ouro imperfeito

Considere novamente os dados sobre rastreamento de câncer de próstata. Recorde os eventos associados aos experimentos:

$$\begin{array}{lll} A = \text{DRE}+ & B = \text{PSA}+ & C = \text{Paciente tem câncer} \\ A^c = \text{DRE-} & B^c = \text{PSA-} & C^c = \text{Paciente não tem câncer} \end{array}$$

Terceiro Exemplo - PSA com um padrão de ouro imperfeito

Suponha que temos uma informação confiável da prevalência de câncer de próstata na população – através de dados do censo do Ministério da Saúde – ou seja, conhecemos $P(C)$.

Definimos também a sensibilidade do PSA (dada por $P(B|C)$), e a especificidade (dada por $P(B^c|C^c)$).

Defina $\text{VPP} = P(C|B)$ o *Valor Preditivo Positivo*, a proporção de indivíduos doentes dentre aqueles que tiveram um resultado positivo no teste. Analogamente, defina $\text{VPN} = P(C^c|B^c)$ o *Valor Preditivo Negativo*.

Terceiro Exemplo - PSA com um padrão de ouro imperfeito

Se um outro teste tem um padrão de ouro perfeito, isto é, se somos capazes de determinar com 100% de segurança se um indivíduo está doente ou não através dele, podemos determinar a especificidade, sensitividade, VPP e VPN do PSA usando esse outro teste como referência, tudo através da tabela 2x2 dos dados observados.

Terceiro Exemplo - PSA com um padrão de ouro imperfeito

Se por outro lado, não temos um padrão de ouro perfeito, apenas um teste convencional (ou standard) que é usado para diagnosticar a doença (mas que normalmente é mais caro ou de mais difícil acesso - como a biópsia ou métodos invasivos para o paciente), então devemos aplicar o Teorema de Bayes para obter os valores preditivos do teste novo.

$$P(C|B) = \frac{P(B,C)}{P(B)} = \frac{P(C)P(B|C)}{P(B)} =$$

$$= \frac{P(C)P(B|C)}{P(C)P(B|C) + P(C^c)P(B|C^c)}$$

Terceiro Exemplo - PSA com um padrão de ouro imperfeito

Neste caso, precisamos do valor da prevalência (probabilidade "a priori"), que pode ser tomado de uma fonte externa – Como o Ministério da Saúde. Suponha que encontramos como prevalência o valor de $0{,}097 = P(C) \Rightarrow P(C^c) = 0{,}903$

$$\text{Sensitividade} = P(B|C) = 481/1881 = 0{,}26$$

$$\text{Especificidade} = P(B^c|C^c) = 16699/17595 = 0{,}949$$

Probabilidade total: $P(B) = 0{,}097 * 0{,}26 + 0{,}903 * 0{,}051 = 0{,}071$

Terceiro Exemplo - PSA com um padrão de ouro imperfeito

Denotaremos o Valor Preditivo Positivo por VPP, e o Valor Preditivo Negativo por VPN, para o caso do câncer de próstata.

$$\text{VPP} = P(C|B) = \frac{0{,}02522}{0{,}071} = 0{,}3542$$

$$\text{VPN} = P(C^c|B^c) = \frac{P(C^c \cap B^c)}{P(B^c)} =$$

$$= \frac{P(C^c)P(B^c|C^c)}{P(C^c)P(B^c|C^c) + P(C)P(B^c|C)} = \frac{0{,}8724}{0{,}9446} = 0{,}9235$$

Terceiro Exemplo - Interpretando os resultados

- Na população de homens com mais de 50 anos, se espera que 9,7% deles tenham câncer de próstata.
- Sensitividade: Dos indivíduos doentes, 26% são positivos no teste.
- Especificidade: Dos indivíduos sãos, 94% são negativos no teste.
- VPP: Dos indivíduos com teste positivo, 35,4% são doentes.
- VPN: Dos indivíduos com teste negativo, 92,3% não são doentes.

Terceiro Exemplo - Combinando Diagnósticos

Podemos analisar a combinação de dois testes diagnósticos.
Seja D_x o evento "pelo menos algum teste é positivo", ou seja,

$$D_x = \begin{cases} \text{DRE+, PSA+} & = & A \cap B \\ \text{DRE+, PSA–} & = & A \cap B^c \\ \text{DRE–, PSA+} & = & A^c \cap B \end{cases}$$

$$D_x^c = \text{DRE–, PSA–} = A^c \cap B^c$$

Terceiro Exemplo - Combinando Diagnósticos

Nos interessa a probabilidade de câncer, dado D_x.

$$P(C|D_x) = \frac{P(C \cap D_x)}{P(D_x)} = \frac{P(C \cap D_x)}{P(C)P(D_x|C) + P(C^c)P(D_x|C^c)}$$

$$\Rightarrow P(D_x|C) = P(A \cap B|C) + P(A \cap B^c|C) + P(A^c \cap B|C) =$$

$$= \frac{189}{1881} + \frac{145}{1881} + \frac{292}{1881} = \frac{626}{1881} = 0{,}33$$

$$\Rightarrow P(D_x|C^c) = P(A \cap B|C^c) + P(A \cap B^c|C^c) + P(A^c \cap B|C^c) =$$

$$= \frac{141}{17595} + \frac{1002}{17595} + \frac{755}{17595} = \frac{1757}{17595} = 0{,}09985$$

Terceiro Exemplo - Combinando Diagnósticos

E portanto, o valor preditivo positivo será:

$$\text{VPP} = P(C|D_x) = \frac{0{,}097 * 0{,}33}{0{,}097 * 0{,}33 + 0{,}903 * 0{,}09985} = 0{,}2637$$

Fica como exercício verificar que o valor preditivo negativo será:

$$\text{VPN} = P(C^c|D_x^c) = 0{,}926$$