Material Teórico - Módulo Probabilidade Condicional

Probabilidade Condicional - Parte 2

Segundo Ano do Ensino Médio

Autor: Prof. Fabrício Siqueira Benevides
Revisor: Prof. Antonio Caminha M. Neto

1 Probabilidade Condicional (continuação)

Nesta aula veremos novos exemplos envolvendo o conceito de probabilidade condicional e independência de eventos, a fim de reforçar a compreensão dos mesmos. Como aplicação, veremos na Seção 4 como interpretar corretamente certos dados estatísticos para que não sejamos vítimas do *paradoxo do falso positivo*.

Exemplo 1. *Uma pessoa joga um par de dados não tendenciosos, um deles de cor azul e o outro de cor branca. Chamemos de A o valor obtido na face do dado azul e de B o valor obtido na face do dado branco. Qual a probabilidade de que:*

(a) A seja igual a 2?

(b) $A+B$ seja menor ou igual a 5?

(c) A seja 2, dado que $A+B$ é menor ou igual a 5?

(d) $A+B$ seja menor ou igual a 5, dado que A é igual a 2?

Solução. Tomamos aqui o espaço amostral que consiste dos 36 pares ordenados (A, B), com $A, B \in \{1, 2, \ldots, 6\}$ e todos equiprováveis.

(a) Dentre os 36 resultados possíveis, há 6 deles em que A é igual a 2, a saber: $(2, 1), (2, 2), \ldots, (2, 6)$. Logo, a probabilidade desejada é igual a $6/36 = 1/6$. Alternativamente, uma vez que os valores de A e B são independentes, também podemos argumentar que os 6 possíveis valores para A são equiprováveis e, portanto, a probabilidade de A ser qualquer um desses valores é igual a $1/6$.

(b) A tabela abaixo indica, para cada escolha de A e B, os possíveis valores de $A+B$ e marca (em vermelho) aqueles em que $A+B \leq 5$. Veja que há 10 casos em que isso acontece. Portanto, a probabilidade desejada é igual a $10/36$.

+		B					
		1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
A	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8		10	11	12

(c) Ao assumir que $A+B \leq 5$, estamos reduzindo o nosso espaço amostral para apenas 10 elementos. Dentre estes, há 3 em que $A = 2$, a saber: $(2, 1)$, $(2, 2)$ e $(2, 3)$. Logo a probabilidade desejada é $3/10$.

(d) Assumindo que $A = 2$, nosso espaço amostral é reduzido ao conjunto $\{(2, 1), (2, 2), \ldots, (2, 6)\}$, ou seja, um total de 6 possíveis resultados. Fazendo $A = 2$, queremos que $A + B = 2 + B \leq 5$. Logo, há três possíveis valores para B (quais sejam, $B = 1$, 2 ou 3), e a probabilidade desejada é igual a $3/6$. □

Observações. *Os itens (c) e (d) do exemplo anterior nos mostram mais uma vez que, dados dois eventos, a probabilidade do primeiro condicionado ao segundo pode ser diferente da probabilidade do segundo condicionado ao primeiro.*

Sobre o item (b), podemos notar que há apenas 11 valores possíveis para a soma $A+B$ (ela varia de 2 a 12) e, dentre estes, há 4 que são menores ou iguais a 5. Contudo, não podemos dizer que a probabilidade desejada em (b) seja igual a $4/11$, pois aqueles 11 valores possíveis não são equiprováveis.

Por fim, veja que as respostas ao exemplo seriam todas idênticas se, no lugar de dois dados de cores diferentes, tivéssemos dois dados idênticos (afinal, a cor dos dados não poderia afetar as probabilidades). O mesmo é válido ainda que tivéssemos um único dado, que fosse lançado duas vezes seguidas.

Problema 2. *Três dados iguais, honestos e com seis faces numeradas de 1 a 6, são lançados simultaneamente. Calcule a probabilidade de que a soma dos resultados de dois quaisquer deles seja igual ao resultado do terceiro dado.*

Terminamos esta seção com uma observação importante sobre o cálculo da probabilidade da interseção de mais de dois eventos. Lembre-se de que, para dois eventos A e B, temos $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A)$.

> Dados três eventos A, B e C de um espaço de probabilidade, podemos escrever:
>
> $$\Pr(A \cap B \cap C) = \\ = \Pr(A) \cdot \Pr(B \mid A) \cdot \Pr(C \mid A \cap B). \tag{1}$$

Para provar que a expressão (1) é válida quando $\Pr(A) \neq 0$ e $\Pr(A \cap B) \neq 0$, basta ver que ela equivale a:

$$\Pr(A \cap B \cap C) = \Pr(A) \cdot \frac{\Pr(A \cap B)}{\Pr(A)} \cdot \frac{\Pr(A \cap B \cap C)}{\Pr(A \cap B)},$$

e esta última igualdade é claramente verdadeira. Por outro lado, quando $\Pr(A) = 0$ ou $\Pr(A \cap B) = 0$, é possível checar que ambos os lados da equação (1) serão iguais a zero.

A expressão (1) é bastante intuitiva para calcular a probabilidade da interseção $A \cap B \cap C$, considerando o seguinte ponto de vista: primeiramente, A deve ocorrer, o que acontece com probabilidade $\Pr(A)$. Além disso, B também deve ocorrer, mas, já sabendo que A ocorre, a probabilidade disso também acontecer é $\Pr(B \mid A)$. C também deve ocorrer, mas, já sabendo que A e B ocorrem, a probabilidade disso também acontecer é $\Pr(C \mid A \cap B)$. Por fim, para que os três eventos aconteçam, tomamos o produto dessas três probabilidades.

Com o mesmo raciocínio, é possível generalizar a expressão (1) para uma quantidade maior de eventos, digamos $A_1, A_2, \ldots, A_k$:

$$\begin{aligned}&\Pr(A_1 \cap A_2 \cap \ldots \cap A_k) = \\ &\quad = \Pr(A_1) \cdot \Pr(A_2 \mid A_1) \cdot \Pr(A_3 \mid A_1 \cap A_2) \cdot \ldots \\ &\qquad \ldots \cdot \Pr(A_k \mid A_1 \cap \ldots \cap A_{k-1}).\end{aligned}$$

2 Outras aplicações

Exemplo 3 (UFRJ, adaptado). *Seja n um inteiro maior ou igual a 1. Um grupo de n homens e n mulheres serão dispostos em uma fila de forma aleatória. Seja p_n a probabilidade de que a primeira mulher da fila ocupe a segunda posição. Calcule p_n e determine a partir de qual valor de n tem-se $p_n \leq 11/40$.*

Solução. Para que a primeira mulher ocupe a segunda posição é preciso que os seguintes eventos aconteçam:
Evento A: "a primeira posição é ocupada por um homem".
Evento B: "a segunda posição é ocupada por uma mulher".
Temos então que $p_n = \Pr(A \cap B)$. Assim:

$$\begin{aligned}p_n &= \Pr(A \cap B) = \Pr(A)\Pr(B \mid A) \\ &= \frac{n}{2n} \cdot \frac{n}{2n-1} = \frac{n}{2(2n-1)}.\end{aligned}$$

Em palavras, $\frac{n}{2(2n-1)}$ é a probabilidade do "primeiro da fila ser um homem" multiplicada pela probabilidade de que "o segundo seja uma mulher dado que o primeiro é um homem".

Resta saber para que valores de n temos que $p_n \leq 11/40$. Basta então resolver a inequação

$$\frac{n}{2(2n-1)} \leq \frac{11}{40}.$$

Que equivale sucessivamente a:

$$\begin{aligned}40n &\leq 22(2n-1) \\ -4n &\leq -22 \\ n &\geq \frac{22}{4} = 5{,}5.\end{aligned}$$

Como n é um numero inteiro, para que p_n seja menor ou igual a $11/40$ devemos ter $n \geq 6$. ☐

Exemplo 4. *Pedro e José jogam um dado não tendencioso, com faces numeradas de 1 a 6. Se o resultado for 6, Pedro vence. Se o resultado for 1 ou 2, José vence. Em qualquer outro caso eles jogam novamente, até que haja um vencedor. A probabilidade de que esse vencedor seja Pedro é igual a:*

(a) $\frac{1}{2}$. | *(b)* $\frac{1}{3}$. | *(c)* $\frac{1}{4}$. | *(d)* $\frac{1}{5}$. | *(e)* $\frac{1}{6}$.

Solução. A probabilidade de que Pedro vença logo no primeiro lançamento é igual a $1/6$. Para que ele vença (exatamente) no segundo lançamento é necessário que nem ele nem José vençam no primeiro, o que acontece com probabilidade $3/6 = 1/2$, e que o resultado do segundo lançamento seja igual a 6, o que acontece com probabilidade $1/6$. Um vez que os resultados do primeiro e do segundo lançamento são independentes, a probabilidade de que Pedro vença no segundo lançamento é igual a $\frac{1}{6} \cdot \frac{1}{2}$.

De forma geral, para todo natural k, os k primeiros lançamentos são independentes. Assim, a probabilidade de que Pedro vença exatamente após k lançamentos é igual a $\frac{1}{6}\left(\frac{1}{2}\right)^{k-1}$ (Pedro obtém 6 no k-ésimo lançamento e, em cada um dos $k-1$ primeiros lançamentos, ambos Pedro e José obtêm 3, 4 ou 5).

Por fim, para que Pedro vença o jogo é necessário que exista um natural k tal que Pedro vence após exatamente k lançamentos (mas o valor de k pode ser qualquer número natural). Assim, a probabilidade de que ele vença é igual à soma (infinita):

$$S = \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{2} + \frac{1}{6} \cdot \left(\frac{1}{2}\right)^2 + \frac{1}{6} \cdot \left(\frac{1}{2}\right)^3 + \cdots.$$

Isto é uma progressão geométrica infinita, de razão $q = 1/2$ e primeiro termo $a_1 = 1/6$. Logo, sua soma é dada pela fórmula $S = \dfrac{a_1}{1-q}$. Temos então que

$$S = \frac{1/6}{1-1/2} = \frac{1/6}{1/2} = \frac{1}{3},$$

e a alternativa correta é o item (b). ☐

Exemplo 5 (VUNESP). *Uma pesquisa publicada pela revista Veja em 07/06/2006 sobre os hábitos alimentares dos brasileiros mostrou que, no almoço, aproximadamente 70% dos brasileiros comem carne bovina e que, no jantar, esse índice cai para 50%. Suponha que a probabilidade de uma pessoa comer carne bovina no jantar, dado que ela comeu carne bovina no almoço, seja 6/10. Calcule a probabilidade de a pessoa comer carne bovina no almoço ou no jantar.*

Solução. Sejam A o evento "comer carne bovina no almoço" e J o evento "comer carne bovina no jantar". O enunciado nos informa que

$$\Pr(A) = 70\%, \quad \Pr(J) = 50\%, \quad \Pr(J \mid A) = 6/10$$

e nos pede para encontrar o valor de $\Pr(A \cup J)$. Para tanto, lembre-se de que:

$$\Pr(A \cup J) = \Pr(A) + \Pr(J) - \Pr(A \cap J),$$

com

$$\Pr(A \cap J) = \Pr(A)\Pr(J \mid A) = \frac{70}{100} \cdot \frac{6}{10} = \frac{42}{100} = 42\%.$$

Logo,

$$\Pr(A \cup J) = 70\% + 50\% - 42\% = 78\%.$$

□

Exemplo 6 (UFPE). *Um economista apresenta propostas de trabalho às empresas X e Y, de modo que a probabilidade dele ser contratado pela empresa X é $0{,}61$, a dele ser contratado pela empresa Y é $0{,}53$, e a dele ser contratado pelas duas é $0{,}27$. Calcule a probabilidade de o economista não ser contratado por nenhuma das duas empresas.*

Solução. Na verdade, não precisamos usar probabilidade condicional para resolver este problema. A observação que faremos após resolvê-lo é o motivo pelo qual ele está aqui nesta aula.

Como de costume, representaremos por X o evento "ser aceito pela empresa X", e por $X^{\complement}$ o evento "não ser aceito pela empresa X". Defina Y e $Y^{\complement}$ de forma semelhante.

Não ser contratado por nenhuma das duas empresas corresponde ao evento $X^{\complement} \cap Y^{\complement}$. Usando a Lei de de Morgan, concluímos que isso é o mesmo que $(X \cup Y)^{\complement}$. Assim, queremos encontrar o valor de $\Pr\left((X \cup Y)^{\complement}\right)$.

O enunciado do problema nos diz que:

$$\Pr(X) = 0{,}61, \quad \Pr(Y) = 0{,}53, \quad \Pr(X \cap Y) = 0{,}27.$$

Sendo assim,

$$\Pr(X \cup Y) = 0{,}61 + 0{,}53 - 0{,}27 = 0{,}87$$

e, daí,

$$\Pr\left((X \cup Y)^{\complement}\right) = 1 - 0{,}87 = 0{,}13.$$

□

Observação. Sejam X e Y como na solução do exemplo anterior. Claramente, temos

$$\Pr(X^{\complement}) = 1 - 0{,}61 = 0{,}39,$$
$$\Pr(Y^{\complement}) = 1 - 0{,}53 = 0{,}47$$

e queremos calcular $\Pr(X^{\complement} \cap Y^{\complement})$. Nesse sentido, seria *errado* calcular

$$\begin{aligned}\Pr(X^{\complement} \cap Y^{\complement}) &= \Pr(X^{\complement})\Pr(Y^{\complement}) \\ &= (0{,}39) \cdot (0{,}47) = 0{,}18,\end{aligned}$$

uma vez que , neste problema, $X^{\complement}$ e $Y^{\complement}$ *não são* eventos independentes. De fato, se $X^{\complement}$ e $Y^{\complement}$ fossem independentes, então X e Y também seriam independentes; por sua vez, isso implicaria que deveríamos ter $\Pr(X \cap Y) = \Pr(X)\Pr(Y)$, o que não condiz com os dados do problema. □

3 Excludentes versus mutuamente independentes

Sejam A e B dois eventos com probabilidades não nulas em um mesmo espaço amostral. Suponha que A e B são disjuntos, o que quer dizer que $A \cap B = \emptyset$. Como sabemos, isso implica que $\Pr(A \cap B) = 0$ e $\Pr(A \cup B) = \Pr(A) + \Pr(B)$. É fácil checar que eles também irão satisfazer $\Pr(A \mid B) = 0 = \Pr(B \mid A)$. Por causa disso, tais eventos também são chamados de *excludentes* (ou ainda *exclusivos*, por alguns autores).

Não se deve confundir a noção de eventos excludentes com aquela de eventos independentes. Note que estes dois conceitos não poderiam estar mais distantes um do outro: se A e B são excludentes e não nulos, então não há como eles serem independentes, uma vez que $\Pr(A \cap B) = 0$ mas $\Pr(A)\Pr(B) \neq 0$.

Considere, agora, uma família de k eventos $\{A_1, \ldots, A_k\}$ de um mesmo espaço de probabilidade. Dizemos que eles são dois a dois excludentes se, para quaisquer $1 \leq i, j \leq k$, com $i \neq j$, tivermos $A_i \cap A_j = \emptyset$. Lembre-se de que, neste caso,

$$\Pr(A_1 \cup \ldots \cup A_k) = \Pr(A_1) + \cdots + \Pr(A_k).$$

A proposição seguinte generaliza este fato.

Proposição 7. *Sejam $A_1, \ldots, A_k$ eventos dois a dois excludentes e E um evento qualquer. Então, vale que:*

$$\Pr\Big(\bigcup_{i=1}^{k} A_i \mid E\Big) = \sum_{i=1}^{k} \Pr(A_i \mid E).$$

Prova (esboço). Façamos a prova no caso em que E é não nulo. Usando a definição de probabilidade condicional, juntamente com o fato de que

$$\Big(\bigcup_{i=1}^{k} A_i\Big) \cap E = \bigcup_{i=1}^{k} (A_i \cap E),$$

obtemos

$$\begin{aligned}\Pr\Big(\bigcup_{i=1}^{k} A_i \mid E\Big) &= \Pr\Big(\Big(\bigcup_{i=1}^{k} A_i\Big) \cap E\Big) / \Pr(E) \\ &= \Pr\Big(\bigcup_{i=1}^{k} (A_i \cap E)\Big) / \Pr(E).\end{aligned}$$

Mas, como $A_1 \cap E, \ldots, A_k \cap E$ também são dois a dois excludentes, segue daí que

$$\begin{aligned}\Pr\Big(\bigcup_{i=1}^{k} A_i \mid E\Big) &= \Big(\sum_{i=1}^{k} \Pr(A_i \cap E)\Big) / \Pr(E) \\ &= \sum_{i=1}^{k} \Pr(A_i \cap E) / \Pr(E) \\ &= \sum_{i=1}^{k} \Pr(A_i | E).\end{aligned}$$

□

O seguinte problema é um exemplo folclórico de exercício sobre probabilidade condicional.

Exemplo 8. *Você entrega a seu amigo uma carta, destinada à sua namorada, para ser colocada no correio. Entretanto, ele pode esquecer-se de enviá-la com probabilidade* 0,1. *Se ele não esquecer, a probabilidade de que o correio extravie a carta é* 0,1. *Finalmente, se a carta for entregue corretamente, a probabilidade de que ela seja interceptada pelo pai ciumento dela é de* 0,1. *Caso não seja interceptada, a carta é, por fim, recebida por sua namorada. Pergunta-se:*

(a) Qual a probabilidade de sua namorada receber a carta?

(b) Dado que sua namorada não recebeu a carta, qual a probabilidade de seu amigo ter esquecido de enviá-la?

Solução. Vamos começar dando nomes para os eventos:
A = "O amigo envia a carta",
B = "O correio não extravia a carta",
C = "O pai não intercepta a carta",
D = "A namorada recebe a carta".

O item (a) nos pede para calcular $\Pr(D) = \Pr(A \cap B \cap C)$, o que pode ser feito usando a equação (1):

$$\Pr(A \cap B \cap C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid A \cap B).$$

Pelos dados do problema, temos que:

$$\Pr(A^{\complement}) = 0{,}1, \quad \Pr(B^{\complement} \mid A) = 0{,}1 \;\text{ e }\; \Pr(C^{\complement} \mid A \cap B) = 0{,}1.$$

Sendo assim, $\Pr(A) = 1 - \Pr(A^{\complement}) = 0{,}9$. Também, a Proposição 7 nos garante que:

$$\Pr(B \mid A) + \Pr(B^{\complement} \mid A) = \Pr\left((B \cup B^{\complement}) \mid A\right) = 1,$$

de sorte que

$$\Pr(B \mid A) = 1 - \Pr(B^{\complement} \mid A) = 1 - 0{,}1 = 0{,}9.$$

Analogamente,

$$\Pr(C \mid A \cap B) = 1 - \Pr(C^{\complement} \mid A \cap B) = 1 - 0{,}1 = 0{,}9.$$

Podemos, então, concluir que:

$$\begin{aligned}\Pr(A \cap B \cap C) &= \Pr(A)\Pr(B \mid A)\Pr(C \mid A \cap B)\\ &= 0{,}9 \cdot 0{,}9 \cdot 0{,}9\\ &= 0{,}729\\ &= 72{,}9\%,\end{aligned}$$

que é a probabilidade da namorada receber a carta.

O item (b) nos pede para calcular $\Pr(A^{\complement} \mid D^{\complement})$. Sabemos que

$$\Pr(A^{\complement} \mid D^{\complement}) = \frac{\Pr(A^{\complement} \cap D^{\complement})}{\Pr(D^{\complement})} = \frac{\Pr(A^{\complement})}{\Pr(D^{\complement})}.$$

(A segunda igualdade acima segue de que $A^{\complement} \subset C^{\complement}$, pois sempre que o amigo não envia a carta a namora não a recebe.) Pelo item (a), temos que

$$\Pr(D^{\complement}) = 1 - \Pr(D) = 1 - 0{,}729 = 0{,}271.$$

Sendo assim,

$$\Pr(A^{\complement} \mid D^{\complement}) = \frac{0{,}1}{0{,}271} \approx 0{,}369.$$

Portanto, sabendo que a namorada não recebeu a carta, a probabilidade de que o amigo não a tenha colocado no correio é de aproximadamente 36,9%. □

Agora, considere uma outra família $\mathcal{F}$ de k eventos $\{B_1, \ldots, B_k\}$, todos em um mesmo espaço de probabilidade. Por definição, dizer que os eventos $B_1, \ldots, B_k$ são dois a dois independentes é o mesmo que dizer que, para quaisquer $1 \le i < j \le k$, os eventos B_i e B_j são independentes.

Um outro conceito relacionado a esse, e que costuma ser confundido com ele, é a noção de *eventos mutuamente independentes*. A exigência que deve ser cumprida para dizermos que os eventos de família $\mathcal{F}$ como acima sejam *mutuamente independentes* é bem maior: para todo $1 \le i \le k$, o evento B_i deve ser independente de todo evento Z que pode ser obtido como a interseção de eventos B_j com $j \ne i$.

Exemplo 9. *Eventos $B_1, \ldots, B_k$ de um mesmo espaço de probabilidade podem ser dois a dois independentes mas não serem mutuamente independentes. Por exemplo, considere o experimento de lançar uma moeda honesta três vezes, e os seguintes eventos:*
B_1 = *"obter cara uma quantidade par de vezes";*
B_2 = *"obter os 2 primeiros lançamentos iguais";*
B_3 = *"obter cara nos últimos 2 lançamentos".*
Encorajamos o leitor a mostrar que B_1, B_2 e B_3 são dois a dois independentes mas não são mutuamente independentes, uma vez que B_1 não é independente de $B_2 \cap B_3$.

Quando os eventos $B_1, \ldots, B_k$ são mutuamente independentes, vale que

$$\Pr(B_1 \cap \ldots \cap B_k) = \Pr(B_1) \ldots \Pr(B_k).$$

4 O paradoxo do falso positivo

Este paradoxo acontece em diversos contextos e pode enganar mesmo a pessoas experientes no assunto. Veremos aqui um exemplo aplicado às ciências médicas. Parte do que é exposto aqui foi inspirado em artigos da Wikipedia.

A fim de identificar pessoas que possuem alguma doença grave em um estágio inicial, quando ainda é possível curá-las, uma ideia comum é submeter um grande número de pessoas a um *teste de triagem*. Em princípio, é claro que isso traz benefícios óbvios. Contudo, testes de triagem usualmente produzem resultados que não são cem por cento confiáveis, já que eles são baratos e, como o nome diz, servem apenas para fazer um triagem, selecionando as pessoas que deverão seguir para uma avaliação mais cuidadosa. Ainda assim, um teste como esse pode afetar negativamente as vidas das pessoas que foram diagnosticadas incorretamente, tanto emocionalmente como em relação aos custos para fazer novos exames. É, pois, interessante entender como calcular a chance de receber um resultado positivo falso.

Dentro de um certo grupo de pessoas onde o teste foi realizado, seja D o conjunto daquelas que, de fato, possuem a doença específica e $S = D^{\complement}$ o conjunto daquelas que não a possuem (que, por simplicidade, chamaremos de *saudáveis*). Seja P o conjunto das pessoas para as quais o teste resulta em *positivo* e $N = P^{\complement}$ o conjunto delas em que o teste resulta em *negativo*. Um teste é ideal quando $P = D$ e $N = S$. Entretanto, na prática é possível que algumas pessoas obtenham resultado positivo mesmo que não estejam doentes, e por essa razão chamaremos tais resultados de *falsos positivos*. Da mesma forma, também é possível que existam pessoas que obtiveram resultado negativo mas estão, de fato, doentes; chamaremos tais resultados de *falsos negativos*.

Algumas estatísticas são comumente utilizadas para saber se um teste é bom. A *sensibilidade* é definida como a probabilidade do teste ser positivo, dado que a pessoa está doente, ou seja, $\Pr(P \mid D)$. Por exemplo, um teste com sensibilidade de 80% que dizer que 80% das pessoas que de fato estão doentes vão receber o resultado correto (positivo). Já a *especificidade* é definida como a probabilidade de que o teste seja negativo dado que a pessoa está saudável, ou seja, $\Pr(N \mid S)$. Por exemplo, uma especificidade de 95% significa que 95% das pessoas saudáveis irão, de fato, receber um resultado correto (negativo). A *acurácia* é comumente definida como a probabilidade de que teste obtenha um resultado correto, ou seja, $\Pr((D \cap P) \cup (S \cap N))$. Contudo, ela nem sempre é um boa maneira de saber se um teste é bom. Por fim, a *prevalência* de uma doença é definida simplesmente como $\Pr(D)$, ou seja, a quantidade de pessoas doentes dividida pelo total da população. Veja que a prevalência não tem relação direta com o teste em si.

Suponha que um determinado teste possui sensibilidade $\Pr(P \mid D) = 100\%$ (o que é bastante raro) e possui especificidade $\Pr(N \mid S) = 95\%$. Isso também quer dizer que a taxa de falsos positivos (entre pessoas que são saudáveis) é de 5%. Para facilitar, vamos considerar um universo de 1.000 pessoas. A pergunta que queremos responder é: dado que o resultado do teste foi positivo, qual a chance de que a pessoa esteja realmente doente? De outra forma, qual o valor de $\Pr(D \mid P)$? As subseções seguintes analisam esse mesmo teste aplicado a populações onde a prevalência da doença é diferente.

4.1 Grupo A: população com alta prevalência

Suponha primeiro que a prevalência seja de 40%, ou seja, que o número de pessoas doentes é $40\% \cdot 1000 = 400$. Assim, o número de pessoas saudáveis é $1.000 - 400 = 600$.

Como a sensibilidade é 100%, todas as pessoas doentes serão corretamente diagnosticadas. Por outro lado, dentre as pessoas saudáveis, $95\% \cdot 600 = 570$ dos testes indicarão resultados negativos (corretos), mas as demais 30 pessoas receberão falsos positivos. A tabela abaixo resume esses dados.

	Doentes (D)	Saudáveis (S)	Total
Positivos (P)	400	30	430
Negativos (N)	0	570	570
Total	400	600	1000

A probabilidade que queremos calcular é:

$$\Pr(D \mid P) = \frac{400}{30 + 400} \cong 93\%.$$

Isso quer dizer que 93% das pessoas que receberam um resultado positivo realmente estão doentes.

4.2 Grupo B: população com baixa prevalência

Suponha agora que a prevalência seja de apenas 2%, ou seja, que apenas $2\% \cdot 1000 = 20$ das pessoas estão realmente doentes. O número de pessoas saudáveis será 980, dentre as quais $95\% \cdot 980 = 931$ receberão resultado negativo (correto), mas $980 - 931 = 49$ receberão falsos positivos. A tabela abaixo indica esses valores.

	Doentes (D)	Saudáveis ($D^{\complement}$)	Total
Positivo (P)	20	49	69
Negativo (N)	0	931	931
Total	20	980	1000

Nesta população, apenas 20 de um total de 69 pessoas cujo resultado do teste foi positivo estão realmente doentes. Assim, $\Pr(D \mid P) = \frac{20}{69} \cong 29\%$. De outra forma, se um paciente recebeu o resultado positivo, suas chances de realmente ter a doença são de apenas 29%. Isso é um resultado ruim para um teste que falaciosamente "parecia funcionar" em 95% dos casos.

4.3 Conclusões

Primeiramente, mais uma vez ressaltamos o fato de que não se pode confundir $\Pr(P \mid D)$ com $\Pr(D \mid P)$. Lembre-se de que essas duas quantidades estão relacionadas pelo

Teorema de Bayes:

$$\Pr(D \mid P) = \frac{\Pr(D)\Pr(P \mid D)}{\Pr(P)}.$$

Um médico que passou anos trabalhando com uma população do Grupo A, onde um resultado positivo indicava uma grande chance do paciente realmente ter a doença, poderia ficar confuso ao trabalhar com uma população do Grupo B, no qual 71% dos resultados positivos estão incorretos. O que dificulta ainda mais é o fato de que a prevalência de uma doença pode mudar ao longo das estações, em períodos de epidemias, ou mesmo com a faixa etária dos pacientes. Por isso é importante considerar a prevalência da doença e interpretar os resultados do teste com cautela. Veja que, mesmo nesse exemplo, ainda estamos considerando o caso em que a sensibilidade do teste é de 100%. Se isto não ocorrer os resultados serão ainda piores, pois existirão também falsos negativos.

Por outro lado, devemos considerar o seguinte: mesmo no Grupo B, em que o teste é menos eficaz, antes de realizar o teste, apenas por fazer parte deste grupo, o paciente tinha uma chance (a priori) de apenas 2% de estar doente (enquanto que no Grupo A tal chance é de 40%). Ao receber um resultado positivo, no Grupo B, tal chance aumenta de 2% para 29% (a posteriori), o que não deve ser desprezado.

Resumidamente, para que um teste seja realmente eficaz, é preciso que quanto menor for a prevalência da doença, maiores sejam a especificidade e a sensibilidade do teste.

Problema 10. *Encontre a acurácia do teste em cada uma das populações acima.*

Dicas para o Professor

Apesar desta aula trazer alguns temas novos, ela basicamente reforça o conteúdo da aula anterior. De fato, incluímos novos exemplos apenas para solidificar o que já foi visto. O material aqui colecionado pressupõe somente a compreensão das definições básicas da aula anterior, não exigindo que o leitor tenha percorrido todos os exemplos da mesma. Em verdade, caso seja conveniente, é possível tratar os conteúdos destes dois materiais em paralelo, seguindo o ritmo da turma e cobrindo mais cuidadosamente os aspectos em relação aos quais os alunos apresentarem maiores dificuldades.

Sugestões de Leitura Complementar

1. T. Apostol. *Cálculo, Volume 2.* Reverté, Porto, 1993.

2. P. C. P. Carvalho, P. Fernandez, A. C. de O. Morgado, J. B. Pitombeira. *Análise Combinatória e Probabilidade.* SBM, Rio de Janeiro, 2000.

3. J. P. O. Santos, M. P. Mello e I. T. Murari. Introdução à Análise Combinatória. Rio de Janeiro, Ciência Moderna, 2007.