

Roteiro de aulas - Introdução à Teoria de Probabilidade

Disciplina: Probabilidade e Estatística

Professor: André Oliveira

29 de setembro de 2025



Probabilidade

O Estudo de probabilidade é anterior ao século XVI. Nesta época, as aplicações eram dirigidas aos jogos de azar. Pessoas se utilizavam do conhecimento de teoria das probabilidades para planejar estratégias de apostas. No entanto, somente no século XX, é que o cálculo de probabilidade teve um desenvolvimento bastante grande e baseado numa teoria matemática através de axiomas rigorosos, definições e teoremas. A probabilidade é o ramo da Matemática com grande aplicação na estatística. Os tópicos abaixo são importantes em relação ao cálculo e fundamentação da teoria de probabilidades.

- Teoria de conjuntos;
- Métodos de contagem;
- Análise combinatória.

Até o presente momento estudamos os fenômenos de forma empírica, isto é, uma justificativa científica (só descrevendo, mas não explicando), o comportamento dos fenômenos através das distribuições freqüências. Temos a partir de agora o interesse em experiências com componente **aleatório**, causais, ou seja, experiências das quais **não podemos prever** os resultados a priori, mesmo que repetido sob as mesmas condições. Por exemplo:

- prever exatamente em quanto tempo um *SSDs* vai apresentar o primeiro erro de leitura;

- prever exatamente quantos *Gigabyte* (GB) de memória RAM terá o *notebook* que chegará para manutenção;
- prever qual a velocidade disponível em dado instante na sua rede wi-fi.

Experimento aleatório

São experimentos que, quando repetido em condições similares idênticas, dão resultado geralmente distintos, ou seja, não é possível predizer o resultado que será obtido.

Experimento determinístico

São experimentos que, quando repetido em condições idênticas, retornam resultado sempre iguais. Não são objetos de estudos na estatística.

Exemplos

- 1) $2\text{H}_2 + \text{O}_2 \longrightarrow 2\text{H}_2\text{O}$: Sempre que este experimento é repetido nas **mesmas condições** o resultado é o mesmo (formação de água). Exemplo clássico de **evento determinístico**.
- 2) Um programa de computador que calcula $2 + 2$: Sempre que este experimento é repetido nas **mesmas condições** o resultado é o mesmo ($2 + 2 = 4$). Exemplo clássico de **evento determinístico**.
- 3) Escolhe-se uma cor do conjunto de cores $S = \{\text{preto}, \text{rosa}, \text{azul}, \text{vermelho}\}$. Sempre que o experimento é repetido **nas mesmas condições**, o resultado pode variar. Exemplo clássico de **evento aleatório**.

Com a construção das suas distribuições de frequências (conhecimento empírico) e com surgimento da teoria da **probabilidade**, foi possível criar modelos probabilísticos (distribuição de probabilidade) que representam **probabilisticamente** muitos fenômenos. Estes modelos probabilísticos são considerados hoje a espinha dorsal da **Estatística**.

Probabilidade

São frequências relativas associadas a uma variável escolhida como descritora de uma característica de uma população. São utilizadas para mensurar a ocorrência de determinado fenômeno aleatório. A probabilidade é uma medida numérica que quantifica a possibilidade de ocorrência de um evento dentro de um espaço amostral.

Chance

Embora a palavra **chance** seja frequentemente utilizada no cotidiano como sinônimo de **probabilidade**, seu conceito formal é **distinto**. A chance de ocorrência de um evento é uma medida que expressa a razão entre a probabilidade de o evento ocorrer e a probabilidade de ele não ocorrer. Seja A um evento com probabilidade $P(A)$. Assim a chance do evento A ocorrer é definida como a comparação entre duas probabilidades.

$$Chance(A) = \frac{P(A)}{P(\bar{A})}$$

Variável aleatória

Variável escolhida como descritora de uma característica de uma população, cujos valores estão associadas a uma probabilidade de ocorrência.

Distribuição de probabilidade

Uma distribuição de probabilidade descreve como as probabilidades se distribuem entre todos os possíveis resultados (pontos) de um experimento aleatório.

Conceitos teóricos de probabilidade

Espaço amostral (Ω)

É o conjunto de todos os resultados possíveis de um experimento. Um espaço amostral pode ser finito ou infinito. Cada um dos elementos de um espaço amostral chama-se ponto amostral.

Evento (A)

É qualquer subconjunto (parte) de um espaço amostral.

Definição

A definição clássica de probabilidade se refere a subconjuntos unitários equiprováveis. No caso enumerável finito.

$$P(A) = \frac{n(A)}{n(\Omega)}$$

Em que $P(A)$ é probabilidade de ocorrência do evento A , $n(A)$ é o número de casos favoráveis a ocorrência do evento A e $n(\Omega)$ é o número de casos possíveis de elementos do espaço amostral.

A definição acima tem apelo da intuição e permanecem sendo usadas para resolver inúmeros problemas. Entretanto, elas não são suficientes para uma formulação matemática mais rigorosa de probabilidade. Sendo assim, Andrei Nikolaevich Kolmogorov, matemático Soviético/Russo 1903 -1987, apresentou um conjunto de axiomas matemáticos para definir probabilidade, permitido incluir as definições anteriores como casos particulares.

- 1) $P(\Omega=1)$
- 2) Se A é um evento pertencente a Ω , então $P(A) \geq 0$.
- 3) Sejam a partição de Ω em $A_1, A_2, A_3, \dots, A_n$ de forma que seja disjuntos, ou seja, $A_i \cap A_j = \emptyset$, para todo $i \neq j$.

Assim $P(A_1 \cup A_2 \cup A_3 + \dots + A_n) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n)$.

São três axiomas que não definem probabilidade, mas sim a definição de como a probabilidade está estruturada. Assim a partir dos axiomas anteriores é possível derivar as propriedades.

- 1) $P(\bar{A}) = 1 - P(A)$
- 2) $P(\emptyset) = 0$
- 3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- 4) $0 \leq P(A) \leq 1$
- 5) Se A e B são eventos independentes, então, $P(A \cap B) = P(A) \cdot P(B)$
- 6) Se A e B são eventos dependentes, então, $P(A \cap B) = P(A) \cdot P(B|A)$ ou seja,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Exemplo resolvidos

- 1) Em um lote, há 3 (três) SSDs SATA de 480 Gigabyte (GB) e 2 (dois) SSDs SATA de 500 *Gigabyte* (GB), todos indistinguíveis visualmente para quem recebeu a encomenda. Retira-se, aleatoriamente, uma **amostra** de **um SSD** deste lote e observa-se a sua capacidade. Pergunta-se: qual a **probabilidade** de que o SSD retirado tenha capacidade de 480 GB?

$$\Omega = \{\text{SSD01}, \text{SSD02}, \text{SSD03}, \text{SSD04}, \text{SSD05}\}$$

- $A=\{\text{SSD01}, \text{SSD02}, \text{SSD03}\} \rightarrow 480 \text{ GB}; \quad n(A) = 3$
- $B=\{\text{SSD04}, \text{SSD05}\} \rightarrow 500 \text{ GB}; \quad n(B) = 2$

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{3}{5} = \frac{60}{100} = 60,00\%$$

2) Em um lote, há 3 (três) SSDs SATA de 480 Gigabyte (GB) e 2 (dois) SSDs SATA de 500 *Gigabyte* (GB), todos indistinguíveis visualmente para quem recebeu a encomenda. Retira-se, aleatoriamente, uma **amostra** de **dois SSDs** **sem reposição** deste lote e observa-se a sua capacidade.

- a) Qual é a **probabilidade** de que os dois *SSDs* retirados tenham capacidade de 500 GB?

Definindo o espaço amostral como Ω e o evento $S = \{\text{Ser sorteado dois SSDs de 500 GB}\}$. Os pontos amostrais são.

$$\Omega = \left\{ \begin{array}{ll} (\text{SSD4801}, \text{SSD4802}) & (\text{SSD4801}, \text{SSD4803}) \\ (\text{SSD4801}, \text{SSD5001}) & (\text{SSD4801}, \text{SSD5002}) \\ (\text{SSD4802}, \text{SSD4801}) & (\text{SSD4802}, \text{SSD4803}) \\ (\text{SSD4802}, \text{SSD5001}) & (\text{SSD4802}, \text{SSD5002}) \\ (\text{SSD4803}, \text{SSD4801}) & (\text{SSD4803}, \text{SSD4802}) \\ (\text{SSD4803}, \text{SSD5001}) & (\text{SSD4803}, \text{SSD5002}) \\ (\text{SSD5001}, \text{SSD4801}) & (\text{SSD5001}, \text{SSD4802}) \\ (\text{SSD5001}, \text{SSD4803}) & (\text{SSD5001}, \text{SSD5002}) \\ (\text{SSD5002}, \text{SSD4801}) & (\text{SSD5002}, \text{SSD4802}) \\ (\text{SSD5002}, \text{SSD4803}) & (\text{SSD5002}, \text{SSD5001}) \end{array} \right\}, \text{ assim } n(\Omega) = 20 \text{ e } n(S) = 2$$

O número de elementos do espaço amostral pode ser determinado pelo **princípio multiplicativo** (PM) sem a necessidade de descrever todos os pontos.

$$\bullet \quad \underbrace{5}_{(\text{cinco SSDs para a primeira escolha})} \quad . \quad \underbrace{4}_{(\text{quatro SSDs para a segunda escolha})} = 5 \times 4 = 20$$

$$n(\Omega) = 20$$

$$\bullet \quad \underbrace{2}_{(\text{dois SSDs para a primeira escolha})} \quad . \quad \underbrace{1}_{(\text{um SSD para a segunda escolha})} = 2$$

$$n(S) = 2$$

$$\text{Assim } P(S) = \frac{n(S)}{n(\Omega)} = \frac{2}{20} = \frac{1}{10} = \frac{1}{10} = 10,00\%$$

Outra forma de desenvolver o cálculo da probabilidade solicitada é por meio da árvore abaixo.

Observando a árvore anterior e usando o conceito de probabilidade condicional

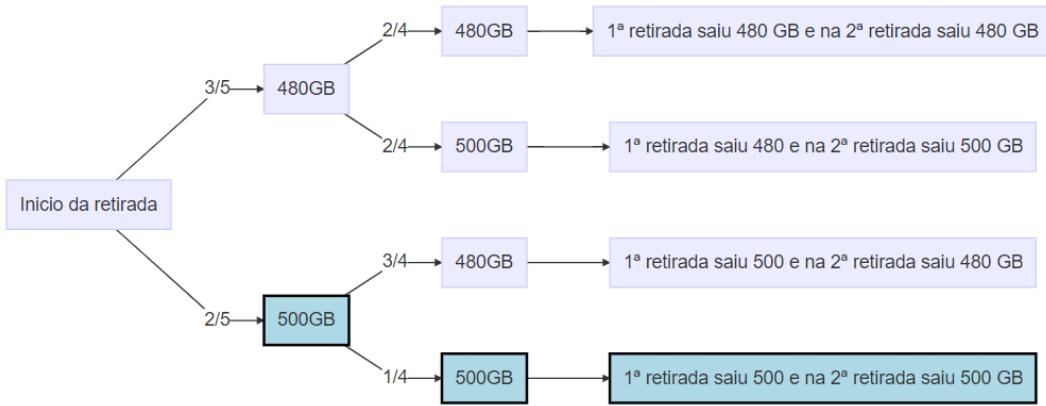


Figura 1: Diagrama em forma de árvore de possibilidades

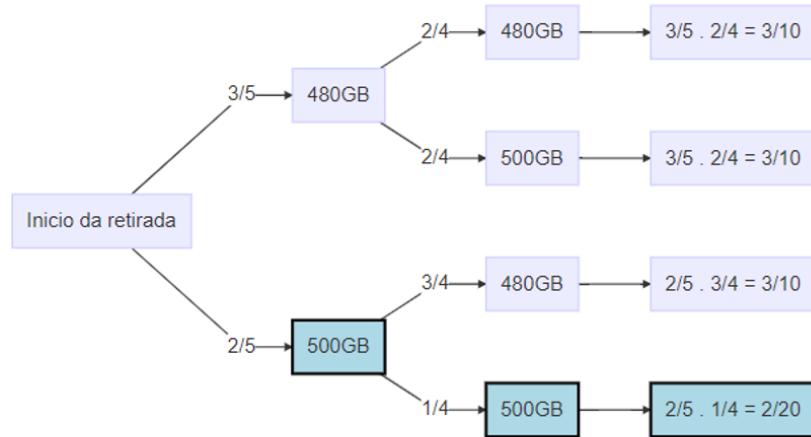


Figura 2: Cálculo das probabilidades

$$P(A \cap B) = P(A) \cdot P(B|A)$$

pode-se desenvolver os cálculos das probabilidades abaixo.

$$P(480 \cap 480) = P(\text{primeiro } 480) \cdot P(\text{segundo } 480|\text{primeiro } 480) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}$$

$$P(480 \cap 500) = P(\text{primeiro } 480) \cdot P(\text{segundo } 500|\text{primeiro } 480) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}$$

$$P(500 \cap 480) = P(\text{primeiro } 500) \cdot P(\text{segundo } 480|\text{primeiro } 500) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}$$

$$P(500 \cap 500) = P(\text{primeiro } 500) \cdot P(\text{segundo } 500|\text{primeiro } 500) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}$$

Assim a probabilidade pedida é $P(500 \cap 500) = \frac{1}{10} = 10,00\%$

- b) Qual é a probabilidade de sair dois *SSDs* com capacidades distintas? **Resposta:** 60,00 %
- c) Qual é a probabilidade de sair **pelo menos** um *SSD* de 500 GB? **Resposta:** 70,00 %
- 3) Em um lote, há 3 (três) *SSDs* SATA de 480 Gigabyte (GB) e 2 (dois) *SSDs* SATA de 500 *Gigabyte* (GB), todos indistinguíveis visualmente para quem recebeu a encomenda. Retira-se, aleatoriamente, uma **amostra** de **dois** *SSDs* com reposição deste lote e observa-se a sua capacidade.
 - a) Qual é a **probabilidade** de que os dois *SSDs* retirados tenham capacidade de 480 GB?

Definindo o espaço amostral como Ω e o evento $A = \{\text{Ser sorteado dois SSDs de 480 GB}\}$. Os pontos amostrais são.

$$\Omega = \left\{ \begin{array}{ll} (\text{SSD4801}, \text{SSD4801}) & (\text{SSD4801}, \text{SSD4802}) \\ (\text{SSD4801}, \text{SSD4803}) & (\text{SSD4801}, \text{SSD5001}) \\ (\text{SSD4801}, \text{SSD5002}) & (\text{SSD4802}, \text{SSD4801}) \\ (\text{SSD4802}, \text{SSD4802}) & (\text{SSD4802}, \text{SSD4803}) \\ (\text{SSD4802}, \text{SSD5001}) & (\text{SSD4802}, \text{SSD5002}) \\ (\text{SSD4803}, \text{SSD4801}) & (\text{SSD4803}, \text{SSD4802}) \\ (\text{SSD4803}, \text{SSD4803}) & (\text{SSD4803}, \text{SSD5001}) \\ (\text{SSD4803}, \text{SSD5002}) & (\text{SSD5001}, \text{SSD4801}) \\ (\text{SSD5001}, \text{SSD4802}) & (\text{SSD5001}, \text{SSD4803}) \\ (\text{SSD5001}, \text{SSD5001}) & (\text{SSD5001}, \text{SSD5002}) \\ (\text{SSD5002}, \text{SSD4801}) & (\text{SSD5002}, \text{SSD4802}) \\ (\text{SSD5002}, \text{SSD4803}) & (\text{SSD5002}, \text{SSD5001}) \\ (\text{SSD5002}, \text{SSD5002}) & \end{array} \right\}, \quad n(\Omega) = 25, n(A) = 9$$

O número de elementos do espaço amostral pode ser determinado pelo **princípio multiplicativo** (PM) sem a necessidade de descrever todos os pontos.

- $\underbrace{5}_{\text{(cinco SSDs para a primeira escolha)}}$. $\underbrace{5}_{\text{(cinco SSDs para a segunda escolha)}}$ = $5 \times 5 = 25$

$$n(\Omega) = 25$$

- $\underbrace{3}_{\text{(três SSDs para a primeira escolha)}}$. $\underbrace{3}_{\text{(três SSDs para a segunda escolha)}}$ = 9

$$n(A) = 9$$

Outra forma de desenvolver o cálculo da probabilidade solicitada é por meio da árvore abaixo.

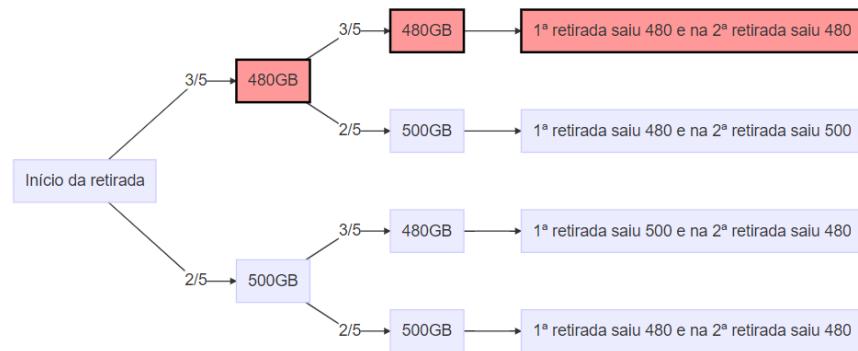


Figura 3: Diagrama em forma de árvore de possibilidades

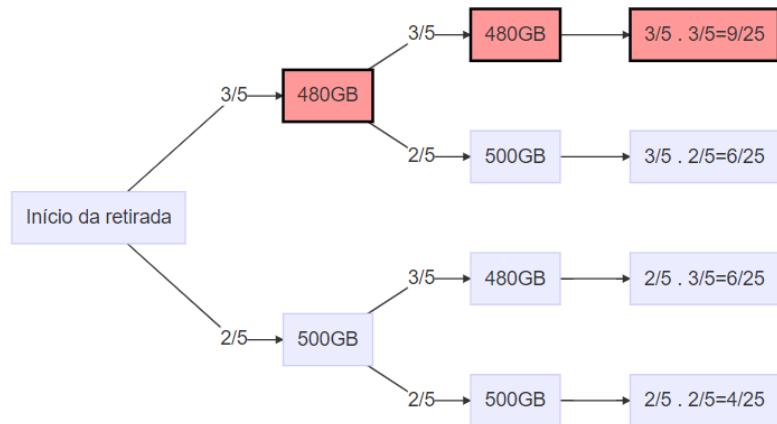


Figura 4: Cálculo das probabilidades

Assim $P(A) = \frac{n(A)}{n(\Omega)} = \frac{9}{25} = \frac{36}{100} = 36,00\%$

Observando a árvore anterior e usando o conceito de probabilidade condicional

$$P(A \cap B) = P(A) \cdot P(B|A)$$

pode-se desenvolver os cálculos das probabilidades abaixo.

$$P(480 \cap 480) = P(\text{primeiro } 480) \cdot P(\text{segundo } 480|\text{primeiro } 480) = \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{25}$$

$$P(480 \cap 500) = P(\text{primeiro } 480) \cdot P(\text{segundo } 500|\text{primeiro } 480) = \frac{3}{5} \cdot \frac{2}{5} = \frac{6}{25}$$

$$P(500 \cap 480) = P(\text{primeiro } 500) \cdot P(\text{segundo } 480|\text{primeiro } 500) = \frac{2}{5} \cdot \frac{3}{5} = \frac{6}{25}$$

$$P(500 \cap 500) = P(\text{primeiro } 500) \cdot P(\text{segundo } 500|\text{primeiro } 500) = \frac{2}{5} \cdot \frac{2}{5} = \frac{4}{25}$$

Assim a probabilidade pedida é $P(480 \cap 480) = 36,00\%$

- b) Qual é a probabilidade de sair dois *SSDs* com capacidades distintas? **Resposta:.. 48,00 %**
c) Qual é a probabilidade de sair **pelo menos** um *SSD* de 480 GB? **Resposta:.. 84,00 %**
- 4) Foi coletada uma amostra de 30 empresas que prestam serviços na área de ciência de dados, e observou-se que entre as linguagens usadas pelas empresas para desenvolver os projetos, estão presentes as *linguagens de programação python* e **R**. Os dados observados foram:

- 15 **usam python**;
- 14 **usam R**;
- 8 **usam as duas linguagens**.

Um cliente dono de uma seguradora de veículos escolhe ao acaso **uma** empresa entre as 30 e contratou um serviço de análise de dados. Considere os eventos definidos como:

- $A = \{\text{A empresa escolhida ao acaso usa } \textit{python} \text{ para desenvolver os projetos}\};$
- $F = \{\text{A empresa escolhida ao acaso usa R para desenvolver os projetos}\};$
- $A \cap F = \{\text{A empresa escolhida ao acaso usa R e } \textit{python} \text{ para desenvolver os projetos}\}.$

- a) Determine a probabilidade do serviço contratado ser realizado por **nenhuma** das duas *linguagens* citadas acima;

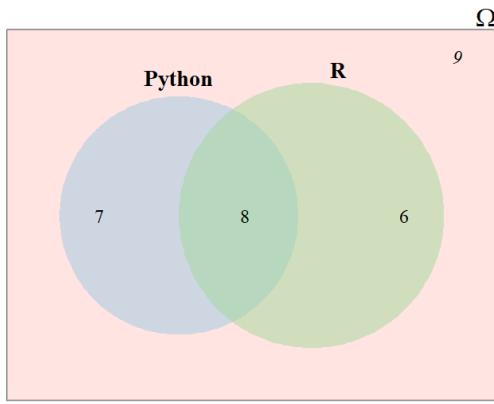


Figura 5: Diagrama de Venn

Resolução

A construção do diagrama de *Venn* abaixo é útil para visualizar os dados.

a)

$$P(A \cup F) = P(A) + P(F) - P(A \cap F)$$

$$P(A \cup F) = \frac{15}{30} + \frac{14}{30} - \frac{8}{30}$$

$$P(A \cup F) = \frac{21}{30}$$

$$P(\overline{A \cup F}) = 1 - P(A \cup F)$$

$$P(\overline{A \cup F}) = 1 - \frac{21}{30}$$

$$P(\overline{A \cup F}) = \frac{30}{30} - \frac{21}{30}$$

$$P(\overline{A \cup F}) = \frac{9}{30}$$

Assim a probabilidade do serviço contratado ser realizado por ***nenhuma*** das duas *linguagens* citadas acima é $P(\overline{A \cup F}) = \frac{9}{30}$.

- b) Determine a probabilidade do serviço contratado ser realizado pela linguagem ***python*** ou pela linguagem **R**. **Resposta:** $P(A \cup F) = \frac{21}{30}$
- c) Determine a probabilidade do serviço contratado **não** ser realizado pela linguagem **R**. **Resposta:** $P(\overline{F}) = \frac{8}{15}$

Lista de exercícios

1) Em uma empresa de tecnologia, foi feita uma análise com 200 computadores para verificar a instalação de dois *softwares* importantes: *antivírus* e *firewall*. Os dados levantados foram:

- 120 computadores têm *antivírus* instalado.
- 90 computadores têm *firewall* instalado.
- 60 computadores têm ambos os *softwares* instalados.

Ao escolher, de forma aleatória, um computador dentre os 200 existentes, qual é a probabilidade de que ele possua **pelo menos um** dos dois *softwares* instalados? **Resposta:..** 75,00%

2) Em uma pesquisa com 150 usuários de um laboratório de informática, constatou-se que:

- 90 utilizam o sistema operacional *Windows*.
- 50 utilizam o sistema operacional *Linux*.
- 20 utilizam ambos os sistemas operacionais.

Ao escolher aleatoriamente um usuário entre os 150, qual é a probabilidade de que ele utilize **pelo menos um** dos dois sistemas operacionais? **Resposta:..** 80,00%

3) Uma equipe de Tecnologia da Informação (TI) precisa selecionar 2 arquivos aleatoriamente de um diretório que contém 30 arquivos, sendo 5 arquivos críticos e 25 arquivos não críticos. Considerando amostragem **sem** reposição calcule a probabilidade de que ambos os arquivos selecionados sejam críticos. **Resposta:..** 2/87

4) Uma equipe de testes seleciona aleatoriamente 2 módulos de um sistema para verificar se há *bugs*. Sabemos que 3 dos 10 módulos possuem *bugs*. Considerando uma amostragem **sem** reposição, calcule a probabilidade de que ambos os módulos selecionados contenham *bugs*? **Resposta:..** 1/15

5) Uma medicação cura um paciente com probabilidade 4/5. Três pacientes foram tratados de forma independente com essa medicação e, após 30 dias, observou-se o número (X) de pacientes curados. Determine a probabilidade de que **pelo menos** um paciente tenha sido curado. **Resposta:..** 124/125

6) Em uma empresa de tecnologia, 40% dos computadores da rede possuem um determinado *software antivírus instalado*. Suponha que **um** computador seja escolhido aleatoriamente.

a) Determine a **probabilidade** de que o computador possua o *antivírus*. **Resposta:..** 2/5

b) Determine a **chance** de que o computador possua o *antivírus*. **Resposta:..** 2/3

Referências

- [1] DEVORE, Jay L. *Probabilidade e Estatística para Engenharia e Ciências*. Tradução da 9^a edição norte-americana. São Paulo: Cengage Learning Brasil, 2018.
- [2] FERREIRA, D. F. *Estatística Básica*. 2. ed. Lavras: Ed. UFLA, 2009.
- [3] FERREIRA, Daniel Furtado. *Fundamentos de Probabilidade*. Lavras: Editora UFLA, 2020.
- [4] MAGALHÃES, M. N. *Probabilidade e Variáveis Aleatórias*. 2^a ed. São Paulo: Edusp, 2006.
- [5] MAGALHÃES, M. N.; LIMA, A. C. P. *Noções de Probabilidade e Estatística*. 6^a ed. São Paulo: USP, 2006.
- [6] MEYER, Paul L. *Probabilidade: Aplicações à Estatística*. Rio de Janeiro: Grupo GEN, 2010.
- [7] MORETTIN, Luiz Gonzaga. *Estatística Básica: Probabilidade e Inferência*. São Paulo: Pearson, 2010.
- [8] MORETTIN, Pedro A. *Estatística Básica*. Rio de Janeiro: Grupo GEN, 2023.