Curso de Estatística e Probabilidade

DPAA-2.339 - Estatística e Probabilidade

Prof. Thiago VedoVatto
thiago.vedovatto@ifg.edu.br
thiagovedovatto.site

Instituto Federal de Educação, Ciência e Tecnologia de Goiás
Campus de Goiânia

2 de março de 2020

Informações sobre a disciplina

Avaliação

A nota final M_F será obtida da seguinte forma:

$$M_F = \frac{1}{3}\left(\sum_{i=1}^{2} A_i + \sum_{i=1}^{12} T_i\right),$$

onde $A_i \in [0, 10]$ é a nota obtida na i-ésima avaliação e $T_i \in [0, 1]$ é a nota obtida no i-ésimo trabalho. Cada avaliação terá valor de 10 pontos e cada trabalho terá valor de 1 ponto.

INSTITUTO FEDERAL Goiás

Informações sobre a disciplina

Livros Recomendados

- Bussab & Morettin (2013)
- Magalhães & Lima (2015)

Entrega dos Trabalhos

1. As datas de entrega dos trabalhos são divulgadas no q-acadêmico.
2. A entrega dos trabalhos deve ser feita exclusivamente por meio digital (email: thiago.vedovatto@ifg.edu.br) com o assunto:
   `Trabalho N° - Aluno`.
   Exemplo: `Trabalho 3 - Joãozinho da Silva Sauro`.
   Os trabalhos devem ser entregues em um único arquivo no formato pdf.
3. Não serão considerados trabalhos entregues
   a. fora do prazo.
   b. fisicamente ou por outros meios digitais como whatsapp, telegram, etc.

FEDERAL
Goiás

Introdução à Probabilidade

Probabilidades

Experimento Aleatório

É qualquer ação cujo resultado não pode ser previsto com certeza (não determinístico).

Variável de Interesse

É a variável que se deseja observar num determinado experimento aleatório.

Espaço Amostral

É o conjunto de todos os possíveis resultados de um experimento aleatório. Notação: Ω.

Eventos

São os subconjuntos do espaço amostral Ω.

Eventos Elementares

São os elementos do espaço amostral que não podem ser expressos como união de outros dois eventos não vazios de Ω.

Exercício - Magalhães & Lima (2015, p. 52)

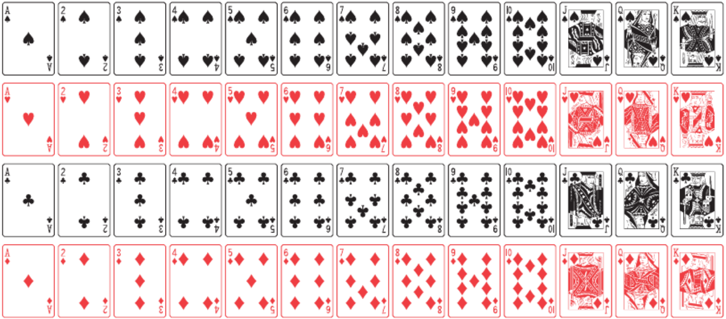
Ref.:LG3L

Para cada um dos casos abaixo, escreva o espaço amostral correspondente e conte seus eventos elementares.

1. Uma moeda é lançada duas vezes e observam-se as faces obtidas.
2. Um dado é lançado duas vezes e a ocorrência de face par ou ímpar é observada.
3. Uma urna contém 10 bolas azuis e 10 vermelhas com dimensões rigorosamente iguais. Três bolas são selecionadas ao acaso com reposição e as cores são anotadas.
4. Dois dados são lançados simultaneamente e estamos interessados na soma das faces observadas.
5. Em uma cidade, famílias com 3 crianças são selecionadas ao acaso, anotando-se o sexo de cada uma.
6. Uma máquina produz 20 peças por hora, escolhe-se um instante qualquer e observa-se o número de defeituosas na próxima hora.
7. Uma moeda é lançada consecutivamente até o lançamento da primeira cara.

Exercício - Weiss (2012, p. 153)

Ref.:4F22



Considere o experimento aleatório de sortear uma carta de baralho. Quais as cartas compõem os eventos:

a. Seleciona-se o rei de copas (coração).
b. Seleciona-se um rei.
c. Seleciona-se uma carta de copas (coração).
d. Seleciona-se uma carta de figura.

Exercício - Bussab & Morettin (2013, p. 108)

Ref.:BG1Z

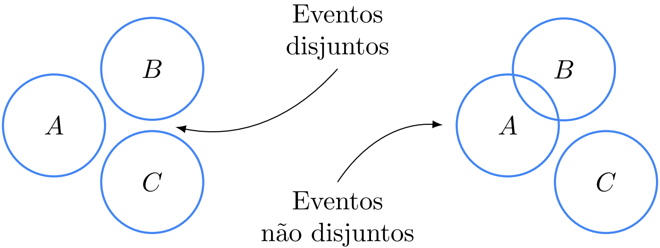
Qual é o espaço amostral e a variável de interesse nos seguintes experimentos? Quais são experimentos aleatórios?

1. Sorteamos um aluno da classe ao acaso e medimos o valor de sua altura.
2. Jogamos um dado e observamos a face que ficou virada para cima.
3. Jogamos um dado de uma determinada altura e observamos o tempo que gasta até chegar ao solo.
4. Jogamos um dado e observamos o número de vezes que ele quica até parar.
5. Em uma escola de futebol observamos um jogador cobrar faltas e contamos a quantidade de vezes que ele acerta.
6. Colocamos 1 litro de água no fogo e medimos a sua temperatura depois de 20 minutos.
7. Jogamos uma moeda e verifica-se o seu valor.

INSTITUTO FEDERAL
Goiás

Eventos disjuntos e complementares

Os eventos $A_1, \ldots, A_n$ são disjuntos (mutuamente excludentes) se $A_i \cap A_j = \emptyset$ para quaisquer i e j distintos.



Dois eventos A e B no espaço amostral Ω são complementares se:

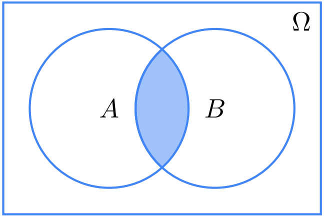
1. $A \cap B = \emptyset$
2. $A \cup B = \Omega$

O complementar de A será denotado por $\overline{A}$.

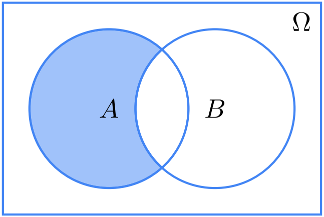
Operações entre eventos

Sejam A e B eventos em um mesmo espaço amostral Ω. Definimos as seguintes operações entre os eventos A e B:

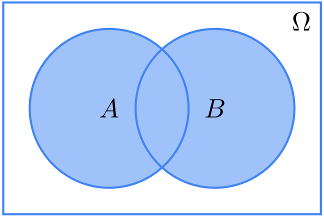
Interseção: $A \cap B$



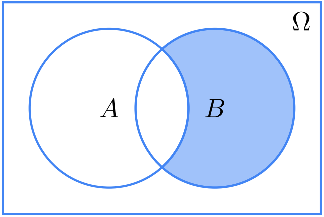
Diferença: $A - B = A \cap \overline{B}$



União: $A \cup B$



Diferença: $B - A = B \cap \overline{A}$



Exercício - Magalhães & Lima (2015, p. 53)

Ref.:HZ9W

Sendo A e B eventos de um mesmo espaço amostral, "traduza" para a linguagem da Teoria dos Conjuntos, as seguintes situações:

1. Pelo menos um dos eventos ocorre.
2. O evento A ocorre mas B não.
3. Nenhum deles ocorre.
4. Exatamente um dos eventos ocorre.

INSTITUTO FEDERAL
Goiás

Exercício

Ref.:BGC1

No experimento de selecionar aleatoriamente uma carta de um baralho comum, considere os eventos:

- A Seleciona-se o rei de copas.
- B Seleciona-se um rei.
- C Seleciona-se uma carta de copas.
- D Seleciona-se uma carta de figura.

Determine adequadamente os eventos:

- a $A \cap B$
- b $A \cup B$
- c $A - B$
- d $B - A$
- e $\bar{D}$
- f $B \cap C$
- g $B \cup C$
- h $C \cap D$

Exercício

Ref.:NGT5

No experimento de selecionar aleatoriamente uma carta de um baralho comum, considere os eventos:

- A seleciona-se uma carta de copas;
- B seleciona-se uma figura;
- C seleciona-se um Ás;
- D seleciona-se um oito;
- E seleciona-se um dez ou um valete

Qual das seguintes coleções de eventos são disjuntos:

- a C e D;
- b D, E e A;
- c C e E;
- d D, E, A e B;
- e D e E

Definição frequentista de probabilidade

Considere o número limite de frequências relativas como o valor da probabilidade. Seja n_A o número de ocorrências do evento A em n repetições independentes do experimento em questão. Então:

$$\mathbb{P}(A) = \lim_{n \to \infty} \frac{n_A}{n}.$$

Lembre-se!!!

Probabilidade não existe!!!

Definição clássica de probabilidade

Seja um espaço amostral Ω composto por um número finito de eventos elementares equiprováveis:

$$\Omega = \{\omega_1, \ldots, \omega_n\}.$$

A probabilidade de ocorrência do evento $A \subset \Omega$ é dada por:

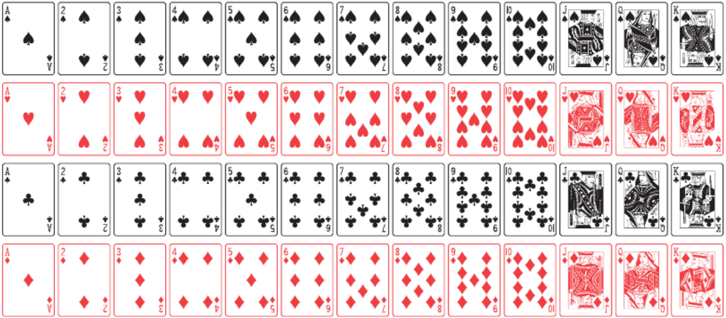
$$\boxed{\mathbb{P}(A) = \frac{n(A)}{n(\Omega)}},$$

onde $n(A)$ é o número de eventos elementares contidos em A e $n(\Omega)$ é o número de eventos elementares no espaço amostral. O caso onde o espaço amostral possui um número infinito de elementos (enumerável ou não) não será foco desse curso introdutório.

INSTITUTO FEDERAL
Goiás

Exercícios

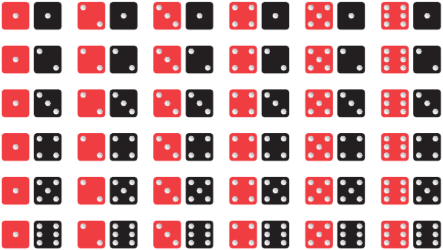
Ref.:BN6G



1. Três moedas são jogadas simultaneamente. Qual a probabilidade de:
    a. Obter exatamente 2 caras?
    b. Obter pelo menos 2 caras?
2. Seleciona-se aleatoriamente uma carta de um baralho comum. Qual a probabilidade de que:
    a. A carta seja um ás e seja vermelha?
    b. A carta seja um ás ou seja vermelha?

Exercícios

Ref.:KZ4C



1. Suponhamos que eu lance simultaneamente um tetraedro (dado de quatro faces) e uma moeda. Qual é a probabilidade de ocorrer uma face par no tetraedro e sair coroa na moeda?
2. Dois dados de cores diferentes são jogados simultaneamente.
   a. Qual a probabilidade de que a soma deles seja maior que sete?
   b. Qual a probabilidade de que o máximo seja maior ou igual à três?
3. De um grupo de n objetos escolhemos r ao acaso com reposição. Qual a probabilidade de não sortearmos objetos repetidos?

1. Uma moeda é lançada duas vezes, sucessivamente. Qual a probabilidade de observarmos 1 cara e 1 coroa?
2. Uma urna contém 10 bolas identificadas como $B_1 \ldots B_{10}$. Qual a probabilidade de que a bola sorteada tenha índice par? E do índice ser primo?
3. Numa classe de 55 alunos, 21 praticam vôlei e basquete, 39 praticam vôlei e 33 praticam basquete. Um aluno da classe é escolhido ao acaso. Qual a probabilidade de o aluno escolhido praticar um e somente um desses esportes?
4. Cinco homens e cinco mulheres estão dispostas em fila indiana. Qual a probabilidade de que:
   a. A primeira pessoa da fila seja homem?
   b. A primeira e a última pessoas da fila sejam homens?
5. Tomando-se, ao acaso, uma das retas determinadas pelos vértices de um pentágono regular. Qual a probabilidade de que a reta tomada ligue dois vértices consecutivos?

Função de probabilidade - axiomas da probabilidade

Uma função $\mathbb{P}(\cdot)$, com domínio no espaço amostral Ω, é denominada probabilidade se satisfaz as seguintes condições:

1. $\mathbb{P}(A) \geq 0, \quad \forall A \subset \Omega;$
2. $\mathbb{P}(\Omega) = 1;$
3. $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$, com $A_i \cap A_j = \emptyset$ sempre que $i \neq j$.

Consequências:

- $\Omega \cap \emptyset = \emptyset$. Portanto, usando as propriedades 2 e 3, $\mathbb{P}(\emptyset) = 0$;
- $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$ (Monotonicidade);
- E da monotonicidade vem naturalmente que $0 \leq \mathbb{P}(A) \leq 1$.

INSTITUTO FEDERAL
Goiás

Regra da Adição de Probabilidades

Sejam $A_1, \ldots, A_n$ eventos de Ω, então:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \\ + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right)$$

Regra da Adição - Casos Particulares

Sejam A e B dois eventos de Ω, então:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Sejam A e B e C três eventos de Ω, então:

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) + \\ - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$$

Exercício

Ref.:BF5Z

1. Dois dados são lançados simultaneamente. Qual a probabilidade de se obter soma dos pontos igual a oito ou dois números iguais?
2. Sejam A e B eventos em um dado espaço amostral Ω, tais que $\mathbb{P}(A) = 2/5$ e $\mathbb{P}(A \cup B) = 4/5$. Determine $\mathbb{P}(B)$ tais que A e B sejam disjuntos.
3. Sejam A e B eventos em um dado espaço amostral Ω, tais que $\mathbb{P}(A) = 1/5$, $\mathbb{P}(B) = p$, $\mathbb{P}(A \cup B) = 1/2$ e $\mathbb{P}(A \cap B) = 1/10$. Determine o valor de p. (Magalhães & Lima, 2015, p. 53)
4. Sejam A e B eventos em um espaço amostral Ω, onde ocorrer B é três vezes mais provável que ocorrer A. Sabendo que $\mathbb{P}(A \cup B) = 4/5$ determine $\mathbb{P}(A)$ e $\mathbb{P}(B)$ quando:
   a. A e B são disjuntos
   b. $\mathbb{P}(A \cap B) = 1/8$.
5. Um torneio é disputado por 4 times A, B, C e D. É três vezes mais provável que A vença do que B, 2 vezes mais provável que B vença do que C e é 3 vezes mais provável que C vença do que B. Quais as probabilidades de cada time vencer? (Morgado et al., 1991, p. 143)
6. Um número entre 1 e 200 é escolhido aleatoriamente. Calcular a probabilidade de que seja divisível por 5 ou por 7. Resp.: $63/200$

Considere os dados tabulados referentes a preferencia esportiva de 285 pessoas entre brasileiros e estrangeiros.

	Brasileiro	Estrangeiro	Total
Futebol	70	15	85
Vôlei	25	30	55
Queimada	10	50	60
Natação	15	70	85
Total	120	165	285

Se uma pessoa for sorteada ao acaso qual a probabilidade de:

a. ser brasileira?
b. preferir futebol?
c. ser estrangeira e preferir natação?
d. ser estrangeira ou preferir queimada?

A Regra dos Complementares

Para quaisquer eventos A e B:

1. $\mathbb{P}(A) = 1 - \mathbb{P}(\overline{A})$
2. $\mathbb{P}(A \cap B) = 1 - \mathbb{P}(\overline{A} \cup \overline{B})$
3. $\mathbb{P}(\overline{A} \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Provas:

1. Note que A e $\overline{A}$ são eventos disjuntos e $A \cup \overline{A} = \Omega$. Aplicando a segunda e a terceira propriedades das funções de probabilidade:

$$\begin{aligned}
\mathbb{P}(A \cup \overline{A}) &= \mathbb{P}(A) + \mathbb{P}(\overline{A}) \\
\mathbb{P}(\Omega) &= \mathbb{P}(A) + \mathbb{P}(\overline{A}) \\
\mathbb{P}(A) &= 1 - \mathbb{P}(\overline{A}).
\end{aligned}$$

2. Note que $\overline{A} \cup \overline{B} = \overline{A \cap B}$, logo $A \cap B$ e $\overline{A} \cup \overline{B}$ são complementares, logo esse resultado decorre diretamente do resultado 1.
3. Note que $B = (A \cap B) \cup (\overline{A} \cap B)$, mas $A \cap B$ e $\overline{A} \cap B$ são disjuntos, logo $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(\overline{A} \cap B)$.

Exercícios

Ref.:M36Q

1. Um dado é lançado três vezes sucessivamente. Qual a probabilidade de não ocorrerem três números iguais?
2. Dois dados são lançados independentemente. Qual a probabilidade de não obtermos dois números ímpares?
3. A probabilidade de um cavalo vencer três ou menos corridas é de 58%; a probabilidade de ele vencer três ou mais corridas é de 71%. Qual é a probabilidade do cavalo vencer exatamente três corridas?
4. Uma caixa contém nove peças das quais três são defeituosas. Sorteamos duas peças. Qual a probabilidade de não escolhermos duas peças defeituosas?

INSTITUTO FEDERAL
Goiás

Uma universidade tem 10 mil alunos dos quais 4 mil são considerados esportistas. Temos, ainda, que 500 alunos são do curso de biologia diurno, 700 da biologia noturno, 100 são esportistas e da biologia diurno e 20 são esportistas e da biologia noturno.Um aluno é escolhido, ao acaso, e pergunta-se a probabilidade de:

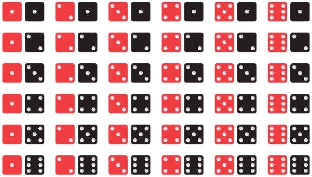
1. Ser esportista.
2. Ser esportista e aluno da biologia e noturno.
3. Não ser da biologia.
4. Ser esportista ou aluno da biologia.
5. Não ser esportista, nem aluno da biologia.

Exercícios - Weiss (2012, p. 147)

Ref.:K3X8

Em um jogo de dados são jogados dois dados honestos de seis faces. Considere os eventos:

- A) A soma das faces é 7;
- B) A soma das faces é 11;
- C) A soma das faces é 2;
- D) A soma das faces é 3;
- E) A soma das faces é 12;
- F) A soma das faces é 8;
- G) As faces são iguais.



- a) Determine a probabilidade de todos os eventos.
- b) O jogador vence esse jogo na primeira rodada se a soma dos dados for 7 ou 11. Calcule a probabilidade desse evento.
- c) O jogador perde esse jogo na primeira rodada se a soma dos dados for 2, 3 ou 12. Calcule a probabilidade desse evento.
- d) Qual a probabilidade das faces serem diferentes?
- e) Qual a probabilidade da soma ser 8 ou das faces serem iguais?

INSTITUTO FEDERAL Goiás

Considere os dados tabulados referentes a preferencia esportiva de 285 pessoas entre brasileiros e estrangeiros.

	Brasileiro	Estrangeiro	Total
Futebol	70	15	85
Vôlei	25	30	55
Queimada	10	50	60
Natação	15	70	85
Total	120	165	285

Se uma pessoa for sorteada ao acaso qual a probabilidade de:

1. não ser brasileira e preferir de futebol?
2. não ser estrangeira e nem preferir de vôlei?
3. não ser estrangeira ou não preferir de queimada?

Probabilidade Condicional

Dados os eventos A e B a probabilidade condicional de A dado B, $\mathbb{P}(A|B)$, é definida como:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{se } \mathbb{P}(B) > 0$$

e

$$\mathbb{P}(A|B) = \mathbb{P}(A), \quad \text{se } \mathbb{P}(B) = 0$$

Onde:

$\mathbb{P}(B)$ é a probabilidade à priori e

$\mathbb{P}(A|B)$ é a probabilidade à posteriori

Regra do Produto de Probabilidades

Dados os eventos A e B:

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

INSTITUTO FEDERAL Goiás

Exercício

Ref.:C7L3

Considere os dados tabulados referentes a preferencia esportiva de 285 pessoas entre brasileiros e estrangeiros.

	Brasileiro	Estrangeiro	Total
Futebol	70	15	85
Vôlei	25	30	55
Queimada	10	50	60
Natação	15	70	85
Total	120	165	285

Se uma pessoa for sorteada ao acaso qual a probabilidade de:

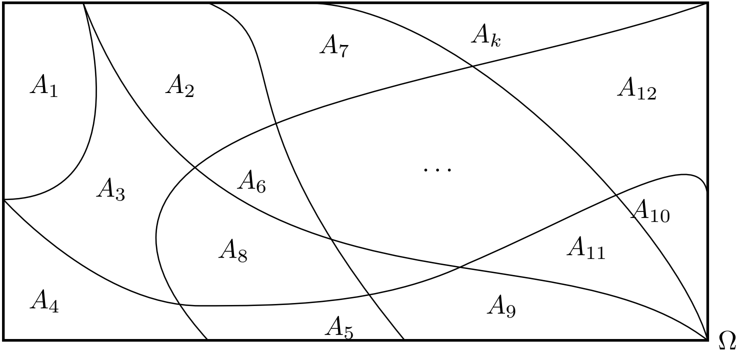
1. sabendo que a pessoa é estrangeira, prefira futebol?
2. sabendo que a pessoa prefere vôlei, seja estrangeira?
3. sabendo que a pessoa não prefere queimada, seja brasileira?

INSTITUTO FEDERAL
Goiás

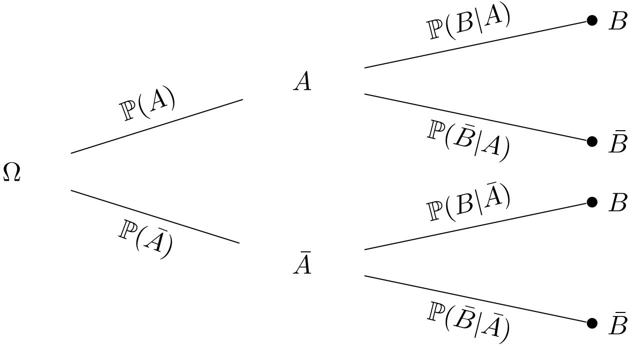
Partição do Espaço Amostral

O conjunto $\mathcal{P} = \{A_1, A_2 \ldots A_n\}$ é uma partição do espaço amostral se seus elementos não têm interseção entre si e se a união de seus elementos é o espaço amostral. Isto é,

1. $\bigcup_{i=1}^{n} A_i = \Omega$
2. $A_i \cap A_j = \emptyset, \forall i \neq j$



Sejam $\mathcal{P}_1 = \{A_1, A_2 \ldots A_n\}$ e $\mathcal{P}_2 = \{B_1, B_2 \ldots B_m\}$ duas partições de Ω. Podemos esquematizar as probabilidades associadas aos eventos que compõem as partições $\mathcal{P}_1$ e $\mathcal{P}_2$ por um diagrama de árvores. No caso particular onde $\mathcal{P}_1 = \{A, \overline{A}\}$ e $\mathcal{P}_2 = \{B, \overline{B}\}$, um diagrama adequado seria:



Exercício - Morgado et al. (1991, p 161)

Ref.:BT1L

Consideremos dois dados: um deles equilibrado e outro viciado com $\mathbb{P}\{1\} = {}^{1}\!/_{2}$ e $\mathbb{P}\{2\} = \mathbb{P}\{3\} = \mathbb{P}\{4\} = \mathbb{P}\{5\} = \mathbb{P}\{6\} = {}^{1}\!/_{10}$. Escolhe-se um dos dados ao acaso e se efetuam dois lançamentos, obtendo-se dois uns. Qual a probabilidade condicional de que o dado escolhido tenha sido o viciado?
Resp.: ${}^{9}\!/_{10}$

INSTITUTO FEDERAL Goiás

Eventos Independentes

Dois eventos A e B são independentes se, e somente se:

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

Intuitivamente dois eventos são independentes quando a ocorrência de um não altera a probabilidade de ocorrência do outro, ou seja:

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

Três eventos A, B e C são independentes se, e somente se:

1. $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
2. $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$
3. $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$
4. $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$

Se apenas as três primeiras condições forem satisfeitas os eventos são mutuamente independentes.

INSTITUTO FEDERAL
Goiás

Exercício

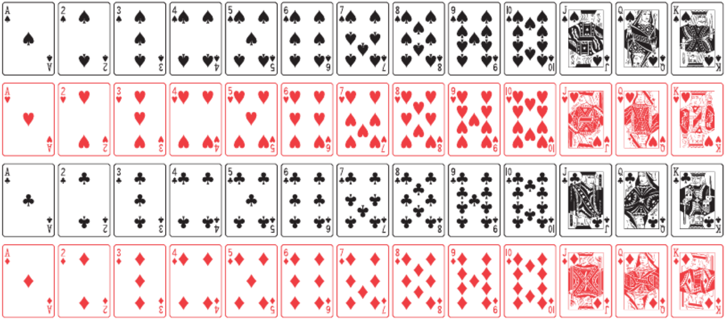
Ref.:NBF9

1. Em uma gaveta temos 12 camisas, das quais, quatro são de gola polo e o restante, de gola normal. Retirando duas camisas sucessivamente ao acaso e sem reposição, qual é a probabilidade de as duas camisas serem de gola polo? E se retirarmos com reposição?
2. Em uma cesta, temos oito bombons de morango, dez bombons de maracujá e quatro bombons de uva. Determine a probabilidade de retiramos sucessivamente com reposição, três bombons de maracujá. E se as retiradas forem sem reposição?
3. Um juiz de futebol possui três cartões no bolso. Um é todo amarelo, outro é todo vermelho e o terceiro é vermelho de um lado e amarelo do outro. Num determinado lance, o juiz retira ao acaso, um cartão do bolso mostrando-o a um jogador. Qual é a probabilidade de a face que o juiz vê ser vermelha, e de a outra face mostrada ao jogador, ser amarela?
4. Em uma urna há 5 bolas amarelas, 3 verdes e 2 roxas. Sorteamos 3 bolas uma após a outra. Qual a probabilidade de obtermos a sequencia amarela-verde-roxa considerando que:
   a. o sorteio é sem reposição
   b. o sorteio é com reposição

INSTITUTO FEDERAL Goiás

Exercício - Morgado et al. (1991, p 166)

Ref.:NX4L



Treze cartas são escolhidas de um baralho comum de 52 cartas. Seja A o evento "O ás de copas está entre as treze cartas" e B o evento "As treze cartas são do mesmo naipe". Verifique se A e B são independentes.

Resp.: Os eventos A e B são independentes.

Exercício

Ref.:MZ2O

1. (Morgado et al., 1991, p 169) Um jogador deve enfrentar, em um torneio, dois outros A e B. Os resultados dos jogos são independentes e as probabilidades dele ganhar de A e de B são $1/3$ e $2/3$ respectivamente. O jogador vencerá o torneio se vencer dois jogos consecutivos, de uma série de três. Que série de jogos é mais favorável ao jogador ABA ou BAB?
   Resp.: Surpreendentemente é a sequência ABA!!!
2. (Morgado et al., 1991, p 157) Numa prova há 7 perguntas do tipo V ou F. Calcular a probabilidade de acertarmos todas as sete se:
   a. escolhermos aleatoriamente as sete respostas;
   b. escolhermos aleatoriamente as respostas sabendo que há mais respostas "verdadeiro" do que "falso".
3. (Morgado et al., 1991, p 162) Marina quer enviar uma carta à Verônica. A probabilidade de que Marina escreva a carta é de $8/10$. A probabilidade de que o correio não a perca é de $9/10$. A probabilidade de que o carteiro a entregue é de $9/10$. Dado que Verônica não recebeu a carta, qual é a probabilidade condicional de que Marina não a tenha escrito?
   Resp.: $25/44$

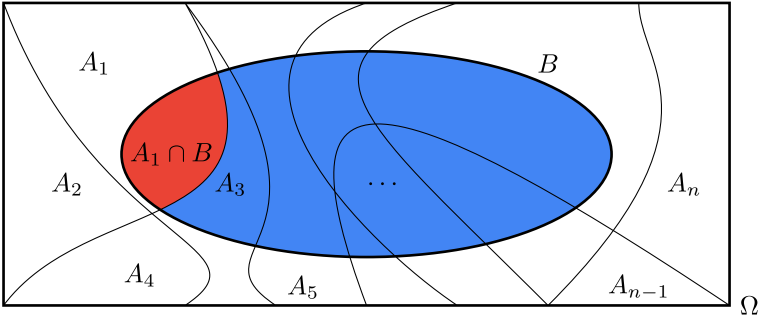
INSTITUTO FEDERAL Goiás

Teorema da Probabilidade Total

Seja B um evento contido numa união de eventos disjuntos $A_1, A_2, \ldots, A_n$ em um espaço amostral Ω tais que $\mathbb{P}(A_1) > 0, \mathbb{P}(A_2) > 0, \ldots, \mathbb{P}(A_n) > 0$. Então:

$$\boxed{\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i)}$$

Para demonstrarmos esse resultado note B pode ser expresso como uma união disjunta de eventos em Ω da seguinte forma:



$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)$$

Aplicando o terceiro axioma da probabilidade:

$$
\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}(A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B) \\
&= \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B) + \ldots + \mathbb{P}(A_n \cap B)
\end{aligned}
$$

Aplicando o teorema da multiplicação de probabilidades:

$$
\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2) + \ldots + \mathbb{P}(A_n)\mathbb{P}(B|A_n)
$$

Portando o teorema da probabilidade total está demonstrado:

$$
\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i)
$$

INSTITUTO FEDERAL
Goiás

1. Um piloto de fórmula 1 tem 50% de probabilidade de vencer determinada corrida, quando esta se realiza sob chuva. Caso não chova durante a corrida, sua probabilidade de vitória é de 25%. Se o serviço de Meteorologia estimar em 30% a probabilidade de que chova durante a corrida, qual é a probabilidade deste piloto ganhar a corrida? Resp.: 0,325
2. (Ross, 2010, p. 66) Uma companhia de seguros acredita que as pessoas podem ser divididas em duas classes: As que já sofreram algum acidente e as que não. Dados estatísticos mostram que uma pessoa que já tenha sofrido algum acidente tem 0,4 de probabilidade de sofrer um novo acidente num período de um ano. As pessoas que nunca sofreram acidente tem 0,2 de probabilidade de sofrer seu primeiro acidente num período de um ano.
   a. Sabendo que 30% da população já sofreu algum acidente, qual é a probabilidade de que um indivíduo escolhido aleatoriamente sofra um acidente no período de um ano? Resp.: 0,26
   b. Supondo que um indivíduo aleatório sofrerá um acidente no período de um ano. Qual a probabilidade de que o indivíduo tenha sofrido seu primeiro acidente? Resp.: 0,46

INSTITUTO FEDERAL
Goiás

3. Três candidatos: João, Maria e Pedro, disputam a presidência do diretório de estudantes de uma universidade. Uma prévia eleitoral mostra que suas chances de vencer são respectivamente 0,5; 0,3 e 0,2. As probabilidades de que eles venham a promover um festival de música se eles forem eleitos são 0,7; 0,6 e 0,9, respectivamente. Qual é a probabilidade de que um festival de música sejam promovido depois da eleição? Resp.: 0,71
4. (Magalhães & Lima, 2015, p. 58) Suponha que um fabricante de sorvetes recebe 20% de todo o leite que utiliza de uma fazenda F_1, 30% de uma fazenda F_2 e 50% de F_3. Um órgão de fiscalização inspecionou as fazendas de surpresa e observou que 20% do leite produzido por F_1 estava adulterado por adição de água, enquanto que para F_2 e F_3, essa proporção era de 5% e 2% respectivamente. Na indústria de sorvete os galões de leite são armazenados em refrigeradores sem identificação das fazendas. Para um galão escolhido ao acaso, Qual a probabilidade do leite estar adulterado?
   Resp.:

INSTITUTO FEDERAL
Goiás

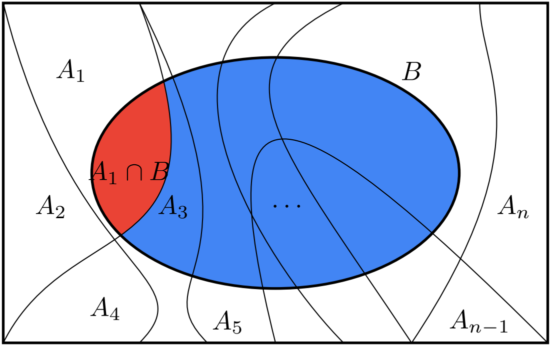
Teorema de Bayes

Se B é um evento contido numa união de eventos disjuntos $A_1, \ldots, A_n$ e $\mathbb{P}(A_1) > 0, \ldots, \mathbb{P}(A_n) > 0$. Então:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i).$$

Nessas condições, se $\mathbb{P}(B) > 0$, então, para $i = 1, \ldots, n$:

$$\begin{aligned}\mathbb{P}(A_i|B) &= \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i)}\end{aligned}$$



Exercícios - Teorema de Bayes I

1. (Morgado et al., 1991, p 159) Sabe-se que 80% dos pênaltis marcados a favor do Brasil são cobrados por jogadores da Jataiense. A probabilidade de um pênalti ser convertido é de 40% se o jogador for da Jataiense e de 70% caso contrário. Um pênalti a favor do Brasil acabou de ser marcado:

   a. Qual a probabilidade do pênalti ser cobrado por um jogador da Jataiense e ser convertido? R.: $0,32$.
   b. Qual a probabilidade do pênalti ser convertido? R.: $0,46$.
   c. Um pênalti a favor do Brasil acabou de ser perdido. Qual a probabilidade do jogador que cobrou o pênalti tenha sido da Jataiense? R.: $0,88$.

2. Suponha que o tratamento do doutor Silva é tal que existe uma chance de que o seu paciente morra, ainda que seu diagnóstico tenha sido correto. A chance de que seu diagnóstico esteja errado é de 10%. A chance de que o paciente morra se o diagnóstico está errado é de 90% e, caso contrário, é de 5%. Sabe-se que um paciente do doutor Silva morreu hoje. Qual a probabilidade de que tenha ocorrido um erro no diagnóstico? R.: $2/3$

INSTITUTO FEDERAL
Goiás

3. (Morgado et al., 1991, p. 164) Durante o mês de agosto a probabilidade de chuva em um dia determinado é de $4/10$. Uma equipe de futebol ganha um jogo em um dia de chuva com probabilidade $6/10$ e em um dia sem chuva com probabilidade de $4/10$.
   a. Qual a probabilidade da equipe ganhar? R.: 0,48
   b. Sabendo que essa equipe ganhou um jogo em um dia do mês de agosto, qual a probabilidade de que choveu nesse dia? R.: $1/2$
4. (Morgado et al., 1991, p. 165) Num exame há 3 respostas para cada pergunta e apenas uma delas é a certa. Portanto, para cada pergunta, um aluno tem probabilidade $1/3$ de escolher a resposta certa se ele está adivinhando a resposta e probabilidade 1 se sabe a resposta. Um estudante sabe 30% das perguntas do exame.
   a. Qual a probabilidade do aluno acertar uma questão em particular? R.: $8/15$
   b. Se ele deu a resposta correta para uma das perguntas, qual é a probabilidade de que a adivinhou? R.: 0,44

INSTITUTO FEDERAL
Goiás

5. Três urnas I, II e III contém respectivamente 1 bola branca e 2 pretas, 2 brancas e 1 preta e 3 brancas e 2 pretas. Uma urna e escolhida ao acaso e dela é retirada uma bola, que é branca. Qual é a probabilidade condicional de que a urna escolhida foi a urna I? R.: $^5/_{24}$
6. (Ross, 2010, p. 102) Em uma certa comunidade, 36% das famílias possuem um cachorro e 22% das famílias que possuem um cachorro também possuem um gato. Sabendo que 30% das famílias possuem um gato determine:
   a. A probabilidade de que uma família selecionada aleatoriamente possua um gato e um cachorro? R.: 0,0792
   b. A probabilidade condicional de que uma família selecionada aleatoriamente possua um cachorro dado que já possui um gato. R.: 0,264
7. Duas máquinas A e B produzem 3 mil peças em um dia. A máquina A produz mil peças, das quais 3% são defeituosas. A máquina B produz as restantes 2 mil, das quais 1% são defeituosas. Da produção total de um dia uma peça é escolhida ao acaso e, examinando-a, constata-se que é defeituosa. Qual é a probabilidade de que a peça tenha sido produzida pela máquina A? R.: $^3/_5$

INSTITUTO FEDERAL Goiás

8. Um estudante resolve um teste com questões do tipo verdadeiro-falso. Ele sabe dar a solução correta para 40% das questões. Quando ele responde uma questão cuja solução conhece ele dá a resposta correta, e nos outros casos decide na cara ou coroa. Se uma questão foi respondida corretamente, qual é a probabilidade de que ele saiba a resposta? R.: $^4/_7$
9. Um geólogo tem em seu laboratório dez amostras de solo tipo A e dez amostras de solo tipo B. Para um experimento ele seleciona ao acaso 15 amostras para serem analisadas.
    a. Quais os possíveis valores para o número de amostras do tipo B que são selecionadas e quais suas probabilidades. R.: $X \in \{5, \ldots, 10\}$
    b. Qual a probabilidade de que a seleção contenha todas as dez amostras do tipo A ou todas as dez amostras do tipo B? R.: 0,0326
    c. Qual a probabilidade de que o número de amostras tipo B selecionadas diste não mais que um desvio padrão da média? R.: 0,6966
10. Dentre os estudantes João, Pedro e Manuel, o professor escolhe ao acaso um deles para fazer uma pergunta. Se cinco perguntas forem feitas. qual a probabilidade de:
    a. Manuel nunca ser escolhido?

INSTITUTO FEDERAL
Goiás

   b. Um (qualquer) dos estudantes não ser solicitado a responder sequer uma pergunta?

11. (Bussab & Morettin, 2013, p. 128) Os colégios A, B e C têm as seguintes porcentagens de rapazes, respectivamente: 40%, 20% e 10%. Um desses colégios é selecionado ao acaso e oito alunos são escolhidos, com reposição. Se o resultado for $RRRMMMMM$ (R para rapaz e M para moça), qual é a probabilidade de ter sido selecionado o colégio C?
12. Uma moeda é jogada 6 vezes. Sabendo-se que no primeiro lançamento deu coroa, calcular a probabilidade condicional de que o número de caras nos seis lançamentos supere o número de coroas.
13. (Bussab & Morettin, 2013, p. 126) Duas lâmpadas queimadas foram acidentalmente misturadas com seis lâmpadas boas. Se vamos testando as lâmpadas, uma por uma, até encontrar duas defeituosas, qual é a probabilidade de que a última defeituosa seja encontrada no quarto teste?

INSTITUTO FEDERAL
Goiás

Exercícios - Teorema de Bayes VI

14. (Ross, 2010, p. 67) Em um teste de múltipla escolha se o estudante não souber a resposta de uma questão ele marca aleatoriamente uma das alternativas. Suponha que cada questão tenha n alternativas e que a probabilidade do estudante saber a resposta de uma questão é p. Qual a probabilidade do estudante saber a resposta de uma questão se ele a respondeu corretamente? R.: $\dfrac{np}{1+(n-1)p}$

INSTITUTO FEDERAL
Goiás

Problema de Monty Hall - Selvin et al. (1975)

Num programa de TV, o objetivo é ganhar um carro como prêmio. O apresentador do programa mostra a você três portas, P_1, P_2 e P_3: atrás de uma porta há um carro e das demais não há nada. Ele pede a você para escolher uma porta, você escolhe P_1, mas essa não é aberta de imediato. Então, o apresentador abre a porta P_3 e ela está vazia (ele sabe onde está o carro!). Então ele pergunta se você quer mudar sua escolha. O que você faria?

Problema do Aniversário - Mckinney (1966)

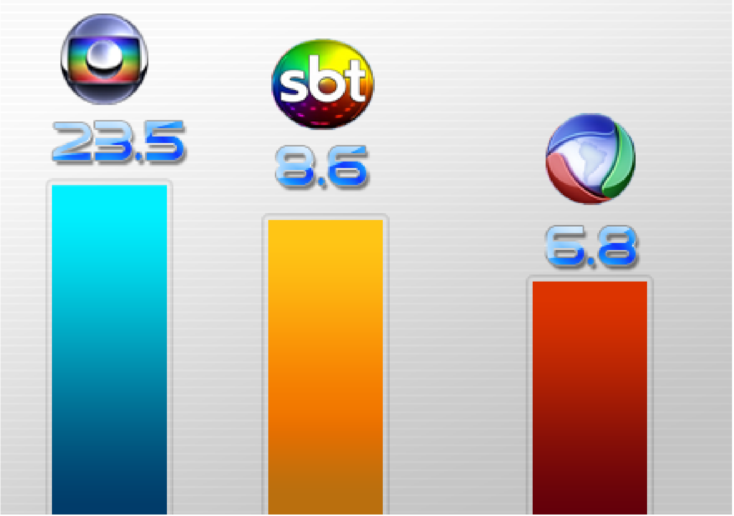
Considere k pessoas numa sala.

1. Qual a probabilidade de que pelo menos duas pessoas façam aniversário no mesmo dia e mês?
2. A partir de qual valor de k essa probabilidade é maior que 0,5?

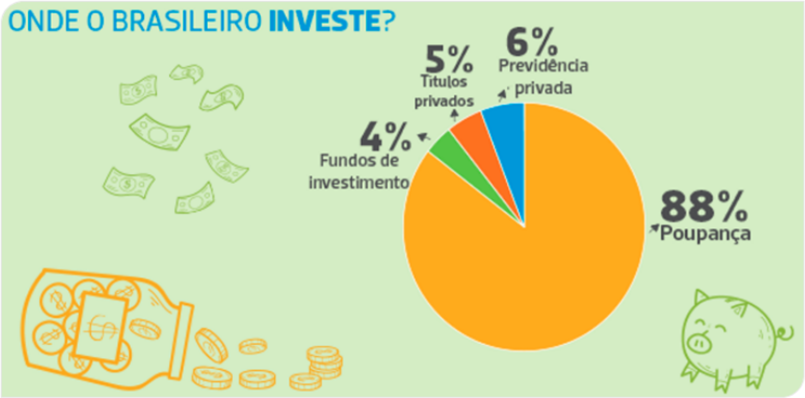
Introdução à Análise Exploratória de Dados

Gráficos muito loucos!!!

Gráficos muito loucos!!!



Gráficos muito loucos!!!



Anbima: Raio X do Investidor Brasileiro

Gráficos muito loucos!!!

Taxa de juros no Brasil


ENTRE 2002 E 2013,
A PRODUÇÃO DE DERIVADOS EM
NOSSAS REFINARIAS AUMENTOU
CERCA DE 30%.
1.641
1.639
1.696
1.735
1.755
1.788
1.780
1.821
1.832
1.896
1.997
2.124
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
PRODUÇÃO DE DERIVADOS BRASIL (MIL BPD)

INSTITUTO FEDERAL
Goiás

Software R - R (2018)

Para usar o R com a IDE RStudio basta fazer o download e instalar os seguintes programas na sequencia:

1. R
2. RStudio

R

`http://cran.at.r-project.org/`

RStudio

`http://www.rstudio.com/`

INSTITUTO FEDERAL Goiás

Distribuição de Frequências

Frequência (Absoluta) n_i

É o número de ocorrências de um valor particular da variável sob investigação.

Distribuição de Frequência

É a tabulação de todos os possíveis valores da variável em questão e suas respectivas frequências.

Frequência Relativa f_i

É a razão entre a frequência n_i e o total n de observações da variável sob investigação:

$$f_i = \frac{n_i}{n}$$

Classes de valor único

Exemplo

Considere o número de televisores em 50 casas aleatórias:

1	1	1	2	6
3	3	4	2	4
3	2	1	5	2
1	3	6	2	2
3	1	1	4	3
2	2	2	2	3
0	3	1	2	1
2	3	1	1	3
3	2	1	2	1
1	3	1	5	1

Em alguns casos o modo mais apropriado de se agrupar dados quantitativos discretos é usar classes que representem um único valor. Estas são chamadas classes de valor único.

	Frequências		
TVs	Absoluta	Relativa	Acumulada
0	1	0,02	0,02
1	16	0,32	0,34
2	14	0,28	0,62
3	12	0,24	0,86
4	3	0,06	0,92
5	2	0,04	0,96
6	2	0,04	1
	50	1	

INSTITUTO FEDERAL Goiás

```r
#Entre com os dados
TVs <- c(1,1,1,2,6,3,3,4,2,4,3,2,1,5,2,1,3,6,2,2,3,1,1,4,3,2,2,
         2,2,3,0,3,1,2,1,2,3,1,1,3,3,2,1,2,1,1,3,1,5,1)

#Frequencia Absoluta
ni <- table(TVs)

#Total de observacoes
n <- sum(ni)

#Frequencia Relativa
fi <- ni/n

#Frequencia Acumulada
fi.acum <- cumsum(fi)

#Constroi tabulacao dos resultados
tabela <- cbind(ni,fi,fi.acum)

#Exibe a tabela com todos os resultados
tabela
```

Gráfico de Barras

- Utiliza-se o plano cartesiano com os valores da variável no eixo das abscissas e as frequências no eixo das ordenadas.
- Para cada valor da variável desenha-se uma barra com altura correspondendo à sua frequência.
- Recomentado para variáveis:
  - Discretas
  - Qualitativas ordinais

INSTITUTO FEDERAL
Goiás

Gráfico de Barras (Freq. Absoluta)

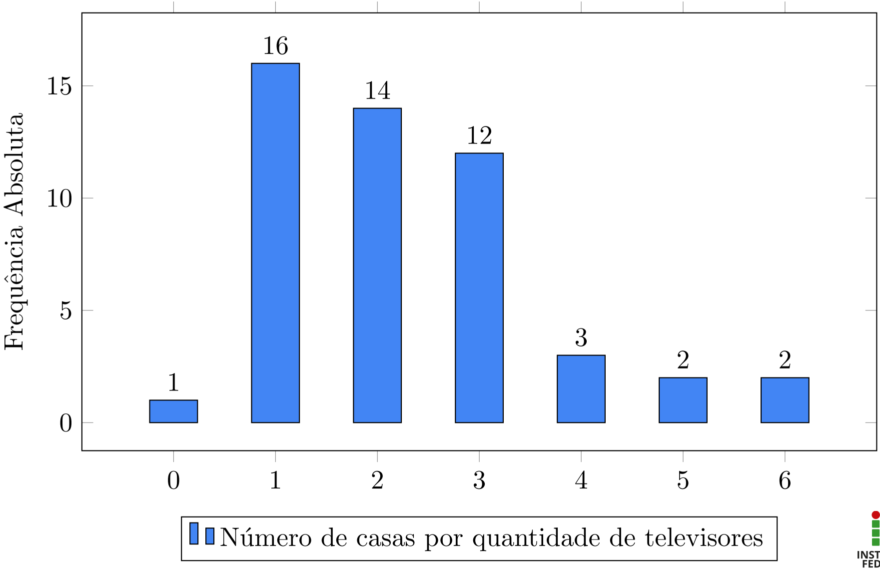


Gráfico de Barras usando o R

```r
TVs <- c(1,1,1,2,6,3,3,4,2,4,3,2,1,5,2,1,3,6,2,2,3,1,1,4,3,2,2,
         2,2,3,0,3,1,2,1,2,3,1,1,3,3,2,1,2,1,1,3,1,5,1)

#Frequência Absoluta
ni <- table(TVs)

#Gráficos de barras com frequências absolutas
barplot(ni)

#Total de observações
n <- sum(ni)

#Frequência Relativa
fi <- ni/n

#Gráficos de barras com frequências relativas
barplot(fi)
```

Gráfico de frequência acumulada vs relativa

```
TVs <- c(1,1,1,2,6,3,3,4,2,4,3,2,1,5,2,1,3,6,2,2,3,1,1,4,3,
         2,2,2,2,3,0,3,1,2,1,2,3,1,1,3,3,2,1,2,1,1,3,1,5,1)
ni <- table(TVs)
n <- sum(ni)
fi <- ni/n
fi.acum <- cumsum(fi)

#Gráfico de frequências acumuladas
barplot(fi.acum,
        col="tomato",
        main="Frequências",
        xlab = "Número␣de␣televisores")

#Gráfico de frequências relativas
barplot(fi,
        add=TRUE,
        col = "brown")

#Adiciona a Legenda
legend("topleft",
       legend = c("Acumulada", "Relativa"),
       fill = c("tomato", "brown"))
```

Gráfico de Setores (Pizza) - Diagrama de Disco

- Consiste em repartir o círculo em setores correspondentes às proporções (frequências relativas) de cada valor.
- Recomentado para variáveis qualitativas em geral.

Exercício - Spiegel & Stephens (2009, p. 74)

Ref.:o9Mi

Cinco moedas foram lançadas mil vezes e, em cada lance, foi anotado o número de caras. O número de lançamentos nos quais foram obtidas 0, 1, 2, 3, 4 e 5 caras estão indicados na tabela a seguir:

Número de caras	Lançamentos
0	38
1	144
2	342
3	287
4	164
5	25

1. Represente em um gráfico os dados da tabela;
2. Construa uma tabela mostrando as percentagens de lançamentos que resultaram em números de caras menores do que 0, 1, 2, 3, 4, 5 ou 6;
3. Represente a tabela construída no item 2 por meio de um gráfico adequado.

INSTITUTO FEDERAL
Goiás

Gráfico de Setores

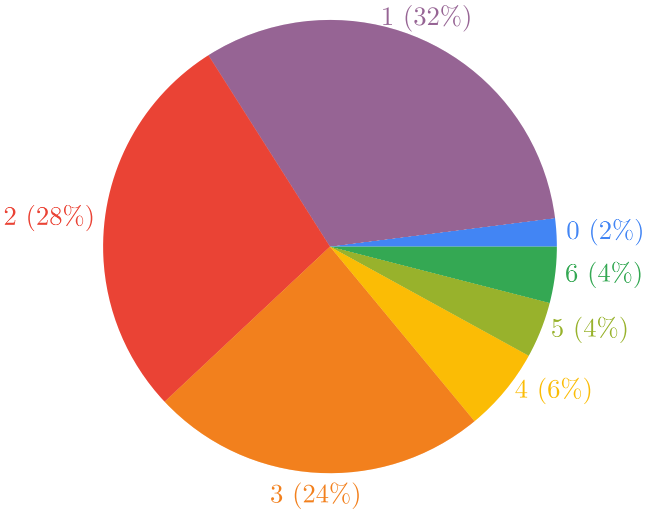
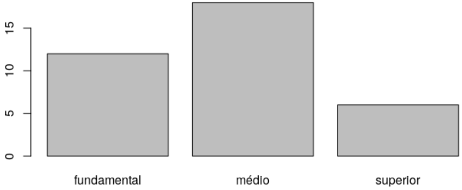


Gráfico de Setores usando o R

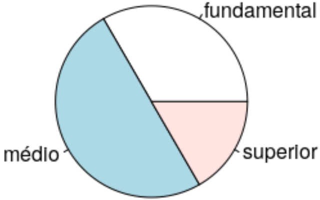
```
TVs <- c(1,1,1,2,6,3,3,4,2,4,3,2,1,5,2,1,3,6,2,2,3,1,1,4,3,
         2,2,2,2,3,0,3,1,2,1,2,3,1,1,3,3,2,1,2,1,1,3,1,5,1)

#Frequência Absoluta
ni <- table(TVs)

#Gráficos de Setores
pie(ni)
```

Considere a variável `Grau de Instrução` do conjunto de dados `CompanhiaMB` disponível no `q-acadêmico`. Uma tabulação e gráficos adequados para esses dados são:

Instrução	Freq.	Freq. Rel.
fundamental	12	0,3
médio	18	0,5
superior	6	0,2
	36	1



Distribuição do Grau de Instrução



Distribuição do Grau de Instrução

Exercício

Ref.:OZ55

Em uma pesquisa eleitoral para verificar a posição de três candidatos à prefeito de uma cidade, 2500 pessoas foram consultadas. Antônio recebeu 350 intensões de voto, o candidato Bento obteve 1200 declarações favoráveis e os demais entrevistados manifestaram preferência pelo atual prefeito Carlos. Construa uma tabela relacionando o total de votos que cada candidato recebeu. Qual candidato está à frente nas pesquisas?

INSTITUTO FEDERAL
Goiás

Exercício

Ref.:B9P7

Considere os conjuntos de dados `Questionário Estudantil` no `q-Acadêmico`. Construa gráficos de barra e setores para todas as variáveis quando possível e interprete-os.

Agrupando dados em intervalos de classe

Para dados quantitativos dividimos a faixa de variação dos dados em intervalos de classes. O menor valor da classe é denominado limite inferior (l_i) e o maior valor da classe é denominado limite superior (L_i).
O intervalo de classe pode ser representado das seguintes maneiras:

$l_i \vdash L_i$ onde o limite inferior da classe é incluído na contagem da frequência absoluta, mas o superior não;

$l_i \dashv L_i$ onde o limite superior da classe é incluído na contagem, mas o inferior não.

Na tabela de distribuição de frequência, acrescentamos uma coluna com os pontos médios de cada intervalo de classe, denotada por x_i. Esta é definida como a média dos limites da classe $x_i = \dfrac{l_i + L_i}{2}$. Estes valores são utilizados na construção de gráficos. Na medida do possível, as classes deverão ter amplitudes iguais.

intervalo de classe = intervalo de frequência

Exemplo

Peso (em libras) de 36 homens com idades entre 18 e 24 anos:

129,2	185,3	218,1
155,2	170,0	151,3
167,3	161,0	178,7
191,1	150,7	187,0
161,7	170,1	165,8
173,6	158,5	136,7
278,8	175,6	188,7
146,4	209,1	175,4
149,9	158,6	172,5
182,5	173,7	132,1
187,5	214,6	182,0
165,0	178,2	145,6

Uma tabulação adequada para esses dados é:

	Frequências		
Peso (lb)	Absoluta	Relativa	Acumulada
$120 \vdash 140$	3	0,083	0,083
$140 \vdash 160$	8	0,222	0,305
$160 \vdash 180$	14	0,388	0,693
$180 \vdash 200$	7	0,194	0,887
$200 \vdash 220$	3	0,083	0,972
$220 \vdash 240$	0	0	0,972
$240 \vdash 260$	0	0	0,972
$260 \vdash 280$	1	0,027	1
	36	1	

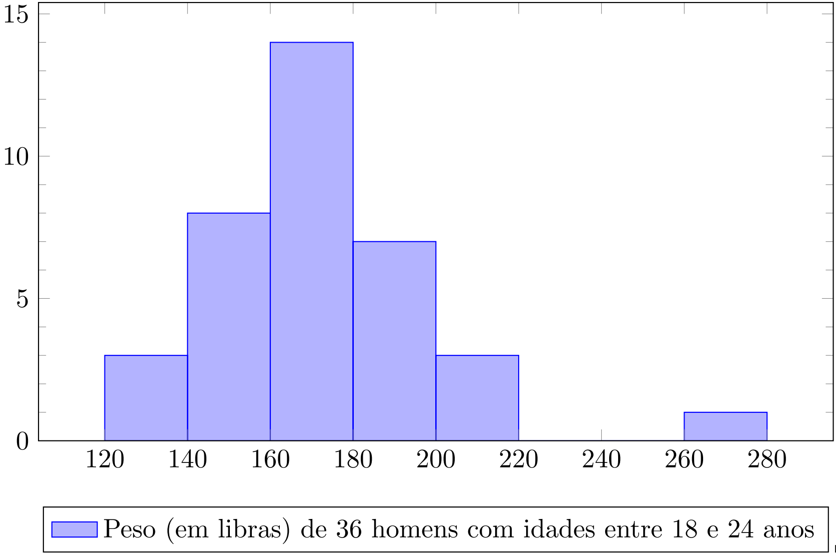
Desta forma os dados estão agrupados em intervalos de classe

INSTITUTO FEDERAL Goiás

Histograma

- Consiste de retângulos contíguos de base dada pelas faixas de valores da variável e área igual à frequência relativa da respectiva faixa.
- A altura de cada retângulo é denominada densidade de frequência e corresponde ao quociente de área pela amplitude da faixa.
- A área sobre histograma é sempre igual 1.
- Recomendado para variáveis
    - discretas;
    - em escala de razão;

Histograma



15
10
5
0
120
140
160
180
200
220
240
260
280
Peso (em libras) de 36 homens com idades entre 18 e 24 anos

Histograma usando o R

Vamos construir um histograma para comparar os preços dos veículos do conjunto de dados `Cars93` do pacote `MASS`.

```
library(MASS) #Carrega a biblioteca MASS

#Histograma simples
hist(Cars93$Price)

#Histograma simples com outros intervalos de classe
limites <- seq(from = 7,
               to = 62,
               length.out = 15)
hist(Cars93$Price,
     breaks = limites)

#Histograma personalizado
hist(Cars93$Price,
     main = "Preço␣dos␣veículos",
     xlab = "Preço",
     ylab = "Frequência",
     col = "brown")
```

INSTITUTO FEDERAL
Goiás

Histogramas sobrepostos usando o R

```
dados1 <- subset(CO2,Treatment=="nonchilled")
dados2 <- subset(CO2,Treatment=="chilled")

#Construção dos limites dos intervalos de classe
limitesIC <- seq(from = min(CO2$uptake),
                 to = max(CO2$uptake),length.out = 15)

#Histograma dos preços máximos
hist(dados1$uptake,
     main = "Emissão␣de␣dióxido␣de␣carbono␣sob␣exposição␣ao␣frio",
     breaks = limitesIC,
     xlab = "Taxa␣(umol/m^2␣sec)", ylab = "Frequências",
     col = rgb(1,0,0,0.5))

#Histograma dos preços mínimos
hist(dados2$uptake,
     breaks = limitesIC, add = TRUE,
     col = rgb(0,0,1,0.5))

#Adiciona a Legenda
legend("topleft",
       legend = c("Não␣exposta", "Exposta"),
       fill = c(rgb(1,0,0,0.5), rgb(0,0,1,0.5)))
```

INSTITUTO FEDERAL
Goiás

Exercício - Spiegel & Stephens (2009, p. 67)

Ref.:IKK6

A seguir são apresentados as alturas de 45 estudantes do sexo feminino de uma universidade com arredondamento para centímetros.

142	150	150	153	155
155	157	157	160	160
163	163	163	163	163
165	165	165	165	165
165	165	168	168	168
168	168	170	170	170
170	170	170	170	170
172	172	173	175	175
178	178	180	188	188

1. Estude a distribuição das alturas por meio de intervalos de classe;
2. Construa um histograma;

INSTITUTO FEDERAL
Goiás

Exercício - Bussab & Morettin (2013, p. 30)

Ref.:V9VN

Esboce o histograma alisado para cada uma das distribuições descritas:

1. salários registrados em carteira de trabalho dos brasileiros;
2. idades dos alunos do IFG;
3. número de óbitos por faixa etária;
4. número de divórcios segundo o numero de anos de casado;
5. número formado pelos dois últimos algarismos do primeiro prêmio da loteria federal, durante os dez últimos anos;
6. alturas dos brasileiros adultos;
7. alturas dos suecos adultos;
8. idades nas quais um grupo de mães tem seu primeiro filho;
9. alturas dos japoneses adultos.

INSTITUTO FEDERAL
Goiás

Exercício - Groebner et al. (2017, p 70)

Ref.:MZ3D

Considere os seguintes dados:

5	3	2	6	6
7	3	3	6	7
7	9	7	5	3
12	6	10	7	2
6	8	0	7	4

Construa uma tabela de frequências apresentando as frequências absolutas e relativas. Construa um gráfico para os dados.

INSTITUTO FEDERAL Goiás

Exercício - Bussab & Morettin (2013, p. 63)

Ref.:9JME

A distribuição de frequências do salário anual dos moradores do bairro A que têm alguma forma de rendimento é apresentada na tabela abaixo:

Faixa salarial (salários mínimos)	Frequência Absoluta
$0 \vdash 2$	10000
$2 \vdash 4$	3900
$4 \vdash 6$	2000
$6 \vdash 8$	1100
$8 \vdash 10$	800
$10 \vdash 12$	2000

1. Construa um histograma da distribuição.
2. Qual a "riqueza total" dos moradores do bairro?

Considere um conjunto de n realizações $x_1, \ldots, x_n$ de uma variável X. A média aritmética desse conjunto é definida como:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Os desvios da média são definidos como:

$$\mathrm{d}(x_i) = x_i - \overline{x}$$

- Representa o "centro de gravidade" da distribuição.
- Não precisa ser um valor contido no conjunto de realizações.
- É definida apenas para dados quantitativos.
- Muito sensível a valores extremos (outliers).
- Depende de todas as n realizações.
- Não informa o número de realizações acima ou abaixo da média.
- Não deve ser usada quando a distribuição é muito assimétrica.

INSTITUTO FEDERAL Goiás

Exercícios

Ref.:JN67

1. Considere um conjunto de n realizações $x_1, \ldots, x_n$ de uma variável X
   a. Prove que a soma dos desvios em relação à média aritmética é zero.
   b. Sejam $Y = X + a$ e $Z = aX$. Mostre que $\overline{y} = \overline{x} + a$ e $\overline{z} = a\overline{x}$.
   c. Mostre que o valor da expressão dos desvios quadráticos em relação a uma constante K

      $$\sum_{i=1}^{n}(x_i - K)^2$$

      é minimizado quando $K = \overline{x}$.
2. A média de 23 números é 14,7. Qual é a somatória total desses números? (Wilcox, 2009, p. 18)

INSTITUTO FEDERAL Goiás

Considere um conjunto de n realizações $x_1, \ldots, x_n$ de uma variável X. A média ponderada pelos fatores de ponderação (pesos) $p_1, \ldots, p_n$ desse conjunto é:

$$\overline{x}_P = \frac{\displaystyle\sum_{i=1}^{n} p_i x_i}{\displaystyle\sum_{i=1}^{n} p_i} = \frac{p_1 x_1 + \ldots + p_n x_n}{p_1 + \ldots + p_n}$$

- Os fatores de ponderação são números inteiros.
- Os fatores de ponderação podem ser interpretados como as frequências absolutas associadas a cada realização.
- Usada quando queremos/precisamos dar mais peso a algumas das realizações do que a outras.

Exercícios

Ref.:N59J

1. Considere um conjunto de n realizações $x_1, \ldots, x_n$ de uma variável X e os correspondentes fatores de ponderação (pesos) $p_1, \ldots, p_n$. Defina $f_j = \dfrac{p_j}{\displaystyle\sum_{i=1}^{n} p_i}$ e mostre que $\overline{x}_p = \displaystyle\sum_{i=1}^{n} f_i x_i = f_1 x_1 + \ldots + f_n x_n$. Como podemos interpretar f_j?
2. Três professores de estatística atribuíram notas médias de 8, 7 e 6 para suas respectivas turmas compostas de 23, 30 e 35 alunos cada. Qual é a nota média geral dos alunos?
3. Um psicólogo observou que 30% de seus pacientes chegam à 110 pontos em um teste de QI, 20% obtém 100 pontos e 50% atingem apenas 90 pontos. Qual é a pontuação média obtida pelos pacientes?
4. O salário anual médio pago aos empregados de uma indústria é de R\$100.000,00. Se os salários médios dos empregados do sexo masculino e feminino são, respectivamente, R\$120.000,00 e R\$30.000,00. Qual é a proporção de empregados de cada sexo?

Considere um conjunto de n realizações $x_1, \ldots, x_n$ de uma variável X, onde $x_i > 0$ para todo i inteiro. A média geométrica desse conjunto é:

$$\overline{x}_G = \left(\prod_{i=1}^{n} x_i \right)^{1/n} = \left(x_1 \cdot \ldots \cdot x_n \right)^{1/n}, \quad x_i > 0 \quad \forall i \in \mathbb{Z}^+$$

- Usada no cálculo de crescimento proporcional médio quando temos crescimentos proporcionais sucessivos.
- Não precisa ser um valor contido no conjunto de realizações.
- É definida apenas para dados quantitativos positivos.
- Depende de todas as n realizações.
- Menos sensível a valores extremos (superiores) que a média aritmética.
- Mais sensível a valores extremos (inferiores) que a média aritmética.
- Não informa o número de realizações acima ou abaixo da média.

INSTITUTO FEDERAL
Goiás

Exercícios

Ref.:N3C7

1. Considere um conjunto de n realizações $x_1, \ldots, x_n$ de uma variável X e um conjunto de m realizações $y_1, \ldots, y_n$ de uma variável Y, onde $x_i, y_i > 0$ para todo i inteiro. Seja $Z = {}^{X}/_{Y}$. Mostre que $z_G = {}^{x_G}/_{y_G}$.
2. Um investimento teve taxas de rendimentos anuais de 10%, 12% e 14% em três anos consecutivos. Qual foi o rendimento médio desse investimento nesses três anos?
3. Uma empresa teve um crescimento de 10%, 5% e -13% em três anos consecutivos. Qual foi o crescimento médio anual dessa empresa?
4. Donaldo tem uma carteira de ações da bolsa. No primeiro mês teve ganho de 5%, no segundo mês teve perdas que totalizaram 7%, no terceiro mês auferiu um lucro de 12% e, finalmente, no último mês terminou com prejuízo de 9%. Qual foi o ganho médio mensal dessa carteira?
5. Uma carteira de investimentos teve os seguintes retornos consecutivos 1%, -5%, 2%, 3%, 4% e 35%. Qual o retorno médio por período da carteira?
6. Se o preço de uma mercadoria dobra num período de 4 anos, qual o aumento médio percentual por ano?
7. Determine dois números cuja média aritmética é 9 e geométrica é 7.

INSTITUTO FEDERAL
Goiás

Considere um conjunto de n realizações $x_1, \ldots, x_n$ de uma variável X, onde $x_i > 0$ para todo i inteiro. A média harmônica desse conjunto é:

$$\overline{x}_H = \frac{n}{\displaystyle\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\dfrac{1}{x_1} + \cdots + \dfrac{1}{x_n}}, \quad x_i > 0 \quad \forall i \in \mathbb{Z}^+$$

- Usada no cálculo de médias de taxas médias.
- Não precisa ser um valor contido no conjunto de realizações.
- É definida apenas para dados quantitativos positivos.
- Depende de todas as n realizações.
- Menos sensível a valores extremos (superiores) que as médias aritmética e geométrica.
- Mais sensível a valores extremos (inferiores) que as médias aritmética e geométrica.
- Não informa o número de realizações acima ou abaixo da média.

INSTITUTO FEDERAL
Goiás

1. Um aluno vai de casa ao colégio, com uma velocidade média de 30km/h e na volta sua velocidade média cai para 25km/h. Qual foi sua velocidade média considerando ida e volta?
2. Um automóvel sobe uma ladeira com uma velocidade média de 30km/h e, em seguida, desce essa mesma ladeira com uma velocidade de 50km/h. Qual foi a velocidade média desse veículo ao longo de todo esse percurso?
3. Um caminhão, que inicialmente estava andando à uma velocidade v, aumenta sua velocidade para $2v$ com uma aceleração média de 20km/h^2. Em seguida acelera para $3v$ agora com aceleração média de 40km/h^2. Qual foi a aceleração média nesse percurso?
4. As cidades A, B e C são equidistantes uma da outra. Um motorista viaja de A para B à 40km/h, depois viaja de B para C à 32km/h e, finalmente, viaja de C para A à 23km/h. Qual foi a velocidade média de toda a viagem?

Considere um conjunto de n realizações $x_1, \ldots, x_n$ de uma variável X. A média quadrática desse conjunto é definida como:

$$\overline{x}_Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} = \sqrt{\frac{x_1^2 + \ldots + x_n^2}{n}}$$

- Não precisa ser um valor contido no conjunto de realizações.
- Sempre será positiva.
- É definida apenas para dados quantitativos.
- Mais sensível a valores extremos que a média aritmética.
- Depende de todas as n realizações.
- Não informa o número de realizações acima ou abaixo da média.
- Não deve ser usada quando a distribuição é muito assimétrica.

INSTITUTO FEDERAL
Goiás

1. Determine a média quadrática dos números 1,5,8,4,7,1,5,6.
2. Determine a média quadrática dos números 1,5,8,4,7,1,5,100.
3. Dados dois números a e b positivos, mostre que a sua média quadrática é sempre maior ou igual a média aritmética que é sempre maior ou igual que a média geométrica que é sempre maior ou igual que a média harmônica, ou seja:

$$\sqrt{\frac{a^2+b^2}{2}} \geq \frac{a+b}{2} \geq \sqrt{ab} \geq \frac{2}{\dfrac{1}{a}+\dfrac{1}{b}}.$$

Desafios
Ref.:NJ6X

Assuma a desigualdade entre as médias:

$$\overline{x}_Q \geq \overline{x} \geq \overline{x}_G \geq \overline{x}_H.$$

Usando a desigualdade das médias mostre

1. Qual é o valor possível para $\sin(x) + \cos(x)$?
2. Uma caixa de base quadrada (sem tampa) é construída com 1200 cm^2 de papelão. Calcule seu volume máximo e indique as dimensões da caixa.
3. Sejam p e q números reais positivos tais que $\dfrac{1}{p} + \dfrac{1}{q} = \dfrac{1}{100}$. Qual é o valor mínimo do produto pq?

```
library(ggplot2)
dados1 <- rnorm(n = 300,mean = 0,sd = 1)
dados2 <- rnorm(n = 300,mean = 2,sd = 1)
dados3 <- rnorm(n = 300,mean = 4,sd = 1)
dados <- data.frame(dados = c(dados1,dados2,dados3),
                    grupo = rep(1:3,each = 300))

ggplot(dados,aes(x=dados)) +
  geom_histogram(data=subset(dados,grupo == 1),
                 fill = "red",
                 alpha = 0.2) +
  geom_histogram(data=subset(dados,grupo == 2),
                 fill = "blue",
                 alpha = 0.2) +
  geom_histogram(data=subset(dados,grupo == 3),
                 fill = "green",
                 alpha = 0.2)

mean(dados1)
mean(dados2)
mean(dados3)
```

Sejam $x_1, \ldots, x_n$ realizações da variável X.

Moda

É a realização (ou realizações) mais frequente do conjunto de valores observados. Notação.: $\text{mo}(X)$

Mediana

$$\text{me}(X) = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{se } n \text{ ímpar} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{se } n \text{ par} \end{cases}$$

Exercício - Spiegel & Stephens (2009, p. 84)

Ref.:SIB6

Determine moda e mediada dos seguintes conjuntos de dados:

Conjunto A

$$2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18$$

Conjunto B

$$3, 5, 8, 10, 12, 15, 16$$

Conjunto C

$$2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9$$

Exercício - Wilcox (2009, p. 19)

Ref.:38PC

Considere os dez valores

3	6	8	12	23
26	37	42	49	63

A média é $\bar{x} = 26,9$.

1. Quais serão os novos valores da média se o maior valor 63 for aumentado para 100, 1000 e 10000?
2. Faça o mesmo que no item anterior para a mediana. Que conclusão você pode tirar comparando os resultados obtidos para as duas medidas?

INSTITUTO FEDERAL Goiás

Exercício - Bussab & Morettin (2013, p. 43)

Ref.:RJS3

Numa pesquisa realizada com 100 famílias, levantamos as informações tabuladas abaixo.

Número de filhos	0	1	2	3	4	5	mais que 5
Frequência de famílias	17	20	28	19	7	4	5

1. Qual a mediana do número de filhos?
2. E a moda?
3. Que problema você enfrentaria para calcular a média? Faça alguma suposição e encontre-a.

```
# Reconstrói o gráfico de histograma sobrepostos
source(file = "ExemploHistogramaSobrepostos2.R")

# Mediana
median(dados1$uptake)
median(dados2$uptake)

library(MASS) #Biblioteca para acessar outros conjunto de dados
#Moda usanda a tabela de frequência
tabela <- table(Cars93$Type)
tabela
subset(tabela,tabela==max(tabela))

table(Cars93$Manufacturer)
table(Cars93$AirBags)
table(Cars93$Cylinders)

table(UScereal$mfr)
table(UScereal$vitamins)

table(Aids2$state)
table(Aids2$T.categ)
```

Estatísticas de ordem

São as observações ordenadas:

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq, \ldots, \leq x_{(n-1)} \leq x_{(n)}$$

$\max(x) = x_{(n)}$ e $\min(x) = x_{(1)}$.

Quartis

Os quartis q_1, q_2 e q_3 são números tais que:

- 1° **Quartil:** 25% dos dados são menores que ou iguais a q_1.
- 2° **Quartil:** 50% dos dados são menores que ou iguais a q_2 (A mediana).
- 3° **Quartil:** 75% dos dados são menores que ou iguais a q_3.

INSTITUTO FEDERAL
Goiás

1° Quartil

É a mediana das 50% menores estatísticas de ordem.

2° Quartil

É a própria mediana.

3° Quartil

É a mediana das 50% maiores estatísticas de ordem.

Exercício - Spiegel & Stephens (2009, p. 87)

Ref.:7CX5

Considere a seguinte tabulação de notas obtidas pelos 60 alunos de uma turma:

25	28	28	28	29	30	32	33	33	33	34	34	35	36	37
38	41	42	42	45	46	47	51	51	53	53	53	55	56	57
57	60	61	62	62	62	67	68	69	71	72	73	73	75	75
79	82	85	86	86	86	88	88	89	91	93	94	96	96	99

Determine o seus quartis.

INSTITUTO FEDERAL
Goiás

Exercício

Ref.:4T4U

Determine a média aritmética, média geométrica, média harmônica, moda, mediana e os quartis para as variáveis `Preço`, `Comprimento` e `Motor` do conjunto de dados `Veículos`. Faça gráficos adequados e interprete-os.

```
#Mínimo e Máximo
min(morley$Speed)
max(morley$Speed)

# Estatísticas de ordem
sort(Cars93$Min.Price)
sort(Cars93$Max.Price)
sort(Nile)

#Não faz sentido calcular estatísticas de ordem para variáveis
#qualitativas nominais!!!
sort(Cars93$Type)
sort(Cars93$Manufacturer)

# Quartis Amostrais
quantile(Cars93$Min.Price)
quantile(Cars93$Max.Price)
quantile(morley$Speed)
```

INSTITUTO FEDERAL Goiás

Quantil de ordem p

É um número $q(p)$, $p \in (0, 1)$, tal que $100p\%$ das observações sejam menores que $q(p)$.

Média Aparada

A média aparada $\overline{x}(p)$ é a média aritmética simples calculada excluindo-se as $100p\%$ menores e as $100p\%$ maiores observações.

O quantil de ordem p, $q(p)$, com $0 < p < 1$, é calculado pela função:

$$q(p) = \begin{cases} x_{(i)} & \text{se } p = p_i = \dfrac{i - 0.5}{n}, i = 1, 2, \ldots, n \\ (1 - f_i)x_{(i)} + f_i x_{(i+1)} & \text{se } p_i < p < p_{i+1} \\ x_{(1)}, & \text{se } p < p_1 \\ x_{(n)}, & \text{se } p > p_n \end{cases}$$

onde:

- $f_i = \dfrac{p - p_i}{p_{i+1} - p_i}$
- $x_{(i)}$ representa a i-ésima estatística de ordem
- p_i representa o p-quantil que corresponde à $x_{(i)}$

Obs.: Há várias outras metodologias para calcular o p-quantil.

Considere a seguinte tabulação de notas obtidas pelos 60 alunos de uma turma:

25	28	28	28	29	30	32	33	33	33	34	34	35	36	37
38	41	42	42	45	46	47	51	51	53	53	53	55	56	57
57	60	61	62	62	62	67	68	69	71	72	73	73	75	75
79	82	85	86	86	86	88	88	89	91	93	94	96	96	99

Determine o 2°, 4°, 6° e 8° decis e calcule a média aparada em 20%.

Decis

Decil	Notação	Quantil
1°	d_1	$q(10\%)$
2°	d_2	$q(20\%)$
$\vdots$	$\vdots$	$\vdots$
9°	d_9	$q(90\%)$
10°	d_{10}	$q(100\%)$

```
# Quartis Amostrais
quantile(Cars93$Min.Price)
quantile(Cars93$Max.Price)

# Decis Amostrais
ordens <- seq(from = 0, to = 1, length.out = 11)
ordens
quantile(Cars93$Min.Price,probs = ordens)
quantile(Cars93$Max.Price,probs = ordens)

# Percentis Amostrais
ordens <- c(0.12,0.36,0.81)
ordens
quantile(Cars93$Min.Price,probs = ordens)
quantile(Cars93$Max.Price,probs = ordens)

#Média Aparada
mean(Cars93$Min.Price,trim = 0.25)
mean(Cars93$Max.Price,trim = 0.25)
```

Exercício - (Wilcox, 2009, p. 29)
Ref.:FHY7

Considere as seguintes taxas de pulsação cardíaca de 21 adultos:

80	85	81	75	76	79	74
86	79	55	82	89	73	79
83	82	88	79	77	81	82

Calcule a meia-média (média aparada a 25%). Dado $p \in (0,1)$ qual a proporção dos dados precisa ser modificada para que o valor da média aparada a $100p\%$ seja alterada em relação à média aritmética convencional?

Medidas de Dispersão

Sejam $x_1, \ldots, x_n$ realizações da variável X

Desvio médio

$$\mathrm{dm}(X) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}|$$

Variância populacional

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Variância amostral

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Desvio padrão populacional

$$\sigma_X = \sqrt{\sigma_X^2}$$

Desvio padrão amostral

$$S_X = \sqrt{S_X^2}$$

O desvio padrão populacional é a média quadrática dos desvios da média.

Medidas de Dispersão

Sejam $x_1, \ldots, x_n$ realizações da variável X

Amplitude

$$A_X = x_{(n)} - x_{(1)}$$

Distância Interquartil

$$d_q = q_3 - q_1$$

Dispersões

Inferior $d_{inf} = q_2 - x_{(1)}$

Superior $d_{sup} = x_{(n)} - q_2$

Coeficiente de variação

$$CV_X = \frac{S_X}{\overline{x}}$$

O coeficiente de variação só deve ser usado se a variável for medida na escala de razão.

Exercício

Ref.:GN68

Foram observados três grupos de viajantes, cada grupo participou de um pacote turístico diferente. Para cada grupo anotou-se a avaliação (0 à 10) que cada integrante fez da sua experiência. As avaliações estão na tabulação a seguir. Compare os grupos usando média, mediana, desvio padrão, amplitude, distância interquartil e o coeficiente de variação. O que você conclui?

Grupo 1	Grupo 2	Grupo 3
0	5	3
1	5	3
2	5	4
3	5	4
4	5	5
5	5	5
6	5	5
7	5	6
8	5	6
9	5	7
10	5	7

INSTITUTO FEDERAL
Goiás

Exercício - (Bussab & Morettin, 2013, p. 62)

Ref.:SNS1

Na companhia A, a média dos salários é R$10.000,00 e o 3^o quartil é R$5.000,00.

1. Se você se apresentasse como candidato a funcionário nessa firma e se o seu salário fosse escolhido ao acaso entre todos os possíveis salários, o que seria mais provável: ganhar mais ou menos que R$5.000,00? Justifique.
2. Suponha que na companhia B, a média dos salários seja de R$7.000,00, a variância praticamente zero e o salário também seja escolhido ao acaso. Em qual companhia você se apresentaria para procurar emprego? Justifique.

Exercício - (Groebner et al., 2017, p. 128)

Ref.:RR5F

Foi solicitado ao registro escolar preparar um relatório sobre os estudantes de graduação em uma determinada universidade. Como a o total de alunos matriculados é muito grande optou-se por coletar os dados de uma amostra de 10 estudantes. Foram coletadas várias informações dentre elas a idade dos entrevistados: 32, 22, 24, 27, 27, 33, 28, 23, 24, 21.

1. Calcule a amplitude, o intervalo interquartil e o desvio padrão para as idades.
2. Sabe-se que a idade média dos estudantes de graduação nessa universidade é de 37,8 anos. Você acha que esses 10 entrevistados formam uma amostra representativa dos alunos? Com base em qual estatística você concluiu isso?

INSTITUTO FEDERAL Goiás

Exercício - Bussab & Morettin (2013, p. 63)

Ref.:QO5H

A distribuição de frequências do salário anual dos moradores do bairro A que têm alguma forma de rendimento é apresentada na tabela abaixo:

Faixa salarial (salários mínimos)	Frequência Absoluta
$0 \vdash 2$	10000
$2 \vdash 4$	3900
$4 \vdash 6$	2000
$6 \vdash 8$	1100
$8 \vdash 10$	800
$10 \vdash 12$	2000

1. Qual a média e o desvio padrão amostral da variável salário?
2. O bairro B apresenta, para a mesma variável, uma média de 7,2 e um desvio padrão amostral de 15,1. Em qual dos bairros a população é mais homogênea quanto à renda?
3. Construa a coluna da distribuição acumulada e determine qual a faixa salarial dos 10% mais ricos da população do bairro.

INSTITUTO FEDERAL
Goiás

Exercício - Bussab & Morettin (2013, p. 62)

Ref.:UWN6

O número de divórcios na cidade, de acordo com a duração do casamento, está representado na tabela abaixo:

Anos de casamento	Nº de divórcios
$0 \vdash 6$	2800
$6 \vdash 12$	1400
$12 \vdash 18$	600
$18 \vdash 24$	150
$24 \vdash 30$	50

1. Qual a duração média dos casamentos? E a mediana?
2. Encontre a variância e o desvio padrão da duração dos casamentos.
3. Construa o histograma da distribuição.
4. Encontre o 1º e o 3º quartis. Qual o intervalo interquartil?

INSTITUTO FEDERAL
Goiás

Exercício - Bussab & Morettin (2013, p. 62)

Ref.:67JR

O que acontece com a mediana, a média e o desvio padrão de uma série de dados quando:

1. cada observação é multiplicada por 2?
2. soma-se 10 a cada observação?
3. subtrai-se a média geral $\bar{x}$ de cada observação?
4. de cada observação subtrai-se a média geral $\bar{x}$ de cada observação e divide-se pelo desvio padrão S_X?

Medidas de Dispersão usando o R

```r
# Reconstrói o gráfico de histograma sobrepostos
source(file = "ExemploHistogramaSobrepostos1.R")

MINP <- Cars93$Min.Price
MAXP <- Cars93$Max.Price

#Desvios da média
MINP - mean(MINP)

#Desvio Médio (Absoluto)
mad(MINP,center = mean(MINP))
mad(MAXP,center = mean(MAXP))

#Desvio Padrão Amostal
sd(MINP)
sd(MAXP)

#Variância Amostral
var(MINP)
var(MAXP)
```

INSTITUTO FEDERAL
Goiás

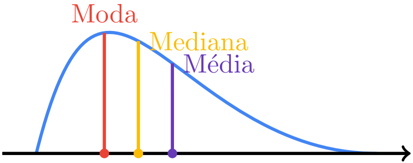
Medidas de Dispersão usando o R

```r
# Reconstrói o gráfico de histograma sobrepostos
source(file = "ExemploHistogramaSobrepostos2.R")

dados1 <- subset(CO2,Treatment=="nonchilled")
dados2 <- subset(CO2,Treatment=="chilled")

#Desvios da média
dados1$uptake - mean(dados1$uptake)

#Desvio Médio (Absoluto)
mad(dados1$uptake,center = mean(dados1$uptake))
mad(dados2$uptake,center = mean(dados2$uptake))

#Desvio Padrão Amostal
sd(dados1$uptake)
sd(dados2$uptake)

#Variância Amostral
var(dados1$uptake)
var(dados2$uptake)
```

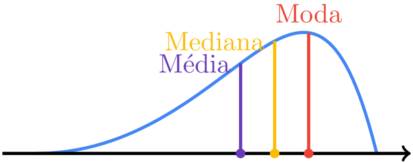
INSTITUTO FEDERAL
Goiás

Assimetria

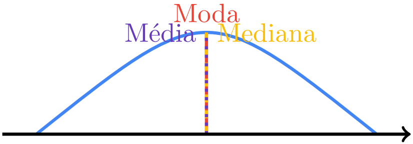
Uma curva é simétrica quando existe uma repartição precisa de valores em torno do ponto central, ou seja, as medidas de tendência central (média, mediana e moda) coincidem.



Assimetria à direita (positiva)



Assimetria à esquerda (negativa)



Distribuição simétria

Medida de Assimetria Amostral

$$A = \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left(\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{3/2}}$$

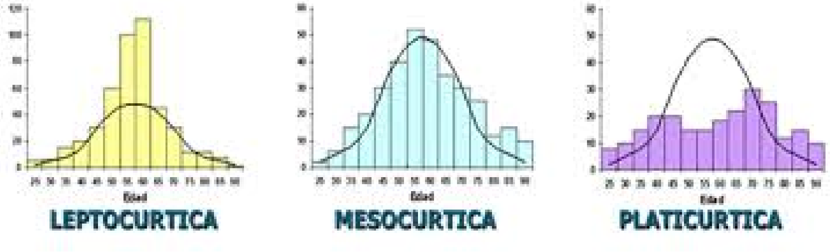
$A = 0$**:** dados simétricos

$A > 0$**:** assimetria positiva (à direita)

$A < 0$**:** assimetria negativa (à esquerda)

Curtose ou Achatamento

A curtose mede grau de “achatamento” da distribuição ou curva de frequência considerada em relação a uma curva normal.



Medida de Curtose Amostral

$$C = \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2} - 3$$

$C = 0$: Mesocúrtica (Curva Normal)

$C > 0$: Platicúrtica

$C < 0$: Leptocúrtica

As taxas de juros recebidas por uma amostra de 25 ações durante certo período foram (medidas em porcentagem):

2,50	2,52	2,53	2,54	2,59
2,60	2,61	2,62	2,64	2,65
2,66	2,68	2,69	2,70	2,72
2,75	2,77	2,80	2,81	2,82
2,83	2,85	2,88	2,89	2,91

Calcule a assimetria e a curtose para esses dados. Classifique a distribuição dos dados a respeito da assimetria e curtose. O que você pode concluir a partir disso?

Para facilitar um projeto de ampliação de rede de esgoto de certa região de Goiânia, as autoridades tomaram uma amostra de tamanho 30 dos 250 quarteirões que compõem a região, e foram encontrados as seguintes quantidade de casas por quarteirão (dados em rol):

10	12	13	13	14	15	15	16	18	18
20	21	22	23	24	25	26	27	28	29
30	31	32	36	40	40	42	42	57	60

Classifique a distribuição dos dados a respeito da assimetria e curtose.

Exercício

Ref.:67JR

Considere o conjunto de dados `Veículos`. Compare os veículos nacionais e importados por meio das assimetrias das variáveis `Preço`, `Comprimento` e `Motor`. Interprete o significados dos valores obtidos e compare-os com as médias e medianas.

Exercício

Ref.:M35Z

Um professor ministra aulas de Karatê para quatro turmas diferentes. Para compreender melhor o perfil de seus alunos ele aplica um questionário aos seus alunos. Uma das questões foi: "Quantas vezes você já sofreu alguma fratura óssea?". As respostas dos alunos nas seis turmas foram as seguintes:

$$\begin{aligned} T_1 &= (0,0,0,1,1,1,2,2,2) \\ T_2 &= (0,0,0,0,0,0,1,1,7) \\ T_3 &= (0,0,0,0,0,0,1,2,3,4) \\ T_4 &= (0,0,0,0,2,2,2,2) \end{aligned}$$

1. Calcule a média, a moda e a mediana, a variância, o desvio padrão, o desvio médio, a amplitude, o mínimo e o máximo para todas as turmas.
2. É possível distinguir as turmas pela média aritmética?
3. Qual das turmas apresentou respostas mais uniformes? Qual apresenta maior contraste entre os alunos com relação a questão acima? Quais estatísticas dão suporte ao seu argumento?
4. Há alguma turma que apresente simetria na distribuição dos dados?

Exercício

Ref.:BH51

Considere duas variáveis $X = \{x_1, \ldots, x_n\}$ e $Y = \{y_1, \ldots, y_m\}$. Esboce os histogramas alisados para X e Y na mesma escala sabendo que:

1. $S_X > S_Y$ e $\overline{x} < \overline{y}$.
2. A média de X é o dobro da média de Y e as variâncias são aproximadamente iguais e próximas de zero.
3. X é platicúrtica, Y é leptocúrtica e suas médias e coeficientes de assimetria são iguais.
4. X é assimétrica à direita, Y é assimétrica a esquerda e além disso $x_{(n)} < y_{(1)}$.

INSTITUTO FEDERAL
Goiás

```
library(MASS)
library(cumstats)
dados <- Cars93$Horsepower

hist(dados)
skewness(dados)
kurtosis(dados)

dados <- rnorm(n = 100,mean = 0,sd = 1)
skewness(dados)
skewness
```

Exercício

Ref.:5BQR

Considere as variáveis a seguir:

$$
\begin{aligned}
A &= \{1,2,3,4,5,6,7,8,9,10\} \\
B &= \{2,2,4,4,6,6,8,8,10,10\} \\
C &= \{1,4,5,6,9\} \\
D &= \{1,1,2,5,8,9,10\} \\
E &= \{1,1,1,5,5,9\} \\
F &= \{5,5,5,5,5,5,5\} \\
G &= \{1,1,1,1,1,3,3,35\} \\
H &= \{3,3,3,4,5,5,5,35\}
\end{aligned}
$$

Calcule para esses conjuntos de dados os valores de todas as estatísticas estudadas. Preencha a tabulação ao lado e compare os resultados.

	A	B	...
Média Aritmética			
Média Geométrica			
Média Harmônica			
1° Quartil			
Mediana			
3° Quartil			
Moda			
Mínimo			
Máximo			
Desvio Médio			
Variância (σ^2)			
Variância (S^2)			
Desvio Padrão (σ)			
Desvio Padrão (S)			
Amplitude			
$\vdots$			

INSTITUTO FEDERAL
Goiás

Boxplot

- Se trata de um retângulo em que a aresta inferior coincide com o primeiro quartil e a superior, com o terceiro quartil.
- A mediana é representada por um traço no interior do retângulo.
- Segmentos de reta (bigodes) são incluídos no boxplot partindo-se do primeiro e terceiro quartis e terminando nos extremos inferior e superior como definidos no slide seguinte.
- As observações que ficam além do intervalo compreendido pelos extremos inferior e superior são considerados valores extremos (aberrantes ou outliers) e são representadas por pontos acima ou abaixo dos extremos.
- Recomendado para dados quantitativos em geral.

Boxplot

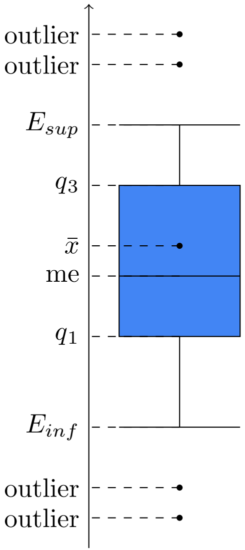
Limites

$$L_{inf} = q_1 - \frac{3d_q}{2}$$
$$L_{sup} = q_3 + \frac{3d_q}{2}$$

Valores Extremos

Observações fora do intervalo $[L_{inf}, L_{sup}]$.

1. Extremo inferior (E_{inf}): É a menor observação maior que L_{inf}.
2. Extremo superior (E_{sup}): É a maior observação menor que L_{sup}.



Exemplo

Ref.:BMS2

Considere a tabulação a seguir com as idades de 38 engenheiros:

14	31	38	39	39	42	42	45	47	48
48	48	52	52	53	54	55	57	59	60
61	64	64	66	66	67	68	68	69	71
71	74	75	77	79	79	79	108		

1. Determine os quartis para os dados;
2. Obtenha o intervalo inter-quartil;
3. Determine o máximo e o mínimo;
4. Calcule os limites superior e inferior para os dados;
5. Identifique potenciais valores extremos;
6. Construa e interprete um boxplot para esses dados.

Exercício - Spiegel & Stephens (2009, p. 67)

Ref.:J2UV

A seguir são apresentados as alturas de 45 estudantes do sexo feminino de uma universidade com arredondamento para centímetros.

142	150	150	153	155
155	157	157	160	160
163	163	163	163	163
165	165	165	165	165
165	165	168	168	168
168	168	170	170	170
170	170	170	170	170
172	172	173	175	175
178	178	180	188	188

1. Encontre os quartis dos dados usando algum dos métodos estudados;
2. Construa um boxplot.

INSTITUTO FEDERAL Goiás

Exercício - Groebner et al. (2017, p. 116)

Ref.:NH56

Considere as seguintes pontuações em um processo seletivo para ingresso em uma Universidade.

23	65	45	19	35	28	39	100	50	26	25	27
24	17	12	106	23	19	39	70	20	18	44	31

1. Calcule os quartis
2. Esboce o boxplot dos dados
3. Calcule a média e a mediana. Com base nos valores obtidos o que você pode afirmar sobre a assimetria dos dados?
4. Construa um histograma para esses dados.

Exercício - Magalhães & Lima (2015, p. 19)

Ref.:DIM9

Suponha que um produtor de laranjas guarde as frutas colhidas de uma árvore em caixas separadas e está interessado em estudar o número de laranjas por caixa (ou por árvore). Após um dia de colheita, 20 caixas foram contadas e os resultados brutos, após a ordenação, são: 22, 29, 33, 35, 35, 37, 38, 43, 43, 44, 48, 48, 52, 53, 55, 57, 61, 62, 67 e 69. Esboce um boxplot para esses dados.

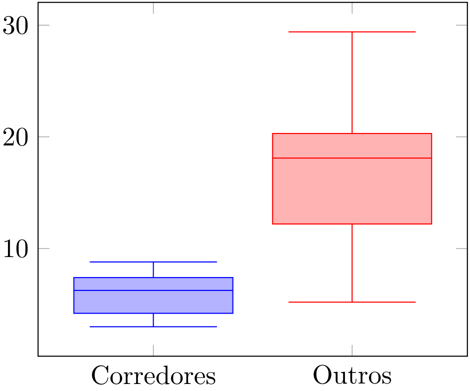
Uma pesquisa mede a espessura da pele de dois grupos de pessoas. O primeiro grupo era de corredores o outro não. Que conclusão podemos tirar?

Corredores

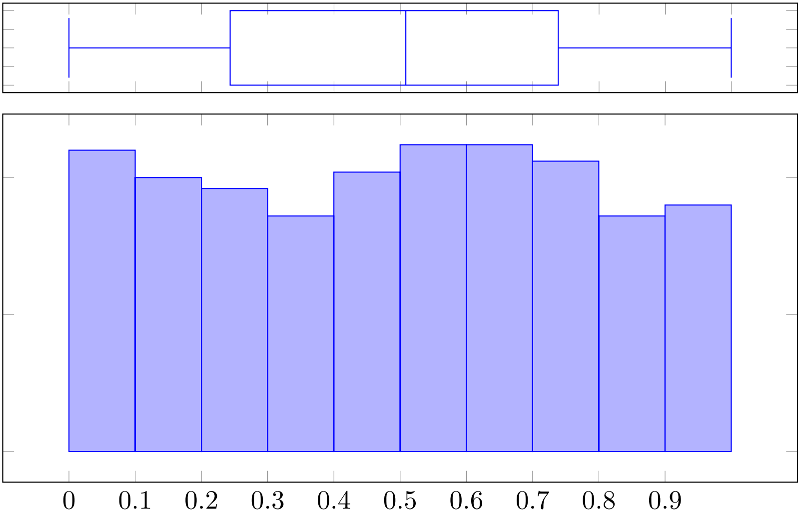
7.3	6.7	8.7
3.0	5.1	8.8
7.8	3.8	6.2
5.4	6.4	6.3
3.7	7.5	4.6

Outros

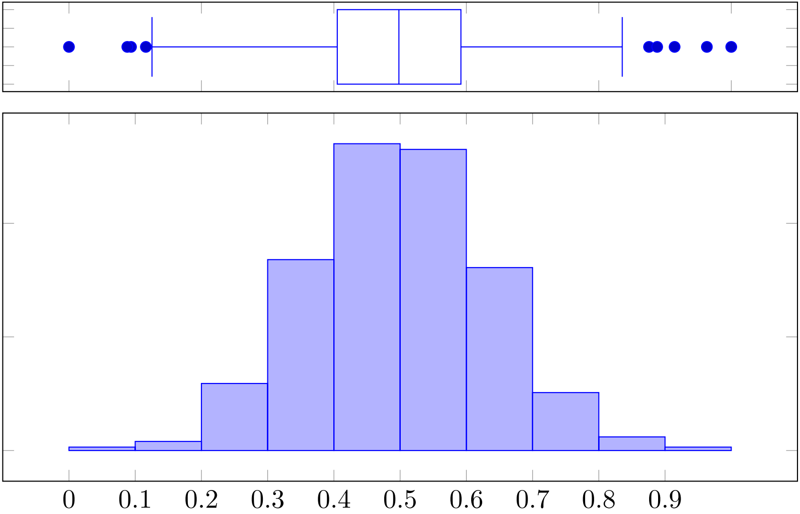
24	19.9	7.5	18.4
28	29.4	20.3	19
9.3	18.1	22.8	24.2
9.6	19.4	16.3	16.3
12.4	5.2	12.2	15.6



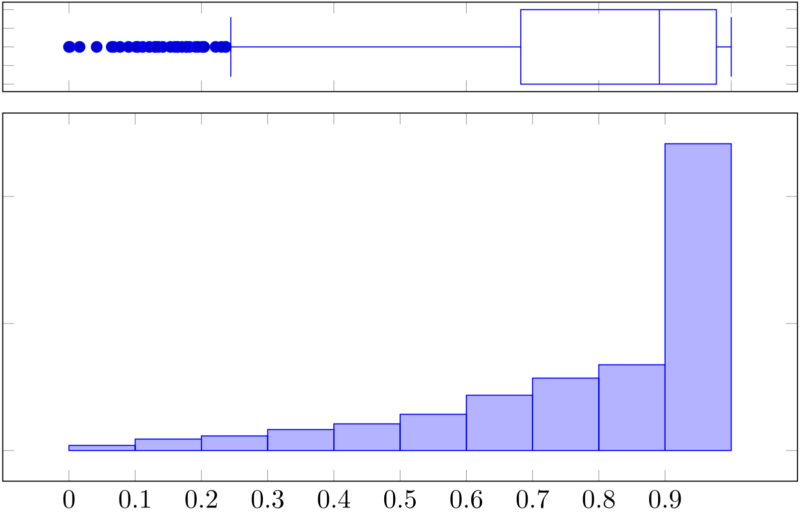
Dados com distribuição aproximadamente uniforme



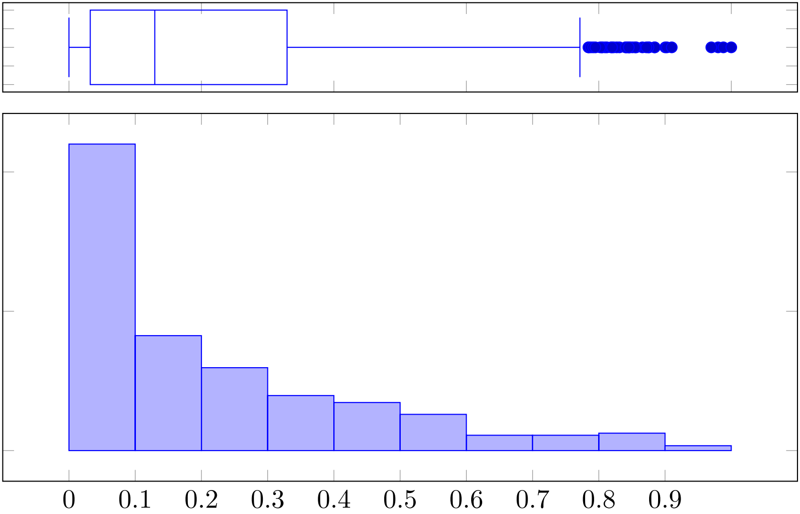
Dados com distribuição aproximadamente simétrica



Dados com distribuição assimétrica à esquerda (negativa)



Dados com distribuição assimétrica à direita (positiva)



Exercício

Ref.:HT6J

As taxas médias de incremento anual (por 100 habitantes) da população dos 30 maiores municípios do Brasil estão dadas abaixo:

3.67	1.82	3.73	4.10	4.30	1.28	8.14	2.43	4.17	5.36
3.96	6.54	5.84	7.35	3.63	2.93	2.82	8.45	5.28	5.41
7.77	4.65	1.88	2.12	4.26	2.78	5.54	0.90	5.09	4.07

1. Construa um histograma.
2. Determine os quartis (q_1, q_2, q_3), o máximo ($x_{(n)}$) e o mínimo ($x_{(1)}$). Podemos considerar que há *outliers* nesses dados?
3. Desenhe o boxplot dos dados.

Exercício - Bussab & Morettin (2013, p. 66)

Ref.:9PRC

Estudando-se a distribuição das idades dos funcionários de duas repartições públicas, obtiveram-se algumas medidas que estão no quadro abaixo.

Repartição	$x_{(1)}$	q_1	q_2	$\bar{x}$	q_3	$x_{(n)}$	S
A	18	27	33	33	39	48	5
B	18	23	32	33	42	48	10

Esboce o boxplot e o histograma alisado das duas distribuições, indicando as medidas descritas no quadro. Comente as principais diferenças entre os dois gráficos.

Exercício - Bussab & Morettin (2013, p. 65)
Ref.:5JEH

Para se estudar o desempenho de duas corretoras de ações, selecionou-se de cada uma delas amostras aleatórias das ações negociadas. Para cada ação selecionada, computou-se a porcentagem de lucro apresentada durante um período fixado de tempo. Os dados estão a seguir.

Corretora A			Corretora B		
45	60	54	57	55	58
62	55	70	50	52	59
38	48	64	59	55	56
55	56	55	61	52	53
54	59	48	57	57	50
65	55	60	55	58	54
			59	51	56

1. Desenhe o boxplot para as duas corretoras na mesma escala.
2. Para verificar a homogeneidade dos dois conjuntos de dados, um analista usou o quociente $F = {}^{S_A^2}/{}_{S_B^2}$ onde S_X^2 é a variância amostral para os dados da corretora X. Que regra de decisão você adotaria para dizer se as amostras são homogêneas ou não?

INSTITUTO FEDERAL
Goiás

Gráfico de quantis (qq-plot)

Sejam as variáveis $X = \{x_1, \ldots, x_n\}$ e $Y = \{y_1, \ldots, y_m\}$. O gráfico de quantis para as variáveis X e Y é obtido da seguinte forma:

1. esboce a reta identidade $y = x$.
2. se $m = n$ marque os pontos

$$(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)}),$$

caso contrário, se $m > n$, marque os pontos

$$(q_x(p_1), q_y(p_1)), \ldots, (q_x(p_n), q_y(p_n)),$$

onde,

$$p_i = \frac{i - 0,5}{n}.$$

Um veterinário cuida de dois grupo de cães com 10 animais cada. Num determinado dia ele anotou o peso de todos os animais dos dois grupos. Os pesos anotados foram:

Grupo 1:	28	30	10	12	12	13	20	22	25	27
Grupo 2:	15	16	16	16	18	19	19	20	22	23

Compare os grupos por meio de um gráfico de quantis. O que você pode concluir por meio desse gráfico?

Exercício

Ref.:153B

1. Considere o conjunto de dados `Brasil`. Compare as variáveis população rural e população urbana usando um qq-plot.
2. Considere o conjunto de dados `Salários`. Compare os salários de professores, mecânicos, administradores e engenheiros usando um qq-plot.
3. Considere o conjunto de dados `Temperaturas`. Compare as temperaturas de Cananéia e Ubatuba usando um qq-plot.

INSTITUTO FEDERAL
Goiás

Exercício - Magalhães & Lima (2015, p. 27)

Ref.:IBH3

Considere que desejamos comparar as alturas em metros de alunos do ensino médio de duas escolas A e B. Uma amostra com 15 estudantes foi sorteada em cada uma dessas escolas e os resultados são representados à seguir:

Escola A			Escola B		
1,60	1,76	1,72	1,62	1,71	1,70
1,69	1,60	1,66	1,62	1,65	1,85
1,85	1,64	1,70	1,57	1,67	1,70
1,85	1,62	1,78	1,65	1,73	1,73
1,58	1,64	1,65	1,61	1,60	1,70

Construa o gráfico de quantis e compare a distribuição das alturas dos alunos nas duas escolas. O que você pode concluir com base nesse gráfico?

Exercício - Magalhães & Lima (2015, p. 42)

Ref.:FBQ5

Um exame vestibular para uma faculdade tem 80 questões, sendo 40 de português e 40 de matemática. Para os 20 melhores classificados apresentamos o número de acertos em cada disciplina, em ordem decrescente do total de pontos.

Aluno	1	2	3	4	5	6	7	8	9	10
Português	35	35	34	32	31	30	26	26	24	23
Matemática	31	29	27	28	28	26	30	28	25	23

Aluno	11	12	13	14	15	16	17	18	19	20
Português	23	12	11	20	17	12	14	20	8	10
Matemática	21	32	31	20	21	25	20	13	23	20

1. Organize uma tabela de frequências para cada variável.
2. Represente graficamente as tabelas obtidas.
3. Construa a tabela de frequências da variável total de pontos.
4. Faça um gráfico de quantis comparando as notas de português e matemática.
5. Os aprovados são melhores em qual disciplina?

Gráfico de quantis no R

```r
X <- c(28,30,10,12,12,13,20,22,25,27)
Y <- c(15,16,16,16,18,19,19,20,22,23)

# qq-plot básico (X e Y com quantidades iquais de observações)
qqplot(x = X,y = Y)
abline(0,1,col = "red")

# qq-plot básico (X e Y com quantidades diferentes de observações)
X <- c(28,30,10,12,12,13,20,22,25,27)
Y <- c(15,16,16,16,18,19,19,20,22,23,20,23)

qqplot(x = X,y = Y)
abline(0,1,col = "red")
```

Referências

Bussab, W. O. & P. A. Morettin (2013). *Estatística Básica*. São Paulo: Saraiva.

Groebner, D. F., P. W. Shannon, & P. C. Fry (2017). *Business Statistics: A Decision-Making Approach* (Global edition ed.). Pearson Education Limited.

Magalhães, M. N. & A. C. P. Lima (2015). *Noções de Probabilidade e Estatística* (7 ed.). São Paulo: EdUSP.

Mckinney, E. H. (1966). Generalized birthday problem. *The American Mathematical Monthly 73*(4), 385–387.

Morgado, A. C., J. B. P. Carvalho, P. C. P. Carvalho, & P. Fernandez (1991). *Análise Combinatória e Probabilidade* (9 ed.). Coleção do Professor de Matemática. SBM.

R, D. C. T. (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. R-3.4.3.

Ross, S. M. (2010). *A First Course in Probability* (8 ed.). New York: Pearson Hall.

INSTITUTO FEDERAL
Goiás

Referências II

Selvin, S., M. Bloxham, A. I. Khuri, M. Moore, R. Coleman, G. R. Bryce, J. A. Hagans, T. C. Chalmers, E. A. Maxwell, & G. N. Smith (1975). Letters to the editor. *The American Statistician 29*(1), 67–71.

Spiegel, M. R. & L. J. Stephens (2009). *Estatística* (4 ed.). Coleção Schaum. Bookman.

Weiss, N. A. (2012). *Introdutory Statistics* (9 ed.). Addison-Wesley.

Wilcox, R. R. (2009). *Basic Statistics: Understanding conventional methods and modern insights*. Oxford University Press.

INSTITUTO FEDERAL
Goiás