PUC MINAS – PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Campus Coração Eucarístico

ESTATÍSTICA E PROBABILIDADE

– Ciência da Computação –

1º semestre de 2025

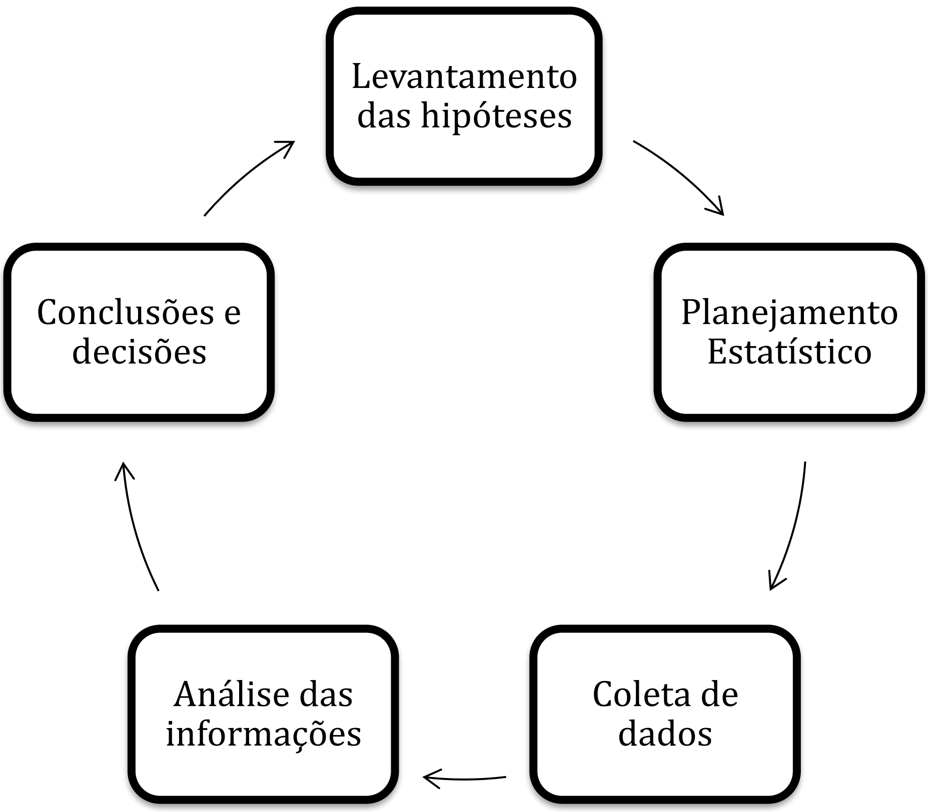
Profa. Mayra Alves Stradioto
stradioto@pucminas.br

I INTRODUÇÃO

*"**ESTATÍSTICA** É A CIÊNCIA QUE NOS AJUDA A TOMAR DECISÕES E TIRAR CONCLUSÕES NA PRESENÇA DE VARIABILIDADE."*

Montgomery (2018)

1.1 Aplicações da Estatística



Conceitos:

- População ou Universo Estatístico (N): é o conjunto constituído por todos os indivíduos (valores, pessoas, medidas, etc) que apresentam pelo menos uma característica comum, cujo comportamento interessa analisar.
- Amostra (n): é um subconjunto, uma parte selecionada da população, através da qual se faz inferência sobre as características da população.

1.2 Classificação de variáveis

Qualitativas (ou categóricas):
São variáveis obtidas por classificação.

Ordinais:
Os valores da variável têm uma ordenação.
Ex: Mês de observação (janeiro, fevereiro, março, ...).

Nominais:
Não é possível estabelecer uma ordem.
Ex: Curso (Engenharia, Sistemas, Nutrição, Enfermagem, ...)

Quantitativas:
São variáveis obtidas sob forma de dados que medem numericamente determinadas características.

Discretas:
Os valores diferem entre si por quantidades fixas; em geral, são resultantes de contagens.
Ex: Número de alunos aprovados.

Contínuas:
Assumem valores em intervalos dos números reais (mensuração).
Ex: Peso.

II ESTATÍSTICA DESCRITIVA

Exemplo: (Adaptado de Werkema, 1995) Uma fábrica de azulejos recentemente começou a receber reclamações de seus clientes. A maioria das reclamações era relativa aos seguintes problemas:

- Os azulejos, ao serem manuseados, quebravam-se facilmente;
- O assentamento dos azulejos não produzia um resultado uniforme em relação ao nível da parede.

Sabe-se que os limites de especificação para a espessura dos azulejos são 5,0 ± 1,5 mm, ou seja, a espessura dos azulejos pode variar entre 3,5 a 6,5 mm, sendo o valor nominal de especificação igual a 5,0 mm. Para avaliar se estavam ocorrendo problemas com a espessura dos azulejos produzidos, o grupo decidiu retirar uma amostra aleatória dos azulejos fabricados pela empresa, medir a espessura destes e comparar os resultados obtidos com as especificações. Como a indústria emprega duas turmas de trabalho (turmas A e B) e pode haver diferença na qualidade dos azulejos produzidos por cada turma, foi utilizada uma estratificação, sendo então retirada uma amostra de 56 azulejos produzidos pela turma A e 56 fabricados pela turma B. Os resultados da turma A são apresentados no Quadro 1.

Quadro 1 – Resultado das espessuras dos azulejos da turma A.

4,5	3,1	2,3	2,8	3,5	2,3	2,8	3,4	3,1	3,1	3,4	3,5
3,3	3,7	3,8	4,6	4,1	4,0	3,5	2,3	4,3	4,5	5,5	5,2
2,9	3,5	3,3	4,2	4,3	3,1	5,4	4,5	3,3	3,4	4,5	4,6
3,9	3,5	4,3	2,7	4,0	4,7	4,0	3,8	3,1	3,4	5,4	5,6
4,3	3,5	5,4	4,9	3,8	5,7	5,6	3,5				

2.1 Tabelas de frequências para variáveis quantitativas

Se a variável é contínua, ou se é discreta, mas assume um grande número de valores distintos, considerar cada valor como uma classe na tabela e no gráfico de frequências fica inviável. A solução é agrupar os valores em classes, ou seja, "categorizar" a variável.

Construção da tabela de frequências para variáveis quantitativas

1°) Determinar o número de classes[1]. Uma regra prática para determinar o *número de classes* é a seguinte:

$$\textit{número de classes} = k = \sqrt{n};$$

2°) Identifique o *maior valor* ($máx$) e o *menor valor* ($mín$) da amostra;

3°) Calcule a *amplitude total dos dados* (a):

$$a = máx - mín;$$

4°) Calcule o comprimento de cada intervalo, ou seja, a *amplitude das classes* (h):

$$h = \frac{a}{k};$$

5°) Arredonde o valor de h conforme as regras de arredondamento. Este número deve ter a mesma quantidade de casas decimais dos dados da amostra;

6°) Calcule os limites de cada intervalo:

- A primeira classe vai de *mín* até *mín + h*;
- A segunda classe vai de *mín + h* até *mín + 2h*;

...

---

[1] É recomendado que o número de classes esteja entre 4 e 10. Também pode ser calculado através da fórmula de Sturges: $k = 1 + log_2 n$.

7°) Construa uma tabela de frequências com as seguintes colunas:

- Limites de cada intervalo: os intervalos são fechados à esquerda e abertos à direita. NOTAÇÃO: ⊢
- Frequência absoluta de cada intervalo (n_i).
- Frequência relativa de cada intervalo (f_i).
- Frequência absoluta acumulada de cada intervalo (N_i);
- Frequência relativa acumulada de cada intervalo (F_i).

Tabela 1 – Distribuição de frequências das espessuras dos azulejos da turma A.

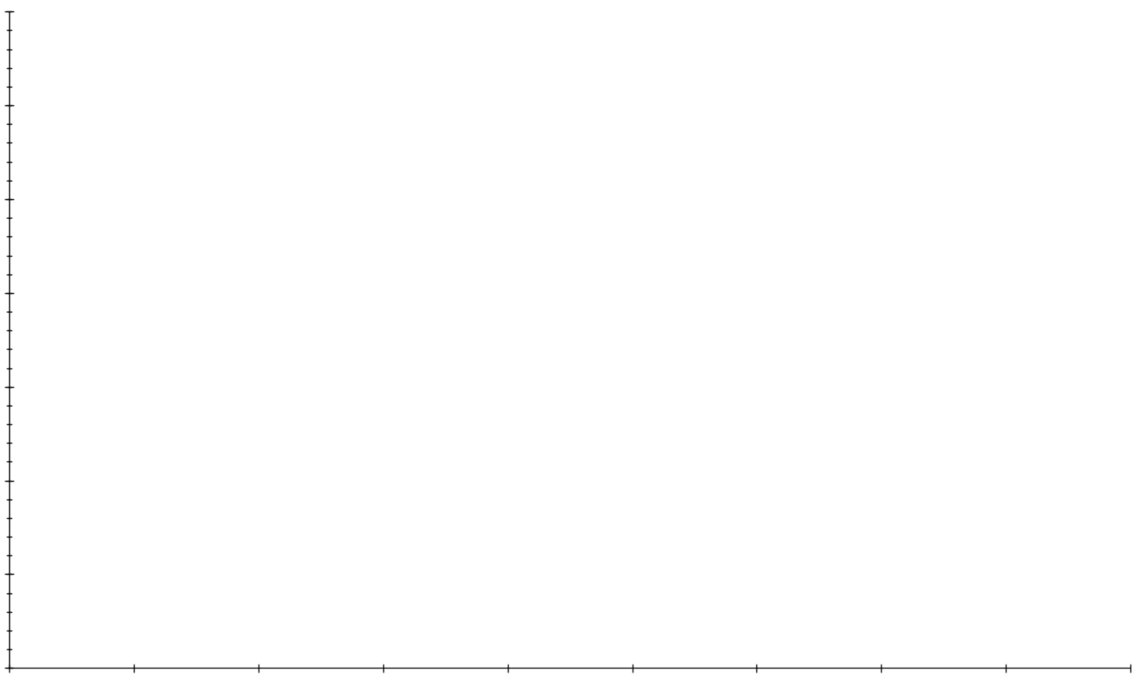
TOTAL				

2.2 Representação gráfica de variáveis quantitativas

I. *Histograma*

O histograma é muito parecido com o gráfico de barras, com uma diferença essencial: os dados apresentados no eixo das abscissas são numéricos e têm uma ordem que deve ser obedecida rigorosamente. Por ser um gráfico resultante da tabela de frequências, essa ordem é a mesma apresentada na tabela. Os números devem, portanto, aparecer em sequência, quer existam ou não dados com determinados valores. A altura das barras pode representar tanto a frequência absoluta quanto a frequência relativa.

Gráfico 1 – Histograma da distribuição de frequências das espessuras dos azulejos da turma A.



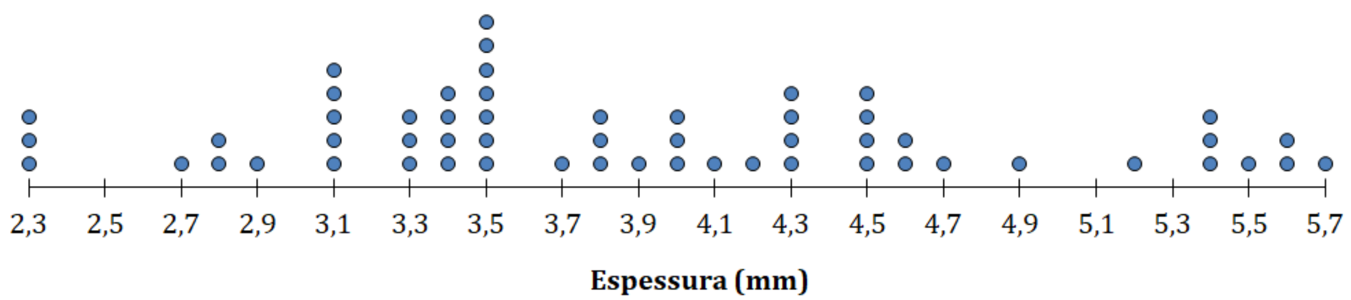
II. Diagrama de pontos

Permite a visualização dos aspectos gerais da distribuição da variável em estudo, tais como valores máximo e mínimo, valores mais frequentes, forma da distribuição (simétrica ou assimétrica), etc.

A construção do diagrama de pontos para um conjunto de dados consiste em:

1. Determinar os valores mínimo e máximo do conjunto de dados;
2. Com base nestes valores, estabelecer um eixo em escala conveniente;
3. Utilizando o eixo estabelecido em (2), para cada valor observado nos dados, marcar com pontos dispostos verticalmente, sua correspondente frequência.

Gráfico 2 – Diagrama de pontos das espessuras dos azulejos da turma A.



III. *Diagrama de Ramo-e-folhas*

O diagrama de ramo-e-folhas fornece praticamente as mesmas informações que o histograma, mas com duas vantagens:

- É mais fácil de construir à mão;
- Apresenta mais informações, porque mostra os dados reais.

A construção do ramo-e-folhas para um conjunto de dados consiste em:

1. Dividir cada número do conjunto de dados em duas partes: uma composta por um ou mais dígitos iniciais (*ramo*) e outra composta pelo dígito restante (*folha*);
2. Traçar uma linha vertical e escrever o ramo, com a escala adotada, à esquerda[1];
3. Colocar à direita da linha vertical o último algarismo de cada número observado. Estes números formarão as folhas do gráfico;
4. Escrever a unidade do ramo-e-folhas[2].

Gráfico 3 – Diagrama de ramo-e-folhas das espessuras dos azulejos da turma A.

2	333
2	7889
3	111113334444
3	555555578889
4	000123333
4	55556679
5	2444
5	5667

Unidade de Folha = 0,1
Chave: 2|3 = 2,3

[1] Essa linha é opcional. Pode ser apenas uma divisão invisível entre o ramo e as folhas.
[2] Também chamada de *chave* do gráfico.

2.3 Medidas de Tendência Central

Os dados apresentados em tabelas ou gráficos fornecem informação sobre o assunto em estudo. Mas pode existir interesse em apresentar essa informação de maneira condensada. Embora você possa “ver” os dados em tabelas e gráficos, é difícil “falar” sobre eles.

I. *Média Aritmética*

A média aritmética de n observações $(x_1, x_2, \cdots, x_n)$ é denotada por $\bar{x}$ e é dada por

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

Exemplo: Considere os seguintes valores, que representam o tempo (em segundos) para atualização de um mesmo aplicativo mobile em diferentes momentos do dia. Calcule e interprete a média.

5,2 6,4 5,7 4,8 6,7 5,4 4,8 6,3 5,5 6,2 4,9 8,3

II. *Mediana*

A mediana (Md) é o valor que ocupa a posição central de um conjunto de dados ordenados, ou seja, a mediana divide a distribuição ao meio (50% das observações estão acima da mediana e 50% abaixo).

Exemplo: Considere os seguintes valores, que representam o tempo (em segundos) para atualização de um mesmo aplicativo mobile em diferentes momentos do dia. Calcule e interprete a mediana. Compare com a média calculada anteriormente.

5,2	6,4	5,7	4,8	6,7	5,4	4,8	6,3	5,5	6,2	4,9	8,3

Observe que a mediana é dada pelo elemento que ocupa a posição $\frac{n+1}{2}$ se n for ímpar, e pela média dos elementos que ocupam as posições $\frac{n}{2}$ e $\frac{n+2}{2}$ se n for par.

III. *Moda*

A moda (Mo) é o valor que ocorre com mais frequência em um conjunto de dados.

Exemplo: Considere os seguintes valores, que representam o tempo (em segundos) para atualização de um mesmo aplicativo mobile em diferentes momentos do dia. Determine a moda.

5,2 6,4 5,7 4,8 6,7 5,4 4,8 6,3 5,5 6,2 4,9 8,3

Um conjunto de dados pode ser:

- Amodal: não tem moda
- Unimodal: tem apenas uma moda
- Bimodal: tem duas modas
- Multimodal: tem mais de duas modas

Exemplo: Para as séries a seguir, encontre a moda e classifique o conjunto de dados quanto ao número de modas:

(a) 7; 8; 9; 10; 10; 10; 11; 12; 13; 15

(b) 3; 5; 8; 10; 12; 13

(c) 2; 3; 4; 4; 4; 5; 6; 7; 7; 7; 8; 9

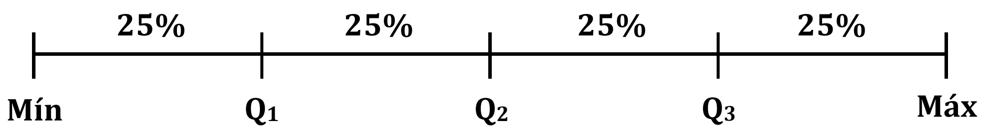
(d) 4; 4; 5; 7; 7; 8; 10; 10; 11; 13; 13

(e) 1; 1; 1; 2; 2; 2; 3; 3; 3; 4; 4; 4; 5; 5; 5

2.4 Medidas Separatrizes

I. *Quartis*

Os quartis são pontos que dividem o conjunto de dados ordenados em quatro partes iguais, originando o *1º quartil* (Q_1), o *2º quartil* (Q_2) *3º quartil* (Q_3). Assim:



Exemplo: Considere os seguintes valores, que representam o tempo (em segundos) para atualização de um mesmo aplicativo mobile em diferentes momentos do dia. Calcular o Q_1, o Q_2 e o Q_3.

4,8 4,8 4,9 5,2 5,4 5,5 5,7 6,2 6,3 6,4 6,7 8,3

Exemplo: (Adaptado de Farias et al., 2003) Considere os dados a seguir, já ordenados do maior para o menor, de 35 observações, em decibéis do nível de ruído de tráfego em certo cruzamento. Determine os valores dos quartis.

52,0	54,0	55,7	55,8	55,9	56,2	56,4
56,7	57,2	57,6	59,4	59,4	59,8	60,0
60,3	60,5	60,8	61,0	61,7	61,8	62,1
62,6	63,1	63,3	64,0	64,6	64,9	65,7
66,2	66,8	67,1	67,9	68,2	69,4	77,1

2.5 Medidas de Variabilidade (Dispersão)

Através das medidas de tendência central, estabelecemos um valor em torno do qual os dados se distribuem. Além de calcular este ponto, devemos calcular um valor de mostre a *variabilidade* dos dados, ou seja, uma medida de dispersão.

Exemplo: Quatro alunos fizeram quatro atividades com valor de 10 pontos cada uma. Suas notas foram:

{5 ; 5 ; 5 ; 5}   {4 ; 6 ; 4 ; 6}   {0 ; 4 ; 6 ; 10}   {0 ; 0 ; 10 ; 10}

I. *Amplitude da amostra*

A amplitude é a medida de dispersão mais fácil de ser calculada e a mais utilizada. A amplitude é simplesmente a diferença entre o maior e o menor valor, ou seja,

$$a = máx - mín$$

Exemplo: Considere um grupo de pessoas com idades 4, 3, 4, 3, 4, 3 e 21. Calcule $\bar{x}$, $\tilde{x}$, Mo e a.

Exemplo: Considere os seguintes valores, que representam o tempo (em segundos) para atualização de um mesmo aplicativo mobile em diferentes momentos do dia. Calcule a amplitude.

5,2   6,4   5,7   4,8   6,7   5,4   4,8   6,3   5,5   6,2   4,9   8,3

II. Variância

Mede a variabilidade dos dados através dos desvios em relação à média. *Desvio em relação à média* é a diferença entre cada valor observado e a média do conjunto de dados.

A notação é s^2 e é calculada como a soma dos desvios ao quadrado, dividida por $n-1$, ou seja,

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}.$$

O denominador $n-1$ é chamado *graus de liberdade da amostra.* Para calcular a variância da população (chamada σ^2), a soma dos desvios ao quadrado é dividida por N. Utilizaremos essa medida nos próximos capítulos.

Exemplo: Considere os seguintes valores, que representam o tempo (em segundos) para atualização de um mesmo aplicativo mobile em diferentes momentos do dia. Calcule a variância.

5,2	6,4	5,7	4,8	6,7	5,4	4,8	6,3	5,5	6,2	4,9	8,3

III. *Desvio-padrão*

O desvio-padrão é um valor que possui a mesma unidade de medida dos dados originais. Por definição, o é a raiz quadrada, com sinal positivo, da variância. O desvio-padrão da amostra é denotado por s e pode ser calculado através da fórmula

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}.$$

Exemplo: Considere os seguintes valores, que representam o tempo (em segundos) para atualização de um mesmo aplicativo mobile em diferentes momentos do dia. Calcule o desvio-padrão.

5,2	6,4	5,7	4,8	6,7	5,4	4,8	6,3	5,5	6,2	4,9	8,3

IV. *Coeficiente de variação*

O coeficiente de variação é um índice relativo de dispersão que compara o desvio-padrão (s) com a média ($\bar{x}$), e fornece uma medida de homogeneidade dos dados. É geralmente expresso em %, e é calculado pela fórmula

$$CV = \frac{s}{\bar{x}} \times 100.$$

O coeficiente de variação é utilizado com maior frequência na comparação de conjuntos de dados.

Exemplo: Considere os seguintes valores, que representam o tempo (em segundos) para atualização de um mesmo aplicativo mobile em diferentes momentos do dia. Calcule o coeficiente de variação.

5,2	6,4	5,7	4,8	6,7	5,4	4,8	6,3	5,5	6,2	4,9	8,3

V. Amplitude interqualtílica[1]

É calculada como a distância entre o Q_3 e o Q_1:

$$\text{AIQ} = Q_3 - Q_1$$

Como existem 50% dos valores entre o Q_1 e o Q_3, quanto maior o valor do IQ, maior a dispersão do conjunto de dados.

A melhor forma de interpretar o IQ é através do diagrama de caixa (Boxplot).

[1] Também chamada de Distância Interquartílica, Amplitude Interquartil, Faixa Interquartil, Desvio Interquartílico...

2.6 Diagrama de caixa (Boxplot)

O boxplot é um gráfico que apresenta a dispersão do conjunto de dados, através dos valores mínimo, máximo, primeiro quartil (Q_1), mediana (Q_2) e terceiro quartil (Q_3). Além disso, através deste gráfico é possível verificar se o conjunto de dados possui valores atípicos (outliers), que podem afetar fortemente algumas medidas como média e desvio-padrão.

A construção do boxplot para um conjunto de dados consiste em:

1. Determinar o valores mínimo, Q_1, Q_2, Q_3, máximo e AIQ do conjunto de dados;
2. Com base nestes valores, estabelecer um eixo em escala conveniente;
3. Construir um retângulo, iniciando no Q_1 e terminando no Q_3;
4. Trace uma linha dentro do retângulo, no valor do Q_2;
5. A partir do valor do Q_1, em direção ao mínimo, trace uma linha perpendicular à linha do Q_1. Essa linha irá terminar no valor mínimo ou no valor de $Q_1 - (1{,}5 \times AIQ)$, o que ocorrer primeiro;
6. A partir do valor do Q_3, em direção ao máximo, trace uma linha perpendicular à linha do Q_3. Essa linha irá terminar no valor máximo ou no valor de $Q_3 + (1{,}5 \times AIQ)$, o que ocorrer primeiro.
7. Marque os *outliers*, se houver[1]. Um valor do conjunto de dados é um *outlier*[2] se ele está:
   i. Abaixo de $Q_1 - (1{,}5 \times AIQ)$;
   ii. Acima de $Q_3 + (1{,}5 \times AIQ)$.

---

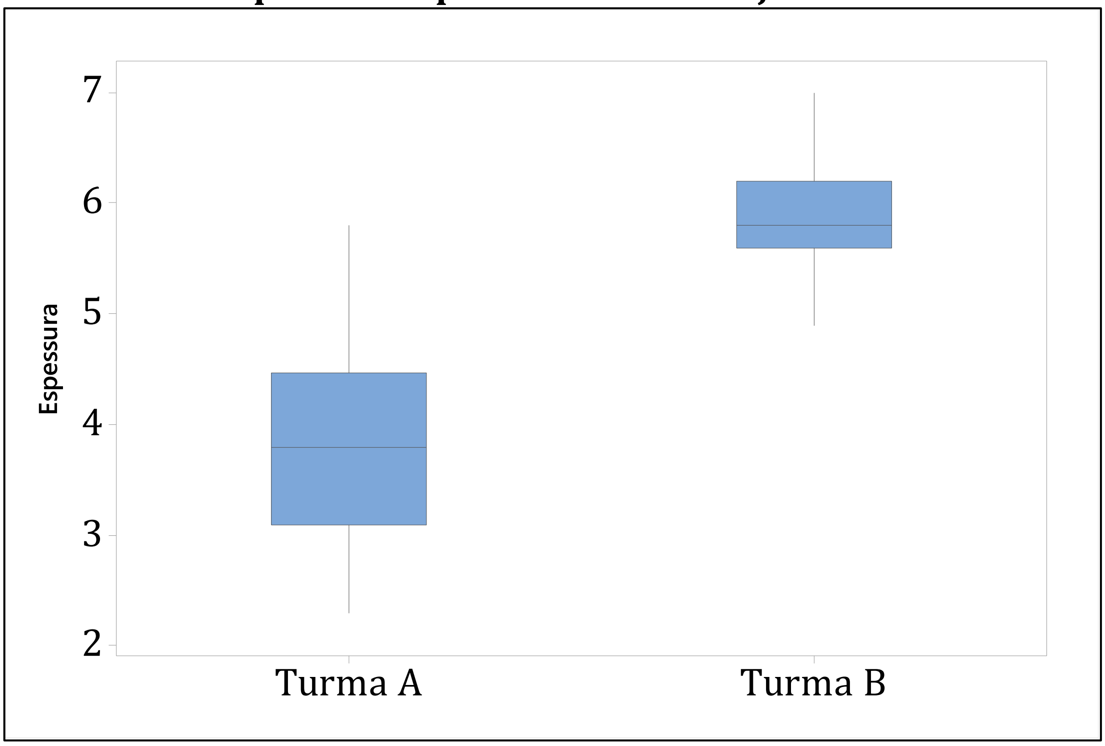
[1] Para representar um outlier, normalmente é utilizado um asterisco (*), um círculo (•) ou um "x".
[2] Um valor que esteja a uma distância de mais de 3 vezes a amplitude interquartílica do Q_1 ou do Q_3 é chamado de *outlier extremo*. Na construção do boxplot, símbolos diferentes podem ser usados para representar os dois tipos de outliers.

Exemplo: (Adaptado de Farias et al., 2003) Considere os dados a seguir, já ordenados do maior para o menor, de 35 observações, em decibéis do nível de ruído de tráfego em certo cruzamento. Construir e interpretar o boxplot do conjunto de dados.

52,0	54,0	55,7	55,8	55,9	56,2	56,4
56,7	57,2	57,6	59,4	59,4	59,8	60,0
60,3	60,5	60,8	61,0	61,7	61,8	62,1
62,6	63,1	63,3	64,0	64,6	64,9	65,7
66,2	66,8	67,1	67,9	68,2	69,4	77,1

Exemplo: (Adaptado de Werkema, 1995) Fábrica de azulejos.
Lembrando: os limites de especificação para a espessura dos azulejos são 5,0 ± 1,5 mm, ou seja, a espessura dos azulejos pode variar entre 3,5 a 6,5 mm, sendo o valor nominal de especificação igual a 5,0 mm.

Gráfico 4 – Boxplot das espessuras dos azulejos das turmas A e B.



III PROBABILIDADE

Em todos os fenômenos estudados pela Estatística, os resultados, mesmo nas mesmas condições de experimentação, variam de uma observação para outra, dificultando a previsão de um resultado futuro.

Para a explicação destes fenômenos, chamados *experimentos aleatórios*, utilizamos o *cálculo das probabilidades.*

3.1 Conceitos

Experimento aleatório

Situação em que os resultados possíveis são conhecidos, mas não sabemos *a priori* qual deles ocorrerá. Além disso, os experimentos aleatórios podem ser repetidos indefinidamente sob as mesmas condições.

Exemplo: Lançar um dado.

Espaço amostral (E)

Conjunto de todos os resultados possíveis de um experimento aleatório.

Exemplo: $E = \{1, 2, 3, 4, 5, 6\}$

Evento (sempre em letras maiúsculas)

Subconjunto do espaço amostral que contém apenas os elementos de interesse. Um evento deve ser representado sempre por letras maiúsculas.

Exemplo 1: Considere o lançamento de dois dados, onde devemos observar as faces superiores. Escreva o espaço amostral quando a ordem dos dados importa e determine:

(a) Os elementos do evento A, onde A = a soma das faces superiores é seis;

(b) Os elementos do evento B, onde B = o valor de uma das faces é exatamente a metade do outro;

(c) Os elementos do evento C, onde C = o valor de uma face multiplicado pelo valor da outra é igual a 1;

(d) Os elementos do evento D, onde D = a soma das faces superiores é maior que 13.

3.2 Cálculo de Probabilidades

Definição Clássica de Probabilidade

Se os eventos são equiprováveis (têm a mesma probabilidade de ocorrer), podemos calcular $P(A)$ (onde A é um evento qualquer) como:

$$\boxed{P(A) = \frac{\text{n° de resultados favoráveis à ocorrência do evento } A}{\text{n° de resultados possíveis}}}$$

Exemplo 1 (continuação): Considere o lançamento de dois dados, onde devemos observar as faces superiores. Determine:

(e) Calcule $P(E)$;

(f) Calcule $P(A)$, $P(B)$, $P(C)$ e $P(D)$.

Definição Frequentista de Probabilidade

Se os eventos simples não são equiprováveis (não têm a mesma probabilidade de ocorrer), podemos calcular $P(A)$ (onde A é um evento qualquer) como:

$$\boxed{P(A) = \frac{\text{n° de vezes que o evento A ocorreu}}{\text{n° total de repetições do experimento}}}$$

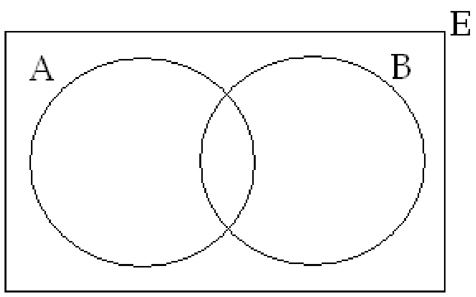
Exemplos:

- Especificar o tempo de garantia de lâmpadas através do cálculo de $P(A_t)$, onde A_t é o evento "a lâmpada funcionar até o tempo t";
- Avaliar a viabilidade da construção de uma usina hidrelétrica utilizando dados históricos de vazões de um rio para calcular a probabilidade de ocorrência de determinados eventos climáticos.

3.3 Tipos especiais de eventos

Evento interseção

É a ocorrência simultânea dos eventos A e $B \Rightarrow A \cap B$.



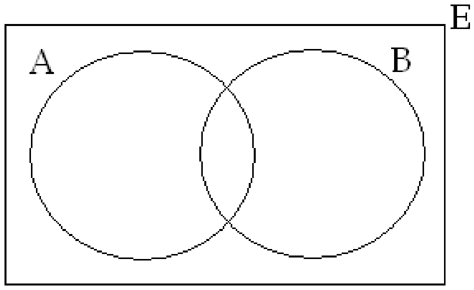
Exemplo 1 (continuação):

(g) Determine os elementos do evento $A \cap B$;

(h) Calcule $P(A \cap B)$.

Evento união

É a ocorrência de A ou de B, ou de ambos $\Rightarrow A \cup B$.



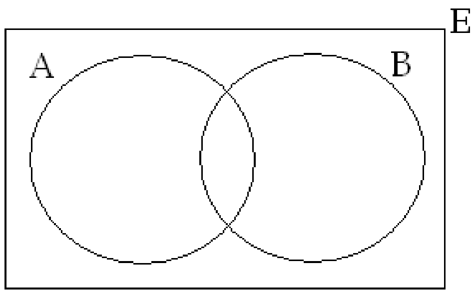
Exemplo 1 (continuação):

(i) Determine os elementos do evento $A \cup B$;

(j) Calcule $P(A \cup B)$.

Evento complementar

O complementar do evento A contém todos os elementos do espaço amostral que não pertencem ao evento $A \Rightarrow \bar{A}\ ou\ A^C$.



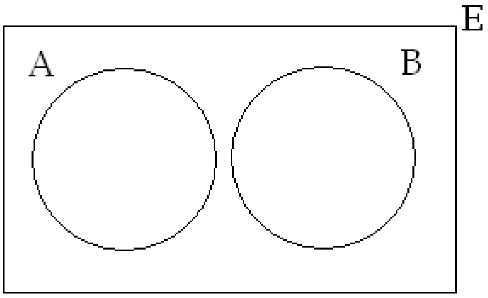
Exemplo 1 (continuação):

(k) Determine os elementos do evento $\bar{A}$;

(l) Calcule $P(\bar{A})$.

Eventos mutuamente excludentes (ou disjuntos)

Quando a ocorrência de um impossibilita a ocorrência do outro, ou seja, $A \cap B = \emptyset$.



Exemplo 1 (continuação):

(m) Determine os elementos do evento $A \cap C$;

(n) Calcule $P(A \cap C)$;

(o) Calcule $P(A \cup C)$.

3.4 Propriedades da Probabilidade

Dado um experimento aleatório com espaço amostral E, a probabilidade de um evento A, dada por $P(A)$, é uma função definida em E que associa a cada evento um número real, satisfazendo as seguintes probabilidades:

i. $0 \leq P(A) \leq 1$ para qualquer evento A.

ii. $P(E) = 1$ onde E é o espaço amostral.

iii. Se A e B são eventos mutuamente excludentes, então

$$\boxed{P(A \cup B) = P(A) + P(B)}$$

Consequentemente:

- Se A e B são dois eventos quaisquer, então

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

- $\boxed{P(\bar{A}) = 1 - P(A)}$

3.5 Probabilidade Condicional

A probabilidade de um evento A ocorrer, dado que se sabe que um evento B ocorreu, é chamada *probabilidade condicional do evento A dado B*, e é denotada por $P(A|B)$. É calculada como

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}} \qquad \text{se } P(B) > 0$$

- **Regra da multiplicação:** $P(A \cap B) = P(A|B)P(B)$
- $\boxed{P(\bar{A}|B) = 1 - P(A|B)}$

Exemplo 1 (continuação): Considere o lançamento de dois dados, onde devemos observar as faces superiores. Determine:

(p) Determine $P(A|B)$;

(q) Determine $P(\bar{A}|B)$.

3.6 Independência Estatística

Os eventos A e B são *independentes* se o fato de um deles ter ocorrido não altera a probabilidade de ocorrência de outro, isto é,

$$\boxed{P(A|B) = P(A)} \qquad \text{e} \qquad \boxed{P(B|A) = P(B)}$$

Pela regra da multiplicação,

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

Exemplo 1 (continuação): Considere o lançamento de dois dados, onde devemos observar as faces superiores. Determine:

(r) Os eventos A e B são independentes? Justifique.

Exemplo 2: Um casal possui 2 filhos(as). Qual a probabilidade de ambos serem do sexo masculino?

Exemplo 3: O depósito da loja de confecções Savanah Ltda. possui 180 calças jeans da marca A, das quais seis são defeituosas, e 200 da marca B, das quais nove são defeituosas. Um funcionário da loja vai ao depósito e retira uma calça jeans. Qual a probabilidade de que a calça jeans seja:

(a) Da marca A? (0,4737)

(b) Defeituosa? (0,0395)

(c) Da marca A e defeituosa? (0,0158)

(d) Da marca A e da marca B? (0)

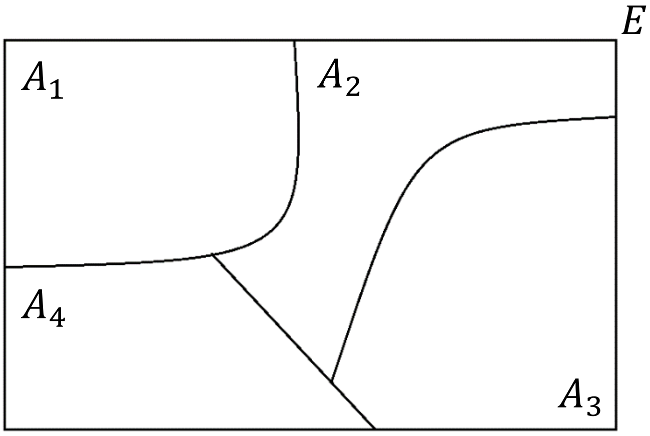
(e) Da marca B ou não defeituosa? (0,9842)

(f) Defeituosa, sabendo-se que é da marca B? (0,045)

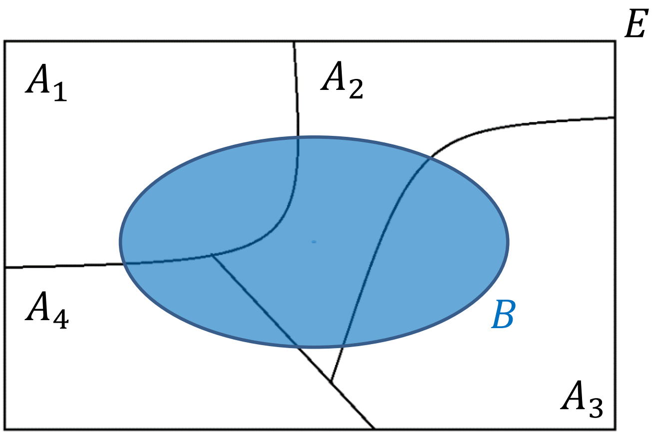
(g) O fato da calça ser defeituosa independe da marca da mesma? Justifique com os cálculos. (Não!...)

3.7 O Teorema de Bayes

Sejam A_1, A_2, A_3, ..., A_n eventos mutuamente excludentes cuja união é o espaço amostral E, ou seja, um dos eventos *necessariamente* deve ocorrer.



Sejam $P(A_i)$ as probabilidades conhecidas dos vários eventos, e B um evento qualquer de E tal que são conhecidas todas as probabilidades condicionais $P(B|A_i)$.



A probabilidade do evento B ocorrer pode ser calculada como a soma das probabilidades das interseções do evento B com cada um dos eventos A_i, ou seja,

$$P(B) = \sum P(B \cap A_i).$$

Essa equação é conhecida como **regra da probabilidade total**.

Pela regra da multiplicação, a regra da probabilidade total pode ser escrita como

$$P(B) = \sum P(A_i)P(B|A_i).$$

Aplicando a regra da probabilidade total para a probabilidade condicional, a probabilidade de cada evento A_i condicional ao evento B pode ser calculada como

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

O resultado acima, conhecido como **Teorema de Bayes**, é bastante importante, pois relaciona probabilidades *a priori* $P(A_i)$ com probabilidades *a posteriori* $P(A_i|B)$.

Exemplo: Três candidatos disputam as eleições para a Presidência da República. O candidato do partido de direita tem 40% da preferência eleitoral, o de centro tem 32% e o da esquerda 28%. Em sendo eleito, a probabilidade de dar, efetivamente, prioridade para Educação e Saúde é de 0,3; 0,5 e 0,6 para os candidatos de direita, centro e esquerda, respectivamente.

(a) Qual é a probabilidade de ser dada prioridade a essas áreas no próximo governo?

(b) Se a área teve prioridade, qual a probabilidade do candidato de direita ter ganho a eleição?

IV DISTRIBUIÇÕES DISCRETAS DE PROBABILIDADE

4.1 Conceitos

Variável aleatória

Uma variável aleatória (v.a.) é uma variável quantitativa que produz valores diferentes quando observada em repetições feitas mesmo sob condições idênticas.

As variáveis aleatórias são denotadas por letras maiúsculas (X, Y, W, etc), e seus respectivos valores por letras minúsculas ($x = 3$, $y = 54$, $w = 0{,}35$).

Variável aleatória Discreta

Se a v.a. assume valores num intervalo de números inteiros, é chamada de *variável aleatória discreta*.

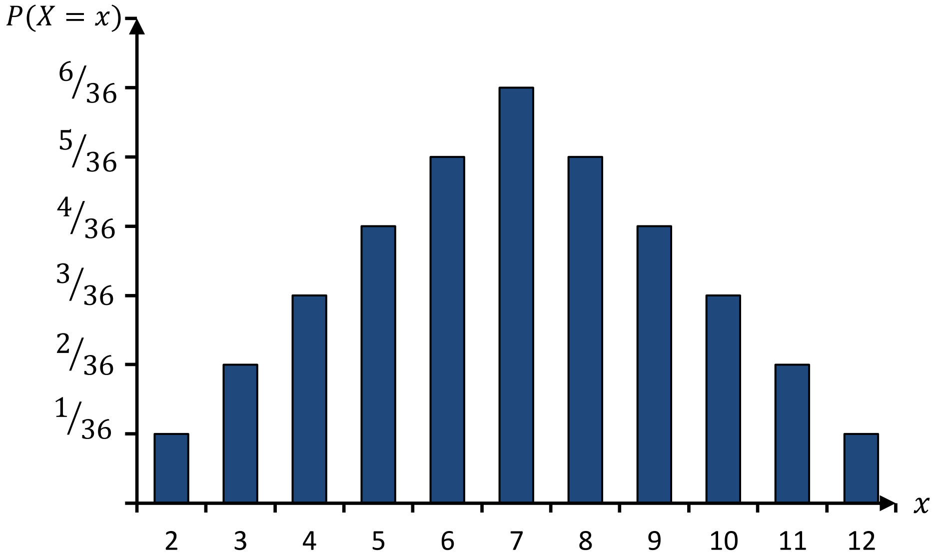
Função de Distribuição de Probabilidade

A função de probabilidade de uma variável aleatória descreve como as probabilidades estão distribuídas para cada valor desta variável. Para variáveis aleatórias discretas, essa é chamada *função de distribuição de probabilidade*, e é denotada por $P(X = x)$ ou $p(x)$.

Exemplo: No lançamento de dois dados onde a ordem importa, a variável aleatória X denota a soma dos pontos das faces superiores.

(a) Determine os valores de X e a distribuição de probabilidade associada.

(b) Construa um gráfico de barras que represente essa distribuição.



Uma função $p(x)$ é uma *função distribuição de probabilidade* se, e somente se:

1. $0 \leq p(x) \leq 1 \quad \forall x$;

2. $\sum p(x) = 1$.

<u>*Função de Distribuição Acumulada*</u>

A *função de distribuição acumulada* (ou *função de probabilidade acumulada*), denotada por $F(x)$, fornece a probabilidade de a v.a. X apresentar valores menores ou iguais a x, ou seja,

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t)$$

Esperança de X: $\underline{E(X)}$

Se X é uma variável aleatória discreta cuja distribuição de probabilidade é dada por

x	x_1	x_2	$\cdots$	x_n
$P(X = x)$	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_n)$

Então, a *média* desta v.a., também chamada *valor esperado*, ou *esperança matemática*, ou simplesmente *esperança de X*, é representada por $E(X)$ ou μ, é calculada como

$$\mu = E(X) = \sum_{i=1}^{n} [\, x_i \times p(x_i)]$$

Variância de X: $\underline{Var(X)}$

Analogamente à fórmula do valor esperado, a variância da v.a. X, denotada por σ^2 ou $Var(X)$, é dada por

$$\sigma^2 = Var(X) = \sum_{i=1}^{n} [\, (\, x_i - \mu)^2 \times p(x_i)]$$

Desenvolvendo essa expressão, podemos escrever a variância de X como

$$\sigma^2 = Var(X) = E(X^2) - \mu^2$$

4.2 Distribuição Binomial

Considere um experimento aleatório realizado n vezes, sob as mesmas condições, com as seguintes características:

- Cada repetição do experimento produz apenas um de dois resultados possíveis, denominados por "*sucesso*" ou "*fracasso*";
- A probabilidade de sucesso $P(S) = p$ é constante em cada repetição do experimento;
- As provas são *independentes*, isto é, o resultado de um ensaio não interfere no resultado do outro.

Se X = *número de sucessos em n realizações do experimento,* então, $X \sim Bin(n,p)$. A probabilidade é dada por

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

onde

- n = número de realizações do experimento,
- p = probabilidade de um sucesso em qualquer dos ensaios,
- $(1 - p)$ = probabilidade de um fracasso em qualquer dos ensaios, e
- $\binom{n}{x} = \frac{n!}{x!(n-x)!}$.

A esperança e a variância de uma v.a. $X \sim Bin$ são dadas por $E(X) = np$ e $Var(X) = np(1-p)$.

Exemplo: Seja X uma v.a. seguindo o modelo Binomial com parâmetros n = 15 e p = 0,4. Calcule:

(a) A esperança e a variância de X.

(b) $P(X = 7)$

(c) $P(X < 3)$

(d) $P(X \geq 3)$

(e) $P(5 \leq X \leq 12)$

(f) $P(X > 13 \text{ e } X < 6)$

(g) $P(X \leq 13 | X \geq 11)$

Exemplo: Uma universidade descobriu que 20% de seus estudantes abandonam a disciplina de Estatística a cada semestre. Considere que 20 estudantes tenham se matriculado para essa disciplina neste semestre.

(a) Qual a probabilidade de que exatamente quatro alunos abandonem a disciplina? (0,2182)

(b) Qual a probabilidade de que dois alunos ou menos abandonem a disciplina? (0,2061)

(c) Qual a probabilidade de que mais de três alunos não abandonem a disciplina? (0,9999999992)

4.3 Distribuição de Poisson

Dizemos que um experimento apresenta uma distribuição de Poisson se for possível observar eventos discretos em um intervalo de meio contínuo (de tempo, área, volume, etc.), e se o experimento possui as seguintes características:

- a probabilidade de observar exatamente um sucesso no intervalo é estável;
- a ocorrência de um sucesso em qualquer intervalo é estatisticamente independente da ocorrência em qualquer outro intervalo.

Na distribuição de Poisson, X = *número de ocorrências do evento de interesse no intervalo.* A notação é $X \sim Po(\lambda)$ e a probabilidade é dada por

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x = 0, 1, 2, \ldots$$

onde λ é o parâmetro da distribuição, ou seja, a taxa média de ocorrência do evento de interesse. A esperança e a variância de uma v.a. $X \sim Po$ são dadas por $E(X) = \lambda$ e $Var(X) = \lambda$.

Exemplo: Seja X uma v.a. seguindo o modelo de Poisson com λ = 3. Calcule:

(a) $P(X = 7)$

(b) $P(X \geq 2)$

(c) $P(X \geq 25)$

Exemplo: Seja p = 0,01 a probabilidade de certo tipo de lâmpada queimar no período de 24 horas. Qual a probabilidade de um luminoso com 10 destas lâmpadas permanecer totalmente aceso durante esse período? E durante 3 dias?

4.4 Distribuição Hipergeométrica

Dizemos que um experimento apresenta uma distribuição Hipergeométrica se desejamos contar o número elementos com determinada característica em uma amostra de tamanho n, retirada de uma população de tamanho N, onde r destes N indivíduos possuem tal característica (sucesso), e se o experimento satisfaz as seguintes condições:

- Cada repetição do experimento produz apenas um de dois resultados possíveis;
- A amostra é retirada sem reposição, ou seja, os resultados *não são independentes*.

Na distribuição Hipergeométrica, X = *número elementos na amostra que possuem a característica de interesse.* A notação é $X \sim Hip(N, r, n)$ e a probabilidade é dada por

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, min(r, n)$$

onde

- N = tamanho da população;
- n = tamanho da amostra;
- r = número de elementos da população com a característica de interesse.

A esperança e a variância de uma v.a. $X \sim Hip$ são dadas por $E(X) = np$ e $Var(X) = np(1-p)\frac{N-n}{N-1}$, onde $p = \frac{r}{N}$ é a probabilidade de se obter um sucesso em uma única retirada.

Exemplo: (Adaptado de Morettin, 2010) Pequenos motores são guardados em caixas de 50 unidades. Um inspetor de qualidade examina cada caixa, antes da posterior remessa, testando 5 motores. Se nenhum for defeituoso, a caixa é aceita. Caso contrário, todos os 50 são testados. Há 6 motores defeituosos numa caixa. Qual a probabilidade de que seja necessário examinar todos os motores desta caixa?

V DISTRIBUIÇÕES CONTÍNUAS DE PROBABILIDADE

5.1 Conceitos

Variável aleatória contínua

Uma v.a. que assume valores no intervalo dos números reais, é chamada de *variável aleatória contínua*.

Função Densidade de Probabilidade

Uma função $f(x)$ é chamada *função densidade de probabilidade* (f.d.p.) da v.a. X, se e somente se:

1. $f(x) \geq 0, \ \forall \ x$;
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Para variáveis aleatórias contínuas, a probabilidade de ocorrerem valores no intervalo $[a, b]$ é calculada através da integral da função densidade de probabilidade, $f(x)$, entre a e b, ou seja,

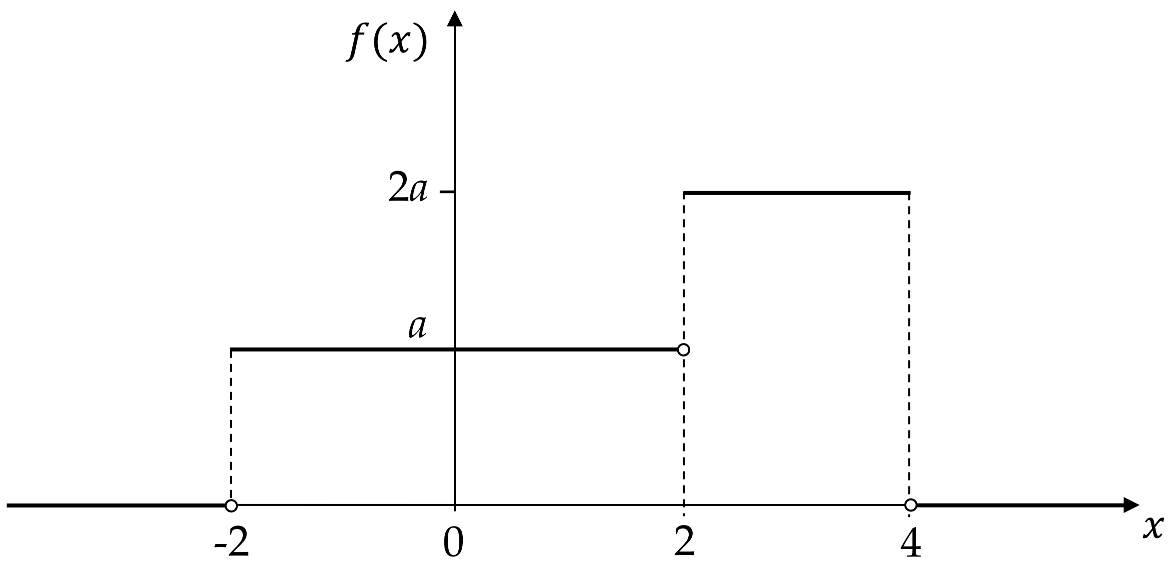
$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Densidade de Probabilidade Acumulada

A função densidade de probabilidade acumulada, denotada por $F(x)$, fornece a probabilidade da v.a. X tomar valores menores ou iguais a x, ou seja,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\, dx$$

Exemplo: O gráfico a seguir representa a densidade de uma variável aleatória X.



(a) Determine o valor de a.

(b) Mostre que esta é uma função densidade de probabilidade.

(c) Determine a função de densidade acumulada de X.

Esperança de X: $E(X)$

Se X é uma variável aleatória contínua cuja função densidade de probabilidade é dada por $f(x)$, então a média desta v.a., representada por $E(X)$ ou μ, é calculada como

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Variância de X: $Var(X)$

Se X é uma variável aleatória contínua cuja função densidade de probabilidade é dada por $f(x)$, então a variância desta v.a., representada por $Var(X)$ ou σ^2, é calculada como

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2$$

Exemplo: Dada a função densidade de probabilidade da v.a. X:

$$f(x) = \begin{cases} \frac{1}{8}, & -2 \leq x < 2 \\ \frac{2}{8}, & 2 \leq x < 4 \\ 0, & c.c. \end{cases}$$

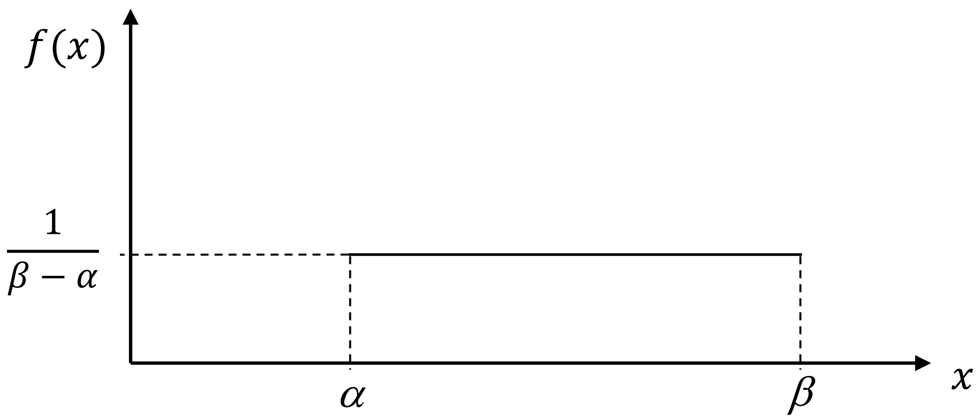
(a) Determine o valor esperado de X.

(b) Determine a variância e o desvio-padrão da v.a. X.

5.2 Distribuição Uniforme

O modelo uniforme é o modelo mais simples para v.a. contínuas. Dizemos que uma v.a. X tem *distribuição uniforme* no intervalo $[\alpha,\beta]$ – ou seja, $X \sim U(\alpha;\beta)$ – se sua função de densidade de probabilidade é dada por

$$f(x) = \begin{cases} \dfrac{1}{\beta-\alpha} & \alpha \leq x \leq \beta \\ 0 & \text{caso contrário} \end{cases}$$



O valor esperado e a variância de uma distribuição Uniforme Contínua são

$$E(X) = \frac{\alpha+\beta}{2} \qquad \text{e} \qquad Var(X) = \frac{(\beta-\alpha)^2}{12}$$

Exemplo: Sabe-se que a variável aleatória X está distribuída uniformemente entre 1,0 e 1,5.

(a) Apresente o gráfico da função densidade de probabilidade.

(b) Calcule $P(X = 1{,}25)$.

(c) Calcule $P(1{,}0 \leq X \leq 1{,}25)$.

(d) Calcule $P(1{,}0 < X < 1{,}25)$.

(e) Calcule $P(1{,}2 \leq X \leq 1{,}4)$.

(f) Calcule $P(1{,}35 \leq X \leq 1{,}55)$;

(g) Calcule $P(X > 1{,}35)$.

Exemplo: Admite-se que uma pane pode ocorrer em qualquer ponto de uma rede elétrica de 10 quilômetros.

(a) Qual é a probabilidade da pane ocorrer nos primeiros 500 metros? E de ocorrer nos 3 km centrais da rede?

(b) O custo de reparo da rede depende da distância do centro de serviço ao local da pane. Considere que o centro de serviço está na origem da rede e que o custo é de R$ 200 para distâncias até 3 km, de R$ 400 entre 3 e 8 km e de R$ 1.000 para as distâncias acima de 8 km. Qual é o custo médio do conserto?

5.3 Distribuição Exponencial

A distribuição Exponencial tem forte relação com a distribuição de Poisson. Enquanto esta última é usada para determinar o número de ocorrências do evento de interesse em um intervalo contínuo, a distribuição Exponencial é usada para determinar o intervalo (de tempo ou de distância) até que o evento de interesse ocorra. Portanto, a distribuição Exponencial somente poderá ser usada se as suposições da distribuição de Poisson forem satisfeitas (taxa média de ocorrência constante no intervalo; independência entre as ocorrências dos eventos).

Na distribuição Exponencial, X = *distância entre ocorrências sucessivas do evento de interesse em um processo Poisson.* A notação é $X \sim Exp(\lambda)$, onde λ é a taxa média de ocorrência por intervalo. Sua função de densidade de probabilidade é dada por

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

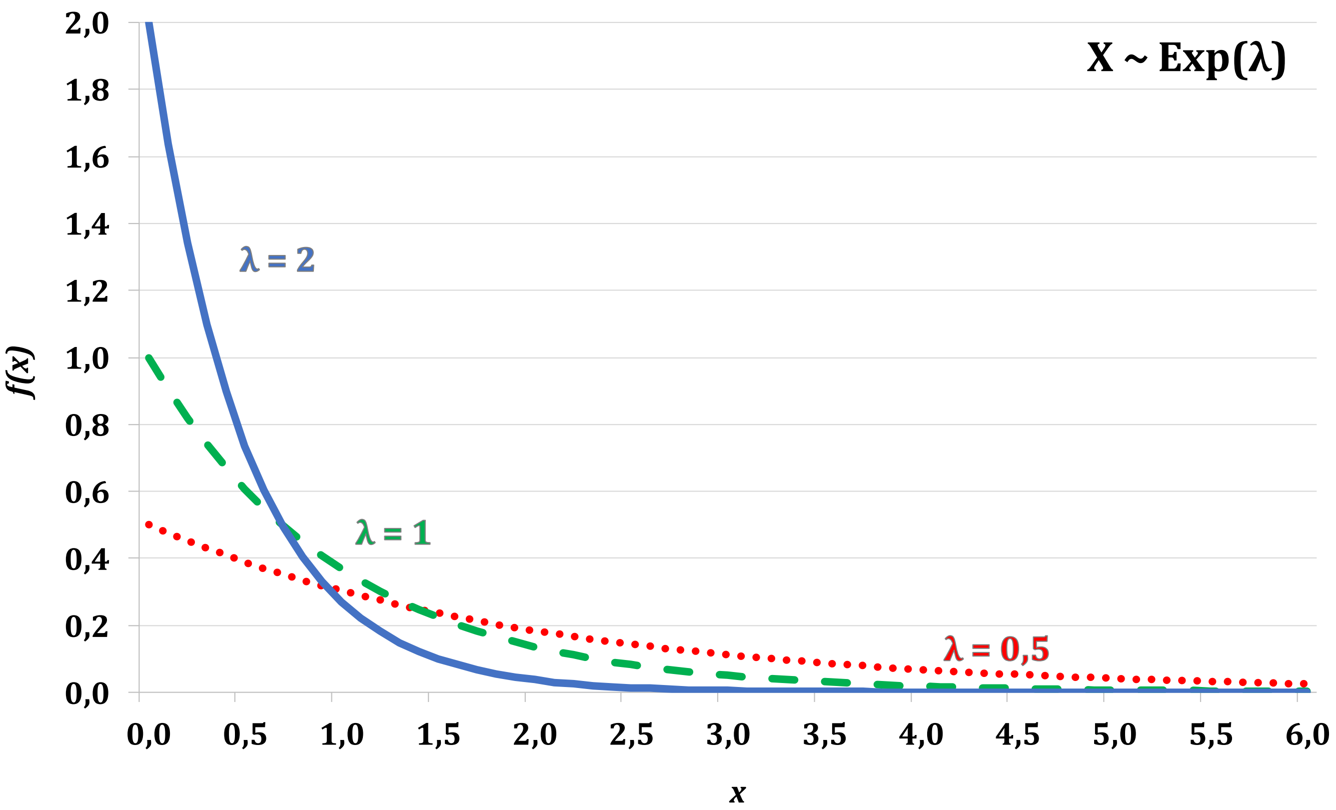
Se a distribuição Exponencial for usada para modelar a vida útil de um equipamento, então λ representa a *taxa de falha*.

O valor esperado e a variância de uma distribuição Exponencial são

$$E(X) = \frac{1}{\lambda} \qquad \text{e} \qquad Var(X) = \frac{1}{\lambda^2}$$

Exemplos de uso da distribuição Exponencial:

- X = distância entre os buracos de uma rodovia;
- X = tempo entre as chegadas de aeronaves em um aeroporto;
- X = tempo até a próxima consulta a uma base de dados;
- X = tempo entre as falhas de um equipamento;
- X = distância entre os defeitos de uma fita magnética;
- X = tempo entre as chegadas de veículos ao posto de pedágio de uma estrada.



Se $X \sim Exp(\lambda)$, o cálculo de probabilidades é facilitado com o uso da distribuição de probabilidade acumulada da distribuição Exponencial, dada por

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Exemplo: (Adaptado de Kokoska, 2013) Uma bomba de aquecimento, projetada para aquecer uma casa nos dias frios de inverno e refrigerar a casa no verão, dura 16 anos em média. Se a vida útil (em anos) desse equipamento pode ser modelada por uma variável aleatória exponencial, responda:

(a) Qual é a probabilidade de que a bomba dure, pelo menos, 5 anos?

(b) Suponha que a bomba já tenha durado 5 anos. Qual é a probabilidade de que dure por, pelo menos, outros 5 anos?

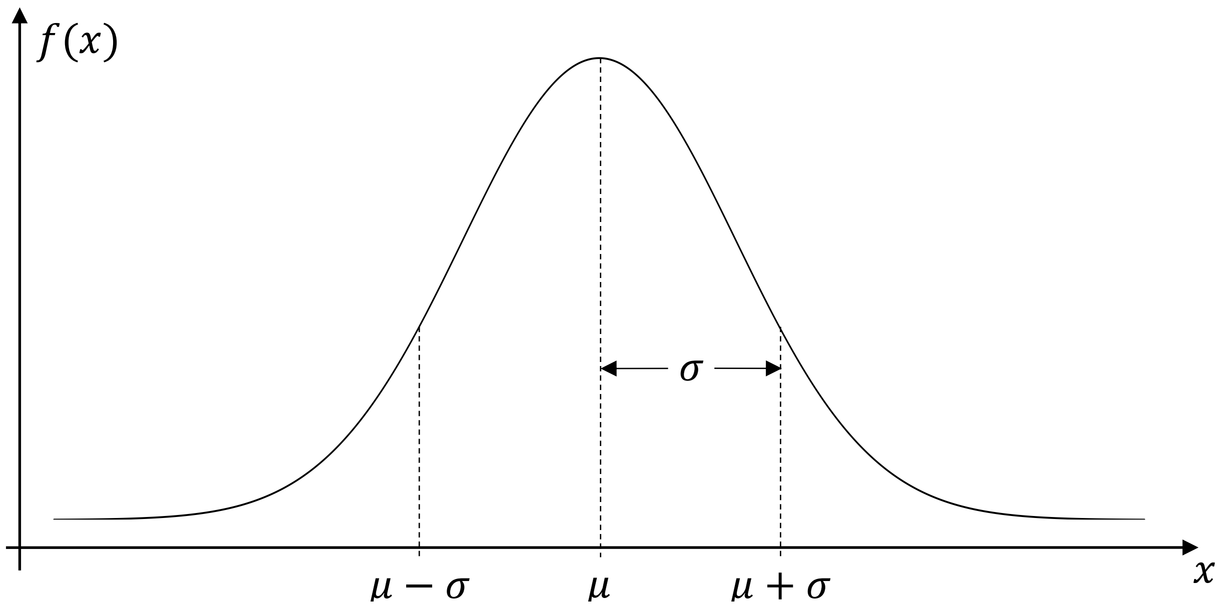
5.4 Distribuição Normal (ou de Gauss)

Se a v.a. X tem distribuição normal com parâmetros μ e σ, sua função densidade de probabilidade é dada por

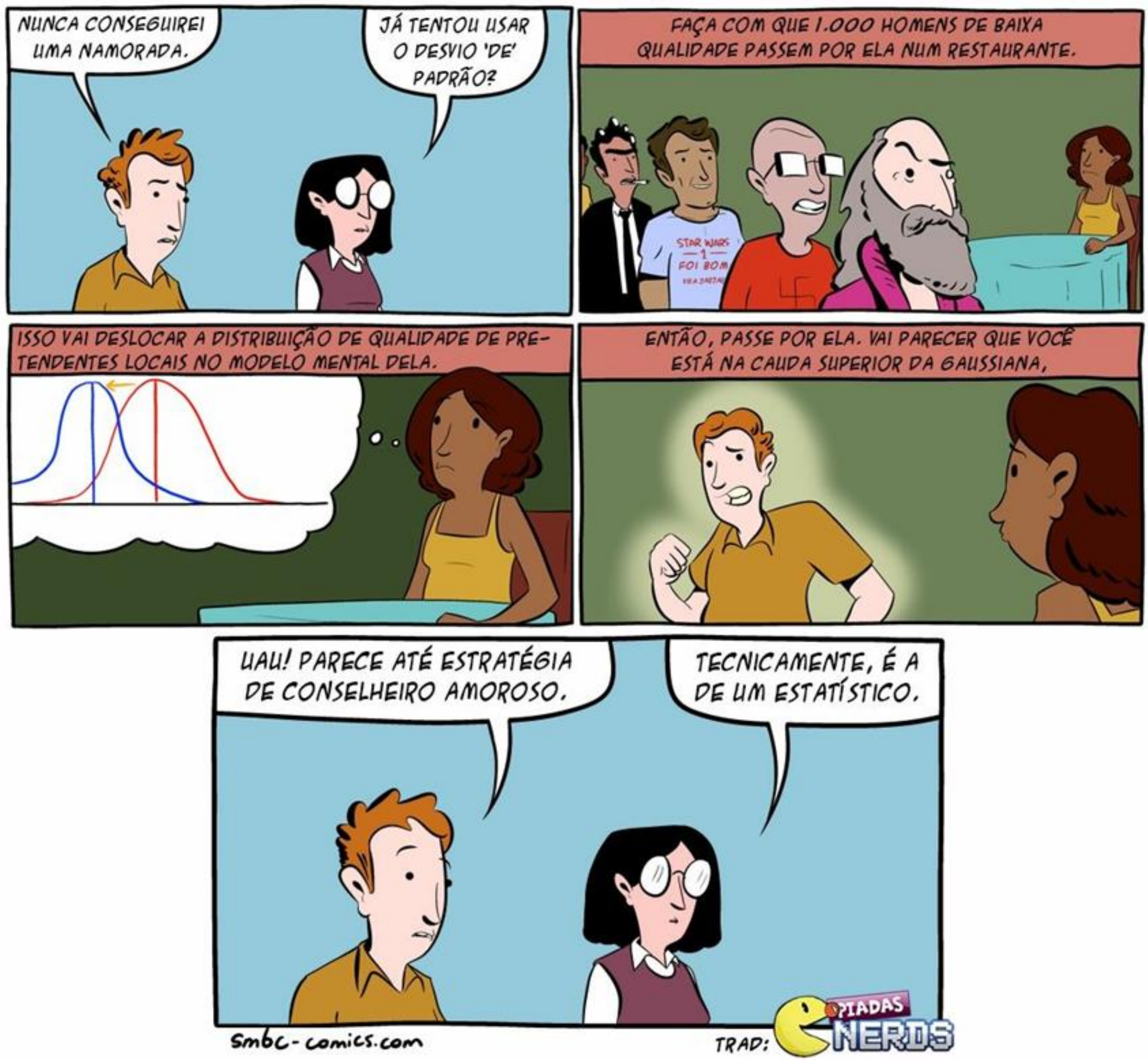
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty$$

Propriedades:

- A curva da distribuição tem forma de sino;



- $f(x)$ é simétrica em relação a μ;
- A área total sob a curva é igual a 1 (ou 100%);
- $f(x) \to 0$ quando $x \to \pm\infty$;
- O valor máximo de $f(x)$ ocorre em $x = \mu$;
- $\mu - \sigma$ e $\mu + \sigma$ são os pontos de inflexão da curva.



<u>*Distribuição Normal Padrão (ou reduzida):*</u>

Quando $\mu = 0$ e $\sigma = 1$, temos uma distribuição normal padrão ou reduzida, ou simplesmente $N(0,1)$. Se $X \sim N(\mu, \sigma)$, então a v.a. definida por

$$Z = \frac{X - \mu}{\sigma}$$

terá média igual a zero e variância igual a um, ou seja, $Z \sim N(0,1)$.

Exemplo: Dada uma v.a. $Z \sim N(0,1)$, calcule as seguintes probabilidades:

(a) $P(0 \leq Z \leq 2{,}01)$

(b) $P(0 < Z < 2{,}01)$

(c) $P(-1{,}93 < Z < 0)$

(d) $P(-0{,}64 < Z < 1{,}32)$

(e) $P(Z > -0{,}99)$

(f) $P(Z < -3{,}31)$

(g) $P(1{,}20 \leq Z \leq 2{,}20)$

Exemplo: A durabilidade de um tipo de pneu da marca *Rodabem* é descrita por uma variável aleatória Normal de média 60.000 km e desvio-padrão de 8.300 km.

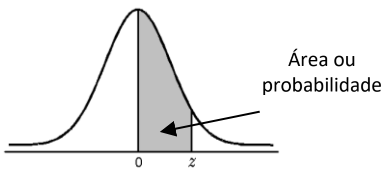
(a) Se a *Rodabem* garante os pneus pelos primeiros 48.000 km, qual a proporção de pneus que deverá ser trocada pela garantia?

(b) O que aconteceria com a proporção do item (a) se a garantia fosse entre 45.000km e 48.000km?

(c) Qual deveria ser a garantia (em km) de tal forma a assegurar que o fabricante trocaria sob garantia no máximo 2% dos pneus?

(d) Se você comprar 4 pneus Rodabem, quantos você esperaria trocar durante a garantia (45.000 km)?

Tabela da Distribuição Normal-Padrão

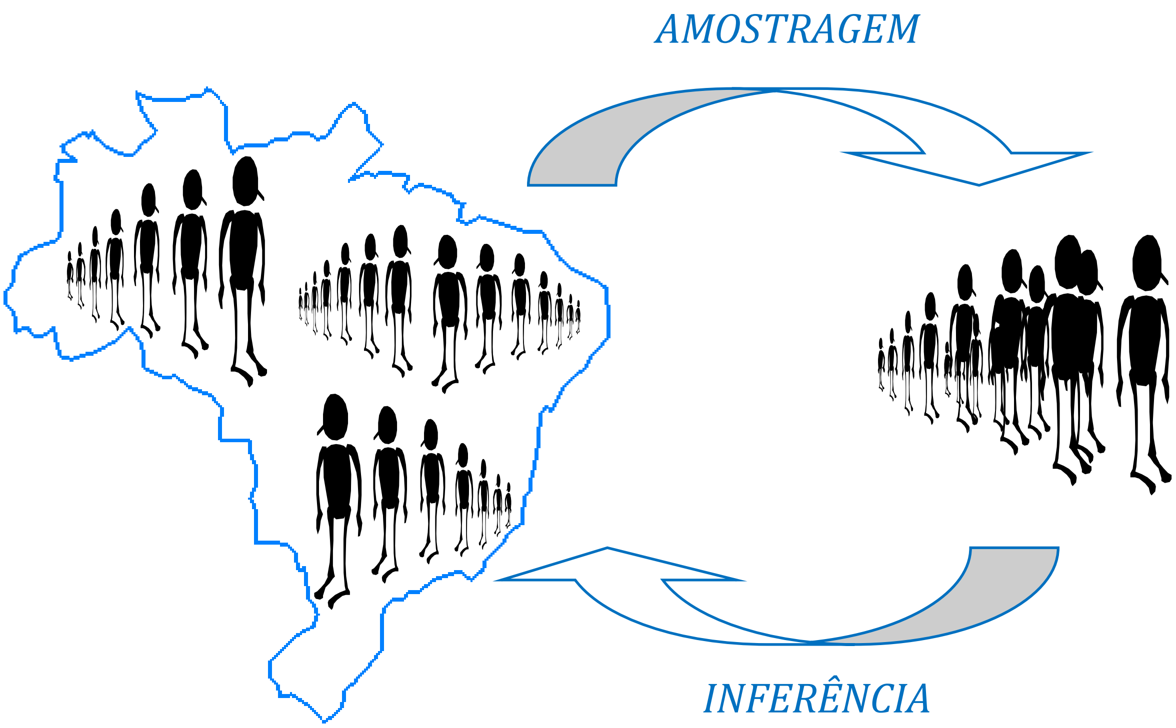


Parte inteira e primeira decimal de *z*	Segunda decimal de *z* 0	1	2	3	4	5	6	7	8	9	Parte inteira e primeira decimal de *z*
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586	**0,0**
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535	**0,1**
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409	**0,2**
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173	**0,3**
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793	**0,4**
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240	**0,5**
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490	**0,6**
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524	**0,7**
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327	**0,8**
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891	**0,9**
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214	**1,0**
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298	**1,1**
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147	**1,2**
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774	**1,3**
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189	**1,4**
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408	**1,5**
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449	**1,6**
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327	**1,7**
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062	**1,8**
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670	**1,9**
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169	**2,0**
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574	**2,1**
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899	**2,2**
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158	**2,3**
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361	**2,4**
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520	**2,5**
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643	**2,6**
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736	**2,7**
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807	**2,8**
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861	**2,9**
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900	**3,0**
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929	**3,1**
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950	**3,2**
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965	**3,3**
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976	**3,4**
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983	**3,5**
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989	**3,6**
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992	**3,7**
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995	**3,8**
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997	**3,9**
4,0	0,49997	0,49997	0,49997	0,49997	0,49997	0,49997	0,49998	0,49998	0,49998	0,49998	**4,0**
4,5	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	**4,5**

VI INFERÊNCIA ESTATÍSTICA

6.1 Estimação de parâmetros

Exemplo: Um fabricante de lâmpadas garante que a vida média de um determinado tipo de lâmpada é de pelo menos 750 horas. Se uma amostra ao acaso com 26 lâmpadas tiver uma vida média de 745 horas e desvio-padrão de 60 horas, você tem evidência suficiente para rejeitar a alegação do fabricante?



Inferência Estatística

Procedimentos para fazer generalizações sobre as características de uma população a partir da informação contida na amostra.

Estimação

É o processo que consiste em utilizar dados amostrais para estimar os valores de parâmetros populacionais desconhecidos. Qualquer característica de uma população pode ser estimada a partir de uma amostra *aleatória*. Exemplo: média, desvio padrão e proporção.

Parâmetros

São as características populacionais, geralmente desconhecidas:

- μ = média populacional;
- σ = desvio padrão populacional;
- p = proporção populacional.

Como o cálculo destes parâmetros é quase impossível, estimamos os valores a partir de uma amostra da população.

Estimadores (ou Estatísticas)

São as características amostrais usadas para estimar os parâmetros:

Parâmetro	**Estimador**
μ	$\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$
σ	$s = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}}$
p	$\hat{p} = \dfrac{\text{nº de elementos com a característica de interesse}}{n}$

Estimativa

Os valores dos estimadores ($\bar{x}$, s e $\hat{p}$), observados na amostra, são chamados de *estimativas pontuais* dos parâmetros.

As estimativas pontuais nem sempre são iguais aos valores populacionais, embora estes possam ser bem próximos.

Devido à variabilidade amostral, deve-se incluir uma *estimativa intervalar* para acompanhar a estimativa pontual.

Estimativa intervalar

Fornece um intervalo de valores possíveis, no qual se admite, com certa confiança, que contenha o verdadeiro valor do parâmetro populacional.

6.2 Distribuições Amostrais dos Estimadores

Quando selecionamos uma amostra aleatória simples, qualquer estimador associado à amostra é uma variável aleatória. A distribuição destas variáveis aleatórias amostrais se aproxima de distribuições contínuas conhecidas à medida que a o tamanho da amostra cresce.

Distribuição Amostral da Média

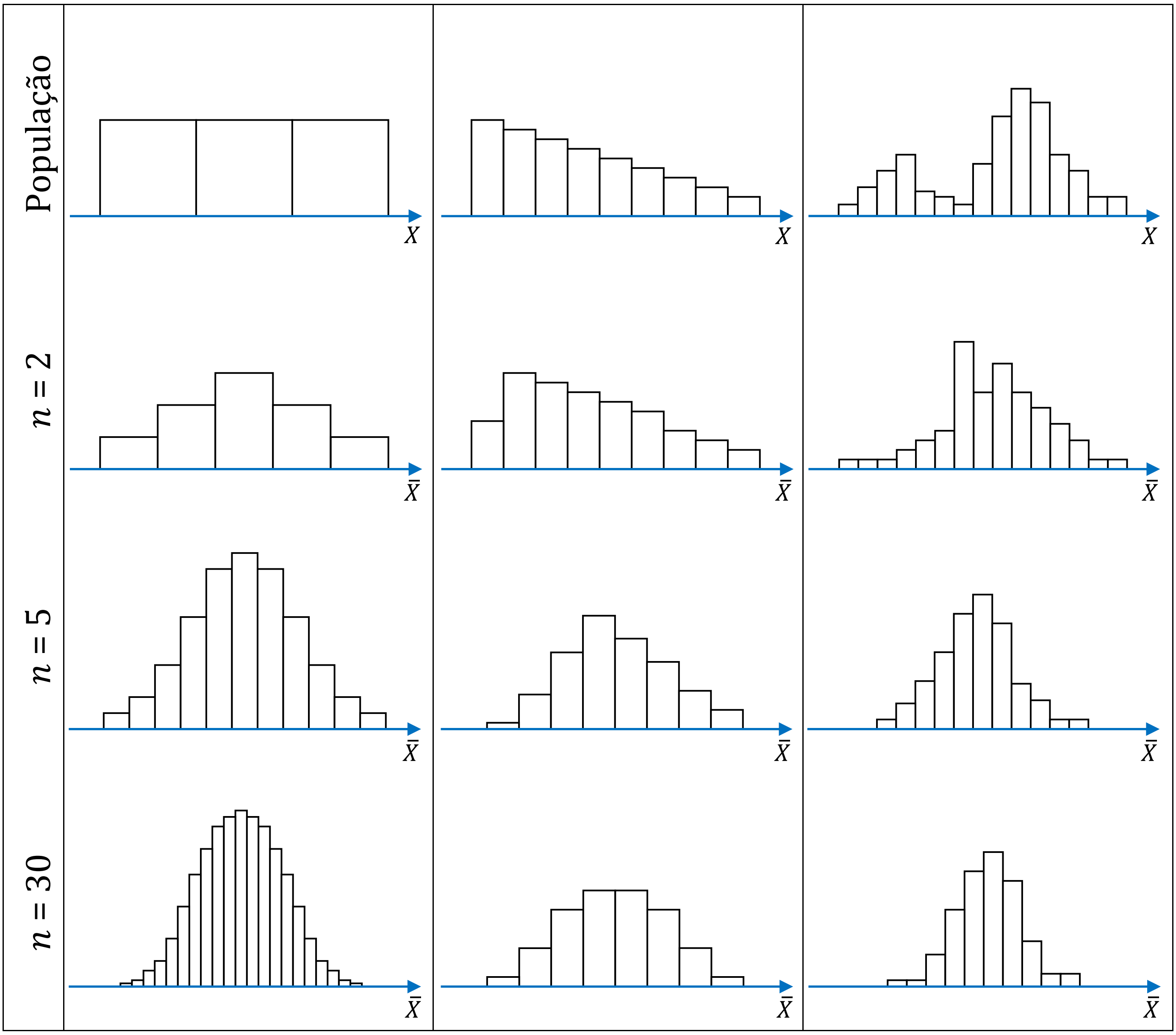
Considere uma amostra de tamanho n de uma variável aleatória X com média $\mu = E(X)$ e desvio-padrão $\sigma = \sqrt{Var(X)}$ que vamos denotar por $X_1, X_2, \cdots X_n$.

Sabemos que tal amostra pode ser proveniente de uma distribuição conhecida (como Binomial ou Normal), ou desconhecida. Mas qual seria a distribuição do estimador $\bar{X}$? A distribuição da média depende da distribuição da amostra. Podemos demonstrar esse fato através do *Teorema Central do Limite*.

Teorema Central do Limite

Distribuição de X	**Tamanho da amostra**	**Distribuição de $\bar{X}$**
$X \sim N(\mu; \sigma)$	Qualquer	$\bar{X} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$
Qualquer	$n \geq 30$	$\bar{X} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$

Distribuições amostrais de três populações diferentes:



Distribuição Amostral da Proporção

Considere uma população em que a proporção de elementos com a característica de interesse é p. Logo, podemos definir uma v.a. X, da seguinte maneira:

$$X = \begin{cases} 1, & \text{se o indivíduo tiver a característica} \\ 0, & \text{se o indivíduo não tiver a característica} \end{cases}$$

Logo,

$$\mu = E(X) = p \quad \text{e} \quad \sigma^2 = Var(X) = p(1-p).$$

Então, pelo Teorema Central do Limite, para n grande podemos considerar que

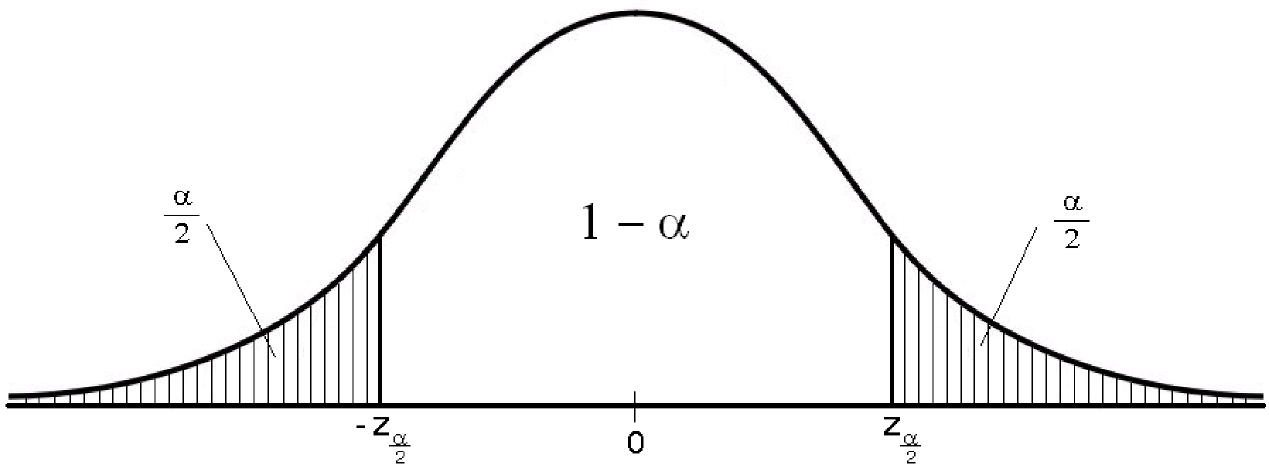
$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

6.3 Intervalos de Confiança

Com a construção de intervalos de confiança, agregamos ao estimador pontual informação sobre sua variabilidade. Um intervalo de confiança consiste em um limite inferior e um limite superior, construídos através de duas partes: uma estimativa pontual para o parâmetro de interesse ± um valor que descreve a precisão da estimativa. Esse valor é chamado de *margem de erro*.

Intervalos de Confiança para μ (média populacional)

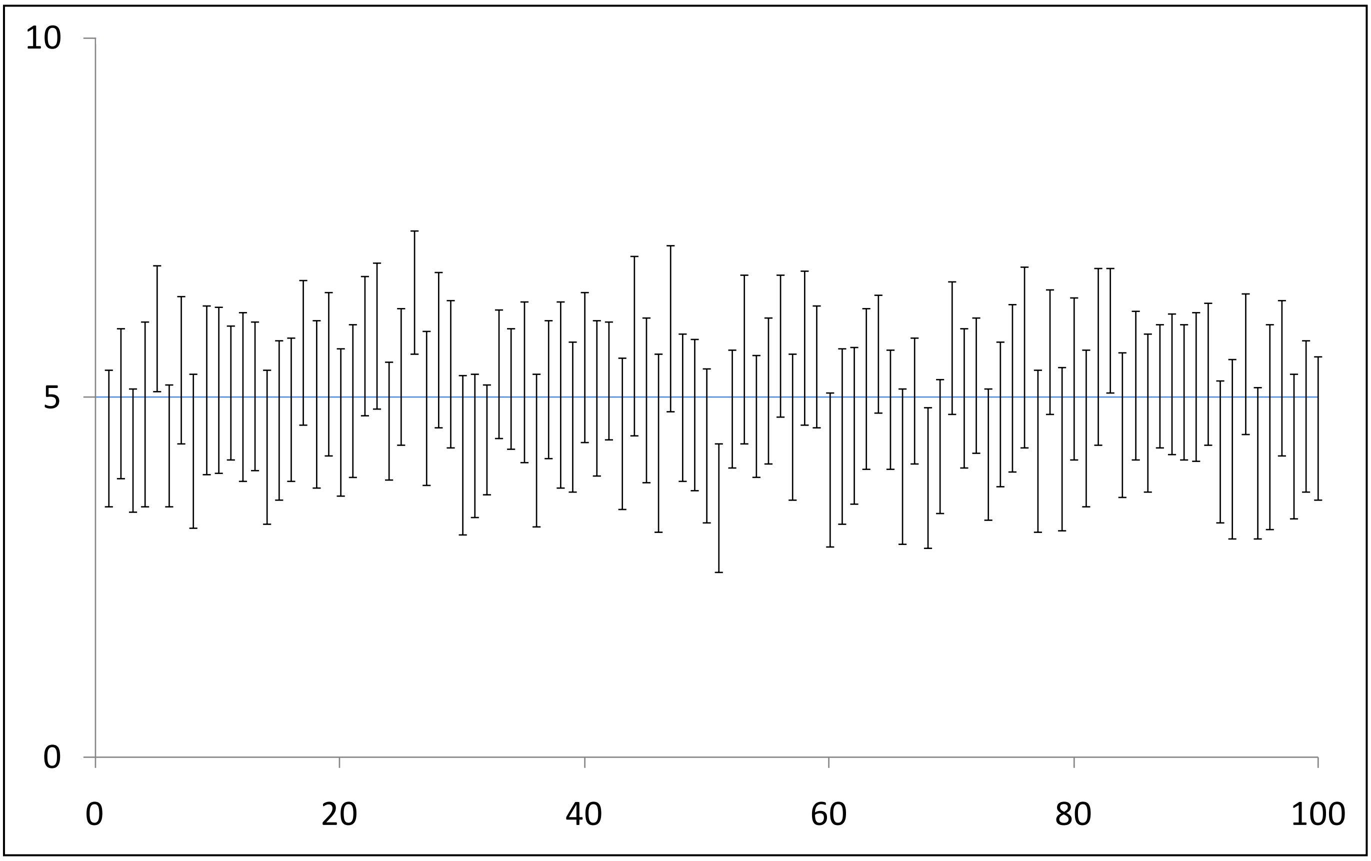
Seja $z_{\frac{\alpha}{2}}$ o valor de z tal que $P\left(Z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$, ou seja,



100(1– α)% é denominado *coeficiente de confiança* ou *nível de confiança*, ou ainda, *índice de confiança (IC)*.

Exemplo: A medida de uma peça usinada tem distribuição Normal com média igual a 5 e desvio-padrão igual a 3. Foram retiradas 100 amostras diferentes de tamanho n = 30 desta população, e calculados intervalos de 95% de confiança para cada uma destas amostras.

Os resultados dos 100 intervalos obtidos são apresentados na figura a seguir:



- ***Caso 1: População Normal, σ conhecido***

Neste caso, um intervalo de $100(1-\alpha)\%$ de confiança para μ é dado por

$$\bar{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

ou simplesmente

$$\boxed{\bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)}$$

Exemplo: Uma máquina automática de refrescos é regulada de modo que a quantidade suprida de cada vez tenha distribuição aproximadamente normal, com desvio padrão de 13ml. Uma amostra de 30 copos de refresco acusou conteúdo médio de 210ml.

(a) Determine um intervalo de 95% de confiança para a quantidade média de todos os refrescos servidos.

(b) Determine um intervalo de 99% de confiança para a quantidade média de todos os refrescos servidos.

(c) Mantendo a confiança de 95%, quantos copos a mais deveriam ser incluídos na amostra para que a margem de erro fosse de, no máximo, 3ml.

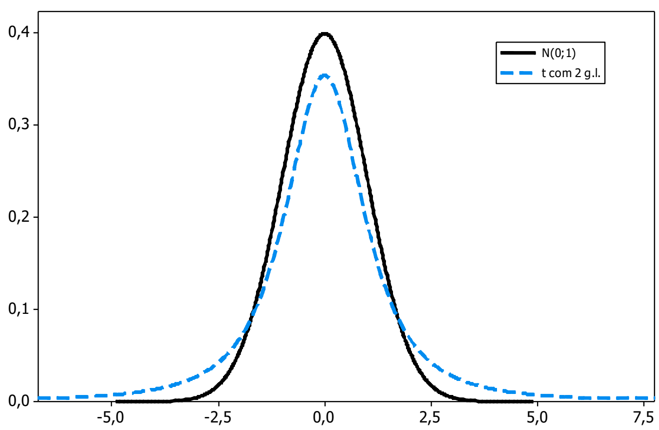
- ***Caso 2: População Normal, σ desconhecido***

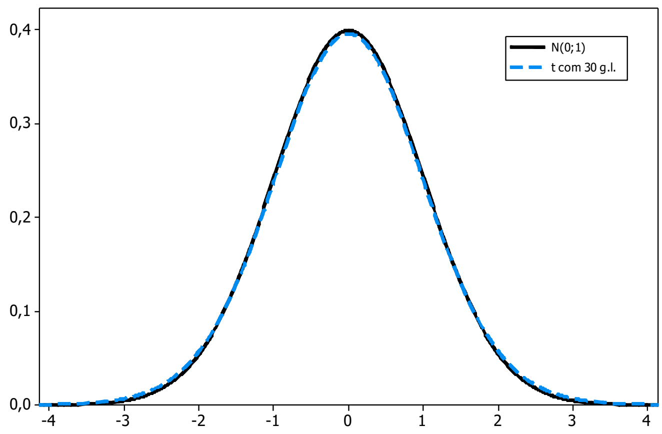
Neste caso, como não conhecemos o σ, podemos estimá-lo pelo desvio-padrão da amostra (s). Neste caso, temos uma nova estatística:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Essa estatística tem distribuição conhecida como *<u>Distribuição t de Student</u>* com $n - 1$ graus de liberdade, onde n é o tamanho da amostra.

A distribuição t é muito parecida com a Normal Padrão (tem forma de sino e é simétrica em zero), mas possui maior dispersão por causa da substituição de σ por s. No entanto, à medida que n cresce, a distribuição t aproxima-se da $N(0; 1)$, pois s aproxima-se de σ.

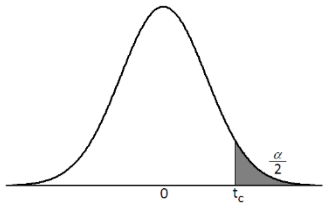




Portanto, no caso de σ desconhecido, podemos estimar o erro padrão da média $\left(\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\right)$ utilizando o s, ou seja,

$$\hat{\sigma}_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}}$$

Tabela da Distribuição t de Student



g.l.	Valores t_C tais que $P(T > t_C) = \frac{\alpha}{2}$															g.l.
	45%	40%	35%	30%	25%	20%	15%	10%	5%	2,5%	2%	1%	0,5%	0,1%	0,05%	
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	15,895	31,821	63,657	318,31	636,62	1
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	4,849	6,965	9,925	22,327	31,599	2
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	3,482	4,541	5,841	10,215	12,924	3
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	2,999	3,747	4,604	7,173	8,610	4
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	2,757	3,365	4,032	5,893	6,869	5
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	2,612	3,143	3,707	5,208	5,959	6
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,517	2,998	3,499	4,785	5,408	7
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,449	2,896	3,355	4,501	5,041	8
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,398	2,821	3,250	4,297	4,781	9
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,359	2,764	3,169	4,144	4,587	10
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,328	2,718	3,106	4,025	4,437	11
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,303	2,681	3,055	3,930	4,318	12
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,282	2,650	3,012	3,852	4,221	13
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,264	2,624	2,977	3,787	4,140	14
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,249	2,602	2,947	3,733	4,073	15
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,235	2,583	2,921	3,686	4,015	16
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,224	2,567	2,898	3,646	3,965	17
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,214	2,552	2,878	3,610	3,922	18
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,205	2,539	2,861	3,579	3,883	19
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,197	2,528	2,845	3,552	3,850	20
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,189	2,518	2,831	3,527	3,819	21
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,183	2,508	2,819	3,505	3,792	22
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,177	2,500	2,807	3,485	3,768	23
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,172	2,492	2,797	3,467	3,745	24
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,167	2,485	2,787	3,450	3,725	25
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,162	2,479	2,779	3,435	3,707	26
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,158	2,473	2,771	3,421	3,690	27
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,154	2,467	2,763	3,408	3,674	28
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,150	2,462	2,756	3,396	3,659	29
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,147	2,457	2,750	3,385	3,646	30
35	0,127	0,255	0,388	0,529	0,682	0,852	1,052	1,306	1,690	2,030	2,133	2,438	2,724	3,340	3,591	35
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,123	2,423	2,704	3,307	3,551	40
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,109	2,403	2,678	3,261	3,496	50
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,099	2,390	2,660	3,232	3,460	60
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,076	2,358	2,617	3,160	3,373	120
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,054	2,326	2,576	3,090	3,291	∞
	45%	40%	35%	30%	25%	20%	15%	10%	5%	2,5%	2%	1%	0,5%	0,1%	0,05%	

Assim, um intervalo de 100(1– α)% de confiança para μ é dado por

$$\bar{x} - t_{\left(n-1;\frac{\alpha}{2}\right)}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{x} + t_{\left(n-1;\frac{\alpha}{2}\right)}\left(\frac{s}{\sqrt{n}}\right)$$

ou simplesmente

$$\boxed{\bar{x} \pm t_{\left(n-1;\frac{\alpha}{2}\right)}\left(\frac{s}{\sqrt{n}}\right)}$$

Exemplo: Um pesquisador está estudando a resistência de um determinado material sob determinadas condições. Ele sabe que essa variável é normalmente distribuída. Utilizando os valores a seguir (em unidades), obtidos de uma amostra de tamanho 9, determine o intervalo de 90% de confiança para a resistência média:

4,9	7,0	8,1	4,5	5,6	6,8	7,2	5,7	6,2

No entanto, quando o tamanho da amostra é grande ($n \geq 30$), o Teorema Central do Limite, permite usar a distribuição Normal para estimar μ.

- ***Caso 3: População Normal, σ desconhecido, grandes amostras***

Neste caso, um intervalo de $100(1-\alpha)\%$ de confiança para μ é dado por

$$\bar{x} - z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)$$

ou simplesmente

$$\boxed{\bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)}$$

Exemplo: De 50.000 válvulas fabricadas por uma companhia retira-se uma amostra de 400 válvulas, e obtém-se a vida média de 800 horas e o desvio padrão de 100 horas.

(a) Qual o intervalo de confiança de 99% para a vida média da população?

(b) Com que confiança pode-se dizer que a vida média é 800 ± 0,98?

(c) Que tamanho deve ter a amostra para que seja de 95% a confiança na estimativa 800 ± 7,84?

Intervalos de Confiança para p (proporção populacional)

Seja X o número de sucessos em n provas de um experimento com distribuição Binomial(n, p). Então, através do TCL podemos afirmar que, para grandes amostras,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0;1).$$

Assim, um intervalo de 100(1– α)% de confiança para p é dado por

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ou simplesmente

$$\boxed{\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Exemplo: Antes de uma eleição, um determinado partido está interessado em estimar a proporção p de eleitores favoráveis ao seu candidato. Uma amostra piloto de tamanho 100 revelou que 60% dos eleitores eram favoráveis ao candidato em questão.

(a) Determine o tamanho da amostra necessário para que o erro cometido na estimação seja de, no máximo, 0,01 com confiança de 80%.

(b) Se na amostra final, com tamanho igual ao obtido em (a), observou-se que 55% dos eleitores eram favoráveis ao candidato em questão, construa um intervalo de 95% de confiança para a proporção p.

(c) Se a pesquisa foi divulgada na TV com margem de erro de 2%, qual a confiança do intervalo divulgado?

Exemplo: Suponha que, em uma amostra de 500 famílias em certa cidade, haja 341 com TV a cabo. Se quisermos estimar o número de famílias que possuem TV a cabo, qual o tamanho da amostra necessário para que tenhamos 95% de confiança em que o erro de nossa estimativa não seja superior a 0,02:

(a) Utilizando a informação das 500 famílias?

(b) Sem utilizar essa informação?

6.4 <u>Determinação do Tamanho da Amostra</u>

<u>*Determinando o tamanho da amostra para estimação de μ:*</u>

$$n = \left(\frac{z_{\frac{\alpha}{2}} \times \sigma}{E} \right)^2$$

onde E é a margem de erro máxima admitida.

<u>*Determinando o tamanho da amostra para estimação de p*</u>

$$n = \hat{p}(1 - \hat{p}) \left(\frac{z_{\frac{\alpha}{2}}}{E} \right)^2$$

onde E é a margem de erro máxima admitida.

6.5 Testes de Hipóteses

Em muitos casos, o pesquisador não está interessado apenas em estimar o parâmetro; está interessado em verificar se o parâmetro de interesse apresenta um valor hipotético, ou se difere em grupos diferentes. Nestes casos, testes de hipóteses são os métodos estatísticos adequados para responder à pergunta do pesquisador.

Exemplos:

(1) O tempo médio de execução de determinada tarefa é de, no máximo, 70 minutos?

(2) A proporção de peças defeituosas produzidas pela turma B é menor que 5%?

(3) A resistência média das peças é de, pelo menos, 100 Mpa?

(4) Uma máquina automática de refrescos é regulada de modo que a quantidade suprida de cada vez tenha distribuição aproximadamente normal, com desvio padrão de 13 ml. Uma amostra de 30 copos de refresco acusou conteúdo médio de 210 ml. Se a máquina é regulada para encher copos de 200 ml, existem evidências estatísticas que comprovem que a máquina está desregulada?

Hipóteses

Hipótese nula (H_0): é a hipótese aceita como verdadeira até que existam evidências estatísticas que comprovem o contrário.

Hipótese alternativa (H_1): é a hipótese de interesse, a hipótese que rejeita a hipótese nula.

	Teste bilateral	Teste unilateral à esquerda	Teste unilateral à direita
Testes para uma média	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$
Testes para uma proporção	$H_0: p = p_0$ $H_1: p \neq p_0$	$H_0: p \geq p_0$ $H_1: p < p_0$	$H_0: p \leq p_0$ $H_1: p > p_0$

Exemplos:

(1) $H_0: \mu \leq 70$
$H_1: \mu > 70$

(2) $H_0: p \geq 5\%$
$H_1: p < 5\%$

(3) $H_0: \mu \geq 100$
$H_1: \mu < 100$

(4) $H_0: \mu = 200$
$H_1: \mu \neq 200$

<u>Erro Tipo I e Erro Tipo II</u>

		A situação real (na população) é que H_0 é:	
		Verdadeira	Falsa
O pesquisador realiza um teste estatístico e decide que H_0 é:	Verdadeira (aceita H_0)	**Decisão correta**	**Erro Tipo II**
	Falsa (rejeita H_0)	**Erro Tipo I**	**Decisão correta**

Exemplos:

(1) *<u>Erro tipo I</u>:* Concluir, através do teste de hipóteses, que o tempo médio de execução de determinada tarefa é maior que 70 minutos, quando na verdade o tempo é igual a 70 minutos;

<u>Erro tipo II</u>: Concluir, através do teste de hipóteses, que o tempo médio de execução de determinada tarefa é igual a 70 minutos, quando na verdade o tempo é maior que 70 minutos.

Nível de significância (α)

É a probabilidade de ocorrência do erro Tipo I, ou seja:

$$\alpha = P(\text{rejeitar } H_0 |\ H_0 \text{ é verdadeira}).$$

Valor Padrão: $\alpha = 0{,}05$ (usual: $\alpha \leq 0{,}10$)

"Aceitamos o risco de que, a cada 100 amostras, a hipótese nula verdadeira será rejeitada 5 vezes."

Estatística de Teste

Valor, calculado com base nos dados amostrais, através do qual decidiremos sobre a rejeição ou não de H_0 após comparação com a região de rejeição.

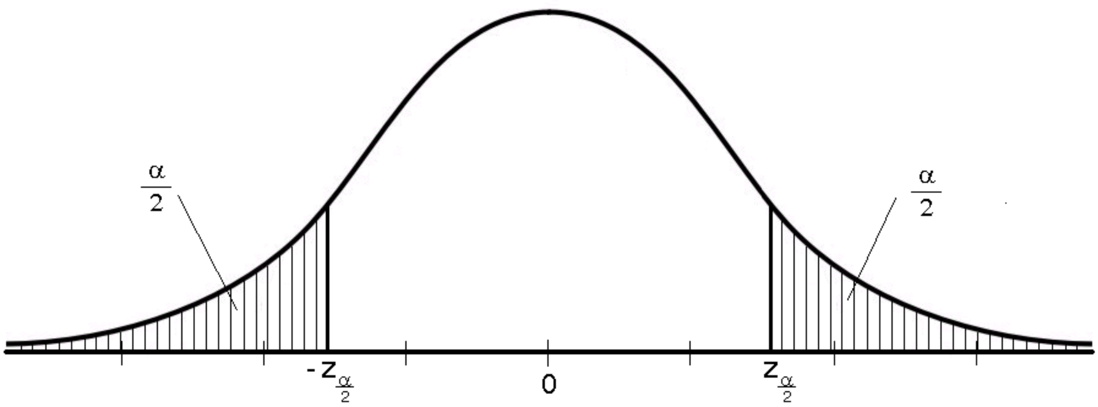
	Testes para uma média	Teste para uma proporção
σ conhecido	$Z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$	
σ desconhecido	$T = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$	
$n \geq 30$	$Z = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$	$Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$

Região de Rejeição (ou Região Crítica)

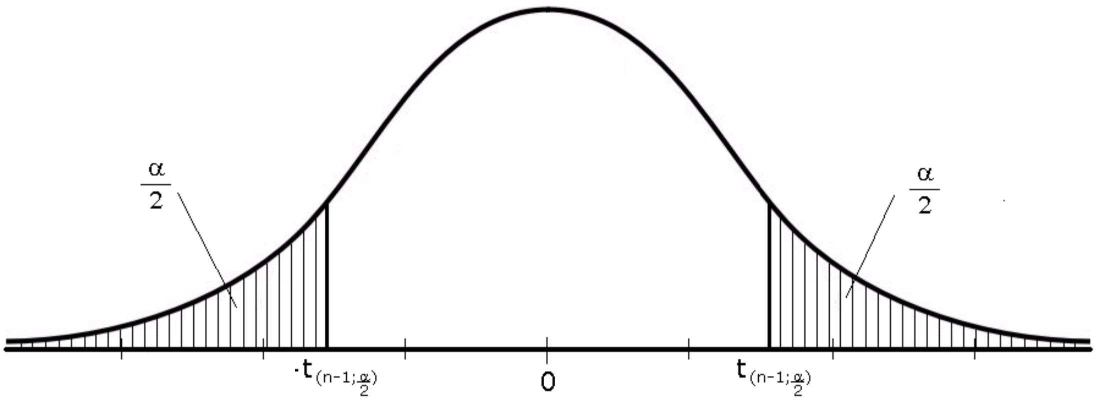
É um valor, obtido através das tabelas das distribuições (Z, T, χ^2, F), que será comparado com o valor da estatística de teste para que se decida sobre a rejeição ou não de H_0.

A região de rejeição depende da hipótese alternativa (bilateral ou unilateral).

- ***No caso dos testes de hipóteses bilaterais, rejeitamos H_0 se $Z < -z_{\frac{\alpha}{2}}$ ou $Z > z_{\frac{\alpha}{2}}$***



ou se $T < -t_{n-1;\frac{\alpha}{2}}$ ou $T > t_{n-1;\frac{\alpha}{2}}$



Exemplo: (adaptado de Montgomery, 2003) O rendimento de um processo químico está sendo estudado. De experiências prévias, sabe-se que o desvio-padrão do rendimento é igual a 3. Os últimos cinco dias de operação da planta resultaram nos seguintes rendimentos:

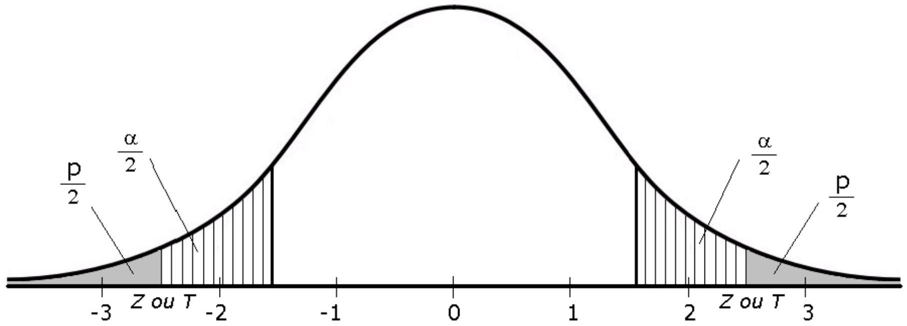
91,6%   88,75%   90,8%   89,95%   91,3%

Usando 10% de significância, há evidência de que o rendimento médio não seja 90%? Calcule o valor-p.

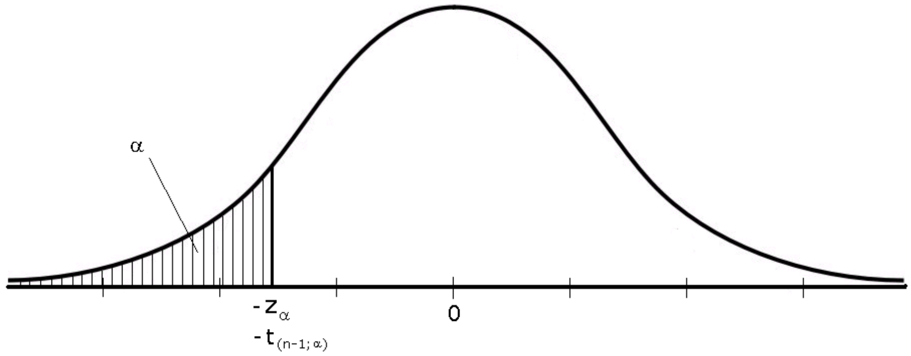
Probabilidade de significância (valor-p)

É a probabilidade de que estejamos cometendo o Erro Tipo I, caso H_0 seja rejeitada. Expressa a probabilidade de ocorrência de valores iguais ou superiores ao assumido pela estatística de teste (em módulo). Se essa probabilidade é menor do que α ($p < \alpha$), existem evidências estatísticas para rejeitar H_0.

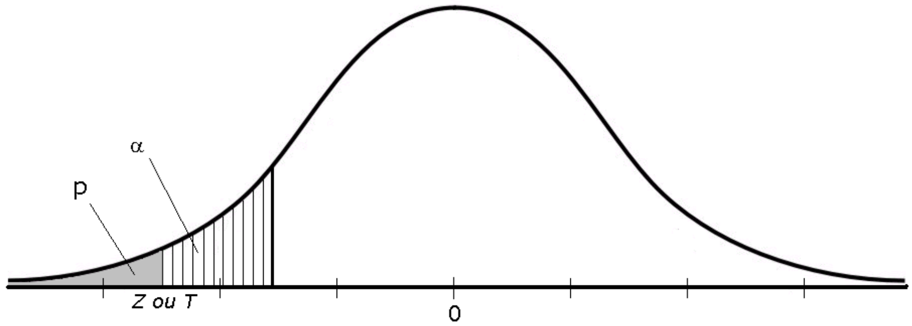
No caso das hipóteses bilaterais, o valor-p é calculado como o dobro da área à esquerda ou à direita de Z ou T.



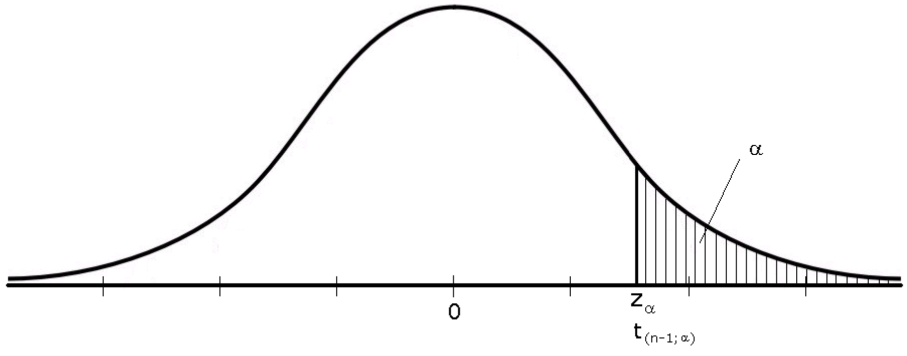
- ***No caso de testes unilaterais à esquerda, rejeitamos H_0 se $Z < -z_\alpha$ ou se $T < -t_{n-1;\alpha}$***



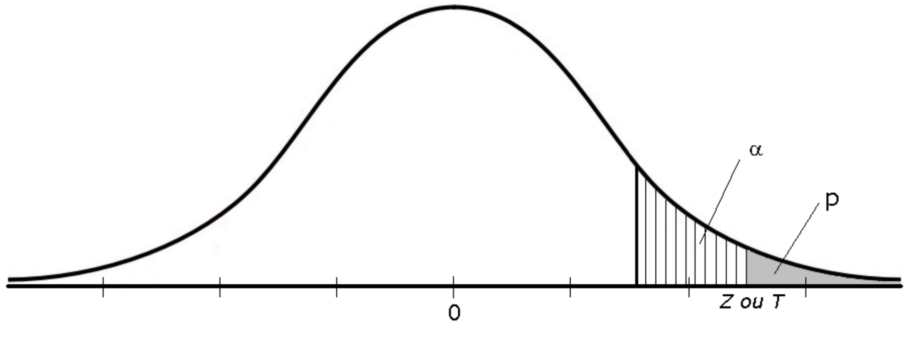
e o valor-p é calculado como a área à esquerda de Z ou T.



- ***De forma análoga, no caso de testes unilaterais à direita, rejeitamos H_0 se $Z > z_\alpha$ ou se $T > t_{n-1;\alpha}$***



e o valor-p é calculado como a área à direita de Z ou T.



Exemplo: (adaptado de Barbetta, 2004) Certo tipo de pneu dura, em média, 50.000 km. O fabricante investiu em uma nova composição de borracha para pneus. Vinte pneus, fabricados com essa nova composição, duraram, em média, 55.000 km, com desvio-padrão de 4.000 km. Supondo que a durabilidade dos pneus segue uma distribuição aproximadamente normal, verificar se os dados provam que os novos pneus são mais duráveis. Tire suas conclusões através da estatística de teste e do valor-p. Use α = 1%.

Exemplo: (adaptado de Montgomery, 2003) Uma amostra aleatória de 50 capacetes de corredores de motos e de automóveis foi sujeita a um teste de impacto, sendo observado algum dano em 18 desses capacetes.

(a) Encontre um intervalo de 95% confiança para a verdadeira proporção de capacetes desse tipo que mostraria algum dano proveniente desse teste.

(b) Usando a estimativa de p, obtida a partir da amostra preliminar de 50 capacetes, quantos capacetes devem ser testados para que estejamos 95% confiantes de que o erro na estimação do verdadeiro valor de p seja menor do que 2%?

(c) Com base nos resultados do teste de impacto, existem evidências de que a proporção de capacetes com danos é menor que 40%?

VII CORRELAÇÃO E REGRESSÃO LINEAR SIMPLES

7.1 Correlação Linear

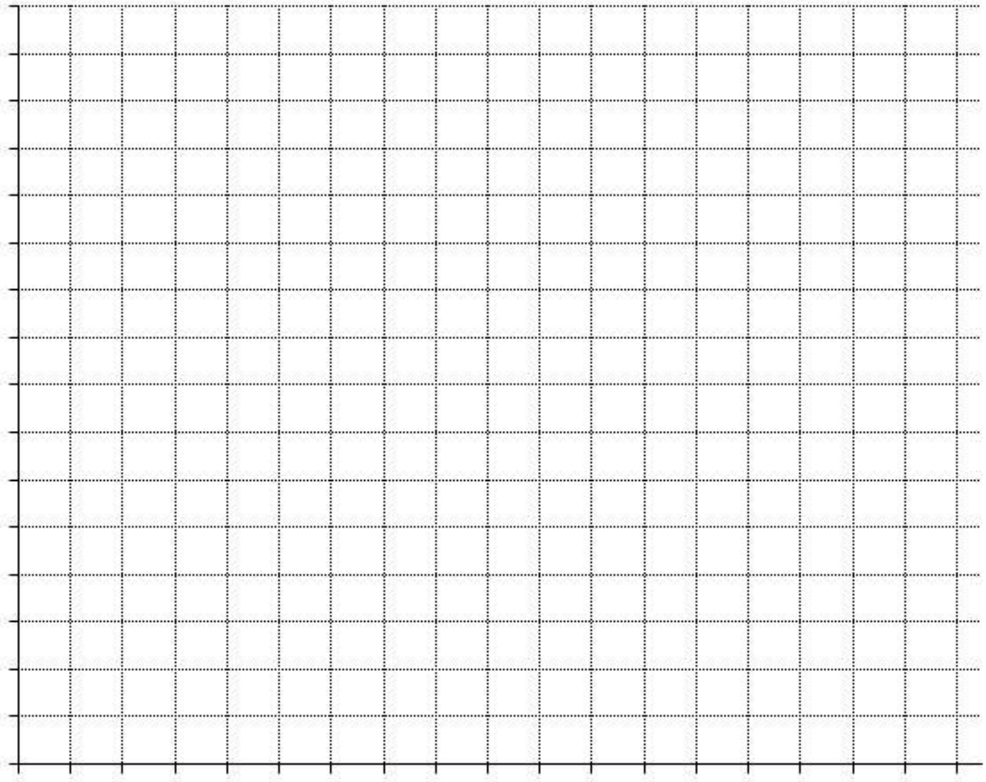
A correlação entre duas variáveis deve ser calculada quando se deseja saber se a variação de uma delas acompanha proporcional ou inversamente a variação da outra.

Diagrama de Dispersão

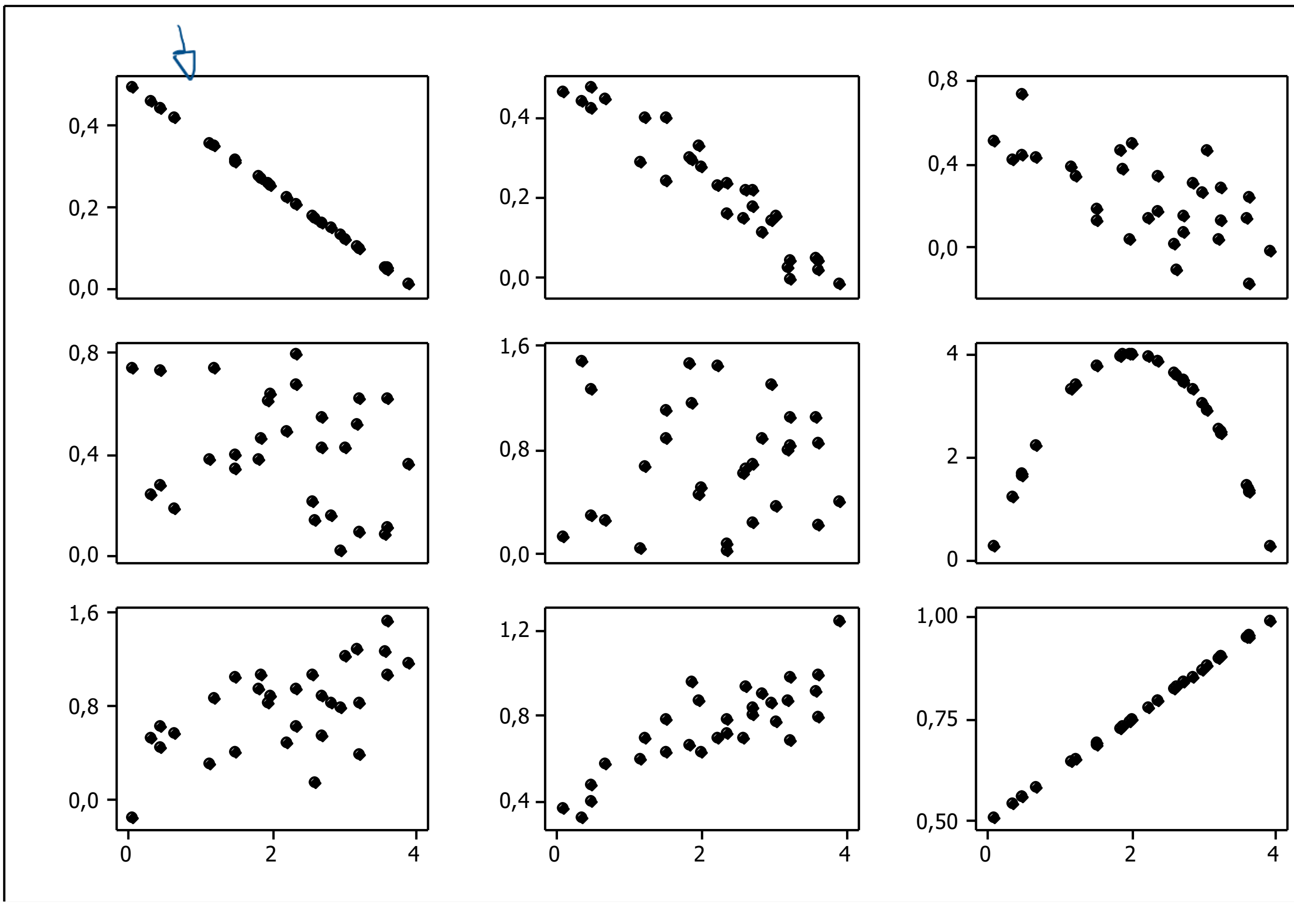
O diagrama de dispersão apresenta os valores das duas variáveis que se deseja correlacionar, cada uma em um eixo (X ou Y), traçados em escalas adequadas para cada variável.

Exemplo: Um estudo do departamento de transportes, sobre a velocidade de condução e o rendimento (medido pela distância percorrida em quilômetros por litro de combustível) para automóveis de tamanho médio, resultou nos seguintes dados:

Velocidade (km/h)	48	80	64	88	48	40	96	40	80	88
Rendimento (km/l)	13,1	10,6	11,9	9,7	12,7	13,5	8,9	14,8	11,0	10,6



Diagramas de dispersão para diferentes associações entre X e Y.



+1

<u>O Coeficiente de Correlação Linear de Pearson</u> ($r_{x,y}$)

Enquanto o diagrama de dispersão é utilizado para visualizar o tipo de relação existente entre as variáveis X e Y, o coeficiente de correlação linear de Pearson ($r_{x,y}$) mede o sentido e a intensidade da relação linear entre elas. O coeficiente de correlação linear de Pearson pode ser calculado pela covariância entre X e Y dividida pelo produto dos desvios-padrão de X e Y, conforme a fórmula que se segue:

$$corr(X,Y) = r_{x,y} = \frac{\text{cov}(X,Y)}{s_x \times s_y}$$

onde

$$cov(X,Y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

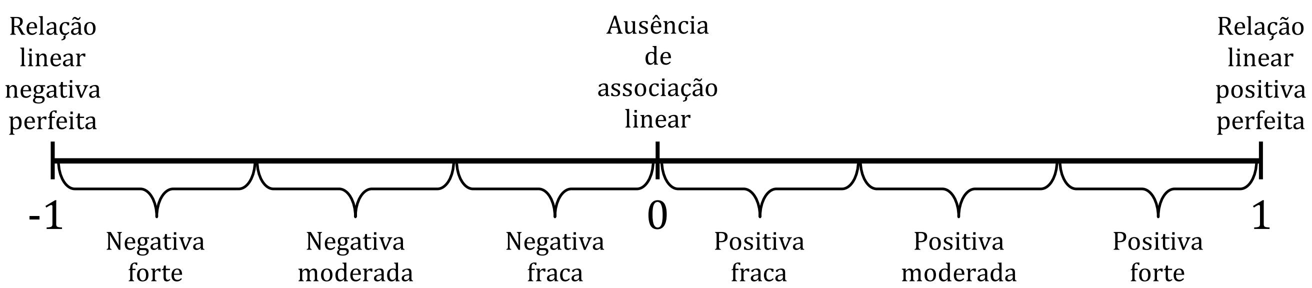
e s_x e s_y são os desvios-padrão de X e Y, respectivamente.

Observe que o sinal de $r_{x,y}$ é determinado pelo sinal da covariância.

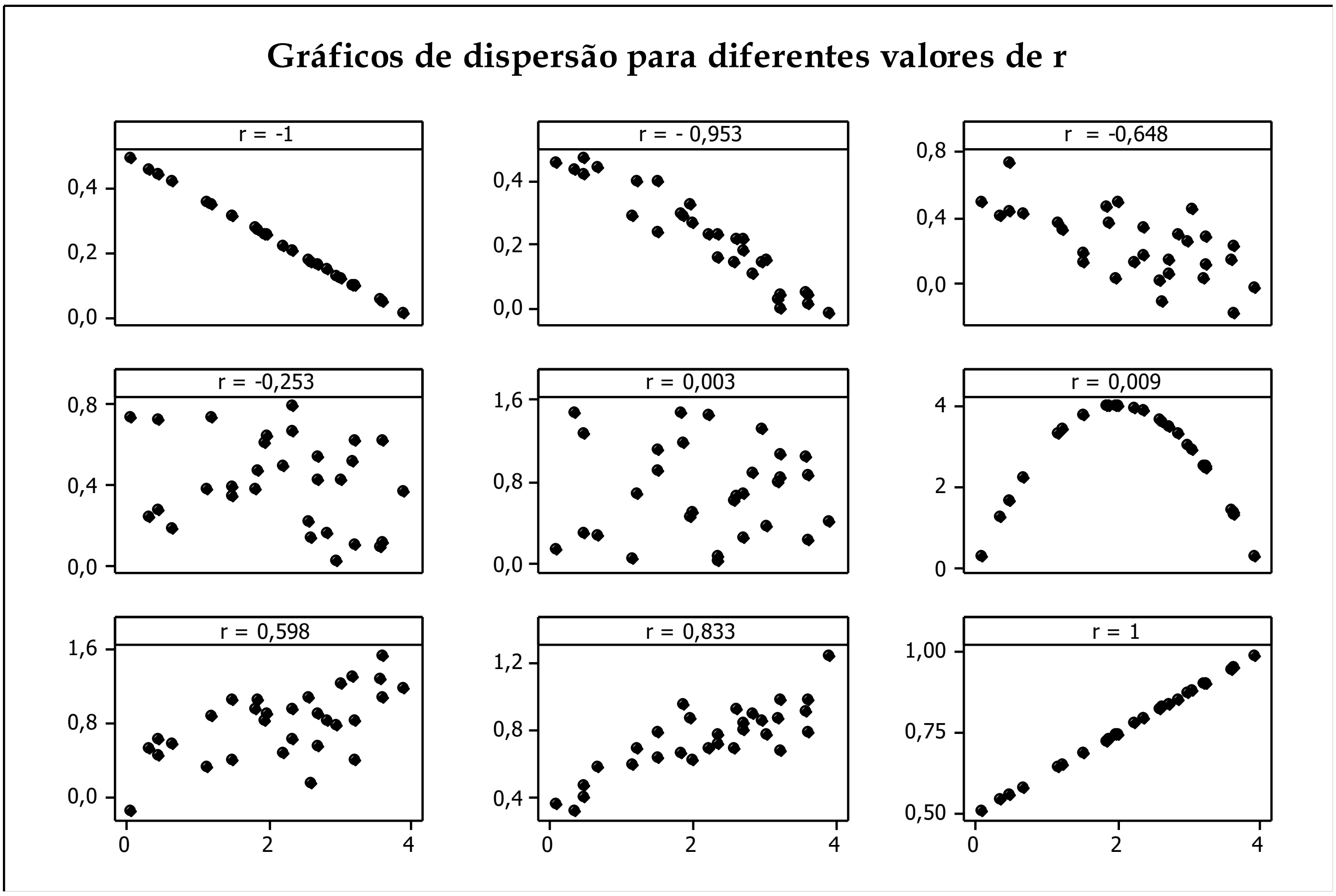
Interpretação do Coeficiente de Correlação:

O coeficiente de correlação satisfaz a condição $-1 \leq r_{x,y} \leq 1$.

- Se $r_{x,y} = -1$, a relação linear entre X e Y é perfeita e inversa;
- Se $r_{x,y} = 0$, não existe associação entre X e Y;
- Se $r_{x,y} = +1$, a relação linear entre X e Y é perfeita e positiva.



Diagramas de dispersão para diferentes valores de $r_{x,y}$.



Exemplo (continuação): Quantifique a associação entre a velocidade de condução e o rendimento para os automóveis de tamanho médio.

7.2 Regressão Linear

Após a análise do diagrama de dispersão e do coeficiente de correlação, se concluirmos que existe uma correlação linear significativa entre duas variáveis, o próximo passo será tentar estimar uma equação que melhor descreva a relação entre essas variáveis. A relação mais simples que conhecemos é aquela descrita pela equação de uma reta.

Uma vez que assumimos a relação linear entre as variáveis, tal relação será descrita pela *equação de regressão*

$$Y = \beta_0 + \beta_1 X$$

onde

Y = variável dependente (ou resposta);

X = variável independente (ou preditora ou explicativa);

β_0 = parâmetro que representa o *coeficiente linear* (ou intercepto) da reta;

β_1 = parâmetro que representa o *coeficiente angular* (ou inclinação) da reta;

Método dos Mínimos Quadrados Ordinais para ajuste da reta

Usando apenas os dados amostrais, não é possível obter os valores exatos dos parâmetros β_0 e β_1. Tais parâmetros serão estimados com base nos dados amostrais. Assim, a *equação de regressão estimada* será dada por

$$\hat{y} = b_0 + b_1 x$$

onde b_0 é uma estimativa de β_0 e b_1 é uma estimativa de β_1.

Os valores de b_0 e b_1 são calculados a partir de

$$b_1 = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

e

$$b_0 = \bar{y} - b_1 \bar{x}.$$

Exemplo (continuação):

(a) Estime a equação de regressão que melhor descreve a associação entre a velocidade de condução e o rendimento para os automóveis de tamanho médio.

(b) Faça uma previsão do *rendimento* (em km/l) para um automóvel que trafega em velocidade média de 45 km/h.

(c) Faça uma previsão do *rendimento* (em km/l) para um automóvel que trafega em velocidade média de 160 km/h.

<u>Coeficiente de Determinação (R^2)</u>

O coeficiente de determinação (R^2) mede a proporção da variação em Y que é explicada pela equação de regressão estimada. Quanto maior for o valor de R^2, maior será a proporção da variação em Y explicada pela equação estimada. É usada para verificar a adequação de um modelo de regressão, e é dada por

$$R^2 = \frac{VE}{VT} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \qquad 0\% \leq R^2 \leq 100\%,$$

onde VE é a variação explicada pelo modelo de regressão estimado e VT é a variação total.

Para uma equação de regressão linear simples, $R^2 = (r_{x,y})^2 \times 100$.

Exemplo (continuação): Calcule e interprete o *coeficiente de determinação* R^2 para a equação de regressão estimada.

ANEXO I – EXERCÍCIOS DE REVISÃO

AI.1 Análise Descritiva de Dados

1.1) A tabela a seguir mostra a remuneração dos altos executivos CEOS – Chief Executive Officer –, a classificação por setor, as vendas anuais e os dados de avaliação da remuneração dos CEOS versus o retorno dos acionistas para 10 empresas. Uma avaliação igual a 1 da remuneração dos CEOS versus o retorno dos acionistas indica que a empresa está no grupo de empresas que têm a melhor relação. Uma avaliação igual a 2 indica que a empresa é similar às empresas que têm uma relação muito boa, mas não a melhor. Empresas com a pior relação têm uma avaliação igual a 5.

Empresa	Remuneração dos CEOS (US$ 1000s)	Setor	Vendas (US$ milhões)	Remuneração dos CEOS *vs.* Retorno dos Acionistas
Bankers Trust	8.925	Bancário	9.565	3
Coca-Cola	2.437	Bebidas	18.546	5
General Mills	1.410	Alimentação	5.567	1
LSI Logic	696	Eletrônico	1.239	2
Motorola	1.847	Eletrônico	27.973	4
Readers Digest	1.490	Gráfico	2.968	3
Sears	3.414	Varejo	38.236	4
Sprint	3.344	Telecomunicações	14.045	4
Walgreen	1.490	Varejo	12.140	2
Wells Fargo	2.861	Bancário	8.723	3

Fonte: *Business Week*, 21 de abril de 1997.

(a) Qual é a população em estudo?
(b) Qual foi o tamanho da amostra utilizado?
(c) Quantas variáveis existem nesse conjunto de dados. Classifique-as.
(d) Qual a porcentagem das empresas que pertencem ao setor bancário?
(e) Que porcentagem das empresas recebeu um valor 3 na avaliação da remuneração dos CEOS *versus* o retorno dos acionistas?
(f) Calcule a média aritmética, a mediana e a moda da remuneração dos altos executivos das empresas.
(g) Calcule a amplitude, a variância, o desvio-padrão e o coeficiente de variação da remuneração dos altos executivos das empresas.

1.2) Uma cidade turística tem 32 hotéis 3 estrelas. Pretende-se conhecer o custo médio de diária para apartamento de casal. Os valores populacionais consistem nos seguintes preços diários (em reais).

Hotel	Preço	Hotel	Preço	Hotel	Preço	Hotel	Preço
1	25	9	30	17	25	25	25
2	20	10	38	18	23	26	23
3	35	11	24	19	20	27	20
4	21	12	20	20	24	28	24
5	22	13	20	21	28	29	28
6	22	14	25	22	24	30	24
7	24	15	20	23	24	31	24
8	25	16	19	24	22	32	22

(a) Qual é a variável em estudo? Classifique-a.
(b) Apresente a distribuição dos dados em um ramo-e-folhas. Use incremento 2.
(c) Para os 32 hotéis, construa uma tabela de frequências adotando classes de mesma amplitude.
(d) Com base na tabela do item (b), construa um histograma dos dados. Utilize as frequências relativas.
(e) Considerando os 32 valores, calcule: Q_1, Q_2, Q_3 e AIQ.

(f) Apresente a distribuição dos dados em um boxplot.
(g) Considerando os 32 valores, calcule: a média, a mediana, a moda, a amplitude, a variância, o desvio-padrão e o coeficiente de variação dos preços diários.
(h) Considere que, para cidades com a mesma característica dessa, a variação dos preços das diárias gira em torno de 40%. Com base neste valor, é possível afirmar que os preços desses 32 hotéis são homogêneos. Justifique.

1.3) Considere os seguintes dados:

14	20	24	22	18	16	22	12
19	24	18	23	16	19	22	25
24	20	17	25	15	21	16	19
19	21	23	25	24	23	16	24
16	22	26	19	21	20	16	20

(a) Organize os dados em uma tabela de frequências, adotando classes iguais. Essa tabela deve conter as classes, as frequências absoluta e relativa, e as frequências acumuladas.
(b) Construa o histograma para os dados da tabela do item (b). Use as frequências relativas.
(c) Calcule as medidas de posição e de dispersão para o conjunto de dados. Calcule também os quartis.

1.4) O Serviço de Recursos Humanos da Roth Young relatou que os salários anuais para os gerentes assistentes de lojas de departamentos variam de R$ 28.000 a R$ 57.000. Assuma que os seguintes dados são uma amostra dos salários anuais de 40 gerentes assistentes de lojas de departamento (os dados estão em mil reais).

31	35	37	39	40	40	41	41	42	42
43	44	44	44	45	45	45	45	45	46
46	47	47	48	48	49	50	50	50	50
51	51	52	52	52	53	54	55	56	57

(a) Qual é a variável em estudo? Classifique-a como categórica nominal ou ordinal, ou quantitativa discreta ou contínua.
(b) Apresente a distribuição dos dados em um ramo-e-folhas. Use incremento 5.
(c) Organize os dados em uma tabela de frequências, adotando classes de amplitudes iguais.
(d) Construa o histograma associado à tabela do item (b). Use as frequências absolutas.

1.5) De acordo com uma pesquisa realizada pelo IBGE sobre o perfil do transporte rodoviário de cargas no Brasil, classificou-se 34.586 empresas de acordo com a quantidade de funcionários empregados no setor de transportes. Os resultados são apresentados na tabela a seguir:

N° de funcionários	n	%	n acumulada	% acumulada
0 a 5	23908			
6 a 19	8192			
20 a 49	1693			
50 a 99	447			
Mais de 99	346			
Total	34586			

(a) Complete a tabela.
(b) Complete:
i. O número de empresas que possuem de 6 a 19 funcionários empregados no setor de transportes é igual a _________ o que equivale a _________ %.

ii. O percentual de empresas que possuem até 49 funcionários é igual a _________ %.
iii. _________ empresas possuem pelo menos 20 funcionários empregados no setor de transportes, o que equivale a _________ %.

1.6) Os dados abaixo representam 40 leituras de temperatura (°C) de um pasteurizador de leite.

74,0	74,4	73,4	73,2	74,1	74,3	72,9	74,5	74,7	77,5	74,4	76,8	73,4	75,6
74,7	74,8	74,7	76,5	75,0	74,9	74,6	77,1	76,0	74,7	75,9	73,6	74,2	73,5
74,7	75,1	75,1	74,8	76,0	77,0	75,0	74,6	75,8	73,3	77,3	74,3		

(a) Construa uma tabela de frequências.
(b) Apresente a distribuição em um histograma.
(c) Calcule a mediana e os quartis.
(d) Apresente a distribuição dos dados em um boxplot.

1.7) Bernardin (UFSC, 1994) realizou um experimento que tinha o objetivo de melhorar a qualidade do processo de formulação de massa cerâmica para pavimento. Os corpos de prova eram "biscoitos" que saíam do processo de queima e a qualidade era avaliada por três variáveis:
- X_1 = *retração linear (%)*
- X_2 = *resistência mecânica*
- X_3 = *absorção de água (%)*
O experimento foi realizado sob 8 condições diferentes. Foram feitos 5 ensaios em cada uma das 8 condições experimentais. Os resultados são apresentados a seguir.

C^1	X_1	X_2	X_3	C^1	X_1	X_2	X_3	C^1	X_1	X_2	X_3	C^1	X_1	X_2	X_3
1	8,9	41,1	5,5	3	9,4	50,0	0,8	5	13,4	60,6	0,5	7	12,9	41,1	0,2
1	9,2	39,0	4,8	3	9,9	48,3	0,6	5	13,4	60,0	0,5	7	12,4	39,0	0,4
1	8,0	36,9	6,2	3	9,6	50,1	0,6	5	13,6	68,4	0,2	7	12,6	36,9	0,5
1	8,7	39,2	5,7	3	9,2	49,9	0,7	5	13,4	60,8	0,7	7	12,6	39,2	0,4
1	8,7	35,9	5,5	3	9,4	56,2	0,5	5	12,4	51,4	1,0	7	12,9	35,9	0,3
2	12,6	52,7	0,9	4	6,6	31,2	9,0	6	9,6	41,2	3,9	8	8,2	40,8	4,4
2	13,6	53,5	0,4	4	6,4	25,3	10,2	6	10,6	53,0	4,5	8	9,2	43,8	3,9
2	11,6	47,0	1,3	4	5,9	22,8	10,5	6	8,9	37,0	3,3	8	9,2	48,6	4,0
2	10,1	31,1	1,8	4	5,9	27,5	10,6	6	7,5	30,1	3,0	8	8,5	46,9	4,3
2	12,1	50,9	1,1	4	6,8	31,9	9,3	6	8,9	41,6	3,5	8	8,7	46,2	4,1

[1]C= condição experimental.

(a) Como as variáveis X_1, X_2 e X_3 podem ser classificadas?
(b) Apresente a distribuição de frequências de X_1, X_2 e X_3 através de histogramas. Comente as formas das distribuições.
(c) Apresente a distribuição dos valores de X_1, X_2 e X_3 através de boxplots. Compare os resultados obtidos.
(d) Calcule a média e o desvio-padrão de X_3 para as condições experimentais 1, 4 e 8. Quais as informações que podem ser extraídas com estas medidas?
(e) Calcule a mediana e os quartis de X_1.
(f) Compare as variáveis X_1, X_2 e X_3 com relação à variabilidade e à homogeneidade.

1.8) Os dados abaixo apresentam a distância (em km) entre a residência e o local de trabalho dos funcionários da empresa AAA.

1,8	2,5	0,4	1,9	4,4	2,2	3,5	1,7	0,8	3,2	15,1	2,1	1,4	0,5	0,9	1,7	0,9	2,1
1,1	1,7	1,2	2,3	1,9	0,8	1,5	0,5	1,4	1,4	1,8	2,0	1,1	1,0	0,8	0,2	1,4	3,7

Na empresa BBB, a distância (em km) até a residência de seus 300 funcionários apresenta as seguintes medidas descritivas:

1° Quartil = 1,6 Mediana = 2,8 3° Quartil = 4,2 Mínimo = 0,8 Máximo = 8,8

(a) Compare as empresas AAA e BBB em termos da distância entre a residência e o local de trabalho dos funcionários.
(b) Construa um diagrama de pontos para representar os dados dos funcionários da empresa AAA. Comente.

1.9) O relatório oficial sobre seguros e acidentes de carros dos Estados Unidos avalia modelos de carros com base no número de reclamações de seguro preenchidas após os acidentes, e utiliza um índice para classificá-los de acordo com a segurança. Os índices avaliados próximos de 100 são considerados médios. Avaliações menores são melhores, indicando um modelo de carro mais seguro. A seguir são mostradas avaliações para 20 carros de tamanho médio e 20 carros de tamanho pequeno:

Carros médios					Carros pequenos				
51	68	81	91	103	73	100	108	118	124
58	75	81	91	119	80	102	108	119	127
60	76	82	93	127	96	103	109	120	133
68	81	82	100	128	100	103	113	122	140

(a) Apresente os resultados obtidos em um diagrama de pontos onde seja possível diferenciar os carros médios dos carros pequenos.

(b) Complete a tabela abaixo.

Carros	n	Média	Mediana	Moda	Variância	Desvio padrão	Amplitude
Médios						21,494	
Pequenos		109,900					

(c) Qual dos dois grupos mostrou-se o mais homogêneo? Justifique com os cálculos.

1.10) Um processo químico de purificação é executado, numa determinada indústria, a quatro temperaturas diferentes. Os resultados obtidos para o teor final da substância de interesse (em %), para cada um dos 48 processos analisados (N = 48), são apresentados abaixo.

Temperatura (ºC)	**Teor final (em %)**											
25 ºC	60	62	62	60	61	63	63	62	62	63	63	63
50 ºC	64	63	63	64	65	64	65	65	65	65	66	66
75 ºC	68	68	68	68	69	68	69	68	70	70	70	69
100 ºC	72	73	72	72	73	73	74	74	75	74	75	74

(a) Qual é a variável em estudo? Classifique-a.
(b) Complete a tabela a seguir:

Temperatura	n	Média	Mediana	Moda	Amplitude	Variância	Desvio-padrão	CV
25°C								
50°C								
75°C								
100°C								

(c) Compare os processos a 100°C e a 50°C quanto à homogeneidade.
(d) Utilize boxplots para comparar a distribuição dos valores de teor final, para as diferentes temperaturas.

AI.2 Probabilidade

2.1) Apresente os espaços amostrais dos seguintes experimentos aleatórios:
(a) Lançamento de uma moeda honesta e observação da face voltada para cima;
(b) Observação da qualidade de peças produzidas, registrando o número de peças defeituosas;
(c) Contagem do número de clientes de um fila única de banco, que chegam durante uma hora;
(d) Medição da velocidade do vento, em km/h, na pista de um aeroporto;
(e) Medição da temperatura, em graus Celsius, numa estação meteorológica da cidade de Florianópolis.

2.2) Considere que você vai cronometrar o tempo, em segundos, para carregar uma página da web.
(a) Represente, em forma de conjuntos, os seguintes eventos:
i. *A = mais do que 5 e, no máximo, 10 segundos;*
ii. *B = mais do que 10 segundos;*
iii. *C = mais do que 8 segundos;*
iv. $D = A \cup B$;
v. $E = A \cap B$;
vi. $F = A \cap C$;
vii. $G = \bar{A}$
(b) Represente geometricamente (como intervalos na reta dos números reais) os conjuntos do item anterior.

2.3) Num estudo das necessidades futuras de um aeroporto, C é o evento de que existirão fundos suficientes para a expansão planejada e E é o evento de que a expansão planejada incluirá um estacionamento suficientemente amplo. Expresse em palavras o que significam as probabilidades, $P(\bar{C})$, $P(\bar{E})$, $P(\bar{C} \cap E)$ e $P(C \cap \bar{E})$.

2.4) Quais dos valores abaixo não podem ser probabilidades? Justifique cada um deles.
(a) $-0{,}2$ (b) $3/2$ (c) 0,000001 (d) $\sqrt{2}$ (e) 1

2.5) Quanto é P(A), se A é o evento "fevereiro tem 30 dias este ano"?

2.6) Retira-se, ao acaso, uma carta de um baralho de 52 cartas. Calcule a probabilidade de:
(a) A carta não ser de ouros;
(b) Ser uma carta de ouros ou uma figura.

2.7) De um conjunto de cinco empresas, deseja-se selecionar, aleatoriamente, uma empresa, mas com probabilidade proporcional ao número de funcionários. O número de funcionários da Empresa A é 20, da Empresa B é 15, da Empresa C é 7, da Empresa D é 5 e da Empresa E é 3.
(a) Qual é a probabilidade de cada uma das empresas ser selecionada?
(b) Qual é a probabilidade de a Empresa A não ser selecionada?

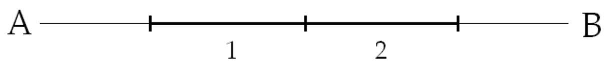
2.8) Considere as seguintes probabilidades para A e B:

$$P(A) = \frac{1}{2} \qquad P(B) = \frac{1}{4} \qquad P(A|B) = \frac{1}{3}$$

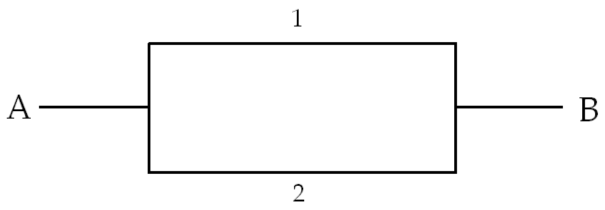
(a) Utilizando as leis de probabilidade, calcular: $P(\bar{A})$, $P(A \cap B)$ e $P(A \cup B)$.
(b) Usar o diagrama de Venn para calcular $P(A \cap \bar{B})$ e $P(\bar{A} \cup \bar{B})$.

2.9) Um sistema tem dois componentes que operam independentemente. Suponhamos que as probabilidades de falha dos componentes 1 e 2 sejam 0,1 e 0,2, respectivamente. Determinar a probabilidade de o sistema funcionar nos dois casos seguintes:

(a) Os componentes são ligados em série (isto é, ambos devem funcionar para que o sistema funcione).



(b) Os componentes são ligados em paralelo (basta um funcionar para que o sistema funcione).



2.10) A e B são duas estações meteorológicas em certo estado. Denotemos por A e B, respectivamente, a ocorrência de chuva em cada uma delas em um dia do mês de novembro. A experiência indica que $P(A) = P(B) = 0{,}45$ e $P(A \cap B) = 0{,}30$. Determine:

(a) $P(A \text{ ou } B)$
(b) $P(A|B)$
(c) $P(B|A)$

2.11) A probabilidade de um ônibus da linha Rio-São Paulo partir no horário é de 0,80 e a probabilidade de o ônibus partir no horário e chegar também no horário é de 0,72.

(a) Qual é a probabilidade de um ônibus que partiu no horário chegar também no horário?
(b) Se há uma probabilidade de 0,75 de um ônibus chegar no horário, qual é a probabilidade de um ônibus que chegou no horário ter partido também no horário?

2.12) O daltonismo é hereditário. Devido ao fato do gen responsável ser ligado ao sexo, o daltonismo ocorre mais frequentemente em pessoas do sexo masculino. Numa amostra de 10.000 pessoas, foi observada a incidência de daltonismo da cor vermelha-verde. Os resultados foram:

Daltonismo	Sexo		Total
	Masculino	Feminino	
Presente	423	65	488
Ausente	4848	4664	9512
Total	5271	4729	10000

Uma pessoa é escolhida ao acaso desta população.

(a) Calcule a probabilidade de ela ser:
   i. Daltônica.
   ii. Daltônica e do sexo feminino.
   iii. Daltônica ou do sexo feminino.
   iv. Daltônica dado que é do sexo feminino.
(b) Os eventos ser daltônico e ser do sexo feminino são independentes? Justifique com os cálculos. O que isto significa na prática?

2.13) Para estudar a forma de transporte preferida para a exportação de seus produtos, 150 indústrias foram pesquisadas, e os resultados encontrados são apresentados a seguir:

Produto exportado	Transporte preferido			Total
	Fluvial	Aéreo	Ferroviário	
Produtos metálicos	12	30	10	52
Máquinas	8	2	4	14
Equipamentos elétricos	11	14	6	31
Equipamentos de transporte	10	3	4	17
Instrumentos de precisão	8	27	1	36
Total	49	76	25	150

Selecionando-se aleatoriamente uma destas indústrias, responda:

(a) Qual a probabilidade de que a empresa exporte produtos metálicos?
(b) Qual a probabilidade de que a empresa prefira o transporte aéreo?
(c) Qual a probabilidade de que a empresa exporte máquinas e prefira o transporte fluvial?
(d) Qual a probabilidade de que a empresa prefira o transporte ferroviário ou exporte instrumentos de precisão?
(e) Sabendo-se que a empresa exporta equipamentos de transporte, qual a probabilidade de que ela prefira o transporte fluvial?
(f) O fato de exportar equipamentos elétricos e preferir o transporte aéreo são eventos independentes? Justifique com os cálculos.

2.14) Os dados de 200 peças usinadas estão resumidos a seguir:

Condição da extremidade	Profundidade do orifício	
	Acima do valor desejado	Abaixo do valor desejado
Grosseira	15	10
Moderada	25	20
Lisa	60	70

(a) Qual é a probabilidade de que uma peça selecionada tenha uma extremidade em condição moderada e uma profundidade do orifício abaixo do valor alvo?
(b) Qual é a probabilidade de que uma peça selecionada tenha uma extremidade em condição moderada ou uma profundidade do orifício abaixo do valor alvo?
(c) Construa um diagrama de Venn, com a representação dos eventos no espaço amostral.

2.15) A Telektronic fabrica computadores, televisores, aparelhos de CD e outros produtos eletrônicos. Quando os itens expedidos são danificados, as causas do dano são categorizadas como água (A), esmagamento (E), perfuração (P) e embalagem danificada (D). A seguir apresentam-se a distribuição dos 200 últimos produtos expedidos pela empresa que voltaram devido a reclamações de danos na expedição.

Produto	Causas do dano			
	Água	Esmagamento	Perfuração	Emb. danificada
Computadores	10	9	18	36
Televisores	12	7	23	40
Aparelhos de CD	3	3	0	14
Outros	5	3	7	10

Escolhido um produto ao acaso:

(a) Qual a probabilidade de que seja um aparelho de CD?
(b) Qual a probabilidade de que seja um computador danificado por perfuração?
(c) Qual a probabilidade que seja um televisor ou um produto danificado por água?
(d) Selecionando-se um produto com embalagem danificada, qual a probabilidade de que seja um aparelho de CD?

2.16) Uma desculpa clássica para a ausência em uma prova é dada por quatro estudantes, que afirmam que o pneu do carro onde os quatro estavam furou. Para confirmar essas alegações, o professor coloca cada um em uma sala e pede que eles escrevam em um papel qual pneu furou. Se nenhum pneu furou e eles escolhem aleatoriamente um pneu que supostamente teria furado, qual é a probabilidade de que eles escolham o mesmo pneu?

2.17) Um programa computacional para detectar fraudes em cartões telefônicos dos consumidores rastreia, todo dia, o número de áreas metropolitanas de onde as chamadas se originam. Sabe-se que 1% dos usuários legítimos fazem suas chamadas de duas ou mais áreas metropolitanas em um único dia. Entretanto, 30% dos usuários fraudulentos fazem suas chamadas de duas ou mais áreas metropolitanas em um único dia. A proporção de usuários fraudulentos é 0,01%. Se um usuário fizer as suas chamadas de duas ou mais áreas metropolitanas em um único dia, qual será a probabilidade de que o usuário seja fraudulento?

2.18) Consumidores são usados para avaliar projetos iniciais de produtos. No passado, 95% dos produtos altamente aprovados recebiam boas críticas, 60% dos produtos moderadamente aprovados recebiam boas críticas e 10% dos produtos ruins recebiam boas críticas. Além disso, 40% dos produtos tinham sido altamente aprovados, 35% tinham sido moderadamente aprovados e 25% tinham sido produtos ruins.

(a) Qual a probabilidade de que um produto atinja uma boa crítica?
(b) Se um projeto recebeu uma boa crítica, qual a probabilidade de que ele se torne um produto altamente aprovado?
(c) Se um projeto não recebeu uma boa crítica, qual a probabilidade de que ele se torne um produto moderadamente aprovado?
(d) Se um projeto recebeu uma crítica ruim, qual a probabilidade de que ele se torne um produto ruim?

2.19) Em uma indústria de enlatados, as linhas de produção I, II e III respondem por 50%, 30% e 20% da produção, respectivamente. As proporções de latas com defeito de produção nas linhas I, II e III são 0,4%, 0,6% e 1,2%. Qual a probabilidade de uma lata defeituosa (descoberta ao final da inspeção do produto acabado) provir da linha I?

2.20) Um estudante é submetido a um teste de múltipla escolha, em que cada questão apresenta cinco respostas, apenas UMA sendo correta. Se o estudante sabe a questão, escolhe a resposta certa. Se ele não sabe, escolhe ao acaso uma das respostas. Suponha que ele saiba 70% das questões.

(a) Qual a probabilidade de ele acertar determinada questão?
(b) Se ele responde corretamente a uma questão, qual a probabilidade de sabê-la?

2.21) Sabe-se que um soro da verdade, quando ministrado a um suspeito, é 90% eficaz quando a pessoa é culpada e 99% eficaz quando é inocente. Em outras palavras, 10% dos culpados são julgados inocentes e 1% dos inocentes são julgados culpados. Se o suspeito foi retirado de um grupo em que 95% jamais cometeram qualquer crime, e o soro indica culpado, qual a probabilidade de o suspeito ser inocente?

2.22) Num certo estado onde os automóveis devem ser testados quanto à emissão de gases poluentes, 25% de todos os automóveis emitem quantidades excessivas de gases poluentes. Ao serem testados, 99% de todos os automóveis que emitem quantidades excessivas de gases poluentes são reprovados, mas 17% dos que não emitem quantidades excessivas de gases poluentes também são reprovados. Qual é a probabilidade de um automóvel que é reprovado no teste efetivamente emitir uma quantidade excessiva de gases poluentes?

2.23) Um jovem amigo meu foi diagnosticado como tendo um tipo de câncer extremamente raro em pessoas jovens. Naturalmente, ele ficou muito abalado. Eu disse a ele que houve um erro, com o seguinte raciocínio. Nenhum teste médico é perfeito: existe sempre incidência de falso positivo[1] e falso negativo[2]. Denotemos por C o evento que ele tem câncer e por + o evento que um indivíduo tenha resposta positiva ao teste. Assuma que a proporção de pessoas com esse tipo de câncer nesta idade é de 1 em 1 milhão, e o teste para detectar este câncer é extremamente bom, dando somente 1% de falsos positivos e 1% de falsos negativos. Encontre a probabilidade de que ele tenha mesmo câncer, dado que seu teste deu resultado positivo para câncer.

AI.3 Distribuições Discretas de Probabilidade

3.1) Considere que a probabilidade de ocorrer k defeitos ortográficos em uma página de jornal é dada por

$$p(k) = \frac{1}{e.k!} \quad (e \approx 2{,}7183)$$

Tomando-se uma página qualquer, calcule a probabilidade de:
(a) Não ocorrer erro;
(b) Ocorrer mais do que dois erros.

3.2) (Martins) Uma variável aleatória tem função distribuição de probabilidade dada pela seguinte fórmula:

$$P(X = x) = \frac{k}{x}, \qquad \text{para } x = 1, 3, 5, 7$$

(a) Determine o valor de k.
(b) Calcule $P(2 \leq X \leq 6)$.
(c) Quanto vale $F(5)$?

3.3) (Martins) Um vendedor calcula que cada contato resulta em venda com probabilidade de 20%. Certo dia, ele contata dois possíveis clientes. Construa a tabela da distribuição de probabilidade para a variável Y = número de clientes que assinam um contrato de venda.

3.4) (Martins) Uma variável aleatória discreta pode assumir cinco valores, conforme a distribuição de probabilidade:

x_i	1	2	3	5	8
$p(x_i)$	0,20	0,25	?	0,30	0,10

(a) Encontrar o valor de $p(3)$.
(b) Qual é o valor da função acumulada para $x = 5$?
(c) Encontrar a média da distribuição.
(d) Calcular a variância e o desvio-padrão.

[1] Falsos positivos: são os testes com resultados positivos em pacientes sabidamente sadios.
[2] Falsos negativos: são os testes com resultados negativos em pacientes sabidamente doentes.

3.5) (Martins) A distribuição de probabilidade de uma variável aleatória discreta X é dada pela fórmula:

$$p(x) = (0{,}8)(0{,}2)^{x-1}, \qquad x = 1, 2, 3, \ldots$$

(a) Calcular $p(x)$ para $x = 1, x = 2, x = 3, x = 4$ e $x = 5$.
(b) Some as probabilidades obtidas no item (a). O que você diria a respeito das probabilidades para valores maiores que 5?

3.6) (Martins) O número de chamadas telefônicas recebidas por uma central e suas respectivas probabilidades para um intervalo de um minuto são:

Número de chamadas (x)	0	1	2	3	4	5
Probabilidades	0,55	0,25	0,10	0,04	0,04	0,02

(a) Calcular $F(2)$.
(b) Determinar $P(1 \leq X \leq 4)$ e $P(X > 1)$.
(c) Qual é o número esperado de chamadas em um minuto?
(d) Avalie o coeficiente de variação para essa distribuição.

3.7) (Martins) Em uma sala, temos cinco rapazes e quatro moças. São escolhidas aleatoriamente três pessoas. Seja X o número de rapazes.
(a) Determinar a distribuição de probabilidade da variável X.
(b) Calcular as seguintes probabilidades:
(b1) $P(X \leq 2)$;
(b2) $P(X \leq 0)$;
(b3) $P(1 < X \leq 3)$;
(b4) $P(2 < X < 3)$;
(b5) $P(X > 2)$;
(b6) $P(X > -1)$;
(b7) $P(X < 5)$;
(c) Determinar:
(c1) $F(2{,}5)$;
(c2) $F(3)$;
(c3) $F(0{,}5)$;
(c4) $F(3{,}5)$;
(c5) $F(2)$;
(c6) $F(1)$;
(c7) $F(6)$;
(c8) $F(-0{,}5)$.

3.8) (Montgomery) O espaço amostral de um experimento aleatório é {*a, b, c, d, e, f*} e cada resultado é igualmente provável. Uma variável aleatória X é definida como segue:

resultado	a	b	c	d	e	f
x	0	0	1,5	1,5	2	3

(a) Determine a função distribuição de probabilidade de X.
(b) Determine a função de distribuição acumulada de X.
(c) Determine a esperança e a variância de X.

3.9) (Montgomery) O setor de comercialização estima que um novo instrumento para análise de amostras de solo terá grande sucesso, moderado sucesso ou não terá sucesso, com probabilidades de 0,3; 0,6 e 0,1, respectivamente. A receita anual associada com um produto de grande sucesso, moderado sucesso ou nenhum sucesso é de R$ 10 milhões, R$ 5 milhões e R$ 1 milhão, respectivamente. Seja a variável aleatória X a renda anual do produto.

(a) Determine a função distribuição de probabilidade de X.
(b) Determine a função de distribuição acumulada de X.
(c) Determine a esperança e a variância X.

3.10) (Montgomery) Em um processo de fabricação de semicondutores, três pastilhas de um lote são testadas. Cada pastilha é classificada como *passa* ou *falha*. Suponha que a probabilidade de uma pastinha passar no teste seja de 0,8 e que as pastilhas sejam independentes.
(a) Qual a probabilidade de que todas as três pastilhas passem no teste?
(b) Determine a função de probabilidade do número de pastilhas de um lote que passam no teste.
(c) Determine a função de distribuição acumulada do número de pastilhas de um lote que passam no teste.

3.11) (Montgomery) Um arranjo consiste em três componentes mecânicos. Suponha que as probabilidades do primeiro, do segundo e do terceiro componentes encontrarem as especificações sejam iguais a 0,95; 0,98 e 0,99. Considere que os componentes sejam independentes.
(a) Qual a probabilidade de que todos os componentes em um arranjo encontrem as especificações?
(b) Determine a função de probabilidade do número de componentes no arranjo que encontram as especificações.

3.12) (Montgomery) Determine a constante c, de modo que a seguinte função seja uma função de probabilidade: $f(x) = cx$, para $x = 1, 2, 3, 4$.

3.13) O presidente da Martin Corporation está considerando duas alternativas de investimento X e Y. Se cada uma das alternativas for levada adiante, há quatro possibilidades de resultado. O lucro líquido e sua respectiva probabilidade de ocorrência são mostrados abaixo:

INVESTIMENTO X

Resultado	Lucro (milhões)	Probabilidade
1	\$ 20	0,2
2	\$ 08	0,3
3	\$ 10	0,4
4	\$ 03	0,1

INVESTIMENTO Y

Resultado	Lucro (milhões)	Probabilidade
1	\$ 12	0,1
2	\$ 09	0,3
3	\$ 16	0,1
4	\$ 11	0,5

(a) Qual é o valor esperado do lucro para os investimentos X e Y? Qual das oportunidades é a mais interessante?
(b) Qual a variância do lucro para os investimentos X e Y? Qual das oportunidades é a mais arriscada?

3.14) Na venda de um produto têm-se duas opções:
i. Cobrar R\$ 1.000,00 por peça sem inspeção;
ii. Classificar o lote em de 1ª ou 2ª qualidade, mediante a seguinte inspeção: retiramos cinco peças do lote e se não encontrarmos mais que uma defeituosa o lote será classificado de 1ª qualidade; sendo de 2ª qualidade o lote que não satisfizer tal condição. Sabe-se que o preço de venda é de R\$ 1.200,00 por peça do lote de 1ª e R\$ 800,00 por peça do lote de 2ª.

Se cerca de 10% das peças produzidas são defeituosas e que serão vendidos 50 lotes contendo 10 peças cada, analisar qual das duas opções é mais vantajosa para o vendedor.

3.15) A Indústria Controlada S.A. tem dois eventuais compradores de seu produto, que pagam preços em função da qualidade:

i. O comprador A paga R$ 150,00 por peça, se em uma amostra de 100 peças não encontrar nenhuma defeituosa e R$ 50,00 por peça, caso contrário;
ii. O comprador B paga R$ 200,00 por peça, desde que encontre no máximo uma peça defeituosa em 120 peças, e R$ 30,00 por peça, caso contrário.

Para qual dos dois compradores o empresário deveria vender se ele sabe que na produção 3% das peças são defeituosas?

3.16) (Martins) De acordo com uma pesquisa do *Data Journal*, 70% das pessoas que trabalham em escritórios utilizam PCs da IBM. Se dois indivíduos que trabalham em escritórios são selecionados, encontrar a distribuição de probabilidade da variável X = número de usuários dos PCs da IBM. Calcular a média e o desvio-padrão dessa variável.

3.17) (Martins) Uma moeda é jogada 10 vezes. Calcule as seguintes probabilidades:

(a) De ocorrer seis caras;
(b) De dar pelo menos duas caras;
(c) De não dar nenhuma coroa;
(d) De dar pelo menos uma coroa;
(e) De não dar cinco caras e cinco coroas.

3.18) (Martins) Um time X tem $\frac{2}{3}$ de probabilidade de vitória sempre que joga. Se X jogar cinco partidas, calcule a probabilidade de:

(a) X vencer exatamente três partidas;
(b) X vencer ao menos uma partida;
(c) X vencer mais da metade das partidas.

3.19) (Martins) A probabilidade de um atirador acertar o alvo é $\frac{1}{3}$. Se ele atirar seis vezes, qual a probabilidade de:

(a) Acertar exatamente dois tiros?
(b) Não acertar nenhum tiro?

3.20) (Martins) Em um teste do tipo certo-errado, com 100 perguntas, qual a probabilidade de um aluno, respondendo às questões ao acaso, acertar 70% das perguntas?

3.21) (Martins) Uma variável aleatória Y com distribuição Binomial tem a função acumulada dada por:

$$F(0) = \frac{1}{243} \qquad F(1) = \frac{11}{243} \qquad F(2) = \frac{51}{243} \qquad F(3) = \frac{131}{243} \qquad F(4) = \frac{211}{243} \qquad F(5) = 1$$

Determinar:

(a) n;
(b) p e $(1-p)$;
(c) Média de Y;
(d) Variância de Y;
(e) $P(Y \geq 1)$;
(f) $P(2 \leq Y \leq 4)$.

3.22) (Martins) Se 5% das lâmpadas de certa marca são defeituosas, encontre a probabilidade de que, numa amostra de 100 lâmpadas escolhidas ao acaso, tenhamos:
(a) Nenhuma defeituosa;
(b) Três defeituosas;
(c) Mais do que uma boa.

3.23) (Montgomery) Determine a função de distribuição acumulada de uma variável aleatória Binomial com $n = 3$ e $p = \frac{1}{4}$.

3.24) (Montgomery) Seja X o número de *bits* recebidos com erros em um canal digital de comunicação, e considere que X seja uma variável aleatória Binomial com $p = 0{,}001$. Se 1.000 *bits* forem transmitidos, determine o seguinte (DICA: use a aproximação pela distribuição de Poisson para calcular as probabilidades):
(a) $P(X = 1)$;
(b) $P(X \geq 1)$;
(c) $P(X \leq 2)$;
(d) A média e a variância de X.

3.25) (Montgomery) Bateladas, que consistem em 50 molas provenientes de um processo de produção, são verificadas com relação à conformidade aos requerimentos dos consumidores. O número médio de molas não conformes em uma batelada é igual a 5. Suponha que o número de molas não conformes em uma batelada, denotada como X, seja uma variável aleatória Binomial.
(a) Quais são os valores de n e p?
(b) Qual é $P(X \leq 2)$?
(c) Qual é $P(X \geq 49)$?

3.26) (Montgomery) Porque nem todos os passageiros de aviões aparecem na hora do embarque, uma companhia aérea vende 125 bilhetes para um voo que suporta somente 120 passageiros. A probabilidade de que um passageiro não apareça é 0,10 e os passageiros se comportam independentemente.
(a) Qual a probabilidade de que cada passageiro que apareça possa embarcar?
(b) Qual é a probabilidade de que o voo decole com assentos vazios?

3.27) (Montgomery) Um teste de múltipla escolha contém 25 questões, cada uma com quatro respostas. Suponha que um estudante apenas tente adivinhar ("chutar") cada questão.
(a) Qual é a probabilidade de que o estudante responda mais de 20 questões corretamente?
(b) Qual é a probabilidade de que o estudante responda menos de 5 questões corretamente?

3.28) (Martins) O pessoal de inspeção de qualidade afirma que os rolos de fita isolante apresentam, em média, uma emenda a cada 50 metros. Admitindo que a distribuição do número de emendas é dada pela Poisson, calcule a probabilidade:
(a) De nenhuma emenda em um rolo de 125 metros;
(b) De ocorrerem no máximo duas emendas em um rolo de 125 metros;
(c) De ocorrer pelo menos uma emenda em um rolo de 100 metros.

3.29) (Martins) Admitindo que X tem distribuição de probabilidade de Poisson, encontrar as probabilidades:
(a) $P(X = 5)$ quando $\lambda = 3{,}0$;
(b) $P(X \leq 2)$ quando $\lambda = 5{,}5$;
(c) $P(X \geq 4)$ quando $\lambda = 7{,}5$;
(d) $P(X = 8)$ quando $\lambda = 4{,}0$.

3.30) (Martins) Suponha que X tenha distribuição Binomial com $n = 100$ e $p = 0{,}02$. Usar a aproximação de Poisson para encontrar:
(a) $P(X \geq 3)$;
(b) $P(X = 5)$;
(c) $P(X = 0)$;
(d) $P(X < 2)$.

3.31) (Martins) Uma fábrica de pneus verificou que, ao testar seus pneus nas pistas, havia em média um estouro de pneu a cada 5.000 km.
(a) Qual a probabilidade de que num teste de 3.000 km haja, no máximo, um pneu estourado?
(b) Qual a probabilidade de que um carro ande 8.000 km sem estourar nenhum pneu?

3.32) (Martins) Certo posto de bombeiros recebe em média três chamadas por dia. Calcular a probabilidade de:
(a) Receber quatro chamadas num dia;
(b) Receber três ou mais chamadas num dia.

3.33) (Martins) Na pintura de paredes, aparecem defeitos em média na proporção de um defeito por metro quadrado. Qual a probabilidade de aparecerem três defeitos numa parede de 2 x 2 m?

3.34) (Martins) Suponha que haja em média dois suicídios por ano numa população de 50.000 habitantes. Em cada cidade de 100.000 habitantes, encontre a probabilidade de que em dado ano tenha havido:
(a) Nenhum suicídio;
(b) Um suicídio;
(c) Dois suicídios;
(d) Dois ou mais suicídios.

3.35) (Martins) Suponha 400 erros de impressão distribuídos aleatoriamente em um livro de 500 páginas. Encontre a probabilidade de que dada página contenha:
(a) Nenhum erro;
(b) Exatamente dois erros.

3.36) (Martins) Certa loja recebe em média cinco clientes por hora. Qual a probabilidade de:
(a) Receber dois clientes em 24 minutos?
(b) Receber pelo menos três clientes em 18 minutos?

3.37) (Martins) A média de chamadas telefônicas em uma hora é três. Qual a probabilidade de:
(a) Receber três chamadas em 20 minutos?
(b) Receber no máximo duas chamadas em 30 minutos?

3.38) (Martins) Um contador Geiser marca em média 40 sinais por minuto, quando nas proximidades de certa substância radioativa. Determinar a probabilidade de que haja dois sinais em um período de seis segundos.

3.39) (Martins) Em uma estrada, passam em média 1,7 carro por minuto. Qual a probabilidade de passarem exatamente dois carros em dois minutos?

3.40) (Martins) Um distribuidor de gasolina tem capacidade de receber, nas condições atuais, no máximo três caminhões por dia. Se chegarem mais de três caminhões, o excesso deve ser enviado a outro distribuidor. Sabendo que, em média, chegam diariamente dois caminhões, qual a probabilidade de, em certo dia, ter que enviar caminhões para outro distribuidor?

3.41) (Martins) Uma fábrica produz tecidos com 2,2 defeitos, em média, por peça. Determinar a probabilidade de haver ao menos dois defeitos em duas peças?

3.42) (Montgomery) Um técnico de instalação de um sistema especializado de comunicação é enviado para uma cidade somente quando existirem três ou mais ordens de serviço. Suponha que as ordens de serviço sigam uma distribuição de Poisson com uma média de 0,25 por semana, para uma cidade com uma população de 100.000 habitantes, e suponha que sua cidade contenha 800.000 habitantes.

(a) Qual é a probabilidade de que um técnico seja requisitado depois de um período de uma semana?
(b) Se você for o primeiro da cidade a pedir uma ordem de serviço, qual será a probabilidade de que você tenha que esperar mais de duas semanas, a partir do tempo do pedido da ordem de serviço, até que o técnico seja despachado?

3.43) (Montgomery) Em uma seção de uma autoestrada, o número de buracos grandes o bastante para que requeiram reparo, é suposto seguir uma distribuição de Poisson com uma média de dois buracos por milha.

(a) Qual é a probabilidade de que não haja buracos que requeiram reparo em 5 milhas da autoestrada?
(b) Qual é a probabilidade de que no mínimo um buraco requeira reparo em 0,5 milha da autoestrada?
(c) Se o número de buracos estiver relacionado à carga do veículo na autoestrada, e algumas seções dessa autoestrada estiverem sujeitas a uma carga leve de veículos, como você se sente a respeito da suposição de distribuição de Poisson para o número de buracos que requerem reparo?

3.44) (Montgomery) O número de insucessos de um instrumento de teste para partículas de contaminação no produto é uma variável aleatória de Poisson com uma média de 0,02 insucesso por hora.

(a) Qual é a probabilidade de que o instrumento não falhe em um turno de 8 horas?
(b) Qual é a probabilidade de, no mínimo, um insucesso em um dia (24 horas)?

3.45) (Montgomery) O número de erros em um livro-texto segue uma distribuição de Poisson com média de 0,01 erro por página. Qual é a probabilidade de que haja três ou menos erros em 100 páginas?

3.46) Um jogo de loteria consiste em selecionar seis dezenas de um conjunto de 60 dezenas, com uma bola para cada dezena, sem reposição. Num volante (cartão aposta), o jogador pode escolher de 6 a 12 dezenas. Qual é a probabilidade de um jogador acertar a quina (5 dezenas) se marcar 10 dezenas no volante?

3.47) (Barros) O dono de uma festa encomendou a um buffet 100 empadinhas de frango e 50 de camarão. Um convidado guloso "sequestra" a bandeja do garçom, que contém 20 empadinhas. O convidado é, além de guloso, alérgico a camarão, e se comer mais de duas empadas de camarão, corre o risco de passar o resto da festa no hospital. Qual a probabilidade disso acontecer?

3.48) A Panini colocou, em seus pacotes de figurinhas do álbum do Campeonato Brasileiro deste ano, figurinhas premiadas que valem 30 pacotes extras de figurinhas. Dentro de cada pacote tem cinco figurinhas comuns, ou quatro comuns e uma premiada. Suponha que seu amigo esteja com 20 pacotinhos, e que em dois deles haja figurinhas premiadas. Ele resolve te dar 6 pacotinhos.

(a) Qual é a probabilidade de que nenhum dos seis tenha a figurinha premiada?
(b) Qual é a probabilidade de você ganhar os dois pacotinhos com figurinha premiada?

(c) Quantos pacotinhos ele deveria te dar para que você pudesse esperar 1 com figurinha premida?

3.49) (Morettin) Uma fábrica de motores para máquinas de lavar roupas separa de sua linha de produção diária de 350 peças uma amostra de 30 itens para inspeção. O número de peças defeituosas é de 14 por dia. Qual a probabilidade de que a amostra contenha pelo menos 3 motores defeituosos?

3.50) (Morettin) Numa cidade, é selecionada uma amostra de 60 adultos e a esses indivíduos é pedido para opinarem se são a favor ou contra determinado projeto Com resultado obtido, observou-se 40 a favor. Se na realidade as opiniões pró e contra são igualmente divididas, qual é a probabilidade de ter obtido tal resultado?

3.51) (Morettin) Um órgão governamental credencia a firma A para fazer vistorias em carros recuperados ou construídos particularmente e dar a aprovação ou não para que determinado carro possa ser lacrado no Detran. Resolve testar se a firma A está trabalhando de acordo com suas especificações. De um lote de 250 carros vistoriados e aprovados por A, escolhe 50 e faz novas vistorias. Se encontrar no mínimo 2 que não mereçam a aprovação, descredencia A. Sabendo-se que no lote de 250 há 8 carros que foram aprovados irregularmente, qual a probabilidade do descredenciamento?

AI.4 Distribuições Contínuas de Probabilidade

4.1) Suponha que o nível salarial em certo setor na economia (em u.m.) tem a seguinte f.d.p.:

$$f(x) = \begin{cases} \dfrac{k}{x^3}, & x > 8 \\ 0, & c.c. \end{cases}$$

(a) Determine o valor de k.
(b) Qual a probabilidade de um salário nesse setor ser superior a 10 u.m.?
(c) Qual o salário médio nesse setor da economia?

4.2) Uma variável aleatória contínua tem f.d.p. dada por:

$$f(x) = \begin{cases} \dfrac{x}{4}, & 1 \leq x \leq 3 \\ 0, & c.c. \end{cases}$$

(a) Faça um gráfico da função acima e verifique que ela satisfaz as condições para ser densidade.
(b) Calcule:
  (b1) $P(X > 2)$
  (b2) $P(X \geq 2)$
  (b3) $P(X = 2)$
  (b4) $P\left(0 \leq X \leq \frac{3}{2}\right)$
  (b5) $P\left(X > \frac{3}{2}\right)$
  (b6) $P\left(X \geq \frac{3}{2}\right)$

4.3) O tempo de corrosão, em anos, de uma certa peça metálica é uma variável com densidade:

$$f(x) = \begin{cases} ax, & 0 \leq x \leq 1; \\ a, & 1 < x \leq 2; \\ -ax + 3a, & 2 < x \leq 3; \\ 0, & c.c. \end{cases}$$

(a) Calcule a constante a.
(b) Uma peça é considerada como tendo boa resistência à corrosão se dura mais que 1,5 anos. Em um lote de 3 peças, qual a probabilidade de termos exatamente 1 delas com boa resistência?

4.4) Uma variável aleatória contínua X é definida pela seguinte função densidade:

$$f(x) = \begin{cases} \frac{3}{2}(x-1)^2, & 0 \leq x < 2 \\ 0, & c.c. \end{cases}$$

Determinar $E(X)$ e $Var(X)$.

4.5) Indique quais das seguintes funções reais podem ser funções de densidades de variáveis aleatórias contínuas:

(a) $f(x) = \begin{cases} 2x - 3, & 1 \leq x \leq 2 \\ 0, & c.c. \end{cases}$

(b) $f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

(c) $f(x) = \begin{cases} x^2, & 2 < x < 3 \\ 0, & c.c. \end{cases}$

4.6) A função f(x) é uma função real definida por:

$$\mathrm{f(x)} = \begin{cases} \frac{4\mathrm{k}^2 - 5}{2\mathrm{k}}(\mathrm{x} - 1), & -\mathrm{k} \leq \mathrm{x} \leq \mathrm{k} \\ 0, & \mathrm{c.\,c.} \end{cases}$$

(a) Para que valor de k a $f(x)$ pode representar uma função densidade de probabilidade da variável aleatória X?
(b) Para o valor de k encontrado, determine as probabilidades: $P(X = 1)$, $P(X \leq 0)$, $P(X < 0)$, $P(0 < X \leq 1)$, $P(0 < X \leq 3)$, $P(X \geq 1)$, $P(3 \leq X \leq 4)$.
(c) Calcule a média e a variância da v.a. X.
(d) Determine a função de distribuição acumulada da v.a. X.

4.7) Depois de terem sido pesadas várias embalagens de 1 kg de café da marca "Apetitoso", chegou-se à conclusão de que, embora a embalagem indique 1 kg, o verdadeiro peso é uma variável aleatória uniformemente distribuída no intervalo [0,85 kg ; 1,05 kg], isto é, a f.d.p. tem a seguinte forma:

$$f(x) = \begin{cases} k, & 0{,}85 \leq x \leq 1{,}05 \\ 0, & c.c. \end{cases}$$

(a) Calcule k e represente graficamente $f(x)$.
(b) Determine a média e a variância do peso das embalagens.
(c) Qual a probabilidade de uma embalagem de café da marca "Apetitoso" pesar menos de 1 kg?
(d) Em 200 embalagens, quantas você esperaria que tivessem o peso superior ao indicado no rótulo?

4.8) (Anderson) A maioria das linguagens de computador contém uma função que pode ser usada para gerar números aleatórios. No Excel, a função ALEATÓRIO pode ser usada para gerar números aleatórios entre 0 e 1. Se admitirmos que x denota um número aleatório gerado pela função ALEATÓRIO, então x é uma variável aleatória contínua com a seguinte função densidade de probabilidade:

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & c.c. \end{cases}$$

(a) Trace o gráfico da função densidade de probabilidade.
(b) Qual é a probabilidade de se gerar um número aleatório entre 0,25 e 0,75?
(c) Qual é a probabilidade de se gerar um número aleatório com valor menor ou igual a 0,30?
(d) Qual é a probabilidade de se gerar um número aleatório com valor maior que 0,60?

4.9) (Anderson) O rótulo de uma garrafa de detergente líquido indica que o conteúdo é de 12 onças por garrafa (1 onça = 29,574 ml). A operação de produção preenche a garrafa uniformemente, de acordo com a seguinte função densidade de probabilidade:

$$f(x) = \begin{cases} 8, & 11{,}975 \leq x \leq 12{,}100 \\ 0, & c.c. \end{cases}$$

(a) Qual é a probabilidade de uma garrafa ser preenchida com volume entre 12 e 12,05 onças?
(b) Qual é a probabilidade de uma garrafa ser preenchida com 12,02 onças ou mais?
(c) O controle da qualidade aceita uma margem de erro de 0,02 onças no preenchimento de uma garrafa em relação ao volume indicado no seu rótulo. Qual é a probabilidade de a garrafa desse detergente líquido deixar de cumprir o padrão estabelecido pelo controle de qualidade?

4.10) (Montgomery) Em uma grande rede corporativa de computadores, as conexões de usuários ao sistema podem ser modeladas como um processo de Poisson com uma média de 25 conexões por hora.
(a) Qual é a probabilidade de não haver conexões em um intervalo de 6 minutos?
(b) Qual é a probabilidade de que o tempo até a próxima conexão esteja entre 2 e 3 minutos?
(c) Determine o intervalo de tempo tal que a probabilidade de nenhuma conexão ocorrer no intervalo seja 0,90.

4.11) (Montgomery) O tempo entre as chamadas para uma loja de suprimentos de encanamentos é distribuído exponencialmente, com um tempo médio de 15 minutos entre as chamadas.
(a) Qual é a probabilidade de não haver chamadas dentro do intervalo de 30 minutos?
(b) Qual é a probabilidade de que no mínimo uma chamada chegue dentro do intervalo de 10 minutos?
(c) Qual é a probabilidade de que a primeira chamada chegue dentro de 5 e 10 minutos depois da loja aberta?
(d) Determine o comprimento de um intervalo de tempo, tal que exista uma probabilidade igual a 0,90 de haver no mínimo uma chamada no intervalo.

4.12) (Montgomery) O tempo entre as chegadas de táxis a uma interseção movimentada é distribuído exponencialmente, com uma média de 10 minutos.
(a) Qual é a probabilidade de você esperar mais de uma hora por um táxi?
(b) Suponha que você já estivesse esperando uma hora por um táxi, qual será a probabilidade de que um táxi chegue dentro dos próximos 10 minutos?
(c) Determine x tal que a probabilidade de você esperar mais de x minutos seja 0,10.
(d) Determine x tal que a probabilidade de você esperar menos de x minutos seja 0,90.
(e) Determine x tal que a probabilidade de você esperar menos de x minutos seja 0,50.

4.13) (Montgomery) O tempo de vida de um arranjo mecânico em um teste vibracional é distribuído exponencialmente, com uma média de 400 horas.
(a) Qual é a probabilidade de que um arranjo em teste falhe em menos de 100 horas?
(b) Qual é a probabilidade de que um arranjo opere por mais de 500 horas antes da falha?

(c) Se um arranjo estiver em teste por 400 horas sem apresentar falha, qual será a probabilidade de uma falha ocorrer nas próximas 100 horas?
(d) Se 10 arranjos estão sendo testados, qual é a probabilidade de que no mínimo um falhe em menos de 100 horas? Considere que os arranjos falhem independentemente.
(e) Se 10 arranjos estão sendo testados, qual é a probabilidade de que todos tenham falhado em 800 horas? Considere que os arranjos falhem independentemente.

4.14) Mensa é a sociedade internacional de indivíduos de alto QI. Para fazer parte da Mensa, uma pessoa precisa ter um QI de 132 ou mais. Se as contagens de QI são distribuídas normalmente com uma média de 100 e um desvio-padrão de 15, que porcentagem da população se qualifica para membro da Mensa?

4.15) Para uma v.a. X com distribuição aproximadamente Normal com média μ e desvio-padrão σ, calcule $P(\mu-\sigma \leq X \leq \mu+\sigma)$, $P(\mu-2\sigma \leq X \leq \mu+2\sigma)$ e $P(\mu-3\sigma \leq X \leq \mu+3\sigma)$.

4.16) Suponha que o diâmetro de uma peça siga uma distribuição Normal com média 2,04 mm e variância de 0,6084 mm².
(a) Determine a probabilidade de uma peça apresentar diâmetro:
  i. menor que 2,81 mm.
  ii. maior que 1,8 mm.
  iii. entre 1,01 e 2,50 mm.
(b) Se considerarmos 200 dessas peças, quantas podemos esperar que tenham o diâmetro entre 2,20 e 3,80 mm?
(c) Qual intervalo, simétrico em torno da média, que abrange 98% dos diâmetros das peças?

4.17) A vida média de certo aparelho é de oito anos, com desvio-padrão de 1,8 anos. O fabricante substitui os aparelhos que acusam defeito dentro do prazo de garantia. Se ele deseja substituir no máximo 5% dos aparelhos que apresentem defeito, qual deve ser o prazo de garantia?

4.18) Uma máquina de encher copos de refrigerantes o faz segundo uma distribuição Normal com média igual a 225 ml e desvio-padrão de 10 ml.
(a) Calcule a probabilidade de um copo apresentar volume líquido:
  i. inferior a 250 ml.
  ii. superior a 200 ml.
  iii. entre 210 e 240 ml.
(b) Se você está enchendo um copo com capacidade de 300 ml, qual a probabilidade do refrigerante transbordar durante o enchimento?

4.19) Um determinado tipo de lâmpada apresenta vida média de 1440 horas com desvio-padrão igual a 72 horas. Considerando que o tempo deste tipo de lâmpada siga uma distribuição aproximadamente Normal, qual a probabilidade de uma lâmpada durar:
(a) Mais de 1650 horas?
(b) Menos de 1300 horas?
(c) Entre 55 e 65 dias?
(d) Determine um intervalo, simétrico em torno da média, que contenha o tempo de vida de 95% destas lâmpadas.

4.20) A precipitação pluviométrica média em certa cidade, no mês de dezembro, é de 8,9 cm. Admitindo a distribuição normal com desvio-padrão de 2,5 cm, determinar a probabilidade de que, no mês de dezembro próximo, a precipitação seja:
(a) Inferior a 1,6 cm.
(b) Superior a 5 cm mas não superior a 7,5 cm.

(c) Superior a 12 cm.
(d) Encontre um intervalo, simétrico em torno da média, que contenha 99% dos valores de precipitação pluviométrica.

4.21) (Martins) O salário semanal dos operários industriais é distribuído normalmente em torno de uma média de R$ 180 com desvio-padrão de R$ 25. Pede-se:
(a) Encontrar a probabilidade de um operário ter salário semanal situado entre R$ 150 e R$ 178.
(b) Dentro de que desvio, de ambos os lados da média, cairão 96% dos salários?

4.22) (Martins) Em uma grande empresa, a avaliação do desempenho profissional nos funcionários acusou média 70 e desvio-padrão 10. Se desejamos atribuir aos 15% superiores o grau A, aos 20% seguintes o grau B, aos 30% médios o grau C, aos próximos 25% o grau D e aos últimos 10% o grau E, quais os intervalos de notas que serão abrangidos por essas classificações.

4.23) (Martins) O tempo necessário em uma oficina para o conserto da transmissão de um tipo de carro segue uma distribuição normal, com média 45 minutos e desvio-padrão de 8 minutos.
(a) O mecânico comunicou a um cliente que o carro estará pronto em até 50 minutos. Qual a probabilidade de que o mecânico esteja enganado?
(b) Qual deve ser a previsão de tempo de trabalho para que haja 90% de probabilidade de que o conserto da transmissão seja efetuado dentro do tempo previsto?

4.24) (Martins) Suponha que as notas de uma prova sejam normalmente distribuídas com média 73 e desvio-padrão 15. Quinze por cento dos alunos mais adiantados recebem a nota A e 12% dos mais atrasados recebem nota F. Encontrar o mínimo para receber A e o mínimo para não receber F.

4.25) (Martins) Constatou-se que o tempo médio para se fazer um teste-padrão de matemática é aproximadamente normal, com média de 80 minutos e desvio-padrão de 20 minutos.
(a) Que porcentagem de candidatos levará menos de 80 minutos?
(b) Que porcentagem não terminará o teste, se o tempo máximo concedido é de duas horas?
(c) Se 200 pessoas fazem o teste, quantas podemos esperar que o terminem na primeira hora?

4.26) (Martins) Uma pessoa tem 20 minutos para chegar ao escritório. Para tal, pode optar entre dois caminhos: A ou B. Sabendo que o tempo para percorrer o caminho A é $N(18;15)$ e que o tempo para percorrer o caminho B é $N(20;2)$, qual a melhor escolha do trajeto?

4.27) Um processo industrial produz peças com diâmetro de 2,00'' e desvio-padrão de 0,01''. As peças com diâmetro que se afaste da média por mais de 0,03'' são consideradas defeituosas. Considerando que o diâmetro das peças tenha uma distribuição aproximadamente Normal, responda:
(a) Qual a porcentagem de peças defeituosas?
(b) Qual a probabilidade de se encontrar duas peças defeituosas em sequência?
(c) Qual a probabilidade de se encontrar duas peças perfeitas em sequência?

AI.5 Intervalos de Confiança e Testes de Hipóteses

5.1) (Adaptado de Montgomery) A resistência à quebra de um fio usado na fabricação de material moldável necessita ser no mínimo 100 psi. Experiência passada indicou que o desvio-padrão da resistência à quebra foi 2 psi. Uma amostra aleatória de nove espécimes é testada e a resistência média à quebra é de 98 psi.

(a) A fibra deve ser julgada como aceitável com $\alpha = 0{,}05$?
(b) Encontre um intervalo bilateral de confiança com 95% para a resistência média verdadeira à quebra.

5.2) (Barbetta) Em uma amostra aleatória simples com 200 edifícios com cinco anos, em certa cidade, 55% apresentaram problemas estéticos relevantes após a entrega da obra. Construir um intervalo de confiança para a proporção de edifícios da cidade que apresentam problemas estéticos relevantes nos cinco primeiros anos. Use nível de confiança de 95%.

5.3) (Barbetta) Uma empresa fabricante de pastilhas para freios efetua um teste para controle de qualidade de seus produtos. Selecionou-se uma amostra de 600 pastilhas, das quais 18 apresentaram níveis de desgaste acima do tolerado. Construir um intervalo de confiança para a proporção de pastilhas com desgaste acima do tolerado, com nível de confiança de 95%. Interpretar o resultado.

5.4) (Adaptado de Barbetta) Uma fundição produz blocos para motor de caminhões. Os blocos têm furos, e deseja-se verificar qual é o diâmetro médio no processo do furo. A empresa retirou uma amostra de 36 blocos e mediu os diâmetros de 36 furos (1 a cada bloco). A amostra acusou média de 98,0 mm e desvio-padrão de 4,0 mm.
(a) Construir um intervalo de confiança para a média do processo, com nível de confiança de 99%. Interpretar o resultado.
(b) Se o processo deveria ter média de 100 mm, há evidência (com 99% de confiança) de que a média do processo não está no valor ideal? Explique.

5.5) (Barbetta) Um pesquisador precisa determinar o tempo médio gasto para perfurar três orifícios em uma peça de metal. Qual deve ser o tamanho da amostra para que a média amostral esteja a menos de 15 s da média populacional? Por experiência prévia, pode-se supor o desvio-padrão em torno de 40 s. Considere também, que a estimação será realizada com nível de confiança de 95%.

5.6) (Adaptado de Montgomery) O diâmetro dos orifícios para arreios de cabo tem um desvio-padrão de 0,01 in. Uma amostra aleatória de tamanho 10 resulta em um diâmetro médio de 1,5045 in.
(a) Teste a hipótese de que o diâmetro médio verdadeiro do orifício seja igual a 1,50 in. Use $\alpha = 0{,}01$.
(b) Qual seria o tamanho necessário da amostra para detectar o diâmetro médio verdadeiro do orifício com um erro menor que 0,005 in, com uma confiança de 90%?

5.7) (Adaptado de Montgomery) Um fabricante produz anéis para pistões de um motor de automóveis. É sabido que o diâmetro do anel é distribuído de forma aproximadamente normal e tem um desvio-padrão de 0,001 mm. Uma amostra aleatória de 15 anéis tem um diâmetro médio de 74,036 mm.
(a) Teste a hipótese de que o diâmetro médio do anel do pistão seja 74,035 mm.
(b) Teste a hipótese de que o diâmetro médio do anel do pistão seja maior que 74,035 mm.

5.8) (Adaptado de Montgomery) De 1.000 casos selecionados aleatoriamente de câncer de pulmão, 823 resultaram em morte.
(a) Construa um intervalo bilateral de confiança de 95% para a taxa de morte de câncer de pulmão.
(b) Quão grande seria a amostra requerida, de modo a se estar pelo menos 95% confiante de que o erro na estimação da taxa de morte de câncer de pulmão seja menor do que 3%?

5.9) (Adaptado de Montgomery) Sabe-se que a vida em horas de um bulbo de uma lâmpada de 75 W é distribuída de forma aproximadamente normal, com desvio-padrão 25 horas. Uma amostra de 20 bulbos apresentou uma vida média de 1.014 horas.
(a) Há alguma evidência que suporte a alegação de que a vida do bulbo excede 1.000 horas? Use $\alpha = 0{,}05$.
(b) Construa um intervalo de 95% de confiança bilateral para a vida média.

(c) Suponha que quiséssemos estar 95% confiantes de que o erro na estimação da vida média fosse menor do que 5 horas. Qual tamanho da amostra deveria ser usado?

5.10) (Adaptado de Montgomery) Um engenheiro de desenvolvimento de um fabricante de pneu está investigando a vida do pneu em relação a um novo componente da borracha. Ele fabricou 16 pneus e testou-os até o final da vida em um teste na estrada. A média e o desvio-padrão da amostra são 60.139,7 e 3.645,94 km, respectivamente.

(a) Encontre um intervalo bilateral de 95% de confiança para a vida média do pneu.

(b) O engenheiro gostaria de demonstrar que a vida média desse novo pneu excede 60.000 km. Formule e teste as hipóteses apropriadas e retire conclusões, usando $\alpha = 0{,}10$.

(c) Suponha que o engenheiro gostaria de detectar um erro de 1.000 km para a vida média com uma confiança de 90%. O tamanho da amostra usado foi adequado? Use s como estimativa de σ para obter sua decisão.

5.11) (Adaptado de Montgomery) Uma máquina produz bastões metálicos usados em um sistema de suspensão de automóveis. Uma amostra aleatória de 15 bastões é selecionada, sendo o diâmetro medido. Os dados resultantes (em mm) são mostrados a seguir:

8,24	8,21	8,23	8,25	8,26
8,23	8,20	8,26	8,19	8,23
8,20	8,28	8,24	8,25	8,24

(a) Existe alguma evidência forte para indicar que o diâmetro médio dos bastões exceda 8,20 mm, usando $\alpha = 0{,}05$.

(b) Encontre um intervalo bilateral de 95% de confiança para o diâmetro médio dos bastões.

5.12) (Adaptado de Montgomery) A espessura da parede de 25 garrafas de 2 litros foi medida por um engenheiro do controle da qualidade. A média da amostra foi 4,05 mm e o desvio-padrão da amostra foi 0,08 mm.

(a) Suponha ser importante demonstrar que a espessura da parede excede 4,00 mm. Formule e teste uma hipótese apropriada, usando esses dados. Obtenha conclusões para $\alpha = 0{,}01$.

(b) Encontre um intervalo de 99% de confiança para a espessura média. Interprete o intervalo que você obteve.

5.13) (Montgomery) Um fabricante de calculadoras eletrônicas está interessado em estimar a fração de unidades defeituosas produzidas. Uma amostra aleatória de 800 calculadoras contém 10 defeitos. Calcule um intervalo de confiança de 99% para a fração defeituosa.

5.14) (Adaptado de Montgomery) Um artigo em *Fortune* (21 de setembro de 1992) afirma que aproximadamente metade de todos os engenheiros continuam seus estudos acadêmicos além do grau de bacharelado, recebendo no final o grau de mestre ou doutor. Dados de um artigo em *Engineering Horizons* (primavera de 1990), indicou que 117 de 484 novos engenheiros graduados estavam planejando fazer uma pós-graduação. Os dados da *Engineering Horizons* são consistentes com a afirmação reportada pela *Fortune*? Use $\alpha = 0{,}05$ para encontrar as suas conclusões.

5.15) A garantia para baterias de fones móveis é estabelecida em 200 horas operacionais, seguindo os procedimentos apropriados de recarga. Um estudo com 5.000 baterias foi executado e 15 pararam de operar antes das 200 horas. Esses experimentos confirmam a afirmação de que mais de 0,2% das baterias da companhia falhará durante o período de garantia, usando os procedimentos apropriados de recarga? Use o procedimento de teste de hipóteses, com $\alpha = 0{,}01$.

5.16) (Barbetta) Uma empresa tem 2.400 empregados. Deseja-se extrair uma amostra de empregados para verificar o grau de satisfação em relação à qualidade da comida no refeitório. Em uma amostra piloto, numa escala de 0 a 10, o grau de satisfação recebeu nota média 6,5 e desvio-padrão 2,0 pontos.

(a) Determine o tamanho mínimo da amostra, supondo amostragem aleatória simples, com erro máximo de 0,5 ponto e nível de confiança de 99%.

(b) Considere que a amostra planejada no item anterior tenha sido realizada e obteve-se média 5,3 e desvio-padrão 1,8 ponto. Construa um intervalo de 99% de confiança para o parâmetro de interesse.

(c) Considerando o resultado do item anterior, você diria, com nível de confiança de 99%, que a nota média seria superior a cinco se a pesquisa fosse aplicada a todos os 2.400 funcionários? Justifique.

(d) Realizada a amostra planejada no item (a), suponha que 70 atribuíram notas iguais ou superiores a cinco. Apresente um intervalo de 90% de confiança para a porcentagem de indivíduos da população que atribuiriam notas iguais ou superiores a cinco.

5.17) (Barbetta) Um fabricante garante que 90% de seus itens estão dentro das especificações. Um comprador examinou uma amostra aleatória de 50 itens e verificou que apenas 84% estavam dentro das especificações. Há evidência de que o nível de qualidade é menor do que o alegado pelo fabricante? Use $\alpha = 2\%$.

AI.6 Correlação e Regressão Linear

6.1) Os pesos (em kg) de oito veículos e a variabilidade de suas distâncias de frenagem (em metros) em uma superfície seca estão dispostos na tabela a seguir.

Peso	1837	1937	2268	2273	2495	2495	2595	2781
Variabilidade na distância de frenagem	0,457	0,485	0,588	0,491	0,506	0,518	0,543	0,582

(a) Escolha adequadamente *X* e *Y*.

(b) Você espera uma relação linear direta ou inversa entre as variáveis *X* e *Y*. Explique.

(c) Construa o diagrama de dispersão entre *X* e *Y*.

(d) Calcule o coeficiente de correlação linear de Pearson. Interprete o valor obtido.

(e) Obtenha a equação de regressão que melhor descreve a associação entre *X* e *Y*. Interprete os valores de b_0 e b_1.

(f) Faça uma previsão da variabilidade da distância de frenagem em superfícies secas para um veículo com 900 kg e para um veículo com 3 ton.

(g) Calcule e interprete o coeficiente de determinação R^2 para a equação de regressão estimada no item (f).

Considere agora as variabilidades nas distâncias de frenagem (em metros) dos mesmos veículos, mas agora em uma superfície molhada:

Peso	1837	1937	2268	2273	2495	2495	2595	2781
Variabilidade na distância de frenagem	0,741	0,841	0,878	0,991	1,042	1,07	1,152	1,411

(h) Escolha adequadamente *X* e *Y*.

(i) Você espera uma relação linear direta ou inversa entre as variáveis *X* e *Y*. Explique.

(j) Construa o diagrama de dispersão entre *X* e *Y*.

(k) Calcule o coeficiente de correlação linear de Pearson. Interprete o valor obtido.

(l) Obtenha a equação de regressão que melhor descreve a associação entre *X* e *Y*. Interprete os valores de b_0 e b_1.
(m) Faça uma previsão da variabilidade da distância de frenagem em superfícies molhadas para um veículo com 900 kg e para um veículo com 3 ton.
(n) Calcule e interprete o coeficiente de determinação R^2 para a equação de regressão estimada no item (n).
(o) Existe diferença entre a variabilidade da frenagem em superfícies secas e molhadas? O que você recomendaria aos motoristas?

6.2) Os dados a seguir referem-se ao número de CDs vendidos por uma determinada gravadora, em milhares de unidades, em 10 semanas consecutivas após o lançamento do mesmo.

Semana	1	2	3	4	5	6	7	8	9	10
CDs (x 1000)	5,0	6,7	6,0	8,7	6,2	8,6	11,0	11,9	10,6	10,8

(a) Escolha adequadamente *X* e *Y*.
(b) Você espera uma relação linear direta ou inversa entre as variáveis X e Y. Explique.
(c) Construa o diagrama de dispersão entre X e Y.
(d) Calcule o coeficiente de correlação linear de Pearson. Interprete o valor obtido.
(e) Obtenha a equação de regressão que melhor descreve a associação entre *X* e *Y*. Interprete os valores de b_0 e b_1.
(f) Faça uma previsão da quantidade de CDs que serão vendidos na 20ª semana.
(g) Calcule e interprete o coeficiente de determinação R^2 para a equação de regressão estimada no item (f).

6.3) As exportações de castanha in natura (em toneladas), processadas pela empresa Yasmin Ltda., no período que se estende de 1983 a 1989, encontram-se na tabela a seguir:

Ano	1983	1984	1985	1986	1987	1988	1989
Quantidade	50	46	36	31	25	11	18

(a) Escolha adequadamente *X* e *Y*.
(b) Construa o diagrama de dispersão entre X e Y.
(c) Calcule o coeficiente de correlação linear de Pearson. Interprete o valor obtido.
(d) Obtenha a equação de regressão que melhor descreve a associação entre *X* e *Y*. Interprete os valores de b_0 e b_1.
(e) Encontre a quantidade de exportações estimada para o ano de 1990.

6.4) Numa pesquisa feita com 10 famílias com renda bruta mensal entre 10 e 60 salários mínimos, mediram-se as seguintes variáveis:
- Renda*: renda bruta mensal* (em salários mínimos);
- %Renda*: porcentagem da renda bruta anual gasta com assistência médica*;

Renda	12	16	18	20	28	30	40	48	50	54
%Renda	7,2	7,4	7,0	6,5	6,6	6,7	6,0	5,6	6,0	5,5

(a) Escolha adequadamente *X* e *Y*.
(b) Construa o diagrama de dispersão entre X e Y.
(c) Calcule o coeficiente de correlação linear de Pearson. Interprete o valor obtido.
(d) Obtenha a equação de regressão que melhor descreve a associação entre *X* e *Y*. Interprete os valores de b_0 e b_1.
(e) Faça uma previsão da porcentagem da renda bruta anual gasta com assistência médica para uma família com renda bruta mensal de 10 salários mínimos.

(f) Calcule e interprete o coeficiente de determinação R^2 para a equação de regressão estimada no item (d).

6.5) Os dados adiante fornecem, para 11 países, o consumo de cigarros per capta em 1930 e as mortes por 1.000.000 de habitantes em 1950 causadas por câncer no pulmão.

País	Consumo de cigarros	Mortes (por 1.000.000 de hab.)
Islândia	240	63
Noruega	255	100
Suécia	340	140
Dinamarca	375	175
Canadá	510	160
Austrália	490	180
Holanda	490	250
Suíça	180	180
Finlândia	1.125	360
Grã-Bretanha	1.150	470
EUA	1.275	200

(a) Escolha adequadamente X e Y.
(b) Construa o diagrama de dispersão entre X e Y.
(c) Calcule o coeficiente de correlação linear de Pearson. Interprete o valor obtido.
(d) Obtenha a equação de regressão que melhor descreve a associação entre X e Y. Interprete os valores de b_0 e b_1.
(e) Se, no ano de 1930, o consumo de cigarros *per capta* no Brasil foi 630, estime o número de mortes causadas por câncer de pulmão no ano de 1950.
(f) Calcule e interprete o coeficiente de determinação R^2 para a equação de regressão estimada no item (d).

ANEXO II – RESPOSTAS DOS EXERCÍCIOS

AII.1 Análise Descritiva de Dados

1.1)

(a) As empresas que possuem executivos CEOS.

(b) n = 10

(c) Quatro variáveis. Remuneração dos CEOS (quantitativa contínua); Setor (qualitativa nominal); Vendas (quantitativa contínua); Remuneração dos CEOS vs. Retorno dos Acionistas (qualitativa ordinal).

(d) 20%.

(e) 30%.

(f) Média aritmética = 2.791,400; Mediana = 2.142,0; Moda = 1.490.

(g) Amplitude = 8.229; Variância = 5446959,156; Desvio-padrão = 2.333,872; CV = 83,6%.

1.2)

(a) O preço da diária para apartamento de casal (quantitativa contínua).

(b)

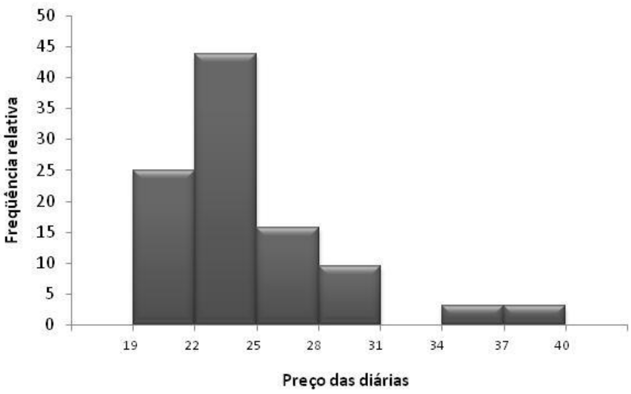
1	1	9
8	2	0000001
14	2	222233
(13)	2	4444444455555
5	2	
5	2	88
3	3	0
2	3	
2	3	5
1	3	
1	3	8

Chave: 1|9 = 19

(c)

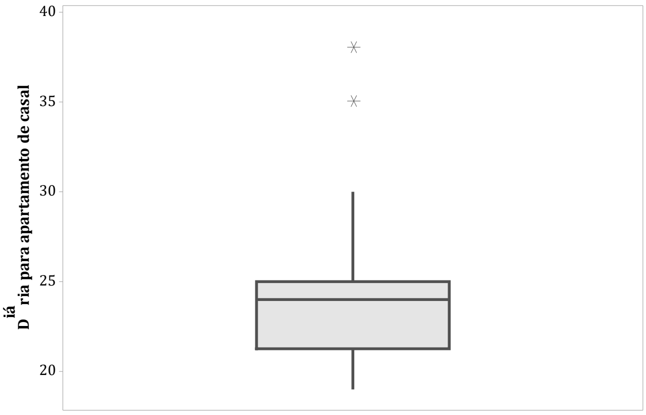
Preço da diária	n	%	n acum.	% acum.
19 ⊢ 22	8	25,000	8	25,000
22 ⊢ 25	14	43,750	22	68,750
25 ⊢ 28	5	15,625	27	84,375
28 ⊢ 31	3	9,375	30	93,750
31 ⊢ 34	0	0,000	30	93,750
34 ⊢ 37	1	3,125	31	96,875
37 ⊢ 40	1	3,125	32	100,000
Total	32	100,000		

(d)



(e) $Q_1 = 21{,}5$ $Q_2 = 24{,}0$ $Q_3 = 25{,}0$ $AIQ = 1{,}5$

(f)



(g) Média = 24,0625 Moda = 24,0 Variância = 17,351 CV = 17,311%

Mediana = 24,0 Amplitude = 19 Desvio-padrão = 4,165

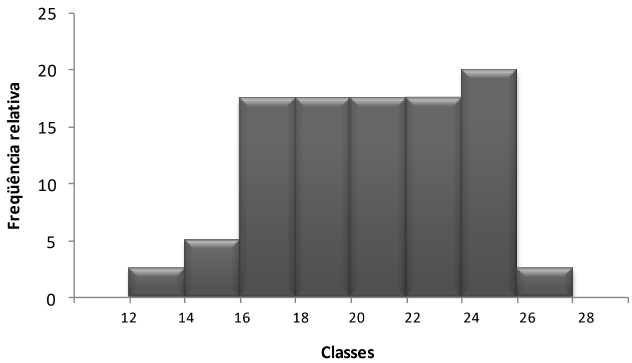
(h) Sim. O CV é menor que 40%.

1.3)

(a)

Classes	n	%	n acum.	% acum.
12 ⊢ 14	1	2,500	1	2,500
14 ⊢ 16	2	5,000	3	7,500
16 ⊢ 18	7	17,500	10	25,000
18 ⊢ 20	7	17,500	17	42,500
20 ⊢ 22	7	17,500	24	60,000
22 ⊢ 24	7	17,500	31	77,500
24 ⊢ 26	8	20,000	39	97,500
26 ⊢ 28	1	2,500	40	100,000
Total	40	100,000		

(b)



(c) Média = 20,150 Mediana = 20,0 Moda = 16 Amplitude = 14,0 Variância = 12,285
Desvio-padrão = 3,505 CV = 17,395% Q_1 = 17,5 Q_2 = 20,0 Q_3 = 23,0

1.4)

(a) Os salários anuais para os gerentes assistentes de lojas de departamentos. Quantitativa contínua.

(b)

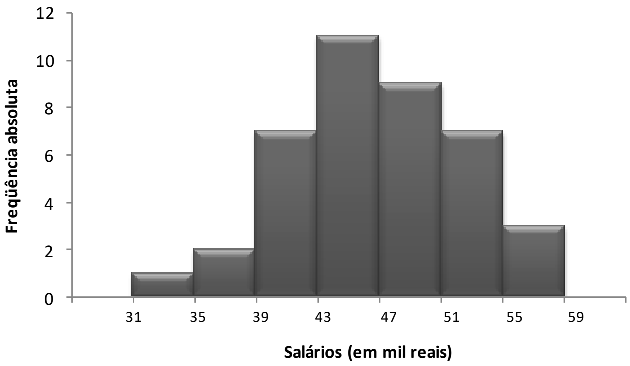
1	3	1
4	3	579
14	4	0011223444
(12)	4	555556677889
14	5	00001122234
3	5	567

Chave: 3|1 = 31

(c)

Salários	n	%	n acum.	% acum.
31 ⊢ 35	1	2,500	1	2,500
35 ⊢ 39	2	5,000	3	7,500
39 ⊢ 43	7	17,500	10	25,000
43 ⊢ 47	11	27,500	21	52,500
47 ⊢ 51	9	22,500	30	75,000
51 ⊢ 55	7	17,500	37	92,500
55 ⊢ 59	3	7,500	40	100,000
Total	40	100,000		

(d)



1.5)

(a)

N° de funcionários	n	%	n acumulada	% acumulada
0 a 5	23908	69,126	23908	69,126
6 a 19	8192	23,686	32100	92,812
20 a 49	1693	4,895	33793	97,707
50 a 99	447	1,292	34240	98,999
Mais de 99	346	1,000	34586	100,000
Total	34586	100,000		

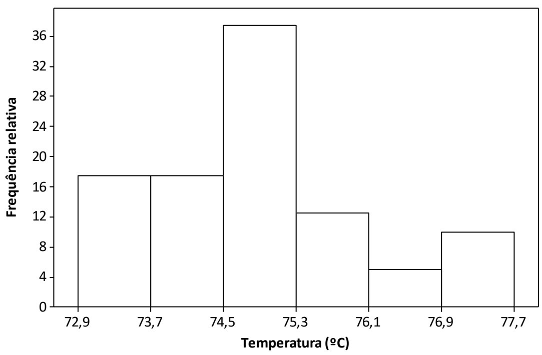
(b)
i. 8192; 23,686%
ii. 97,707%
iii. 2486; 7,188%

1.6)

(a)

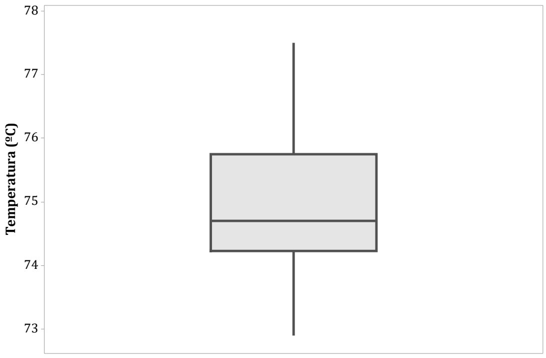
Temperatura (°C)	n	%	n acum.	% acum.
72,9 ⊢ 73,7	7	17,500	7	17,500
73,7 ⊢ 74,5	7	17,500	14	35,000
74,5 ⊢ 75,3	15	37,500	29	72,500
75,3 ⊢ 76,1	5	12,500	34	85,000
76,1 ⊢ 76,9	2	5,000	36	90,000
76,9 ⊢ 77,7	4	10,000	40	100,000
Total	40	100,000		

(b)



(c) $Q_1 = 74{,}25$ $Q_2 = 74{,}7$ $Q_3 = 75{,}7$

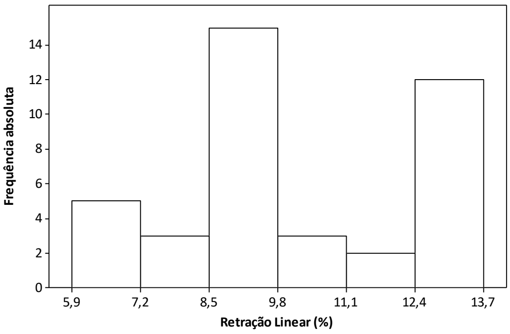
(d)

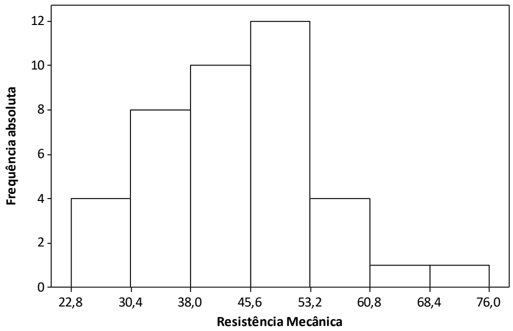


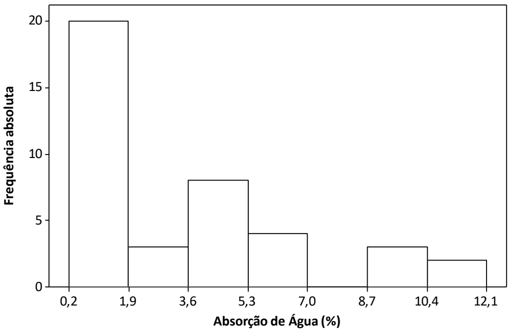
1.7)

(a) X_1: Quantitativa contínua, X_2: Quantitativa contínua, X_3: Quantitativa contínua.

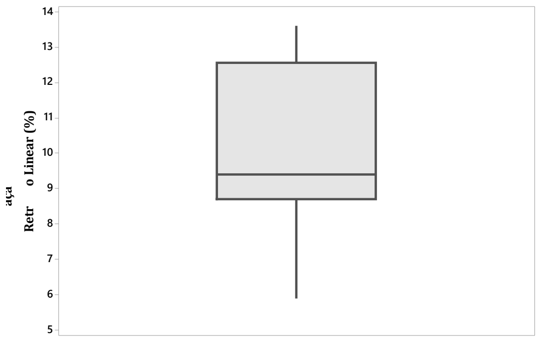
(b)

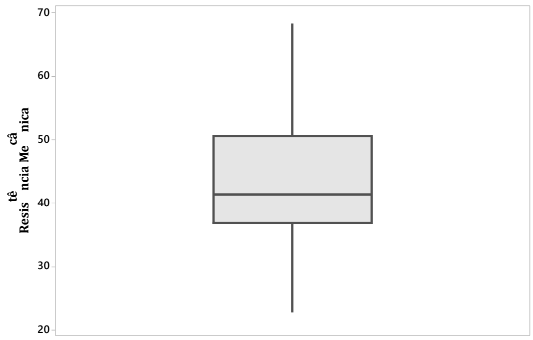


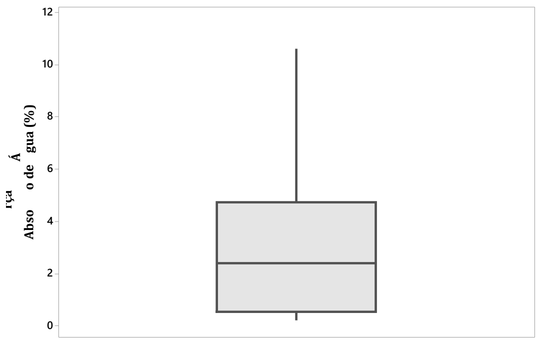




(c)







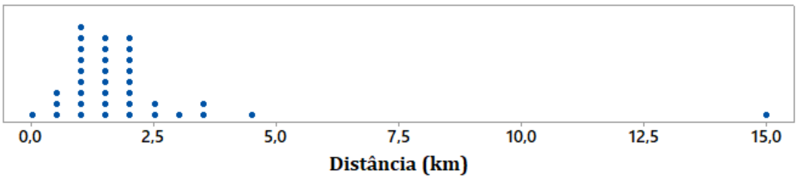
(d) Condição 1 — $\bar{x}_3$ = 5,54; s_3 = 0,503
Condição 4 — $\bar{x}_3$ = 9,920; s_3 = 0,726
Condição 8 — $\bar{x}_3$ = 4,140; s_3 = 0,207

(e) Q_1 = 8,7 Q_2 = 9,4 Q_3= 12,5

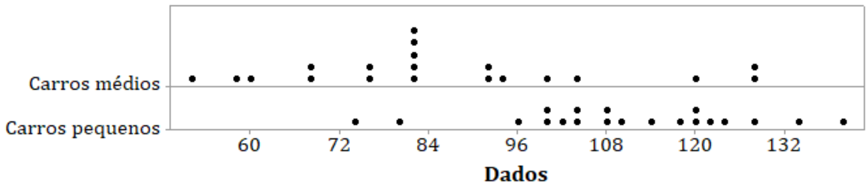
(f) $\bar{x}_1$ = 10,037; s_1 = 2,337; CV_1 = 23,279%
$\bar{x}_2$ = 43,575; s_2 = 10,413; CV_2 = 23,898%
$\bar{x}_3$ = 3,240; s_3 = 3,179; CV_3 = 98,115%

1.8) (a) Q_1 = 0,95 Q_2 = 1,6 Q_3= 2,1 Máx = 15,1 Mín = 0,2

(b)



1.9)
(a)



(b)

Carros	n	Média	Mediana	Moda	Variância	Desvio padrão	Amplitude
Médios	20	85,75	81,5	81	461,987	21,494	77
Pequenos	20	109,9	108,5	100, 103, 108	270,937	16,460	67

(c) $CV_{Médios}$ = 25,066%
$CV_{Pequenos}$ = 14,977%

O grupo dos carros pequenos é mais homogêneo, porque possui menor coeficiente de variação.

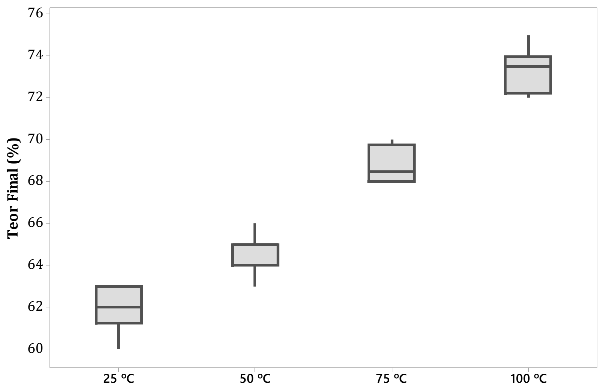
1.10)
(c) O teor final da substância de interesse (quantitativa contínua).

(b)

Temperatura	n	Média	Mediana	Moda	Amplitude	Variância	Desvio-padrão	CV
25°C	12	62,000	62,0	63	3	1,273	1,128	1,819%
50°C	12	64,583	65,0	65	3	0,992	0,996	1,542%
75°C	12	68,750	68,5	68	2	0,750	0,866	1,260%
100°C	12	73,417	73,5	74	3	1,174	1,084	1,476%

(c) O processo a 100°C é mais homogêneo (possui menor CV).

(d)

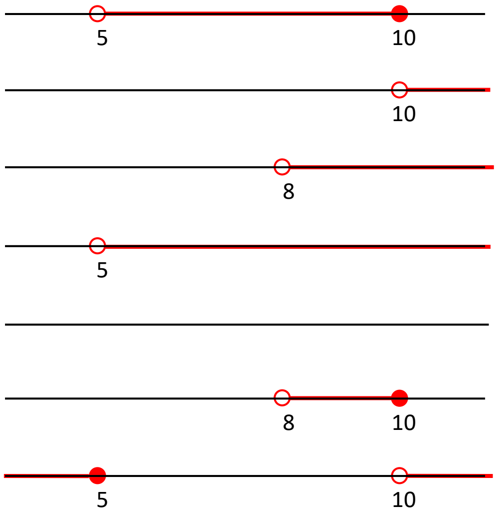


AII.2 Probabilidade

2.1) (a) $E = \{cara, coroa\}$
(b) $E = \{0, 1, 2, \dots\}$
(c) $E = \{0, 1, 2, \dots\}$
(d) $E = \{v, \text{tal que } v \geq 0\}$
(e) $E = \{t, \text{tal que} -\infty \leq t \leq +\infty\}$

2.2) (a)
i. $A = \{t, \text{tal que } 5 < t \leq 10\}$
ii. $B = \{t, \text{tal que } t > 10\}$
iii. $C = \{t, \text{tal que } t > 8\}$
iv. $D = \{t, \text{tal que } t > 5\}$
v. $E = \emptyset$
vi. $F = \{t, \text{tal que } 8 < t \leq 10\}$
vii. $G = \{t, \text{tal que } t \leq 5 \text{ e } t > 10\}$

(b) i. ii. iii. iv. v. vi. vii.



2.3) $P(\bar{C})$ = não existirão fundos suficientes para a expansão planejada
$P(\bar{E})$ = a expansão planejada não incluirá um estacionamento suficientemente amplo
$P(\bar{C} \cap E)$ = não existirão fundos suficientes para a expansão planejada e a expansão planejada incluirá um estacionamento suficientemente amplo
$P(C \cap \bar{E})$ = existirão fundos suficientes para a expansão planejada e a expansão planejada não incluirá um estacionamento suficientemente amplo

2.4) (a) Porque não existe probabilidade negativa.
(b) Porque não existe probabilidade maior que um.
(d) Porque não existe probabilidade maior que um.

2.5) $P(A) = 0$

2.6) (a) 0,750 (b) 0,423

2.7) (a) $P(A) = 0{,}400$ $P(B) = 0{,}300$ $P(C) = 0{,}140$ $P(D) = 0{,}100$ $P(E) = 0{,}060$
(b) $P(\bar{A}) = 0{,}600$

2.8) (a) $P(\bar{A}) = 0{,}5$ $P(A \cap B) = 0{,}083$ $P(A \cup B) = 0{,}667$
(d) $P(A \cap \bar{B}) = 0{,}417$ $P(\bar{A} \cup \bar{B}) = 0{,}917$

2.9) (a) $P(\bar{1} \cap \bar{2}) = 0{,}72$ $P(\bar{1} \cup \bar{2}) = 0{,}98$

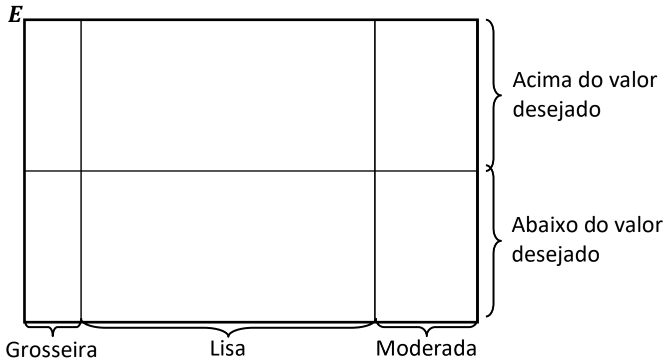
2.10) (a) $P(A \text{ ou } B) = 0{,}600$ (b) $P(A|B) = 0{,}667$ (c) $P(B|A) = 0{,}667$

2.11) (a) $P(C|P) = 0{,}900$ (b) $P(P|C) = 0{,}960$

2.12) (a) i. 0,0488 ii. 0,0065 iii. 0,5152 iv. 0,0137
(b) Não.

2.13) (a) 0,347 (b) 0,507 (c) 0,053 (d) 0,400 (e) 0,588 (f) Não

2.14) (a) 0,100 (b) 0,625 (c)



2.15) (a) 0,100 (b) 0,090 (c) 0,500 (d) 0,140

2.16) 0,0039

2.17) 0,00299

2.18) (a) 0,615 (b) 0,618 (c) 0,364 (d) 0,584

2.19) 0,3226

2.20) (a) 0,760 (b) 0,921

2.21) 0,174

2.22) 0,660

2.23) 0,00009899

AII.3 Distribuições Discretas de Probabilidade

3.1) (a) $P(X = 0) = 0{,}368$ (b) $P(X > 2) = 0{,}080$

3.2) (a) 0,5966 (b) 0,3182 (c) 0,9148

3.3)

y	0	1	2
$p(y)$	0,64	0,32	0,04

3.4) (a) 0,15 (b) 0,90 (c) μ = 3,45 (d) σ^2 = 4,5475 e σ = 2,132

3.5) (a) $p(1)$ = 0,8 $p(2)$ = 0,16 $p(3)$ = 0,032 $p(4)$ = 0,0064 $p(5)$ = 0,00128
(b) $\sum_{i=1}^{5} p(x_i)$ = 0,99968

3.6) (a) 0,9 (b) 0,43 e 0,2 (c) 0,83 (d) CV = 145,673%

3.7)
(a)

x	0	1	2	3
$p(x)$	0,0476	0,3571	0,4762	0,1190

(b) (b1) 0,8810 (b2) 0,476 (b3) 0,5952 (b4) 0,000 (b5) 0,1190 (b6) 1,000 (b7) 1,000
(c) (c1) 0,8810 (c2) 1,000 (c3) 0,0476 (c4) 1,000 (c5) 0,8810 (c6) 0,4048 (c7) 1,000 (c8) 0,000

3.8)
(a)

x	0	1,5	2	3
$p(x)$	$1/3$	$1/3$	$1/6$	$1/6$

(b)

x	0	1,5	2	3
$F(x)$	$1/3$	$2/3$	$5/6$	1

(c) μ = 1,333 e σ^2 = 1,14

3.9)
(a)

x	10	5	1
$p(x)$	0,3	0,6	0,1

(b)

x	10	5	1
$F(x)$	0,3	0,9	1,0

(c) μ = 6,1 milhões e σ^2 = 7,89 milhões

3.10) (a) 0,512
(b)

x	0	1	2	3
$p(x)$	0,008	0,096	0,384	0,512

(c)

x	0	1	2	3
$F(x)$	0,008	0,104	0,488	1,000

3.11) (a) 0,92169
(b)

x	0	1	2	3
$p(x)$	0,00001	0,00167	0,07663	0,92169

3.12) c = 0,1

3.13) (a) μ_X = 10,7 e μ_Y = 11,0. Resposta: Y é mais interessante
(b) σ_X^2 = 25,61 e σ_Y^2 = 3,8. Resposta: X é mais arriscada

3.14) μ_i = R$ 500.000 e μ_{ii} = R$ 583.708. Resposta: opção *ii* (com inspeção).

3.15) μ_A = R$ 50 e μ_B = R$ 30. Resposta: comprador A.

3.16)

x	0	1	2
$p(x)$	0,09	0,42	0,49

μ = 1,4 e σ = 0,648

3.17) (a) 0,205 (b) 0,989 (c) 0,00098 (d) 0,9990 (e) 0,754

3.18) (a) 0,329 (b) 0,9959 (c) 0,790

3.19) (a) 0,329 (b) 0,088

3.20) 0,000023

3.21) (a) $n = 5$ (b) $p = 0,667$ e $(1 - p) = 0,333$ (c) $\mu = 3,333$ (d) $\sigma^2 = 1,111$ (e) 0,996 (f) 0,823

3.22) (a) 0,0059 (b) 0,140 (c) 1,000

3.23)

x	0	1	2	3
$F(x)$	0,4219	0,8438	0,9844	1,000

3.24) (a) 0,368 (b) 0,632 (c) 0,920 (d) $\mu = 1$ e $\sigma^2 = 0,999$

3.25) (a) $n = 50$ e $p = 0,1$ (b) 0,112 (c) 0,000

3.26) (a) 0,996 (b) 0,989

3.27) (a) 0,00000000096769 (Obs: Menor do que a probabilidade de ganhar na Mega-sena!!!)
(b) 0,2137

3.28) (a) 0,082 (b) 0,544 (c) 0,865

3.29) (a) 0,101 (b) 0,088 (c) 0,941 (d) 0,030

3.30) (a) 0,323 (b) 0,036 (c) 0,135 (d) 0,406

3.31) (a) 0,878 (b) 0,202

3.32) (a) 0,168 (b) 0,577

3.33) 0,195

3.34) (a) 0,018 (b) 0,073 (c) 0,147 (d) 0,908

3.35) (a) 0,449 (b) 0,144

3.36) (a) 0,271 (b) 0,191

3.37) (a) 0,061 (b) 0,809

3.38) 0,147

3.39) 0,193

3.40) 0,143

3.41) 0,934

3.42) (a) 0,6767 (b) 0,092

3.43) (a) 0,00005 (b) 0,632

3.44) (a) 0,852 (b) 0,381

3.45) 0,981

3.46) 1/3973 = 0,00025

3.47) 0,999852 (Nunca faça isso se você for alérgico a camarão!)

3.48) (a) 0,4789 (b) 0,0789 (c) 10

3.49) 0,108453

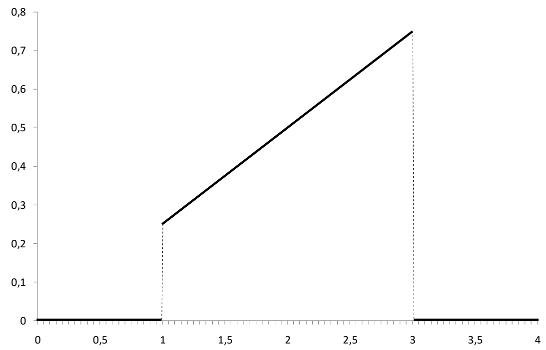
3.50) 0,00364

3.51) 0,499071

AII.4 Distribuições Contínuas de Probabilidade

4.1) (a) k = 128 (b) 0,64 (c) 16

4.2)
(a)



(b) (b1) 0,625 (b2) 0,625 (b3) 0,000 (b4) 0,156 (b5) 0,844 (b6) 0,844

4.3) (a) 0,5 (b) 0,375

4.4) $E(X)$ = 1 $Var(X)$ = 0,6

4.5) Nenhuma.

4.6)
(a) k = 1

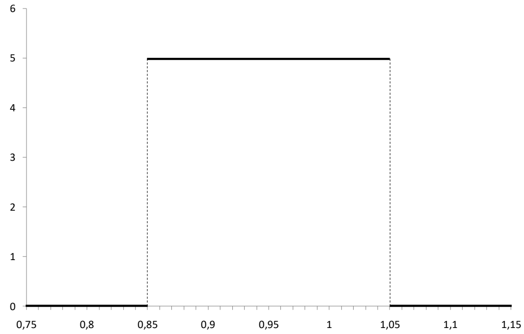
(b) $P(X = 1) = 0{,}000$ $P(X \leq 0) = 0{,}750$ $P(X < 0) = 0{,}750$ $P(0 < X \leq 1) = 0{,}25$

$P(0 < X \leq 3) = 0{,}25$ $P(X \geq 1) = 0{,}000$ $P(3 \leq X \leq 4) = 0{,}000$

(c) $E(X) = -\frac{1}{3}$ e $Var(X) = \frac{1}{3}$

(d) $F(X) = \begin{cases} 0, & x < 1 \\ -\left(\frac{x^2-2x-3}{4}\right), & -1 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$

4.7) (a) $k = 5$



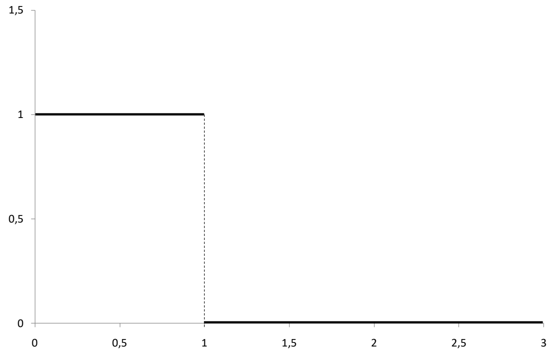
(b) $E(X) = 0{,}95$ e $Var(X) = 0{,}003$

(c) 0,75

(d) 50

4.8)

(a)



(b) 0,50

(c) 0,30

(d) 0,40

4.9) (a) 0,40 (b) 0,64 (c) 0,68

4.10) (a) 0,082 (b) 0,148 (c) 0,00421 hora = 0,25 minuto

4.11) (a) 0,1353 (b) 0,4866 (c) 0,2031 (d) 34,54

4.12) (a) 0,0025 (b) 0,6321 (c) 23,026 (d) 23,026 (e) 6,931

4.13) (a) 0,2212 (b) 0,2865 (c) 0,2212 (d) 0,9179 (e) 0,2336

4.14) 1,66%

4.15) 0,6826; 0,9545; 0,9973

4.16) (a) i. 0,8389 ii. 0,6217 iii. 0,629 (b) 80,984 (c) [0,223;3,857]

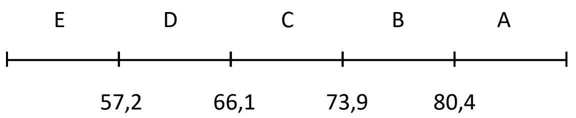
4.17) 5 anos

4.18) (a) i. 0,9938 ii. 0,9938 iii. 0,8664 (b) 0,000

4.19) (a) 0,00175 (b) 0,02619 (c) 0,90508 (d) [1298,88;1581,12]

4.20) (a) 0,00175 (b) 0,22836 (c) 0,10749 (d) [2,45;15,35]

4.21) (a) 0,3530 (b) 2,05

4.22) 

4.23) (a) 0,26763 (b) 55,24 minutos

4.24) 88,6 e 55,3

4.25) (a) 50% (b) 2,28% (c) 32

4.26) Caminho A

4.27) (a) 0,27% (b) 0,00000729 (c) 0,994607

AII.5 Intervalos de Confiança e Testes de Hipóteses

5.1) (a) $H_0: \mu \geq 100$ $Z = -3{,}000$ p = 0,00135
$H_1: \mu < 100$ $-z_\alpha = -1{,}64$ Resposta: Como -3,0 está na região de rejeição (ou ainda, como 0,00135 < 0,05), rejeitamos H_0. Logo, existem evidências de que a fibra não deve ser julgada como aceitável com 5% de significância.

(b) $IC_{95\%}$: [96,693; 99,307]

5.2) $IC_{95\%}$: [0,481; 0,619]

5.3) IC$_{95\%}$: [0,016; 0,044] – Com 95% de confiança, o processo está produzindo entre 1,6% e 4,4% das pastilhas com nível de desgaste acima do tolerado.

5.4) (a) IC$_{99\%}$: [96,28; 99,72]

(b) $H_0: \mu = 100$ $\quad Z = -3{,}000 \quad$ p = 0,0027

$H_1: \mu \neq 100$ $\quad z_{\frac{\alpha}{2}} = 2{,}58$

Resposta: Como -3,0 está na região de rejeição (ou ainda, como 0,0027 < 0,01), rejeitamos H_0. Logo, existem evidências de que a média do processo não está no valor ideal com 1% de significância.

5.5) n = 28

5.6) (a) $H_0: \mu = 1{,}50$ $\quad Z = 1{,}423 \quad$ p = 0,1556

$H_1: \mu \neq 1{,}50$ $\quad z_{\frac{\alpha}{2}} = 2{,}58$

Resposta: Como 1,423 não está na região de rejeição (ou ainda, como 0,1556 > 0,01), não rejeitamos H_0. Logo, existem evidências de que o diâmetro médio verdadeiro dos orifícios é igual a 1,50 in com 1% de significância.

(b) n = 11

5.7) (a) $H_0: \mu = 74{,}035$ $\quad Z = 3{,}873 \quad$ p = 0,0001

$H_1: \mu \neq 74{,}035$ $\quad z_{\frac{\alpha}{2}} = 1{,}96$

Resposta: Como 3,873 está na região de rejeição (ou ainda, como 0,0001 < 0,05), rejeitamos H_0. Logo, existem evidências de que o diâmetro médio dos anéis de pistão é diferente de 74,035 mm com 5% de significância.

(b) $H_0: \mu \leq 74{,}035$ $\quad Z = 3{,}873 \quad$ p = 0,00005

$H_1: \mu > 74{,}035$ $\quad z_{\alpha} = 1{,}64$

Resposta: Como 3,873 está na região de rejeição (ou ainda, como 0,00005 < 0,05), rejeitamos H_0. Logo, existem evidências de que o diâmetro médio dos anéis de pistão é maior que 74,035 mm com 5% de significância.

5.8) (a) IC$_{95\%}$: [0,799; 0,847]

(b) n = 622

5.9) (a) $H_0: \mu \leq 1.000$ $\quad Z = 2{,}504 \quad$ p = 0,00621

$H_1: \mu > 1.000$ $\quad z_{\alpha} = 1{,}64$

Resposta: Como 2,504 está na região de rejeição (ou ainda, como 0,00621 < 0,05), rejeitamos H_0. Logo, existem evidências de que a vida do bulbo excede 1.000 horas com 5% de significância.

(b) IC$_{95\%}$: [1.003,043; 1.024,957]

(c) n = 97

5.10) (a) IC$_{95\%}$: [58.197,325; 62.082,075]

(b) $H_0: \mu \leq 60.000$ $\quad T = 0{,}153 \quad$ 0,40 < p < 0,45

$H_1: \mu > 60.000$ $\quad t_{n-1;\alpha} = 1{,}341$

Resposta: Como 0,153 não está na região de rejeição (ou ainda, como p > 0,10), não rejeitamos H_0. Logo, existem evidências de que a vida média desse novo pneu não excede 60.000 km com 10% de significância.

(c) Não. O tamanho da amostra deveria ser de, pelo menos, n = 36 pneus.

5.11) (a) $H_0: \mu \leq 8{,}20$ $\quad T = 5{,}267 \quad$ p < 0,0005

$H_1: \mu > 8{,}20$ $\quad t_{n-1;\alpha} = 1{,}761$

Resposta: Como 5,267 está na região de rejeição (ou ainda, como p < 0,05), rejeitamos H_0. Logo, existem evidências para indicar que o diâmetro médio dos bastões excede 8,20 mm com 5% de significância.

(b) IC$_{95\%}$: [8,220; 8,248]

5.12) (a) $H_0: \mu \leq 4{,}00$ $\quad T = 3{,}125 \quad$ 0,001 < p < 0,005

$H_1: \mu > 4{,}00$ $t_{n-1;\alpha} = 2{,}492$ Resposta: Como 3,125 está na região de rejeição (ou ainda, como p < 0,01), rejeitamos H_0. Logo, existem evidências de que a espessura da parede da garrafa excede 4,00 mm com 1% de significância.

(b) $IC_{99\%}$: [4,005; 4,095] – Com 99% de confiança, é possível afirmar que a espessura média das garrafas está entre 4,005 e 4,095 mm.

5.13) $IC_{99\%}$: [0,0024; 0,0226]

5.14) $H_0: p = 0{,}5$ $Z = -11{,}352$ p ≈ 0,000
$H_1: p \neq 0{,}5$ $z_{\frac{\alpha}{2}} = 1{,}96$ Resposta: Como -11,352 está na região de rejeição (ou ainda, como 0,000 < 0,05), rejeitamos H_0. Logo, existem evidências de que os dados da *Engineering Horizons* não são consistentes com a afirmação reportada pela *Fortune* com 5% de significância.

5.15) $H_0: p \leq 0{,}002$ $Z = 1{,}583$ p = 0,057
$H_1: p > 0{,}002$ $z_\alpha = 2{,}33$ Resposta: Como 1,583 não está na região de rejeição (ou ainda, como 0,057 > 0,01), não rejeitamos H_0. Logo, não existem evidências de que mais de 0,2% das baterias da companhia falhará durante o período da garantia com 1% de significância.

5.16) (a) n = 107

(b) $IC_{99\%}$: [4,851; 5,749]

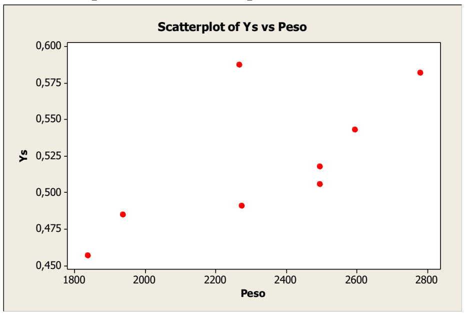
(c) $H_0: \mu \leq 5$ Resposta: Não, porque o intervalo de confiança construído apresenta valores menores que 5.
$H_1: \mu > 5$

(d) $IC_{90\%}$: [0,579; 0,729]

5.17) $H_0: p \geq 0{,}90$ $Z = -1{,}414$ p = 0,07927
$H_1: p < 0{,}90$ $-z_\alpha = -2{,}05$ Resposta: Como -1,414 não está na região de rejeição (ou ainda, como 0,07927 > 0,01), não rejeitamos H_0. Logo, não existem evidências de que o nível de qualidade é menor do que o alegado pelo fabricante, com 1% de significância.

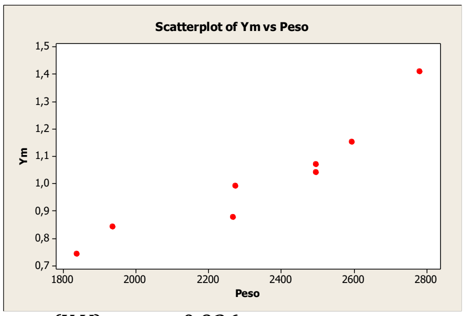
AII.6 Intervalos de Confiança e Testes de Hipóteses

6.1)
(a) X = Peso ; Y = Variabilidade na distância de frenagem em superfícies secas
(b) Direta: quanto maior o peso, maior a variabilidade
(c)



(d) corr(X,Y) = $r_{X,Y}$ = 0,697
(e) Y = 0,287 + 0,0001(X)
(f) 900 kg = 0,377 metros 3 ton = 0,587 metros
(g) R^2 = 48,615%
(h) X = Peso ; Y = Variabilidade na distância de frenagem em superfícies molhadas
(i) Direta: Quanto maior o peso, maior a variabilidade.

(j)



(k) corr(X,Y) = $r_{X,Y}$ = 0,936

(l) Y = -0,393 + 0,0006(X)

(m) 900 kg = 0,147 metros 3 ton = 1,407 metros

(n) R^2 = 87,621%

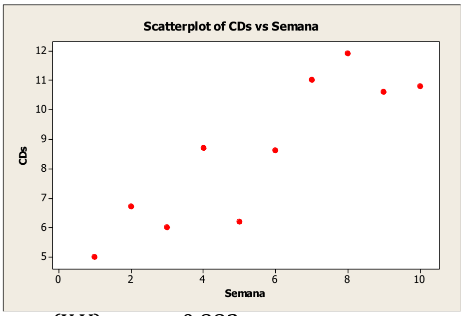
(o) Sim

6.2)

(a) X = Semana ; Y = Número de CDs vendidos

(b) Direta: conforme o tempo passa, o número de CDs vendidos aumenta

(c)



(d) corr(X,Y) = $r_{X,Y}$ = 0,882

(e) Y = 4,607 + 0,717 (X)

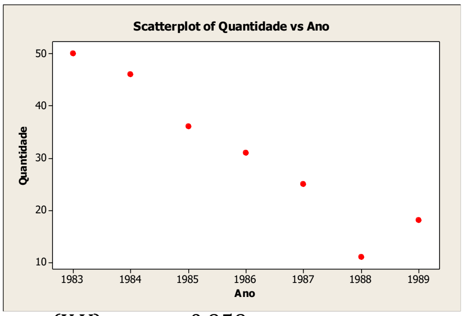
(f) 18,947 milhares de unidades = 18.947 unidades

(g) R^2 = 77,721%

6.3)

(a) X = ano ; Y = quantidade de castanha in natura exportada.

(b)



(c) corr(X,Y) = $r_{X,Y}$ = -0,959

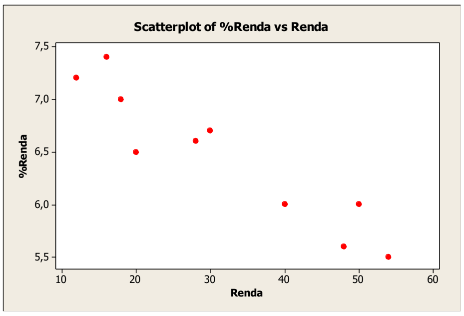
(d) Y = 12585,357 – 6,321 (X)

(e) 6,21 toneladas

6.4)

(a) X = renda ; Y = %Renda.

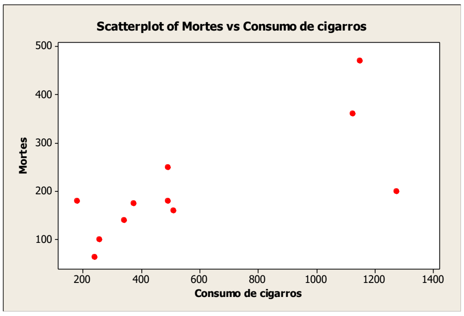
(b)



(c) corr(X,Y) = $r_{X,Y}$ = -0,940

(d) Y =7,716 – 0,040 (X)

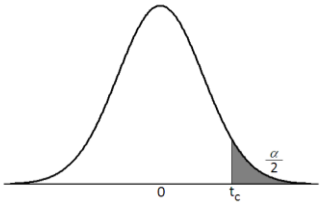
(e) 7,316%
(f) R^2 = 88,447%

6.5)
(a) X = consumo de cigarros ; Y = mortes.
(b)


(c) corr(X,Y) = $r_{X,Y}$ = 0,744
(d) Y = 80,68 + 0,216 (X)
(e) 216,76 mortes por 1.000.000 de habitantes
(f) R^2 = 55,313%

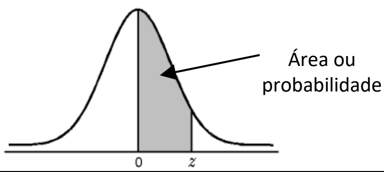
ANEXO III – TABELAS

Tabela da Distribuição t de Student



g.l.	Valores t_C tais que $P(T > t_C\ ) \ = \ \frac{\alpha}{2}$															g.l.
	45%	40%	35%	30%	25%	20%	15%	10%	5%	2,5%	2%	1%	0,5%	0,1%	0,05%	
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	15,895	31,821	63,657	318,31	636,62	1
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	4,849	6,965	9,925	22,327	31,599	2
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	3,482	4,541	5,841	10,215	12,924	3
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	2,999	3,747	4,604	7,173	8,610	4
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	2,757	3,365	4,032	5,893	6,869	5
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	2,612	3,143	3,707	5,208	5,959	6
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,517	2,998	3,499	4,785	5,408	7
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,449	2,896	3,355	4,501	5,041	8
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,398	2,821	3,250	4,297	4,781	9
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,359	2,764	3,169	4,144	4,587	10
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,328	2,718	3,106	4,025	4,437	11
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,303	2,681	3,055	3,930	4,318	12
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,282	2,650	3,012	3,852	4,221	13
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,264	2,624	2,977	3,787	4,140	14
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,249	2,602	2,947	3,733	4,073	15
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,235	2,583	2,921	3,686	4,015	16
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,224	2,567	2,898	3,646	3,965	17
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,214	2,552	2,878	3,610	3,922	18
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,205	2,539	2,861	3,579	3,883	19
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,197	2,528	2,845	3,552	3,850	20
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,189	2,518	2,831	3,527	3,819	21
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,183	2,508	2,819	3,505	3,792	22
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,177	2,500	2,807	3,485	3,768	23
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,172	2,492	2,797	3,467	3,745	24
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,167	2,485	2,787	3,450	3,725	25
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,162	2,479	2,779	3,435	3,707	26
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,158	2,473	2,771	3,421	3,690	27
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,154	2,467	2,763	3,408	3,674	28
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,150	2,462	2,756	3,396	3,659	29
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,147	2,457	2,750	3,385	3,646	30
35	0,127	0,255	0,388	0,529	0,682	0,852	1,052	1,306	1,690	2,030	2,133	2,438	2,724	3,340	3,591	35
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,123	2,423	2,704	3,307	3,551	40
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,109	2,403	2,678	3,261	3,496	50
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,099	2,390	2,660	3,232	3,460	60
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,076	2,358	2,617	3,160	3,373	120
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,054	2,326	2,576	3,090	3,291	∞
	45%	40%	35%	30%	25%	20%	15%	10%	5%	2,5%	2%	1%	0,5%	0,1%	0,05%	

Tabela da Distribuição Normal-Padrão



Parte inteira e primeira decimal de z	Segunda decimal de z										Parte inteira e primeira decimal de z
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586	**0,0**
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535	**0,1**
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409	**0,2**
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173	**0,3**
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793	**0,4**
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240	**0,5**
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490	**0,6**
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524	**0,7**
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327	**0,8**
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891	**0,9**
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214	**1,0**
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298	**1,1**
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147	**1,2**
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774	**1,3**
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189	**1,4**
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408	**1,5**
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449	**1,6**
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327	**1,7**
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062	**1,8**
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670	**1,9**
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169	**2,0**
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574	**2,1**
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899	**2,2**
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158	**2,3**
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361	**2,4**
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520	**2,5**
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643	**2,6**
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736	**2,7**
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807	**2,8**
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861	**2,9**
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900	**3,0**
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929	**3,1**
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950	**3,2**
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965	**3,3**
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976	**3,4**
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983	**3,5**
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989	**3,6**
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992	**3,7**
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995	**3,8**
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997	**3,9**
4,0	0,49997	0,49997	0,49997	0,49997	0,49997	0,49997	0,49998	0,49998	0,49998	0,49998	**4,0**
4,5	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	**4,5**

ANEXO IV – REGRAS DE ARREDONDAMENTO

O arredondamento pode ser feito de diversas maneiras, porém há norma nacional (ABNT NBR 5891:1977) e internacional (ISO 31-0:1992, Anexo B). O arredondamento, conforme essas normas, deve ser feito segundo o seguinte critério:

Se o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é menor que 5, o algarismo da posição para a qual será feito o arredondamento fica inalterado.

Exemplos:

58,43 arredondado a 1 decimal passa a ser 58,4 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 3); 234,9876432 arredondado a 4 decimais passa a ser 234,9876 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 4); 432,391 arredondado a 2 decimais passa a ser 432,39 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 1); 123,6702 arredondado a 3 decimais passa a ser 123,670 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 2).

Se o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é maior que 5 ou, sendo 5, há pelo menos um algarismo subsequente diferente de zero, o algarismo da posição para a qual será feito o arredondamento deve ser aumentado de uma unidade.

Exemplos:

58,46 arredondado a 1 decimal passa a ser 58,5 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 6); 234,9876732 arredondado a 4 decimais passa a ser 234,9877 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 7); 432,36512 arredondado a 2 decimais passa a ser 432,37 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 5 e este é seguido de pelo menos um algarismo diferente de zero); 123,670501 arredondado a 3 decimais passa a ser 123,671 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 5 e este é seguido de pelo menos um algarismo diferente de zero).

Se o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é igual a 5 e não há algarismos subsequentes ou, sendo igual a 5, os algarismos subsequentes são constituídos de zeros sem nenhum algarismo diferente de zero, o arredondamento deve ser feito para o número par mais próximo. Em outras palavras, se o algarismo da posição para a qual deve ser feito o arredondamento é par, este será mantido e se for ímpar a ele deve ser somada uma unidade.

Exemplos:

123,465 arredondado a 2 decimais passa a ser 123,46 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 5, sem nenhum algarismo subsequente e o algarismo da posição para a qual deve ser feito o arredondamento é par); 123,425000 arredondado a 2 decimais passa a ser 123,42 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 5 seguido de zeros, sem nenhum algarismo subsequente diferente de zero e o algarismo da posição para a qual deve ser feito o arredondamento é par); 123,4915 arredondado a 3 decimais passa a ser 123,492 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 5, sem nenhum algarismo subsequente e o algarismo da posição para a qual deve ser feito o arredondamento é ímpar, sendo a ele somada uma unidade); 123,435 000 arredondado a 2 decimais passa a ser 123,44 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 5 seguido de zeros, sem nenhum algarismo subsequente diferente de zero e o algarismo da posição para a qual deve ser feito o arredondamento é ímpar, sendo a ele somada uma unidade); 129,5000 arredondado a inteiro passa a ser 130 (o algarismo imediatamente à direita da posição para a qual será feito o arredondamento é 5 seguido de zeros, sem nenhum algarismo subsequente diferente de zero e o algarismo da posição para a qual deve ser feito o arredondamento é ímpar, sendo a ele somada uma unidade).