UFSJ
UNIVERSIDADE FEDERAL
DE SÃO JOÃO DEL-REI

Estatística & Probabilidade

Ben Dêivide

Sumário

1

Conceitos básicos da estatística

A Estatística é uma Ciência que fornece métodos para coleta, organização, descrição, análise e interpretação de dados (observacionais ou experimentais) e para a utilização dos mesmos na tomada de decisões.

A utilização de técnicas, destinadas à análise de situações complexas ou não, tem aumentado e faz parte do nosso cotidiano. O que tem levado a essa qualificação de nossas vidas no dia a dia?

Um fator importante é a popularização dos computadores. No passado, tratar uma grande massa de números era uma tarefa custosa e cansativa, que exigia horas de trabalho tedioso. Recentemente, no entanto, grandes quantidades de informações podem ser analisadas rapidamente com um computador pessoal e programas adequados. Desta forma o computador contribui, positivamente, na difusão e uso de métodos estatísticos. Por outro lado, o computador possibilita uma automação que pode levar um indivíduo sem preparo específico a utilizar técnicas inadequadas para resolver um dado problema. Assim, é necessária a compreensão dos conceitos básicos da Estatística, bem como as suposições necessárias para o seu uso de forma criteriosa.

A grosso modo podemos dividir a Estatística em três áreas:

- Estatística Descritiva;
- Probabilidade;
- Inferência Estatística.

Definimos a Estatística descritiva como,

Definição 1.1: Estatística Descritiva

Um conjunto de técnicas destinadas a descrever e resumir os dados, a fim de podermos entender características de interesse da população.

Em geral utilizamos a Estatística Descritiva na etapa inicial da análise quando tentamos entender as informações contidas nos dados. Uma forma de entender essas características seria observar os dados coletados, e daí buscando algum padrão de conhecimento. Entretanto, ao depararmos com uma grande massa de valores, percebemos imediatamente que a tarefa parece não ser simples.

A finalidade da Estatística Descritiva é tornar as informações, contidas nos dados, mais fáceis de entender, de relatar e discutir. Por exemplo, a média industrial Dow-Jones, a taxa de poluição, o custo de vida, o índice pluviométrico, a quilometragem média por litro de combustível, as médias de estudantes, peso médio, são exemplos de dados tratados pela Estatística Descritiva.

1.1 Inferência Estatística

Definição 1.2: Inferência Estatística

O estudo de técnicas que possibilitam a extrapolação, a um grande conjunto de dados, das informações e conclusões obtidas a partir de subconjuntos de valores, usualmente de dimensões muito menores, é chamado de inferência estatística.

Deve-se notar que se tivermos acesso a todos os elementos que desejamos estudar, não é necessário o uso das técnicas de inferência estatística; entretanto, elas são indispensáveis quando existe a impossibilidade de acesso a todo o conjunto de dados, por razões de natureza econômica, ética ou física.

1.2 Probabilidade

A **Probabilidade** pode ser pensada como a teoria matemática utilizada para estudar a incerteza oriunda de fenômenos que envolvem o acaso (fenômenos aleatórios). A decisão de um fabricante de cola de empreender uma grande campanha de propaganda visando aumentar sua participação no mercado, a decisão de parar de imunizar pessoas com menos de vinte anos contra determinada doença, a decisão de arriscar-se a atravessar uma rua no meio do quarteirão, todas utilizam a probabilidade consciente ou inconscientemente.

1.3 Fases do Trabalho Estatístico

O trabalho estatístico é um método científico, que consiste das cinco etapas básicas seguintes:

1. Coleta e crítica de dados;
2. Tratamento dos dados;
3. Apresentação dos dados;
4. Análise e interpretação dos resultados;
5. Conclusão.

Coleta e crítica dos dados: Após definirmos cuidadosamente o problema que se quer pesquisar, elabora-se um delineamento e damos início á coleta dos dados numéricos necessários à sua descrição. Obtidos os dados, eles devem ser cuidadosamente criticados, à procura de possíveis falhas e imperfeições, a fim de não incorrermos em erros grosseiros ou certo vulto, que possam influir sensivelmente nos resultados. A crítica é externa quando visa às causas dos erros por parte do informante, por distração ou má interpretação das perguntas que lhe foram feitas; é interna quando visa observar os elementos originais dos dados da coleta.

Tratamento dos dados: Nada mais é do que a soma e o processamento dos dados obtidos e a disposição mediante critérios de classificação. Pode ser manual ou eletrônica.

Apresentação dos dados: Por mais diversa que seja a finalidade que se tenha em vista, os dados devem ser apresentados sob forma adequada (tabelas e gráficos) tornando mais fácil o exame daquilo que está sendo objeto de tratamento estatístico.

Análise dos resultados: Após a apresentação dos dados devemos calcular as medidas típicas convenientes para fazermos uma análise dos resultados obtidos, através dos métodos da Estatística Indutiva ou Inferencial, e tirarmos desses resultados conclusões e previsões.

Interpretação e Conclusão: É de responsabilidade de um especialista no assunto que está sendo pesquisado, que não é necessariamente um estatístico, relatar as conclusões de maneira que sejam facilmente entendidas por quem as for usar na tomada de decisões.

OBS: A ciência Estatística é aplicável em qualquer ramo do conhecimento em que se manipulem dados, experimentais ou observacionais.

Para ilustrar as fases do trabalho estatístico, considere o seguinte:

Exemplo 1. *Para verificar o nível de desnutrição de crianças de 4 anos de idade numa determinada região do sul da Bahia, fez-se a pesagem de 500 crianças. Para tal, o trabalho estatístico poderia estar organizado no seguinte esquema:*

- Coleta: *16,3; 18,2; 14,0; 12,8; etc (em kg); (Como são 500 crianças, será possível analisar a informação?)*
- Organização e apresentação: *Tabelas e Gráficos;*
- Análise: *Qual é o peso médio?*
- Interpretação: *Porque tão baixo (ou tão alto) esse peso médio?*

1.4 Definições básicas

Algumas definições básicas:

Definição 1. ***População:*** *É um conjunto de elementos com pelo menos uma característica comum. Essa(s) característica(s) comum(s) deve(m) delimitar inequivocamente quais elementos que pertencem ou não à população. A notação usual para o número de elementos da população é "N".*

A população pode ser: Finita *(quando pode ser enumerada) ou* Infinita *(quando não pode ser enumerada).*

Definição 2. ***Amostra:*** *É um subconjunto de uma população. É necessariamente finita, pois todos os seus elementos serão examinados para efeito da realização do estudo estatístico desejado. A notação usual para o número de elementos da amostra é "n".*

Definição 3. ***Variável:*** *Característica pela qual deseja-se que a população seja descrita. Essa característica pode assumir diferentes valores de elemento para elemento. A notação usual é X, Y, Z, ou X_i, Y_i, Z_i para um particular elemento amostral (geralmente utiliza-se as últimas letras do alfabeto.*

Definição 4. ***Dados ou Observação:*** *É o valor que assume a variável para um elemento em particular.*

Definição 5. *As variáveis podem ser:*

(a) ***Variável qualitativa:*** *são aquelas variáveis que correspondem a atributos ou categorias, podem ser:*

(a1) ***v. qualitativa Nominal:*** *Quando os atributos (ou categorias) não são possíveis de ordenação. Ex: sexo dos alunos de uma turma.*

(a2) ***v. qualitativa Ordinal:*** *Quando os atributos (ou categorias) são possíveis de ordenação. Ex: Nível de conhecimento de inglês dos alunos de uma turma. (baixo, médio, alto).*

(b) ***Variável quantitativa:*** *são aquelas variáveis que correspondem a números resultantes de contagem ou medidas. Podem ser:*

(b1) ***v. quantitativa Discreta:*** *São próprias de dados de contagem, isto é, estão definidas em um conjunto enumerável. Ex: Número de dentes com cárie.*

(b2) ***v. quantitativa Contínua:*** *São aquelas em que as realizações resultam de uma medida (uma mensuração) que podem assumir qualquer valor real entre dois extremos. (em um intervalo real). Ex: peso e altura dos alunos de uma turma.*

Pensando sobre uma variável!

- *Uma variável originalmente quantitativa pode ser coletada de forma qualitativa. Por exemplo, a variável idade, medida em anos completos, é quantitativa (contínua); mas, se for informada apenas a faixa etária (0 a 5 anos, 6 a 10 anos, etc...), é qualitativa (ordinal). Outro exemplo é o peso dos lutadores de boxe, uma variável quantitativa (contínua) se trabalhamos com o valor obtido na balança, mas qualitativa (ordinal) se o classificarmos nas categorias do boxe (peso-pena, peso-leve, peso-pesado, etc.);*
- *Outro ponto importante é que nem sempre uma variável representada por números é quantitativa. O número do telefone de uma pessoa, o número da casa, o número de sua identidade. Às vezes o sexo do indivíduo é registrado na planilha de dados como 1 se macho e 2 se fêmea, por exemplo. Isto não significa que a variável sexo passou a ser quantitativa!*
- *Dependendo da limitação do equipamento em que é fornecido o valor da variável, podemos tornar uma variável contínua em discreta.*

1.5 Técnicas de Somatório

Apresentaremos a seguir algumas das principais propriedades que envolvem somatório devida sua vasta aplicação em Estatística.

O símbolo X_i (Leia-se X índice "i") representa qualquer um dos n valores $x_1, x_2, \ldots, x_n$, assumidos por uma variável X. A letra "i" é o"índice" da variável X.

A letra grega sigma maiúsculo (Σ) indica a soma de determinados valores. O símbolo utilizado para representar a soma de todos os valores x_i desde x_1 até x_n é o "somatório" dos valores x_i com "i" variando de 1 até n, representado por: $\sum_{i=1}^{n} X_i$.

Teorema 1. *Considere a, b e k constantes e X e Y variáveis. Então as seguintes propriedades envolvendo somatório são válidas:*

(i) $\sum_{i=1}^{n} aX_i = a \sum_{i=1}^{n} X_i$

Demonstração. $\sum_{i=1}^{n} aX_i = aX_1 + aX_2 + \ldots + aX_n = a\left(X_1 + \ldots + X_n\right) = a \sum_{i=1}^{n} X_i$ □

(ii) $\sum_{i=1}^{n} X_i Y_i \neq \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i$

Demonstração. Basta observar que $\sum\limits_{i=1}^{n} X_iY_i = X_1Y_1 + X_2Y_2 + \;...\; + \; X_nY_n$ é diferente de $\sum\limits_{i=1}^{n} X_i \sum\limits_{i=1}^{n} Y_i = (X_1 + \ldots + X_n)(Y_1 + \ldots + Y_n)$, uma vez que irão aparecer termos com índices diferentes. □

(iii) $\sum\limits_{i=1}^{n} (aX_i + bY_i) = a\sum\limits_{i=1}^{n} X_i + b\sum\limits_{i=1}^{n} Y_i$.

Demonstração.

$$\begin{aligned}
\sum_{i=1}^{n} (aX_i + bY_i) &= aX_1 + bY_1 + aX_2 + bY_2 + \ldots + aX_n + bY_n \\
&= aX_1 + aX_2 + aX_n + bY_2 + bY_1 + \ldots + bY_n \\
&= a(X_1 + \ldots + X_n) + b(Y_1 + \ldots + Y_n) \\
&= a\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} Y_i
\end{aligned}$$

□

(iv) $\sum\limits_{i=1}^{n} k = nk$

Demonstração. $\sum\limits_{i=1}^{n} k = \underbrace{k + k + \ldots + k}_{n\,vezes} = nk$ □

(v) $\sum\limits_{i=1}^{n} \left(X_i - \bar{X}\right) = 0$, *em que* $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$.

Demonstração.

$$\begin{aligned}
\sum_{i=1}^{n} \left(X_i - \bar{X}\right) &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X} = \sum_{i=1}^{n} X_i - n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i = 0
\end{aligned}$$

□

(vi) $\sum_{i=1}^{n} X_i^2 \neq \left(\sum_{i=1}^{n} X_i\right)^2$

Outra forma de apresentarmos a notação de somatório de variáveis é por meio de tabela de dupla entrada. Po

1.6 Normas de Arredondamento

A seguir algumas normas úteis para arredondamento segundo a norma ABNT NBR 5891 de Dezembro de 1977.

Esta norma tem por objetivo estabelecer as regras de arredondamento na numeração decimal. São elas:

1. Quando o algarismo imediatamente seguinte ao último algarismo a ser conservado for inferior a 5, o último algarismo a ser conservado permanecerá sem modificação. Exemplo:
   - 1,3333 arredondado à primeira decimal tornar-se-á 1,3.
2. Quando o algarismo imediatamente seguinte ao último algarismo a ser conservado for superior a 5, ou, sendo 5, for seguido de no mínimo um algarismo diferente de zero, o último algarismo a ser conservado deverá ser aumentado de uma unidade. Exemplo:
   - 1,666 6 arredondado à primeira decimal tornar-se-á: 1,7.
   - 4,850 5 arredondados à primeira decimal tornar-se-ão : 4,9.
3. Quando o algarismo imediatamente seguinte ao último algarismo a ser conservado for 5 seguido de zeros, dever-se-á arredondar o algarismo a ser conservado para o algarismo par mais próximo. Conseqüentemente, o último a ser retirado, se for ímpar, aumentará uma unidade. Exemplo:
   - 4,550 0 arredondados à primeira decimal tornar-se-ão: 4,6.
4. Quando o algarismo imediatamente seguinte ao último a ser conservado for 5 seguido de zeros, se for par o algarismo a ser conservado, ele permanecerá sem modificação. Exemplo:
   - 4,850 0 arredondados à primeira decimal tornar-se-ão: 4,8.

OBS.: O arredondamento deve ser feito na resposta final, **NUNCA** nas contas intermediárias.

1.7 Lista de Exercícios

Exercício proposto 1.7.1

Classifique os seguintes tipos de variáveis:

a) Velocidade do vento;

b) Tipos de títulos oferecidos por uma universidade;

c) Nível de extroversão;

d) Marcas de carros;

e) Times de futebol;

f) Números de peças de xadrez capturadas em um jogo;

g) Peso de pandas gigantes;

h) Número de pinturas expostas em galerias de arte.

Exercício proposto 1.7.2

Seja a seguinte tabela:

	Localidade (j)		
Nível de Instrução (i)	1	2	3
1	6	14	18
2	11	14	13
3	23	15	6
4	7	13	10

Calcule:

a) $\sum_{i=1}^{4}\sum_{j=1}^{3} X_{ij}^2$;

b) $\sum_{j=1}^{3} X_{ij}^2$, para $i = 1, 2, 3, 4$;

c) $\sum_{i=1}^{4} X_{ij}^2$, para $j = 1, 2, 3$.

Exercício proposto 1.7.3

Seja a média aritmética $\bar{X} = \frac{\sum_{j=1}^{n} X_j}{n}$ e a variância $S^2 = \frac{1}{n-1}\left[\sum_{j=1}^{n} X_j^2 - \frac{\left(\sum_{j=1}^{n} X_j\right)^2}{n}\right]$. Dado um conjunto de dados $X = \{2, 3, 5, 6, 1, 8\}$, calcule a média e a variância.

Exercício proposto 1.7.4

Sejam as amostras de tamanho n = 5 dadas por:

$$X = \{2, 4, 4, 3, 2\},$$
$$Y = \{1, 2, 3, 6, 7\}.$$

Obtenha:

a) $\sum_{i=1}^{4} X_i$;

b) $\sum_{i=1}^{5} 4 \times X_i^2$;

c) $\sum_{i=2}^{n} X_i$;

d) $\sum_{i=1}^{n} X_i \times Y_i$;

e) $\sum_{i=1}^{n}(3X_i + 2Y_i)$;

f) $\sum_{i=1}^{n} X_iY_i + \sum_{i=1}^{n} Y_i^2$.

Exercício proposto 1.7.5

Sejam as amostras de tamanho n = 5 dadas por:

$$X = \{2, 4, 4, 3, 2\},$$
$$Y = \{1, 2, 3, 6, 7\}.$$

Obtenha:

a) $\sum_{i=1}^{4} X_i$;

b) $\sum_{i=1}^{5} 4 \times X_i^2$;

c) $\sum_{i=2}^{n} X_i$;

d) $\sum_{i=1}^{n} X_i \times Y_i$;

e) $\sum_{i=1}^{n}(3X_i + 2Y_i)$;

f) $\sum_{i=1}^{n} X_i Y_i + \sum_{i=1}^{n} Y_i^2$.

Exercício proposto 1.7.6

Sejam as amostras de tamanho n = 5 dadas por:

$$X = \{2, 4, 4, 3, 2\},$$
$$Y = \{1, 2, 3, 6, 7\}.$$

Obtenha:

a) $\sum_{i=1}^{4} X_i$;

b) $\sum_{i=1}^{5} 4 \times X_i^2$;

c) $\sum_{i=2}^{n} X_i$;

d) $\sum_{i=1}^{n} X_i \times Y_i$;

e) $\sum_{i=1}^{n}(3X_i + 2Y_i)$;

f) $\sum_{i=1}^{n} X_i Y_i + \sum_{i=1}^{n} Y_i^2$.

Exercício proposto 1.7.7

Complete a tabela:

X	Y	X^2	Y^2	$X - Y$	XY
10	3				
15	4				
12	1				
9	1				
10	1				

Calcule:

a) $\sum_{i=1}^{n} X_i$;

b) $\sum_{i=1}^{n} Y_i$;

c) $\sum_{i=1}^{n} X_i^2$;

d) $\sum_{i=1}^{n} Y_i^2$;

e) $\sum_{i=1}^{n}(Y_i X_i)$;

f) $(\sum_{i=1}^{n} Y_i)^2$;

g) $(\sum_{i=1}^{n} X_i)^2$;

h) $\sum_{i=1}^{n}(X_i - \frac{\sum_{i=1}^{n} X_i}{n})^2$;

i) $\sum_{i=1}^{n}(Y_i - \frac{\sum_{i=1}^{n} Y_i}{n})^2$;

j) $\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}$;

k) $\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$;

l) A que conclusão se pode chegar sobre os itens (h) e (j), bem como (i) e (k).

1.8 Organização e apresentação dos dados, e distribuição de frequência

Uma das fases em um projeto de pesquisa é a coleta dos dados. É com base nos dados que encontraremos as características da população em estudo. Daí a importância dessa fase no processo de pesquisa, e será o foco dessa subseção.

Os dados coletados numa forma sem ordenação e sem nenhum tipo de arranjo sistemático são chamados **dados brutos**. Observe que é difícil ter noção da temperatura representativa dos dados, ou saber aonde os dados se concentram ou se estão muito dispersos. Não conseguimos detectar facilmente se existem valores discrepantes ou lacunas.

Tabela 1.1: Dados brutos obtidos numa amostra de 14 plantas da geração F_2 do cruzamento de uma planta de ervilha com sementes amarelas e lisas (AL) com outra de sementes verdes e rugosas (VR).

AL	AL	VL	AL	AR	VL	VR	AL	VL	AL	AL	AR	AR	AL

Os dados da Tabela 1.1 são apresentados na forma com que foram coletados, sendo as siglas AL, AR, VL e VR significando plantas com sementes amarelas lisas, amarelas rugosas, verdes lisas e verdes rugosas, respectivamente. A Tabela 2.1 mostra agora, os dados quantitativos contínuos de uma outra amostra.

Tabela 1.2: Dados brutos de grãos em g/planta obtidos numa amostra de n=20 plantas de feijão da geração F_2 dos cruzamentos das cultivares Flor de Maio e Carioca.

3,65	21,26	3,87	24,57	1,38
5,67	9,79	12,56	4,54	6,79
13,19	4,14	3,78	15,6	6,23
12,13	17,12	19,68	5,64	8,21

Essa representação de dados apresentados nas Tabelas 1.1 e 2.1 é pouco informativa. Para melhorá-las um pouco é possível ordenar os dados em uma sequência crescente ou decrescente ou agrupá-los quanto as suas categorias ou atributos. As Tabelas 1.3 e 1.4, contêm os dados das Tabelas 1.1 e 2.1, respectivamente, nessa nova organização. Os dados nessa forma de apresentação recebem a denominação de **dados elaborados** ou **dados em Rol**.

Na Tabela 1.3 são apresentados os atributos agrupados por tipos, das respectivas plantas que os possuem.

Na Tabela 1.4, encontra-se os dados de produção de grãos em g/planta ordenados de forma crescente por coluna. É interessante observar que essa representação facilita a obtenção de algumas características desses dados de imediato, quais sejam, a menor produtividade (1,38g) e a maior produtividade (24,57g).

Tabela 1.3: Dados brutos obtidos numa amostra de 14 plantas da geração F_2 do cruzamento de uma planta de ervilha com sementes amarelas e lisas (AL) com outra de sementes verdes e rugosas (VR).

AL	AL	AL	AL	AL	AL	AL	AR	AR	AR	VL	VL	VL	VR

Tabela 1.4: Dados elaborados de grãos em g/planta obtidos numa amostra de n=20 plantas de feijão da geração F_2 dos cruzamentos das cultivares Flor de Maio e Carioca.

1,38	4,14	6,23	12,13	17,12
3,65	4,54	6,79	12,56	19,68
3,78	5,64	8,21	13,19	21,26
3,87	5,67	9,79	15,6	24,57

1.8.1 Distribuição de frequência

A evidente necessidade de resumir os dados, sem perda de muita informação contida neles, ficou bastante explicitada nos exemplos precedentes apresentados. Sendo assim, numa primeira instância, para os dados qualitativos nominais, para os quantitativos discretos e contínuos, percebe-se prontamente que eles poderiam ser resumidos agrupando suas categorias e apresentando-os em tabelas.

1.8.1.1 Dados agrupados sem intervalo de classe

Esse tipo de agrupamento será feito para os dados qualitativos nominais e para os dados quantitativos discretos.

Os dados qualitativos nominais das classes fenotípicas das sementes de ervilhas estão apresentados na Tabela 1.5

Tabela 1.5: Dados brutos obtidos numa amostra de 14 plantas da geração F_2 do cruzamento de uma planta de ervilha com sementes amarelas e lisas (AL) com outra de sementes verdes e rugosas (VR).

Classe Fenotípica	Frequências (F_i)
AL	7
AR	3
VL	3
VR	1

Na Tabela 1.6, estão apresentados os dados referentes ao número de ovos danificados em uma inspeção feita em 30 embalagens de um dúzia cada, em um carregamento para o mercado municipal de Lavras/MG. Esses dados podem ser agrupados de uma forma similar aos dados da ervilha por existir apenas 6 categorias (0, 1, ..., 5). No caso de um número excessivo de categorias, algumas delas devem ser agrupadas e a frequência computada para esses subconjuntos de categorias.

Tabela 1.6: Número de ovos danificados em uma inspeção feita em 30 embalagens de um dúzia cada, em um carregamento para o mercado municipal de Lavras/MG, proveniente de uma cidade distante.

Número de ovos quebrados (x_i)	Frequências F_i
0	13
1	9
2	3
3	3
4	1
5	1

1.8.1.2 Dados agrupados com intervalo de classe

Os dados quantitativos apresentados na Tabela 1.4, verificam-se que não é possível efetuar o mesmo tipo de tratamento dispensado aos dados qualitativos e aos dados discretos. Para resolver o problema de apresentar a distribuição de dados quantitativos contínuos de uma forma resumida e manter o máximo da informação contida nela, será apresentada a distribuição de frequência para esse tipo de dados.

Nesse tipo de representação, os dados quantitativos contínuos agrupado em classe de valores, das quais as frequências e os limites são apresentados em uma tabela.

O número de classes (k) que são formados para sumariar os dados serão enfocados apenas dois métodos, tendo em vista que há várias metodologias para esse critério.

1. **Critério empírico**
   Um critério empírico para isso que tem sido muito usado é o de considerar um número de classes entre 5 e 20, em função do conhecimento do investigador sobre os dados de sua pesquisa. Um outro critério empírico usado, baseia-se numa função do tamanho amostral. Esse critério está apresentado na Tabela 1.7.

Tabela 1.7: Critério empírico para determinar o número de classes (k) em função do tamanho amostral (n).

Tamanho da amostra (n)	Número de classes (k)
Até 100	$\sqrt{n}$ (inteiro mais próximo)
Acima de 100	$5\log_{10} n$ (inteiro mais próximo)

2. **Critério de Scott**

   O critério de Scott é dado pela expressão (1.1).

$$k = 1 + \frac{A\sqrt[3]{n}}{3,49\ S}, \tag{1.1}$$

em que A é a amplitude, n o tamanho da amostra e S o desvio padrão. As estatísticas A e S são definidas nas equações (1.2) e (1.3).

$$A = X_n - X_1 \tag{1.2}$$

e,

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]}, \tag{1.3}$$

onde X_1 é a estatística do menor valor, e X_n é a estatística do maior valor.

OBS.: O valor de k deve ser tomado como o inteiro mais próximo ao valor encontrado.

Usando como exemplo os dados referentes da Tabela 1.4, calcularemos o número de classes (k), utilizando o critério de Scott. Obtendo a amplitude e o desvio padrão,

$$A = X_n - X_1 = 24,57 - 1,38 = 23,19 \text{ e } S = 6,714717.$$

Com esses resultados, aplicando a fórmula (1.1), é possível obter o número de classes, que é $k = 3,68 \approx 4$.

O próximo passo é determinar a amplitude de cada classe, que é dado por

$$c = \frac{A}{k-1}. \tag{1.4}$$

Com base nisso, o limite inferior da primeira classe (LI1) é definido a seguir.

$$LI1 = X_1 - \frac{c}{2} \tag{1.5}$$

O limite superior da primeira classe é então obtido somando-se ao limite inferior dessa classe, a amplitude da classe. O limite inferior da segunda classe é igualando ao limite superior da primeira classe. O limite superior dessa classe é obtido somando-se a amplitude da classe ao limite inferior. O processo é repetido para formar as demais classes, devendo parar quando a última classe k for formada. Os resultados da distribuição de frequência dos dados da Tabela 1.4 estão apresentados na Tabela 2.2.

Tabela 1.8: Distribuição de frequências das produções de grãos em g/planta obtidas numa amostra de $n = 20$ plantas de feijão da geração F_2 do cruzamento das cultivar Flor de Maio e Carioca.

Classes de pesos	$\overline{X}_i$	F_i	F_{ri}	$F_{pi}(\%)$
$-02,49 \vdash 05,25$	1,38	6	0,3	30
$05,25 \vdash 12,98$	9,11	8	0,4	40
$12,98 \vdash 20,71$	16,84	4	0,2	20
$20,71 \vdash 28,44$	24,57	2	0,1	10

Um outro tipo de representação dos dados contínuos é o do acúmulo das frequências para uma leitura rápida da proporção de dados que superem um determinado valor ou de quantos são inferiores a esse valor. Na Tabela 1.9 estão apresentadas as distribuições de frequências acumuladas dos dados de produtividade em g/plantas da Tabela 1.4.

1.8.2 Gráficos - Melhorar esse tópico

O gráfico também é uma forma alternativa para apresentar tais tipos de dados, possibilitando uma fácil e rápida inspeção por parte do interessado naquele estudo. A Figura **??** e **??** ilustra graficamente os dados representados na Tabela 1.5.

Tabela 1.9: Distribuição de frequências acumuladas das produções de grãos em g/planta obtidas numa amostra de $n = 20$ plantas de feijão da geração F_2 do cruzamento das cultivar Flor de Maio e Carioca.

Limites (X_i)	$FC_i(X < X_i)$	$FC_i(X > X_i)$
-2,49	0	20
5,25	6	14
12,98	14	6
20,71	18	2
28,44	20	0

Código R:

```
####################################################
               #gráfico de pizza
####################################################
dados<-c(rep("AL",7),rep(c("AR","VL"),3),"VR");dados
pie(table(dados), main="Gráfico de pizza",
    col=c("red","green","yellow","cyan"))
#Inserindo as porcentagens:
#tem que clicar no local do gráfico aonde
#deseja a informação
text(locator(1),paste(round((prop.table(table(dados))[1]),
     2)*100,"%"))
text(locator(1),paste(round((prop.table(table(dados))[2]),
     2)*100,"%"))
text(locator(1),paste(round((prop.table(table(dados))[3]),
     2)*100,"%"))
text(locator(1),paste(round((prop.table(table(dados))[4]),
     2)*100,"%"))

####################################################
                 #gráfico de colunas
####################################################
library(fBasics)
barplot(table(dados),main="Gráfico de colunas",
        xlab="Classe fenotípica",ylab="Frequência",
             col="red")

## Discrete Distribution Plot:
plot(table(rpois(100,5)), type = "h", col = "red", lwd=30,
     main="rpois(100,lambda=5)")
```

A representação gráfica dos dados apresentados na Tabela 2.2 é feita por meio do histograma e polígono de frequências, Figura ??. O histograma e o polígono de frequências são importantes para a determinação da forma da distribuição dos dados.

A representação gráfica das frequências acumuladas da Tabela 1.9, está na Figura ??.

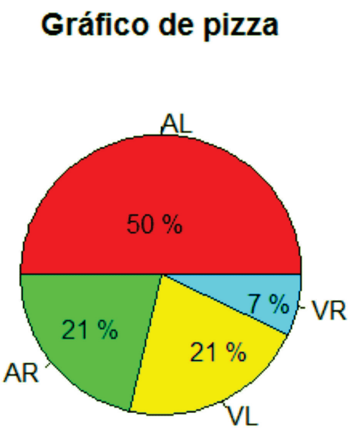


Figura 1.1: Representação do gráfico de pizza ou de setores, mostrando formas alternativas para representar as classe fenotípicas da segregação F_2 do cruzamento de plantas de ervilha de sementes amarelas e lisas com plantas de ervilha de sementes verdes e rugosas.

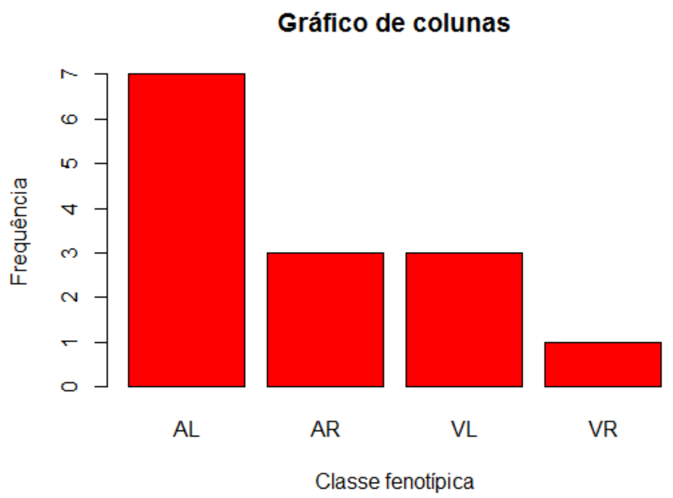


Figura 1.2: Representação do gráfico de de coluna, mostrando formas alternativas para representar as classe fenotípicas da segregação F_2 do cruzamento de plantas de ervilha de sementes amarelas e lisas com plantas de ervilha de sementes verdes e rugosas.

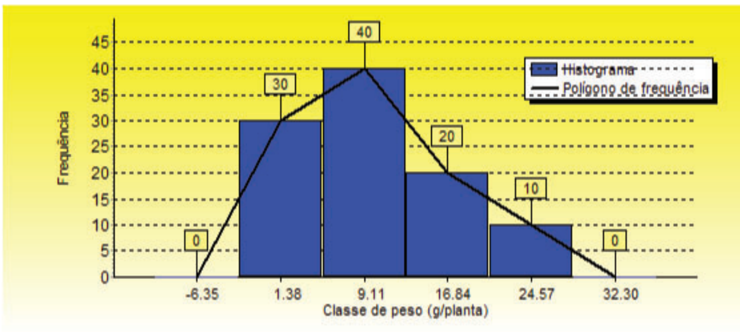


Figura 1.3: Histograma e polígono de frequências das produções de grãos em g/planta obtidas numa amostra de $n = 20$ plantas de feijão da geração F_2 do cruzamento das cultivar Flor de Maio e Carioca.

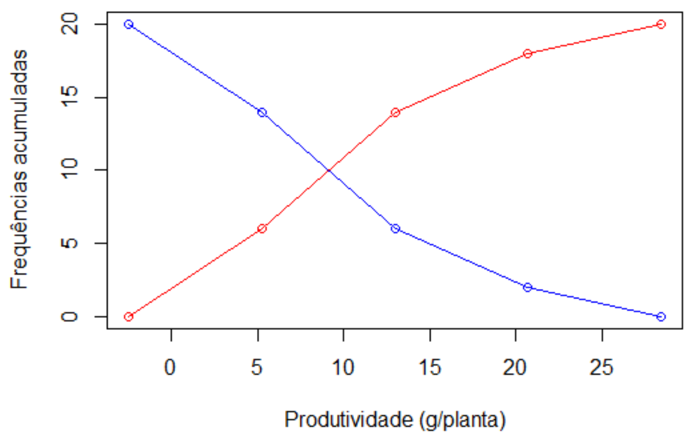


Figura 1.4: Ogivas das produções de grãos em g/planta obtidas numa amostra de $n = 20$ plantas de feijão da geração F_2 do cruzamento das cultivar Flor de Maio e Carioca.

1.9 Lista de Exercícios

Exercício proposto 1.9.1

Considere os dados de um levantamento de uma amostra de 40 famílias de um conjunto residencial, com respeito ao nível de instrução (Código: 1 - Nenhum; 2 - Fundamental; 3 - Médio)

3	3	2	2	3	1	3	3	3	2	2	1	2	2	3	2	3	3	3	3
3	3	3	2	2	3	1	3	2	3	3	2	3	1	1	1	3	3	3	3

a) Qual a natureza da variável nível de instrução do chefe da casa;

b) Organize esses dados em distribuição de frequências;

c) Calcule a frequência absoluta (f_i), frequência relativa (f_r), frequência acumulada absoluta acima ($f_{i_{ac\uparrow}}$), frequência acumulada abaixo ($f_{i_{ac\downarrow}}$), frequência percentual ($f_{\%}$), frequência acumulada percentual acumulada abaixo ($f_{\%\downarrow}$), frequência percentual acima ($f_{\%\uparrow}$);

d) Qual o grau de instrução com maior porcentagem?

Exercício proposto 1.9.2

Considere a ddistribuição de frequências do último nível de instrução completado pelo chefe de casa, numa amostra de 120 famílias, dividida segundo as localidades de três bairros diferentes de uma determinada cidade, apresentada na seguinte tabela:

	Localidade		
Nível de Instrução	Bairro 1	Bairro 2	Bairro 3
Nenhum	6	14	18
Fundamental	11	14	13
Ensino Médio	23	15	6
Total	40	43	37

a) Qual a natureza da variável nível de instrução do chefe da casa;

b) Organize esses dados em distribuição de frequências;

c) Calcule a frequência absoluta (f_i), frequência relativa (f_r), frequência acumulada absoluta acima ($f_{i_{ac\uparrow}}$), frequência acumulada abaixo ($f_{i_{ac\downarrow}}$), frequência percentual ($f_{\%}$), frequência acumulada percentual acumulada abaixo ($f_{\%\downarrow}$), frequência percentual acima ($f_{\%\uparrow}$);

d) Qual o bairro que apresentou a maior porcentagem de chefes de casa com maior grau de instrução? E o que apresentou o menor grau de instrução?

e) Apresente a representação gráfica desses dados.

Exercício proposto 1.9.3

Considere os dados de taxas de afalbetismo de uma amostra de 40 municípios brasileiros, apresentados a seguir.

57,25	76,85	92,90	89,07	75,49	65,28	94,59	71,20	71,20	82,30
72,81	66,01	90,52	87,94	58,88	45,37	81,15	94,83	94,83	81,42
54,70	67,95	69,91	95,02	77,62	91,22	64,65	85,70	85,70	81,34
59,07	68,04	73,22	95,34	83,52	64,19	64,17	95,34	95,34	84,66

a) Qual a natureza da variável taxa de alfabetismo;

b) Organize esses dados em distribuição de frequências;

c) Determine o número de classes pelo critério empírico;

d) Determine a amplitude;

e) Determine a amplitude de cada classe;

f) Detemine o ponto médio de cada classe;

g) Calcule a frequência absoluta (f_i), frequência relativa (f_r), frequência acumulada absoluta acima ($f_{i_{ac\uparrow}}$), frequência acumulada abaixo ($f_{i_{ac\downarrow}}$), frequência percentual ($f_{\%}$), frequência acumulada percentual acumulada abaixo ($f_{\%\downarrow}$), frequência percentual acima ($f_{\%\uparrow}$);

h) Qual a porcentagem de municípios com taxa de afalbetismo superior a 80?

i) Obtenha a taxa em que 20% dos municípios são inferiores a este valor.

j) Apresente a representação gráfica desses dados.

2

Medidas de posição

Pela concentração de dados de um conjunto de mensurações nas proximidades de alguns valores, verifica-se que esse valores podem ser usados para representar todos os dados. entre outras palavras, é possível afirmar que alguns valores podem ser representantes do conjunto de mensurações. Eles são denominados de medidas de posição ou medidas de tendência central.

2.1 Média Aritmética

A medida de posição mais comum, intensa e extensivamente utilizada, é a média aritmética, geralmente denominada de média. A unidade da média aritmética é a mesma de cada mensuração.

2.1.1 Média para uma população

Seja uma população de tamanho N, ou seja, X_1, X_2, ..., X_N. Então, a média aritmética populacional (μ) é dada por:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X_1 + X_2 + \ldots + X_N}{N}. \tag{2.1}$$

2.1.2 Média para uma amostra

2.1.2.1 Dados não-agrupados

Seja uma amostra de tamanho n, sendo X_1, X_2, ..., X_n. Então, a média aritmética amostral ($\overline{X}$) é dada por:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}. \tag{2.2}$$

2.1.2.2 Dados agrupados sem intervalo de classe

A média amostral para dados agrupados sem intervalo de classe, é obtida ponderando-se os valores de X_i pela sua respectiva frequência F_i, isto é,

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i F_i}{n} = \frac{X_1F_1 + X_2F_2 + \ldots + X_nF_n}{n}, \tag{2.3}$$

em que $n = \sum_{i=1}^{n} F_i$.

2.1.2.3 Dados agrupados com intervalo de classe

A média amostral para dados agrupados com intervalo de classe, é obtida ponderando-se o valor médio da classe pela sua respectiva frequência. A expressão é dada por:

$$\overline{X} = \frac{\sum_{i=1}^{k} \overline{X}_i F_i}{n} = \frac{\overline{X}_1 F_1 + \overline{X}_2 F_2 + \ldots + \overline{X}_k F_k}{n}, \tag{2.4}$$

em que $\overline{X}_i$ é o ponto médio, $n = \sum_{i=1}^{k} F_i$ e F_i é a frequência da classe i, para $i = 1, 2, \ldots, k$; k é o número de classes.

Exemplo 2. *Ilustrar o cálculo da média para os dados das Tabelas 2.1 e 2.2. Faça o cálculo da média, usando as expressões* (2.2) *e* (2.4)*:*

Tabela 2.1: Dados brutos de grãos em g/planta obtidos numa amostra de n=20 plantas de feijão da geração F_2 dos cruzamentos das cultivares Flor de Maio e Carioca.

3,65	21,26	3,87	24,57	1,38
5,67	9,79	12,56	4,54	6,79
13,19	4,14	3,78	15,6	6,23
12,13	17,12	19,68	5,64	8,21

Tabela 2.2: Distribuição de frequências das produções de grãos em g/planta obtidas numa amostra de $n = 20$ plantas de feijão da geração F_2 do cruzamento das cultivar Flor de Maio e Carioca.

Classes de pesos	$\overline{X}_i$	F_i	F_{ri}	$F_{pi}(\%)$
$-02,49 \vdash 05,25$	1,38	6	0,3	30
$05,25 \vdash 12,98$	9,11	8	0,4	40
$12,98 \vdash 20,71$	16,84	4	0,2	20
$20,71 \vdash 28,44$	24,57	2	0,1	10

Dados brutos da Tabela 2.1:

$$\overline{X} = \frac{3,65 + 5,67 + \ldots + 8,21}{20} = 9,99 \text{ g/planta.}$$

Dados agrupados da Tabela 2.2:

$$\overline{X} = \frac{1,38 \times 6 + 9,11 \times 8 + 16,84 \times 4 + 24,57 \times 2}{20} = \frac{197,66}{20} = 9,883 \text{ g/planta.}$$

A estimativa mais precisa é obviamente a primeira, uma vez que, no segundo caso, os pontos médios das classes, obtidas pela média dos limites dessas classes, foram usados para representá-las. Essa é a principal razão da diferença e é conhecida como hipótese tabular básica (HTB[1]). Apesar das diferenças encontradas, é possível utilizar o estimador de dados agrupados em distribuições de frequências na ausência dos dados brutos ou elaborados, uma vez que a perda de precisão, na maioria das situações, é considerada desprezível.

[1]HIPÓTESE TABULAR BÁSICA: todas as observações contidas numa classe são consideradas iguais ao ponto médio da classe.

2.1.3 Propriedades e caracterísitcas

A média possui algumas propriedades e características:

i. É possível verificar que $\sum\limits_{i=1}^{n} X_i = n\overline{X}$.

   Demonstração:

   Sabendo que $\overline{X} = \frac{\sum\limits_{i=1}^{n} X_i}{n}$, então,

$$n\overline{X} = n\left[\frac{\sum\limits_{i=1}^{n} X_i}{n}\right] = \sum_{i=1}^{n} X_i.$$

ii. A soma dos desvios em relação a média é igual a zero para qualquer amostra: $\sum\limits_{i=1}^{n}(X_i - \overline{X}) = 0$.

   Demonstração:

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \overline{X}) &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \overline{X}, \\
&= \sum_{i=1}^{n} X_i - n\overline{X}, \\
&= \sum_{i=1}^{n} X_i - \cancel{n}\left[\frac{\sum\limits_{i=1}^{n} X_i}{\cancel{n}}\right], \\
&= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i = 0.
\end{aligned}$$

iii. A soma de quadrados de desvios em relação a uma constante arbitrária A, qualquer, será um valor mínimo se $A = \overline{X}$.

   Demonstração:

   Fazendo:

$$D = \sum_{i=1}^{n}(X_i - A)^2.$$

   Expandindo o somatório e derivando D em relação a “A"tem-se:

$$D = \sum_{i=1}^{n}(X_i - A)^2 = \sum_{i=1}^{n}(X_i^2 - 2AX_i + A^2) = \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} 2AX_i + \sum_{i=1}^{n} A^2$$

$$\frac{dD}{dA} = -2\sum_{i=1}^{n} X_i + 2nA$$

Igualando a derivada a zero, e resolvendo em A, tem-se:

$$\frac{dD}{dA} = -2\sum_{i=1}^{n} X_i + 2nA = 0,$$

$$2nA = 2\sum_{i=1}^{n} X_i,$$

$$A = \frac{\sum\limits_{i=1}^{n} X_i}{n}.$$

Certificando se o ponto é de máximo ou de mínimo,

$$\frac{d^2D}{dAdA} = 2n > 0.$$

Como a segunda derivada é maior que zero, fica provado que o ponto é de mínimo.

iv. A soma ou subtração de uma constante (k) aos dados, altera a média de tal forma que a nova média fica adicionada ou subtraída pela constante.

Demonstração:

Sejam

$$Y_i = X_i \pm k \;\; \text{e} \;\; \overline{X} = \frac{\sum\limits_{i=1}^{n} X_i}{n},$$

então,

$$\overline{Y} = \frac{\sum\limits_{i=1}^{n} X_i \pm k}{n} = \frac{\sum\limits_{i=1}^{n} X_i}{n} \pm \frac{\sum\limits_{i=1}^{n} k}{n} = \overline{X} \pm \frac{nk}{k} = \overline{X} \pm k.$$

v. A multiplicação dos dados ou divisão por uma constante (k), aos dados, altera a média de tal forma que a nova média fica multiplicada ou dividida pela constante.

Demonstração:

Sejam

$Y_i = kX_i$, com $k \in \mathbb{R}$ e $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$,

então,

$$\overline{Y} = \frac{\sum\limits_{i=1}^{n} Y_i}{n} = \frac{\sum\limits_{i=1}^{n} (kX_i)}{n} = \frac{k\sum\limits_{i=1}^{n} X_i}{n} = k\overline{X}.$$

vi. A média é influenciada por valores extremos. A média tenderá a ser grande, se existirem alguns poucos valores que são maiores que a maioria das mensurações realizadas, ou a ser pequena, se existirem na amostra alguns poucos valores menores que a maioria das mensurações.

vii. É muito influenciada pelos valores extremos da distribuição;

viii. Localiza-se, em geral, na classe de maior frequência;

ix. Na sua determinação são considerados todos os dados da distribuição;

x. É única para um conjunto de dados;

xi. Não pode ser calculada para dados agrupados que apresentam classes extremas abertas.

Para tentar contornar a deficiência da propriedade (vi) da média, quando existirem na amostra alguns valores extremos ou existirem suspeitas de sua presença, uma média que é mais robusta a essa violação pode ser apresentada. Essa média é conhecida como **média aparada**.

Exemplo 3. *Ilustrar o cálculo da média aparada em que um conjunto de dados ($n = 20$) fictícios apresenta o maior e menos valores como suspeitos de serem atípicos quanto as suas ocorrências:*

$$1, 4, 5, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 9, 9, 10, 10, 40$$

A média aparada original é:

$$\overline{X} = \frac{1 + 4 + 5 + \ldots + 10 + 40}{20} = 8,80 \ unid.$$

Removendo as observações $X_1 = 1$ e $X_{20} = 40$, tem-se a média aparada,

$$\overline{X}_A = \frac{4 + 5 + \ldots + 10}{18} = 7,50 \ unid.$$

2.2 Mediana

A mediana é uma medida típica de tendência central, sendo definida em um conjunto de dados ordenados como o valor central, ou seja, o valor para a qual há tantas mensurações que o superam quanto são superados por ele. A unidade da mediana é a mesma de cada mensuração.

2.2.1 Dados não-agrupados e agrupados sem intervalos de classes

Para a estimação da mediana (Md), é necessário ordena os dados (dados elaborados). Nesse caso a ordenação será feita de forma crescente. A mediana, como a média, possui a mesma unidade de cada observação individual. Assim, temos

$$Md = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{se } n \text{ for ímpar;} \\ \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n+2}{2}\right)}}{2} & \text{se } n \text{ for par.} \end{cases} \tag{2.5}$$

2.2.2 Dados agrupados com intervalo de classe

Para estimar a mediana a partir dos dados arranjados em uma tabela de distribuição de frequência com intervalo de classe, é necessário definir a classe mediana e em seguida encontrar a mediana interpolando os resultados. A posição mediana é obtida acumulando-se frequências das classes 1, 2, etc., até se encontrar o valor que seja igual ou imediatamente superior a $n/2$.

A dedução do cálculo da mediana será baseado na Figura 2.1, sendo:

- LI_{Md}: Limite inferior da classe da mediana;
- LS_{Md}: Limite superior da classe da mediana;
- f_{Md}: Frequência absoluta da classe da Mediana;
- f_{ant}: Frequência acumulada (abaixo de) anterior à classe da Mediana;

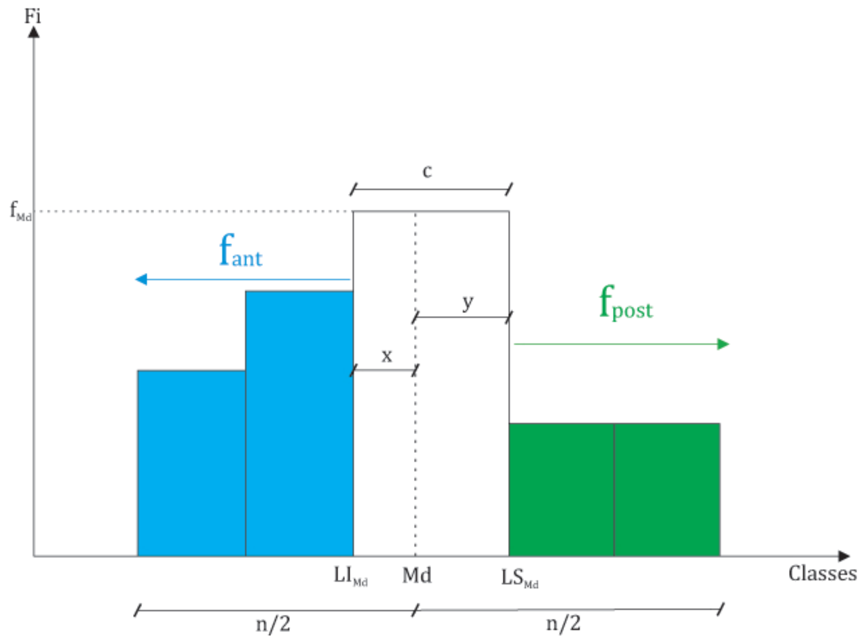


Figura 2.1: Histograma de frequência para a dedução do cálculo da mediana.

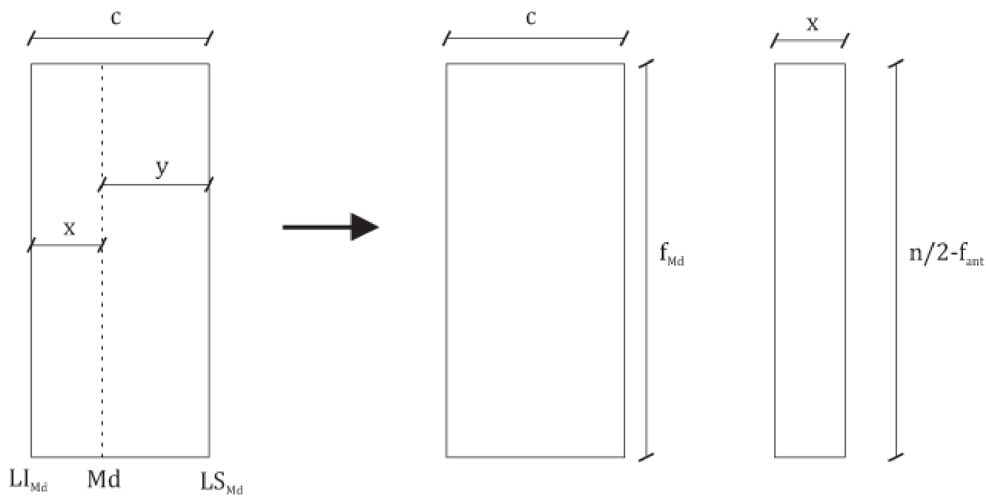
- f_{post}: Frequência acumulada (acima de) posterior à classe da Mediana;
- c: Amplitude da classe da Mediana.

Primeiro Método

Observa-se pela Figura 1, que o valor da mediana é:

$$Md = LI_{Md} + x,$$

sendo necessário encontrar o valor x. Assim, faremos uma regra de três simples:



Variação		Frequência
c	$\rightarrow$	f_{Md}
x	$\rightarrow$	$n/2 - f_{ant}$

Determinando x,

$$x = \left\{ \frac{\frac{n}{2} - f_{ant}}{f_{Md}} \right\} c.$$

Como $Md = LI_{Md} + x$, então

$$Md = LI_{Md} + \left\{ \frac{\frac{n}{2} - f_{ant}}{f_{Md}} \right\} c. \tag{2.6}$$

Segundo Método

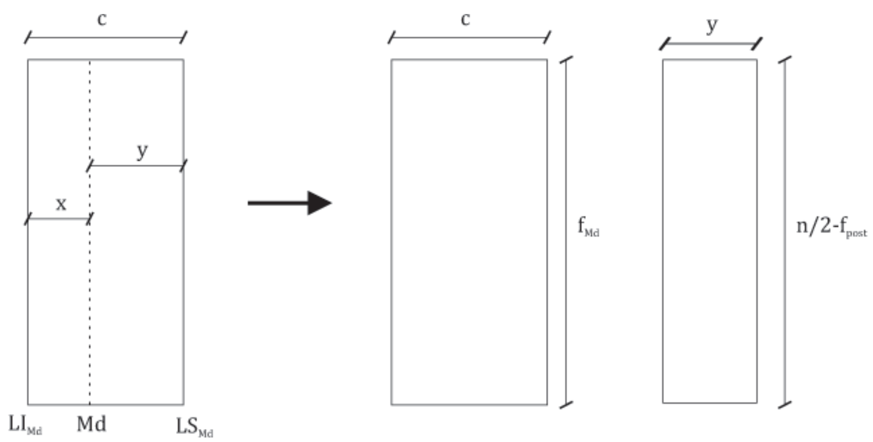
Sendo necessário encontrar o valor x. Assim, faremos uma regra de três simples:

Variação		Frequência
c	$\rightarrow$	f_{Md}
x	$\rightarrow$	$n/2 - f_{post}$

Determinando x,

$$x = \left\{ \frac{\frac{n}{2} - f_{post}}{f_{Md}} \right\} c.$$

Observa-se pela Figura 1, que o valor da mediana é:



$$Md = LS_{Md} - \left\{ \frac{\frac{n}{2} - f_{post}}{f_{Md}} \right\} c. \tag{2.7}$$

2.2.3 Propriedades e características para a mediana

A mediana possui as seguintes propriedades e características:

i. A soma dos módulos dos desvios em relação a uma constante arbitrária A, será um valor mínimo se $A = Md$.

$$h(A) = \sum_{i=1}^{n} |x_i - A|, \text{ será um ponto de mínimo se } A = Md;$$

Demonstração:

Sejam x_1, x_2, $\ldots$, x_n, os valores observados. Deseja-se mostrar que o valor de A que minimiza a função $h(A)$.

Derivando $h(A)$ em relação a A, temos:

$$\begin{aligned} \frac{dh(A)}{dA} &= \frac{d}{dA} \sum_{i=1}^{n} |x_1 - A|, \\ &= \sum_{i=1}^{n} \frac{d}{dA} |x_1 - A|. \end{aligned}$$

Usando o fato de que a derivada da função módulo é a função sinal representada por $sign()$, temos:

$$\frac{dh(A)}{dA} = -sign(x_i - A).$$

Igualando a zero a derivada, obtemos:

$$\sum_{i=1}^{n} sign(x_i - A) = 0. \tag{2.8}$$

Lembrando da função sinal:

$$\begin{array}{lll} sign(x) = -1, & \text{se} & x < 0 \\ sign(x) = 0, & \text{se} & x = 0 \\ sign(x) = 1, & \text{se} & x < 0, \end{array}$$

que tem a propriedade de que

$$\frac{d|x|}{dx} = sign(x), \quad \text{para} \quad x \neq 0.$$

Com a definição, a equação (2.8) diz que temos uma soma de números -1, 0 e 1, que é nula. Isso só ocorre quando a quantidade de -1 é igual a quantidade de 1.

O resultado $sign(x_i - A) = -1$ ocorre quando $x_i < A$, e ocorre $sign(x_i - A) = 1$ quando $x_i > A$. Dessa forma, a quantidade de valores de x_i menores que A é igual a quantidade de valores maiores que A. Isso só ocorre quando A é a mediana dos x_i.

Portanto o valor de A que minimiza $\sum_{i=1}^{n} |x_i - A|$ é a mediana.

Acrescentar a dedução feita por Daniel Furtado.

ii. A soma ou subtração de uma constante (k) aos dados, altera a mediana de tal forma que a nova mediana fica adicionada ou subtraída pela constante.

Sejam

$$Y_i = X_i \pm k,$$

então

$$Md_y = Md_x \pm k;$$

iii. A multiplicação dos dados ou divisão por uma constante (k) aos dados, altera a mediana de tal forma que a nova mediana fica multiplicada ou dividida pela constante.

Sejam

$$Y_i = kX_i, \text{ com } k \in \mathbb{R},$$

então

$$Md_y = kMd_x;$$

iv. A mediana populacional nunca é maior que um desvio padrão da média, ou seja,

$$\mu - \sigma \leq Md \leq \mu + \sigma;$$

v. A mediana não é influenciada por valores extremos, e sim, pelo número de observações;

vi. Pode ser obtida em distribuições de frequências que apresentam classes com limites indefinidos;

vii. É muito empregada em pesquisas nas quais os valores extremos têm pouca importância.

2.3 Moda

A moda é uma outra medida de tendência central, sendo definida de uma forma mais grosseira em um conjunto de dados como o valor mais frequente. Uma melhor definição poderia ser dada por aquele valor da variável em que há mais densa concentração de valores na sua proximidade. O processo de estimação da moda depende da natureza dos dados. A unidade da moda é a mesma de cada mensuração.

Se todas as classes tiverem as mesmas frequências, a distribuição não terá moda. Se duas classes ou mais apresentarem frequências mais elevadas e idênticas, então a distribuição será multimodal (bimodal, trimodal, etc.).

2.3.1 Dados qualitativos nominais ou ordinais, e dados quantitativos

para dados quantitativos nominais ou ordinais e para dados quantitativos discretos, a definição da moda, valor mais frequente da amostra.

2.3.2 Dados quantitativos contínuos

A moda de Czuber pode ser facilmente obtida pela semelhança de triângulos ABC e DCE no esquema seguinte. A moda, refere-se ao valor da abscissa correspondente ao vértice C comum aos dois triângulos. É fácil perceber que os segmentos de retas AB e DE correspondem aos valores Δ_1 e Δ_2.

Observa-se pela Figura 2.2, que o valor da moda é $Mo = LI_{Mo} + x$, bastando determinar o valor x pela semelhança de triângulos, sendo:

$$\frac{x}{c-x} = \frac{\Delta_1}{\Delta_2} \Rightarrow x = \left\{\frac{\Delta_1}{\Delta_1 + \Delta_2}\right\} c,$$

assim, a moda é determinada por:

$$Mo = LI_{Mo} + \left\{\frac{\Delta_1}{\Delta_1 + \Delta_2}\right\} c, \tag{2.9}$$

sendo:

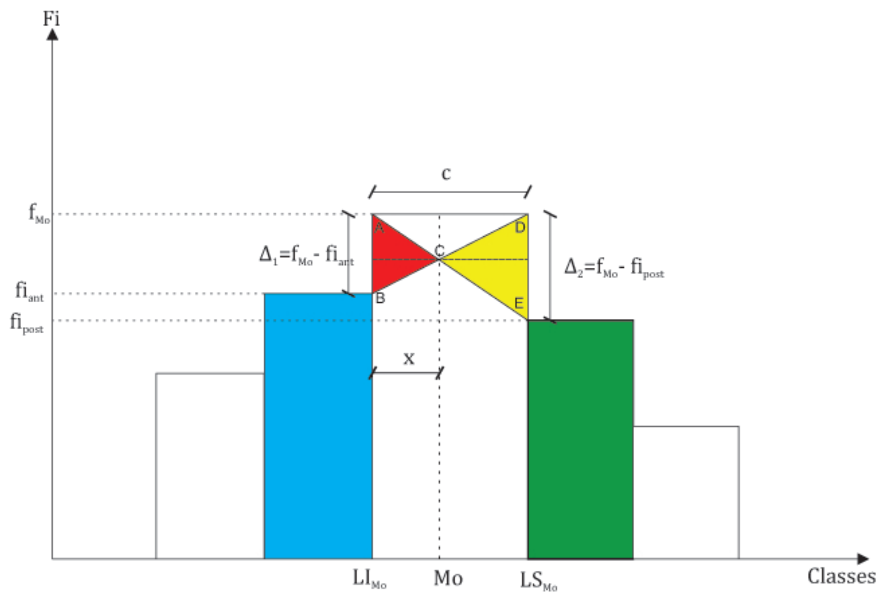


Figura 2.2: Histograma de frequência para a deduação do cálculo da moda.

Karl Pearson, observou a existência de uma relação empírica que permite calcular a moda quando são conhecidas a média ($\overline{X}$) e a mediana (Md) de uma distribuição assimétrica. Essas condições satisfazem a relação empírica,

$$Mo = 3Md - 2\overline{X}. \tag{2.10}$$

2.3.3 Propriedades e características

i. A soma ou subtração de uma constante (k) aos dados, altera a moda de tal forma que a nova moda fica adicionada ou subtraída pela constante.

   Sejam
   $$Y_i = X_i \pm k, \text{ com } k \in \mathbb{R},$$
   então
   $$Mo_y = Mo_x \pm k.$$

ii. A multiplicação dos dados ou divisão por uma constante (k), altera a moda de tal forma que a nova moda fica complicada ou dividida pela constante.

   Sejam
   $$Y_i = kX_i, \text{ com } k \in \mathbb{R},$$
   então
   $$Mo_y = kMo_x.$$

iii. Não é afetada por valores extremos, a não se que estes constituam a classe modal;

2.4 Relações empíricas entre média, mediana e moda

As relações empíricas entre a média, mediana e moda, podem ser observados na Tabela 3.3, podendo fornecer grandes informações sobre o formato da distribuição da variável.

Tabela 2.3: Relação empírica entre a média, mediana e moda.

Distribuição	Relação
Simétrica	$\overline{X} = Md = Mo$
Assimétrica à direita (assimétrica positiva)	$\overline{X} > Md > Mo$
Assimétrica à esquerda (assimétrica positiva)	$\overline{X} < Md < Mo$

Exemplo 4. *Um estudante está procurando um estágio para o próximo ano. As companhias A e B têm programas de estágios e oferecem uma remuneração por 20 horas semanais com as seguintes características (em salários mínimos):*

Companhia	A	B
Média	2,5	2,0
Mediana	1,7	1,9
Moda	1,5	1,9

Qual a companhia mais adequada?

Inicialmente vamos discutir as informações fornecidas supondo que o estudante terá seu salário “escolhido” de acordo com uma política salarial cuja tabela acima é um resumo. A companhia A tem 50% dos seus estagiários recebendo até 1,7 salários mínimos e o valor com mais chance de ocorrência é 1,5. Como a média é 2,5 devem haver alguns poucos estagiários com salário bem mais alto. A companhia B tem as três medidas bem próximas indicando uma razoável simetria entre salários altos e baixos. A opção do estudante dependerá de sua qualificação. Se ele for bem qualificado, deve preferir a companhia A pois terá maior chance de obter um dos altos salários. se tiver qualificação próxima ou abaixo dos outros estudantes, deve preferir B que parece ter uma política mais homogênea de salários.

2.5 Outras medidas

2.5.1 Midrange

A amplitude média é dada por:

$$\overline{R} = \frac{X_1 + X_n}{2}. \tag{2.11}$$

Um exemplo de seu uso é o da temperatura, média diária. Outro fato interessante, é que estamos testando essa media para um futuro teste de comparação múltipla, tal qual como utilizado por Tukey com a amplitude estudentizada.

2.5.2 Média geométrica

A média geométrica ($\overline{X}_G$), outra medida de posição, é definida como sendo a raiz n-ésima do produto dos n dados amostrais, sendo expressa por:

$$\overline{X}_G = \sqrt[n]{X_1 \times X_2 \times \ldots \times X_n}, \; X_i > 0, \; \forall i = 1, 2, \ldots, n. \tag{2.12}$$

A tomada de logaritmos pode evitar problemas computacionais de se ter que trabalhar com números de elevada magnitude. A expressão alternativa em (2.12) está apresentada em (2.13) considerando o uso do logaritmo neperiano (ln), cuja base é o neperiano e (2,71828...).

$$\overline{X}_G = \exp\left\{ \frac{\sum\limits_{i=1}^{n} \ln X_i}{n} \right\}, \; X_i > 0, \; \forall i = 1, 2, \ldots, n. \tag{2.13}$$

A média geométrica é apresentada para calcular as taxas de variações, média de razões, médias econômicas e taxas de crescimento de microrganismos.

2.5.3 Média harmônica

A média harmônica ($\overline{X}_H$) é usada para obter médias de razões e em algumas técnicas da estatística, é usado em alguns processos conhecidos como “estimação de componentes de variância".

$$\overline{X}_H = \frac{n}{\sum\limits_{i=1}^{n} \frac{1}{X_i}}, \; X_i > 0, \; \forall i = 1, 2, \ldots, n. \tag{2.14}$$

2.5.4 Média quadrática

A média quadrática usada em física é a raiz média quadrática ou média quadrática ($\overline{X}_{MQ}$), dada por:

$$\overline{X}_{MQ} = \sqrt{\frac{X_1^2 + X_2^2 + \ldots + X_n^2}{n}}. \tag{2.15}$$

2.5.5 Relação entre a média, média geométrica e média harmônica

A relação entre a média, média geométrica e média harmônica é dada por:

$$\overline{X}_H \leq \overline{X}_G \leq \overline{X}.$$

2.6 Lista de Exercícios

Exercício proposto 2.6.1

Com base nos Exercícios Propostos da subseção 1.9, apresente para todas as questões os valores das média, mediana e moda.

Exercício proposto 2.6.2

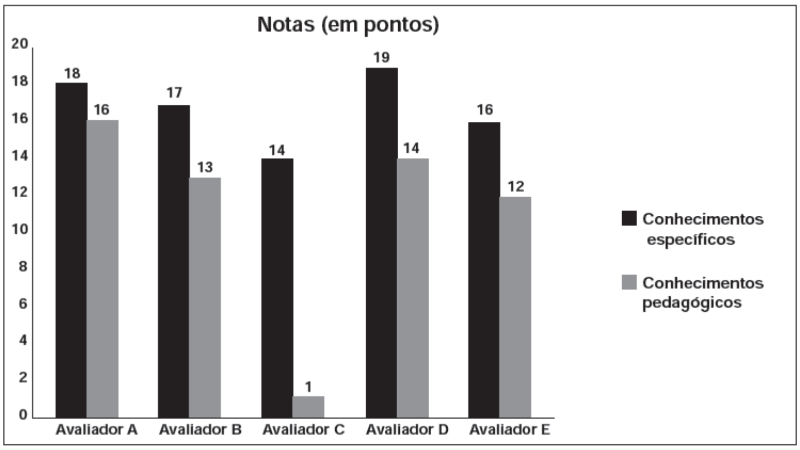
A tabela a seguir mostra a evolução da receita bruta anual nos três últimos anos de cinco microempresas (ME) que se encontram à venda.

ME	2009 (em milhares de reais)	2010 (em milhares de reais)	2011 (em milhares de reais)
Alfinetes V	200	220	240
Balas W	200	230	200
Chocolates X	250	210	215
Pizzaria Y	230	230	230
Tecelagem Z	160	210	245

Um investidor deseja comprar duas das empresas listadas na tabela. Para tal, ele calcula a média da receita bruta anual dos últimos três anos (de 2009 até 2011) e escolhe as duas empresas de maior média anual. Qual das duas empresas esse investidor deveria escolher. Justifique e apresente os cálculos.

Exercício proposto 2.6.3

As notas de um professor que participou de um processo seletivo, em que a banca avaliadora era composta por cinco membros, são apresentadas no gráfico. Sabe-se que cada membro da banca atribuiu duas notas ao professor, uma relativa aos conhecimentos específicos da área de atuação e outra, aos conhecimentos pedagógicos, e que a média final do professor foi dada pela média aritmética de todas as notas atribuídas pela banca avaliadora.



Utilizando um novo critério, essa banca avaliadora resolveu descartar a maior e a menor notas atribuídas ao professor. A nova média, em relação à média anterior, é:

a) 0,25 ponto maior.

b) 1,00 ponto maior.

c) 1,00 ponto menor.

d) 1,25 ponto maior.

e) 2,00 pontos menor.

3

Medidas de dispersão

Vimos no Capítulo 2 que as medidas de posição caracterizam o conjunto de dados, porém a informação contida nestas medidas não é suficiente. Vejamos o seguinte exemplo.

Exemplo 3.1

Considere três turmas de estatística nos cursos de engenharia. Selecionamos uma amostra de 10 alunos em cada uma entre três turmas e coletamos as notas da primeira prova da disciplina de Estatística e Probabilidade na UFSJ, de mesmo grau de dificuldade, notas de 0 a 100 pontos, campus Alto Paraopeba (CAP). Segue os dados,

Turma 1	Turma 2	Turma 3
81.00	100.00	70.00
83.00	55.00	79.00
76.00	100.00	97.00
78.00	78.00	93.00
79.00	65.00	81.00
80.00	86.00	67.00
78.00	100.00	69.00
80.00	80.00	79.00
79.00	58.00	73.00
86.00	78.00	92.00
$\bar{X} = 80$ pontos	$\bar{X} = 80$ pontos	$\bar{X} = 80$ pontos

Apesar de todas as turmas, em média, terem apresentado notas iguas, percebemos que existem diferenças entre as amostras, mais ainda, existe diferenças entre os elementos de uma dada amostra. A turma 2 apresentou três notas máximas, em contrapartida, dois alunos ficarm com nota abaixo da nota de corte de aprovação (nota 60). Na turma 3, três alunos tiveram notas superiores a 90, algo que não ocorreu na turma 1. Contudo, dois alunos tiveram notas próximas a nota de corte. Percebemos assim, que os alunos da turma 1 foram mais homogênios quanto a nota, apesar de não ter apresentado ninguém próximo da nota máxima, mas também ninguém próximo a nota de corte. Isso é o que chamamos de variabilidade entre os dados. Dessa forma, precisamos complementar a informação dos dados, além da informação contida na média (medida de tendência central). Na estatística, enfatizamos sempre a variabilidade em relação a medida de tendência central, em particular, a média.

Nesse contexto, definimos

Definição 3.1: Medidas de Dispersão ou de Variabilidade

As medidas que quantificam a variabilidade ou dispersão em um conjunto de dados, é chamada de Medida de Dispersão ou Medida de Variabilidade.

Geralmente as medidas de dispersão quantificam a distância entre os dados de uma amostra e uma medida de tendência central, em particular a média aritmética. Contudo, iniciamos por medidas mais simples, definidas a seguir.

Definição 3.2: Amplitude total

Considere um conjunto de dados de tamanho k, isto é, X_1, X_2, $\ldots$, X_k, e o ordenamos em ordem de magnitude[a] para $X_{(1)}$, $X_{(2)}$, $\ldots$, $X_{(k)}$, então definimos a amplitude total, denotada A_t, por

$$A_t = X_{(k)} - X_{(1)}, \tag{3.1}$$

sendo $X_{(k)}$ e $X_{(1)}$ o maior e o menor, respectivamente, dentre o k dados.

[a]Termo usado quando organizamos os dados de modo crescente.

Em relação a dados não agrupados, quando k for o tamanho amostral, n, diremos que $A_t = X_{(n)} - X_{(1)}$ será a amplitude total amostral. Quando $k = N$, sendo N o tamanho populacional, teremos a amplitude total populacional, $A_t = X_{(N)} - X_{(1)}$. Para os dados agrupados, quando k representar o número de grupos, isto é, teremos para A_t, o valor $X_{(k)}$ correspondente ao dado

Tabela 3.1: Dados agrupados sem intervalo de classe (Variáveis quantitativas discretas).

Grupo	Frequência
X_1	f_1
X_2	f_2
$\vdots$	$\vdots$
X_k	f_k
Total	$\sum_{i=1}^{k} f_i$

do k-ésimo grupo e o valor $X_{(1)}$ como o dado do primeiro grupo. Para dados agrupados com intervalo de classe, isto é, teremos para A_t, o valor $X_{(k)}$ correspondente ao ponto médio do k-ésimo grupo e o valor $X_{(1)}$ como o ponto médio do primeiro grupo. Usamos o ponto médio, uma vez que não sabemos quais os valores contidos em cada classe. Alguns autores convencionam assumir a amplitude total como, $At = LS_{k^{a}} - LI_{1^{a}}$. Entretanto, não temos a garantia que estes valores existam. Assim, para uma melhor proteção, utilizamos os pontos médios.

Tabela 3.2: Dados agrupados com intervalo de classe (Variáveis quantitativas contínuas).

Classe	Ponto Médio	Frequência
$LI_{1^{a}} \mid\!-\!- LS_{1^{a}}$	X_1	f_1
$LI_{2^{a}} \mid\!-\!- LS_{2^{a}}$	X_2	f_2
$\vdots$	$\vdots$	$\vdots$
$LI_{k^{a}} \mid\!-\!-\!\mid LS_{k^{a}}$	X_k	f_k
Total	-	$\sum_{i=1}^{k} f_i$

Essa medida é bastante simples, fácil de ser obtida e calculada. No entanto, ela é uma pobre medida de dispersão por não considerar todas as mensurações, e levando em conta apenas os valores extremos (mínimo e máximo). Além disso, como é improvável que uma amostra contenha os valores mínimo e máximo da população, a amplitude geralmente subestima a amplitude populacional. Deve ser considerada ainda, a influência de possíveis “outliers", que são mensurações discrepantes, no estimador da amplitude. Apesar das limitações dessa medida de dispersão, a amplitude é usada para se ter uma indicação rápida e fácil da variabilidade em diversas áreas.

3.0.0.1 Propriedades

i. A soma ou subtração de uma constante (k) aos dados, não altera o resultado da amplitude.

   Demonstração:

   Sejam
   $$Y_i = X_i \pm k, \text{ com } k \in \mathbb{R} \text{ e } A_X = X_n - X_1,$$
   então
   $$A_Y = (X_n \pm k) - (X_1 \pm k) \Rightarrow A_Y = X_n - X_n \Rightarrow A_Y = A_X.$$

ii. A multiplicação ou divisão de uma constante (k) aos dados, altera a amplitude de tal forma que a nova amplitude fica multiplicada ou dividida pela constante.

   Demonstração:

   Sejam
   $$Y_i = X_i \times k, \text{ com } k \in \mathbb{R} \text{ e } A_X = X_n - X_1,$$
   então
   $$A_Y = (X_n \times k) - (X_1 \times k) \Rightarrow A_Y = k(X_n - X_1) \Rightarrow A_Y = kA_X.$$

3.0.1 Desvio médio

Como a amplitude não considera todos os valores amostrais no seu cálculo, ela pode ser considerada deficiente. A amplitude é uma medida que não informa como a distribuição está concentrada ou arranjada em torno do seu valor central. É possível considerar conjunto de dados com a mesma amplitude, mas com diferentes estruturas de variação de seus valores intermediários. É possível expressar a variabilidade de um conjunto de dados em termos de desvio da média.

Sabe-se que a soma dos desvios em torno da média é zero, $\sum\limits_{i=1}^{n}(X_i - \overline{X}) = 0$, par toda a amostra, ou seja,

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \overline{X}) &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \overline{X}, \\
&= \sum_{i=1}^{n} X_i - n\overline{X}, \\
&= \sum_{i=1}^{n} X_i - \cancel{n} \left[\frac{\sum\limits_{i=1}^{n} X_i}{\cancel{n}} \right], \\
&= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i = 0.
\end{aligned}$$

Dessa forma essa medida se torna inútil como medida de variabilidade ou dispersão. Em virtude de o interesse residir na magnitude dos desvios em relação a média e não em saber se eles são positivos ou negativos, pode ser definida uma estatística de medida da variabilidade considerando essa idéia. Assim, o desvio médio ($S_{|\overline{X}|}$) será definido como sendo a média dos desvios absolutos em relação a média da amostra, sendo expressa pela equação (3.2). Este tipo de medida está na mesma unidade dos dados e pode ser definido também como sendo a média dos desvios absolutos em relação à mediana.

$$S_{|\overline{X}|} = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n} \tag{3.2}$$

Apesar do seu aspecto atrativo, essa medida, em razão dos valores absolutos, conduz a sérias dificuldades em problemas de inferência e é, por isso, raramente usada.

3.0.2 Variância

Uma outra forma de contornar o problema de a soma dos desvios, em relação a média aritmética, ser sempre igual a zero é usar a soma de quadrados de desvios, já que, a soma de quadrados em relação a média é um valor mínimo quando comparada com a soma de quadrados em relação a qualquer outra constante diferente de $\overline{X}$.

Fazendo:

$$D = \sum_{i=1}^{n} (X_i - A)^2.$$

Expandindo o somatório e derivando D em relação a “A"tem-se:

$$D = \sum_{i=1}^{n} (X_i - A)^2 = \sum_{i=1}^{n} (X_i^2 - 2AX_i + A^2) = \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} 2AX_i + \sum_{i=1}^{n} A^2$$

$$\frac{dD}{dA} = -2\sum_{i=1}^{n} X_i + 2nA$$

Igualando a derivada a zero, e resolvendo em A, tem-se:

$$\frac{dD}{dA} = -2\sum_{i=1}^{n} X_i + 2nA = 0,$$

$$2nA = 2\sum_{i=1}^{n} X_i,$$

$$A = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Certificando se o ponto é de máximo ou de mínimo,

$$\frac{d^2D}{dAdA} = 2n > 0.$$

Como a segunda derivada é maior que zero, fica provado que o ponto é de mínimo.

Essa propriedade torna essa soma de quadrados uma medida atraente para ser usada e se relaciona a um dos principais métodos usados na estatística, conhecido como método de mínimos quadrados. As somas de quadrados populacionais (população finita) e amostrais estão apresentados nas equações (3.3) e (3.4), respectivamente. As expressões das variâncias populacional e amostral diferem em relação ao índice, que vai de 1 a N para a população e de 1 a n para a amostra, e também em razão de os desvios serem tomados em relação a média populacional e amostral, respectivamente.

$$SQ_p = \sum_{i=1}^{N}(X_i - \mu)^2 \tag{3.3}$$

$$SQ_p = \sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{3.4}$$

A variância populacional (σ^2) é definida dividindo-se a soma de quadrados de desvios pelo tamanho da população (expressão 3.5). A variância pode ser considerada como um valor dos médio dos desvios ao quadrado, portanto, sendo também, como quadrado médio.

$$\sigma^2 = \frac{SQ_p}{N} = \frac{\sum\limits_{i=1}^{N}(X_i - \mu)^2}{N} \tag{3.5}$$

A variância amostral (S^2) poderia ser definida da mesma forma que a variância populacional, substituindo-se N por n e μ por $\overline{X}$. A razão para isso é justificada por uma propriedade importante do estimador (característica da amostra) denominada de viés, do qual, um estimador não-viesado é aquele em que a esperança do estimador é igual ao parâmetro. A soma de quadrados será dividida por $n-1$ ao invés de usar n. Isso tornaria o estimador da variância populacional não viesado. Esse estimador está apresentado em (3.6).

A quantidade $n-1$, usada como divisora, é conhecida como graus de liberdade.

$$S^2 = \frac{SQ}{n-1} = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \tag{3.6}$$

Essas expressões (3.5) e (3.6) são pouco usadas para o cálculo da variância. As razões disso são:

i. pouca precisão, uma vez que a média deve ser calculada e possivelmente arredondada e, ainda, ter que se trabalhar com arredondamentos dos próprios desvios;

ii. duas passagens pelos dados são requeridas, a primeira para calcular a média e a segunda para obter a variância.

Para contornar isso, expressões equivalentes são usadas. Para a equação (3.5), podemos deduzir,

$$\begin{aligned}
\sigma^2 &= \frac{\sum\limits_{i=1}^{N}(X_i - \mu)^2}{N}, \\
&= \frac{1}{N}\sum_{i=1}^{N}(X_i^2 - 2X_i\mu + \mu^2), \\
&= \frac{1}{N}\left[\sum_{i=1}^{N}X_i^2 - 2\sum_{i=1}^{N}X_i\frac{\sum\limits_{i=1}^{N}X_i}{N} + N\left(\frac{\sum\limits_{i=1}^{N}X_i}{N}\right)^2\right],
\end{aligned}$$

$$= \frac{1}{N}\left[\sum_{i=1}^{N} X_i^2 - 2\frac{\left(\sum\limits_{i=1}^{N} X_i\right)^2}{N} - N\frac{\left(\sum\limits_{i=1}^{N} X_i\right)^2}{N^2}\right],$$

$$= \frac{1}{N}\left[\sum_{i=1}^{N} X_i^2 - 2\frac{\left(\sum\limits_{i=1}^{N} X_i\right)^2}{N} - \frac{\left(\sum\limits_{i=1}^{N} X_i\right)^2}{N}\right],$$

$$\sigma^2 = \frac{1}{N}\left[\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum\limits_{i=1}^{N} X_i\right)^2}{N}\right]. \tag{3.7}$$

Para a equação (3.6), podemos deduzir de forma semelhante a expressão (3.7), obtendo

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum\limits_{i=1}^{n} X_i\right)^2}{n}\right]. \tag{3.8}$$

É importante definir a variância para dados agrupados com intervalo de classe, apresentado na expressão (3.9). Para dados quantitativos discretos, agrupados por categorias, esse mesmo estimador pode ser usado substituindo $\overline{X}_i$, ponto médio da classe i, por X_i, valor da categoria i.

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} F_i\overline{X}_i^2 - \frac{\left(\sum\limits_{i=1}^{k} F_i\overline{X}_i\right)^2}{n}\right]. \tag{3.9}$$

A unidade da variável não é a mesma de cada mensuração. Essa unidade não tem significado físico por estar ao quadrado. Assim, se os dados estão em kg, a variância é expressa em kg^2, se os dados estão em cm, a variância estará em cm^2, etc. Apesar desse fato, a variância é extremamente útil como medida de variabilidade, sendo igual a zero quando todas as mensurações são iguais entre si e crescendo à medida que se aumentam as diferenças (dispersão) entre os elementos do conjunto mensurado. As propriedades matemáticas e a grande quantidade de técnicas estatísticas, nas suas diferentes áreas, que usam essa medida, a tornam a mais popular e útil medida de variabilidade.

3.0.2.1 Propriedades

i. A variância de uma constante (k) é nula. Demonstração:

   Seja

$$X_1 = X_2 = \ldots = X_n = k \Rightarrow \overline{X} = k,$$

portanto,

$$\begin{aligned}S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,\\ &= \frac{1}{n-1}\sum_{i=1}^{n}(k-k)^2,\\ &= 0.\end{aligned}$$

ii. Somando ou subtraindo uma constante (k) aos dados, a variância não se altera.

Demonstração:

Sejam

$$Y = X \pm k,$$
$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

com $k \in \mathbb{R}$, então

$$\begin{aligned}S_Y^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2\\ &= \frac{1}{n-1}\sum_{i=1}^{n}[(X_i \pm k) - (\overline{X} \pm k)]^2\\ &= \frac{1}{n-1}\sum_{i=1}^{n}[X_i - \overline{X}]^2\\ &= S_X^2\end{aligned}$$

iii. Multiplicando ou dividindo uma constante (k) aos dados, altera a variância de tal forma que a nova variância fica multiplicada ou dividida por k^2.

Demonstração:

Sejam

$$Y = X \times k,$$
$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

com $k \in \mathbb{R}$, então

$$\begin{aligned}S_Y^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2\\ &= \frac{1}{n-1}\sum_{i=1}^{n}[(X_i \times k) - (\overline{X} \times k)]^2\\ &= \frac{1}{n-1}\sum_{i=1}^{n}[k(X_i - \overline{X})]^2\\ &= k^2\frac{1}{n-1}\sum_{i=1}^{n}[(X_i - \overline{X})]^2\\ &= k^2 S_X^2\end{aligned}$$

3.0.3 Desvio padrão

O desvio padrão é definido tomando-se a raiz quadrada da variância. Dessa forma o desvio padrão é expresso na mesma unidade dos dados e por essa razão possui significado físico e é preferido pelos investigadores, por ser mais fácil de interpretar. O desvio populacional (σ) é definido na equação (3.10).

$$\sigma^2 = \sqrt{\frac{1}{N}\left[\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum\limits_{i=1}^{N} X_i\right)^2}{N}\right]} \tag{3.10}$$

O estimador amostral do desvio padrão populacional σ é um estimador viesado, embora seja derivado de um estimador não-viesado, sendo obtido pela extração da raiz quadrada da variância amostral. O maior viés ocorre principalmente em pequenas amostras. O desvio padrão está apresentado na equação (3.11)

$$S^2 = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum\limits_{i=1}^{n} X_i\right)^2}{n}\right]} \tag{3.11}$$

Para dados agrupados, o desvio padrão é apresentado em (3.12). Para relatar o desvio padrão, deve-se considerar o mesmo número de casas decimais usadas para a média e para apresentar a variância, o dobro.

$$S^2 = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} \overline{X}_i^2 - \frac{\left(\sum\limits_{i=1}^{n} F_i\overline{X}_i\right)^2}{n}\right]}, \tag{3.12}$$

em que, $\overline{X}_i$ é o ponto médio da classe i.

3.0.3.1 Propriedades

i. Somando ou subtraindo uma constante (k) aos dados, não altera o desvio padrão.

   Demonstração:

   Sejam

$$Y = X \pm k,$$

$$S_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2},$$

com $k \in \mathbb{R}$, então

$$\begin{aligned}
S_Y &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \\
&= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[(X_i \pm k) - (\overline{X} \pm k)]^2} \\
&= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[X_i - \overline{X}]^2} \\
&= S_X
\end{aligned}$$

ii. Multiplicando ou dividindo uma constante (k) aos dados, altera o desvio padrão de tal forma que o novo desvio padrão fica multiplicado por k. Demonstração:

Sejam

$$Y = X \times k,$$

$$S_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2},$$

com $k \in \mathbb{R}$, então

$$\begin{aligned}
S_Y &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \\
&= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[(X_i \times k) - (\overline{X} \times k)]^2} \\
&= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[k(X_i - \overline{X})]^2} \\
&= \sqrt{k^2\frac{1}{n-1}\sum_{i=1}^{n}[(X_i - \overline{X})]^2} \\
&= kS_X
\end{aligned}$$

3.0.3.2 Teorema de Tchebichev

Quando o desvio padrão é pequeno, próximo de zero, existirá uma grande concentração de dados em torno da média. Por outro lado, se o desvio padrão for grande, os valores se concentrarão com tal intensidade. Essa idéia pode ser expressa mais formalmente pelo seguinte teorema, chamado de teorema de Tchebichev, devido ao matemático russo P.L. Tchebichev.

Para um conjunto de dados (população ou amostra) e qualquer constante $k > 1$, a proporção dos dados que podem estar a menos de k desvios padrões da média (para qualquer dos dois lados) é pelo menos: $1 - 1/k^2$, isto é,

$$P(\mu - k\sigma < X_i < \mu + k\sigma) \geq 1 - \frac{1}{k^2} \text{ ou } P(\overline{X} - k\sigma < X_i < \overline{X} + k\sigma) \geq 1 - \frac{1}{k^2}.$$

Para ilustrar o teorema de Tchebichev, por exemplo, é possível afirmar que ao menos:

$$1 - \frac{1}{2^2} = \frac{3}{4} = 75\%$$

dos valores de qualquer conjunto de dados, devem estar a menos de dois desvios padrões da média, de qualquer lado dela.

3.0.4 Coeficiente de Variação

O desvio padrão e a variância são medidas absolutas da variabilidade dos dados. Essas medidas são dependentes da grandeza, escala ou unidade de medida empregada para mensurar os dados. Conjuntos de dados com diferentes unidades de medidas não podem ter suas dispersões comparadas pela variância ou pelo desvio padrão. Mesmo para uma única unidade, se os conjuntos possuem médias de diferentes magnitudes, suas variabilidades não podem ser comparadas por essas medidas de dispersão apresentadas. Fica evidente uma medida de dispersão que não dependente desses fatores. Essa avaliação da variabilidade é conhecida por medida da variabilidade relativa de uma amostra ou população. O coeficiente de variação (CV) é usado para esse propósito. O coeficiente de variação populacional (CV_p) está apresentado na expressão (3.13).

$$CV_p = \frac{\sigma}{\mu} \times 100\% \tag{3.13}$$

O coeficiente de variação amostral (CV) está apresentado na expressão (3.14).

$$CV = \frac{S}{\overline{\overline{X}}} \times 100\% \tag{3.14}$$

O coeficiente de variação é a expressão do desvio padrão como porcentagem da média de conjunto de dados. É uma medida adimensional da variabilidade, ou seja, não possui unidade de medida. Não deve ser calculado para dados de temperatura, exceto se for utilizada a escala Kelvin. Esse tipo de dado, temperatura, pertence a uma classificação dos dados, denominada de dados intervalares. Dados intervalares possuem um intervalo definido, mas não uma escala com zero verdadeiro, ou de referência real. Outros exemplos de dados intervalares na Biologia, que nesse caso são denominados de circulares, são os dados de tempo em dias ou de tempo em anos.

Exemplo 5. *A média e o desvio padrão da produtividade de duas cultivares de milho são:* $\overline{X}_A = 4,0t/ha$ *e* $S_A = 0,8t/ha$ *para a variedade de polinização aberta A e* $\overline{X}_B = 8,0t/ha$ *e* $S_B = 1,2t/ha$ *para o híbrido simples B. Qual das cultivares possui maior uniformidade de produção?*

Se ao inspecionar as estatísticas, apresentadas você fosse induzido a responder que a variedade de polinização aberta A seria a que possui maior uniformidade e que a razão seria o menor desvio padrão apresentado por ela $(0,8t/ha)$, você teria cometido um erro. O fundamento usado aqui para comparar a variabilidade das cultivares não foi correto, uma vez que o desvio padrão é uma medida de variabilidade absoluta. Embora as unidades não sejam diferentes, as médias das amostras o são. O procedimento adequado seria o de estimar o CV para ambas as cultivares e compará-los. Coeficiente de variação são:

$CV_A = \frac{S_A}{\overline{X}_A} \times 100 = \frac{0,8}{4,0} \times 100 = 20\%$

$CV_B = \frac{S_B}{\overline{X}_B} \times 100 = \frac{1,2}{8,0} \times 100 = 15\%$

É fácil observar que o milho híbrido simples (B) é o mais uniforme, possui um menor CV que a variedade de polinização aberta A. Geneticamente, explica-se isso considerando que todas as plantas de um milho híbrido simples tem a mesma constituição genética, o que não com a variabilidade de polinização.

3.0.5 Erro padrão da média

Para definir o erro padrão da média suponha que amostras aleatórias de tamanho n são retiradas de uma população e que em cada amostra seja estimada a média. Se for computado o desvio padrão da população formada por todas as estimativas de médias obtidas, o valor encontrado é conhecido como erro padrão da média. O erro padrão da média ($\sigma_{\overline{X}}$) é dado pela razão entre o desvio padrão populacional e a raiz do tamanho da amostra (3.15).

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \tag{3.15}$$

O desvio padrão amostral é dado pela expressão (3.16). As razões da necessidades desse estimador são:

1. não se conhece, em geral, o desvio padrão populacional;
2. na maioria das situações reais não é possível retirar todas as amostras de uma população;
3. em geral, apenas uma amostra é extraída da população.

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{n}} \tag{3.16}$$

3.1 Estatísticas descritivas da distribuição

As medidas de posição e de dispersão fornecem importantes informações de locação e de variabilidade da distribuição de referência.

Nas distribuição simétricas, os mesmos valores são obtidos para a moda, para a mediana e para a média. Já as distribuições assimétricas essas medidas são diferentes e podem ser usadas para avaliar indiretamente a forma da distribuição.

3.1.1 Momentos

Os momentos populacionais centrados na média (μ_r) são definidos na equação (**??**). O coeficiente r da expressão é a ordem do momento. Assim, para $r = 1$ tem-se o momento de primeira ordem, o qual é sempre igual a zero; para $r = 2$ o momento de ordem 2, que é a variância da população; para $r = 3$ o momento de assimetria de ordem 3; para $r = 4$ o momento de curtose. É conveniente salientar que essa definição se refere à população finita.

$$\mu_r = \frac{\sum_{i=1}^{N}(X_i - \mu)^r}{N} \tag{3.17}$$

3.1.2 Medidas de assimetria

3.1.2.1 Coeficiente de assimetria

O coeficiente de assimetria populacional ($\sqrt{\beta_1}$) é uma forma padronizada do estimador do momento de assimetria ($r = 3$). Seu estimador ($\sqrt{b_1}$)é dado pela razão do momento amostral de ordem 3 pelo de ordem 2, na potência de 3/2, é apresentado na equação (3.18).

$$\sqrt{b1} = \frac{m_3}{(m_2)^{3/2}} \tag{3.18}$$

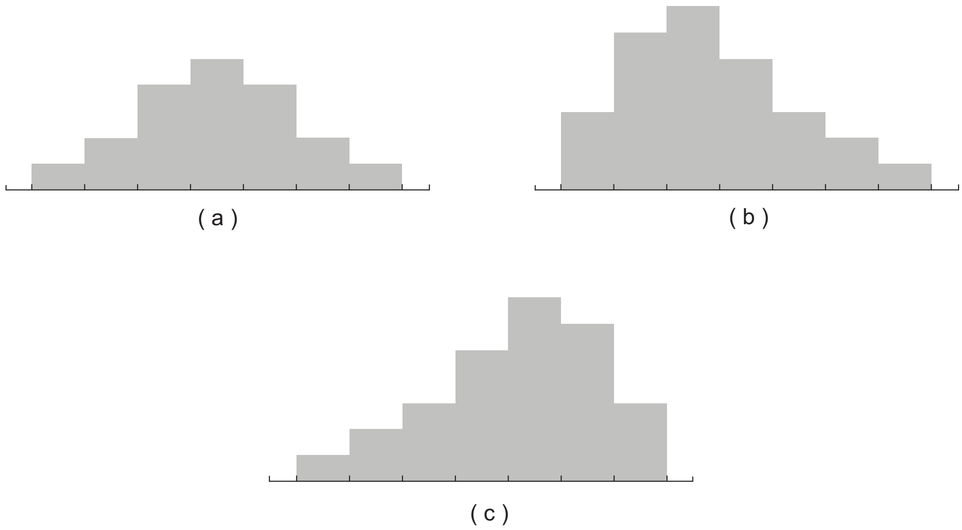


Figura 3.1: Formas das distribuição de frequência em situações reais: (a) distribuição em forma simétrica; (b) distribuição assimétrica à direita; (c) distribuição assimétrica à esquerda.

As populações cuja distribuição é simétrica apresentam valor de $\sqrt{\beta_1} = 0$ (Figura 3.1a). As distribuições assimétricas à direita (assimetria positiva) apresentam $\sqrt{\beta_1} > 0$ (Figura 3.1b), e as assimétricas à esquerda (assimetria negativa) apresentam $\sqrt{\beta_1} < 0$ (Figura 3.1c).

Estimativas amostrais diferentes de zero devem ser analisadas com cuidado para serem realizadas inferências sobre a natureza da distribuição da população que propiciou aquela estimativa. É possível extrair amostras de distribuições populacionais perfeitamente simétricas que apresentam desvios de assimetria.

3.1.2.2 Relação empírica da assimetria

As relações empíricas entre a média, mediana e moda, podem ser observados na Tabela 3.3, podendo fornecer grandes informações sobre o formato da distribuição da variável.

Tabela 3.3: Relação empírica entre a média, mediana e moda.

Distribuição	Figura	Relação
Simétrica	Figura 3.1a	$\overline{X} = Md = Mo$
Assimétrica à direita (assimétrica positiva)	Figura 3.1b	$\overline{X} > Md > Mo$
Assimétrica à esquerda (assimétrica positiva)	Figura 3.1c	$\overline{X} < Md < Mo$

3.1.2.3 Coeficiente de assimetria de Pearson

1. Uma relação empírica encontrada por Pearson sobre a assimetria é apresentada na equação (3.19).

$$AS = \frac{\overline{X} - Mo}{S}, \tag{3.19}$$

   que $\overline{X}$ é a média amostral, Mo é a moda amostral e S é o desvio padrão.

2. Outra relação empírica encontrada por Pearson sobre a assimetria é apresentada na equação (3.20).

$$AS = \frac{3(\overline{X} - Md)}{S}, \tag{3.20}$$

em que $\overline{X}$ é a média amostral, Md é a mediana amostral e S é o desvio padrão.

Se :

- $AS = 0$, diz-se que a distribuição é simétrica;
- $AS > 0$, diz-se que a distribuição é assimétrica à positiva;
- $AS < 0$, diz-se que a distribuição é assimétrica negativa.

3.1.3 Coeficiente de curtose

O grau de achatamento de uma distribuição é denominada de curtose. É fácil perceber, pela própria definição, que a curtose de uma distribuição deve ser analisada sob alguma referência. Como já se comentou anteriormente, a distribuição normal de probabilidade será considerada a distribuição de referência. Para medir a curtose, define-se o estimador ($b2$) do coeficiente de curtose (β_2) na expressão (3.21). A distribuição normal possui coeficiente de curtose igual a 3.

$$b2 = \frac{m_4}{(m_2)^2} \tag{3.21}$$

As distribuições que possuem valor de curtose igual a 3 são denominadas mesocúrticas. Aquelas que possuem $\beta_2 > 3$ são denominadas de leptocúrticas e as que possuem $\beta_2 < 3$ são as platicúrticas. As distribuições leptocúrticas são aquelas que possuem uma concentração de valores (mensurações) próxima ao valor central maior que a da distribuição normal (mesocúrtica). Nas distribuições platicúrticas, por sua vez, ocorre ao contrário, ou seja, uma menor concentração de valores em torno do centro da distribuição. A Figura 3.2 ilustra os três tipos de curvas quanto ao grau de achatamento.

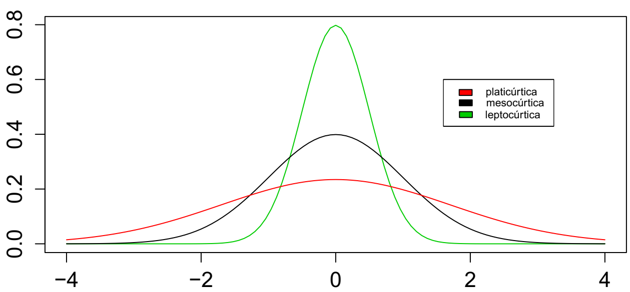


Figura 3.2: Formas das distribuição de frequência quanto ao grau de achatamento mostrando as curvas leptocúrticas, mesocúrticas e platicúrticas.

3.2 Lista de Exercícios

Exercício proposto 3.2.1

Cronometrando o tempo para várias provas de uma gincana automobilística, encontramos:

- Equipe 1: 40 provas; Tempo médio: 45 segundos; Variância: 400 segundos ao quadrado.
- Equipe 2:

Tempo:	20	40	50	80
Nº de provas:	10	15	30	5

a) Qual o coeficiente de variação relativo à equipe 1?

b) Qual a média da equipe 2?

c) Qual o desvio padrão da equipe 2?

d) Qual a equipe que apresentou resultados mais homogêneos? Justifique.

Exercício proposto 3.2.2

Os dados a seguir referem-se ao número de empresas falidas/ano observadas em $n = 85$ anos. A amostra foi obtida em Lavras, MG.

Empresas falidas	Frequências
0	36
1	19
2	16
3	7
4	4
5	2
6	1

a) Calcular: média, mediana e moda.

b) Calcular a amplitude, variância, desvio padrão, erro padrão da média e coeficiente de variação;

c) quanto a simetria qual a natureza da distribuição de frequências das empresas falidas.

Exercício proposto 3.2.3

A seguir, estão apresentadas estimativas do coeficiente de assimetria e de curtose. Classificá-las quanto à simetria e grau de achatamento da distribuição de frequência.

Coeficiente de Assimetria	Coeficiente de Curtose	Classificação da Simetria	Classificação do Grau de Achatamento
0,5	3,0		
-2,0	1,0		
2,0	2,0		
3,0	3,0		
0,0	3,0		
0,0	3,5		
-3,0	4,5		

Exercício proposto 3.2.4

A tabela a seguir apresenta algumas estatísticas das notas dos alunos de determinado curso que participaram do ENADE 2005.

Estatísticas Básicas da prova de Componente Específico por grupo de estudantes

Estatísticas	Total	Grupo	
		Ingressantes	Concluintes
População	550	326	224
Tamanho da amostra	385	214	171
Presentes	344	182	162
Média	29,0	26,8	32,2
Erro padrão da média	0,3	0,4	0,4
Desvio padrão	10,5	9,3	11,3
Nota mínima	0,0	0,0	0,0
Mediana	28,1	27,2	31,5
Nota máxima	68,2	54,4	68,2

MEC/INEP/DEAES - ENADE2005

Com base na tabela acima, pode-se afirmar que a(s):

I) menor dispersão das notas ocorre no grupo dos alunos concluintes;

II) amplitude total das notas é menor no grupo dos concluintes;

III) variância das notas é menor no grupo de ingressantes;

IV) medidas de posição na distribuição de notas são menores no grupo dos ingressantes.

São verdadeiras APENAS as afirmações:

a) I e III;

b) I e IV;

c) II e III;

d) II e IV;

e) III e IV.

4

Probabilidade

4.1 Noções iniciais sobre teoria de conjuntos

Quando desejamos compreender algo da natureza, tentamos estudá-la por meio de um processo de observação chamado experimento. Esta na realidade é uma situação que envolve incertezas. Daí podemos definir um experimento aleatório.

Definição 4.1: Experimento Aleatório

Todo experimento cujo resultado não pode ser previsto antes de sua execução, é chamado de experimento aleatório.

Podemos apresentar alguns exemplos.

Exemplo 6. *Lançar um dado equilibrado e observar o resultado obtido na face superior do dado.*

Exemplo 7. *Observar o número de chamadas telefônicas que chegam a uma central telefônica em um determinado intervalo de tempo.*

Exemplo 8. *Para a escolha ao acaso de uma lâmpada que acabou de sair do processo de fabricação, verificar o tempo de duração da lâmpada em funcionamento.*

Por mais que não seja possível prever o resultado antes de sua execução, sabemos que diversos resultados possíveis podem ocorrer. Assim, definimos,

Definição 4.2: Espaço amostral

O conjunto de todos os resultados possíveis de um experimento, denotado por Ω, é chamado de espaço amostral.

Um exemplo para a compreensão dessa definição, será apresentado a seguir.

Exemplo 4.1

Um experimento lança três moedas honestas, e deseja-se verificar a face superior dessas moedas. Sabe-se que cada moeda apresenta duas faces: cara (H) e coroa (T). Dessa forma, o espaço amostral é dado por:

$$\begin{aligned} \Omega &= \{(H,H,H),(H,H,T),(H,T,H),(H,T,T),\\ &\quad (T,H,H),(T,H,T),(T,T,H),(T,T,T)\}. \end{aligned}$$

Definição 4.3: Subconjunto

Se todo elemento do conjunto A é também elemento do conjunto B, então A é definido como um subconjunto de B, sendo representado $A \subset B$ ou $B \supset B$ (A está contido em B ou B contém A).

Essa definição pode ser aplicada também a subconjuntos e Ω, como no exemplo a seguir.

Exemplo 4.2

Sejam os conjuntos:

$$B = \{1, 2, 3, 4, 5, 6\} \text{ e } A = \{1, 2, 3\},$$

então A é um subconjunto de B, pois, os elementos que contém em A, também contém em B.

Definição 4.4: Evento

É o subconjunto do espaço amostral (Ω), representado por letras maiúsculas, A, B, ...

Exemplo 4.3

Um evento retirado do espaço amostral do Exemplo 4.1 seria $A = \{(H, H, H), (H, H, T), (H, T, T)\}$, ou seja, o evento em que dos três arremessos de moedas, tenha saído “cara"na primeira moeda.

Definição 4.5: Evento Disjunto

Dois eventos A e B são disjuntos ou mutuamente exclusivos, quando não tem elementos em comum.

Exemplo 4.4

Sejam os eventos $A = \{1, 2, 3, 4\}$ e $B = \{5, 6\}$, então $A \cap B = \varnothing$

Exemplo 4.5: Eventos coletivamente exaustivos

Se ao menos um evento ocorrer durante um dado experimento.

Definição 4.6: Conjunto equivalente

Dois conjuntos A e B são definidos equivalentes, ou iguais, se $A \subset B$ e $B \subset A$.

Exemplo 4.6

Sejam os conjuntos:

$$B = \{1, 2, 3\} \text{ e } A = \{1, 2, 3\},$$

então A é igual a B, pois $A \subset B$ e $B \subset A$.

Definição 4.7: Conjunto Vazio

Se o conjunto A não contém nenhum elemento, então A é chamado de conjunto nulo ou conjunto vazio, ou seja, $A = \varnothing$ ou $A = \{\ \}$.

Definição 4.8: Complemento

O complemento do conjunto A com respeito ao espaço Ω, denotado por $\overline{A}$, A^c, ou $\Omega - A$, é o subconjunto dos elementos de Ω exceto os elementos do conjunto A.

Exemplo 4.7

Seja o espaço amostral Ω do experimento que consiste em arremessar três moedas honestas. Diremos que H consiste na face superior da moeda ser cara, e T coroa. Assim

$$\begin{aligned}\Omega &= \{(H,H,H),(H,H,T),(H,T,H),(H,T,T),\\ &\quad (T,H,H),(T,H,T),(T,T,H),(T,T,T)\}.\end{aligned}$$

e um subconjunto de Ω, cujo evento será aparecer cara na primeira moeda, dado por

$$A = \{(H,H,H),(H,H,T),(H,T,H),(H,T,T)\}.$$

Então o complemento de A será:

$$\overline{A} = \{(T,H,H),(T,H,T),(T,T,H),(T,T,T)\}.$$

Definição 4.9: União

eja A e B, dois subconjunto quaisquer de Ω, então o conjunto todos os elementos que estão em A ou B ou em ambos, é definido conjunto união de A e B, denotado por $A \cup B$.

Exemplo 4.8

Sejam os conjuntos:

$$A = \{1,2,3\} \text{ e } B = \{3,4,5,6\},$$

então

$$A \cup B = \{1,2,3,4,5,6\}$$

Definição 4.10: Intersecção

Seja A e B, dois subconjuntos quaisquer de Ω, então o conjunto que contém todos os elementos que estão em A e B, é definido a intersecção de A e B, e escrito $A \cap B$ ou AB.

Exemplo 4.9

Sejam os conjuntos:

$$A = \{1,2,3\} \text{ e } B = \{2,3\},$$

então

$$A \cap B = \{2,3\}.$$

Definição 4.11: Conjunto diferença

Seja A e B dois conjuntos de Ω. O conjunto de todos os elementos de A que não estão em B, serão denotados por $A - B$, sendo definido por conjunto diferença.

Exemplo 4.10

Sejam os conjuntos $A = \{1, 2, 3, 4\}$ e $B = \{3, 4\}$, então $A - B = \{1, 2\}$.

4.1.1 Leis básicas da teoria de conjuntos

- Lei comutativa: $A \cup B = B \cup A$ e $A \cap B = B \cap A$.
- lei associativa: $A \cup (B \cup C) = (A \cup B) \cup C$.
- Lei distributiva: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ e $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

4.1.2 Teoremas clássicos da teoria de conjuntos

Teorema 4.1

$(A^C)^C = \overline{(\overline{A})} = A$, em outras palavras, o complemento de $\overline{A}$ é igual a A.

Prova:

Seja Ω, o espaço amostral de todos os elementos de um experimento, e A, um subconjunto de Ω. Assim, se $A^C = \Omega - A$, então

$$\overline{(\overline{A})} = \cancel{\Omega} - (\cancel{\Omega} - A) = A.$$

Teorema 4.2

Seja Ω, o espaço amostral de todos os elementos de um experimento, e A, um subconjunto de Ω, então $A\Omega = A$, $A \cup \Omega = \Omega$, $A\varnothing = \varnothing$ e $A \cup \varnothing = A$.

Prova:

1. $A\Omega = A$

$$\begin{aligned} A\Omega = A \cap (A \cup A^C) &= (A \cap A) \cup (A \cap A^C) \\ &= A \cup \varnothing \\ &= A. \end{aligned}$$

2. $A \cup \Omega = \Omega$

$$\begin{aligned} A \cup \Omega = (A \cup A) \cup \Omega &= (A \cup \Omega) \cup (A \cup \Omega) \\ &= \Omega \cup \Omega \\ &= \Omega. \end{aligned}$$

3. $A\varnothing = \varnothing$

$$\begin{aligned} A\varnothing = \varnothing = A \cap (A \cap A^C) &= (A \cup A) \cap (A \cap A^C) \\ &= A \cap \varnothing \\ &= \varnothing. \end{aligned}$$

4. $A \cup \varnothing = A$

$$\begin{aligned} A \cup \varnothing = A \cup (A \cap A^C) &= A \cap (A \cup A^C) \\ &= A \cap \Omega \\ &= A. \end{aligned}$$

4.2 Probabilidade

Probabilidade é uma função $P(.)$ que atribui valores numéricos, entre 0 e 1, aos eventos do espaço amostral satisfazendo os seguintes axiomas de Kolmogorov:

(i). $0 \leq P(A) \leq 1, \forall\ A \subset \Omega$;

(ii). $P(\Omega) = 1$;

(iii). $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} p(A_i)$, com $A_i's$ disjuntos.

Há duas formas para atribuir probabilidades aos elementos do espaço amostral:

1. A primeira delas, consiste na atribuição de probabilidades, baseando-se em características teóricas da realização do fenômeno, chamado de *probabilidade clássica ou a priori*; formalmente, se um experimento aleatório obtiver resultados multuamente exclusivos e igualmente prováveis, e se n_A desses resultados têm um atributo A, então, a probabilidade de acontecer A é a fração n_A/n. Mais ainda, Laplace define a probabilidade de um acontecimento como sendo o quociente entre o número de casos favoráveis e o número de casos possíveis, supondo todos equiprováveis. A principal limitação é que os eventos tenham que ser igualmente possíveis.

2. uma outra maneira de obter probabilidades é através das frequências de ocorrências, também conhecido como *probabilidade frequentista ou a posteriori*, em que a probabilidade de um dado acontecimento pode ser medida observando a frequência relativa do mesmo acontecimento numa sucessão numerosa de provas ou experiências, idênticas e independentes. A principal limitação é que os eventos possam repetir-se indefinidamente nas mesmas circunstâncias.

Exemplo 4.11

Um lançamento de um dado, temos o espaço amostral $\Omega = \{1, 2, 3, 4, 5, 6\}$. Se admitirmos que o dado foi construído de forma homogênea e com medidas rigorosamente simétricas, não temos nenhuma razão para privilegiar essa ou aquela face. Logo, podemos considerar $P(X = 1) = P(X = 2) = P(X = 3) = \ldots = P(X = 6) = 1/6$, fato que se enquadra na probabilidade clássica ou a priori.

Exemplo 4.12

Suponha que seja conhecida a frequência de cada possível elemento do espaço amostral. Se sorteamos aleatoriamente um elemento dessa população, a probabilidade de sortear esse ou aquele elemento será a sua respectiva frequência relativa, fato que se enquadra no tipo de probabilidade frequentista ou a posteriori.

4.2.1 Regra de adição de probabilidades

A probabilidade de ocorrência do evento A ou do evento B é igual a:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{4.1}$$

Caso os eventos A e B sejam mutuamente exclusivos, isto é, $P(A \cap B) = 0$, então:

$$P(A \cup B) = P(A) + P(B). \tag{4.2}$$

Essa regra pode ser estendida para n eventos mutuamente exclusivos: A_1, A_2, ..., A_n, isto é,

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n). \tag{4.3}$$

4.2.2 Probabilidade de um evento complementar

Se A^C é um evento complementar de A, então:

$$P(A^C) = 1 - P(A). \tag{4.4}$$

Essa situação é consequência da regra da adição para $A \subset \Omega$ na expressão (4.2), substituindo B por A^C, temos

$$\begin{aligned} P(A \cup A^C) &= P(A) + P(A^C) - P(A \cap A^C), \\ P(\Omega) &= P(A) + P(A^C) - 0, \\ 1 &= P(A) + P(A^C), \end{aligned}$$

logo segue a expressão (4.4).

As próximas Definições e Teoremas iremos apresentar com base em uma motivação:

Motivação 1. *Paulo é um jovem empreendedor e quer abrir seu próprio negócio. Ele observou que o mercado de sandálias era lucrativo. Então resolveu abrir uma fábrica de sandálias. Devido a dificuldade financeira, resolveu comprar três máquinas de sandálias usadas. As informações anteriores sobre estas máquinas dadas pelo proprietário foram:*

Máquina	***Produto***	***Total da produção***	***Produto com defeito***
M1	*Pantufas*	*50%*	*1%*
M2	*Sandálias baixas*	*40%*	*2%*
M3	*Sandálias de couro*	*10%*	*3%*

Surgiu as seguintes indagações:

- *Do total de sandálias produzidas, qual a probabilidade de Paulo produzir uma sandália com defeito?*
- *Pensando em aumentar o lucro da fábrica, Paulo pensa e substituir uma das máquinas, qual seria sua decisão?*
  - *Será que a máquina M1 que produz mais sandálias e consequentemente tem maior desgaste, deve ser trocada primeiro?*
  - *Ou será que apesar da máquina M3 ter menor produção, é a que gera mais defeito por sandália, deve ser trocada primeiro?*

Muitas vezes nos deparamos com situações em que antes da realização de algum experimento, temos alguma informação adicional. Queremos saber o quanto que essa informação pode afetar a medida de probabilidade.

Definição 6 (Probabilidade condicional)**.** *Dados dois eventos A e B definidos em $\mathcal{F}$, então a probabilidade condicional do evento A dado que ocorreu o evento B, denotado por $P(A|B)$, é definida por:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \tag{4.5}$$

para $P(B) > 0$. □

Com base no problema de Paulo, denote o evento D as sandálias produzidas com defeitos, M_1 o evento que representa as sandálias produzidas pela máquina M1, M_2 o evento que representa as sandálias produzidas pela máquina M2 e M_3 o evento que representa as sandálias produzidas pela máquina M3. Assim,

$$P(D|M_1) = 0,01,$$
$$P(D|M_2) = 0,02,$$
$$P(D|M_3) = 0,03.$$

Teorema 2 ($P(A|B)$ é uma medida de probabilidade)**.** *Sejam A e B eventos aleatórios, tal que $P(B) > 0$, e considere $Q : \mathcal{F} \to [0, 1]$ definida por*

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

que é a probabilidade condicional de A dado B. Então $(\Omega, \mathcal{F}, Q)$ é também um espaço de probabilidade. □

Demonstração. Verificar se $Q(.)$ é uma medida de probabilidade, é assumir que Q satisfaz os axiomas de Kolmogorov, isto é,

- Axioma 1:

$$\begin{aligned} Q(\Omega) &= \frac{P(\Omega \cap B)}{P(B)} \\ &= \frac{P(B)}{P(B)} \\ &= 1. \end{aligned}$$

- Axioma 2:
  Como P é uma medida de probabilidade, então

$$\forall A \in \Omega : \quad Q(A) \geq 0.$$

- Axioma 3: Seja uma sequência A_1, A_2, ..., disjunta, então

$$\begin{aligned} Q(\bigcup_{i=1}^{\infty} A_i) &= \frac{P(\bigcup_{i=1}^{\infty} A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} Q(A_i). \end{aligned}$$

  Prova concluída.

□

Teorema 3 (Regra do produto de probabilidade)**.** *Seja os eventos* $A_1, A_2, \ldots, A_n$ *em* $(\Omega, \mathcal{F}, P)$*, com* $P(\bigcap_{i=1}^{n} A_i) > 0$*, então a probabilidade do produto desses eventos é dado por*

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)\ldots P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

Demonstração. Por indução, consideremos $n = 2$. Assim, pela Definição 6, temos

$$\begin{aligned} P(A_2|A_1) &= \frac{P(A_1 \cap A_2)}{P(A_1)} \\ P(A_1 \cap A_2) &= P(A_1)P(A_2|A_1), \end{aligned}$$

pois $P(A_1) > 0$. Agora para $n = k$, generalizamos a indução,

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) \;=\; P[(A_1A_2 \ldots A_{k-1}) \cap A_k],$$

pela definição 6, temos

$$P[(A_1A_2 \ldots A_{k-1}) \cap A_k] \;=\; P[(A_1A_2 \ldots A_{k-1})]P(A_k|A_1A_2 \ldots A_{k-1})$$

podendo ser reescrito como

$$\begin{aligned} P[(A_1A_2 \ldots A_{k-1}) \cap A_k] \;=\; & \underbrace{P(A_1A_2 \ldots A_{k-2})P(A_{k-1}|A_1A_2 \ldots A_{k-2})}_{P(A_1A_2 \ldots A_{k-1})} \times \\ & \times P(A_k|A_1A_2 \ldots A_{k-1}). \end{aligned}$$

Assim, usando a indução sucessivas vezes, chegaremos a expressão

$$\begin{aligned} P(A_1A_2 \ldots \cap A_k) \;=\; & P(A_1)P(A_2|A_1)\ldots P(A_{k-1}|A_1A_2 \ldots A_{k-2}) \times \\ & \times P(A_k|A_1A_2 \ldots A_{k-1}). \end{aligned}$$

Observe que por hipótese, todos os condicionamentos da expressão do lado direito, têm probabilidades positivas, pois contém $\bigcap_{i=1}^{n} A_i$. Portanto, teorema provado. □

Antes de falarmos sobre o teorema da lei da probabilidade total, será interessante fazer a definição sobre a partição de Ω.

Definição 7 (Partição de Ω)**.** *Se a sequência* $A_1, A_2, \ldots,$ *são disjuntos dois a dois e* $\bigcup_{i=1}^{\infty} A_i = \Omega$*, então dizemos que essa sequência forma uma partição de* Ω*.*

Entretanto, para calcular a probabilidade de uma sandália está com defeito, isto é $P(D)$, apresentamos o seguinte Teorema,

Teorema 4 (Teorema da probabilidade total)**.** *Considere uma sequência de eventos* $A_1, A_2, \ldots, A_n$ *de* $\mathcal{F}$*, disjuntos, tal que* $\bigcup_{i=1}^{n} A_i = \Omega$*, e* B *um evento de* $\mathcal{F}$*, então a probabilidade de* B *é dada por:*

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i), \tag{4.6}$$

para $P(A_i) > 0$*, sendo* $i = 1, 2, \ldots, n$*.* □

Demonstração. Sabendo que $A_i \cap_{i \neq j} A_j$ e que $\bigcup_{i=1}^{n} B \cap A_i = \Omega$, então $\bigcup_{i=1}^{n} B \cap A_i = B$ e os $B \cap A_i$ são também disjuntos. Dessa forma,

$$\begin{aligned} P(B) =& P\left(\bigcup_{i=1}^{n} B \cap A_i\right) \\ =& \sum_{i=1}^{n} P(B \cap A_i) \qquad (B \cap A_i \text{ disjuntos}) \\ =& \sum_{i=1}^{n} P(B|A_i)P(A_i) \qquad (\text{Teorema 3}) \end{aligned}$$

como queríamos provar. □

Exemplo 9. *Voltando ao problema de Paulo, como* $P(M_1) = 0,50$*,* $P(M_2) = 0,40)$ *e* $P(M_3) = 0,10$*, então a probabilidade de uma sandália ter defeito é*

$$\begin{aligned} P(D) =& \sum_{i=1}^{3} P(D|M_i)P(M_i) \\ =& P(D|M_1)P(M_1) + P(D|M_2)P(M_2) + P(D|M_3)P(M_3) \\ =& 0,01 \times 0,50 + 0,02 \times 0,40 + 0,03 \times 0,10 \\ =& 0,016. \end{aligned}$$

Uma outra Definição interessante é a independência de eventos, apresentada a seguir.

Definição 8 (Independência de dois eventos)**.** *Considere o espaço amostral* Ω*. Dois eventos A e B de* Ω *são independentes se satisfaz ao menos uma das seguintes condições:*

I) $P(A \cap B) = P(A)P(B)$*;*

II) $P(A|B) = P(A)$*, para* $P(B) > 0$*;*

III) $P(B|A) = P(B)$*, para* $P(A) > 0$*.*

□

É fácil mostrar que (I) implica em (II), (II) implica em (III), e (III) implica em (I). $(i) \to (ii)$: Se $P(AB) = P(A)P(B)$, então

$$P(A|B) = \frac{P(AB)}{P(A)} = \frac{P(A)P(B)}{P(B)} = P(A), \quad \text{para } P(B) > 0;$$

$(ii) \to (iii)$: Se $P(A|B) = P(A)$, então

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = P(B), \quad \text{para } P(A) > 0;$$

$(iii) \to (i)$: Se $P(B|A) = P(B)$, então

$$P(AB) = P(B|A)P(A) = P(B)P(A), \text{para } P(A) > 0.$$

Esse teorema mostra o cunho filosófico da independência entre A e B, que consiste em dizer que na presença ou ausência do evento B, a probabilidade do evento A é a mesma, não se altera.

Exemplo 10. *Sejam os resultados possível de um dado honesto, cujo espaço amostral é* $\Omega = \{1, 2, 3, 4, 5, 6\}$. *Considere um evento que representa o conjunto dos números ímpares desse espaço amostral,* $A = \{1, 3, 5\}$, *e outro evento que consite nos múltiplos de 3,* $B = \{3, 6\}$. *A probabilidade de A é* $P(A) = 1/2$, *a probabilidade de B é* $P(B) = 1/3$, *e a probabilidade da interseção entre A e B é* $P(A \cap B) = 1/6$. *Veja que dado que o evento B ocorra, ou não ocorra* $B^c = \{1, 2, 4, 5\}$, *a probabilidade do evento A é a mesma, veja:*

$$P(A|B) = \frac{1/6}{1/3} = 1/2$$
$$P(A|B^c) = \frac{2/6}{4/6} = 1/2.$$

Que é o mesmo que entender que $P(A) \times P(B) = 1/6 = P(A \cap B)$.

A independência entre dois eventos não implica em independência coletiva. Vejamos,

Exemplo 11. *Seja um experimento cujo objetivo é verificar a face superior de um tetraedro, isto é,* $\Omega = \{1, 2, 3, 4\}$. *Sejam os eventos em* Ω, $A = \{1, 4\}$, $B = \{2, 4\}$ *e* $C = \{3, 4\}$. *Considerando o tetraedro honesto e que cada valor é equiprovável, assim* $P(A) = P(B) = P(C) = 1/2$. *Observamos que estes eventos são independentes dois a dois, isto é,* $P(A \cap B) = 1/4 = P(A)P(B)$, $P(A \cap C) = 1/4 = P(A)P(C)$ *e* $P(B \cap C) = 1/4 = P(B)P(C)$. *Porém,* $P(A \cap B \cap C) = 1/4 \neq P(A)P(B)P(C)$. *Logo, os eventos A, B e C não são independentes três a três.*

Para uma definição mais geral, temos

Definição 9 (Independência de eventos)**.** *Considere o espaço amostral* Ω. *Uma sequência de eventos* $A_1, A_2, \ldots, A_n$ *de* Ω *são independentes se e somente se:*

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j), \quad \text{para } i \neq j; \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k), \quad \text{para } i \neq j \neq k; \\ &\vdots \\ P(\cap_{i=1}^n A_i) &= \prod_{i=1}^n P(A_i). \end{aligned} \tag{4.7}$$

□

Paulo poderia indagar, os M_i e D são independentes ou dependentes? Pela Definição 8, temos que

$$P(D|M_i) \neq P(D) = 0,016 \Rightarrow D \text{ e } M_i \text{ não são independentes, para } i = 1, 2, 3.$$

A grande questão agora é qual a máquina que Paulo deveria substituir com o propósito de aumentar seu lucro na empresa. A ideia será calcular $P(M_i|D)$, isto é, dado um defeito na sandália qual a probabilidade de vindo da máquina i? A maior probabilidade será a máquina substituída. Entretanto, ainda não temos ferramenta para resolver essa resposta. Para isso, apresentamos o seguinte Teorema:

Teorema 5 (Teorema de Bayes)**.** *Considere o espaço amostral* Ω. *Considere uma sequência de eventos* $A_1, A_2, \ldots, A_n$ *de* Ω, *disjuntos, tal que* $\bigcup_{i=1}^n A_i = \Omega$, *e B um evento de* $\mathcal{F}$, *então a probabilidade de* A_k, *para* $k = 1, 2, \ldots, n$, *dado que ocorreu o evento B, denotado por* $P(A_k|B)$, *é dado por:*

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}, \qquad k = 1, 2, \ldots, n, \tag{4.8}$$

para $P(A_k) > 0$ *e* $P(A_i) > 0$, *sendo* $i = 1, 2, \ldots, n$. □

Demonstração. Para um i qualquer, temos

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

ou

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

Isto implica que

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

Pela lei da probabilidade total, a probabilidade de B pode ser dada por $P(B) = \sum_{i=1}^{\infty} P(A_i)$ $P(B|A_i)$. Portanto,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{\infty} P(A_i)P(B|A_i)},$$

prova concluída. $\square$

Tal é a sua importância, que um dos ramos de estudo da inferência estatística é baseado nesse teorema. O Teorema de Bayes fornece uma atualização do conhecimento já existente $P(A_k)$, conhecido como "*a priori*", por meio da ocorrência do evento B. Essa atualização é a probabilidade " *a posteriori*" $P(A_k|B)$.

Com esse resultado, Paulo agora pode tomar uma decisão mais plausível, isto é, dado um defeito numa determinada sandália produzida na fábrica, qual a probabilidade desta ter sido produzida em cada uma das máquinas?

$$P(M_1|D) = \frac{0,01 \times 0,50}{0,016} = 0,3125$$
$$P(M_2|D) = \frac{0,02 \times 0,40}{0,016} = 0,5000$$
$$P(M_3|D) = \frac{0,03 \times 0,10}{0,016} = 0,1875$$

A tomada de decisão será substituir a máquina M2. Poderíamos ter tomado uma decisão equivocada se não fosse o teorema de Bayes.

Devemos abrir uma discussão que ocorre muito frequente entre as Definições 10 e 8, isto é, eventos disjuntos e independência. Nas próprias definições percebemos a distinção clara entre as características. A primeira se remete a eventos (conjuntos), e a segunda é uma condição probabilística dos eventos. Contudo, em determinados problemas ainda há muita confusão ao tentar resolvê-los. Assim, apresentemos os seguintes teoremas,

Teorema 6 (Eventos disjuntos e independentes)**.** *Considere A e B, dois eventos Ω. Se $A \cap B = \emptyset$ (eventos disjuntos), então A e B são independentes apenas, se e somente se, um dos eventos tiver probabilidade 0.*

Demonstração. Considerando que o evento A tenha probabilidade 0, isto é, $P(A) = 0$, implica que $A = \emptyset$. Assim, $P(A \cap B) = P(\emptyset \cap B) = P(\emptyset) = 0$. A condição de independência entre os dois eventos existe se $P(A)P(B) = P(A \cap B)$, e isso ocorre de fato, $P(A)P(B) = P(\emptyset)P(B) = 0 \times P(B) = 0 = P(A \cap B)$, o que completa a prova. $\square$

Caso esses eventos não tenham probabilidade 0, a condição $A \cap B = \emptyset$ implica que estes são dependentes. Vejamos o seguinte exemplo para elucidar essas definições.

Exemplo 12 (Eventos disjuntos e independência de eventos). *A tabela abaixo dá a distribuição das probabilidade dos quatro tipos sanguíneos, numa certa comunidade.*

Tipo sanguíneo	*A*	*B*	*AB*	*O*
Probabilidade de ter o tipo especificado	*0,2*			
Probabilidade de não ter o tipo especificado		*0,9*	*0,95*	

Calcular a probabilidade de que:

a) um indivíduo, sorteado ao acaso nessa comunidade, tenha o tipo O;

b) dois indivíduos, sorteados ao acaso nessa comunidade, tenham tipo A e tipo B, nessa ordem;

c) um indivíduo, sorteado ao acaso nessa comunidade, não tenha o tipo B ou não tenha o tipo AB.

Vejamos que os tipos sanguíneos são multuamente exclusivos e formam a partição do espaço amostral, uma vez que não existe outro tipo sanguineo além dos informados e que não há indivíduo com dois tipos sanguíneos. Assim,

a) $P(\Omega) = P(A) + P(B) + P(AB) + P(O) \Rightarrow 1 = 0,2000 + 0,1000 + 0,0500 + P(O) \Rightarrow P(O) = 0,6500.$

b) Este ítem merece uma atenção. Como os eventos A e B são multumente exclusivos, estes não são independentes pois nenhum tem probabilidade 0. Uma vez determinada as probabilidades de especificação em indivíduos diferentes, a probabilidade de especificar o tipo sanguíneo A em um indivíduo da comunidade não interfere em nada na probabilidade de especificar o tipo sanguíneo de um outro indivíduo dessa mesma comunidade. Assim, a probablidade de especificar o tipo sanguíneo desses dois indivíduos simultaneamente é $P(A) \times P(B) = 0,2000 \times 0,1000 = 0,0200.$

c) Agora os eventos "não ter o tipo sanguíneo especificado" não implica que os eventos sejam multuamente exclusivos pelo fato dos eventos "ter o tipo sanguíneo especificado" terem sido disjuntos. Veja, o evento não ter o tipo sanguínio AB e o evento não ter o tipo sanguíneo B, pode existir indivíduos comum a estes dois eventos, por exemplo, um indivíduo do tipo sanguíneo A ou O, e a probabilidade destes não é zero, logo, os eventos não ter o tipo sanguínio AB e não ter o tipo sanguíneo B não são disjuntos. Entretanto, esses eventos são independentes, pois a probabilidade de um evento não influencia na probabilidade do outro. Assim,

$$\begin{aligned} P[(AB)^c \cup B^c] &= P[(AB)^c] + P(B^c) - P[(AB)^c \cap B^c] \\ &= 0,9000 + 0,9500 - P[(AB)^c \cap B^c]. \end{aligned} \tag{4.9}$$

Vejamos o evento $(AB)^c = A \cup B \cup O$ *e o evento* $B^c = A \cup AB \cup O$*. A interseção entre estes é* $(AB)^c \cap B^c = A \cup O$*, em que A e O são disjuntos, assim,*

$$\begin{aligned} P[(AB)^c \cap B^c] &= P(A \cup O) = P(A) + P(O) \\ &= 0,20 + 0,65 = 0,85. \end{aligned} \tag{4.10}$$

Substituindo (4.10) *em* (4.9)*, segue que*

$$P[(AB)^c \cup B^c] = 0,90 + 0,95 - 0,85 = 1.$$

Ao final, temos a tabela completada da seguinte forma:

Tipo Sanguíneo	*A*	*B*	*AB*	*O*
Prob. tipo esp.	*0,20*	*0,10*	*0,05*	*0,65*
Prob. não tipo esp	*0,80*	*0,90*	*0,95*	*0,35*

Vale a pena discutirmos sobre a independência nessa situação. Quando falamos na especificação do tipo sanguínio é fato que um mesmo elemento não pode ser especificado em dois ou mais tipos sanguíneo. Fica claro que os eventos A, AB, B e O, são disjuntos. Agora, será que a probabilidade de especificar, por exemplo, o tipo sanguíneo A, não interfere na probabilidade do tipo sanguíneo B, ou qualquer um outro tipo sanguíneo? Observe que uma vez especificado a probabilidade de um determinado tipo sanguíneo, por exemplo, tipo A, não haverá mais chances de ele ter o tipo sanguíneo B, logo a probabilidade de B ocorrer é 0. Assim, a condição de ter especificado o tipo sanguíneo A alterou a probabilidade de especificar o tipo sanguíneo B. Logo estes eventos são dependentes.

Podemos ainda expressar mais três teoremas para complementar as afirmações feitas no Exemplo 12.

Teorema 7. *Se A e B são eventos independentes no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, então*

a) A e B^c também são independentes;

b) A^c e B também são independentes;

c) A^c e B^c também são independentes.

Demonstração. Usando as seguintes equivalências:

$$P(A) = P(A \cap B) + P(A \cap B^c) \tag{4.11}$$
$$P(A^c) = P(A^c \cap B) + P(A^c \cap B^c) \tag{4.12}$$
$$P(B) = P(B \cap A) + P(B \cap A^c) \tag{4.13}$$
$$P(B^c) = P(B^c \cap A) + P(B^c \cap A^c) \tag{4.14}$$

e a condição de que $P(A \cap B) = P(A)P(B)$ (independentes), então usando (4.11) temos

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A)P(B) \quad \text{(Independência)} \\ &= P(A)[1 - P(B)] = P(A)P(B^c), \end{aligned}$$

o que prova o ítem (a). Usando (4.13) pelo mesmo raciocínio, provamos o ítem (b). Usando o resultado do ítem (a), já provado, e a condição de independência na expressão (4.14), temos

$$\begin{aligned} P(A^c \cap B^c) &= P(B^c) - P(B^c)P(A) \\ &= P(B^c)[1 - P(A)] = P(B^c)P(A^c), \end{aligned}$$

o que prova o ítem (c), concluindo assim, a prova do teorema. □

Definição 10 (Eventos Disjuntos ou multuamente exclusivos)**.** *Sejam A e B, dois eventos quaisquer de Ω, então estes são disjuntos ou mutuamente exclusivos quando não existir elementos em comum entre A e B, isto é, $A \cap B = \emptyset$.*

Teorema 8. *Sejam dois eventos A e B em Ω. Se $A \cap B = \emptyset$, então $A^c \cap B^c \neq \emptyset$, a menos que A e B sejam partição do espaço amostral.*

Demonstração. Considere $A \cap B = \emptyset$ e que

$$\begin{aligned} A \cup B &= (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) \\ &= (A \cap B^c) \cup (A^c \cap B) \\ &\neq \Omega \quad \text{(Pelo fato de A e B não serem partição do espaço amostral).} \end{aligned} \tag{4.15}$$

Usando a Lei de Morgan $A^c \cap B^c = (A \cup B)^c$, logo percebemos pela expressão (4.15) que $A^c \cap B^c \neq \emptyset$, o que completa a prova. □

4.3 Lista de Exercícios

Questão 1. Defina o que é um experimento aleatório e exemplifique.

Questão 2. Defina o que é um espaço amostral e exemplifique.

Questão 3. Defina o que é um evento aleatório e exemplifique.

Questão 4. Para cada um dos casos abaixo, escreva o espaço amostral correspondente e conte seus elementos.

a) Uma moeda é lançada duas vezes e observam-se as faces obtidas.
b) Um dado é lançado duas vezes e a ocorrência de face par ou ímpar é observado.
c) Dois dados são lançados simultaneamente e estamos interessados na soma das faces observadas.
d) Em uma cidade, famílias com 3 crianças são selecionadas ao acaso, anotando-se o sexo de cada uma.

Questão 5. Uma Universidade tem 10 mil alunos dos quais 4 mil são considerados esportistas. Temos ainda que 500 alunos são do curso de Administração noturno, 700 de Ciências contábeis noturno, 100 são esportistas e da Administração noturno e 200 são esportistas e da Ciências contábeis noturno. Qual a probabilidade de:

a) Ser esportista.
b) Ser esportista e aluno da Administração.
c) Ser esportista ou aluno da Ciências contábeis.
d) Não ser esportista nem aluno da Administração.

Questão 6. Sejam A e B dois eventos em um dado espaço amostral, tais que $P(A) = 0,2$, $P(B) = p$, $P(A \cup B) = 0,5$ e $P(A \cap B) = 0,1$. Determine o valor de p.

Questão 7. Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{4}$, e A e B são multuamente exclusivos, calcular:

a) $P(A^c)$.
b) $P(B^c)$.
c) $P(A \cap B)$.
d) $P(A \cup B)$.
e) $P(A^c \cap B^c)$.

Questão 8. Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{3}$ e $P(A \cap B) = \frac{1}{3}$. Calcule:

a) $P(A \cup B)$.

b) $P(A^c \cup B)$.

c) $P(A^c \cap B^c)$.

Questão 9. Dois dados são lançados simultaneamente. Qual a probabilidade de:

a) a soma ser menor que 4.

b) a soma ser 9.

c) o primeiro resultado ser maior que o segundo.

Questão 10. Qual a probabilidade de sair um rei ou uma carta de copas, quando retiramos uma carta de um baralho?

Questão 11. As probabilidades de três jogadores marcarem um pênalti são respectivamente: $\frac{2}{3}$, $\frac{4}{5}$ e $\frac{7}{10}$. Se cada um "cobrar" uma única vez, qual a probabilidade de:

a) todos acertarem?

b) apenas um acertar?

c) todos errarem?

Questão 12. Dois armários guardam as bolas de voleibol e basquete. O armário 1 tem três bolas de voleibol e 1 de basquete, enquanto o armário 2 tem 3 de voleibol e 2 de basquete. Escolhendo-se ao acaso um armário e, em seguida, uma de suas bolas, calcule a probabilidade dela ser:

a) De voleibol, sabendo-se que o armário 1 foi escolhido.

b) De basquete, sabendo-se que o armário 2 foi escolhido.

c) De basquete.

Questão 13. Em certo colégio, 5% dos homens, 2% das mulheres têm mais do que 1,80m. Por outro lado, 60% dos estudantes são homens. Se um estudante é selecionado aleatoriamente e tem mais de 1,80m de altura, qual a probabilidade de que o estudante seja mulher?

Questão 14. A probabilidade de uma mulher está viva daqui a 30 anos é $\frac{3}{4}$ e a de seu marido, $\frac{3}{5}$. Calcular a probabilidade de:

a) apenas o homem está vivo;

b) somente a mulher está;

c) ambos estarem vivos.

Questão 15. Verifique se as afirmações são válidas:

a) Se $P(A) = 1/3$ e $P(B|A) = 3/5$, então A e B não podem ser disjuntos.

b) Se $P(A) = 1/2$ e $P(B|A) = 1$ e $P(A|B) = 1/2$ então A não pode estar contido em B.

Questão 16. Se $P(A \cup B) = 0,8$, $P(A) = 0,5$ e $P(B) = x$, determine o valor de x no caso de:

a) A e B serem multuamente exclusivos.

b) A e B serem independentes.

Questão 17. Se P(B) = 0,4, P(A) = 0,7 e $P(A \cap B) = 0,3$, calcule $P(A|B^c)$.

Questão 18. O São Paulo Futebol Clube ganha com probabilidade 0,7 se chove e 0,8 se não chove. Em Dezembro, a probabilidade de chuva é de 0,3. O São Paulo ganhou uma partida em Dezembro, qual a probabilidade de ter chovido nesse dia?

Questão 19. Três máquinas A, B, e C produzem, respectivamente 30%, 50% e 20% do total de peças de uma fábrica. As porcentagens e peças defeituosas nas respectivas máquinas são 3%, 5% e 2%. Uma peça é sorteada ao acaso, e verifica-se que é defeituosa. Qual a probabilidade de que a peça tenha vindo da máquina B?

Questão 20. Numa certa população, a probabilidade de gostar de teatro é de 1/3, enquanto que a de gostar de cinema é 1/2. Determine a probabilidade de gostar de teatro e não de cinema, nos seguintes casos:

a) Gostar de teatro e gostar de cinema são eventos disjuntos.

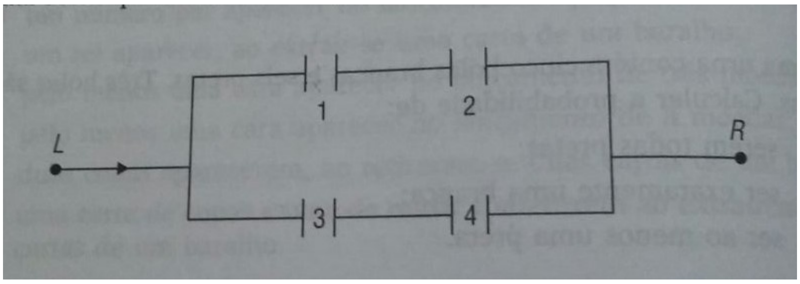
b) Gostar de teatro e gostar de cinema são eventos independentes.

c) Todos que gostam de teatro gostam de cinema.

d) A probabilidade de gostar de teatro e de cinema é de 1/8.

e) Dentre os que não gostam de cinema, a probabilidade de não gostar de teatro é de 3/4.

Questão 21. A probabilidade de fechamento de cada relé do circuito apresentado a seguir é dada por p. Se todos os relés funcionarem independentemente, qual será a probabilidade de que haja corrente entre os terminais L e R?



5

Variáveis aleatórias e distribuição de probabilidade

5.1 Introdução

Definição 11 (Variável Aleatória)**.** *Uma quantidade X, associada a cada possível resultados do espaço amostral, é denominada de variável aleatória discreta se assume valores num conjunto enumerável, com certa probabilidade. Por outro lado, será denominada variável aleatória contínua se seu conjunto de valores for qualquer intervalo de números reais, o que seria um conjunto não enumerável.*

Para explicar o conceito de uma variável aleatória será considerado o exemplo, no qual duas variedades de uma espécie A (A_1, A_2) e três de outra espécie B (B_1, B_2 e B_3) são disponibilizados para uma pesquisa. Uma amostra de duas variedades ($n = 2$) é extraída. O espaço amostral dos resultados desse experimento está apresentado a seguir.

$$\Omega = \left\{ \begin{array}{ccccc} (A_1, A_2) & (A_1, B_1) & (A_1, B_2) & (A_1, B_3) & (A_2, B_1) \\ (A_2, B_2) & (A_2, B_3) & (B_1, B_2) & (B_1, B_3) & (B_2, B_3) \end{array} \right\}$$

Se for considerado o número de variedades da espécie A na amostra sorteada, percebemos que os valores encontrados são: 0, 1 e 2. É possível associar a esses valores alguns pontos do espaço amostral Ω, formando subconjuntos, quais sejam:

$$\begin{array}{cc} 0 & \{(B_1, B_2)\ (B_1, B_3)\ (B_2, B_3)\} \\ 1 & \left\{ \begin{array}{ccc} (A_1, B_1) & (A_1, B_2) & (A_1, B_3) \\ (A_2, B_1) & (A_2, B_2) & (A_2, B_3) \end{array} \right\} \\ 2 & \{(A_1, A_2)\} \end{array}$$

Com essa associação de números aos pontos de Ω, espaço amostral, pretendemos definir uma função sobre os elementos desse espaço. Essas funções dos pontos de Ω são as variáveis aleatórias, que na realidade, são funções e não variáveis.

Para o exemplo do experimento do sorteio das duas variedades, definindo-se X como sendo a variável aleatória relativa a contagem de variedades da espécie A na amostra, verificamos que podem ser assumidos pela variável X: $x = 0, 1, 2$. É comum representar a variável por X (maiúsculo) e os seus valores por x (minúsculo). Considerando ainda, que para esse exemplo os pontos de Ω são equiprováveis, então a probabilidade de X assumir um dados valor x será denotado por $P(X = x)$, $p(x)$ ou p_i com $i = 1, 2, \ldots$, sendo denominada também de função de probabilidade de X.

A função $(x, p(x))$ é denominada de função de probabilidade da variável aleatória X. Para um mesmo espaço amostral, é possível associar outras variáveis aleatórias. No exemplo que está sendo considerado, poderia se pensar em uma variável aleatória Y que representasse o número de variedade da espécie B. A função de probabilidade de X define a distribuição de probabilidade dessa variável aleatória.

Dessa forma, a distribuição de probabilidade pode ser vista como uma correspondência que associa as probabilidades aos valores de uma variável aleatória, que é função do espaço amostral.

5.2 Variável aleatória discreta

5.2.1 Distribuição de probabilidade

A função que atribui a cada valor da variável aleatória sua probabilidade é denominada de função discreta de probabilidade, ou simplesmente distribuição de probabilidade. Observe:

X	x_1	x_2	$\dots$
p_i	p_1	p_2	$\dots$

Uma distribuição de probabilidade satisfaz as seguintes condições:

1. $0 \leq p_i \leq 1$;
2. $\sum_{i=1} p_i = 1$.

Retornando ao exemplo das variedades, podemos apresentar a distribuição de probabilidade da variável X, número de variedades da espécie A na amostra sorteada, $n = 2$. Cada ponto do espaço amostral amostral foi considerado como equiprovável.

X: número de variedades de A	$P(X = x)$: probabilidade
0	3/10
1	6/10
2	1/10

5.2.2 Função de distribuição

A função de distribuição ou função acumulada de probabilidade de uma variável aleatória X é definida, para qualquer número real, pela seguinte expressão:

$$F(X) = P(X \leq x) = \sum_{\{j|x_j<x\}} p(x_j) \tag{5.1}$$

Exemplo 13. *Uma população de 1000 crianças foi analisada num estudo para determinar a efetividade de uma vacina contra um tipo de alergia. No estudo, as crianças recebiam uma dose de vacina e após um mês passavam por um novo teste. Caso ainda tivessem algum reação alérgica, recebiam outra dose da vacina, Ao fim de 5 doses, todas as crianças foram consideradas imunizadas.*

Os resultados completos estão na Tabela 5.1.

Tabela 5.1: Número de frequências e doses recebidas.

Doses	1	2	3	4	5
Frequência	245	288	256	145	66

Supondo que uma criança é sorteada ao acaso dessa população qual a chance dela ter recebido 2 doses? Utilizando a ideia de atribuir probabilidade através da frequência de ocorrência, podemos imediatamente responder a questão. A probabilidade desejada é de 288/1000=0,288. Podemos assim obter a função de probabilidade para a variável aleatória número de doses recebidas.

Doses	1	2	3	4	5
p_i	0,245	0,288	0,256	0,145	0,066

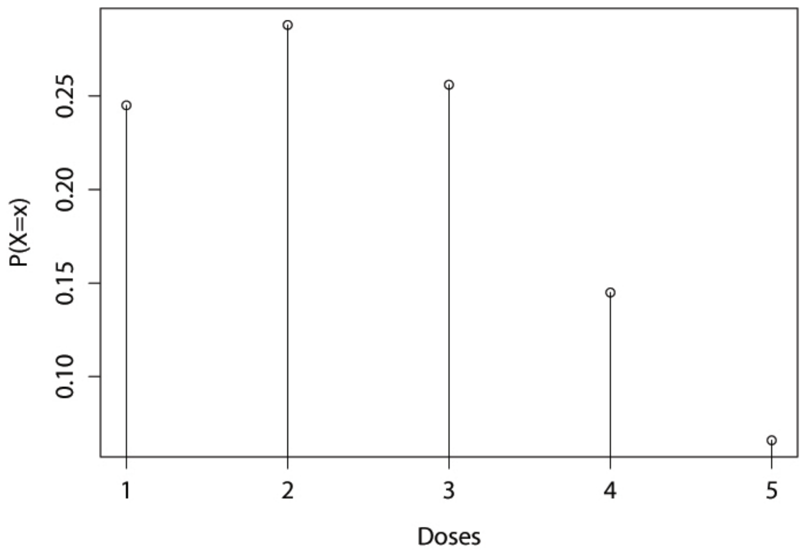


Figura 5.1: Gráfico da distribuição de probabilidade.

Suponha agora que desejamos calcular a probabilidade de uma criança ter recebido até duas vacinas. O que precisamos obter é a função de distribuição no ponto 2, ou seja, calculamos a probabilidade acumulada de ocorrência de valores menores ou iguais a 2. Assim,

$$F(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = 0,533.$$

Note que, tendo em vista que a variável só assume valores inteiros, esse valor fica inalterado no intervalo $[2, 3)$. Isto é, $F(2,3)$, $F(2,45)$ ou $F(2,99)$ têm todos o mesmo valor acima. Por essa razão escrevemos:

$$F(X) = P(X \leq x) = 0,533, \quad \text{para} \;\; 2 \leq x < 3.$$

Os valores completos da função de distribuição são os seguintes:

$$F(X) = \begin{cases} 0, & \text{se} \;\; x < 1; \\ 0,245, & \text{se} \;\; 1 \leq x < 2; \\ 0,533, & \text{se} \;\; 2 \leq x < 3; \\ 0,789, & \text{se} \;\; 3 \leq x < 4; \\ 0,934, & \text{se} \;\; 4 \leq x < 5; \\ 1, & \text{se} \;\; x \geq 5. \end{cases}$$

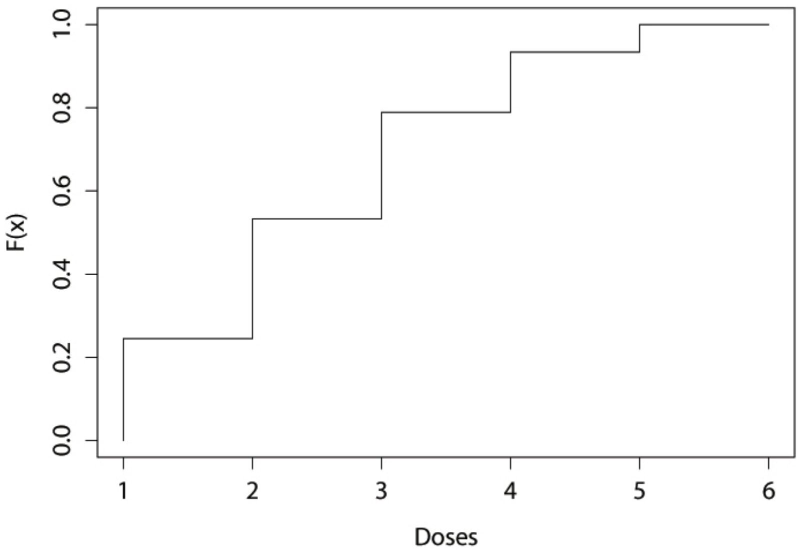


Figura 5.2: Gráfico da função de distribuição.

5.3 Variável aleatória contínua

Dizemos que $f(x)$ é uma função contínua de probabilidade ou função densidade de probabilidade, uma variável aleatória contínua X, que satisfaz duas condições:

1. $f(x) \geq 0$, para todo $x \in (-\infty, \infty)$;
2. A área definida por $f(x)$ é igual a 1.

Com o auxílio do cálculo diferencial e integral, podemos caracterizar a condição 2 por meio de

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Da mesma forma, para calcular probabilidade, temos que $a \leq b$,

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

A integral acima indica a área sob a função f definida pelo intervalo $[a, b]$. Note que, pela forma como atribuímos as probabilidade no caso contínuo, teremos área zero sob qualquer valor individual, isto é, $P(X = x_k)$, para qualquer k. Portanto, em se tratando de variáveis aleatórias contínuas, a probabilidade de ocorrência de um valor isolado é sempre zero e, consequentemente, as probabilidade calculadas sobre os intervalos $[a, b]$, $[a, b)$, $(a, b]$ e (a, b) são os mesmos para quaisquer valores de a e b.

Exemplo 14. *Verifica se* $f(x) = \frac{1}{128}e^{-x/128}$ *é uma densidade de probabilidade para* $x \geq 0$*. A variável aleatória* X *representa o tempo de vida de uma espécie vegetal arbórea dada em anos.*

i) $f(x) \geq 0$, pois $e^{-x/128} \geq 0 \;\; \forall \;\; x \geq 0$, e $1/128$ é uma constante sempre positiva.

ii) Verificar se $\displaystyle\int_0^{\infty} f(x)dx = 1$.

Inicialmente vamos relembrar a derivada de uma função $h(x) = e^{-mx}$:

$$h'(x) = -me^{-mx},$$

integrando em ambos os lados, temos:

$$\begin{aligned}
\int h'(x)dx &= \int -me^{-mx}dx, \\
h(x) &= -m\int e^{-mx}dx, \\
\frac{h(x)}{-m} &= \int e^{-mx}dx,
\end{aligned}$$

como $h(x) = e^{-mx}$, então,

$$\int e^{-mx}dx = \frac{-e^{-mx}}{m} + c. \tag{5.2}$$

Assim, se considerarmos $m = 1/128$, então começamos a prova do item (ii).

$$\begin{aligned}
\int_0^\infty \frac{1}{128}e^{-\frac{1}{128}x}dx &= \frac{1}{128}\underbrace{\int_0^\infty e^{-\frac{1}{128}x}dx}_{\text{idêntica a } eq.(5.2)}, \\
&= \frac{1}{128}\left[\frac{-e^{-(1/128)x}}{1/128}\right]_0^\infty, \\
&= \frac{1}{128}\left[0 - \left(\frac{-e^{-(1/128)\times 0}}{1/128}\right)\right], \\
&= \frac{1}{128}\left[128e^0\right] \\
&= \frac{1}{128} \times 128 \times 1 = 1.
\end{aligned}$$

Portanto, $f(x)$ é uma função densidade.

Exemplo 15. *Seja*

$$f(x) = \begin{cases} 1/6x + k, & se\ 0 \leq x \leq 3 \\ 0, & em\ qualquer\ outro\ caso. \end{cases}$$

Encontrar o valor de "k"na função para que $f(x)$ seja uma densidade de probabilidade (f.d.p.).

Para encontrar o valor de k, então

$$\int_0^3 (1/6x + k)dx = 1.$$

Assim,

$$\begin{aligned}
\int_0^3 (1/6x + k)dx &= \left[1/6\frac{x^2}{2} + xk\right]_0^3 \\
&= 1/6\frac{3^2}{2} + 3k.
\end{aligned} \tag{5.3}$$

Igualando (5.3) a 1, temos:

$$1/6\frac{3^2}{2} + 3k = 1 \Rightarrow k = 1/12.$$

Portanto, $f(x) = 1/6x + 1/12$ é uma f.d.p..

5.3.1 Função de distribuição

A função de distribuição ou função acumulada de uma variável aleatória contínua de X é definida para qualquer número real x, pela seguinte expressão

$$F(x) = \int_{-\infty}^{x} f(x)dx. \tag{5.4}$$

5.3.1.1 Propriedades

i) $F(-\infty) = \lim\limits_{x\to-\infty} F(x) = 0$ e $F(\infty) = \lim\limits_{x\to\infty} F(x) = 1$;

ii) Se $a < b \Rightarrow F(a) \leq F(b)$;

iii) F(x) é contínua à direita, $\lim\limits_{x\to a^+} F(x) = F(a)$

Essas propriedades podem ser melhor observadas pela Figura 5.3,

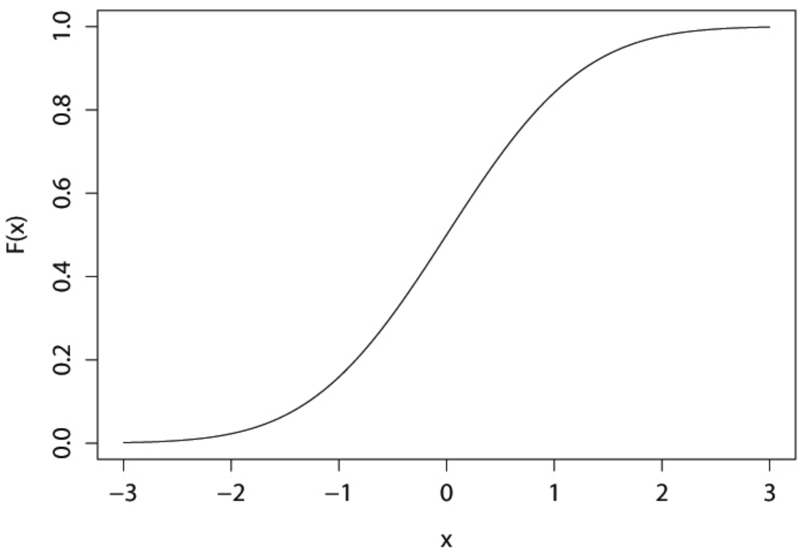


Figura 5.3: Gráfico da função de distribuição.

5.3.1.2 Relação entre $f(x)$ e $F(x)$

i) Se $f(x)$ é conhecido, então $F(x) = \int_{-\infty}^{x} f(x)dx$

ii) Se $F(x)$ é conhecido, então $f(x) = \frac{d}{dx}F(x)$ $\forall x$ em que a derivada exista.

Essa relação vale também para as variáveis aleatórias discretas, em que no lugar da integral, usa-se o somatório.

5.4 Medidas de posição e dispersão

5.4.1 Esperança matemática

Esperança matemática, valor esperado ou média da variável X e dado pela expressão,

$$E[X] = \sum_{i=1}^{n} x_i p(x_i), \tag{5.5}$$

para o caso das v.a. discretas, e

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx, \tag{5.6}$$

para o caso das v.a. contínuas.

5.4.1.1 Propriedades

i) $E[k] = k$, sendo k uma constante;

ii) $E[kX] = kE[X]$;

iii) $E[X \pm Y] = E[X] \pm E[Y]$;

iv) $E[X \pm k] = E[X] \pm k$;

v) $E[XY] = E[X]E[Y]$ se X e Y forem independentes.

5.4.2 Mediana

A mediana é o valor que satisfaz a seguinte expressão

$$P(X \geq Md) = P(X \leq Md) = 0,50, \tag{5.7}$$

isto é, existe igual probabilidade de ocorrência de valores acima e abaixo da mediana. Em algumas situações, a igualdade é satisfeita por qualquer valor num certo intervalo e, nesse caso, tomamos a mediana como ponto médio de intervalo.

5.4.3 Moda

A moda é o valor (ou valores) da variável aleatória que tem maior probabilidade de ocorrência, representando-a por Mo, temos

$$P(X = Mo) = \max(p_1, p_2, \ldots, p_n). \tag{5.8}$$

Exemplo 16. *Considere a v.a. X com a seguinte função discreta de probabilidade.*

x	-5	10	15	20
$p(x)$	0,3	0,2	0,4	0,1

Então,

$$E[X] = \sum_{i=1}^{n} x_i p(x_i) = (-5) \times 0,3 + 10 \times 0,2 + \ldots + 20 \times 0,1 = 8,5 \text{ unid.}.$$

A moda é o valor com maior probabilidade e, portanto segue imediatamente que $Mo = 15$ unid.. Por outro lado, a mediana poderá ser qualquer número entre 10 e 15, pois,

$$P(X \leq 10) = P(X \geq 15) = 0,50.$$

Por conveniência adotada, tomamos $Md = 15$unid. (ponto médio do intervalo).

5.4.4 Variância e desvio padrão

Seja uma v.a. X com $P(X_i = x_i) = p_x(x_i)$, $i = 1, 2, \dots, n$ e média μ. A variância de X é a ponderação pelos respectivas probabilidade, dos desvios relativos à média elevados ao quadrado, isto é,

$$Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i). \tag{5.9}$$

Extraindo-se a raiz quadrada da variância, obtemos o desvio padrão que é representado por $DP(X)$, dado pela expressão

$$DP(X) = \sqrt{Var(X)}. \tag{5.10}$$

Se considerarmos o valor esperado da função $g(.)$ da variável aleatória X, denotado $E[g(x)]$, definido por:

$$E[g(x)] = \sum_{i=1}^{n} g(x_i)p(x_i),$$

para o caso discreto, e

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx,$$

para o caso contínuo.

Então, para $g(x) = x$, temos

$$E[g(x)] \sum_{i=1}^{n} x_i p(x_i) = E[X] = \mu_x,$$

para o caso discreto, e

$$E[g(x)] = \int_{-\infty}^{infty} xf(x)dx = E[X] = \mu_X,$$

para o caso contínuo.

Para o caso $g(x) = (x - \mu)^2$, temos

$$E[g(x)] = E[(x - \mu_X)^2] \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx = Var(X), \tag{5.11}$$

para o caso contínuo, e

$$E[g(x)] = E[(x - \mu_X)^2] = \sum_{i=1}^{n} (x_i - \mu_X)^2 p(x_i) = Var(X),$$

como visto pela equação (5.9).

Portanto, podemos redefinir a variância como $E[(X - \mu_X)^2]$. Como $\mu_X = E[X]$, então

$$\begin{aligned} Var(X) &= E[(X - E[X])^2], \\ &= E[X^2 - 2XE[X] + E^2[X]], \\ &= E[X^2] - 2E[X]E[X] + E^2[X], \\ &= E[X^2] - E^2[X]. \end{aligned} \tag{5.12}$$

5.4.4.1 Propriedades

i) $Var(k) = 0$, sendo k uma constante;

ii) $Var(kX) = k^2 Var(X)$;

iii) $Var(k \pm X) = Var(X)$;

iv) $Var(X \pm Y) = Var(X) \pm Var(Y)$ se X e Y forem independentes.

6

Distribuições especiais de probabilidade

O estudo das distribuições especiais abrange as principais distribuições discretas (Binomial e Poisson), contínuas (Normal), bem como as distribuições amostrais (Qui-quadrado e t de Student). Essas distribuições são de tal importância, pois grande parte das variáveis podem ser modeladas ou derivadas por meio delas.

6.1 Distribuição Bernoulli

A variável aleatória discreta X é definida ter distribuição Bernoulli se a distribuição de probabilidade de X de parâmetro p, é definida por

$$P(X = x) = \begin{cases} p^x(1-p)^{1-x}, & \text{para } x = 0 \text{ ou } 1, \\ 0, & \text{caso contrário} \end{cases} \tag{6.1}$$

em que p satisfaz $0 \leq p \leq 1$.

Em outras palavras, define-se uma v.a. X que assume dois possíveis resultados: o valor 1, se ocorrer o evento desejado (sucesso); e o valor 0, se ocorrer o valor não desejado (fracasso). Definindo-se ainda, p como sendo a probabilidade de ocorrer o sucesso, e $q = (1-p)$, a probabilidade de ocorrer o fracasso, tem distribuição de probabilidade Bernoulli.

Graficamente, temos:

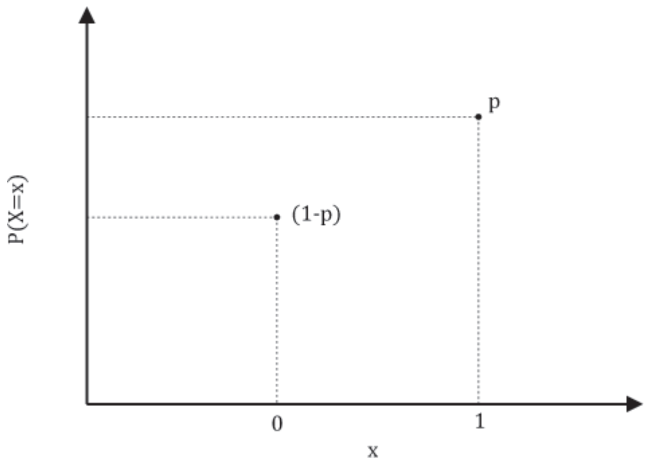


Figura 6.1: Função de probabilidade da distribuição Bernoulli com parâmetro p.

6.1.1 Validando a distribuição de probabilidade

i) Mostrar que $P(X = x) \geq 0$;

ii) Mostrar que $\sum_{x=0}^{n} P(X = x) = 1$.

Prova:

i) Como os valores de $x = 0$ ou 1 e $0 \leq p \leq 1$, então $p^x(1-p)^{1-x} \geq 0$;

ii) $\sum_{x=0}^{n} P(X = x) = P(X = 0) + P(X = 1) = (1-p) + p = 1$.

Exemplo 17. *Uma urna contém 30 bolas brancas e 20 verdes. Retira-se uma bola dessa urna. Seja X: número de bolas verdes, determine $P(x)$.*

Solução:

$$X = \begin{cases} 0 \Rightarrow q = 30/50 = 3/5, \\ 1 \Rightarrow p = 20/50 = 2/5, \end{cases} \quad \therefore P(X = x) = (2/5)^x (3/5)^{1-x}.$$

6.1.2 Esperança, Variância e Desvio Padrão

Sendo a esperança definida por:

$$E[X] = \sum_{x=0}^{n} xP(X = x),$$

e considerando o experimento da v.a. X com distribuição Bernoulli, em que $x = 1$, sucesso para o experimento com probabilidade p e $x = 0$, o fracasso, com probabilidade $(1-p)$, então a esperança será,

$$E[X] = 1 \times p + 0 \times (1-p) = p, \tag{6.2}$$

e a variância sendo $Var(X) = E[X^2] - E^2[X]$, então,

$$E[X^2] = 1^2 \times p + 0^2 \times (1-p) = p$$

e

$$E^2[X] = p^2,$$

portanto,

$$Var(X) = p - p^2 = p(1-p) = pq. \tag{6.3}$$

O desvio padrão é dado por:

$$DP(X) = \sqrt{pq}. \tag{6.4}$$

6.2 Distribuição Binomial

Considere a repetição de n ensaios de Bernoulli independentes e todos com a mesma probabilidade de sucesso p. A v.a. que conta o número total de sucessos tem distribuição Binomial com parâmetros n e p, definida por:

$$P[X = x] = C_x^n p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, \tag{6.5}$$

com C_x^n representando o coeficiente binomial calculado por

$$C_x^n = \frac{n!}{x!(n-x)!}$$

A esperança de X com distribuição binomial com parâmetros p e n, então

$$E[X] \quad = \quad np.$$

A variância de X é

$$Var(X) \quad = \quad = np(1-p). \tag{6.6}$$

O desvio padrão é dado por:

$$DP(X) = \sqrt{np(1-p)}. \tag{6.7}$$

6.2.1 Características

1. Uma distribuição binomial fica caracterizada pelos parâmetros n e p.
2. Se n for pequeno, os cálculos serão relativamente fáceis. Contudo, se n for relativamente grande, os cálculos tornam-se cansativos. felizmente, dispomos de calculadoras, algoritmos computacionais, e também poderemos aproximar a distribuição binomial pela distribuição de poisson.
3. Para qualquer n, a distribuição será simétrica, se $p = q = 0,5$, será assimétrica à direita, se $p > q$, e assimétrica à esquerda, se $p < q$.

6.3 Distribuição de Poisson

Uma variável aleatória X tem distribuição de Poisson com parâmetro $\lambda > 0$ se sua distribuição de probabilidade é dada por:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \tag{6.8}$$

Com o parâmetro λ sendo usualmente requerido como a taxa de ocorrência. O modelo tem sido muito utilizado em experimentos físicos e biológicos. Nesses casos, λ é a frequência média ou esperada de ocorrências num determinado intervalo de tempo.

A esperança matemática de X é

$$E[X] \quad = \quad \lambda.$$

A variância de X é

$$Var(X) \quad = \quad \lambda. \tag{6.9}$$

o desvio padrão é dado por

$$DP(X) = \sqrt{\lambda}. \tag{6.10}$$

6.4 Relação entre Binomial e Poisson

Seja $X \sim Binom(n,p)$, vamos discutir as condições de aproximação com distribuição de Poisson.

Temos

$$\begin{aligned} P(X=x) &= \binom{n}{x} p^x (1-p)^{n-x}, \\ &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \end{aligned}$$

sendo $n!/[x!(n-x)!] = (n)_x = n(n-1)\ldots(n-x+1)$, e definindo $\lambda = np$, então

$$P[X=x] = \frac{(n)_x}{x!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}.$$

Observando que $\left(1-\frac{\lambda}{n}\right)^{n-x} = \left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-x}$, temos

$$\begin{aligned} P[X=x] &= \frac{(n)_x}{x!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-x}, \\ &= \frac{(n)_x}{x!}\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-x}, \\ &= \frac{(n)_x}{n^x}\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-x}. \end{aligned}$$

Considerando agora que a probabilidade p seja de tal modo pequena, quando n se aproxima do infinito, λ passa a ser considerado como aproximadamente constante.

Na aplicação do limite observe que

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} = 1 \quad \text{e} \quad \lim_{n\to\infty}\frac{(n)_x}{n^x} = 1,$$

e por um resultado clássico de limite, segue também que

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}.$$

Nestes termos:

$$\lim_{n\to\infty} P(X=x) \simeq \frac{\lambda^x e^{-\lambda}}{x!}, \tag{6.11}$$

que corresponde a função de probabilidade Poisson.

6.5 Distribuição Normal

Dizemos que uma v.a. contínua X tem distribuição normal com parâmetros μ e σ^2 se sua função densidade é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad \text{para} \quad -\infty < x < \infty. \tag{6.12}$$

Algumas propriedade da densidade da normal podem ser facilmente observadas de seu gráfico:

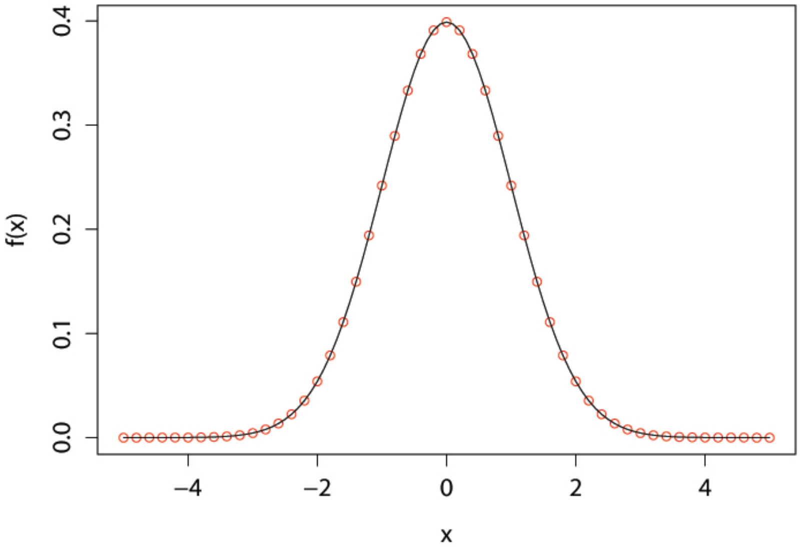


Figura 6.2: Função densidade da distribuição normal.

i) $f(x)$ é simétrica em relação a μ;

ii) $f(x) \to 0$ quando $x \to \pm\infty$;

iii) o valor máximo de x se dar para $x = \mu$.

A esperança de X é dada por:

$$E[X] \quad = \quad \mu. \tag{6.13}$$

A variância é dada por:

$$Var(X) \quad = \quad \sigma^2.$$

O desvio padrão é dado por:

$$DP(X) = \sqrt{\sigma^2} = \sigma. \tag{6.14}$$

6.5.1 *Normal padrão*

Seja X uma variável aleatória com distribuição normal com média μ e variância σ^2. Então,

$$Z = \frac{X - \mu}{\sigma}, \tag{6.15}$$

tem distribuição normal com média igual a zero e variância igual a um, ou seja,

$$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}E[X-\mu] = \frac{E[X]-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0,$$

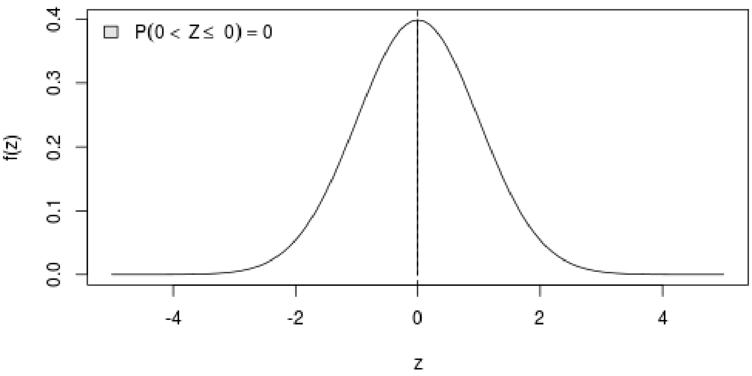
e

$$Var[Z] = Var\left[\frac{X-\mu}{\sigma}\right] = \frac{Var[X-\mu]}{\sigma} = \frac{Var[X]}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1.$$

A distribuição da variável aleatória Z é conhecida como normal padrão.

Tabela da distribuição normal padrão

Ben Dêivide – DEFIM/UFSJ



Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997

6.5.2 Exercícios Propostos

Questão 1. Defina variável aleatória (V.A.), variável aleatória discreta (V.A.D.), variável aleatória contínua (V.A.C.) e descreva diversos exemplos na sua área de estudo.

Questão 2. Suponha que 2% dos itens produzidos por uma fábrica sejam defeituosos. Encontre a probabilidade P de existirem 3 defeituosos em uma amostra de 100.

Questão 3. Usando a curva normal padronizada, determinar as áreas subtendidas entre os valores abaixo, com representação gráfica.

a) 0,35 e 0,0;
b) 0,0 e 1,52;
c) -0,34 e 1,9;
d) à direita de -1,91;
e) à esquerda de 1,13;
f) à esquerda de -2,13.

Questão 4. Dada uma distribuição normal com $\mu = 40$ e $\sigma = 6$, calcular:

a) $P(X \leq 33)$;
b) $P(X \geq 29)$;
c) $P(39 \leq X \leq 45)$;
d) Ponto que tem 58% de área acima dele;
e) Ponto que tem 5% de área acima;
f) $P(Z > 0)$.

Questão 5. Em um exame vestibular de matemática as notas distribuíram-se normalmente com média 6 o desvio padrão 1,5. Calcular o número de aprovados entre os 120 candidatos, sabendo-se que a nota mínima de aprovação é 5.

Questão 6. Uma máquina de empacotar determinado produto apresenta variação de peso com desvio padrão de 20g. Em quanto deve ser regulado o peso médio do pacote para que apenas 10% tenham menos de 400g. Super distribuição normal dos pesos dos pacotes.

Questão 7. Na 1ª prova de Estatística a média foi 4,5 e o desvio padrão 2,3. Considerando o método científico de aprovação:

- Conceito A: nota média $\geq \mu + \sigma$;
- Conceito B: $\mu \leq$ média $\leq \mu + \sigma$;
- Conceito C: $\mu - \sigma \leq$ média $\leq \mu + \sigma$;
- Conceito C: $\mu - \sigma \leq$ média;

a) Quantos alunos receberam cada um dos conceitos?
b) Quantos alunos foram aprovados (Conceito A, B e C);
c) Quantos alunos foram aprovados com distinção (média $\geq \mu + 2\sigma$);
d) Considere o método comum de aprovação ($X \geq 5,0$), quantos foram reprovados.

OBSERVAÇÃO: 100 alunos fizeram essa prova.

Questão 8. Ao investir num determinado negócio, um cidadão pode ter um lucro anual de US$ 60.000 com probabilidade 0,3 ou tomar um prejuízo de US$ 20.000 com probabilidade 0,7. Qual é a sua esperança matemática?

7

Amostragem

A amostragem ilizada no nosso dia a dia inconscientemente. Quando preparamos uma refei e vamos prova antes de servi-la, estamos na realidade fazendo uma amostragem, ao passo que a refei representa o todo (popula) e a parte retirada para provar se a refei estava pronta para servi-la mostra. Dessa forma, queremos, por meio de uma amostra, saber algumas caractericas (partro) sobre a popula. As medidas obtidas da amostra, com as quais nos indicam alguma caracterica da popula, chamamos de estimador. Com base nisso, para obtermos bons estimadores, precisamos garantir antecipadamente que a amostra coletada da popula represente a popula, de modo a preservar as caractericas relevantes da popula.

Definição 12 (Popula)**.** *O conjunto de elementos (unidades) para os quais desejamos que as concluses de pesquisa sejam vdas, com a restri de que esses elementos possam ser observados ou mensurados sob as mesmas condis amado de popula, universo ou popula objetivo. O tamanho da popula serpresentado por N.*

A popula pode ser formada por pessoa, famas, estabelecimentos industriais, ou qualquer tipo de elementos, dependendo basicamente dos objetivos da pesquisa. Podemos dizer ainda que uma popula pode ser finita ou infinita. Uma popula nita quando se consegue enumerar todos os elementos que a formam. Refere-se a um universo limitado em uma dada unidade de tempo. Como exemplo podemos dizer que a quantidade de automs produzidos em um m a popula de uma cidade, o nmero de alunos de uma sala de aula sxemplos de popula finita. Uma popula ta infinita quando nodemos enumerar todos os elementos. Refere-se a um universo nelimitado. Os resultados (cara ou coroa) obtidos em sucessivos lances de uma moeda, o conjunto dos nmeros inteiros, reais ou naturais sxemplos de popula infinita.

Definição 13 (Elemento, unidade ou unidade elementar)**.** *Indivo ou objeto a ser medido ou observado na pesquisa, do qual ntidade portadora de informas que pretende-se coletar amado de unidade elementar (UE).*

Um outro tipo de elemento da popula nidade amostral.

Definição 14 (Unidade amostral)**.** *Um conjunto formado por uma unidade elementar ou vas unidades elementares do qual seja de interesse para pesquisa amada de unidade amostral (UA).*

Exemplo 18. *Uma unidade amostral pode ser os conclomerados pelo mdo de amostragem por conglomerados. Exemplos de conglomerados s quarteres, ruas, departamentos, prateleiras, caixas, lotes de produtos, etc..*

O interesse na pesquisa est determinar caractericas da popula relevantes para o estudo. Definimos,

Definição 15 (Partro)**.** *Qualquer caracterica atribu a popula amada de partro.*

Exemplo 19. *Numa pesquisa epidemiola, a popula pode ser definida como todas as pessoas (unidade elementar) da regim estudo, no momento da pesquisa. Um partro de interesse pode ser a porcentagem de pessoas contaminadas.*

Como nem sempre ssl coletar informas de todas as unidades elementares da popula, faz-se necesso retirarmos informas sobre uma parte da popula acessl, e assim, por meio destas informas, consigamos obter informas sobre o partro de interesse na pesquisa.

Definição 16 (Amostragem)**.** *O mecanismo (tica) que consistem em selecionar parte de uma popula para observar, de modo que seja possl preservar as principais informas sobre toda a popula.*

A parte da popula chamamos de amostra.

Definição 17 (Amostra)**.** *Qualquer subconjunto da popula obtido por meio de um processo de sele adequado amado de amostra. O tamanho da amostra serpresentado por n.*

Dizemos que quando uma amostra preserva as principais caractericas de uma popula, diz-se que esta a amostra representativa. Quais pro n estiver de N mais representativo serta amostra de tamanho n.

As informas obtidas por meio de uma amostra shamadas de estaticas. Quando uma estatica estima um partro da popula, temos um estimador.

Definição 18 (Estimador)**.** *Um estimador a medida, fun da amostra (estatica) que representarartro desconhecido da popula.*

Exemplo 20. *Uma pesquisa deseja saber a altura dos estudantes de uma determinada universidade. Se usarmos uma medida de tendia central para representar a altura dos estudantes, poderos escolher a ma. Logo, um partro para a popula amapopulacional.Geralmenteμ ou qualquer outro partro sconhecido, entomo estimador para μ podemos considerar $\bar{X}$, a ma amostral. Observe que este estimador n da amostra. Dizemos que o resultado de um estimador stimativa.*

A estima nada mais que criarmos mecanismos para estrapolar as medidas amostrais para que estas possam representar os partros desconhecidos. Como o estimador n da amostra, se esta n representativa, com certeza estaremos comentendo algum erro em afirmar que esse estimador poderia representar o partro desconhecido. Damportia de obtermos metodologias para coletar amostras representativas da popula.

Definição 19 (Erro amostral)**.** *O erro amostral iferenntre a estimativa e o partro que se quer estimar.*

Esse erro reflete a tendia da amostra. Querendo ou n o fato de jmarmos decisom base em uma amostra, jtamos cometendo erro. Contudo, o que a estatica tenta mostrar, por meio da amostragem, e podemos tomar concluses com base em uma amostra sobre a popula de modo a cometer o mmo de erro amostral possl. Assim, dizemos que o objetivo da pesquisa amostral sernhecer caractericas sobre a popula, pesquisando (estudando) a amostra, de modo a cometer o mmo de erro amostral possl.

Se a pesquisa envolve a observa de todas as unidades da popula, o mdo de pesquisa nominado censo ou pesquisa exaustiva. Se nduzida sobre uma amostra da popula, o mdo de pesquisa nominado levantamento por amostragem. O censo somente licl na situa que a popula seja finita e suas unidades sejam identificis e disponis para a coleta da amostra. Por essa raz o levantamento por amostragem ito mais frequentemente utilizado. Por que fazer amostragem ao inve um censo?

- Vantagens: Pesquisa por amostragem em rela ao censo:
  - ais barata;
  - ais rda;
  - ais fl de ser controlada por envolver operas menores.
- Desvantagens: Pesquisa por amostragem em rela ao censo:
  - O censo pode ser mais vantajoso quando a popula quena e/ou as informas se fl obten;
  - Os resultados da pesquisa por amostragem contrros amostrais;
  - Se a popula for muito heteroga o erro pode ser muito grande, dependendo do mdo de amostragem que seja utilizado.

Numa fase inicial dos levantamentos amostrais cesso **formular o problema** e aventar **hipes** sobre o objeto de estudo ou expectativas sobre os possis resultados. Ainda nessa fase inicial, o investigador deve **definir a popula de estudo**, parte identificl e acessl da popula objeto, os **objetivos** e as **variis observadas**. Numa segunda etapa alizado o planejamento, elaborado o **plano de amostragem** ou determinando o caminho a ser percorrido para atingir os objetivos propostos. O plano de amostragem devem ter bem definidos:

1. Unidade amostral: indivos ou grupos de indivos (conglomerados);
2. Sistema de referia: lista completa das unidades amostrais;
3. Tamanho da popula (N), finido pelo nmero de indivos da popula objetivo;
4. Tamanho da amostra (n), definido pelo nmero de indivos selecionados na amostra, tal que $n < N$.

Os planos ou mdos de amostragem podem ser classificados em amostragem probabilica e amostragem nrobabilica. Dizemos que se um mdo de amostragem jetivo e estabelece uma probabilidade conhecida a cada unidade da popula objetivo ser inclu na amostra, esse nominado amostragem probabilica ou amostragem aleat; caso contro, nominado nrobabilica ou amostragem nleat. Os principais esquemas amostrais podem ser observados na Figura 7.1.

Nas aulas em sala explanamos sobre os mdos de amostragem probabilica. Iremos falar sobre os mdos de amostragem nprobabilica.

O primeiro mdo amostragem a esmo. Imagine uma caixa com 1.000 parafusos. A enumera destes parafusos ficaria muito difl, e a amostragem aleat simples se tornaria invil. Ent em situas deste tipo, supondo que a popula de parafusos seja homoga, escolhemos a esmo a quantidade relativa ao tamanho a amostra. Quanto mais homoga for a popula, mais podemos supor a equivalia com uma amostragem simples ao acaso. Desta forma, os parafusos escolhidos para compor a amostra de um determinado tamanho sem nenhuma norma ou a esmo. Por isso, do nome ao mdo de amostragem.

Outro mdo amostragem intencional que corresponde ela em que o amostrador deliberadamente escolhe certos elementos para pertencer ostra, por julgar tais elementos bem representativos da popula. Um exemplo deste tipo de amostragem corresponde tua em que deseja saber a aceita em rela a uma nova marca de um determinado produto a ser inserida no mercado de uma cidade. Somente entrarara compor a amostra pessoas que fa uso do produto e que tenham condis financeiras de comprar esta nova marca (classe social de maior poder aquisitivo).

O ltimo mdo falado amostragem por cotas. Nesse tipo de amostragem, a popula vidida em grupos, e seleciona-se uma cota proporcional ao tamanho de cada grupo. Entretanto, dentro de cada grupo n feito sorteio,e sim os elementos srocurados ate a cota de cada grupo seja cumprida.

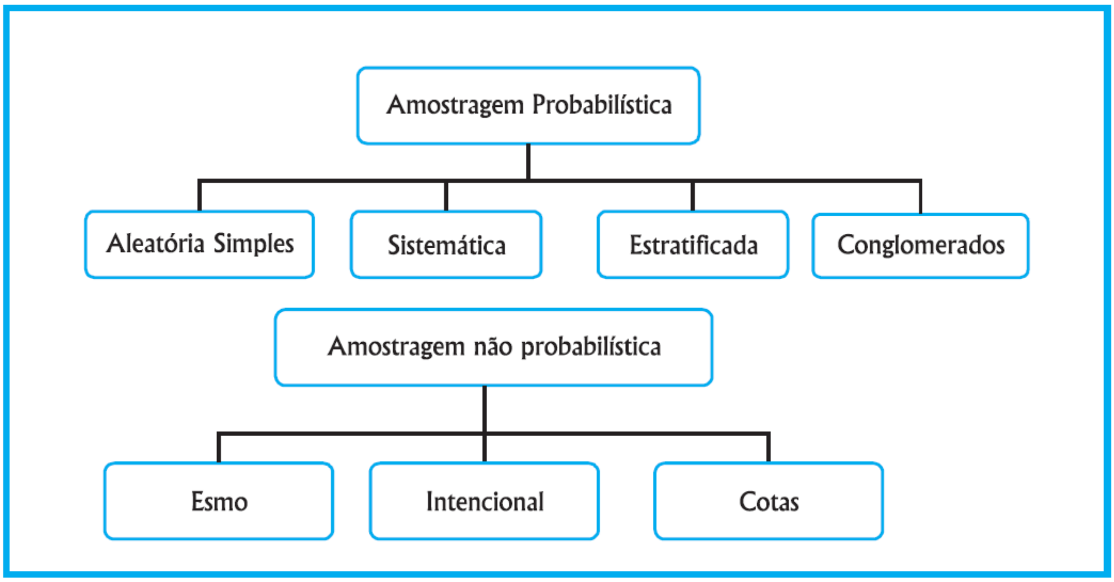


Figura 7.1: Mdos de amostragem probabilica e nprobabilica.

Em pesquisas eleitorais, a divise uma popula em grupos (considerando, por exemplo, o sexo, o nl de escolaridade, a faixa eta e a renda) pode servir de base para a defini dos grupos, partindo da suposi de que estas variis definem grupos com comportamentos diferenciados no processo eleitoral.

Referências Bibliográficas

DEVORE, J. L. **Probabilidade e estatística**: para engenharia e ciências. São Paulo: Cengage Learning, 2006. 692 p.