Universidade Federal do Piauí
Centro de Educação Aberta a Distância

PROBABILIDADE E ESTATÍSTICA I

Juarez Rodrigues Martins

Ministério da Educação - MEC
Universidade Aberta do Brasil - UAB
Universidade Federal do Piauí - UFPI
Centro de Educação Aberta a Distância - CEAD

Probabilidade e Estatística I

Juarez Rodrigues Martins

UFPI
2013

M386p Martins, Juarez Rodrigues.
Probabilidade e estatística I / Juarez Rodrigues Martins.
– Teresina : CEAD/UFPI, 2011.
131 p.

ISBN: 978-85-7463-492-0

1. Estatística. 2. Probabilidade. 3. Estatística Descritiva.
I. Título.

C.D.D. - 519.2

Apresentação

Este livro é destinado aos alunos aprendizes que participam do Programa de Educação a Distância da Universidade Aberta do Piauí (UAPI), vinculada ao consórcio formado pela Universidade Federal do Piauí (UFPI), Universidade Estadual do Piauí (UESPI) e Instituto Federal de Educação, Ciência e Tecnologia do Piauí – IFET, com apoio do Governo do Estado do Piauí, através da Secretaria de Educação.

Este material é composto de quatro unidades, contendo subunidades, estruturadas de modo sequencial, que discorrem sobre a Estatística Básica; Fundamentos de Análise Combinatória; Principais Conceitos de Probabilidades e seus principais Teoremas e Variáveis Aleatórias.

Na Unidade 1, apresentamos os principais conceitos de Estatística e distribuições de frequências. Exemplificamos o rol, dados brutos, passos para a construção de uma tabela de distribuição de frequências, além de uma lista de exercícios no final da unidade.

Na Unidade 2, apresentamos as Medidas de posição, também conhecidas como Medidas de tendência central e as separatrizes. As principais Medidas de Tendência Central são: a média aritmética e a média ponderada, a moda e a mediana, como também as separatrizes, como: a mediana, que também é considerada uma medida separatriz, os quartis, os decis e os centis.

Na Unidade 3, apresentamos as Medidas de dispersão ou variabilidades; divididas em Dispersão absoluta e Dispersão relativa. As principais Medidas de Dispersão Absoluta são: Amplitude total, desvio médio, a variância e o desvio padrão. As principais Medidas de Dispersão Relativas são: Coeficiente de dispersão de Pearson e o Coeficiente de dispersão de Thorndike.

Na Unidade 4, são expostos conteúdos sobre Análise Combinatória; principais conceitos de Probabilidades e seus principais Teoremas, como também Variáveis aleatórias.

Sumário

09 **UNIDADE 1**
DISTRIBUIÇÃO DE FREQUÊNCIAS
Conceitos Básicos de Estatística..................................11
Distribuição de Frequências.......................................14

33 **UNIDADE 2**
MEDIDAS DE POSIÇÃO
Média..33
Moda...48
Mediana..53
Quartis – decis – centis..64

47 **UNIDADE 3**
MEDIDAS DE DISPERSÃO
Medidas de dispersão absoluta.................................79
Medidas de dispersão relativa...................................94

99 **UNIDADE 4**
FUNDAMENTOS DE ANÁLISE COMBINATÓRIA, PRINCIPAIS CONCEITOS PROBABILÍSTICOS E SEUS PRINCIPAIS TEOREMAS E VARIÁVEL ALEATÓRIA
Fundamentos de Análise Combinatória.....................103
Probabilidade...108

UNIDADE 1

Distribuição de frequência

Objetivos:

- Identificar os principais conceitos básicos de estatística;
- Construir tabelas de frequências e gráficos;
- Identificar os principais elementos de uma distribuição de frequência;
- Diferenciar a estatística descritiva da estatística indutiva;
- Identificar as fases do método estatístico.

1 DISTRIBUIÇÃO DE FREQUÊNCIA

Conceitos Básicos de Estatística

Estatística

No plural, Estatísticas indicam qualquer coleção consistente de dados numéricos, reunidos com o objetivo de fornecer informações acerca de uma atividade qualquer. Por exemplo, as estatísticas demográficas referem-se aos dados numéricos sobre nascimentos, falecimentos, matrimônios, desquites, etc.

No singular, Estatística indica a atividade humana especializada ou um corpo de técnicas ou ainda uma metodologia desenvolvida para a coleta, a classificação, a apresentação, a análise e interpretação de dados quantitativos, e a utilização desses dados para a tomada de decisões. Nos dias atuais, ela já é considerada como a ciência que estuda os processos de obtenção, organização e análise de dados sobre uma população, e os métodos de tirar conclusões ou fazer predições com base nesses dados.

Estatística descritiva

É aquela que se preocupa com a organização e descrição de dados experimentais.

Estatística indutiva ou inferencial

É aquela que se preocupa com a análise e interpretação de dados experimentais.

Fases do método estatístico

Definição do problema - saber exatamente o que se pretende pesquisar, ou seja, definir corretamente o problema.

Planejamento - determinar o procedimento necessário para resolver o problema; como por exemplo, levantar informações sobre o assunto objeto de estudo. É importante a escolha das perguntas em um questionário, que devem ser fechadas, na medida do possível.

Coleta de dados - o levantamento dos dados pode ser feito de duas maneiras: Censitário e Amostragem. E pode ser classificado, quanto ao tempo, em:

- Contínua (Ex: inflação, desemprego);
- Periódica (Ex: Censo);
- Ocasional (Ex: pesquisa de mercado, eleitoral).

Crítica dos dados - objetiva a eliminação de erros capazes de promover futuros enganos. Faz-se uma revisão critica dos dados, suprimindo os valores estranhos ao levantamento.

Apresentação dos dados - a organização dos dados denomina-se “série estatística”. Sua apresentação pode ocorrer por meio de tabelas e gráficos.

Análise e interpretação dos dados - consiste em tirar conclusões que auxiliem o pesquisador a resolver seu problema, descrevendo o fenômeno através do cálculo de medidas estatísticas, especialmente as de posição e as de dispersão.

População ou universo

É um conjunto de todos os elementos (pessoas, objetos, etc.) que possuem pelo menos uma característica em comum, que os relaciona ao problema que está sendo estudado.

Amostra

São os objetos selecionados da população, os quais se tiverem a mesma chance de serem selecionados, teremos uma amostra aleatória.

Variável

É uma característica ou dado associado a cada elemento da população ou amostra. A Variável apresenta diferentes valores, e, em geral, é denotada por letras maiúsculas: X - peso, Y- altura.

Qualitativa - é aquela que resulta de uma classificação por tipos ou atributos, como nos exemplos:

População: moradores de uma cidade.

Variável: cor dos olhos (pretos, castanhos, verdes, etc.).

População: estudantes de um Estado.

Variável: escolaridade (1º Grau completo, 2º Grau completo, Superior, Pós- Graduação).

Quantitativa - é aquela que pode assumir valores numéricos, os quais podem ser obtidos através de uma contagem ou mensuração.

Discreta: seus valores são obtidos por um processo de contagem. Por exemplo, quantidade de aluno por sala, quantidade de falhas por peça em determinada linha de produção.

Contínua: seus valores são obtidos por um procedimento de mensuração, podendo assumir quaisquer valores, em um intervalo dos números reais, como, por exemplo: temperatura, altura, peso, etc.

Tabelas ou séries

Consiste em dispor os dados em linhas e colunas, distribuídos de modo ordenado, segundo algumas regras práticas obedecendo à resolução nº 886/66, de 26 de setembro de 1966, do Conselho Nacional de Estatística. De acordo com o fenômeno, local e época, classificam-se em:

a) Série temporal (histórica ou cronológica): os dados são observados segundo a época de sua ocorrência. Ex: óbitos infantis no Piauí de 2001 a 2004.

b) Série geográfica (espaciais, territoriais ou de localização): os dados são observados, segundo o local onde ocorreram. Ex: óbitos no Piauí, segundo microrregiões, em 2004.

c) Série especificada (categórica): os dados são agrupados, segundo a modalidade (espécie) de ocorrência.

d) Série mista ou de dupla entrada: é a fusão de duas ou mais séries simples. Ex: óbitos infantis no Piauí, segundo microrregiões, no período de 2001 a 2004.

Somatório

Indicado pela letra grega sigma maiúscula (∑), facilita a indicação e a formulação de medidas, bem como algumas operações algébricas desenvolvidas pela Estatística.

Define-se genericamente:

$$\sum_{i=1}^{n} X_i = X_1+X_2+...+X_n.$$

Parâmetro

Trata-se de medida numérica que descreve uma característica da população.

Estimador

É uma medida numérica, que descreve uma característica da amostra.

População	Amostra
μ	$\bar{x}$
σ^2	S^2
σ	s
Parâmetro populacional desconhecido	Estimadores amostrais

Como você pode organizar os dados de uma forma mais eficiente e com maior quantidade de informações?

Distribuição de Frequência

Dados brutos ou tabela primitiva

São os dados apresentados desordenadamente, ou seja, da forma que foram coletados; ou ainda tabela, cujos elementos não foram numericamente organizados.

Vamos supor que fizemos uma coleta de dados relativos às estaturas de 40 alunos do Curso de Sistema de Informação da UAPI.

Tabela 1.1 - Estatura de 40 Alunos do Curso de Sistema de Informação da UAPI

66	60	61	50	62	60	65	67	64	160
62	61	68	63	56	73	60	55	64	168
55	52	63	60	55	55	69	51	70	164
54	61	56	72	53	57	56	58	58	161

Fonte: Martins (2009).

a) Rol

É uma lista em que os dados estão dispostos em ordem crescente, ou seja, é a tabela obtida após a ordenação dos dados.

Tabela 1.2 - Estatura de 40 Alunos do Curso de Sistema de Informação da UAPI

50	154	155	157	160	161	162	164	166	169
51	155	156	158	160	161	162	164	167	170
152	155	156	158	160	161	163	164	168	172
153	155	156	160	160	161	163	165	168	173

Fonte: Martins (2009).

b) Tabela de frequências

As tabelas de frequências podem representar tanto valores individuais como valores agrupados em classes.

Distribuição de frequência de dados tabulados não agrupados em classes

Aparecem individualmente, ou seja, uma tabela de frequências de valores não agrupados em classes*. Ex: Uma empresa fabricante de instrumento de precisão está interessada em saber o número de aparelhos defeituosos, rejeitados pela seção encarregada do controle de qualidade. Os dados, fornecidos por essa seção referem-se ao período de 1971 a 1974.

Tabela 1.3 – Empresa X - Número Mensal de Aparelhos Defeituosos

Mês / Ano	J	F	M	A	M	J	J	A	S	O	N	D
1971	6	2	5	6	0	8	7	6	3	4	5	8
1972	10	9	7	6	3	4	6	4	5	4	0	1
1973	3	6	7	9	3	1	4	6	5	3	5	4
1974	7	2	5	8	6	4	2	5	1	6	5	2

Fonte: Toledo (1985)

Abaixo estão os valores da Variável nº mensal de aparelhos defeituosos, dispostos em uma tabela de distribuição de frequências.

Tabela 1.4 – X – Número Mensal de Aparelhos Defeituosos

i	Nº de Aparelhos com defeito	Número de Meses (fi)
1	0	2
2	1	3
3	2	4
4	3	5
5	4	7
6	5	8
7	6	9
8	7	4
9	8	3
10	9	2
11	10	1
		n = 48

Fonte:Toledo (1985).

Na primeira coluna está o índice i, onde aparecem os números correspondentes à ordem dos valores da variável. Na segunda coluna está Xi, correspondente aos valores da variável; e na última coluna está fi, correspondente às frequências, que são os resultados numéricos provenientes da contagem. A soma das frequências é sempre igual ao número total de valores observados.

$$\sum_{i=1}^{k} fi = n.$$

Onde, k é o extremo superior do intervalo de valores do índice i,
fi é o número de observações de um valor,
n é o número total de valores observados.

Distribuição de frequências de dados agrupados em classes

Primeiramente, iremos estudar os elementos de uma distribuição de frequências; posteriormente, os passos para a sua construção. São estes os principais elementos:

Frequência simples absoluta – é o número de observações correspondentes a uma classe. Essa frequência é simbolizada por (fi).

- Amplitude total – é a diferença entre o maior e o menor valor observado da Variável em estudo. Essa amplitude é indicada por (At).
- Classe – é cada um dos grupos de valores em que se subdivide a amplitude total do conjunto de valores observados da variável. Essa classe é indicada por (k).

Para se determinar o nº de classes há diversos métodos. A regra de Sturges é um dos métodos que estabelece o número de classes, que é igual a:

$$k = 1 + 3{,}3.\log n$$

k= é o nº de classes
n = nº total de observações

Truman L. Kelley, em The Grouping Data for Graphic Portrayal,

PARA REFLETIR

É verdade que essas fórmulas não nos levam a uma decisão final; esta vai depender, na realidade, de um julgamento pessoal, que deve estar ligado à natureza dos dados, da unidade usada para expressá-los, do objetivo que se tem em vista, procurando, sempre que possível, evitar classe com frequência nula ou com frequência relativa muito exagerada.

sugere os seguintes números de classe, com base no número total de observações, para efeito de representação gráfica:

n	5	10	25	50	100	200	500	1000
k	2	4	6	8	10	12	15	15

$K = \sqrt{n}$, essa última é também muito utilizada.

- Limites de classes – Os limites de classes são seus valores extremos. O limite inferior da classe (li) e o limite superior (Li).

- Amplitude de um intervalo de classe – É a medida que define a classe. Ela é obtida pela diferença entre os limites superior e inferior dessa classe e indicada por (Ac), assim:

$$Ac = Li - li$$

Usaremos esta notação: li←Li (limite inferior incluso ao limite superior excluso)

- Amplitude total da amostra (At) – É a diferença entre o maior e o menor valor da amostra.

At = X(máx) – X(mín), onde X(máx) é o maior valor da amostra e X(mín) é o menor valor da amostra.

- Amplitude total da distribuição – É a diferença entre o limite superior máximo Li(máx) e o limite inferior mínimo li(mín).

At = Li(Max) – li(mín), onde Li(máx) é o limite superior máximo da distribuição e li(mín) é o limite inferior mínimo da distribuição.

- Ponto médio de uma classe (xi) – É o ponto que divide o intervalo de classe em duas partes iguais.

xi = (Li + li) / 2, onde xi indica o ponto médio da classe, para dados agrupados em classes ou o valor da variável para dados não agrupados em classes.

- Tipos de frequências

- Frequência simples absoluta (fi) é o número de repetições de um valor individual ou de uma classe de valores da variável. A soma das frequências simples absoluta em uma tabela é chamada frequência total e corresponde ao número total de observações.

- Frequência simples relativa (fri) é aquela que representa a proporção de observações de um valor individual ou de uma classe, com relação ao número total de observações. Para calcular a frequência relativa, basta dividir a frequência absoluta da classe ou do valor individual pelo número total de observações.

- Frequência acumulada “abaixo de” (Fi abaixo) é a soma da frequência simples absoluta de uma classe ou de um valor individual de uma variável com as frequências simples absolutas das classes ou dos valores individuais anteriores. Toda vez que se procura saber quantas observações existem até uma determinada classe ou valor individual recorre-se à frequência acumulada “abaixo de”.

- Frequência relativa acumulada “abaixo de” (Fri) é aquela que se obtém dividindo a frequência acumulada de uma classe ou de um valor individual de uma variável pela frequência total da distribuição.

- Frequência acumulada “acima de” (Fi acima) é aquela que representa o número de observações existentes além do valor de uma variável ou de uma classe; e obtém-se somando a frequência simples absoluta da classe ou do valor individual de uma variável às frequências simples absolutas das classes ou dos valores individuais posteriores.

- Frequência relativa acumulada “acima de” (Fri acima) é aquela que se obtém dividindo a frequência acumulada “acima de” de uma classe ou de um valor individual de uma variável pela frequência total da distribuição.

- Passos para a construção de uma distribuição de frequências de dados agrupados em classes:

1º) Primeiramente, pegam-se os dados brutos e constrói-se o rol;
2º) Calcula-se a amplitude total das amostras(At);
3º) Calcula-se o nº de classes, usando uma das regras estudadas

anteriormente; aqui será usada a Regra de Sturges;

4º) Calcula-se a amplitude do intervalo de classe (Ac);

5º) Constrói-se uma tabela que contenha uma coluna para i (ordem das classes), e outra para os intervalos, para as frequências simples absoluta, para o ponto médio das classes(xi), para as frequências simples relativas(fri), frequências acumuladas “abaixo e acima de” e as frequências acumuladas relativas “abaixo e acima de”.

Agora temos uma tabela de frequências que contêm todas as informações da variável em estudo. Vejamos o exemplo referente à Estatura de 40 Alunos do Curso de Sistema de Informação da UAPI (tabela 1.5).

Cálculo da amplitude total da Amostra:

At = Li(máximo) – li(mín) = 173 – 150 = 23 cm

Cálculo do nº de classes (i) pela Regra de Sturges:

i = 1 + 3,3. logn, como n= 40, temos:

i = 1 + 3,3. log40

i = 1 + 3,3. 1,6020599

i = 1 + 5,2867979

i = 6,28679

aproximamos para i $\cong$ 6

Cálculo da amplitude de classe (Ac)

Ac = 23/6 = 3,8333...

aproximamos para Ac $\cong$ 4 , isto é, 6 classes de intervalos iguais a 4.

Podemos agora dar a distribuição de frequências das estaturas dos 40 alunos do Curso de Sistema de Informação da UAPI à seguinte representação tabular técnica:

Tabela 1.5 - Estaturas de 40 Alunos do Curso de Sistema de Informação da UAPI

i	ESTATURAS (cm)	Nº de Alunos (fi)
1	150 ← 154	4
2	154 ← 158	9
3	158 ← 162	11

SAIBA MAIS

Os intervalos de classe devem ser escritos, de acordo com a Resolução 886/66 do IBGE, em termos de desta quantidade até menos aquela, empregando, para isso, o símbolo ← (incluso de li e exclusão de Li). Ex: no intervalo 154←158 é um intervalo fechado à esquerda e aberto à direita, tal que: 154≤ x < 158.

i	ESTATURAS (cm)	Nº de Alunos (fi)
4	162 ⊢ 166	8
5	166 ⊢ 170	5
6	170 ⊢ 174	3
		$\sum fi = 40$

Fonte: Martins (2009).

Na 1ª classe, contamos no rol, quantos alunos têm estaturas entre 150 incluso e 154 excluso, e o resultado foram 4;

Na 2ª, entre 154 incluso a 158 excluso, 9;

Na 3ª, entre 158 incluso e 162 excluso, 11;

Na 4ª, entre 162 incluso e 166 excluso, 8;

Na 5ª, entre 166 incluso e 170 excluso, 5;

Na 6ª, entre 170 incluso e 174 excluso, 3.

Considerando a tabela 1.5 acima, podemos montar a seguinte tabela com as frequências estudadas:

Tabela 1.6 - Estaturas de 40 Alunos do Curso de Sistema de Informação da UAPI.

i	Estatura (cm)	Nº de Alunos(fi)	xi	fri	Fi ↓	Fri ↓	Fi ↑	Fri ↑
1	150⊢154	4	152	4/40=0,100	4	4/40=0,100	40	40/40=1,00
2	154⊢158	9	156	9/40=0,225	13	13/40=0,325	36	36/40=0,900
3	158⊢162	11	160	11/40=0,275	24	24/40=0,600	27	27/40=0,675
4	162⊢166	8	164	8/40=0,200	32	32/40=0,800	16	16/40=0,400
5	166⊢170	5	168	5/40=0,125	37	37/40=0,925	8	8/40=0,200
6	170⊢174	3	172	3/40=0,075	40	40/40=1,00	3	3/40=0,075
		$\sum fi=40$		$\sum fri=1,000$				

Fonte: Martins (2009).

Para obtermos o ponto médio de cada classe, pegamos a amplitude de classe e dividimos por 2, obtemos 2. Somamos 2 a cada limite inferior de cada classe. Mas para obtermos as frequências acumuladas abaixo (Fi↓), pegamos F_1↓=4; F_2↓=4+9=13; F_3↓=13+11=24; F_4↓=24+8=32; F_5↓=32+5=37 e F_6↓=37+3=40; e as frequências acumuladas acima de (Fi↑), assim: F_1↑=40; F_2↑=40-4=36; F_3↑=36-9=27; F_4↑=27-11=16; F_5↑=16-8=8; e F_6↑=8-5=3.

O conhecimento dos vários tipos de frequência ajuda-nos a responder muitas questões com relativa facilidade, como as seguintes:

a) Quantos alunos têm estatura entre 154 cm, inclusive, e 158 cm?

Resposta: 9 alunos; são os valores que formam a segunda classe (f2=9).

b) Qual a percentagem de alunos cuja estatura é inferior a 154 cm?

Esses valores são os que formam a primeira classe (fri=0,100); obtemos a resposta multiplicando a frequência relativa por 100:

$$0{,}100.\ 100 = 10.$$

Logo, a percentagem de alunos é 10%.

c) Quantos alunos têm estatura abaixo de 162 cm?

Basta procurarmos a frequência acumulada abaixo (Fi↓), referente à 3ª classe, ou seja, Fi↓= 24. Portanto, 24 alunos têm estatura abaixo de 162 cm.

d) Quantos alunos têm estatura acima de 158 cm, inclusive?

Basta procurarmos a frequência acumulada "acima" (Fi↑), referente à 3ª classe, ou seja, Fi↑=27. Portanto, 27 alunos.

Representação gráfica de uma distribuição de frequência

Uma distribuição de frequência pode ser representada graficamente pelo histograma, pelo polígono de frequência e pelo polígono de frequência acumulada.

Construímos qualquer um dos gráficos mencionados utilizando o primeiro quadrante do sistema de eixos coordenados cartesianos ortogonais. Na linha horizontal (eixo das abscissas) colocamos os valores da variável e na linha vertical (eixo das ordenadas), as frequências.

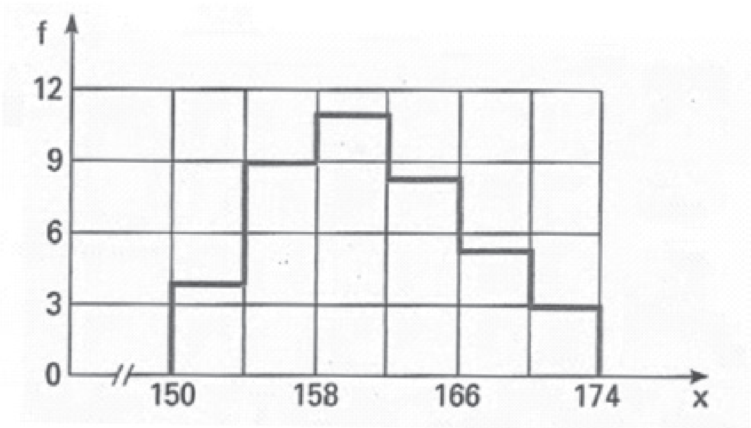
Histograma

O histograma é formado por um conjunto de retângulos justapostos, cujas bases se localizam sobre o eixo horizontal, de tal modo que seus pontos médios coincidam com os pontos médios dos intervalos de classe.

As larguras dos retângulos são iguais às amplitudes dos intervalos de classe.

As alturas dos retângulos devem ser proporcionais às frequências das classes, sendo a amplitude dos intervalos igual. Isso nos permite tomar as alturas numericamente iguais às frequências.

À distribuição da tabela 1.5 (p. 8) corresponde o seguinte histograma:



Importante:

O histograma goza de uma propriedade da qual faremos considerável uso: a área de um histograma é proporcional à soma das frequências.

No caso de usarmos as frequências relativas, obtemos um gráfico de área unitária.

Quando queremos comparar duas distribuições, o ideal é fazê-lo pelo histograma de frequências relativas.

Polígono de frequência

Para realmente obtermos um polígono (linha fechada), devemos completar a figura ligando os extremos da linha obtida aos pontos médios da classe anterior à primeira, e da posterior à última da distribuição.

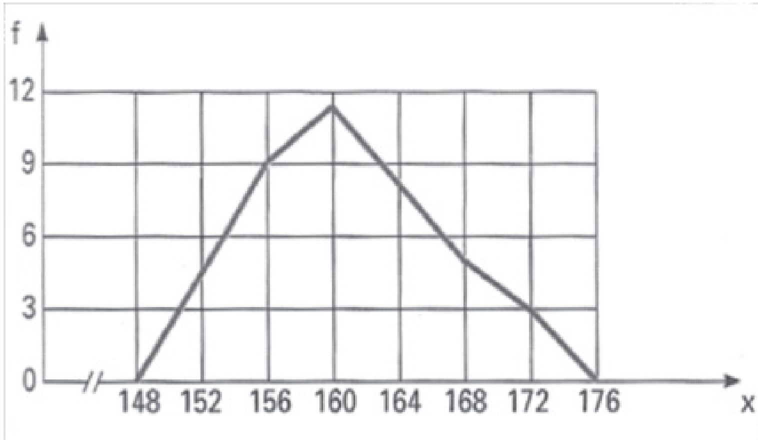
SAIBA MAIS

Veja a construção do histograma com Excel, p. 48, do livro Estatística, usando Excel do Lapponi.

SAIBA MAIS

Polígono de frequência* é um gráfico de análise no qual as frequências das classes são localizadas sobre perpendiculares levantadas nos pontos médios das classes.

A distribuição da tabela 1.5 corresponde ao seguinte polígono de frequência*:



Importante:

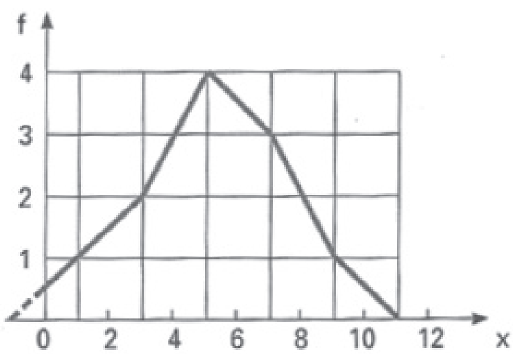
No caso de termos uma variável essencialmente positiva, cuja distribuição se inicia no valor zero, devemos considerar um intervalo anterior localizado no semieixo negativo. Porém, consideraremos apenas a parte positiva do segmento que liga o ponto médio desse intervalo com a frequência do intervalo 0←... .

Exemplo:

Tabela 1.7 – Classes.

i	classes	fi
1	0 ← 2	1
2	2 ← 4	2
3	4 ← 6	4
4	6 ← 8	3
5	8 ← 10	1
Total		**11**

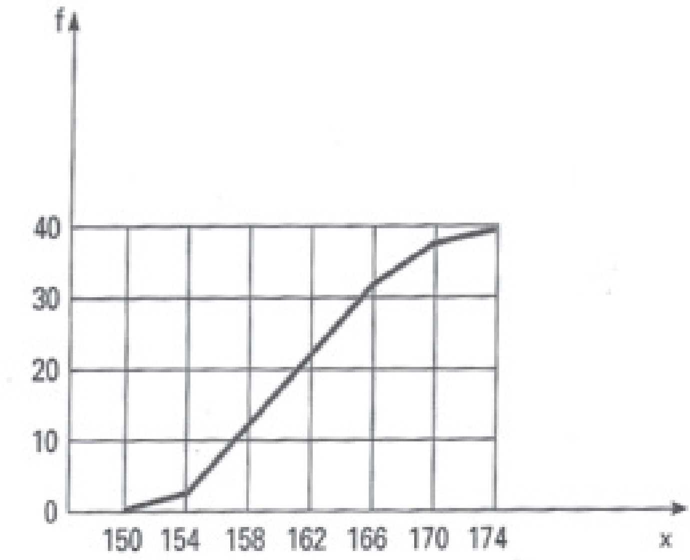
Fonte: CRESPO (2002).



Polígono de frequência acumulada

O polígono de frequência acumulada é traçado marcando-se as frequências acumuladas sobre perpendiculares ao eixo horizontal, levantadas nos pontos correspondentes aos limites superiores dos intervalos de classe.

Assim, à distribuição da tabela 1.6 (Fi↓) corresponde o seguinte polígono de frequência acumulada:



Exercícios Propostos 01

Nas questões de 01 a 08, assinale a alternativa correta:

1. População ou Universo é:

a) Conjunto de pessoas.

b) Conjunto de indivíduos apresentando uma característica especial.

c) Conjunto de todos os indivíduos apresentando uma característica comum objeto de estudo.

2. Os fenômenos de massa são:

a) Aqueles em que as características observadas não valem para os indivíduos.

b) Aqueles que não podem ser definidos por uma simples observação.

c) Aqueles que irão compor os fenômenos coletivos.

3. A variável é discreta quando:

a) Dados dois valores reais, podemos encontrar pelo menos um valor entre eles.

b) A menor diferença não nula entre dois valores dados for finita.

c) Dados dois valores reais, a diferença entre eles é zero.

4. As fases principais do método estatístico são:

a) A coleta dos dados, amostragem, apresentação tabular e apresentação gráfica e definição dos problemas.

b) Amostragem, apresentação tabular, apuração dos dados, interpretação dos dados e planejamento.

c) Definição do problema, planejamento, coleta dos dados, apuração, apresentação dos dados, análise e interpretação dos dados.

5. A série estatística é chamada cronológica quando:

a) O elemento variável é o tempo.

b) O elemento variável é o local.

c) Não tem elemento variável.

6. A amplitude total é:

a) A diferença entre dois valores quaisquer de um conjunto de valores.

b) A diferença entre o maior e o menor valor observado da variável dividido por 2.

c) A diferença entre o maior e o menor valor observado da variável.

7. Para obter o ponto médio de uma classe:

a) Soma-se ao seu limite superior metade de sua amplitude.

b) Soma-se ao seu limite inferior metade de sua amplitude.

c) Soma-se ao seu limite inferior metade de sua amplitude e divide o resultado por 2.

8. Frequência simples absoluta de um valor da variável é:

a) O número de repetições desse valor.

b) A porcentagem de repetições desse valor.
c) O número de observações acumuladas até esse valor.

9. A tabela abaixo apresenta as vendas diárias de um determinado aparelho elétrico, durante um mês, por uma firma comercial:

14	12	11	13	14	13
12	14	13	14	11	12
12	14	10	13	15	11
15	13	16	17	14	14

Fonte: Martins (2009).

Forme uma distribuição de frequência sem intervalos de classe.

10. A tabela seguinte representa as alturas (em cm), de 40 alunos de uma classe.

162		163	148	166	169	154	170	166
164		165	159	175	155	163	171	172
170	157	176	157	157	165	158	158	
160	158	163	165	164	178	150	168	
166	169	152	170	172	165	162	164	

Fonte: Toledo (1985).

a) Calcular a amplitude total.
a) Admitindo-se 6 classes, qual a amplitude do intervalo de classe?
c) Construir uma tabela de frequência das alturas dos alunos, admitindo que o limite da 1ª classe seja 148 cm.
d) Determinar os pontos médios das classes.

11. A tabela a seguir representa os salários pagos a 100 operários da empresa GLT & Cia.

Nº de Salários Mínimos	Nº de Operários (fi)
0 ← 2	40
2 ← 4	30
4 ← 6	10
6 ← 8	15
8 ← 10	5
Total	**100**

Fonte: Toledo (1985).

a) Determinar frequências absolutas acumuladas (abaixo e acima de); frequências simples relativas e frequências relativas acumuladas (abaixo e acima de).

b) Quantos operários ganham até dois salários mínimos?

c) Quantos operários ganham até 5 salários mínimos?

d) Qual a porcentagem de operários com salário entre 6 e 8 salários mínimos?

e) Qual a porcentagem de operários com salários inferior a 4 salários mínimos?

12. Os dados seguintes representam 20 observações relativas ao índice pluviométrico em determinados municípios do Estado.

144			152	159	166	160	151
157	146	154	145				
141	150	142	146	142	141	141	150
143	158						

Fonte: Toledo (1985).

a) Determinar o número de classes pela regra de Sturges.

b) Construir a tabela de frequências absolutas simples.

c) Determinar as frequências acumuladas (abaixo e acima de).

d) Determinar as frequências simples relativas.
e) Determinar as frequências relativas acumuladas (abaixo e acima de).
13. Considere a seguinte tabela:

Classes	Frequências (fi)
2,75 ← 2,80	2
2,80 ← 2,85	3
2,85 ← 2,90	10
2,90 ← 2,95	11
2,95 ← 3,00	24
3,00 ← 3,05	14
3,05 ← 3,10	9
3,10 ← 3,15	8
3,15 ← 3,20	6
3,20 ← 3,25	3
Total	**90**

Fonte: Toledo (1985).

Identificar os seguintes elementos da tabela:

a) Frequência simples absoluta da quinta classe.
b) Frequência total.
c) Limite inferior da sexta classe.
d) Limite superior da quarta classe.
e) Amplitude do intervalo de classe.
f) Amplitude total.
g) Ponto médio da terceira classe.
h) Número total de classe.
i) Frequência absoluta acumulada além da sexta classe.
j) Porcentagem de valores iguais ou maiores que 3,20.

Resumindo

Nesta unidade iremos abordar os principais conceitos básicos de Estatística e um tipo de tabelas que condensa uma coleção e dados, conforme as frequências ou repetições de seus valores. Uma das vantagens das tabelas estatísticas é a de condensar, de forma consistente, as informações necessárias ao estudo desejado. Isto porque, frequentemente, o estudo de um determinado fenômeno requer a coleta de uma grande massa de dados numéricos, difícil de ser tratada, se esses dados não forem organizados e condensados em uma tabela. Agora, seguiremos os passos para a construção de uma tabela de distribuição de frequências.

UNIDADE 2

Medidas de Posição

Objetivos:

- Identificar as medidas de posição;
- Calcular a média aritmética de dados não tabulados e tabulados e interpretar os resultados;
- Calcular a moda, mediana de dados não tabulados e tabulados e interpretar os resultados;
- Identificar as medidas de tendência central;
- Definir e calcular os Quartis, Decis e Centis.

2 MEDIDAS DE POSIÇÃO

Média

A medida de tendência central mais comumente usada para descrever de modo resumido uma distribuição de frequência é a média, ou mais apropriadamente, a média aritmética. Há vários tipos de médias: média aritmética, média geométrica, média harmônica, média quadrática, média cúbica, média biquadrática, porém daremos maior ênfase à média aritmética.

Média aritmética

Símbolo $\overline{x}$
(lê-se "x traço" ou "x barra")

A média aritmética* de um conjunto de números pode ser de dois tipos: simples ou ponderada.

Média aritmética simples

a_1) dados não agrupados

A média aritmética simples de um conjunto de números é igual ao quociente entre a soma dos valores do conjunto e o número total de valores.

Suponhamos que em um escritório de Consultoria a Empresas há 5 contínuos que recebem os seguintes salários mensais: R$ 800,00; R$ 780,00; R$ 820,00; R$ 810,00 e R$ 790,00. A média aritmética dos salários ou o salário médio mensal dos contínuos desse escritório será de 800,00 reais, de acordo com a definição.

$$\overline{x} = (800+780+820+810+780)/5 = 800$$

SAIBA MAIS

A média aritmética*, ou simplesmente média de um conjunto de n observações, x_1, x_2, ..., x_n é definida como:

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

onde:
xi= valor genérico da observação
n= número de observações

A média aritmética simples será calculada sempre que os valores não estiverem tabulados, ou seja, quando aparecerem representados individualmente, como é o caso, por exemplo, dos dados brutos. No exemplo acima, a variável x representa os salários dos contínuos. Consequentemente,

$X_1 = 800$
$X_2 = 780$
$X_3 = 820$ $\quad$ $i = 1, 2, 3, 4, 5$
$X_4 = 810$ $\quad$ $n = 5$
$X_5 = 790$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{\sum_{i=1}^{5} x_i}{5}$$

Média aritmética ponderada

Exemplo: Um professor de Estatística realiza quatro provas por semestre em sua matéria, atribuindo a cada uma delas os seguintes pesos: 1, 2, 3, 4. Suponhamos que um dos alunos recebeu as seguintes notas 8, 7, 9 e 9, nessa ordem, sua nota final foi a média aritmética ponderada* 8,5, obtida da seguinte maneira:

Média final = [(8x1) + (7x2) + 9x3) + (9x4)] / (1+2+3+4)
= [8 + 14 + 27 + 36] / 10 = 85/10 = 8,5

No exemplo apresentado, os pesos dos valores da variável são fixados previamente para efeito de cálculo. Tratando-se, todavia, de distribuições de frequências, os pesos dos valores da variável não são atribuídos arbitrariamente, mas correspondem ao número de vezes que cada valor ocorre.

Assim, por exemplo, admitamos que as notas atribuídas a vinte alunos em um teste de Estatística sejam as seguintes, dispostas em ordem crescente:

SAIBA MAIS

A média é considerada ponderada* quando os valores do conjunto tiverem pesos diferentes. Obtém-se através do quociente entre o produto dos valores da variável pelos respectivos pesos e a soma dos pesos.

4	5	5	5	5
5	6	6	6	6
6	6	7	7	7
7	7	8	8	8

A nota obtida mediante a utilização da fórmula abaixo será:

$$\bar{x} = \left(\sum_{i=1}^{n} x_i\right) / n$$

$\bar{x}$ = (4+5+5+5+5+5+6+6+6+6+6+6+7+7+7+7+7+8+8+8)/20
= 124 / 20 = 6,2, logo $\bar{x}$ = 6,2

Portanto, a média aritmética simples das notas é 6,2. Como os valores da variável aparecem repetidos, é possível adotar o número de observações ou frequência de cada um deles como peso ou fator de ponderação. Assim, por exemplo, a nota 7 aparece cinco vezes. É indiferente, portanto, para efeito de cálculo da média somar o número 7 cinco vezes ou multiplicar esse valor por cinco.

$$7+7+7+7+7 = 5 \times 7 = 35.$$

É possível proceder da mesma forma para os demais valores de variável, como se observa no desenvolvimento seguinte:

$$\bar{x} = \frac{(4+5+5+5+5+5+6+6+6+6+6+6+7+7+7+7+7+8+8+8)}{1 + 5 + 6 + 5 + 3},$$

ou

$$\bar{x} = \frac{(4 \times 1) + (5 \times 5) + (6 \times 6) + (7 \times 5) + (8 \times 3)}{20}$$

O numerador da primeira expressão representa a operação necessária para o cálculo da média aritmética simples, de acordo com a fórmula (1). O denominador é o número total de observações.

Já o numerador da segunda expressão apresenta o procedimento para o cálculo da média ponderada, onde cada valor da variável é multiplicado pela respectiva frequência. O número 4 apareceu uma vez, o

5 e o 7 apareceram cinco vezes, o 6 apareceu seis vezes e o 8 três vezes. Ao invés de considerar cada nota do aluno individualmente, como é feito para o cálculo da média aritmética simples, toma-se o valor tantas vezes quantas ele houver ocorrido. O denominador da segunda expressão é calculado pela soma das frequências de cada valor da variável, o que equivale à frequência total, ou número total de valores observados. Retomando-se os cálculos:

$$\overline{x} = \frac{(4 \times 1) + (5 \times 5) + (6 \times 6) + (7 \times 5) + (8 \times 3)}{1 + 5 + 6 + 5 + 3}$$

$$\overline{x} = ((4+25+36+35+24))/(1+5+6+5+3) = 124/20 = 6{,}2$$

Esse resultado é o mesmo que o obtido pelo emprego da fórmula (1), porque o princípio dos métodos é o mesmo, diferindo apenas a forma de cálculo.

Genericamente, se os valores $x_1, x_2, ..., x_k$ ocorrerem $f_1, f_2, ..., f_k$ vezes, respectivamente, a média aritmética do conjunto será calculado por:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i f_i}{n} \qquad (2)$$

Em que:

Xi = valores da variável ou pontos médios de classe

$\sum_{i=1}^{K}$ = n = número total de observações

k = número de valores individuais diferentes da variável ou número de classes.

Na tabela 2.1, condensamos os resultados do exemplo acima.

Tabela 2.1 - Notas de 20 Alunos de Estatística (fictícias)

x_i	f_i	$X_i.f_i$
4	1	4 x 1 = 4
5	5	5 x 5 = 25
6	6	6 x 6 = 36
7	5	7 x 5 = 35
8	3	8 x 3 = 24
	$\sum_{i=1}^{5} f_i = 20 = n$	$\sum_{i=1}^{5} x_i . f_i = 124$

Fonte: Martins (2009).

Normalmente, para o cálculo da média aritmética ponderada, recorre-se a uma tabela desse tipo, o que possibilita maior rapidez de operação e organização dos valores.

Quando os valores estão agrupados em classes, a tabela requer mais uma coluna necessária para dispor os pontos médios de classes, que veremos a seguir.

b_1) Dados agrupados em classes

Neste caso, convencionamos que todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio, e determinamos a média aritmética ponderada por meio da fórmula (2).

$$\overline{x} = \frac{\sum x_i f_i}{n} \quad (3)$$

Onde xi é o ponto médio da classe.

Consideremos os dados da tabela 1.5 referentes à estatura de 40 alunos do Curso Sistemas de Informação da UAP:

Tabela 2.2 - Estatura de 40 Alunos do Curso de Sistema de Informação da UAPI

i	ESTATURAS (cm)	Nº de Alunos (fi)
1	150 ← 154	4
2	154 ← 158	9
3	158 ← 162	11
4	162 ← 166	8
5	166 ← 170	5
6	170 ← 174	3
		$\sum fi = 40$

Fonte: Martins (2009).

Pela mesma razão do caso anterior, vamos, inicialmente, abrir uma coluna para os pontos médios (x_i) e outra para os produtos $x_i f_i$:

Tabela 2.3 - Estatura de 40 Alunos do Curso de Sistema de Informação da UAPI.

i	ESTATURA (cm)	f_i	x_i	$x_i.f_i$
1	150←154	4	152	608
2	154←158	9	156	1404
3	158←162	11	160	1760
4	162←166	8	164	1312
5	166←170	5	168	840
6	170←174	3	172	516

i	ESTATURA (cm)	f_i	x_i	$x_i.f_i$
		$\sum f_i =$ 40		$\sum x_i.f_i = 6440$

Fonte: Martins (2009).

Como, neste caso:

o $\sum x_i.f_i = 6440$ e o $\sum f_i = n = 40$ e $\overline{x} = \dfrac{\sum x_i.f_i}{n}$

temos:

$$\overline{x} = \frac{6440}{40} = 161 \text{ cm.}$$

Isto quer dizer que, a estatura média dos alunos do Curso de Sistema de Informação da UAPI é de 161 cm.

Propriedades da média

P1) A soma algébrica dos desvios de um conjunto de números tomados em relação à média aritmética é zero. Simbolicamente:

$$\sum d_i = \sum (x_i - \overline{x}) = 0 \qquad (4)$$

$$\sum_{i=1}^{n} d_i .f_i = \sum (x_i - \overline{x}) .fi = 0 \qquad (5)$$

Em que,

$d_i = x_i -- \overline{x}$

e i = {1,2,3,..., n} e xi é o valor da variável para dados brutos e o ponto médio das classes para dados tabulados.

A comprovação dessa propriedade é simples:

Para dados brutos:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (xi - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x}$$

Como $\overline{x}$ é uma constante, para um dado conjunto de valores, então:

$$\sum_{i=1}^{n} \overline{x} = n\,\overline{x}$$

mas

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{e, assim, } n\overline{x} = n\,\frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i$$

Então,

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

Exemplo 1:

Consideremos os salários recebidos pelos cinco contínuos.

X = {780, 790, 800, 810, 820}

A média dos salários é $\overline{x}$ = 800, e a soma dos desvios será zero, como na tabela seguinte:

Tabela 2.4

x_i	$d_i = x_i - \overline{x}$
780	780 – 800 = - 20
790	790 – 800 = - 10

x_i	$d_i = x_i - \bar{x}$
800	800 – 800 = 0
810	810 – 800 = 10
820	820 – 800 = 20
	$\sum_{i=1}^{5} d_i = \sum_{i=1}^{5} (x_i - 800) = 0$

Fonte: Toledo (1985)

Logo, o $\sum_{i=1}^{5} d_i = 0$.

Exemplo 2:

Com os dados da tabela 2.1, comprovar a primeira propriedade da média.

Notas x_i	Número de alunos f_i
4	1
5	5
6	6
7	5
8	3
	n = 20

FFonte: Toledo (1985)

A média aritmética das notas é $\bar{x} = 6{,}2$, conforme visto anteriormente. A tabela 2.5 permite operar mais rapidamente.

Tabela 2.5

x_i	f_i	$x_i f_i$	$d_i = x_i - 6,2$	$d_i f_i = (x_i - 6,2) \cdot f_i$
4	1	4	4 – 6,2 = - 2,2	(-2,2). 1 = - 2,2
5	5	25	5 – 6,2 = -1,2	(-1,2). 5 = -6,0
6	6	36	6 – 6,2 = -0,2	(-0,2). 6 = -1,2
7	5	35	7 – 6,2 = 0,8	(0,8). 5 = 4,0
8	3	24	8 – 6,2 = 5,4	(1,8). 3 = 5,4
	n = 20	$\sum x_i f_i = 124$		$\sum_{i=1}^{5} d_i f_i = 0$

Fonte: Toledo (1985)

P2) Segunda Propriedade

A soma dos quadrados dos desvios tomados em relação à média aritmética é um mínimo.

Para dados brutos

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 < \sum_{i=1}^{n} (x_i - x_0)^2 \qquad (6)$$

Para dados tabulados

$$\sum_{i=1}^{k} (x_i - \overline{x})^2 . fi < \sum_{i=1}^{k} (x_i - x_0)^2 . fi \qquad (7)$$

Em que x0 é um valor arbitrário qualquer, muitas vezes chamado de média arbitrária. Neste caso, $x_0 \neq \overline{x}$.

A comprovação da propriedade será feita apenas para a fórmula (6).

Devemos ter:

$$S1 = \sum(x_i - \overline{x})^2 = \sum(x_i^2 + \overline{x}^2 - 2x_i\overline{x}) = \sum x_i^2 + n\overline{x}^2 - 2\sum x_i \overline{x} =$$

$$= \sum x_i^2 + n\,\overline{x}^2 - 2\,(\sum x_i)\,((\sum x_i)/n)\; n/n =$$

$$= \sum x_i^2 + n\,\overline{x}^2 - 2n\,\overline{x}^2 = \sum x_i^2 - n\,\overline{x}^2$$

$$S2 = \sum(x_i - x_0)^2 = \sum(x_i^2 + x_0^2 - 2\,x_i x_0) =$$
$$= \sum x_i^2 + n\,x_0^2 - 2x_0\sum x_i =$$

$$= \sum x_i^2 + nx_0^2 - 2x_0 n\overline{x}$$

Mas

$$S2 - S1 = (\sum x_i^2 + nx_0^2 - 2nx_0\overline{x}\,) - (\sum x_i^2 - n\,\overline{x}^2) =$$

$$= n(x_0^2 - 2x_0\overline{x} + \overline{x}^2) = n\,(x_0 - \overline{x})^2$$

Como $n > 0$ e $(x_0 - \overline{x})^2 > 0$ implica que $n(x_0 - \overline{x})^2 > 0$.

Assim sendo, a diferença

$S2 - S1 = n(x_0 - \overline{x})^2$ é sempre positiva, para $x_0 \neq x$, ou seja,
$S2 - S1 > 0$

Daí, tiramos:

$$S1 < S2 \text{ ou } \sum(x_i - \overline{x}\,)^2 < \sum(xi - x_0)^2$$

Exemplo:

Considere o conjunto de números X = {3, 5, 7, 9, 10}. A média aritmética do conjunto é:

$$\overline{x} = \frac{\sum_{i=1}^{5} x_i}{5} = (3+5+7+9+10)/5 = 34/5 = 6{,}8$$

Considere agora os seguintes valores arbitrários para x_0 = 7 e 5. Os cálculos das somas dos quadrados dos desvios serão feitos com recurso da tabela auxiliar 2.6.

Tabela 2.6

x_i	$(x_i - \bar{x})$ $(x_i - 6{,}8)$	$(x_i - 6{,}8)^2$	$X_0 = 7$ $(x_i - 7)$	$(x_i - 7)^2$	$X_0 = 5$ $(x_i - 5)$	$(x_i - 5)^2$
3	- 3,8	14,44	-4	16	-2	4
5	- 1,8	3,24	-2	4	0	0
7	0,2	0,04	0	0	2	4
9	2,2	4,84	2	4	4	16
10	3,2	10,24	3	9	5	25
		$\sum_{i=1}^{5}(x_i-6{,}8)^2 = 32{,}8$		$\sum_{i=1}^{5}(x_i-7)^2 = 33$		$\sum_{i=1}^{5}(x_i-5)^2 = 45$

Fonte: Toledo (1985)

Como pode ser observado,

$$\sum(x_i - 6{,}8)^2 < \sum(x_i - 7)^2 < \sum(x_i - 5)^2$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$32{,}8 \quad < \quad 33 \quad < \quad 45$$

Para $\bar{x}$ = 6,8.

P3) Terceira propriedade

Média aritmética ponderada de todas as médias. Se n1 números tem média $\bar{x}_1$, n_2 números tem média $\bar{x}_2$, ..., n_k números tem média $\bar{x}_k$, a média do conjunto formado por todos os números é dada pela expressão:

$$\bar{x} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \ldots + \bar{x}_k n_k}{N_1 + n_2 + \ldots + n_k}$$

$$\overline{x} = \frac{\sum_{i=1}^{5} x_i f_i}{\sum_{i=1}^{k} n_i} \qquad (08)$$

A comprovação dessa propriedade é imediata, por isso não a demonstraremos.

Exemplo:

Considere os seguintes conjuntos de números:

$X_1 = \{10, 40, 70, 80\}$
$X_2 = \{5, 15, 20, 30, 40\}$
$X_3 = \{1, 2, 3, 10, 12, 20\}$

Devemos ter:

$$\overline{x}_1 = \frac{(10+40+70+80)}{4} = \frac{200}{4} = 50$$

$$\overline{x}_2 = \frac{(5+15+20+30+40)}{5} = \frac{110}{5} = 22$$

$$\overline{x}_3 = \frac{(1+2+3+10+12+20)}{6} = \frac{48}{6} = 8$$

i = 1, 2, 3.

$$\overline{x} = \frac{\sum_{i=1}^{3} x_i n_i}{3} = \frac{((50x4)+(22x5)+8x6))}{(4+5+6)} =$$

$$\frac{(200+110+48)}{15} =$$

$$\frac{\sum_{i=1} n_i}{} = \frac{358}{15} = 23{,}87$$

Supondo agora que as três séries de números constituem uma única série, teríamos ordenados os valores:

1, 2, 3, 5, 10, 10, 12, 15, 20, 20, 30, 40, 40, 70, 80

Daí calculamos:

$$\overline{x} = \frac{(1+2+3+5+10+10+12+15+20+20+30+40+40+70+80)}{15} =$$

$$= \frac{358}{15} = 23{,}87$$

P4) Quarta propriedade

Somando-se (ou subtraindo-se) um valor constante e arbitrário a cada um dos elementos de um conjunto de números, a média aritmética fica somada (ou subtraída) por essa constante.

Seja $X = \{x_1, x_2, \ldots, x_n\}$ e seja x_0 uma constante escolhida arbitrariamente.

Façamos $Y = X – X_0$

Assim sendo,

$$\overline{x} = \frac{\sum_{i=1}^{n} xi}{n}$$

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} (x_i - x_0)}{n} = \frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sum_{i=1}^{n} x_0}{n}$$

$$= \bar{x} - \frac{nx_0}{n} = \bar{x} - x_0 \text{ , logo}$$

$$\bar{y} = \bar{x} - x_0 \qquad (09)$$

Como pode ser observado, $\bar{x}$, a média de x, ficou subtraída pela constante. Da mesma forma, poderia ser demonstrada que para:

$Y = x + x_0$ teríamos:

$$\bar{y} = \bar{x} + x_0 \qquad (10)$$

P5) Quinta propriedade

Multiplicando-se (ou dividindo-se) cada elemento de um conjunto de números por um valor constante e arbitrário, a média fica multiplicada (ou dividida) por essa constante.

Sejam, $X = \{x_1, x_2, \ldots , x_n\}$ e c uma constante arbitrária. Façamos $Y = c\,X$.

Devemos ter:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} cx_i}{n} = c\,\frac{\sum_{i=1}^{n} x_i}{n} = c\bar{x}$$

$$\bar{y} = c\,\bar{x} \qquad (11)$$

De igual modo poderia ser demonstrado que, para $y = \frac{X}{c}$, teríamos

$$\bar{y} = \frac{\bar{x}}{c} \qquad (12)$$

Exemplo:

Comprovar a quinta propriedade, com os dados da tabela 2.7, fazendo c = 5 e Y = X/5. A comprovação é feita com o auxílio da tabela 2.7.

Tabela 2.7

Classes	f_i	x_i	$x_i.f_i$	$Y_i = \frac{x}{5}$	$Y_i f_i$
10 ⊢ 20	5	15	75	$\frac{15}{5}=3$	3x5 = 15
20 ⊢ 30	10	25	250	$\frac{25}{5}=5$	5x10 = 50
30 ⊢ 40	15	35	525	$\frac{35}{5}=7$	7x15 = 105
40 ⊢ 50	10	45	450	$\frac{45}{5}=9$	9x10 = 90
50 ⊢ 60	5	55	275	$\frac{55}{5}=11$	11x5 = 55
	$\sum_{i=1}^{5} f_i = 45$		$\sum_{i=1}^{5} x_i.f_i = 1575$		$\sum_{i=1}^{5} y_i.f_i = 315$

Fonte: Toledo (1985)

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i.f_i}{45} = \frac{1575}{45} = 35$$

De acordo com a quinta propriedade, devemos ter:

$$\bar{y} = \frac{\bar{x}}{c} = \frac{35}{5} = 7$$

A média dos valores da variável transformada Y é, segundo a definição:

$$\bar{y} = \frac{\sum_{i=1}^{k} y_i.f_i}{} = \frac{315}{45} = 7,$$ onde 7 é igual à média aritmética dos valores de x_0 dividida pela constante arbitrária 5.

SAIBA MAIS

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

Moda

Símbolo: M_o ou $\hat{X}$

A moda* é outra medida de tendência central, mas há outras denominações para designá-la: norma, valor dominante, valor típico. Quando afirmamos que o salário modal de uma empresa é igual a dois mil reais, queremos dizer que esse é o salário percebido pelo maior número de pessoas dessa empresa.

Determinação da moda de valores não tabulados

Considerando um conjunto ordenado de valores, a moda será o valor predominante, o valor mais frequente desse conjunto. Evidentemente, um conjunto de valores pode não apresentar moda, sendo denominado conjunto amodal, caso em que todos os valores da variável em estudo correm com a mesma frequência. Por outro lado, podemos ter conjuntos plurimodais, quando houver mais de um valor predominante.

Exemplo:

Calcular a moda dos seguintes conjuntos de valores:

X = {4, 5, 5, 6, 6, 6, 7, 7, 8, 8}

Y = {4, 4, 5, 5, 6, 6}

Z = {1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6}

W = {1, 2, 3, 4, 5}

A moda de cada um dos conjuntos será:

Moda de X: $M_o = 6$, o valor 6 é o mais frequente.

Moda de Y: Esse conjunto é amodal, pois seus três valores apareceram duas vezes cada um. Não há, portanto, predominância de nenhum valor do conjunto sobre os outros.

Moda de Z: $Mo_1 = 2$ e $Mo_2 = 5$. Trata-se de um conjunto bimodal, uma vez que tanto o valor 3 como o 5 apresentaram o maior número de observações.

Moda de W: Esse é outro conjunto amodal.

Determinação da moda para valores tabulados

a) Dados não agrupados em classes:

Os valores da variável dispostos em uma tabela de frequências podem se apresentar individualmente ou agrupados em classes. No primeiro caso, a determinação da moda é imediata, bastando, para isso, consultar a tabela, localizando o valor que apresenta a maior frequência. Esse valor será a moda do conjunto. Assim, por exemplo, a moda do conjunto apresentado na tabela abaixo é Mo = 3. Esse resultado indica que a rejeição de 3 peças defeituosas por mês foi o resultado mais observado.

Tabela 2.8 – Empresa X - Número de peças de precisão defeituosas devolvidas mensalmente pelo Controle de Qualidade.

Número de Peças com Defeitos x_i	Número de meses f_i
0	2
1	4
2	6
3	8
4	4
5	2
6	1
	$\sum_{i=1}^{7} f_i = 27$

Fonte: Toledo (1985)

b) Dados agrupados em classes:

Tratando-se de uma tabela de frequências com valores tabulados e agrupados em classes, o procedimento não é imediato, sendo disponíveis alguns métodos de cálculo distintos. Qualquer que seja o método adotado, o primeiro passo para determinar a moda é localizar a classe que apresenta a maior frequência, comumente chamada de classe modal.

Moda Bruta

O método mais rudimentar de cálculo da moda em tabelas de frequências com valores agrupados em classes consiste em tomar o ponto médio da classe modal. Esse valor recebe o nome de moda bruta. Examinando os dados da tabela 4.7, por exemplo, podemos dizer que a terceira classe é a classe modal e a moda bruta será seu ponto médio: Mo = 35.

Os dois métodos que apresentaremos a seguir são mais elaborados e baseiam-se não apenas na frequência da classe modal, mas também nas frequências das classes adjacentes.

Método de King

O método de King, para o cálculo da moda elaborada, baseia-se na influência das frequências das classes adjacentes sobre a classe modal. Considerando essas frequências, admite-se que a moda se desloca dentro do intervalo de classe para um determinado ponto (valor), de tal sorte que as distâncias desse ponto aos limites de classe sejam inversamente proporcionais às frequências das respectivas classes adjacentes. Então, quanto maior for à frequência da classe adjacente, menor será a distância do ponto a essa classe. Isto se comprova com a fórmula seguinte:

$$Mo = l_i^* + A^* . \frac{f_{post}}{f_{ant} + f_{pos}} \quad (13)$$

Onde,

l_i^* = limite inferior da classe modal

A^* = amplitude da classe modal

f_{ant} = frequências simples inferior à classe modal

f_{pos} = frequência simples superior à classe modal

Exemplo:

Calcular pelo método de King, a moda dos valores constantes da tabela abaixo:

Tabela 2.9

Classes	Frequências (f_i)
10←20	2
20←30	4
30←40	8
40←50	5
50←60	1
	n = 20

Fonte: Martins (2009).

Solução:

A classe modal é a terceira: 30←40, pois apresenta a maior frequência 8. A moda, segundo a fórmula de King, será:

$$Mo = l_i^* + \frac{f_{post}}{f_{ant} + f_{post}} = 30 + \frac{5}{(4+5)} \cdot 10 = 30 + \frac{50}{9} = 30 + 5{,}556$$

$= 35{,}56$

$Mo = 35{,}56$

A moda bruta desses valores é ligeiramente menor: $Mo_{bruta} = 35$.

Método de Czuber

O método de Czuber, para o cálculo da moda elaborada, leva em consideração não apenas as frequências das classes adjacentes, mas também a frequência da classe modal. O ponto que corresponde à moda divide o intervalo da classe modal em duas partes, as quais são proporcionais às diferenças entre a frequência da classe modal e as das respectivas classes adjacentes. Isto se comprova com a seguinte fórmula:

$$Mo = l_i^* + \frac{d_1}{d_1 + d_2} \qquad (14)$$

Onde,

Mo = moda

$d_1 = f^* - f_{ant}$ e f^* = frequência simples da classe modal.

$d_2 = f^* - f_{post}$

Exemplo:

Determinar a moda, pelo método de Czuber, usando os dados do da tabela 2.9.

De acordo com a fórmula 14,

$l_i^* = 30$

$A^* = L_i^* - l_i^* = 40 – 30 = 10$

$d_1 = f^* - f_{ant} = 8 – 4 = 4$

$d_2 = f^* - f_{pos} = 8 – 5 = 3$

$$Mo = l_i^* + A^* \cdot \frac{d_1}{d_1 + d_2} =$$

$$= 30 + 10 \cdot \frac{4}{4 + 3} = 30 + 5{,}7 = 35{,}7$$

Mo = 35,7

Mediana

Símbolo: Md ou X̃.

A mediana* é a terceira medida de tendência central, ela é considerada uma separatriz, por ser um promédio que divide a distribuição ou conjunto de dados em partes iguais. Trata-se de uma medida muito utilizada na análise de dados estatísticos, especialmente quando se atribui pouca importância aos valores extremos da variável.

SAIBA MAIS

A mediana* (Md) é o valor que divide uma série ordenada de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos, ou seja, 50% dos dados são superiores à mediana e 50% são inferiores.

Elemento mediano

Por ser uma separatriz, isto é, em virtude de a mediana se constituir em um valor que separa a distribuição em partes de tal sorte que uma fração (0,5 ou 1/2) de valores lhe seja inferior e os restantes superiores, podemos concluir que essa medida apresenta um número de ordem. Assim é que, ordenando os valores da série, a mediana é um valor que ocupa uma determinada ordem ou posição na série ordenada. O número que indica a ordem em que se encontra o valor correspondente à mediana é denominado elemento mediano, cujo símbolo é E_{md}.

Determinação da mediana de valores não tabulados

A determinação da mediana de valores não tabulados processa-se a partir de um rol ou lista ordenada dos dados. Podem ocorrer duas hipóteses com relação ao número de observações n: que ele seja ímpar ou par. Veremos os dois casos.

O número de observações é impar

O procedimento para o cálculo da mediana, quando a lista de valores contiver um número ímpar de observações, requer, em primeiro lugar, que se determine a ordem em que se encontra a mediana na série. Deve-se, então, encontrar o valor do elemento mediano, o que é feito da seguinte forma:

$$E_{md} = \frac{n+1}{2} \qquad (15)$$

O passo seguinte será localizar a mediana na lista de valores, de acordo com o resultado obtido no cálculo do elemento mediano.

Exemplo:

Calcular a mediana do seguinte conjunto de números:

$$X = \{2, 3, 6, 12, 15, 23, 30\}$$

Solução:

A primeira providência a ser adotada seria a de ordenar os valores. Neste exemplo, os valores da série já se encontram ordenados.

Em seguida, determinaremos o valor do elemento mediano, utilizando a fórmula (14), uma vez que o número de observações é ímpar (n = 7).

$$E_{md} = \frac{n+1}{2} = \frac{7+1}{2} = 4 \text{ (quarto elemento na série)}$$

Ressalte-se o fato de que o valor 4 é um número ordinal. Assim, Emd = 4 indica que a mediana é o valor que se encontra na quarta posição da lista ordenada de valores, é o quarto número na série.

Finalmente, procuraremos no conjunto qual o valor que se encontra no quarto lugar da lista. Esse número corresponde à mediana do conjunto.

No exemplo:

Md = 12

Observe que existem três valores menores do que doze (2, 3 e 6) e três valores maiores (15, 23 e 30), o que corresponde a cinqüenta por cento (ou metade) de itens maiores e menores do que a mediana.

O número de observações é par

O procedimento para calcular a mediana de um número par de observações é ligeiramente diferente do adotado para o caso em que n é ímpar. O elemento mediano será determinado, agora, através da seguinte expressão:

$$E_{md} = n/2 \qquad (16)$$

Exemplo:

Calcular a mediana do conjunto:

X = {3, 6, 9, 12, 14, 15, 17, 20}.

Solução:

Como vemos n = 8.

O elemento mediano será, de acordo com a fórmula (15),

$$E_{md} = \frac{n}{2} = \frac{8}{2} = 4$$

Seguindo a mesma linha de raciocínio do exemplo anterior, identificaríamos a mediana como o quarto elemento da lista, ou seja, Md = 12. Entretanto, esse procedimento de identificação resultaria em um valor da mediana que contraria a definição, uma vez que não teríamos a mesma proporção de valores menores e maiores do que o valor doze. Haveria, então, três valores menores do que doze (3, 6 e 9) e quatro maiores (14, 15, 17 e 20).

Toda vez que houver um número par de observações, a lista apresentará dois valores centrais e a mediana será determinada calculando a média aritmética deles. No exemplo, teríamos:

$$Md = \frac{12+14}{2} = 13$$

Percebe-se, agora, a ocorrência de igual número de valores maiores (14, 15, 17 e 20) e menores (3, 6, 9 e 12) do que a mediana.

Determinação da mediana de valores tabulados não agrupados em classes

Quando os valores da variável estiverem já tabulados, o procedimento a ser adotado será praticamente idêntico ao anterior.

Em primeiro lugar, deve-se verificar se o número de observações é ímpar ou par, e conforme o caso, aplicar as fórmulas (14) e (15) para o cálculo do elemento mediano.

Em seguida, acrescentamos uma coluna à tabela de frequência original, onde serão determinadas as frequências acumuladas. Comparando o resultado obtido no cálculo do elemento mediano com os valores constantes dessa coluna, determinaremos a mediana.

Exemplo:

Calcular a mediana dos valores apresentados nas tabelas 2.10 e 2.11.

Tabela 2.10

Valores (x_i)	Frequências (f_i)
2	5
3	10
4	15
5	12
6	5
7	3
	n = 50

Fonte: Toledo (1985)

Tabela 2.11

Valores (x_i)	Frequências (f_i)
3	3
4	6
5	9
6	8
7	6
8	3
	n = 35

Fonte: Toledo (1985)

Solução:

O número de observações da variável, conforme a tabela 2.10, é n=50. Assim sendo, fazemos:

$$E_{md}\ \frac{n}{2} = \frac{50}{2} = 25$$

A mediana deverá ser o vigésimo quinto elemento, se levarmos em consideração os valores do conjunto ordenados, ou seja:

{2, 2, 2, 2, 2, 3, 3, 3, ..., 7, 7, 7}

O passo seguinte era o cálculo das frequências acumuladas, o que será feito com o recurso da tabela auxiliar 2.12.

Tabela 2.12

Valores (x_i)	Frequências (f_i)	$F_i \downarrow$
2	5	5
3	10	15
4	15	30
5	12	42
6	5	47
7	3	50
	n = 50	

Fonte: Toledo (1985)

Examinando a coluna das frequências acumuladas, verificaremos até o valor dois, inclusive, existem cinco observações, o que significa dizer que o quinto elemento da lista é igual a dois. Portanto, a mediana não pode ser dois, uma vez que ela equivale ao vigésimo quinto valor, e até dois temos apenas cinco observações. Assim, iremos percorrendo a coluna até encontrar um valor (frequência acumulada) igual ou maior que o elemento mediano. A frequência acumulada seguinte é quinze, que por ser inferior a vinte e cinco, indica que três não é a mediana. Já a frequência acumulada até o valor quatro da variável é superior a vinte e cinco. Portanto, a mediana desse conjunto é:

$$Md = 4$$

Observação: Em virtude de o número de observações serem par, teremos dois valores centrais, que no caso são iguais. Assim,

$$Md = \frac{4+4}{2} = 4$$

O conjunto de observações, neste caso, é ímpar:
n = 35.
O elemento mediano será:

$$E_{md} = (n+1)/2\ (35+1)/2 = 18$$

A mediana será o décimo oitavo valor da lista. A tabela 2.13 mostra que até o cinco, inclusive, temos dezoito observações. Consequentemente,

$$Md = 5$$

Tabela 2.13

Valores (x_i)	Frequências (f_i)	$F_i \downarrow$
3	3	3
4	6	9
5	9	18
6	8	26
7	6	32
8	3	35
	n = 35	

Fonte: Toledo (1985)

SAIBA MAIS

OBS:
Quando n/2 coincidir com a frequência acumulada, neste caso, a mediana será a média aritmética do valor correspondente a tal frequência acumulada e o seguinte.

Neste caso, basta calcular 35/2 = 17,5

A mediana será o valor correspondente à frequência acumulada imediatamente superior a 17,5. Nesse caso é 18,

Logo, a mediana será:

Md = 5

Exemplo:

Calcular a mediana dos valores constantes da tabela 2.14.

Tabela 2.14

Valores (x_i)	Frequências (f_i)	$F_i \downarrow$
3	3	3
4	6	9
5	9	18
6	8	26
7	6	32
8	3	36
	n = 36	

Fonte: Toledo (1985)

Solução:

Temos n = 36 → como n é par, então 36/2 = 18.

Observe que coincide com a frequência acumulada, logo a mediana será calculada pela média aritmética do elemento correspondente a 18 que é o 5, e o seguinte é 6, logo a mediana será:

$$Md = \frac{5+6}{2} = 5,5$$

Determinação da mediana de valores tabulados agrupados em classes. Quando os valores da variável estiverem agrupados em classes, o cálculo da mediana será realizado por interpolação. Tratando-se de dados agrupados, admite-se que os valores da variável na distribuição de frequências se distribuam continuamente. A mediana será, neste caso, o valor da variável, para o qual metade ou cinquenta por cento da frequência total (N/2) ficam situados abaixo e a outra metade acima dele. Geometricamente, isso significa dizer que a mediana é o valor de x (eixo das abscissas) que corresponde à perpendicular que divide o histograma em duas partes que apresentam áreas iguais. Vejamos alguns métodos para o cálculo da mediana, quando os valores estiverem agrupados em classes:

Primeiro método: resolução por fórmula

Neste caso, o problema consiste em determinar o ponto do intervalo em que está compreendida a mediana.

Para tanto, temos inicialmente que determinar a classe na qual se acha a mediana – classe mediana. Tal classe será, evidentemente, aquela correspondente à frequência acumulada imediatamente superior a $\frac{\sum fi}{2}$.

Feito isto, um problema de interpolação resolve a questão, admitindo-se, agora, que os valores se distribuem uniformemente em todo o intervalo de classe. Assim, considerando a distribuição da tabela 2.15, acrescida das frequências acumuladas:

Tabela 2.15

i	Estaturas (cm)	f_i	$F_i \downarrow$
1	150 ⊢ 154	4	4
2	154 ⊢ 158	9	13
3	**158 ⊢ 162**	**11**	**24**
4	162 ⊢ 166	8	32
5	166 ⊢ 170	5	37
6	170 ⊢ 174	3	40
		$\sum f_i = 40$	

Fonte: Martins (2009).

Temos:

$$\frac{\sum fi}{2} = \frac{40}{2} = 20$$

Tendo em vista que há 24 valores incluídos nas três primeiras classes da distribuição, e como pretendemos determinar o valor que ocupa o 20º lugar, a partir do inicio da série, vemos que este deve estar localizado na terceira classe (i=3), supondo que as frequências dessas classes estejam uniformemente distribuídas.

Como há 11 elementos nessa classe e o intervalo de classe é igual a 4, devemos tomar, a partir do limite inferior, a distância:

$$\frac{20-13}{11} \times 4 = \frac{7}{11} \times 4$$

E a mediana será dada por:

$$Md = 158 + \frac{7}{11} \times 4 = 158 + \frac{28}{11} = 158 + 2{,}54$$

$$= 160{,}54$$

Logo:

$$Md = 160{,}5 \text{ cm}$$

Na prática, executamos os seguintes passos:

1º) Determinamos as frequências acumuladas.

2º) Calculamos $\frac{\sum f_i}{2}$.

3º) Marcamos a classe correspondente à frequência acumulada imediatamente superior à $\frac{\sum f_i}{2}$ – classe mediana – e, em seguida, empregamos a fórmula:

$$Md = l_i^* + \frac{\left[\frac{\sum f_i}{2} - Fi\downarrow(ant)\right]}{f^*} \cdot A^* \qquad (15)$$

na qual:

l_i^* é o limite inferior da classe mediana;

$Fi\downarrow$ é a frequência acumulada da classe anterior à classe mediana;

f^* é a frequência simples da classe mediana;

A^* é a amplitude do intervalo da classe mediana.

Tomando como exemplo a distribuição anterior, temos:

$$\frac{\sum f_i}{2} = \frac{40}{2} = 20$$

Logo, a classe mediana é a de ordem 3. Então:

$$Md = 158 + \frac{20\text{-}13.4}{11} = 158 + \frac{28}{11} = 158 + 2{,}54 =$$

$$= 160{,}54,$$

Isto é:

$$Md = 160{,}5 \text{ cm}.$$

Segundo método: por interpolação

Admite-se que as alturas se distribuem continuamente e a altura mediana é o valor tal que metade dos 40 valores lhe seja inferior e metade superior. Para obter a mediana, devemos localizar o vigésimo elemento, já que $E_{md} = 20$. Percorrendo a tabela 4.14 na coluna de frequências acumuladas, verificamos que o vigésimo elemento encontra-se na terceira classe: 158←162. Como até a segunda classe, inclusive, acumulamos 13 observações, precisamos de mais 7 para completar as 20 necessárias. Assim sendo, se tomarmos todo o intervalo da terceira classe, estaremos considerando todas as observações dessa classe (26). Entretanto, interessam-nos apenas 7 observações para atingir o ponto que corresponde à mediana. Para saber qual a parcela do intervalo da classe mediana (c = 4) deve levar em consideração a seguinte regra de três simples:

$$\begin{cases} 11 \_\_\_\_\_\_\_\_ 4 \\ 7 \_\_\_\_\_\_\_\_\_\_\_\_ X \end{cases} \quad \text{implica que } x = \frac{4x7}{11} = \frac{28}{11} = 2{,}54$$

Acrescentando-se esse valor ao limite inferior da classe mediana, chegaremos a:

$$Md = 158 + 2{,}54 = 160{,}54 \cong 160{,}5 \text{ cm}$$

Valor igual conseguido com o uso da fórmula (15). Na realidade, esse último é apenas o resultado de interpolação. Assim, de acordo com o que vimos acima,

$$x = \frac{4x7}{11} = 4\,\frac{28}{11}, \text{ onde}$$

$4 = A^*$

$7 = 20 – 13$

$11 = f_i^*$ (frequência simples da classe mediana)

Substituindo os números pelos símbolos que os representam, temos:

$$X = A^*\,\frac{E_{md} - F_i{\downarrow}_{(ant)}}{fi^*},$$

que somado ao limite inferior da classe mediana (li*), reproduz a fórmula (15).

$$Md = l_i^* + A^*\,\frac{\left[\frac{\sum f_i}{2} - F_i{\downarrow}_{(ant)}\right]}{f_i^*}$$

SAIBA MAIS

Quartis* divide a distribuição em quatro partes iguais quanto ao número de elementos de cada parte.
Decis* divide a distribuição em dez partes iguais quanto ao número de elementos de cada parte.
Percentis* divide a distribuição em cem partes iguais quanto ao número de elementos de cada parte.

Quartis – Decis – Percentis (ou Centis)

Há uma série de medidas de posição semelhantes na sua concepção à mediana, embora não sejam medidas de tendência central. Como se sabe, a mediana divide a distribuição em duas partes iguais, quanto ao número de elementos de cada parte. São elas: os quartis*, os decis* e os centis*.

Para simbolizar cada uma dessas medidas separatrizes, faremos:

Q_i = quartis   i = 1, 2, 3

D_i = decis   i = 1, 2 , 3, ... , 9

C_i = centis   i = 1, 2, 3, ... , 99

Assim, para dividir uma série ordenada de valores em quatro partes iguais, precisaremos de três separatrizes (quartis); para dividi-la em dez, iremos recorrer a nove separatrizes (decis); em cem, recorremos a noventa e nove separatrizes (centis). O gráfico a seguir ilustra melhor o que acabamos de dizer em relação aos quartis e decis:

A |________|________|_______|________| B

Q_1 Q_2 Q_3

A |________________|_________________| B

D_5

$Q_2 = D_5 = Md$

Quartis

- **Primeiro Quartil**

Símbolo: Q1

Definição: Dado um conjunto ordenado (ordem crescente) de valores, o primeiro quartil, Q1 é o valor que divide o conjunto em duas partes tais que um quarto ou vinte e cinco por cento dos valores sejam menores do que ele e três quartos ou setenta e cinco por cento dos restantes sejam maiores. O elemento que indica a ordem ou posição do primeiro quartil é determinado, para dados agrupados em classes, pela seguinte expressão:

$$E_{Q1} = \frac{n}{4},$$

Onde n é o número de valores do conjunto ou o número de observações.

Segundo Quartil

Símbolo: Q_2 ou Md

Definição: Dado um conjunto ordenado de valores, o segundo quartil ou mediana é o valor que divide em duas partes iguais quanto ao número de elementos, isto é, cinquenta por cento ou dois quartos dos valores do conjunto são menores, e os dois quartos restantes são maiores do que ele. O elemento mediano é calculado, conforme veremos, através da seguinte expressão:

$$E_{md} = E_{Q2} = \frac{2n}{4} = \frac{n}{2}$$

Terceiro Quartil

Símbolo: Q_3

Definição: Dado um conjunto ordenado (ordem crescente) de valores, o terceiro quartil é o valor que divide o conjunto em duas partes tais que setenta e cinco por cento ou três quartos dos valores sejam menores e vinte e cinco por cento ou um quarto sejam maiores do que ele. O elemento que indica a ordem em que n encontra o terceiro quartil é calculado, para dados tabulados, conforme segue:

$$E_{Q3} = \frac{3n}{4} \bullet,$$

Onde n é o número de valores observados.

Genericamente, para determinar a ordem ou posição do quartil a ser calculado, usaremos a seguinte expressão:

$$E_{Qi} = \frac{in}{4} \qquad (16)$$

Onde:

i= número do quartil a ser calculado
n = número de observações.

Decis

Símbolo: D_i $\quad i = 1, 2, 3, ..., 9$

A definição dos decis obedece ao mesmo princípio da definição dos quartis, com a modificação da porcentagem de valores que ficam aquém e além do decil que se pretende calcular. Assim, por exemplo:

Primeiro Decil: D_1

O primeiro decil de um conjunto ordenado (ordem crescente) de valores é o valor que divide um conjunto em duas partes tais que dez por cento ou um décimo dos valores sejam menores e nove décimos ou noventa por cento sejam maiores do que ele. O elemento que indica a posição do primeiro decil é calculado pela expressão:

$$E_{D1} = \frac{n}{10}$$

Segundo decil: D_2

Trata-se do valor que divide o conjunto em duas partes, tais que vinte por cento ou dois décimos dos valores sejam menores e oitenta por cento ou oito décimos dos valores sejam maiores; para saber a ordem do segundo decil, usamos a expressão:

$$E_{D2} = \frac{2n}{10}$$

De especial interesse é o quinto decil, que divide o conjunto em duas partes, tais que cinco décimos ou cinquenta por cento dos valores sejam menores do que cinco décimos ou cinquenta por cento dos valores restantes maiores do que ele. Assim sendo, o quinto decil é igual ao segundo quartil que, por sua vez, é igual à mediana. O elemento que indica a ordem do quinto decil é igual ao elemento mediano, ou seja:

$$E_{D5} = \frac{5n}{10} = \frac{n}{2} = \frac{2n}{4}$$

Podemos, então, afirmar que:

$Md = D_5 = Q_2$

De modo geral, para calcular os decis, recorremos à seguinte expressão que define a ordem em que o decil se encontra:

$$E_{Di} = \frac{in}{10} \qquad (17)$$

Onde:

n = número de valores observados
i = número que identifica o decil a ser calculado

2.4.3 Percentis ou centis

Símbolo: C_i $\qquad$ i = 1, 2, 3, ... , 99

Neste caso, cada parte em que foram subdivididos os valores do conjunto, através dos noventa e nove centis, contará com um centésimo ou um por cento dos valores do conjunto.

O elemento que definirá a ordem do centil, em uma distribuição de frequências de valores tabulados agrupados em classes, será encontrado pelo emprego da expressão:

$$E_{Ci} = \frac{in}{100} \qquad (18)$$

Onde:
i = número identificador do centil
n = número total de observações

É oportuno lembrar que os centis englobam todos os decis e quartis.

Assim, por exemplo:

Décimo centil: C_{10}

O décimo centil dividirá o conjunto ordenado (ordem crescente) de valores em duas partes, tais que dez por cento ou dez centésimos dos valores do conjunto sejam menores e noventa por cento ou noventa centésimos sejam maiores do que ele. Teremos, então:

$$E_{C10} = \frac{10n}{100} = \frac{n}{10} = 0{,}1n \qquad \text{ou}$$

$$E_{D1} = \frac{1n}{10} = \frac{n}{10} = 0,1n$$

Vigésimo Centil: C_{20}

O vigésimo centil é igual ao segundo decil, porque

$$EC_{20} = \frac{20n}{100} = 0,2n = E_{D2} = \frac{2n}{10} = 0,2n$$

O vigésimo quinto centil é igual ao primeiro quartil, porque

$$E_{C25} = \frac{25n}{100} = \frac{n}{4} = 0,25n$$

, logo $E_{C25} = E_{Q1}$

$$E_{Q1} = \frac{1n}{4} = 0,25n$$

A fórmula do cálculo dos centis será:

$$C_i = l_i^* + A^* \frac{\left[\frac{in}{100} - F_i\downarrow_{(ant)}\right]}{fi^*} \qquad (19)$$

Exemplo:

Utilizando os dados da tabela 2.15, calcular as seguintes medidas:

Trigésimo Centil: C_{30}

Quinquagésimo Centil: C_{50}

Septuagésimo quinto Centil: C_{75}

Décimo Quinto Centil: C_{15}

Tabela 2.16

i	Estaturas (cm)	f_i	$F_i\downarrow$
1	150 ⊢ 154	4	4
2	154 ⊢ 158	9	13
3	**158 ⊢ 162**	**11**	**24**
4	162 ⊢ 166	8	32
5	166 ⊢ 170	5	37

i	Estaturas (cm)	f_i	$F_i\downarrow$
6	170←174	3	40
		$\sum f_i= 40$	

Fonte: Martins (2009).

Solução:

Trigésimo Centil: C_{30}

Inicialmente, calcularemos E_{C30}, ou seja, o elemento que define a ordem ou posição do trigésimo centil:

$$E_{C30} = \frac{in}{100} = \frac{30.40}{100} = 12$$

Agora, basta procurar a frequência acumulada imediatamente superior a 12 que é igual a 13; logo, a classe do trigésimo centil é a segunda: 154←158, aplicando a fórmula (19), obtemos:

$$C_i = l_i^* + A^* \frac{[E_{Ci} - F_i\downarrow_{(ant)}]}{fi^*} = 154 + 4 \frac{12-4}{9} =$$

$$= 154 + \frac{32}{9} = 154 + 3,55 = 157,55 \cong 157,6$$

C_{30} = 157,6 cm

Interpretação: o valor 157,6 refere-se à estatura dos alunos em cm. Como esse valor corresponde ao trigésimo centil, podemos afirmar que trinta por cento dos quarenta alunos têm estatura até 157,6 cm, enquanto que setenta por cento têm mais de 157,6 cm de altura.

Quinquagésimo Centil: C_{50}

Inicialmente calcularemos o E_{c50}, ou seja, o elemento que define a ordem ou posição do quinquagésimo centil:

$$E_{c50} = \frac{in}{100} = \frac{50.40}{100} = 20$$

Agora basta procurar a frequência acumulada imediatamente superior a 20, que é igual a 24; logo, a classe do quinquagésimo centil é a terceira: 158←162, aplicando a fórmula (19), obtemos:

$$C_{50} = 158 + 4.\ \frac{20\text{-}13}{11} = 158 + \frac{28}{11} = 158 + 2,54 =$$

$$= 160,54 \cong 160,5 = Md$$

$$C_{50} = 160,5 \text{ cm}$$

Interpretação: Neste caso, metade ou cinquenta por cento dos alunos tem estatura até 160,5 cm e os outros cinquenta por cento ou metade tem estatura superior a 160,5 cm.

Septuagésimo Quinto centil: C_{75}

Segue o mesmo procedimento das letras: a e b:

$$E_{c75} = \frac{75.40}{100} = 30$$

A frequência acumulada imediatamente superior a 30 é igual a 32; logo, a classe do septuagésimo quinto centil é a quarta: 162←166, aplicando a fórmula (19), obtemos:

$$C_{75} = 162 + 4.\ \frac{30\text{-}24}{8} = 162 + \frac{24}{8} = 162 + 3 =$$

$$= 165$$

$$C_{75} = 165 \text{ cm}$$

Interpretação: Neste caso, podemos dizer que três quartos ou setenta e cinco por cento dos alunos têm estatura até 165 cm, enquanto que um quarto ou vinte e cinco por cento possuem altura superior a 165 cm.

Décimo Quinto Centil: C_{15}

$$E_{C15} = \frac{15.40}{100} = 6$$

A frequência acumulada imediatamente superior a 6 é igual a 13; logo, a classe do décimo quinto centil é a segunda: 154←158, aplicando a fórmula (19), obtemos:

$$C_{15} = 154 + 4.\frac{(6\text{-}4)}{9} = 154 + 8/9 = 154 + 0{,}888 =$$

$$\cong 154{,}9$$

$$C_{15} = 154{,}9 \text{ cm}$$

Interpretação: o valor 154,9 refere-se à estatura dos alunos em cm. Como esse valor corresponde ao décimo quinto centil, podemos afirmar que quinze por cento dos 40 alunos têm estatura até 154,9 cm, enquanto que oitenta e cinco por cento têm estatura maior que 154,9 cm.

Exercícios Propostos 01

1) A tabela abaixo representa os salários pagos a 100 operários da empresa GLT & Cia.

Nº de Salários Míninos (x_i)	**Nº de Operários (f_i)**
0 ← 2	40
2 ← 4	30
4 ← 6	10
6 ← 8	15
8 ←10	5
	n = 100

Fonte: Toledo (1985)

Determinar:

a) Salário médio.

b) Salário modal (Czuber).

c) Salário mediano.

2) Com os dados do exercício 01, determinar o primeiro e terceiro Quartil.

3) Seja a distribuição das estaturas de 100 alunos de uma classe.

Estaturas (m)	Nº de alunos (f_i)
1,40←1,50	5
1,50←1,60	10
1,60←1,70	30
1,70←1,80	40
1,80←1,90	10
1,90←2,00	5
	n = 100

Fonte: Martins (2009).

Determinar:

a) A estatura média.

b) A estatura modal (Czuber).

c) A estatura mais frequente de (King).

d) A estatura mediana.

4) Um comerciante atacadista vende determinado produto em sacas que deveriam conter 16,50 kg. A pesagem de 40 sacas revelou os resultados representados na tabela abaixo:

Pesos (xi)	**Nº de sacas (f_i)**
14,55←15,05	1
15,05←15,55	3
15,55←16,05	8

16,05←16,55	9
16,55←17,05	10
17,05←17,55	6
17,55←18,05	3
	n=40

Fonte: Toledo (1985)

Pede-se:

a) A média da distribuição.

b) A mediana.

c) A moda (Czuber).

d) O septuagésimo quinto centil.

e) O terceiro decil.

f) A porcentagem de sacas entre a mediana e o seputagésimo quinto centil.

5) A tabela abaixo apresenta a distribuição das exportações de empresas eletrônicas em 1972.

Volume exportado em ($)	**Nº de empresas (f_i)**
50.000←60.000	5
60.000←70.000	10
70.000←80.000	20
80.000←90.000	10
90.000←100.000	5
	n = 50

Fonte: Toledo (1985)

Pede-se:

a) A média.

b) A moda.

c) A mediana.
d) Primeiro quartil.
e) Terceiro quartil.
f) Quinto decil.
g) Septuagésimo quinto centil.

Resumindo

As medidas de posição, objeto de estudo desta unidade, podem-se apresentar de várias formas, dependendo daquilo que se pretende conhecer a respeito dos dados estatísticos. As mais importantes são as medidas de tendência central ou promédias, as quais são assim denominadas, em virtude de a tendência dos dados observados se agruparem em torno desses valores centrais. A moda, a média aritmética e a mediana são as três medidas de tendência central ou promédios mais utilizados para resumir o conjunto de valores representativos dos fenômenos que se deseja estudar. Outros promédios menos usados são a média geométrica, harmônica, quadrática, cúbica e biquadrática e ainda as separatrizes, como os quartis, os centis.

UNIDADE 3

Medidas de Dispersão

Objetivos:

- Diferenciar medidas de posição de modelo de dispersão;
- Identificar e calcular as medidas de dispersão e interpretar os resultados;
- Identificar e calcular a medida de dispersão absoluta e relativa;
- Fazer uma análise completa de dados, através das medidas de dispersão.

3 MEDIDAS DE DISPERSÃO

Medidas de dispersão absoluta

Amplitude total ou intervalo total

Símbolo: A_t

Exemplo 1:
Calcular a amplitude total dos seguintes conjuntos de números:
A ={10,12, 13, 20, 25, 34, 45}
B = {17,18, 19, 20, 21, 22, 23}

Solução:
Para o conjunto A, temos: At = 45 – 10 = 35
Para o conjunto B, temos At = 23 – 17 = 6

Se os dados vierem dispostos em uma tabela de frequências, com os valores agrupados em classes, que basta calcular a diferença entre o limite superior da última classe e o limite inferior da primeira classe.

A_t = Limite superior da última classe – limite inferior da primeira classe.

Exemplo 2:
Calcular a amplitude total dos valores dispostos na tabela 3.1

SAIBA MAIS

Amplitude Total - é a diferença entre os valores extremos do conjunto.

Tabela 3.1

Classes	f_i	x_i
10⊢20	5	15
20⊢30	12	25
30⊢40	20	35
40⊢50	14	45
50⊢60	10	55
60⊢70	4	65
	n = 65	

Fonte: Toledo (1985)

Solução:

$$A_t = 70 - 10 = 60$$

Restrições ao uso da amplitude total

Embora a amplitude total seja a mais simples das medidas de dispersão, há uma forte restrição ao uso em virtude de sua grande instabilidade, uma vez que ela leva em conta apenas os valores extremos da série. Comparemos os conjuntos A e B do exemplo 1:

A média aritmética de cada um desses conjuntos é igual a 20. Portanto, no que diz respeito a uma medida de posição, ambos os conjuntos podem ser considerados idênticos. Ao calcular a amplitude total, verificaremos que os valores do conjunto A apresentam maior dispersão. Todavia, no cálculo da amplitude total não são levados em consideração os valores da série que se encontram entre os extremos, o que poderia conduzir o analista a interpretações equivocadas. Muitas vezes, um valor particularmente anormal poderá afetar de maneira acentuada a medida. O conjunto A, por exemplo, apresenta o último valor (45) sensivelmente distante do penúltimo (25), fato que talvez tenha provocado uma amplitude total de tal magnitude (35).

SAIBA MAIS

Desvio médio* é igual à média aritmética dos valores absolutos dos desvios tomados em relação à média ou mediana.

Desvio Médio

Símbolo: D_m

.

Desvio médio* para dados brutos

Quando os valores não vierem dispostos em uma tabela de frequência, o desvio médio será calculado, de acordo com a definição, através do emprego de uma das seguintes fórmulas:

$$D_m = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} \qquad (20)$$

Onde di =(x_i - $\bar{x}$) = desvio em relação à média aritmética.

$$D_m = \frac{\sum_{i=1}^{n} |x_i - Md|}{n} \qquad (21)$$

Onde $d_i = x_i - Md$ = desvio em relação à mediana.

Exemplo:

Calcular o desvio médio do conjunto de números

A = {10,12, 13, 20, 25, 34, 45}

Os dados necessários para o cálculo dos desvios são:

$$\bar{X}A = \frac{10+12+13+20+25+34+45}{7} = \frac{159}{7} = 22{,}714$$

Md=20

Faremos uma tabela com cinco colunas, uma para x_i, outra para $x_i - \bar{x}$, para $|x_i - \bar{x}|$, outra para x_i – Md e a última para $|x_i - \bar{x}|$.

Tabela 3.2

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$	x_i Md	x_i - Md $\rvert$
10	- 12,714	12,714	-10	10
12	-10,714	10,714	-8	8
13	-9,714	9,714	-7	7
20	-2,714	2,714	0	0
25	2,286	2,286	5	5
34	11,286	11,286	14	14
45	22,286	22,286	25	25
	$\sum(x_i - \bar{x}) = 0$	$\sum \lvert x_i - \bar{x} \rvert = 71{,}714$		$\sum \lvert x_i\text{-Md} \rvert = 69$

Fonte: Toledo (1985)

Usando as fórmulas (20) e (21), chegaremos a:

Pela Mé

$$D_m = \frac{\sum_{i=1}^{7} \lvert x_i - 22{,}714 \rvert}{7} = \frac{71{,}714}{7} = 10{,}245$$

$D_m = 10{,}245$

Pela Mediana

$$D_m = \frac{\sum_{i=1}^{7} \lvert x_i - 20 \rvert}{7} = \frac{69}{7} = 9{,}857$$

$D_m = 9{,}857$

O desvio médio, neste caso, é menor quando tomado em relação à mediana do que em relação à média aritmética.

Desvio médio para dados tabulados

Se os valores vierem dispostos em uma tabela de frequências, agrupados ou não em classes, serão usadas as seguintes fórmulas:

Cálculo pela média

$$D_m = \frac{\sum_{i=1}^{k} |x_i - \overline{x}| f_i}{n} \qquad (22)$$

Onde x_i representa um valor individual ou um ponto médio da classe.

Cálculo pela mediana

$$D_m = \frac{\sum_{i=1}^{k} |x_i - Md|}{n} \qquad (23)$$

Exemplo:

Calcular o desvio médio dos valores representativos das alturas dos alunos do Curso de Sistema de Informação da UAPI.

Tabela 3.3

Alturas (cm)	f_i	x_i	xi.fi	$x_i - x$	$\vert x_i - x \vert f_i$
150 ⊢ 154	4	152	608	-9	36
154 ⊢ 158	9	156	1404	-5	45
158 ⊢ 162	11	160	1760	-1	11
162 ⊢ 166	8	164	1312	3	24

Alturas (cm)	f_i	x_i	xi.fi	x_i - x	| x_i - x | f_i
166⊢170	5	168	840	7	35
170⊢174	3	172	516	11	33
	40		∑= 6440		∑ = 184

Fonte: Juarez Martins (2009)

A média aritmética foi calculada no exemplo da pag.?? ($\bar{x}$ = 161 cm) e a mediana (Md = 160,5) do capítulo precedente. Teremos, então:

Cálculo pela média

$$D_m = \frac{\sum_{i=1}^{6} |x_i - 161| f_i}{40} = \frac{184}{40} = 4,6$$

$D_m = 4,6$

O cálculo do desvio médio pela mediana segue o mesmo raciocínio usado para o cálculo pela média.

Cálculo pela Mediana

Usando a fórmula (23), obtemos:

$D_m = 4,5$

Temos novamente neste exemplo o desvio médio, calculado com base na mediana, menor que o calculado com base na média aritmética.

Observações:

- O desvio médio resulta mais vantajoso que a medida de dispersão precedente pelo fato de, em seu cálculo, levar em consideração todos os valores da distribuição;

- O desvio médio calculado, levando-se em consideração os desvios em torno da mediana, é mínimo, ou seja, é menor do que qualquer desvio médio calculado com base em qualquer outra medida de tendência central.

Apesar de o desvio médio expressar aceitavelmente a dispersão de uma amostra, não é tão frequentemente empregado como o desvio padrão, o qual será descrito mais adiante, pois este se adapta melhor a uma ampla gama de aplicações. Além disso, o desvio médio despreza o fato de alguns desvios serem negativos e outros positivos. Todavia, será preferido o uso do desvio médio em lugar do desvio padrão, quando esse for indevidamente influenciado pelos desvios extremos.

Desvio padrão

Símbolo: S

O desvio padrão é a medida de dispersão mais usada, tendo em comum com o desvio médio o fato de ambos serem os desvios com relação à $\overline{x}$. Só que, no cálculo do desvio padrão, em lugar de serem usados os valores absolutos das discrepâncias ou desvios, calculam-se os quadrados desses.

Desvio padrão de dados brutos

Seja o seguinte conjunto de números: X = {x1, x2, x3,... ,xn}. O desvio padrão ou a média quadrática dos desvios ou afastamento em relação à média aritmética desse conjunto será definido por:

$$S = \sqrt{\frac{\sum_{i=1}^{n} d^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} (x - \bar{x})^2}{n}} \qquad (24)$$

Onde $d_i = (x_i - \overline{x})$.

Se desenvolvermos o numerador da expressão sob o radical, chegaremos à fórmula desenvolvida do desvio padrão:

SAIBA MAIS

Desvio padrão(s) é igual à raiz quadrada da variância.

S = √s2

$$S=\sqrt{\frac{1}{n}\left[\sum xi-\frac{(\sum xi)^2}{n}\right]} \qquad (25)$$

Observação:

Quando o desvio padrão representar uma descrição da amostra e não da população, caso mais frequente em Estatística, o denominador das expressões 24 e 25 será igual a n – 1, em vez de n. A razão desse procedimento reside no fato de que, utilizando o divisor (n – 1), obtém-se uma estimativa melhor no parâmetro de população. Além do mais, apenas (n – 1) das discrepâncias (x_i - $\overline{x}$) são independentes, uma vez que essas (n – 1) discrepâncias determinam automaticamente a n-ésima. Para valores grandes de n (n > 30) não há grande diferença entre os resultados proporcionados pela utilização de qualquer dos dois divisores, n ou n – 1. Entretanto, daremos preferência à fórmula que proporciona uma estimativa mais justa do desvio padrão da população, ou seja:

$$S=\sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} xi^2-\frac{(\sum xi)^2}{n}\right]} \qquad (26)$$

Exemplo:

Calcular o desvio padrão de cada um dos conjuntos de números do conjunto A = {10, 12, 13, 20, 25, 34, 45}.

Tabela 3.4

Fórmula Original		Fórmula Desenvolvida	
x_i	$X_i - \overline{X}$	$d_i^2 = (x_i - \overline{x})^2$	X_i^2
10	0,714	161,646	100
12	-9,714	114,790	144
13	-2,714	94,362	169
20	2,286	7,366	400
25	11,286	5,226	625
34	22,286	127,374	1156
45		496,666	2025

Fórmula Original		Fórmula Desenvolvida	
$\sum x_i = 159$		$\sum d_i^2 = 1007{,}430$	$\sum x_i^2 = 4619$

Fonte: Toledo (1985)

Resolvendo pela fórmula original, devemos ter:

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} xi^2 - \frac{(\sum xi)^2}{n}\right]} \qquad (26)$$

Onde n = 7 e $\overline{x}$ = 22,714.

$$S = \sqrt{\frac{1}{7-1}\sum (x_i - 22{,}714)^2} = \sqrt{\frac{1}{6}.\ 1007{,}430} =$$

$$= \sqrt{167{,}905} = 12{,}958$$

S_A = 12,958

Utilizando a fórmula desenvolvida:

$$S = \sqrt{\frac{1}{n-1}\left[\sum x^2 - \frac{(\sum xi)^2}{n}\right]} =$$

$$S = \sqrt{\frac{1}{7-1}\left[4619 - \frac{(159)^2}{7}\right]} = \sqrt{\frac{1}{6}\left[4619 - \frac{25281}{7}\right]}$$

$$= \sqrt{\frac{1}{6}[4619 - 3611{,}571]} = \sqrt{\frac{1007{,}429}{6}} = \sqrt{167{,}905}$$

S_A = 12,958

Conforme se pode observar, as duas expressões para o cálculo do desvio padrão são equivalentes.

Desvio padrão de dados tabulados

Quando os valores vierem dispostos em uma tabela de frequências, o cálculo do desvio padrão far-se-á através de uma das seguintes fórmulas:

$$S = \sqrt{\frac{\sum_{i=1}^{k} di^2 fi}{n}} = \sqrt{\frac{\sum_{i=1}^{k} (xi - \bar{x})^2 fi}{n}} \quad (27)$$

Ou, pela fórmula desenvolvida:

$$S = \sqrt{\frac{1}{n}\left[\sum xi^2 fi - \frac{(\sum xifi)^2}{n}\right]} \quad (28)$$

Onde di = $x_i - \bar{x}$ = desvio em torno da média.

X_i = valor isolado da variável, ou ponto médio da classe, se os valores vierem agrupados em classe.

Pelos mesmos motivos expostos anteriormente, será dada preferência ao divisor n – 1, em vez de n, no cálculo do desvio padrão para dados tabulados.

$$S = \sqrt{\frac{\sum_{i=1}^{k} di^2 fi}{n-1}} = \sqrt{\frac{\sum_{i=1}^{k} (xi - \bar{x})^2 . fi}{n-1}} \quad (29)$$

Ou

$$S = \sqrt{\frac{\left[\sum_{i=1}^{k} xi^2 fi - \frac{(\sum xifi)^2)}{n}\right]}{n-1}} \quad (30)$$

Ou

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{k} xi^2.fi - \frac{(\sum xi.fi)^2}{n}\right]} \qquad (31)$$

Exemplo:

Com os dados da tabela 2.2, calcular o desvio padrão da distribuição de frequências da altura dos alunos em (cm):

Tabela 3.5

Alturas (cm)	f_i	x_i	x_if_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2f_i$	$x_i^2f_i$
150 ⊢ 154	4	152	608	-9	324	92416
154 ⊢ 158	9	156	1404	-5	225	219024
158 ⊢ 162	11	160	1760	-1	11	281600
162 ⊢ 166	8	164	1312	3	72	215168
166 ⊢ 170	5	168	840	7	245	141120
170 ⊢ 174	3	172	516	11	363	88752
	40		**6440**		**1240**	**1038080**

Fonte: Martins (2009).

A média aritmética já foi calculada anteriormente:

$$\bar{x} = \frac{(\sum x_if_i)}{n} = \frac{6440}{40} = 161$$

Cálculo do desvio padrão pela fórmula original (29)

$$S = \sqrt{\frac{\sum_{i=1}^{k}(xi - \bar{x})fi}{n-1}} =$$

$$= \sqrt{\frac{1240}{39}} = \sqrt{31{,}79487179} \cong \mathbf{5{,}64}$$

S = 5,64 cm

O desvio padrão da altura dos alunos é de 5,64 cm. Recorde-se que o desvio médio, calculado anteriormente, resultou em Dm = 4,6 cm.

Pela fórmula desenvolvida (30):

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{k} xi^2 fi - \frac{(\sum xifi)^2}{n}\right]} =$$

$$S = \sqrt{\frac{1}{40-1}\left[\,1038080 - \frac{(6440)^2}{40}\right]} =$$

$$= \sqrt{\frac{1}{39}\left[\,1038080 - 1036840\,\right]} =$$

$$= \sqrt{\frac{1}{39}.(1240)} = \sqrt{31{,}79487179} =$$

= 5,64

S = 5,64 cm

Interpretação do desvio padrão

O desvio padrão não tem uma interpretação física, como ocorre com a média, mediana, a moda e os quartis. Contudo, é possível interpretá-lo de forma analítica. Consideremos por exemplo, que dois estudantes tenham obtido os seguintes resultados em 5 provas de Estatística, realizadas ao longo do ano letivo:

Estudante A: 40 50 60 70 80
Estudante B: 20 40 60 80 100

Ambos os estudantes foram aprovados na disciplina, pois suas médias foram iguais a 60:

$$\overline{x}A = \frac{40+50+60+70+80}{5} = \frac{300}{5} = 60$$

$$\bar{x}B = \frac{20+40+60+80+100}{5} = \frac{300}{5} = 60$$

Entretanto, a variação das notas em torno das respectivas médias difere do aluno A para o aluno B; este último apresentando maior dispersão do que aquele. É possível perceber ainda que a diferença entre pares sucessivos de notas do aluno B é igual a duas vezes a do aluno A, que a amplitude total das notas de B é igual ao dobro da de A. Assim sendo, podemos afirmar que o aluno B apresentou resultados com uma variação média igual ao dobro da do aluno A. O desvio padrão das notas permite comprovar o que foi dito:

$$S_A = \sqrt{\frac{\sum(xi-\bar{x})^2}{n-1}} =$$

$$= \sqrt{\frac{(40-60)^2+(50-60)^2+(60-60)^2+(70-60)^2+(80-60)^2}{5-1}}$$

$$= \sqrt{\frac{400+100+0+100+400}{4}} = \sqrt{250} = 5\sqrt{10}$$

$$S_A = 5\sqrt{10}$$

$$S_B = \sqrt{\frac{(20-60)^2+(40-60)^2+(60-60)^2+(80-60)^2+(100-60)^2}{5-1}}$$

$$= \sqrt{\frac{1600+400+0+400+1600}{4}} =$$

$$= \sqrt{1000} = 10\sqrt{10}$$

$$S_B = 10\sqrt{10}$$

Ou seja,

$$S_B = 2S_A$$

Deste modo, o desvio padrão é realmente uma média satisfatória de dispersão, embora não se possa afirmar muita coisa quanto à sua magnitude.

Variância

Simbolo: S^2

SAIBA MAIS

Variância* é definida como sendo o quociente entre a soma dos quadrados dos desvios e o número de elementos.

Conforme se pode perceber pelo símbolo, a variância* é o quadrado do desvio padrão, ou se preferir, o desvio padrão é a raiz quadrada da variância. Dessa forma, pode-se dizer que a fórmula da variância é igual à expressão do desvio padrão, sem o sinal do radical. Adiantamos que as definições que se seguem aparecem com a correção de Bessel, isto é, com o divisor n -1.

Variância de dados brutos

Fórmula original

$$S_2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n - 1} \qquad (31)$$

Fórmula desenvolvida

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} xi^2 - \frac{\left(\sum_{i=1}^{i} xi\right)^2}{n}\right] \qquad (32)$$

Ou

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum xi)^2}{n}}{n - 1} \qquad (33)$$

Exemplo:
Calcular a variância dos seguintes conjuntos de números:

A = {10, 12, 13, 20, 25, 34, 45}

$$B = \{17, 18, 19, 20, 21, 22, 23\}$$

Anteriormente, já foi calculado o desvio padrão de cada conjunto. Os dados para a determinação de variância foram extraídos desse exemplo.

Sabemos que

$\overline{x}A = 22{,}714$

$\overline{x}B = 20$

Então,

$$S_2A = \frac{\sum(xi-22{,}714)^2}{(7-1)} = 167.905$$

$$S_2B = \frac{\sum(xi-20)^2}{7-1} = 4{,}667$$

Variância de dados tabulados

Fórmula original

$$S_2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2.f_i}{n - 1} \qquad (34)$$

Fórmula desenvolvida

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} xi^2.fi - \frac{(\sum_{i=1}^{k} xi.fi)^2}{n}\right] \qquad (35)$$

Exemplo:

Calcular a variância da altura dos 40 alunos com os dados da tabela 5.5.

Usando a fórmula (34) temos:

$$S_2 = \frac{\sum_{i=1}^{K}(x_i - \overline{x})^2.f_i}{n - 1} = \frac{\sum_{i=1}^{6}(x_i - 161)^2.f_i}{40 - 1} =$$

$$= \frac{1240}{39} = 31,79487 \cong 31,795$$

$S_2 = 31,795(cm)^2$

Conforme pode ser observado, a variância expressa à unidade de medida sempre ao quadrado, ao contrário do desvio padrão.

Medidas de dispersão relativa

Para determinadas classes de problemas, as medidas de variabilidade ou dispersão relativa, em uma distribuição de frequências, proporcionam uma avaliação mais apropriada do grau de dispersão da variável do que as de dispersão absoluta. A dispersão relativa permite ainda comparar duas ou mais distribuições, mesmo que essas se refiram a diferentes fenômenos e sejam expressas em unidades de medidas distintas. As medidas de dispersão resultam, em geral, de comparação entre uma medida de dispersão absoluta e um promédio, sendo seu resultado expresso em termos percentuais.

Desvio quartil reduzido

Símbolo: Dq_r

O desvio quartil é uma medida de dispersão relativa à resultante do quociente entre o desvio quartil reduzido e a mediana. Simbolicamente,

$$Dq_r = \frac{Dq}{Md} = \frac{\frac{Q3-Q1}{2}}{Md} = \frac{Q3-Q1}{2Md} \qquad (36)$$

O

$$Dq = \frac{Q3 - Q1}{2Md} \times 100 \qquad (37)$$

Coeficiente de variação

O coeficiente de variação ou coeficiente de variação relativa é uma porcentagem cujo cálculo resulta da comparação entre o desvio padrão ou o desvio médio e a média ou a mediana. Definiremos os seguintes coeficientes de variação:

Coeficiente de Variação de Pearson

Símbolo: CVp

O coeficiente de variação de Pearson é igual ao quociente entre o desvio padrão e a média aritmética.

$$CVp = \frac{S}{\bar{x}}$$

Ou

$$CVp = \frac{S}{\bar{x}} \times 100 \qquad (38)$$

Exemplo:

Suponhamos que uma empresa fabricante de pneumáticos desenvolveu um novo produto, com um cordel que proporciona maior resistência às flexões repetidas e maior resistência à tração do que o original. Tendo submetido esse componente à prova, chegou-se à conclusão de que a resistência às flexões repetidas, testada em um aparelho de dobrar cordéis, foi em média de 139 minutos com desvio padrão de 15 minutos contra a média de 88 minutos e desvio padrão de 14 minutos do cordel comum. Comparando os valores dos desvios padrões, parece não haver diferença significativa quanto à dispersão do tempo de resistência à flexão. Entretanto, deve-se ter presente que o desvio padrão para o novo cordel refere-se a uma maior resistência média às flexões repetidas; e é neste aspecto que se baseia o conceito de dispersão relativa refletido pelo resultado do índice CVp. Calculemos, então, o CVp para os dois casos.

Para o novo cordel:

$$CVp = \frac{15}{139} = 0{,}108 \qquad \text{Ou } 10{,}8\%$$

Para o cordel antigo:

$$CVp = \frac{14}{88} = 0{,}159 \qquad \text{Ou } 15{,}9\%$$

Comparando-se os resultados vê-se que a variação relativa é muito menor para o novo cordel do que para o antigo.

Coeficiente de Variação de Thorndike

Símbolo: CV_T

O coeficiente de Thorndike é igual ao quociente entre o desvio padrão e a mediana.

$$CV_T = \frac{S}{Md}$$

Ou

$$CV_T = \frac{S}{Md} \times 100 \qquad (39)$$

Exemplo:

Calcular o coeficiente de Thonrndike com os dados do exemplo relativo à altura dos 40 alunos do Curso de Sistema de Informação da UAPI.

Os dados para resolver o problema já são conhecidos:

S = 5,64

Md = 160,5

$$CV_T = \frac{5,64}{160,5} = 0,03514 \cong 0,035 \qquad \text{ou } 3,5\%$$

O coeficiente de Pearson para o mesmo problema é

$$CV_T = \frac{5,64}{161} = 0,03503 \cong 0,035 \qquad \text{ou } 3,5\%$$

Exercícios Propostos 01

1. A tabela abaixo representa os salários pagos a 100 operários da empresa GLT & Cia.

Nº de Salários Míninos (x_i)	**Nº de Operários (f_i)**
$0 \vdash 2$	40
$2 \vdash 4$	30
$4 \vdash 6$	10
$6 \vdash 8$	15
$8 \vdash 10$	5
	n = 100

Fonte: Toledo (1985)

Determinar:

a) Desvio quartil.

b) Desvio médio.

c) Desvio padrão.

2. Dados os conjuntos A = {1000, 1001. 1002, 1003, 1004, 1005} e B = {0, 1, 2, 3, 4, 5}, podemos afirmar que:

a) O desvio padrão A é igual a 1000 vezes o desvio padrão de B.

b) O desvio padrão de A é igual ao desvio padrão de B.

c) O desvio padrão de A é igual ao desvio padrão de B multiplicado pelo quadrado de 1000.

d) O desvio padrão de A é igual ao desvio padrão de B dividido por 1000.

A tabela abaixo representa a vida útil de postes telefônicos de madeira:

Anos	Nº de postes substituídos
0.5←2,5	11
2,5←4,5	47
4,5←6,5	87
6,5←8,5	134
8,5←10,5	200
10,5←12,5	198
12,5←14,5	164
14,5←16,5	102
16,5←18,5	48
18,5←20,5	6
20,5←22,5	3
	1000

Fonte: Toledo (1985)

Pede-se:

a) O desvio padrão.

b) A variância;

c) O coeficiente de variação de Pearson.

Resumindo

Nesta unidade, veremos a análise completa dos dados, que requer não apenas sua apresentação, através de gráficos e tabelas ou o cálculo de promédios ou outras medidas de posição, pois caracterizar um conjunto de valores apenas através de uma média, por exemplo, é descrevê-lo inadequadamente, uma vez que os dados diferem entre si, em maior ou menor grau.

UNIDADE 4

Fundamentos de Análise Combinatória, Principais Conceitos Probalísticos e seus Principais Teoremas e Variáveis

Objetivos:

- Resolver problemas de contagem, utilizando os princípios aditivo e multiplicativo ou noções de permutações, arranjes simples e combinações simples;
- Identificar os principais conceitos probabilísticos;
- Calcular a probabilidade de um evento, aplicando o resultado.

4 FUNDAMENTOS DE ANÁLISE COMBINATÓRIA, PRINCIPAIS CONCEITOS PROBALÍSTICOS E SEUS PRINCIPAIS TEOREMAS E VARIÁVEIS

Fundamentos de análise combinatória

Principio multiplicativo

Vamos aprender agora a determinar o número de possibilidades de ocorrência de um evento, sem a necessidade de descrever todas as possibilidades.

Exemplo 1:

André tem duas bermudas (preta e cinza) e quatro camisetas (branca, verde, amarela e roxa). De quantas maneiras diferentes ele poderá se vestir usando uma bermuda e uma camiseta?

Solução:

Há duas possibilidades de escolher uma bermuda. Para cada uma delas, quatro possibilidades de escolher uma camiseta. Logo, o número total de maneiras diferentes de André se vestir é 2. 4 = 8.

Se um acontecimento ocorrer por várias etapas sucessivas e independentes, de tal modo que:

P_1 é o número de possibilidades da 1º etapa
P_2 é o número de possibilidades da 2º etapa
.
.
.
P_k é o número de possibilidades da k-ésima etapa,

Então $p_1 \cdot p_2 \ldots p_k$ é o número total de possibilidades de o acontecimento ocorrer.

Exemplo 2:

Os números dos telefones de uma cidade têm 8 algarismos. Determinar a quantidade máxima de telefones a serem instalados, sabendo que os números não devem começar com zero.

Solução:

Como o número do telefone tem 8 algarismos, ele apresenta a seguinte forma:

$$P_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5 \cdot p_6 \cdot p_7 \cdot p_8 = 9.10.10.10.10.10.10.10 =$$
$$= 9.10^7 = 90.000.000$$

Fatorial

Sendo n um número inteiro maior que 1, define-se fatorial de n como o produto dos n números naturais consecutivos de n a 1. Indica-se n!.

$n! = n(n-1).(n-2)....3 \cdot 2 \cdot 1$, sendo $n \in N$ e $n > 1$.

$8! = 8.\ 7.\ 6 \cdot 5.\ 4 \cdot 3 \cdot 2.\ 1 = 40320$

Arranjos Simples

Denomina-se arranjo simples dos n elementos tomados p a p, com ($p \leq n$) são os agrupamentos ordenados diferentes que se podem formar com p dos n elementos dados.

$$A_{n,p} = \frac{n!}{(n-p)!} \qquad (40)$$

Exemplo:

Calcule o A10,4 .

$$A_{10,4} = \frac{10!}{(10-4)!} = \frac{10!}{6!} = \frac{10 .9 .8 .7!}{7!} = 5040$$

Exemplo 2:

Quantos números de dois algarismos diferentes podemos escrever com os algarismos 1, 2, 3, 4, 5, 6, 7, 8 e 9?

Solução:

Procuramos agrupamentos de 2 elementos em que a ordem é importante, pois, por exemplo, 12 ≠ 21. Temos 9 elementos para serem arranjados 2 a 2. Assim, temos:

$$A_{9,2} = \frac{9!}{(9-2)!} = \frac{9!}{7!} = \frac{9 .8 .7!}{7!} = 72$$

Portanto, existem 72 números de 2 algarismos distintos que podem ser escritos com os algarismos de 1 a 9.

Exemplo 3:

Quantos números pares de quatro algarismos obtemos com os algarismos 0, 1, 2, 3, 4, 5 e 6, sem repeti-los?

Solução:

São sete algarismos disponíveis para formar números de quatro algarismos, com a condição que sejam números pares; portanto devem terminar em 0, 2, 4 ou 6.

$$\begin{array}{l} \_\_\_\ \_\_\_\ \_\_\_\ 0 \rightarrow A_{6,3} \\ \_\_\_\ \_\_\_\ \_\_\_\ 2 \rightarrow A_{6,3} \\ \_\_\_\ \_\_\_\ \_\_\_\ 4 \rightarrow A_{6,3} \\ \_\_\_\ \_\_\_\ \_\_\_\ 6 \rightarrow A_{6,3} \\ \hline \qquad\qquad\quad 4 \,.\, A_{6,3} \end{array}$$

Quando os números terminam em 2, 4 ou 6, eles não podem começar por zero.

0 ___ ___ $2 \rightarrow A5,2$
0 ___ ___ $4 \rightarrow A5,2$
0 ___ ___ $6 \rightarrow A5,2$

$3 \cdot A_{5,2}$

Portanto, o total de números é: $4 \cdot A_{6,3} - 3 \cdot A_{5,2} =$
$= 4 \cdot 6 \cdot 5 \cdot 4 - 3 \cdot 5 \cdot 4 =$
$= 480 - 60 = 420$

Obtemos 420 números

Permutação simples

Se tivermos n elementos distintos, então o número de agrupamentos ordenados que possam obter com todos esses elementos, chamamos de permutações simples, e indicamos por:

$$P_n = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1 \qquad (41)$$

Exemplo 1:

Quantos números de 5 algarismos distintos podem ser formados, usando os algarismos 1, 3, 5, 7 e 8?

Solução:

Queremos formar números (agrupamentos de 5 algarismos com os 5 algarismos dados(1,3,5, 7, 8).

___ ___ ___ ___ ___ $\rightarrow P_5 = A_{5,5}$
$P5 = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$

Podem ser formados 120 números.

Exemplo 2:

Quantos anagramas há na palavra MITO?

Solução:

Um anagrama é qualquer ordenação (sequência) das letras de uma palavra. Como a palavra MITO tem 4 letras, vamos formar anagramas de

4 letras com M, I, T e O. Assim, temos:

$$\_\_\_\_ \ \_\_\_\_ \ \_\_\_\_ \ \_\_\_\_ \rightarrow P_4 = 4! =$$
$$= 4.3.2.1 = 24$$

A palavra MITO tem 24 anagramas.

Combinações simples

Combinações simples de n elementos tomados p a p(p≤n) são os subconjuntos com exatamente p elementos que se podem formar com os n elementos dados. Indicamos por:

$$C_{n,p} = \binom{n}{p} = \frac{n!!}{p!(n-p)!} \qquad (42)$$

Exemplo 1:

Quantas comissões de 3 participantes podem ser formadas com 5 pessoas?

Solução:

No exemplo, as comissões devem ter 3 participantes, isto é, não usaremos todas as pessoas. Vamos chamar de A, B, C, D e E as 5 pessoas que podem ser indicadas para a comissão. Dessas escolhemos 3.

A B C

Invertendo-se a ordem, temos B A C

Como A, B e C são a mesma comissão de B, A, C, o problema é de combinação. Logo:

$$C_{5,3} = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{5 .4 .3!}{3! .2 .1} = 10$$

Podemos formar 10 comissões.

Exemplo 2:

Em uma classe tem 10 alunos e 5 alunas. Formam-se comissões de 4 alunos e duas alunas. Determine o número de comissões em que participa o aluno x e não participa a aluna y.

SAIBA MAIS

Para Classificar um agrupamento como arranjo ou combinação, procedemos da seguinte forma:
1º) Forma-se o agrupamento sugerido pelo problema;
2º) Em seguida, muda-se a ordem de seus elementos.
3º) Se, com essa mudança de ordem, obtivermos um agrupamento diferente do inicial, esses agrupamentos serão de arranjos,porém se o agrupamento for igual ao inicial, esses agrupamentos serão combinações.

SAIBA MAIS

Experimentos aleatórios são aqueles que, mesmos experimentos repetidos várias vezes, sob condições semelhantes, apresentam resultados imprevisíveis.

SAIBA MAIS

Ao conjunto desses resultados possíveis* damos o nome de espaço amostral ou conjunto universo, representado por S.

Solução:

A comissão deve ter 6 pessoas.

___ ___ ___ ___ e ___ ___

4 alunos e 2 alunas

Devemos escolher:

→ 3 alunos entre os 9 restantes: $C_{9,3}$

→ 2 alunas entre as 4 restantes: $C_{4,2}$

A escolha dos alunos pode ser efetuada de $C_{9,3}$ maneiras e a escolha das alunas de $C_{4,2}$. Então, pelo principio fundamental da contagem, o número total de comissões é dado por:

$$C_{9,3} \cdot C_{4,2} = \frac{9!}{3!6!} \cdot \frac{4!}{2!2!} = 504$$

Serão 504 comissões.

Probabilidade

Experimento aleatório

Em quase tudo, em maior ou menor grau, vislumbramos o acaso. Assim, da afirmação, “é provável que meu time ganhe a partido de hoje”, pode resultar:

_Que, apesar do favoritismo, ele perca:

_Que, como pensamos, ele ganhe;

_Que empate.

Como vimos, o resultado final depende do acaso. Fenômenos como esses são chamados fenômenos aleatórios ou experimentos aleatórios.

Espaço amostral

A cada experimento correspondem, em geral, vários resultados possíveis. Assim, ao lançarmos uma moeda, há dois resultados possíveis: ocorrer cara(k) ou ocorrer coroa(c). Já ao lançarmos um dado há seis resultados possíveis*: 1, 2, 3, 4, 5 ou 6.

Os dois experimentos citados anteriormente têm os seguintes espaços amostrais:

- Lançamento de uma moeda: S = {k, c}
- Lançamento de um dado: S = {1, 2, 3, 4, 5, 6}.

Do mesmo modo, como em dois lançamentos sucessivos de uma moeda, podemos obter cara nos dois lançamentos, ou cara no primeiro e coroa no segundo; ou coroa no primeiro e cara no segundo; ou coroa nos dois lançamentos, o espaço amostral é:

S = {(kk), (kc), (ck), (cc)}.

Eventos

Sendo as operações com conjuntos, podem-se formar novos eventos.

Assim:

I) A ∪B → é o evento que ocorre se A ocorre ou B ocorre ou ambos ocorrem;

II) A∩B → é o evento que ocorre se A e B ocorrem;

III) $\overline{A}$ → é o evento que ocorre se A não ocorre.

Exemplos:

1) Seja o experimento E: jogar três moedas e observar os resultados:

S = {(kkk), (kkc), (kck), (ckk), (ccc), (cck), (kcc), (ckc)}

Seja A o evento: ocorrer pelo menos 2 caras.

Então, a = {(kkk), (kkc), (kck), (ckk)}

2) Seja o experimento E: Lançar um dado e observar o número de cima.

Então S = {1, 2, 3, 4, 5, 6}

Seja B o evento: ocorrer múltiplo de 2.

Então, B = {2, 4, 6}

Sendo S espaço amostral finito, com n elementos podem verificar que 2n fornecem o número total de eventos extraídos de S.

SAIBA MAIS

Evento é um subconjunto de S. Em particular, S e Φ(conjunto vazio) são eventos, S é dito o evento certo e Φ o evento impossível.

SAIBA MAIS

Dois eventos A e B são denominados mutuamente exclusivos, se eles não puderem ocorrer simultaneamente, isto é, A∩B = Φ

Eventos Mutuamente Exclusivos

Exemplo:

Seja E: jogar um dado e observar o resultado.
S = {1, 2, 3, 4, 5, 6}
Sejam os eventos: A = ocorrer nº par, e
B = ocorrer nº ímpar.
Então: A = {2, 4, 6} e B = {1, 3, 5}, então
$A \cap B = \Phi$.

A e B são mutuamente exclusivos, pois a ocorrência de um nº par e ímpar não pode ser verificada como decorrência da mesma experiência.

Definição de probabilidade

Chamamos de probabilidade de um evento A ($A \subset S$) o número real P(A), tal que:

$$P(A) = \frac{n(A)}{n(S)} \qquad (43)$$

Onde: n(A) é o número de elementos de A;
n(S) é o número de elementos de S.

Satisfazendo os seguintes axiomas:
I) $0 \leq P(A) \leq 1$
II) $P(S) = 1$
III) Se A e B forem eventos mutuamente exclusivos,
($A \cap B = \Phi$), então $P(A \cup B) = P(A) + P(B)$.

Principais teoremas

1 Se Φ é um conjunto vazio, então $P(\Phi) = 0$

Demonstração:

Seja A um evento qualquer.
A e Φ são disjuntos, pois $A \cap \Phi = \Phi$
$P(A \cup \Phi) = P(A) + P(\Phi)$ por III.

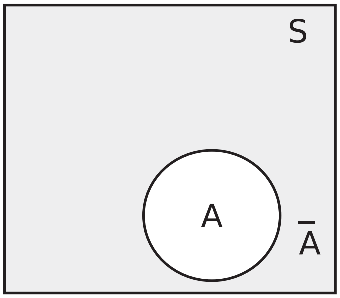
$P(A) = P(A) + P(\Phi)$ pois $A \cup \Phi = A$
Portanto $P(\Phi) = 0$.

2 Se $\bar{A}$ é o complemento do evento A, então

$$P(\bar{A}) = 1 - P(A). \qquad (44)$$

Demonstração:

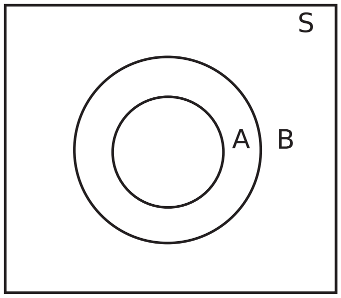
Pode-se escrever $S = A \cup \bar{A}$. (veja diagrama).



Ora, $A \cap \bar{A} = \Phi$ (são mutuamente exclusivos)
$P(A \cup \bar{A}) = P(A) + P(\bar{A})$
$P(S) = P(A) + P(\bar{A})$
$1 = P(A) + P(\bar{A})$ por II.
Logo: $P(\bar{A}) = 1 - P(A)$

3 Se $A \subset B$, então $P(A) \leq P(B)$

Demonstração:
Pode-se escrever $B = A \cup (\bar{A} \cap B)$
Ora, A e $(\bar{A} \cap B)$ são mutuamente exclusivos.
Logo, $P(B) = P(A) + P(\bar{A} \cap B)$
$P(\bar{A} \cap B) = P(B) - P(A)$
$P(B) - P(A) \geq 0$ por I
Portanto, $P(A) \leq P(B)$.



4 Teorema da soma: Se A e B são dois eventos quaisquer, então:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (45)$$

Demonstração:

a) A e B são mutuamente exclusivos $P(\bar{A} \cap B) = 0$ recai no axioma III.

b) Para o caso em que $\bar{A} \cap B \neq \Phi$.

Os eventos A e $(\bar{A} \cap B)$ são mutuamente exclusivos. Logo, pelo axioma III:
$P[A \cup (\bar{A} \cap B)] = P(A \cup B) = P(A) + P(\bar{A} \cap B)$. Mas, B é a união dos eventos mutuamente exclusivos $(B \cap \bar{A})$ e $(B \cap \bar{A})$; logo $P(B) = P(\bar{A} \cap B) + P(\bar{A} \cap B)$.
Substituindo o valor de $P(\bar{A} \cap B) = P(B) - P(\bar{A} \cap B)$ na expressão anterior resulta:

$$P(A \cup B) = P(A) + P(B) - P(\bar{A} \cap B) \qquad (46)$$

Probabilidades Finitas dos Espaços Amostrais Finitos

Seja S um espaço amostral finito $S = \{a_1, a_2, ..., a_n\}$. Considere-se o evento formado por um resultado simples
$A = \{a_i\}$.

A cada evento simples $\{a_i\}$ associa-se um número pi denominado probabilidade de {ai} satisfazendo as seguintes condições:
a) $p_i \geq 0 \quad i = 1, 2, ..., n$
b) $p_1 + p_2 + ... + p_n = 1$

A probabilidade P(A) de cada evento composto (mais de um elemento) é então definida pela soma das probabilidades dos pontos de A.

Exemplo:

Três cavalos, A< B e C, estão em uma corrida; A tem duas vezes mais probabilidade de ganhar que B que tem duas vezes mais probabilidade de ganhar que C. Quais são as probabilidades de vitória de cada um, isto é, P(A), P(B) e P(C)?

Resolução:

Faça $P(C) = p$; como B tem duas vezes mais probabilidade de ganharque C, logo $P(B) = 2p$ e assim

$P(A) = 2P(B)= 2.\ 2p = 4p$. Como a soma das probabilidades é 1, então:

$$p + 2p + 4p = 1 \text{ ou } 7p = 1 \text{ ou } p = 1/7 \text{ ; logo temos:}$$

$$P(A) = 4p = 4/7 \text{ ; } P(B) = 2p= 2/7 \text{ e } P(C) = p = 1/7 .$$

Qual seria a probabilidade de B ou C ganhar?

$$P(A \cup B) = P(A) + P(B), \text{ o axioma III com } A \cap B = \Phi.$$

Logo, $P(B \cup C) = P(B) + P(C) = 2/7 + 1/7 = 3/7$.

Espaços Amostrais Finitos Equiprováveis

Quando a mesma probabilidade se associa a cada ponto amostral, o espaço amostral chama-se equiprovável ou uniforme. Em particular, se S contém n pontos, então, a probabilidade de cada ponto será $1/n$.

Por outro lado, se um evento A contém r pontos, então:

$$P(A) = r \cdot \left(1/n \right) = r/n .$$

Este método de avaliar P(A) é frequentemente enunciado da seguinte maneira:

$$P(A)=\frac{\text{nº de vezes em que o evento A pode ocorrer}}{\text{nº de vezes em que o espaço amostral S ocorre}} \qquad (47)$$

$$P(A) = \frac{\text{nº de casos favoráveis}}{\text{nº total de casos}} \qquad (48)$$

Como se observa, o cálculo de probabilidades de um evento reduz-se a um problema de contagem (visto anteriormente).

Combinação de r elementos tomados (combinados) p a p (p≤ r), calcula-se por:

$$C_{r,p} = \binom{r}{n} = \frac{r!}{p!(r-p)!} \qquad (49)$$

Onde: r! = r (r – 1) (r – 2) . . . 1
p! = p (p – 1) (p – 2) . . . 1
Admite-se que o! = 1

Exemplo 1:

Quantas comissões de três pessoas podem-se formar com um grupo de 10 pessoas?

$$C_{10,3} = \frac{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10!}{3!7!} = \frac{10 .9 .8 .7!}{3.2.1.7!} = 120$$

Exemplo 2:

Em um lote de 12 peças, 4 são defeituosas; 2 peças são retiradas aleatoriamente. Calcule:

a) A probabilidade de ambas serem defeituosas;
b) A probabilidade de ambas não serem defeituosas;
c) A probabilidade de ao menos uma peça ser defeituosa.

Solução:
A = {ambas são defeituosas}

A pode ocorrer $\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4 .3.2!}{2.1.2!} =$ 6 vezes

S pode ocorrer $\binom{12}{2} = \frac{12!}{2!(12-2)!} = \frac{12 .11.10!}{2 .1.10!} =$ 66 vezes

Logo, $P(A) = \frac{N.C.F.}{N.T.C.} = \frac{6}{66} = \frac{1}{11}$

B = {ambas não defeituosas}

B pode ocorrer $\binom{8}{2} = \frac{8!}{2!(8-2)!} = \frac{8.7.6!}{2.6!}$ = 28 vezes

S pode ocorrer $\binom{12}{2}$ = 66 vezes

Logo, $P(B) = \frac{28}{66} = \frac{14}{33}$

C = {ao menos uma é defeituosa}
C = é o complemento de B,C = $\overline{B}$
Logo, $p(C) = 1 - P(B) = 1 - \frac{14}{33} = \frac{33-14}{33} = \frac{19}{33}$

Probabilidade condicional

Dados dois eventos, A e B, denota-se P(A/B) a probabilidade condicionada do evento A, quando B tiver ocorrido, por:

$$P(A) = \frac{P(A \cap B)}{P(B)} \qquad (50)$$

Com $P(B) \neq 0$, pois B já ocorreu.

Pode-se constatar que P(A/B), assim definida, satisfaz os axiomas da probabilidade já mencionados.

Para aplicações, convém usar uma fórmula mais prática para o cálculo da probabilidade condicional, assim:

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

$$P(A/B) = \frac{n(A \cap B)}{n(B)} \qquad (51)$$

Exemplo:

Dois dados são lançados. Consideremos os eventos:
$A = \{(x_1,x_2) / x_1 + x_2 = 10\}$ e $B = \{(x_1,x_2 / x_1 > x_2\}$
Onde $x_1 \in D_1$ e $x_2 \in D_2$. Calcular P(A), P(B); P(A/B) e P(B/A).

Solução:

$S = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2),..., (6,6)\}$
Evento $A = \{(4,6), (5,5), (6,4)\}$
Logo, $P(A) = \dfrac{n(A)}{n(S)} = \dfrac{3}{36} = \dfrac{1}{12}$.

Evento $B= \{(2,1), (3,1), (3,2), (4,1), (4,2), (4,3), (5,1), (5,2)\ (5,3), (5,4), (6,1), (6,2), (6,3), (6,4), (6,5)\}$
Logo, $P(B) = \dfrac{n(B)}{n(S)} = \dfrac{15}{36} = \dfrac{5}{12}$

$(A \cap B) = \{(5,5)\}$. Logo $P(A/B) = \dfrac{n(A \cap B)}{n(B)} = \dfrac{1}{15}$

$P(B/A) = \dfrac{n(B \cap A)}{n(A)} = \dfrac{1}{3}$

Teorema do produto

A probabilidade da ocorrência simultânea de dois eventos, A e B, do mesmo espaço amostral, é igual ao produto da probabilidade de um deles pela probabilidade condicional do outro, dado o primeiro.

Assim:
$P(A/B) = (P(A \cap B))/(P(B)) \quad \Rightarrow \quad P(A \cap B)=P(B).P(A/B) \quad (52)$

$P(B/A) = (P(B \cap A))/(P(A)) \quad \Rightarrow \quad P(B \cap A) = P(A).P(B/A) \quad (53)$

Exemplo:

Em um lote de 12 peças, 4 são defeituosas, 2 peças são retiradas uma após a outra sem reposição. Qual a probabilidade de que ambas sejam boas?

Solução:
A = {a primeira peça é boa}
B = {a segunda peça é boa}

$$P(A \cap B) = P(A).\ P(B/A) = \frac{8}{12} \cdot \frac{7}{11} = \frac{56}{132} = \frac{14}{33}$$

Independência estatística

Um evento A é considerado independente de outro evento B se a probabilidade de A é igual à probabilidade condicional de A dada B, isto é, se:

$$P(A) = P(A/B) \qquad (54)$$

É evidente que, se A é independente de B, B é independente de A; assim:

$$P(B) = P(B/A) \qquad (55)$$

Considerando o teorema do produto, pode-se afirmar que: se A e B são independentes, então:

$$P(A \cap B) = P(A).\ P(B) \qquad (56)$$

Teorema de Bayes

Seja A_1, A_2,... , Na, n eventos mutuamente exclusivos tais que $A_1 \cup A_2 \cup ... An = S$. Sejam P(Ai) as probabilidades conhecidas dos vários eventos, e b um evento qualquer de S tal que são conhecidas todas as probabilidades condicionais $P(B/A_i)$.

Então, para cada i, tem-se:

$$P(A_i/B) = \frac{P(A_i).\ P(B/A_i)}{P(A_1)P(B/A_1)+P(A_2)P(B/A_2)+...+P(Na)P(B/A_n)} \qquad 57$$

O resultado acima é muito importante, pois relaciona probabilidades a priori $P(A_i)$ com probabilidades a posteriori $P(A_i/B)$, probabilidade de Ai depois que ocorrer B.

Exemplo:

Admita a seguinte configuração:

Escolheu-se uma urna ao acaso, e dela extraiu-se uma bola ao

Urnas	U_1	U_2	U_3
Cores			
Pretas	3	4	2
Brancas	1	3	3
Vermelhas	5	2	3

acaso, verificando-se que a bola é branca. Qual a probabilidade de a bola ter vindo da urna 2? E da urna 3?

Solução:

$$P(U_1) = \frac{1}{3}\ ;P(U_2) = \frac{1}{3}\ \text{ e } P(U_3) = \frac{1}{3}$$

$$P(br/U_1) = \frac{1}{9};\ P(br/U_2) = \frac{3}{9} = \frac{1}{3} \text{ e } P(br/U_3) = \frac{3}{8}$$

Apicando o Teorema de Bayes, temos:

$$P(U_2/br) = \frac{P(U_2)\ P(br/U_2)}{P(U_1)P(br/U_1)+P(U_2)P(br/U_2)+P(U_3)P(br/U_3)}$$

Ou seja, probabilidade a priori de U_2 era 1/3. Dada a informaçãoao

$$= \frac{\frac{1}{3}\cdot\frac{1}{3}}{\frac{1}{3}\cdot\frac{1}{9}+\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{3}{8}} = \frac{24}{59}$$

que saiu uma bola branca, a probabilidade a posteriori de U2 será 24/59.

$$P(U_3/br) = \frac{P(U_3).\ P(br/U_3)}{P(U_1)P(br/U_1)+P(U_2)P(br/U_2)+P(U_3)P(br/U_3)}$$

Como $P(U_1/br)+P(U_2/br)+P(U_3/br) = 1$

$$= \frac{\frac{1}{3}\cdot\frac{3}{8}}{\frac{1}{3}\cdot\frac{1}{9}+\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{3}{8}} = \frac{27}{59}$$

Tem-se que: $P(U_1/br) = 1 - (24/59 + 27/59) = 8/59$.

Variáveis aleatórias

Definição*

Exemplo:

Seja E: lançamento de duas moedas;

X: nº de caras obtidas nas duas moedas;

S = {k,k), (k,c), (c,k), (c,c)}

X = 0 → corresponde ao evento (c, c) com probabilidade 1/4;

X = 1 → corresponde ao evento (k, c), (c, k) com probabilidade 2/4

X = 2 → corresponde ao evento (k, k) com probabilidade 1/4

Observações importantes:

Observe que é uma função cujo domínio é S e contradomínio é R;

Nas aplicações, é conveniente trabalhar com números e não com eventos; daí o uso da variável aleatória;

Se S é numérico, então X(s) = s;

Uma variável aleatória X será discreta se o número de valores possíveis de X (contradomínio) for finito ou infinito enumerável. Caso seu contradomínio seja um intervalo ou uma coleção de intervalos, ela será uma variável contínua.

SAIBA MAIS

Variável aleatória é a função que associa a todo evento, pertencente a uma partição do espaço amostral, um único número real. Indicaremos por: X: x_1 x_2, ... , x_n.

Função de probabilidades

Ao conjunto $\{(x_i, p(x_i)), i = 1, ..., n\}$ damos o nome de distribuição de probabilidades da variável aleatória X como no quadro resumo abaixo:

Tabela 4.1

x	0	1	2
P(x)	1/4	1/2	1/4

Fonte: Fonseca, 2006.

É importante verificar que para haver uma distribuição de

probabilidades de uma variável aleatória X é necessário que:

SAIBA MAIS

Função de probabilidades

É a função que associa a cada valor assumido pela variável aleatória, a probabilidade do evento correspondente, isto é, $P(X=x_i) = P(A_i)$, i=1,2,...,n.

$$\sum_{i=1}^{n} P(Xi) = 1$$

Exemplos:

$X \rightarrow$ V. A. pontos de um dado
$Y = X + X \rightarrow$ V. A. soma dos pontos de dois lançamentos
$Z = \text{Max}\{(x_1, x_2)\}$ pontos de dois dados.
Determinar a distribuição de probabilidade de X, Y e Z.

Solução:

- A distribuição de probabilidade de X dada por uma tabela será:

x	1	2	3	4	5	6
P(x)	1/6	1/6	1/6	1/6	1/6	1/6

Fonte: Fonseca, 2006.

- Para se obter a distribuição de probabilidade de Y convém construir o espaço amostral para Y:

S = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1),..., (6,6)} $\rightarrow$ n(S) = 36

Para y = 2, temos P{(1,1)} = $\frac{1}{36}$

Para y = 3, temos P{(1,2), (2,1)} = $\frac{2}{36}$

.
.
.

Para y = 8, temos P{(2,6), (3,5), (4,4), (5,3), (6,2)} = 5/36
Para y = 9, temos P(y=9) = 4/36 e assim por diante até chegar em
Para y = 12, teremos P(y=12) = 2/36, logo a distribuição de probabilidade de y será:

y	2	3	4	5	6	7	8	9	10	11	12
P(y)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Fonte: Fonseca, 2006.

- Enquanto a distribuição de Z poderá ser expressa pela tabela obtida da seguinte maneira:

Para z=1, temos P{(1,1)} = 1/36

Para z=2, temos P{(1,2), (2,1), (2,2)} = 3/36

Para z=3, temos P{(1,3), (2,3), (3,1), (3,2), (3,3)} = 5/36

Para z=4, temos P{(1,4), (2,4), (3,4), (4,1), (4,2), (4,3), (4,4)} = 7/36

Para z=5, temos P{z=5} = 9/36

Para z=6, temos P{z=6} = 11/36

Logo, a tabela de z será:

z	1	2	3	4	5	6
P(z)	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Fonte: Fonseca, 2006.

Exercícios Propostos 01

1. Usando-se 5 dos algarismos 1, 2, 3, 4, 5, 6 e 7, sem repeti-los, quantos números pares podemos formar?

2. Resolva a equação (n – 4)! = 120.

3. Quantos números de três algarismos, sem repetição, podemos formar com os algarismos 1, 2, 3, 4, 5, 6, 7, 8, e 9, incluindo sempre o algarismo 4?

4. Quantos números de 4 algarismos distintos podemos formar com os algarismos 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9?

5. Quantos anagramas da palavra EDITORA:
a) Começam com A?
b) Começam com A e terminam com E?

6. Em relação à palavra ESCOLA:
a) Quantos anagramas podem ser formados?
b) Quantos anagramas não possuem duas consoantes ou duas vogais juntas?
c) Quantos anagramas possuem as 3 vogais sempre juntas?

7. Com 10 espécies de frutas, quantos tipos de saladas, contendo 6 espécies diferentes, podem ser feitas?

8. Em uma sala temos 5 rapazes e 6 moças. Quantos grupos de 2 rapazes e 3 moças podemos formar?

9. Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{4}$ e A e B mutuamente exclusivos, calcular:
a) $P(\overline{A})$ b) $P(\overline{B})$ c) $P(A \cup B)$ d) $P(A \cap B)$

10. Um número é escolhido aleatoriamente dentre os números 1, 2, 3, ..., 50. Qual a probabilidade de:
a) O número ser divisível por 5.
b) Terminar em 3.
c) Ser primo.
d) Ser divisível por 6 ou 8.

11. Um lote é formado por 10 peças boas, 4 com defeitos e 2 com defeitos graves. Uma peça é escolhida ao acaso. Calcule a probabilidade de que:
a) Ela não tenha defeitos graves.
b) Ela não tenha defeitos.
c) Ela não tenha defeitos ou tenha defeitos graves.

12. Uma urna contém 5 bolas brancas e 6 pretas. Três bolas são retiradas. Calcular a probabilidade de:
a) Todas pretas.
b) Exatamente uma branca.

c) Ao menos uma preta.

13. Dado $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{3}$; $P(A \cap B) = \frac{1}{4}$ calcular:

a) $P(A \cup B)$

b) $P(A/B)$

c) $P(B/A)$

d) $P[(A \cap B)/B]$

14. Consideremos 250 alunos que cursam o primeiro período do Curso de Sistema de Informações da UAPI. Destes alunos, 100 são homens (H) e 150 mulheres (M), 110 cursam Física (F) e 140 fazem o curso de

Disciplina / Sexo	F	Q	Total
H	40	60	100
M	70	80	150
Total	110	140	250

Química (Q). A distribuição dos alunos é a seguinte:

Um aluno é sorteado ao acaso. Qual a probabilidade de que esteja cursando Química, dado que é mulher?

15. A e B jogam 120 partidas de xadrez, das quais A ganha 60, B ganha 40 e 20 terminam empatadas. A e B concordam em jogar 3 partidas. Determinar a probabilidade de:

a) A ganhar todas as três;

b) A e B ganharem alternadamente.

16. Lançam-se três moedas. Seja X: nº de ocorrências da face cara. Determinar Distribuição de probabilidade de X.

17. Uma urna tem 4 bolas brancas e 3 pretas. Retiram-se 3 bolas sem reposição. Seja X: nº de bolas brancas. Determinar a distribuição de Probabilidade de X.

Resumindo

As primeiras atividades matemáticas da humanidade estavam ligadas à contagem de objetos de um conjunto, enumerando seus elementos. A análise combinatória é a área de Matemática que trata dos problemas de contagem. Vamos estudar algumas técnicas para a descrição e a contagem de todos os casos possíveis de um acontecimento. Todas as vezes que se estudam fenômenos de observação, cumpre-se distinguir o próprio fenômeno e o modelo matemático (determinístico ou probabilístico) que melhor o explique. Os fenômenos estudados pela Estatística são fenômenos cujo resultado, mesmo em condições normais de experimentação, varia de uma observação para outra, dificultando, dessa maneira, a previsão de um resultado futuro. Para a explicação desses fenômenos – fenômenos aleatórios – adota-se um modelo matemático probabilístico. Nesta unidade abordaremos sobre variável aleatória discreta, função de probabilidade e distribuição de probabilidades.

RESPOSTAS DOS EXERCÍCIOS

1ª Lista de Exercícios UNIDADE I

1. C
2. B
3. C
4. C
5. A
6. C
7. B
8. A
9.10.

x_i	10	11	12	13	14	15	16	17
f_i	1	3	4	5	7	2	1	1

a) 30
b) 5
c) 3; 6; 6; 13; 9; 3
d) 150,5; 155,5; 160,5; 165,5; 170,5; 175,5.
11.
a) Fi↓ : 40; 70; 80; 95; 100.
b) 40
c) 80
d) 0,15 ou 15%
Fi↑ : 100; 60; 30; 20; 5
e) 0,70 ou 70%
12.
a) i = 5 e At = 19
b)

Mm(chuvas)	f_i
140→ 144	7
144→148	3
148→152	4
152→156	1
156→160	5
Total	20

c) Fi↓ : 7; 10; 14; 15; 20   Fi↑: 20; 13; 10; 6;5
d) falta
e) falta
13.
a) 24
b) 90
c) 3,00
d) 2,95
e) 0,05
f) 0,5
g) 2,875
h) i = 10   i) 26   j) 3,33%

2ª Lista de Exercícios UNIDADE II

1.
a) 3,30
b) 1,60
c) 2,67
2. $Q_1 = 0,625$ e $Q_3 = 4,5$
3.
a) 1,705m
b) 1,725m
c) 1,725m
d) 1,713m
4.
a) 16,475
b) 16, 494
c) 16,650
d) 17,00
e)
f) 25%

5.
a) 75000
b) 75000
c) 67500
d) 82500
e) 75000
f) 82500

3ª Lista de Exercícios UNIDADE III

1.
a) 1,88
b) 2,02
c) 2,48
2. b
3.
a) 3,767
b) 14,190289

c) 0,353 ou 35,3%

4ª Lista de Exercícios UNIDADE IV

1. 1080
2. n = 9
3. 168
4. 4536
5.
a) 720
b) 120
6.
a) 720
b) 72
c) 144
7. 210
8. 200
9.
a) 1/2
b) 3/4
c) 3/4
d) 0
10.
a) 1/5
b) 1/10
c) 3/10
d) 0,24
11.
a) 7/8
b) 5/8
c) 3/4
12.
a) 4/33
b) 5/11
c) 31/33
13.
a) 7/12
b) 3/4
c) 1/2

d) 1

14. 80/150

15.

a) 1/8

b) 5/36

16.

x	P(x)
0	1/8
1	3/8
2	3/8
2	1/8
	1

17.

x	P(x)
0	27/343
1	108/343
2	144/343
2	64/343

Referências

COSTA NETO, Pedro de Oliveira. **Estatística**. 2. ed. São Paulo: Blucher, 2002.

CRESPO, Antonio Arnout. **Estatística fácil**. 17. ed. São Paulo: Saraiva, 2002.

FONSECA, Jairo Simon da. **Curso de Estatística**. 10. ed. Reimp. Martins, Gilberto de Andrade. São Paulo: Atlas, 2006.

LAPPONI, Juan Carlos. **Estatística usando Excel**. 4. ed. Reimpressão. Rio de Janeiro: Elsevier, 2005.

MEYER, Paul L. **Probabilidade: aplicações à Estatística**. 2. ed. Rio de Janeiro: LTC, 2000.

MORETIN, Pedro Alberto. **Estatística básica**. 5. ed. São Paulo: Saraiva, 2006.

SPIEGEL, M. R. **Probabilidade e Estatística**. 1. ed. São Paulo: McGraw-Hill, 2001.

TOLEDO, Geraldo Luciano. **Estatística básica**. 2. ed. São Paulo: Atlas, 1985.

Minicurrículo

Juarez Rodrigues Martins

Especialista em Matemática (2001) e em Estatística (2008) pela Universidade Federal do Piauí. Graduou-se em Biologia pela Universidade Estadual do Piauí (1992) e em Matemática pela Universidade Federal do Piauí (1995). Foi professor substituto na Universidade Federal do Piauí, durante quatro anos, nos períodos de março de 2003 a março de 2005, e abril de 2007 a abril de 2009. Foi professor do Ensino Médio da rede pública estadual do Piauí e foi professor da rede particular de ensino de Teresina. Atualmente é professor efetivo permanente, com dedicação exclusiva, da Universidade Federal do Piauí. lotado no CABJ. Tem experiência na área de matemática, com ênfase em matemática, atuando principalmente nos seguintes temas: estatística, probabilidade e matemática.

Email: martins-juarez@bol.com.br.
Web site: http://www.famat.ufu.br/prof/marcelo/exercicios.htm

www.uapi.ufpi.br