

Estatística Vol. I



SUZI SAMÁ PINTO
CARLA SILVA DA SILVA

Estadística

Volume I



UNIVERSIDADE FEDERAL DO RIO GRANDE-FURG

Reitora

CLEUZA MARIA SOBRAL DIAS

Vice-Reitor

DANILO GIROLDO

Chefe do Gabinete da Reitora

JACIRA CRISTIANE PRADO DA SILVA

Pró-Reitor de Extensão e Cultura

DANIEL PORCIUNCULA PRADO

Pró-Reitor de Planejamento e Administração

MOZART TAVARES MARTINS FILHO

Pró-Reitor de Infraestrutura

MARCOS ANTONIO SATTE DE AMARANTE

Pró-Reitor de Graduação

RENATO DURO DIAS

Pró-Reitora de Assuntos Estudantis

DAIANE TEIXEIRA GAUTÉRIO

Pró-Reitora de Gestão e Desenvolvimento de Pessoas

ALINE RODRIGUES DE AVILA

Pró-Reitor de Pesquisa e Pós-Graduação

EDUARDO RESENDE SECCHI

EDITORA DA FURG

Coordenadora

CLEUSA MARIA LUCAS DE OLIVEIRA

COMITÊ EDITORIAL

Presidente

DANIEL PORCIUNCULA PRADO

Titulares

ANDERSON ORESTES CAVALCANTE LOBATO

ANDRE ANDRADE LONGARAY

ANGELICA CONCEIÇÃO DIAS MIRANDA

CARLA AMORIM NEVES GONÇALVES

CLEUSA MARIA LUCAS DE OLIVEIRA

EDUARDO RESENDE SECCHI

ELIANA BADIALE FURLONG

GIONARA TAUCHEN

LUIZ EDUARDO MAIA NERY

MARCELO GONÇALVES MONTES D'OCA

MARCIA CARVALHO RODRIGUES

RAÛL ANDRÉS MENDOZA SASSI

Editora da FURG

Campus Carreiros

CEP 96203 900 – Rio Grande – RS – Brasil

editora@furg.br

Integrante do PIDL

Editora Associada à



Suzi Samá Pinto e Carla Silva da Silva

Estadística

Volume I



Rio Grande
2020

© **Suzi Samá Pinto e Carla Silva da Silva**

2020

Capa: Anderson Mendonça

Formatação eletrônica da capa: Anderson Mendonça

Ajustes da capa: Anael Macedo

Formatação e diagramação:

João Balansin

Gilmar Torchelsen

Revisão: João Reguffe

Dados Internacionais de Catalogação na Publicação (CIP)

Ficha catalográfica elaborada pela Bibliotecária Marcia Carvalho Rodrigues,
CRB 10/1411.

P659e Pinto, Suzi Samá
Estatística [recurso eletrônico] : volume I / Suzi Samá
Pinto e Carla Silva da Silva. – Dados eletrônicos. – Rio
Grande: Ed. da FURG, 2020.

Modo de acesso:

<http://repositorio.furg.br/handle/1/8845>

Edições anteriores em formato de livro impresso.

ISBN: 978-65-5754-009-1 (eletrônico)

1. Estatística matemática. 2. Probabilidade. I. Silva,
Carla Silva da. II. Título.

CDU, 2. ed.: 519.2

Índice para o catálogo sistemático:

1. Estatística matemática	519.2
2. Probabilidade	519.2

SUMÁRIO

PREFÁCIO	9
1. Introdução à Estatística	11
1.1 Campos da Estatística	12
1.2 População e amostra	13
1.3 Métodos de amostragem	15
1.3.1 Métodos de amostragem probabilísticos	17
1.3.2 Métodos de amostragem não-probabilísticos	20
1.4 Variáveis e dados estatísticos	22
1.5 Planejamento de um estudo estatístico	24
Exercícios complementares	25
Respostas	27
2. Apresentação e organização de dados	30
2.1 Tabelas para dados qualitativos	31
2.2 Gráficos estatísticos para dados qualitativos	33
2.3 Tabelas para dados quantitativos	37
2.3.1 Outra forma de montar uma distribuição de frequência com intervalos de classe	43
2.3.2 Tipos de frequência	44
2.4 Gráficos para dados quantitativos	46
Exercício resolvido	51
Exercícios complementares	58
Respostas	60
3. Medidas estatísticas para dados não-agrupados	64
3.1 Medidas de posição ou tendência central	64

3.1.1 Média	65
3.1.2 Mediana	67
3.1.3 Moda	69
3.2 Medidas de dispersão	70
3.2.1 Amplitude total	71
3.2.2 Desvio médio	72
3.2.3 Variância	73
3.2.4 Desvio-padrão	75
3.2.5 Coeficiente de variação de Pearson	76
3.2.6 Quartis e percentis	78
3.2.6.1 Gráfico de caixa ou o esquema dos cinco números	80
3.3 Assimetria	83
3.4 Curtose	84
Exercícios resolvidos	87
Exercícios complementares	93
Respostas	96
4. Medidas estatísticas para dados agrupados	99
4.1 Dados agrupados sem intervalos de classe	99
4.2 Dados agrupados com intervalos de classe	102
Exercícios resolvidos	105
Exercícios complementares	110
Respostas	112
5. Introdução à probabilidade	115
5.1 Experimento	116
5.2 Espaço amostral	116
5.3 Evento	117
5.4 Tipos de evento	117
5.5 Operação com eventos	117
5.6 Tipos de probabilidade	118
5.6.1 Probabilidade clássica	118
5.6.2 Probabilidade empírica ou frequentista	120

5.7 Teorema da soma	121
5.8 Probabilidade do complementar de um evento	124
5.9 Probabilidade condicional	125
5.10 Eventos independentes	127
5.11 Teorema do produto	128
5.12 Tabela de contingência	130
Exercícios resolvidos	133
Exercícios complementares	135
Respostas	141
6. Variáveis aleatórias	146
6.1 Variável aleatória discreta	146
6.1.1 Média ou valor esperado	148
6.1.2 Variância e desvio-padrão	150
6.1.3 Exemplo de aplicação	150
Exercícios resolvidos	152
Exercícios complementares	154
Respostas	155
7. Distribuição de probabilidade para variáveis aleatórias discretas	157
7.1 Distribuição binomial	157
7.1.1 Parâmetros da distribuição binomial	160
7.2 Distribuição de Poisson	161
7.2.1 Parâmetros da distribuição de Poisson	162
Exercícios resolvidos	165
Exercícios complementares	167
Respostas	170
8. Distribuição de probabilidade para variáveis aleatórias contínuas	171
8.1 Distribuição normal	173
8.1.1 Características da curva normal	174

8.1.2 Variável normal padrão ou variável padronizada	174
8.1.3 A curva normal padrão	175
8.1.4 Uso da tabela da curva normal padrão	176
8.1.5 Exemplos de aplicação	179
Exercício resolvido	180
Exercícios complementares	184
Respostas	187
Bibliografia	189
Anexos I	190
Anexos II	192
Anexos III	194

Prefácio

Esta obra visa a disponibilizar um material adequado à ementa de várias disciplinas oferecidas na Universidade Federal do Rio Grande (FURG), bem como propiciar aos estudantes de graduação e pós-graduação um material de apoio que auxilie no seu processo de ensino-aprendizagem.

Este livro foi iniciado com base no material que já vínhamos construindo no contexto da sala de aula na modalidade presencial. Devido à necessidade de preparar um material didático para as disciplinas de Estatística oferecidas pela FURG na modalidade a distância, foram feitas adaptações. Os textos foram redigidos de forma a facilitar a compreensão dos temas tratados, possibilitando ao estudante uma aprendizagem independente e autônoma.

A contribuição dos estudantes e dos tutores a distância ao longo do curso possibilitaram que este trabalho fosse revisado e melhorado.

O primeiro capítulo apresenta alguns conceitos de estatística, métodos de amostragem, os tipos de variáveis e planejamento de um estudo estatístico. No segundo capítulo, é discutida a forma de organização dos dados. São apresentadas as tabelas e os gráficos para dados qualitativos, bem como a distribuição de frequência e o histograma, que consistem num tipo de organização particular em que os dados quantitativos são agrupados segundo a frequência com que ocorrem. Além disso, traz os cuidados na montagem e interpretação dos gráficos, pois estes podem distorcer a informação quando mal elaborados. O terceiro e o quarto capítulo tratam das medidas estatísticas, como média, desvio-padrão, variância,

assimetria e curtose. Ainda o terceiro capítulo trata das medidas estatísticas para dados não-agrupados e o quarto capítulo, das medidas estatísticas para dados agrupados. O quinto capítulo começa com uma abordagem dos conceitos básicos de probabilidade, indo até a discussão do teorema da soma, teorema do produto e a probabilidade condicional. O sexto capítulo apresenta as variáveis aleatórias com ênfase nas variáveis aleatórias discretas. O sétimo capítulo discute dois modelos probabilísticos de variáveis aleatórias discretas: Binomial e Poisson. O oitavo capítulo apresenta a distribuição normal. Essa distribuição é importante em estatística, pois muitas das técnicas de inferência partem da premissa de que as variáveis apresentam distribuição normal.

Em cada capítulo foram elaborados exemplos e exercícios resolvidos. Também são disponibilizados exercícios complementares, com a finalidade de que o estudante avalie sua aprendizagem a cada etapa. As respostas desses exercícios são apresentadas ao final de cada capítulo.

Ficam aqui registrados nossos agradecimentos a todas as pessoas que contribuíram de diferentes formas para o desenvolvimento deste livro. E um agradecimento muito especial aos nossos familiares pela compreensão nas horas perdidas de convívio para a elaboração deste livro.

Desde já agradecemos as contribuições e sugestões futuras a fim de aperfeiçoar este trabalho.

As autoras

1 Introdução à Estatística

A Estatística é a ciência que se ocupa de coletar, organizar, analisar e interpretar dados a fim de subsidiar a tomada de decisões. Sua grande importância reside no fato de ter um vasto campo de aplicação, na administração, educação, medicina, engenharia, política, entre outras ciências e atividades.

Todos os dias nos são apresentados dados estatísticos, seja pela televisão, jornais, rádio ou Internet. Basta a divulgação de pesquisas sobre taxas de desemprego, índice de popularidade de governantes, índice de audiência de programas de televisão, para que algumas pessoas já contestem sua validade. É claro que pesquisas feitas de forma incorreta, assim como dados mal coletados, podem levar a decisões equivocadas. Mas, quando um estudo estatístico é feito de forma criteriosa, seus resultados permitem obter conclusões confiáveis.

No passado, o grande obstáculo para o uso da Estatística eram os cálculos que exigiam horas de trabalho cansativo. Com o avanço da tecnologia esse obstáculo praticamente desapareceu. Entretanto, não basta ter um computador ou uma calculadora que elabore gráficos, tabelas e realize cálculos: é necessário que tenhamos conhecimento para compreendê-los, bem como capacidade para construí-los. Portanto, é importante o conhecimento de conceitos básicos para que utilizemos de forma adequada as diferentes técnicas estatísticas.

1.1 Campos da Estatística

A Estatística está dividida em três campos: Estatística Descritiva, Probabilidade e Estatística Inferencial.

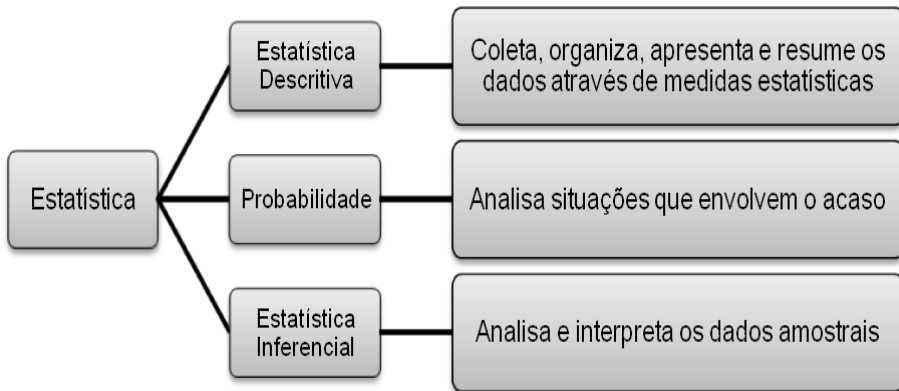


A **Estatística Descritiva** tem por objetivo a coleta, a organização, a apresentação dos dados em tabelas ou gráficos, assim como a obtenção de algumas medidas como média, proporção, dispersão que auxiliam na descrição de um conjunto de dados.

Embora a Estatística Descritiva seja amplamente usada em várias áreas do conhecimento, não vai além da informação fornecida pelos dados; nenhuma inferência é feita além dos próprios dados. Por exemplo, se cinco alunos de uma turma obtêm as seguintes notas numa avaliação: 8, 9, 7, 7, 9, é possível, por meio da Estatística Descritiva, resumir esta informação calculando a média das notas destes cinco alunos, que neste exemplo será 8, mas não podemos concluir que a média de toda a turma seja 8.

A análise de amostras com a finalidade de inferir sobre a população exige generalizações que vão além dos dados coletados. Como as amostras são formadas por apenas alguns elementos da população, há incertezas com relação às conclusões que a podemos chegar sobre as características de uma população por meio da análise da amostra. Como a **probabilidade** é uma ferramenta importante para a análise de qualquer situação que envolva incerteza, ela é fundamental na inferência estatística. A **Estatística Inferencial**

é um conjunto de técnicas que obtém informações sobre uma população a partir de resultados observados numa amostra.



1.2 População e amostra

Num estudo estatístico, não interessa analisar unidades individuais, mas sim grupos ou conjuntos, pois seu objetivo é o estudo de toda a população. Em estatística, **população** é definida como o conjunto de indivíduos, objetos ou produtos que apresentam, pelo menos, uma característica em comum. Assim, nossa população pode ser tanto toda a população de uma cidade, como todas as peças produzidas por uma fábrica. Uma população pode ser pequena, grande ou infinita. O número de observações em uma população finita é designado pela letra **N**.

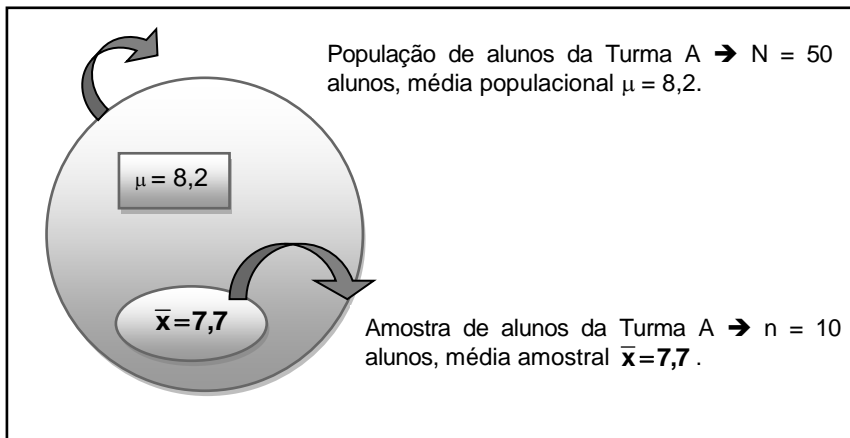
Apesar de algumas populações terem um número finito de elementos, muitas vezes é inviável fazer-se uma análise de todos eles, seja pela distribuição geográfica, pelo custo elevado, pelo tempo que a pesquisa demandaria, ou em estudos em que o elemento é destruído no experimento. Por isso, geralmente os pesquisadores trabalham com amostras que consistem de uma parte

da população selecionada para o estudo. Neste caso, escolhe-se uma parte da população para o estudo, chamada **amostra**, com **n** elementos, com as mesmas características da população. O objetivo de selecionar uma amostra é obter conclusões que possam ser generalizadas para a população, isto é, possam ser inferidas.

Exemplo: Uma pesquisa pretende estudar a opinião de estudantes com relação à mudança do sistema de avaliação em uma universidade. Neste caso, a população é formada por todos os estudantes matriculados na instituição. Para compor a amostra os pesquisadores entrevistaram 200 estudantes.

As medidas estatísticas obtidas com base em uma população são denominadas **parâmetros populacionais**, representadas, geralmente, por letras gregas. Por exemplo, a média da população é representada pela letra grega μ (lê-se mi). As medidas estatísticas obtidas com base em amostras são denominadas **estimadores ou estatísticas amostrais**. Os valores numéricos assumidos pelos estimadores ou estatísticas amostrais são chamados **estimativas**. Em uma definição mais simplista, o parâmetro é obtido a partir da análise de toda a população e o estimador ou estatística amostral é obtido a partir da análise da amostra.

Exemplo: Um professor calcula a média das notas obtidas entre os 50 alunos da Turma A. Se considerarmos a Turma A como a população a ser considerada, a média obtida é um parâmetro populacional. No entanto, se o professor seleciona 10 alunos entre os 50 alunos da Turma A e calcula a média das notas destes 10, a média é considerada uma estatística amostral ou um estimador.



Como, em geral, os parâmetros são desconhecidos, pois dificilmente é possível fazer uma análise de toda a população, as estatísticas amostrais são usadas para se ter uma idéia do valor dos parâmetros populacionais. Portanto, é muito importante que a amostra seja representativa da população, pois amostras tendenciosas levarão a resultados questionáveis. Não existe técnica estatística capaz de salvar uma amostra mal-coletada, sendo assim, nenhuma inferência estatística poderá ser realizada com base nesta amostra. Antes da coleta dos dados, é preciso definir o método de amostragem mais adequado para escolher os elementos da população que irão compor a amostra.

1.3 Métodos de amostragem

Na maioria dos problemas de inferência estatística, é impossível ou impraticável observar a população inteira. Logo, dependemos de uma amostra de observações da população que ajude a tomar decisões acerca desta. Para que as inferências sejam válidas, é necessário que a amostra seja representativa da população.

Existem dois tipos de métodos de amostragem: os **métodos probabilísticos** e os **não-probabilísticos**. Os métodos probabilísticos exigem que cada elemento ou indivíduo da população tenha uma probabilidade conhecida (não-nula) de ser selecionado para compor a amostra. Além disso, a amostragem é realizada com seus elementos selecionados independentemente uns dos outros (um por vez). Os resultados podem ser inferidos para a população. Nos métodos não-probabilísticos, a escolha destes elementos depende em parte do julgamento do pesquisador ou do entrevistador – há escolha deliberada dos elementos da amostra. Uma vez que os elementos da população não têm a mesma probabilidade de compor a amostra, na amostragem não-probabilística não é garantida a representatividade da população.

Para a escolha do método de amostragem, o pesquisador deve levar em conta o tipo de pesquisa, a acessibilidade e disponibilidade dos elementos da população, a variabilidade da população, a disponibilidade de tempo para a realização da pesquisa, e os recursos financeiros e humanos disponíveis, entre outros fatores.

Do ponto de vista estatístico, o método probabilístico é o preferível, pois constitui a base da maior parte das técnicas estatísticas.

Quadro 1 – Vantagens e desvantagens dos métodos de amostragem

Método de amostragem	Vantagens	Desvantagens
Probabilístico	Há garantia de representatividade da população	Custo alto Demorado
Não-probabilístico	Conveniência Rapidez Baixo custo	Falta de exatidão devido a tendência na seleção. Não há garantia de representatividade da população.

1.3.1 Métodos de amostragem probabilísticos

Dentre **os métodos de amostragem probabilísticos**, podemos citar a amostragem aleatória simples, amostragem sistemática, amostragem estratificada e amostragem por agrupamento ou conglomerados.

Na **amostragem aleatória simples** (AAS), todos os elementos da população têm a mesma probabilidade de compor a amostra. A amostra pode ser selecionada por sorteio (nomes em pedaços de papel), ou, quando a população for muito grande, os elementos podem ser numerados e em seguida selecionados por meio de calculadoras ou computadores.

Exemplo 1: O professor deseja selecionar cinco estudantes para apresentar um trabalho. Escreve o nome dos 30 estudantes em pedaços de papel e seleciona os estudantes por sorteio.

Exemplo 2: O dono de uma locadora de filmes deseja fazer uma pesquisa de satisfação entre seus clientes. A amostra será composta por 50 clientes. A locadora tem 3.000 associados, por isso fica inviável colocar o nome de cada um num pedaço de papel e realizar o sorteio. Como o cadastro dos sócios está organizado por números, o proprietário da locadora utiliza um computador e gera aleatoriamente 50 números de 1 a 3000. Os clientes com o número gerado pelo computador integrarão a amostra.

A **amostragem sistemática** é utilizada quando a população apresenta-se organizada segundo algum critério, de tal modo que cada um de seus elementos possa ser unicamente identificado pela posição (por exemplo, fichas, lista telefônica etc.). Neste método de amostragem, supõe-se que a distribuição dos elementos da população, em uma lista, é aleatória. Neste caso, a amostragem é

realizada por intervalos fixos. Seleciona-se, aleatoriamente, o primeiro elemento, que deve estar entre 1 e o fator de sistematização, depois escolhem-se os membros da amostra a intervalos regulares. O fator de sistematização é obtido por meio da divisão do número de elementos da população (N) pelo número de elementos da amostra (n):

$$\text{fator de sistematização} = \frac{N}{n}$$

Arredonda-se o fator de sistematização para o número inteiro mais próximo.

Exemplo: Ao selecionar-se uma amostra sistemática com 30 elementos de uma população de 600 indivíduos, o fator de sistematização será de $600/30 = 20$ intervalos. Um número entre 1 e 20 será escolhido aleatoriamente entre os primeiros indivíduos da população. Suponha-se que tenha sido escolhido o número 7. O sétimo elemento será o primeiro elemento da amostra. As seleções subseqüentes serão 27, 47, 67, 87, ..., 567 e 587.

A **amostragem estratificada** é indicada quando a população encontra-se dividida em grupos distintos (população heterogênea). Dependendo dos objetivos do estudo, a população será dividida em dois ou mais subgrupos, denominados estratos, que compartilham uma característica comum, como sexo, grau de instrução e classe social. Depois que uma população é dividida em estratos apropriados, podemos fazer uma amostra aleatória simples em cada estrato. Os resultados da amostragem podem então ser ponderados e combinados em estimativas apropriadas da população. Com a estratificação obtemos estratos homogêneos internamente e heterogêneos em relação aos outros estratos. Nessa situação, a estratificação gera amostras mais representativas da população.

O número de elementos de cada estrato que constituirão a amostra é calculado com base em duas informações: (1) o tamanho que deve ter a amostra total e (2) como a amostra total deve ser alocada entre os estratos. As amostras dentro de cada estrato podem ser

proporcionais ou desproporcionais ao tamanho do estrato em relação à população.

Exemplo: Uma comunidade universitária é formada por 8.000 indivíduos, entre professores, estudantes e funcionários. Na tabela 1.2 é apresentado o número de indivíduos em cada um destes estratos, proporcional ao seu número na população, considerando uma amostra com 5% dos elementos da população.

Tabela 1.2 – Amostragem estratificada

Estratos	População	Amostra
Professores	800	40
Funcionários	1.200	60
Estudantes	6.000	300
Total	8.000	400

Na **amostragem por agrupamento ou conglomerado**, os elementos da população são divididos em grupos, de forma que cada grupo seja representativo da população total. Uma amostra aleatória simples dos grupos é então obtida, e todos os elementos dentro de cada grupo são analisados. Podemos citar como agrupamentos agências, quarteirões, edifícios ou bairros. A amostragem por agrupamentos resulta em economia de custo, particularmente se a população estiver dispersa por uma extensa área geográfica, pois em um agrupamento muitas observações da amostra podem ser obtidas em tempo relativamente curto, o que possibilita obter um tamanho de amostra maior, com custo total significativamente mais baixo.

Exemplo: Numa pesquisa sobre a desnutrição dos estudantes nas escolas públicas de um município, foram sorteadas aleatoriamente algumas escolas (conglomerados) e todos os estudantes de cada escola participaram da pesquisa.

Quadro 2 – Resumo dos métodos de amostragem probabilística

Tipo	Descrição
Aleatória simples	A seleção pode ser feita por uma lista aleatória de elementos ou por sorteio.
Sistemática	População organizada sob algum critério. Começa com um início aleatório e depois a amostragem é realizada por intervalos fixos.
Estratificada	A população é dividida em estratos homogêneos e amostras aleatórias são selecionadas de cada estrato.
Agrupamento	A população é dividida em seções ou grupos e é obtida uma amostra aleatória dos grupos. Todos os elementos de cada grupo são analisados.

1.3.2 Métodos de amostragem não-probabilísticos

Dentre **os métodos de amostragem não-probabilísticos**, podemos citar a amostragem por conveniência, por julgamento ou por quotas.

Na **amostragem por conveniência**, os elementos ou indivíduos são selecionados com base na sua semelhança presumida com a população e na sua disponibilidade imediata. Frequentemente são entrevistados clientes de estabelecimentos comerciais, possibilitando ao pesquisador fazer contato com grande número de pessoas em curtos períodos de tempo e a baixo custo. Portanto, a amostragem por conveniência tem a vantagem de ser rápida e barata pela fácil seleção da amostra e coleta dos dados, no entanto é difícil avaliar quão representativa da população é essa amostragem.

Exemplo 1: Um programa de televisão libera um número de telefone para que os telespectadores possam ligar e dar sua opinião sobre determinado assunto.

Exemplo 2: Um repórter de TV faz entrevistas na rua.

Na **amostragem por julgamento**, a pessoa mais conhecedora do assunto a ser pesquisado escolhe intencionalmente os indivíduos ou elementos que ela considera representativos da população para comporem a amostra. Com frequência este é um modo relativamente fácil de selecionar uma amostra. No entanto, a qualidade dos resultados da amostra depende do julgamento da pessoa que faz a seleção.

Exemplo 1: Em estudos sobre o assédio sexual no trabalho, o pesquisador pode entrevistar apenas aqueles que sofreram assédio sexual no trabalho e/ou pessoas que trabalham e desenvolvem pesquisas sobre este assunto.

Exemplo 2: Antes de lançar um novo produto no mercado, algumas empresas o testam entre seus funcionários. Isso porque acredita-se que os funcionários terão reações mais favoráveis em relação ao novo produto do que o público. Dessa forma, se o produto não passar por esse grupo, não tem perspectiva de sucesso no mercado em geral.

Na **amostragem por quotas**, o pesquisador procura obter uma amostra que seja similar à população sob determinado(s) aspecto(s) ou dimensão(ões) considerando as características da população, como sexo, idade, classe social, entre outras. A amostra deve possuir proporções similares de pessoas com as mesmas características na população. Se acreditarmos que a resposta a uma pergunta pode variar dependendo do sexo da pessoa, então devemos buscar respostas proporcionais de homens e mulheres. Podemos achar também que as pessoas da classe média têm opinião diferente das pessoas da classe baixa sobre determinado assunto, então isso seria um outro aspecto a ser considerado na coleta da amostra. Portanto, podemos pedir ao entrevistador para encontrar pessoas da classe média, sexo feminino e de determinada faixa etária.

Exemplo: Pesquisas de opinião de mercado e pesquisas de marketing.

Quadro 3 – Resumo dos métodos de amostragem não-probabilística

Tipo	Descrição
Conveniência	Os elementos são selecionados com base na sua semelhança presumida com a população e na sua disponibilidade imediata.
Julgamento	Pesquisador usa o seu julgamento para escolher intencionalmente os indivíduos ou elementos que ele considera representativos da população.
Quotas	O pesquisador entrevista um número predefinido de pessoas segundo determinados aspectos.

1.4 Variáveis e dados estatísticos

Definimos como **variável** a característica que cada indivíduo, objeto ou produto assume (p. ex., sexo, idade, peso) e **dados estatísticos** os diferentes valores ou categorias que cada variável pode assumir (ex., masculino e feminino). As variáveis podem ser qualitativas ou quantitativas.

As **variáveis qualitativas ou categóricas** podem ser separadas em diferentes categorias que se distinguem por alguma característica não-numérica. Podem ser do tipo nominal ou ordinal. Em uma **variável qualitativa nominal**, os dados observados são meramente classificados em várias categorias distintas, usando-se nomes, marcas ou qualidades – por exemplo, cores, religião, disciplinas. Uma **variável qualitativa ordinal** envolve dados, que podem ser classificados em categorias distintas, podendo ser arranjados em ordem, como, por exemplo, classes sociais, grau de instrução, opinião do usuário.

As **variáveis quantitativas ou numéricas** consistem em números que representam contagens ou medidas. Podem ser do tipo discreta ou contínua. Em uma **variável quantitativa discreta** os dados observados assumem apenas valores inteiros, como, por exemplo, número de livros, número de passageiros. Em uma **variável quantitativa contínua** os dados observados podem assumir qualquer valor num certo intervalo, por exemplo, altura, peso, temperatura do ar.

Quadro 4 – Resumo dos tipos de variáveis

Variável	Tipo	Exemplos
Qualitativa	Nominal	Sexo: masculino e feminino Disciplina: matemática, português, história
	Ordinal	Notas: A, B, C, D, E Opinião do usuário: bom, regular ou ruim
Quantitativa	Discreta	Número de peças produzidas, número de alunos, número de erros em faturas
	Contínua	Altura de pessoas, temperatura do ar, peso Notas variando de 0 a 10

Como as variáveis têm naturezas diferentes quanto aos possíveis valores que podem assumir, a escolha do processo a ser utilizado na análise ou descrição de dados estatísticos depende do tipo de variável considerada. Por exemplo, numa pesquisa de opinião em que é perguntado ao usuário se ele está satisfeito com o serviço recebido, se as opções de resposta forem: muito bom, bom, regular ou ruim, não poderemos calcular a média, pois esta variável não é numérica e não poderá ser usada nenhuma técnica estatística baseada na média.

1.5 Planejamento de um estudo estatístico

O objetivo de um estudo estatístico é **coletar dados** para subsidiar a **tomada de decisões**. Se o processo for planejado de forma inadequada, as conclusões não serão confiáveis.

Um bom planejamento fornece respostas para questões como:

- Por quê
- Quem
- Quando
- O quê
- Onde
- Como

Dessa forma, evitam-se surpresas indesejáveis ao final da pesquisa por dados coletados incorretamente, medidas não adequadas às técnicas estatísticas de interesse, entre outras.

Um bom estudo estatístico também depende da correta execução e finalização, que devem seguir as seguintes etapas:

- Coleta dos dados.
- Organização e apresentação dos dados em tabelas ou gráficos que facilitem sua visualização. Resumo dos dados através de medidas estatísticas como média, desvio-padrão, coeficiente de variação etc.
- Análise e interpretação dos dados.
- Elaboração de um relatório de forma que seja facilmente compreendido por quem for fazer uso dos resultados da pesquisa na tomada de decisões.

Muitas pessoas jamais precisarão desenvolver um estudo estatístico, mas provavelmente terão que interpretar os resultados de

algum desses estudos, seja no exercício de sua cidadania, no trabalho ou na sua vida pessoal. Por isso, é importante que todos estejam familiarizados com as etapas do planejamento de um estudo estatístico, para ter condições de analisá-los com criticidade.

Exercícios complementares

1) O que é Estatística?

2) Explique a diferença entre uma variável qualitativa e uma variável quantitativa. Dê um exemplo de cada.

3) Explique a diferença entre uma variável numérica discreta e uma variável numérica contínua. Dê um exemplo de cada.

4) Defina população e amostra.

5) O questionário abaixo foi aplicado aos estudantes de uma universidade que frequentaram a livraria do campus durante a primeira semana de aula de 2008. A pesquisa foi realizada pela Secretaria de Avaliação Institucional da universidade. Identifique o tipo de variável.

a) Idade (em anos)

b) Sexo

c) Número de livros que você costuma ler por ano.

d) Forma de pagamento utilizada na compra:

() em dinheiro () cheque () cartão de crédito () cartão de débito

e) Você classificaria o serviço prestado pela livraria como:

() ótimo () bom () péssimo

f) Valor gasto na compra do(s) livro(s)

6) Determine se cada afirmação é verdadeira ou falsa. Se for falsa, reescreva-a em sua forma verdadeira.

- a) Dados do tipo nominal são qualitativos.
- b) Dados do tipo ordinal são quantitativos.
- c) A estatística descritiva trata de tirar conclusões sobre uma população, a partir de uma amostra.
- d) Uma amostra é um subconjunto de uma população.
- e) Um estimador é uma medida que descreve uma característica de uma população.
- f) Dados categóricos consistem em números que representam contagem ou medidas.

7) Faça a distinção entre métodos de amostragem probabilísticos e não-probabilísticos.

8) Explique rapidamente as características:

- a) da amostragem aleatória simples;
- b) da amostragem por quotas;
- c) da amostragem sistemática.

9) Que considerações um pesquisador deve levar em conta, para escolher o melhor método de amostragem?

10) Elabore um exemplo, em sua área, para um tipo de **amostragem probabilística** e outro exemplo para um tipo de **amostragem não-probabilística**. Primeiro, descreva o objetivo da pesquisa. Com base nesta informação, defina a população. Escolha o **método de amostragem** (probabilístico ou não-probabilístico) e o **tipo de amostragem** (estratificada, conglomerados, julgamento etc.). Justifique sua escolha.

Respostas

1. Estatística é a ciência que se ocupa de coletar, organizar, analisar e interpretar dados, a fim de subsidiar a tomada de decisões.

2. Variáveis qualitativas podem ser separadas em diferentes categorias que se distinguem por alguma característica não-numérica. Exemplo: sexo, cidade, nível de escolaridade. Variáveis quantitativas consistem em números que representam contagens ou medidas. Exemplo: número de alunos na escola, número de livros, temperatura do ar, altura de pessoas.

3. Nas variáveis numéricas discretas, os dados observados assumem apenas valores inteiros, como em contagens, por exemplo: número de filhos, número de carros no estacionamento. Nas variáveis numéricas contínuas, os dados observados podem assumir qualquer valor num certo intervalo; por exemplo: peso, tempo cronometrado numa corrida de carros, estatura.

4. População é todo conjunto de indivíduos, objetos ou produtos que apresentam pelo menos uma característica em comum. Amostra é um subconjunto da população que preserva as mesmas características observadas na população.

- 5. a)** variável quantitativa discreta;
b) variável qualitativa nominal;
c) variável quantitativa discreta;
d) variável qualitativa nominal;
e) variável qualitativa ordinal;
f) variável quantitativa contínua.

6.a) Verdadeira

b) Falsa. Dados do tipo ordinal são qualitativos.

c) Falsa. A estatística descritiva coleta, organiza, apresenta em tabelas e gráficos e resume os dados, através de medidas estatísticas [ou] A estatística inferencial trata de tirar conclusões sobre uma população a partir de uma amostra.

d) Verdadeira

e) Falsa. Um parâmetro é uma medida que descreve uma característica da população.

f) Falsa. Dados quantitativos consistem em números que representam contagem ou medidas.

7. Os métodos probabilísticos exigem que cada elemento ou indivíduo da população possua uma probabilidade conhecida (não-nula) de ser selecionado para compor a amostra. Seus resultados podem ser inferidos para a população. Nos métodos não-probabilísticos, os elementos da população não possuem a mesma probabilidade de compor a amostra.

8. Resposta individual

9. Para a escolha do melhor método de amostragem, o pesquisador deve levar em conta: o tipo de pesquisa; a acessibilidade e disponibilidade dos elementos da população; a representatividade necessária para atingir os objetivos da pesquisa; a disponibilidade de tempo para a realização da pesquisa; os recursos financeiros e humanos disponíveis, entre outros fatores.

10. Exemplo 1 – Método não-probabilístico:

Objetivo da pesquisa: verificar a opinião dos ciclistas de uma determinada cidade quanto às condições de uso das vias públicas e o que fazer para melhorar a circulação nas mesmas.

População: ciclistas que utilizam as vias públicas.

Método de amostragem: tendo em vista a inexistência de um cadastro de ciclistas, o que impossibilita determinar o número de ciclistas na cidade, sugerimos uma amostragem não-probabilística.

Tipo de amostragem: neste caso, adotaremos uma amostragem por conveniência, em que os entrevistadores serão distribuídos em 10 pontos estratégicos da cidade para entrevistar os ciclistas que concordarem em participar da pesquisa.

Exemplo 2 – Método probabilístico:

Objetivo da pesquisa: verificar a opinião dos estudantes de graduação de uma universidade com relação ao processo de avaliação da aprendizagem.

População: estudantes regularmente matriculados em cursos de graduação da instituição.

Método de amostragem: tendo em vista que a universidade possui um cadastro de todos os estudantes organizados ordenadamente por número de matrícula e curso, sugerimos a realização de uma amostragem probabilística.

Tipo de amostragem: para manter as mesmas características da população na amostra, estas deverão ser proporcionais ao número de alunos por curso, devendo ser realizada uma amostragem estratificada.

2 Apresentação e organização de dados

A apresentação e a organização dos dados de um estudo estatístico podem ser feitas por meio de tabelas e gráficos.

Uma tabela é um meio bastante eficiente de apresentar um conjunto de dados, facilitando sua compreensão e interpretação. Uma tabela deve apresentar: título, cabeçalho, corpo e fonte. A tabela ainda pode apresentar elementos complementares, como notas e chamadas, colocadas de preferência no rodapé.

Tabela 2.1 – Exemplo de tabela

Frota de Veículos do Município de Rio Grande/RS 2007	
Tipo de Veículo	Quantidade
Automóvel	34.571
Caminhão	1.696
Caminhão trator	508
Caminhonete	1.950
Microônibus	106
Motocicleta	16.865
Motoneta	3.870
Ônibus	421
Trator de rodas	66

Fonte: IBGE

TÍTULO (aponta para o título da tabela)

CABEÇALHO (aponta para o cabeçalho da tabela)

CORPO (aponta para o corpo da tabela)

COLONA (aponta para a coluna da tabela)

CASA OU CÉLULA (aponta para a célula da tabela)

O **título**, localizado no topo da tabela, é formado por um conjunto de informações que identificam o tipo de variável que está sendo tratada, o local e o período em que os dados foram coletados. Basicamente, o título deve responder três questões: “o que”, “onde”, “quando”.

O **cabeçalho** especifica o conteúdo das colunas. Em uma coluna coloca-se uma variável do tipo: tempo, local ou espécie, e na outra a quantidade ou frequência observada.

O **corpo** da tabela é formado por um conjunto de linhas e colunas com as informações sobre a variável em estudo. O espaço destinado a um único dado ou número é denominado casa ou célula.

A **fonte** especifica quem coletou originalmente os dados.

A forma dos gráficos e o tipo de tabela dependem da natureza dos dados coletados. A seguir, serão apresentadas as tabelas ou séries para dados qualitativos com seus respectivos gráficos, e na sequência, as distribuições de frequência, que são séries específicas para dados quantitativos e sua representação gráfica.

2.1 Tabelas para dados qualitativos

Um conjunto de dados estatísticos dispostos em tabelas em função do tempo, do local e da espécie é denominado **série estatística**. As séries estatísticas podem ser classificadas em temporal, geográfica e específica, dependendo do tipo de variável que apresentam. As séries mistas ou tabelas de dupla entrada são séries que apresentam mais de uma variável.

As **séries específicas** ou **categóricas** são constituídas por dados obtidos nas diferentes espécies ou categorias de uma mesma variável.

Tabela 2.2 – Série estatística específica ou categórica

Número de cursos participantes por área – ENADE 2006	
Área	Quantidade
Administração	1.475
Ciências Contábeis	772
Ciências Econômicas	216
Direito	811
Total	3.274

Fonte: MEC/INEP/DEAES – ENADE 2006

As **séries temporais** ou **cronológicas** apresentam os dados de uma variável tomados em intervalos obtidos ao longo do tempo.

Tabela 2.3 – Série estatística temporal ou cronológica

Exportação de frutas, cascas de cítricos e melões	
Brasil – 1997 a 2002	
Anos	Valor (US\$ 1.000)
1997	301.005
1998	293.030
1999	323.973
2000	369.181
2001	346.456
2002	369.177

Fonte: ABIA – CESEX¹

A **série geográfica** apresenta dados provenientes de diferentes regiões geográficas ou locais.

Tabela 2.4 – Série geográfica

Ocupação territorial das regiões no Brasil – 2008	
Estados	Porcentagem (%)
Norte	45
Nordeste	18
Centro-Oeste	19
Sudeste	11
Sul	7

Fonte: IBGE

As **séries mistas** ou de dupla entrada combinam dois tipos de séries onde os dados podem referir-se a locais, espécies ou tempos diferentes. Por apontar mais de uma característica dos dados ao mesmo tempo, a série mista exige sempre mais de uma coluna.

¹ABIA. Associação Brasileira das Indústrias da Alimentação. <<http://www.abia.org.br>>.

Tabela 2.5 – Série mista

Número de cursos participantes por categoria administrativa segundo as áreas avaliadas – ENADE 2006

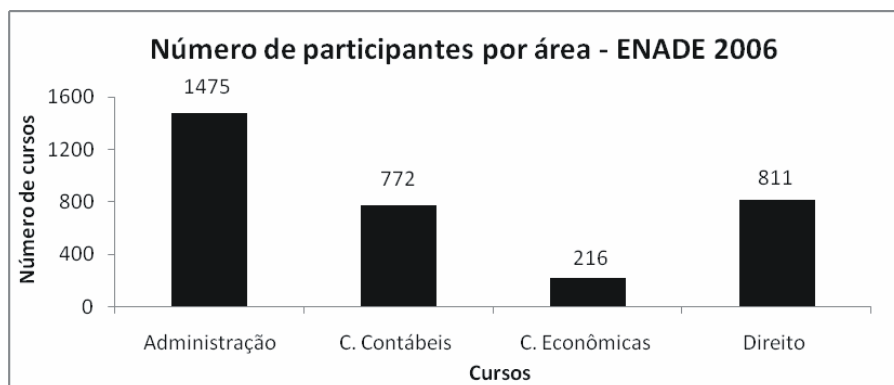
Área	Categoria administrativa		
	Pública	Privada	Total
Administração	163	1.312	1.475
Ciências Contábeis	122	650	772
Ciências Econômicas	75	141	216
Direito	94	717	811
Total	454	2.820	3.274

Fonte: MEC/INEP/DEAES – ENADE 2006

2.2 Gráficos estatísticos para dados qualitativos

Os dados estatísticos também podem ser apresentados por **gráficos**, cujo objetivo é auxiliar na visualização e compreensão dos dados, visto que proporcionam rápida interpretação das principais características do fenômeno em estudo. Para tal, são utilizados diversos tipos de gráficos, como gráfico de linhas, setores, colunas ou barras. A escolha do tipo de gráfico depende dos dados a serem apresentados.

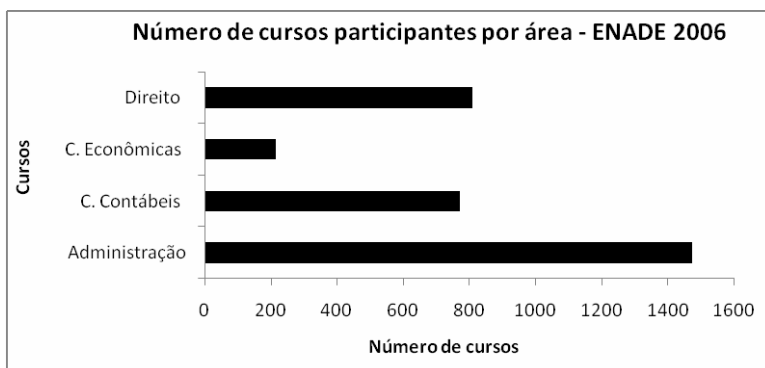
Gráfico de colunas – os retângulos têm a mesma base e as alturas representam o número ou a frequência observada. Os dados da tabela 2.2 estão representados na figura 2.1.



Fonte: MEC/INEP/DEAES – ENADE 2006

Figura 2.1 – Gráfico de colunas

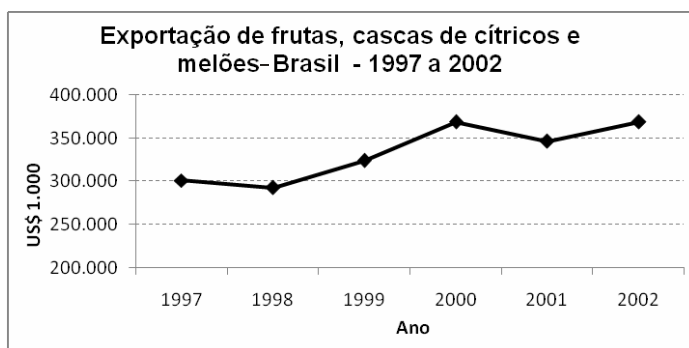
Gráfico de barras – Os retângulos têm a mesma altura e as bases representam o número ou a frequência observada. O gráfico de barras é mais indicado quando as legendas são extensas; o de colunas, quando as legendas são curtas. Os dados da tabela 2.2 são apresentados na Figura 2.2, num gráfico de barras.



Fonte: MEC/INEP/DEAES – ENADE 2006

Figura 2.2 – Gráfico de barras

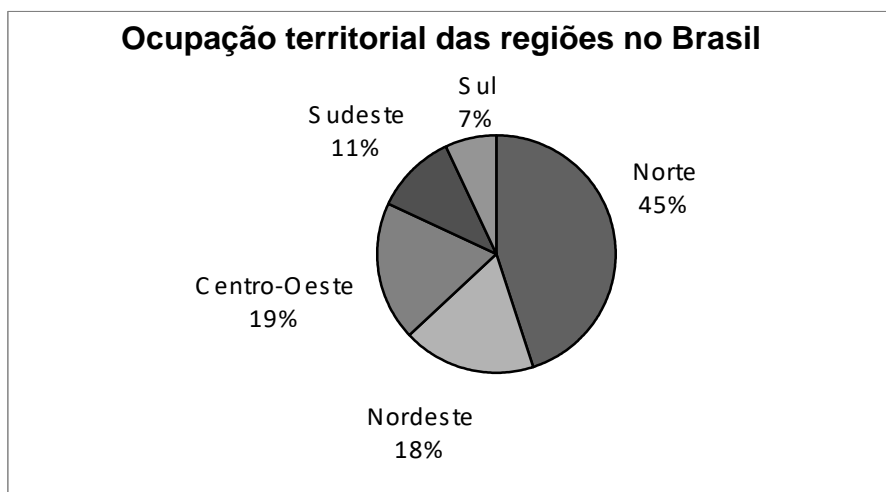
Gráfico de linhas – Na representação gráfica de uma série temporal, o gráfico de linhas geralmente é o mais indicado, principalmente quando o objetivo é mostrar as observações em função do tempo em que foram medidas. As linhas são particularmente mais eficientes do que as colunas quando existem intensas flutuações nos dados. Os dados da tabela 2.3 estão representados no gráfico de linhas da figura 2.3.



Fonte: ABIA – CESEX

Figura 2.3 – Gráfico de linhas para a série temporal

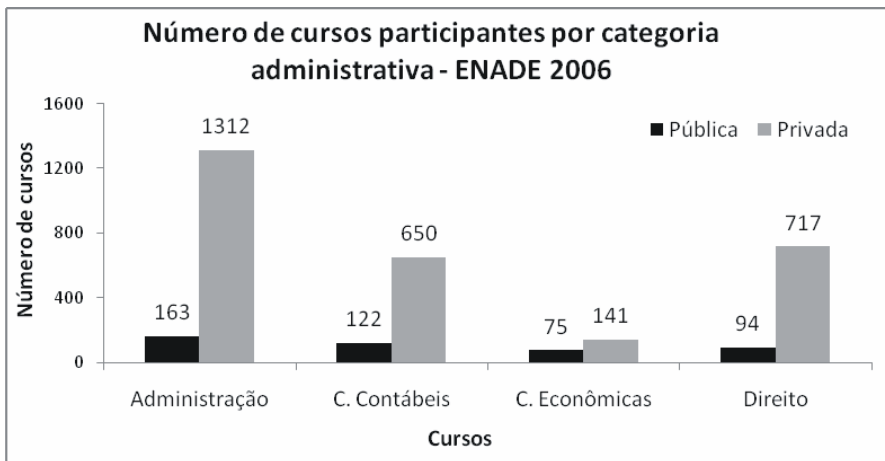
Gráfico de setores – Ajuda a visualizar a porção que cada grupo representa no todo. O total é representado pelo círculo que fica dividido em tantos setores quantas são as partes. Obtemos cada setor por meio de uma regra de três simples, lembrando que o total da série corresponde a 360° e/ou 100%. Por exemplo, nos dados da tabela 2.4 a fatia relativa à Região Sul corresponde a 7% do círculo, (figura 2.4)



Fonte: IBGE

Figura 2.4 – Gráfico de setores

Gráfico de colunas múltiplas – Na representação gráfica de uma série mista, podemos utilizar um gráfico de colunas múltiplas ou linhas múltiplas, dependendo do tipo de informação a ser transmitida. Os dados da tabela 2.5 estão representados na figura 2.5.



Fonte: MEC/INEP/DEAES – ENADE 2006

Figura 2.5 – Gráfico de colunas múltiplas

Gráficos enganosos – Os gráficos constituem excelente forma de apresentar dados, mas também podem ser enganosos. Na Figura 2.6 são apresentados dois gráficos que representam a evolução nas vendas de uma loja durante os meses de janeiro a junho. O gráfico da Figura 2.6(a) aparentemente apresenta maior evolução do que o gráfico da Figura 2.6(b), no entanto os dois representam o mesmo conjunto de dados – a diferença está na escala do eixo vertical. No primeiro gráfico, os valores do eixo vertical iniciam em 20, e no segundo gráfico, iniciam em 10, o que modifica a linha do gráfico. Portanto, deve-se prestar atenção nos gráficos, porque a configuração pode induzir a leitura equivocada.

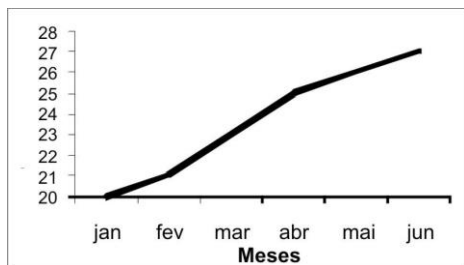


Figura 2.6 (a)

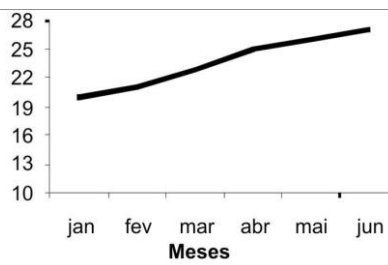


Figura 2.6 (b)

2.3 Tabelas para dados quantitativos

Quando descrevemos os dados estatísticos resultantes de variáveis numéricas ou quantitativas – por exemplo, altura, peso, salários –, utilizamos as **distribuições de frequência**.

Uma distribuição de frequência é uma série estatística na qual os dados são organizados em grupos de classes ou categorias convenientemente estabelecidas. Na apresentação dos dados numéricos, as distribuições de frequência podem ser divididas em dois tipos: distribuição de frequência sem intervalos de classe e distribuição de frequência com intervalos de classe.

Quando os dados quantitativos estão listados na ordem em que foram coletados, sem nenhum tipo de ordenação, são denominados **dados brutos**. A seguir, são apresentados os dados brutos das idades de 36 alunos da disciplina Probabilidade e Estatística, FURG–1998.

Dados brutos:

18	19	19	19	20	20	21	19	18
21	22	20	18	19	19	18	21	23
19	20	20	21	22	20	19	19	18
23	18	24	19	20	19	22	20	19

Observando os dados acima, verifica-se que é difícil determinar se os valores estão concentrados ou dispersos, e qual é o menor ou o maior valor. Mas, se os dados forem organizados em ordem crescente ou decrescente, em **rol**, fica mais fácil ter uma idéia do comportamento dos valores.

Dados em rol:

18	18	18	18	18	18	19	19	19
19	19	19	19	19	19	19	19	19
20	20	20	20	20	20	20	20	21
21	21	21	22	22	22	23	23	24

Com os dados ordenados (rol) podemos verificar facilmente que o aluno mais jovem da turma tem 18 anos e o mais velho tem 24; que a amplitude de variação foi de $24 - 18 = 6$ anos; que há uma concentração maior de alunos entre 18 e 20 anos.

Se um conjunto de dados consiste em relativamente poucos valores, muitos dos quais estão repetidos, como no exemplo das idades, simplesmente contamos quantas vezes cada um desses valores ocorre e então apresentamos o resultado em uma tabela de frequência.

Tabela de frequência para as idades

Idade	Frequência
18	6
19	12
20	8
21	4
22	3
23	2
24	1
Total	36

Acrescentando título e fonte à tabela, teremos uma distribuição de frequência sem intervalos de classe, não havendo perda de informação.

Tabela 2.6 – Distribuição de frequência sem intervalos de classe

Idade dos alunos da disciplina Probabilidade e Estatística FURG-2005		
Classes (i)	Idade (x_i)	Frequência
1	18	6
2	19	12
3	20	8
4	21	4
5	22	3
6	23	2
7	24	1
-	Total	36

Fonte: Sala de aula

Em uma distribuição de frequência, as classes são intervalos de variação da variável (x_i). Na tabela 2.6 há sete classes de variação. A idade de 20 anos está na terceira classe da distribuição de frequência. As classes são representadas pela letra i , sendo $i = 1, 2, 3, \dots, k$ (k é o número total de classes da distribuição).

Quando um conjunto de dados possui muitos valores com poucas repetições, a construção de uma distribuição de frequência sem intervalos de classes pode ficar muito extensa, dificultando a visualização do comportamento do grupo como um todo. Nesse caso, é mais conveniente agrupar os valores em intervalos de classe, o que resultará em uma distribuição de frequência **com** intervalos de classe.

Exemplo:

Dados brutos

Notas de 40 alunos									
5,5	10,0	9,9	7,8	8,0	8,9	2,0	3,0	8,9	3,8
4,5	7,3	3,8	8,8	7,5	3,4	1,0	4,0	7,5	5,8
8,0	6,9	5,0	3,0	6,5	9,9	1,1	7,4	5,5	7,5
6,4	10,0	6,0	9,9	10,0	8,9	8,0	10,0	8,6	9,5

Rol

Notas de 40 alunos									
1,0	3,0	4,0	5,5	6,5	7,5	8,0	8,8	9,5	10,0
1,1	3,4	4,5	5,8	6,9	7,5	8,0	8,9	9,9	10,0
2,0	3,8	5,0	6,0	7,3	7,5	8,0	8,9	9,9	10,0
3,0	3,8	5,5	6,4	7,4	7,8	8,6	8,9	9,9	10,0

Organizando os dados em uma tabela, temos:

Nota	Frequência	Nota	Frequência	Nota	Frequência
1,0	1	5,8	1	8,6	1
1,1	1	6,0	1	8,8	1
2,0	1	6,4	1	8,9	1
3,0	2	6,5	1	9,0	1
3,4	1	6,9	1	9,5	2
3,8	2	7,3	1	9,9	3
4,0	1	7,4	1	10,0	4
4,5	1	7,5	3		
5,0	1	7,8	1	Total	40
5,5	2	8,0	3		

Se observarmos a tabela acima, podemos verificar que esta ficou extensa devido a grande variação dos dados, dificultando sua análise. Neste caso, torna-se válido o agrupamento dos dados em intervalos de classe. Costuma-se usar o menor valor da série como limite inferior da primeira classe da distribuição de frequência – por exemplo, 1,0 (menor valor). Supondo que a amplitude do intervalo de classe seja 1,5, o limite superior da primeira classe será 2,5 ($1,0 + 1,5 = 2,5$). Dessa forma, em vez de dizer que um aluno tirou 1,0, um aluno tirou 1,1 e um aluno tirou 2,0; podemos dizer que três alunos tiraram notas entre 1,0 e 2,5. A tabela fica organizada da seguinte forma:

Tabela 2.7 – Montagem da distribuição de frequência com intervalos de classe

Notas de 40 alunos		
Notas	Tabulação	Frequência
1,0 † 2,5	┌	3
2,5 † 4,0	┌ └	5
4,0 † 5,5	┌	3
5,5 † 7,0	┌ └ ┌	7
7,0 † 8,5	┌ ┌	9
8,5 † 10,0	┌ └ ┌	13
Total		40

Inserindo título e fonte na tabela 2.7, teremos uma distribuição de frequência como na tabela 2.8.

Tabela 2.8 – Distribuição de frequência com intervalos de classe

Notas de 40 alunos	
Notas	Frequência
1,0 † 2,5	3
2,5 † 4,0	5
4,0 † 5,5	3
5,5 † 7,0	7
7,0 † 8,5	9
8,5 † 10,0	13
Total	40

Fonte: Caderno de chamada

Cada intervalo de classe possui um limite inferior da classe (l_i) e um limite superior da classe (L_i). Na tabela 2.8, na terceira classe, o limite inferior é $l_3 = 4,0$ e o limite superior é $L_3 = 5,5$.

Montagem da distribuição de frequência com intervalos de classe

Para montar a distribuição de frequência com intervalos de classe das notas dos alunos (tabela 2.8), foram realizados os seguintes cálculos:

▪ **Amplitude total (AT):** é a diferença entre o maior e o menor valor do conjunto de dados.

$$AT = V(\text{máx.}) - V(\text{mín.}) = 10,0 - 1,0 = 9 \rightarrow \mathbf{AT = 9}$$

▪ **Número de intervalos de classe (k):** para ter uma idéia do número de intervalos de classes mais adequado pode-se usar a regra de Sturges, que nos dá o número de classes em função do número de dados coletados: se os dados coletados forem relativos a uma amostra, o número de elementos na regra de Sturges será representado pela letra “n”:

$$k = 1 + 3,3 \log n$$

Se os dados forem relativos a uma população, o número de elementos é representado por “N”.

$$k = 1 + 3,3 \log N$$

No exemplo acima, temos:

$$k = 1 + 3,3 \cdot \log 40 = 1 + 3,3 \cdot 1,60206 = 6,2868 \rightarrow \mathbf{k = 6,29}$$

Na montagem de uma distribuição de frequência, procura-se usar, geralmente, de 5 a 15 classes. Usando menos de 5 classes há muita perda de informação; mais do que 15, a tabela fica muito extensa e dificulta a interpretação.

▪ **Amplitude do intervalo de classe (h):** calcula-se a amplitude de cada classe dividindo a amplitude total (AT) pelo número de classes (k). A amplitude de classe é definida como o intervalo entre o limite inferior e o limite superior de cada classe. O valor de h pode ser arredondado, para facilitar a análise e interpretação dos dados.

$$h = \frac{AT}{k} = \frac{V_{\max} - V_{\min}}{k} \quad \rightarrow \quad h = \frac{AT}{k} = \frac{9}{6,29} = 1,43$$

No exemplo, como são notas, para facilitar a interpretação da tabela a amplitude de cada classe foi arredondada para **h = 1,5**.

Por último, conta-se a frequência de ocorrência de cada valor que pertence a cada intervalo de classe da distribuição de frequência.

No exemplo, a distribuição de frequência terá aproximadamente seis intervalos de classe ($k \approx 6$) com a amplitude de cada classe igual a $h = 1,5$, ou seja, se começarmos com a menor nota, a primeira classe conterá todas as notas maiores ou iguais a 1,0 e menores do que 2,5 ($1,0 + h = 2,5$). O valor 2,5 é o limite superior da primeira classe. O limite superior não será contado nesta classe e sim na próxima, por isso o sinal entre os números na distribuição é fechado à esquerda e aberto à direita. O intervalo de $4,0 \text{ } \vdash \text{ } 5,5$ tem três alunos que tiraram: 4,0; 4,5 e 5,0. A nota 5,5 só é contada na próxima classe.

Numa distribuição de frequência com intervalos de classe, ganha-se em simplicidade, mas perde-se informação. Por exemplo, na tabela 2.8, pode-se verificar facilmente que cinco alunos tiraram entre 2,5 e 4,0, excluindo o 4,0, sem no entanto saber exatamente a nota de cada um dos cinco alunos.

Geralmente inicia-se a primeira classe pelo menor valor do conjunto de dados, somando-se o valor da amplitude de classe para encontrar o limite superior, mas pode-se iniciar com outro valor mais conveniente.

2.3.1 Outra forma de montar uma distribuição de frequência com intervalos de classe

Algumas distribuições de frequência são montadas utilizando classes do tipo “menos de” ou “mais de”, denominadas classes abertas, usadas, geralmente, para reduzir o número de classes de uma distribuição quando o conjunto de dados possui alguns valores muito menores ou muito maiores do que os demais. A desvantagem no uso de distribuições com classes abertas é a impossibilidade de calcular medidas estatísticas como, por exemplo, média e desvio padrão.

Tabela 2.9 – Distribuição de frequência com classes abertas

Distribuição de frequência das idades dos alunos	
Idades	Frequências
Menos de 20	8
20 – 25	13
25 – 30	10
30 – 35	5
35 – 40	2
Mais de 40	2
Total	40

Fonte: Fictícia

Além dos exemplos apresentados aqui, existem outras formas de se montar uma distribuição de frequência. O importante é escolher a melhor maneira de descrever o conjunto de dados facilitando a compreensão e análise da variável sob estudo sem muita perda de informação.

2.3.2 Tipos de frequência

▪ **Frequência simples ou absoluta (f_i):** são os valores que representam o número de elementos de cada classe, como no exemplo da tabela 2.10.

Exemplo: A frequência simples da segunda classe é igual a cinco: $f_2 = 5$.

▪ **Frequência relativa (fr_i):** é a razão entre a frequência simples (f_i) de cada classe e a frequência total ou soma das frequências simples ($\sum f_i$):

$$fr_i = \frac{f_i}{\sum f_i}$$

Exemplo: A frequência relativa da segunda classe da tabela 2.10 é:

$$fr_2 = \frac{f_2}{\sum f_i} = fr_2 = \frac{5}{40} = 0,125$$

Obs. 1: A soma das frequências relativas é sempre igual a um (1): $\sum fr_i = 1$

Obs. 2: O uso da frequência relativa facilita as comparações entre mais de um conjunto de dados com diferentes números de elementos, o que seria mais difícil com a frequência simples.

▪ **Frequência percentual (fp_i):** para obter a frequência percentual, basta multiplicar a frequência relativa por 100:

$$fp = fr \times 100$$

Exemplo: A frequência percentual da terceira classe é:

$$fp_2 = fr_2 \times 100 = 0,125 \times 100 = 12,5\%$$

Tabela 2.10 – Frequência simples (f_i), frequência relativa (fr_i) e frequência percentual (fp_i)

Classe (i)	NOTAS	f_i	fr_i	fp_i
1	1,0 – 2,5	3	0,075	7,5%
2	2,5 – 4,0	5	0,125	12,5%
3	4,0 – 5,5	3	0,075	7,5%
4	5,5 – 7,0	7	0,175	17,5%
5	7,0 – 8,5	9	0,225	22,5%
6	8,5 – 10,0	13	0,225	32,5%
	Total	40	1	100%

■ **Frequência acumulada (F)**: é a soma da frequência simples da classe com as frequências simples das classes anteriores. A frequência acumulada da última classe é igual ao número total de observações.

Exemplo: Na tabela 2.11, a frequência acumulada correspondente à terceira classe é:

$$F_3 = f_1 + f_2 + f_3 \quad \Rightarrow \quad F_3 = 3 + 5 + 3 = 11$$

Portanto, existem 11 alunos com nota inferior a 5,5 (limite superior do intervalo da terceira classe).

Da mesma forma que na frequência simples, também podemos calcular a frequência relativa ou percentual da frequência acumulada.

■ **Frequência acumulada relativa (Fr_i)**: é a frequência acumulada da classe, dividida pela frequência total do conjunto de dados:

$$Fr_i = \frac{F_i}{\sum f_i}$$

Exemplo: A frequência acumulada relativa da terceira classe é:

$$Fr_3 = \frac{F_3}{\sum f_i} \Rightarrow Fr_3 = \frac{11}{40} = 0,275 \Rightarrow Fr_3 = 0,275$$

Tabela 2.11 – Frequência acumulada (F_i), Frequência acumulada relativa (Fr_i) e Frequência acumulada percentual (Fp_i)

Classe (i)	NOTAS	f_i	F_i	Fr_i	Fp_i
1	1,0 – 2,5	3	3	0,075	7,5%
2	2,5 – 4,0	5	8	0,20	20,0%
3	4,0 – 5,5	3	11	0,275	27,5%
4	5,5 – 7,0	7	18	0,45	45,0%
5	7,0 – 8,5	9	27	0,675	67,5%
6	8,5 – 10,0	13	40	1,0	100%
	Total	40	-	-	-

Os vários tipos de frequência auxiliam a responder a questões como:

1) Quantos alunos obtiveram nota entre 7,0 (inclusive) até 8,5?

Esses são os valores da variável nota que formam a quinta classe da distribuição de frequência. Como $f_5 = 9$, a resposta é: 9 alunos.

2) Quantos alunos têm nota abaixo de 7,0?

Notas inferiores a sete estão nas quatro primeiras classe ($i = 1, 2, 3$ e 4). Assim, o número de alunos é dado por: $f_1 + f_2 + f_3 + f_4 = F_4 = 3 + 5 + 3 + 7 = 18$. Logo, 18 alunos têm nota abaixo de 7,0.

3) Qual a percentagem de alunos cujas notas são inferiores a 7,0?

Notas inferiores a sete estão nas quatro primeiras classes. $Fp_4 = 45\%$, portanto 45% dos alunos obtiveram notas inferiores a sete.

2.4 Gráficos para dados quantitativos

As distribuições de frequência podem ser representadas graficamente pelo histograma e pelo polígono de frequência. Quando a distribuição de frequência é sem intervalos de classe, o **histograma** é um conjunto de hastes ou colunas, representadas em um sistema de coordenadas cartesianas que tem por base os valores da distribuição e por altura a frequência simples, relativa ou

percentual correspondente. Os dados da tabela 2.6 são representados graficamente por um histograma na figura 2.7(a).

Quando a distribuição de frequência é com intervalos de classe, o **histograma** é formado por um conjunto de retângulos justapostos, cujas bases representam os intervalos de classe e cujas alturas representam as frequências simples, relativas ou percentuais de cada intervalo de classe. Os dados da tabela 2.8 são representados graficamente por um histograma na figura 2.7(b).

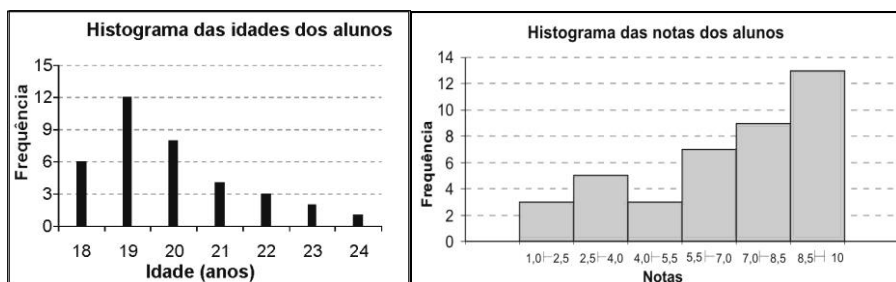


Figura 2.7 – (a) Histograma da distribuição sem intervalos de classe
(b) Histograma da distribuição com intervalos de classe

Ao analisar a representação gráfica de um conjunto de dados numéricos, é possível observar-se seus aspectos principais, como forma, valores máximos e mínimos, picos mais destacados, valores extremos.

Na Figura 2.7(a), observa-se que os alunos mais jovens têm 18 anos e o mais velho, 24; há maior concentração de alunos com idades de 18 a 21 anos; à medida que as idades aumentam, o número de alunos diminui. Na Figura 2.7(b) observa-se que a menor nota foi igual a 1,0 e a maior nota foi 10,0; à medida que aumentam as notas, aumenta também o número de alunos em cada intervalo de classe. A maior frequência foi observada no intervalo de 8,5 a 10,0 inclusive.

O **polígono de frequência** é um gráfico em linha, sendo as frequências marcadas sobre os pontos médios dos intervalos de classe e unidas por segmentos de retas. Uma vez que o gráfico deve começar e terminar sobre o eixo horizontal, crie uma classe

fictícia antes da primeira classe da distribuição e marque seu ponto médio. Faça o mesmo no final da distribuição. Estenda a linha do gráfico até estes dois últimos pontos médios calculados.

Observação: O ponto médio (PM) de uma classe é a média aritmética entre os limites inferior e superior da classe.

$$\text{Ponto médio} = \frac{\text{limite inferior da classe} + \text{limite superior da classe}}{2} = \frac{l_i + L_i}{2}$$

Exemplo: Monte o histograma e o polígono de frequência dos dados da tabela 2.12.

Tabela 2.12 – Estaturas dos alunos

Estaturas dos Alunos	
Estaturas (cm)	f _i
154 – 158	8
158 – 162	10
162 – 166	12
166 – 170	9
170 – 174	7
174 – 178	4
Total	50

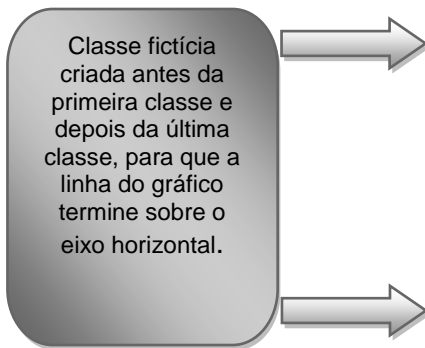
Fonte: Sala de aula

Solução:

Tabela 2.13 – Cálculo do ponto médio

Estaturas dos alunos		
Estaturas (cm)	PM	f _i
150 – 154	152	0
154 – 158	156	8
158 – 162	160	10
162 – 166	164	12
166 – 170	168	9
170 – 174	172	7
174 – 178	176	4
178 – 182	180	0
Total		50

Fonte: Sala de aula



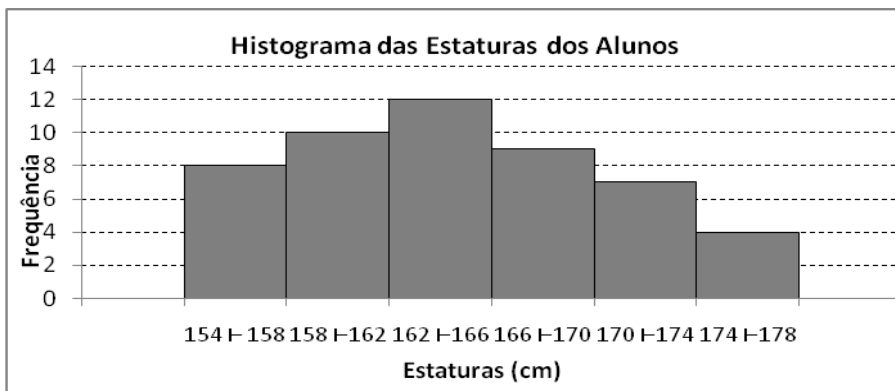


Figura 2.8 – Histograma das estaturas dos alunos da tabela 2.12.

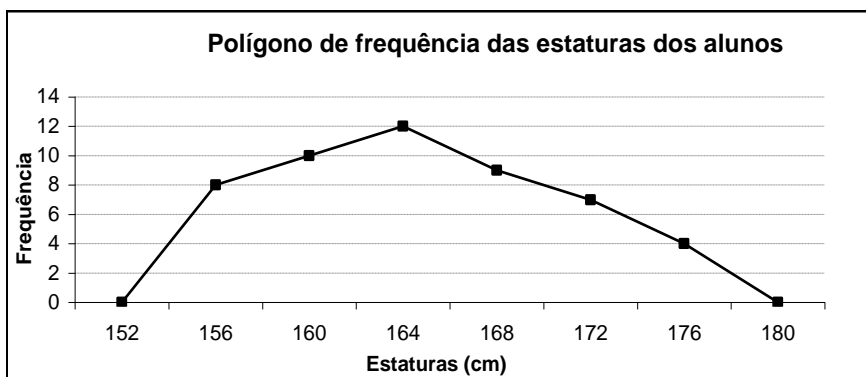


Figura 2.9 – Polígono de frequência das estaturas dos alunos da tabela 2.12.

OBS: Para melhor compreensão da montagem do polígono de frequência, a Figura 2.10 apresenta os dois gráficos juntos.

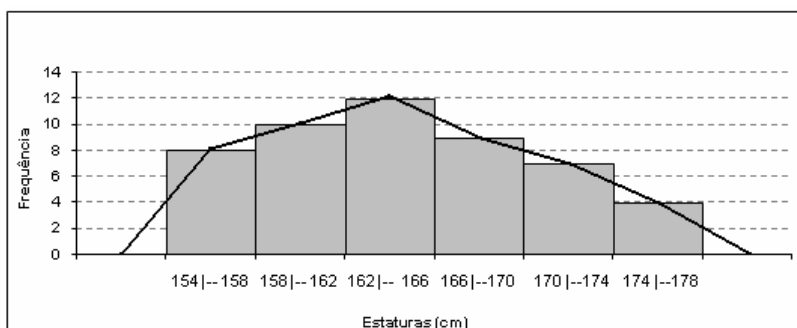


Figura 2.10 – Histograma e polígono de frequência das estaturas dos alunos da tabela 2.13.

Quadro 2 – Resumo dos conceitos básicos

Estatística descritiva: coleta, organiza, apresenta e resume os dados através de medidas estatísticas.

Probabilidade: analisa situações que envolvem o acaso.

Estatística inferencial: analisa e interpreta dados amostrais.

População: conjunto de indivíduos ou objetos que apresentam pelo menos uma característica em comum.

Amostra: parte da população.

Parâmetro populacional: medida estatística obtida com base em uma população.

Estatística ou estimador: medida estatística obtida com base em uma amostra.

Estimativa: valor numérico assumido pelos estimadores ou estatísticas amostrais.

Variável qualitativa: separada em diferentes categorias que se distinguem por alguma característica não-numérica. Pode ser do tipo nominal ou ordinal.

Variável quantitativa: consiste em números que representam contagem ou medidas. Pode ser do tipo discreta ou contínua.

Série estatística: conjunto de dados estatísticos dispostos em tabelas em função do tempo, do local e da espécie.

Distribuição de frequência: relaciona cada valor ou classes de valores com a frequência com que cada um ocorre, identificando o local, o tempo e a variável em estudo.

Dados brutos: dados quantitativos listados fora da ordem sem nenhum tipo de ordenação.

Rol: conjunto de dados ordenados em ordem crescente ou decrescente.

Frequência acumulada: soma da frequência simples da classe com as frequências simples das classes anteriores.

Histograma: representação gráfica de uma distribuição de frequência.

Polígono de frequência: gráfico em linha que representa uma distribuição de frequência.

Exercício resolvido

1) Uma empresa realiza pesquisa para verificar a qualidade do serviço de entregas a domicílio em suas 48 filiais. A tabela abaixo apresenta o número de reclamações recebidas quanto às entregas, por cada uma das filiais, durante o período de um mês.

11	13	10	9	0	16	28	15	14	1	2	17
18	12	18	8	9	22	16	20	34	15	21	13
12	14	13	19	0	2	17	11	18	16	13	6
19	8	8	12	13	21	11	19	1	14	4	16

a) Monte uma distribuição de frequência com intervalos de classe.

Comece a primeira classe pelo zero (menor valor), ou seja, limite inferior da 1ª classe = zero;

b) Calcule as frequências relativas (fr_i);

c) Calcule as frequências percentuais (fp_i);

d) Calcule as frequências acumuladas (F_i);

e) Calcule as frequências acumuladas relativas (Fr_i);

f) Calcule as frequências acumuladas percentuais (Fp_i);

g) Interprete a quarta classe da distribuição de frequência;

h) Calcule o ponto médio de cada classe;

i) Faça o histograma;

j) Faça o polígono de frequência.

Solução do item a: Para montar a distribuição de frequência, o primeiro passo é colocar os dados em ordem crescente (rol).

Rol

0	2	8	9	11	13	13	15	16	18	19	21
0	2	8	10	12	13	14	15	16	18	19	22
1	4	8	11	12	13	14	16	17	18	20	28
1	6	9	11	12	13	14	16	17	19	21	34

A seguir, calcule o número de intervalos de classe da distribuição de frequência, usando a regra de Sturges:

$$k = 1 + 3,3 \log n \quad k = 1 + 3,3 \log 48 \quad k = 1 + 3,3 \times 1,6812 \\ k = 6,5481 \rightarrow \text{arredondando, teremos aproximadamente sete classes.}$$

Com o valor do número de intervalos de classe (k), pode-se calcular a amplitude do intervalo de classe (h).

$$V_{\min} = 0 \quad V_{\max} = 34$$

$$h = \frac{AT}{k} = \frac{V_{\max} - V_{\min}}{k} = \frac{34 - 0}{6,5481} = 5,19 \rightarrow \text{arredondando, teremos a amplitude do intervalo de classe: } h = 5.$$

Por fim, monte a distribuição de frequência com intervalos de classe. Iniciando pelo menor valor, zero, soma-se a amplitude do intervalo de classe (h=5), até incluir o maior valor dos dados (34). Podemos observar na tabela 2.14 que temos sete classes. Na segunda coluna estão os intervalos de classe (ex: 0 – 5 ; 5 – 10...). Na terceira coluna, a frequência simples, obtida pela contagem dos dados que estão dentro de cada classe. Lembre-se que os intervalos são fechados nos limites inferiores (ex.: limite inferior da primeira classe é o zero) e abertos nos limites superiores (ex.: limite superior da primeira classe é o cinco), o que significa que valores de zero até o limite de cinco (o cinco não entra porque o intervalo é aberto à direita) devem ser contados e suas frequências devem ser somadas.

Na primeira classe, temos dois números “0”, dois números “1”, dois números “2” e um número “4”, totalizando um total de sete números contidos nessa classe. Faça isso para todas as classes.

Tabela 2.14 – Montagem da distribuição de frequência

Classe (i)	Número de reclamações	f_i
1	0 – 5	7
2	5 – 10	6
3	10 – 15	15
4	15 – 20	14
5	20 – 25	4
6	25 – 30	1
7	30 – 35	1
	Total	48

Solução do item b: As frequências relativas (tabela 2.15) são obtidas dividindo-se a frequência simples pelo total de dados, neste caso, $\sum f_i = 48$.

Solução do item c: As frequências percentuais (fp_i) (tabela 2.15) são calculadas multiplicando-se as frequências relativas por 100.

Tabela 2.15 – Cálculo da frequência relativa e frequência percentual

Classe (i)	Número de reclamações	f_i	fr_i	fp_i
1	0 – 5	7	$7/48 = 0,1458$	$0,1458 \times 100 = 14,58 \%$
2	5 – 10	6	$6/48 = 0,125$	$0,125 \times 100 = 12,50 \%$
3	10 – 15	15	$15/48 = 0,3125$	$0,3125 \times 100 = 31,25 \%$
4	15 – 20	14	$14/48 = 0,2917$	$0,2917 \times 100 = 29,17 \%$
5	20 – 25	4	$4/48 = 0,0833$	$0,0833 \times 100 = 8,33 \%$
6	25 – 30	1	$1/48 = 0,0208$	$0,0208 \times 100 = 2,08 \%$
7	30 – 35	1	$1/48 = 0,0208$	$0,0208 \times 100 = 2,08 \%$
	Total	48	$\Sigma 1$	100%

Solução do item d: A frequência acumulada (F_i) (tabela 2.16) é obtida somando-se a frequência simples desta classe com a frequência simples das classes anteriores. Portanto, na primeira classe, devemos repetir o valor da frequência simples ($f_i = 7$); nas outras classes, devemos acumular os valores.

Tabela 2.16 – Cálculo da frequência acumulada

Classe (i)	Número de reclamações	f_i	F_i
1	0 – 5	7	7
2	5 – 10	6	$7+6 = 13$
3	10 – 15	15	$7+6+15 = 28$
4	15 – 20	14	$7+6+15 +14 = 42$
5	20 – 25	4	$7+6+15 +14 +4 = 46$
6	25 – 30	1	$7+6+15 +14 +4 +1 = 47$
7	30 – 35	1	$7+6+15 +14 +4 +1+1 = 48$
	Total	48	$\Sigma 1$

Solução do item e: A frequência acumulada relativa (Fr_i) (tabela 2.17) é obtida dividindo-se cada frequência acumulada pelo número total de dados: $\Sigma f_i = 48$.

Solução do item f: A frequência acumulada percentual (Fp_i), (tabela 2.17) é calculada multiplicando-se a frequência acumulada relativa por 100.

Tabela 2.17 – Cálculo da frequência acumulada relativa e percentual

Classe (i)	Número de reclamações	F_i	Fr_i	Fp_i
1	0 – 5	7	$7/48 = 0,1448$	$0,1448 \times 100 = 14,58\%$
2	5 – 10	13	$13/48 = 0,2708$	$0,2708 \times 100 = 27,08\%$
3	10 – 15	28	$28/48 = 0,5833$	$0,5833 \times 100 = 58,33\%$
4	15 – 20	42	$42/48 = 0,8750$	$0,8750 \times 100 = 87,50\%$
5	20 – 25	46	$46/48 = 0,9583$	$0,9583 \times 100 = 95,83\%$
6	25 – 30	47	$47/48 = 0,9792$	$0,9792 \times 100 = 97,92\%$
7	30 – 35	48	$48/48 = 1,0$	$1,0 \times 100 = 100\%$

Solução do item g: Para fazer a interpretação da quarta classe da distribuição de frequência, observe a tabela 2.18: quatorze filiais receberam de 15 a menos de 20 reclamações com relação ao seu serviço de entregas, no último mês, o que representa 29,17% das reclamações de toda a empresa. Quarenta e duas filiais receberam menos de 20 reclamações, o que corresponde a 87,50% das filiais.

Tabela 2.18 – Distribuição de frequência e os tipos de frequência

Classe (i)	Número de reclamações	f_i	fr_i	fp_i	F_i	Fr_i	Fp_i
1	0 – 5	7	0,1458	14,58%	7	0,1448	14,58%
2	5 – 10	6	0,1250	12,50%	13	0,2708	27,08%
3	10 – 15	15	0,3125	31,25%	28	0,5833	58,33%
4	15 – 20	14	0,2917	29,17%	42	0,8750	87,50%
5	20 – 25	4	0,0833	8,33%	46	0,9583	95,83%
6	25 – 30	1	0,0208	2,08%	47	0,9792	97,92%
7	30 – 35	1	0,0208	2,08%	48	1,0	100%
	Total	48	$\Sigma 1$	100%			

Solução do item h: O ponto médio (tabela 2.19) é calculado fazendo-se a média de cada classe, isto é, somando o limite inferior e superior de cada classe e dividindo o resultado por 2.

$$1^{\text{a}} \text{ classe} = \frac{0 + 5}{2} = 2,5$$

$$2^{\text{a}} \text{ classe} = \frac{5 + 10}{2} = 7,5$$

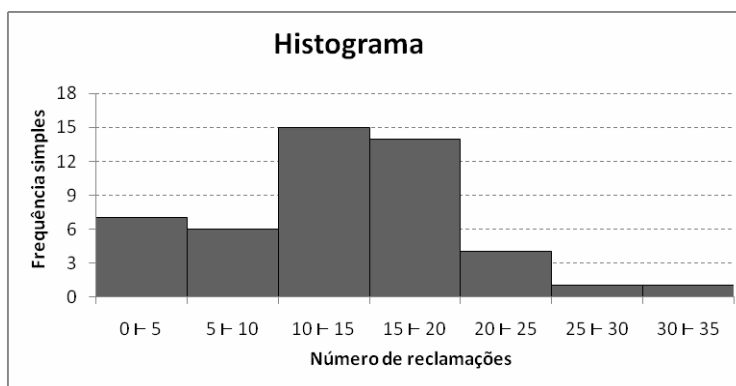
⋮

$$7^{\text{a}} \text{ classe} = \frac{30 + 35}{2} = 32,5$$

Tabela 2.19 – Cálculo do ponto médio

Classe (i)	Número de reclamações	f _i	Ponto médio
1	0 – 5	7	2,5
2	5 – 10	6	7,5
3	10 – 15	15	12,5
4	15 – 20	14	17,5
5	20 – 25	4	22,5
6	25 – 30	1	27,5
7	30 – 35	1	32,5
	Total	48	

Solução do item i:



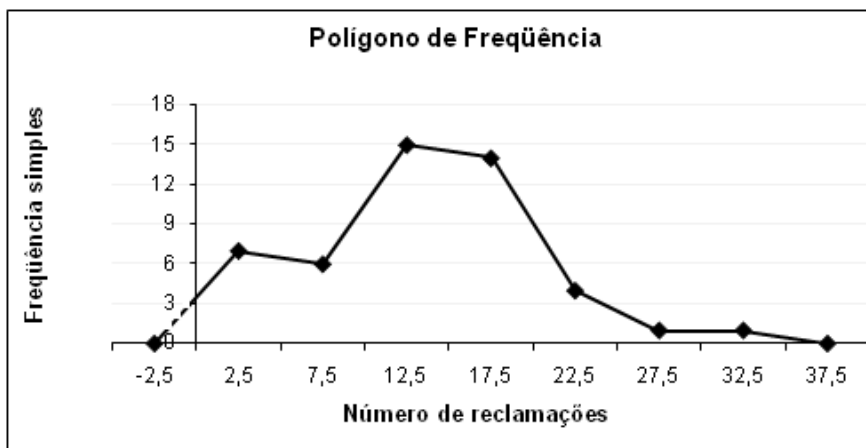
Solução do item j: Uma vez que o gráfico deve começar e terminar sobre o eixo horizontal, crie uma classe antes da primeira classe da distribuição e marque seu ponto médio. Faça o mesmo no final da distribuição. Estenda a linha do gráfico até esses dois últimos pontos médios calculados.



Classe (i)	Nº de reclamações	f_i	Ponto médio
Classe fictícia	-5 – 0	0	-2,5
1	0 – 5	7	2,5
2	5 – 10	6	7,5
3	10 – 15	15	12,5
4	15 – 20	14	17,5
5	20 – 25	4	22,5
6	25 – 30	1	27,5
7	30 – 35	1	32,5
Classe fictícia	35 – 40	0	37,5
	Total	48	



OBSERVE as linhas marcadas. Estas classes não existem, são fictícias, por esse motivo a frequência simples é igual a zero. Usaremos o ponto médio dessas classes, na montagem do polígono de frequência.



Exercícios complementares

1) No questionário do exercício 5, capítulo 1, aplicado aos estudantes que freqüentaram a livraria do campus universitário, foi perguntado como classificariam o serviço prestado pela livraria. Na tabela abaixo estão as respostas dadas a esta questão.

a) Monte uma série estatística com os dados coletados;

b) Represente graficamente os dados, escolhendo o gráfico que mais se adapta a esta variável. Justifique sua escolha.

Ótimo	Bom	Bom	Péssimo	Bom	Bom	Ótimo	Ótimo
Bom	Ótimo	Bom	Ótimo	Bom	Bom	Ótimo	Bom
Péssimo	Bom	Péssimo	Bom	Péssimo	Bom	Bom	Bom
Bom	Ótimo	Bom	Péssimo	Ótimo	Ótimo	Bom	Péssimo

2) As notas coletadas no caderno de chamada do primeiro bimestre de 36 alunos da disciplina Probabilidade e Estatística da FURG em 1998 estão na tabela abaixo:

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

a) Monte uma distribuição de frequência com intervalos de classe, iniciando a primeira classe pela nota 4,0.

b) Qual a classe com maior frequência?

c) Construa o histograma.

d) Observando a distribuição de frequência ou o histograma, escreva um breve relato sobre as notas dos alunos desta turma na disciplina Probabilidade e Estatística.

3) O setor de atendimento ao cliente do Supermercado Esplanada realizou uma pesquisa para verificar o tempo de espera que cada cliente permanece na fila do caixa. O tempo, em minutos,

registrado por cada cliente, está na tabela abaixo. Os dados foram coletados aleatoriamente, ao longo do mês de janeiro de 2008:

3	5	2	2	6	4	7	2	8	2
7	2	8	5	6	6	4	7	4	5
8	8	5	3	2	5	3	4	6	4
2	7	8	5	2	5	6	6	4	5
2	7	6	6	7	3	5	2	2	7

a) Monte uma distribuição de frequência e o histograma.

b) Calcule a frequência simples percentual, frequência acumulada, frequência acumulada percentual. Interprete os resultados da terceira classe, juntamente com as frequências calculadas.

4) A tabela a seguir mostra os salários de 200 empregados de uma empresa.

Salários dos empregados da empresa FGNick – Março de 2007	
Salário (R\$)	f_i
200,00 – 400,00	24
400,00 – 600,00	36
600,00 – 800,00	50
800,00 – 1000,00	65
1000,00 – 1200,00	15
1200,00 – 1400,00	10
TOTAL	200

Fonte: Recursos Humanos

a) faça o polígono de frequência desta distribuição;

b) calcule a frequência simples percentual;

c) calcule a frequência acumulada;

d) calcule a frequência acumulada percentual.

e) interprete a terceira classe da distribuição de frequência.

Respostas

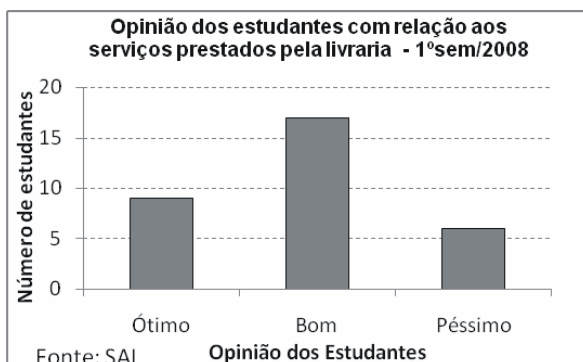
1. a)

Opinião dos estudantes da Universidade com relação aos serviços prestados pela Livraria – 1º sem/2008

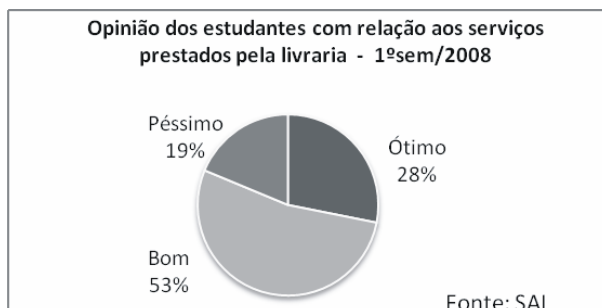
Opinião	f _i
Ótimo	9
Bom	17
Péssimo	6
Total	32

Fonte: Secretaria de Avaliação Institucional (SAI)

b) Como os dados são qualitativos e a série é específica, podemos utilizar um gráfico de barras, colunas ou setores.



ou



2. a)

$$k = 1 + 3,3 \log n$$

$$k = 1 + 3,3 \log 36$$

$$k = 6,1357$$

$$h = \frac{AT}{k} = \frac{10 - 4,5}{6,1357} = 0,896 \cong 1$$

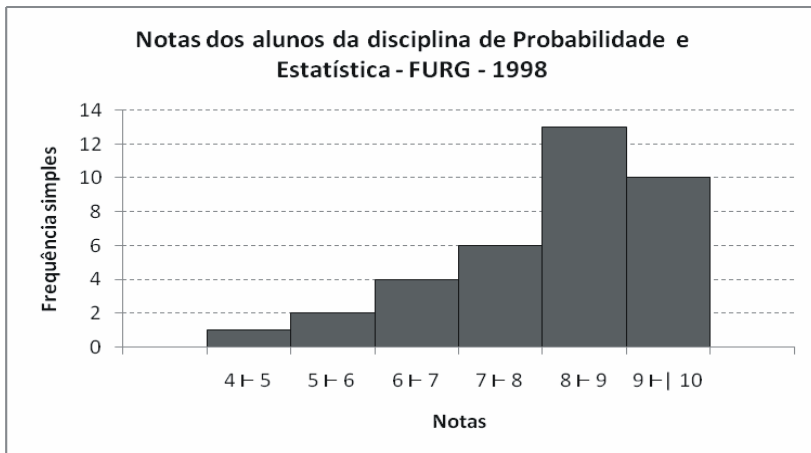
Notas dos alunos da disciplina Probabilidade e Estatística -
FURG – 1998

Classe(i)	Notas	f _i
1	4 – 5	1
2	5 – 6	2
3	6 – 7	4
4	7 – 8	6
5	8 – 9	13
6	9 – 10	10
–	–	Σ 36

Fonte: Caderno de chamada

b) A classe 5 é a que tem maior frequência (f₅ = 13).

c)



Fonte: Caderno de chamada

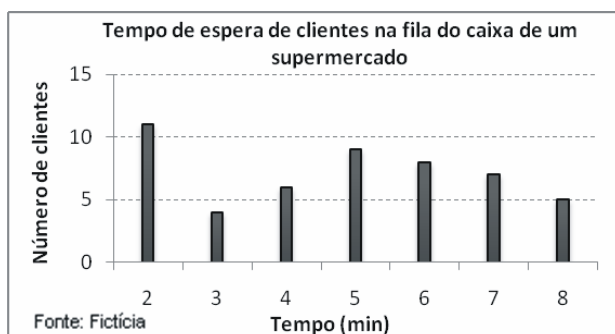
d) Resposta individual

3.a)

Tempo de espera de clientes na fila do caixa
de um supermercado

TEMPOS	f_i
2	11
3	4
4	6
5	9
6	8
7	7
8	5
Total	50

Fonte: Fictícia



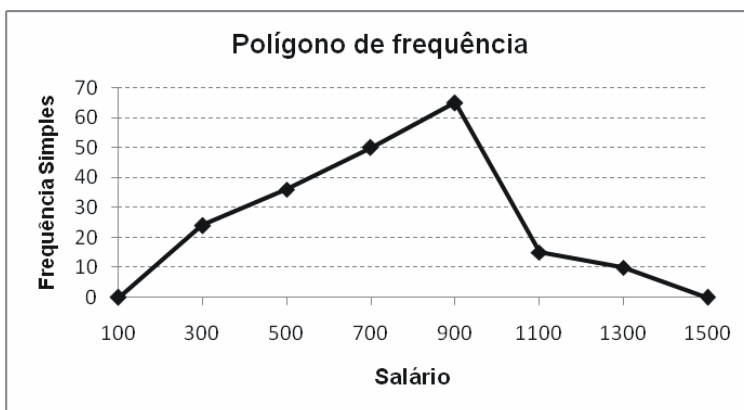
Fonte: Fictícia

b)

TEMPOS	f_i	fp_i	F_i	Fp_i
2	11	22%	11	22%
3	4	8%	15	30%
4	6	12%	21	42%
5	9	18%	30	60%
6	8	16%	38	76%
7	7	14%	45	90%
8	5	10%	50	100%
	50	1		

Interpretação da terceira classe: Seis clientes esperaram 4 minutos na fila do caixa, o que representa 12% do total de clientes observados. Vinte e um clientes aguardaram 4 minutos ou menos, o que representa 42% do total de clientes.

4. a)



Salário (R\$)	f_i	f_{p<i>i</i>}	F_i	F_{p<i>i</i>}	Ponto médio x_i
0 – 200,00					100
200,00 – 400,00	24	12%	24	12%	300
400,00 – 600,00	36	18%	60	30%	500
600,00 – 800,00	50	25%	110	55%	700
800,00 – 1000,00	65	32,5%	175	87,5%	900
1000,00 – 1200,00	15	7,5%	190	95%	1.100
1200,00 – 1400,00	10	5%	200	1	1.300
1400,00 – 1600,00					1500
TOTAL	200	1			

e) Cinquenta funcionários recebem salários entre R\$ 600,00 e R\$ 800,00 (exclusive). Cento e dez funcionários ganham menos que R\$ 800,00, o que representa 55% dos funcionários.

3 Medidas estatísticas para dados não-agrupados

Neste capítulo estudaremos as medidas estatísticas para dados não-agrupados, que são: as medidas de posição; as medidas de dispersão e a medida de assimetria. Nas medidas de posição ou tendência central, estudaremos a média, a mediana e a moda. Nas medidas de dispersão, a amplitude total, o desvio médio, a variância, o desvio-padrão e o coeficiente de variação de Pearson. Além das medidas de posição e dispersão, a forma da distribuição é uma importante fonte de informação sobre o comportamento dos dados; assim, veremos ainda neste capítulo o coeficiente de assimetria e curtose.

3.1 Medidas de posição ou tendência central

As medidas de posição ou medidas de tendência central indicam um valor que melhor representa todo o conjunto de dados, ou seja, indicam a tendência da concentração dos valores observados.

As principais medidas de posição são: a média, a mediana e a moda. A **média aritmética** é uma das principais medidas de posição, representando o ponto de equilíbrio de um conjunto de dados. A **mediana** é o valor central de um rol, ou seja, é o valor que fica no centro da série, quando os dados são arranjados na ordem crescente ou decrescente. A **moda** é definida como o valor ou valores que ocorrem com maior frequência. Evidentemente, um conjunto de dados pode não apresentar moda, portanto é denominado amodal.

3.1.1 Média

Média da amostra (\bar{x} , lê-se x-barra): a média de um conjunto de dados é a medida de tendência central encontrada pela soma de todos os valores, e esta soma é dividida pelo número total de valores. A média é considerada o ponto de equilíbrio no conjunto de dados. Se as observações em uma amostra de tamanho n são x_1, x_2, \dots, x_n , a média amostral é calculada pela seguinte expressão:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{que pode ser representada por:}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

onde x_i é o valor da observação i , n o número de observações e Σ (a letra sigma maiúscula do alfabeto grego) que, na fórmula, indica o símbolo de somatório.

Exemplo: Calcule a média dos seguintes dados: 2, 4, 6, 7, 11

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2+4+6+7+11}{5} = 6$$

A forma de calcular a média é a mesma tanto para uma amostra como para uma população finita, mas usa-se uma notação diferente, para indicar que estamos trabalhando com uma população. O número de elementos em uma população é denotado por N e a média da população por μ .

Média da população (μ , lê-se mi): a média de uma população finita é encontrada somando-se todos os valores da população e dividindo pelo tamanho N da mesma.

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

Exemplo: Os valores abaixo são os salários de seis empregados de uma pequena empresa. Calcule o salário médio dos empregados dessa empresa.

R\$ 860,00 – R\$ 750,00 – R\$ 980,00 – R\$ 1.200,00 – R\$ – R\$ 790,00 – R\$ 950,00

Solução: Como a empresa possui apenas seis empregados, podemos calcular a média dos salários considerando todos os empregados. Neste caso, estamos trabalhando com toda a população.

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{860 + 750 + 980 + 1200 + 790 + 950}{6} = \frac{5530}{6} = 921,67$$

O salário médio nessa empresa é de $\mu = \text{R\$ } 921,67$

A média nem sempre representa a tendência central dos dados, além disso ela pode ser influenciada por valores extremos, como na série 2 do exemplo a seguir. Essa é uma das diferenças marcantes entre a mediana e a média: a mediana depende da posição, e não dos valores dos elementos na série ordenada.

Exemplo:

Série 1: 5 – 7 – 10 – 13 – 15 → $\bar{x} = 10$ e $Me = 10$

Série 2: 5 – 7 – 10 – 13 – 65 → $\bar{x} = 20$ e $Me = 10$

Vejamos outro exemplo: na tabela 3.1, a turma A apresenta uma distribuição simétrica, ou seja, os valores estão distribuídos de forma homogênea em torno do centro do conjunto de dados. Nesse

caso, a média é uma boa medida de tendência central. A turma B apresenta um valor extremo que desvia a média mais para a esquerda do conjunto de dados. Neste caso, a mediana é mais indicada como medida de tendência central, pois reflete melhor a tendência dos dados.

Tabela 3.1 – Notas dos alunos com as respectivas médias e medianas

Turma	Notas dos alunos									\bar{X}	Me
A	7	7,5	7,5	8	8	8	8,5	8,5	9	8	8
B	0	7	7,5	7,8	8	8	8,2	8,5	9	7,1	8

3.1.2 Mediana

Mediana (Me): é o valor cuja posição separa o conjunto de dados em duas partes iguais; metade do número de elementos está acima do valor mediano e a outra metade abaixo do valor mediano.

Para obter o valor mediano de uma distribuição de dados, primeiro ordene os valores. Isso poderá ser feito tanto em ordem crescente quanto em ordem decrescente. Depois, determine a posição que o valor mediano ocupa, pela seguinte expressão:

$$\text{Pos Me} = \frac{n + 1}{2}$$

Esta fórmula não fornece o valor mediano, mas sim sua localização no conjunto de dados. A forma de determinar o valor mediano depende se o número de observações que compõe o conjunto de dados é par ou ímpar.

– **Número ímpar de elementos:** o valor mediano é a observação que ocupa a posição $(n+1)/2$ desse conjunto de dados.

Exemplo: Calcule a mediana do seguinte conjunto de dados:

2, 4, 6, 7, 11

Solução: o número de observações $n=5$ é ímpar. O valor mediano é a observação central desse conjunto de dados. Pela fórmula da posição da mediana, tem-se $\text{Pos Me} = \frac{5+1}{2} = 3^{\text{a}}$ posição.

A mediana é o valor seis que se encontra na terceira posição no conjunto de dados. O número seis possui duas observações à sua esquerda e duas observações à sua direita, ou seja, 50% dos valores do conjunto de dados são inferiores a seis e 50% dos valores são superiores a seis.

- Número par de elementos: quando o número de observações no conjunto de dados é par, a posição $(n+1)/2$ não será um número inteiro. A mediana será dada pela média aritmética das duas observações centrais dos dados ordenados.

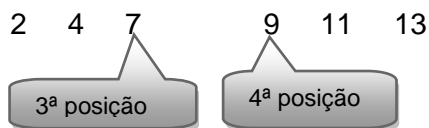
Exemplo: Calcule a mediana do seguinte conjunto de dados:

2, 4, 7, 9, 11, 13.

Solução: Como o número de observações $n = 6$ é par, não existe um valor central. Pela fórmula da posição da mediana, tem-se:

$$\text{Pos Me} = \frac{6+1}{2} = 3,5^{\text{a}} \text{ posição.}$$

O valor mediano está entre a 3^a e a 4^a posição. Nesses casos, o valor mediano não será um dos valores da distribuição e sim a média aritmética dos valores que se encontram nessas duas posições. A terceira posição é ocupada pelo valor sete e a quarta posição é ocupada pelo valor nove.



$$\text{Me} = \frac{7+9}{2} = 8$$

A mediana é o valor oito ($\text{Me} = 8$). Este valor possui três observações à sua esquerda e três observações à sua direita, ou

seja, 50% dos valores do conjunto de dados são inferiores a oito e 50% dos valores são superiores a oito.

A vantagem da mediana é que ela não é influenciada por valores extremos, pois ela depende da posição e não dos valores das observações no conjunto de dados.

3.1.3 Moda

Moda (Mo): é o valor que ocorre com maior frequência em um conjunto de dados.

Exemplo 1: Determine a moda do seguinte conjunto de dados:

2, 5, 7, 9, 13, 15, 22

Solução: Este conjunto de dados não possui moda, pois todos os valores ocorrem uma única vez. Nesse caso, dizemos que o conjunto de dados apresenta uma distribuição amodal.

Exemplo 2: Determine a moda do seguinte conjunto de dados:

16, 19, 19, 21, 21, 21, 23, 27

Solução: $Mo = 21$, e a distribuição é unimodal, pois possui apenas uma moda.

Exemplo 3: Determine a moda do seguinte conjunto de dados:

2, 7, 7, 13, 15, 15, 22

Solução: Esta distribuição apresenta duas modas, $Mo_1 = 7$ e $Mo_2 = 15$, sendo denominada distribuição bimodal.

Quando a distribuição apresenta mais de uma moda, o histograma tem mais de um pico. Quando o conjunto de dados apresenta três modas, denomina-se trimodal, e quatro ou mais, multimodal.

3.2 Medidas de dispersão

As medidas de dispersão são medidas estatísticas que caracterizam o quanto um conjunto de dados está disperso em torno de sua tendência central.

Não há nenhuma razão para se calcular a média de um conjunto de dados em que não haja variação desses elementos (Exemplo: 5 5 5 5; $\bar{x}=5$). No entanto, se a variabilidade dos dados for muito grande, sua média terá um grau de confiabilidade tão pequeno que será inútil calculá-la, como discutido no capítulo anterior, na série 2 do exemplo. Portanto, não é possível analisar um conjunto de dados apenas através de uma medida de tendência central, também é necessário analisar de que forma os valores observados estão espalhados em torno de seu centro. Além disso, dois conjuntos de dados podem possuir a mesma média e, no entanto, os valores podem estar distribuídos de forma diferente.

Exemplo 1: A tabela abaixo apresenta a nota de oito alunos da turma A e turma B. Embora as duas turmas de alunos possuam a mesma média, 5, diferem bastante na variabilidade das notas. Enquanto a turma A apresenta notas mais dispersas, na turma B observam-se pequenas variações nas notas obtidas pelos alunos. Dessa forma, para descrever adequadamente um conjunto de dados, além de uma medida que descreva sua tendência central, é necessário uma medida que descreva sua dispersão.

Tabela 3.2 – Notas de oito alunos de duas turmas

Turma A	0	2	4	5	5	6	8	10
Turma B	4	4,5	5	5	5	5	5,5	6

Exemplo 2: Em duas cidades A e B, foram coletados dados sobre temperatura diária, durante um mês, verificando-se que a média das duas cidades foi de $\bar{x} = 20^{\circ}\text{C}$. Em qual das duas cidades é mais agradável de se viver, em termos de clima? Apenas conhecendo a média da temperatura do ar, fica difícil de responder. Nesse caso, as medidas de dispersão podem nos auxiliar a verificar qual cidade possui menor variação em temperatura, e assim, apontar a que possui o clima mais agradável.

Para avaliar o grau de dispersão ou variabilidade dos valores de um conjunto de dados, usaremos dois tipos de medidas de dispersão: **absoluta** (amplitude total, desvio médio, variância e desvio-padrão) e **relativa** (coeficiente de variação de Pearson).

3.2.1 Amplitude total

Para uma rápida medida da variabilidade, podemos calcular a amplitude total (AT), que é a diferença entre o mais alto e o mais baixo valor em uma distribuição.

$$AT = V_{\max} - V_{\min}$$

A amplitude total considera apenas o valor máximo e o valor mínimo, ignorando todos os outros valores no conjunto de dados. Além disso, esses valores podem ser valores extremos ou atípicos. Podemos aperfeiçoar nossa descrição da dispersão, por meio de outras medidas, como o desvio médio.

Exemplo: Calcule a amplitude total das notas dos alunos da turma B:

Turma B	4	4,5	5	5	5	5	5,5	6
----------------	---	-----	---	---	---	---	-----	---

Solução: $AT = V_{\max} - V_{\min} = 6 - 4 = 2 \rightarrow AT = 2$

3.2.2 Desvio médio

Para levar em consideração todos os valores da distribuição, além dos extremos, subtrai-se a média aritmética de cada elemento do conjunto de dados e somam-se as diferenças, calculando, dessa forma, o desvio de cada elemento à média. Como essa soma é sempre igual a zero, pois alguns valores são negativos e outros positivos, considera-se apenas o módulo das diferenças. Portanto, o desvio médio é definido como a média aritmética dos desvios em módulo.

$$\text{Desvio médio} \quad DM = \frac{\sum_{i=1}^N |x_i - \mu|}{N}$$

Exemplo: Calcule o desvio médio para as notas dos alunos da Turma A da tabela 3.2.

Solução: Podemos utilizar uma tabela para realizar os cálculos.

Tabela 3.3 – Notas dos alunos da turma A.

Aluno	Nota	$(x_i - \bar{x})$	$ x_i - \bar{x} $
1	0	-5	5
2	2	-3	3
3	4	-1	1
4	5	0	0
5	5	0	0
6	6	+1	1
7	8	+3	3
8	10	+5	5
Total	40	0	18

Substituindo na fórmula, temos: $DM = \frac{\sum_{i=1}^N |x_i - \mu|}{N} = \frac{18}{8} = 2,25$

Também podemos resolver direto pela fórmula:

$$DM = \frac{\sum_{i=1}^n |x_i - \mu|}{N} = \frac{|0-5| + |2-5| + |4-5| + |5-5| + |5-5| + |6-5| + |8-5| + |10-5|}{8} =$$

$$DM = \frac{5+3+1+0+0+1+3+5}{8} = \frac{18}{8} = 2,25$$

O desvio médio já não é tão usado, pois ele utiliza a função módulo que nem sempre é viável em algumas análises estatísticas. Apesar de não ser muito utilizado na inferência estatística, o desvio médio é considerado uma boa medida de dispersão, quando o objetivo é apenas descrever o conjunto de dados. Além disso, auxilia na compreensão de outras medidas de dispersão como a variância e o desvio-padrão.

3.2.3 Variância

Para o cálculo da **variância**, em vez de considerar o módulo da diferença, eleva-se esta diferença ao quadrado, eliminando, assim, o problema do sinal negativo. A variância é definida como a média aritmética dos quadrados dos desvios.

$$\text{Variância populacional} \quad \sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

$$\text{Variância amostral} \quad s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Se fosse usado n no denominador da fórmula da variância amostral, estaríamos estimando uma medida de variabilidade menor do que a variância da população. Portanto, no cálculo da variância amostral, utiliza-se $(n-1)$ no denominador. Para grandes amostras, não há diferença entre os resultados proporcionados pela utilização de qualquer dos dois divisores, n ou $n - 1$.

Exemplo 1: Calcule a variância para as notas dos alunos da Turma A.

Solução: Pela tabela,

Tabela 3.4 – Notas dos alunos da Turma A.

Turma A	Nota	$(x_i - \mu)$	$(x_i - \mu)^2$
1	0	-5	25
2	2	-3	9
3	4	-1	1
4	5	0	0
5	5	0	0
6	6	+1	1
7	8	+3	9
8	10	+5	25
Total	40	0	70

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \frac{70}{8} = 8,75$$

Pela fórmula:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \frac{(0-5)^2 + (2-5)^2 + (4-5)^2 + (5-5)^2 + (5-5)^2 + (6-5)^2 + (8-5)^2 + (10-5)^2}{8} =$$

$$\sigma^2 = \frac{25 + 9 + 1 + 0 + 0 + 1 + 9 + 25}{8} = \frac{70}{8} = 8,75$$

Exemplo 2: Calcule a variância para a idade de cinco empregados, aleatoriamente selecionados, para comporem a amostra:

$$35 - 42 - 53 - 41 - 48$$

Solução: Como se trata de uma amostra, usaremos a seguinte fórmula:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{(35-43,8)^2 + (42-43,8)^2 + (53-43,8)^2 + (41-43,8)^2 + (48-43,8)^2}{5-1} =$$

$$= \frac{77,44 + 3,24 + 84,64 + 7,84 + 17,64}{4} = 47,7 \rightarrow s^2 = 47,7$$

Se desenvolvermos o numerador da expressão sob o radical, chegaremos a uma fórmula mais prática da variância:

$$\text{Variância populacional} \quad \sigma^2 = \frac{1}{N} \left[\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{N} \right]$$

$$\text{Variância amostral} \quad s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right]$$

3.2.4 Desvio-padrão

A desvantagem da variância consiste no fato de suas unidades normalmente não terem sentido. A variância para as notas dos alunos, por exemplo, é medida em “notas ao quadrado”. Pode-se retomar a unidade original dos dados, extraindo a raiz quadrada da variância, denominada desvio-padrão.

O desvio-padrão é definido como a raiz quadrada da média aritmética dos quadrados dos desvios, ou seja, a raiz quadrada da variância.

$$\text{Desvio-padrão populacional} \quad \sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

$$\text{Desvio-padrão amostral} \quad s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

O desvio-padrão calculado usando todos os elementos da população é simbolizado por σ ; o desvio-padrão populacional é um parâmetro. Se o desvio-padrão é calculado a partir de uma amostra, este é representado pelo símbolo s , chamado desvio-padrão amostral, e é considerado um estimador.

Exemplo 1: Calcule o desvio-padrão para as notas dos alunos da Turma A.

Solução: Como no exemplo anterior a variância já foi calculada, basta extrair sua raiz quadrada:

$$\sigma = \sqrt{\sigma^2} = \sqrt{8,75} = 2,958$$

Exemplo 2: Calcule o desvio-padrão para os seguintes dados amostrais:

$$25 - 26 - 33 - 21 - 30$$

Solução: Como se trata de uma amostra, usaremos a seguinte fórmula:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(25-27)^2 + (26-27)^2 + (33-27)^2 + (21-27)^2 + (30-27)^2}{5-1}} =$$

$$= \sqrt{\frac{4+1+36+36+9}{4}} = \sqrt{21,5} = 4,64 \rightarrow s = 4,64$$

3.2.5 Coeficiente de variação de Pearson

O **coeficiente de variação de Pearson (CV)** é uma medida de dispersão relativa que mede a dispersão dos dados em relação à média aritmética. É calculado dividindo-se o desvio-padrão pela média e multiplicando por 100, para expressar o resultado em porcentagem, em vez de se utilizar a unidade de medida da variável em análise.

$$\text{População} \rightarrow CV = \frac{\sigma}{\mu} \cdot 100 \quad \text{Amostra} \rightarrow CV = \frac{s}{\bar{x}} \cdot 100$$

Exemplo 1: Calcule o coeficiente de variação de Pearson das notas dos alunos da turma A e B do exemplo anterior.

Turma A	0	2	4	5	5	6	8	10
Turma B	4	4,5	5	5	5	5	5,5	6

$$CV_A = \frac{2,96}{5} \cdot 100 = 59,2\% \qquad CV_B = \frac{0,56}{5} \cdot 100 = 11,2\%$$

A turma B apresenta menor dispersão relativa do que a turma A, o que indica que o desempenho dos alunos da turma B foi mais homogêneo.

A dispersão relativa também permite comparar duas ou mais distribuições, mesmo que essas se refiram a diferentes fenômenos e sejam expressas em unidades de medida diferentes.

Exemplo 2: Na tabela abaixo são apresentados os valores do desvio-padrão e da média da altura e peso de um grupo de pessoas.

	Média	Desvio-padrão
Altura	174cm	7cm
Peso	78kg	12kg

Embora a diferença nas unidades de medida impossibilite comparar o desvio-padrão de 7cm com o desvio-padrão de 12kg, podemos comparar os coeficientes de variação, que não têm unidades de medida. A variável altura apresenta $CV_{\text{altura}} = 4\%$ e a variável peso, $CV_{\text{peso}} = 15,4\%$. Portanto, a variável peso apresenta maior dispersão relativa do que a variável altura.

Observação: Para facilitar a interpretação do coeficiente de variação, usaremos os seguintes intervalos:

$CV \geq 30\% \rightarrow$ Alta dispersão

$15\% < CV < 30\% \rightarrow$ Média dispersão

$CV \leq 15\% \rightarrow$ Baixa dispersão

No exemplo 1, podemos verificar que a turma A apresenta alta dispersão e a turma B, baixa dispersão.

$$CV_A = \frac{2,96}{5} \cdot 100 = 59,2\% \qquad CV_B = \frac{0,56}{5} \cdot 100 = 11,2\%$$

3.2.6 Quartis e percentis

Como a média e o desvio-padrão são afetados por valores extremos, não são considerados adequados para indicar a tendência central e a dispersão de distribuições assimétricas. Assim, é necessário considerar outras medidas como a mediana, os quartis e os percentis.

Os quartis, denotados por Q_1 , Q_2 e Q_3 , dividem os dados em quatro partes iguais. Cada parte possui 25% das observações.

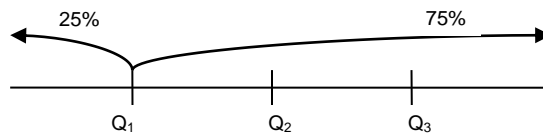


Figura 3.1 – Representação gráfica dos quartis

O primeiro quartil separa a distribuição de tal forma que 25% dos dados são iguais ou inferiores ao Q_1 e 75% dos dados são iguais ou superiores ao Q_1 . Os percentis dividem o conjunto de dados em 100 partes iguais, cada uma com 1%. A mediana é igual a Q_2 e ao P_{50} . Da mesma forma, $Q_1 = P_{25}$ e $Q_3 = P_{75}$. Quando a mediana é a medida de tendência central adotada, o uso dos quartis ou percentis, como medida de dispersão, é mais apropriado.

Para determinar o valor dos quartis ou percentis, assim como no cálculo da mediana, basta encontrar sua posição e verificar no conjunto de dados ordenados qual elemento da distribuição ocupa essa posição. A posição dos quartis e percentis é dada por:

$$\text{Pos}Q_i = \frac{i(n+1)}{4} \text{ com } i = 1, 2, 3 \quad \text{Pos}P_i = \frac{i(n+1)}{100} \text{ com } i = 1, 2, \dots, 99$$

Exemplo 1: Calcule o Q_1 dos dados abaixo:

35 – 46 – 57 – 65 – 74 – 87 – 92

Solução: $\text{Pos}Q_1 = \frac{1 \times (7 + 1)}{4} = 2^{\text{a}} \text{ posição} \rightarrow Q_1 = 46$

Podemos afirmar que 25% dos valores do conjunto de dados são inferiores a 46 ou 75% dos valores são superiores a 46.

Exemplo 2: Calcule o P_{30} e o P_{65} dos dados abaixo:

15 – 21 – 28 - 25 – 30 – 11 – 17 – 12 – 25 – 20 - 16 – 23 – 12 – 10

Solução: primeiro devemos ordenar o conjunto de dados:

10 – 11 – 12 – 12 – 15 – 16 – 17 – 20 – 21 – 23 – 25 – 25 – 28 – 30

$$\text{PosP}_{30} = \frac{30 \times (14 + 1)}{100} = 4,5^{\text{a}} \text{ posição}$$

O valor procurado está entre a 4ª e 5ª posição. A 4ª posição é ocupada pelo valor 12 e a 5ª pelo valor 15. O percentil 30 é a média desses dois valores.

$$P_{30} = 13,5$$

$$\text{PosP}_{65} = \frac{65 \times (14 + 1)}{100} = 9,75^{\text{a}} \text{ posição}$$

A 9ª posição é ocupada pelo valor 21. Como estamos procurando a posição 9,75ª precisamos calcular a parte decimal que ainda falta e somá-la e esse valor.

Temos que a diferença entre o elemento da 9ª e 10ª posição é: $23 - 21 = 2$. Calculando 0,75 de 2 obtemos 1,5. Portanto, o $P_{65} = 21 + 1,5 = 22,5$.

Interpretando: Podemos afirmar que 30% dos valores são inferiores a 13,5 (P_{30}) ou 70% dos valores do conjunto de dados é superior a 13,5 (P_{30}). E 65% dos valores são inferiores a 22,5 (P_{65}) ou 35% dos valores são superiores a 22,5.

Nos casos em que a posição do quartil ou do percentil não é um número inteiro existem regras ligeiramente diferentes para determinar seus valores. Dependendo da bibliografia ou pacote computacional utilizado os resultados podem diferir um pouco.

3.2.6.1 Gráfico de caixa ou o Esquema dos cinco números

O gráfico de caixa dá uma idéia da posição, dispersão e assimetria de um conjunto de dados. Este gráfico admite várias configurações em sua montagem. Aqui usaremos o esquema dos cinco números, que consistem do valor mínimo, primeiro quartil, mediana, terceiro quartil e valor máximo de um conjunto de dados.

A caixa central do gráfico de caixa tem suas extremidades nos quartis e, assim, abrange a metade interna dos dados. O quadrado dentro da caixa assinala a mediana. A partir da caixa, em ambos os extremos, segue uma linha até o valor máximo e mínimo. Os gráficos de caixa podem ser traçados horizontal ou verticalmente.

A Figura 3.2(a) apresenta o gráfico de caixa para as alturas, medidas em centímetros, de 50 empregados de uma empresa. O valor mediano, $Me=170$, indica o centro da distribuição. Os quartis mostram a dispersão da metade central dos dados, $Q_1 = 166\text{cm}$ e $Q_3 = 175\text{cm}$. As extremidades são representadas pelo valor máximo, 185cm , e o valor mínimo, 156cm , indicando a dispersão do conjunto de dados como um todo. As alturas dos empregados desta empresa apresentam maior dispersão dos quartis aos extremos.

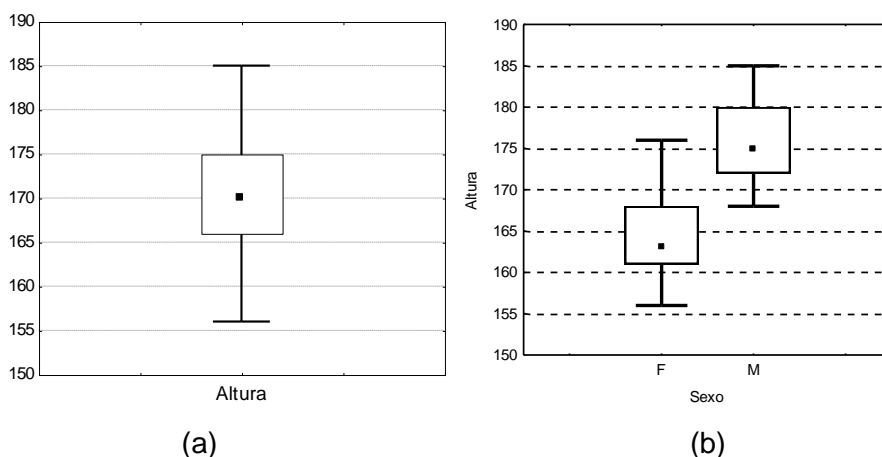


Figura 3.2 – Gráfico de caixa das alturas 50 empregados.

O gráfico de caixa também pode ser usado para se detectar e compreender melhor as diferenças entre dois grupos. Na Figura 3.2(b) são apresentadas as alturas separadas por sexo. Os homens apresentam altura mediana superior à das mulheres, além de menor variabilidade.

Exemplo 1: Com o objetivo de medir a paciência das pessoas, um grupo de pesquisadores desenvolveu o seguinte experimento: as pessoas discam para um número gratuito dado em um anúncio de jornal sobre a promoção do tipo “bom demais para ser verdade” de um novo aparelho de televisão de alta definição. Os que fazem perguntas sobre a “liquidação” são deixados esperando indefinidamente pela telefonista, até que resolvem desligar, por frustração. Monte o diagrama de caixa para os dados abaixo que representam o tempo de espera (em minutos) em resposta ao experimento:

3 – 7 – 2 – 1 – 4 – 3 – 3 – 5 – 6 – 4 – 5 – 6 – 7 – 6 – 5 – 2 – 3 – 9 – 5 – 7

Solução: Primeiro precisamos ordenar os dados:

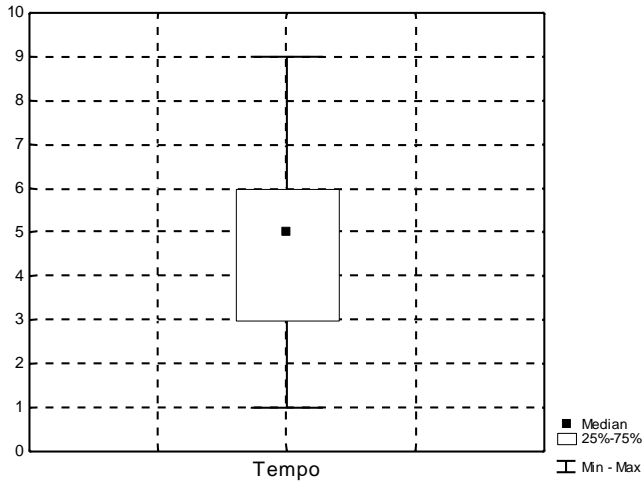
1 – 2 – 2 – 3 – 3 – 3 – 3 – 4 – 4 – 5 – 5 – 5 – 5 – 6 – 6 – 6 – 7 – 7 – 7 – 9

$$V_{\text{máx}} = 9; V_{\text{mín}} = 1;$$

$$\text{Posição da mediana} \rightarrow \text{PosMe} = \frac{20+1}{2} = 10,5^{\text{a}} \text{ posição} \rightarrow \text{Me} = 5$$

$$\text{Posição do } Q_1 \rightarrow \text{PosQ}_1 = \frac{1 \times (20+1)}{4} = 5,25^{\text{a}} \text{ posição} \rightarrow Q_1 = 3$$

$$\text{Posição do } Q_3 \rightarrow \text{PosQ}_3 = \frac{3 \times (20+1)}{4} = 15,75^{\text{a}} \text{ posição} \rightarrow Q_3 = 6$$



Exemplo 2: Com base no desdobramento de tempos de espera para homens e mulheres, monte dois diagramas de caixa no mesmo gráfico.

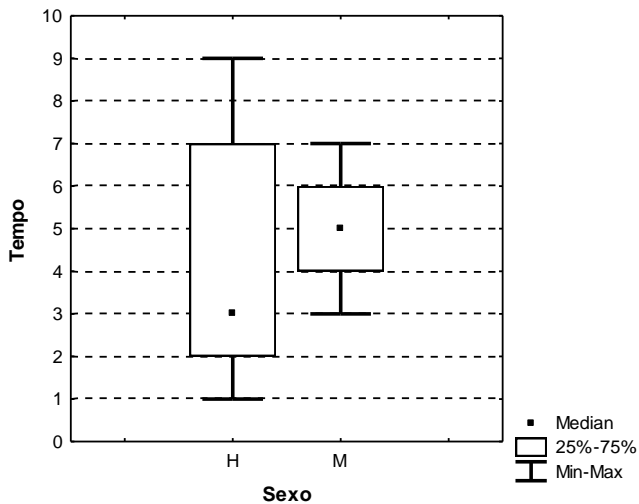
Homens: 5 – 2 – 7 – 9 – 3 – 2 – 3 – 1 – 3 – 7

Mulheres: 3 – 5 – 7 – 4 – 5 – 6 – 6 – 6 – 5 – 4

Solução:

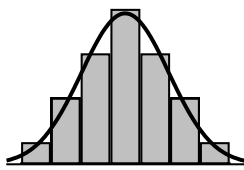
Homens: $V_{\min} = 1$; $Q_1 = 2$; $Me = 3$; $Q_3 = 7$; $V_{\max} = 9$

Mulheres: $V_{\min} = 3$; $Q_1 = 4$; $Me = 5$; $Q_3 = 6$; $V_{\max} = 7$

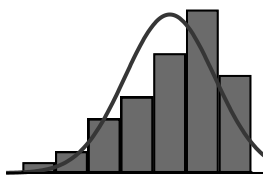


3.3 Assimetria

Além das medidas de posição e dispersão, a forma da distribuição é uma importante fonte de informação sobre o comportamento dos dados. Algumas distribuições podem apresentar forma simétrica ou assimétrica.



$Mo = Me = \bar{x}$
 $As = 0 \rightarrow$ simétrica
Figura 3.3(a)



$\bar{x} < Me < Mo$
 $As < 0 \rightarrow$ assimetria negativa
Figura 3.3(b)



$Mo < Me < \bar{x}$
 $As > 0 \rightarrow$ assimetria positiva
Figura 3.3(c)

Uma distribuição é simétrica [Figura 3.3(a)] se o gráfico apresenta o mesmo comportamento à direita e à esquerda da média. Nesse caso, moda, média e mediana são iguais ou muito próximas. A distribuição de dados apresenta assimetria negativa [Figura 3.3(b)] quando sua cauda esquerda afasta-se mais do pico do que a cauda direita, sendo a moda maior do que a mediana, e esta maior do que a média. A distribuição apresenta assimetria positiva [Figura 3.3(c)] quando sua cauda direita afasta-se mais do pico do que a cauda esquerda, o que faz com que a média seja maior do que a mediana, e esta maior do que a moda.

Existe mais de uma forma de se determinar o coeficiente de assimetria. Aqui, usaremos o **coeficiente de assimetria** pelo segundo critério de Pearson:

$$As = \frac{3 \cdot (\bar{x} - Me)}{s}$$

Quadro 3.1 – Interpretação do coeficiente de assimetria de Pearson

Assimétrica negativa	$As \leq -1$
Assimétrica negativa moderada	$-1 < As < -0,15$
Simétrica	$-0,15 \leq As \leq +0,15$
Assimétrica positiva moderada	$+0,15 < As < +1$
Assimétrica positiva	$As \geq 1$

Exemplo: Os dados a seguir referem-se a uma amostra dos salários recebidos em uma determinada empresa. Determine o coeficiente de assimetria e interprete.

830 920 920 1020 1100 1150 1300 1340 2600 2950

Solução: Para determinar o coeficiente de assimetria, precisamos calcular a média, a mediana e o desvio-padrão.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 1413 \qquad s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = 740,65$$

$$\text{PosMe} = \frac{n+1}{2} = \frac{10+1}{2} = 5,5^{\text{a}} \text{ posição} \qquad \text{Me} = \frac{1100+1150}{2} = 1125$$

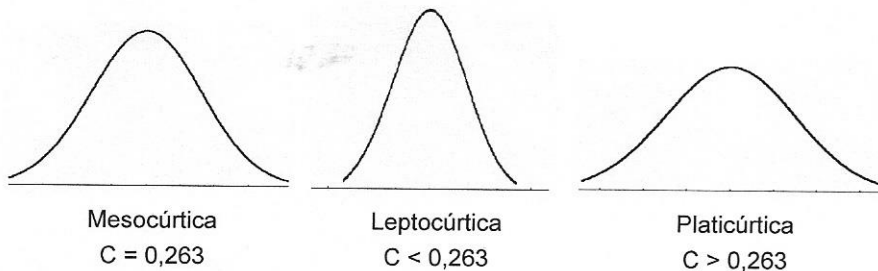
$$As = \frac{3 \cdot (\bar{x} - \text{Me})}{s} = \frac{3 \cdot (1413 - 1125)}{740,65} = 1,17$$

Interpretação: Os salários dos empregados apresentam assimetria positiva ($As=1,17$). Esse resultado evidencia a presença de valores extremos à direita do conjunto de dados que puxam a média para cima, mas não influenciam a mediana.

3.4 Curtose

Denomina-se curtose o grau de achatamento ou alongamento de uma distribuição em relação a uma distribuição padrão, denominada curva normal. A curva normal tem o coeficiente de curtose igual a 0,263

e é classificada como mesocúrtica. Quando a curva é mais alongada que a curva normal, seu coeficiente de curtose é inferior a 0,263 e esta curva é classificada como leptocúrtica. Quando é mais achatada que a curva normal, denomina-se platocúrtica e apresenta coeficiente de curtose maior que 0,263.



O cálculo deste coeficiente é dado por:

$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

Exemplo: Os dados a seguir referem-se aos salários recebidos em uma determinada empresa. Calcule a curtose.

610	612	620	630	638	640	653	668	670	680
680	690	700	710	718	720	735	740	755	770
774	780	782	785	786	792	796	797	810	820
860	890	920	931	954	960	980	1010	1070	1075

Solução:

$$\text{Posição do } Q_1 \rightarrow \text{Pos}Q_1 = \frac{1 \times (40 + 1)}{4} = 10,25^{\text{a}} \text{ posição} \rightarrow Q_1 = 680$$

$$\text{Posição do } Q_3 \rightarrow \text{Pos}Q_3 = \frac{3 \times (40 + 1)}{4} = 30,75^{\text{a}} \text{ posição} \rightarrow Q_3 = 850$$

$$\text{Posição do } P_{10} \rightarrow \text{Pos}P_{10} = \frac{10 \times (40 + 1)}{100} = 4,1^{\text{a}} \text{ posição} \rightarrow P_{10} = 630,8$$

$$\text{Posição do } P_{90} \rightarrow \text{Pos}P_{90} = \frac{90 \times (40 + 1)}{100} = 36,9^{\text{a}} \text{ posição} \rightarrow P_{90} = 978$$

$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{850 - 680}{2(978 - 630,8)} = 0,245$$

Interpretação: C=0,245, leptocúrtica, ou seja, os salários apresentam uma distribuição mais alongada do que a curva normal.

Quadro 3.2 – Resumo das medidas de posição

Medida	Definição	Vantagens	Desvantagens
Média	$\bar{x} = \frac{\sum x_i}{n}$	Usada em muitos métodos estatísticos.	Afetada por valores extremos.
Mediana	Valor central	Apropriada quando há valores extremos ou distribuições assimétricas.	Usada em poucos métodos estatísticos
Moda	Valor mais freqüente	Apropriada para dados qualitativos.	Nem sempre existe. Pode haver mais de uma moda. Não se presta à análise matemática.

Quadro 3.3 – Resumo das medidas de dispersão

<p>Variância populacional</p> $\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$ <p>Desvio-padrão populacional</p> $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$ <p>Coef. de variação populacional</p> $CV = \frac{\sigma}{\mu} . 100$	<p>Variância amostral</p> $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$ <p>Desvio-padrão amostral</p> $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$ <p>Coeficiente de variação amostral</p> $CV = \frac{s}{\bar{x}} . 100$
---	--

Exercícios resolvidos

1) Os dados abaixo representam o número de erros no preenchimento de apólices de seguro, no mês de janeiro, em uma seguradora. Calcule a média, moda e mediana. Interprete os resultados.

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

Solução: Para o cálculo da mediana é necessário ordenar os dados:

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

Diagrama de posição para a mediana:

$$\text{Mediana} \rightarrow \text{PosMe} = \frac{n+1}{2} = 30,5^{\text{a}} \text{ posição}$$

Procuramos os elementos que estão na 30ª e 31ª posição e calculamos a média desses elementos:

$$\text{Me} = \frac{1+1}{2} = 1 \rightarrow \text{Me} = 1$$

Moda \rightarrow $\text{Mo} = 1$ (valor que ocorre com a maior frequência)

$$\text{Média} \rightarrow \bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{79}{60} = 1,32$$

Interpretação: No preenchimento de apólices de seguro observa-se em média 1,32 erro por apólice; 50% dos erros são maiores ou iguais a 1, sendo esse, também, o número de erros que ocorre com a maior frequência.

2) Dados os dois conjuntos de dados amostrais, calcule:

Grupo A:

25 28 34 40 10 21 75 23 27 15 24 76 15 49 18

Grupo B:

28 39 15 43 41 20 35 17 29 24

a) a média de cada grupo.

b) a mediana de cada grupo.

c) Compare os valores obtidos para a média e a mediana no grupo A e no grupo B. A média é maior ou menor que a mediana? Explique o porquê.

Solução:

a)

Grupo A

$$\bar{x} = \frac{25+28+34+40+10+21+75+23+27+15+24+24+76+15+49+18}{15} = 32$$

Grupo B

$$\bar{x} = \frac{28+39+15+43+41+20+35+17+29+24}{10} = 29$$

b)

Grupo A

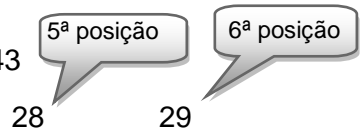
10 15 15 18 21 23 24 25 27 28 34 40 49 75 76

$$\text{PosMe} = \frac{15+1}{2} = 8^{\text{a}} \text{ posição} \quad \text{Me} = 25$$

Grupo B

15 17 20 24 28 29 35 39 41 43

$$\text{PosMe} = \frac{10+1}{2} = 5,5^{\text{a}} \text{ posição}$$



$$\text{Me} = \frac{28+29}{2} = 28,5 \quad \text{Me} = 28,5$$

c)

No grupo A, a média e mediana estão mais afastadas, pois a média é influenciada por valores extremos. Nesse grupo, há dois

valores discrepantes, o 75 e o 76, o que fez com que a média ficasse afastada do centro dos dados. Já no grupo B, a média e mediana estão bem mais próximas, pois esse conjunto de dados não apresenta valores extremos.

3) A distribuição de frequência a seguir apresenta os níveis de escolaridade de 50 empregados de uma empresa. Considerando o tipo de variável, calcule a medida de tendência central mais adequada para descrever a distribuição dada.

Nível de escolaridade	Número de funcionários
Pós-graduação	4
Superior completo	9
Superior incompleto	13
Ensino Médio completo	15
Ensino Médio incompleto	7
Ensino Fundamental completo	2
Total	50

Solução: Moda, porque os dados são qualitativos. A variável com maior frequência é Ensino Médio completo, $f_i = 15$. A mediana também poderia ser usada, pois a variável é qualitativa ordinal. Pos $Me = 25,5^a$ posição. $Me =$ Superior completo.

4) Na tabela abaixo, são apresentadas amostras de salários mensais, em reais, de professores de escolas do Ensino Fundamental da rede pública e privada.

Professores da rede pública	586	581	587	568	548	559	599	562
Professores da rede privada	418	384	403	376	397	383	394	408

a) Obtenha a média, o desvio-padrão e o coeficiente de variação dos salários dos professores da rede pública e privada.

b) Compare os dois grupos e interprete os resultados.

Solução a):

x_i (Salários rede pública)	x_i^2	x_i (Salários rede privada)	x_i^2
586	343.396	418	174.724
581	337.561	384	147.456
587	344.569	403	162.409
568	322.624	376	141.376
548	300.304	397	157.609
559	312.481	383	146.689
599	358.801	394	155.236
562	315.844	408	166.464
Σ 4.590	Σ 2.635.580	Σ 3.163	Σ 1.251.963

Rede pública

$$\text{Média} \rightarrow \bar{x} = \frac{\sum x_i}{n} = \frac{4.590}{8} = 573,75$$

$$\text{Desvio-padrão} \rightarrow s = \sqrt{\frac{1}{n-1} \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right]} = \sqrt{\frac{1}{8-1} \left[2.635.580 - \frac{(4.590)^2}{8} \right]} = 17,19$$

$$\text{Coeficiente de variação} \rightarrow CV = \frac{s}{\bar{x}} \times 100 = \frac{17,19}{573,75} \times 100 = 2,99\%$$

Rede privada

$$\text{Média} \rightarrow \bar{x} = \frac{3.163}{8} = 395,38$$

$$\text{Desvio-padrão} \rightarrow s = \sqrt{\frac{1}{8-1} \left[1.251.963 - \frac{(3.163)^2}{8} \right]} = 14,10$$

$$\text{Coeficiente de variação} \rightarrow CV = \frac{14,10}{395,38} \times 100 = 3,57\%$$

b) A média salarial da rede pública é maior do que a média salarial da rede privada. Apesar de o desvio-padrão da rede pública ser maior do que o da rede privada, os salários da rede pública apresentam dispersão relativa (CV) menor do que na rede privada.

5) Os dados a seguir referem-se aos salários recebidos em uma determinada empresa. Calcule a média, mediana, desvio-padrão e coeficiente de variação. Interprete os resultados.

610	612	620	630	638	640	653	668	670	680
680	690	700	710	718	720	735	740	755	770
774	780	782	785	786	792	796	797	810	820
860	890	920	931	954	960	980	1010	1070	1075

Solução:

Média → $\bar{x} = \frac{31.211}{40} = 780,28$

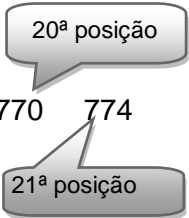
Desvio-padrão → $s = \sqrt{\frac{1}{40-1} \left[24.981.007 - \frac{(31.211)^2}{40} \right]} = 126,88$

Coefficiente de variação → $CV = \frac{126,88}{780,27} \times 100 = 16,26\%$

Mediana

Posição da mediana → $PosMe = \frac{40+1}{2} = 20,5^{a} \text{ posição} \rightarrow 770 \quad 774$

$Me = \frac{770+774}{2} = 772 \rightarrow Me = 772$



Interpretação: O salário médio na empresa é R\$780,28, sendo que 50% dos empregados recebem acima de R\$772,00. O desvio-padrão é de R\$126,88 o que representa uma média dispersão (CV = 16,26%).

6) A seguir, estão os tempos de espera, em minutos, para atendimento no caixa de 19 clientes de uma agência bancária. Monte o gráfico de caixa.

6,5	6	11	9	4	13	3	7	12	8	12	5	9,5	5	11	12	7,5	7	6
-----	---	----	---	---	----	---	---	----	---	----	---	-----	---	----	----	-----	---	---

Solução: Primeiro precisamos ordenar os dados:

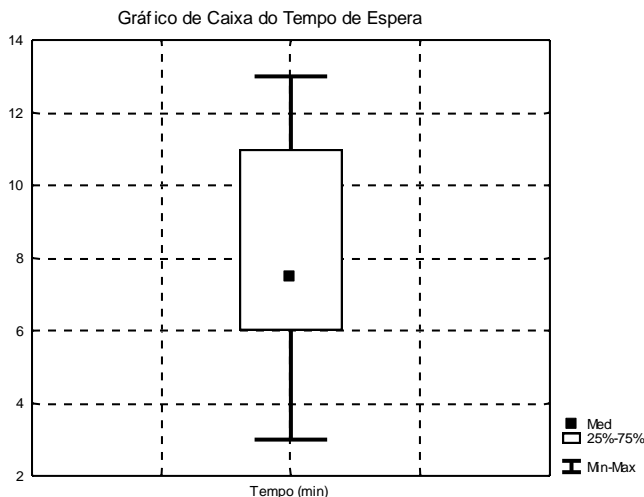
3	4	5	5	6	6	6,5	7	7	7,5	8	9	9,5	11	11	12	12	12	13
---	---	---	---	---	---	-----	---	---	-----	---	---	-----	----	----	----	----	----	----

$V_{\text{máx}} = 13; V_{\text{mín}} = 3;$

Posição da mediana → $PosMe = \frac{19+1}{2} = 10^{a} \text{ posição} \rightarrow Me = 7,5$

Posição do $Q_1 \rightarrow \text{Pos}Q_1 = \frac{1 \times (19 + 1)}{4} = 5^{\text{a}} \text{posição} \rightarrow Q_1 = 6$

Posição do $Q_3 \rightarrow \text{Pos}Q_3 = \frac{3 \times (19 + 1)}{4} = 15^{\text{a}} \text{posição} \rightarrow Q_3 = 11$

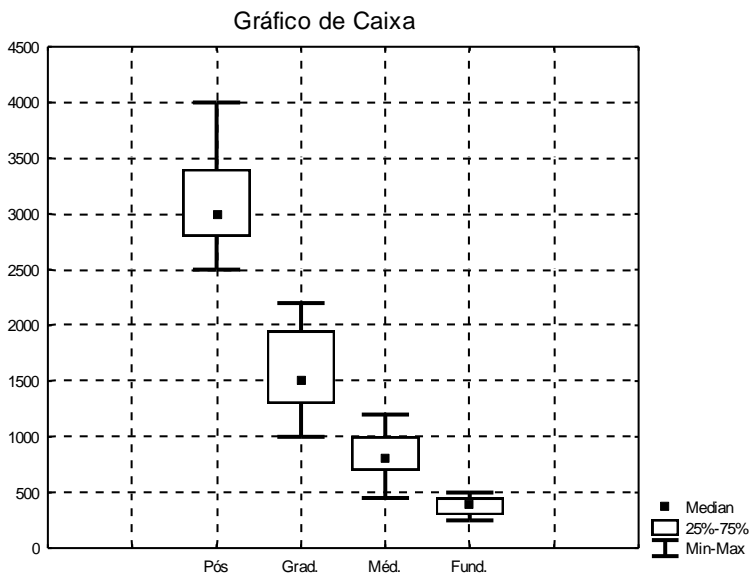


7) Foi feita uma pesquisa para verificar a relação entre escolaridade e renda. A tabela a seguir apresenta o grau de escolaridade máximo de cada pessoa e sua renda. Para a variável renda são apresentados os valores máximo, mínimo, primeiro quartil, terceiro quartil e mediana. Com estes dados monte gráficos de caixa que comparem os quatro grupos de escolaridade, todos no mesmo gráfico. Escreva um breve resumo da relação observada entre educação e renda.

Grau de escolaridade máximo	V_{\min}	Q_1	Me	Q_3	V_{\max}
Pós-graduação	2.500	2.800	3.000	3.400	4.000
Graduação	1.000	1.300	1.500	1.950	2.200
Ensino Médio	450	700	800	1.000	1.200
Ensino Fundamental	250	300	400	450	500

Solução: como podemos observar no gráfico, a renda dos indivíduos com Ensino Fundamental é a que apresenta menor valor mediano e menor variação. A renda dos indivíduos com pós-

graduação é a que tem maior valor mediano e maior dispersão, principalmente nos valores acima da mediana.



Exercícios complementares

1) Os dados a seguir representam a percentagem de calorias provenientes de gordura em lanches de carne e frango, a partir de uma amostra em uma cadeia de lanchonetes:

Carne → 43; 47; 48; 51; 51; 55; 55; 57; 59

Frango → 20; 22; 45; 46; 53; 54; 56; 57

a) Calcule a média, a mediana e a moda para a percentagem de gordura em lanches de carne. Que medida de tendência central melhor representa esse conjunto de dados? Por quê?

b) Calcule a média, a mediana e a moda para a percentagem de gordura em lanches de frango. Que medida de tendência central melhor representa esse conjunto de dados? Por quê?

c) Calcule a variância, o desvio-padrão e o coeficiente de

variação para a percentagem de gordura em lanches de carne.

d) Calcule a variância, o desvio-padrão e o coeficiente de variação para a percentagem de gordura em lanches de frango.

e) Com relação à percentagem de gordura nos lanches de frango e nos de carne, qual tipo de lanche possui a menor **dispersão relativa**?

2) Um fabricante de pilhas extraiu uma amostra de 15 pilhas, de um mesmo lote, e as utilizou até que descarregassem. O número de horas que cada uma demorou até descarregar foi:

425	341	317	535	276	441	917	1.068
509	254	497	571	285	479	468	

Calcule a média, a mediana, desvio-padrão, coeficiente de variação. Interprete os resultados obtidos.

3) O peso de uma população de 8 pessoas foi obtido utilizando dois tipos de balanças: analógica e digital. Na balança analógica, a média de peso foi de $\mu = 74,5\text{kg}$ com desvio-padrão de $\sigma = 14,5\text{kg}$; na balança digital, a média de peso foi de $\mu = 76,2\text{kg}$ com desvio-padrão de $\sigma = 8,7\text{kg}$. Qual das duas balanças é relativamente mais precisa?

4) Que medida de tendência central representa o ponto de frequência máxima em um conjunto de dados?

5) Que medida de tendência central divide um conjunto de dados ao meio, quando os valores se dispõem em ordem crescente ou decrescente?

6) Que medida de tendência central é considerada o ponto de equilíbrio de um conjunto de dados?

7) Explique a relação entre a variância e o desvio-padrão. Alguma dessas medidas pode ser negativa? Justifique dentro do contexto da Estatística.

8) Obtenha um conjunto de dados para o qual $n = 6$, $\bar{x} = 9$ e $s = 0$.

9) Por que o desvio-padrão é usado com mais freqüência do que a variância?

10) Um técnico de laboratório estudou medidas recentes efetuadas com dois instrumentos diferentes. Com o primeiro instrumento, obteve média de 27,4cm e desvio-padrão de 1,1cm. Com o segundo instrumento, obteve média de 25,32cm e desvio-padrão de 0,7cm. Qual dos dois instrumentos é **relativamente** mais preciso? Justifique sua resposta.

Obs.: A precisão está associada à variação.

11) Durante um período de uma hora, foi anotado o valor gasto, em reais, por 20 fregueses. Calcule a média, mediana, moda, o desvio-padrão e coeficiente de variação. Interprete os resultados, elaborando um texto coeso.

1,50 2,75 1,50 5,50 3,50 4,00 3,00 6,50 4,50 2,75
1,50 1,50 4,00 6,50 4,75 3,50 1,50 4,75 2,75 1,50

12) A seguir estão os tempos de espera, em minutos, para atendimento no caixa de 18 clientes de uma agência bancária:

9	6	12	11	6	12	3	12	6	9	5	8	13	5	6	4	8	7
---	---	----	----	---	----	---	----	---	---	---	---	----	---	---	---	---	---

Determine a média, mediana, variância, o desvio-padrão e coeficiente de variação. Interprete os resultados.

13) Em que casos a mediana é mais indicada como medida de tendência central do que a média? Cite um exemplo.

14) Os valores a seguir se referem aos salários recebidos em uma determinada empresa. Determine:

520	522	530	545	548	550	560	570	570	580
580	590	600	612	618	618	635	640	655	670
674	680	682	685	689	694	736	747	780	782
786	789	792	793	795	830	880	910	970	975

a) Qual o salário acima do qual estão os 10% mais bem-remunerados?

b) Qual o salário abaixo do qual se encontram os 15% mais mal-remunerados?

15) Dão-se, a seguir, os números de alarmes falsos recebidos por um grupamento de bombeiros durante 17 semanas consecutivas:

3 12 5 6 12 6 12 6 3 11 8 7 5 6 8 8 4

a) Determine a mediana, o primeiro quartil, o terceiro quartil, o valor máximo e o valor mínimo;

b) Desenhe o gráfico de caixa.

Respostas

1.

a) $\bar{x} = 51,6$; $Me = 51$; $Mo_1 = 51$; $Mo_2 = 55$. A percentagem de gordura nos lanches de carne é mais bem representada tanto pela média quanto pela mediana, pois nesse conjunto de dados não há valores extremos.

b) $\bar{x} = 44,1$; $Me = 49,5$; $Mo = \text{amodal}$. A percentagem de gordura nos lanches de frango é mais bem representada pela mediana, pois esse conjunto de dados apresenta valores extremos à sua esquerda, o que faz com que a média seja desviada mais para baixo.

c) $s^2 = 27,04$; $s = 5,2$; $C.V. = 10,04\%$

d) $s^2 = 222,7$; $s = 14,92$; $C.V. = 33,83\%$

e) O lanche de frango apresenta a menor dispersão relativa ($C.V._{\text{frango}} = 10,04\%$).

2. $\bar{x} = 492,2$; $Me = 468$; $s = 228,07$; $C.V. = 46,34\%$. Em média as pilhas duram 492,2 horas, sendo que 50% das pilhas duram 468 horas ou mais. O desvio-padrão é de 228,07 horas o que representa uma alta dispersão ($C.V. = 46,34\%$) no tempo de duração das pilhas.

3. $C.V._{\text{analógica}} = 19,46\%$; $C.V._{\text{digital}} = 11,42\%$. A balança que apresenta a menor dispersão relativa é a digital, pois tem o menor coeficiente de variação (medida de dispersão relativa).

4. Moda

5. Mediana

6. Média

7. O desvio-padrão é a raiz quadrada da variância. Nenhuma destas medidas pode ser negativa, pois medem o grau de afastamento (distância) entre a média e cada elemento do conjunto de dados, e não existe medida de distância negativa.

8. Resposta: 9 9 9 9 9 9

9. A variância é uma medida na qual a unidade de seu resultado está elevada ao quadrado e, assim, não nos possibilita uma análise direta.

10. A precisão está associada à menor variação. Como é solicitado o instrumento **relativamente** mais preciso, precisamos considerar uma medida de dispersão relativa: o coeficiente de variação.

Instrumento 1: $CV = 4,01\%$ Instrumento 2: $CV = 2,76\%$

Com base nos resultados do coeficiente de variação, podemos verificar que o instrumento que possui menor dispersão relativa é o instrumento 2; logo, ele é mais preciso do que o instrumento 1.

11.

1,50	1,50	1,50	1,50	1,50	1,50	2,75	2,75	2,75	3,00
3,50	3,50	4,00	4,00	4,50	4,75	4,75	5,50	6,50	6,50

Média = R\$3,39; Mediana = R\$3,25; Moda = R\$1,50;

Desvio-padrão = R\$1,66; $CV = 48,9\%$

Com base nos resultados, podemos verificar que o valor que é gasto com mais frequência é de R\$1,50. O valor médio gasto pelos fregueses é R\$3,39. O desvio-padrão é de R\$1,66 e indica que há variação nos dados; pelo coeficiente de variação, podemos verificar que é uma alta dispersão (CV=48,9%).

12. Média = 7,89min; mediana = 7,5min; variância = 9,36min²; desvio-padrão = 3,06min e CV = 38,84%

Interpretação: Em média, os clientes esperam 7,89 minutos, sendo que 50% dos clientes esperam mais de 7,5 minutos. O tempo de espera apresenta alta dispersão (CV=38,84%).

13. A mediana é mais indicada do que a média quando os dados apresentam valores extremos.

14.

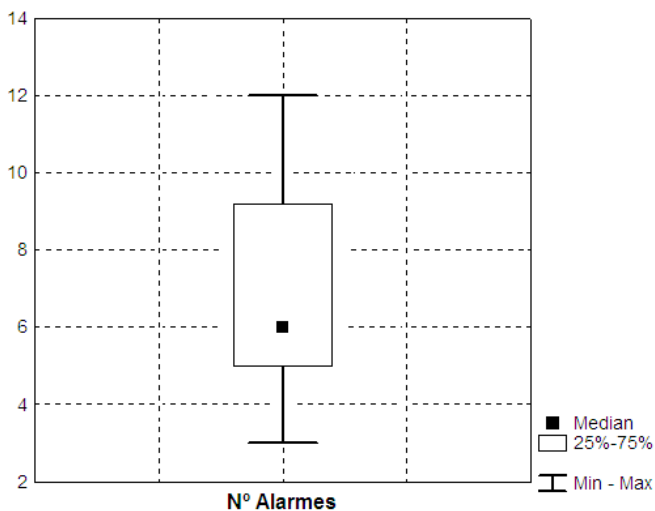
a) $P_{90} = R\$875,00$

b) $P_{15} = R\$551,50$

15.

a) $V_{\min} = 3; Q_1 = 5; Me = 6; Q_3 = 9,5; V_{\max} = 12$

b)



4 Medidas estatísticas para dados agrupados

No cálculo das medidas estatísticas, o ideal é trabalhar com os dados originais não-agrupados, ou seja, os dados brutos. Caso estes não estejam disponíveis, podemos fazer esses cálculos através de uma distribuição de frequência. Nas distribuições de frequência sem intervalo de classes, não há perda de informação, apenas algumas fórmulas são adaptadas para essa nova forma de apresentação dos dados. Já nas distribuições de frequência com intervalo de classe, há perda de informação, o que torna os resultados meras aproximações, nem sempre precisas.

4.1 Dados agrupados sem intervalos de classe

Para calcular a **média aritmética**, **variância** e **desvio-padrão** em cada uma das classes, multiplica-se o valor pelo número de vezes que ele ocorre (frequência).

$$\text{Média amostral} \rightarrow \bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i}$$

$$\text{Variância amostral} \rightarrow s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 \cdot f_i - \frac{\left(\sum_{i=1}^n x_i \cdot f_i \right)^2}{n} \right]$$

Exemplo 1: Calcule a média, o desvio-padrão, a variância e o coeficiente de variação de Pearson para os seguintes dados amostrais:

Distribuição de frequência sem intervalo de classe

x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
2	5	10	20
3	10	30	90
4	15	60	240
5	12	60	300
6	5	30	180
7	3	21	147
Total	50	211	977

Solução:

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i} = \frac{211}{50} = 4,22$$

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 \cdot f_i - \frac{\left(\sum_{i=1}^n x_i \cdot f_i \right)^2}{n} \right] = \frac{1}{50-1} \left[977 - \frac{(211)^2}{50} \right] = 1,7669$$

$$s = \sqrt{1,7669} = 1,33$$

$$CV = \frac{1,33}{4,22} \cdot 100 = 31,52\%$$

Para os dados agrupados sem intervalo de classe, a **moda** é calculada da mesma forma que para os dados não-agrupados. A moda será o valor da variável de maior frequência. Neste exemplo a $Mo = 4$.

A **mediana** para os dados agrupados sem intervalo de classe será calculada do mesmo modo que para dados não-agrupados. Após o cálculo da posição da mediana, devemos calcular a coluna das frequências acumuladas.

Exemplo 2: Calcule a mediana para os dados abaixo:

Variável (x_i)	f_i	F_i
0	2	2
1	6	8
Me → 2	10	18
3	12	30
4	4	34
	Σ 34	

$$\text{PosMe} = \frac{n+1}{2} = \frac{35}{2} = 17,5$$

Neste exemplo, a posição da mediana está entre a décima sétima e décima oitava posição. Observe a coluna das frequências acumuladas: a terceira linha contém os elementos dessas duas posições.

Podemos verificar, na coluna das frequências acumuladas, que o elemento que está na décima sétima posição é igual a 2 ($x = 2$) e o elemento na décima oitava posição também é igual a dois. Portanto, a mediana será:

$$\text{Me} = \frac{2+2}{2} = 2$$

Exemplo 3: Calcule a mediana para os dados abaixo:

Variável (x_i)	f_i	F_i
12	1	1
14	2	3
Me=15,5 ← 15	1	4
16	2	6
17	1	7
20	1	8
	Σ 8	

$$\text{PosMe} = \frac{n+1}{2} = \frac{9}{2} = 4,5$$

Neste exemplo, o valor mediano é o valor que se encontra entre a quarta e a quinta posição. Observe, na coluna das frequências acumuladas, que o elemento que se encontra na quarta posição (terceira linha) é igual a 15 e o elemento que se encontra na quinta posição (quarta linha) é igual a 16. Portanto, a mediana será:

$$Me = \frac{15+16}{2} = 15,5$$

A mediana é igual a 15,5.

4.2 Dados agrupados com intervalo de classe

No caso de distribuições de frequência com intervalos de classe, como há perda de informação, os valores obtidos para as medidas estatísticas serão meras aproximações, em alguns casos, bastante grosseiras. Esses métodos justificavam-se quando não havia calculadoras e computadores, o que tornava bastante cansativo o cálculo das medidas estatísticas quando o conjunto de dados era muito grande. No entanto, com o avanço da tecnologia, é possível realizar esses cálculos para um grande conjunto de dados, com o auxílio de programas estatísticos ou planilhas eletrônicas. Caso esses não estejam disponíveis, podemos fazer estes cálculos através de uma distribuição de frequência.

Para o cálculo da média, variância e desvio-padrão, calcula-se o ponto médio da classe e este valor passa a representar a classe. Para continuar usando as mesmas fórmulas da média e variância, para distribuições de frequência sem intervalo de classe, usamos x_i para representar o ponto médio.

$$\text{Ponto médio} \rightarrow x_i = \frac{L_i + L_s}{2}$$

Exemplo1: Calcule a média, variância e desvio-padrão dos seguintes dados:

Distribuição de frequência com intervalo de classe

Notas	f_i	x_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
0 - 2	6	1	6	1	6
2 - 4	12	3	36	9	108
4 - 6	7	5	35	25	175
6 - 8	4	7	28	49	196
8 - 10	2	9	18	81	162
Total	31		123		647

$$\text{Média} \rightarrow \bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i} = \frac{123}{31} = 3,97$$

$$\text{Variância} \rightarrow s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 \cdot f_i - \frac{\left(\sum_{i=1}^n x_i \cdot f_i \right)^2}{n} \right] = \frac{1}{31-1} \left[647 - \frac{(123)^2}{31} \right] = 5,3$$

$$\text{Desvio-padrão} \rightarrow s = \sqrt{5,3} = 2,30$$

Para o cálculo da **mediana** usamos a seguinte fórmula:

$$\text{Mediana} \rightarrow Me = L_i Me + \left[\frac{P - F_{ant}}{f_i} \right] \cdot h \quad \text{onde } P = \frac{\sum f_i}{2}$$

$L_i Me$ = limite inferior da classe que contém a mediana

P = posição da mediana

F_{ant} = frequência acumulada anterior a classe que contém a mediana

f_i = frequência simples da classe que contém a mediana

h = amplitude do intervalo de classe

Exemplo 2: Calcule a mediana.

Notas	f_i	F_{\downarrow}
0 - 2	6	6
2 - 4	12	18
4 - 6	7	25
6 - 8	4	29
8 - 10	2	31
Total	31	

Na 1ª classe estão os elementos que ocupam da 1ª até a 6ª posição

Na 2ª classe estão os elementos que ocupam da 7ª até a 18ª posição

$$\text{Posição da mediana} \rightarrow P = \frac{\sum f_i}{2} = \frac{31}{2} = 15,5^{\text{a}} \text{ posição}$$

Para determinar a classe que contém a mediana, calcule a frequência acumulada 'abaixo de'. A 15,5ª posição está na segunda classe da distribuição.

$$Me = L_i + \left[\frac{P - F_{\text{ant}}}{f_i} \right] \cdot h = 2 + \left[\frac{15,5 - 6}{12} \right] \cdot 2 = 3,58 \quad \mathbf{Me = 3,58}$$

Obs.: Não calcule as medidas estatísticas com base em uma distribuição de frequência com intervalos de classe, se você tiver acesso aos dados brutos.

Exercícios resolvidos

1) Os dados da distribuição de frequência abaixo representam os salários de 35 funcionários de uma empresa.

Salários (R\$)	f_i
830,00	2
850,00	5
920,00	8
960,00	10
1.020,00	6
1.300,00	3
1.540,00	1
Total	35

Calcule a moda, média, mediana, o desvio-padrão, a variância, o coeficiente de variação e o coeficiente de assimetria. Interprete os resultados.

Solução

Cálculo da moda

Moda é o valor que tem maior frequência. Podemos verificar na coluna 2 que a maior frequência é 10, correspondente à variável salário 960,00. $M_o = R\$ 960,00$

Cálculo da média

Para calcular a média, vamos criar a coluna 3, esta coluna é gerada pela multiplicação das colunas 1 e 2.

Coluna 1	Coluna 2	Coluna 3	Coluna 4	Coluna 5	Coluna 6
Salários (R\$) (x_i)	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$	F_i
830,00	2	1660	688900	1377800	2
850,00	5	4250	722500	3612500	7
920,00	8	7360	846400	6771200	15
→ 960,00	10	9600	921600	9216000	25
1.020,00	6	6120	1040400	6242400	31
1.300,00	3	3900	1690000	5070000	34
1.540,00	1	1540	2371600	2371600	35
Total	$\Sigma 35$	$\Sigma 34.430$		$\Sigma 34.661.500$	

O somatório da coluna 3 será usado na equação que define a média, como é mostrado a seguir:

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i} = \frac{34.430}{35} = 983,71$$

Cálculo da variância e do desvio-padrão

Vamos calcular, em primeiro lugar, a variância. O desvio-padrão é a raiz quadrada da variância.

Para tanto, precisamos calcular os valores da variável ao quadrado (x_i^2), coluna 4. Na coluna 5, foram calculados os valores de $x_i^2 \cdot f_i$.

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 \cdot f_i - \frac{\left(\sum_{i=1}^n x_i \cdot f_i \right)^2}{n} \right] = \frac{1}{35-1} \left[34.661.500 - \frac{(34.430)^2}{35} \right] = 23.300,50$$

$$s = \sqrt{23.300,50} = \text{R\$}152,64$$

Cálculo da mediana

Para calcular a mediana, achamos a posição onde ela se encontra:

$$\text{PosMe} = \frac{n+1}{2} = \frac{35+1}{2} = 18$$

A mediana encontra-se na décima oitava posição. Calculamos, agora, as frequências acumuladas (coluna 6).

A mediana estará na quarta linha, onde a frequência acumulada é 25, isto é, o elemento que se encontra na décima oitava posição está contido nessa classe. Portanto, a mediana é igual a $Me = \text{R\$} 960,00$.

Cálculo do coeficiente de variação

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{152,64}{983,71} \cdot 100 = 15,5\%$$

Cálculo do coeficiente de assimetria

$$A_s = \frac{3 \cdot (\bar{x} - Me)}{s} = \frac{3 \cdot (983,71 - 960,00)}{152,64} = 0,466$$

Assimetria positiva moderada

Interpretação: o salário médio nesta empresa é de R\$983,71. Cinquenta por cento dos empregados ganham acima de R\$960,00, sendo este, também, o salário que ocorre com a maior frequência. O desvio-padrão é de R\$152,64, o que resulta em uma média dispersão (CV=15,5%). A distribuição dos salários apresenta uma assimetria moderada positiva.

2) Os dados da distribuição de frequência abaixo representam as alturas de 50 estudantes. Calcule a média, mediana, desvio-padrão, coeficiente de variação e a assimetria.

Altura (cm)	f_i
160 – 165	2
165 – 170	5
170 – 175	10
175 – 180	15
180 – 185	9
185 – 190	6
190 – 195	3
Total	$\Sigma 50$

Solução:

Os dados acima descritos estão agrupados com intervalos de classe. A primeira coisa a ser feita é calcular o ponto médio (coluna 3) de cada intervalo de classe. Usaremos os pontos médios dos intervalos, ao longo dos nossos cálculos.

Coluna1	Coluna 2	Coluna 3	Coluna 4	Coluna 5	Coluna 6	Coluna 7
Altura	f_i	x_i	$x_i \cdot f_i$	F_i	x_i^2	$x_i^2 \cdot f_i$
160 – 165	2	162,5	325	2	26406,25	52812,5
165 – 170	5	167,5	837,5	7	28056,25	140281,3
170 – 175	10	172,5	1725	17	29756,25	297562,5
175 – 180	15	177,5	2662,5	32	31506,25	472593,8
180 – 185	9	182,5	1642,5	41	33306,25	299756,3
185 – 190	6	187,5	1125	47	35156,25	210937,5
190 – 195	3	192,5	577,5	50	37056,25	111168,8
Total	$\Sigma 50$		$\Sigma 8.895$			$\Sigma 1.585.113$

Cálculo da média

Para o cálculo da média, devemos montar a coluna 4, que é obtida multiplicando-se a coluna 2 e a coluna 3.

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i} = \frac{8.895}{50} = 177,90 \text{ cm}$$

Cálculo da mediana

Para o cálculo da mediana, devemos calcular a coluna das frequências acumuladas (coluna 5) e a posição onde esta se encontra:

$$\text{PosMe} = \frac{n+1}{2} = \frac{50+1}{2} = 25,5$$

A mediana encontra-se entre as posições vigésima quinta e vigésima sexta. Na coluna da frequência acumulada, devemos procurar a linha que contém essas posições. Podemos verificar que a quarta linha, que tem frequência acumulada de 32, é a que contém essas posições. Portanto, a mediana deve estar nessa linha.

A quarta classe conterá a mediana.

$$Me = L_i Me + \left[\frac{P - F_{ant}}{f_i} \right] \cdot h$$

$L_i Me$ = limite inferior da classe que contém a mediana = 175

P = posição da mediana = 25

F_{ant} = frequência acumulada anterior à classe que contém a mediana = 17

f_i = frequência simples da classe que contém a mediana = 15

h = amplitude do intervalo de classe = 5

Substituindo os valores:

$$Me = 175 + \left[\frac{25 - 17}{15} \right] \cdot 5 = 177,66 \text{ cm}$$

$$Me = 177,66 \text{ cm}$$

Cálculo do desvio-padrão

Para calcular o desvio-padrão, calcula-se a variância e, depois, extrai-se a raiz quadrada. Para o cálculo da variância, calculam-se os valores da variável ao quadrado (x_i^2), coluna 6. Na coluna 7, foram calculados os valores de $x_i^2 \cdot f_i$.

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 \cdot f_i - \frac{\left(\sum_{i=1}^n x_i \cdot f_i \right)^2}{n} \right] = \frac{1}{50-1} \left[1.585.113 - \frac{(8.895)^2}{50} \right] = 54,95 \text{ cm}^2$$

O desvio-padrão será dado pela raiz quadrada da variância.

$$s = \sqrt{54,95} = 7,41 \text{ cm}$$

Cálculo do coeficiente de variação

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{7,41}{177,90} \cdot 100 = 4,16\%$$

Cálculo do coeficiente de assimetria

$$A_s = \frac{3 \cdot (\bar{x} - Me)}{s} = \frac{3 \cdot (177,90 - 177,66)}{7,41} = 0,097 \text{ Distribuição simétrica}$$

Interpretação dos resultados: A altura média dos estudantes é 177,90cm, sendo que 50% têm estatura superior a 177,66cm. O desvio-padrão é de 7,41cm, o que resulta em baixa dispersão (CV=4,16%). As alturas dos estudantes apresentam uma distribuição simétrica ($A_s = 0,097$).

Exercícios complementares

1) Uma auditoria feita em 60 apólices de seguro revelou os seguintes números de erros no preenchimento do formulário da Corretora FGNick, no mês de março de 2005:

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

- construa uma distribuição de frequência adequada;
- monte o histograma;
- calcule a média, mediana, desvio-padrão, coeficiente de variação e coeficiente de assimetria.

2) Calcule a média, desvio-padrão e coeficiente de variação das idades dos alunos da tabela abaixo.

Distribuição de frequência das idades dos alunos	
Idades	Frequências
20 – 25	13
25 – 30	10
30 – 35	5
35 – 40	2
40 – 45	2
Total	32

Fonte: Fictícia

3) O setor de atendimento ao cliente do Supermercado Esplanada realizou uma pesquisa, para verificar o tempo de espera que cada cliente permanece na fila do caixa. O tempo, em minutos, registrado por cada cliente, está na tabela abaixo. Os dados foram coletados aleatoriamente, ao longo do mês de janeiro de 2008.

Tempo de espera de clientes na fila do caixa de um supermercado	
TEMPOS	f_i
2	11
3	4
4	6
5	9
6	8
7	7
8	5
Total	50

Fonte: Fictícia

Calcule a média, moda, mediana, desvio-padrão, coeficiente de variação e coeficiente de assimetria.

4) A tabela a seguir mostra os salários de 200 empregados de uma empresa. Calcule a média e o coeficiente de variação.

Salários dos Empregados da Empresa FGNick – Março de 2007	
Salário (R\$)	f_i
200,00 – 400,00	24
400,00 – 600,00	36
600,00 – 800,00	50
800,00 – 1000,00	65
1000,00 – 1200,00	15
1200,00 – 1400,00	10
TOTAL	200

Fonte: Recursos Humanos

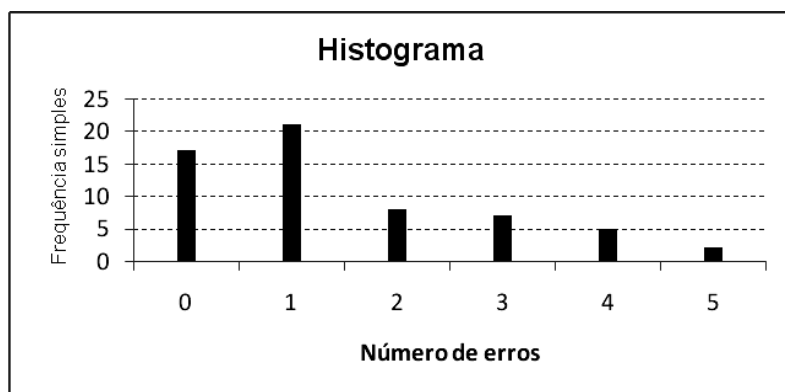
Respostas

1. a)

Número de erros no preenchimento de apólices de seguro da corretora FGNick – março 2005	
X_i	f_i
0	17
1	21
2	8
3	7
4	5
5	2
	$\Sigma 60$

Fonte: Corretora FGNick

b)



c)

Média = 1,47

Desvio-padrão = 1,40

Mediana = 1

Coefficiente de variação = 95,77 %

Coefficiente de assimetria = 1,007

2.

Idades	Frequências	Ponto médio	$x_i f_i$	$x_i^2 f_i$
20 ┆ 25	13	22,5	292,5	6.581,25
25 ┆ 30	10	27,5	275,0	7.562,5
30 ┆ 35	5	32,5	162,5	5.281,25
35 ┆ 40	2	37,5	75,0	2.812,5
40 ┆ 45	2	42,5	85,0	3.612,5
Total	32		Σ 890,0	Σ 25.850

$$\text{Média} = 27,81$$

$$\text{Desvio-padrão} = 5,95$$

$$\text{Coeficiente de variação} = 21,14 \%$$

3. Média:

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} = \frac{240}{50} = 4,8 \text{ minutos}$$

Mediana

$$\text{Pos Me} = \frac{n+1}{2} = \frac{50+1}{2} = 25,5^{\text{a}} \text{ posição}$$

$$\text{Me} = \frac{5+5}{2} = 5 \quad \text{Me} = 5$$

5
25ª posição

5
26ª posição

Moda → Mo = 2

$$\text{Desvio-padrão } s = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 f_i - \frac{\left(\sum_{i=1}^n x_i f_i \right)^2}{n} \right]} = \sqrt{\frac{1}{50-1} \left[1.352 - \frac{240^2}{50} \right]} = 2,02$$

Coeficiente de variação

$$\text{CV} = \frac{s}{\bar{x}} \cdot 100 = \frac{2,02}{4,8} \cdot 100 = 42,09\%$$

Coefficiente de assimetria

$$As = \frac{3 \cdot (\bar{x} - Me)}{s} = \frac{3 (4,8 - 5)}{2,02} = -0,297$$

4. Média

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{148.200}{200} = 741$$

Desvio-padrão

$$s = \sqrt{\frac{1}{n-1} \left[\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n} \right]} = \sqrt{\frac{1}{200-1} \left[123.360.000 - \frac{148.200^2}{200} \right]} = 260,88$$

Coefficiente de variação

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{260,88}{741} \cdot 100 = 35,21\%$$

5 Introdução à Probabilidade

Na Estatística Descritiva, coletamos informações sobre variáveis e descrevemos o que observamos, estudando suas distribuições através de tabelas, gráficos e de medidas estatísticas que nos permitem avaliar a tendência central, a variabilidade e a forma da curva. Imagine que um pesquisador observou o tempo de frenagem de carros de passeio que estavam a 60km/h em asfalto seco. Os dados obtidos nesse experimento, podem ser apresentados em tabelas e gráficos e resumidos por medidas estatísticas como média e desvio-padrão. Mas, se este pesquisador tiver a intenção de estender suas conclusões a outros carros, além daqueles que observou, ele terá que fazer uso da inferência estatística. Para tal, é necessária a aplicação de técnicas que exigem o conhecimento de probabilidade.

A probabilidade permite ao estatístico utilizar as informações de uma amostra, para realizar inferência a respeito da população. Para tal, precisamos de dois conceitos, o de amostragem e o da distribuição de uma variável. Quando a amostra é selecionada aleatoriamente, não podemos determinar ou prever os resultados. Contudo, podemos construir modelos probabilísticos que permitem calcular as chances de ocorrência dos possíveis resultados, através da teoria de probabilidades. Neste capítulo, veremos alguns conceitos de probabilidade que auxiliarão na compreensão das técnicas de inferência estatística.

5.1 Experimento

No estudo de probabilidades, lidamos com diversos experimentos, tais como lançar uma moeda e observar a face voltada para cima; captura de um animal para determinar seu sexo, comprimento e peso; número de funcionários que faltam ao trabalho em determinado dia e seleção de um eleitor para indicar seu candidato nas próximas eleições.

Um experimento é qualquer processo que permite fazer observações. Os experimentos podem ser determinísticos ou aleatórios.

▪ **Experimento determinístico** é aquele cujo resultado é sempre o mesmo, apesar de ser repetido, várias vezes, em condições semelhantes.

Exemplo: a água ferve a 100°C.

▪ **Experimento aleatório** é todo experimento que, mesmo repetido várias vezes, sob condições semelhantes, apresenta resultados imprevisíveis entre os resultados possíveis.

Exemplo: lançamento de uma moeda.

5.2 Espaço amostral

Espaço amostral (S) de um experimento aleatório (ε) é o conjunto de todos os resultados possíveis desse experimento.

Exemplo: No lançamento de um dado honesto, todos os resultados possíveis são:

$$S = \{1, 2, 3, 4, 5, 6\}$$

Observação: Um dado é dito honesto se foi construído de forma homogênea e com medidas simétricas.

5.3 Evento

Evento é qualquer subconjunto do espaço amostral (S).

Exemplo:

A: observar face ímpar no lançamento de um dado. $A = \{1, 3, 5\}$

B: observar face maior do que 4. $B = \{5, 6\}$

C: observar face par. $C = \{2, 4, 6\}$

5.4 Tipos de evento

▪ **Evento impossível** é o evento igual ao conjunto vazio (\emptyset).

Exemplo: D: observar face maior do que 6. $D = \{ \}$

▪ **Evento certo** é o evento igual ao espaço amostral S.

Exemplo: E: observar face menor do que 7. $E = \{1, 2, 3, 4, 5, 6\}$

▪ **Evento simples** é um evento que não comporta decomposição em coleções menores, ou seja, ele tem um único elemento.

Exemplo: F: observar face maior do que cinco. $F = \{6\}$

5.5 Operação com eventos

▪ **União:** $A \cup B$ é o evento que ocorre se, e somente se, A ocorre ou B ocorre ou ambos ocorrem simultaneamente. Podemos representar usando também a nomenclatura (A ou B) = $(A \cup B)$

▪ **Intersecção:** $A \cap B$ é o evento que ocorre se, e somente se, A e B ocorrem simultaneamente. A intersecção de eventos também pode ser representada usando a seguinte nomenclatura: (A e B) = $(A \cap B)$

Observação: se $A \cap B = \emptyset$, então A e B são ditos mutuamente exclusivos ou disjuntos.

▪ **Complementar** (\bar{A}): é o evento que ocorre se, e somente se, A não ocorrer. Dizemos que A e B são complementares se sua união é igual ao espaço amostral, ou seja, $(A \text{ ou } \bar{A}) = S$, e sua intersecção é vazia, $(A \text{ e } \bar{A}) = \emptyset$.

5.6 Tipos de probabilidade

Há três tipos de probabilidade: probabilidade clássica, probabilidade empírica e probabilidade subjetiva. Neste livro, estudaremos os dois primeiros tipos de probabilidade.

5.6.1 Probabilidade clássica

A **probabilidade clássica** aplica-se às situações em que os resultados que compõem o espaço amostral têm a mesma possibilidade de ocorrerem, ou seja, os eventos simples são considerados equiprováveis e o espaço amostral é finito.

$$P(A) = \frac{n(A)}{n(S)} \quad \text{onde}$$

$P(A)$ é a probabilidade de ocorrer o evento A,

$n(A)$ é o número de elementos de A;

$n(S)$ é o número de elementos de S.

Exemplo 1: Lança-se um dado honesto e observa-se a face voltada para cima. Determine a probabilidade de ocorrer face 4.

Solução:

Experimento: lançar um dado.

Espaço amostral: $S = \{1, 2, 3, 4, 5, 6\}$ $n(S) = 6$

Evento A: ocorrer face 4. $A = \{4\}$ $n(A) = 1$

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{6} = 0,1667$$

Interpretação: Com base nesse resultado, podemos afirmar que, no lançamento de um dado honesto, a cada seis lançamentos, um deles será a face 4? Não, porque não se tem certeza do que acontecerá nos vários lançamentos. No entanto, pode-se afirmar que, se repetirmos o lançamento do dado muitas vezes, a proporção de faces 4 se aproximará de 0,1667.

Exemplo 2: Lança-se um dado honesto e observa-se a face voltada para cima. Determine a probabilidade de ocorrer face maior do que 6.

Solução:

Experimento: lançar um dado.

Espaço amostral: $S = \{1, 2, 3, 4, 5, 6\}$ $n(S) = 6$

Evento B: ocorrer face maior do que 6. $B = \{ \}$ $n(B) = 0$

$$P(B) = \frac{n(B)}{n(S)} = \frac{0}{6} = 0$$

Evento impossível!

Exemplo 3: Lança-se um dado honesto e observa-se a face voltada para cima. Determine a probabilidade de ocorrer face menor do que 7.

Solução:

Experimento: lançar um dado.

Espaço amostral: $S = \{1, 2, 3, 4, 5, 6\}$ $n(S) = 6$

Evento C: ocorrer face menor do que 7. $C = \{1, 2, 3, 4, 5, 6\}$

$$n(C) = 6$$

$$P(C) = \frac{n(C)}{n(S)} = \frac{6}{6} = 1$$

Evento certo!

Exemplo 4: Um baralho é constituído de 52 cartas. Retirando desse baralho uma carta ao acaso, calcule a probabilidade de ser do naipe de espadas.

Solução:

Experimento: retirar uma carta do baralho.

Espaço amostral: $n(S) = 52$

Evento A: retirar uma carta de espadas $n(A) = 13$

$$P(A) = \frac{n(A)}{n(S)} = \frac{13}{52} = \frac{1}{4} = 0,25$$

Observação:

▪ Um **evento impossível** possui probabilidade de ocorrência igual a 0 (zero), enquanto um **evento certo** possui probabilidade igual a 1. Logo, a probabilidade de um evento A qualquer é maior ou igual a zero ou menor ou igual a um:

$$0 \leq P(A) \leq 1$$

▪ A soma das probabilidades de um experimento é sempre igual a um.

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

5.6.2 Probabilidade empírica ou frequentista

Na probabilidade clássica, o número de casos favoráveis e o número de casos possíveis são previamente conhecidos. Por exemplo, quando lançamos um dado, há apenas seis resultados possíveis. Já na probabilidade empírica, os resultados são baseados em dados observados. A frequência de resultados no espaço amostral é estimada a partir de um experimento. A **probabilidade empírica** de um evento A ocorrer é a frequência relativa desse evento.

$$P(B) = \frac{\text{frequência do evento B}}{\text{número de vezes em que o experimento foi repetido}} = \frac{f_i}{\sum f_i}$$

Exemplo: A tabela a seguir apresenta grau de instrução dos 40 funcionários do departamento de pessoal de uma pequena empresa. Determine a probabilidade de selecionar aleatoriamente um funcionário dessa empresa e ele ter Ensino Médio completo.

Grau de instrução	f_i
Pós-graduação	3
Superior completo	5
Superior incompleto	11
Ensino Médio completo	14
Ensino Médio incompleto	7
Total	40

Solução:

Experimento: selecionar aleatoriamente um funcionário.

$$n(S) = \sum f_i = 40$$

Evento B: ter Ensino Médio completo. $n(B) = f_i = 14$

$$P(B) = \frac{f_i}{\sum f_i} = \frac{14}{40} = 0,35 = 35\%$$

As probabilidades calculadas com base em freqüências relativas são estimativas da verdadeira probabilidade. À medida que o número total de observações aumenta, as estimativas correspondentes tendem a se aproximar do valor exato. Essa probabilidade é estabelecida pela **Lei dos Grandes Números**.

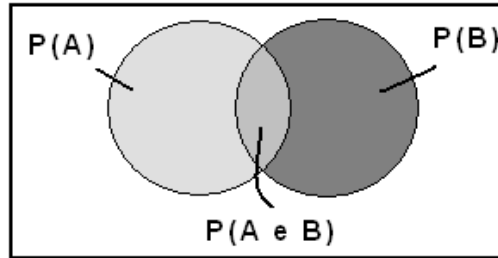
Para saber mais, veja o Anexo I

5.7 Teorema da soma

Seja S um espaço amostral, A e B eventos de S. A probabilidade de que o evento A ocorra ou o evento B ou ambos é dada por:

$$P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B)$$

O teorema da soma consiste em somar a probabilidade do evento A com a probabilidade de ocorrência do evento B; o evento A intersecção B deve, então, ser subtraído desse total, uma vez que já foi incluído duas vezes no cálculo da probabilidade de A e da probabilidade de B.



Exemplo: Numa comunidade de 1.000 habitantes, 400 pessoas são assinantes do jornal A, 300 são assinantes do jornal B e 200 de ambos. Calcule a probabilidade de uma pessoa escolhida ao acaso ser assinante de A ou B.

Solução:

\mathcal{E} : seleccionar um habitante da comunidade $n(S) = 1.000$

A: ser assinante do jornal A $n(A) = 400$

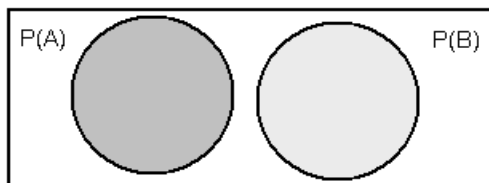
B: ser assinante do jornal B $n(B) = 300$

$P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B) =$

$$= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \text{ e } B)}{n(S)} = \frac{400}{1000} + \frac{300}{1000} - \frac{200}{1000} = 0,5$$

Interpretação: A probabilidade de um habitante seleccionado aleatoriamente ser assinante do jornal A ou do jornal B é de 50%.

Observação: O teorema da soma é simplificado sempre que A e B não podem ocorrer simultaneamente, de modo que $P(A \text{ e } B)$ se torna zero. Nesse caso, dizemos que A e B são eventos mutuamente exclusivos, então **$P(A \text{ ou } B) = P(A) + P(B)$** .



Exemplo 2: Uma urna contém quatro bolas amarelas, duas brancas e três vermelhas. Retirando-se uma bola, calcular a probabilidade de ela ser amarela ou branca.

Solução:

\mathcal{E} : retirar uma bola da urna $n(S) = 9$

A: retirar uma bola amarela $n(A) = 4$

B: retirar uma bola branca $n(B) = 2$

C: retirar uma bola vermelha $n(C) = 3$

$$P(A \text{ ou } B) = P(A) + P(B) = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} = \frac{4}{9} + \frac{2}{9} = \frac{6}{9} = 0,6667$$

Interpretação: A probabilidade de que a bola retirada seja amarela ou branca é de 66,67%.

Exemplo 3: Retirando-se uma carta de um baralho de 52 cartas, qual é a probabilidade de:

a) retirar um ás ou uma carta de copas?

b) retirar um valete ou uma dama?

Solução do item a:

\mathcal{E} : retirar uma carta do baralho $n(S) = 52$

A: retirar um ás $n(A) = 4$

A carta retirada pode ser do naipe de copas e ser um ás.

C: retirar uma carta de copas $n(C) = 13$

$P(A \text{ ou } C) = P(A) + P(C) - P(A \text{ e } C) =$

$$= \frac{n(A)}{n(S)} + \frac{n(C)}{n(S)} - \frac{n(A \text{ e } C)}{n(S)} = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13} = 0,31$$

Solução do item b:

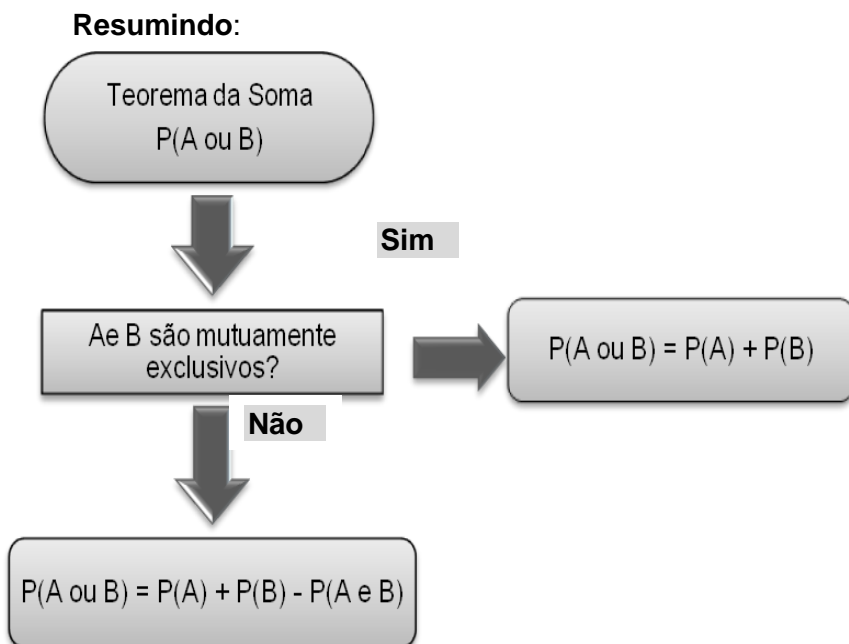
\mathcal{E} : retirar uma carta do baralho $n(S) = 52$

D: retirar uma dama $n(D) = 4$

Eventos mutuamente exclusivos.

V: retirar um valete $n(V) = 4$

$$P(D \text{ ou } V) = P(D) + P(V) = \frac{n(D)}{n(S)} + \frac{n(V)}{n(S)} = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = 0,15$$



5.8 Probabilidade do complementar de um evento

O **complementar de um evento** consiste de todos os resultados para os quais A não ocorre. Os eventos A e \bar{A} são mutuamente exclusivos, pois é impossível um evento e seu complementar ocorrerem ao mesmo tempo. Do teorema da soma de eventos mutuamente exclusivos, temos:

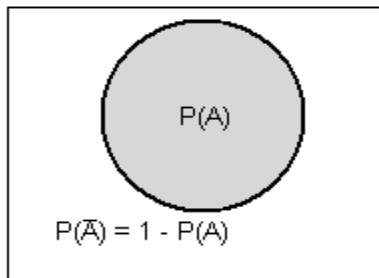
$$P(A \text{ ou } \bar{A}) = P(A) + P(\bar{A})$$

A soma das probabilidades de todos os resultados de um espaço amostral é 1 ou 100%; logo,

$$P(A) + P(\bar{A}) = 1$$

Isolando $P(\bar{A})$, temos:

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



Exemplo: A probabilidade de nascer uma menina é 0,56. Determine a probabilidade de nascer um menino.

Solução: Como os eventos são mutuamente exclusivos, temos:

$$P(\text{menino}) = 1 - P(\overline{\text{menino}}) = 1 - 0,56 = 0,44$$

Interpretação: A probabilidade de nascer um menino é de 44%.

5.9 Probabilidade condicional

Para entender a idéia de probabilidade condicional, vamos trabalhar com o seguinte exemplo: suponha que um dado foi lançado. A probabilidade de ter ocorrido face 4 no lançamento do dado é $1/6$. Suponha, agora, que o dado foi jogado e já se sabe que ocorreu face par. Determine a probabilidade de ter ocorrido face 4.

Solução: Se já saiu face par, só podem ter ocorrido as faces 2, 4 ou 6. O espaço amostral foi reduzido para apenas três casos possíveis. Portanto, a probabilidade de ter ocorrido face 4, sabendo-se que já ocorreu face par, é $1/3$.

Definição: Sejam A e B dois eventos de um mesmo espaço amostral. A probabilidade condicional de A dado que B ocorreu, representada por A/B , é definida como:

$$P(A/B) = \frac{P(A \text{ e } B)}{P(B)} \quad \text{se } P(B) \neq 0$$

Como também a probabilidade condicional de B dado que A ocorreu (B/A):

$$P(B/A) = \frac{P(B \text{ e } A)}{P(A)} \quad \text{se } P(A) \neq 0$$

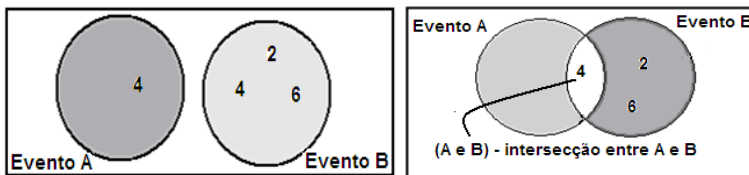
Observação: A expressão $P(B/A)$ indica a probabilidade de ocorrer o evento B, dado que tenha ocorrido o evento A. Note que B/A não é uma fração.

No exemplo acima, temos:

\mathcal{E} : lançar um dado $S = \{1,2,3,4,5,6\}$ $n(S) = 6$

A: ocorrer face 4 $A = \{4\}$ $n(A) = 1$

B: ocorrer face par $B = \{2,4,6\}$ $n(B) = 3$



A probabilidade de ocorrer a face 4, sabendo-se que já ocorreu face par, é:

$$P(A/B) = \frac{P(A \text{ e } B)}{P(B)}$$

onde $P(A \text{ e } B) = \frac{n(A \text{ e } B)}{n(S)} = \frac{1}{6}$ e $P(B) = \frac{n(B)}{n(S)} = \frac{3}{6}$

Substituindo, temos:

$$P(A/B) = \frac{P(A \text{ e } B)}{P(B)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{6} \cdot \frac{6}{3} = \frac{1}{3} \quad P(A/B) = \frac{1}{3} = 0,33$$

5.10 Eventos independentes

Dois eventos A e B são ditos independentes se a probabilidade de ocorrência de um não influenciar a probabilidade de ocorrência do outro.

Exemplo: Considere o experimento de lançar uma moeda e um dado. Determine a probabilidade de ocorrer face 4 no lançamento do dado, sabendo que ocorreu cara no lançamento da moeda.

Espaço amostral associado a esse experimento é:

$$S = \{(1, \text{Ca}), (2, \text{Ca}), (3, \text{Ca}), (4, \text{Ca}), (5, \text{Ca}), (6, \text{Ca}), (1, \text{Co}), (2, \text{Co}), (3, \text{Co}), (4, \text{Co}), (5, \text{Co}), (6, \text{Co})\}$$

$$n(S) = 12$$

Evento A: ocorrer cara no lançamento da moeda

$$A = \{(1, \text{Ca}), (2, \text{Ca}), (3, \text{Ca}), (4, \text{Ca}), (5, \text{Ca}), (6, \text{Ca})\} \quad n(A) = 6$$

Evento B: ocorrer 4 no lançamento do dado

$$B = \{(4, \text{Ca}), (4, \text{Co})\} \quad n(B) = 2$$

O fato de B ocorrer não fornece qualquer informação sobre a ocorrência de A. Os eventos A e B são não relacionados.

$$P(A) = \frac{n(A)}{n(S)} = \frac{6}{12} = \frac{1}{2}$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{2}{12} = \frac{1}{6}$$

$$P(A \text{ e } B) = \frac{n(A \text{ e } B)}{n(S)} = \frac{1}{12}$$

$$P(A/B) = \frac{P(A \text{ e } B)}{P(B)} = \frac{\frac{1}{12}}{\frac{1}{6}} = \frac{1}{12} \cdot \frac{6}{1} = \frac{6}{12} = \frac{1}{2} \quad \Longrightarrow \quad P(A/B) = P(A)$$

Podemos verificar que a probabilidade de ocorrer cara no lançamento da moeda não foi modificada pela ocorrência do número 4 no lançamento do dado. Portanto, dois eventos A e B são independentes se $P(A) = P(A/B)$ ou $P(B) = P(B/A)$.

Logo, podemos concluir que se dois **eventos são independentes**, então:

$$P(A \text{ e } B) = P(A) \cdot P(B)$$

No exemplo acima, também podemos verificar a independência dos eventos da seguinte forma:

$$P(A \text{ e } B) = P(A) \cdot P(B)$$

$$\frac{n(A \text{ e } B)}{n(S)} = \frac{n(A)}{n(S)} \cdot \frac{n(B)}{n(S)}$$

$$\frac{1}{12} = \frac{1}{2} \cdot \frac{1}{6} \quad \longrightarrow \quad \frac{1}{12} = \frac{1}{12}$$

Os eventos A e B são ditos independentes.

Observação: Se A e B **não são** independentes, então A e B são dependentes.

5.11 Teorema do produto

Do cálculo da probabilidade condicional segue o chamado **teorema do produto**:

$$P(A \text{ e } B) = P(A) \cdot P(B/A)$$

Se os eventos A e B são independentes a $P(B/A) = P(B)$ o teorema do produto pode ser simplificado para:

$$P(A \text{ e } B) = P(A) \cdot P(B)$$

Exemplo: Uma urna contém quatro bolas amarelas, cinco bolas brancas e sete bolas vermelhas. Retirando duas bolas da urna, aleatoriamente, qual a probabilidade da primeira e segunda bola serem amarelas?

Solução:

A: a primeira bola é amarela

B: a segunda bola é amarela

Primeiramente, calcule essa probabilidade, supondo que as bolas da urna são retiradas **sem reposição**.

A probabilidade de a primeira bola ser amarela é $4/16$. Como as bolas são retiradas sem reposição, a segunda bola é retirada sem a colocação da primeira. Portanto, se a primeira bola retirada é amarela, na urna restam apenas três bolas amarelas dentre as 15 bolas restantes. A probabilidade para o segundo evento leva em conta o fato de que o primeiro evento já ocorreu; logo, temos:

$$P(A \text{ e } B) = P(A) \cdot P(B/A) = \frac{4}{16} \cdot \frac{3}{15} = \frac{12}{240} = \frac{1}{20} = 0,05$$

Neste caso, dizemos que os eventos são dependentes.

Agora, recalcule essa probabilidade, supondo que as bolas são retiradas **com reposição**.

A probabilidade de a primeira bola ser amarela é $4/16$. Como as bolas são retiradas com reposição, a primeira bola é recolocada na urna antes da retirada da segunda; logo, o número de bolas amarelas continua sendo quatro e o número total de bolas 16. Nesse caso, esses dois eventos são independentes e $P(B/A) = P(B)$. Substituindo, temos:

$$P(A \text{ e } B) = P(A) \cdot P(B) = \frac{4}{16} \cdot \frac{4}{16} = \frac{16}{256} = \frac{1}{16} = 0,0625$$

A distinção entre as duas partes desse problema é muito importante em estatística. Na primeira parte, a amostragem foi feita **sem reposição**. Nesse caso, a ocorrência do segundo evento é afetada pela ocorrência do primeiro evento. Na segunda parte, a amostragem foi feita **com reposição**. As duas retiradas são independentes, porque o segundo evento não é afetado pela ocorrência do primeiro.

A regra da multiplicação pode ser estendida para mais do que dois eventos.

Exemplo: Nos dois exemplos anteriores, suponha que são retiradas três bolas da urna. Calcule a probabilidade de as três bolas serem amarelas.

Solução:

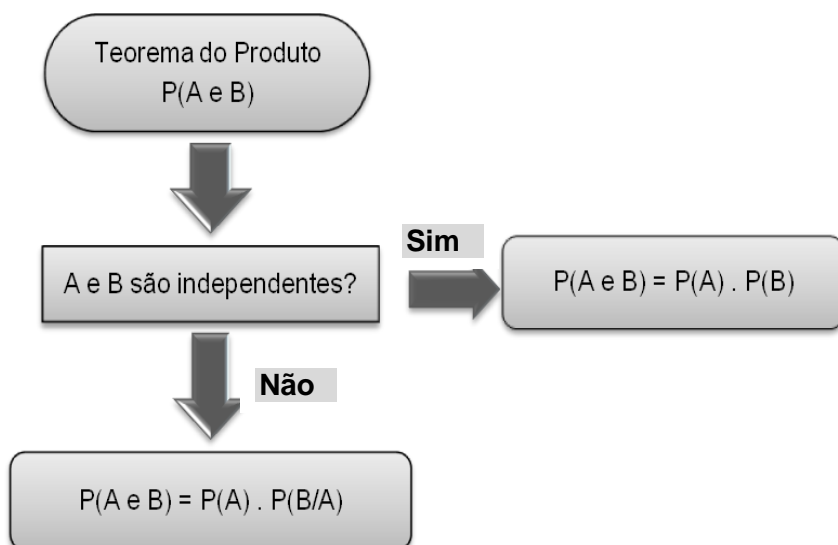
Retirada **sem reposição**:

$$P(A \text{ e } B \text{ e } C) = P(A) \cdot P(B/A) \cdot P(C/A \text{ e } B) = \frac{4}{16} \cdot \frac{3}{15} \cdot \frac{2}{14} = \frac{24}{3.360} = 0,0071$$

Retirada **com reposição**:

$$P(A \text{ e } B \text{ e } C) = P(A) \cdot P(B) \cdot P(C) = \frac{4}{16} \cdot \frac{4}{16} \cdot \frac{4}{16} = \frac{64}{4.096} = 0,0156$$

Resumindo:



5.12 Tabela de contingência

Existem várias formas pelas quais um determinado espaço amostral pode ser visualizado. As tabelas de contingência ou tabela de tabulação cruzada é uma delas. Os valores nas células da tabela são obtidos através da subdivisão do espaço amostral. A tabela 5.1 apresenta os resultados de uma pesquisa sobre o sexo e a opção do estudante em cursos na área de Ciências Sociais e Humanas.

Tabela 5.1 – Tabela de contingência

Sexo	Opção		Total
	Ciências Sociais	Ciências Humanas	
Masculino	60	55	115
Feminino	50	35	85
Total	110	90	200

Em uma **tabela de contingência**, a probabilidade simples também é conhecida como probabilidade marginal, uma vez que o número total de sucessos pode ser obtido através da margem da tabela. A probabilidade simples refere-se à probabilidade de ocorrência de um evento simples. As probabilidades que envolvem dois eventos simultâneos são denominadas probabilidades conjuntas ou combinadas.

Tabela 5.2 – Tabela de probabilidades combinadas e marginais

Sexo	Opção		Total
	Ciências Sociais (A)	Ciências Humanas (B)	
Masculino (M)	$P(M \text{ e } A)$	$P(M \text{ e } B)$	$P(\text{masculino})$
Feminino (F)	$P(F \text{ e } A)$	$P(F \text{ e } B)$	$P(\text{feminino})$
Total	$P(\text{Ciências Sociais})$	$P(\text{Ciências Humanas})$	1,0

Probabilidade marginal

Probabilidade conjunta

Probabilidade marginal

Exemplo: A tabela 5.1 apresenta os resultados de um levantamento, no qual perguntou-se a 115 estudantes do sexo masculino e 85 estudantes do sexo feminino de uma universidade a área de seu curso.

a) Determine a probabilidade de um estudante, aleatoriamente escolhido, ser do sexo feminino. Interprete.

b) Determine a probabilidade de um estudante, aleatoriamente escolhido, ter feito a opção por um curso na área de Ciências Sociais.

c) Determine a probabilidade de um estudante, aleatoriamente escolhido, ter feito a opção por um curso na área de Ciências Humanas e ser do sexo feminino. Interprete.

d) Determine a probabilidade de um estudante, aleatoriamente escolhido, ser do sexo masculino e ter feito a opção por um curso na área de Ciências Sociais.

e) Determine a probabilidade de um estudante, aleatoriamente escolhido, ser do sexo masculino, sabendo que ele fez a opção por um curso na área de Ciências Sociais. Interprete.

f) Monte a tabela de probabilidades combinadas e marginais.

g) Com base nas probabilidades calculadas no item f, determine a probabilidade de selecionarmos um estudante, aleatoriamente, e ele ser do sexo masculino.

h) Com base nas probabilidades calculadas no item f, determine a probabilidade de um estudante, aleatoriamente selecionado, ser do sexo feminino, sabendo que ele fez a opção pelo curso na área de Ciências Humanas.

Solução do item a)

Probabilidade marginal

$$P(\text{Feminino}) = \frac{\text{número de estudantes do sexo feminino}}{\text{número total de estudantes}} = \frac{85}{200} = 0,425$$

Interpretação: Existe uma probabilidade de 42,5% de que um estudante, aleatoriamente escolhido, seja do sexo feminino.

Solução do item b)

Probabilidade marginal

$$P(\text{ciências sociais}) = \frac{110}{200} = 0,55$$

Solução do item c)

Probabilidade conjunta

$$P(\text{feminino e ciências humanas}) = \frac{35}{200} = 0,175$$

Interpretação: Existe uma probabilidade de 17,5% de que um estudante, aleatoriamente escolhido, seja do sexo feminino e tenha escolhido um curso na área de Ciências Humanas.

Solução do item d)

$$P(\text{masculino e ciências sociais}) = \frac{60}{200} = 0,3$$

Probabilidade conjunta

Solução do item e)

$$P(\text{masculino/ciências sociais}) = \frac{60/200}{110/200} = \frac{60}{110} = 0,545$$

Probabilidade condicional

Interpretação: Sabendo-se que o estudante fez a opção por um curso na área de Ciências Sociais, há uma probabilidade de 66,67% de que ele seja do sexo masculino.

Solução do item f)

Tabela de probabilidades combinadas e marginais

Sexo	Opção		Total
	Ciências Sociais (A)	Ciências Humanas (B)	
Masculino (M)	0,3	0,275	0,575
Feminino (F)	0,25	0,175	0,425
Total	0,55	0,45	1,0

Solução do item g)

$$P(\text{masculino}) = 0,575$$

Probabilidade marginal

Solução do item h)

$$P(\text{feminino/Ciências Humanas}) = P(F/B) = \frac{P(F \text{ e } B)}{P(B)} = \frac{0,175}{0,45} = 0,3889$$

Probabilidade condicional

Exercícios resolvidos

1) De 100 pessoas que solicitaram emprego durante o ano passado em uma grande empresa, 60 possuíam experiência anterior e 40 possuíam diploma de Ensino Superior. Vinte e cinco dos candidatos

possuíam tanto experiência anterior como diploma de Ensino Superior.

a) Qual a probabilidade de que um candidato aleatoriamente escolhido tenha experiência ou diploma?

b) Qual a probabilidade de que um candidato aleatoriamente escolhido tenha um diploma, dado que ele tem alguma experiência anterior.

c) Verifique se a posse de diploma e experiência anterior são eventos independentes.

Solução:

$$a) P(E \text{ ou } D) = \frac{60}{100} + \frac{40}{100} - \frac{25}{100} = 0,75$$

$$b) P(D/E) = \frac{P(D \text{ e } E)}{P(E)} = \frac{0,25}{0,60} = 0,4167$$

$$c) P(E \text{ e } D) = P(E) \times P(D) = 0,25 = 0,60 \times 0,40$$

$0,25 \neq 0,24$ Logo, os eventos não são independentes.

2) Duas bolas são retiradas, **sem reposição**, de uma urna que contem duas bolas brancas, três pretas e quatro verdes. Qual a probabilidade de que ambas:

a) sejam verdes?

b) sejam da mesma cor?

Solução:

$$a) P(V \text{ e } V) = \frac{4}{9} \times \frac{3}{8} = 0,1667$$

$$b) P(B \text{ e } B) \text{ ou } P(P \text{ e } P) \text{ ou } P(V \text{ e } V) = \\ = \frac{2}{9} \times \frac{1}{8} + \frac{3}{9} \times \frac{2}{8} + \frac{4}{9} \times \frac{3}{8} = 0,2778$$

3) Se há 0,70 de probabilidade de uma pessoa entrevistada ser a favor do candidato A em uma eleição, qual a probabilidade de:

- a) as duas primeiras pessoas entrevistadas serem contra?
- b) as duas primeiras pessoas entrevistadas serem contra, mas a terceira ser a favor?

Solução:

- a) $P(C \text{ e } C) = 0,3 \cdot 0,3 = 0,09$
- b) $P(C \text{ e } C \text{ e } A) = 0,3 \cdot 0,3 \cdot 0,7 = 0,063$

4) Numa certa cidade, 40% da população é sócia do clube A, 25% é sócia do clube B e 10% são sócias de ambos os clubes. Uma pessoa da cidade é selecionada aleatoriamente.

- a) Qual a probabilidade de ser sócia do clube A ou do clube B?
- b) Qual a probabilidade de não ser sócia do clube A nem do clube B?

Solução:

- a) $P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B) = 0,40 + 0,25 - 0,10 = 0,55$
- b) $P(\text{não ser sócia de A nem B}) = 1 - P(A \text{ ou } B) = 1 - 0,55 = 0,45$

Exercícios complementares

1) Foi realizada uma pesquisa para obter informações quanto ao hábito dos estudantes no uso de cartões bancários (crédito e débito). Dentre os 300 estudantes pesquisados, 80 têm cartão de crédito, 125 têm cartão de débito e 45 têm ambos. Calcule a probabilidade de um estudante escolhido ao acaso:

- a) ter cartão de crédito ou cartão de débito.
- b) não ter cartão de crédito nem de débito.

2) Um baralho é constituído de 52 cartas. Retira-se uma carta ao acaso. Qual a probabilidade de que seja um rei, sabendo que a

carta retirada é do naipe de espadas?

3) A tabela abaixo apresenta os resultados de uma pesquisa realizada com 857 funcionários de uma empresa com relação à área de formação de cada um.

Área	Sexo		Total
	Masculino (M)	Feminino (F)	
Ciências Exatas (E)	265	150	415
Ciências Humanas (H)	150	100	250
Ciências Sociais (S)	134	58	192
Total	549	308	857

- Monte a tabela de probabilidades combinadas e marginais.
- Qual a probabilidade de que um funcionário entrevistado seja da área de Ciências Humanas?
- Qual a probabilidade de que um funcionário entrevistado seja do sexo masculino ou da área de Ciências Sociais?
- Qual a probabilidade de que um funcionário entrevistado seja da área de Ciências Exatas e do sexo feminino?
- Qual a probabilidade de que um funcionário entrevistado seja do sexo masculino, dado que seja da área de Ciências Humanas?
- Qual a probabilidade de que um funcionário entrevistado seja da área de Ciências Exatas ou de Ciências Sociais?

4) A probabilidade de que certo filme obtenha um prêmio de melhor fotografia é 0,35; a probabilidade de que obtenha um prêmio relativo de melhor trilha sonora é 0,27, e a probabilidade de que obtenha ambos é 0,05.

- Qual a probabilidade de ganhar o prêmio de fotografia ou de trilha sonora?
- Qual a probabilidade de que não ganhe prêmio nem pela fotografia nem pela trilha sonora?

5) Um baralho é constituído de 52 cartas. Retirando uma carta ao acaso, calcule as probabilidades dos seguintes eventos:

- a) A carta é de espadas.
- b) A carta é valete ou de copas.
- c) Sabendo que a carta é de ouros, qual a probabilidade que seja uma figura?

6) Uma universidade tem seis mil alunos, dos quais dois mil trabalham. Temos ainda que: 500 alunos são do curso de Administração; 700 do curso de Direito; 100 alunos trabalham e são do curso de Administração; 200 alunos trabalham e são do curso de Direito, e 1700 alunos são de outros cursos e trabalham.

- a) Monte a tabela de contingência.
- b) Monte a tabela de probabilidades combinadas e marginais.
- c) Qual a probabilidade de que um aluno, aleatoriamente selecionado, trabalhe?
- d) Qual a probabilidade de que um aluno, aleatoriamente selecionado, trabalhe e seja aluno do curso de Direito?
- e) Qual a probabilidade de que um aluno, aleatoriamente selecionado, trabalhe ou seja aluno do curso de Direito?
- f) Qual a probabilidade de que um aluno, aleatoriamente selecionado, trabalhe dado que ele é do curso de direito?
- g) Se o aluno é de outros cursos, qual a probabilidade de que ele trabalhe?

7) A tabela abaixo é uma tabela de contingência que descreve o status de promoção de funcionários masculinos e femininos de uma grande empresa. Depois de rever o registro de promoções, um grupo de funcionárias levantou um caso de discriminação com base no fato de que 250 funcionários receberam promoções, mas somente 50 funcionárias foram promovidas. A administração da empresa argumentou que o número relativamente baixo de promoções para as funcionárias não foi devido a discriminação, mas ao fato de que há relativamente

poucas funcionárias na empresa. Mostre como a probabilidade condicional poderia ser usada para analisar a acusação de discriminação.

	Sexo		Totais
	Masculino	Feminino	
Promovidos	250	50	300
Não Promovidos	600	200	800
Totais	850	250	1100

8) Qual a probabilidade de se obter dois ases em seguida, quando se extraem duas cartas de um baralho comum de 52 cartas, se:

- a primeira carta é repostada no baralho antes da extração da segunda carta?
- a primeira carta extraída não é repostada antes da extração da segunda carta?

9) No lançamento de dois dados, qual a probabilidade de:

- Obter o par (1;1)?
- Obter como soma das duas faces um número maior que 7?
- Obter como soma das duas faces um número menor que 5?

10) Retirando-se uma carta de um baralho normal de 52 cartas, qual é a probabilidade de sair um ás ou uma carta de ouros?

11) Sejam $P(A) = 0,30$, $P(B) = 0,8$ e $P(A \text{ e } B) = 0,15$.

- Determine $P(\bar{B})$
- Determine $P(A \text{ ou } B)$

12) Tira-se aleatoriamente uma carta de um baralho de 52 cartas.

- Qual a probabilidade de sair uma carta de número 5?
- Qual a probabilidade de selecionarmos um valete ou uma dama ou um rei?

13) Uma moeda é lançada duas vezes. Calcule a probabilidade de não ocorrer cara nenhuma vez.

14) A tabela a seguir apresenta o sexo e o número de horas de atividade física por semana de cada indivíduo.

Sexo	Tempo de atividade física por semana			Total
	Menos de 2 horas (A)	De 2 a 6 horas (B)	Mais de 6 horas (C)	
Feminino (F)	24	11	9	44
Masculino (M)	16	15	5	36
Total	40	26	14	80

a) Qual é a probabilidade de se sortear aleatoriamente uma pessoa do sexo feminino e que pratique atividade física semanal de 2 a 6 horas?

b) Calcule a probabilidade de selecionar aleatoriamente uma pessoa que pratique mais de 6 horas semanais de atividade física.

c) Se escolhermos aleatoriamente uma pessoa que dedica mais de 6 horas à atividade física, qual a probabilidade de que ela seja do sexo masculino?

d) Qual a probabilidade de escolhermos aleatoriamente uma pessoa do sexo feminino e que dedique menos de 2 horas de atividade física por semana?

15) Um time de futebol consiste de oito jogadores naturais do local e três naturais de outras cidades. Se dois dos jogadores são selecionados aleatoriamente para uma entrevista, qual é a probabilidade de serem ambos naturais de outra cidade?

16) Num lote de 12 peças, 4 são defeituosas. Três peças são retiradas aleatoriamente, uma após a outra, sem reposição. Encontre a probabilidade de todas essas três peças serem não-defeituosas.

17) Um baralho é constituído de 52 cartas. Retirando uma carta ao acaso deste baralho, calcule as probabilidades dos seguintes eventos:

- a) A carta é do naipe de copas.
- b) A carta é valete ou do naipe de copas.
- c) A carta é uma dama ou um valete.
- d) A carta é do naipe de espadas.
- e) A carta não é do naipe de espadas.

18) Você coleta uma amostra de 60 alunos do curso de Psicologia e registra a nota de cada um na disciplina de Estatística. Se selecionar ao acaso um aluno, qual será a probabilidade de a nota dele estar entre 6 (inclusive) e 8 (exclusive)?

Notas	Freqüência
0 – 2	3
2 – 4	5
4 – 6	11
6 – 8	25
8 – 10	16
	Σ 60

19) O serviço meteorológico informa que, para o fim de semana, a probabilidade de chover é de 60%, a de fazer frio é de 70% e a de chover e fazer frio é de 50%. Calcule a probabilidade de que, no fim de semana:

- a) chova ou faça frio;
- b) não chova e não faça frio.

20) Um levantamento entre 50 estudantes do curso de Administração, modalidade a distância, sobre o número de atividades extracurriculares, resultou nos dados a seguir:

Número de atividades extracurriculares	Frequência
0	8
1	20
2	12
3	6
4	3
5	1
	Σ 50

- a) Seja A o evento em que um estudante participa de pelo menos uma atividade. Ache a $P(A)$.
- b) Seja B o evento em que um estudante participa de três atividades ou mais. Ache $P(B)$.
- c) Qual é a probabilidade de que o estudante participe exatamente de duas atividades?

Respostas

1. a) $P(C \text{ ou } D) = P(C) + P(D) - P(C \text{ e } D) = 0,5333$
 b) $P(\text{não tem crédito, nem débito}) = 1 - 0,5333 = 0,4667$

$$2. P(\text{rei/espadas}) = \frac{P(\text{rei e espadas})}{P(\text{espadas})} = \frac{\frac{1}{52}}{\frac{13}{52}} = \frac{1}{13} = 0,0769$$

3. a)

Área	Sexo		Probabilidade marginal
	Masculino (M)	Feminino (F)	
Ciências Exatas (E)	$P(E \text{ e } M)=0,3092$	$P(E \text{ e } F)=0,1750$	$P(E)=0,4842$
Ciências Humanas (H)	$P(H \text{ e } M)=0,1750$	$P(H \text{ e } F)=0,1167$	$P(H)=0,2917$
Ciências Sociais (S)	$P(S \text{ e } M)=0,1564$	$P(S \text{ e } F)=0,0677$	$P(S)=0,2240$
Probabilidade marginal	$P(M)=0,6406$	$P(F)=0,3594$	1

b) $P(H) = 0,2917$

c) $P(M \text{ ou } S) = 0,6406 + 0,2240 - 0,1564 = 0,7082$

d) $P(E \text{ e } F) = 0,1750$

e) $P(M/H) = \frac{P(M \text{ e } H)}{P(H)} = \frac{0,175}{0,2917} = 0,6$

f) $P(E \text{ ou } S) = 0,4842 + 0,2240 = 0,7082$

4. a) $P(F \text{ ou } T) = P(F) + P(T) - P(F \text{ e } T) = 0,35 + 0,27 - 0,05 = 0,57$

b) $P(\text{não ganhar nenhum prêmio}) = 1 - P(F \text{ ou } T) = 1 - 0,57 = 0,43$

5. a) $P(E) = \frac{13}{52} = 0,25$

b) $P(V \text{ ou } C) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = 0,3077$

c) $P(F/O) = \frac{P(F \text{ e } O)}{P(O)} = \frac{3}{13} = 0,2308$

6. a) Tabela de contingência

Situação	Cursos			Total
	Administração (A)	Direito (D)	Outros (O)	
Trabalham (T)	100	200	1700	2000
Não trabalham (NT)	400	500	3100	4000
Total	500	700	4800	6000

b) Tabela de probabilidades combinadas e marginais

Situação	Cursos			Probabilidade marginal
	Administração (A)	Direito (D)	Outros (O)	
Trabalham (T)	$P(T \text{ e } A) = 0,0167$	$P(T \text{ e } D) = 0,0333$	$P(T \text{ e } O) = 0,2833$	$P(T) = 0,3333$
Não trabalham (NT)	$P(NT \text{ e } A) = 0,0667$	$P(NT \text{ e } D) = 0,0833$	$P(NT \text{ e } O) = 0,5167$	$P(NT) = 0,6667$
Prob. marginal	$P(A) = 0,0833$	$P(D) = 0,1167$	$P(O) = 0,8000$	1

c) $P(T) = 0,3333$

d) $P(T \text{ e } D) = 0,0333$

$$\begin{aligned} \text{e) } P(T \text{ ou } D) &= P(T) + P(D) - P(T \text{ e } D) = \\ &= 0,3333 + 0,1167 - 0,0333 = 0,4167 \end{aligned}$$

$$\text{f) } P(T/D) = \frac{P(T \text{ e } D)}{P(D)} = \frac{0,0333}{0,1167} = 0,2853$$

$$\text{g) } P(T/O) = \frac{P(T \text{ e } O)}{P(O)} = \frac{0,2833}{0,8} = 0,3541$$

$$\text{7.a) } P(P/F) = \frac{P(P \text{ e } F)}{P(F)} = \frac{\frac{50}{1100}}{\frac{250}{1100}} = 0,2$$

$$\text{b) } P(P/M) = \frac{P(P \text{ e } M)}{P(M)} = \frac{\frac{250}{1100}}{\frac{850}{1100}} = 0,2941$$

Pelo resultado obtido no cálculo dos itens a e b, podemos verificar que a probabilidade de promoção é maior para um funcionário do sexo masculino do que para uma funcionária.

$$\text{8. a) } P(\text{ás e ás}) = \frac{4}{52} \times \frac{4}{52} = 0,0059$$

$$\text{b) } P(\text{ás e ás}) = \frac{4}{52} \times \frac{3}{51} = 0,0045$$

9. Espaço amostral: $n(S) = 36$

(1;1)	(2;1)	(3;1)	(4;1)	(5;1)	(6;1)
(1;2)	(2;2)	(3;2)	(4;2)	(5;2)	(6;2)
(1;3)	(2;3)	(3;3)	(4;3)	(5;3)	(6;3)
(1;4)	(2;4)	(3;4)	(4;4)	(5;4)	(6;4)
(1;5)	(2;5)	(3;5)	(4;5)	(5;5)	(6;5)
(1;6)	(2;6)	(3;6)	(4;6)	(5;6)	(6;6)

a) Evento (1 e 1), tamanho do evento $n(A) = 1$,

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{36} = 0,0278$$

b) Evento a soma das duas faces ser maior que 7, tamanho do evento $n(B) = 15$, $P(B) = \frac{n(B)}{n(S)} = \frac{15}{36} = 0,4167$

c) Evento a soma das duas faces ser menor que 5, tamanho do evento $n(C) = 6$, $P(C) = \frac{n(C)}{n(S)} = \frac{6}{36} = 0,1667$

$$10. P(A \text{ ou } B) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = 0,3077$$

$$11.a) P(\bar{B}) = 1 - P(B) = 1 - 0,8 = 0,2$$

$$b) P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B) = 0,3 + 0,8 - 0,15 = 0,95$$

$$12.a) P(4) = \frac{4}{52} = 0,0769$$

$$b) P(\text{valetes ou dama ou rei}) = \frac{4}{52} + \frac{4}{52} + \frac{4}{52} = \frac{12}{52} = 0,2308$$

13. Espaço amostral:

$$S = \{ (Ca, Ca) ; (Co, Co); (Ca, Co); (Co, Ca) \} \quad n(S) = 4$$

$$\text{Evento: } E = \{(Co, Co)\} \quad n(A) = 1 \quad P(A) = \frac{n(A)}{n(S)} = \frac{1}{4} = 0,25$$

$$14.a) P(F \text{ e } B) = 0,1375$$

$$b) P(C) = 0,1750$$

$$c) P(M/C) = 0,3571$$

$$d) P(F \text{ e } A) = 0,3$$

$$15. P(J1 \text{ e } J2) = P(J1) \times P(J2/J1) = \frac{3}{11} \times \frac{2}{10} = \frac{6}{110} = 0,0545$$

$$16. P(\text{ND e ND e ND}) = \frac{8}{12} \times \frac{7}{11} \times \frac{6}{10} = 0,2545$$

$$17. \text{ a) } P(\text{copas}) = \frac{13}{52} = 0,25$$

$$\text{ b) } P(\text{valetes ou copas}) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = 0,3077$$

$$\text{ c) } P(\text{dama ou valetes}) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = 0,1538$$

$$\text{ d) } P(\text{espadas}) = \frac{13}{52} = 0,25$$

$$\text{ e) } P(\text{não ser espadas}) = \frac{39}{52} = 0,75$$

$$18. P(6 \text{ ou } 8) = \frac{25}{60} = 0,4167$$

$$19. \text{ a) } P(C \text{ ou } F) = P(C) + P(F) - P(C \text{ e } F)$$

$$P(C \text{ ou } F) = 0,60 + 0,70 - 0,50 = 0,80$$

$$\text{ b) } P(\text{não chova e não faça frio}) = 1 - 0,80 = 0,20$$

$$20. \text{ a) } P(A) = 0,84$$

$$\text{ b) } P(B) = 0,20$$

$$\text{ c) } P(2) = 0,24$$

6 Variáveis aleatórias

Uma **variável aleatória** é uma função que associa um valor numérico a cada ponto do espaço amostral de um experimento. As variáveis aleatórias podem ser discretas ou contínuas. Uma **variável aleatória discreta** tem um número finito de valores que podem ser associados a um processo de contagem, como o número de alunos em sala de aula. Uma **variável aleatória contínua** tem infinitos valores, e esses valores podem ser associados com medidas em uma escala contínua, como o tempo, o peso ou a altura.

No estudo de variáveis aleatórias, em geral, estamos interessados na probabilidade de cada valor que a variável aleatória assume, ou seja, na sua **distribuição de probabilidade** que pode ser representada por uma tabela ou gráfico.

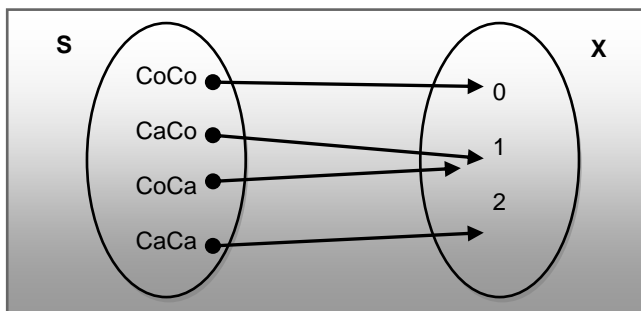
Para melhor compreensão do conceito de variável aleatória, apresentaremos com mais detalhes as variáveis aleatórias discretas.

6.1 Variável aleatória discreta

Considere o experimento de lançar uma moeda duas vezes. O espaço amostral correspondente a este experimento é:

$S = \{CaCa, CaCo, CoCa, CoCo\}$; “Ca” representa cara e “Co” coroa

Seja X a variável aleatória discreta que conta o número de caras que ocorrem no lançamento de uma moeda duas vezes.



Podemos montar uma distribuição de frequência para o número de eventos que corresponde à ocorrência de nenhuma, uma ou duas caras respectivamente:

X	Frequência (f_i)
0	1
1	2
2	1
Total	4

Podemos também associar a cada um dos valores assumidos pela variável aleatória X as probabilidades correspondentes:

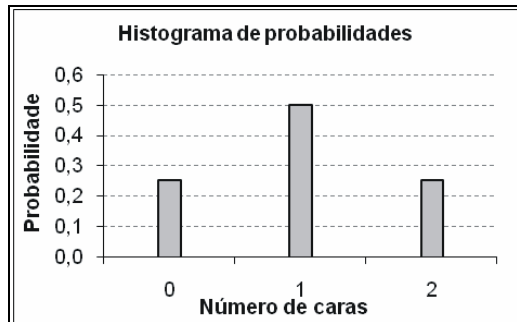
$$P(X = 0) = 1/4 = 0,25$$

$$P(X = 1) = 2/4 = 0,50$$

$$P(X = 2) = 1/4 = 0,25$$

A **distribuição de probabilidade** da variável aleatória X pode ser representada por uma tabela ou pelo histograma de probabilidades.

Distribuição de probabilidade	
X	P(X)
0	0,25
1	0,50
2	0,25
Total	1,0



A distribuição de probabilidade de uma variável aleatória discreta deve satisfazer as seguintes condições:

- A probabilidade de cada valor da variável aleatória deve estar entre 0 e 1:

$$0 \leq P(x) \leq 1$$

- A soma de todas as probabilidades é igual a:

$$\sum P(X) = 1$$

6.1.1 Média ou valor esperado

É possível determinar a tendência central de uma distribuição de probabilidade pela sua média, representada por μ ou $E(x)$. A média de uma distribuição de probabilidade é o valor esperado de sua variável aleatória. A média é obtida multiplicando-se cada valor da variável aleatória x por sua correspondente probabilidade $P(x)$, e somando-se os produtos resultantes:

$$\mu = \sum [x_i \cdot P(x_i)]$$

O valor esperado é o valor médio que esperaríamos se o experimento continuasse sendo repetido várias vezes.

Exemplo: Determine o valor esperado de solicitações de empréstimos empresariais aprovadas por semana. A variável aleatória x representa o número de empréstimos empresariais aprovados por semana.

x	P(x)
0	0,10
1	0,10
2	0,20
3	0,30
4	0,15
5	0,10
6	0,05
Total	1,0

Solução: Para calcular o número esperado aplica-se a seguinte fórmula:

$$\mu = \sum [x_i \cdot P(x_i)] \quad \text{onde:}$$

$$\sum [x_i \cdot P(x_i)] = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + \dots + x_n \cdot P(x_n)$$

Podemos utilizar a tabela para os cálculos de $x \cdot P(x)$:

x	P(x)	$x \cdot P(x)$
0	0,10	0
1	0,10	0,10
2	0,20	0,40
3	0,30	0,90
4	0,15	0,60
5	0,10	0,50
6	0,05	0,30
Total	1,0	$\mu = 2,8$

Ou podemos substituir os valores da variável aleatória x e das probabilidades correspondentes na própria fórmula:

$$\begin{aligned} \mu &= \sum [x \cdot P(x)] = \\ &= 0 \cdot 0,10 + 1 \cdot 0,10 + 2 \cdot 0,20 + 3 \cdot 0,30 + 4 \cdot 0,15 + 5 \cdot 0,10 + 6 \cdot 0,05 = \\ &= 0 + 0,10 + 0,40 + 0,90 + 0,60 + 0,50 + 0,30 = 2,8 \end{aligned}$$

Interpretação: O valor esperado ou média do número de empréstimos aprovados por semana é 2,8, ou seja, espera-se que em média sejam aprovados 2,8 empréstimos por semana.

6.1.2 Variância e desvio-padrão

A variabilidade de uma distribuição de probabilidade pode ser determinada pela variância ou desvio-padrão.

$$\text{Variância} \rightarrow \sigma^2 = \sum [x^2 \cdot P(x)] - \mu^2$$

$$\text{Desvio-padrão} \rightarrow \sigma = \sqrt{\sigma^2}$$

Exemplo: Calcule a variância e o desvio-padrão da distribuição de probabilidade de solicitações de empréstimos empresariais aprovados por semana do exemplo anterior.

Solução: Utilizando a tabela para realizar os cálculos, temos:

x	P(x)	x ²	x ² · P(x)
0	0,10	0	0
1	0,10	1	0,10
2	0,20	4	0,80
3	0,30	9	2,70
4	0,15	16	2,40
5	0,10	25	2,50
6	0,05	36	1,80
Total	1,0		10,3

$$\sigma^2 = \sum [x^2 \cdot P(x)] - \mu^2 = 10,3 - 2,8^2 = 10,3 - 7,84 = 2,46$$

O desvio-padrão é a raiz quadrada da variância, assim:

$$\sigma = \sqrt{\sigma^2} = \sqrt{2,46} = 1,5684$$

Interpretação: O desvio-padrão de solicitações de empréstimos empresariais aprovadas por semana é de 1,5684.

6.1.3 Exemplo de aplicação

Suponha que você está indeciso quanto ao melhor investimento entre duas opções. A tabela abaixo apresenta as estimativas dos retornos para cada investimento (em R\$1.000,00), sob três condições econômicas com as respectivas probabilidades.

Condição econômica	Probabilidade	Investimento A	Investimento B
Recessão	0,25	+200	-100
Economia estável	0,45	+50	+100
Economia em expansão	0,30	-100	+250

a) Para verificar qual investimento maximiza o lucro esperado, calcule a média ou o valor esperado para cada um dos investimentos.

b) Para verificar qual dos investimentos minimiza o risco ou a incerteza, calcule o desvio-padrão para os dois investimentos.

Solução do item a)

Investimento A

$$\begin{aligned}\mu = E(x) &= \sum [x \cdot P(x)] = 200 \cdot 0,25 + 50 \cdot 0,45 + (-100) \cdot 0,30 = \\ &= 50 + 22,5 - 30 = 42,5\end{aligned}$$

O valor esperado para o investimento A é de R\$42.500,00.

Investimento B

$$\begin{aligned}\mu = E(x) &= \sum [x \cdot P(x)] = -100 \cdot 0,25 + 100 \cdot 0,45 + 250 \cdot 0,30 = \\ &= -25 + 45 + 75 = 95\end{aligned}$$

O valor esperado para o investimento B é de R\$95.000,00.

Interpretação do item a: O investimento B maximiza o lucro esperado, pois possui o maior valor esperado.

Solução do item b) Investimento A

$$\begin{aligned}\sigma^2 &= \sum [x^2 \cdot P(x)] - \mu^2 = \\ &= [200^2 \cdot 0,25 + 50^2 \cdot 0,45 + (-100)^2 \cdot 0,30] - 42,5^2 = \\ &= [10.000 + 1.125 + 3.000] - 1.806,25 = 14.125 - 1.806,25 \\ \sigma^2 &= 12.318,75 \quad \rightarrow \quad \sigma = \sqrt{12.318,75} = 110,99\end{aligned}$$

Investimento B

$$\begin{aligned}\sigma^2 &= \sum [x^2 \cdot P(x)] - \mu^2 = \\ &= [(-100)^2 \cdot 0,25 + 100^2 \cdot 0,45 + 250^2 \cdot 0,30] - 95^2 = \\ &= [2.500 + 4.500 + 18.750] - 9.025 = 25.750 - 9.025 \\ \sigma^2 &= 16.725 \quad \rightarrow \quad \sigma = \sqrt{16.725} = 129,33\end{aligned}$$

Interpretação do item b: O investimento A minimiza o risco ou a incerteza, pois possui o menor desvio-padrão.

Exercícios resolvidos

1) Utilizando os registros do departamento de trânsito de uma pequena cidade, foi feita uma relação do número de acidentes por dia nos últimos 200 dias.

Número de acidentes por dia (x)	f _i
0	20
1	40
2	90
3	30
4	10
5	10

a) Monte uma distribuição de probabilidade para a variável aleatória x número de acidentes por dia. Faça o gráfico da distribuição de probabilidade.

b) Obtenha o número esperado de acidentes por dia.

c) Qual a probabilidade de ter ocorrido três acidentes por dia?

d) Qual a probabilidade de ter ocorrido no máximo dois acidentes por dia?

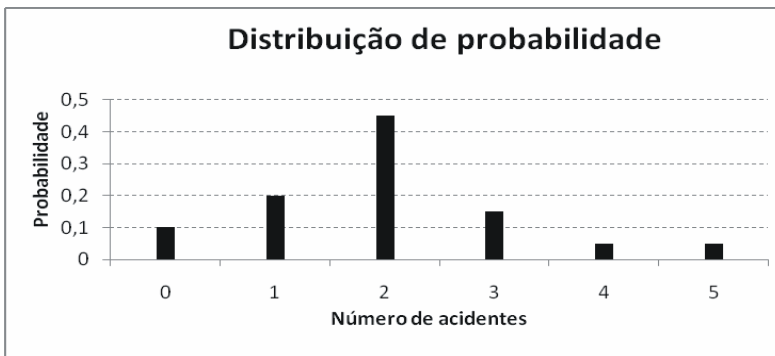
Solução do item a):

Aproveitando a tabela para calcular as probabilidades:

Número de acidentes por dia (X)	f _i	P(x)
0	20	20/200 = 0,1
1	40	40/200 = 0,2
2	90	90/200 = 0,45
3	30	30/200 = 0,15
4	10	10/200 = 0,05
5	10	10/200 = 0,05
Total ou Σf_i	200	1,0

$\Sigma P(x) = 1$

Distribuição de probabilidade	
Número de acidentes por dia (x)	P(x)
0	0,1
1	0,2
2	0,45
3	0,15
4	0,05
5	0,05
Total	1,0



Solução do item b):

Podemos utilizar a própria tabela para calcular o valor de $x \cdot p(x)$:

Distribuição de probabilidade		
Número de acidentes por dia (x)	P(x)	$x \cdot P(x)$
0	0,1	0
1	0,2	0,2
2	0,45	0,9
3	0,15	0,45
4	0,05	0,2
5	0,05	0,25
Total	1,0	2,0

Também podemos obter o número esperado de acidentes por dia substituindo os valores diretamente na fórmula:

$$\begin{aligned} \mu &= \sum [x \cdot P(x)] = \\ &= 0 \cdot 0,1 + 1 \cdot 0,2 + 2 \cdot 0,45 + 3 \cdot 0,15 + 4 \cdot 0,05 + 5 \cdot 0,05 = \\ &= 0 + 0,2 + 0,9 + 0,45 + 0,2 + 0,25 = 2,0 \end{aligned}$$

Interpretação: Em média ocorrem dois acidentes por dia nessa cidade.

Solução do item c): A probabilidade de terem ocorrido três acidentes por dia é:

$$P(x=3) = 0,15 \times 100 = \mathbf{15\%}$$

Solução do item d): A probabilidade de terem ocorrido no máximo dois acidentes por dia é:

$$P(x=0) + P(x=1) + P(x=2) = 0,1 + 0,2 + 0,45 = 0,75 \times 100 = \mathbf{75\%}$$

Exercícios complementares

1) Um jogador lança três moedas não viciadas. Ganha R\$6,00 se ocorrerem três caras; R\$3,00 se ocorrerem duas caras; R\$1,00 se ocorrer somente uma cara. Por outro lado, perde R\$10,00 se ocorrerem três coroas. Encontre o valor esperado do jogo e o desvio-padrão.

Obs.: A variável aleatória X conta o valor do jogo e as probabilidades são dadas pela probabilidade de ocorrer cada um dos eventos (ex.: probabilidade de ocorrerem três caras é 1/8).

	3 caras	2 caras	1 caras	3 coroas
Valor do jogo (X)	+ R\$ 6,00	+ R\$ 3,00	+ R\$ 1,00	- R\$ 10,00
P(x)	1/8	3/8	3/8	1/8

2) Um produto deve ser lançado no mercado no próximo ano. A expectativa do departamento de marketing de que o lançamento seja bem-sucedido é de 80%. Nesse caso, o lucro será de R\$60.000,00. Se isso não acontecer, o prejuízo deve chegar a R\$20.000,00. Calcule o valor esperado no lançamento deste produto. Monte a distribuição de probabilidade.

3) As probabilidades de um fiscal de construções observar de 0 a 5 infrações do código de obras em um prédio estão representadas na tabela abaixo. Determine a média e o desvio-padrão dessa distribuição.

Número de infrações	0	1	2	3	4	5
p(x)	0,48	0,25	0,14	0,08	0,04	0,01

4) Uma empresa pretende expandir sua produção com um novo tipo de produto. Como a empresa tem dúvidas quanto à demanda do novo produto, não sabe se a expansão deve se dar em média ou grande escala. Na tabela abaixo são apresentadas as estimativas de lucro anual em R\$1.000,00 com as respectivas probabilidades.

Demanda	Lucros de expansão de média escala		Lucros de expansão de grande escala	
	x	p(x)	y	P(y)
Alta	25	0,20	0	0,20
Média	75	0,50	50	0,50
Baixa	100	0,30	150	0,30

a) Calcule o valor esperado para o lucro associado às duas alternativas de expansão. Que decisão é preferida para o objetivo de maximizar o lucro esperado?

b) Calcule a variância para o lucro associado às duas alternativas de expansão. Que decisão é preferida para o objetivo de minimizar o risco ou a incerteza?

Respostas

1. $\mu = R\$1,00$ (a favor do jogador) $\sigma = R\$4,44$

2. valor esperado = $\mu = R\$44.000,00$

3. $\mu = 0,98$ e $\sigma = 1,21$

4.a) $\mu = 72,5$ (média escala) e $\mu = 70$ (grande escala). Se o objetivo é maximizar o lucro, a expansão de média escala é mais adequada, pois apresenta o maior valor esperado.

b) $\sigma^2 = 681,25$ (média escala) e $\sigma^2 = 3.100$ (grande escala). Se o objetivo é minimizar o risco ou a incerteza, a expansão de média escala é mais indicada, pois apresenta menor variabilidade.

7 Distribuição de probabilidade para variáveis aleatórias discretas

Muitas das distribuições de probabilidade surgiram de variáveis que foram estudadas em problemas práticos ou teóricos. Dentre as diferentes distribuições discretas de probabilidade, trataremos aqui apenas de duas: **distribuição binomial** e **distribuição de Poisson**.

7.1 Distribuição binomial

Há muitos experimentos para os quais a conclusão de cada tentativa pode ser reduzida a dois resultados: sucesso ou fracasso. Quando lançamos uma moeda, pode ocorrer cara ou coroa; no nascimento de uma criança, menino ou menina. Experimentos probabilísticos como esses são chamados binomiais.

Um experimento binomial tem as seguintes características:

- as n tentativas de um mesmo experimento são independentes;
- cada tentativa admite apenas dois resultados: sucesso ou fracasso;
- a probabilidade de sucesso (p) em cada tentativa é constante.

Portanto, se em n tentativas independentes de um experimento o resultado só puder ser um sucesso (p) ou um fracasso (q), e a probabilidade de sucesso for constante ao longo do experimento, a probabilidade de a variável aleatória x ter k = número de sucessos nas n tentativas será obtida da seguinte forma:

$$P(x = k) = C_{n,k} \cdot p^k \cdot q^{n-k} \quad \text{com} \quad C_{n,k} = \frac{n!}{k!(n-k)!}$$

Para saber mais sobre combinação, veja o Anexo II

Onde:

$k \rightarrow$ o número de sucessos, k , variando de 0, 1, 2, ..., n ;

$n \rightarrow$ o número de tentativas ou observações;

$p \rightarrow$ probabilidade de sucesso em cada tentativa;

$q \rightarrow$ probabilidade de fracasso em cada tentativa.

Exemplo 1: Determine a probabilidade de ocorrer exatamente duas caras nos três lançamentos de uma mesma moeda.

Solução: Podemos resolver esse problema usando o conceito clássico de probabilidade. Para tal, é necessário conhecer o espaço amostral e o evento:

$S = \{\text{CaCaCa}, \text{CaCaCo}, \text{CaCoCa}, \text{CaCoCo}, \text{CoCaCa}, \text{CoCaCo}, \text{CoCoCa}, \text{CoCoCo}\}$

$A = \{\text{CaCaCo}, \text{CaCoCa}, \text{CoCaCa}\}$

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{8} = 0,375$$

O experimento de lançar três moedas é um experimento binomial, pois o resultado de um lançamento não altera o resultado do próximo lançamento, tentativas independentes; há apenas dois resultados possíveis: cara (sucesso) ou coroa (fracasso); a probabilidade de sucesso (cara) é constante ao longo do experimento. Logo, podemos resolver esse problema usando a distribuição binomial:

A variável aleatória x conta o número de caras: $x =$ número de caras.

Serão realizadas três tentativas: $n = 3$

A probabilidade de ocorrer cara (sucesso) em cada tentativa é de 0,5: $p=0,5$. A probabilidade de **não** ocorrer cara (fracasso) é de 0,5: $q=0,5$.

O número de sucessos é $k=2$. Substituindo na fórmula da distribuição binomial, temos:

$$P(x = k) = C_{n,k} \cdot p^k \cdot q^{n-k}$$

$$P(x = 2) = C_{3,2} \cdot 0,5^2 \cdot 0,5^{3-2} = \frac{3!}{2! (3-2)!} \cdot 0,5^2 \cdot 0,5^1 = 3 \cdot 0,25 \cdot 0,5 = 0,375$$

Multiplicando por 100, teremos o resultado em porcentagem:
 $0,375 \times 100 = 37,5\%$.

Interpretação: A probabilidade de ocorrerem exatamente duas caras no lançamento de três moedas é de 0,375, ou seja, há 37,5% de chance de isso ocorrer.

Exemplo 2: Suponha que no exemplo acima fossem lançadas dez moedas. Qual a probabilidade de ocorrerem exatamente duas caras?

Solução: Nesse caso, fica muito trabalhosa a solução pelo conceito clássico de probabilidade. Portanto, conhecendo as características de um experimento, podemos classificá-lo entre as diferentes distribuições ou modelos de probabilidade facilitando sua resolução.

$$x = \text{número de caras}; \quad n = 10; \quad p = 0,5; \quad q = 0,5; \quad k = 2$$

$$P(x = k) = C_{n,k} \cdot p^k \cdot q^{n-k}$$

$$\begin{aligned} P(x = 2) &= C_{10,2} \cdot 0,5^2 \cdot 0,5^{10-2} = \frac{10!}{2! (10-2)!} \cdot 0,5^2 \cdot 0,5^8 = \\ &= 45 \cdot 0,25 \cdot 0,0039 = 0,0439 \end{aligned}$$

Multiplicando por 100, teremos o resultado em porcentagem:
 $0,0439 \times 100 = 4,39\%$

$$\mathbf{P(x = 2) = 4,39\%}$$

Exemplo 3: Numa empresa, 40% dos contratos são pagos em dia. Qual a probabilidade de que, entre 12 contratos, três ou menos sejam pagos em dia?

Solução: X: número de contatos pagos em dia

$$n = 12, \quad p = 0,4, \quad q = 0,6$$

$$P(x \leq 3) = P(3 \text{ ou menos}) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$$

Precisamos calcular a probabilidade para cada um dos valores acima:

$$P(x=0) = C_{12,0} \cdot 0,4^0 \cdot 0,6^{12-0} = \\ = \frac{12!}{0! (12-0)!} \cdot 0,4^0 \cdot 0,6^{12} = 1 \cdot 1 \cdot 0,00218 = 0,0022$$

$$P(x=1) = C_{12,1} \cdot 0,4^1 \cdot 0,6^{12-1} = \\ = \frac{12!}{1! (12-1)!} \cdot 0,4^1 \cdot 0,6^{11} = 12 \cdot 0,4 \cdot 0,0036 = 0,0173$$

$$P(x=2) = C_{12,2} \cdot 0,4^2 \cdot 0,6^{12-2} = \\ = \frac{12!}{2! (12-2)!} \cdot 0,4^2 \cdot 0,6^{10} = 66 \cdot 0,16 \cdot 0,00605 = 0,0639$$

$$P(x=3) = C_{12,3} \cdot 0,4^3 \cdot 0,6^{12-3} = \\ = \frac{12!}{3! (12-3)!} \cdot 0,4^3 \cdot 0,6^9 = 220 \cdot 0,064 \cdot 0,01008 = 0,1419$$

$$P(x \leq 3) = P(3 \text{ ou menos}) = P(x=0) + P(x=1) + P(x=2) + P(x=3) \\ = 0,0022 + 0,0173 + 0,0639 + 0,1419 = 0,2253$$

$$\mathbf{P(x \leq 3) = P(3 \text{ ou menos}) = 0,2253}$$

7.1.1 Parâmetros da distribuição binomial

As propriedades de uma distribuição binomial permitem o cálculo de seus parâmetros por fórmulas mais simples.

Média: $\mu = E(X) = n \cdot p$

Variância: $\sigma^2 = n \cdot p \cdot q$

Desvio-padrão: $\sigma = \sqrt{n \cdot p \cdot q}$

Exemplo: Sabe-se que a eficiência de uma vacina é de 70%. Um grupo de 20 indivíduos é sorteado, dentre a população vacinada, e submetido a testes para averiguar se a imunização foi efetiva. Seja a variável X: o número de indivíduos imunizados nesse grupo. Calcule o número esperado de imunizados e o desvio-padrão.

Solução:

X: conta o número de indivíduos imunizados

$$n = 20 \quad p = 0,7 \quad q = 0,3$$

$$\mu = n.p = 20 \cdot 0,7 = 14$$

$$\sigma = \sqrt{n.p.q} = \sqrt{20 \cdot 0,7 \cdot 0,3} = \sqrt{4,2} = 2,05$$

Interpretação: Espera-se que em média sejam imunizados 14 indivíduos, com desvio-padrão de 2,05.

7.2 Distribuição de Poisson

A distribuição de Poisson é usada para descrever a probabilidade do número de ocorrências num intervalo contínuo (tempo ou espaço), como número de pessoas em uma fila, defeitos por metro em determinado tecido, chamadas telefônicas recebidas em uma hora de expediente. A unidade de medida (tempo ou espaço) é uma variável contínua, mas a variável aleatória (número de ocorrências) é discreta. Além disso, as falhas não são contáveis – por exemplo, não podemos determinar a probabilidade do número de carros (n) que deixaram de passar num cruzamento. Na distribuição de Poisson, os valores possíveis da variável aleatória x são: 0, 1, 2, ..., sem qualquer limite superior.

Características de uma distribuição de Poisson:

- as ocorrências são independentes;
- a variável aleatória x é o número de ocorrências de um evento ao longo de um intervalo (tempo ou espaço);
- as ocorrências são aleatórias.

A função de probabilidade da variável x será obtida pela relação:

$$P(x = k) = \frac{e^{-\mu} \cdot \mu^k}{k!} \quad \text{onde } \mu = \lambda \cdot t$$

λ = coeficiente de proporcionalidade e t = tempo ou espaço

7.2.1 Parâmetros da distribuição de Poisson

Média: $E(X) = \mu$

Variância: $\sigma^2 = \mu$

Desvio-padrão: $\sigma = \sqrt{\mu}$

Exemplo 1: As chamadas telefônicas chegam a uma delegacia de polícia à razão de 8 por hora em dias úteis.

a) Quantas chamadas de emergência são esperadas num período de 15 minutos?

b) Qual a probabilidade de nenhuma chamada num período de 15 minutos?

c) Qual a probabilidade de ocorrer pelo menos duas chamadas no período de 15 minutos?

d) Qual a probabilidade de ocorrer exatamente duas chamadas em vinte minutos?

Solução do item a)

x: número de chamadas

$$\lambda = \frac{8 \text{ chamadas}}{1 \text{ hora}} \quad t = 15 \text{ minutos}$$

Passando o tempo para minutos em λ também:

$$\mu = \lambda \cdot t = \frac{8 \text{ chamadas}}{60 \text{ minutos}} \cdot 15 \text{ minutos} = 2 \text{ chamadas} \rightarrow \mu = 2 \text{ chamadas}$$

Solução do item b)

$$P(x = 0) = \frac{e^{-2} \cdot 2^0}{0!} = 0,1353 \rightarrow P(x = 0) = 0,1353$$

Interpretação: A probabilidade de a delegacia não receber chamadas num período de 15 minutos é de 0,1353, ou seja, 13,53%.

Solução do item c): pelo menos duas chamadas é o mesmo que no mínimo duas chamadas, ou seja, duas ou três ou quatro ou ...

$$P(x \geq 2) = P(x=2) + P(x=3) + P(x=4) + \dots$$

Como a distribuição de Poisson não tem limite superior, fica impossível calcular dessa forma. Nesse caso, a solução é pelo complementar, ou seja, calculamos o que não nos interessa e diminuímos da soma das probabilidades que é 1:

$$P(x=0) + P(x=1) + \underbrace{P(x=2) + P(x=3) + P(x=4) + \dots}_{P(x \geq 2)} = 1; \text{ logo:}$$

$$\text{Substituindo, temos: } P(x=0) + P(x=1) + P(x \geq 2) = 1$$

Isolando a probabilidade de x igual ou maior que 2, temos:

$$P(x \geq 2) = 1 - [P(x=0) + P(x=1)]$$

Como o período de tempo é o mesmo, a média continua sendo $\mu = 2$ chamadas:

$$P(x=0) = \frac{e^{-2} \cdot 2^0}{0!} = 0,1353 \qquad P(x=1) = \frac{e^{-2} \cdot 2^1}{1!} = 0,2706$$

$$P(x \geq 2) = 1 - [P(x=0) + P(x=1)] = 1 - [0,1353 + 0,2706] \\ = 1 - 0,4059 = 0,5941 \qquad \mathbf{P(x \geq 2) = 0,5941}$$

Interpretação: A probabilidade de a delegacia receber ao menos duas chamadas num período de 15 minutos é de 0,5941, ou seja, 59,41%.

Solução do item d): como o período mudou, é necessário recalcular a média: $\lambda = \frac{8 \text{ chamadas}}{60 \text{ min.}} \quad t=20 \text{ min}$

$$\mu = \lambda \cdot t = \frac{8 \text{ chamadas}}{60 \text{ min}} \cdot 20 \text{ min} = 2,67 \text{ chamadas}$$

$$P(x=2) = \frac{e^{-2,67} \cdot 2,67^2}{2!} = 0,2468$$

Interpretação: A probabilidade de a delegacia receber exatamente duas chamadas num período de 20 minutos é de 24,68%.

Exemplo 2: Suponha que 150 erros de impressão são distribuídos aleatoriamente em um livro de 200 páginas. Encontre a probabilidade de que duas páginas contenham:

- a) nenhum erro de impressão;
- b) três erros de impressão;
- c) um ou mais erros de impressão.

Solução do item a):

$$x: \text{número de erros, } \lambda = \frac{150 \text{ erros}}{200 \text{ páginas}} \quad t = 2 \text{ páginas}$$

$$\mu = \lambda \cdot t = \frac{150 \text{ erros}}{200 \text{ páginas}} \cdot 2 \text{ páginas} = 1,5 \text{ erro} \rightarrow \mu = 1,5 \text{ erro}$$

$$P(x = 0) = \frac{e^{-1,5} \cdot 1,5^0}{0!} = 0,2231$$

Interpretação: A probabilidade de que duas páginas não contenham nenhum erro de impressão é de 0,2231, ou seja, 22,31%.

Solução do item b):

$$P(x = 3) = \frac{e^{-1,5} \cdot 1,5^3}{3!} = \frac{0,2231 \cdot 3,375}{6} = \frac{0,7530}{6} = 0,1255$$

$$P(x=3) = 0,1255$$

Interpretação: A probabilidade de que duas páginas contenham exatamente três erros de impressão é de 0,1255, ou seja, 12,55%.

Solução do item c): um ou mais erros é igual a:

$$P(x \geq 1) = P(x=1) + P(x=2) + P(x=3) + P(x=4) + \dots$$

Calculando pelo complementar, temos:

$$P(x \geq 1) = 1 - P(x=0)$$

$$\text{Calculando a } P(x = 0) = \frac{e^{-1,5} \cdot 1,5^0}{0!} = 0,2231$$

Substituindo, temos:

$$P(x \geq 1) = 1 - P(x=0) = 1 - 0,2231 = 0,7769$$

$$\mathbf{P(x \geq 1) = 0,7769}$$

Interpretação: A probabilidade de que duas páginas contenham um ou mais erros de impressão é de 0,7768, ou seja, 77,68%.

Exercícios resolvidos

1) A probabilidade de que um cliente aleatoriamente escolhido faça uma compra é 20%. Se um vendedor visita seis clientes, determine:

- a) a probabilidade de que o vendedor realize exatamente quatro vendas;
- b) a probabilidade de que o vendedor realize quatro ou mais vendas;
- c) o valor esperado do número de vendas e o desvio-padrão.

Solução do item a):

X: número de vendas $n=6$ $p=0,2$ $q=0,8$ $k=4$

$$P(x = 4) = C_{6,4} \cdot 0,2^4 \cdot 0,8^{6-4} =$$

$$\frac{6!}{4! (6-4)!} \cdot 0,2^4 \cdot 0,8^2 = 15 \cdot 0,0016 \cdot 0,64 = 0,0154$$

$$\mathbf{P(x=4) = 1,54\%}$$

Solução do item b):

$$P(4 \text{ ou mais}) = P(x \geq 4) = P(x = 4) + P(x = 5) + P(x = 6)$$

$$P(x=4) = 0,0154 \rightarrow \text{calculado no item a}$$

$$P(x = 5) = C_{6,5} \cdot 0,2^5 \cdot 0,8^{6-5} =$$

$$\frac{6!}{5! (6-5)!} \cdot 0,2^5 \cdot 0,8^1 = 6 \cdot 0,00032 \cdot 0,8 = 0,0015$$

$$P(x = 6) = C_{6,6} \cdot 0,2^6 \cdot 0,8^{6-6} =$$

$$\frac{6!}{6! (6-6)!} \cdot 0,2^6 \cdot 0,8^0 = 1 \cdot 0,000064 \cdot 1 = 0,000064$$

Substituindo:

$$\begin{aligned} P(x \geq 4) &= P(x = 4) + P(x = 5) + P(x = 6) \\ &= 0,0154 + 0,0015 + 0,000064 = 0,0169 \times 100 = 1,69\% \end{aligned}$$

$$\mathbf{P(x \geq 4) = 1,7\% \text{ (arredondado)}}$$

Solução do item c)

$$\mu = E(X) = n \cdot p = 6 \cdot 0,2 = 1,2$$

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{6 \cdot 0,2 \cdot 0,8} = \sqrt{0,96} = 0,98$$

2) Em uma transportadora chegam caminhões à razão de 2,8 caminhões/hora. Determine a probabilidade de chegarem três ou mais caminhões:

- a) num período de trinta minutos;
- b) num período de 2h.

Solução do item a):

x: número de caminhões

$$\lambda = \frac{2,8 \text{ caminhões}}{1 \text{ hora}} \quad t = 30 \text{ minutos}$$

Passando o tempo para minutos em λ também:

$$\mu = \lambda \cdot t = \frac{2,8 \text{ caminhões}}{60 \text{ min.}} \cdot 30 \text{ min.} = 1,4 \text{ caminhões} \rightarrow \mu = 1,4 \text{ caminhões}$$

$$P(3 \text{ ou mais}) = P(x \geq 3) = P(x=3) + P(x=4) + \dots$$

Neste caso, teremos que resolver pelo complementar:

$$P(x \geq 3) = 1 - P(x < 3) =$$

$$P(x \geq 3) = 1 - [P(x=0) + P(x=1) + P(x=2)] = \quad (1)$$

$$P(x = 0) = \frac{e^{-1,4} \cdot 1,4^0}{0!} = 0,2466$$

$$P(x = 1) = \frac{e^{-1,4} \cdot 1,4^1}{1!} = 0,3452$$

$$P(x = 2) = \frac{e^{-1,4} \cdot 1,4^2}{2!} = 0,2417$$

Substituindo em (1):

$$P(x \geq 3) = 1 - [P(x=0) + P(x=1) + P(x=2)] = \\ = 1 - [0,2466 + 0,3452 + 0,2417] = 1 - 0,8335 = 0,1665$$

$$P(x \geq 3) = 16,65\%$$

Solução do item b):

$$X: \text{número de caminhões} \quad \lambda = \frac{2,8 \text{ caminhões}}{1 \text{ hora}} \quad t = 2 \text{ horas}$$

$$\mu = \lambda \cdot t = \frac{2,8 \text{ caminhões}}{1 \text{ hora}} \cdot 2 \text{ horas} = 5,6 \text{ caminhões}$$

$$P(x \geq 3) = 1 - P(x < 3) = 1 - [P(x=0) + P(x=1) + P(x=2)] \quad (2)$$

$$P(x = 0) = \frac{e^{-5,6} \cdot 5,6^0}{0!} = 0,0037$$

$$P(x = 1) = \frac{e^{-5,6} \cdot 5,6^1}{1!} = 0,0207$$

$$P(x = 2) = \frac{e^{-5,6} \cdot 5,6^2}{2!} = 0,0580$$

Substituindo em (2):

$$P(x \geq 3) = 1 - [P(x=0) + P(x=1) + P(x=2)] =$$

$$P(x \geq 3) = 1 - [0,0037 + 0,0207 + 0,0580] = \\ = 1 - 0,0824 = 0,9176$$

$$P(x \geq 3) = 91,76\%$$

Exercícios complementares

1) Numa empresa, 60% das reclamações recebidas são resolvidas. Qual a probabilidade de que entre cinco reclamações:

- nenhuma seja resolvida?
- exatamente três sejam resolvidas?
- no mínimo três sejam resolvidas?

2) Um departamento recebe oito chamadas por hora. Qual a probabilidade de que:

- a) em uma hora ocorram três chamadas?
- b) em 30 minutos ocorram pelo menos duas chamadas?
- c) em 15 minutos não ocorram chamadas?

3) Suponha que $\frac{2}{5}$ da população tenham sangue tipo O⁺. Escolhidas aleatoriamente oito pessoas, qual a probabilidade de que:

- a) quatro delas tenham sangue O⁺?
- b) nenhuma tenha sangue O⁺?

4) Suponha que determinado tratamento aplicado a portadores de certa patologia cure 90% dos casos. Qual a probabilidade de que, entre 10 indivíduos submetidos a esse tratamento,

- a) oito tenham sido curados?
- b) mais de nove tenham sido curados?
- c) menos de nove tenham sido curados?

5) Em média, seis pessoas por hora utilizam os serviços de caixa automático de um banco. Qual a probabilidade de que:

- a) exatamente cinco pessoas usem os serviços do caixa-automático durante o intervalo de uma hora?
- b) menos de três pessoas usem os serviços do caixa-automático durante o intervalo de uma hora?
- c) nenhuma pessoa use o caixa automático durante um intervalo de dez minutos?
- d) nenhuma pessoa use os serviços durante o intervalo de trinta minutos?

6) Suponha que o manuscrito de um livro-texto tenha um total de 50 erros nas 500 páginas de material. Se os erros estão distribuídos aleatoriamente ao longo do texto, qual a probabilidade de que:

- a) um capítulo cobrindo 25 páginas tenha dois ou mais erros?

- b) um capítulo cobrindo 50 páginas tenha no mínimo dois erros?
- c) uma página aleatoriamente selecionada não tenha nenhum erro?

7) O setor de atendimento ao cliente de uma empresa resolve 80% das reclamações recebidas. Qual a probabilidade de, entre cinco reclamações recebidas:

- a) todas serem resolvidas pelo setor?
- b) pelo menos duas serem resolvidas?
- c) exatamente três serem resolvidas?

8) Se a probabilidade de um eleitor, escolhido aleatoriamente, não comparecer para votar em uma eleição, é 25%, qual a probabilidade de, entre 10 eleitores:

- a) exatamente seis não comparecerem para votar?
- b) no máximo dois não comparecerem para votar?
- c) Qual o valor esperado de eleitores que não comparecerão para votar?

9) Se um posto de gasolina recebe em média três cheques sem fundos por dia, qual é a probabilidade de receber:

- a) quatro cheques sem fundos em um dia qualquer?
- b) pelo menos dois cheques sem fundos em dois dias?

10) Cada rolo de tecido de 500m contém, em média, duas imperfeições. Qual a probabilidade de que um segmento específico de:

- a) 100m de tecido não contenha nenhuma imperfeição?
- b) 200m de tecido contenha no máximo duas imperfeições?
- c) 300m de tecido não contenha três imperfeições?

Respostas

1. a) $P(x=0) = 0,0102$
b) $P(x=3) = 0,3456$
c) $P(x \geq 3) = 0,6826$
2. a) $P(x=3) = 0,0286$
b) $P(x \geq 2) = 0,9084$
c) $P(x=0) = 0,1353$
3. a) $P(x=4) = 23,22\%$
b) $P(x=0) = 1,68\%$
4. a) $P(x=8) = 0,1937$
b) $P(x > 9) = 0,3487$
c) $P(x < 9) = 0,2639$
5. a) $P(x=5) = 16,06\%$, $\mu = 6$ pessoas
b) $P(x < 3) = 6,2\%$, $\mu = 6$ pessoas
c) $P(x=0) = 36,79\%$, $\mu = 1$ pessoa
d) $P(x=0) = 4,98\%$, $\mu = 3$ pessoas
6. a) $P(x \geq 2) = 71,26\%$, $\mu = 2,5$ erros
b) $P(x \geq 2) = 95,96\%$, $\mu = 5$ erros
c) $P(x=0) = 90,48\%$, $\mu = 0,1$ erro
7. a) $P(x = 5) = 32,77\%$
b) $P(x \geq 2) = 99,33\%$
c) $P(x = 3) = 20,48\%$
8. a) $P(x = 6) = 1,62\%$
b) $P(x \leq 2) = 52,56\%$
c) $\mu = E(x) = 2,5$
9. a) $P(x = 4) = 16,81\%$, $\mu = 3$ cheques
b) $P(x \geq 2) = 98,26\%$, $\mu = 6$ cheques
10. a) $P(x = 0) = 67,03\%$, $\mu = 0,4$
b) $P(x \leq 2) = 95,25\%$, $\mu = 0,8$
c) $1 - P(x = 3) = 1 - 0,0867 = 0,9133$, $\mu = 1,2$

8 Distribuição de probabilidade para variáveis aleatórias contínuas

Os valores assumidos por uma variável aleatória contínua podem ser associados com medidas em uma escala contínua, como por exemplo, a altura de uma pessoa: 1,7m; 1,72m; 1,723m; pois o resultado depende da precisão do instrumento utilizado na medição. Na prática, não nos resta outra escolha senão arredondar as medidas para o inteiro mais próximo ou para algumas casas decimais.

No caso das variáveis aleatórias discretas, as probabilidades são representadas pelas alturas dos retângulos do histograma. Nas variáveis aleatórias contínuas, representamos probabilidades por áreas, como na Figura 8.1, que apresenta os dados referentes à altura dos empregados de uma empresa.

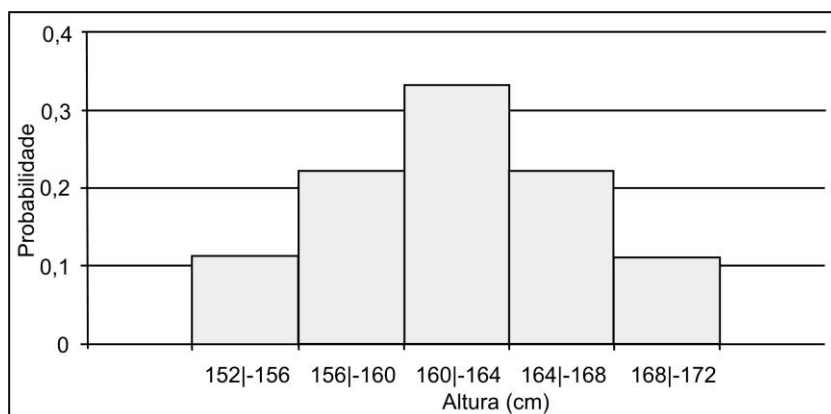


Figura 8.1 – Histograma da variável contínua altura

O gráfico de uma variável aleatória contínua pode apresentar diversas formas, como nos gráficos a seguir:

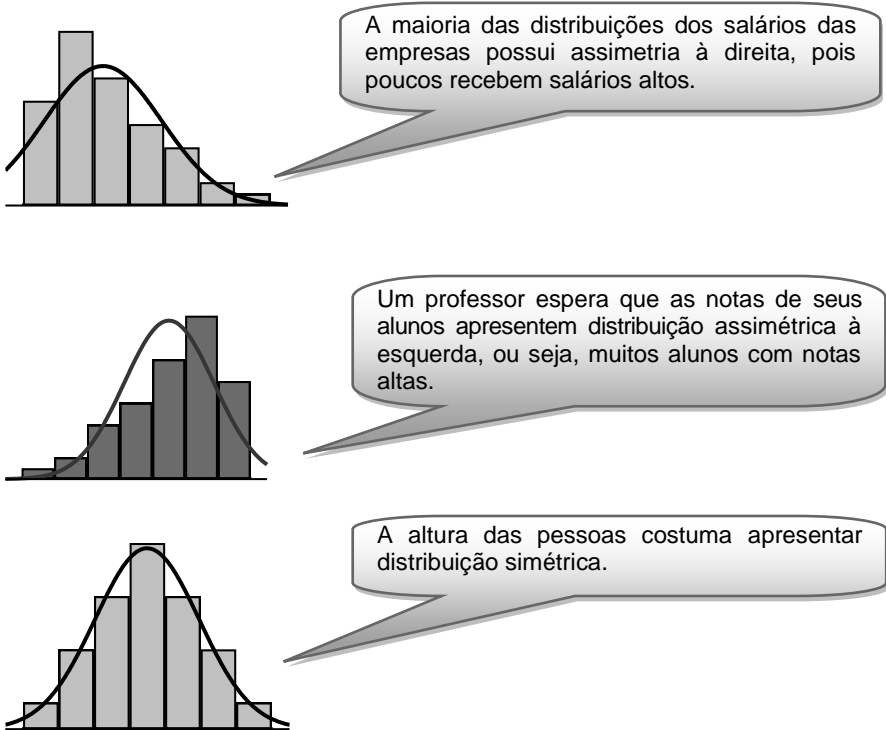


Figura 8.2 – Diferentes distribuições para uma variável aleatória contínua.

A probabilidade de uma variável aleatória assumir um valor no intervalo de **a** a **b** é dada pela área sob a curva entre esses dois valores. A probabilidade ou a área abaixo da curva é calculada pela seguinte expressão:

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

onde $f(x)$ é a função que gera a curva.

Existem várias distribuições de probabilidade contínuas que variam de acordo com suas características e forma da curva. Neste livro, estudaremos especificamente a **distribuição normal**.

8.1 Distribuição normal

Muitos fenômenos sociais, psicológicos, físicos, entre outros, apresentam uma distribuição normal. Uma das propriedades das distribuições normais é a simetria da curva, como na Figura 8.1. No que diz respeito às características humanas, a maioria dos adultos se enquadra entre 155 e 185cm de altura; poucos são muito baixos (menos de 155cm) ou muito altos (mais de 185cm). A maior parte dos coeficientes de inteligência (QI) está entre 85 e 115. Há uma queda gradativa dos escores (medida) em ambas as extremidades – poucos “gênios” obtendo QI maior que 145 e poucas pessoas com QI inferior a 55.

Diversas variáveis contínuas apresentam comportamento semelhante independentemente do experimento de que são provenientes. As distribuições normais são importantes em estatística, pois muitas das técnicas de inferência estatística baseadas nas distribuições normais também funcionam bem quando aplicadas a outras distribuições que são aproximadamente simétricas. Ao estudar esse comportamento, Gauss chegou à seguinte expressão:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{para } -\infty < x < \infty$$

onde

σ = desvio-padrão populacional

μ = média da população

e = número de Euler

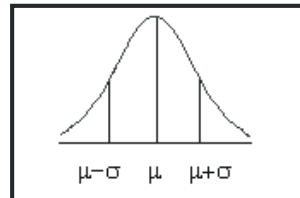
π = constante pi

Notação: $X:N(\mu, \sigma^2)$ lê-se: a variável aleatória x tem distribuição normal com média μ e variância σ^2 .

Exemplo: $X:N(200, 25)$ lê-se: a variável aleatória x apresenta distribuição normal com média igual a 200 e variância igual a 25.

8.1.1 Características da curva normal

- É simétrica em relação à média μ ;
- O ponto máximo (moda) de $f(x)$ é o ponto $x = \mu$;
- Os pontos de inflexão da função são: $\mu - \sigma$ e $\mu + \sigma$;



- A área total limitada pela curva e pelo eixo das abscissas é 1 ou 100%;
- A curva é assintótica em relação ao eixo dos x , ou seja, prolonga-se em ambas as direções, aproximando-se do eixo horizontal, sem jamais tocá-lo.

Para obter a probabilidade indicada na figura acima, devemos calcular:

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Como os valores de probabilidade são obtidos por integração numérica, seria muito trabalhoso utilizar essa função. Para facilitar usaremos a variável normal padrão ou variável padronizada, z .

8.1.2 Variável normal padrão ou variável padronizada

A distribuição normal padrão é uma distribuição com forma normal (de sino), de média 0 (zero) e desvio-padrão igual a 1 (um).

Devido a essas e outras características, a distribuição normal padrão é muito útil. Para podermos utilizar a distribuição normal padrão na análise de dados, basta calcular o escore padrão ou a variável padronizada (z) que indica a direção e o quanto qualquer escore padrão se desvia da média de uma distribuição em uma escala de unidades de desvios-padrões.

$$z = \frac{x - \mu}{\sigma}$$

Exemplo: A média para valores de QI em testes de inteligência é 100, com um desvio-padrão de 15. Se você tem um QI de 130, qual seria seu valor z ? Interprete este resultado.

Solução: $\mu = 100$ $\sigma = 15$

$$z = \frac{130 - 100}{15} = 2,0$$

Uma pessoa com QI igual a 130 está dois desvios-padrões acima da média, pois o valor de z é positivo.

8.1.3 A curva normal padrão

A função normal padrão é dada pela seguinte expressão:

$$f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2}$$

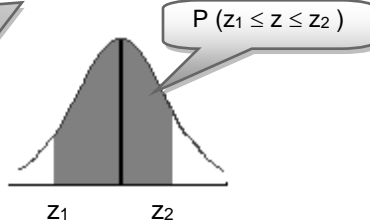
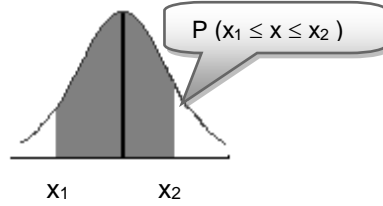
Para obter a probabilidade, nesse caso, devemos calcular:

$$P(a \leq z \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2} dz$$

A vantagem de se usar a variável padronizada z é a possibilidade de calcular e organizar os valores da área, ou as probabilidades, em uma única tabela.

A probabilidade de um determinado valor x estar entre dois limites, x_1 e x_2 , é igual à probabilidade de um determinado valor z estar entre dois limites z_1 e z_2 :

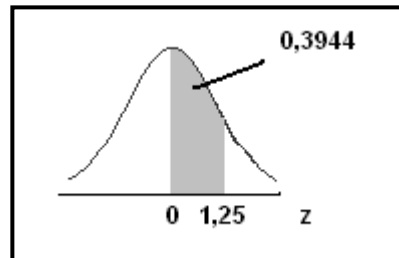
$$P(x_1 \leq x \leq x_2) = P(z_1 \leq z \leq z_2)$$



8.1.4 Uso da tabela da curva normal padrão

Para usar a tabela z , Anexo III, da curva normal padrão, deve-se considerar o fato de que a curva é simétrica e centrada na média. O corpo da tabela é constituído das probabilidades (área sob a curva entre os limites de zero e z). Os valores de z estão nas margens da tabela; na primeira coluna está o valor inteiro e a primeira casa decimal; na primeira linha está a segunda casa decimal. Por exemplo, o valor de $z=1,25$ é obtido pela intersecção da linha que contém o valor 1,2 e a coluna que contém a segunda casa decimal, o 5, do valor de z procurado:

A área abaixo da curva entre zero e o valor de z igual a 1,25 é de 0,3944, como mostra a tabela a seguir:



z	Segunda casa decimal							
	0	1	2	3	4	5	6	7
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486
0,7	0,2580	0,2613	0,2645	0,2677	0,2709	0,2740	0,2771	0,2801
0,8	0,2881	0,2912	0,2942	0,2972	0,3001	0,3031	0,3061	0,3091
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418

Valor inteiro e a primeira casa decimal

Segunda casa decimal

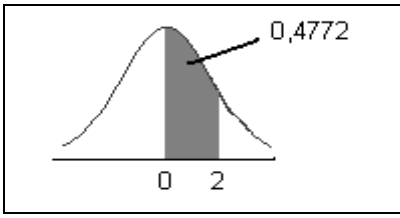
Probabilidade entre a média zero e o valor z = 1,25

Exemplo de uso da tabela:

Dados os valores de z, determine as probabilidades indicadas abaixo:

a) $P(0 \leq z \leq 2)$

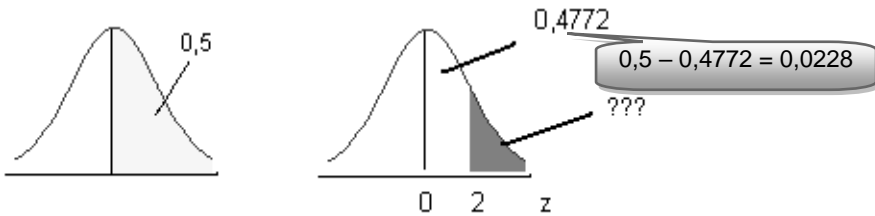
Estamos procurando a probabilidade de um valor z estar entre zero e dois. A tabela fornece a área entre a média de z (zero) e o valor de z procurado. Obtemos esse valor procurando na primeira coluna da tabela o valor de z = 2,0 e na primeira linha o valor "0" que é a segunda casa decimal. O valor da área é 0,4772.



$P(0 \leq z \leq 2) = 0,4772$

b) $P(z \geq 2)$

Como a curva é simétrica e a área abaixo da curva é igual a um, cada metade vale 0,5. Estamos procurando a probabilidade de um valor z ser maior ou igual a 2. A tabela fornece a área entre a média de z (zero) e o valor de z procurado. Portanto, é necessário subtrair o valor encontrado na tabela de 0,5.



$$P(z \geq 2) = 0,5 - 0,4772 = 0,0228$$

c) $P(z \leq -2)$

Como a curva é simétrica, a área entre $z=0$ e $z=2$ é a mesma entre $z=-2$ e $z=0$. Com raciocínio análogo ao item anterior, temos:

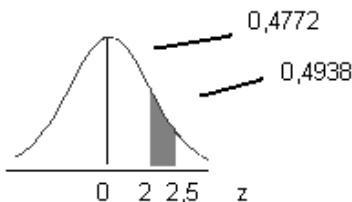


$$P(z \leq -2) = 0,5 - 0,4772 = 0,0228$$

$$P(z \leq -2) = \mathbf{0,0228}$$

d) $P(2,0 \leq z \leq 2,5)$

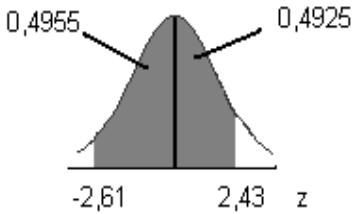
De $z=0$ até o ponto $z=2,5$ a área é igual a 0,4938; subtraindo desse valor a área menor, que é 0,4772 (área de $z=0$ até o ponto $z=2$), teremos a área procurada.



$$P(2,0 \leq z \leq 2,5) = 0,4938 - 0,4772$$

$$P(2,0 \leq z \leq 2,5) = \mathbf{0,0166}$$

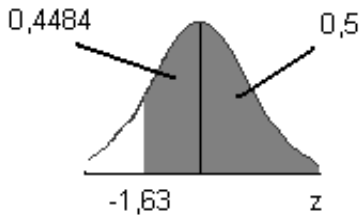
e) $P(-2,61 \leq z \leq 2,43)$



$$P(-2,61 \leq z \leq 2,43) = 0,4955 + 0,4925$$

$$P(-2,61 \leq z \leq 2,43) = 0,9880$$

f) $P(z \geq -1,63)$



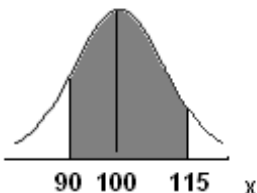
$$P(z \geq -1,63) = 0,5 + 0,4484 = 0,9484$$

$$P(z \geq -1,63) = 0,9484$$

8.1.5 Exemplos de aplicação

Exemplo 1: Voltando ao exemplo dos QIs: qual a probabilidade de selecionarmos ao acaso uma pessoa e ela ter QI entre 90 e 115? Considere média populacional 100 e desvio-padrão igual a 15.

Solução:

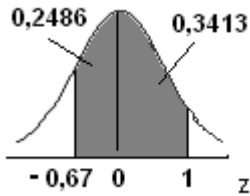


$$P(90 \leq x \leq 115) = ?$$

Padronizando a variável x_1 , temos:

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{90 - 100}{15} = -0,67$$

Para $x_1 = 90$ o valor de z_1 correspondente é $z = -0,67$



Padronizando a variável x_2 , temos:

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{115 - 100}{15} = 1,0$$

Para $x_2 = 115$ o valor de z_2 correspondente é $z = 1,0$

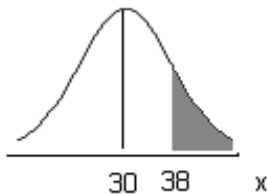
$$P(90 \leq x \leq 115) = P(0,67 \leq z \leq 1,0) = 0,2486 + 0,3413$$

$$P(90 \leq x \leq 115) = 0,5899$$

$$\mathbf{P(90 \leq x \leq 115) = 0,5899}$$

Exemplo 2: Em uma distribuição normal com média 30 e desvio-padrão 4, determine a probabilidade de um indivíduo apresentar valor maior ou igual a 38.

Solução:



Padronizando a variável x , temos:

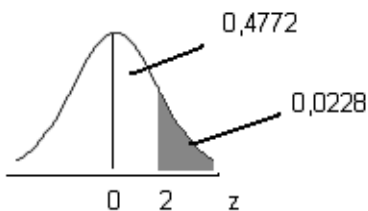
$$z = \frac{x - \mu}{\sigma} = \frac{38 - 30}{4} = 2,0$$

Para $x = 38$ o valor de z correspondente é $z = 2,0$

$$P(x \geq 38) = P(z \geq 2) = 0,5 - 0,4772 = 0,0228$$

$$P(x \geq 38) = 0,0228$$

$$\mathbf{P(x \geq 38) = 0,0228}$$



Exercício resolvido

A duração de um pneu de automóvel, em quilômetros rodados, apresenta distribuição normal com média 70.000km e desvio-padrão de 10.000km.

a) Qual a probabilidade de um pneu escolhido ao acaso durar mais de 85.000km?

b) Qual a probabilidade de um pneu durar entre 68.500km e 75.000km?

c) Qual a probabilidade de um pneu durar entre 55.000km e 65.000km?

d) O fabricante deseja fixar uma garantia de quilometragem rodada, de tal forma que, se a duração de um pneu for inferior à da garantia, o pneu será trocado. De quanto deve ser essa garantia para que somente 1% dos pneus sejam trocados?

e) De acordo com o item anterior, a probabilidade de que um pneu seja trocado é de 1%. Se o fabricante vende 5.000 pneus por mês, quantos pneus espera trocar?

Solução do item a) $P(x > 85.000)$

Padronizando a variável x , temos: $z = \frac{x - \mu}{\sigma} = \frac{85.000 - 70.000}{10.000} = 1,5$



$$P(x > 85.000) = P(z > 1,5) = 0,5 - 0,4332 = 0,0668$$

$$P(x > 85.000) = 0,0668 \times 100 = 6,68\%$$

$$\mathbf{P(x > 85.000) = 6,68\%}$$

Solução do item b) $P(68.500 \leq x \leq 75.000)$

Padronizando a variável x_1 , temos:

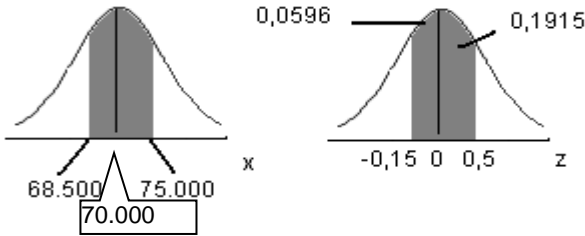
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{68.500 - 70.000}{10.000} = -0,15$$

Para $x_1 = 68.500$ o valor de $z = -0,15 \rightarrow$ tabela = 0,0596

Padronizando a variável x_2 , temos:

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{75.000 - 70.000}{10.000} = 0,5$$

Para $x_2 = 75.000$ o valor de $z = 0,5 \rightarrow$ tabela = 0,1915



$$P(68.500 \leq x \leq 75.000) = P(-0,15 \leq z \leq 0,5) = 0,1915 + 0,0596 = 0,2511$$

$$P(68.500 \leq x \leq 75.000) = 0,2511 \times 100 = 25,11\%$$

$$\mathbf{P(68.500 \leq x \leq 75.000) = 25,11\%}$$

Solução do item c) $P(55.000 \leq x \leq 65.000)$

Padronizando a variável x_1 , temos:

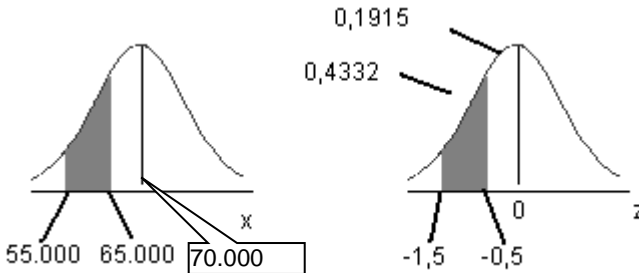
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{55.000 - 70.000}{10.000} = -1,5$$

Para $x_1 = 55.000$ o valor de $z = -1,5 \rightarrow$ tabela = 0,4332

Padronizando a variável x_2 , temos:

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{65.000 - 70.000}{10.000} = -0,5$$

Para $x_2 = 65.000$ o valor de $z = -0,5 \rightarrow$ tabela = 0,1915



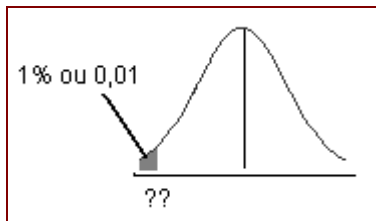
$$P(55.000 \leq x \leq 65.000) = P(-1,5 \leq z \leq -0,5) = 0,4332 - 0,1915 = 0,2417$$

$$P(55.000 \leq x \leq 65.000) = 0,2417 \times 100 = 24,17\%$$

$$\mathbf{P(55.000 \leq x \leq 65.000) = 24,17\%}$$

Solução do item d) $P(x \leq ??) = 1\%$

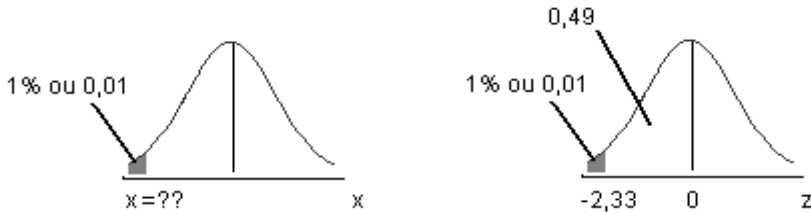
O valor da variável x (quilometragem rodada do pneu) procurado está à esquerda da curva, pois a área ou probabilidade é 1% para valores iguais ou menores do que x . Construindo a curva normal padrão, temos que: se a área hachurada é igual a 0,01 a área até o centro da curva na média $z = 0$ é igual a 0,49. Procurando no corpo da tabela o valor da área 0,49 encontraremos o valor de z correspondente:



z	0	1	2	3	4	5	6	...
∴	∴	∴	∴	∴	∴	∴	∴	∴
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	...
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	...
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	...
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	...
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	...
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	...
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	...
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	...
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	...
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	...
∴	∴	∴	∴	∴	∴	∴	∴	∴



O valor mais próximo de 0,49 na tabela é 0,4901, portanto o valor de z correspondente a essa área é $z = -2,33$, negativo, pois está à esquerda de zero. Logo, temos:



Substituindo o valor de z na fórmula da variável padronizada:

$$z = \frac{x - \mu}{\sigma}$$

$$-2,33 = \frac{x - 70.000}{10.000} =$$

$$x - 70.000 = - 2,33 \cdot 10000$$

$$x = - 23.300 + 70.000$$

$$x = 46.700$$

$$\mathbf{P(x \leq 46.700) = 1\%}$$

A quilometragem fixada pelo fabricante é de 46.700km, ou seja, se um pneu durar menos do que 46.700km, será trocado. O fabricante tem a garantia de que a probabilidade de um pneu ser trocado é de 1%.

Solução do item e) 1% de 5.000 pneus = 50 pneus

Exercícios complementares

1) O salário médio dos funcionários de uma empresa é de R\$800,00, com desvio-padrão de R\$300,00. Qual a probabilidade de que, selecionando-se aleatoriamente um funcionário, ele receba:

- entre R\$650,00 e R\$1.100,00;
- entre R\$550,00 e R\$750,00;
- acima de R\$500,00;
- acima de R\$1.100,00;

2) A dosagem de nicotina encontrada nos cigarros de uma determinada marca apresenta uma distribuição normal com média de 12mg e desvio-padrão de 0,5mg.

a) qual a probabilidade de que, selecionando-se ao acaso um desses cigarros, nele encontrem-se mais de 13,5mg de nicotina?

b) qual a probabilidade de que, selecionando-se ao acaso um desses cigarros, a dosagem de nicotina esteja entre 11,5 e 12,5mg?

c) em um maço com 20 cigarros, quantos cigarros apresentam os níveis de nicotina do item b acima?

3) O aproveitamento dos alunos em determinada disciplina segue uma distribuição normal com nota média de 7,2 e desvio-padrão de 3,0.

a) qual o percentual de alunos cuja nota seja igual ou superior a 6,0?

b) qual o percentual de alunos que poderão ser reprovados por possuírem grau inferior a 5,0? (a nota mínima para aprovação é 5,0).

c) que percentual de alunos espera-se possuam nota entre 8,0 e 10,0?

4) O salário médio dos funcionários de uma empresa é de R\$800,00, com desvio-padrão de R\$300,00 (não se consideram os salários da direção da empresa). Qual o salário que é mínimo para 30,5% dos funcionários que mais recebem?

5) Um conjunto de escores de QI tem distribuição normal, com média $\mu = 100$ e desvio-padrão $\sigma = 15$. Com base nessa distribuição, determine a probabilidade de escolher aleatoriamente uma pessoa com:

a) QI de 110 ou mais;

b) QI entre 88 e 120.

6) Se a altura média dos estudantes de uma universidade é 1,75m e o desvio-padrão 0,08m, qual a probabilidade de um estudante sorteado ao acaso ter altura entre 1,60m e 1,70m?

7) Um entregador de pizza tem até 20 minutos para fazer uma entrega, ou ele paga a conta. Para tanto ele dispõe de dois itinerários: o itinerário A, que tem distribuição $N(18; 25)$, e o itinerário B, que tem distribuição $N(20; 4)$. Qual a melhor escolha?

Obs.: Realize os cálculos para justificar sua resposta.

8) A produção de determinado produto apresenta distribuição normal. As peças com mais de 22mm e menos de 18mm são consideradas defeituosas. O processo atual produz 30% de peças defeituosas. Foi proposta a troca por um processo com média de 21mm e desvio-padrão de 0,9mm. Deve ser feita a troca?

Obs: Realize os cálculos para justificar sua resposta.

9) Um estudo das modificações percentuais dos preços, no atacado, de produtos industrializados, mostrou que há distribuição normal com média de 50% e desvio-padrão de 10%. Qual a porcentagem dos artigos que:

a) sofreram aumentos superiores a 75%?

b) sofreram aumentos entre 30% e 80%?

10) Um fabricante de baterias sabe, por experiência, que as baterias de sua fabricação têm vida média de 600 dias e desvio-padrão de 100 dias, sendo que a duração tem aproximadamente distribuição normal. Oferece garantia de 312 dias, isto é, troca as baterias que apresentem falhas nesse período. Fabrica 10.000 baterias mensalmente. Quantas baterias deverá trocar, mensalmente, em garantia?

11) Duas fábricas produzem o mesmo tipo de circuito integrado. A fábrica A produz circuitos integrados cujo tempo de vida médio é de 12.000 horas, com desvio-padrão de 2.000 horas. Os circuitos integrados produzidos pela fábrica B possuem vida média de 13.000 horas, com desvio-padrão de 1.500 horas. Qual dos circuitos integrados tem maior probabilidade de durar pelo menos 14.000 horas – os produzidos pela fábrica A ou pela fábrica B?

12) Qual das afirmações abaixo não é válida para uma curva normal?

- a) É assimétrica.
- b) É uma distribuição de probabilidades.
- c) A área total abaixo da curva é igual a 1,0.
- d) A moda, a mediana e a média são iguais.

Respostas

1. a) $P(650 \leq x \leq 1.100) = 0,5328$

b) $P(550 \leq x \leq 750) = 0,2292$

c) $P(x > 500) = 0,8413$

d) $P(x > 1.100) = 0,1587$

2. a) $P(x > 13,5) = 0,0013$

b) $P(11,5 \leq x \leq 12,5) = 0,6826$

c) 13,652 → arredondando, 14 cigarros.

3. a) $P(x \geq 6,0) = 0,6554$

b) $P(x < 5,0) = 0,2327$

c) $P(8,0 \leq x \leq 10,0) = 0,2174$

4. $P(x > R\$ 953,00) = 30,5\%$

5. a) $P(x \geq 110) = 0,2514$

b) $P(88 \leq x \leq 120) = 0,6963$

6. $P(1,60m \leq x \leq 1,70m) = 0,2342$

7. O primeiro itinerário, com $N(18; 25)$, pois apresenta a maior probabilidade de ser completado no máximo em 20 minutos, $P(x \leq 20) = 65,54\%$.

8. Sim, pois a probabilidade de produzir peças defeituosas no novo processo é de 13,39%, ou seja, $P(x < 18) + P(x > 22) = 13,39\%$, menor do que o antigo processo que produzia 30% de peças defeituosas.

9. a) $P(x > 75\%) = 0,62\%$

b) $P(30\% \leq x \leq 80\%) = 97,59\%$

10. A probabilidade de uma bateria durar 312 dias ou menos é: $P(x \leq 312) = 0,2\%$. Como o fabricante fabrica 10.000 baterias mensalmente, deverá trocar 20 baterias por mês (0,2% de 10.000).

11. Os circuitos integrados da fábrica B, pois apresentam maior probabilidade de durar pelo menos 14.000 horas.

Fábrica A $\rightarrow P(x \geq 14.000) = 15,87\%$

Fábrica B $\rightarrow P(x \geq 14.000) = 25,14\%$

12. a) É assimétrica.

BIBLIOGRAFIA

ANDERSON, David R.; SWEENEY Dennis J.; WILLIAMS Thomas A. Estatística aplicada a administração e economia. São Paulo: Pioneira Thomson Learning, 2003.

BUSSAB, Wilton de O.; MORETTIN, Pedro A. Estatística básica. 5. ed. São Paulo: Saraiva, 2006.

LARSON, Ron; FARBER Betsy. Estatística aplicada. São Paulo: Prentice Hall, 2004.

LEVIN, Jack; FOX, James Alan. Estatística para Ciências Humanas. 9 ed. São Paulo: Prentice Hall, 2004.

LEVINE, David L.; STEPHAN, David; KREHBIEL, Timothy C.; BERENSON, Mark L. Estatística: teoria e aplicações usando o Microsoft Excel em português. Rio de Janeiro: Livros Técnicos e Científicos, 2005.

MAGALHÃES, Marcos Nascimento; LIMA, Antonio Carlos Pedroso. Noções de probabilidade e estatística. 6. ed. São Paulo: EDUSP, 2004.

MILONE, Giuseppe. Estatística geral e aplicada. São Paulo: Pioneira Thomson Learning, 2004.

MOORE, David S.; McCABE, George P.; DUCKWORTH, William M.; SCLOVE, Stanley L. A prática da estatística empresarial: como usar os dados para tomar decisões. Rio de Janeiro: LTC, 2006.

MORETTIN, Pedro A. Estatística básica. São Paulo: Makron Books, 1999. v. 1: Probabilidade.

SILVA, E. M.; SILVA, E. M.; GONÇALVES, V.; MUROLO, A . C. Estatística. São Paulo: Atlas, 1996.

TANAKA, O. K.; PEREIRA, W. Estatística: conceitos básicos. São Paulo: McGraw-Hill, 1990.

TRIOLA, Mario F. Introdução à estatística. 9. ed. Rio Janeiro: LTC, 2005.

Anexo I – A Lei dos Grandes Números

Para entender melhor a lei dos grandes números, vamos trabalhar com o seguinte exemplo:

Experimento: Lançar um dado $S=\{1, 2, 3, 4, 5, 6\}$ $n(S) = 6$

Evento A: ocorrer face 6 $A=\{6\}$ $n(A) = 1$

Pela probabilidade clássica, temos que a probabilidade de ocorrer face 6 no lançamento de um dado é:

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{6} = 0,167$$

Agora, em vez de uma única vez, o dado foi lançado 10 vezes e registrado o número de vezes que ocorreu o evento 6. Depois, o dado foi lançado 20 vezes e anotado o número de vezes que ocorreu o evento 6. Esse experimento foi repetido várias vezes, aumentando-se o número de lançamentos para 30, 50, 100, 200... 1000. A frequência relativa foi calculada para cada lançamento. Na figura 1, é apresentado o número de lançamentos do dado e a frequência relativa do evento 6. Podemos verificar que, à medida que aumenta o número de repetições do experimento, as frequências relativas tendem para 0,167, que é aproximadamente 1/6 (probabilidade clássica).

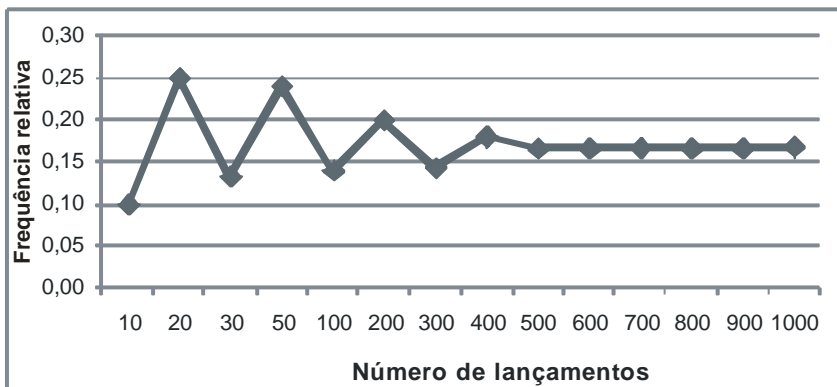


Figura 1 – Lei dos grandes números

Lei dos grandes números: à medida que um experimento é repetido várias vezes, a probabilidade dada pela frequência relativa de um evento tende a se aproximar da probabilidade clássica.

Anexo II – Combinação

Há muitos problemas em que precisamos saber o número de maneiras pelas quais podemos escolher k objetos de um conjunto de n objetos, em que a troca de posição entre esses objetos não gera um agrupamento diferente, ou seja, a ordem em que a escolha é feita não importa. Por exemplo, podemos estar interessados em conhecer o número de maneiras de escolher seis dentre 40 declarações de imposto de renda para a malha fina ou o número de maneiras de escolher três estudantes para representar a turma no comitê estudantil.

Para obter uma fórmula que se aplique a problemas como esses, examinemos primeiro o seguinte exemplo:

Exemplo 1: Quantos grupos de duas estudantes podem ser formados a partir de um grupo de três estudantes: Ana, Beatriz e Catia.

Ana – Beatriz	Ana – Catia
Beatriz – Ana	Beatriz – Catia
Catia – Ana	Catia – Beatriz

Obtemos seis grupos, no entanto o grupo formado por Ana – Beatriz é o mesmo que Beatriz – Ana, pois a ordem em que as estudantes são selecionadas não muda a natureza do grupo. Portanto, podemos formar apenas três grupos diferentes com seis estudantes.

O número de maneiras pelas quais k objetos podem ser escolhidos de um conjunto de n objetos é obtido pela seguinte fórmula:

$$C_{n,k} = \frac{n!}{k!(n-k)!}$$

Combinções de n objetos tomados k de cada vez.

No exemplo anterior $n=3$ e $k=2$, substituindo na fórmula, temos:

$$C_{n,k} = \frac{n!}{k!(n-k)!} = C_{3,2} = \frac{3!}{2!(3-2)!} = \frac{3 \times 2 \times 1}{2 \times 1 \times 1} = \frac{6}{2} = 3$$

Podemos formar três grupos diferentes com dois elementos cada entre três elementos.

Exemplo 2: Determine o número de maneiras de escolher quatro declarações dentre 30 declarações de imposto de renda para a malha fina.

Solução: $n = 30$ $k = 4$

$$\begin{aligned} C_{30,4} &= \frac{30!}{4!(30-4)!} = C_{30,4} = \frac{30!}{4!(30-4)!} = \\ &= \frac{30!}{4!(26)!} = \frac{30 \times 29 \times 28 \times 27 \times 26!}{4 \times 3 \times 2 \times 1 (26!)} \\ &= \frac{30 \times 29 \times 28 \times 27}{4 \times 3 \times 2 \times 1} = 27.405 \end{aligned}$$

Simplificamos o 26! do numerador com o 26! do denominador

Anexo III

Tabela da distribuição normal padrão



z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986

Continuação

z	0	1	2	3	4	5	6	7	8	9
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

EDITORA E GRÁFICA DA FURG
CAMPUS CARREIROS
CEP 96203 900
editora@furg.br

Suzi Samá Pinto é professora do Instituto de Matemática, Estatística e Física da Universidade Federal do Rio Grande (IMEF/FURG); licenciada em Matemática, mestre em Engenharia Oceânica e doutoranda em Educação em Ciência pela FURG. Membro do grupo de pesquisa em Educação Estatística e do grupo de pesquisa em Educação a distância e tecnologia. Atuou como professora pesquisadora e tutora à distância junto a Secretaria de Educação a Distância da FURG.

suzisama@furg.br

Carla Silva da Silva é professora da Escola de Engenharia da Universidade Federal do Rio Grande (EE/FURG). Bacharel em Engenharia Civil, mestre em Engenharia Oceânica pela FURG e doutora em Engenharia Civil pela UNICAMP. Suas áreas de interesse são Ensino de Estatística, Hidráulica, Mecânica, Gestão de Águas e Educação a Distância. Atuou como professora pesquisadora e formadora na modalidade a distância junto a UAB/FURG.

carlasilva@furg.br

ISSN 978-65-5759-009-3



9 786557 590093