

Estatística & Probabilidade

aplicadas às Engenharias e Ciências

Ben Dêivide de Oliveira Batista



Democratizando
Conhecimento

ESTATÍSTICA & PROBABILIDADE

BEN DÊIVIDE DE OLIVEIRA BATISTA

ESTATÍSTICA & PROBABILIDADE

BEN DÊIVIDE DE OLIVEIRA BATISTA



Democratizando
Conhecimento

Ouro Branco, MG, 26 de abril de 2022

© 2021 by Ben Dêivide de Oliveira Batista



Esse material está licenciado com uma Licença Creative Commons - Atribuição - Não Comercial 4.0 Internacional. Usamos também a filosofia de trabalho com o Selo Democratizando Conhecimento (DC). O leitor é livre para compartilhar, redistribuir, transformar ou adaptar essa obra, desde que não venha a utilizá-la em nenhuma atividade de propósito comercial. Por fim, a única exigência é a atribuição dos dos créditos aos autores dessa obra.

Direitos de publicação reservados ao seu conhecimento.

Impresso no Brasil - **ISBN** (Digital):

Impresso no Brasil - **ISBN** (Impresso):

Projeto Gráfico: Ben Dêivide de Oliveira Batista

Revisão técnica e textual: Daniel Furtado Ferreira

Revisão de Referências Bibliográficas:

Editoração Eletrônica: Ben Dêivide de Oliveira Batista

Capa: Ben Dêivide de Oliveira Batista

Como citar essa obra:

BATISTA, B. D. O.. **Estatística e Probabilidade:** aplicadas às Engenharias e Ciências. 1ed. Ouro Branco, MG:[sn]. 2022. Disponível em: <<https://bendeivide.github.io/book-epaec/>>

Autor correspondente e mantenedor da obra:

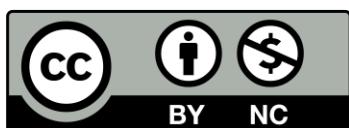
Ben Dêivide de Oliveira Batista

Contato: <ben.deivide@ufs.br>

Site pessoal: <<http://bendeivide.github.io/>>

Licença

Todos os direitos autorais contidos nesse livro são reservados ao seu conhecimento, usufrua-o, pois é totalmente de graça. Use-o com responsabilidade e saiba valorizar.



Este trabalho está licenciado com uma Licença Creative Commons - Atribuição - Não Comercial 4.0 Internacional. Usamos também a filosofia de trabalho com o Selo Democratizando Conhecimento (DC).



Epígrafe

*O pulsar de minha existência está
intensa e totalmente no presente da vida!*
Ben Dêivide

Sumário

Licença	i
Epígrafe	iii
Prefácio	vii
1 Definições gerais da Estatística e técnicas de somatório	1
1.1 Introdução	1
1.2 Definições gerais da Estatística	2
1.2.1 Estatística na pesquisa científica	5
1.3 Definições básicas	7
1.4 Técnicas de Somatório	11
Exercícios propostos	16
2 Coleta, organização e apresentação dos dados	19
2.1 Introdução	19
2.2 Representação tabular	20
2.3 Representação gráfica	29
Exercícios propostos	30
3 Medidas de Posição	32
3.1 Introdução	32
3.2 Média	32
3.3 Mediana	36
3.4 Moda	44
Exercícios propostos	48
4 Medidas de dispersão	49
4.1 Introdução	49
4.2 Amplitude total ou Amplitude	50
4.3 Variância	53
4.4 Desvio padrão	58
4.5 Coeficiente de Variação	60
4.6 Erro padrão da média	62
Exercícios propostos	64
5 Probabilidades	66
5.1 Introdução à teoria de conjuntos no contexto probabilístico	66
5.2 Definições de probabilidades	76
5.3 Propriedades	78
5.3.1 Regra de adição de probabilidades	78

5.3.2 Probabilidade de um evento complementar	79
5.4 Eventos independentes e probabilidade condicional	79
5.5 Teorema de Bayes	85
5.6 Variável Aleatória	89
5.7 Distribuição de X	91
5.8 Função de probabilidade (FP)	91
5.9 Função densidade de probabilidade (FDP)	91
5.10 Função de distribuição acumulada	93
5.10.1 FDA para X discreta	93
5.10.2 FDA para X contínua	94
5.11 Medidas resumo	96
5.12 Exercícios	99
6 Distribuições de probabilidades	104
6.1 Introdução	104
6.2 Distribuições discretas de probabilidades	104
6.3 Distribuições contínuas de probabilidades	122
6.4 Exercícios	135
7 Amostragem (Em desenvolvimento!)	138
7.1 Introdução	138
7.2 Amostragem não-probabilística e probabilística	142
7.3 Técnicas de amostragem probabilística	142
8 Distribuição de amostragem	143
8.1 Introdução	143
8.2 Distribuição de amostragem da média	143
8.3 Distribuição de amostragem de proporções	143
8.4 Distribuição de amostragem de diferença entre médias	143
8.5 Distribuições amostrais (qui-quadrado, t e F)	143
9 Teoria da estimação	144
9.1 Introdução	144
9.2 Conceitos básicos	144
9.3 Estimação pontual	145
9.3.1 Métodos para obtenção de estimadores pontuais	146
9.3.2 Propriedades de um estimador pontual	154
9.4 Estimação por intervalo	154
9.4.1 Intervalo de confiança para a média	155
9.4.2 Intervalo de confiança para a variância	157
9.4.3 Intervalo de confiança para a diferença entre médias	158
9.5 Dimensionamento de amostras	158
9.6 Exercícios	159
10 Teoria da decisão	161
10.1 Introdução	161
10.2 Testes de hipóteses	162
10.3 Erros tipo I e II	163
10.4 Teste unilateral e bilateral	166
10.5 Passos para a construção de um teste de hipótese	166
10.6 Teste de hipóteses para a média	166
10.7 Testes de hipóteses para a proporção	166
10.8 Testes de hipóteses para a variância	166

10.9 Testes de hipóteses para a diferença entre médias	166
10.10 Exercícios	167
11 Correlação e regressão linear simples	169
11.1 Introdução	169
11.2 Correlação linear	170
11.2.1 Coeficiente de correlação linear	170
11.2.2 Teste de hipóteses acerca do coeficiente de correlação linear	170
11.3 Regressão linear simples	170
11.3.1 Modelo	171
11.3.2 Estimação dos parâmetros do modelo	171
11.3.3 Somas de quadrados	173
11.3.4 Teste de hipóteses para o modelo de regressão	174
11.3.5 Medidas de adequação do modelo	177
11.4 Exercícios	179
12 Apêndice A - Introdução ao R	180
13 Apêndice B - Tabelas	181
14 Apêndice C - Coordenadas polares	203
15 Gabarito dos Exercícios	206
Referências Bibliográficas	239
Índice Remissivo	240

Prefácio

Esse é um **livro digital** da 1^a edição do “Estatística & Probabilidade aplicado às Engenharias e Ciências”, um livro com o **selo Democratizando Conhecimento (DC)**. O Livro é voltado para quem deseja iniciar no estudo sobre a estatística. Daremos as bases e fundamentos, de modo aplicado a problemas na área das Engenharias e Ciências, de assuntos desde o que é uma população, amostra, até estudos sobre a teoria de decisão, estudo de regressão, dentre outros assuntos básicos, para que assim, a partir desse material, o leitor tenha base para ler, assuntos mais aprofundados na área da estatística.

Capítulo 1

Definições gerais da Estatística e técnicas de somatório

1.1 Introdução

Em pleno século XXI, passamos por um processo de transformação na era digital. Uma grande massa de informações surge instantaneamente a cada momento sobre os mais diversos temas possíveis. Por exemplo, nas redes sociais quando percebemos o número de curtidas de uma determinada declaração, número de downloads de um determinado vídeo, a repercussão que determinada declaração proporciona nas redes sociais, o número de propagandas, etc, tudo isso cria um grande banco de dados sobre os usuários, que hoje se torna mais valiosa do que a própria moeda local. Isso é a nova revolução chamada “Big Data”. Por meio de grande banco de dados, podemos por exemplo, traçar um perfil dos usuários, como eles se comportam, quais as suas preferências, escolhas, diversão, etc. Contudo, o entendimento dessas informações podem não ser tão claras, ou devido a complexidade do problema, ou pela quantidade de informações recebidas rapidamente, ou outros fatores. Diversos outros exemplos poderiam ser citados, tudo isso para mostrar a necessidade de entender o que está por trás desses dados, cuja compreensão é o grande objetivo nessa era global.

Nesse enfoque, temos a Estatística como Ciência que fornece métodos para coleta, organização, descrição, análise e interpretação de dados (observacionais ou experimentais) e para a utilização dos mesmos na tomada de decisões. Os dados são informações retiradas de um conjunto de elementos de interesse. Podemos estar interessados na Produção anual de Gás Natural não associado com o petróleo (GASN) de um determinado país, e ao longo dos anos, coletarmos informações para ao final, por exemplo, termos informações que nos indique o potencial energético desse recurso natural em um local ou consequências dessa fonte energética na economia do país, por exemplo.

Assim, por meio da utilização de técnicas estatísticas, tentamos entender as informações contidas nos dados. Devido a complexidade dessas informações em algumas situações, o estudo sobre essas técnicas têm aumentado, fazendo parte do nosso cotidiano. Na era digital, a grande quantidade de informações é gigantesca e valiosa, e as empresas tentam entender o que está por trás desses dados, ou você acha que o Facebook foi criado simplesmente para gerar entretenimento às pessoas? Ou você também acha que o Google criou uma plataforma de pesquisa simplesmente para facilitar a vida das pessoas? Algo muito nobre está por trás de tudo isso, os dados.

No passado, tratar uma grande massa de dados era uma tarefa custosa e cansativa, que exigia horas de trabalho tedioso. Porém, hoje, esse volume de informações pode ser analisado rapidamente por meio de um computador pessoal e programas adequados. O computador contribui positivamente na difusão e uso dos métodos estatísticos. Já se perguntou como é que lojas virtuais lhe oferecem produtos sendo que nunca acessou aquele site antes? Já percebeu que a Netflix quando lhe oferece uma série a tela de entrada, às vezes, se altera? Tudo isso é fruto das técnicas de máquinas de aprendizagem (do inglês, *Machine Learning*), uma área da inteligência artificial. Juntamente com a estatística, essas ferramentas estão presentes em várias das tecnologias que utilizamos hoje.

Por outro lado, o computador possibilita uma automação que pode levar um indivíduo sem preparo específico a utilizar técnicas inadequadas para resolver um dado problema. Assim, é necessário a compreensão dos conceitos básicos de Estatística, bem como as suposições necessárias para o seu uso de forma criteriosa. Com base nisso, tentaremos expor ao longo desse livro, as principais técnicas da Estatística, desde o seu fundamento teórico às aplicações.

1.2 Definições gerais da Estatística

Inicialmente, podemos dividir a Estatística em três ramos:

- Estatística Descritiva;
- Probabilidade;
- Estatística Inferencial.

Definimos cada uma dessas áreas a seguir. A primeira delas é a Estatística Descritiva, apresentada na Definição 1.1.

Definição 1.1: Estatística Descritiva ou Estatística Dedutiva

Um conjunto de técnicas estatísticas destinadas a coleta, descrição e sintetização dos dados, a fim de podermos entender características de interesse da população^a, é chamada de Estatística Descritiva ou Estatística Dedutiva.

^aVer Definição 1.6.

As técnicas mencionadas na Definição 1.1 são: coleta, organização, tabulação, representação gráfica, bem como medidas que sintetizam todas as informações contidas nos dados.

As quatro primeiras técnicas serão abordadas no Capítulo 2. As medidas serão estudadas nos Capítulos 3 e 4. Essa fase é de grande relevância, pois com base na Estatística Descritiva podemos sintetizar as informações contidas nos dados, e torná-las mais compreensíveis que de outro modo seriam complexas de serem entendidas.

No Capítulo 7, daremos uma maior ênfase sobre a definição de uma população. De modo simples, podemos definir como um conjunto de elementos com uma característica em comum. Uma vez definida a característica que delimita essa população (Ver Subseção 1.2.1), e também a característica de interesse (chamada de variável) da pesquisa, faremos a coleta dos valores observados da variável em cada elemento da população ou de um subconjunto (amostra). Os valores observados são chamados de dados.

Definição 1.2: Dado ou valor observado

A característica de interesse observada em cada elemento da população é definida como valor observado ou dado.

Todas as técnicas mencionadas anteriormente auxiliam na descrição dos dados, uma não necessariamente sobrepõe a outra. Vejamos, a Company (2018) lançou o relatório técnico de 2018 sobre os diversos tipos de produção de energias dos países, e na Figura 1.1 é mostrado um gráfico que sintetiza a produção e o consumo de petróleo do Brasil, em milhões de barris por dia (MMbbl/d), nos períodos de 2007 a 2017.

O gráfico nos revela que o Brasil produz petróleo abaixo do que necessitaria para o consumo, de tal modo que a produção é 27,71% a menos do que o consumo. Isso explica o porquê do Brasil como grande produtor de petróleo, ainda assim, necessita importar essa fonte de energia. Contudo, o gráfico não apresenta um resumo perfeito. Por exemplo, mesmo a produção de petróleo sendo mais baixa do que o consumo, o coeficiente de variação (assunto abordado no Capítulo 4) dessas duas variáveis, são 12,99% e 10,93%, respectivamente, calculados de acordo com a Tabela 1.1. Isso implica, que as variações da produção de petróleo são maiores do que as do consumo, no Brasil. Observe que essas últimas informações não podem ser vistas facilmente na Figura 1.1, mas juntamente com o auxílio das medidas numéricas (medidas de posição e dispersão) e as medidas gráficas, podemos complementar as informações, e assim, obter uma melhor descrição sobre essas informações.

Para tentar refazer o gráfico da Figura 1.1, usamos o Código R 1.1, apresentado em seguida.

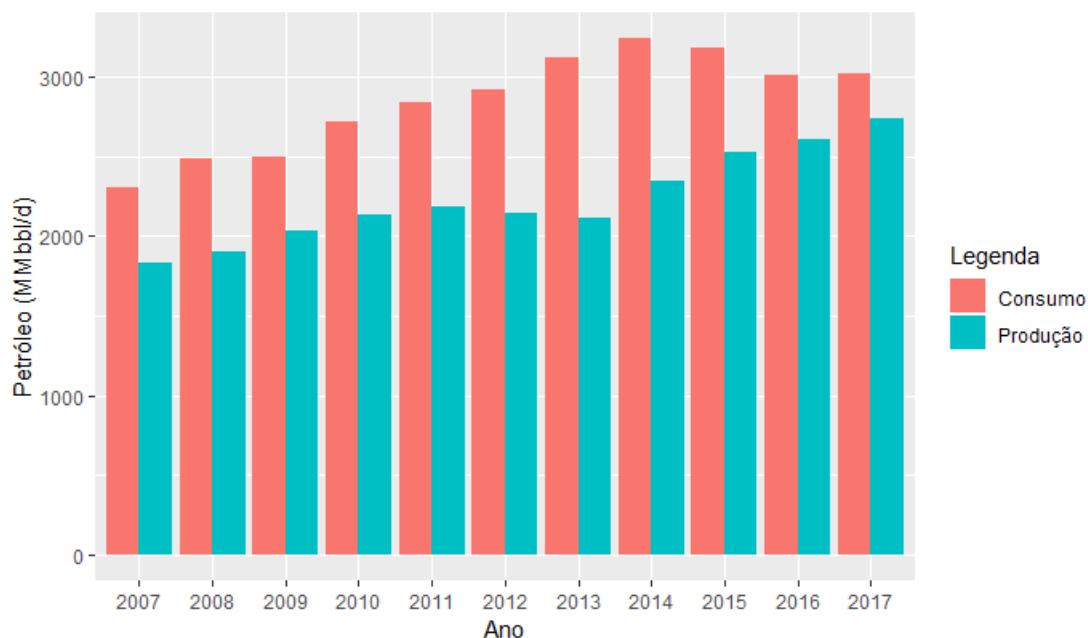


Figura 1.1: Produção e consumo de petróleo do Brasil nos períodos de 2007 a 2017.

Tabela 1.1: Produção e consumo de petróleo do Brasil nos períodos de 2007 a 2017.

Ano	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Produção	1831	1897	2029	2137	2179	2145	2110	2341	2525	2608	2734
Consumo	2308	2481	2498	2716	2839	2915	3124	3242	3181	3013	3017

Código R 1.1

Script:

```

1 # BP Statistical Review 2018
2
3 # Producao e consumo de petroleo no brasil (Milhares de barris por dia)
4 ano <- as.factor(c(2007:2017, 2007:2017))
5 prodcons <- c(1831, 1897, 2029, 2137, 2179, 2145, 2110, 2341, 2525, 2608,
       2734, 2308, 2481, 2498, 2716, 2839, 2915, 3124, 3242, 3181, 3013, 3017)
6 id <- c(rep("Produção", 11), rep("Consumo", 11))
7
8 # Objeto que armazena as informações
9 dados <- data.frame(ano, Legenda = id, prodcons)
10
11 # Pacote utilizado
12 install.packages("ggplot2")
13 library(ggplot2)
14
15 # Funcao para criacao do grafico de barras
16 ggplot(dados, aes(x=ano, y=prodcons, fill=Legenda)) +
17   geom_bar(position="dodge", stat="identity") +
18   xlab("Ano") + ylab(``Petróleo (Milhões de barris por dia)'')

```

Quando precisamos extender as informações contidas em um subconjunto (amostra) de dados para todo o conjunto (população), necessitamos de técnicas específicas dentro da estatística para garantir que estas informações sejam o mais verossímil possível. Técnicas estas são chamadas de

Estatística Inferencial, definida a seguir.

Definição 1.3: Estatística Inferencial ou Estatística Indutiva

O estudo de técnicas que visam extender (extrapolar) a informação contida na amostra à população, chamamos de Estatística Inferencial ou Estatística Indutiva.

As técnicas abordadas na Estatística Inferencial estão relacionadas a determinar características (parâmetros) populacionais desconhecidas, ou até mesmo fazer afirmações sobre esses parâmetros. A determinação de parâmetros por meio de características amostrais (estimadores) que chamamos de Estimação será abordado no Capítulo 9. As afirmações realizadas sobre estes parâmetros, chamadas de hipóteses, serão estudadas no Capítulo 10.

A Definição 1.3 nos mostra que por meio da Estatística, podemos tomar decisões sobre uma população através da amostra. Isso se faz necessário muitas vezes em uma pesquisa, devido a duas coisas preciosas: tempo e dinheiro. Muito embora, se tivermos acesso a todos os elementos de uma população, não se faz necessário o uso de técnicas da inferência estatística, e daí realizamos o que chamamos de Censo.

A forma de como se obter uma amostra é um dos passos mais importante em todo o processo da análise, uma vez que não adianta está com todo o aparato técnico se as informações contidas nessas amostra não são representativas da população. Para isso, temos uma área na estatística chamada Amostragem, que será responsável pelo desenvolvimento de métodos de como selecionar os elementos populacionais para compor a amostra de modo que as principais características contidas na população sejam preservadas na amostra. Esse assunto será estudado no Capítulo 7.

Contudo, sabemos que entender uma população por um subconjunto desta, gera-se uma incerteza ou erro. A estatística tenta minimizar esse erro o máximo possível, isto é, reduzir as incertezas das informações contidas na amostra e extrapolar essas informações para a população. Para isso, usamos a probabilidade como suporte, assunto estudado nos Capítulos 5, 6 e 8.

Definição 1.4: Probabilidade

A teoria matemática que estuda a incerteza de fenômenos aleatórios é chamada de probabilidade.

Os fenômenos aleatórios estão relacionados a situações que dificilmente saberemos com certeza do que pode acontecer. Por exemplo, se arremessarmos um dado de seis faces de tamanhos iguais e desejarmos saber a face superior desse dado antes do arremesso, não temos como afirmar com certeza qual o valor, se considerarmos as faces numerados de 1 a 6. Observe que, mesmo sabendo quais os valores das faces, não podemos afirmar com exatidão qual o valor da face superior antes do arremesso. Mas, por meio da probabilidade, podemos minimizar essa incerteza e dizer que há aproximadamente 17% de chance de um número escolhido ocorrer.

Em nosso cotidiano, a probabilidade auxilia na decisão de um fabricante de cola de empreender uma grande publicidade no seu produto visando aumentar sua participação no mercado, ou na decisão de parar de imunizar pessoas com menos de vinte anos contra determinada doença, ou ainda na decisão de arriscar-se a atravessar uma rua no meio do quarteirão. Esses pequenos exemplos mostram a relação que a probabilidade tem com a inferência estatística, pois ela nos auxiliará a tomar decisões em procedimento inferenciais tentando traduzir para a nossa linguagem do dia-a-dia.

Ao final dessas definições gerais, podemos mostrar uma ilustração, Figura 1.2, que facilitará a compreensão do que abordamos nessa seção. Por fim, um último assunto estudado nesse livro, Capítulo 11, será o estudo da correlação e regressão linear, quando estamos interessado em estudar a forma e o grau relação entre duas ou mais variáveis.

Todos esses assuntos estudaremos nos capítulos seguintes com um certo detalhamento, dando ênfase a exemplos práticos estudados em nosso campo de trabalho. Alguns Capítulos poderão conter uma seção chamada *Aprofundamento*, com o intuito de apresentar uma maior profundidade sobre o tema estudado. Alguns apêndices serão criados nesse livro para dar suporte ao conteúdo.

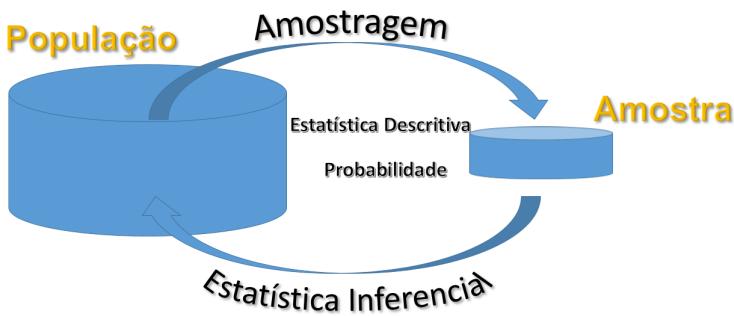


Figura 1.2: Ilustração animada sobre as disposições gerais da Estatística.

1.2.1 Estatística na pesquisa científica

O trabalho estatístico é parte integrante do método científico. Segundo Silva (2007), definimos,

Definição 1.5: Método científico

Um conjunto de regras e procedimentos para a obtenção de resultados, isto é, uma conclusão, sobre um determinado problema de uma pesquisa científica, é denominado de Método científico.

A pesquisa científica por sua vez, desenvolve conhecimentos para um saber mínimo de um determinado fenômeno estudado. A pesquisa científica se inicia a partir de um problema dentro da população em estudo. Por meio desse problema surgem diversas indagações.

Exemplo 1.1:

No Estado de Minas Gerais houveram dois acidentes, de grandes proporções nos últimos anos, envolvendo barragens que armazenam rejeitos de mineração. Os acidentes ocorreram na cidade de Mariana e Brumadinho, vitimando centenas de pessoas e um impacto ambiental imenso com o arrombamentos dessas barragens. Indagamos:

- *Quem* são os responsáveis por essas duas tragédias?
- *Quanto* será o custo o impacto dessas tragédias?
- *Quando* começou esse problema trágico?
- *Que* medidas poderiam ter sido tomadas para que isso não acontecesse?
- *Onde* estão os órgãos de fiscalização para coibir esses acontecimentos, uma vez que no intervalo de três anos ocorreram duas catástrofes dessas?

Com essas indagações lançadas para o estudo do problema, e definido a questão inicial dentre as citadas ou outras que possam surgir, o método científico se encarregará de estruturar a pesquisa de modo preciso e sistemático. A resposta a essas indagações resultam em um plano de pesquisa que consiste em:

1. Identificar o problema e o objetivo da pesquisa

A identificação do problema é o norte da pesquisa. Por meio das perguntas iniciais, procuraremos entender as possíveis causas e efeitos da situação, formulando assim o problema. Nessa fase, devemos identificar a população e os elementos que a compõe, como também os demais procedimentos da pesquisa, inclusive o objetivo do trabalho, no qual se estipula a finalidade do presente estudo. Esse passo é o combustível que impulsiona a pesquisa científica.

2. Formular a hipótese estudada

A hipótese é uma afirmação atribuída pelo pesquisador sobre a população, com o intuito de responder a(s) indagação(ões) do problema, atingindo o objetivo da pesquisa. Essa afirmação pode ser sugerida pela literatura ou até mesmo construída pelo próprio pesquisador. De toda forma, a elaboração dessa hipótese deve ser bem formulada para que sua não rejeição ou rejeição consiga responder a indagação inicial e o objetivo seja atingido, ou desencadeie novas dúvidas e outras pesquisas possam surgir.

Exemplo 1.2:

Tentando inspecionar, como controle de qualidade, uma remessa de peças enviadas por um fornecedor de uma determinada fábrica de peças para indústria de automóveis, garante que essa remessa não há mais de 8% de peças com defeito. Dessa forma o problema foi lançado com a seguinte indagação: será que essa remessa não há mais de 8% de peças com defeito? Lancemos a hipótese a ser testada^a:

$$\{ H_0 : \text{A porcentagem de peças com defeito é igual ou superior a } 8\%.$$

Perguntamos aos leitores: essa indagação elucida a indagação do problema? A avaliação dessa hipótese será realizada na *análise e interpretação de dados* do qual aplicaremos técnicas específicas para refutar ou não a hipótese estudada H_0 . Supondo que não tenhamos evidências estatísticas para a rejeição dessa hipótese, e decidimos não rejeitá-la, a dúvida que fica é: será que a hipótese estudada foi não rejeitada porque a quantidade de peças é igual ou superior a 8%? Observe, se foi 8% a afirmação inicial do fornecedor das peças está correta. Entretanto, se o número de peças foi superior a 8%, o que o fornecedor afirmou está equivocado. Concluímos, que a hipótese não foi bem elaborada para responder a indagação inicial no problema levantado. A forma correta deveria ser:

$$\{ H_0 : \text{A porcentagem de peças com defeito é menor ou igual a } 8\%.$$

A importância do desenvolvimento das hipóteses é muito importante, uma vez que podemos tomar decisões totalmente equivocadas, e assim, todo o trabalho estudado ser desperdiçado em vão.

^aA forma de como construir as hipóteses de uma pesquisa científica será abordada no Capítulo 10, do qual dissertaremos sobre a Teoria da decisão.

3. Revisão de literatura

Um passo importante na pesquisa é a confirmação ou o não das respostas encontradas no estudo. É por meio, dos trabalhos já publicados que embasamos nossas argumentações, corroborando-as ou refutando-as. Com isso, surge o progresso da ciência, não havendo uma verdade absoluta.

4. Formular um Plano amostral e Identificar as variáveis de interesse para a pesquisa (Capítulo 7).

5. Coleta, crítica e tratamento dos dados

Após definirmos cuidadosamente o problema que se quer pesquisar, elabora-se um delineamento e damos início à coleta dos dados necessários à sua descrição. Obtidos os dados, eles devem ser cuidadosamente criticados, à procura de possíveis falhas e imperfeições, a fim de não incorrermos em erros grosseiros ou certo vulto, que possam influir sensivelmente os resultados. Por fim, o tratamento dos dados que consiste no processamento dessas informações e a disposição mediante critérios de classificação, Podendo ser manual ou eletrônica.

6. Apresentação dos dados

Por mais diversa que seja a finalidade, os dados devem ser apresentados sob forma adequada (tabelas e gráficos) tornando o mais fácil e simples a sua descrição.

7. Análise e interpretação dos resultados

Após a apresentação dos dados devemos calcular as medidas típicas convenientes para fazermos uma análise dos resultados obtidos, através de métodos estatísticos (Estatística inferencial ou indutiva), e tirarmos desses resultados conclusões e previsões.

8. Conclusão e derivação da conclusão que poderá rejeitar ou não a hipótese estudada, gerando assim, uma confirmação ou indagações para outros problemas

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

9. Apresentação dos resultados por meio de trabalhos científicos para a propagação do conhecimento sobre o problema estudado.

Esses pontos do plano de pesquisa podem sofrer alterações em algumas metodologias científicas. Contudo, elas estão envolvidas direta ou indiretamente nas metodologias estudadas, sendo que não necessariamente elas ocorrem em todas as pesquisas nessa ordem, e que podem ser ilustrados na Figura 1.3.

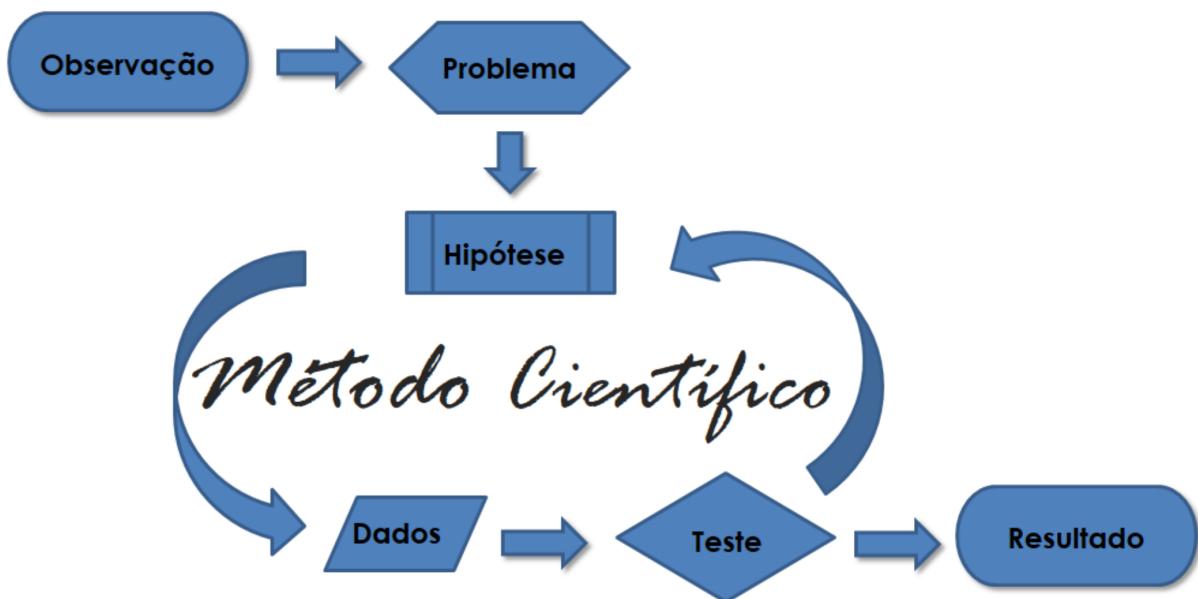


Figura 1.3: Fases do plano da pesquisa científica.

1.3 Definições básicas

Ao ser discutido na seção anterior sobre as definições gerais da Estatística, iniciaremos agora ao que chamamos de definições básicas, que consiste em definir formalmente alguns termos tais como população e amostra, como também os termos variável, dado ou valor observado. Essas definições serão importantes para o desenvolvimento do conteúdo do livro.

O conjunto de todos os elementos dos quais temos o interesse de suas informações, chamamos esse conjunto de população. A palavra população, em nosso cotidiano, está sempre relacionado a um conjunto de pessoas que habitam um determinado local (país, cidade, etc.). Contudo, na estatística ampliamos a definição de população da seguinte forma,

Definição 1.6: População

O conjunto finito ou infinito de todos os elementos com pelo menos uma característica comum, dos quais é de interesse para a pesquisa, denominamos de População. O número de elementos é denominado tamanho da população, denotado por N .

Percebemos pela Definição 1.6 que a idéia sobre população é mais geral. Podemos dizer que o conjunto de peças com defeitos fabricados por uma determinada empresa constitui uma população. Um outro exemplo é a concentração de metais pesados no Rio, sendo que o rio constitui a população. No primeiro caso, a população constitui a empresa que fabrica essas peças com defeitos, que por sua vez, essas peças representam os elementos dos quais a característica em comum a todas as peças é que foi fabricada por essa empresa e apresenta defeito. No segundo caso, a especificação dos elementos poderá não ser muito claro, pois é um caso de população infinita. Daremos mais detalhes sobre isso no Capítulo 7.

Essa(s) característica(s) comum(s) deve(m) delimitar inequivocamente quais elementos que pertencem ou não à população. A notação usual para o número de elementos da população é " N ". A população pode ser *Finita* (quando pode ser enumerada) ou *Infinita* (quando não pode ser enumerada).

Definição 1.7: Amostra

Um subconjunto de elementos da população é denominado amostra. O número de elementos da amostra é chamado de tamanho da amostra, sendo denotado por " n ".

A amostra é necessariamente finita, pois todos os seus elementos serão examinados para efeito da realização do estudo estatístico desejado. Esse estudo está baseado em características de interesse da população para tentar responder as indagações iniciais do problema da pesquisa (Ver Seção 1.2.1). Definimos essa característica como variável.

Definição 1.8: Variável

A característica pela qual desejamos que a população seja descrita é denominada de variável.

A variável representa o mecanismo pelo qual podemos atingir o objetivo da pesquisa. Será por meio dos dados observados, isto é, do valor observado dessa variável assumido por cada elemento da população (ou da amostra), que faremos as análises específicas para se chegar a uma conclusão. Muitas vezes não trabalhamos apenas com uma única variável, dependendo da complexidade da pesquisa, poderemos estudar diversas variáveis ao mesmo tempo.

A variável pode assumir diferentes valores de elemento para elemento, chamado de dado ou valor observado, como foi apresentado na Definição 1.2. A notação usual para a variável é X, Y, Z , ou X_i, Y_i, Z_i para um particular elemento amostral, em que $i = 1, 2, \dots, n$. Definimos a natureza das variáveis, a seguir.

Definição 1.9: Natureza de uma variável

Definimos o tipo de variável pela sua natureza, isto é, pelo valor assumido em cada elemento da população ou amostra como:

- 1) **Variável qualitativa:** é a variável cujo valor observado assume um valor com natureza de atributo ou categoria. Esta ainda se subdivide:
 - a) **Nominal:** Quando os valores não são possíveis de ordenação;
 - b) **Ordinal:** Quando os valores são possíveis de ordenação, segundo um critério quantitativo.

2) **Variável quantitativa:** é a variável cujo valor observado assume um valor com natureza numérica (enumerável ou não). Ainda podem ser divididas:

- a) **Discreta:** Quando os valores são dados de contagem, isto é, descrevem uma quantidade contável, cujos potenciais valores dessa variável podem ser enumerados em um conjunto de valores;
- b) **Contínua:** Quando os valores resultam de uma medida (ou mensuração), podendo assumir qualquer valor real entre dois extremos, e dessa forma não podemos enumerar seus valores.

Vejamos o Exemplo 1.3, para elucidar todas essas definições mencionadas anteriormente.

Exemplo 1.3: Desmatamento da Amazônia Legal

O Brasil vem passando por um processo de desmatamento na Amazônia legal, que o mundo vem acompanhando nesses últimos anos. O Instituto Nacional de Pesquisas Espaciais (INPE) desenvolveu o PRODES¹ (Programa de Monitoramento do Desmatamento da Amazônia) que vem acompanhando desde 1988, as taxas de desmatamento na região. Apresentamos a seguir, a Tabela 1.2, que apresenta algumas informações sobre esse tema em cada estado do qual compreende a Amazônia Legal, em que os dados da taxa acumulada de desmatamento por estado, estão disponíveis na página do INPE, em <http://terrabrasilis.dpi.inpe.br/app/dashboard/deforestation/biomes/legal_amazon/rates>.

Podemos verificar que apresentamos diversas variáveis para um melhor detalhamento da taxa de desmatamento na Amazônia legal, em relação a algumas informações sobre os estados que compõe essa região. Observemos que, quanto a natureza das variáveis, Região, UF, e Classificação, são variáveis qualitativas, pois os valores observados representam uma categoria. Apesar de representar uma qualidade, percebemos que classificação apresenta um ordenamento, referente a quantidade de desmatamento acumulado de cada estado, em que, o estado do Pará está em primeiro lugar, por ter sido o estado que mais desmatou, desde 1988. No caso, as variáveis Região e UF representam apenas categorias que não apresentam qualquer forma de ordenamento. Muito embora, os valores de Classificação estejam representados por números, a natureza é qualitativa. Portanto, Região e UF são variáveis qualitativas nominais, e Classificação, variável qualitativa ordinal.

As demais variáveis são Número de cidades, Desmatamento acumulado, Área total e População estimada. Conseguimos observar que Número de cidades e População estimada apresentam dados de contagem, logo, essas variáveis são quantitativas discretas. Isso significa, que entre dois valores consecutivos, há uma discretização, ou seja, o estado do Amapá está dentro da região da Amazônia legal, e tem 14 município. Já o estado de Roraima apresenta 15 municípios dentro da Amazônia Legal. Dessa forma, não há um potencial valor para a variável Número de Cidades, entre esses dois valores, isto é, 14,5. De outro modo, podemos ordenar em um conjunto enumerável todos os valores de uma variável quantitativa discreta.

Agora, para o caso das variáveis Desmatamento e Área total, percebemos que os valores assumidos por essas variáveis não são dados de contagem, mas de medição, isto significa que não contamos área ou taxa de Desmatamento acumulado, mas sim, medimos. De outro modo, teoricamente nós não conseguimos identificar os potenciais valores de uma variável quantitativa contínua em certo certo conjunto enumerável, porque observe o valor da área total do estado do Pará, 1.245.870,00 km^2 , se tivéssemos instrumentos de medidas mais precisos, esse valor não seria exatamente esse, poderia ter sido 1.245.870,001 km^2 , 1.245.870,0001 km^2 , 1.245.870,00001 km^2 , e assim por diante. Dessa forma, em uma determinada ordem nós não conseguiríamos saber qual o próximo valor ordenado para a área, após observarmos a área do estado do Pará.

Tabela 1.2: Taxa de desmatamento acumulado, por estado, na Amazônia Legal, compreendido desde 1988 a 07/12/2020.

Região	UF	Nº de cidades ²	Desmat. acum. (km ²)	Área total (km ²)	Clas. ³	Pop. estimada ⁴
Norte	Pará	144	157.667,00	1.245.870,00	1º	8.690.745
Centro-Oeste	Mato Grosso	141	147.926,00	903.207,02	2º	3.526.220
Norte	Rondônia	52	62.936,00	237.765,20	3º	1.796.460
Norte	Amazonas	62	28.493,00	1.559.167,89	4º	4.207.714
Nordeste	Maranhão	181	25.707,00	276.419,84	5º	7.114.598
Norte	Acre	22	15.725,00	164.123,96	6º	894.470
Norte	Tocantins	139	8.727,00	277.466,76	7º	1.590.248
Norte	Roraima	15	8.597,00	223.644,53	8º	631.181
Norte	Amapá	14	1.696,00	142.470,76	9º	861.773

Pensando sobre uma variável



- Uma variável originalmente quantitativa pode ser coletada de forma qualitativa. Por exemplo, a variável idade, medida em anos completos, é quantitativa (discreta). Porém, se considerarmos uma nova variável como faixa etária, do qual os valores possíveis são: “criança” (0 a 12 anos), “adolescente” (12 a 18 anos), “adulto” (18 a 60 anos) e “idoso” (acima de 60 anos), cujos valores foram originais da variável idade estão entre parênteses, a variável faixa etária passa a ser considerada uma variável qualitativa (ordinal). Outro exemplo é o peso dos lutadores de boxe, uma variável quantitativa (contínua) se trabalhamos com o valor obtido na balança, mas qualitativa (ordinal) se o classificarmos nas categorias do boxe (peso-pena, peso-leve, peso-pesado, etc.);
- Outro ponto importante é que nem sempre uma variável representada por números é quantitativa. O número do telefone de uma pessoa, o número da casa, o número de sua identidade. Às vezes o sexo do indivíduo é registrado na planilha de dados como 1 se macho e 2 se fêmea, por exemplo. Isto não significa que a variável sexo passou a ser quantitativa!
- É fato que a rigor, as variáveis quantitativas seriam todas discretizadas devido a limitação dos nossos instrumentos de medidas. Observe que o tamanho de uma pessoa não está limitado às escalas de metros, centímetros, milímetros, etc.. Porém, os instrumentos de medida que obtém essa informação, estão limitados a essas escalas. De toda forma, precisamos dessa limitação para que as análises sejam possíveis, e possamos tomar decisões a partir dos dados.

¹Para quem desejar entender com detalhes a metodologia baseada para o programa, acesse: <<http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>>.

²Dados coletados da página do IBGE, edição 2019, <https://geoftp.ibge.gov.br/organizacao_do_territorio/estrutura_teritorial/amazonia_legal/2019/lista_de_municipios_da_amazonia_legal_2019.ods>

³Essa variável se refere a classificação do estado que obteve maior taxa de desmatamento acumulado, desde 1988 a 2020.

⁴Essa variável se refere a população estimada de cada estado, e os dados foram retirados do IBGE, disponível em <<https://www.ibge.gov.br/cidades-e-estados>>

1.4 Técnicas de Somatório

Um tipo de notação muito importante para a Estatística é o uso de técnicas de somatório, muito usado, por exemplo, na notação de medidas estatísticas. A ideia da técnica de somatório é simplificar a notação da soma de dados, de modo que possamos representar essas operações por meio de notação matemática de modo simplificado.

Como já falado anteriormente, representamos por X uma determinada variável. Nesse caso, não fará sentido falar de variáveis qualitativas, uma vez que o objetivo nessa notação é a representação de operações matemáticas. Assim, estaremos restritos as variáveis quantitativas.

Baseado nos dados da Tabela 1.2, supomos que X representa o número de cidades pertencentes a Amazônia legal de um terminado estado, então nesse caso, como temos a representação de todos os estados, estamos diante de dados populacionais. Assim, $N = 9$, podemos representar a variável com um índice para se referir ao número de cidade de um determinado estado. Por exemplo, X_1 representa a variável número de cidades do Pará, o seu valor observado $x_1 = 144$ cidades. A variável X pode ser representada nos demais elementos da seguinte forma: X_1, X_2, \dots, X_9 . Podemos estar interessados em saber o total de cidades na Amazônia legal. Em notação, podemos calcular esse total da seguinte forma:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 &= 144 + 141 + 52 + 62 + 181 + 22 + 139 + 15 + 14 \\ &= 770 \text{ cidades.} \end{aligned}$$

Percebemos que com apenas nove observações, a notação para essa simples operação se torna extensa. E isso acaba aumentando à medida que o número de observações aumentam. Também, quando realizamos operações mais complexas, essa representação também se tornam mais complexas. Pensando nisso, surgem as técnicas de somatórios, para simplificar essas representações. Representaremos um somatório pela letra grega sigma maiúsculo (Σ), que indica a soma de determinados valores. Agregado ao símbolo do somatório, usaremos uma (ou algumas) indexação(ões) para representar qual(is) os elementos fazem parte desta operação, seguido da(s) variável(is) de interesse, isto é,

$$\sum_{i=1}^m X_i = X_1 + X_2 + \dots + X_m,$$

sendo que m pode representar o tamanho amostral, n , ou o tamanho populacional, N . No caso da representação anterior, podemos simplificar a representação da soma de um conjunto de valores usando as técnicas de somatório, apresentadas a seguir,

$$X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 = \sum_{i=1}^9 X_i.$$

Tornamos a notação mais simples de ser representada. Isso será muito importante, quando formos definir medidas estatísticas, provas de teoremas, etc.

De modo similar, podemos realizar as mesmas alterações com transformações na(s) variável(is). Por exemplo, quando formos estudar medidas de dispersão, no Capítulo 4, será útil as seguintes operações, considerando uma amostra de tamanho n ,

$$\sum_{i=1}^n X_i^2 = X_1^2 + X_2^2 + \dots, X_n^2,$$

isto é, a soma do quadrado da variável. Outra operação interessante, é o quadrado da soma, apresentado a seguir,

$$\left(\sum_{i=1}^n X_i \right)^2 = (X_1 + X_2 + \dots, X_n)^2. \quad (1.1)$$

Uma técnica muito utilizada na Estatística é o estudo da Regressão linear, que estuda a relação entre duas ou mais variáveis, e será abordado no Capítulo 11. Assim, uma das operações utilizadas é a soma do produto entre duas variáveis, por exemplo, X e Y , do qual podemos representar esta soma para um conjunto de pares (X_i, Y_i) , de tamanho n , da seguinte forma,

$$\sum_{i=1}^n X_i Y_i = X_1 Y_1 + X_2 Y_2 + \dots + X_n Y_n. \quad (1.2)$$

Outra forma é o produto das somas de variáveis, que nesse caso, nos limitaremos as duas variáveis X e Y . Por exemplo, para um amostra de pares (X_i, Y_i) de tamanho n , pode ser representada por:

$$\begin{aligned} \left(\sum_{i=1}^n X_i \right) \times \left(\sum_{i=1}^n Y_i \right) &= (X_1 + X_2 + \dots + X_n) \times (Y_1 + Y_2 + \dots + Y_n) \\ &= \sum_{i=1}^n X_i Y_i + \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n X_i Y_j, \end{aligned} \quad (1.3)$$

em que o primeiro dessa último resultado é dado pela notação expressa em (1.2). O segundo termo, apresenta uma nova notação que é o duplo somatório. A ideia dessa notação é simples, fixaremos o índice no primeiro somatório e percorremos a soma dos valores usando o segundo índice. Após ter realizado toda a operação, passaremos para o próximo índice no primeiro somatório e realizamos o mesmo procedimento para o índice no segundo somatório. Toda a operação será finalizada, quando tivermos percorrido a soma em todos os valores. Vejamos para um caso de duplo somatório, com um par de variáveis (X_i, Y_i) , para $n = 3$,

$$\begin{aligned} \sum_{i=1}^3 \sum_{j=1}^3 X_i Y_j &= \sum_{j=1}^3 X_1 Y_j + \sum_{j=1}^3 X_2 Y_j + \sum_{j=1}^3 X_3 Y_j \\ &= (X_1 Y_1 + X_1 Y_2 + X_1 Y_3) + (X_2 Y_1 + X_2 Y_2 + X_2 Y_3) + (X_3 Y_1 + X_3 Y_2 + X_3 Y_3). \end{aligned}$$

Para uma amostra de tamanho n , podemos generalizar essa notação da seguinte forma,

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n X_i Y_j &= \sum_{j=1}^n X_1 Y_j + \sum_{j=1}^n X_2 Y_j + \dots + \sum_{j=1}^n X_n Y_j \\ &= (X_1 Y_1 + X_1 Y_2 + \dots + X_1 Y_n) + (X_2 Y_1 + X_2 Y_2 + \dots + X_2 Y_n) + \dots \\ &\quad + (X_n Y_1 + X_n Y_2 + \dots + X_n Y_n). \end{aligned} \quad (1.4)$$

Porém, observe que o resultado em (1.4), soma todos os produtos X e Y . Como desejamos fazer uma relação entre a expressão (1.3) e a expressão (1.2), separamos a soma de produtos X e Y com índices iguais das demais situações. Para isso, impomos a restrição no segundo termo, depois da igualdade na expressão (1.4), para enfatizar que somaremos o produto de todos os $X_i \times Y_i$, tais que $i \neq j$, resultando na expressão (1.3).

Podemos também representar a notação $\sum_{i=1}^n \sum_{j=1}^n X_i Y_j$ da seguinte forma,

$$\sum_{i=1}^n \sum_{j=1}^n X_i Y_j = \sum_{\substack{i=1 \\ j=1}}^n X_i Y_j. \quad (1.5)$$

Uma outra notação que pode ser apresentada para o duplo somatório é abrindo o quadrado da

soma no resultado da expressão (1.1), dada da seguinte forma,

$$\begin{aligned}
 (X_1 + X_2 + \dots + X_n)^2 &= (X_1 + X_2 + \dots + X_n) \times (X_1 + X_2 + \dots + X_n) \\
 &= X_1 X_1 + X_1 X_2 + \dots + X_1 X_n + \\
 &\quad + X_2 X_1 + X_2 X_2 + \dots + X_2 X_n + \dots \\
 &\quad \dots + X_n X_1 + X_n X_2 + \dots + X_n X_n \\
 &= \sum_{i=1}^n X_i^2 + 2 \sum_{j>i=1}^n X_i X_j.
 \end{aligned} \tag{1.6}$$

Nesse caso, impomos também uma outra restrição no somatório do segundo termo após a igualdade, que foi somar todos os produtos $X_i \times X_j$ s, exceto aqueles com ele mesmo. Assim, observamos que situações do tipo $X_1 \times X_2 = X_2 \times X_1$, e desse modo, podemos representar $X_1 \times X_2 + X_2 \times X_1 = 2X_1 X_2$ que o resultado será o mesmo, e simplifica a notação. Generalizando a soma para os demais casos, temos $2 \sum_{j>i=1}^n X_i X_j$, como verificado na expressão (1.6).

Por fim, queremos enfatizar uma última situação que é usar um indexador não como a identificação da variável para um determinado elemento da população ou amostra, mas como valor observado. Essas situações serão muito utilizadas em notações no Capítulo 5 e 6, do qual somaremos as probabilidades da variável assumir valores em um determinado conjunto. A ideia de variável nesses capítulos será entendida como uma função, mas isso é assunto mais para frente. Nesse caso, vamos entender que $P(\cdot)$ é uma função que mede a chance de determinado X assumir um determinado valor, isto é, $P(\cdot)$ assume um valor entre 0 e 1. Essa função será chamada mais a frente de probabilidade. Assuma também que os valores possíveis de X assumam valores em um conjunto $A = \{1, 2, 3, 4, 5\}$, e estamos interessados em representar a chance de X assumir valor 3. Nesse caso, usamos $P(X = 3)$. Agora desejarmos representar a chance de X assumir valores, no mínimo, igual a 3. Dessa forma, representamos essa chance da seguinte forma,

$$P(X \geq 3) = \sum_{x=3}^5 P(X = x) = P(X = 3) + P(X = 4) + P(X = 5). \tag{1.7}$$

Observamos que o indexador no somatório agora é o valor assumido pela variável, e não a identificação da variável a um determinado elemento da amostra ou população. Se desejarmos somar todas as chances que X assume, podemos apresentar duas notações diferentes, apresentadas na sequência,

$$\sum_{x=1}^5 P(X = x) = \sum_{x \in A} P(X = x). \tag{1.8}$$

O índice no somatório indica agora que iremos somar as chances de X assumir todos os valores pertencentes ao conjunto A . Claro que, muitas outras formas de apresentar as técnicas de somatório podem ocorrer ao longo do texto, uma vez que outras formas podem ser abordadas, dependendo do assunto, como também da área estudada. De todo modo, tentamos passar parte da notação que será utilizada ao longo do livro, para que o leitor possa se ambientar nesse tipo de representação matemática.

Para complementar essas informações, o Teorema 1.1 apresenta algumas propriedades sobre técnicas de somatório que serão importantes para os próximos capítulos.

Teorema 1.1: Propriedades de somatório

Considere a , b e k constantes, e que X e Y são variáveis quantitativas, então as seguintes propriedades envolvendo somatório são válidas:

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

$$\text{II)} \sum_{i=1}^n X_i Y_i \leq \sum_{i=1}^n X_i \sum_{i=1}^n Y_i;$$

$$\text{III)} \sum_{i=1}^n (aX_i \pm bY_i) = a \sum_{i=1}^n X_i \pm b \sum_{i=1}^n Y_i;$$

$$\text{IV)} \sum_{i=1}^n k = nk;$$

$$\text{V)} \sum_{i=1}^n (X_i - \bar{X}) = 0, \text{ em que } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i;$$

$$\text{VI)} \sum_{i=1}^n X_i^2 \leq \left(\sum_{i=1}^n X_i \right)^2;$$

$$\text{VII)} n\bar{X}^2 = \frac{(\sum_{i=1}^n X_i)^2}{n}, \text{ em que } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i;$$

$$\text{VIII)} \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - \frac{1}{n} \sum_{i=1}^n X_i^2;$$

$$\text{IX)} \sum_{i=1}^n Y_i(X_i - \bar{X}) = \sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})$$

Prova:

$$(i) \sum_{i=1}^n aX_i = aX_1 + aX_2 + \dots + aX_n = a(X_1 + \dots + X_n) = a \sum_{i=1}^n X_i;$$

(ii) Observando as expressões (1.2) e (1.3), claramente que $\sum_{i=1}^n X_i Y_i < \sum_{i=1}^n X_i \sum_{i=1}^n Y_i$. A única condição de igualdade acontece se $n = 1$. Porém, em termos práticos para contexto estatístico, essa informação é inútil, uma vez que com apenas uma observação na amostra ou população não haverá condições apresentarmos alguma informação sobre a mesma.

(iii) Segue,

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

$$(iv) \sum_{i=1}^n k = \underbrace{k + k + \dots + k}_{n \text{ vezes}} = nk;$$

(v) Segue,

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

(vi) Verificando a expressão (1.6), percebemos claramente que $\sum_{i=1}^n X_i^2 < \left(\sum_{i=1}^n X_i \right)^2$. A única condição de igualdade acontece se $n = 1$, e em termos práticos para uso estatístico, usamos a mesma justificativa dada na propriedade (I);

$$(vii) n\bar{X}^2 = n \left(\frac{\sum_{i=1}^n X_i}{n} \right)^2 = \frac{(\sum_{i=1}^n X_i)^2}{n};$$

(viii) Vejamos a seguinte dedução,

$$\begin{aligned}\sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + \sum_{i=1}^n \bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - 2\frac{1}{n} \sum_{i=1}^n X_i \times \sum_{i=1}^n X_i + n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - 2\frac{(\sum_{i=1}^n X_i)^2}{n} + n\frac{(\sum_{i=1}^n X_i)^2}{n^2} \\ &= \sum_{i=1}^n X_i^2 - 2\frac{(\sum_{i=1}^n X_i)^2}{n} + \frac{(\sum_{i=1}^n X_i)^2}{n} \\ &= \sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n};\end{aligned}$$

(ix) Antes de mostrarmos a prova da propriedade (IX), vejamos que

$$\sum_{i=1}^n \bar{Y}(X_i - \bar{X}) = \bar{Y} \underbrace{\sum_{i=1}^n (X_i - \bar{X})}_{=0, \text{ Propriedade (V)}} = 0.$$

Desse modo, temos que

$$\sum_{i=1}^n Y_i(X_i - \bar{X}) = \sum_{i=1}^n Y_i(X_i - \bar{X}) - \underbrace{\sum_{i=1}^n \bar{Y}(X_i - \bar{X})}_{=0, \text{ Propriedade (VIII)}},$$

logo,

$$\sum_{i=1}^n Y_i(X_i - \bar{X}) = \sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X}).$$

Exercícios propostos

Exercício 1.1: Baseado no Exemplo 1.1, identifique um problema para essa pesquisa bem como um objetivo, e desenvolva hipóteses a serem estudadas, de modo que estas possam responder as indagações do problema e atinja o objetivo proposto.

Solução na página 206

Exercício 1.2: Usando os resultados do Teorema 1.1, mostre em quais aplicações na estatística poderemos utilizar esses resultados.

Solução na página 206

Exercício 1.3: Baseado em Devore (2006), um famoso experimento executado em 1882, Michelson e Newcomb fizeram 66 observações do tempo levado pela luz para percorrer a distância entre dois locais em Washington, D.C. Algumas das medidas (codificadas de certa forma) foram 31, 23, 32, 36, -2, 26, 27 e 31. Por que essas medidas não são idênticas?

Solução na página 206

Exercício 1.4: Sejam as amostras de tamanho $n = 5$ de duas variáveis, dadas por:

$$\begin{aligned} X &= \{2, 4, 5, 1, 2\}, \\ Y &= \{1, 2, 3, 5, 8\}. \end{aligned}$$

Obtenha:

- | | | |
|---|---|--|
| a) $\sum_{i=1}^4 X_i;$ | h) $\sum_{i=1}^n Y_i;$ | o) $\sum_{i=1}^n (Y_i - \frac{\sum_{i=1}^n Y_i}{n})^2;$ |
| b) $\sum_{i=1}^5 4 \times X_i^2;$ | i) $\sum_{i=1}^n X_i^2;$ | p) $\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n};$ |
| c) $\sum_{i=2}^n X_i;$ | j) $\sum_{i=1}^n Y_i^2;$ | q) $\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n};$ |
| d) $\sum_{i=1}^n X_i \times Y_i;$ | k) $\sum_{i=1}^n (Y_i X_i);$ | r) Qual conclusão se pode chegar sobre os itens (n) e (p), bem como (o) e (q)? |
| e) $\sum_{i=1}^n (3X_i + 2Y_i);$ | l) $(\sum_{i=1}^n Y_i)^2;$ | |
| f) $\sum_{i=1}^n X_i Y_i + \sum_{i=1}^n Y_i^2;$ | m) $(\sum_{i=1}^n X_i)^2;$ | |
| g) $\sum_{i=1}^n X_i;$ | n) $\sum_{i=1}^n (X_i - \frac{\sum_{i=1}^n X_i}{n})^2;$ | |

Solução na página 207

Exercício 1.5: Forneça uma amostra possível, de tamanho 5, de cada uma das populações a seguir:

- a) todos os jornais publicados no Brasil;
- b) todas as empresas na área de telecomunicações;
- c) todos os alunos da Universidade Federal de São João del-Rei;
- d) todas as notas, pontuados de 0 a 100, dos alunos da disciplina de Estatística e Probabilidade;

Solução na página 207

Exercício 1.6: Observou-se o tempo, em minutos, que 10 atendimentos de clientes da empresa telefônica A demoraram para serem atendidos, que seguem: 5, 10, 2, 13, 7, 15, 8, 12, 6 e 5. O objetivo do estudo foi verificar se o tempo médio, em minutos, do atendimento era superior a 10 minutos. Pergunta-se:

- Qual a população em estudo?
- Qual o problema indagado?
- Qual(is) a(s) variável(is) em estudo do trabalho, como também a natureza dessa(s) variável(is)?
- Podemos identificar o tamanho da população e da amostra, com essas informações?

Solução na página 207

Exercício 1.7: Como podemos relacionar os ramos da Estatística com os ítems do plano de pesquisa científica, presentes nesse capítulo?

Solução na página 208

Exercício 1.8: Os dados retirados de Tavares e Anjos (1999), representam a distribuição percentual do estado nutricional em homens idosos brasileiros (idade ≥ 60 anos), segundo Índice de Massa Corporal (IMC^a), por macrorregião e situação de domicílio, Pesquisa Nacional sobre Saúde e Nutrição, 1989, que seguem,

Regiões	Número	Estado Nutricional (%) ^b			
		Magreza	Adequado	Sobre peso I	Sobre peso II e III
Norte	223	4,4	60,6	29,4	5,6
Nordeste	586	8,8	68,3	19,8	3,1
Urbano	267	7,1	62,3	26,6	4,0
Rural	319	10,7	74,6	12,5	2,2
Sudeste	46,3	7,9	59,0	26,7	6,4
Urbano	197	5,6	56,4	30,2	7,8
Rural	232	6,4	66,4	22,0	5,2
Sul	429	5,1	56,5	29,2	9,2
Urbano	197	4,5	51,2	33,0	11,3
Rural	232	6,4	66,4	22,0	5,2
Centro-Oeste	327	10,7	60,6	22,8	5,9
Urbano	154	10,6	55,2	27,3	6,9
Rural	173	11,0	71,4	13,7	3,9
Brasil	2.028	7,8	61,8	24,7	5,7
Urbano	1.038	6,0	57,2	29,5	7,3
Rural	990	11,7	71,7	14,2	2,4

Como poderíamos, em notação usando as técnicas de somatório, representar a soma de todos os valores de IMC do Brasil, levando em consideração as demais variáveis? Se desejássemos, calcular o total dos valores observados de IMC dos homens do nordeste, considerando as

demais condições? E se fosse do nordeste e da zona urbana, como representaríamos esse somatório?

Solução na página 208

^aA unidade de IMC em kg/m^2 .

^bA classificação do estado nutricional em relação ao IMC foi: Magreza (todas as formas - $IMC < 18,5$); adequado ($18 \leq IMC < 25,0$); sobrepeso I ($25 \leq IMC < 30,0$); sobrepeso I e II ($IMC \geq 30,0$).

Exercício 1.9:

Considere a expressão $\sum_{i=1}^n (X_i - A)^2$. Qual o valor de A para que essa expressão seja minizada?

Solução na página 209

Capítulo 2

Coleta, organização e apresentação dos dados

2.1 Introdução

Após selecionado a população de interesse, definindo os elementos que a compõe, bem como as variáveis que serão estudadas, fazemos o processo de coleta dos dados. Os dados são os valores assumidos de uma variável em um determinado elemento da população, que pode estar sendo estudado por meio de uma amostra ou coletado diretamente da população. Neste último caso, a pesquisa realizada é um Censo.

Ao termos um primeiro contato com os dados, percebemos que algumas informações prévias podem não ser facilmente obtidas, devido a desorganização dessas observações. Isso ocorre principalmente quando temos um grande número de dados.

Definição 2.1: Dados brutos

Os dados coletados numa forma sem ordenação e sem nenhum tipo de arranjo sistemático são chamados dados brutos.

Os dados da Tabela 2.1, retirado de Montgomery e Runger (2016, p. 188), representam o número de erros em um conjunto de caracteres (*strings*) de 1.000 *bits*, que foram monitorados por um canal de comunicação. No total, foram coletados dados de 20 conjuntos de caracteres.

Tabela 2.1: Dados brutos sobre o número de erros encontrados em 20 conjuntos de caracteres monitorado em um canal de comunicação.

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

Podemos observar pela Tabela 2.1, que estes representam um tipo de dados brutos, pois não há qualquer ordenamento sobre os seus valores, e que a interpretação desses dados poderá se complicar à medida que o tamanho da amostra aumenta. Quando ordenamos os dados brutos podemos obter algumas informações mais facilmente, como por exemplo, valores mínimos e máximos desses conjunto de dados.

Definição 2.2: Dados em rol ou elaborados

Os dados brutos, Definição 2.1, ordenados de modo crescente ou decrescente alfanumericamente, são chamados de dados em rol ou elaborados.

Agora, podemos transformar os dados brutos da Tabela 2.1, em dados elaborados (em Rol), apresentados na Tabela 2.2. Em termos de notação, iremos representar um conjunto de variáveis ordenadas dessa forma, X_1, X_2, \dots, X_n , de tamanho n . Com o ordenamento, usaremos um parêntese no índice, $X_{(1)}, X_{(2)}, \dots, X_{(n)}$, de modo que, $X_{(1)} = \min_i(X_i)$ e $X_{(n)} = \max_i(X_i)$ para $i = 1, 2, \dots, n$. Da mesma forma, vale para a representação dos valores observados dessas respectivas variáveis, isto é, valores observados sem ordenação denotados por $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, e valores observados ordenados $x_{(1)} = \min_i(x_i)$ e $x_{(n)} = \max_i(x_i)$. Este último é a representação, em notação, dos dados elaborados.

Percebemos, com os dados elaborados, que os valores extremos representam, respectivamente, os valores mínimo e máximo do conjunto de dados, independente do número de elementos. Isso

Tabela 2.2: Dados elaborados sobre o número de erros encontrados em 20 conjunto de caracteres monitorado em um canal de comunicação.

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

facilita a percepção de algumas informações, porém ainda limitado, uma vez que a quantidade de valores pode ser simplificada, sem perda de informações, por meio de tabulações agrupadas em distribuição de frequências. Além de simplificar, podemos obter mais informações do que se estes dados tivessem expressos sem tabulação, do qual, trataremos na próxima seção.

2.2 Representação tabular

A frequência simples ou frequência absoluta, representa o número de vezes que determinado valor foi observado, que em notação, denotaremos por F_i i-ésima frequência de determinada variável, em que a frequência observada será denotada por f_i . Vejamos o agrupamento dos dados da Tabela 2.3, em distribuição de frequência, a seguir.

Tabela 2.3: Distribuição de frequência do número de erros encontrados em 20 conjunto de caracteres monitorado em um canal de comunicação.

Número de erros (x_i)	Frequência simples (f_i)
0	3
1	7
2	4
3	5
4	1
Total	20

De forma mais fácil, podemos por meio da Tabela 2.3 saber quantas vezes um determinado valor foi observado, sem grandes esforços, bastando apenas olhar para a coluna de frequências absolutas. Se desejarmos, apresentar uma forma relativa dessa frequência em relação ao número total de observações, podemos utilizar a frequência relativa, denotada por F_r , em que f_r representa o seu respectivo valor observado, e que essa frequência será um valor entre 0 e 1. Calculamos a frequência relativa, de acordo com a expressão (2.1),

$$F_{r_i} = \frac{F_i}{\sum_{i=1}^k F_i}, \quad i = 1, 2, \dots, k, \quad (2.1)$$

sendo k o número de grupos ou classes. No caso, o cálculo da frequência relativa baseado na Tabela 2.3, a representação de k se refere ao número de grupos, uma vez que os dados são discretizados. Portanto, esse tipo de agrupamento é válido tanto para as variáveis qualitativas quanto para a variável quantitativa discreta. No caso da variável quantitativa contínua, agrupamos os seus valores em classes, uma vez que sua natureza não é discretizada. O modo de como criar essas classes, aprenderemos mais a frente. Desse modo, a Tabela 2.4 apresenta o agrupamento dos dados do número de erros encontrados em 20 conjunto de caracteres monitorado em um canal de comunicação, juntamente com a frequência relativa de seus valores.

Percebemos que $\sum_{i=1}^k f_i = n$, uma vez que os dados são amostrais. A frequência relativa passa a ter sentido prático quando usamos o resultado em porcentagem, surgindo então, a frequência percentual, denotada por $F_{\%}$, cujo valor observado é dado por $f_{\%}$, de modo que essa frequência é calculada pela expressão (2.2).

$$F_{\%_i} = F_{r_i} \times 100, \quad i = 1, 2, \dots, k, \quad (2.2)$$

Tabela 2.4: Frequência relativa do número de erros encontrados em 20 conjunto de caracteres monitorado em um canal de comunicação.

Número de erros (x_i)	Frequência simples (f_i)	Frequência relativa (f_{r_i})
0	3	$3/20 = 0,15$
1	7	$7/20 = 0,35$
2	4	$4/20 = 0,20$
3	5	$5/20 = 0,25$
4	1	$1/20 = 0,05$
Total	20	1

em que F_{r_i} é dado pela expressão (2.1), e k é igual ao número de grupos ou classes.

Assim, podemos acrescentar a frequência percentual aos dados da Tabela 2.4, que pode ser apresentado na Tabela 2.5.

Tabela 2.5: Frequência percentual do número de erros encontrados em 20 conjunto de caracteres monitorado em um canal de comunicação.

Número de erros (x_i)	Frequência simples (f_i)	Frequência relativa (f_{r_i})	Frequência percentual ($f_{\%_i}$)
0	3	0,15	$0,15 \times 100 = 15$
1	7	0,35	$0,35 \times 100 = 35$
2	4	0,20	$0,20 \times 100 = 20$
3	5	0,25	$0,25 \times 100 = 25$
4	1	0,05	$0,05 \times 100 = 5$
Total	20	1	100

Podemos observar que, 35% do grupo de caracteres apresentava apenas 1 erro, 15% dos grupos não apresentaram erros. Porém, se perguntassemos, no mínimo, quantos grupos apresentaram 2 erros? Em quantos grupos tivemos, no máximo, 3 erros? A primeira pergunta, seria respondida somando as frequências simples (ou absolutas) $4 + 5 + 1 = 10$ grupos. Para a segunda pergunta, responderíamos $3 + 7 + 4 + 5 + 1 = 19$ grupos. Isso poderia tornar mais oneroso, à medida que o nome de grupos fosse aumentando. Ao invés, usaremos as frequências acumuladas, para auxiliar indagações desse tipo aos dados.

Temos dois tipos de frequências acumuladas, a frequência acumulada *abaixo de*, denotada por $F_{ac\downarrow}$, cujo valor calculado é denotado por $f_{ac\downarrow}$, dado pela expressão (2.3),

$$F_{ac\downarrow i} = \sum_{j=1}^i F_j, \quad i = 1, 2, \dots, k, \quad (2.3)$$

sendo k o número de grupos ou classes, e F_j representando j -ésima frequência absoluta.

A outra é a frequência acumulada *acima de*, denotada por $F_{ac\uparrow}$, cujo valor calculado é denotado por $f_{ac\uparrow}$, dado pela expressão (2.4),

$$F_{ac\uparrow i} = \sum_{j=i}^k F_j, \quad i = 1, 2, \dots, k, \quad (2.4)$$

sendo k o número de grupos ou classes, e F_j representando j -ésima frequência absoluta. Na Tabela 2.6, complementamos as informações com as frequências acumuladas os dados apresentados da Tabela 2.5.

É importante notar que sempre o último valor da coluna da frequência acumulada *abaixo de* é o número total de elementos, e que o último valor da frequência acumulada *acima de* coincide com o seu respectivo valor da frequência simples (f_i), como pode ser observado na Tabela 2.6. Uma outra coisa interessante, é que podemos ter uma forma alternativa de calcular a frequência acumulada *acima de*, dado pela expressão (2.5),

Tabela 2.6: Frequência percentual do número de erros encontrados em 20 conjunto de caracteres monitorado em um canal de comunicação.

Número de erros (x_i)	f_i	f_{r_i}	$f_{\%_i}$	Frequência acumulada ($f_{ac\downarrow_i}$)	Frequência acumulada ($f_{ac\uparrow_i}$)
0	3	0,15	15	3	$3 + 7 + 4 + 5 + 1 = 20$
1	7	0,35	35	$3 + 7 = 10$	$7 + 4 + 5 + 1 = 20 - 3 = 17$
2	4	0,20	35	$3 + 7 + 4 = 14$	$4 + 5 + 1 = 17 - 7 = 10$
3	5	0,25	25	$3 + 7 + 4 + 5 = 19$	$5 + 1 = 10 - 4 = 6$
4	1	0,05	5	$3 + 7 + 4 + 5 + 1 = 20$	$1 = 6 - 5 = 1$
Total	20	1	100	-	-

$$F_{ac\uparrow_i} = \begin{cases} \sum_{i=1}^k F_i, & i = 1, \\ F_{ac\uparrow_{i-1}} - F_{i-1}, & \text{demais casos,} \end{cases} \quad (2.5)$$

sendo k o número de grupos ou classes. Vamos tentar entender as equivalências entre as expressões (2.4) e (2.5). Indagamos, quantos grupos apresentam, no mínimo, 1 erro? Observe que $x_2 = 1$, isto é, os valores observados iguais a 1, estão no segundo grupo. Assim, pela expressão (2.4), temos

$$f_{ac\uparrow_2} = \sum_{j=2}^5 f_j = 7 + 4 + 5 + 1 = 17.$$

Da mesma forma, podemos utilizar a expressão (2.5), e de modo equivalente, temos

$$\begin{aligned} f_{ac\uparrow_2} &= f_{ac\uparrow_{2-1}} - f_{2-1} \\ &= f_{ac\uparrow_1} - f_1 \\ &= 20 - 3 = 17. \end{aligned}$$

A expressão (2.5) pode parecer no primeiro momento mais trabalhoso o cálculo. Porém, perceberemos com a prática de exercícios que esse processo é mais rápido do que calcular usando a expressão (2.4).

Por fim, podemos apresentar a forma relativa e percentual das frequências acumuladas, usando de modo similar, quando calculamos as frequências relativas e percentuais, dadas nas expressões (2.1) e (2.2), respectivamente. Denotaremos a frequência relativa acumulada *abaixo de*, por $Fr_{ac\downarrow}$, e a frequência relativa acumulada *acima de*, por $Fr_{ac\uparrow}$. As expressões dessas duas frequências são dadas, respectivamente, por

$$Fr_{ac\downarrow} = \frac{F_{ac\downarrow_i}}{\sum_{i=1}^k F_i} \quad i = 1, 2, \dots, k, \quad (2.6)$$

e

$$Fr_{ac\uparrow} = \frac{F_{ac\uparrow_i}}{\sum_{i=1}^k F_i} \quad i = 1, 2, \dots, k, \quad (2.7)$$

em que $F_{ac\downarrow_i}$ expresso em (2.3), $F_{ac\uparrow_i}$ expresso em (2.4), e k é igual ao número de grupos ou classes. Já para as frequências percentuais acumuladas *abaixo de* e *acima de*, denotando-as por $F_{ac\downarrow}\%$ e $F_{ac\uparrow}\%$, respectivamente, e sendo expressas por

$$F_{ac\downarrow}\% = Fr_{ac\downarrow} \times 100 \quad i = 1, 2, \dots, k, \quad (2.8)$$

e

$$F_{ac\uparrow \%} = Fr_{ac\uparrow} \times 100 \quad i = 1, 2, \dots, k, \quad (2.9)$$

respectivamente, em que k é o número de grupos ou classes.

Voltando ao conjunto de dados iniciado na Tabela 2.1, podemos finalizar a sua tabulação com todas as frequências mencionadas anteriormente, em um quadro resumo na Tabela 2.7.

Tabela 2.7: Distribuição de frequências do número de erros encontrados em 20 conjunto de caracteres monitorado em um canal de comunicação.

Número de erros (x_i)	f_i	f_{r_i}	$f\%_i$	$f_{ac\downarrow_i}$	$f_{ac\uparrow_i}$	$fr_{ac\downarrow}$	$fr_{ac\uparrow}$	$f_{ac\downarrow \%}$	$f_{ac\uparrow \%}$
0	3	0,15	15	3	20	0,15	1	30	100
1	7	0,35	35	10	17	0,50	0,85	33	85
2	4	0,20	35	14	10	0,70	0,50	47	50
3	5	0,25	25	19	6	0,95	0,20	0,95	20
4	1	0,05	5	20	1	1	0,05	100	5
Total	20	1	100	-	-	1	1	-	-

Essa representação tabular como falado anteriormente, pode ser usado para todas as variáveis discretizadas, como as variáveis qualitativas quanto a variável quantitativa discreta. Porém, para o caso da variável qualitativa nominal, não faz sentido o uso da frequência acumulada, uma vez que esse tipo de variável não tem ordenamento no sentido quantitativo. Para o caso da variável quantitativa contínua, precisamos agrupar os valores observados em intervalos de classe, isso porque a discretização de seus valores se devem aos instrumentos de medida, e não a natureza da variável. Por exemplo, quando medimos uma altura, 1,78m, de fato a altura real não está limitado a segunda casa decimal. Então, a melhor forma será criar regiões (intervalos), de modo que possamos contemplar determinados valores.

Existem diversas formas de como agrupar as variáveis quantitativas contínuas, isto é, metodologias de como desenvolver a criação de classes. Contudo, iremos nos restringir a um critério empírico, que se baseia no número de elementos, seja na amostra ou população. A primeira indagação que surge é, qual o número de classes para agrupar esses dados? Denotaremos por k o número de classes, em que sua expressão por:

$$k \approx \begin{cases} \sqrt{\text{número de elementos}}, & \text{Caso o tamanho seja igual ou inferior a 100} \\ 5\log_{10}(\text{número de elementos}), & \text{Caso o tamanho seja superior a 100.} \end{cases} \quad (2.10)$$

Nesse caso, o cálculo de k é uma aproximação, e devemos aproximar a um número inteiro mais próximo. Pode ocorrer situações em que o número de classes resulte em um agrupamento em que tenhamos classes com frequência 0, isto é, com nenhum elemento dentro dessa classe. Não faz sentido criar uma classe sem elementos. Dessa forma, ao final do processo da criação de tabelas com intervalos de classes e verificado classes sem elementos, o processo deve ser reiniciado e alterado o valor de k , ou um número inteiro para baixo ou para cima. Após isso, todo o processo que será apresentado na sequência, deverá seguir. Caso esse problema se repita, novamente, voltaremos a fase de determinação de k até encontrar um inteiro, do qual se obtenha agrupamento de dados com intervalo de classes, com frequência em suas classes superior a 0. Em todo esse processo, devemos evitar que o número de classes seja inferior a 3, uma vez que para $k < 3$ não será necessário um agrupamento de dados em tabela, para uma quantidade tão pequena de valores.

Dando sequência, após a determinação do número de classes, determinaremos a amplitude total, denotada por A_t , sendo definida pela expressão (2.11),

$$A_t = \max_i(X_i) - \min_i(X_i), \quad (2.11)$$

para $i \in \mathbb{N}^+$.

Posteriormente, deveremos determinar a amplitude da classe, denotada por c e expressa como:

$$c = \begin{cases} \frac{A_t}{k-1}, & \text{Amostra} \\ \frac{A_t}{k}, & \text{População.} \end{cases} \quad (2.12)$$

em A_t é expresso em (2.11) e k dada pela expressão (2.10). O fato de o denominado ter o valor de k subtraído de 1, ao invés de k para o caso dos dados amostrais, é devido a uma correção realizada no cálculo do limite inferior da primeira classe, que será apresentada a seguir. Segundo Ferreira (2009, p. 13) a justificativa se deve a suposição de que uma amostra de tamanho n tem grande chance de não conter o valor mínimo da população, isto é, à medida que o tamanho da amostra aumenta, tem-se uma maior chance de obter elementos menores que o valor mínimo que foi encontrado para uma amostra de tamanho menor.

Por fim, apresentamos o cálculo para se obter o limite inferior da primeira, denotado por Li_{1a} , sendo dado pela expressão (2.13),

$$Li_{1a} = \begin{cases} X_{(1)} - c/2, & \text{Amostra} \\ X_{(1)}, & \text{População.} \end{cases} \quad (2.13)$$

Realizando esses quatro passos, iremos criar as classes, iniciando pelo limite inferior da primeira classe, e para essa mesma classe, o seu limite superior será denotado por Ls_{1a} , cujo cálculo é dado por $Ls_{1a} = Li_{1a} + c$. Representaremos em notação a primeia classe da seguinte forma:

Classe
$Li_{1a} — Ls_{1a}$

Em termos de conjunto, diremos que Classe 1 = $\{x \in \mathbb{R} : Li_{1a} \leq x < Ls_{1a}\}$, isto é, os valores observados pertencerão a essa classe se forem iguais ou superiores a Li_{1a} e inferiores a Ls_{1a} . Como o valor do limite superior não pertence a essa classe, será contabilizado para a próxima classe. Nesse caso, o limite inferior da segunda classe será dado por $Li_{2a} = Ls_{1a}$ e seu limite superior $Ls_{2a} = Ls_{1a} + c$, em que c é a amplitude da classe. Inserindo agora a segunda classe na tabela, temos:

Classe
$Li_{1a} — Ls_{1a}$
$Li_{2a} — Ls_{2a}$

Mais uma vez, o valor do limite superior dessa classe não pertence, mas pertencerá a próxima classe. Esse processo continua, até chegar a k -ésima classe, que ao final teremos uma tabela da seguinte forma,

Classe
$Li_{1a} — Ls_{1a}$
$Li_{2a} — Ls_{2a}$
⋮
$Li_{ka} — Ls_{ka}$

No caso da última classe, nós contemplamos os valores dos limites a essas classes, caso existam no banco de dados. Assim, a frequência absoluta é calculada verificando os valores que estão dentro da amplitude dos intervalos, e as demais frequências seguem o mesmo raciocínio falado anteriormente. Algumas notações vistas na literatura não contemplam o limite superior da última classe, e assim, essa classe pode ser também representada da forma $Li_{ka} |— Ls_{ka}$. Um problema surge com os dados, porque ao serem agrupados nas classes nós perdemos essa informação. Vejamos a representação de uma classe com a sua frequência absoluta,

Classe	F_i
Li_{1a}	Ls_{1a}
Li_{2a}	Ls_{2a}
⋮	
Li_{ka}	Ls_{ka}
	f_k

Observe que sabemos quantos valores existem em cada classe, mas sem a informação dos dados brutos ou elaborados, nós não sabemos quais são os valores pertencentes a cada classe. Assim, uma alternativa de valor para representar as f_i , para $i = 1, 2, \dots, k$, em cada classe é usar o ponto médio. Esse critério é chamado hipótese tabular básica. Essa hipótese sugere que assumir o ponto médio como um potencial representante dos valores de uma determinada classe, assume um menor erro do que escolher qualquer outro valor dentro desse intervalo para representar essas observações. O ponto médio, denotado por \tilde{X}_i , será dado por:

$$\tilde{X}_i = \frac{Li_{ia} + Ls_{ia}}{2}, \quad i = 1, 2, \dots, k. \quad (2.14)$$

Percebemos, que ocorre uma perda de precisão nos dados quando agrupamos em intervalo de classes, uma vez que o ponto médio passa a representar esses valores em cada classe. De todo modo, se observa que essa perda de informação é pequena para o ganho que se obtém ao representar esses dados em tabulação com intervalo de classe, no sentido não só de organização, mas de entendimento das informações. Em resumo, podemos dizer que o algoritmo para criar um agrupamento de dados em intervalo de classes pode ser dados em sete passos:

- 1) Calcular k ,
- 2) Calcular A_t ,
- 3) Calcular c ,
- 4) Calcular Li_{1a} ,
- 5) Determinar as classes,
- 6) Calcular o ponto médio, e
- 7) Calcular as frequências como apresentadas no início dessa seção.

Os dados do Exemplo 2.1 foram retirados de Presidential Comission (1986, p. 129-131), e apresentados a seguir.

Exemplo 2.1:

Os dados representam a temperatura ($^{\circ}\text{F}$) do anel de vedação de cada teste de acionamento ou lançamento real do motor do foguete Challenger, isso porque, em 1986, houve nos Estados Unidos um dos maiores acidentes com ônibus espaciais, vitimando em 8 astronautas que estavam na tripulação. Foram realizados diversos estudos pela NASA para identificar as causas da falha. A primeira atenção se voltou para a temperatura do anel de vedação, que é apresentado a seguir.

84	49	61	40	83	67	45	66	70	69	80	58
68	60	67	72	73	70	57	63	70	78	52	67
53	67	75	61	70	81	76	79	75	76	58	31

Diante dessas informações, para melhor apresentar essas informações, vamos agrupar esses dados em uma tabela com intervalo de classes, uma vez que a temperatura do anel de vedação é uma variável quantitativa contínua. Inicialmente, vamos calcular o número de classes, $k = \sqrt{36} = 6$ classes. Observando os valores, percebemos que $x_{(1)} = 31^{\circ}\text{F}$ e $x_{(36)} = 84^{\circ}\text{F}$, logo a amplitude total é $A_t = 53^{\circ}\text{F}$. Na sequência, calculamos amplitude da classe, $c = 53/(6 - 1) = 10,6^{\circ}\text{F}$, por

fim, o limite inferior da primeira classe, $Li_{1a} = 31 - 10,6/2 = 25,7^{\circ}F$. Assim, começaremos pela primeira classe, em que já temos o limite inferior dela e seu limite superior será $Li_{1a} = 25,7 + 10,6 = 36,3^{\circ}F$. As demais classes, segue o procedimento descrito anteriormente. Por fim, verificaremos quais os valores pertencentes em cada classe para computar a frequência absoluta, o cálculo do ponto médio de acordo com a expressão (2.14) e as demais frequências calculadas como descritas no início dessa seção. Assim, temos o quadro geral de uma tabela com intervalo de classes para esses dados, apresentados a seguir.

	Classe	f_i	\tilde{x}_i	f_{r_i}	$f_{ac\downarrow_i}$	$f_{ac\uparrow_i}$	$f_{\%_i}$	$f_{ac\downarrow\%}$	$f_{ac\uparrow\%}$
1	25,7 — 36,3	1,00	31,00	0,03	1,00	36,00	3,00	2,78	100,00
2	36,3 — 46,9	2,00	41,60	0,06	3,00	35,00	6,00	8,33	97,22
3	46,9 — 57,5	4,00	52,20	0,11	7,00	33,00	11,00	19,44	91,67
4	57,5 — 68,1	12,00	62,80	0,33	19,00	29,00	33,00	52,78	80,56
5	68,1 — 78,7	12,00	73,40	0,33	31,00	17,00	33,00	86,11	47,22
6	78,7 — 89,3	5,00	84,00	0,14	36,00	5,00	14,00	100,00	13,89

Criamos o Código R 2.1, para que seja possível calcular de forma otimizada o agrumento de dados em intervalos de classe, para variáveis quantitativas contínuas.

Código R 2.1**Script:**

```

1 # Tabela em distribuicao de frequencias
2 tabfreq <- function(dados, k = NULL){
3   # Dados em Rol (Elaborados)
4   sort(dados)
5   # tamanho da amostra
6   n <- length(dados); n
7   if (is.null(k)) {
8     # Numero de classes
9     if (n <= 100) {
10       k <- round(sqrt(n))
11       # OBS.: O valor de k nao necessariamente precisa ser
12       #       sqrt(n). Esse eh um valor base
13     }
14     if (n > 100) {
15       k <- 5 * log10(n)
16     }
17   } else {
18     if(!is.numeric(k)) stop("O argumento deve ser numérico!")
19   }
20   # Amplitude total
21 At <- diff(range(dados))
22 # Valor minimo
23 x1 <- min(dados)
24 # Amplitude da classe
25 c <- round(At/(k - 1), 2)
26 # Limite inferior da primeira classe
27 LI1 <- x1 - c/2; LI1
28 vi <- c(LI1, rep(0, k - 1))
29 vs <- c(LI1 + c, rep(0, k - 1))
30 # Calculando os limites inferior e superior
31 for (i in 2:k) {
32   vi[i] <- vi[i - 1] + c
33   vs[i] <- vs[i - 1] + c
34 }
35 vi <- round(vi, 2)
36 vs <- round(vs, 2)

```

Script:

```

32 # Continuacao...
33 # Funcao para calculo das frequencias das classes
34 freq <- function(x, vi, vs, k) {
35   freq <- rep(0, k)
36   for (i in 1:(k - 1)) {
37     freq[i] <- length(x[x >= vi[i] & x < vs[i]])}
38   }
39 freq[k] <- length(x[x >= vi[k] & x <= vs[k]])
40 return(freq)
41 }
42 # Frequencia absoluta
43 fi <- freq(dados, vi, vs, k)
44 # Construindo as classes
45 classe <- paste(round(vi, 2), "|---_",
46 classe[k] <- paste(round(vi[k], 2), "|---|",
47 classe
48 # Ponto medio
49 pm <- (vi + vs)/2
50 # Frequencia relativa
51 fr <- round(fi/n, 2)
52 # Frequencia acumulada (abaixo de)
53 fac1 <- cumsum(fi)
54 # Frequencia acumulada (acima de)
55 # x: representa as frequencias absolutas
56 fac2 <- function(x) {
57   f1 <- sum(x)
58   n <- length(x)
59   vet <- c(f1, rep(0,n - 1))
60   for (j in 2:n) {
61     vet[j] <- vet[j - 1] - x[j - 1]
62   }
63   return(vet)
64 }
65 fac22 <- fac2(fi)
66 # Frequencia percentual
67 fp <- round(fr*100, 2)
68 # Fac (abaixo de) percentual
69 fac1p = round((fac1/n)*100, 2)
70 # Fac (acima de) percentual
71 fac22p <- round((fac22/n)*100, 2)
72 # Estatisticas
73 estat <- list(
74   "Número_de_classes" = k,
75   "Amplitude_total" = At,
76   "Valor_mínimo" = x1,
77   "Amplitude_da_classe" = c,
78   "LI_da_1ª_Cla" = LI1
79 )
80 # Tabela de frequencias
81 tabela <- data.frame(Classe = classe, Fi = fi,
82                       PM = pm, Fr = fr,
83                       Fac1 = fac1, Fac2 = fac22,
84                       Fp = fp, Fac1p = fac1p,
85                       Fac2p = fac22p)

```

Script:

```
86 # Continuacao...
87 # Fi => Frequencia absoluta # PM => Ponto medio da classe
88 # Fr => Frequencia relativa # Fac1 => Frequencia acumulada (abaixo de)
89 # Fac2 => Frequencia acumulada (acima de)
90 # Fp => Frequencia percentual
91 # Fac1p => Fac1 percentual # Fac2p => Fac2 percentual
92 listres <- list(tabela = tabela, estat = estat)
93 return(listres)
94 }
```

2.3 Representação gráfica

Por fim, nesta última seção, mostraremos mais uma forma de apresentar os dados, usando gráficos.

Exercícios propostos

Exercício 2.1: Observamos nas expressões (2.3) e (2.4), a forma de se calcular as frequências acumuladas *acima de* e *abaixo de*, quando indagamos questões que envolvem situações do tipo, quantas vezes observamos, *no máximo*, $X = x$? Para isso, usamos a expressão (2.3). Em outra situação, indagamos, quantas vezes observamos, *no mínimo*, $X = x$? Para isso, usamos a expressão (2.4). Percebemos que a condição limiar está inclusa na situação. Por exemplo, na Tabela 2.4, podemos estar interessados em saber quantos grupos de caracteres monitorados^a em um canal de comunicação, foram encontrados, *no mínimo*, 2 erros? Isto significa, que desejamos saber todos os grupos, tais que $X \geq 2$. Ou seja, o grupo que continha dois erros estava incluso na contagem. Para isso, usamos a expressão (2.4). Porém, podemos estar interessados na situação em que a condição limiar não esteja inclusa na contagem de elementos. Por exemplo, refazendo a indagação anterior, quantos grupos de caracteres apresentam *acima de* 2 erros. Observe que os grupos apresentam dois erros não entram na contagem. Isso vale também para a outra situação. Logo, não será possível utilizar as expressões das frequências acumuladas (2.3) e (2.4).

Desenvolva as expressões, para essas últimas situações, de modo similar ao apresentado para as expressões (2.3) e (2.4), fazendo as adaptações devidas.

Solução na página 210

^aEntenda nessa situação que grupo de caracteres é um elemento da amostra.

Exercício 2.2: Os dados retirados de Tavares e Anjos (1999, p. 763), representam a distribuição percentual do estado nutricional em homens idosos brasileiros (idade ≥ 60 anos), segundo Índice de Massa Corporal (IMC^a), por macrorregião e situação de domicílio, Pesquisa Nacional sobre Saúde e Nutrição, 1989, que seguem,

Regiões	Número	Estado Nutricional (%) ^b			
		Magreza	Adequado	Sobrepeso I	Sobrepeso II e III
Norte	223	4,4	60,6	29,4	5,6
Nordeste	586	8,8	68,3	19,8	3,1
Urbano	267	7,1	62,3	26,6	4,0
Rural	319	10,7	74,6	12,5	2,2
Sudeste	46,3	7,9	59,0	26,7	6,4
Urbano	197	5,6	56,4	30,2	7,8
Rural	266	17,3	69,5	12,4	0,8
Sul	429	5,1	56,5	29,2	9,2
Urbano	197	4,5	51,2	33,0	11,3
Rural	232	6,4	66,4	22,0	5,2
Centro-Oeste	327	10,7	60,6	22,8	5,9
Urbano	154	10,6	55,2	27,3	6,9
Rural	173	11,0	71,4	13,7	3,9
Brasil	2.028	7,8	61,8	24,7	5,7
Urbano	1.038	6,0	57,2	29,5	7,3
Rural	990	11,7	71,7	14,2	2,4

Considere a variável estado nutricional como estudo, então como seria desenvolvido a distri-

buição de frequência baseado nas informações da tabela? Apresente-a(s).

Solução na página 210

^aA unidade de IMC em kg/m^2 .

^bA classificação do estado nutricional em relação ao IMC foi: Magreza (todas as formas - IMC < 18,5); adequado ($18 \leq IMC < 25,0$); sobrepeso I ($25 \leq IMC < 30,0$); sobrepeso II ($IMC \geq 30,0$).

Exercício 2.3:

Considerando os dados do Exercício 2.2, como poderíamos representá-los graficamente, considerando apenas um gráfico?

Solução na página 210

Exercício 2.4: Dados retirados de Montgomery e Runger (2016), mostram os dados retirados planta de fabricação de semicondutores. Nessa planta, o semicondutor é um fio colado a uma estrutura. A variável em estudo representa a resistência à tração (lbf/pol^2), isto é, uma força requerida para romper a cola. Os dados são apresentados a seguir.

9,95	24,45	31,75	35,00	25,02	16,86	14,38	9,60	24,35	27,50
17,08	37,00	41,95	11,66	21,65	17,89	69,00	10,30	34,93	46,59
44,88	54,12	56,63	22,13	21,15					

Apresente uma tabela de distribuição de frequências para esses dados, informando as frequências: absoluta, relativa, acumulada *abaixo de*, acumulada *acima de*, percentual, percentual acumulada *abaixo de* e percentual acumulada *acima de*.

Solução na página 211

Medidas de Posição

3.1 Introdução

Após tabularmos os dados ou apresentarmos graficamente, percebemos que ainda assim a quantidade de informações pode ser muito grande para descrevê-los. Desse modo, surgem algumas medidas que podem resumir tudo isso, de modo a preservar as principais características contidas nessas observações, são as denominadas medidas de posição ou tendência central, e as medidas de dispersão ou de variabilidade, que tem a propriedade de localizar a distribuição dos dados e também caracterizar sua variabilidade, respectivamente. Nesse capítulo trataremos das medidas de posição, e no Capítulo 4, as medidas de dispersão.

As medidas de posição representam o ponto central da massa de dados, de modo que o seu valor indica que as observações estão em torno dele, mas que não necessariamente, o valor dessa medida central existe no conjunto de dados. A escolha das medidas de posição apresentadas, dependerá da natureza das variáveis, bem como algumas peculiaridades existentes nos dados, como por exemplo, a existência de dados discrepantes. Vamos apresentar na sequência, a primeira medida de tendência central e a mais conhecida e utilizada na estatística, a média aritmética.

3.2 Média

Quando iniciamos uma conversa e percebemos que alguém está no meio termo em um determinado posicionamento, dizemos que a pessoa está fazendo “média”, vulgarmente, dizemos que está em cima do muro. Nesse mesmo raciocínio, é a média aritmética, uma medida em que o seu valor representa o valor central das observações. Podemos comparar a média como um ponto de equilíbrio em um sistema de pesos, do qual se cada observação pode ser representada com uma certa massa no ponto no eixo X de um plano cartesiano, então o ponto que representa a média equilibrará esse sistema de pesos. Definimos,

Definição 3.1: Média aritmética

Seja uma amostra X_1, X_2, \dots, X_n , de uma população X_1, X_2, \dots, X_N , de tamanhos n e N , respectivamente, definimos a média aritmética por:

$$\mu = \frac{\sum_{i=1}^N X_i}{N}, \quad (\text{População}) \quad (3.1)$$

e

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}. \quad (\text{Amostra}) \quad (3.2)$$

Em notação, dizemos que μ é uma característica amostral, isto é, representa a média populacional e chamamos de parâmetro. Na prática, essa informação é desconhecida e a representamos por uma medida amostral, que chamamos de estimador, uma função que depende apenas dos dados

amostrais. Um estimador para μ representa a média aritmética \bar{X} . O valor observado de \bar{X} pode ser representado por \bar{x} , em termos de notação. Vejamos um exemplo a seguir.

Exemplo 3.1:

Considerando os dados da Tabela 2.1, podemos calcular a média amostral da seguinte forma:

$$\bar{X} = \frac{3 + 1 + \dots + 1}{20} = \frac{34}{20} = 1,7 \text{ erros.}$$

Portanto, o número de erros encontrados em um conjunto de caracteres, podem ser representados por uma única medida, que é a média amostral. A interpretação é que, em média, ocorreram 1,7 erros nos caracteres monitorados em um meio de comunicação, e significa, que os 20 conjuntos de caracteres apresentam um número de erros em torno desse valor.

A Definição 3.1 é utilizada para dados sem agrupamento, isto é, dados brutos ou dados elaborados. Para o caso de dados agrupados em distribuição de frequência, definimos,

Definição 3.2: Média aritmética em dados agrupados

Seja uma amostra X_1, X_2, \dots, X_n , de tamanho n , agrupados em k grupos com variáveis X_i e frequência F_i , ou k classes com pontos médios \tilde{X}_i e F_i frequências, para $i = 1, 2, \dots, k$ e $\sum_{i=1}^k F_i = n$, então a média aritmética de uma amostra, é definida por:

$$\bar{X} = \begin{cases} \frac{\sum_{i=1}^k X_i \times F_i}{\sum_{i=1}^k F_i}, & \text{agrupados sem intervalo de classe,} \\ \frac{\sum_{i=1}^k \tilde{X}_i \times F_i}{\sum_{i=1}^k F_i}, & \text{agrupados com intervalo de classe,} \end{cases} \quad (3.3)$$

sendo \tilde{X}_i o ponto médio das classes.

Podemos representar a Definição 3.2 em termos populacionais, substituindo o tamanho n por N , como também representar a expressão em termos de valor observado. Porém, em termos de notação, preferimos usar dessa forma. Vejamos mais um exemplo a seguir.

Exemplo 3.2:

$$\bar{X} = \frac{0 \times 3 + 1 \times 7 + \dots + 4 \times 1}{20} = \frac{34}{20} = 1,7 \text{ erros.}$$

Notamos que o resultado para a média amostral é o mesmo obtido no Exemplo 3.1, porque mesmo agrupando os dados, o cálculo da média para esse tipo de dado, se baseia nos próprios valores observados.

Porém, para o caso da variáveis quantitativas contínuas isso não ocorre, porque usamos o ponto médio para representar as observações de cada classe. Vejamos o próximo exemplo, a seguir.

Exemplo 3.3:

Consideremos agora os dados agrupados do Exemplo 2.1. Trata-se de uma variável quantitativa contínua e, portanto, a média é baseada de acordo com a expressão (3.3), para o caso de dados agrupados com intervalo de classe, que segue:

$$\bar{X} = \sum_{i=1}^k \frac{\tilde{X}_i F_i}{\sum_{i=1}^k F_i} = \frac{31 \times 1 + 41,60 \times 2 + \dots + 84,00 \times 5}{1 + 2 + \dots + 5} = 66,04^\circ\text{F}. \quad (3.4)$$

Se calculássemos a média sem agrupamento, o valor seria $\bar{X} = 65,86^\circ\text{F}$. Observamos uma perda de

precisão com os dados quando agrupados com intervalo de classe. Mas isso pode ser justificado por exemplo, se nesse experimento a diferença em 0,18°F não altera os resultados da pesquisa, e assim, podemos apresentar os dados de forma mais organizada.

Nos exemplos anteriores, observamos que a média leva em consideração a todas as observações, em seu cálculo. Apesar dessa ideia ser interessante, uma vez que conseguimos captar as informações de cada elemento da amostra ou população, qualquer alteração que houver em alguma observação, pode alterar completamente o resultado da média aritmética. É caso dos dados discrepantes, isto é, observações muito distante da grande parte dos dados. Isso pode ocorrer por diversas situações, como erro humano, ao digitar errado em uma planilha, elementos mal amostrados, de modo que, determinado elemento selecionado para a amostra não pertencia a população de interesse, ou até mesmo, uma condição atípica na realização da coleta dos dados. Vejamos mais algumas características da média aritmética:

- a unidade da média está na mesma escala da variável em estudo;
- a média é uma das medidas mais conhecidas e utilizadas, devido as suas propriedades estatísticas que serão vistas nos capítulos seguintes;
- é única para cada conjunto de dados;
- usada apenas para variáveis quantitativas;
- não pode ser calculada para dados agrupados que apresentam classes extremas abertas;
- é influenciada por dados discrepantes.

Uma saída para contornar o problema dos dados discrepantes, pode ser abordado no exemplo a seguir.

Exemplo 3.4:

Considere um conjunto de dados ($n = 17$) fictícios que apresentam a maior e a menor observação como suspeitos de serem atípicos quanto as suas ocorrências:

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

Para representar esse conjunto de dados, usamos a média aritmética para representá-los:

$$\bar{X} = \frac{1 + 5 + \dots + 40}{17} = 9,67 \text{ und.}$$

Observamos que as observações $x_1 = 1 \text{ unid.}$ e $x_{17} = 40 \text{ unid.}$ podem ter influenciado o resultado, e como suspeitamos desses valores, vamos usar uma medida mais robusta a essa violação, isto é, que não será influenciado por esses valores. Chamamos de média aparada, denotada por \bar{X}_{ap} , que para uma amostra de tamanho n , temos:

$$\bar{X}_{ap} = \frac{\sum_{i=2}^{n-1} X_{(i)}}{n - 2}, \quad (3.5)$$

em que $X_{(i)}$ é a (i) -ésima variável em ordem crescente de magnitude, tal que $X_{(1)} = \min_i X_i$ e $X_{(n)} = \max_i X_i$.

Usando a expressão (3.5), apresentamos a média aparada:

$$\bar{X}_{ap} = \frac{5 + 5 + \dots + 10}{15} = 7,65 \text{ unid.}$$

Observamos pelo resultado, que os valores extremos acabam não influenciando no resultado da média aparada, e portanto, pode ser uma alternativa de medida de posição, para representar o conjunto de dados.

Complementando as características da média, apresentamos algumas propriedades pelo Teorema 3.1 a seguir, do qual iremos usar a Definição 3.1 como base, e as demais seguem de forma similar.

Teorema 3.1: Propriedades da Média aritmética

Baseado na Definição 3.1, e considerando c uma constante, então:

- I) Se para uma amostra X_1, X_2, \dots, X_n , a média aritmética é dada por $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, a nova média aritmética é dada por $\bar{Y} = \bar{X} \pm c$;
- II) Se para uma amostra X_1, X_2, \dots, X_n , a média aritmética é dada por $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, então para uma transformação de $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$, a nova média aritmética é dada por $\bar{Y} = \bar{X} \times c$. Esse resultado vale também para a transformação $Y_i = X_i/m$, sendo m também uma constante. Basta usar $c = 1/m$ e o resultado é o mesmo.
- III) A soma de quadrado de desvios dos dados em relação a uma constante c , é minimizada se $c = \bar{X}$.

Prova:

- I) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, então a média aritmética de Y_i é dado por:

$$\begin{aligned}\bar{Y} &= \frac{\sum_{i=1}^n Y_i}{n} \\ &= \frac{\sum_{i=1}^n X_i \pm c}{n} \\ &= \frac{\sum_{i=1}^n X_i}{n} \pm \frac{\sum_{i=1}^n c}{n} \\ &= \frac{\sum_{i=1}^n X_i}{n} \pm \frac{n \times c}{n} \\ &= \bar{X} \pm c. \quad \text{c.q.d.}\end{aligned}$$

- II) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$, então a amplitude de Y_i é dado por:

$$\begin{aligned}\bar{Y} &= \frac{\sum_{i=1}^n Y_i}{n} \\ &= \frac{\sum_{i=1}^n X_i \times c}{n} \\ &= \frac{X_1 \times c + X_2 \times c + \dots + X_n \times c}{n} \\ &= c \times \frac{X_1 + X_2 + \dots + X_n}{n} \\ &= c \times \bar{X}. \quad \text{c.q.d.}\end{aligned}$$

- III) Fazendo:

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

Expandindo o somatório e derivando D em relação a "c", temos que

$$D = \sum_{i=1}^n (X_i - c)^2 = \sum_{i=1}^n (X_i^2 - 2cX_i + c^2) = \sum_{i=1}^n X_i^2 - \sum_{i=1}^n 2cX_i + \sum_{i=1}^n c^2,$$

e que

$$\frac{dD}{dc} = -2 \sum_{i=1}^n X_i + 2nc.$$

Igualando a derivada a zero, e resolvendo em A , temos:

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

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

$$c = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}.$$

Certificando se o ponto é de máximo ou de mínimo,

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

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

3.3 Mediana

Uma outra alternativa para contornarmos os problemas de dados discrepantes encontrados na média aritmética, pode ser apresentada por meio da medida de posição chamada de mediana, do qual leva em consideração a posição ordenada dos dados ao invés de usar os próprios valores observados. Mas especificamente, o valor da mediana é o ponto central dos dados, em que abaixo desse valor, representa as 50% menores observações, ao passo que, os valores acima da mediana representam as 50% maiores observações. De outro modo, dizemos que a mediana representa um ponto central no conjunto de dados em que a quantidade de elementos abaixo ou acima desse valor, não supera 50%. Essa última definição representa melhor o que significa a mediana, pois podemos ter valores centrais repetidos, e dessa forma isso ocorrendo, a primeira afirmação não será válida para a definição da mediana. Formalmente, definimos,

Definição 3.3: Mediana

Seja uma amostra X_1, X_2, \dots, X_n , de uma população X_1, X_2, \dots, X_N , de tamanhos n e N , respectivamente, definimos a mediana por:

$$\mu_d(X) = \begin{cases} \frac{X_{(\frac{N}{2})} + X_{(\frac{N+1}{2})}}{2}, & \text{se } N \text{ for um número par} \\ X_{(\frac{N+1}{2})}, & \text{se } N \text{ for um número ímpar} \end{cases}, \quad (\text{População}) \quad (3.6)$$

sendo $\mu_d(X)$ a mediana populacional e que $X_{(i)}$ é a (i) -ésima variável em ordem crescente de magnitude, tal que $X_{(1)} = \min_i X_i$ e $X_{(n)} = \max_i X_i$. De modo similar,

$$Md(X) = \begin{cases} \frac{X_{(\frac{n}{2})} + X_{(\frac{n+1}{2})}}{2}, & \text{se } n \text{ for um número par} \\ X_{(\frac{n+1}{2})}, & \text{se } n \text{ for um número ímpar} \end{cases}, \quad (\text{Amostra}) \quad (3.7)$$

sendo $Md(X)$ a mediana amostral e que $X_{(i)}$ é a (i) -ésima variável em ordem crescente de magnitude, tal que $X_{(1)} = \min_i X_i$ e $X_{(n)} = \max_i X_i$.

A mediana amostral é o melhor estimador para a mediana populacional, e pode ser considerado também como um estimador para a média populacional (μ). Detalhes sobre a escolha de um melhor estimador para um determinado parâmetro, será estudado no Capítulo 9. Como a mediana leva em consideração a posição das observações, a condição do tamanho amostral ou populacional acaba sendo importante para essa medida, de modo que, se o tamanho for um número par ou ímpar, teremos condições diferentes para o cálculo. Uma outra informação importante para o cálculo da mediana, é que será necessário ordenar as observações de modo crescente. Em notação para o caso de uma amostra de tamanho n , dizemos que $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ representa uma amostra em ordem crescente de magnitude, isto é, $X_{(1)} = \min_i X_i$ e $X_{(n)} = \max_i X_i$, e precisaremos desse ordenamento para obter o valor da mediana, baseados na expressões da Definição 3.3. Se utilizarmos o Exemplo 3.4, perceberemos que não é necessário eliminar as observações extremas em ordem de magnitude, como foi realizado com a média aparada. Isso demonstra que a mediana é uma outra alternativa de medida robusta para a escolha de uma medida de posição de modo a representar um conjunto de dados. Vejamos o exemplo a seguir.

Exemplo 3.5:

Considerando o Exemplo 3.4, como $n = 17$ é ímpar, a mediana amostral desse conjunto de dados fictícios é dado por:

$$Md(X) = X_{(\frac{17+1}{2})} = X_{(9)} = 8 \text{ unid.}$$

Esse valor representa uma medida central do qual os 50% menores valores dos dados estão abaixo de 8 unid., e que os 50% maiores valores dos dados estão acima de 8 unid.. Porém, percebemos que os valores $x_{(1)} = 1$ e $x_{(17)} = 40$ não influenciaram nesse resultado. Isso mostra, a robustez da mediana quanto a esse aspecto.

Para o caso de variáveis quantitativas contínuas sem agrupamento, o procedimento é o mesmo realizado no Exemplo 3.6. Para os dados da Tabela 2.3, isto é, dados agrupados sem intervalo de classe (variáveis quantitativas discretas), podemos calcular a mediana usando a Definição 3.3. Precisaremos apenas complementar as informações com o acréscimo da frequência acumulada *abaixo de* ($f_{ac,i}$), que foi apresentada na Tabela 2.6. Vejamos o próximo exemplo.

Exemplo 3.6:

Vejamos os dados do número de erros de caracteres em 20 conjuntos, descritos na Tabela 2.6, em que simplificamos os resultados, que segue:

Número de erros (X_i)	F_i	Frequência acumulada ($F_{ac\downarrow i}$)
0	3	3
1	7	10
2	4	14
3	5	19
4	1	20
Total	20	-

O valor da mediana será dado da seguinte forma:

$$\begin{aligned} Md(X) &= \frac{X_{(\frac{20}{2})} + X_{(\frac{20}{2}+1)}}{2} \\ &= \frac{X_{(10)} + X_{(11)}}{2}. \end{aligned}$$

Para sabermos qual o valor observado para a variável $X_{(10)}$ e $X_{(11)}$, marcamos os grupos 2 (linhas 2 de vermelho) e 3 (linha 3 de amarelo). No grupo 2, temos sete elementos que correspondem as variáveis $X_{(4)}, X_{(5)}, \dots, X_{(10)}$, uma vez que os três menores valores estão no grupo 1. Assim o $X_{(10)} = 1$ erros. No grupo 3, nós temos quatro elementos que correspondem as variáveis $X_{(11)}, X_{(12)}, \dots, X_{(14)}$, uma vez que abaixo desse grupo nós temos as dez primeiras observações. Assim, o $X_{(11)} = 2$ erros. Usamos as frequências simples (F_i) e acumulada $F_{ac\downarrow i}$, para obter essas informações. Retornando ao cálculo da mediana, temos:

$$\begin{aligned} Md(X) &= \frac{X_{(10)} + X_{(11)}}{2} \\ &= \frac{1+2}{2} = 1,5 \text{ erros.} \end{aligned}$$

Caso os dados estivessem em rol, o resultado seria o mesmo.

No caso de dados agrupados com intervalo de classe (variáveis quantitativas contínuas), vamos definir um estimador para a mediana populacional, usando uma dedução geométrica por meio do histograma de frequências e as ogivas. Para isso, vamos usar os dados do Exemplo 2.1 para facilitar a explicação, em que apresentamos na Figura 3.1 o histograma e as ogivas desses dados agrupados.

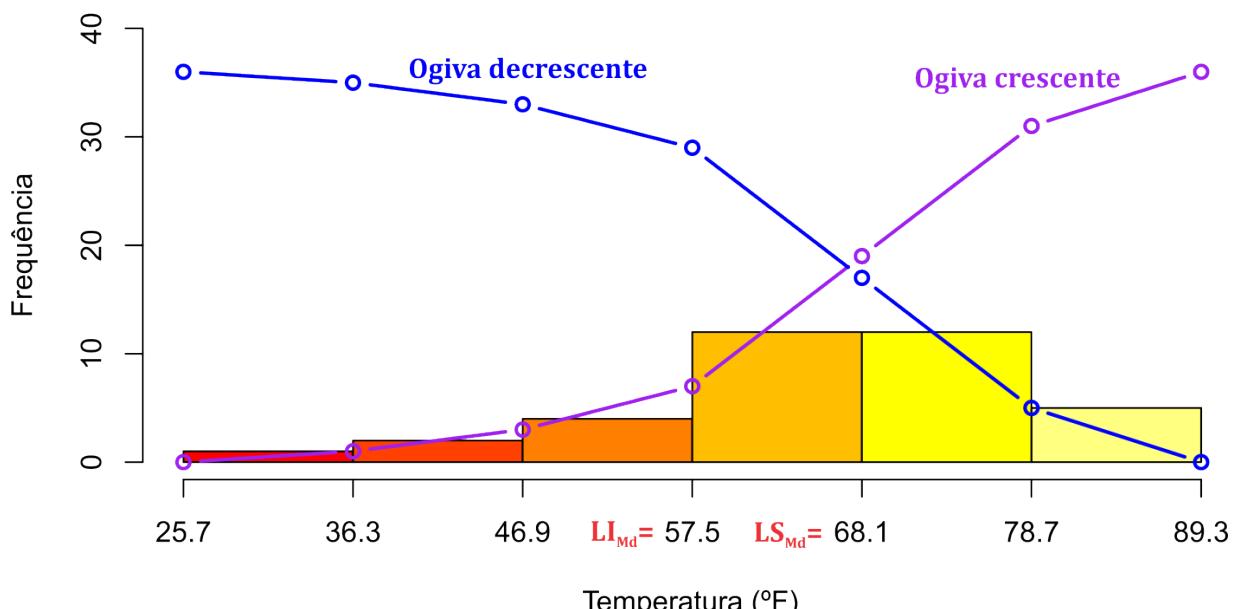


Figura 3.1: Histograma de frequência e ogivas para a dedução do cálculo da mediana.

Para estimar a mediana a partir dos dados arranjados em uma tabela de distribuição de frequência com intervalo de classe, é necessário definir a classe mediana e em seguida encontrar a mediana interpolando os resultados. A posição da mediana é obtida acumulando-se frequências das classes 1, 2, etc., até se encontrar o valor que seja igual ou imediatamente superior a $n/2$. Apresentamos algumas notações importantes para o entendimento da dedução do estimador de $\mu_d(X)$, que segue:

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

Com essa notação apresentamos a Figura 3.2 para facilitar a compreensão da dedução. Iremos apresentar dois métodos, o primeiro baseado no limite inferior da classe da mediana, e o segundo baseado no limite superior da classe da mediana. Nesse tipo de natureza de dados, desprezaremos se o número de elementos é par ou ímpar. Entenderemos que a classe da mediana é aquela que contempla o valor observado para a variável $X_{(n/2)}$. Para isso, podemos observar esse valor na coluna da frequência acumulada (abaixo de), $f_{ac\downarrow i}$. Nos dados do Exemplo 2.1, a classe da mediana é 57,5 — 68,1 porque $f_{ac\downarrow 4} = 19$, isto é, abaixo de 68,1 °F temos as primeiras 19 observações, e nessa classe contemplamos as observações ordenadas $x_{(8)}, x_{(9)}, \dots, x_{(19)}$, que contém $x_{(n/2)} = x_{(36/2)} = x_{(19)}$. Temos essas observações na classe 4 (classe da mediana), porque a frequência acumulada (abaixo de) anterior a classe da mediana, $f_{ac\downarrow 3} = 7$. Isso significa que a partir do oitavo elemento ordenado até o décimo nono temos elementos pertencentes a classe da mediana.

Feito essas considerações, apresentamos o primeiro método de dedução da expressão da mediana, a seguir.

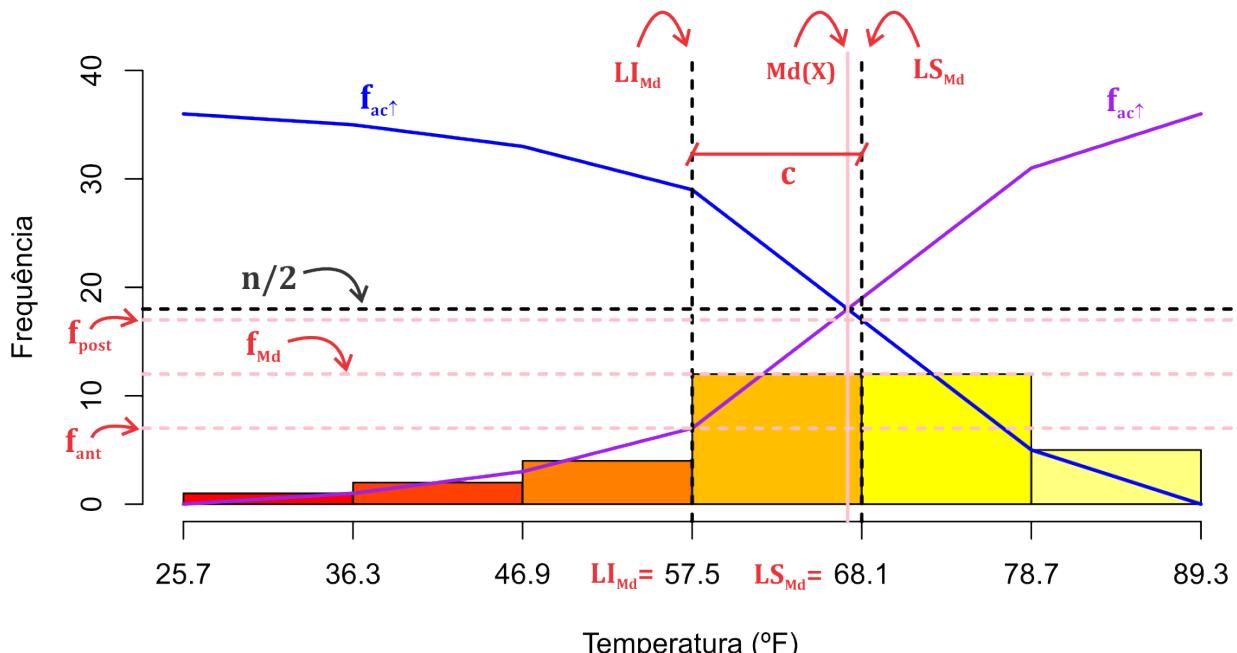


Figura 3.2: Histograma de frequência e ogivas para a dedução do cálculo da mediana com as notações.

1º Método

Uma vez que sabemos a classe da mediana, pela Figura 3.3 podemos determinar o valor da mediana por:

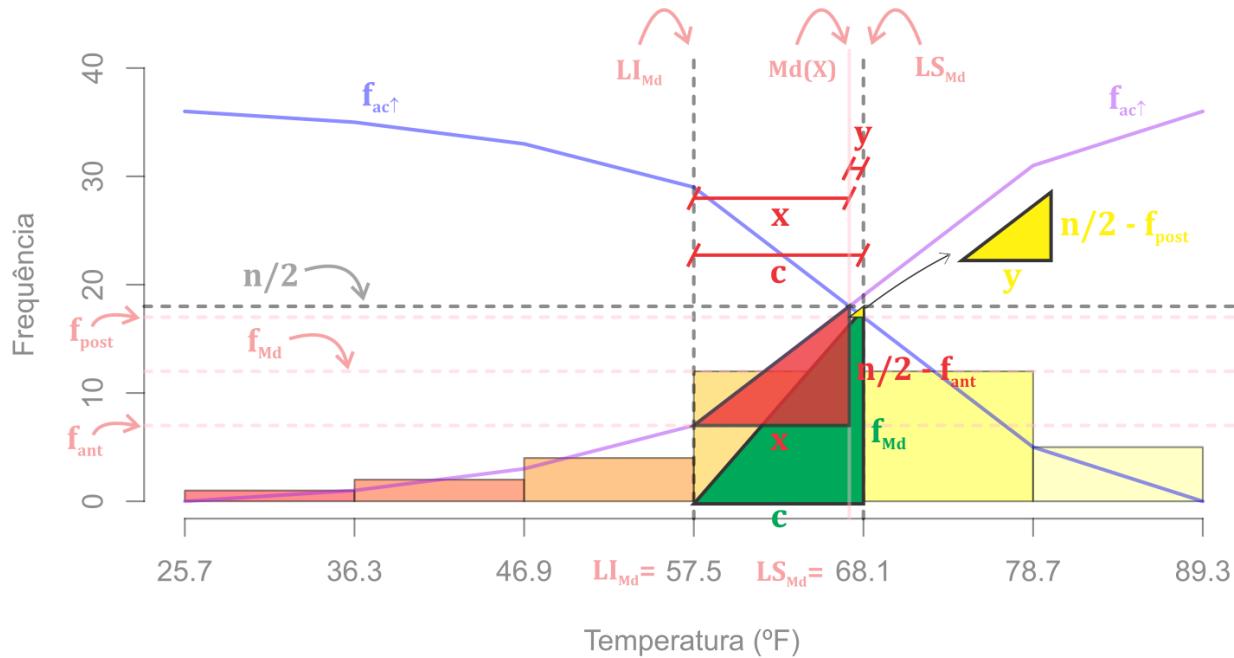


Figura 3.3: Detalhamento do Histograma de frequência e ogivas para a dedução do cálculo da mediana com as notações.

$$Md(X) = LI_{Md} + x, \quad (3.8)$$

sendo necessário encontrar o valor x . Assim, faremos uma regra de três simples pela semelhança de triângulos (triângulo verde e vermelho) que pode ser observado pela Figura 3.4.

Assim, temos

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

Determinando x ,

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

Como $Md(X) = LI_{Md} + x$, então

$$Md(X) = LI_{Md} + \left\{ \frac{\frac{n}{2} - f_{ant}}{f_{Md}} \right\} \times c. \quad (3.9)$$

2º Método

Uma vez que sabemos a classe da mediana, pela Figura 3.3 podemos determinar o valor da mediana pelo segundo método, sendo necessário encontrar o valor y na seguinte expressão (3.10).

$$Md(X) = LS_{Md} - y. \quad (3.10)$$

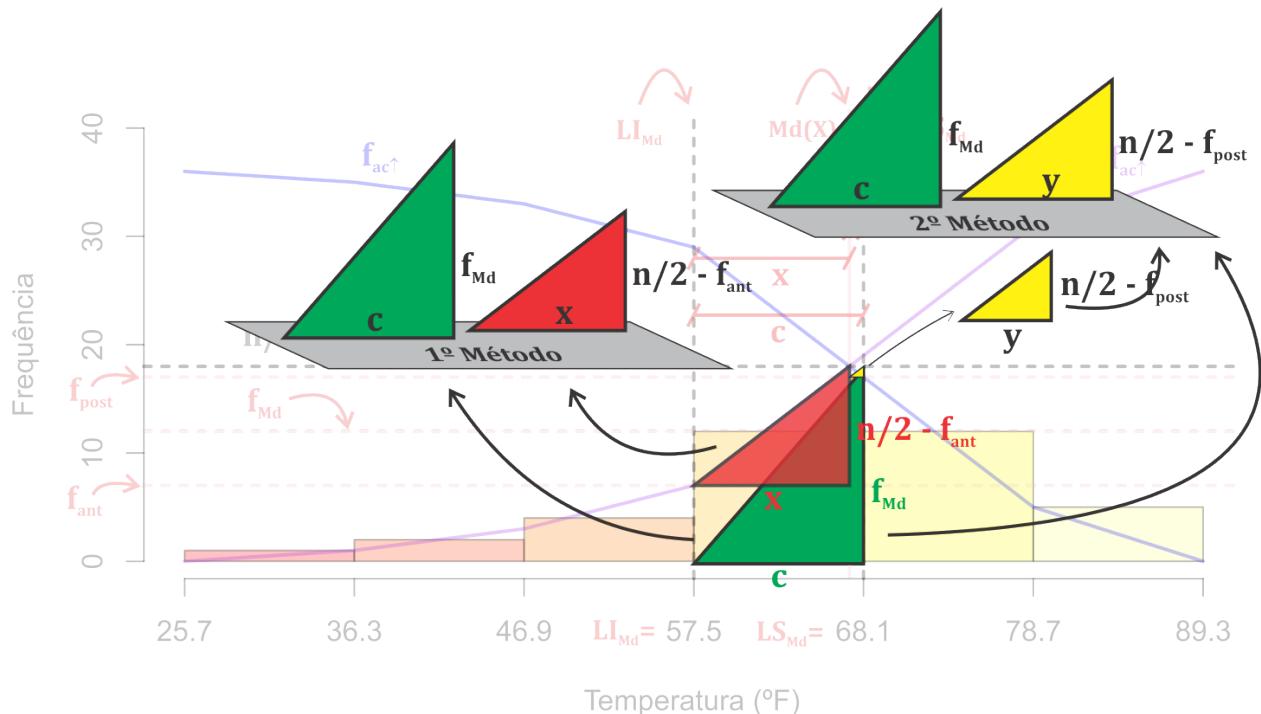


Figura 3.4: Semelhança de triângulos para a dedução do cálculo da mediana com as notações.

Assim, faremos uma regra de três simples usando a semelhança de triângulos (triângulo verde e amarelo). Assim,

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

Determinando y ,

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

a mediana amostral pode ser expressa como

$$Md(X) = LS_{Md} - \left\{ \frac{\frac{n}{2} - f_{post}}{f_{Md}} \right\} \times c. \quad (3.11)$$

Formalizando essas ideias, definimos um estimador da mediana amostral para dados agrupados com intervalo de classe da seguinte forma,

Definição 3.4: Mediana em dados agrupados com intervalo de classe

Seja uma amostra $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ em ordem crescente de magnitude, de tamanho n , agrupados em k classes com pontos médios \tilde{X}_i e F_i frequências, para $i = 1, 2, \dots, k$ e $\sum_{i=1}^k F_i = n$, então a mediana amostral é definida por:

$$Md(X) = LI_{Md} + \left\{ \frac{\frac{n}{2} - f_{ant}}{f_{Md}} \right\} \times c. \quad (3.12)$$

em que LI_{Md} é o limite inferior da classe da mediana, f_{ant} é a frequência acumulada (*abaixo de*) anterior a classe da mediana, f_{Md} frequência absoluta da classe da mediana, c a amplitude da classe da mediana, ou de forma similar,

$$Md(X) = LS_{Md} - \left\{ \frac{\frac{n}{2} - f_{post}}{f_{Md}} \right\} \times c. \quad (3.13)$$

em que LS_{Md} é o limite superior da classe da mediana e f_{post} é a frequência acumulada (*acima de*) posterior a classe da mediana.

Vamos apresentar o resultado da mediana para os dados do Exemplo 2.1 a seguir.

Exemplo 3.7:

Retornando aos dados do Exemplo 2.1, vamos calcular a mediana pelos dois métodos, do qual temos o primeiro resultado, usando a expressão (3.12),

$$Md(X) = 57,5 + \left\{ \frac{18 - 7}{12} \right\} \times 10,6 = 67,22^{\circ}\text{F}.$$

Usando o segundo método, expressão (3.13), temos,

$$Md(X) = 68,1 - \left\{ \frac{18 - 17}{12} \right\} \times 10,6 = 67,22^{\circ}\text{F}.$$

Os resultados são equivalentes como era de se esperar.

Vejamos algumas características sobre a mediana, que seguem:

- A mediana não é influenciada por valores extremos;
- Uma medida que pode ser obtida em distribuições de frequências que apresentam classe com limites indefinidos;
- o resultado da mediana é obtida na mesma escala da variável em estudo;
- a mediana é menos informativa que a média, por não levar em consideração os valores observados, mas as posições dessas observações;
- a mediana pode ser calculada em variáveis qualitativas ordinais, cuja média não pode ser obtida;
- a mediana ainda pode ser obtida em um conjunto de dados em que alguns valores ainda não foram registrados, caso em que a média não pode ser obtida.

Para complementarmos essas características, vamos apresentar algumas propriedades da mediana no Teorema 3.2. Iremos a Definição 3.3, bem como a expressão (3.7) para n ímpar. Para os demais casos, os resultados são similares.

Teorema 3.2: Propriedades da mediana

Baseado na Definição 3.3, e considerando c uma constante, então:

- I) Se para uma amostra $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ em ordem crescente de magnitude, a mediana é dada por $MdX = X_{(\frac{n+1}{2})}$, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, a mediana aritmética é dada por $Md(Y) = Md(X) \pm c$;
- II) Se para uma amostra $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ em ordem crescente de magnitude, a mediana é dada por $MdX = X_{(\frac{n+1}{2})}$, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, a nova mediana é dada por $MdY = MdX \times c$. Esse resultado vale também para a transformação $Y_i = X_i/m$, sendo m também uma constante. Basta usar $c = 1/m$ e o resultado é o mesmo.
- III) A soma do módulo dos desvios dos dados em relação a uma constante arbitrária c , terá um valor mínimo se $c = Md(X)$.

Prova:

- I) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, então a mediana de Y_i é dado por:

$$\begin{aligned} Md(Y) &= Y_{(\frac{n+1}{2})} \\ &= X_{(\frac{n+1}{2})} \pm c \\ &= Md(X) \pm c. \quad \text{c.q.d.} \end{aligned}$$

- II) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$, então a mediana de Y_i é dado por:

$$\begin{aligned} Md(Y) &= Y_{(\frac{n+1}{2})} \\ &= X_{(\frac{n+1}{2})} \times c \\ &= Md(X) \times c. \quad \text{c.q.d.} \end{aligned}$$

- III) Considerando uma amostra X_1, X_2, \dots, X_n , e a soma do módulo dos desvios entre os dados e uma constante c , por $h(c) = \sum_{i=1}^n |X_i - c|$, tal que,

$$h_i(c) = |X_i - c| = \begin{cases} (X_i - c) & \text{se } X_i > c, \\ -(X_i - c) & \text{se } X_i < c. \end{cases}$$

Para minimizar $h(c)$ em relação a c , temos que

$$\frac{dh_i(c)}{dc} = -I_{\{X_i > c\}}(x) + I_{\{X_i < c\}}(x),$$

que resulta em

$$\frac{dh(c)}{dc} = \sum_{i=1}^n [I_{\{X_i < c\}}(x) - I_{\{X_i > c\}}(x)],$$

em que $I(x)$ representa a função indicadora. Logo,

$$\frac{dh_i(c)}{dc} = 0,$$

se

$$\begin{aligned} \sum_{i=1}^n I_{\{X_i < c\}}(x) &= \sum_{i=1}^n I_{\{X_i > c\}}(x) \\ n_- &= n_+. \end{aligned}$$

De modo que a igualdade $n_- = n_+$ só ocorrerá se c for igual a $Md(X)$, pois a quantidade de valores menores a c é igual a quantidade de valores maiores que c . Portanto, para $n = n_- + n_+$ um número par, c deverá ser um valor entre $X_{(n/2)}$ e $X_{(\frac{n+2}{2})}$, isto é,

$$c = Md(X) = \frac{X_{(n/2)} + X_{(\frac{n+1}{2})}}{2}, \quad \text{c.q.d.}$$

e se $n = n_- + n_+$ for um número ímpar,

$$c = Md(X) = X_{(\frac{n+1}{2})}. \quad \text{c.q.d.}$$

3.4 Moda

As medidas de posição até agora apresentadas não foram aplicadas para as variáveis qualitativas de um modo geral, apenas a mediana para o caso de variável quantitativa ordinal. Contudo, podemos apresentar um medida mais simples, que seja possível ser aplicada para todas as naturezas de variáveis apresentadas, definida a seguir.

Definição 3.5: Moda para natureza de dados discretizados

Seja uma amostra X_1, X_2, \dots, X_n , de uma população X_1, X_2, \dots, X_N , de tamanhos n e N , respectivamente, cuja natureza da variável é discretizada^a. Então a moda representa o valor que mais se repete em um conjunto de dados. Denotamos μ_o a moda populacional, e $Mo(X)$ a moda amostral.

^aEntendemos que uma variável tem natureza discretizada quando seus potenciais valores assumem em um conjunto enumerável ou categorizado, isto é, variável quantitativa discreta e variáveis qualitativas.

Dessa forma, podemos perceber que um conjunto de dados poderá ter mais de uma moda, isto é, se observarmos dois valores mais frequentes e iguais, teremos uma distribuição bimodal, três valores mais frequentes iguais, teremos uma distribuição trimodal, mais de três, uma distribuição multimodal, ou até mesmo uma distribuição amodal, quando todos os valores se repetem apenas uma vez.

Exemplo 3.8:

Observando as variáveis da Tabela 1.2, percebemos que para a variável Região a moda é Norte. Já para UF e Número de cidades a distribuição é amodal.

Para o caso das variáveis quantitativas contínuas, essa definição não se aplica, porque dificilmente dois valores serão iguais para esse tipo de variável. O que faz pensar que dois valores sejam iguais em uma variável quantitativa contínua é a limitação do instrumento de medida. Basta perceber que dois valores possivelmente iguais, se mensurados por outros instrumentos de medidas mais precisos, os valores serão diferentes à medida que o número de dígitos nas casas decimais aumentam. Assim, faz-se necessário pensarmos em uma definição para a moda como sendo o valor com alta densidade de observações em sua proximidade. Uma forma de determinarmos um estimador para μ_o em variáveis contínuas é por meio do histograma de frequências. Inicialmente, determinamos a classe de maior frequência para os dados agrupados com intervalo de classe, para determinarmos a moda. Se todas as classes apresentarem mesma frequência, não haverá moda. A classificação quanto a distribuição segue a mesma mencionada anteriormente, isto é, amodal, unimodal, bimodal, trimodal ou multimodal. A moda baseada no histograma de frequência é também chamada de moda de Czuber.

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

Observamos pela Figura 3.5, que o valor da moda é $Mo = LI_{Mo} + x$, bastando determinar o valor x pela semelhança de triângulos, isto é,

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

Assim, a moda é determinada por:

$$Mo(X) = LI_{Mo} + \left\{ \frac{\Delta_1}{\Delta_1 + \Delta_2} \right\} \times c,$$

sendo LI_{Mo} o limite inferior da classe da moda, $\Delta_1 = f_{Mo} - f_{i_{ant}}$, $\Delta_2 = f_{Mo} - f_{i_{post}}$, f_{Mo} é a frequência absoluta da classe da moda, $f_{i_{ant}}$ frequência absoluta anterior à classe da moda, $f_{i_{post}}$ frequência posterior à classe da moda, e c a amplitude da classe.

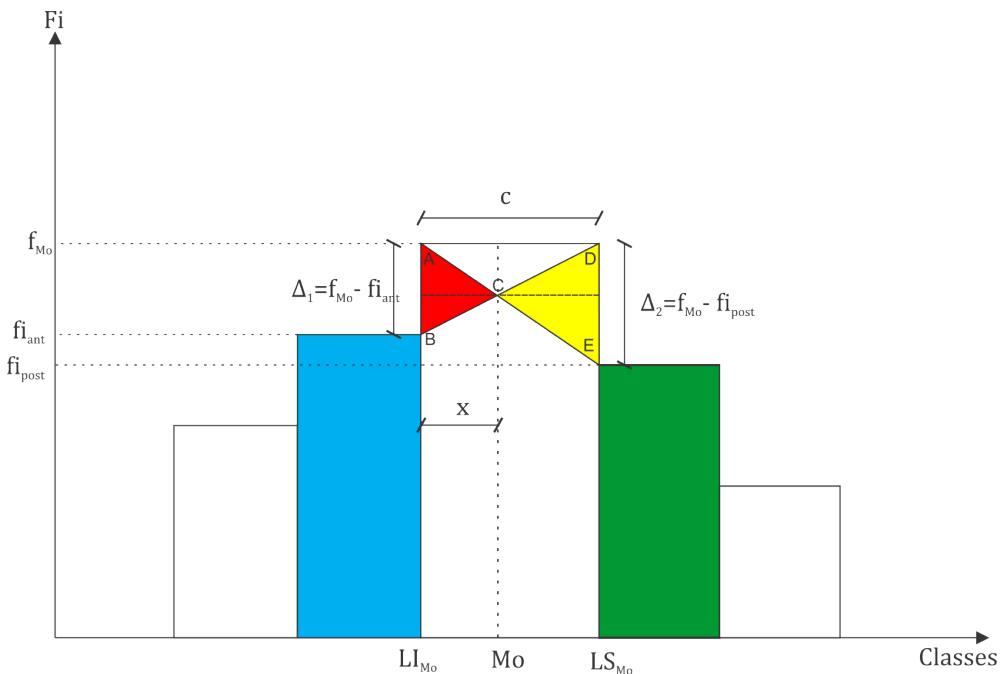


Figura 3.5: Histograma de frequência para a deduação do cálculo da moda.

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

$$Mo(X) = 3Md(X) - 2\bar{X}. \quad (3.14)$$

Formalmente, definimos

Definição 3.6: Moda para dados agrupados com intervalo de classe

Seja uma amostra $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ em ordem crescente de magnitude, de tamanho n , agrupados em k classes com pontos médios \tilde{X}_i e F_i frequências, para $i = 1, 2, \dots, k$ e $\sum_{i=1}^k F_i = n$, então a moda amostral é definida por:

$$Mo(X) = LI_{Mo} + \left\{ \frac{\Delta_1}{\Delta_1 + \Delta_2} \right\} \times c, \quad (3.15)$$

em que LI_{Mo} o limite inferior da classe da moda, $\Delta_1 = f_{Mo} - f_{i_{ant}}$, $\Delta_2 = f_{Mo} - f_{i_{post}}$, f_{Mo} é a frequência absoluta da classe da moda, $f_{i_{ant}}$ frequência absoluta anterior à classe da moda, $f_{i_{post}}$ frequência posterior à classe da moda, e c a amplitude da classe.

Vejamos algumas características sobre a moda, que seguem:

- A mediana não é influenciada por valores extremos, desde que estes não pertençam a classe modal;
- Uma medida que pode ser obtida em distribuições de frequências que apresentam classe com limites indefinidos;
- o resultado da moda é obtida na mesma escala da variável em estudo;

- a moda é menos informativa que a média, por não levar em consideração os valores observados;
- a moda pode ser calculada para todas as naturezas de variáveis;
- a moda é a medida mais simples dentre as apresentadas;

Exemplo 3.9:

Observando a tabela de frequência dos dados de temperatura do anel de vedação do foguete Challenger no Exemplo 2.1, percebemos que há duas classes de maior frequência, porém classes vizinhas. Perceberemos pelo cálculo que na realidade haverá apenas uma moda nesse caso. Vejamos, o cálculo da moda para a classe 4,

$$Mo(X) = 57,5 + \left\{ \frac{12 - 4}{(12 - 4) + (12 - 12)} \right\} \times 10,6 = 68,1^{\circ}\text{F}.$$

Vejamos a moda para a classe 5,

$$Mo(X) = 67,1 + \left\{ \frac{12 - 12}{(12 - 5) + (12 - 12)} \right\} \times 10,6 = 68,1^{\circ}\text{F}.$$

O que poderia ser feito para que não gerasse confusão quanto a distribuição ser unimodal ou bimodal, era reagrupar os dados com um outro número de classes.

Para complementarmos essas características, vamos apresentar algumas propriedades da moda no Teorema 3.2. Iremos usar a Definição 3.5 como referência, porém, para os demais casos, os resultados são similares.

Teorema 3.3: Propriedades da moda

Baseado na Definição 3.5, e considerando c uma constante, então:

- I) Se para uma amostra X_1, X_2, \dots, X_n em ordem crescente de magnitude, a moda representa o valor de maior frequência e representado por $Mo(X)$, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, a moda é dada por $Mo(Y) = Mo(X) \pm c$;
- II) Se para uma amostra X_1, X_2, \dots, X_n em ordem crescente de magnitude, a moda representa o valor de maior frequência e representado por $Mo(X)$, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, a nova moda é dada por $Mo(Y) = Mo(X) \times c$. Esse resultado vale também para a transformação $Y_i = X_i/m$, sendo m também uma constante. Basta usar $c = 1/m$ e o resultado é o mesmo.

Prova:

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

Seja

$$Y_i = X_i \pm c,$$

então

$$Mo(Y) = Mo(X) \pm c.$$

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

Sejam

$$Y_i = cX_i,$$

então

$$Mo(Y) = cMo(X).$$

Por fim, apresentamos o Exemplo 3.10 de Magalhães e Lima (2015), para termos uma noção sobre essas três medidas de posição, a seguir.

Exemplo 3.10: Retirado de Magalhães e Lima (2015)

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

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

Qual a companhia mais adequada?

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

Exercícios propostos

Exercício 3.1: A tabela abaixo apresenta a distribuição de frequências das notas (em pontos) obtidas num teste de matemática, realizado por 50 estudantes.

Notas	F_i
0 — 2	4
2 — 4	12
4 — 6	15
6 — 8	13
8 — 10	6

Apresente o cálculo para todas as medidas de posição estudadas e as interprete.

Solução na página 213

Exercício 3.2: Observando as expressões $\sum_{i=1}^n (X_i - \bar{X})^2$ e $\sum_{i=1}^n (X_i - \mu)^2$, em que $\bar{X} = \sum_{i=1}^n X_i/n$ e $\mu = \sum_{i=1}^N X_i/N$, representam a média amostral e a média populacional, respectivamente, qual das duas somas de quadrados representa o menor valor para soma?

Solução na página 213

Exercício 3.3: Considere uma amostra X_1, X_2, \dots, X_n de tamanho n , do qual conseguimos computar as medidas de posição $\bar{X}, Md(X)$ e $Mo(X)$, isto é, a média, mediana e moda, respectivamente, bem como a variância S_X^2 . Considere também as transformações $Y_i = X_i - \bar{X}$ e $Z_i = (X_i - \bar{X})/S$, para $i = 1, 2, \dots, n$, sendo $S = \sqrt{S^2}$. Apresente as medidas de posição para essas transformações, Y e Z .

Solução na página 213

Exercício 3.4: Uma empresa de telecomunicações concedeu 10% de aumento de salário a todos os seus funcionários. A média salarial foi de 1.500 reais antes do reajuste. Qual será a nova média salarial após o reajuste?

Solução na página 214

Exercício 3.5: Uma prova de estatística foi aplicada com pontuação de 0 a 10 pontos, sendo que a média das notas de todos os alunos de uma turma foi de 5,8 pontos. Se a média das mulheres é 6,3 pontos e a dos rapazes é 4,3 pontos, então qual a porcentagem de mulheres na turma?

Solução na página 214

Exercício 3.6: Uma prova de estatística foi aplicada com pontuação de 0 a 10 pontos, numa turma com 30 alunos, sendo uma média de 70 pontos. Nenhum dos alunos obteve nota inferior a 60 pontos. Dessa forma, qual o número máximo de alunos que podem ter obtido nota igual a 90 pontos?

Solução na página 215

Capítulo 4

Medidas de dispersão

4.1 Introdução

Iniciamos esse capítulo, motivados por três conjuntos de dados que resultaram em mesma média aritmética propositalmente. Vejamos o Código R 4.1 a seguir. Esse código apresenta uma simulação de três grupos com 10 observações cada um. Os grupos formados foram gA, gB e gC, com suas respectivas médias $\bar{X}_{gA} = 9,018686$ unid., $\bar{X}_{gB} = 9,018686$ unid. e $\bar{X}_{gC} = 9,018686$ unid., respectivamente. Dessa forma, poderíamos concluir que os grupos de dados são iguais? A resposta é não. Isso significa, que essa medida de posição apenas não consegue caracterizar completamente os grupos de dados.

Código R 4.1

Script:

```
1 # Semente de simulação
2 # set.seed(10)
3
4 # Dados dos grupos gerados
5 set.seed(10); gA <- rnorm(10, mean = 10, sd = 2)
6 set.seed(10); gB <- rnorm(9, mean = 10, sd = 4)
7 set.seed(10); gC <- rnorm(9, mean = 10, sd = 6)
8
9 # Media do grupo gA
10 (media.gA <- mean(gA))
11
12 # gerando gB[10] e media de gB
13 gB[10] <- gB10 <- length(gA) * media.gA - sum(gB)
14 (media.gB <- mean(gB))
15
16
17 # gerando gC[10] e media de gC
18 gC[10] <- gc10 <- length(gA) * media.gA - sum(gC)
19 (media.gC <- mean(gC))
20
21 # Apresentando os dados
22 gA; gB; gC
```

Os dados para os três grupos são apresentados na Tabela 4.1. Ao observarmos com mais detalhes os valores observados nos grupos, percebemos que estes não são iguais. Logo, poderíamos erroneamente afirmar que estes grupos eram semelhantes, simplesmente olhando para as suas médias. O que ocorre é uma variabilidade nos dados diferenciada em cada grupo, e que a medida de dispersão não consegue caracterizá-la. Dizemos que variabilidade é a dispersão com que ocorre nos dados, e as medidas responsáveis em expressar essa variabilidade, chamamos na estatística descritiva de medidas de dispersão. Geralmente, usamos a representação dessa dispersão em torno de um valor central nos dados, que em nosso caso, será a média aritmética, devido a algumas propriedades matemáticas e estatísticas que essa medida tem, e que será vista ao longo de todo o livro.

A seguir, mostraremos algumas medidas de dispersão que auxiliarão na caracterização dos da-

Tabela 4.1: Dados dos três grupos simulados pelo Código R 4.1.

gA	gB	gC
10,037492	10,074985	10,1124770
9,631495	9,262990	8,8944847
7,257339	4,514678	1,7720167
8,801665	7,603329	6,4049937
10,589090	11,178181	11,7672708
10,779589	11,559177	12,3387658
7,583848	5,167695	2,7515429
9,272648	8,545296	7,8179439
6,746655	3,493309	0,2399639
9,487043	18,787223	28,0874035
$\bar{X}_{gA} = 9,018686$		$\bar{X}_{gB} = 9,018686$
$\bar{X}_{gC} = 9,018686$		

dos, bem como no auxílio a fundamentação de temas tão importantes como a inferência estatística e teoria de decisão, abordado em capítulos posteriores.

4.2 Amplitude total ou Amplitude

A primeira medida de dispersão que definiremos é a amplitude ou amplitude total, denotada por A ou A_t . Iremos apresentar, três definições sobre a amplitude baseadas nos valores observados da população, da amostra, e em dados agrupados sem e com intervalo de classe. Vejamos o primeiro caso, pensando em uma população, apresentada na Definição 4.1.

Definição 4.1: Amplitude em uma população

Seja uma população X_1, X_2, \dots, X_N , de tamanho N , e em ordem crescente de magnitude temos $X_{(1)} = \min_i(X_i)$, $X_{(2)}, \dots, X_{(N)} = \max_i(X_i)$, para $i = 1, 2, \dots, N$. Então a amplitude de uma população, denotada por A_p , é definida por:

$$A_p = X_{(N)} - X_{(1)}. \quad (4.1)$$

Se desejarmos representar essa notação em termos de valor observado, temos $a_t = x_{(N)} - x_{(1)}$. Já usamos uma referência sobre a amplitude total ou amplitude, expressão (2.11), quando agrupamos os dados em intervalo de classes, para o caso das variáveis quantitativas contínuas. Vejamos o Exemplo 4.1 sobre os dados da taxa de desmatamento na Amazônia legal, compreendido entre 1988 a 07/12/2020.

Exemplo 4.1: Desmatamento da Amazônia Legal

Já mencionamos anteriormente, na Tabela 1.2, os dados de desmatamento da Amazônia legal. Se considerarmos que os elementos da população sejam os estados, portanto, temos as informações de todos os elementos, e assim, estamos diante de dados populacionais. Vamos assim, calcular a amplitude para a variável desmatamento acumulado, em km^2 , de acordo com a expressão (2.11), isto é,

$$A = 157.667,00 - 1.696,00 = 155971 \text{ } km^2.$$

Isso representa, uma variação de $155971 \text{ } km^2$. Observe que essa medida está na mesma escala da variável, e que se houve um outro conjunto de dados, em mesma unidade, poderíamos comparar qual a que apresentou maior dispersão.

Podemos representar a amplitude em termos amostras, como será apresentado na Definição 4.2, a seguir.

Definição 4.2: Amplitude em uma amostra

Seja uma amostra X_1, X_2, \dots, X_n , de tamanho n , e em ordem crescente de magnitude temos $X_{(1)} = \min_i(X_i)$, $X_{(2)}, \dots, X_{(n)} = \max_i(X_i)$, para $i = 1, 2, \dots, n$. Então a amplitude de uma população, denotada por A , é definida por:

$$A = X_{(n)} - X_{(1)}. \quad (4.2)$$

O que vai diferenciar a representação nas expressões (4.1) e (4.2), e o número de elementos, isto é, o tamanho populacional representado por “ N ”, e o tamanho amostral, representado por “ n ”. Porém, percebemos que a realização do cálculo é a mesma. Da mesma forma que a representação populacional, podemos representar a amplitude em uma amostra, em termos de valor observado como $a = x_{(n)} - x_{(1)}$. Vejamos o Exemplo 4.2, a seguir.

Exemplo 4.2:

Retornando aos dados amostrais simulados na Tabela 4.1, calculamos as amplitudes,

Grupo	Amplitude
gA	$A = 10,77959 - 6,746655 = 4,032934$ und.
gB	$A = 18,787223 - 3,493309 = 15,29391$ und.
gC	$A = 28,0874035 - 0,2399639 = 15,29391$ und.

Por esse resultado, podemos observar os grupos gB e gC que suas amplitudes foram iguais. Será que podemos, então afirmar que esses dois grupos são iguais? Quando observamos os dados, percebemos que os valores não são iguais. Isso ocorre por uma limitação nessa medida de dispersão. Vamos deixar para explorar essa situação mais a frente.

Para a amplitude em termos de dados agrupados, temos a situação em que as variáveis podem ser discretas ou contínuas. No caso, das variáveis quantitativas contínuas, os grupos são classes, e os valores passam a ser representados pelos seus pontos médios de cada classe. Assim, apresentamos na Definição 4.3, a amplitude para dados agrupados com e sem intervalo de classe.

Definição 4.3: Amplitude em dados agrupados

Seja uma amostra X_1, X_2, \dots, X_n , de tamanho n , agrupados em k grupos com variáveis X_i e F_i frequências, ou k classes com pontos médios \tilde{X}_i e F_i frequências, para $i = 1, 2, \dots, k$ e $\sum_{i=1}^k F_i = n$, então a amplitude de uma amostra, denotada por A , é definida por:

$$A = \begin{cases} X_{(k)} - X_{(1)}, & \text{Agrupados sem intervalo de classe,} \\ \tilde{X}_{(k)} - \tilde{X}_{(1)}, & \text{Agrupados com intervalo de classe,} \end{cases} \quad (4.3)$$

em que $X_{(k)} = \max_i(X_i)$, $X_{(1)} = \min_i(X_i)$, $\tilde{X}_{(k)} = \max_i(\tilde{X}_i)$, $\tilde{X}_{(1)} = \min_i(\tilde{X}_i)$, sendo \tilde{X}_i o ponto médio das classes.

Podemos representar a Definição 4.3 em termos populacionais, substituindo o tamanho n por N , como também representar a expressão em termos de valor observado, como mencionado na definições anteriores. Vejamos o Exemplo 4.3, a seguir.

Exemplo 4.3:

Retornando aos dados da Tabela 2.3, podemos calcular a amplitude do número de erros encontra-

dos em 20 conjunto de caracteres, usando a expressão (4.3), da seguinte forma:

$$A = 4 - 1 = 3 \text{ erros.}$$

Para os dados agrupados com intervalo de classe, apresentados no Exemplo 2.1, podemos calcular a amplitude da temperatura do anel de vedação de cada teste de acionamento do motor do foguete Challenger, da seguinte forma:

$$A = 84 - 31 = 52 \text{ }^{\circ}\text{F}.$$

No primeiro caso, usamos as próprias observações para o cálculo da amplitude. No segundo caso, usamos os pontos médios.

Podemos ainda apresentar algumas características sobre a amplitude, dos quais temos:

- o resultado da amplitude é dado na mesma unidade da variável em estudo;
- uma medida de dispersão facilmente calculada;
- limitada apenas as variáveis quantitativas;
- essa medida é muito utilizada em comparações múltiplas, cartas de controle em estatística de qualidade, dentre outras áreas;
- a amplitude pode ser utilizada como medida de dispersão para comparar a variabilidade de dados de dois ou mais grupos diferentes;
- a amplitude é sensível a dados discrepantes¹;
- a amplitude é limitada por levar em consideração apenas os valores extremos, e nada sobre as demais observações. Nesse caso, podem ocorrer situações como os apresentados no Exemplo 4.2, em que poderíamos erroneamente concluir que os grupos de dados gB e gC são iguais, uma vez que apresentam amplitude e média aritmética iguais;
- segundo Ferreira (2009, p. 36), a amplitude amostral, expressão (4.2), substima a amplitude populacional, expressão (4.1), uma vez que é pouco provável que uma amostra contenha os valores mínimo e máximo da população, portanto, a amplitude amostral é um estimador² viesado³ e ineficiente.

Complementando as características da amplitude, apresentamos algumas propriedades pelo Teorema 4.1 a seguir, do qual iremos usar a Definição 4.2 como base, e as demais seguem de forma similar.

¹Entendemos por dados discrepantes, as observações que estão distantes da massa de dados (maior parte dos dados). Esses dados quando influenciam as análises estatísticas, dizemos que estes dados são influentes.

²Entendemos por estimador uma função que depende apenas dos dados amostrais e que irá representar um parâmetro (característica populacional) desconhecida.

³Dizemos que um estimador é viesado se a esperança matemática desse estimador é diferente do parâmetro de interesse.

Teorema 4.1: Propriedades da Amplitude

Baseado na Definição 4.2, e considerando c uma constante, então:

- I) Se para uma amostra X_1, X_2, \dots, X_n , a amplitude é dada por $A_X = X_{(n)} - X_{(1)}$, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, a nova amplitude não se altera, isto é, $A_Y = A_X$.
- II) Se para uma amostra X_1, X_2, \dots, X_n , a amplitude é dada por $A_X = X_{(n)} - X_{(1)}$, então para uma transformação de $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$, a nova amplitude é dada por $A_Y = A_X \times c$. Esse resultado vale também para a transformação $Y_i = X_i/m$, sendo m também uma constante. Basta usar $c = 1/m$ e o resultado é o mesmo.

Prova:

- I) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, então a amplitude de Y_i é dado por:

$$\begin{aligned} A_Y &= Y_{(n)} - Y_{(1)} \\ &= (X_{(n)} \pm c) - (X_{(1)} \pm c) \\ &= X_{(n)} - X_{(1)} \\ &= A_X. \quad \text{c.q.d.} \end{aligned}$$

- II) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$, então a amplitude de Y_i é dado por:

$$\begin{aligned} A_Y &= Y_{(n)} - Y_{(1)} \\ &= (X_{(n)} \times c) - (X_{(1)} \times c) \\ &= (X_{(n)} - X_{(1)}) \times c \\ &= A_X \times c. \quad \text{c.q.d.} \end{aligned}$$

Devido ao problema encontrado no Exemplo 4.2, vamos apresentar algumas outras medidas que levem em consideração as demais variáveis bem como uma referência da posição central dos dados, que em nosso caso será a média aritmética.

4.3 Variância

Diante do Exemplo 4.2, percebemos que complementar a caracterização dos dados com a amplitude, se torna uma medida muito simples. Observamos que os grupos gB e gC apresentam mesmas médias e amplitudes. Assim, poderíamos dizer que os grupos são semelhantes. Mas quando observamos a Tabela 4.1, percebemos que estes são diferentes. Assim, vamos apresentar mais algumas medidas que englobem as demais variáveis e o valor central desses dados em seu cálculo, para apresentarmos medidas mais explicativas para dispersão de dados.

Considerando uma população X_1, X_2, \dots, X_N e sua respectiva amostra X_1, X_2, \dots, X_n , podemos considerar inicialmente o desvio médio como outra medida de dispersão, dada por:

$$DM_p = \sum_{i=1}^N (X_i - \mu), \quad (\text{Populacional}) \tag{4.4}$$

em que $\mu = \sum_{i=1}^N X_i / N$, e seu respectivo estimador é dado por:

$$DM = \sum_{i=1}^n (X_i - \bar{X}), \quad (\text{Amostral}) \quad (4.5)$$

em que $\bar{X} = \sum_{i=1}^n X_i / n$. Observamos agora, que diferentemente da amplitude, essa medida leva em consideração todos os elementos, seja da amostra ou da população, em relação a uma medida central, como preconizamos inicialmente a definição de uma medida de dispersão no início desse capítulo. O problema é que a expressão (4.5), como mostrado no Teorema 1.1, propriedade (V), sempre resulta em valor nulo para qualquer grupo amostral. De modo similar, a expressão (4.4) também $\sum_{i=1}^N (X_i - \mu) = 0$. Isso significa que essa medida não traz ganho algum a descrição dos dados, porque os desvios positivos anulam-se com os desvios negativos no somatório, sendo pois uma questão de problema algébrico. Para isso, podemos contornar essa situação inserindo uma função modular nessa medida anterior, e criar o módulo do desvio, dada por:

$$S_{|\mu|} = \sum_{i=1}^N |X_i - \mu|, \quad (\text{Populacional}) \quad (4.6)$$

e

$$S_{|\bar{X}|} = \sum_{i=1}^n |X_i - \bar{X}|. \quad (\text{Amostral}) \quad (4.7)$$

Desse modo, sabemos que $\sum_{i=1}^n |X_i - \bar{X}| \geq 0$, e agora temos uma medida que represente a dispersão com o qual os dados estão em torno da média. Quanto maior o módulo do desvio, mais disperso é o conjunto de dados. A questão do uso do módulo para resolver o problema da medida do desvio médio, nos gera uma outra dificuldade que poderemos ter mais a frente quando formos estudar inferência estatística. Tem situações que iremos precisar integrar, derivar, etc., dentre outras ferramentas matemáticas, que se torna mais fácil ao invés de usar o módulo, usarmos uma função quadrática na medida. Daí, surge uma outra medida de variabilidade que é a soma de quadrados, dada por:

$$SQ_p = \sum_{i=1}^N (X_i - \mu)^2, \quad (\text{Populacional}) \quad (4.8)$$

e

$$SQ = \sum_{i=1}^n (X_i - \bar{X})^2. \quad (\text{Amostral}) \quad (4.9)$$

Percebemos que a soma de quadrados amostral pode ser também expressa por:

$$SQ = \sum_{i=1}^n X_i^2 - \frac{1}{n} \sum_{i=1}^n X_i^2, \quad (4.10)$$

como pode ser provado no Teorema 1.1. Nesse último caso, podemos trabalhar sem o uso da informação da média, mas sim, apenas com as informações das observações. Essa medida apresenta uma outra informação interessante que é penalizar as observações quanto mais estiver distante do valor central. Observe que quando elevamos ao quadrado um alto desvio, esse valor se torna maior ainda, mas quando elevamos ao quadrado um desvio pequeno, esse valor não cresce tanto. Assim, conseguimos compreender quais os dados que estão mais dispersos em torno da média.

Baseado nessas informações, surge a variância populacional que é a média da soma de quadrados, denotada por σ^2 , definida a seguir.

Definição 4.4: Variância de uma população

Seja uma população X_1, X_2, \dots, X_N , de tamanho N , com parâmetro conhecido $\mu = \sum_{i=1}^N X_i / N$, então a variância populacional, denotada por σ^2 , é definida por:

$$\sigma^2 = \frac{SQ_p}{N}, \quad (4.11)$$

em que SQ_p é dado pela expressão (4.8), ou de forma similar,

$$\sigma^2 = \frac{\sum_{i=1}^N X_i^2 - \frac{1}{N} \sum_{i=1}^N X_i^2}{N}. \quad (4.12)$$

Podemos de forma intuitiva, pensar no estimador para σ^2 simplesmente substituindo “ N ” por “ n ” e SQ_p por SQ , usando as mesmas expressões do que foram usados na Definição 4.4, isto é,

$$\hat{\sigma}^2 = \frac{SQ}{n}. \quad (4.13)$$

Porém, existe uma propriedade nos estimadores, vista mais a frente, que é o seu viés. Dizemos que estimadores são viesados quando a sua esperança matemática não é igual ao parâmetro de interesse. Significa dizer em termos práticas, que mesmo se nós retirássemos todas as k amostras possíveis de uma população e para cada uma dessas amostras calculássemos a variância amostral, expressão (4.13), e posteriormente a média dessas variâncias, ou seja, $(\hat{\sigma}_1^2 + \hat{\sigma}_2^2 + \dots + \hat{\sigma}_k^2)/k$, esse valor não seria igual a σ^2 . Logo, $\hat{\sigma}^2$ é um estimador viesado. De outro modo, é um estimador defeituoso. Para contornar esse problema, usamos a seguinte definição para uma variância amostral não viesada, denotada por S^2 , e apresentada na Definição 4.5.

Definição 4.5: Variância de uma amostra

Seja uma população X_1, X_2, \dots, X_n , de tamanho n , com $\bar{X} = \sum_{i=1}^n X_i / n$, então a variância amostral, denotada por S^2 , é definida como:

$$S^2 = \frac{SQ}{n-1}, \quad (4.14)$$

em que SQ é dado pela expressão (4.9), ou de forma similar,

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - \frac{1}{n} \sum_{i=1}^n X_i^2}{n-1}. \quad (4.15)$$

Para elucidar essas informações, vejamos o Exemplo 4.5.

Exemplo 4.4:

Retornando aos dados amostrais simulados na Tabela 4.1, podemos calcular a variância amostral para cada um dos grupos. Vamos usar a expressão (4.15) para isso, que segue:

- Variância amostral para o grupo gA:

$$\begin{aligned} S_{gA}^2 &= \frac{10,037492^2 + \dots + 9,487043^2 - 1/9 \times (10,037492 + \dots + 9,487043)^2}{9-1} \\ &= \frac{831,0017 - 8133.67/9}{8} \\ &= 1,959404 \text{ und}^2 \end{aligned}$$

- Variância amostral para o grupo gB:

$$\begin{aligned} S_{\text{gB}}^2 &= \frac{10,074985^2 + \dots + 18,787223^2 - 1/9 \times (10,074985 + \dots + 18,787223)^2}{9 - 1} \\ &= \frac{988,9577 - 8133.67/9}{8} \\ &= 19,51007 \text{ und}^2 \end{aligned}$$

- Variância amostral para o grupo gC:

$$\begin{aligned} S_{\text{gC}}^2 &= \frac{10,1124770^2 + \dots + 28,0874035^2 - 1/9 \times (10,1124770 + \dots + 28,0874035)^2}{9 - 1} \\ &= \frac{1373,903 - 8133.67/9}{8} \\ &= 62,28176 \text{ und}^2 \end{aligned}$$

Podemos perceber que de fato os grupos gB e gC não são iguais, como podem verificar pelas suas variâncias amostrais, uma vez que isso poderia ter ocorrido pelo resultado do Exemplo 4.2. A dispersão das informações se torna mais detalhado, porque agora a medida leva em consideração a todas as observações.

Para o caso de dados agrupados, apresentamos a seguir a notação para o cálculo da variância, pela Definição 4.6.

Definição 4.6: Variância em dados agrupados

Seja uma amostra X_1, X_2, \dots, X_n , de tamanho n , agrupados em k grupos com variáveis X_i e frequência F_i , ou k classes com pontos médios \tilde{X}_i e F_i frequências, para $i = 1, 2, \dots, k$ e $\sum_{i=1}^k F_i = n$, então a variância de uma amostra, denotada por S^2 , é definida por:

$$S^2 = \begin{cases} \frac{\sum_{i=1}^k (X_i - \bar{X})^2 \times F_i}{\sum_{i=1}^k F_i - 1}, & \text{agrupados sem intervalo de classe,} \\ \frac{\sum_{i=1}^k (\tilde{X}_i - \bar{X})^2 \times F_i}{\sum_{i=1}^k F_i - 1}, & \text{agrupados com intervalo de classe,} \end{cases} \quad (4.16)$$

sendo \tilde{X}_i o ponto médio das classes, $\bar{X} = \sum_{i=1}^k X_i F_i / \sum_{i=1}^k F_i$ e $\tilde{\bar{X}} = \sum_{i=1}^k \tilde{X}_i F_i / \sum_{i=1}^k F_i$, ou se forma similar,

$$S^2 = \begin{cases} \frac{\sum_{i=1}^k X_i^2 \times F_i - \frac{1}{\sum_{i=1}^k F_i} (\sum_{i=1}^k X_i F_i)^2}{\sum_{i=1}^k F_i - 1}, & \text{agrupados sem intervalo de classe,} \\ \frac{\sum_{i=1}^k \tilde{X}_i^2 \times F_i - \frac{1}{\sum_{i=1}^k F_i} (\sum_{i=1}^k \tilde{X}_i F_i)^2}{\sum_{i=1}^k F_i - 1}, & \text{agrupados com intervalo de classe.} \end{cases} \quad (4.17)$$

Podemos representar a Definição 4.6 em termos populacionais, substituindo o tamanho n por N e considerando o denominador apenas como $\sum_{i=1}^k F_i - 1$ ao invés de $\sum_{i=1}^k F_i - 1$, tal que $\sum_{i=1}^k F_i = N$. Podemos também representar a expressão em termos de valor observado, como mencionado na definições anteriores.

Vejamos algumas características da variância:

- a unidade da variância está na escala ao quadrado da unidade da variável;
- limitada apenas as variáveis quantitativas;

- a variância é sempre uma medida positiva, exceto quando todos os valores são iguais que resultam em uma variância nula;
- quanto mais próximo de zero a variância for, mais concentrado os dados estão em torno da média, ao passo que, à medida que a variância se distancia de zero, mas disperso os dados estão em torno da média;
- devido as suas propriedades matemáticas, algumas mencionadas anteriormente, bem como a quantidade de técnicas estatísticas que empregam essa medida, a torna como a mais conhecida dentre as medidas de dispersão;
- uma vez que a média é sensível aos dados, a variância também é sensível, uma vez que esta depende da média.

Pelo Teorema 4.2, apresentamos algumas propriedades da variância a seguir, do qual iremos usar a Definição 4.5 como base, e as demais seguem de forma similar.

Teorema 4.2: Propriedades da Variância

Baseado na Definição 4.5, e considerando c uma constante, então:

- I) Se para uma amostra X_1, X_2, \dots, X_n , a variância é dada por S_X^2 , expressão (4.14), então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, então a nova variância não se altera, isto é, $S_Y^2 = S_X^2$.
- II) Se para uma amostra X_1, X_2, \dots, X_n , a variância é dada por S_X^2 , expressão (4.14), então para uma transformação de $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$, então a nova variância é dada por $S_Y^2 = c^2 \times S_X^2$. Esse resultado vale também para a transformação $Y_i = X_i/m$, sendo m também uma constante. Basta usar $c = 1/m$ e o resultado é o mesmo.

Prova:

- I) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, então a variância de Y_i é dado por:

$$\begin{aligned} S_Y^2 &= \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1} \\ &= \frac{\sum_{i=1}^n [(X_i \pm c) - (\bar{X} \pm c)]^2}{n-1} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \\ &= S_X^2. \quad \text{c.q.d.} \end{aligned}$$

- II) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$, então a variância de Y_i é dado por:

$$\begin{aligned} S_Y^2 &= \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1} \\ &= \frac{\sum_{i=1}^n [(X_i \times c) - (\bar{X} \times c)]^2}{n-1} \\ &= \frac{\sum_{i=1}^n c^2 (X_i - \bar{X})^2}{n-1} \\ &= c^2 \times \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}, \quad (\text{Teorema 1.1, (I)}) \\ &= c^2 \times S_X^2. \quad \text{c.q.d.} \end{aligned}$$

4.4 Desvio padrão

A variância apesar de ter resolvido alguns dos problemas mencionados anteriormente, para uma medida de dispersão, apresenta sua unidade ao quadrado da unidade da variável em estudo, isso significa que se tivermos usando uma variável na escala de metros, a dispersão dada pela variância estará na escala de área, isto é, em metros ao quadrado. Isso se torna difícil a percepção de dispersão quando observamos os dados. Dessa forma, surge a medida do desvio padrão, definida a seguir.

Definição 4.7: Desvio padrão

Seja uma amostra X_1, X_2, \dots, X_n , de uma população X_1, X_2, \dots, X_N , de tamanhos n e N , respectivamente, com parâmetro $\mu = \sum_{i=1}^N X_i / N$ e seu estimador $\bar{X} = \sum_{i=1}^n X_i / n$, então o desvio padrão é definido por:

$$\sigma = \sqrt{\sigma^2}, \quad (\text{População}) \quad (4.18)$$

em que σ é apresentado na Definição 4.4,

$$S = \sqrt{S^2}, \quad (\text{Amostra}) \quad (4.19)$$

em que S é apresentado na Definição 4.5.

Com o desvio padrão, podemos verificar a medida de variabilidade na mesma unidade da variável. Cabe destacar que a expressão (4.18) mede a variabilidade das observações em torno da média populacional. Porém na prática, não conhecemos o parâmetro μ nem muito menos temos informações de todas as observações. Com isso usamos como estimador de σ , o desvio padrão amostral dado na expressão (4.19), que se baseia em apenas uma amostra. Vejamos o exemplo a seguir.

Exemplo 4.5:

Retornando ao Exemplo 4.2, podemos então calcular os desvios padrões dos grupos, que segue:

- Desvio padrão amostral para o grupo gA:

$$S_{gA} = \sqrt{1,959404} = 1,399787 \text{ unid.}$$

- Desvio padrão amostral para o grupo gB:

$$S_{gB} = \sqrt{19,51007} = 4,41702 \text{ und.}$$

- Desvio padrão amostral para o grupo gC:

$$S_{gC} = \sqrt{62,28176} = 7,89188 \text{ und.}$$

Considerando que as unidades dos grupos são iguais, bem como as suas médias, podemos concluir que o grupo gA apresenta menor dispersão. Claro que esse resultado, poderia ter sido observado pela variância. A diferença é que conseguimos entender na unidade da variável essa dispersão.

Contudo, quando iremos comparar grupos de dados e verificar qual grupo apresenta maior variabilidade, devemos ter muito cuidado ao usar o desvio padrão ou a variância, sob dois aspectos:

- 1) Os grupos de observações devem estar na mesma unidade de mensuração;
- 2) A média desses grupos devem ser iguais.

O primeiro aspecto está muito claro, uma vez que não temos, por exemplo, como comparar uma unidade em gramas e saber se a dispersão desses dados é maior ou menor quando se compara com outro conjunto de dados cuja unidade esteja na escala de comprimento. O segundo aspecto está limitado devido a forma de como foram calculados o desvio padrão e a variância. A soma de seus desvios levam em consideração a média. Assim, quando comparamos dois desvios padrões de duas amostras de uma população, em que temos o desvio padrão $S_1^2 = 10 \text{ unid}$ para a amostra 1, e $S_2^2 = 20 \text{ unid}$ para a amostra 2. Não podemos afirmar que a amostra 2 apresenta maior dispersão que a amostra 1, isso porque não sabemos o quanto esse valor representa em relação a média. Supomos que a média da amostra 1 seja $\bar{X}_1 = 100 \text{ unid}$ e para a amostra 2, seja $\bar{X}_2 = 50 \text{ unid}$. Desse modo, observemos que para a amostra 1, o desvio padrão representa apenas 10% do valor da média. Já na amostra 2, o desvio padrão representa 40% da média, uma variação muito mais considerável, isto é, os dados na amostra 2 são muito mais dispersos em torno da média. Isso justifica então, a criação de uma medida de dispersão relativa à média, que será definida na próxima seção.

Vejamos algumas características do desvio padrão, que segue:

- a unidade do desvio padrão está na mesma escala da unidade da variável em estudo;
- limitada apenas as variáveis quantitativas;
- uma vez que a média é sensível aos dados, o desvio padrão também é sensível, uma vez que esta depende da média;
- embora a variância amostral, S^2 seja um estimador não viesado para a variância populacional σ^2 , o desvio padrão amostral S , que é derivado de S^2 , é um estimador viesado do desvio padrão populacional σ ;
- assim como a variância, o desvio padrão é sempre uma medida positiva, exceto quando todos os valores são iguais que resultam em uma variância nula;
- assim como na variância, quanto mais próximo de zero o desvio padrão for, mais concentrado os dados estão em torno da média, ao passo que, à medida que o desvio padrão se distancia de zero, mais disperso os dados estão em torno da média.

Complementando as características do desvio padrão, apresentamos algumas propriedades no Teorema 4.1, do qual iremos usar a Definição 4.2 como base, e as demais seguem de forma similar.

Teorema 4.3: Propriedades do Desvio Padrão

Baseado na Definição 4.2, e considerando c uma constante, então:

- I) Se para uma amostra X_1, X_2, \dots, X_n , a amplitude é dada por $A_X = X_{(n)} - X_{(1)}$, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$, então o novo desvio padrão não se altera, isto é, $S_Y = S_X$.
- II) Se para uma amostra X_1, X_2, \dots, X_n , a amplitude é dada por $A_X = X_{(n)} - X_{(1)}$, então para uma transformação de $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$, então a nova amplitude é dada por $A_Y = A_X \times c$. Esse resultado vale também para a transformação $Y_i = X_i/m$, sendo m também uma constante. Basta usar $c = 1/m$ e o resultado é o mesmo.

Prova:

- I) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \pm c$, para i

$= 1, 2, \dots, n$, então o desvio padrão de Y_i é dado por:

$$\begin{aligned} S_Y &= \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \\ &= \sqrt{\frac{\sum_{i=1}^n [(X_i \pm c) - (\bar{X} \pm c)]^2}{n-1}} \\ &= \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \\ &= S_X. \quad \text{c.q.d.} \end{aligned}$$

II) Considerando uma amostra X_1, X_2, \dots, X_n , e c uma constante, e que $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$, então a variância de Y_i é dado por:

$$\begin{aligned} S_Y &= \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \\ &= \sqrt{\frac{\sum_{i=1}^n [(X_i \times c) - (\bar{X} \times c)]^2}{n-1}} \\ &= \sqrt{\frac{\sum_{i=1}^n c^2 (X_i - \bar{X})^2}{n-1}} \\ &= \sqrt{c^2 \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}, \quad (\text{Teorema 1.1, (I)}) \\ &= c \times \sqrt{S_X^2}, \\ &= c \times S_X. \quad \text{c.q.d.} \end{aligned}$$

4.5 Coeficiente de Variação

As medidas de variabilidade tais como a variância e desvio padrão, são conhecidas como medidas de dispersão absoluta. Diante do que foi exposto no fim da seção anterior sobre alguns problemas do desvio padrão, apresentamos mais uma medida de dispersão, Definição 4.8, agora uma medida relativa chamada de Coeficiente de Variação (CV), do qual pode ser usada para comparar a variabilidade entre quaisquer grupo de dados.

Definição 4.8: Coeficiente de Variação

Seja uma amostra X_1, X_2, \dots, X_n , de uma população X_1, X_2, \dots, X_N , de tamanhos n e N , respectivamente, com parâmetros $\mu = \sum_{i=1}^N X_i / N$ e $\sigma^2 = \sum_{i=1}^N (X_i - \mu)^2 / (N - 1)$ e seus estimadores $\bar{X} = \sum_{i=1}^n X_i / n$ e $S = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1)}$, respectivamente. Então o coeficiente de variação é definido por:

$$CV_p = \frac{\sigma}{\mu} \times 100, \quad (\text{População}) \quad (4.20)$$

e

$$CV = \frac{S}{\bar{X}} \times 100. \quad (\text{Amostra}) \quad (4.21)$$

Essa medida permite com que possamos comparar a dispersão com dois ou mais grupos com características completamente diferentes e com médias diferentes. Vejamos o Exemplo 4.6, para ilustrar essa característica.

Essa medida permite com que possamos comparar a dispersão com dois ou mais grupos com características completamente diferentes e com médias diferentes. Vejamos o seguinte exemplo a seguir.

Exemplo 4.6:

Com a medida do coeficiente de variação, podemos comparar a dispersão dos dados do Exemplo 2.1 com a dispersão do grupo gA da Tabela 4.1. Para o primeiro conjunto de dados, podemos calcular a média e o desvio padrão do número de erros encontrados em 20 conjuntos de caracteres, dados por:

$$\bar{X}_e = \frac{3 + 1 + \dots + 1}{20} = 1,7 \text{ erros,}$$

e

$$\begin{aligned} S_e &= \sqrt{\frac{3^2 + 1^2 + \dots + 1^2 - 1/20 \times (3 + 1 + \dots + 1)^2}{19}} \\ &= 1,174286 \text{ erros,} \end{aligned}$$

respectivamente. No caso dos dados do grupo gA, nós já temos os resultados da média e desvio padrão, dados na Tabela 4.1 e no Exemplo 4.5, respectivamente. Desse modo, comparando a dispersão dos dois grupos pelo coeficiente de variação, temos:

Dados	Coeficiente de Variação (CV)
gA	$CV_{gA} = \frac{1,399787}{9,018686} \times 100 = 15,52\%$
Número de erros	$CV_e = \frac{1,174286}{1,7} \times 100 = 69,08\%$

Nesse caso, percebemos que os dados de gA tem menor variabilidade do que os dados do número de erros, e com isso, esses dados são melhor representado pela sua média amostral quando se comparado com o outro grupo de dados.

Mesmo resolvendo alguns problemas existentes nas medidas variância e desvio padrão, o coeficiente de variação apresenta algumas características importantes, que seguem:

- O CV é adimensional e uma medida de dispersão relativa;
- Essa medida pode ser utilizada para comparar a dispersão entre grupos diferentes de dados;
- Como o CV é uma medida de dispersão relativa, isto é, o desvio padrão ponderado pela média. Isso significa que o CV calcula o quanto representa a dispersão (o desvio padrão) representa à média. Dessa forma, o CV se torna limitado a variáveis em que a escala de mensuração das observações em que fornece um zero absoluto ou uma origem significativa. Por exemplo, a temperatura em graus celsius (°C), uma observação igual a 0°C não significa ausência de temperatura, logo, o CV para esse tipo de variável não faz sentido. Já o a variável peso em quilos, isto é, o valor 0kg representa ausência de peso, de outro modo, esse tipo de variável permite magnitudes de valores na escala, tais como, uma observação de 40kg é o dobro de uma observação de 20kg. Assim, podemos utilizar o coeficiente de variação para verificar a dispersão da variável peso;
- O CV pode superar o 100%. Isso ocorre quando o desvio padrão é maior do que a média. Dizemos que esses superdispersos, um exemplo, são dados de contagem que seguem uma distribuição de Poisson.

As propriedades do CV levam em consideração as propriedade de \bar{X} e S , que já foram demonstradas. Assim, ficam para estudo no Exercício proposto 4.1, a demonstração para as propriedades do CV .

4.6 Erro padrão da média

Para iniciarmos uma última ideia sobre medidas de dispersão, dentre as medidas básicas, vamos iniciar como motivação, a Definição 4.9.

Definição 4.9: Erro da média amostral

Seja uma amostra X_1, X_2, \dots, X_n , de uma população com parâmetro μ , que representa a média populacional, e seu estimador $\bar{X} = \sum_{i=1}^n X_i/n$, então definimos o erro da média amostral, denotado por $EA_{\bar{X}}$, da seguinte forma:

$$EA_{\bar{X}} = \bar{X} - \mu. \quad (4.22)$$

A medida expressa em (4.22) representa o erro de assumirmos a média amostral como um representante da média populacional. O desvio padrão de $EA_{\bar{X}}$ é o que chamamos de erro padrão da média, definido a seguir.

Definição 4.10: Erro padrão da média (Populacional)

Seja uma amostra X_1, X_2, \dots, X_n , de uma população cujos parâmetros μ e σ , representam a média e o desvio padrão populacional, respectivamente, então o erro padrão da média, denotada por $\sigma_{\bar{X}}$, é definido como:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}, \quad (4.23)$$

em que n representa o tamanho da amostra.

Quando fazemos um comparativo entre o desvio padrão amostral e o erro padrão da média, entendemos que a primeira medida reflete a variabilidade de cada observação em torno da média amostral. Já o erro padrão da média representa a variabilidade de cada média amostral de todas amostra possíveis, em relação a média populacional.

Dessa forma, surgem alguns problemas para determinar a variabilidade da média amostral em torno da média populacional usando a expressão (4.24). Primeiro, é praticamente impossível realizar todas as amostras possíveis de uma população para computar a sua média. Se isso é possível, não precisaremos de amostra, uma vez que temos todas as informações da população, e então, estamos diante de um censo. Os outros fatores, podemos destacar que na prática, realizamos apenas uma amostra para análise das informações, e que o desvio padrão populacional geralmente é desconhecido, e assim, torna-se inviável o cálculo de $\sigma_{\bar{X}}$. Uma alternativa é usar o estimador S ao invés de σ , surgindo então um estimador para o erro padrão da média populacional, definido a seguir.

Definição 4.11: Erro padrão da média (Amostral)

Seja uma amostra X_1, X_2, \dots, X_n , de uma população cujos parâmetros μ e σ , representam a média e o desvio padrão populacional, respectivamente, então o erro padrão da média, denotada por $\sigma_{\bar{X}}$, é definido como:

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}, \quad (4.24)$$

em que n representa o tamanho da amostra, e S é o desvio padrão da Definição 5.25.

É fácil observar que à medida que $n \rightarrow N$, isto é, à medida que n aumenta, a média amostral tende a μ , logo o $EA_{\bar{X}} \rightarrow 0$. Isso significa que a média amostral é mais precisa porque se aproxima cada vez mais da média populacional. Assim, com apenas uma amostra poderemos ter uma estimativa do erro padrão da média, em que veremos um exemplo a seguir.

Exemplo 4.7:

Retornando ao Exemplo 4.2, podemos então calcular os erros padrões da média para as três amostras, que segue:

- Erro padrão da média amostral para o grupo gA:

$$S_{\bar{X}_{gA}} = 1,399787 / \sqrt{9} = 0,4665957 \text{ unid.}$$

- Erro padrão da média amostral para o grupo gB:

$$S_{\bar{X}_{gB}} = 4,41702 / \sqrt{9} = 1,47234 \text{ und.}$$

- Erro padrão da média amostral para o grupo gC:

$$S_{\bar{X}_{gC}} = 7,89188 / \sqrt{9} = 2,630627 \text{ und.}$$

Percebemos que a média de gA estima melhor o parâmetro μ , uma vez que o erro padrão da média foi o menor dentre os demais.

Desse modo, observamos que o erro padrão da média representa uma precisão com que a média amostral estimou o parâmetro μ . Além do erro padrão da média, há diversos outros erros padrões para outros estimadores de parâmetros diversos, sendo abordado mais a frente. Além do mais, essa medida será largamente usada na teoria de estimação e de decisão, tanto para a construção de intervalos de confiança, como também no desenvolvimento de testes de hipóteses, sendo também abordado nos próximos capítulos.

Exercícios propostos

Exercício 4.1: Com relação as propriedades do Coeficiente de Variação (CV), prove que:

- Se para uma amostra X_1, X_2, \dots, X_n , o coeficiente de variação, Definição 4.8, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$ e c uma constante, então o novo coeficiente de variação é igual a $CV_Y = S_X / (\bar{X} \pm c) \times 100$, em que \bar{X} e S_X são a média e o desvio padrão de X_i , $i = 1, 2, \dots, n$;
- Se para uma amostra X_1, X_2, \dots, X_n , o coeficiente de variação, Definição 4.8, então para uma transformação de $Y_i = X_i \times c$, para $i = 1, 2, \dots, n$ e c uma constante, então o novo coeficiente de variação não se altera, isto é $CV_Y = CV_X$. Esse resultado vale também para a transformação $Y_i = X_i/m$, sendo m também uma constante. Basta usar $c = 1/m$ e o resultado é o mesmo.

Solução na página 216

Exercício 4.2: A tabela abaixo apresenta a distribuição de frequências das notas (em pontos) obtidas num teste de matemática, realizado por 50 estudantes.

Notas	F_i
0 — 2	4
2 — 4	12
4 — 6	15
6 — 8	13
8 — 10	6

Apresente o cálculo para todas as medidas de dispersão estudadas e as interprete.

Solução na página 216

Exercício 4.3: Para uma amostra X_1, X_2, \dots, X_n de tamanho n . Desejamos usar a transformação $Y_i = \beta_0 + \beta_1 X_i$, para $i = 1, 2, \dots, n$, considerando β_0 e β_1 constantes. Então, apresente as relação entre as médias e os desvios padrões de X e Y .

Solução na página 216

Exercício 4.4: Sabemos que a conversão da temperatura de graus Celsius ($^{\circ}\text{C}$) para a escala de Fahrenheit ($^{\circ}\text{F}$) é dada por: $F = 9/5 \times C + 32$, considerando a variável F a temperatura em Fahrenheit, e C a temperatura em graus Celsius. Se medíssemos a temperatura de um peça no momento de fabricação, e verificássemos que em média a peça é fabricada com temperatura de 70°C e variância de $2(^{\circ}\text{C})^2$, como poderíamos representar essas medidas na escala Fahrenheit?

Solução na página 216

Exercício 4.5: Se tivéssemos estudando a variável temperatura em três escalas: graus Celsius, Fahrenheit e Kelvin, poderíamos calcular o coeficiente de variação para as três escalas? Explique.

Solução na página 216

Exercício 4.6: Considere uma amostra X_1, X_2, \dots, X_n de tamanho n , do qual conseguimos computar a média aritmética e variância amostral, sendo representadas por \bar{X}_n e S_n^2 , respectivamente, e que esses índices representam que estas medidas foram calculadas baseadas em um tamanho de amostra n . Por alguma situação precisamos adicionar mais uma variável a esta amostra, isto é, a variável X_{n+1} . Como poderíamos calcular as medidas \bar{X}_{n+1} e S_{n+1}^2 , partindo do pressuposto que só sabemos das informações \bar{X}_n , S_n^2 e X_{n+1} ? Em um segundo momento considere o Exemplo 2.1, os dados sem agrupamento de classes, de modo que foi realizado uma nova medição da temperatura do anel de vedação no acionamento do foguete Challenger, sendo aferido o valor $x_{37} = 63^\circ\text{F}$, use os resultados obtidos e determine \bar{x}_{n+1} e s_{n+1}^2 após desse novo dado as observações.

Solução na página 216

Exercício 4.7: Considere uma amostra X_1, X_2, \dots, X_n de tamanho n , do qual conseguimos computar as medidas de posição \bar{X} , $Md(X)$ e $Mo(X)$, isto é, a média, mediana e moda, respectivamente, bem como a variância S_X^2 . Considere também as transformações $Y_i = X_i - \bar{X}$ e $Z_i = (X_i - \bar{X})/S$, para $i = 1, 2, \dots, n$, sendo $S = \sqrt{S^2}$. Apresente as medidas de dispersão, desvio padrão e variância, para essas transformações, Y e Z .

Solução na página 216

Probabilidades

Após finalizarmos as principais ideias sobre a Estatística Descritiva, Capítulos 1 a 4, iniciamos por meio da probabilidade, os passos iniciais para a tomada de decisão por meio dos dados. Para isso, usaremos a Estatística Inferencial (Teoria da Estimação e Teoria da decisão), assuntos vistos nos Capítulos 9 e 10. Contudo, é imprescindível uma fundamentação teórica sobre a probabilidade, que é a base para a tomada de decisão.

A probabilidade vem aparecer como ramo da matemática no século XV, embora tenha surgido antes desse período. Entretanto, somente no século XVI, é que a teoria da probabilidade passa a ser estudada com profundidade, quando Jerônimo Cardano (1501-1576) passa a estudar problemas com os jogos de azar: cartas, dados, etc. Os jogadores de cassinos, tentavam encontrar meios de obter chances maiores de por exemplo ganhar um jogo, acertar um número ou uma carta. Daí surge a probabilidade para resolver esses problemas por meio dos matemáticos.

Já a estatística inicialmente, tentava identificar determinados problemas do Estado, como o número de nascidos e de mortos, determinação do número de pessoas do sexo masculino e feminino, etc. Entretanto, apenas no início do século XX, é que a probabilidade e a estatística passam a ser interligadas, isto é, a estatística agora necessita de técnicas probabilísticas para o estudo de dados.

Hoje, a estatística tem como um dos objetivos entender características atribuíveis a população de estudo. Por meio de um subconjunto (amostra) da população, a estatística tenta se aproximar dessas características (parâmetros) por meio da inferência, através dos estimadores (características atribuíveis a amostra). Entretanto, se basear numa amostra para entender a população, gera uma incerteza. E essa incerteza é medida por meio da teoria da probabilidade, pela qual toda a estatística é desenvolvida.

Inicialmente, faremos uma revisão sobre Teoria de conjuntos, já usando termos específicos dentro da probabilidade, como por exemplo, a definição de um Experimento aleatório, dentre outras. Isso porque se faz necessário o entendimento sobre o agrupamento de elementos, e a chance com que esses elementos podem ocorrer em um experimento.

5.1 Introdução à teoria de conjuntos no contexto probabilístico

Quando desejamos compreender algum fenômeno da natureza, tentamos estudá-lo por meio de um processo de observação chamado experimento. Para isso, definimos um experimento aleatório, Definição 5.1, a seguir.

Definição 5.1: Experimento Aleatório

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

Vejamos os Exemplos 5.1, 5.2 e 5.3, todos esses para exemplificar um experimento aleatório.

Exemplo 5.1:

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

Exemplo 5.2:

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

Exemplo 5.3:

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

Em um contexto aplicado, podemos estar interessados em estudar a resistência de um fio de cobre a uma determinada corrente. Para isso, replicamos diversas vezes esse fenômeno para medirmos a resistência, e assim, temos o que chamamos de experimento. Para que esse experimento não tenha resultados inconsistentes, usamos muitas vezes um laboratório para tentar controlar outras variáveis que possam perturbar o experimento, isto é, medimos a resistência do fio, de modo que a maior influência dessa variável para o experimento, seja devida a corrente aplicada ao final. Por mais que controlemos as condições possíveis do experimento, surgem sempre variáveis não controláveis ao sistema que foge do controle do pesquisador nesses casos, que são as variáveis não controláveis, Figura 5.1. Isso mostra que por mais que repliquemos o experimento, em mesmas condições, veremos que a medida da resistência do fio não será igual, devido a essas variáveis não controláveis, e que isso reflete em um componente aleatório, que por consequência, dizemos que estes tipos de experimentos são chamados de experimentos aleatórios.



Figura 5.1: Componente aleatória de um experimento aleatório.

Baseado, nos exemplos anteriores, percebemos pelo Exemplo 5.1, que não sabemos de fato qual o número da face superior que ocorrerá após o lançamento do dado. Mas sabemos, quais os resultados possíveis, que são: 1, 2, 3, 4, 5 e 6. O conjunto de todos esses resultados, chamaremos de Espaço amostral, apresentado na Definição 5.2, a seguir.

Definição 5.2: Espaço amostral

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

Cada um dos elementos do espaço amostral é representado por ω . Na Definição 5.5 apresentaremos o significado de evento. Contudo, podemos antecipar como um subconjunto de Ω . Assim, diremos que um determinado evento ocorrerá, se o resultado do experimento estiver nesse evento. Existem duas relações entre eventos que usaremos constantemente ao longo do conteúdo, que são:

- Continência: $A \subset B \Leftrightarrow \omega \in A \Rightarrow \omega \in B$;
- Equivalência: $A = B \Leftrightarrow A \subset B$ e $B \subset A$.

E que fique claro também, que a relação de elemento para conjunto é uma de pertinência, isto é, $\omega \in A$, que significa que ω é um elemento pertencente (ou membro) de A ; e que a relação entre conjuntos é uma relação de continência, isto é, $A \subset B$, significa que todo elemento de A é também elemento de B .

Retornando a Definição 5.2, podemos apresentar um outro espaço amostral, para o experimento dado no Exemplo 5.4, a seguir.

Exemplo 5.4:

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

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

Contudo, assim como definimos a natureza das variáveis no Capítulo 1, definimos também a natureza dos espaços amostrais de acordo os seus resultados, do qual podemos apresentá-la na Definição 5.3.

Definição 5.3: Espaços amostrais discretos e contínuos

Um espaço amostral é discreto se o conjunto dos possíveis resultados são finito ou infinito contável (ou enumerável). Um espaço amostral é dito contínuo se o conjunto dos possíveis resultados são infinitos não contável (ou não enumerável).

Vejamos o Exemplo 5.5, retirado de Montgomery e Runger (2016), para distinguir espaços amostrais discretos e contínuos, apresentado a seguir.

Exemplo 5.5: Câmera Flash

Considere um experimento em que você seleciona uma câmera de telefone celular e registra o tempo de recarga de um *flash*. Os resultados possíveis para o tempo dependem da resolução do temporizador e dos tempos máximo e mínimo de recarga. Entretanto, podemos definir inicialmente o espaço amostral em termos da reta real positiva (\mathbb{R}_+), isto é,

$$\Omega = \mathbb{R}_+ = \{x : x > 0\}.$$

Se soubermos que os tempos de recarga estão entre 1,5 e 5 segundos, podemos definir o espaço amostral da seguinte forma:

$$\Omega = \{x : 1,5 \leq x \leq 5\}.$$

Caso, consideremos o tempo de recarga como *baixo*, *médio* ou *alto*, reescrevemos o espaço amostral como:

$$\Omega = \{\text{baixo}, \text{médio}, \text{alto}\}.$$

Por fim, podemos considerar apenas o fato da câmera satisfazer ou não as especificações do tempo de recarga mínimo, e assim, podemos assumir como resultados para esse espaço amostral: *sim* ou *não*, isto é,

$$\Omega = \{\text{sim}, \text{não}\}.$$

Para as duas primeiras situações, temos exemplos de espaços amostrais contínuos, e nos dois últimos, exemplos de espaços amostrais discretos.

Entretanto, também podemos ter um conjunto qualquer A , que contém parte do elementos de Ω , isto é, $A \subset \Omega$, e que A passa a ser chamado de subconjunto de Ω , apresentado na Definição 5.4.

Definição 5.4: Subconjunto

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

$$A \subset B \Leftrightarrow A \subseteq B \text{ e } A \neq B.$$

Essa definição pode ser aplicada também a subconjuntos de Ω , como apresentado no Exemplo 5.6, a seguir.

Exemplo 5.6:

Sejam os subconjuntos de Ω do experimento aleatório apresentado no Exemplo 5.1, dos quais temos:

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

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

Definição 5.5: Evento

Todo subconjunto do espaço amostral (Ω), representado por letras latinas em maiúsculo, A , B , ..., é chamado de evento.

Vejamos o Exemplo 5.7, para um entendimento inicial sobre um evento, apresentado a seguir.

Exemplo 5.7:

Um evento retirado do espaço amostral do Exemplo 5.4 seria $A = \{(H, H, H), (H, H, T), (H, T, T)\}$, ou seja, o evento em que dos três arremessos de moedas, tenha saído "cara"na primeira moeda.

Um outro exemplo abordado em James (2004), pode exemplificar um evento dentro do círculo unitário, apresentado no Exemplo 5.8, a seguir.

Exemplo 5.8:

Escolher ao acaso um ponto no círculo de raio 1 centrado na origem. Então

$$\Omega = \text{círculo unitário} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

Vejamos alguns eventos para esse exemplo:

$$A = \text{"distância entre o ponto escolhido e a origem é" } \leq 1/2$$

$$B = \text{"distância entre o ponto escolhido e a origem é" } \geq 15$$

$$C = \text{"1ª Coordenada do ponto escolhido é maior que a 2ª".}$$

Se $\omega = (x, y)$ for um resultado do experimento, então ω pertencerá a A se, e somente se, $x^2 + y^2 \leq 1/4$. Pertencerá ao evento C se, e somente se, $x > y$. Nenhum ponto ω pertencerá a B , como pode ser observado pela Figura 5.2. Logo, temos:

$$A = \{(x, y) \in \Omega : \sqrt{x^2 + y^2} \leq 1/2\},$$

$$B = \emptyset = \text{conjunto vazio},$$

$$C = \{(x, y) \in \Omega : x > y\}.$$

Então, todo evento associado a este experimento pode ser identificado por um subconjunto do espaço amostral.

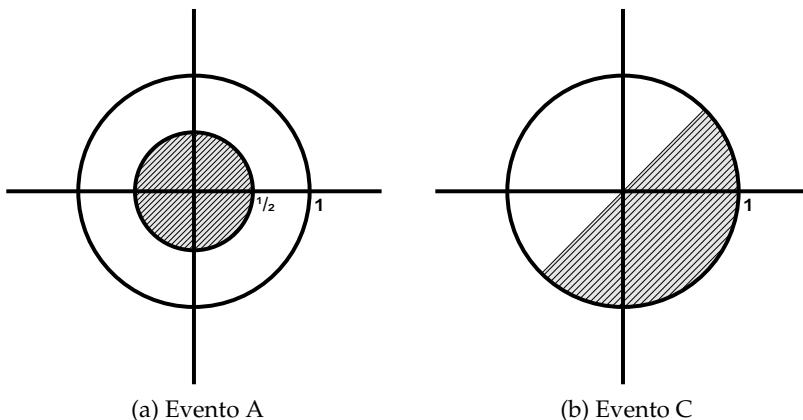


Figura 5.2: Escolha do ponto em um círculo unitário.

Dante, do que falamos sobre a definição de evento, podemos apresentar três eventos básicos: o evento certo, impossível e o elementar, apresentados na Definição 5.6, a seguir.

Definição 5.6: Evento certo, impossível e elementar

Seja Ω o espaço amostral do experimento. Então dizemos que Ω é o evento certo, e \emptyset é o evento impossível, e o evento $\{\omega\}$ é dito elementar.

Uma outra forma de definir o evento impossível é representá-lo como um conjunto vazio, apresentado na Definição 5.7, a seguir.

Definição 5.7: Conjunto Vazio

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

$$A = \{\omega \in \Omega : \omega \neq \omega\}. \quad (5.1)$$

Podemos perceber que todo conjunto vazio é um subconjunto de qualquer evento não vazio do espaço amostral, como pode ser apresentado no Teorema 5.1.

Teorema 5.1

Considere o conjunto vazio, \emptyset , e um evento não vazio, A , definido no espaço amostral, Ω . Então $\emptyset \subseteq A$.

Prova:

Vamos realizar a prova por contradição. Supomos que $\emptyset \not\subseteq A$. Isso significa que $\exists \omega : \omega \in \emptyset$ e $\omega \notin A$, porém $\omega \in \emptyset$ é um absurdo, logo $\emptyset \subseteq A$, o que conclui a prova.

E ainda podemos concluir que se existe um conjunto vazio, ele é único, como pode ser apresentado no Corolário 5.1.

Corolário 5.1

Existe somente um conjunto vazio.

Prova:

Suponha que exista dois conjuntos vazios, \emptyset_1 e \emptyset_2 . Pelo Teorema 5.1, sabemos que $\emptyset_1 \subseteq \emptyset_2$, uma vez que \emptyset_1 é um conjunto vazio. Mas, também sabemos que $\emptyset_2 \subseteq \emptyset_1$, uma vez que \emptyset_2 é um conjunto vazio. Logo, pela equivalência de conjuntos (eventos), Definição 5.12, $\emptyset_1 = \emptyset_2$, o que conclui a prova.

Em algumas situações, podemos apresentar alguns eventos a partir da combinação de outros eventos. Dessa forma, se faz necessário apresentar algumas operações elementares de conjuntos e suas consequências, tais como a união, interseção, complemento, dentre outras definições abordadas a seguir. Inicialmente, apresentamos na Definição 5.8, a união de dois eventos.

Definição 5.8: União de dois eventos

Sejam A e B, dois eventos quaisquer de Ω , então o conjunto de todos os elementos que estão em A ou B ou em ambos, é definido como o conjunto união de A e B, denotado por $A \cup B$, tal que,

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}. \quad (5.2)$$

Vejamos o Exemplo 5.9, sobre a união de dois eventos, a seguir.

Exemplo 5.9:

Sejam os conjuntos:

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

então

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

A Definição 5.9 apresenta a próxima propriedade de conjuntos, que é a interseção de de eventos, apresentada a seguir.

Definição 5.9: Interseção de dois eventos

Sejam A e B, dois eventos quaisquer de Ω , então o conjunto que contém todos os elementos que estão em A e B, é definido como a interseção de A e B, denotado por $A \cap B$ ou AB , tal que,

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ e } \omega \in B\}. \quad (5.3)$$

Do Exemplo 5.9, temos que a intersecção de $AB = \{3\}$.

Definição 5.10: Eventos Disjuntos ou mutuamente exclusivos

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

Vejamos o Exemplo 5.10, para entendermos sobre eventos disjuntos, apresentado a seguir.

Exemplo 5.10:

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

Em seguida, apresentamos mais duas definições interessantes, que são os eventos coletivamente exaustivos (Definição 5.11) e eventos equivalentes (Definição 5.12), apresentados na sequência.

Definição 5.11: Eventos coletivamente exaustivos

Considere um conjunto de eventos em Ω , se ao menos um evento ocorrer durante um dado experimento, dizemos que esses eventos são coletivamente exaustivos.

Na sequência, segue a definição sobre eventos equivalentes.

Definição 5.12: Eventos equivalentes

Dois eventos A e B são definidos equivalentes, ou iguais, se $A \subseteq B$ e $B \subseteq A$.

Exemplo 5.11:

Sejam os conjuntos:

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

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

Uma relação de eventos que será muito importante para o estudo da teoria da probabilidade, é a definição de complemento, abordado a seguir.

Definição 5.13: Evento Complementar

Seja A um evento de Ω . Então o complemento do evento A com respeito a Ω , denotado por \bar{A} , A^c , ou $\Omega - A$, é o subconjunto dos elementos de Ω exceto os elementos do evento A , isto é,

$$A^c = \{\omega \in \Omega : \omega \notin A\}. \quad (5.4)$$

Exemplo 5.12:

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

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

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

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

Então o complemento de A será:

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

Definição 5.14: Diferença de dois eventos

Sejam A e B dois eventos de Ω . O conjunto de todos os elementos de A que não estão em B , serão denotados por $A - B$ ou $A \setminus B$, sendo definido por conjunto diferença, isto é,

$$A - B = \{\omega \in \Omega : \omega \in A \text{ e } \omega \notin B\}. \quad (5.5)$$

A Definição 5.15 pode ser confundida com a Definição 5.12, porém a segunda definição se remete ao espaço amostral, e a diferença entre dois eventos se refere apenas a existência dos elementos de um evento que não estão no outro evento. Vejamos o Exemplo 5.13, e depois compare com o Exemplo 5.12, para elucidar essas duas definições.

Exemplo 5.13:

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

Por fim, uma última definição é a partição de conjuntos, apresentado na Definição XX, a seguir.

Definição 5.15: Partição

Considerando uma sequência de eventos A_1, A_2, \dots, A_n , não vazios, é uma partição do evento A , se e somente se,

- I) $A = \bigcup_{i=1}^n A_i$;
- II) A_1, A_2, \dots, A_n são mutuamente disjuntos, tais que $A_i \cap A_j, \forall i \neq j$.

Considerando que $A = \Omega$, dizemos que temos uma partição do espaço amostral. A seguir, apresentamos algumas leis importantes para a teoria de conjuntos, que estabelece algumas propriedades. Vejamos o Teorema 5.2.

Teorema 5.2

Considere três eventos A, B , e C definidos em Ω , então segue que:

- I) Lei comutativa: $A \cup B = B \cup A$ e $A \cap B = B \cap A$;
- II) Lei associativa: $A \cup (B \cup C) = (A \cup B) \cup C$;
- III) Lei distributiva: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ e $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
- IV) Lei DeMorgan: $(A \cup B)^c = A^c \cap B^c$ e $(A \cap B)^c = A^c \cup B^c$.

Prova:

Para provar que dois conjuntos são iguais, devemos demonstrar que todo elemento que está em um conjunto, também está no outro, e vice-versa.

- I) Supomos que $\omega \in A \cup B$, portanto, $\omega \in A$ ou $\omega \in B$. Logo, $\omega \in B \cup A$. Da mesma forma, supomos que $\omega \in B \cup A$, portanto, $\omega \in B$ ou $\omega \in A$. Logo, $\omega \in A \cup B$. De forma resumida, podemos expressar essa prova da seguinte forma: $\omega \in A \cup B \Leftrightarrow \omega \in A$ ou $\omega \in B \Leftrightarrow \omega \in B \cup A$; Para a outra parte da prova, supomos que $\omega \in A \cap B$, portanto, $\omega \in A$ e $\omega \in B$. Logo, $\omega \in B \cap A$. Da mesma forma, supomos que $\omega \in B \cap A$, portanto, $\omega \in B$ e $\omega \in A$. Logo, $\omega \in A \cap B$. De forma resumida, podemos expressar essa prova da seguinte forma: $\omega \in A \cap B \Leftrightarrow \omega \in A$ e $\omega \in B \Leftrightarrow \omega \in B \cap A$, o que finaliza a prova da Lei comutativa;
- II) Supomos que $\omega \in A \cup (B \cup C)$, portanto, $\omega \in A$ ou $\omega \in B \cup C$, e que isso implica em $\omega \in A \cup B$ ou $\omega \in C$. Logo, $\omega \in (A \cup B) \cup C$. Da mesma forma, supomos que $\omega \in (A \cup B) \cup C$, portanto, $\omega \in A \cup B$ ou $\omega \in C$, e que isso implica em $\omega \in A$ ou $\omega \in B \cup C$. Logo, $\omega \in A \cup (B \cup C)$. De forma resumida, podemos expressar essa prova da seguinte forma: $\omega \in A \cup (B \cup C) \Leftrightarrow \omega \in A$ ou $\omega \in B \cup C \Leftrightarrow \omega \in A \cup B$ ou $\omega \in C \Leftrightarrow \omega \in (A \cup B) \cup C$, o que finaliza a prova da Lei associativa;
- III) Supomos que $\omega \in A \cup (B \cap C)$, portanto, $\omega \in A$ ou $\omega \in (B \cap C)$. Considerando que $\omega \in A$, então $\omega \in (A \cup B)$ e $\omega \in (A \cup C)$, logo, $\omega \in (A \cup B) \cap (A \cup C)$. Considerando que $\omega \in (B \cap C)$, então $\omega \in B$ e $\omega \in C$. Logo, $\omega \in (A \cup B) \cap (A \cup C)$. Agora, assumimos que $\omega \in (A \cup B) \cap (A \cup C)$, portanto, $\omega \in (A \cup B)$ e $\omega \in (A \cup C)$. Isso significa que, ou $\omega \in A$ ou $\omega \in (B \cap C)$, logo, $\omega \in A \cup (B \cap C)$. Para a segunda parte, assumimos que $\omega \in A \cap (B \cup C)$. Isso implica que $\omega \in A$ e $\omega \in (B \cup C)$. Como $\omega \in A$, então

$\omega \in (A \cap B)$ e $\omega \in (A \cap C)$. Logo, $\omega \in (A \cap B) \cup (A \cap C)$. Agora, assumimos que $\omega \in (A \cap B) \cup (A \cap C)$, portanto, $\omega \in (A \cap B)$ ou $\omega \in (A \cap C)$. Se $\omega \in (A \cap B)$, então ω está em A e B . Como $\omega \in B$, então $\omega \in (B \cup C)$. Se $\omega \in (A \cap C)$, então ω está em A e C . Como $\omega \in C$, então $\omega \in (B \cup C)$. Logo, $\omega \in A \cap (B \cup C)$, o que finaliza a prova da Lei distributiva.

- IV) Supomos que $\omega \in (A \cup B)^c$, então $\omega \notin (A \cup B)$, isto é, $\omega \notin A$ ou $\omega \notin B$. Como $\omega \notin A$ ou $\omega \notin B$, então $\omega \in A^c$ e $\omega \in B^c$. Logo, $\omega \in A^c \cap B^c$. Agora, assumimos que $\omega \in A^c \cap B^c$. Isso implica que $\omega \in A^c$ e $\omega \in B^c$, de modo que ou $\omega \notin A$ ou $\omega \notin B$. Assim, $\omega \notin (A \cup B)$, e pela definição de evento complementar, concluímos que $\omega \in (A \cup B)^c$. De forma resumida, podemos expressar essa prova da seguinte forma: $\omega \in (A \cup B)^c \Leftrightarrow \omega \notin A \text{ ou } \omega \notin B \Leftrightarrow \omega \in A^c \text{ e } \omega \in B^c \Leftrightarrow \omega \in A^c \cap B^c$. Na segunda parte, assumimos que $\omega \in (A \cap B)^c$, então $\omega \notin (A \cap B)$. Assim, $\omega \notin A$ e nem $\omega \notin B$. Como consequência, $\omega \in A^c$ ou $\omega \in B^c$, logo $\omega \in A^c \cup B^c$. Da mesma forma, assumimos que $\omega \in A^c \cup B^c$, isto é, $\omega \in A^c$ ou $\omega \in B^c$. Isso implica que $\omega \notin (A \cap B)$. Usando a definição de evento complementar, logo $\omega \in (A \cap B)^c$, o que conclui a prova para a Lei DeMorgan. De forma resumida, podemos expressar essa prova da seguinte forma: $\omega \in (A \cap B)^c \Leftrightarrow \omega \notin (A \cap B) \Leftrightarrow \omega \notin A \text{ e } \omega \notin B \Leftrightarrow \omega \in A^c \text{ ou } \omega \in B^c \Leftrightarrow \omega \in A^c \cup B^c$.

Para finalizar, apresentamos pelo Teorema 5.3, algumas identidades que serão importantes na teoria de conjuntos para o estudo sobre a probabilidade.

Teorema 5.3

Sejam os eventos A e B definidos no espaço amostral Ω , não vazio. Então, apresentamos as seguintes identidades:

- I) $A \cap A^c = \emptyset$;
- II) $A \cup A^c = \Omega$;
- III) $\Omega^c = \emptyset$;
- IV) $\emptyset^c = \Omega$;
- V) $(A^c)^c = \overline{\overline{A}} = A$, em outras palavras, o complemento de \overline{A} é igual a A ;
- VI) $A\Omega = A$ (Elemento neutro);
- VII) $A \cup \Omega = \Omega$;
- VIII) $A \cap A = A$ (Idempotência);
- IX) $A\emptyset = \emptyset$ (Elemento absorvente);
- X) $A \cup \emptyset = A$;
- XI) $A - B = A - (A \cap B) = A \cap B^c$;
- XII) $B = (B \cap A) \cup (B \cap A^c)$;
- XIII) $B - A = B \cap A^c$;
- XIV) $A \cup B = A \cup (B \cap A^c)$;
- XV) $A \cup B = (A^c \cap B) \cup (A \cap B) \cup (A \cap B^c)$.

Prova:

Para provar que dois conjuntos são iguais, devemos demonstrar que todo elemento que está em um conjunto, também está no outro, e vice-versa.

- I) Vamos apresentar a prova por contradição. Suponha que $\omega \in A \cap A^c \neq \emptyset$, então $\exists \omega \in A \cap A^c : \omega \in A \text{ e } \omega \in A^c$. Mas por definição $A^c = \{\omega \in \Omega : \omega \notin A\}$, então a afirmação $\exists \omega \in A \cap A^c : \omega \in A \text{ e } \omega \in A^c$, é absurdo. Logo, $A \cap A^c = \emptyset$;
- II) $\omega \in A \cup A^c \Leftrightarrow \omega \in A \text{ ou } \omega \in A^c \Leftrightarrow \omega \in A \text{ ou } \omega \notin A \Leftrightarrow \omega \in \Omega$;
- III) Vamos apresentar a prova por contradição. Supomos que $\Omega^c \neq \emptyset$. Isso significa que $\exists \omega : \omega \in \Omega^c \text{ e } \omega \notin \Omega$, e isso é absurdo, pois Ω representa o conjunto de todos os resultados possíveis em um experimento. Logo, $\Omega^c = \emptyset$;
- IV) Vamos provar por contradição. Supomos que $\emptyset^c \neq \Omega$. Isso implica que, $\exists \omega : \omega \in \Omega \text{ e } \omega \notin \emptyset^c$. Então, isso implica que $\omega \in \emptyset$, que é absurdo. Logo, $\emptyset^c = \Omega$. Uma outra forma de apresentar essa prova é usar a definição de evento complementar, isto é, $A^c = \Omega - A$. Considerando $A = \emptyset$, então $\emptyset^c = \Omega - \emptyset$. Logo, $\emptyset^c = \Omega$;
- V) $\omega \in (A^c)^c \Leftrightarrow \omega \notin A^c \Leftrightarrow \omega \in A$;
- VI) $\omega \in A \cap \Omega \Leftrightarrow \omega \in A \text{ e } \omega \in \Omega \Leftrightarrow \omega \in A$;
- VII) $\omega \in A \cap A \Leftrightarrow \omega \in A \text{ e } \omega \in A \Leftrightarrow \omega \in A$;
- VIII) $\omega \in A \cup \Omega \Leftrightarrow \omega \in A \text{ ou } \omega \in \Omega \Leftrightarrow \omega \in \Omega$;
- IX) Vamos apresentar a prova por contradição, isto é, vamos supor que $A \cap \emptyset \neq \emptyset$, então $\exists \omega : \omega \in A \cap \emptyset$. Assim, $\omega \in A \text{ e } \omega \in \emptyset$. Porém, $\omega \in \emptyset$ é falso. Então $\omega \in A \cap \emptyset$ é falso. Logo, $A \cap \emptyset = \emptyset$;
- X) Considerando que $\omega \in A \cup \emptyset$, então ou $\omega \in A$ ou $\omega \in \emptyset$. Mas $\omega \in \emptyset$ é falso, logo $\omega \in A$. Do mesmo modo se $\omega \in A$, então $\omega \in A \cup \emptyset$;
- XI) $\omega \in (A - B) \Leftrightarrow \omega \in A \text{ e } \omega \notin B \Leftrightarrow \omega \in A \text{ e } \omega \notin (A \cap B) \Leftrightarrow A - (A \cap B)$. Da mesma forma, $\omega \in (A - B) \Leftrightarrow \omega \in A \text{ e } \omega \notin B \Leftrightarrow \omega \in A \text{ e } \omega \in B^c \Leftrightarrow \omega \in A \cap B^c$;
- XII) Considerando que $\omega \in B$, então ou $\omega \in A$ ou $\omega \in A^c$. Se $\omega \in A$, então $\omega \in (B \cap A)$. Se $\omega \in A^c$, então $\omega \in (B \cap A^c)$. Logo, $\omega \in (B \cap A) \cup (B \cap A^c)$. Agora, considerando que $\omega \in (B \cap A) \cup (B \cap A^c)$, sabemos que $\omega \in (B \cap A)$ ou $\omega \in (B \cap A^c)$. Se $\omega \in (B \cap A)$, então $\omega \in B$. Se $\omega \in (B \cap A^c)$, então $\omega \in B$. Logo, $\omega \in B$. Uma outra forma de provar mais facilmente essa identidade é usar a Lei distributiva (Teorema 5.2, III), isto é, $(B \cap A) \cup (B \cap A^c) = B \cap (A \cup A^c) = B \cap \Omega = B$;
- XIII) Basta usar a propriedade VI desse teorema;
- XIV) $\omega \in (A \cap B) \Leftrightarrow \omega \in A \text{ ou } \omega \in B \Leftrightarrow \omega \in A \text{ ou } \omega \in (B \cap A^c) \Leftrightarrow \omega \in A \cup (B \cap A^c)$;
- XV) $\omega \in A \cup B \Leftrightarrow \omega \in A \text{ ou } \omega \in B \Leftrightarrow \omega \in (A \cap B) \text{ ou } \omega \in \underbrace{(A^c \cap B) \cup (A \cap B)}_{prop. (XII)} \Leftrightarrow \omega \in (A \cap B) \cup (A^c \cap B) \cup (A \cap B)$.

Baseado em tudo o que foi estudado sobre uma introdução à teoria de conjuntos, iremos a partir da próxima seção, contextualizar todas essas informações com o estudo sobre a medida de probabilidade.

5.2 Definições de probabilidades

Após um contexto sobre a teoria de conjuntos, iniciamos o contexto probabilístico, com o interesse de saber a chance de determinado elemento de um evento, ocorrer como resultado de um experimento, ao invés de estar interessado nesse resultado. Isso tem total importância prática, pois é dessa forma, por exemplo, que prevemos determinados resultados de um determinado fenômeno de interesse. Consideremos um evento A contido no espaço amostral Ω , e desejamos associar ao evento A uma medida que assume valores entre 0 e 1, que chamamos de medida de probabilidade de A , denotada por $P(A)$. Assim, diremos que $P(A)$ é a probabilidade de que o evento A ocorra no espaço amostral Ω . De outro modo, a probabilidade do evento A representa a chance de ao menos um de seus elementos ocorrerem como resultado de um experimento. Voltando ao Exemplo 5.1 retirado de James (2004), considerando que esse dado é equilibrado, e o evento $A \subset \Omega$, então poderemos atribuir uma probabilidade para A da seguinte forma:

$$P(A) = \frac{\#A}{6} = \frac{\text{número de resultados favoráveis a } A}{\text{número de resultados possíveis}}.$$

Esta é a definição clássica de probabilidade quando Ω é finito. Entretanto, a probabilidade que o evento A ocorra no espaço amostral nem sempre é possível, devido a complexidade desses eventos. Retornando ao Exemplo 5.8, podemos interpretar a probabilidade de $A \subset \Omega$ como:

$$P(A) = \frac{\text{área } A}{\text{área } \Omega} = \frac{\text{área } A}{\pi},$$

sendo a área de A bem definida. Segundo um teorema profundo da teoria da medida, não se pode definir $P(A)$ para $A \subset \Omega$ de modo que a área de A não esteja bem definida. A prova disso depende do **Axioma da escolha**. Um exemplo clássico desses eventos são os **conjuntos de Vitali de \mathbb{R}** , os quais não podemos atribuir nenhuma medida quando ela generaliza o comprimento de intervalos de \mathbb{R} . De fato é impossível atribuir comprimento a todos subconjuntos de \mathbb{R} preservando a aditividade e invariância por translação.

Dessa forma, estaremos apenas interessados em eventos cuja área está bem definida, apresentada na Definição 5.16, a seguir.

Definição 5.16: Evento Aleatório

Todo evento de Ω que podemos atribuir uma probabilidade, chamamos de evento aleatório.

Dessa forma, definimos a medida de probabilidade apresentada na Definição 5.17, a seguir.

Definição 5.17: Medida de Probabilidade

Seja Ω o espaço amostral, então uma função P , tal que $P : \Omega \rightarrow \mathbb{R}$, é chamada de medida de probabilidade ou probabilidade, aos eventos do espaço amostral satisfazendo os seguintes axiomas de Kolmogorov:

- (i). $P(\Omega) = 1$;
- (ii). $0 \leq P(A) \leq 1, \forall A \subset \Omega$;
- (iii). $P(A_1 \cup A_2) = P(A_1) + P(A_2)$, com $A_1 \cap A_2 = \emptyset$, para $A_1, A_2 \subset \Omega$.

Assim como mencionado por Montgomery e Runger (2016), os axiomas não determinam probabilidades, mas capacitam a calcular facilmente as probabilidades de alguns eventos, a partir do conhecimento de outras probabilidades. Na realidade, a probabilidade se baseia no conhecimento do sistema em estudo.

Vejamos o Exemplo 5.14, para elucidar essas definições, a seguir.

Exemplo 5.14: Retirado de Montgomery e Runger (2016)

Uma peça moldada de injeção é igualmente provável de ser obtida, a partir de qualquer uma das oito cavidades de um molde.

- Qual é o espaço amostral?
- Qual é a probabilidade de a peça ser proveniente da cavidade 1 ou 2?
- Qual é a probabilidade de a peça não ser proveniente nem da cavidade 3 nem da 4?

Nesse caso, (a) o espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$. Como a peça moldada de injeção é igualmente provável, então (b) a probabilidade de a peça ser proveniente da cavidade 1 ou 2, é dada por:

$$\begin{aligned} P(\{1\} \cup \{2\}) &= P(\{1\}) + P(\{2\}), \quad (\text{Eventos disjuntos}) \\ &= 1/8 + 1/8 \\ &= 2/8. \end{aligned}$$

Por fim, (c) a probabilidade de a peça não ser proveniente nem da cavidade 3 nem da 4, é dada por:

$$\begin{aligned} P(\{3\}^c \cap \{4\}^c) &= P[(\{3\} \cap \{4\})^c], \quad (\text{Lei DeMorgan, Teorema 5.2 prop. III}) \\ &= 1 - P(\{3\} \cap \{4\}), \quad (\text{Evento complementar}) \\ &= 1 - [P(\{3\}) + P(\{4\})], \quad (\text{Eventos disjuntos}) \\ &= 1 - [1/8 + 1/8] \\ &= 1 - 2/8 \\ &= 6/8. \end{aligned}$$

O ítem (c) do Exemplo 5.14 exigiria um conhecimento sobre algumas propriedades da medida de probabilidade como consequência das propriedades da teoria de conjuntos abordadas na seção anterior, mas que serão abordadas a seguir.

Apesar da definição formal sobre a probabilidade, há duas formas para atribuir probabilidades aos elementos do espaço amostral, que em algumas situações são aplicáveis, que seguem:

- A primeira delas, consiste na atribuição de probabilidades, baseando-se em características teóricas da realização do fenômeno, chamado de *probabilidade clássica ou a priori*; formalmente, se um experimento aleatório obtiver resultados mutuamente exclusivos e igualmente prováveis, e se n_A desses resultados têm um atributo A , então, a probabilidade de acontecer A é a fração n_A/n . Mais ainda, Laplace define a probabilidade de um acontecimento como sendo o quociente entre o número de casos favoráveis e o número de casos possíveis, supondo todos equiprováveis. A principal limitação é que os eventos tenham que ser igualmente possíveis.
- uma outra maneira de obter probabilidades é através das frequências de ocorrências, também conhecido como *probabilidade frequentista ou a posteriori*, em que a probabilidade de um dado acontecimento pode ser medida observando a frequência relativa do mesmo acontecimento numa sucessão numerosa de provas ou experiências, idênticas e independentes. A principal limitação é que os eventos possam repetir-se indefinidamente nas mesmas circunstâncias.

Exemplo 5.15:

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

clássica ou a priori.

Exemplo 5.16:

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

5.3 Propriedades

Vejamos algumas propriedades da medida de probabilidade, consequências dos Teoremas 5.2 e 5.3, porém sem a apresentação das devidas provas.

5.3.1 Regra de adição de probabilidades

Considere um espaço amostral e dois eventos não vazios A e B , então a probabilidade de ocorrência do evento A ou do evento B é igual a:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (5.6)$$

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

$$P(A \cup B) = P(A) + P(B). \quad (5.7)$$

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

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n). \quad (5.8)$$

Vejamos o Exemplo 5.17, a seguir.

Exemplo 5.17: Retirado de Devore (2006)

Uma empresa de eletricidade oferece uma taxa vitalícia de energia a qualquer lar cuja utilização de energia esteja abaixo de 240 kWh durante um determinado mês. Represente por A o evento de um lar selecionado aleatoriamente em um comunidade que não excede a utilização da taxa vitalícia em janeiro e por B o evento análogo para o mês de julho (A e B se referem ao mesmo lar). Suponha que $P(A) = 0,8$, $P(B) = 0,7$ e $P(A \cup B) = 0,9$. Calcule:

- a) $P(A \cap B)$;
- b) A probabilidade de a quantia da taxa vitalícia ser excedida em exatamente um dos dois meses. Descreva esse evento em termos de A e B .

Considere,

$$\begin{aligned} A &= \{\omega \in \Omega : \omega = \text{"lar X que não excede } 240\text{kWh em janeiro"}\}, \\ B &= \{\omega \in \Omega : \omega = \text{"lar X que não excede } 240\text{kWh em julho"}\}. \end{aligned}$$

Então, para o primeiro ítem (a), temos que:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (\text{equação 5.6}) \quad (5.9)$$

Isso ocorre porque os eventos não são disjuntos, uma vez que os dois eventos consistem no mesmo lar X, em ser selecionado. Assim, pela expressão 5.10, podemos obter $P(A \cap B)$, dado por:

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0,8 + 0,7 - 0,9 \\ &= 0,6. \end{aligned} \quad (5.10)$$

No caso do ítem (b), o evento que representa o lar X de a quantia vitalícia ser excedida em exatamente um dos dois meses por ser representado por: $(A^c \cap B) \cup (A \cap B^c)$, uma vez que,

$$\begin{aligned} A^c &= \{\omega \in \Omega : \omega = \text{"lar X que excede 240kWh em janeiro"}\}, \\ B^c &= \{\omega \in \Omega : \omega = \text{"lar X que excede 240kWh em julho"}\}. \end{aligned}$$

Podemos ainda observar pelo Teorema 5.3(prop. XV), que $A \cup B = (A^c \cap B) \cup (A \cap B) \cup (A \cap B^c)$, e que cada um dos eventos dentro do parêntese são disjuntos dois a dois, logo,

$$\begin{aligned} P(A \cap B) &= P[(A^c \cap B) \cup (A \cap B) \cup (A \cap B^c)] \\ &= P[(A^c \cap B) \cup (A \cap B^c)] + P(A \cap B). \end{aligned}$$

Desse modo, percebemos que $P[(A^c \cap B) \cup (A \cap B^c)] = P(A \cap B) - P(A \cap B)$, logo,

$$P[(A^c \cap B) \cup (A \cap B^c)] = 0,9 - 0,6 = 0,3.$$

5.3.2 Probabilidade de um evento complementar

Considere um espaço amostral e o evento, não vazio, A , então a probabilidade do evento complementar A^c ocorrer, é dado por:

$$P(A^C) = 1 - P(A). \quad (5.11)$$

Essa situação é consequência da regra da adição para $A \subset \Omega$ na expressão (5.7), substituindo B por A^C , temos

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

logo segue a expressão (5.11).

Exemplo 5.18: Retirado de Devore (2006)

Usando os resultados do Exemplo 5.17, podemos calcular a probabilidade de A^c , como:

$$P(A^c) = 1 - 0,8 = 0,2.$$

Isso representa a chance do lar X exceder a energia acima de 240kWh, em janeiro.

5.4 Eventos independentes e probabilidade condicional

Nessa seção, iremos apresentar iniciar com uma motivação, por meio do Exemplo 5.19, uma abordagem sobre dois assuntos muito interessantes na probabilidade, que são a independência de eventos

e a modificação do espaço amostral, dada uma informação antecipada, e qual a implicância dessas informações para a probabilidade de um evento ocorrer.

Exemplo 5.19:

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

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

Surgiu as seguintes indagações:

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

Muitas vezes nos deparamos com situações em que antes da realização de algum experimento, temos alguma informação adicional. Queremos saber o quanto que essa informação pode afetar a medida de probabilidade. Assim, apresentamos a Definição 5.18, que define a probabilidade condicional, a seguir.

Definição 5.18: Probabilidade condicional

Dados dois eventos A e B definidos em Ω , então a probabilidade condicional do evento A dado que ocorreu o evento B , denotado por $P(A|B)$, é definida por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad (5.12)$$

para $P(B) > 0$.

Baseado no problema de Paulo, denotemos o evento D as sandálias produzidas com defeitos pela empresa, M_1 o evento que representa as sandálias produzidas pela máquina M_1 , M_2 o evento que representa as sandálias produzidas pela máquina M_2 e M_3 o evento que representa as sandálias produzidas pela máquina M_3 . Assim, percebemos que a probabilidade do evento D não pode ser observada facilmente, pois o defeito dos produtos produzidos pelas máquinas está condicionado a cada máquina. Desse modo podemos representar, a probabilidade desses defeitos da seguinte forma: $P(D|M_1) = 0,01$, $P(D|M_2) = 0,02$, e $P(D|M_3) = 0,03$. Essas probabilidades apresentam uma alteração no espaço amostral para cada evento, porque esses resultados mostram a chance de defeito do produto, dado o conhecimento de que máquina foi produzido, é o que chamamos de restrição do espaço amostral.

Essa restrição do espaço amostral, pode nos questionar se de fato a probabilidade condicional, de fato, é uma medida de probabilidade. Para isso, apresentamos o Teorema 5.4, na sequência.

Teorema 5.4: $P(A|B)$ é uma medida de probabilidade

Sejam A e B eventos aleatórios, tal que $P(B) > 0$, e considere $Q : \Omega \rightarrow [0, 1]$ definida por

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

que é a probabilidade condicional de A dado B . Então Q é também uma medida de probabilidade. \square

Prova:

Para verificarmos se $Q(\cdot)$ é uma medida de probabilidade, devemos assumir que Q satisfaz os axiomas de Kolmogorov, isto é,

- Axioma 1:

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

- Axioma 2:

Como P é uma medida de probabilidade, então

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

- Axioma 3: Sejam dois eventos A_1 e A_2 , disjuntos, então

$$\begin{aligned} Q(A_1 \cup A_2) &= \frac{P[(A_1 \cup A_2) \cap B]}{P(B)} \\ &= \sum_{i=1}^2 \frac{P[(A_1 \cap B) \cup (A_2 \cap B)]}{P(B)}, \quad (\text{Lei distributiva}) \\ &= Q(A_1) + Q(A_2). \end{aligned}$$

o que conclui a prova.

Como sequência de importantes resultados sobre propriedades da probabilidade, apresentamos o Teorema 5.5, que será importante para resultados muito utilizados na área aplicada, como o Teorema de Bayes, apresentado na sequência.

Teorema 5.5: Regra do produto de probabilidade

Seja os eventos não vazios A_1, A_2, \dots, A_n em Ω , com $P(\bigcap_{i=1}^n A_i) > 0$, então a probabilidade do produto desses eventos é dado por

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

Prova:

Por indução, consideremos $n = 2$. Assim, pela Definição 5.18, temos

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

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

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

pela definição 5.18, temos

$$P[(A_1 A_2 \dots A_{k-1}) \cap A_k] = P[(A_1 A_2 \dots A_{k-1})]P(A_k|A_1 A_2 \dots A_{k-1})$$

podendo ser reescrito como

$$\begin{aligned} P[(A_1 A_2 \dots A_{k-1}) \cap A_k] &= \underbrace{P(A_1 A_2 \dots A_{k-2})P(A_{k-1}|A_1 A_2 \dots A_{k-2})}_{P(A_1 A_2 \dots A_{k-1})} \times \\ &\quad \times P(A_k|A_1 A_2 \dots A_{k-1}). \end{aligned}$$

Assim, usando a indução sucessivas vezes, chegaremos a expressão

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

Observe que por hipótese, todos os condicionamentos da expressão do lado direito, têm probabilidades positivas, pois contém $\bigcap_{i=1}^n A_i$, o que conclui a prova.

Antes de falarmos sobre o teorema da lei da probabilidade total, será interessante fazer a definição sobre a partição de Ω , apresentada na Definição 5.19, a seguir.

Definição 5.19: Partição de Ω

Se a sequência A_1, A_2, \dots , são disjuntos dois a dois, não vazios, e $\bigcup_{i=1}^{\infty} A_i = \Omega$, então dizemos que essa sequência forma uma partição de Ω .

Entretanto, para calcular a probabilidade de uma sandália está com defeito, isto é $P(D)$, independente de qual máquina a produziu, pode ser calculada por meio do Teorema 5.6, a seguir.

Teorema 5.6: Teorema da probabilidade total

Seja uma sequência de eventos A_1, A_2, \dots, A_n de Ω , disjuntos, tal que $\bigcup_{i=1}^n A_i = \Omega$, e B um evento de Ω , não vazio, então a probabilidade de B é dada por:

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i), \tag{5.13}$$

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

Prova:

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

disjuntos. Dessa forma,

$$\begin{aligned} P(B) &= P\left(\bigcup_{i=1}^n B \cap A_i\right) \\ &= \sum_{i=1}^n P(B \cap A_i) \quad (B \cap A_i \text{ disjuntos}) \\ &= \sum_{i=1}^n P(B|A_i)P(A_i) \quad (\text{Teorema 5.5}) \end{aligned}$$

como queríamos provar.

Retornando ao cálculo da probabilidade $P(D)$, no Exemplo 5.20 apresentamos como o Teorema 5.6 soluciona esse problema, a seguir.

Exemplo 5.20:

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

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

Nesse momento, surge uma importante definição na Estatística e Probabilidade, que é a independência de eventos, apresentada na Definição 5.20. A ideia da independência é uma característica probabilística, e isto significa, que se dois eventos forem independentes, então a probabilidade de um evento ocorrer não é influenciado pela ocorrência ou não do outro evento. A implicância da pressuposição da independência em problemas práticos na estatística podem ser resolvidos de forma trivial, devido as técnicas probabilísticas serem resolvidas de forma mais facilmente.

Definição 5.20: Independência de dois eventos

Considere o espaço amostral Ω . Dois eventos A e B de Ω são independentes se satisfaz ao menos uma das seguintes condições:

- I) $P(A \cap B) = P(A)P(B)$;
- II) $P(A|B) = P(A)$, para $P(B) > 0$;
- III) $P(B|A) = P(B)$, para $P(A) > 0$.

□

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

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

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

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

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

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

A intuição para independência na Definição 5.20 fica justificada pelo fato de que A é independente de B tanto na ocorrência quanto a não ocorrência de B e isso não muda em nada a probabilidade da ocorrência de A , isto é, $P(A|B) = P(A)$ e $P(A|B^c) = P(A)$. Essas duas expressões significam que

$$\begin{aligned} P(A \cap B) &= P(B)P(A|B) = P(B)P(A) \\ P(A \cap B^c) &= P(B^c)P(A|B) = P(B^c)P(A). \end{aligned}$$

Entretanto, a independência entre dois eventos não implica em independência coletiva. Vejamos o Exemplo 5.21, a seguir.

Exemplo 5.21:

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

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

Que é o mesmo que entender que $P(A) \times P(B) = 1/6 = P(A \cap B)$. Logo, A e B são eventos independentes.

Exemplo 5.22:

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

Para uma definição mais geral sobre a independência de eventos, apresentamos a Definição 5.21, a seguir.

Definição 5.21: Independência de eventos

Considere o espaço amostral Ω . Uma sequência de eventos A_1, A_2, \dots, A_n de Ω são independentes se e somente se:

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j), \quad \text{para } i \neq j; \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k), \quad \text{para } i \neq j \neq k; \\ &\vdots \\ P(\cap_{i=1}^n A_i) &= \prod_{i=1}^n P(A_i). \end{aligned} \tag{5.14}$$

Paulo poderia indagar, se os eventos M_i e D são independentes ou dependentes. Contudo, pela Definição 5.20, temos que

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

logo, não são independentes.

5.5 Teorema de Bayes

A grande questão agora é qual a máquina que Paulo deveria substituir com o propósito de aumentar seu lucro na empresa. A ideia será calcular $P(M_i|D)$, isto é, dado um defeito na sandália qual a probabilidade de vindo da máquina i ? A maior probabilidade será a máquina substituída. Entretanto, ainda não temos ferramenta para resolver essa resposta. Para isso, apresentamos o seguinte Teorema 5.7 a seguir.

Teorema 5.7: Teorema de Bayes

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

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}, \quad k = 1, 2, \dots, n, \quad (5.15)$$

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

Prova:

Para um i qualquer, temos

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

ou

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

Isto implica que

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

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

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

prova concluída.

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

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

cada uma das máquinas?

$$P(M_1|D) = \frac{0,01 \times 0,50}{0,016} = 0,3125$$

$$P(M_2|D) = \frac{0,02 \times 0,40}{0,016} = 0,5000$$

$$P(M_3|D) = \frac{0,03 \times 0,10}{0,016} = 0,1875$$

A tomada de decisão será substituir a máquina M_2 . Poderíamos ter tomado uma decisão equivocada se não fosse o teorema de Bayes.

Devemos abrir uma discussão que ocorre muito frequente entre as Definições 5.10 e 5.20, isto é, eventos disjuntos e independência. Nas próprias definições, percebemos a distinção clara entre as características. A primeira se remete a eventos (conjuntos), e a segunda é uma condição probabilística dos eventos. Contudo, em determinados problemas ainda há muita confusão ao tentar resolvê-los. Assim, apresentemos o Teorema 5.8, a seguir.

Teorema 5.8: Eventos disjuntos e independentes

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

Prova:

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

Caso esses eventos não tenham probabilidade 0, a condição $A \cap B = \emptyset$ implica que estes são dependentes. Vejamos o Exemplo 5.23 adaptado de Morettin (2010), para elucidar essas definições.

Exemplo 5.23: Adaptado de Morettin (2010)

A tabela abaixo dá a distribuição das probabilidades dos quatro tipos sanguíneos, numa certa comunidade.

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

Calcular a probabilidade de que:

- um indivíduo, sorteado ao acaso nessa comunidade, tenha o tipo O;
- um indivíduo, sorteado ao acaso nessa comunidade, tenha o tipo A e tipo B ao mesmo tempo. Podemos afirmar, que estes são independentes?
- dois indivíduos, sorteados ao acaso nessa comunidade, tenham tipo A e tipo B, nessa ordem;
- um indivíduo, sorteado ao acaso nessa comunidade, não tenha o tipo B ou não tenha o tipo AB.

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

- a) Consideremos o evento A , os indivíduos da comunidade especificado do tipo sanguíneo A, tal que, $A = \{\omega_A \in \Omega : \omega_A \in A\}$; o evento B , os indivíduos da comunidade especificado do tipo sanguíneo B, tal que, $B = \{\omega_B \in \Omega : \omega_B \in B\}$; o evento AB , os indivíduos da comunidade especificado do tipo sanguíneo AB, tal que, $AB = \{\omega_{AB} \in \Omega : \omega_{AB} \in AB\}$; o evento O , os indivíduos da comunidade especificado do tipo sanguíneo O, tal que, $O = \{\omega_O \in \Omega : \omega_O \in O\}$. Desse modo, o espaço amostral é dado por $\Omega = \{\omega_A, \omega_B, \omega_{AB}, \omega_O\}$, cujos elementos de ω devem estar apenas em um dos eventos anteriores, e que a união de todos os elementos desses formam o espaço amostral, logo esses eventos formam uma partição do espaço amostral (Definição 5.19). Observe também que os elementos não são equiprováveis, como mostrado na tabela de probabilidades na própria questão. Assim, $P(\Omega) = P(A) + P(B) + P(AB) + P(O) \Rightarrow 1 = 0,2000 + 0,1000 + 0,0500 + P(O) \Rightarrow P(O) = 0,6500$;
- b) Este ítem merece uma atenção. Como os eventos A e B são multuamente exclusivos, logo $P(A \cap B) = 0$, e estes não são independentes pois nenhum tem probabilidade 0 (Teorema 5.8), logo A e B são eventos dependentes;
- c) Diferentemente do espaço amostral anterior, neste temos uma combinação de 16 possibilidades a cardinalidade de Ω , uma vez que temos 4 possibilidades para o primeiro indivíduo e mais 4 possibilidades para o segundo indivíduo, isto é, $\Omega = \{(\omega_A, \omega_A), (\omega_A, \omega_B), (\omega_A, \omega_{AB}), (\omega_A, \omega_O), (\omega_B, \omega_A), (\omega_B, \omega_B), (\omega_B, \omega_{AB}), (\omega_B, \omega_O), (\omega_{AB}, \omega_A), (\omega_{AB}, \omega_B), (\omega_{AB}, \omega_{AB}), (\omega_{AB}, \omega_O), (\omega_O, \omega_A), (\omega_O, \omega_B), (\omega_O, \omega_{AB}), (\omega_O, \omega_O)\}$. Uma vez determinada as probabilidades de especificação em indivíduos diferentes, a probabilidade de especificar o tipo sanguíneo A em um indivíduo da comunidade não interfere em nada na probabilidade de especificar o tipo sanguíneo de um outro indivíduo dessa mesma comunidade (Uma ressalva é válida, no sentido também que estamos desconsiderando indivíduos consanguíneos, isto é, com grau de parentesco). Assim, a probabilidade de especificar o tipo sanguíneo desses dois indivíduos simultaneamente é $P(A) \times P(B) = 0,2000 \times 0,1000 = 0,0200$. De outro modo, temos que:

$$\begin{aligned} P(A \cap B) &= P(\{\omega_A, \omega_B\}) \\ &= P(\underbrace{\omega_A \in A}_{\text{Indivíduo I}} \text{ e } \underbrace{\omega_B \in B}_{\text{Indivíduo II}}) \\ &= P(\omega_A \in A) \times P(\omega_B \in B) \\ &= P(A) \times P(B) \\ &= 0,2000 \times 0,0100 \\ &= 0,0200; \end{aligned}$$

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

$$\begin{aligned} P[(AB)^c \cup B^c] &= P[(AB)^c] + P(B^c) - P[(AB)^c \cap B^c] \\ &= 0,9500 + 0,9000 - P[(AB)^c \cap B^c]. \end{aligned} \tag{5.16}$$

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

$$\begin{aligned} P[(AB)^c \cap B^c] &= P(A \cup O) = P(A) + P(O) \\ &= 0,20 + 0,65 = 0,85. \end{aligned} \tag{5.17}$$

Substituindo (5.17) em (5.16), segue que

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

Ao final, temos a tabela completada da seguinte forma:

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

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

Podemos ainda expressar mais dois teoremas para complementar as afirmações feitas no Exemplo 5.23, e suas implicâncias com relação aos eventos serem independentes e eventos disjuntos. Inicialmente, apresentamos o Teorema 5.9 como uma implicância da independência de eventos, a seguir.

Teorema 5.9

Se A e B são eventos independentes, não vazio, definidos em Ω , então

- a) A e B^c também são independentes;
- b) A^c e B também são independentes;
- c) A^c e B^c também são independentes.

Prova:

Usando as seguintes equivalências:

$$P(A) = P(A \cap B) + P(A \cap B^c), \quad (5.18)$$

$$P(A^c) = P(A^c \cap B) + P(A^c \cap B^c), \quad (5.19)$$

$$P(B) = P(B \cap A) + P(B \cap A^c), \text{ e} \quad (5.20)$$

$$P(B^c) = P(B^c \cap A) + P(B^c \cap A^c), \quad (5.21)$$

e a condição de que $P(A \cap B) = P(A)P(B)$ (independentes), então usando (5.18) temos

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

o que prova o ítem (a). Usando (5.20) pelo mesmo raciocínio, provamos o ítem (b). Usando o resultado do ítem (a), já provado, e a condição de independência na expressão (5.21), temos

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

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

Por fim, o Teorema 5.10 apresenta uma implicância sobre eventos disjuntos, a seguir.

Teorema 5.10

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

Prova:

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

$$\begin{aligned} A \cup B &= (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) \\ &= (A \cap B^c) \cup (A^c \cap B) \\ &\neq \Omega \quad (\text{Pelo fato de A e B não serem partição do espaço amostral}). \end{aligned} \quad (5.22)$$

Usando a Lei de Morgan $A^c \cap B^c = (A \cup B)^c$, logo percebemos pela expressão (5.22) que $A^c \cap B^c \neq \emptyset$, o que completa a prova.

5.6 Variável Aleatória

Em estatística, avaliamos um experimento não pelos eventos em si, mas por uma função definida no espaço amostral, que associa o evento a um número real. Essa função chamamos de variável aleatória, denotada por uma letra maiúscula, X ou $X(\cdot)$. Alguns autores criticam o termo “variável aleatória”, já que a mesma é uma função. Como essa definição ficou conhecida com esse nome, seria um equívoco tentar renomeá-la, do qual é apresentada na Definição 5.22, a seguir.

Definição 5.22: Variável Aleatória

Seja o espaço amostral Ω de um experimento, então uma função $X : \Omega \rightarrow \mathbb{R}$ é chamada de variável aleatória, isto é, considerando $\omega \in \Omega$, então a variável aleatória, $X(\omega)$, é uma função com domínio em Ω e imagem no conjunto dos reais B , tal que $B = \{x \in \mathbb{R} : X(\omega) = x, \omega \in \Omega\}$.

Consideramos X uma variável aleatória discreta, quando B representa um conjunto contável (ou enumerável) de valores (finito ou infinito). Por outro lado, se B for um conjunto não contável (ou não enumerável), X será denominada de variável aleatória contínua. O fato é que, independente da natureza da variável aleatória, ela induz a um novo espaço amostral na reta real.

Exemplo 5.24:

Para explicar a definição de uma variável aleatória será considerado o exemplo, no qual duas variedades de uma espécie A (A_1, A_2) e três de outra espécie E (E_1, E_2 e E_3) são disponibilizados para uma pesquisa. Uma amostra de duas variedades ($n = 2$) é extraída. O espaço amostral dos resultados desse experimento, segue,

$$\Omega = \left\{ \begin{array}{lllll} (A_1, A_2) & (A_1, E_1) & (A_1, E_2) & (A_1, E_3) & (A_2, E_1) \\ (A_2, E_2) & (A_2, E_3) & (E_1, E_2) & (E_1, E_3) & (E_2, E_3) \end{array} \right\}.$$

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

$X(\omega)$	Eventos (C_i)
0	$C_1 = \{(E_1, E_2), (E_1, E_3), (E_2, E_3)\}$
1	$C_2 = \{(A_1, E_1), (A_1, E_2), (A_1, E_3), (A_2, E_1), (A_2, E_2), (A_2, E_3)\}$
2	$C_3 = \{(A_1, A_2)\}$

No exemplo 5.24, criamos uma partição do espaço amostral (Ω) e associamos com um outro espaço amostral induzido (Ω_X), por meio da variável aleatória X , como pode ser apresentado na Figura 5.3.

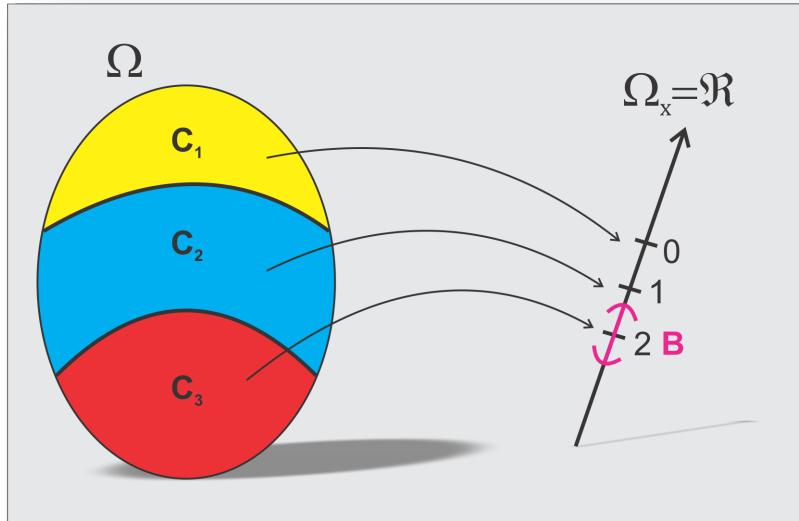


Figura 5.3: Espaço amostral e espaço amostral induzido pela variável aleatória X .

Para o exemplo do experimento do sorteio das duas variedades, definindo-se X como sendo a variável aleatória relativa a contagem de variedades da espécie A na amostra, verificamos que podem ser assumidos pela variável X , isto é, $x = 0, 1, 2$. É comum representar a variável por X (maiúsculo) e os seus valores por x (minúsculo). Considerando ainda, que para esse exemplo os pontos de Ω são equiprováveis, então a probabilidade de X assumir um dado valor x será denotado por $P(X = x)$, $p_X(x)$ ou p_i com $i = 1, 2, \dots$, sendo denominada também de função de probabilidade de X , para o caso da variável discreta. Supondo que desejamos calcular a probabilidade de C_3 ocorrer, temos:

$$P(C_3) = P(\{\omega \in \Omega : \omega \in C_3\}) = P(\{(A_1, A_2)\}) = \frac{\#\{(A_1, A_2)\}}{\#\Omega} = \frac{1}{10}. \quad (5.23)$$

Vamos observar que de modo equivalente iremos calcular a probabilidade do evento C_3 agora olhando para a variável X , tal que $P(C_3) = p_X(2) = P(X = 2)$, que segue,

$$\begin{aligned} p_X(2) &= P(X = 2) = P_X(B \in \Omega_X), \quad (B = \{2\}, \text{Figura 5.3}) \\ &= P_X(\{x \in \Omega_X : X(x) \in B, x \in \Omega\}) \\ &= P(X^{-1}(2)) \\ &= P(\{\omega \in \Omega : X(\omega) = 2\}) \\ &= P(\{\omega \in \Omega : \omega \in C_3\}) \\ &= P(\{(A_1, A_2)\}) \\ &= P(C_3) \\ &= \frac{1}{10}, \quad (\text{resultado (5.23)}). \end{aligned} \quad (5.24)$$

Portanto, a partir de agora em diante, iremos calcular as probabilidades dos eventos a partir da variável aleatória. Para um mesmo espaço amostral, é possível associar outras variáveis aleatórias. No exemplo que estava sendo considerado, poderíamos pensar em uma variável aleatória Y

que representasse o número de espécies da variedade E . A função de probabilidade de X define a distribuição de probabilidade dessa variável aleatória.

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

5.7 Distribuição de X

Consideramos que a distribuição de X como o conjunto das probabilidades $P(X(w) \in B) = P(\{\omega \in \Omega : X(\omega) \in B\})$, em que B representa todos os subconjuntos dos reais, isto é, $B \in \mathbb{R}$. Assim, especificaremos nas subseções seguintes, o detalhamento da distribuição de X para as variáveis aleatórias discretas e contínuas.

5.8 Função de probabilidade (FP)

Considerando os valores possíveis x_1, x_2, \dots , de uma variável aleatória discreta X , e suas respectivas probabilidades, p_1, p_2, \dots , então a função que associa x_i com sua respectiva p_i é denominada de função de probabilidade, representada como:

X	x_1	x_2	...
p_X	p_1	p_2	...

Em notação, dizemos que a função de probabilidade é dada da seguinte forma:

$$P(X = x_i) = p_X(x_i) = p_i, \quad i = 1, 2, \dots \quad (5.25)$$

Uma distribuição de probabilidade satisfaz os seguintes axiomas:

1. $0 \leq p_i \leq 1$;
2. $\sum_i p_i = 1$.

Vejamos o Exemplo 5.25, para o entendimento da função de probabilidade, apresentada a seguir.

Exemplo 5.25:

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

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

5.9 Função densidade de probabilidade (FDP)

Podemos iniciar afirmando que uma variável aleatória X é contínua se $P_X(X = x) = 0$. Assim, como poderíamos atribuir probabilidade às variáveis aleatórias? Com o intuito de resolver esse problema, definimos a função densidade de probabilidade, que servirá como uma “probabilidade pontual” da variável aleatória contínua. Assim, dizemos que uma função densidade de probabilidade, denotada por $f_X(x)$, é uma função $f : \mathbb{R} \rightarrow \mathbb{R}$, que deve satisfazer as seguintes condições:

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

Da mesma forma, para calcular a probabilidade de X , tal que $a \leq X \leq b$, para $a < b$, é dada por:

$$P(a \leq X \leq b) = \int_a^b f(x)dx. \quad (5.26)$$

Portanto, como $P_X(X = x) = 0$, em se tratando de variáveis aleatórias contínuas, como consequência, as probabilidades calculadas sobre os intervalos $[a, b]$, $[a, b)$, $(a, b]$ e (a, b) serão sempre as mesmas para quaisquer valores de a e b .

Vamos verificar se uma f_X pode ser considerada uma função densidade de probabilidade, pelo Exemplo 5.26.

Exemplo 5.26:

Verifica se $f(x) = \frac{1}{128}e^{-x/128}$ é uma densidade de probabilidade para $x \geq 0$. A variável aleatória X representa o tempo de vida de uma espécie vegetal arbórea dada em anos. Para verificarmos se f_X é uma função densidade de probabilidade, devemos provar as condições:

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

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

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

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

integrando em ambos os lados, temos:

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

como $h(x) = e^{-mx}$, então,

$$\int e^{-mx}dx = \frac{-e^{-mx}}{m} + c. \quad (5.27)$$

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

$$\begin{aligned} \int_0^\infty \frac{1}{128}e^{-\frac{1}{128}x}dx &= \frac{1}{128} \underbrace{\int_0^\infty e^{-\frac{1}{128}x}dx}_{\text{idêntica a eq.(5.27)}}, \\ &= \frac{1}{128} \left[\frac{-e^{-(1/128)x}}{1/128} \right]_0^\infty, \\ &= \frac{1}{128} \left[0 - \left(\frac{-e^{-(1/128)\times 0}}{1/128} \right) \right], \\ &= \frac{1}{128} [128e^0] \\ &= \frac{1}{128} \times 128 \times 1 = 1. \end{aligned}$$

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

Vejamos o Exemplo 5.27, uma outra situação para determinarmos uma função densidade de probabilidade.

Exemplo 5.27:

Seja

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

Encontrar o valor de “k” na função para que $f(x)$ seja uma densidade de probabilidade (f.d.p.). Para determinar o valor de k , então

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

Assim,

$$\begin{aligned} \int_0^3 (1/6x + k)dx &= \left[1/6 \frac{x^2}{2} + xk \right]_0^3 \\ &= 1/6 \frac{3^2}{2} + 3k. \end{aligned} \quad (5.28)$$

Igualando (5.28) a 1, temos:

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

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

5.10 Função de distribuição acumulada

Na seção anterior, afirmamos que a distribuição de X é dada por $P(X(\omega) \in B)$, em que $B \in \mathbb{R}$. Existe um caso especial, em que $B = (-\infty, x]$, em que $x \in \mathbb{R}$. Assim, dizemos que $P(X(\omega) \in (\infty, x])$ representa a função de distribuição ou função de distribuição acumulada (FD ou FDA), e em notação temos $F_X(x) = P(X(\omega) \in (\infty, x])$. A seguir, apresentamos a função de distribuição para as variáveis aleatórias discretas e contínuas.

5.10.1 FDA para X discreta

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

$$F_X(x) = P(X \leq x) = \sum_{\{j|x_j < x\}} p(x_j) \quad (5.29)$$

Vejamos o Exemplo 5.28, para apresentarmos o cálculo da função de distribuição de uma variável aleatória discreta, a seguir.

Exemplo 5.28:

Considere um estudo hipotético do qual desejamos imunizar 1000 pessoas em uma comunidade rural da doença da COVID-19, por meio de uma determinada vacina. Supomos que sejam aplicados 5 doses em cada pessoa, em períodos espaçados, dessa vacina. A cada dose aplicada, as pessoas passam por uma série de avaliações para a verificar se adquiriu imunidade ou não. Caso se verifique a imunidade em uma determinada dose aplicada, esta pessoa não irá tomar a dose subsequente; caso contrário, seguirá tomando as doses subsequentes, até a 5ª dose. Os resultados completos, são apresentados a seguir.

Doses	1	2	3	4	5
Frequência	230	270	300	120	80

Considerando uma pessoa dessa comunidade sorteada ao acaso, poderíamos estar interessados em saber qual a probabilidade dela ter recebido 2 doses? usando a ideia da probabilidade frequentista, a probabilidade desejada é de $270/1000=0,27$. Podemos assim obter a função de probabilidade para a variável aleatória número de doses recebidas, que também pode ser observado pela Figura 5.4.

Doses	1	2	3	4	5
p_i	0,23	0,27	0,30	0,12	0,08

Pela função de distribuição, podemos responder, por exemplo, a chance de uma determinada pessoa dessa população ter tomado até duas doses, que segue,

$$F_X(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = 0,50.$$

Apesar da escolha de x ter sido sempre um número inteiro até agora, esse valor fica inalterado no intervalo $[2, 3)$. Isto significa que, $F_X(2, 3)$, $F_X(2, 45)$ ou $F_X(2, 99)$ têm os mesmos valores que $F_X(2)$. Por isso, escrevemos

$$F_X(x) = P(X \leq x) = 0,50, \text{ para } 2 \leq x < 3.$$

Por fim, apresentamos a função de distribuição para todo x , como também o gráfico dessa função, Figura 5.5.

$$F(X) = \begin{cases} 0, & \text{se } x < 1; \\ 0,23, & \text{se } 1 \leq x < 2; \\ 0,50, & \text{se } 2 \leq x < 3; \\ 0,80, & \text{se } 3 \leq x < 4; \\ 0,92, & \text{se } 4 \leq x < 5; \\ 1, & \text{se } x \geq 5. \end{cases}$$

5.10.2 FDA para X contínua

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

$$F(x) = \int_{-\infty}^x f(t)dt, \quad x \in \mathbb{R}. \quad (5.30)$$

5.10.2.1 Propriedades

Apresentamos algumas propriedades da função de distribuição:

i) $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ e $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$;

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

iii) $F(x)$ é contínua à direita, $\lim_{x \rightarrow a^+} F(x) = F(a)$

Essas propriedades podem ser melhor observadas pela Figura 5.6.

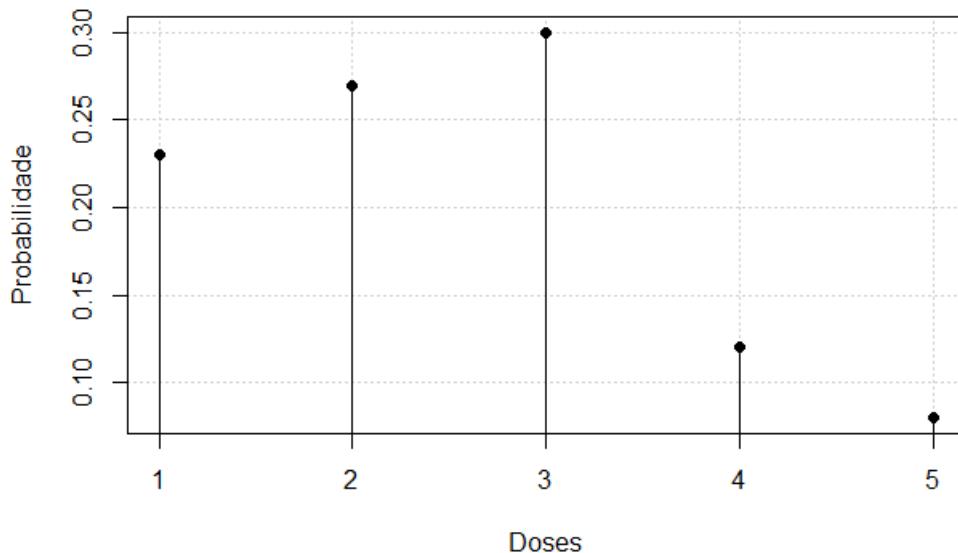


Figura 5.4: Gráfico da função de probabilidade.

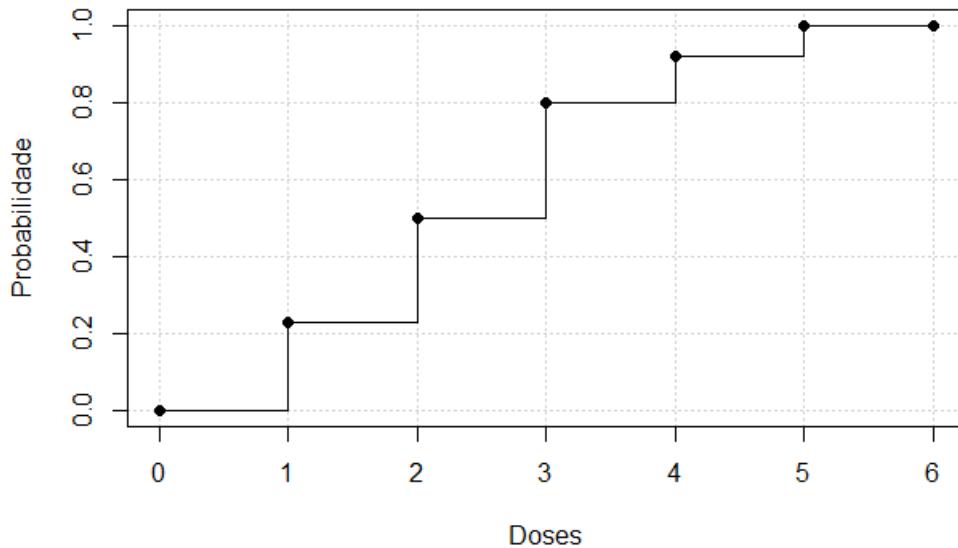


Figura 5.5: Gráfico da função de distribuição.

5.10.2.2 Relação entre $f_X(x)$ e $F(x)$

Em algumas situações podemos determinar a função densidade de probabilidade pelo conhecimento da função de distribuição, e vice-versa, como pode ser observados nas relações que seguem:

- Se $f(x)$ é conhecido, então $F(x) = \int_{-\infty}^x f(x)dx$
- Se $F(x)$ é conhecido, então $f(x) = \frac{d}{dx}F(x) \quad \forall x$ em que a derivada existe.

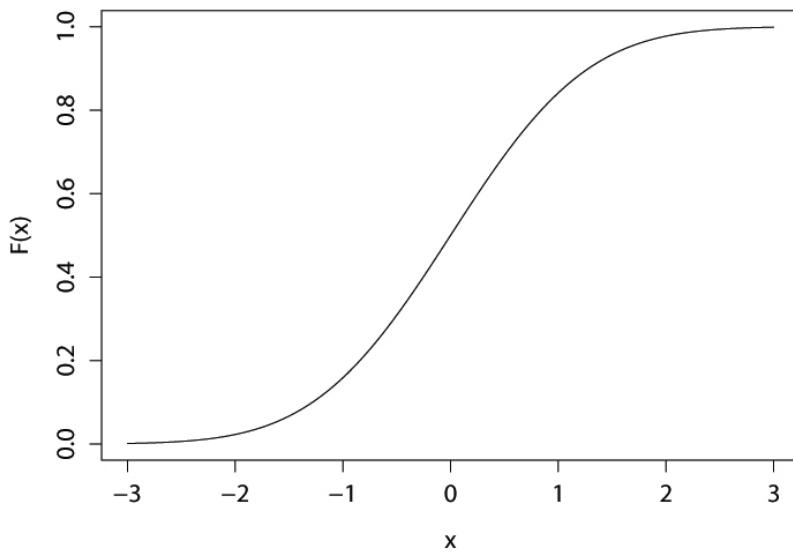


Figura 5.6: Gráfico da função de distribuição.

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

5.11 Medidas resumo

Assim como estudado na Estatística descrita, podemos apresentar as medidas de posição em termos de variável aleatória, do qual temos a esperança matemático ou valor esperado (que é o correspondente para a média aritmética na estatística descritiva), mediana e moda. Iniciaremos pela esperança matemática, apresentada na Definição 5.23.

Definição 5.23: Esperança matemática

Seja X uma variável aleatória, então a esperança matemática (ou média) de X , denotada por μ_X ou $E[X]$, é definida:

- i) se X for discreta,

$$E[X] = \sum_i x_i p_X(x_i), \quad (5.31)$$

para um conjunto de pontos $x_1, x_2, \dots, x_i, \dots$;

- ii) se X for contínua,

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx, \quad (5.32)$$

sendo $f_X(x)$ uma função densidade de probabilidade.

Vejamos algumas propriedades da esperança matemática:

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

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

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

- iv) $E[X \pm k] = E[X] \pm k;$
- v) $E[XY] = E[X]E[Y]$ se X e Y forem independentes.

Uma outra medida resumo é a mediana, denotada por μ_d , que é o valor que satisfaz:

$$P(X \geq \mu_d) = P(X \leq \mu_d) = 0,50, \quad (5.33)$$

Isto é, a probabilidade X ser maior ou menor que μ_d é igual. Em algumas situações, a igualdade é satisfeita considerando um valor em um certo intervalo, tal que μ_d seja o ponto médio de intervalo.

Por fim, apresentamos a moda, como sendo o valor (ou valores) da variável aleatória que apresenta maior probabilidade de ocorrência. A moda é denotada por μ_o , do qual pode ser definida como:

$$P(X = \mu_o) = \max(p_1, p_2, \dots, p_n). \quad (5.34)$$

Exemplo 5.29:

Considere X com a seguinte função de probabilidade,

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

Então,

$$E[X] = \sum_{i=1}^n x_i p(x_i) = (0) \times 0,3 + 5 \times 0,2 + \dots + 20 \times 0,1 = 9 \text{ unid.}$$

A moda é o valor com maior probabilidade e, portanto segue imediatamente que $\mu_o = 15$ unid.. No caso da mediana, podemos usar qualquer número entre 5 e 15, pois,

$$P(X \leq 5) = P(X \geq 15) = 0,50.$$

Por conveniência adotada, tomamos $\mu_d = 10$ unid. como o valor da mediana.

A próxima medida apresentada, representa a variabilidade dos valores assumidos a variável aleatória X , em torno do valor central μ , que representa a esperança matemática, que é a variância, Definição 5.24.

Definição 5.24: Variância

Seja X uma variável aleatória e sua esperança matemática dada por $E[X] = \mu$, então a variância de X , denotada por σ_X^2 ou $Var[X]$, é definida:

- i) se X for discreta, por:

$$\sigma_X^2 = \sum_i (x_i - \mu)^2 P_X(x_i), \quad (5.35)$$

para um conjunto de pontos $x_1, x_2, \dots, x_i, \dots$;

- ii) se X for contínua, por:

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx, \quad (5.36)$$

sendo $f_X(x)$ uma função densidade de probabilidade.

Assim, como observamos em estatística descritiva, a variância apresenta um problema que é a unidade ao quadrado da unidade da variável, e isso acaba dificultando a interpretação prática de seu resultado. Como solução, usamos a Definição 5.25, para apresentar o desvio padrão de X .

Definição 5.25: Desvio padrão

Seja X uma variável aleatória, então o desvio padrão de X , denotado por σ_X , é definido por:

$$\sigma_X = +\sqrt{\sigma_X^2}, \quad (5.37)$$

em que σ_X^2 é a variância de X .

Podemos obter de forma alternativa a variância de X , usando o Teorema 5.11, a seguir.

Teorema 5.11

Se X um variável aleatória em (Ω, \mathcal{F}, P) , então

$$Var[X] = E[X^2] - (E[X])^2, \quad (5.38)$$

dado que $E[X^2]$ existe. □

Prova:

Observe inicialmente que se $E[X^2]$ existe, então $E[X]$ também existe. Logo,

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

Uma forma alternativa para a expressão (5.38) pode ser dado da seguinte forma:

$$\begin{aligned} Var[X] &= E[X^2] + (E[X] - E[X]) - (E[X])^2 \\ &= E[X(X - 1)] + E[X] - (E[X])^2. \end{aligned} \quad (5.39)$$

Esse resultado é importante para demonstrar a variância de algumas distribuições, como por exemplo a distribuição geométrica. Por fim, apresentamos algumas propriedades da variância de X , que segue:

- i) $Var(k) = 0$, sendo k uma constante;
- ii) $Var(kX) = k^2Var(X)$;
- iii) $Var(k \pm X) = Var(X)$;
- iv) $Var(X \pm Y) = Var(X) \pm Var(Y)$ se X e Y forem independentes.

5.12 Exercícios

Exercício 5.1: Defina o que é um experimento aleatório e exemplifique.

Solução na página 217

Exercício 5.2: Defina o que é um espaço amostral e exemplifique.

Solução na página 217

Exercício 5.3: Para cada um dos casos abaixo, escreva o espaço amostral correspondente e conte seus elementos.

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

Solução na página 217

Exercício 5.4: Duas pessoas A e B , lançam três moedas de modo independente, e nessa ordem, primeiro a pessoa A , depois a pessoa B , alternadamente. Nesse jogo, ganha quem conseguir três faces iguais, sendo finalizado com a vitória de um dos competidores. Assim, perguntamos:

- Qual a probabilidade de A ganhar?
- Qual a probabilidade de B ganhar?

Solução na página 217

Exercício 5.5: Uma Universidade tem 10 mil alunos dos quais 4 mil são considerados esportistas. Temos ainda que 500 alunos são do curso de Administração noturno, 700 de Ciências contábeis noturno, 100 são esportistas e da Administração noturno e 200 são esportistas e da Ciências contábeis noturno. Qual a probabilidade de:

- Ser esportista.
- Ser esportista e aluno da Administração.
- Ser esportista ou aluno da Ciências contábeis.
- Não ser esportista nem aluno da Administração.

Solução na página 218

Exercício 5.6: Sejam A e B dois eventos em um dado espaço amostral, tais que $P(A) = 0,2$, $P(B) = p$, $P(A \cup B) = 0,5$ e $P(A \cap B) = 0,1$. Determine o valor de p .

Solução na página 218

Exercício 5.7: Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{4}$, e A e B são mutuamente exclusivos, calcular:

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

Solução na página 219

Exercício 5.8: Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{3}$ e $P(A \cap B) = \frac{1}{3}$. Calcule:

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

Solução na página 219

Exercício 5.9: Dois dados são lançados simultaneamente. Qual a probabilidade de:

- a) a soma ser menor que 4.
- b) a soma ser 9.
- c) o primeiro resultado ser maior que o segundo.

Solução na página 219

Exercício 5.10: Qual a probabilidade de sair um rei ou uma carta de copas, quando retiramos uma carta de um baralho?

Solução na página 219

Exercício 5.11: As probabilidades de três jogadores marcarem um pênalti são respectivamente: $\frac{2}{3}$, $\frac{4}{5}$ e $\frac{7}{10}$. Se cada um “cobrar” uma única vez, qual a probabilidade de:

- a) todos acertarem?
- b) apenas um acertar?
- c) todos errarem?

Solução na página 219

Exercício 5.12: Dois armários guardam as bolas de voleibol e basquete. O armário 1 tem três bolas de voleibol e 1 de basquete, enquanto o armário 2 tem 3 de voleibol e 2 de basquete. Escolhendo-se ao acaso um armário e, em seguida, uma de suas bolas, calcule a probabilidade dela ser:

1. De voleibol, sabendo-se que o armário 1 foi escolhido.
2. De basquete, sabendo-se que o armário 2 foi escolhido.
3. De basquete.

Solução na página 220

Exercício 5.13: Em certo colégio, 5% dos homens, 2% das mulheres têm mais do que 1,80m. Por outro lado, 60% dos estudantes são homens. Se um estudante é selecionado aleatoriamente e tem mais de 1,80m de altura, qual a probabilidade de que o estudante seja mulher?

Solução na página 220

Exercício 5.14: A probabilidade de uma mulher está viva daqui a 30 anos é $\frac{3}{4}$ e a de seu marido, $\frac{3}{5}$. Calcular a probabilidade de:

- a) apenas o homem está vivo;
- b) somente a mulher está;
- c) ambos estarem vivos.

Solução na página 221

Exercício 5.15: Se $P(A \cup B) = 0,8$, $P(A) = 0,5$ e $P(B) = x$, determine o valor de x no caso de:

1. A e B serem mutuamente exclusivos.
2. A e B serem independentes.

Solução na página 221

Exercício 5.16: Se $P(B) = 0,4$, $P(A) = 0,7$ e $P(A \cap B) = 0,3$, calcule $P(A|B^c)$.

Solução na página 221

Exercício 5.17: O São Paulo Futebol Clube ganha com probabilidade 0,7 se chove e 0,8 se não chove. Em Dezembro, a probabilidade de chuva é de 0,3. O São Paulo ganhou uma partida em Dezembro, qual a probabilidade de ter chovido nesse dia?

Solução na página 221

Exercício 5.18: Três máquinas A, B, e C produzem, respectivamente 30%, 50% e 20% do total de peças de uma fábrica. As porcentagens de peças defeituosas nas respectivas máquinas são 3%, 5% e 2%. Uma peça é sorteada ao acaso, e verifica-se que é defeituosa. Qual a probabilidade de que a peça tenha vindo da máquina B?

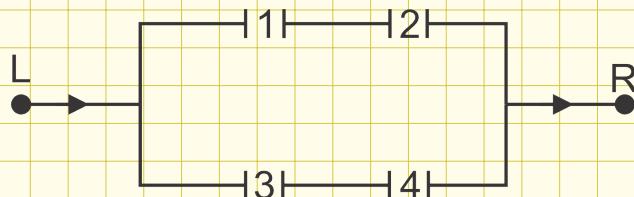
Solução na página 222

Exercício 5.19: Numa certa população, a probabilidade de gostar de teatro é de $1/3$, enquanto que a de gostar de cinema é $1/2$. Determine a probabilidade de gostar de teatro e não de cinema, nos seguintes casos:

- a) Gostar de teatro e gostar de cinema são eventos disjuntos.
- b) Gostar de teatro e gostar de cinema são eventos independentes.
- c) Todos que gostam de teatro gostam de cinema.
- d) A probabilidade de gostar de teatro e de cinema é de $1/8$.
- e) Dentre os que não gostam de cinema, a probabilidade de não gostar de teatro é de $3/4$.

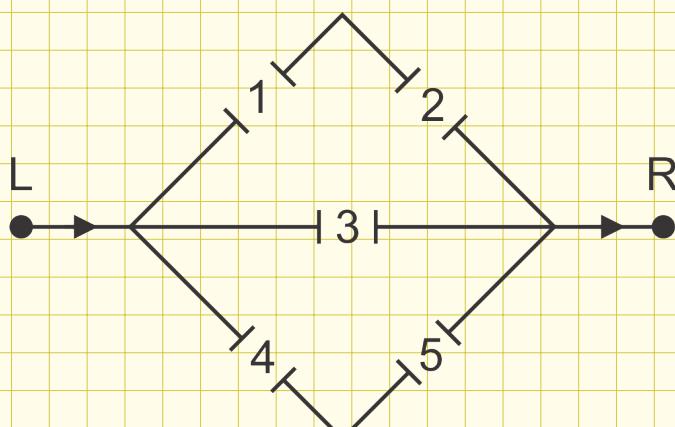
Solução na página 222

Exercício 5.20: A probabilidade de fechamento de cada relé do circuito apresentado a seguir é dada por p . Se todos os relés funcionarem independentemente, qual será a probabilidade de que haja corrente entre os terminais L e R?



Solução na página 222

Exercício 5.21: A probabilidade de fechamento de cada relé do circuito apresentado a seguir é dada por p . Se todos os relés funcionarem independentemente, qual será a probabilidade de que haja corrente entre os terminais L e R?



Solução na página 223

Exercício 5.22: Considerando a população brasileira vacinada de covid-19 até o dia 13/06/2021, de acordo com o Ministério da Saúde, de uma população vacinável de 160.044.909 milhões de pessoas, observou-se que 826.158 mil pessoas tomaram apenas a primeira dose da vacina, 1.950.914 milhões de pessoas tomaram a primeira e a segunda dose, sendo que 157.267.837 milhões de pessoas ainda não tomaram vacina. Considerando essas informações apresente a função de distribuição de X , sendo X o número de doses aplicadas na população vacinável.

Solução na página 223

Exercício 5.23: Considere X uma variável aleatória discreta, cuja sua função de distribuição é dada por:

$$F_X(x) = \begin{cases} 0, & \text{se } x < 5, \\ 0,2, & \text{se } 5 \leq x < 7, \\ 0,5, & \text{se } 7 \leq x < 8, \\ 0,9, & \text{se } 8 \leq x < 20, \\ 1, & \text{se } x \geq 20. \end{cases}$$

Determine:

- a) a função de probabilidade de X ;
- b) $P(X \leq 7)$;
- c) $P(X < 7)$;
- d) $P(8 \leq X \leq 18)$;
- e) $P(X \geq 15)$;
- f) qual é o valor da esperança matemática? E a mediana? E a moda?

Solução na página 223

Exercício 5.24: Sejam dois eventos não vazios A e B de Ω , e que $\omega \in C = (A \cap B^c) \cup (A^c \cap B)$ representam os elementos que estão exatamente em um dos dois eventos, isto é, não há elementos em comum entre esses dois eventos. Assim, prove que

$$P(C) = P(A) + P(B) - 2P(A \cap B).$$

Solução na página 224

Distribuições de probabilidades

6.1 Introdução

O estudo das distribuições especiais abrange as principais distribuições discretas (Bernoulli, Binomial e Poisson) e contínuas (Normal). Essas distribuições são de tal importância, pois grande parte das variáveis podem ser modeladas ou derivadas por meio delas.

6.2 Distribuições discretas de probabilidades

Considere um experimento cujo resultado apresenta apenas duas possibilidades *sucesso* ou *fracasso*. Por exemplo, uma peça produzida em um indústria automobilística pode ter sido fabricada com duas possibilidade, defeituosa ou não, assim como um *bit* transmitido por um canal digital pode ter sido realizado com erro ou não. Situações como essas, dizemos que houve *sucesso* ou *fracasso*. Mais cuidado, o *sucesso* não representa necessariamente o que é *bom* ou *favorável* para a situação, mas o que é de interesse para o pesquisador. Matematicamente, dizemos a variável aleatória X , que contará esse sucesso, assumirá $X = 1$, e o fracasso $X = 0$. Se considerarmos p como a probabilidade do sucesso, teremos o complemento dessa probabilidade, $1 - p$, como a probabilidade do fracasso.

Esse tipo de experimento é chamado de experimento de Bernoulli , devido aos primeiros estudos sobre esses tipos de experimentos terem sido feitos pelo matemático suíço Jacob Bernoulli. Desse modo, o modelo probabilístico de X que conta o sucesso desse experimento é modelado pela distribuição Bernoulli apresentado na Definição 6.1, a seguir.

Definição 6.1: Distribuição Bernoulli

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

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

em que p satisfaz $0 < p < 1$. Em notação, $X \sim Bernoulli(p)$ representa que X tem distribuição Bernoulli com parâmetro p .

Em outras palavras, define-se uma v.a. X que assume dois possíveis resultados: o valor 1, se ocorrer o evento desejado (sucesso); e o valor 0, se ocorrer o valor não desejado (fracasso). Definindo-se ainda p como sendo a probabilidade de ocorrer o sucesso, e $(1 - p)$, a probabilidade de ocorrer o fracasso, então X tem distribuição de probabilidade Bernoulli. Pelo Teorema 6.1, provamos que a distribuição Bernoulli é uma função de probabilidade.

Teorema 6.1: Função de probabilidade Bernoulli

A distribuição Bernoulli expressa em (6.1), é uma função de probabilidade.

Prova:

Para mostrarmos que a $p_X(x)$ da expressão (6.1), devemos demonstrar que:

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

Prova: Como os valores de $x = 0$ ou 1 e $0 < p < 1$, então $p^x(1 - p)^{1-x} \geq 0$.

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

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

Graficamente, apresentamos na Figura 6.1, a distribuição de X , modelado por uma distribuição Bernoulli.

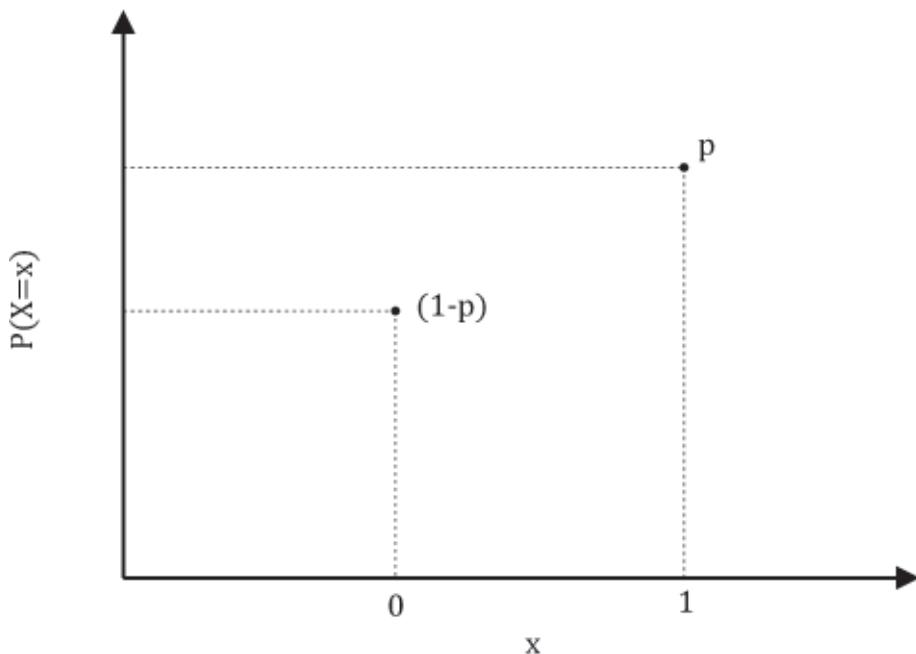


Figura 6.1: Função de probabilidade da distribuição Bernoulli com parâmetro p .

Vejamos o Exemplo 6.1, uma aplicação da distribuição Bernoulli, a seguir.

Exemplo 6.1:

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

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

Para complementarmos a caracterização de uma variável aleatória X modelada por uma distribuição Bernoulli, apresentamos no Teorema 6.2, a esperança, variância e desvio padrão, a seguir.

Teorema 6.2: Esperança, Variância e Desvio padrão da distribuição de Bernoulli

Se a variável aleatória X possui distribuição Bernoulli, cuja função de probabilidade é expressa em (6.1), a esperança de X é dada por:

$$\mu_X = p, \quad (6.2)$$

a variância é dada por:

$$\sigma_X^2 = p(1 - p), \quad (6.3)$$

e o desvio padrão é dado por:

$$\sigma_X = \sqrt{p(1 - p)}. \quad (6.4)$$

□

Prova:

Sendo a esperança definida por:

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

e considerando o experimento de uma variável aleatória X com distribuição Bernoulli, em que $x = 1$, sucesso para o experimento com probabilidade p e $x = 0$, o fracasso, com probabilidade $(1 - p)$, então a esperança é dada por

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

e a variância sendo $\sigma_X^2 = E[X^2] - E^2[X]$, então,

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

e

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

portanto,

$$\sigma_X^2 = p - p^2 = p(1 - p).$$

O desvio padrão é dado por:

$$\sigma_X = \sqrt{p(1 - p)},$$

o que conclui a prova.

Exemplo 6.2:

Usando o Exemplo 6.1, temos que $E[X] = 2/5$ bolas verdes, $\sigma_X^2 = (2/5) \times (3/5) = 6/25$ (bolas verdes)², e $\sigma_X = 0,4899$ bolas verdes.

Considere agora n experimentos de Bernoulli que satisfaça as seguintes condições:

- i) os n experimentos são independentes, de modo que a chance do resultado de um experimento não interfere na chance de ocorrência de um determinado resultado em outro experimento;
- ii) designamos um certo resultado de *sucesso*, e quando esse resultado não ocorre, dizemos que resultou em *fracasso*;
- iii) A probabilidade de *sucesso* de qualquer resultado de um experimento é sempre a mesma para todos os n experimentos.

Nesse contexto, generalizamos a distribuição Bernoulli, pela distribuição Binomial, apresentada na Definição 6.2, a seguir.

Definição 6.2: Distribuição Binomial

Uma variável aleatória X discreta, tem distribuição Binomial , se sua função de probabilidade é dada por

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x}, & \text{para } x = 0, 1, 2, \dots, n, \\ 0, & \text{caso contrário,} \end{cases} \quad (6.5)$$

em que $\binom{n}{x} = n!/[x!(n - x)!]$, e p é o parâmetro que indica o número de sucessos do evento ocorrer, sendo $0 < p < 1$. Em notação, $X \sim \text{Binomial}(n, p)$ representa que X tem distribuição Binomial com parâmetros n e p .

Vejamos o Exemplo 6.3, para apresentar uma aplicação da distribuição Binomial.

Exemplo 6.3:

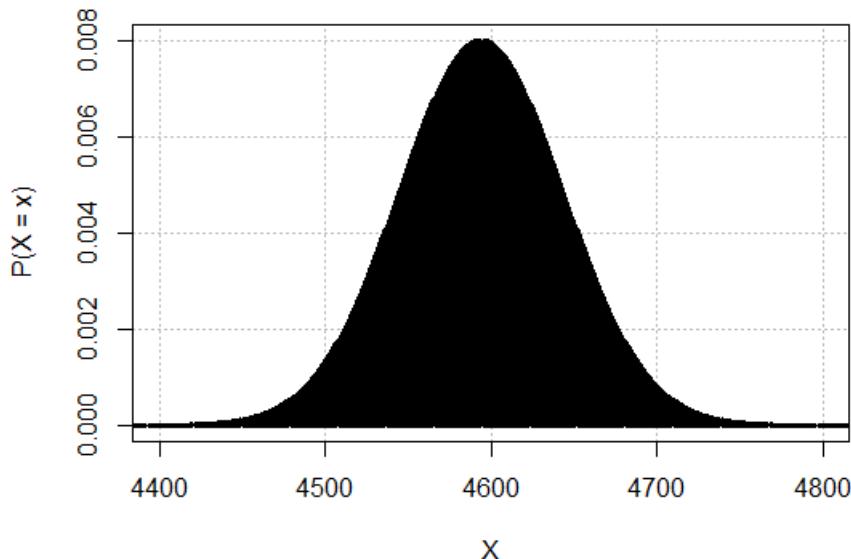
Considere o sistema de segurança de acesso a conta de usuários de um determinado banco, por meio de uma senha de 6 dígitos, com 26 caracteres (a-z) ou números (0-9), e que esse banco tem em sua carteira 10.000 clientes com senhas únicas. Um *hacker* tentou invadir o sistema, e selecionou aleatoriamente (com reposição) um bilhão de senhas únicas do potencial conjunto de possibilidades do sistema de segurança. Então o *hacker* irá testar o seu conjunto de senhas em cada uma das senhas dos clientes, de modo que X representará o número de êxitos com que o hacker conseguiu detectar cada senha. Indagamos:

- qual a distribuição de X ?
- qual a probabilidade do *hacker* acertar ao menos uma senha?

Inicialmente, sabemos que a quantidade de senhas possíveis são de 36^6 , uma vez que os dígitos poderão ser repetidos, com caracteres ou números. Considere que o experimento consiste no *hacker* verificar se uma determinada senha de um cliente está ou não no seu conjunto de senhas. Isto consiste em um experimento de Bernoulli de parâmetro $p = 10^9/36^6 = 0,4593937$, tal que, $Y \sim \text{Bernoulli}(p)$, isto é, a probabilidade de $Y = 1$ representa o êxito do *hacker* em ter encontrado determinada senha em seu banco de senhas, e $Y = 0$, caso contrário. Porém, isso será verificado para todas as $n = 10.000$ senhas dos clientes, assumindo que essas n tentativas são independentes. Então $X = \sum_{i=1}^{10.000} Y_i$ é a variável aleatória discreta que conta o número de êxitos do *hacker* em ter encontrado as senhas dos clientes. Logo, (a) dizemos que $X \sim \text{Binomial}(n, p)$. Para o segundo ítem (b), calculamos $P(X \geq 1)$, da seguinte forma:

$$\begin{aligned} P(X \geq 1) &= \sum_{i=1}^{10.000} \binom{10.000}{x} (0,4593937)^x (1 - 0,4593937)^{10.000-x} \\ &= 1 - P(X = 0) \\ &= 1, \end{aligned}$$

e no R, isso foi calculado pelo código: `1 - dbinom(0, 10000, 0.4593937)` ou `pbinom(0, 10000, 0.4593937, lower.tail = FALSE)`. Esse resultado pode ser melhor entendido quando verificamos o gráfico da distribuição na Figura 6.2, que individualmente a probabilidade de $X = x$ é praticamente 0, porém, a probabilidade de X estar em uma determinada região passa a assumir um determinado valor diferente de zero. O gráfico da distribuição binomial pode ser obtida pela Código 6.1.

Figura 6.2: Distribuição de $X \sim \text{Binomial}(10,000, 10^9/36^6)$.

Código R 6.1

Script:

```

1 # Valores
2 n <- 10000; x <- 0:n; p <- 0.459394
3 # Probabilidade
4 px <- dbinom(x, size = n, prob = p)
5 # Gráfico
6 plot(x, px, type = "h", xlab = "X", ylab = expression("P(X=x)"),
      panel.first = grid(col="gray"), ylim = c(0, 0.008), xlim = c(4400, 4800),
      lwd = 3)

```

Observamos que a expressão (6.5) reduz-se a expressão (6.1) quando $n = 1$. Isso reforça que a distribuição Binomial é uma generalização para a distribuição Bernoulli. Vamos demonstrar, pelo Teorema 6.5, que a distribuição Binomial também é uma função de probabilidade, a seguir.

Teorema 6.3: Função de probabilidade Binomial

A distribuição Binomial expressa em (6.5), é uma função de probabilidade .

Prova:

Para mostrarmos que a distribuição Binomial é uma função de probabilidade, devemos mostrar as seguintes propriedades:

- i) Mostrar que $P(X = x) \geq 0$;
- ii) Mostrar que $\sum_{x=0}^n P(X = x) = 1$.
- iii) Como $n > 0, x \geq 0, n > x$ e $0 < p < 1$, então $\binom{n}{x} \geq 0, p^x \geq 0$ e $(1 - p)^{n-x} \geq 0$.

Portanto,

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \geq 0.$$

- ii) Antes de chegarmos ao resultado, vamos entender uma propriedade do binômio de Newton. Para $n = 3$, temos

$$\begin{aligned} x = 0 &\Rightarrow \binom{3}{0} p^0 (1-p)^{3-0} = (1-p)^3 = q^3, \\ x = 1 &\Rightarrow \binom{3}{1} p^1 (1-p)^{3-1} = 3p(1-p)^2 = 3pq^2, \\ x = 2 &\Rightarrow \binom{3}{2} p^2 (1-p)^{3-2} = 3p^2(1-p)^1 = 3p^2q, \\ x = 3 &\Rightarrow \binom{3}{3} p^3 (1-p)^{3-3} = p^3(1-p)^0 = p^3. \end{aligned}$$

Agora observe,

$$(p+q)^3 = p^3 + 3p^2q + 3pq^2 + q^3,$$

então, podemos concluir que

$$\begin{aligned} \sum_{x=0}^3 P(X = x) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= (p+q)^3. \end{aligned}$$

Expandindo esse resultado para n , podemos observar que

$$\sum_{x=0}^n P(X = x) = (p+q)^n. \quad (6.6)$$

Sabendo que $q = 1 - p$, então

$$\begin{aligned} \sum_{x=0}^n P(X = x) &= [p + (1-p)]^n, \\ &= 1^n = 1. \end{aligned} \quad (6.7)$$

Portanto, $P(X = x)$ é uma função de probabilidade.

Apresentamos, algumas características da distribuição Binomial:

1. Uma distribuição binomial fica caracterizada pelos parâmetros n e p .
2. Para qualquer n , a distribuição será simétrica, se $p = q = 0,5$, será assimétrica à direita, se $p > q$, e assimétrica à esquerda, se $p < q$, como podemos perceber na Figura 6.3;
3. Para um determinado valor p , à medida que n a distribuição Binomial se torna cada vez mais simétrica, Figura 6.4.

A criação das Figuras 6.3 e 6.4 foram desenvolvidas usando os Códigos 6.2 e 6.4.

Código R 6.2

Script:

```

1 #-----
2 # Consideramos a alteracao do valor do parametro p
3 #
4 # Grafico da distribuicao binomial para n = 30 e p = 0.5
5 x <- 0:30; n <- 30; p <- 0.5
6 px <- dbinom(x, size = n, prob = p)
7 plot(x, px, type = "h", xlab = "X", ylab = expression("P(X=x)"), panel.first
      = grid(col="gray"), ylim = c(0, 0.5), lwd = 2)
8 # n = 30 e p = 0.1
9 x <- 0:30; n <- 30; p <- 0.1
10 px <- dbinom(x, size = n, prob = p)
11 lines(x, px, col = "red", type = "h", lwd = 2)
12 # n = 30 e p = 0.9
13 x <- 0:30; n <- 30; p <- 0.9
14 px <- dbinom(x, size = n, prob = p)
15 lines(x, px, col = "green", type = "h", lwd = 2)
16 # Legendas
17 legend(20, 0.5, legend = c("n=30;p=0.5",
18                           "n=30;p=0.1",
19                           "n=30;p=0.9"),
20        col = c("black", "red", "green"),
21        bg = "antiquewhite1", lty = 1, cex = 0.8, lwd = 2)

```

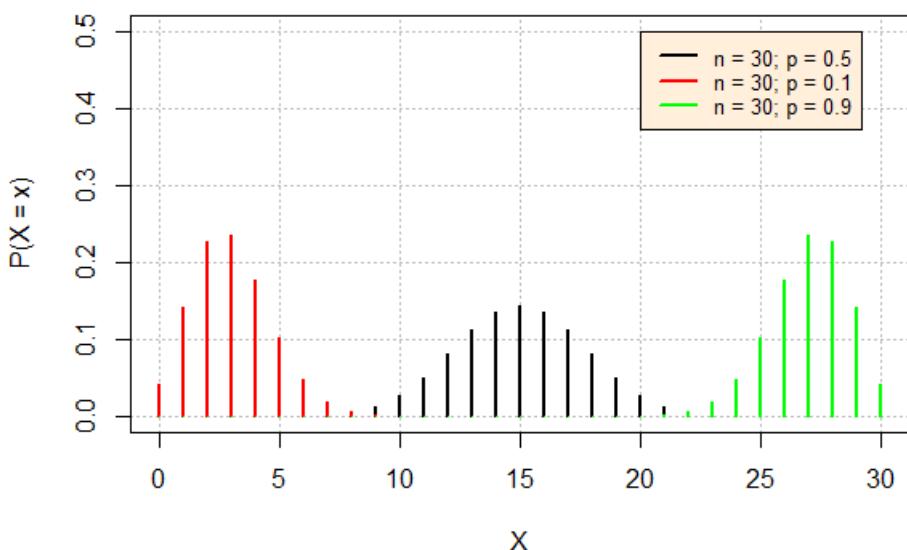


Figura 6.3: Gráfico da função de probabilidade da distribuição binomial, de parâmetros $n = 30$ e diferentes valores de p .

Para finalizar, podemos obter pelo Teorema 6.4, a esperança, variância e desvio padrão da distribuição Binomial.

Teorema 6.4: Esperança, Variância e Desvio padrão da distribuição de Binomial

Se a variável aleatória X possui distribuição Bernoulli, cuja função de probabilidade é expressa em (6.1), a esperança de X é dada por:

$$E[X] = np, \quad (6.8)$$

a variância é dada por:

$$\sigma_X^2 = np(1 - p). \quad (6.9)$$

e o desvio padrão é dado por:

$$\sigma_X = \sqrt{np(1 - p)}. \quad (6.10)$$

Prova:

Vamos apresentar duas provas. Na primeira, usaremos as propriedades da esperança e variância e da independência de variáveis aleatórias, abordados no Capítulo 5, seção 5.11. Na segunda prova, iremos apresentar os resultados por meio da definição de esperança e variância. Dessa forma, temos:

- (I) Sabemos que se X_1, X_2, \dots, X_n são variáveis aleatórias independentes e identicamente distribuídas^a (*iid*) com distribuição de Bernoulli de parâmetro p , então $Y = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$. Logo,

$$\mu_Y = E[Y] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = nE[X] = np,$$

considerando $E[X]$ em (6.2). A variância pode ser obtida pelo mesmo raciocínio feito para obter a esperança, isto é,

$$\sigma_Y^2 = Var[Y] = Var\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n Var[X_i] = nVar[X] = np(1 - p),$$

considerando $Var[X]$ em (6.3). O desvio é a raiz da variância, logo

$$\sigma_Y = \sqrt{np(1 - p)},$$

o que conclui a primeira prova.

(II) Para a segunda prova, iniciaremos pela esperança matemática:

$$\begin{aligned}
 \mu = E[X] &= \sum_{x=0}^n x p_X(x) \\
 &= 0 \times p_X(0) + \sum_{x=1}^n x p_X(x) \\
 &= \sum_{x=1}^n x \binom{n}{x} p^x (1-p)^{n-x} \\
 &= \sum_{x=1}^n \frac{n!x}{x!(n-x)!} pp^{x-1} (1-p)^{n-x} \\
 &= \sum_{x=1}^n \frac{n(n-1)!}{(x-1)!(n-x)!} pp^{x-1} (1-p)^{n-x} \\
 &= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}.
 \end{aligned}$$

Usando a transformação $y = x - 1$, então

$$\begin{aligned}
 \mu = E[X] &= np \underbrace{\sum_{y=0}^{n-1} \frac{(n-1)!}{y![(n-1)-y]!} p^y (1-p)^{(n-1)-y}}_{=1, \text{ consequência do resultado (6.7)}} \\
 \mu = E[X] &= np,
 \end{aligned} \tag{6.11}$$

como queríamos demonstrar. Para o cálculo da variância, calcularemos inicialmente $E[X^2]$, que segue:

$$\begin{aligned}
 E[X^2] &= \sum_{x=0}^n x^2 p_X(x) \\
 &= 0 \times p_X(0) + \sum_{x=1}^n x^2 p_X(x) \\
 &= \sum_{x=1}^n x^2 \binom{n}{x} p^x (1-p)^{n-x} \\
 &= \sum_{x=1}^n \frac{n!x^2}{x!(n-x)!} pp^{x-1} (1-p)^{n-x} \\
 &= \sum_{x=1}^n \frac{n(n-1)!x}{(x-1)!(n-x)!} pp^{x-1} (1-p)^{n-x} \\
 &= np \sum_{x=1}^n \frac{(n-1)!x}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}.
 \end{aligned}$$

Usando a transformação $y = x - 1$, então

$$\begin{aligned}
 E[X^2] &= np \sum_{y=0}^{n-1} (y+1) \frac{(n-1)!}{y![(n-1)-y]!} p^y (1-p)^{(n-1)-y} \\
 &= np \left[\underbrace{\sum_{y=0}^{n-1} y \frac{(n-1)!}{y![(n-1)-y]!} p^y (1-p)^{(n-1)-y}}_{E[Y]=(n-1)p} + \underbrace{\sum_{y=0}^{n-1} \frac{(n-1)!}{y![(n-1)-y]!} p^y (1-p)^{(n-1)-y}}_{=1} \right] \\
 &= np[(n-1)p + 1] \\
 &= np + (np)^2 - np^2.
 \end{aligned} \tag{6.12}$$

Usando a Definição 5.24 e substituindo os termos pelas expressões (6.11) e (6.12), então

$$\begin{aligned}
 \sigma_X^2 = Var[X] &= np + (np)^2 - np^2 - (np)^2 \\
 &= np - np^2 \\
 &= np(1-p),
 \end{aligned} \tag{6.13}$$

como queríamos demonstrar. Por fim, desse resultado obtemos o desvio padrão, dado por:

$$\sigma_X = \sqrt{np(1-p)},$$

o que concluímos toda a prova do teorema.

^aO termo independente está relacionado a chance de uma variável aleatória assumir um determinado valor não interfere em nada na chance de outra variável aleatória assumir um determinado valor. Já o termo identicamente distribuído se refere as variáveis aleatórias que são modeladas por uma mesma distribuição e mesmos parâmetros.

Vejamos o Exemplo, para o cálculo da esperança, variância e desvio padrão de uma distribuição Poisson.

Exemplo 6.4:

Considere o Exemplo 6.3 e apresente a esperança e o desvio padrão de X que representa o número de êxitos da coincidência das senhas selecionada com a dos usuários do banco. Inicialmente, calculamos a esperança matemática (média),

$$E[X] = np = 10.000 \times (10^9 / 36^6) = 4593,937 \text{ êxitos.}$$

Para o cálculo da variância, temos

$$Var[X] = 10.000 \times (10^9 / 36^6) \times (1 - 10^9 / 36^6) = 2483,511 \text{ êxitos}^2,$$

em que o desvio padrão segue,

$$\sigma_X = \sqrt{2483,511} = 49,83484 \text{ êxitos.}$$

Uma distribuição discreta amplamente aplicada em diversos tipos de experimentos, é a distribuição Poisson. Eventos do tipo, fenômenos que esperamos dado um tempo ou um espaço, como o número de telefonemas de uma atendente em uma hora, ou o número de acidentes por 15 km em uma rodovia, dentre outros, podem ser modelados por essa distribuição. Uma das ideias por trás da distribuição Poisson é que para pequenos intervalos de tempo/espaço a probabilidade de

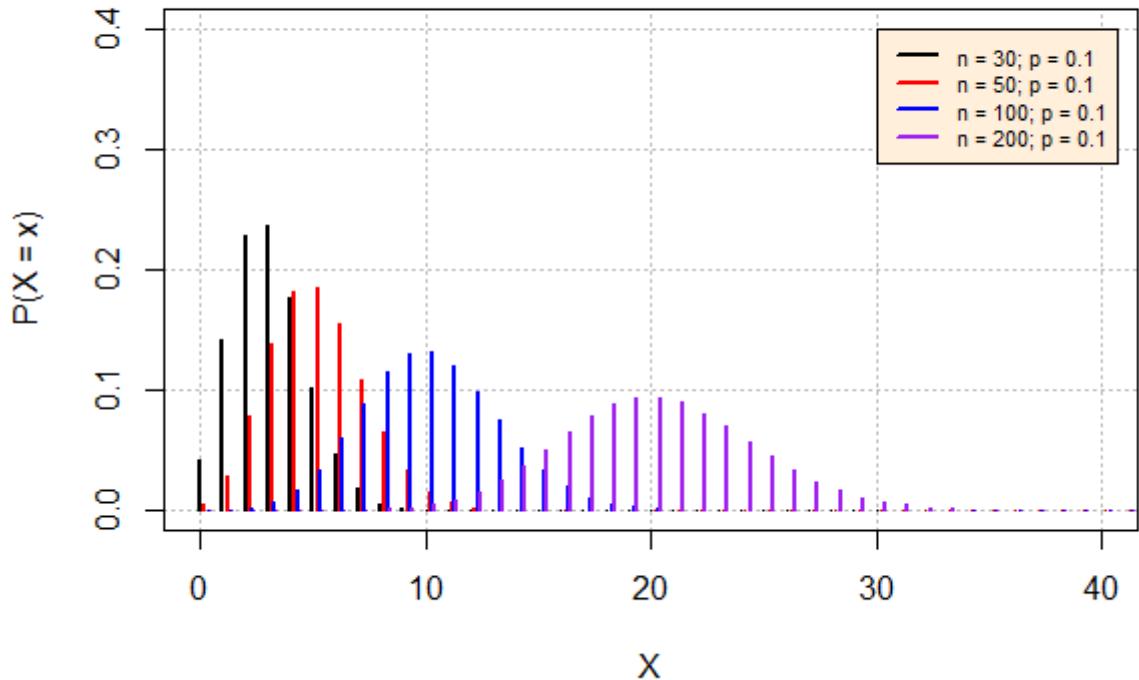


Figura 6.4: Gráfico da função de probabilidade da distribuição binomial, de parâmetros diferentes de n e $p = 0.1$.

ocorrência de uma determinada ocorrência é proporcional ao tempo/espaço de espera, isto é, podemos assumir que quanto mais longo for a espera ou o espaço, mais provável é a ocorrência. De outro modo, suponha que X seja uma variável aleatória que corresponde ao número de ocorrências (sucesso) em um determinado tempo/espaço T , e suponha que o número médio de ocorrências seja dado por λ . Assumimos também que a esperança de X é $E[X] = \lambda$. Ainda, considere ainda que o tempo/espaço pode ser subdividido em n subintervalos suficientemente pequenos $\Delta t = T/n$, de modo que a existência de mais de uma ocorrência em Δt é desprezível, e a probabilidade p será a mesma em cada subintervalo de modo independente da probabilidade de ocorrência em outros subintervalos.

Código R 6.3**Script:**

```

1  #-----
2  # Consideramos a alteracao do valor do parametro n
3  #-----
4  # Grafico da distribuicao binomial para n = 30 e p = 0.1
5  n <- 30; x <- 0:n; p <- 0.1
6  px <- dbinom(x, size = n, prob = p)
7  plot(x, px, type = "h", xlab = "X", ylab = expression("P(X=x)"), panel.first
      = grid(col = "gray"), ylim = c(0, 0.4), xlim = c(0, 40), lwd = 2)
8  # n = 30 e p = 0.1
9  n <- 50; x <- 0:n; p <- 0.1
10 px <- dbinom(x, size = n, prob = p)
11 lines(x + 0.2, px, col = "red", type = "h", lwd = 2)
12 # n = 100 e p = 0.1
13 n <- 100; x <- 0:n; p <- 0.1
14 px <- dbinom(x, size = n, prob = p)
15 lines(x + 0.3, px, col = "blue", type = "h", lwd = 2)
16 # n = 200 e p = 0.1
17 n <- 200; x <- 0:n; p <- 0.1
18 px <- dbinom(x, size = n, prob = p)
19 lines(x + 0.4, px, col = "purple", type = "h", lwd = 2)
20 # Legendas
21 legend(30, 0.4, legend = c("n=30;p=0.1",
22                           "n=50;p=0.1",
23                           "n=100;p=0.1",
24                           "n=200;p=0.1"),
25                           col = c("black", "red", "blue", "purple"),
26                           bg = "antiquewhite1", lty = 1, cex = 0.7, lwd = 2)

```

Assim, a distribuição de X tem distribuição Poisson, derivada de uma distribuição Binomial assintoticamente, isto é, para $n \rightarrow \infty$ e $p \rightarrow 0$, mantendo-se np constante. Para isso, temos

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

sendo $n!/[x!(n-x)!] = (n)_x/x!$, tal que $(n)_x = n(n-1)\dots(n-x+1)$, e definindo $\lambda = np$, temos

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

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

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

Considerando agora que a probabilidade $p \rightarrow 0$ com $n \rightarrow \infty$, λ passa a ser considerado como aproximadamente constante. Na aplicação do limite observe que

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

e por um resultado clássico de limite, segue também que

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

Desse modo, concluímos que

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

do qual essa distribuição é a função de probabilidade Poisson. Vejamos o Exemplo 6.5, para mostrar essa aproximação.

Exemplo 6.5:

Considere uma revisão tipográfica de um livro, e foi observado que existe um erro a cada 500 palavras impressas. Uma página apresenta 250 palavras. Desse modo, qual é a probabilidade de que não ocorram mais do que 1 erro a cada quatro páginas revisadas? Se considerarmos que as palavras impressas são experimentos de Bernoulli com probabilidade $p = 1/500$. Então X representa o número de erros que ocorre em quatro páginas, e $X \sim \text{Binomial}(250 \times 4, 1/500)$. Desse modo,

$$P(X \leq 1) = \binom{1000}{x} \left(\frac{1}{500}\right)^x \left(1 - \frac{1}{500}\right)^{1000-x} = 0,6766765.$$

Ao invés, podemos aproximar essa probabilidade usando a distribuição Poisson com $\lambda = np = 1000 \times (1/500) = 2$ erros a cada quatro páginas, que segue,

$$P(X \leq 1) = \frac{2^x e^2}{x!} = 0,6766764.$$

Observamos que a diferença ocorreu na sétima casa decimal, sendo uma aproximação muito boa, levando em consideração que o cálculo da última probabilidade foi relativamente mais simples. Vejamos na Figura 6.5, a representação gráfica dessas duas distribuições mostrando praticamente equivalência entre elas. Lembrando que esse resultado se torna tão melhor quanto mais $n \rightarrow \infty$ e $p \rightarrow 0$.

Formalmente, apresentamos a distribuição Poisson na Definição 6.3, a seguir.

Definição 6.3: Distribuição Poisson

Uma variável aleatória X discreta, tem distribuição Poisson , se sua função de probabilidade é dada por

$$P(X = x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!}, & \text{para } x = 0, 1, 2, \dots, \\ 0, & \text{caso contrário,} \end{cases} \quad (6.14)$$

em que $\lambda > 0$. Em notação, $X \sim \text{Poisson}(\lambda)$ representa que X tem distribuição Poisson com parâmetro λ .

Vejamos o Exemplo 6.6, para apresentar uma aplicação dessa distribuição.

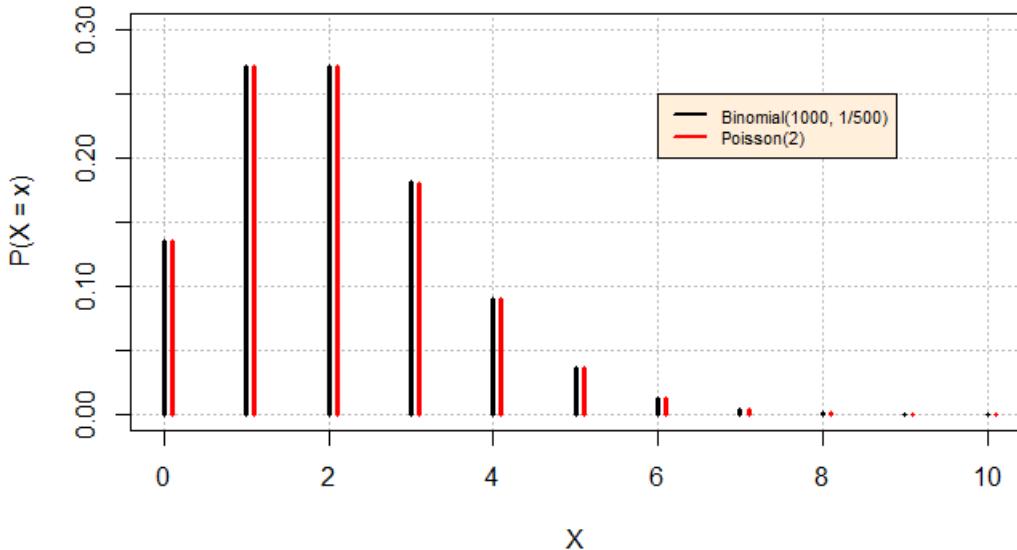


Figura 6.5: Relação entre as distribuições Poisson e Binomial para $n = 1000$, $p = 1/500$ e $\lambda = np$

Exemplo 6.6:

O número de falhas em parafusos de máquinas da indústria de papeis segue uma distribuição Poisson, em que ocorre 0,2 falhas por metro quadrado da produção.

- Qual a probabilidade de não ocorrer falha em um metro quadrado de papel?
- Qual a probabilidade de ocorrer no mínimo 1 erro por metro quadrado de papel?
- Qual a probabilidade de ocorrer 1 erro por $1,5 \text{ m}^2$?

Para o primeiro caso (a), temos,

$$P(X = 0) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{0,2^0 e^{-0,2}}{0!} = e^{-0,2} = 0,81873.$$

Usando o R, calculamos essa probabilidade usando `dpois(0, 0.2)`. Para o segundo caso (b), temos,

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0), \quad (\text{Evento complementar}) \\ &= 1 - 0,81873 \\ &= 0,18127, \end{aligned}$$

em que, usando o R calculamos essa probabilidade pelo código: `1 - dpois(0, 0.2)` ou `ppois(0, 0.2, lower.tail = FALSE)`. Por fim, o último ítem (c), tem uma peculiaridade porque a área com o qual se pergunta não é a mesma para o parâmetro λ . Assim, devemos fazer uma atualização dessa média espacial, isto é,

$$\left\{ \begin{array}{lcl} \lambda^* & \rightarrow & 1,5 \\ 0,2 & \rightarrow & 1 \end{array} \right. \therefore \lambda^* = 0,3 \text{ falhas por } 1,5 \text{ m}^2.$$

Concluímos que,

$$P(X = 1) = \frac{\lambda^{*x} e^{-\lambda^*}}{x!} = \frac{0,3^1 e^{-0,3}}{1!} = 0,22224.$$

Usando o R, calculamos essa probabilidade dessa forma: `dpois(1, 0.3)`.

Para verificar que a distribuição Poisson é uma função de probabilidade, apresentamos o Teorema 6.4, a seguir.

Teorema 6.5: Função de probabilidade Poisson

A distribuição Poisson expressa em (6.14), é uma função de probabilidade .

Prova:

Para provar que a distribuição de Poisson é uma função de probabilidade, temos que demonstrar os itens:

- Mostrar que $P(X = x) \geq 0$;
- Mostrar que $\sum_{x=0}^{\infty} P(X = x) = 1$,
- Sabe-se que $\lambda > 0$, $x = 0, 1, 2, \dots$, e a constante $e = 2,718281\dots$, então

- $e^{-\lambda} > 0$,
- $\lambda^x > 0$,
- $x! > 0$,

portanto, $P(X = x) \geq 0$.

- Tendo,

$$\begin{aligned}\sum_{x=0}^{\infty} P(X = x) &= \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!}, \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}, \\ &= e^{-\lambda} \left[\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^x}{x!} + \dots \right], \\ &= e^{-\lambda} \underbrace{\left[1 + \lambda + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^x}{x!} + \dots \right]}_{\text{série de Taylor}},\end{aligned}$$

e sabendo que $e^x = 1 + x + \frac{x^2}{2!} + \dots$, é a expressão da série de Taylor, portanto

$$\sum_{x=0}^{\infty} P(X = x) = e^{-\lambda} e^{\lambda} = e^0 = 1.$$

Assim, provamos que a distribuição Poisson é uma função de probabilidade.

Podemos apresentar uma representação gráfica da função de probabilidade Poisson, observando o comportamento do parâmetro λ . Na Figura 6.6, observamos que à medida que λ aumenta, mais simétrica a distribuição se torna como observado, e para o desenvolvimento desse gráfico usando o R, segue o Código

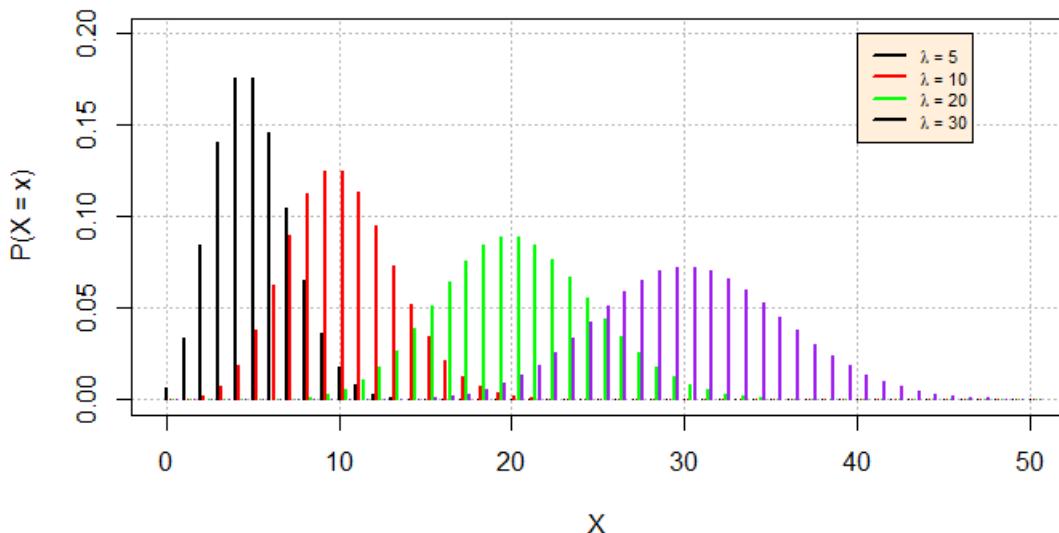


Figura 6.6: Comportamento da distribuição Poisson com relação ao parâmetro λ .

Código R 6.4

Script:

```

1  #-----
2  # Consideramos a alteracao do valor do parametro lambda
3  #-----
4  # Grafico da distribuicao poisson para lambda=5
5  x <- 0:50; lambda = 5
6  px <- dpois(x, lambda = lambda)
7  plot(x, px, type = "h", xlab = "X", ylab = expression("P(X=x)"), panel.first
      = grid(col="gray"), ylim = c(0, 0.2), xlim = c(0, 50), lwd = 2)
8  # n = 30 e lambda = 10
9  x <- 0:50; lambda = 10
10 px <- dpois(x, lambda = lambda)
11 lines(x+0.2, px, col = "red", type = "h", lwd = 2)
12 # n = 30 e lambda = 20
13 x <- 0:50; lambda = 20
14 px <- dpois(x, lambda = lambda)
15 lines(x+0.4, px, col = "green", type = "h", lwd = 2)
16 # n = 30 e lambda = 30
17 x <- 0:50; lambda = 30
18 px <- dpois(x, lambda = lambda)
19 lines(x+0.6, px, col = "purple", type = "h", lwd = 2)
20 # Legendas
21 legend(40, 0.2, legend = c(expression(lambda^"=5"),
22                           expression(lambda^"=10"),
23                           expression(lambda^"=20"),
24                           expression(lambda^"=30")),
25         col = c("black", "red", "green"),
26         bg = "antiquewhite1", lty = 1, cex = 0.7, lwd = 2)

```

Para completarmos a caracterização dessa distribuição, apresentamos o Teorema 6.6 para a espe-

rança, variância e desvio padrão da distribuição Poisson, a seguir. Vamos observar que a distribuição Poisson é um caso de distribuição superdispersos, pois mostraremos que a esperança e variância têm valores iguais a λ .

Teorema 6.6: Esperança, Variância e Desvio padrão da distribuição Poisson

Se a variável aleatória X possui distribuição Poisson expressa em (6.14), então a esperança de X é dada por:

$$\mu_X = E[X] = \lambda, \quad (6.15)$$

a variância é dada por:

$$\sigma_X^2 = Var[X] = \lambda, \quad (6.16)$$

e o desvio padrão é dado por:

$$\sigma_X = \sqrt{\lambda}. \quad (6.17)$$

Prova:

A dedução da esperança matemática pode ser apresentada da seguinte forma:

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!}, \\ &= 0 \times \frac{e^{-\lambda} \lambda^0}{0!} + \sum_{x=1}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!}. \end{aligned}$$

Fazendo $y = x - 1$, então,

$$\begin{aligned} E[X] &= \sum_{y=0}^{\infty} (y+1) \frac{e^{-\lambda} \lambda^{y+1}}{(y+1)!}, \\ &= \sum_{y=0}^{\infty} (y+1) \frac{e^{-\lambda} \lambda^{y+1}}{(y+1)y!}, \\ E[X] &= \sum_{y=0}^{\infty} \frac{e^{-\lambda} \lambda^y \lambda}{y!}, \\ &= \lambda \underbrace{\sum_{y=0}^{\infty} \frac{e^{-\lambda} \lambda^y}{y!}}_{=1}, \\ &= \lambda. \end{aligned}$$

A variância pode ser desenvolvida da forma a seguir.

$$\begin{aligned} E[X^2] &= \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!}, \\ &= 0^2 \times \frac{e^{-\lambda} \lambda^0}{0!} + \sum_{x=1}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!}, \\ &= \sum_{x=1}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!} \end{aligned}$$

Como $Var(X) = E[X^2] - E^2[X]$, então,

$$Var(X) = \sum_{x=1}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!} - \lambda^2,$$

fazendo $y = x - 1$, então,

$$\begin{aligned}
 Var(X) &= \sum_{y=0}^{\infty} (y+1)^2 \frac{e^{-\lambda} \lambda^{(y+1)}}{(y+1)!} - \lambda^2, \\
 &= \sum_{y=0}^{\infty} (y+1)^2 \frac{e^{-\lambda} \lambda^{(y+1)}}{(y+1)y!} - \lambda^2, \\
 &= \sum_{y=0}^{\infty} (y+1) \frac{e^{-\lambda} \lambda^{(y+1)}}{y!} - \lambda^2, \\
 &= \sum_{y=0}^{\infty} y \frac{e^{-\lambda} \lambda^y \lambda}{y!} + \sum_{y=0}^{\infty} \frac{e^{-\lambda} \lambda^y \lambda}{y!} - \lambda^2, \\
 &= \lambda \left[\underbrace{\sum_{y=0}^{\infty} y \frac{e^{-\lambda} \lambda^y}{y!}}_{=E[X]=\lambda} + \underbrace{\sum_{y=0}^{\infty} \frac{e^{-\lambda} \lambda^y}{y!}}_{=1} - \lambda \right], \\
 &= \lambda(\lambda + 1 - \lambda) = \lambda.
 \end{aligned}$$

o desvio padrão é dado por

$$\sigma_X = \sqrt{\lambda},$$

em que concluímos a prova.

Exemplo 6.7:

Considerando o Exemplo 6.6, podemos observar que $E[X] = 0,2$ falhas/ m^2 , a variância $\sigma_X^2 = 0,2(falhas/m^2)^2$ e desvio padrão $\sigma_X = \sqrt{0,2} = 0,44721$ falhas/ m^2 . Para o ítem (c) desse mesmo exemplo, como houve uma alteração espacial do valor médio de falhas, então a ideia do cálculo das medidas resumo será a mesma, assumindo apenas as medidas com o novo parâmetro, isto é, $\lambda^* = 0,3$ falhas/ m^2 .

O número de falhas em parafusos de máquinas da indústria de papéis segue uma distribuição Poisson, em que ocorre 0,2 falhas por metro quadrado da produção.

- a) Qual a probabilidade de não ocorrer falha em um metro quadrado de papel?
- b) Qual a probabilidade de ocorrer no mínimo 1 erro por metro quadrado de papel?
- c) Qual a probabilidade de ocorrer 1 erro por $1,5 m^2$?

Para o primeiro caso (a), temos,

$$P(X = 0) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{0,2^0 e^{-0,2}}{0!} = e^{-0,2} = 0,81873.$$

Usando o R, calculamos essa probabilidade usando `dpois(0, 0.2)`. Para o segundo caso (b), temos,

$$\begin{aligned}
 P(X \geq 1) &= 1 - P(X = 0), \quad (\text{Evento complementar}) \\
 &= 1 - 0,81873 \\
 &= 0,18127,
 \end{aligned}$$

em que, usando o R calculamos essa probabilidade pelo código: `1 - dpois(0, 0.2)` ou `ppois(0, 0.2, lower.tail = FALSE)`. Por fim, o último ítem (c), tem uma peculiaridade porque a área com

o qual se pergunta não é a mesma para o parâmetro λ . Assim, devemos fazer uma atualização dessa média espacial, isto é,

$$\begin{cases} \lambda^* & \rightarrow 1,5 \\ 0,2 & \rightarrow 1 \end{cases} \therefore \lambda^* = 0,3 \text{ falhas por } 1,5 \text{ m}^2.$$

Concluímos que,

$$P(X = 1) = \frac{\lambda^{*x} e^{-\lambda^*}}{x!} = \frac{0,3^1 e^{-0,3}}{1!} = 0,22224.$$

Usando o R, calculamos essa probabilidade dessa forma: `dpois(1, 0.3)`.

6.3 Distribuições contínuas de probabilidades

Iniciamos essa seção falando sobre um problema datado no século XVIII na ciência, que eram os problemas de medição entre os cientistas, na área da matemática e física, que para uma mesma situação, resultavam em valores diferentes por diversos fatores, como por exemplo, os instrumentos de medidas, variações do tempo, dentre outros. Por exemplo, nessa época era notório o problema existente em conciliar a Lei da Gravitação de Newton com a observação celeste. Assim, para resolver esse problema em 7 de julho de 1795, surge na França a teoria da medição, por meio de uma sanção de um sistema métrico de pesos e medidas, para compreender e quantificar essas variações encontradas na medidas pelos cientistas, que mais a frente, essas variações representam os erros aleatórios.

Apesar desses erros serem aleatórios, alguns cientistas, acreditavam que existiriam um padrão nesses erros independente do que fosse medido. Então, em 1733, Abraham de Moivre observando os números do triângulo de Pascal, Código R 6.5, para encontrar aproximação para os números que estão nas linhas das centenas de milhares, percebeu uma curva quando se assumia o coeficiente das linhas do triângulo como uma barra no gráfico de barras, Figura 6.7, se assemelhava a uma forma de sino, à medida que se vai plotando os coeficientes das linhas das centenas de milhares, também chamada de curva normal ou de Gauss. O termo normal deve ao fato que suavizando ainda mais essa curva, se chega a uma expressão matemática, que hoje conhecemos como a distribuição normal, e esta distribuição que está presente na maior parte da modelagem dos erros aleatórios nos sistemas de medidas. Esta distribuição também pode ser chamada de distribuição Gaussiana, devido a fundamentação teórica que o matemático, físico e astrônomo Francês Carl Friedrich Gauss conseguiu demonstrar.

Desse modo, apresentamos formalmente a distribuição normal na Definição 6.4, a seguir.

Definição 6.4: Distribuição Normal

Uma variável aleatória X contínua, tem distribuição normal se sua função densidade de probabilidade é dada por

$$f_X(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, & \text{para } -\infty < x < \infty, \\ 0, & \text{caso contrário,} \end{cases} \quad (6.18)$$

em que os parâmetros μ e σ^2 satisfazem $\mu \in \mathbb{R}$ e $\sigma^2 > 0$. Em notação, $X \sim N(\mu, \sigma^2)$ representa que X tem distribuição Normal com parâmetros μ e σ^2 .

Código R 6.5**Script:**

```

1 # Triangulo pascal
2
3 # Codigo para linhas acumuladas. Por exemplo, 'linhas = 10',
4 # imprime as linhas de 1 a 10 do trianbulo de Pascal
5 tpacum <- function(linhas) {
6   lapply(0:linhas, function(i) choose(i, 0:i))
7 }
8
9 # Codigo para imprimir uma determinada linha do triangulo
10 # de Pascal. Por exemplo, 'linha = 10', imprime apenas a
11 # linha 10
12 tpind <- function(linha) {
13   choose(linha, 0:linha)
14 }
15
16 # Gerando os graficos das linhas do triangulo de pascal
17 # como barras do gráfico de barras. Vamos gerar os graficos
18 n <- c(10, 25, 50, 100, 500, 1000)
19 par(mfrow = c(3, 2))
20 for(i in n) {
21   x <- tpind(i)
22   plot(x, type = "h", xlab = paste("Ordem_dos_coeficientes_da",
23                                     "linha", i), ylab = "Valor_do_coeficiente", panel.first =
24                                     grid(col="gray"), lwd = 3)
25 }
```

Algumas propriedade da densidade da normal podem ser facilmente observadas na Figura 6.8, que segue:

- i) $f_X(x)$ é simétrica em relação a μ ;
- ii) $f_X(x) \rightarrow 0$ quando $x \rightarrow \pm\infty$;
- iii) o valor máximo de x se dar para $x = \mu$.

Com relação aos parâmetros, percebemos que μ é um parâmetro de escala, à medida que se altera esse parâmetro, a distribuição se translada em torno do eixo X , Figura 6.9. Já a alteração do parâmetro σ^2 , proporciona uma alteração na forma da distribuição, Figura 6.10. Por isso, que dizemos que σ^2 é um parâmetro de forma. Uma outra característica interessante na distribuição normal apresenta alguns resultados úteis, tais como:

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= 0,6827 \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &= 0,9545 \\ P(\mu - 3\sigma < X < \mu + 3\sigma) &= 0,9973, \end{aligned}$$

e que podem ser verificados na Figura 6.11.

Para complementar essas características, precisamos demonstração que a distribuição normal é uma função densidade de probabilidade, pelo Teorema 6.7, a seguir.

Teorema 6.7: Função densidade de probabilidade Normal

A distribuição Normal expressa em (6.18), é uma função densidade de probabilidade .

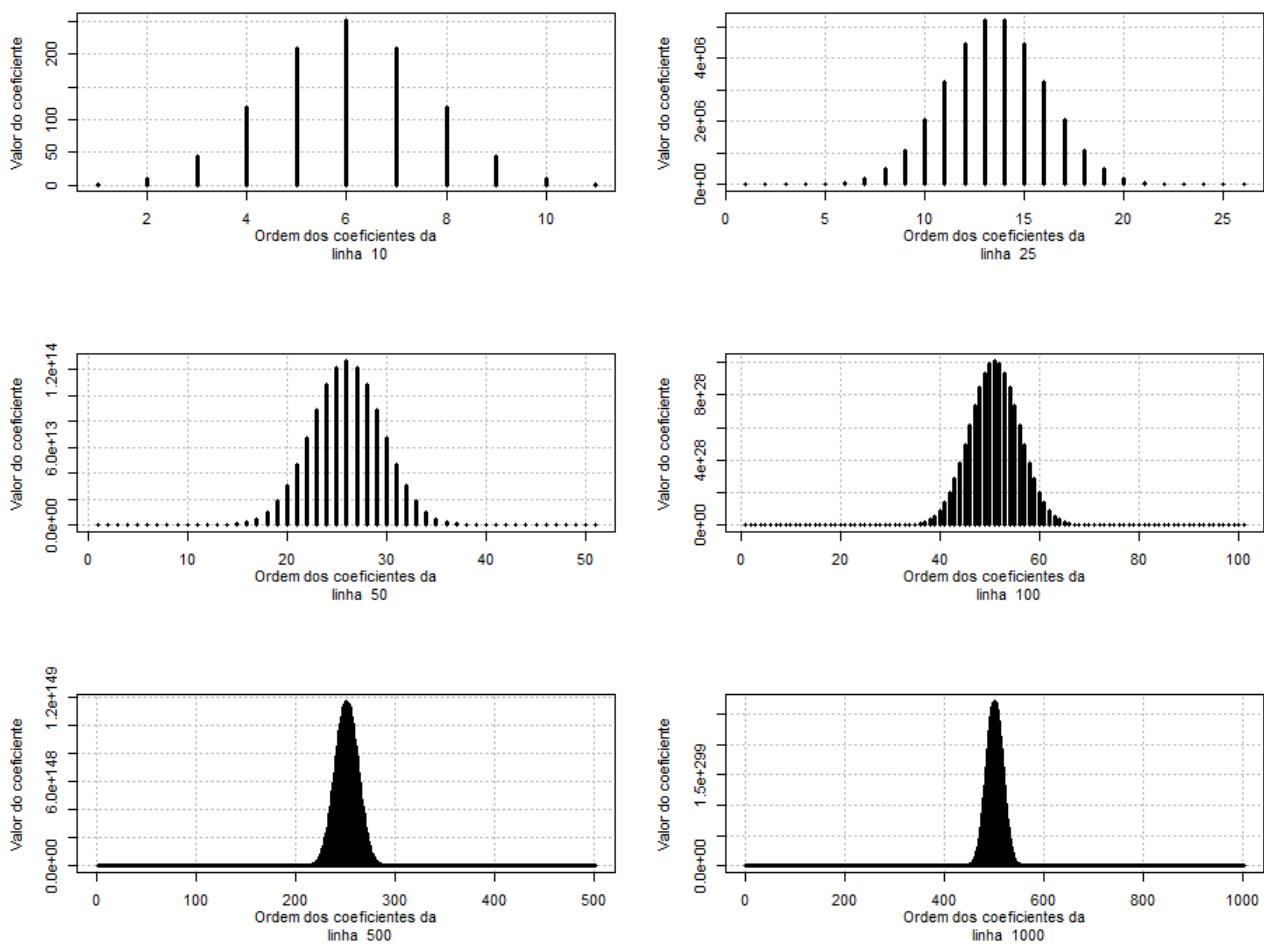


Figura 6.7: Gráfico de barras dos coeficientes para determinadas linhas do triângulo Pascal.

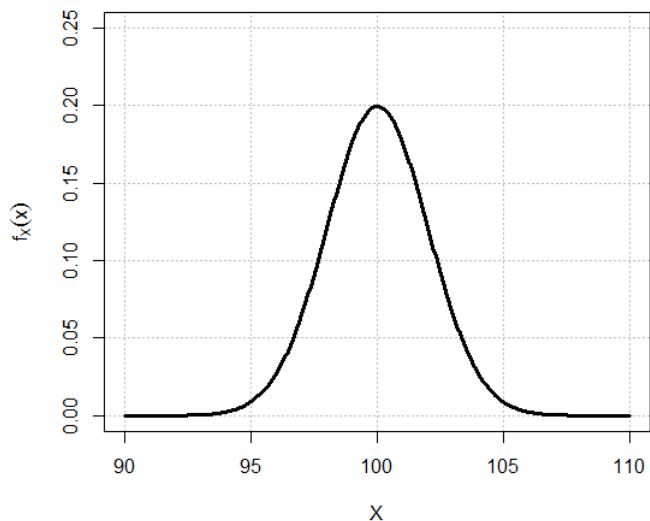


Figura 6.8: Função densidade da distribuição normal, tal que $X \sim N(100, 4)$.

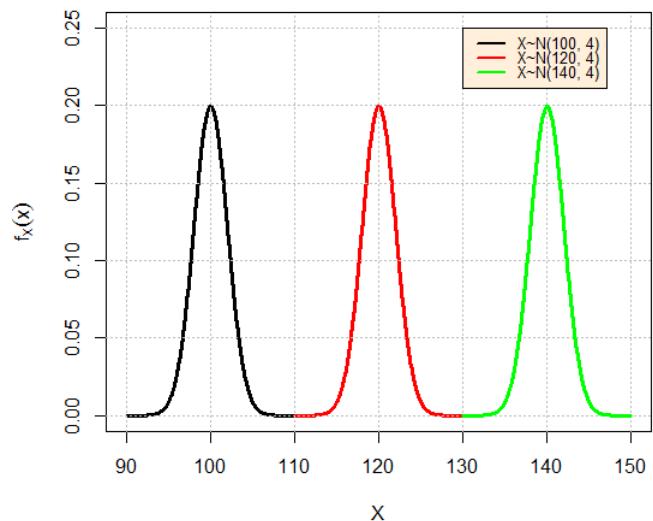


Figura 6.9: Função densidade da distribuição normal, para três variações do parâmetro μ .

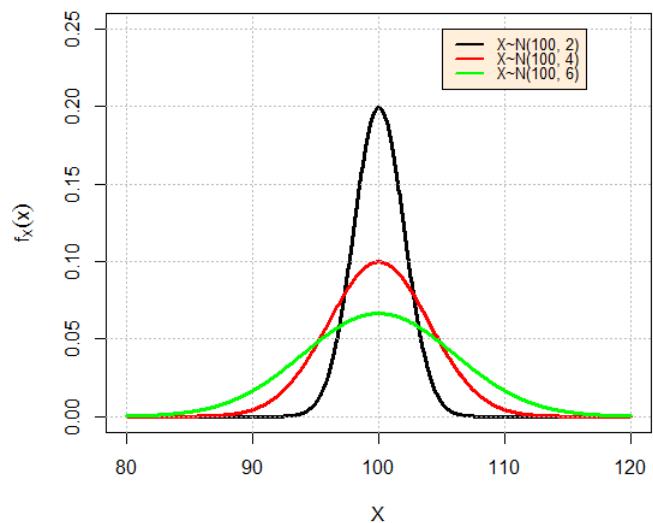


Figura 6.10: Função densidade da distribuição normal, para três variações do parâmetro σ^2 .

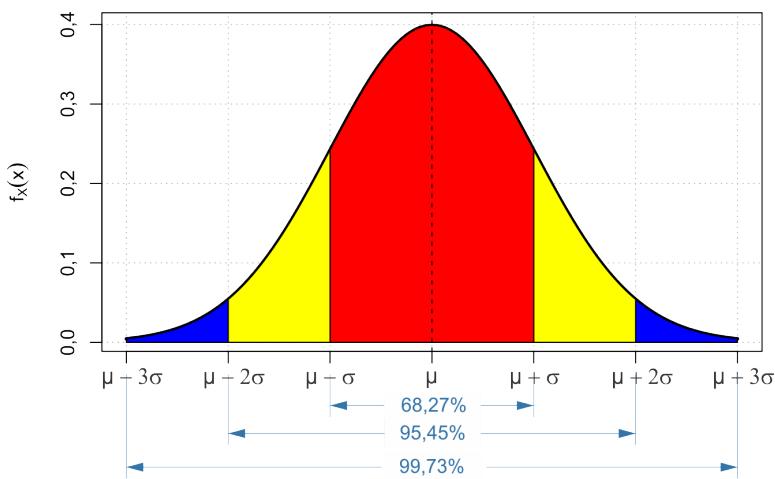


Figura 6.11: Probabilidade da distribuição normal sob três desvios padrões da média populacional.

Prova:

Para verificar se $f(x)$ é uma função densidade de probabilidade, temos que mostrar que:

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

Prova:

- Vamos fazer as seguintes observações:

- a constante $\frac{1}{\sqrt{2\pi\sigma^2}} \geq 0$;
- todo valor e elevado a um valor negativo, sempre retorna um valor positivo.

Dessa forma, $f(x) \geq 0$.

- Para mostrar que $A = \int_{-\infty}^{\infty} f(x)dx = 1$, vamos partir do seguinte raciocínio. Se $A = 1$, então $A^2 = 1$. Assim,

$$\begin{aligned} A^2 &= A \times A, \\ &= \int_{-\infty}^{\infty} f(x)dx \times \int_{-\infty}^{\infty} f(x)dx, \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \times \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx. \end{aligned}$$

Fazendo uma mudança de variável, $y = \frac{(x-\mu)}{\sigma}$ e $dy = \frac{1}{\sigma}dx$, então

$$A^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy.$$

Mudando uma variável y de uma das densidades, simplesmente para diferenciá-las, chamando-a de w , temos

$$\begin{aligned} A^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} dw, \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2\pi}} e^{-\frac{1}{2}(y^2+w^2)} dy dw. \end{aligned}$$

Usando a técnica do cálculo da integral de coordenadas polares (Ver **Apêndice C**), temos

$$\begin{aligned}y &= r \sin \theta \\w &= w \cos \theta.\end{aligned}$$

Fazendo,

$$\begin{aligned}y^2 + w^2 &= r^2 \sin^2 \theta + r^2 \cos^2 \theta, \\&= r^2 (\sin^2 \theta + \cos^2 \theta) = r^2,\end{aligned}$$

considerando $dA = rdrd\theta$, e que o intervalos das variáveis r sendo $[0, \infty]$ e θ sendo $[0, 2\pi]$, então A^2 pode ser expresso da seguinte forma

$$\begin{aligned}A^2 &= \int_0^\infty \int_0^{2\pi} \frac{1}{\sqrt{2\pi}\sqrt{2\pi}} e^{-\frac{1}{2}r^2} d\theta r dr, \\&= \int_0^\infty \int_0^{2\pi} \frac{1}{2\pi} e^{-\frac{1}{2}r^2} d\theta r dr, \\&= \frac{1}{2\pi} \int_0^\infty e^{-\frac{1}{2}r^2} r [2\pi] dr, \\&= \frac{1}{2\pi} \int_0^\infty e^{-\frac{1}{2}r^2} r(2\pi) dr - \underbrace{\frac{1}{2\pi} \int_0^\infty e^{-\frac{1}{2}r^2} r(0) dr}_{=0}, \\&= \int_0^\infty e^{-\frac{1}{2}r^2} r dr.\end{aligned}$$

Faremos um artifício de multiplicar $1 = (-1) \times (-1)$ na expressão, que não alterará o resultado, ou seja,

$$\begin{aligned}A^2 &= 1 \times \int_0^\infty e^{-\frac{1}{2}r^2} r dr, \\&= (-1) \times (-1) \times \int_0^\infty e^{-\frac{1}{2}r^2} r dr, \\&= (-1) \times \int_0^\infty (-1)e^{-\frac{1}{2}r^2} r dr.\end{aligned}$$

Usando mais uma mudança de variável $u = -(1/2)r^2$ e $du = -rdr$, podemos concluir

$$\begin{aligned}A^2 &= (-1) \times \int_0^\infty e^u du, \\&= -[e^u]_0^\infty, \\&= -[0 - 1] = 1.\end{aligned} \tag{6.19}$$

Se $A^2 = 1$, então $A = 1$. Assim, fica concluído que a distribuição normal é uma função densidade de probabilidade.

Percebemos que analiticamente o cálculo da probabilidade de X , tendo X uma distribuição normal, está em uma região da distribuição não é nada fácil. Para isso, usamos valores tabelados da probabilidade da distribuição normal padrão, que é um caso particular da distribuição normal, apresentados no **Apêndice B**, e a definição dessa distribuição é apresentada na Definição 6.5.

Definição 6.5: Distribuição Normal Padrão

Uma variável aleatória X contínua, tem distribuição normal padrão se sua função densidade de probabilidade é dada por

$$f_X(x) = \phi_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, & \text{para } -\infty < x < \infty, \\ 0, & \text{caso contrário,} \end{cases} \quad (6.20)$$

em que a expressão 6.20 é igual a expressão 6.18, considerando os parâmetros $\mu = 0$ e $\sigma^2 = 1$. Em notação, $X \sim N(0, 1)$ representa que X tem distribuição Normal Padrão.

Como consequência, se provamos que uma distribuição normal é uma função densidade de probabilidade para qualquer valor assumidos pelos parâmetros, pelo Teorema 6.7, segue que a distribuição normal padrão também é uma função densidade de probabilidade. A função de distribuição de uma normal padrão, denotamos por $\Phi(x)$, tal que,

$$\Phi_X(x) = \int_{-\infty}^x \phi_X(t) dt. \quad (6.21)$$

Para realizar a transformação de uma variável com distribuição normal para uma variável aleatória com distribuição normal padrão, apresentamos o Teorema 6.8.

Teorema 6.8: Transformação de $X \sim N(\mu, \sigma^2)$ para $Z \sim N(0, 1)$

Considerando $X \sim N(\mu, \sigma^2)$, então a transformação $Z = (X - \mu)/\sigma$ tem distribuição normal padrão, isto é, segue $Z \sim N(0, 1)$.

Para a prova do Teorema 6.8, precisamos de algumas informações adicionais, como o Jacobiano da transformação, que para esse momento se torna um assunto mais complexo. Contudo, iremos abordar assuntos como esse, na seção Aprofundamento do presente capítulo. Para tanto, vamos tomar como verdade que a transformação Z terá distribuição normal e que a esperança e variância são, respectivamente,

$$\mu_Z = E[Z] = E\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma}E[X - \mu] = \frac{E[X] - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0, \quad (6.22)$$

e

$$\sigma_Z^2 = Var[Z] = Var\left[\frac{X - \mu}{\sigma}\right] = \frac{Var[X - \mu]}{\sigma^2} = \frac{Var[X]}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1. \quad (6.23)$$

Esses resultados são consequências das propriedades da esperança e variância, apresentadas na seção 5.11 do Capítulo 5, bem como o seguinte Teorema 6.9, apresentado a seguir.

Teorema 6.9: Esperança, Variância e Desvio padrão da distribuição Normal

Se a variável aleatória X possui distribuição Normal expressa em (6.18), então a esperança de X é dada por:

$$\mu_X = E[X] = \mu, \quad (6.24)$$

a variância é dada por:

$$\sigma_X^2 = Var[X] = \sigma^2, \quad (6.25)$$

e o desvio padrão é dado por:

$$\sigma_X = \sigma. \quad (6.26)$$

Prova:

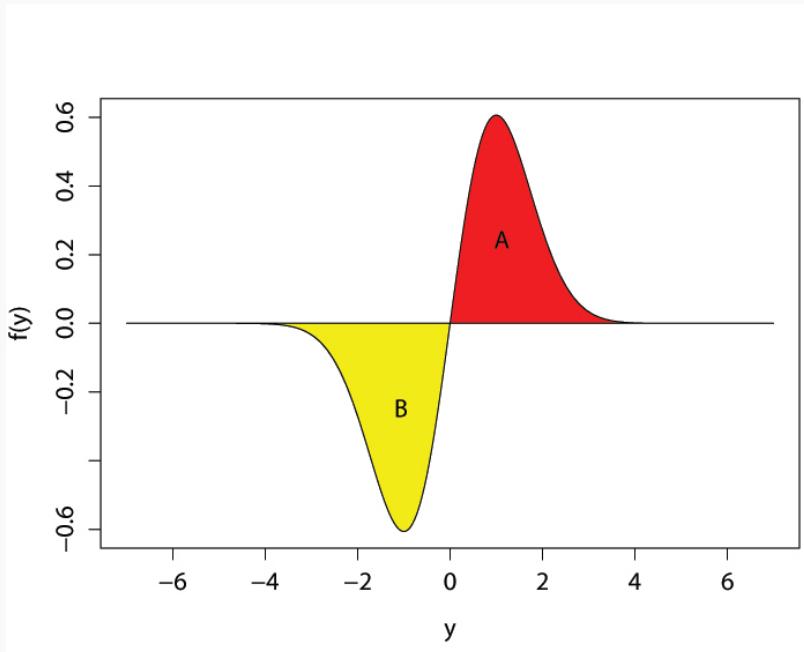
A esperança pode ser deduzida da seguinte forma:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xf(x)dx, \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx, \end{aligned}$$

fazendo uma mudança de variável, $y = (x - \mu)/\mu \Rightarrow x = y\sigma + \mu$ e $du = (1/\sigma)dx \Rightarrow dx = \sigma dy$, então

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} \left(\frac{y\sigma + \mu}{\sigma\sqrt{2\pi}} \right) e^{-\frac{1}{2}y^2} \sigma dy, \\ &= \int_{-\infty}^{\infty} \left(\frac{y\sigma + \mu}{\sqrt{2\pi}} \right) e^{-\frac{1}{2}y^2} dy, \\ &= \int_{-\infty}^{\infty} \left(\frac{y\sigma}{\sqrt{2\pi}} \right) e^{-\frac{1}{2}y^2} dy + \int_{-\infty}^{\infty} \left(\frac{\mu}{\sqrt{2\pi}} \right) e^{-\frac{1}{2}y^2} dy, \\ &= \frac{\sigma}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} ye^{-\frac{1}{2}y^2} dy}_{=0} + \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy}_{=1}, \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ye^{-\frac{1}{2}y^2} dy + \mu. \end{aligned}$$

Algo interessante acontece com a integral $\int_{-\infty}^{\infty} ye^{-\frac{1}{2}y^2} dy$. Observe graficamente a seguir. Como $A = B$, a integral se anula por simetria. Assim,



$$\begin{aligned} E[X] &= \frac{\sigma}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} ye^{-\frac{1}{2}y^2} dy}_{=0} + \mu, \\ &= \mu. \end{aligned}$$

A variância será deduzida a seguir.

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx,$$

usando a mudança de variável $y = (x - \mu)/\sigma \Rightarrow x = \mu + \sigma y$ e $dy = (1/\sigma)dx \Rightarrow dx = \sigma dy$, então

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} \frac{(\mu + \sigma y)^2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} \sigma dy, \\ &= \int_{-\infty}^{\infty} \frac{\mu^2 + 2\mu\sigma y + \sigma^2 y^2}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy, \\ &= \int_{-\infty}^{\infty} \frac{\mu^2}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy + \int_{-\infty}^{\infty} \frac{2\mu\sigma y}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy + \int_{-\infty}^{\infty} \frac{\sigma^2 y^2}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy, \\ &= \underbrace{\mu^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy}_{=1} + 2\mu\sigma \underbrace{\int_{-\infty}^{\infty} \frac{y}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy}_{=0} + \underbrace{\sigma^2 \int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}(y)^2} dy}. \end{aligned}$$

A primeira integral da expressão de $E[X^2]$ é igual a 1 como provado pela expressão (6.19), uma distribuição normal com parâmetros $\mu = 0$ e $\sigma = 1$. A segunda integral foi mostrado pela figura anterior que a integral se anula por simetria. Agora a terceira integral pode ser observado o gráfico abaixo, mostrando que é uma função par (simétrica no eixo y), $A = B$, ou seja,

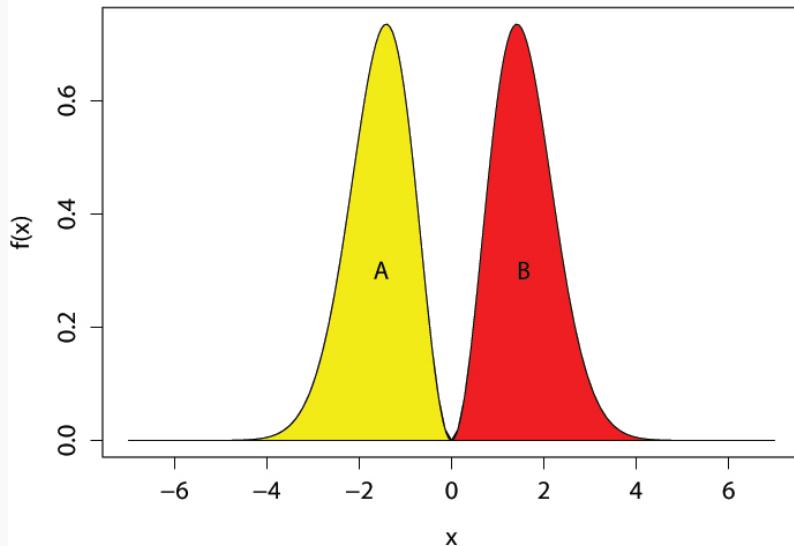
$$\text{Área} = A + B = A + A = 2A,$$

e portanto,

$$\int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}y^2} dy = 2 \int_0^{\infty} y^2 e^{-\frac{1}{2}y^2} dy.$$

Assim,

$$E[X^2] = \mu^2 + 2 \int_0^{\infty} \frac{\sigma^2}{\sqrt{2\pi}} y^2 e^{-\frac{1}{2}y^2} dy,$$



fazendo uma mudança de variável $z = y^2/2 \Rightarrow y = \sqrt{z}$ e $dz = ydy \Rightarrow dy = dz/y$, temos

$$\begin{aligned} E[X^2] &= \mu^2 + 2\sigma^2 \int_0^\infty \frac{(\sqrt{2z})^2}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{2z})^2} \frac{dz}{\sqrt{dz}}, \\ E[X^2] &= \mu^2 + 2\sigma^2 \int_0^\infty \frac{2z}{\sqrt{2\pi}} e^{-z} \frac{dz}{\sqrt{2z}}, \end{aligned}$$

sabendo que $2 = \sqrt{2} \times \sqrt{2}$ e $z = \sqrt{z} \times \sqrt{z}$, então

$$\begin{aligned} E[X^2] &= \mu^2 + 2\sigma^2 \int_0^\infty \frac{\sqrt{2} \times \sqrt{2} \times \sqrt{z} \times \sqrt{z}}{\sqrt{2} \times \sqrt{\pi}} e^{-z} \frac{dz}{\sqrt{2} \times \sqrt{z}}, \\ &= \mu^2 + 2\sigma^2 \int_0^\infty \frac{\sqrt{z}}{\sqrt{\pi}} e^{-z} dz, \\ &= \mu^2 + \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty z^{1/2} e^{-z} dz, \end{aligned}$$

A função gama é definida por:

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx, \quad t > 0.$$

Reescrevendo $1/2 = 3/2 - 1$, então

$$\begin{aligned} E[X^2] &= \mu^2 + \frac{2\sigma^2}{\sqrt{\pi}} \underbrace{\int_0^\infty z^{3/2-1} e^{-z} dz}_{=\Gamma(3/2)}, \\ &= \mu^2 + \frac{2\sigma^2}{\sqrt{\pi}} \Gamma(3/2). \end{aligned}$$

Uma outra propriedade da função gama é $\Gamma(t+1) = t\Gamma(t)$, e um caso particular dessa propriedade é que $\Gamma(1/2) = 2\Gamma(3/2) = \sqrt{\pi}$. Assim,

$$\begin{aligned} E[X^2] &= \mu^2 + \frac{2\sigma^2}{\sqrt{\pi}} \left(\frac{\sqrt{\pi}}{2} \right), \\ &= \mu^2 + \sigma^2. \end{aligned}$$

Como a $Var(X) = E[X^2] - E^2[X]$, então

$$Var(X) = [\mu^2 + \sigma^2] - \mu^2 = \sigma^2.$$

O desvio padrão é dado por:

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

o que conclui a prova.

Com essas informações, vamos iniciar verificar como podemos calcular a probabilidade uma variável aleatória normal assumir em um determinado da reta real usando os valores da probabilidade de uma distribuição normal padrão. Por exemplo, considerando $X \sim N(100, 4)$, podemos calcular a probabilidade $P(X > 102)$ usando a transformação de X pela variável aleatória Z , da seguinte

forma:

$$P(X > 102) = P\left(\frac{X - \mu}{\sigma} > \frac{102 - 100}{2}\right) = P(Z > 1).$$

Pelo **Apêndice B**, usando a tabela devida podemos chegar ao resultado $P(X > 102) = P(Z > 1) = 0,1586553$, como pode ser complementado pela Figura 6.12.

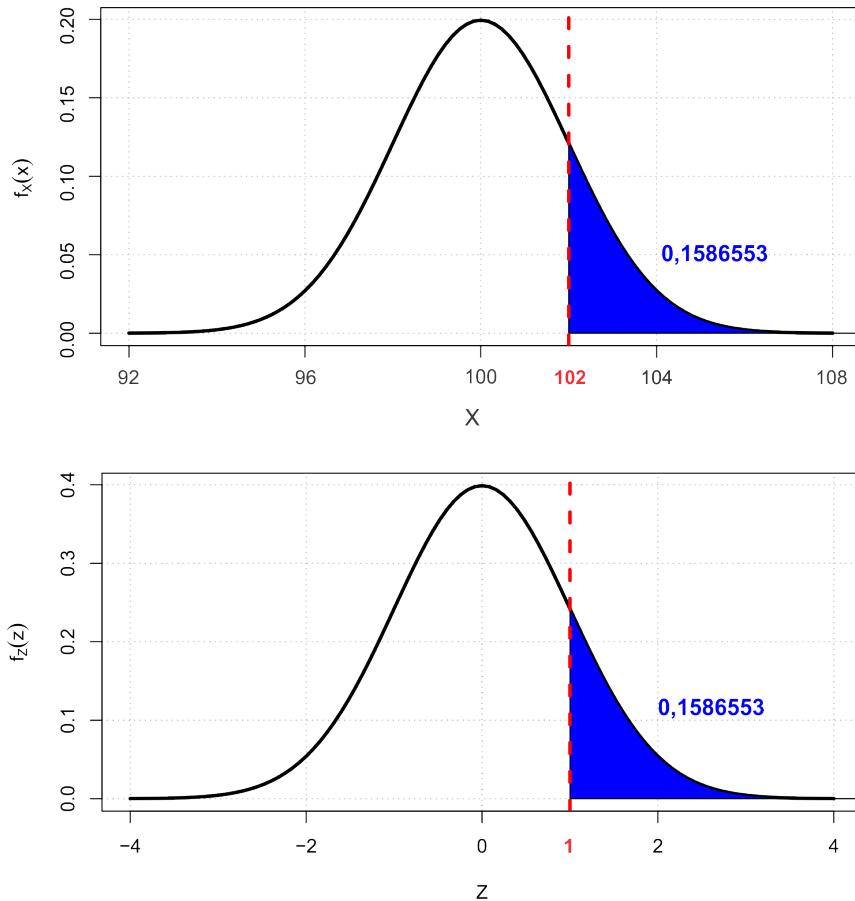


Figura 6.12: Gráfico da equivalência do cálculo das probabilidades entre a distribuição normal e normal padrão.

Dissemos anteriormente que devemos usar as tabelas devidas da distribuição normal padrão, uma vez que apresentamos três tipos de tabelas no **Apêndice B**, a primeira tabela desse apêndice representa o cálculo das probabilidades para $P(Z > z)$, a segunda tabela para $P(0 < Z < z)$, e a terceira tabela para $P(Z < z)$. Antes de apresentarmos como determinar as probabilidades na tabela da normal padrão, devemos observar que o fato da distribuição normal ser simétrica em torno de μ , então $P(X > \mu) = P(Z > 0) = 1/2 = P(X < \mu) = P(Z < 0)$. Outra coisa interessante é que $P(Z < z) = P(Z < 0) + P(0 < Z < z) = 0,5000 + P(0 < Z < z)$.

Desse modo, podemos determinar a probabilidade $P(X > 102) = P(Z > 1)$, usando a tabela da normal padrão para $P(Z > z)$. Nesse caso, usamos a forma de acessar a tabela para o cálculo da probabilidade devida como:

Tabela para $P(Z > z)$			
Z	...	0,00	
:		:	
1,0	...	0,15866	

Desse modo, a primeira coluna representa os dois primeiros dígitos para o valor de Z , e a primeira linha representa o último dígito. No caso, como a probabilidade desejada foi $P(Z > 1,00)$,

então na primeira coluna observamos o valor 1,0, e na primeira linha o valor 0,00. Após isso, cruzamos esses valores para o encontro desejado da probabilidade, que é 0,1586553. Outro exemplo, se desejarmos calcular a probabilidade $P(Z > 1,15)$, observando nessa mesma tabela, veremos que:

Tabela para $P(Z > z)$

Z	...	0,15	
:		:	
1,0	...	0,12507	

o valor é 0,1250719.

Vejamos o Exemplo 6.8, para ilustrar uma aplicação da distribuição normal, a seguir.

Exemplo 6.8:

Considere X uma variável aleatória que representa o ruído na detecção de um sinal digital, medido em *volt*s, tal que $X \sim N(0; 0,66^2)$, então sabendo que o sistema verifica que um sinal digital foi transmitido se a voltagem exceder 0,95, qual é a probabilidade de se detectar um sinal digital sem informação alguma, isto é, apenas o ruído?

Para isso, calculamos a probabilidade $P(X > 0,95)$, que segue,

$$P(X > 0,95) = P\left(\frac{X - \mu}{\sigma} > \frac{0,95 - 0}{0,66}\right) = P(Z > 1,44).$$

Consultando a tabela da normal padrão para $P(Z > z)$, temos,

Tabela para $P(Z > z)$

Z	...	0,04	
:		:	
1,4	...	0,07493	

Logo, a probabilidade $P(Z > 1,44) = 0,07493$. De modo equivalente, poderíamos utilizar a tabela da normal padrão para $P(0 < Z < z)$ e também calcular essa mesma probabilidade desejada, da seguinte forma:

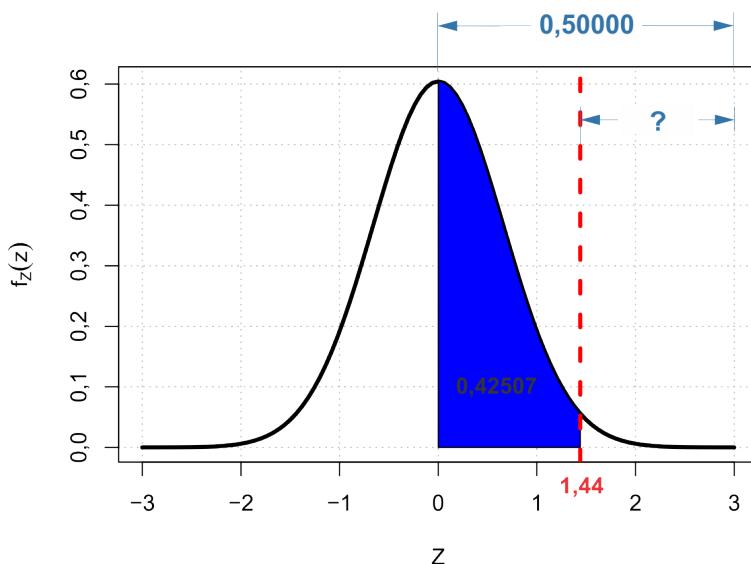
Tabela para $P(0 < Z < z)$

Z	...	0,04	
:		:	
1,4	...	0,42507	

Logo,

$$\begin{aligned} P(Z > 1,44) &= P(Z > 0) - P(0 < Z < 1,44) \\ &= 0,50000 - 0,42507 \\ &= 0,07493, \end{aligned}$$

como pode ser observado no gráfico abaixo.



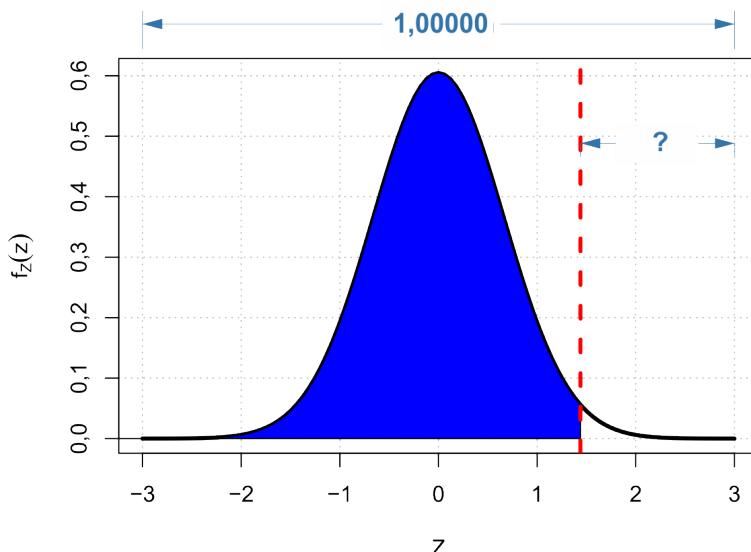
Por fim, para obter o mesmo resultado, podemos utilizar a tabela da normal padrão para $P(Z < z)$, da seguinte forma:

Tabela para $P(Z < z)$		
Z	...	0,04
:		:
1,4	...	0,92507

Logo,

$$\begin{aligned}
 P(Z > 1,44) &= 1 - P(Z < 1,44) \\
 &= 1 - 0,92507 \\
 &= 0,07493,
 \end{aligned}$$

como pode ser observado no gráfico abaixo.



6.4 Exercícios

Exercício 6.1: Defina variável aleatória (V.A.), variável aleatória discreta (V.A.D.), variável aleatória contínua (V.A.C.) e descreva diversos exemplos na sua área de estudo.

Solução na página 225

Exercício 6.2: Suponha que 2% dos itens produzidos por uma fábrica sejam defeituosos. Encontre a probabilidade P de existirem 3 defeituosos em uma amostra de 100.

Solução na página 225

Exercício 6.3: Usando a curva normal padronizada, determinar as áreas subtendidas entre os valores abaixo, com representação gráfica.

1. 0,35 e 0,0;
2. 0,0 e 1,52;
3. -0,34 e 1,9;
4. à direita de -1,91;
5. à esquerda de 1,13;
6. à esquerda de -2,13.

Solução na página 225

Exercício 6.4: Dada uma distribuição normal com $\mu = 40$ e $\sigma = 6$, calcular:

1. $P(X \leq 33)$;
2. $P(X \geq 29)$;
3. $P(39 \leq X \leq 45)$;
4. Ponto que tem 58% de área acima dele;
5. Ponto que tem 5% de área acima;
6. $P(Z > 0)$.

Solução na página 226

Exercício 6.5: Em um exame vestibular de matemática as notas distribuíram-se normalmente com média 6 o desvio padrão 1,5. Calcular o número de aprovados entre os 120 candidatos, sabendo-se que a nota mínima de aprovação é 5.

Solução na página 227

Exercício 6.6: Uma máquina de empacotar determinado produto apresenta variação de peso com desvio padrão de 20g. Em quanto deve ser regulado o peso médio do pacote para que apenas 10% tenham menos de 400g. Super distribuição normal dos pesos dos pacotes.

Solução na página 227

Exercício 6.7: Na 1^a prova de Estatística a média foi 4,5 e o desvio padrão 2,3. Considerando o método científico de aprovação:

- Conceito A: nota média $\geq \mu + \sigma$;
 - Conceito B: $\mu \leq$ média $\leq \mu + \sigma$;
 - Conceito C: $\mu - \sigma \leq$ média $\leq \mu$;
 - Conceito D: média $\leq \mu - \sigma$;
- Quantos alunos receberam cada um dos conceitos?
 - Quantos alunos foram aprovados (Conceito A, B e C);
 - Quantos alunos foram aprovados com distinção (média $\geq \mu + 2\sigma$);
 - Considere o método comum de aprovação ($X \geq 5,0$), quantos foram reprovados.

OBSERVAÇÃO: 100 alunos fizeram essa prova.

Solução na página 227

Exercício 6.8: Ao investir num determinado negócio, um cidadão pode ter um lucro anual de US\$ 60.000 com probabilidade 0,3 ou tomar um prejuízo de US\$ 20.000 com probabilidade 0,7. Qual é a sua esperança matemática?

Solução na página 228

Exercício 6.9: Considere que existem defeitos aleatórios na superfície de um *chip* semicondutor, e o fabricante informa que 5% de sua produção apresenta defeito. Desse modo, em uma amostra de 35 *chips*, qual a probabilidade de encontrarmos:

- nenhum defeito;
- apenas um defeito;
- não mais que um defeito;
- acima de dois defeitos.

Solução na página 228

Exercício 6.10: Considere uma avaliação de câmeras de uma determinada marca de celular, e que 85% dessas câmeras passaram no teste de avaliação, de modo que os celulares foram avaliados de modos independentes. Qual o tamanho da amostra necessário para que a probabilidade de no mínimo uma câmera não tenha passado no teste seja no mínimo de 90%?

Solução na página 228

Exercício 6.11: Considere que o número de alterações em uma página *web* de entretenimentos seja modelada por uma distribuição de Poisson, e que ocorre em média 0,30 alterações por dia.

- Qual a probabilidade de que não haja alterações em um dia?
- Qual a probabilidade de ocorra mais do que uma alteração em 8 horas?

- c) Qual a probabilidade de ocorrer no máximo duas alterações em 2 dias?
- d) Qual o valor esperado dessas alterações por dia? E a variabilidade dessas alterações em torno de 0,30 alterações por dia?

Solução na página 229

Exercício 6.12: A probabilidade de um consumidor responder ao questionário de grau de satisfação em um site de compras, após a finalização da compra é de 5%. Considerando que 100 consumidores apresentam comportamentos independentes quanto ao interesse pelas compras nesse site, determine:

- a) nenhum consumidor responder ao questionário;
- b) mais de cinco consumidores responderem ao questionário;
- c) exatamente 15 pessoas responderem o questionário.
- d) a esperança de X ;
- e) a variância e o desvio padrão de X ;

Solução na página 230

Exercício 6.13: Considere o sistema de segurança de acesso a conta de usuários de um determinado banco, por meio de uma senha de 6 dígitos, com 26 caracteres (a-z) ou números (0-9). Um *hacker* tentou invadir o sistema, e com informações privilegiadas percebeu algumas informações sobre as senhas de clientes:

- (I) Alguns clientes da carteira do banco apresentavam senhas com cinco letras e um número;
- (II) Alguns clientes da carteira do banco apresentavam senhas com quatro letras seguidas por dois números.

Com isso, indagamos:

- a) foi selecionado ao acaso 10 clientes, qual a probabilidade de 2 clientes apresentarem senhas do tipo (I)?
- b) foi selecionado ao acaso 15 clientes, qual a probabilidade de nenhum cliente apresentar senha do tipo (II)?

Solução na página 230

Amostragem (Em desenvolvimento!)

7.1 Introdução

A amostragem já é utilizada no nosso dia a dia inconscientemente. Quando preparamos uma refeição e vamos prová-la antes de servi-la, estamos na realidade fazendo uma amostragem, ao passo que a refeição representa o todo (população) e a parte retirada para provar se a refeição estava pronta para servi-la é a amostra. Dessa forma, queremos, por meio de uma amostra, saber algumas características (parâmetro) sobre a população. As medidas obtidas da amostra, com as quais nos indicam alguma característica da população, chamamos de estimador. Com base nisso, para obtermos bons estimadores, precisamos garantir antecipadamente que a amostra coletada da população represente a população, de modo a preservar as características relevantes.

Porém, o uso inadequado dos métodos de amostragem podem nos levar a equívocos como os encontrados por Lohr (2019) no livro de Hite (1987), do qual ela apresentou os seguintes resultados:

- 84% das mulheres não estão satisfeitas emocionalmente com seus relacionamentos.
- 70% de todas as mulheres casadas há cinco ou mais anos fazem sexo fora de seus casamentos.
- 95% das mulheres relatam formas de assédio emocional e psicológico de homens com quem mantêm relacionamentos amorosos.
- 84% das mulheres relatam formas de condescendência dos homens em seus relacionamentos amorosos.

Dessa forma, para que não cometamos viéses nos resultados como os encontrados por Hite (1987), precisamos entender os princípios básicos da Teoria de Amostragem. Algo interessante, ocorre nos livros de Estatística básica que apresentam de forma muito introdutória essa teoria. Nos livros sobre inferência toda a teoria se baseia nas informações sobre uma amostra. Entretanto, em muitos desses livros, em nada se discute como essa amostra é obtida, apenas é assumido que as observações foram obtidas de forma independente e com igual probabilidade de uma determinada população. Com essa motivação nos materiais didáticos presentes na literatura, pretendemos nesse capítulo apresentar a Teoria de Amostragem dando um maior detalhamento teórico e prático.

Definição 7.1: População ou População alvo

O conjunto de elementos (ou unidade de observação) para os quais desejamos que as conclusões de uma pesquisa sejam válidas, com a restrição de que esses elementos possam ser observados ou mensurados sob as mesmas condições é chamado de população, universo ou população alvo. O número total de elementos representa o tamanho da população, denotado por N .

A população pode ser formada por pessoa, famílias, estabelecimentos industriais, ou qualquer tipo de elementos, dependendo basicamente dos objetivos da pesquisa. Podemos dizer ainda que uma população pode ser finita ou infinita. Uma população é finita quando se consegue enumerar todos os elementos que a formam. Refere-se a um universo limitado em uma dada unidade de tempo. Como exemplo podemos dizer que a quantidade de automóveis produzidos em um mês, a população de uma cidade, o número de alunos de uma sala de aula são exemplos de **população finita**. Uma população é dita infinita quando não podemos enumerar todos os elementos. Refere-se

a um universo não delimitado. Os resultados (cara ou coroa) obtidos em sucessivos lances de uma moeda, o conjunto dos números inteiros, reais ou naturais são exemplos de **população infinita**.

Na realidade, a identificação de uma população depende dos objetivos de uma pesquisa (ver Subseção 1.2.1). E isso é algo crucial e um processo extremamente oneroso. Para isso, precisamos de algumas outras definições.

Considerando uma primeira definição sobre uma amostra apresentada na Definição 1.7

Definição 7.2: População amostral

O conjunto de unidades de observação

Uma outra característica interessante da população é que as unidades de observação nem sempre são individualizados...

Exemplo 7.1:

Problema prático para população!

Definição 7.3: Elemento, unidade ou unidade elementar

Indivíduo ou objeto a ser medido ou observado na pesquisa, do qual é a entidade portadora de informações que pretende-se coletar é chamado de unidade elementar (UE).

Um outro tipo de elemento da população é a unidade amostral.

Definição 7.4: Unidade amostral

Um conjunto formado por uma unidade elementar ou várias unidades elementares do qual seja de interesse para pesquisa é chamada de unidade amostral (UA).

Exemplo 7.2:

Uma unidade amostral pode ser os conglomerados pelo método de amostragem por conglomerados. Exemplos de conglomerados são: quarterões, ruas, departamentos, prateleiras, caixas, lotes de produtos, etc..

O interesse na pesquisa está em determinar características da população relevantes para o estudo. Definimos,

Definição 7.5: Parâmetro

Qualquer característica atribuída a população é chamada de parâmetro.

Exemplo 7.3:

Numa pesquisa epidemiológica, a população pode ser definida como todas as pessoas (unidade elementar) da região em estudo, no momento da pesquisa. Um parâmetro de interesse pode ser a porcentagem de pessoas contaminadas.

Como nem sempre é possível coletar informações de todas as unidades elementares da população, faz-se necessário retirarmos informações sobre uma parte da população acessível, e assim, por meio destas informações, consigamos obter informações sobre o parâmetro de interesse na pesquisa.

Definição 7.6: Amostragem

O mecanismo (técnica) que consistem em selecionar parte de uma população para observar, de modo que seja possível preservar as principais informações sobre toda a população.

A parte da população chamamos de amostra.

Definição 7.7: Amostra

Qualquer subconjunto da população obtido por meio de um processo de seleção adequado é chamado de amostra. O tamanho da amostra será representado por n .

Dizemos que quando uma amostra preserva as principais características de uma população, diz-se que esta é uma amostra representativa. Quão mais próximo n estiver de N mais representativo será esta amostra de tamanho n .

As informações obtidas por meio de uma amostra são chamadas de estatísticas. Quando uma estatística estima um parâmetro da população, temos um estimador.

Definição 7.8: Estimador

Um estimador é uma medida, função da amostra (estatística) que representará o parâmetro desconhecido da população.

Exemplo 7.4:

Uma pesquisa deseja saber a altura dos estudantes de uma determinada universidade. Se usarmos uma medida de tendência central para representar a altura dos estudantes, poderíamos escolher a média. Logo, um parâmetro para a população é μ a média populacional. Geralmente μ ou qualquer outro parâmetro é desconhecido, então como estimador para μ podemos considerar \bar{X} , a média amostral. Observe que este estimador é função da amostra. Dizemos que o resultado de um estimador é a estimativa.

A estimativa nada mais é do que criarmos mecanismos para estender as medidas amostrais para que estas possam representar os parâmetros desconhecidos. Como o estimador é função da amostra, se esta não é representativa, com certeza estaremos cometendo algum erro em afirmar que esse estimador poderia representar o parâmetro desconhecido. Daí a importância de obtermos metodologias para coletar amostras representativas da população.

Definição 7.9: Erro amostral

O erro amostral é a diferença entre a estimativa e o parâmetro que se quer estimar.

Esse erro reflete a tendência da amostra. Querendo ou não, o fato de já tomarmos decisão com base em uma amostra, já estamos cometendo erro. Contudo, o que a estatística tenta mostrar, por meio da amostragem, é que podemos tomar conclusões com base em uma amostra sobre a população de modo a cometer o mínimo de erro amostral possível. Assim, dizemos que o objetivo da pesquisa amostral será conhecer características sobre a população, pesquisando (estudando) a amostra, de modo a cometer o mínimo de erro amostral possível.

Se a pesquisa envolve a observação de todas as unidades da população, o método de pesquisa é denominado **censo** ou **pesquisa exaustiva**. Se é conduzida sobre uma amostra da população, o método de pesquisa é denominado **levantamento por amostragem**. O censo somente é aplicável na situação que a população seja finita e suas unidades sejam identificáveis e disponíveis para a coleta da amostra. Por essa razão, o levantamento por amostragem é muito mais frequentemente utilizado. Por que fazer amostragem ao invés de um censo?

- Vantagens: Pesquisa por amostragem em relação ao censo:
 - É mais barata;
 - É mais rápida;
 - É mais fácil de ser controlada por envolver operações menores.

- Desvantagens: Pesquisa por amostragem em relação ao censo:

- O censo pode ser mais vantajoso quando a população é pequena e/ou as informações são de fácil obtenção;
- Os resultados da pesquisa por amostragem contém erros amostrais;
- Se a população for muito heterogênea o erro pode ser muito grande, dependendo do método de amostragem que seja utilizado.

Numa fase inicial dos levantamentos amostrais é necessário **formular o problema** e aventar **hipóteses** sobre o objeto de estudo ou expectativas sobre os possíveis resultados. Ainda nessa fase inicial, o investigador deve **definir a população de estudo**, parte identificável e acessível da população objeto, os **objetivos** e as **variáveis observadas**. Numa segunda etapa é realizado o planejamento, elaborado o **plano de amostragem** ou determinando o caminho a ser percorrido para atingir os objetivos propostos. O plano de amostragem devem ter bem definidos:

1. Unidade amostral: indivíduos ou grupos de indivíduos (conglomerados);
2. Sistema de referência: lista completa das unidades amostrais;
3. Tamanho da população (N), é definido pelo número de indivíduos da população objetivo;
4. Tamanho da amostra (n), definido pelo número de indivíduos selecionados na amostra, tal que $n < N$.

Os planos ou métodos de amostragem podem ser classificados em amostragem probabilística e amostragem não probabilística. Dizemos que se um método de amostragem é objetivo e estabelece uma probabilidade conhecida a cada unidade da população objetivo ser incluída na amostra, esse é denominado amostragem probabilística ou amostragem aleatória; caso contrário, é denominado não probabilística ou amostragem não aleatória. Os principais esquemas amostrais podem ser observados na Figura 7.1.

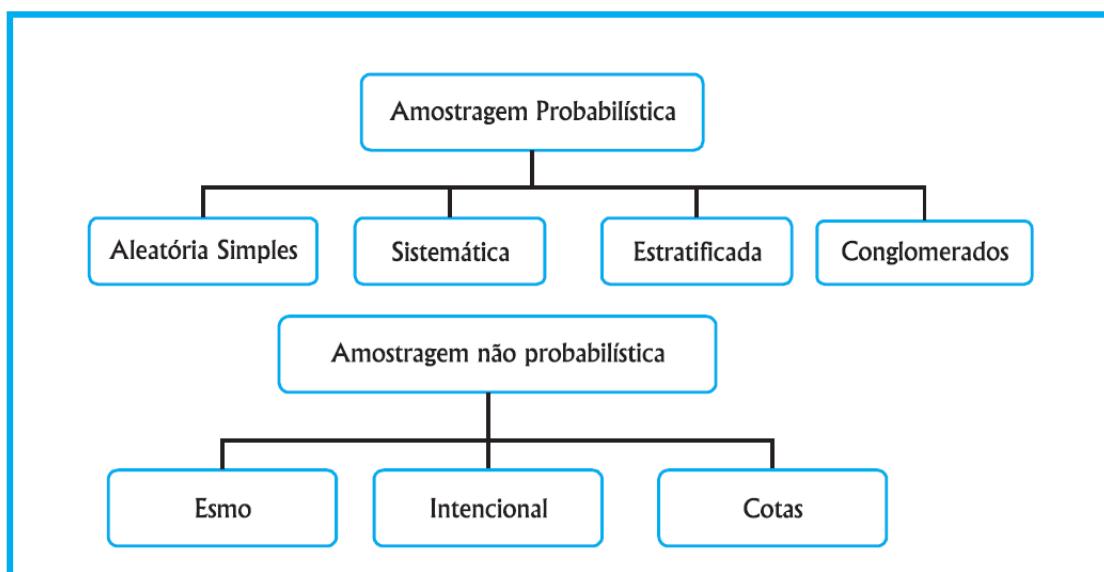


Figura 7.1: Métodos de amostragem probabilística e não-probabilística.

7.2 Amostragem não-probabilística e probabilística

Nas aulas em sala explanamos sobre os métodos de amostragem probabilística. Iremos falar sobre os métodos de amostragem não-probabilística.

O primeiro método é a **amostragem a esmo**. Imagine uma caixa com 1.000 parafusos. A enumeração destes parafusos ficaria muito difícil, e a amostragem aleatória simples se tornaria inviável. Então, em situações deste tipo, supondo que a população de parafusos seja homogênea, escolhemos a esmo a quantidade relativa ao tamanho a amostra. Quanto mais homogênea for a população, mais podemos supor a equivalência com uma amostragem simples ao acaso. Desta forma, os parafusos escolhidos para compor a amostra de um determinado tamanho sem nenhuma norma ou a esmo. Por isso, do nome ao método de amostragem.

Outro método é a **amostragem intencional** que corresponde àquela em que o amostrador deliberadamente escolhe certos elementos para pertencer à amostra, por julgar tais elementos bem representativos da população. Um exemplo deste tipo de amostragem corresponde à situação em que deseja saber a aceitação em relação a uma nova marca de um determinado produto a ser inserida no mercado de uma cidade. Somente entrarão para compor a amostra pessoas que façam uso do produto e que tenham condições financeiras de comprar esta nova marca (classe social de maior poder aquisitivo).

O último método falado é a **amostragem por cotas**. Nesse tipo de amostragem, a população é dividida em grupos, e seleciona-se uma cota proporcional ao tamanho de cada grupo. Entretanto, dentro de cada grupo não é feito sorteio, e sim os elementos são procurados até que a cota de cada grupo seja cumprida. Em pesquisas eleitorais, a divisão de uma população em grupos (considerando, por exemplo, o sexo, o nível de escolaridade, a faixa etária e a renda) pode servir de base para a definição dos grupos, partindo da suposição de que estas variáveis definem grupos com comportamentos diferenciados no processo eleitoral.

7.3 Técnicas de amostragem probabilística

Em sala de aula!

Distribuição de amostragem

8.1 Introdução

8.2 Distribuição de amostragem da média

8.3 Distribuição de amostragem de proporções

8.4 Distribuição de amostragem de diferença entre médias

8.5 Distribuições amostrais (qui-quadrado, t e F)

Teoria da estimação

9.1 Introdução

Chegamos em um dos momentos mais importantes do nosso estudo, que consiste em uma tomada de decisão sobre a população a partir dos dados. Isso é que falamos anteriormente, na extração da informação contida nos dados amostrais para a população em estudo. Para isso, identificamos inicialmente o modelo estatístico/probabilístico que caracteriza a(s) variável(is) em estudo. Contudo, sabemos que esses modelos dependem de características populacionais (parâmetro), que na maioria dos casos, é desconhecida. Desse modo, se considerarmos que um ruído de um sinal digital, segue uma distribuição normal, tal que um ruído ocorre quando a voltagem da transmissão não excede 0,92 volts, estamos limitados na chance de um determinado sinal digital não ter enviado nada, porque apesar de sabermos a distribuição do ruído, não temos informações sobre os seus parâmetros. Uma saída será encontrarmos características (funções) da amostra e que estas possam representar os parâmetros desconhecidos. A forma de encontrarmos as funções amostrais (estimadores) é a área da inferência estatística ou estatística inferencial.

9.2 Conceitos básicos

Assumimos X_1, X_2, \dots, X_n uma amostra aleatória com função densidade de probabilidade (fdp) ou função de probabilidade (fp) $f_X(x; \theta)$, em que a forma de $f_X(x; \theta)$ é conhecida, mas o parâmetro θ é desconhecido. Até agora, usamos a notação p_X para denotar a função de probabilidade e $f_X(x)$ para denotar a função densidade de probabilidade. Por questão de simplificação, podemos usar apenas a notação $f_X(x)$ ou $f_X(x; \theta)$ para representar qualquer uma das funções. Esta última, ainda acrescemos a dependência que essas funções têm de seus parâmetros, e que representa o objetivo de estudo do presente capítulo.

Considere o termo amostra aleatória como um conjunto de variáveis aleatórias independentes e identicamente distribuídas (iid), e que em notação temos $X \stackrel{\text{iid}}{\sim} f_X(x; \theta)$. Ainda podemos assumir que θ pode ser um vetor de parâmetros $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_k]'$. Considere Θ o espaço paramétrico, denotando o conjunto dos possíveis valores que o parâmetro θ pode assumir.

O objetivo é encontrar funções da amostra X_1, X_2, \dots, X_n para serem usadas como estimadores de θ_j , $j = 1, 2, \dots, k$. Ou mais geral, nosso objetivo será tentar encontrar estimadores de certas funções, ditas $\tau_1(\boldsymbol{\theta}), \tau_2(\boldsymbol{\theta}), \dots, \tau_r(\boldsymbol{\theta})$ de $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_k]'$. O ramo da estatística do qual buscamos essas funções da amostra é a inferência estatística.

Definição 9.1: Inferência estatística

Seja uma amostra aleatória X_1, X_2, \dots, X_n com função densidade de probabilidade (fdp) ou função de probabilidade (fp) $f_X(x; \theta)$, em que o parâmetro $\theta \in \Theta$ é desconhecido. Chamamos de inferência estatística o problema que consiste em especificar um ou mais valores para θ , baseado em um conjunto de valores observados x_1, x_2, \dots, x_n de X_1, X_2, \dots, X_n .

Definição 9.2: Estatística

Seja uma amostra aleatória X_1, X_2, \dots, X_n com fdp ou fp $f_X(x; \theta)$, com $\theta \in \Theta$ em que Θ é o espaço paramétrico. Então qualquer função do tipo $T = t(X_1, X_2, \dots, X_n)$ que não depende de θ é chamado de estatística.

Entendemos que o espaço paramétrico é o conjunto de todos os possíveis valores assumidos a um determinado parâmetro. Algumas estatísticas conhecidas são: \bar{X} , S^2 , $X_{(1)}$, $X_{(n)}$, $X_{(n)} - X_{(1)}$, \bar{X}/S^2 . Então a afirmação "...não depender de θ ...", significa que $\theta\bar{X}$ não é uma estatística. Entretanto, as estatísticas têm distribuições que podem depender do parâmetro θ desconhecido, do qual pode ser observado no Exemplo 9.1.

Exemplo 9.1: Estatística e sua distribuição

Consideremos uma população com distribuição normal de parâmetros μ e σ^2 , tal que, a variável aleatória de interesse pode ser representada, em notação $X \sim N(\mu, \sigma^2)$. Então, dada uma amostra aleatória X_1, X_2, \dots, X_n , e uma estatística $\bar{X} = \sum_{i=1}^n X_i/n$, pois \bar{X} depende apenas da amostra, mas a sua distribuição, tal que $\bar{X} \sim N(\mu, \sigma^2/n)$, depende de parâmetros.

A seguir, iremos apresentar dois tipos de estimadores para representar parâmetros desconhecidos, dos quais apresentam aplicações das mais diversas possíveis.

9.3 Estimação pontual

A necessidade de encontrarmos estimadores para determinados parâmetros se deve a diversas aplicações. Por exemplo, os estimadores apresentados até agora, nos informam um ponto amostral (estimador) que representará um ponto populacional (parâmetro). A este, denominamos de estimador pontual, formalizado na Definição 9.3.

Definição 9.3: Estimador pontual

Qualquer estatística cujos valores são usados para estimar $\tau(\theta)$, em que $\tau(\cdot)$ é alguma função do parâmetro θ , é definida ser um estimador pontual de $\tau(\theta)$.

Fazemos a distinção entre estatística e estimador, para chamar a atenção que um estimador deve ser alguma função que esteja relacionado ao parâmetro de interesse. Por exemplo, se desejamos encontrar um estimador para a média populacional μ , em uma amostra dessa população podemos obter diversas estatísticas, \bar{X} , S^2 , S , dentre outras. Porém, é intuitivo percebermos um estimador para μ , dentre esses citados, é a média amostral \bar{X} , uma vez que a característica de interesse populacional é uma medida de posição.

Um estimador é sempre uma estatística que é uma função de uma amostra aleatória e que portanto, também é uma variável aleatória. Porém, uma estatística nem sempre é um estimador para um parâmetro de interesse. Usaremos como notação $\hat{\theta}$ para representar um estimador de θ , ou mais geral, $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$ é um vetor de estimadores de $(\theta_1, \theta_2, \dots, \theta_k)$. Diremos também que o valor do estimador é chamado de estimativa pontual.

Exemplo 9.2:

Considere o tempo de vida útil de uma certa bateria automotiva de uma determinada marca, do qual retirou-se uma amostra de três baterias, cujos valores foram: $X_1 = 1$ ano, $X_2 = 1,6$ anos e $X_3 = 1,4$ anos. Então, em busca de saber a média (μ) do tempo de vida útil da bateria dessa marca, é intuitivo pensar na média amostral $\bar{X} = 1,33$ anos como um valor plausível para representar o parâmetro μ desconhecido, com base nas informações disponíveis na amostra. Assim, o estimador para μ é \bar{X} e 1,33 anos é sua estimativa.

Como observado no Exemplo 9.2 a estimativa representa o valor observado do estimador \bar{X} . Muito embora, as funções de parâmetros são funções identidades desses parâmetros, em algumas situações podemos estar interessados em estimadores pontuais de funções de parâmetros, como apresentado no Exemplo 9.3.

Exemplo 9.3:

Considere uma amostra aleatória X_1, X_2, \dots, X_n de uma população Bernoulli de parâmetro p . Considerando que $\tau(p) = p(1 - p)$ que representa a variância de X , e que é uma função do parâmetro p , então um estimador poderia ser $\tau(\hat{p}) = \hat{p}(1 - \hat{p})$, em que $\hat{p} = \bar{X}$.

Apesar de intuitivo pensarmos na média amostral como um estimador para a média populacional, existem métodos de como obter estimadores pontuais, abordados a seguir.

9.3.1 Métodos para obtenção de estimadores pontuais

Apesar de existir diversos métodos de obtenção de estimadores pontuais, iremos apresentar três métodos largamente utilizados na teoria da Estimação, que são: método dos momentos, método da máxima verossimilhança e método dos mínimos quadrados.

Inicialmente, iremos apresentar como obter um estimador pontual, por meio do método dos momentos. Para isso, definimos pela Definição 9.4, o que é um momento amostral, a seguir.

Definição 9.4: Momentos amostrais

Seja uma amostra aleatória X_1, X_2, \dots, X_n com fdp ou fp $f_X(x; \theta)$, com $\theta = [\theta_1, \theta_2, \dots, \theta_k]' \in \Theta$ em que Θ é o espaço paramétrico. O k -ésimo momento amostral, denotado por M_k , é definido por

$$M_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad (9.1)$$

e o k -ésimo momento amostral em torno da média amostral \bar{X} , denotado por M'_k , é definido por

$$M'_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k. \quad (9.2)$$

Também, para entendermos o método dos momentos, precisamos definir um momento populacional, apresentado na Definição 9.5, na sequência.

Definição 9.5: Momentos populacionais

Seja uma amostra aleatória X_1, X_2, \dots, X_n com fdp ou fp $f_X(x; \theta)$, com $\theta = [\theta_1, \theta_2, \dots, \theta_k]' \in \Theta$ em que Θ é o espaço paramétrico. O k -ésimo momento populacional, denotado por μ_k , é definido por

$$\mu_k = E[X^k], \quad (9.3)$$

e o k -ésimo momento populacional em torno da média populacional $\mu = E[X]$, denotado por μ'_k , é definido por

$$\mu'_k = E[(X - \mu)^k]. \quad (9.4)$$

Geralmente μ_k ou μ'_k é uma função conhecida e função dos k parâmetros $\theta_1, \theta_2, \dots, \theta_k$. Portanto, poderíamos reescrever $\mu_k = \mu_k(\theta_1, \theta_2, \dots, \theta_k)$ e $\mu'_k = \mu'_k(\theta_1, \theta_2, \dots, \theta_k)$.

Definição 9.6: Método dos momentos

O método dos momentos consiste em igualar (9.1) e (9.3) ou (9.2) e (9.4), isto é,

$$M_j = \mu_j(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k), \text{ para } j = 1, 2, \dots, k, \quad (9.5)$$

ou

$$M'_j = \mu'_j(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k), \text{ para } j = 1, 2, \dots, k, \quad (9.6)$$

e obter a solução para as k equações. Diremos que $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ são os estimadores (EMM) de $\theta_1, \theta_2, \dots, \theta_k$ pelo método dos momentos.

Por fim, para complementar o método dos momentos, apresentamos uma sequência de exemplos de como se obter um estimador pontual.

Exemplo 9.4:

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma normal com média μ e variância σ^2 . Denote $(\theta_1, \theta_2) = (\mu, \sigma^2)$. Estimando os parâmetros μ e σ^2 pelo método dos momentos, igualamos

$$\begin{aligned} M_1 &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} = \mu_1 = E[X] = \mu \\ \bar{X} &= \mu. \end{aligned}$$

Igualando o segundo momento usando (9.2) e (9.4), temos

$$M'_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \mu'_2 = E[(X - \mu)^2] = \hat{\sigma}^2.$$

Assim, os estimadores de μ e σ^2 pelo método dos momentos são $\hat{\mu} = \bar{X}$ e $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

Exemplo 9.5:

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma distribuição Poisson com parâmetro λ . Há somente um parâmetro, então há somente uma equação, que é

$$\begin{aligned} M_1 &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} = \mu_1 = E[X] = \lambda \\ \bar{X} &= \lambda. \end{aligned}$$

Então o estimador de λ pelo método dos momentos é $\hat{\lambda} = \bar{X}$.

Exemplo 9.6:

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma distribuição exponencial com densidade $f_X(x; \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$. O estimador pelo método dos momentos é

$$\begin{aligned} M_1 &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} = \mu_1 = E[X] = \frac{1}{\theta} \\ \bar{X} &= \frac{1}{\theta}. \end{aligned}$$

Assim, $\hat{\theta} = \frac{1}{\bar{X}}$.

O outro método apresentado é o da máxima verossimilhança. Inicialmente, definimos uma função de verossimilhança, apresentada na Definição 9.7.

Definição 9.7: Função de Verossimilhança

Seja uma amostra aleatória X_1, X_2, \dots, X_n com fdp ou fp conjunta $f_X(\mathbf{x}; \theta)$, com $\theta \in \Theta$ em que Θ é o espaço paramétrico. Considere ainda x_1, x_2, \dots, x_n a realização da amostra aleatória X_1, X_2, \dots, X_n , então a função de verossimilhança é definida por

$$L(\theta; x_1, x_2, \dots, x_n) = L(\theta; \mathbf{x}) = f_X(\mathbf{x}; \theta) = \prod_{i=1}^n f_X(x_i; \theta). \quad (9.7)$$

Na sequência, Definição 9.8, apresentamos o método da máxima verossimilhança.

Definição 9.8: Método da máxima Verossimilhança

Seja uma função de verossimilhança $L(\theta; x_1, x_2, \dots, x_n)$, Definição 9.7, para uma amostra aleatória X_1, X_2, \dots, X_n . Então o método da máxima verossimilhança é a forma de encontrar um $\hat{\theta} = \vartheta(x_1, x_2, \dots, x_n)$, uma função das observações x_1, x_2, \dots, x_n , que é o valor estimado de $\theta \in \Theta$ que maximiza $L(\theta; x_1, x_2, \dots, x_n)$. Dizemos que $\hat{\theta} = \vartheta(X_1, X_2, \dots, X_n)$ é o estimador de máxima verossimilhança (EMV) de θ .

Para maximizar $L(\theta; x_1, x_2, \dots, x_n)$, tomamos a sua derivada em relação a θ , igualamos a zero e resolvemos o sistema para obtenção de $\hat{\theta} = \vartheta(X_1, X_2, \dots, X_n)$. Posteriormente, devemos identificar se a segunda derivada de $L(\theta; x_1, x_2, \dots, x_n)$ é negativa para saber se $\hat{\theta}$ é um ponto de máximo. Muitas vezes esse processo torna-se complicado. Uma solução é usar a função de Log-verossimilhança, apresentada na Definição 9.9.

Definição 9.9: Função de Log-verossimilhança

Se $L(\theta; x_1, x_2, \dots, x_n)$, expressão (9.7), é a função de verossimilhança, então

$$l(\theta; \mathbf{x}) = l(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x}), \quad (9.8)$$

é a função de log-verossimilhança, para $\mathbf{x} = [x_1, x_2, \dots, x_n]'$.

Como a função logaritmo é monótona crescente, então $l(\theta; \mathbf{x})$ e $L(\theta; \mathbf{x})$ levam ao mesmo máximo de θ . Se denotarmos $h(\theta) = L(\theta; \mathbf{x})$ e $g(y) = \log(y)$, temos

$$0 = \frac{d}{d\theta}(g \circ h(\theta)) = g'(h(\theta))h'(\theta) = \frac{h'(\theta)}{h(\theta)},$$

logo as raízes dessa expressão são as mesmas que $h'(\theta) = 0$. Assim, como é mais fácil fazer manipulações algébricas com a função logaritmo, o problema antes intratável, agora pode ser resolvido.

Exemplo 9.7: EMV de μ de uma normal

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma normal com média μ e variância 1. Vamos determinar o estimador do parâmetro μ pelo método de máxima verossimilhança. A função de

log-verossimilhança pode ser dada por:

$$\begin{aligned} l(\mu; \mathbf{x}) &= \log \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) \\ &= \log \left(\left[\frac{1}{2\pi} \right]^{n/2} e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2}} \right) \\ &= \frac{n}{2} \log \left(\frac{1}{2\pi} \right) - \frac{\sum_{i=1}^n (x_i-\mu)^2}{2}. \end{aligned}$$

Tomando a primeira derivada em relação a μ , temos

$$0 = \frac{d}{d\mu} l(\mu; \mathbf{x}) = \frac{\sum_{i=1}^n (x_i - \hat{\mu})}{2}.$$

Isolando $\hat{\mu}$ segue que

$$\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}.$$

Para verificar se $\hat{\mu}$ é um estimador de máxima verossimilhança, calculemos a segunda da log-verossimilhança, isto é,

$$\frac{d^2}{d\mu^2} l(\mu; \mathbf{x}) = -\frac{n}{2} < 0.$$

Como $\hat{\mu}$ é um ponto de máximo, logo, é um estimador de máxima verossimilhança.

Exemplo 9.8: EMV de p de uma distribuição Bernoulli

Seja X_1, X_2, \dots, X_n uma amostra aleatória de uma distribuição Bernoulli de parâmetro p , então um estimador de máxima verossimilhança de p pode ser obtida a partir da função de verossimilhança, dada por:

$$\begin{aligned} L(p; \mathbf{x}) &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^n x_i} (1-p)^{\sum_{i=1}^n (1-x_i)} \\ &= p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}. \end{aligned}$$

A função de log-verossimilhança será mais fácil, para determinarmos o estimador pontual de p , que segue,

$$\begin{aligned} l(p; \mathbf{x}) &= \log \left(p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i} \right) \\ &= \log(p) \sum_{i=1}^n x_i + \left(n - \sum_{i=1}^n x_i \right) \log(1-p) \\ &= \log(p) \sum_{i=1}^n x_i + n \log(1-p) - \log(1-p) \sum_{i=1}^n x_i. \end{aligned} \tag{9.9}$$

Derivando (9.9) em relação a p e igualando a zero, temos

$$\begin{aligned} 0 &= \frac{\sum_{i=1}^n x_i}{\hat{p}} + \frac{n}{1-\hat{p}} - \frac{\sum_{i=1}^n x_i}{1-\hat{p}} \\ &= \frac{(1-\hat{p})\sum_{i=1}^n x_i + n\hat{p} - \hat{p}\sum_{i=1}^n x_i}{\hat{p}(1-\hat{p})} \\ &= (1-\hat{p})\sum_{i=1}^n x_i + n\hat{p} - \hat{p}\sum_{i=1}^n x_i \\ &= \sum_{i=1}^n x_i + n\hat{p}. \end{aligned}$$

Logo, $\hat{p} = \bar{X}$. Verificando se \hat{p} é um ponto de máximo, calculamos a segunda derivada da função de log-verossimilhança com relação a p , isto é,

$$\frac{d^2 l(p; \mathbf{x})}{d^2 p} = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{(n - \sum_{i=1}^n x_i)}{(1-p)^2} < 0,$$

uma vez que $0 < p < 1$, e $0 \leq \sum_{i=1}^n x_i \leq n$. Considerando que \hat{p} é um ponto de máximo, logo é um estimador de máxima verossimilhança.

Exemplo 9.9: EMV de p de uma distribuição Binomial

Seja Y_1, Y_2, \dots, Y_k uma amostra aleatória de uma distribuição Binomial de parâmetro n e p , de modo que n é conhecido. Considere também que cada variável Y_i representa a soma de sucessos de uma amostra aleatória de tamanho n de uma distribuição Bernoulli de parâmetro p , isto é,

$$\begin{aligned} X_{11}, X_{12}, \dots, X_{1n} &\Rightarrow Y_1 = \sum_{i=1}^n X_{1i} \\ X_{21}, X_{22}, \dots, X_{2n} &\Rightarrow Y_2 = \sum_{i=1}^n X_{2i} \\ &\vdots \\ X_{k1}, X_{k2}, \dots, X_{kn} &\Rightarrow Y_k = \sum_{i=1}^n X_{ki}. \end{aligned}$$

Então um estimador de máxima verossimilhança de p pode ser obtido a partir da função de verossimilhança, dada por:

$$\begin{aligned} L(p; \mathbf{y}) &= \prod_{j=1}^k \binom{n}{y_j} p^{y_j} (1-p)^{n-y_j} \\ &= \binom{n}{y}^k p^{\sum_{j=1}^k y_j} (1-p)^{\sum_{j=1}^k (n-y_j)} \\ &= \binom{n}{y}^k p^{\sum_{j=1}^k y_j} (1-p)^{nk - \sum_{j=1}^k y_j}. \end{aligned}$$

Iremos usar a função de log-verossimilhança para determinarmos o estimador pontual de p , que

segue,

$$\begin{aligned} l(p; \mathbf{y}) &= \log \left(\binom{n}{y}^k p^{\sum_{j=1}^k y_j} (1-p)^{nk - \sum_{j=1}^k y_j} \right) \\ &= k \log \binom{n}{y} + \log(p) \sum_{j=1}^k y_j + \left(nk - \sum_{j=1}^k y_j \right) \log(1-p) \end{aligned} \quad (9.10)$$

Derivando (9.10) em relação a p e igualando a zero, temos

$$\begin{aligned} 0 &= \frac{\sum_{j=1}^k y_j}{\hat{p}} + \frac{nk}{1-\hat{p}} - \frac{\sum_{j=1}^k y_j}{1-\hat{p}} \\ &= \frac{(1-\hat{p}) \sum_{j=1}^k y_j + nk\hat{p} - \hat{p} \sum_{j=1}^k y_j}{\hat{p}(1-\hat{p})} \\ &= (1-\hat{p}) \sum_{j=1}^k y_j + nk\hat{p} - \hat{p} \sum_{j=1}^k y_j \\ &= \sum_{j=1}^k y_j + nk\hat{p}. \end{aligned}$$

Logo, $\hat{p} = \bar{y}/n$. Verificando se \hat{p} é um ponto de máximo, calculamos a segunda derivada da função de log-verossimilhança com relação a p , isto é,

$$\frac{d^2 l(p; \mathbf{y})}{d^2 p} = -\frac{\sum_{j=1}^k y_j}{p^2} - \frac{\left(nk - \sum_{j=1}^k y_j \right)}{(1-p)^2} < 0,$$

uma vez que $0 < p < 1$, e $0 \leq \sum_{j=1}^k y_j \leq nk$. Considerando que \hat{p} é um ponto de máximo, logo $\hat{p} = \bar{Y}/n$ é um estimador de máxima verossimilhança.

Sendo $\hat{p}_j = \bar{X}_j = \sum_{i=1}^n X_{ji}/n = Y_j/n$, $j = 1, 2, \dots, k$, para a j -ésima amostra de uma população Bernoulli de parâmetro p . Logo, temos que o estimador de máxima verossimilhança de p de uma população Binomial, encontrado nesse exercício, pode ser reescrito como

$$\begin{aligned} \hat{p} &= \frac{\bar{Y}}{n} = \frac{1}{nk} \sum_{j=1}^k Y_j = \frac{1}{nk} \sum_{j=1}^k \sum_{i=1}^n X_{ji} \\ &= \frac{1}{nk} \sum_{i=1}^n \sum_{j=1}^k X_{ji} = \frac{1}{k} \sum_{j=1}^k \left(\frac{1}{n} \sum_{i=1}^n X_{ji} \right) \\ &= \frac{1}{k} \sum_{j=1}^k \hat{p}_j, \end{aligned}$$

em que o estimador de máxima verossimilhança de p de uma Binomial é a média dos estimadores das k amostras de tamanho n de uma distribuição Bernoulli.

Por fim, apresentamos o método dos mínimos quadrados, que é um método específico para determinarmos estimadores de um modelo de regressão linear, assunto estudado no Capítulo 11. A ideia é relacionar duas ou mais variáveis de modo que, por meio de um modelo de regressão, consigamos explicar as variações de uma variável, por meio de outra. Considere uma amostra de tamanho n , do qual cada elemento da amostra obtemos duas variáveis (X_i, Y_i), por exemplo, altura e peso de um determinado indivíduo, e nesse caso, apresentamos uma relação funcional entre essas duas variáveis, apresentadas na Definição XX, a seguir.

Definição 9.10: Modelo de Regressão Linear Simples - MRLS

Seja uma variável aleatória Y e X_1, X_2, \dots, X_n um conjunto de variáveis regressoras, então a relação

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (9.11)$$

é chamada de Modelo de Regressão Linear Simples (MRLS), sendo $\beta_j \in \mathbb{R}$, para $j = 0, 1$, com $p' = p + 1$ parâmetros desconhecidos do MRLS, e ε é o erro aleatório não observável.

Observamos que a relação entre X e Y não é uma reta perfeita, devido a adição de ε , que representa a variável erro. Isso porque, ao coletarmos os dados, sempre incorremos em erros de mensuração, seja pela natureza da própria variável, seja pelos instrumentos de medida.

Desse modo, surge o método dos mínimos quadrados dos erros como forma de determinar os estimadores dos parâmetros de modelos lineares, de tal modo, que os estimadores desses parâmetros representam a solução do sistema de equações normais. Dito isso, o que desejamos pelo método dos mínimos quadrados é determinar estimadores para β_0 e β_1 , sendo apresentado, por meio, do Teorema 9.1, a seguir.

Teorema 9.1: Estimador de mínimos quadrados de MRLS

Seja o modelo expresso em 11.4, apresentado na Definição 11.4, então os valores $\hat{\beta}_0$ e $\hat{\beta}_1$, que minimizam $\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$ é igual a

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \quad (9.12)$$

e

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i X_i - \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n}}{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}}, \quad (9.13)$$

respectivamente. Então $\hat{\beta}_0$ e $\hat{\beta}_1$ são os estimadores de mínimos quadrados de β_0 e β_1 , respectivamente.

Prova:

Seja $\varepsilon_i = Y_i - \beta_0 - \beta_1 X_i$ a partir da expressão (11.4), então para minimizar a soma de quadrados dos erros ao longo de todos os n pares (X_i, Y_i) , temos

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2.$$

Derivando Q em relação aos parâmetros β_0 e β_1 , e igualando a zero as funções resultantes, obtemos assim os estimadores de mínimos quadrados. Assim,

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i),$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) X_i.$$

Igualando a zero, teremos o que chamamos de sistema de equações normais (SEN). Assim,

$$\begin{aligned} -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) &= 0, \\ -2 \left[\sum_{i=1}^n Y_i - \sum_{i=1}^n \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n X_i \right] &= 0, \\ \sum_{i=1}^n Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n X_i &= 0, \\ n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i, \end{aligned} \quad (9.14)$$

e

$$\begin{aligned} -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i &= 0, \\ \sum_{i=1}^n Y_i X_i - \hat{\beta}_0 \sum_{i=1}^n X_i - \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= 0, \\ \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n Y_i X_i. \end{aligned} \quad (9.15)$$

As equações (11.6) e (11.7) forma o SEN,

$$\begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i \\ \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i X_i. \end{cases}$$

Isolando β_0 na primeira equação do SEN, temos

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i, \\ n\hat{\beta}_0 &= \sum_{i=1}^n Y_i - \hat{\beta}_1 \sum_{i=1}^n X_i, \\ \hat{\beta}_0 &= \frac{\sum_{i=1}^n Y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n X_i}{n}, \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}. \end{aligned}$$

Substituindo esse resultado na segunda equação do SEN, temos

$$\begin{aligned} (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n Y_i X_i, \\ \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n X_i \sum_{i=1}^n X_i}{n} + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n Y_i X_i, \\ \hat{\beta}_1 \left[\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right] &= \sum_{i=1}^n Y_i X_i - \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n}, \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n Y_i X_i - \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n}}{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}}, \\ \hat{\beta}_1 &= \frac{SPXY}{SQX}. \end{aligned}$$

Portanto o modelo estimado é dado por:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X. \quad (9.16)$$

Desse modo, por meio do Teorema 9.1, obtemos o modelo estimado da relação entre X e Y , dado por:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i.$$

9.3.2 Propriedades de um estimador pontual

Após apresentar alguns métodos de estimação pontual de parâmetros, nos perguntamos, qual dos estimadores de θ é o melhor? Sabemos que muitos desses métodos, apresentam os mesmos estimadores para um determinado parâmetro, ver Exemplos 9.4 e 9.7. Outros apresentam resultados completamente diferentes. Assim, precisamos de algum critério que possa nos informar qual dos estimadores é o que melhor estima θ .

Para uma amostra aleatória X_1, X_2, \dots, X_n com fdp ou fp $f_X(x; \theta)$, $f_X(x; \theta)$, com $\theta \in \Theta$ em que Θ é o espaço paramétrico, poderíamos obter uma escolha inicial do melhor estimador $T = t(X_1, X_2, \dots, X_n)$ de $\tau(\theta)$ com base na Definição 9.11, apresentada a seguir.

Definição 9.11: Estimador não-viesado

Um estimador $T = t(X_1, X_2, \dots, X_n)$ é definido um estimador não-viesado de $\tau(\theta)$ se e somente se

$$E_\theta[T] = E[t(X_1, X_2, \dots, X_n)] = \tau(\theta), \forall \theta \in \Theta. \quad (9.17)$$

9.4 Estimação por intervalo

Na estimativa pontual não temos a ideia da margem do erro que é cometido ao estimarmos o parâmetro. A estimativa intervalar visa preencher esta lacuna criando um intervalo de possíveis valores para o parâmetro θ com margem de erro conhecido.

Definição 9.12: Intervalo de Confiança

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma população com $f_X(x; \theta)$, parametrizada por θ . Considere ainda duas estatísticas $T_1 = t_1(X_1, X_2, \dots, X_n)$ e $T_2 = t_2(X_1, X_2, \dots, X_n)$ satisfazendo $T_1 \leq T_2$, tal que $P_\theta(T_1 \leq \theta \leq T_2) = \gamma$, em que γ não depende de θ . Então o intervalo aleatório (T_1, T_2) é o estimador intervalar de confiança $100\gamma\%$ (ou o intervalo de confiança) de θ , sendo γ o coeficiente de confiança, e T_1 e T_2 são os limites inferior e superior, respectivamente.

Exemplo 9.10: Estimador intervalar

Para uma amostra X_1, X_2, X_3, X_4 de uma população normal com média μ e variância 1, isto é, $X \sim N(\mu, 1)$. Um estimador intervalar possível de μ é $[\bar{X} - 1, \bar{X} + 1]$. Isto significa que o intervalo contém o valor do parâmetro μ com uma certa confiança. \square

Qual a vantagem de ter usado um estimador com menor precisão do que um estimador pontual? Pontualmente, estimamos μ por \bar{X} . E agora estimamos por $[\bar{X} - 1, \bar{X} + 1]$. Apesar de abrirmos mão de uma certa precisão, ganhamos alguma confiança, isto é, uma garantia de que nossa assertão está correta.

Outra importante observação é que o intervalo $[T_1, T_2]$, Definição 9.12, é a quantidade aleatória e não o parâmetro θ . Assim, não podemos afirmar que θ está dentro do intervalo $[T_1, T_2]$ com uma confiança $100\gamma\%$, e sim, que o intervalo $[T_1, T_2]$ contém o parâmetro θ com uma confiança $100\gamma\%$.

Diversos métodos são encontrados na literatura para obter um intervalo de confiança, tais como: inverter uma estatística de teste, intervalos bayesianos, método da quantidade pivotal, dentre outros. Nos restringiremos ao último método, cuja Definição 9.13, é apresentada a seguir.

Definição 9.13: Quantidade pivotal

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma população com $f_X(x; \theta)$, parametrizada por θ . Considere ainda uma medida $Q = q(X_1, X_2, \dots, X_n; \theta)$ uma função de X_1, X_2, \dots, X_n e θ . Se Q tem distribuição independente de θ , então Q é chamado de quantidade pivotal. \square

Exemplo 9.11:

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma população $X \sim N(\theta, 9)$. Assim, podemos afirmar que $Q = \bar{X} - \theta$ é uma quantidade pivotal, pois sua distribuição é normal, tal que $Q \sim N(0, 9/n)$. Agora $Q_2 = \bar{X}/\theta$ não é uma quantidade pivotal, pois $Q_2 \sim N(0, 9/(n\theta^2))$ que depende de θ . \square

Definição 9.14: Método da quantidade pivotal

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma população com $f_X(x; \theta)$, parametrizada por θ . Considere ainda uma quantidade pivotal $Q = q(X_1, X_2, \dots, X_n; \theta)$ com fp ou fdp $f_Q(x)$ que não depende de θ . Fixado $0 < \gamma < 1$ existem q_1 e q_2 que dependem de γ tal que $P(q_1 < Q < q_2) = \gamma$. Se para cada amostra observada (x_1, x_2, \dots, x_n) , $q_1 < q(x_1, x_2, \dots, x_n; \theta) < q_2$ pode ser pivotado em $t_1(x_1, x_2, \dots, x_n) < \tau(\theta) < t_2(x_1, x_2, \dots, x_n)$ para t_1 e t_2 , que não dependem de θ , então (T_1, T_2) é o intervalo de confiança $100\gamma\%$ para $\tau(\theta)$, sendo τ uma função de θ , em que $T_i = t_i(X_1, X_2, \dots, X_n)$, $i = 1, 2$. \square

Dessa definição, devemos fazer algumas observações sobre esse método:

- i) q_1 e q_2 são independentes de θ , uma vez que a distribuição de Q também é;
- ii) Para um valor fixado de γ , há muitos valores possíveis para q_1 e q_2 tal que $P(q_1 < Q < q_2) = \gamma$;
- iii) Diferentes pares de q_1 e q_2 produzem diferentes pares de t_1 e t_2 ;

Dessa forma, precisamos de algum critério para escolher q_1 e q_2 .

Definição 9.15: Tamanho do intervalo

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma população com $f_X(x; \theta)$, parametrizada por θ . Se (T_1, T_2) é o intervalo de confiança $100\gamma\%$ de θ , Definição 9.12, então o tamanho do intervalo, denotado por C , é definido por

$$C_{T_1, T_2} = T_2 - T_1, \quad (9.18)$$

em que $T_2 > T_1$. \square

Dessa forma, devemos escolher o par q_1 e q_2 que resulte em menor comprimento C_{T_1, T_2} .

9.4.1 Intervalo de confiança para a média

Considerando agora uma amostra aleatória X_1, X_2, \dots, X_n de tamanho n de uma população tenha distribuição normal, tal que, $X \sim N(\mu, \sigma^2)$, então apresentamos o seguinte Teorema

Teorema 9.2

Se $X \sim N(\mu, \sigma^2)$, então

- i) \bar{X} e S^2 são independentes;
- ii) $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$, para σ^2 conhecido;
- iii) $\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$;
- iv) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$;

sendo $\bar{X} = (\sum_{i=1}^n X_i/n)$, $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, χ^2_{n-1} é a variável aleatória de uma distribuição de qui-quadrado com $n-1$ graus de liberdade e t_{n-1} é a variável aleatória de uma distribuição de t de Student com $n-1$ graus de liberdade. \square

Observamos que os itens (ii), (iii) e (iv) são todas quantidades pivotais, já que suas distribuições não dependem dos parâmetros de interesse. Dessa forma, poderemos construir intervalo de confiança para essas quantidades.

Teorema 9.3: Intervalo de confiança para μ

Considerando o Teorema 9.2, o intervalo de confiança $100(1-\alpha)\%$ para μ é dado por:

- I) $P\left[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$, considerando a quantidade pivotal do item (iii), Teorema 9.2, para σ^2 conhecido. O quantil superior $100(\alpha/2)\%$ $z_{\frac{\alpha}{2}}$ tem distribuição normal padrão;
- II) $P\left[\bar{X} - t_{\frac{\alpha}{2}, n-1} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}, n-1} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$, considerando a quantidade pivotal do item (iv), Teorema 9.2, sendo S o desvio padrão amostral. O quantil superior $100(\alpha/2)\%$ $t_{\frac{\alpha}{2}, n-1}$ tem distribuição t de Student com $n-1$ graus de liberdade. \square

Prova:

Seja

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

uma quantidade pivotal. Vamos agora pivotar, isto é,

$$\begin{aligned} q_1 &\leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq q_2 \\ q_1 \frac{\sigma}{\sqrt{n}} &\leq \sqrt{n}(\bar{X} - \mu) \leq q_2 \frac{\sigma}{\sqrt{n}} \\ -\bar{X} + q_1 \frac{\sigma}{\sqrt{n}} &\leq -\mu \leq -\bar{X} + q_2 \frac{\sigma}{\sqrt{n}} \\ \bar{X} - q_1 \frac{\sigma}{\sqrt{n}} &\geq \mu \geq \bar{X} - q_2 \frac{\sigma}{\sqrt{n}} \\ T_1 = \bar{X} - q_2 \frac{\sigma}{\sqrt{n}} &\leq \mu \leq T_2 = \bar{X} - q_1 \frac{\sigma}{\sqrt{n}} \end{aligned}$$

O tamanho do intervalo é:

$$\begin{aligned} C_{T_1, T_2} &= \left(\bar{X} - q_1 \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{X} - q_2 \frac{\sigma}{\sqrt{n}} \right) \\ &= (q_2 - q_1) \frac{\sigma}{\sqrt{n}}, \end{aligned}$$

que é o mesmo que

$$1 - \alpha = \int_{q_1}^{q_2} f_Z(k) dk = F_Z(q_2) - F_Z(q_1), \quad (9.19)$$

em que $f_Z(z)$ é a função densidade da normal padrão. Diferenciando (9.19) em relação a q_1 , temos

$$\frac{d}{dq_1} (F_Z(q_2) - F_Z(q_1)) = \frac{dq_2}{dq_1} f_Z(q_2) - f_Z(q_1) = 0. \quad (9.20)$$

Isto implica em

$$\frac{dq_2}{dq_1} = \frac{f_Z(q_1)}{f_Z(q_2)}. \quad (9.21)$$

Para minimizar C_{T_1, T_2} , portanto fazemos $dC_{T_1, T_2}/dq_1$, isto é,

$$\frac{d}{dq_1} (q_2 - q_1) \frac{\sigma}{\sqrt{n}} = \left(\frac{dq_2}{dq_1} - 1 \right) \frac{\sigma}{\sqrt{n}} = 0. \quad (9.22)$$

Substituindo (9.21) em (9.22), temos

$$\left(\frac{f_Z(q_1)}{f_Z(q_2)} - 1 \right) \frac{\sigma}{\sqrt{n}} = 0. \quad (9.23)$$

Isso só ocorrerá se $q_2 = q_1$. Entretanto, $\int_{q_1}^{q_1} f_Z(k) dk \neq 1 - \alpha$. Ou pode ser $q_1 = -q_2$, já que $f_Z(q_1) = f_Z(-q_1)$ pela distribuição ser simétrica, daí poderemos obter $\int_{q_1}^{-q_1} f_Z(k) dk = 1 - \alpha$. Portanto, para minimizar C_{T_1, T_2} , $q_1 = -q_2$. Considerando $q_1 = z_{\frac{\alpha}{2}}$, então $-q_2 = -z_{\frac{\alpha}{2}}$, e segue o resultado do item (I). O resultado do item (II) segue nos mesmos moldes, substituindo apenas σ por S .

9.4.2 Intervalo de confiança para a variância

Teorema 9.4: Intervalo de confiança para σ^2

Considerando o Teorema 9.2, o intervalo de confiança $100(1 - \alpha)\%$ para σ^2 é dado por:

$$P \left[\frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}} \right] = 1 - \alpha, \quad (9.24)$$

sendo $\chi^2_{1-\alpha/2}$ o quantil superior $100(1 - \alpha/2)\%$ e $\chi^2_{\alpha/2}$ o quantil superior $100(\alpha/2)\%$ da distribuição de qui-quadrado com $n - 1$ graus de liberdade. \square

Prova:

9.4.3 Intervalo de confiança para a diferença entre médias

9.5 Dimensionamento de amostras

9.6 Exercícios

Exercício 9.1: Um experimento objetivou detectar a tensão máxima necessária (em MPa) até o aparecimento de trincas em parabrisas de caminhões pesados de uma determinada marca, do qual verificou-se a tensão máxima até o aparecimento numa amostra de 10 parabrisas, cujos valores são: 9,02; 8,82; 7,63; 8,40; 9,29; 9,39; 7,79; 8,64; 7,37; 8,74. Considerando que os dados têm distribuição normal, apresente um estimador intervalar para a média populacional μ para a tensão máxima necessária (em MPa) até o aparecimento de trincas em parabrisas de caminhões pesados de uma determinada marca, com um nível de confiança de 95% de confiança.

Solução na página 233

Exercício 9.2: Considere X_1, X_2, \dots, X_n uma amostra aleatória de uma população com distribuição exponencial de parâmetro λ , tal que sua função densidade de probabilidade é dada por:

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

em que $\lambda > 0$. Determine um estimador para λ pelo método de máxima verossimilhança e pelo método dos momentos.

Solução na página 233

Exercício 9.3: Considere X_1, X_2, \dots, X_n uma amostra aleatória de uma população com distribuição uniforme no intervalo $[0, \theta]$, tal que sua função densidade de probabilidade é dada por:

$$f_X(x) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0 & \text{caso contrário.} \end{cases}$$

Determine um estimador para θ pelo método de máxima verossimilhança.

Solução na página 233

Exercício 9.4: Considere um experimento do qual se observa o desempenho de um sistema operacional implantado em uma determinada indústria de robótica. Foi verificado que o tempo de resposta (em milissegundos) de um comando do operador a uma determinada atividade usando esse sistema operacional, tem distribuição normal com desvio padrão de 20 milissegundos. Retirado uma amostra de tamanho $n = 100$, com média de 52,05 milissegundos, determine um estimador intervalar para a média populacional do tempo de resposta do comando do operador a uma determinada atividade, para verificar o desempenho do sistema operacional implantado, com um nível de confiança de 95% de probabilidade.

Solução na página 233

Exercício 9.5: Considere duas máquinas que produzem peças, dos quais estamos avaliando a resistência à tensão dessas máquinas (MPa). Retirou-se uma amostra de 8 peças de cada máquina, e obtendo as seguintes resistências:

Máquina 1	161	147	162	161	154	136	142	147
Máquina 2	140	162	147	133	130	137	137	136

Considere que a resistência à tensão apresenta distribuição normal e que as variâncias populacionais são desconhecidas para os dados obtidos pelas duas máquinas, porém iguais, encontre um estimador intervalar para a diferença das médias dos dois grupos, com um nível de confiança de 99% de probabilidade.

Solução na página 233

Exercício 9.6: Verifique se o estimador de máxima verossimilhança de $\hat{p} = \bar{Y}/k$, de uma amostra aleatória Y_1, Y_2, \dots, Y_k de uma distribuição Binomial de parâmetros n (conhecido) e p , é um estimador não viesado de p . Verifique também se esse estimador é consistente.

Solução na página 233

Capítulo 10

Teoria da decisão

10.1 Introdução

Consideramos uma amostra aleatória X_1, X_2, \dots, X_n de tamanho $n \geq 2$ definida pelo seu modelo probabilístico $f_X(x)$ que é dependente de um parâmetro θ (pode ser um vetor), em que o conjunto dos valores que θ pode assumir é o espaço paramétrico, denotado por Θ . Denote \mathcal{X} o espaço amostral das observações, isto é, $\mathcal{X} = \{(x_1, x_2, \dots, x_n) : (x_1, x_2, \dots, x_n)\}$ são as possíveis realizações de $(X_1, X_2, \dots, X_n)\}$.

Uma das etapas da inferência estatística é testar a veracidade de algumas afirmações impostas ao parâmetro θ . Essa etapa é a teoria de decisão conhecida também como teste de hipóteses.

Definição 10.1: Hipótese estatística

Uma hipótese estatística, denotada por \mathcal{H} , é uma afirmação ou conjectura feita ao parâmetro θ . Se a hipótese especifica completamente o valor de θ , diz-se uma hipótese simples, caso contrário, é chamada de hipótese composta. \square

Quando a forma da distribuição de probabilidade da população é conhecida e as afirmações feitas são relacionadas aos parâmetros dessa distribuição, dizemos que estamos diante de hipóteses paramétricas. Caso contrário, temos as hipóteses não paramétricas.

Assim, com base numa amostra da população e num teste de hipóteses se decide qual das hipóteses complementares é verdadeira.

Definição 10.2: Hipóteses complementares

As hipóteses complementares de uma hipótese estatística são a hipótese nula, denotada por \mathcal{H}_0 , e a hipótese alternativa, denotada por \mathcal{H}_1 . O formato geral das hipóteses complementares é

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_0^c, \end{cases} \quad (10.1)$$

em que $\Theta_0 \subset \Theta$, e Θ_0^c é seu complemento. \square

Definição 10.3: Hipótese unilateral e hipótese bilateral

Uma hipótese \mathcal{H} que faz tais afirmações para um parâmetro univariado θ ,

$$\mathcal{H} : \theta \geq \theta^* \quad \text{ou} \quad \mathcal{H} : \theta < \theta^*, \quad (10.2)$$

é chamada de hipótese unilateral. Uma hipótese \mathcal{H} que faz a seguinte afirmação de um parâmetro univariado θ ,

$$\mathcal{H} : \theta \neq \theta^*, \quad (10.3)$$

é chamada de hipótese bilateral.

Exemplo 10.1:

Seja uma amostra aleatória X_1, X_2, \dots, X_n , tal que $X \sim N(\theta, 25)$. Se denotarmos $\mathcal{H} : \theta \leq 17$, \mathcal{H} é chamada de hipótese composta. Caso $\mathcal{H} : \theta = 17$, \mathcal{H} é chamada de hipótese simples. \square

10.2 Testes de hipóteses

A Hipótese avaliada pelo pesquisador é a hipótese \mathcal{H}_0 , e ele deve tomar a seguinte decisão: "Rejeitar" ou "Não rejeitar" a hipótese nula. Essa decisão será baseada por um teste.

Definição 10.4: Teste de uma hipótese

Um teste de hipótese, denotado por Y , é uma regra ou procedimento para rejeitar a hipótese \mathcal{H}_0 . \square

A decisão de um teste Y está baseada nas observações x_1, x_2, \dots, x_n da amostra X_1, X_2, \dots, X_n . E a região para qual o teste rejeita a hipótese \mathcal{H}_0 é definida a seguir.

Definição 10.5: Região crítica

Seja Y um teste que avalia a hipótese estatística \mathcal{H}_0 . Então se a decisão for: rejeitar \mathcal{H}_0 se $(x_1, x_2, \dots, x_n) \in C_Y$, em que C_Y é um subconjunto de \mathcal{X} , dizemos que C_Y é a região crítica do teste Y . \square

O teste Y será baseado em uma estatística $W(X_1, X_2, \dots, X_n)$ que é função da amostra X_1, X_2, \dots, X_n com $f_X(x; \theta)$ parametrizada por $\theta \in \Theta$. A decisão por meio da estatística do teste $W(X_1, X_2, \dots, X_n)$ será: se $(X_1, X_2, \dots, X_n) \in C_Y$ a hipótese \mathcal{H}_0 é rejeitada, caso contrário, não haverá evidências para a sua rejeição. Por exemplo, um teste Y pode especificar que H_0 será rejeitada se a média amostral (\bar{X}) for maior que 3. Nesse caso a estatística do teste é $W(X_1, X_2, \dots, X_n) = \bar{X}$ e a região crítica é $C_Y = \{(x_1, x_2, \dots, x_n) : \bar{X} > 3\}$.

Exemplo 10.2:

Tentando verificar a eficiência de um robô desenvolvido para o mercado de ações, supondo que esse robô tem como estratégia operar apenas comprando. Para isso, aplicamos um experimento para o robô sobre 10 ativos. Será que ele está adivinhando na compra desses ativos? Nessa situação, vamos entender que o adivinhar do robô é quando ele realiza a compra de um ativo e não há um retorno efetivo dessa compra, isto é, a decisão da compra de determinado ativo pode ter concretizado em retorno ou prejuízo.

Hipóteses:

H_0 : O robô compra os ativos ao acaso;

H_1 : O robô não compra os ativos ao acaso (compra efetiva);

Considerando p a probabilidade (desconhecida) do robô ter acertado na compra de um ativo, formulamos as hipóteses complementares:

$$\begin{cases} H_0 : p \leq 1/2 \\ H_1 : p > 1/2 \end{cases}$$

Com base nos 10 ativos do experimento e que a decisão da compra desses ativos pelo robô, atuam de forma independente, dizemos que X é uma variável aleatória que assume o número de compras efetivas dentre os 10 ativos, isto é, $X \sim Bin(10, p)$.

Vamos assumir ainda, pelo nosso empirismo que o robô não compra ao acaso se ele realizar 8 ou mais compras efetivas, isto é, rejeitamos H_0 se $X \geq 8$ (Região de rejeição) e não rejeitamos H_0 se $X < 8$ (Região de não rejeição).

Assim,

$$\alpha = P(X \geq 8 | p = 1/2) = \sum_{x=8}^{10} 0.5^x (1 - 0.5)^{10-x} \approx 0,055.$$

Essa probabilidade é chamada de nível de significância ou erro tipo I, denotada por α . Fica claro que o valor de α depende da regra de decisão, por exemplo se a região crítica fosse $X \geq 7$ teríamos $\alpha \approx 0,171875$.

10.3 Erros tipo I e II

Dessa forma, percebemos que a tomada de decisão do pesquisador sobre θ não está livre de erros.

Definição 10.6: Erro tipo I

Seja um teste Y que avalia uma hipótese verdadeira \mathcal{H}_0 . A decisão de rejeitar a hipótese \mathcal{H}_0 , dado que ela é verdadeira, é chamada de erro tipo I. O tamanho do erro tipo I é a probabilidade com que o erro é cometido, denotado por α . \square

Veremos como usar α para construir uma regra de decisão.

Definição 10.7: Erro tipo II

Seja um teste Y que avalia uma hipótese falsa \mathcal{H}_0 . A decisão de aceitar a hipótese \mathcal{H}_0 , dado que ela é falsa, é chamada de erro tipo II. O tamanho do erro tipo II é a probabilidade com que o erro é cometido, denotado por β . \square

Em situações reais esses dois erros podem ocorrer, porém o que se deseja é minimizá-los, uma vez que o valor de α é inversamente proporcional ao valor de β . Assim, fixado um tamanho da amostra (n), a baixa probabilidade de se incorrer no erro tipo I implica numa alta probabilidade de se incorrer no erro tipo II, e o único modo de se causar redução simultânea de ambos é aumentar o tamanho da amostra (n). No desenvolvimento de um teste de hipóteses o que se faz é fixar o valor de α e daí tentar minimizar o valor do erro tipo II. Do ponto de vista teórico, o erro tipo II pode ser minimizado com uma série de ações:

- escolha apropriada do teste e da avaliação do atendimento das pressuposições da utilização do teste. Se o teste é adequado e as pressuposições para sua aplicação estão satisfeitas, existe uma garantia de um maior poder;
- determinação de um tamanho de amostra para que o teste tenha um maior poder possível e que não aumente em demasia o custo da pesquisa a ser realizada; (exemplo: Casella port p. 334)
- a fixação de α entre 0,10 e 0,01, é sempre que possível, uma boa medida por causa da relação inversa entre as taxas do erro tipo I e do erro tipo II.

Se por um lado a probabilidade de se cometer o erro tipo I é conhecida e fixada pelo investigador, por outro a probabilidade do erro tipo II não é conhecida e nem pode ser especificada, uma vez que a hipótese alternativa não tem o seu parâmetro especificado, a menos que estejamos diante de hipóteses simples, algo muito raro nas situações práticas.

A forma de mensurar o tamanho desses erros é avaliado pela função poder.

Definição 10.8: Função Poder

Seja um teste Y com região de rejeição C_Y , então a função poder, denotada por $\pi_Y(\theta)$, é a probabilidade $\pi_Y(\theta) = P_\theta[(X_1, X_2, \dots, X_n) \in C_Y]$ parametrizado por θ . \square

Em outras palavras, a função poder é a probabilidade de rejeitar a hipótese \mathcal{H}_0 quando a distribuição de que a amostra foi obtida é parametrizada por θ . Dessa forma, o teste Y com região C_Y para as hipóteses

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_0^c, \end{cases} \quad (10.4)$$

dado que a hipótese \mathcal{H}_0 é verdadeira, tem probabilidade de erro tipo I dada por $\alpha = \pi_Y(\theta \in \Theta_0) = P_{\theta \in \Theta_0}((X_1, X_2, \dots, X_n) \in C_Y)$. Considerando a probabilidade do erro tipo II dada por β , a probabilidade de rejeitar a hipótese \mathcal{H}_0 , dado que $\theta \in \Theta_0^c$, é dado por $\pi_Y(\theta \in \Theta_0^c) = P_{\theta \in \Theta_0^c}((X_1, X_2, \dots, X_n) \in C_Y)$, do qual chamamos de poder do teste que é na realidade o complemento da probabilidade do erro tipo II, isto é, $1 - \beta$.

Definição 10.9: Tamanho do teste

Seja um teste Y que avalia a hipótese $\mathcal{H}_0 : \theta \in \Theta_0$, $\Theta_0 \subset \Theta$, sendo Θ o espaço paramétrico. Então para $0 \leq \alpha \leq 1$ o tamanho do teste é definido por

$$\alpha = \sup_{\theta \in \Theta_0} \pi_Y(\theta). \quad (10.5)$$

□

Definição 10.10: Nível do teste

Seja um teste Y que avalia a hipótese $\mathcal{H}_0 : \theta \in \Theta_0$, $\Theta_0 \subset \Theta$, sendo Θ o espaço paramétrico. Então para $0 \leq \alpha \leq 1$ o nível do teste é definido por

$$\alpha \geq \sup_{\theta \in \Theta_0} \pi_Y(\theta). \quad (10.6)$$

□

A escolha do nível de significância para um teste é completamente arbitrária. Além disso, quando a distribuição da estatística do teste é discreta, o nível escolhido pode nem mesmo ser atingido. Vejamos o seguinte exemplo,

Exemplo 10.3:

Uma empresa de desenvolvimento de tecnologias garante que 90% dos seus robôs produzidos são eficientes (retorno) no mercado de ações. Baseado nessa informação, selecionamos ao acaso 10 desses robôs e fizemos uma simulação de treinamento para verificar se as estratégias desses robôs eram realmente eficientes ou não. É intuitivo decidir em não fazer parceria com essa empresa no mercado de ações se o número de robôs observados eficientes for muito pequeno. A questão é o quão pequeno devemos assumir.

Nesse caso X assume o número de robôs eficientes na amostra de 10, isto é, $X \sim Bin(10, p)$, cujas hipóteses complementares são:

$$\begin{cases} H_0 : p \geq 0,90 \\ H_1 : p < 0,90 \end{cases}$$

Suponha que decidimos manter $\alpha \geq 0,025$ e a partir deste valor vamos estabelecer a nossa regra de decisão, ou seja obter o valor da constante c tal que H_0 é rejeitada se $X \geq c$. Para isso

vamos calcular α para diferentes regiões críticas,

$$P(X \geq 5 | p = 0,9) \sum_{k=0}^5 0,9^k (1 - 0,9)^{10-k} = 0,002$$

$$P(X \geq 6 | p = 0,9) \sum_{k=0}^6 0,9^k (1 - 0,9)^{10-k} = 0,013$$

$$P(X \geq 7 | p = 0,9) \sum_{k=0}^7 0,9^k (1 - 0,9)^{10-k} = 0,002$$

Portanto, devemos usar a região crítica $X \geq 6$. Isto é, vamos rejeitar o uso desses robôs para o mercado de ações se no mínimo 4 desses robôs forem ineficientes.

Na maioria das aplicações práticas o valor escolhido é 0,05 ou 0,01, mas não há nada que justifique o uso destes valores em particular. Um enfoque consiste em calcular uma quantidade chamada nível crítico, probabilidade de significância ou valor-p.

Definição 10.11: Valor-p

Seja uma amostra aleatória X_1, X_2, \dots, X_n com $f_X(x; \theta), \theta \in \Theta$, sendo Θ o espaço paramétrico. Considere um teste de hipótese Y que avalia $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$, em que $\Theta_0 \subset \Theta$, com região de crítica C_Y e tamanho α . Então o valor-p é uma estatística, denotada por $p(\mathbf{X})$, que satisfaz $0 \leq p(\mathbf{X}) \leq 1$ tal que

$$p(\mathbf{X}) = \inf\{\alpha : \mathbf{X} \in C_Y\}, \quad (10.7)$$

em que $\mathbf{X} = [X_1, X_2, \dots, X_n]'$.

Em geral, o *valor-p* é o menor nível de significância α para que rejeite a hipótese H_0 com base em um teste de tamanho α .

O cálculo do *valor-p* pode ser dado como segue. Considere que o teste Y é da forma: “Rejeite H_0 se $W(\mathbf{X}) > c$ ”, sendo W uma estatística do teste, então sob H_0 , o *valor-p* é a probabilidade $P(W > w)$, sendo w o valor observado de W .

Exemplo 10.4:

Do Exemplo ?? suponha que o número observado de questões certas foi $X = 9$. Então o valor-p será

$$P(X \geq 9 | p = 1/2) = \binom{10}{9} 0,5^9 (1 - 0,5)^{10-9} + \binom{10}{10} 0,5^{10} (1 - 0,5)^{10-10} = 0,0107,$$

e rejeitamos H_0 para todo nível de significância maior do que este valor.

Portanto, o *valor-p* é a probabilidade de observar resultados tão extremos quanto aqueles que foram obtidos se a hipótese nula for verdadeira. A ideia é que se o *valor-p* for grande ele fornece evidência de que H_0 é verdadeira, enquanto que um *valor-p* pequeno indica que existe evidência nos dados contra H_0 .

Da forma como a metodologia clássica de testes de hipóteses foi desenvolvida, podemos ter a impressão de que estamos calculando probabilidades a respeito de uma hipótese. De fato, algumas vezes é incorretamente afirmado que rejeitar H_0 ao nível α indica que a probabilidade de H_0 ser verdadeira é menor do que α . Essa interpretação não é válida e o *valor-p* calculado em um teste não fornece nenhuma indicação sobre qualquer probabilidade a respeito de H_0 .

Por exemplo, um *valor-p* próximo de zero nos fornece (do ponto de vista clássico) muita evidência contra H_0 , porém não significa de maneira alguma $P(H_0 \text{ ser verdadeira})$ seja próximo de

zero. Esta última afirmação probabilística sequer faz sentido na inferência clássica, embora seja exatamente isto que gostaríamos de calcular.

Para que esta interpretação fosse válida, teríamos que usar a abordagem Bayesiana. Basicamente, teríamos que atribuir uma probabilidade *a priori*, isto é, antes de observar os dados, para a hipótese H_0 . Após a observação dos dados amostrais esta probabilidade seria atualizada, segundo regras da inferência Bayesiana, e teríamos uma probabilidade *a posteriori* para a hipótese H_0 .

10.4 Teste unilateral e bilateral

10.5 Passos para a construção de um teste de hipótese

10.6 Teste de hipóteses para a média

10.7 Testes de hipóteses para a proporção

10.8 Testes de hipóteses para a variância

10.9 Testes de hipóteses para a diferença entre médias

10.10 Exercícios

Exercício 10.1: Um professor aplica um teste do tipo certo-errado com 12 questões. Queremos testar a hipótese de que o aluno está advinhando. A hipótese nula é que o aluno acerta ao acaso as questões. A hipótese alternativa é que o aluno tem algum conhecimento. Sendo p a probabilidade (desconhecida) do aluno acertar cada questão, então a hipótese estatística de interesse pode ser dada como $H_0 : p = 1/2$. Como hipótese alternativa $H_1 : p > 1/2$, isto é, o aluno tem algum conhecimento para resolver as questões. Determine:

- A probabilidade do erro tipo I, supondo que adotamos a seguinte regra de decisão: o aluno não está advinhando se acertar 8 ou mais questões.
- Qual seria a regra de decisão, se assumíssimos um nível de significância $\alpha \leq 0,08$?

Solução na página 234

Exercício 10.2: Considere um experimento do qual se observa o desempenho de um sistema operacional implantado em uma determinada indústria de robótica. Foi verificado que o tempo de resposta (em milissegundos) de um comando do operador a uma determinada atividade usando esse sistema operacional, tem distribuição normal com desvio padrão de 20 milissegundos. Retirado uma amostra de tamanho $n = 100$, com média de 52,05 ms. O fabricante nos informou que o tempo médio de resposta da máquina, devido ao sistema operacional, não ultrapassa 50 ms. Diante disso, use um teste de hipótese para verificar a informação do fabricante, ao nível de significância de 5% de probabilidade.

Solução na página 234

Exercício 10.3: Considere duas máquinas que produzem peças, dos quais estamos avaliando a resistência à tensão dessas máquinas (MPa). Retirou-se uma amostra de 8 peças de cada máquina, e obtendo as seguintes resistências:

Máquina 1	161	147	162	161	154	136	142	147
Máquina 2	140	162	147	133	130	137	137	136

Considere que a resistência à tensão apresenta distribuição normal e que as variâncias populacionais são desconhecidas para os dados obtidos pelas duas máquinas, porém iguais. Verifique por meio de um teste de hipóteses, se a resistência à tensão média da máquina 1 é superior a da máquina 2, ao nível de significância de 10% de probabilidade.

Solução na página 234

Exercício 10.4: Um estudo realizado sobre dois tipos de concreto para verificar a taxa de infiltração de água em sua estrutura foi realizado. Os dois tipos de concreto (C1 e C2) diferiam na porcentagem de material retido na peneira de abertura no processo de fabricação desses concretos, em que para fabricação de C1 os agregados retidos na peneira foram 100% de 6,3mm, ao passo que C2 os agregados retidos na peneira foram de 6,3mm (50%) e 4,75mm (50%). Um experimento (ensaio com carga variável) avaliou 40 corpos de provas desses dois tipos de concreto, e mediu a condutividade hidráulica (taxa de infiltração de água em cm/s), dos quais os dados são obtidos na tabela abaixo:

Concreto	Amostra	Média amostral (cm/s)	Desvio padrão amostral (cm/s)
C1	40	0,523	0,054
C2	40	0,641	0,061

Considerando que a condutividade hidráulica tem distribuição normal, verifique se há diferenças estatísticas entre as médias populacionais das condutividades hidráulicas desses dois tipos de concreto, ao nível de significância de 10% de probabilidade. Considere que as variâncias populacionais são homocedásticas.

Solução na página 234

Exercício 10.5: Um estudo realizado por Seker *et al.* (2002) avaliaram os efeitos da energia eletromagnética de celulares, no cérebro de ratos. O experimento foi conduzido *in vivo*, sob dois grupos de ratos, um primeiro grupo controle sem exposição da energia eletromagnética provindo dos celulares, e outro grupo teste apresentando essa exposição. Algumas variáveis foram analisadas, dentre as quais, a pressão sanguínea arterial desses ratos ($mmHg$), coletadas durante o experimento. Os resultados são apresentados na tabela abaixo:

Grupo	Amostra	Média amostral ($mmHg$)	Desvio padrão amostral ($mmHg$)
Teste (Grupo 1)	9	115	10
Controle (Grupo 2)	8	90	5

Dessa forma, indagamos:

- Podemos afirmar que os efeitos da energia eletromagnética expostos aos ratos, proveniente de celulares, podem ter elevado a pressão sanguínea arterial dos ratos? Considere a realização do teste usando os níveis de significância a 1% e 5% de probabilidade, e que a variável mensurada para ambos os grupos apresentam distribuição normal com variâncias desconhecidas e diferentes (heterocedasticidade).
- Determine o *valor – p* para o teste realidade no ítem anterior.
- Por meio de um intervalo de confiança, responda a indagação do ítem (a)
- Podemos afirmar que, em média, a pressão sanguínea do grupo teste é no mínimo 16 $mmHg$ maior do que a média do grupo controle, ao nível de significância de 5% de probabilidade, baseado nas mesmas suposições do ítem (a)?
- Como poderíamos responder responder o ítem (d) por meio de um intervalo de confiança?

Solução na página 234

Exercício 10.6: Baseado nos ítems (d) e (e) do Exercício Proposto 10.5, justifique o que levou a formular as hipóteses complementares desse problema.

Solução na página 234

Correlação e regressão linear simples

11.1 Introdução

Um problema comum na estatística é tentar verificar a associação entre duas variáveis X e Y . Após isto, estamos interessados na forma como ocorre esta associação. Procuramos então uma relação funcional entre essas variáveis, tal que

$$(X, Y) \rightarrow Y \simeq f(X). \quad (11.1)$$

A relação em (11.1) não é perfeita. Os pontos não se situam perfeitamente sobre a função que relaciona as duas variáveis. Mesmo se existisse uma relação exata entre as variáveis como temperatura e pressão, flutuações em torno da curva aparecerão devido a erros de medidas. Frequentemente, o tipo de curva a ser ajustada é sugerido por evidência empírica ou por argumentos teóricos. O modelo a ser adotado depende de vários fatores, por exemplo, natureza das variáveis, relação linear ou não, homogeneidade de variâncias ou não, tipos de erros, independência dos erros etc.

Diremos que a variável Y é aleatória sendo chamada de variável resposta ou dependente. A variável X é fixa, sendo chamada de variável explicativa, regressora ou independente. Os valores da variável X são selecionados pelo pesquisador, não havendo variação aleatória associada. A seleção dos X s pode envolver um conjunto específico de valores ou valores que estão simplesmente dentro de uma amplitude de variação.

Portanto, na estatística é a Teoria de Regressão que tenta encontrar a relação funcional em (11.1), que chamaremos modelo de regressão. Essa teoria teve origem no século XIX com Francis Galton. Em um de seus trabalhos, Galton estudou a relação da altura dos pais (X_i) e dos filhos (Y_i), procurando saber como a altura dos pais influenciava a altura dos filhos. Ele notou que se o pai fosse muito alto ou muito baixo, o filho teria uma altura tendendo à média, isto é, existia uma tendência dos dados regredirem à média. Por isso do nome “regressão”. Os objetivos desse modelo são:

- **Predição:** Esperamos que a maior variação ocorrida na variável Y seja explicada por X . Assim, por meio do modelo de regressão podemos obter valores de Y correspondente aos valores de X que não estavam entre os dados. Em geral, sempre se seleciona valores de X dentro do intervalo das observações de X utilizadas para ajustar o modelo. Muito embora possa ser possível utilizar valores de X fora desse intervalo, dizemos pois que ocorreu uma extrapolação nos dados. A extrapolação deve ser utilizada com muito cuidado, pois o modelo adotado não garante um valor correto para Y fora do intervalo estudado para X . A predição, talvez, seja o uso mais comum para o estudo dos modelos de regressão.
- **Seleção de variáveis:** Muitas vezes a variável Y pode ser explicada por mais de uma variável X , e a relação funcional (11.1) pode ser expressa por: $Y \simeq f(X_1, X_2, \dots, X_p)$. Ou seja, Y é explicada por p variáveis explicativas. Entretanto, nem sempre essas p variáveis apresentam tanta influência na explicação dos valores para Y . Dessa forma, a análise de regressão poderá auxiliar no processo de seleção das p variáveis X , eliminando aquelas que não sejam importantes.
- **Estimação de parâmetros:** Em (11.1), f é dependente de parâmetros, e estes são desconhecidos. Então dada uma amostra e um modelo, a estimativa pontual tenta encontrar, por meio de algum método, estimativas que possam representar os parâmetros desconhecidos. Após substituí-los, temos um modelo ajustado.

- Inferência:** Após realizarmos a estimativa pontual, as estimativas mudam para cada amostra realizada. Por isso, precisamos de uma confiança sobre estas estimativas. Daí fazemos a inferência sobre os parâmetros, por meio dessas estimativas, realizando testes de hipóteses e intervalos de confiança.

11.2 Correlação linear

Inicialmente, estamos interessados em estudar a relação que existe entre as variáveis X e Y . Para isso, caracterizaremos essa relação por meio da correlação linear, Definição XX.

11.2.1 Coeficiente de correlação linear

11.2.2 Teste de hipóteses acerca do coeficiente de correlação linear

11.3 Regressão linear simples

Após estudarmos a relação entre X e Y , podemos também estar interessados em saber a forma dessa relação funcional, que resulta no modelo de regressão, Definição 11.1, apresento a seguir.

Definição 11.1: Modelo de Regressão

Seja uma variável aleatória Y e X_1, X_2, \dots, X_p um conjunto de variáveis regressoras, então a relação

$$Y = f(X_1, X_2, \dots, X_p; \beta_0, \beta_1, \dots, \beta_p) + \varepsilon, \quad (11.2)$$

é chamada de Modelo de Regressão, sendo $\beta_j \in \mathbb{R}$, para $j = 0, 1, 2, \dots, p$, os $p' = p + 1$ parâmetros desconhecidos da função $f(\cdot)$, e ε é o erro aleatório não observável. \square

Definição 11.2: Modelo de Regressão Linear

Um modelo de regressão, expresso em (11.2), tal que

$$\frac{\partial f}{\partial \beta_j} = h(X_1, X_2, \dots, X_p),$$

para $j = 0, 1, 2, \dots, p$, e $h(\cdot)$ uma função qualquer que dependa apenas das variáveis regressoras, é chamado de modelo de regressão linear. \square

Se pelo menos uma das derivadas parciais $\partial f / \partial \beta_j$ depender de algum dos parâmetros, então $f(\cdot)$ é dita função ou modelo de regressão não-linear. Esse ponto não fará parte desse estudo.

Exemplo 11.1: Regressão linear e não-linear

Exemplos de modelos de regressão linear:

- $f(X, \beta_0) = \beta_0$, pois $\frac{df}{d\beta_0} = 1$;
- $Y = f(X, \beta_0, \beta_1) = \beta_0 + \beta_1 X$, pois $\frac{\partial f}{\partial \beta_0} = 1$ e $\frac{\partial f}{\partial \beta_1} = X$.

Exemplos de modelo de regressão não-linear:

- $f(X, \beta_0, \beta_1, \beta_2) = \frac{\beta_0 + \beta_1 X}{1 + \beta_2 X}$, pois $\frac{\partial f}{\partial \beta_0} = \frac{1}{1 + \beta_2}, \frac{\partial f}{\partial \beta_1} = \frac{X}{1 + \beta_2}$, e $\frac{\partial f}{\partial \beta_2} = -\frac{(\beta_0 + \beta_1 X)X}{(1 + \beta_2)^2}$;
- $f(X, \beta_0, \beta_1, \beta_2) = \beta_0 + \beta_1 e^{\beta_2 X}$, pois $\frac{\partial f}{\partial \beta_0} = 1, \frac{\partial f}{\partial \beta_1} = e^{\beta_2 X}$, e $\frac{\partial f}{\partial \beta_2} = \beta_1 X e^{\beta_2 X}$.

Definição 11.3: Modelo de Regressão Linear Múltipla - MRLM

Considerando as Definições 11.1 e 11.2, um modelo de regressão linear múltipla para n observações independentes de Y associados com os X s pode ser definido por

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad (11.3)$$

sendo $i = 1, 2, \dots, n$, supondo $E[\varepsilon_i] = 0$, $\text{var}[\varepsilon_i] = \sigma^2$ e $\text{cov}[\varepsilon_i, \varepsilon_j] = 0$ para todo $i \neq j$. \square

Esse modelo descreve um hiperplano no espaço p -dimensional de variáveis regressoras. Podemos interpretar β_j , para $j = 1, 2, \dots, p$, como a mudança esperada em Y_i devido ao aumento de uma unidade em X_j , estando as outras variáveis regressoras constantes. Por consequência de $E[\varepsilon_i] = 0$ e $\text{var}[\varepsilon_i] = \sigma^2$, temos que $E[Y_i] = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$ e $\text{var}[Y_i] = \sigma^2$.

11.3.1 Modelo

Quando temos apenas uma variável regressão como forma de explicar a variável resposta, estamos diante de um modelo de regressão linear simples, apresentado na Definição 11.4, a seguir.

Definição 11.4: Modelo de Regressão Linear Simples - MRLS

Considerando a Definição 11.3 para $p = 1$, o modelo expresso em (11.3), reduz-se a

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (11.4)$$

também conhecido como modelo de regressão linear simples. \square

Para a expressão (11.4) o parâmetro β_0 representa o coeficiente linear e o β_1 o coeficiente angular para o modelo.

Percebemos que a Definição 11.4 é um caso particular da Definição 11.3. Assim, quando falarmos sobre regressão linear, sempre estaremos nos referindo ao caso geral, Definição 11.3, que englobará todos os outros modelos de regressão linear.

Inicialmente, pensemos num conjunto de pares (X_i, Y_i) , $i = 1, 2, \dots, n$ de duas variáveis e plotemos um gráfico de dispersão para que possamos obter alguma ideia sobre a forma de associação entre X e Y , Figura 11.1. Esse gráfico também ajuda a detectar pontos discrepantes (*Outliers*). Contudo, o gráfico de dispersão deve ser olhado com muito cuidado, uma vez que este não leva em consideração a correlação entre duas ou mais variáveis regressoras, caso exista.

Após averiguar essas informações iniciais, sabemos que mesmo tendo indicativos de um modelo preliminar, a equação (11.3) constitui uma família de funções de parâmetros β_j s. Assim, o objetivo é encontrar aquele que melhor descreva o comportamento de Y .

11.3.2 Estimação dos parâmetros do modelo

Definindo $\varepsilon_i = Y_i - \beta_0 - \beta_1 X_i$ a partir do modelo (11.4), então, um critério interessante para determinar esses estimadores seria minimizar a soma de quadrados desses resíduos ao longo de todos os n pares (X_i, Y_i) .

Essa soma de quadrados é dada por

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2. \quad (11.5)$$

Esse método é chamado de mínimos quadrados e é facilmente obtido derivando esse último relação aos parâmetros β_0 e β_1 igualando esses derivadas a zero.

Considerando,

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2,$$

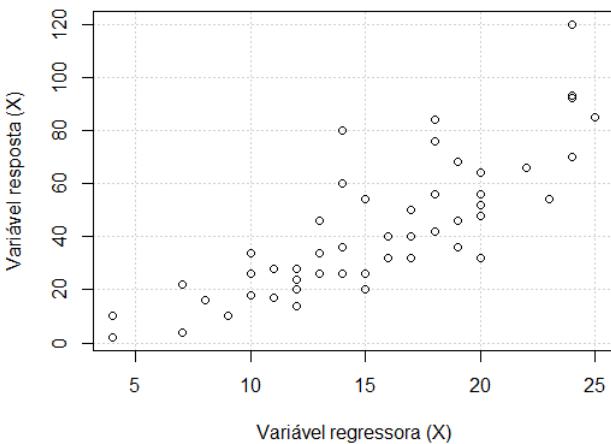


Figura 11.1: Gráfico de dispersão.

então, derivando Q em relação aos parâmetros β_0 e β_1 , e igualando a zero as funções resultantes, obtemos assim os estimadores de mínimos quadrados. Assim,

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i),$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) X_i.$$

Igualando a zero, teremos o que chamamos de sistema de equações normais (SEN). Assim,

$$\begin{aligned} -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) &= 0, \\ -2 \left[\sum_{i=1}^n Y_i - \sum_{i=1}^n \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n X_i \right] &= 0, \\ \sum_{i=1}^n Y_i - n \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n X_i &= 0, \\ n \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i, \end{aligned} \tag{11.6}$$

e

$$\begin{aligned} -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i &= 0, \\ \sum_{i=1}^n Y_i X_i - \hat{\beta}_0 \sum_{i=1}^n X_i - \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= 0, \\ \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n Y_i X_i. \end{aligned} \tag{11.7}$$

As equações (11.6) e (11.7) formam o SEN,

$$\begin{cases} n \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i \\ \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n Y_i X_i. \end{cases}$$

Isolando β_0 na primeira equação do SEN, temos

$$\begin{aligned}
 n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i, \\
 n\hat{\beta}_0 &= \sum_{i=1}^n Y_i - \hat{\beta}_1 \sum_{i=1}^n X_i, \\
 \hat{\beta}_0 &= \frac{\sum_{i=1}^n Y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n X_i}{n}, \\
 \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}.
 \end{aligned} \tag{11.8}$$

Substituindo esse resultado na segunda equação do SEN, temos

$$\begin{aligned}
 (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n Y_i X_i, \\
 \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n X_i \sum_{i=1}^n X_i}{n} + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n Y_i X_i, \\
 \hat{\beta}_1 \left[\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right] &= \sum_{i=1}^n Y_i X_i - \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n}, \\
 \hat{\beta}_1 &= \frac{\sum_{i=1}^n Y_i X_i - \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n}}{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}}, \\
 \hat{\beta}_1 &= \frac{SPXY}{SQX}.
 \end{aligned} \tag{11.9}$$

Portanto o modelo estimado é dado por:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X. \tag{11.10}$$

11.3.3 Somas de quadrados

A função Q representa a soma de quadrados. Logo admitindo que $\epsilon_i = Y_i - \hat{Y}_i$ é o estimador de ϵ , então

$$\sum_{i=1}^n \epsilon_i^2 = SQ_{res} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2,$$

sendo $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$, então

$$\begin{aligned}
\sum_{i=1}^n \epsilon_i^2 &= \sum_{i=1}^n (Y_i - (\bar{Y} - \hat{\beta}_1 \bar{X}) - \hat{\beta}_1 X_i)^2, \\
&= \sum_{i=1}^n (Y_i - \bar{Y} + \hat{\beta}_1 \bar{X} - \hat{\beta}_1 X_i)^2, \\
&= \sum_{i=1}^n [(Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X})]^2, \\
&= \sum_{i=1}^n (Y_i - \bar{Y})^2 - 2\hat{\beta}_1 \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) + \hat{\beta}_1^2 \sum_{i=1}^n (X_i - \bar{X})^2, \\
&= \sum_{i=1}^n (Y_i - \bar{Y})^2 - 2 \left(\frac{SPXY}{SQX} \right) SPXY + \left(\frac{SPXY}{SQX} \right)^2 SQX, \\
&= \sum_{i=1}^n (Y_i - \bar{Y})^2 - 2 \frac{SPXY^2}{SQX} + \frac{SPXY^2}{SQX}, \\
SQ_{res} &= \sum_{i=1}^n (Y_i - \bar{Y})^2 - \frac{SPXY^2}{SQX}. \tag{11.11}
\end{aligned}$$

A soma de quadrados totais é dado por:

$$SQ_{tot} = SQY = SQ_{reg} + SQ_{res}, \tag{11.12}$$

ou

$$\begin{aligned}
SQ_{tot} &= \sum_{i=1}^n (Y_i - \bar{Y})^2, \\
&= \sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} \tag{11.13}
\end{aligned}$$

Assim, sabendo a informação da equação (11.11), a soma de quadrados da regressão pode ser obtida isolando-a da expressão (11.12), dada por:

$$SQ_{reg} = \frac{SPXY^2}{SQX} = \hat{\beta}_1 SPXY = \hat{\beta}_1^2 SQX \tag{11.14}$$

Essas ideias podem ser mais facilmente compreendidas analisando a Figura 11.2. Nessa figura, representamos a decomposição da variação total, mensurada pelo desvio $Y_i - \bar{Y}$ em um desvio devido a regressão $\hat{Y}_i - \bar{Y}$ e pelo erro $Y_i - \hat{Y}_i$, em que \hat{Y}_i é dado pela expressão (11.10). Depreendemos assim, que a variação total pode ser decomposta em uma variação devida à regressão e outra devida a fatores não explicados pelo modelo de regressão adotado. Esses resultados são muitos importantes para a inferência.

11.3.4 Teste de hipóteses para o modelo de regressão

Consideremos $(y_1, x_1), \dots, (y_n, x_n)$ uma amostra aleatória observada na expressão

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \tag{11.15}$$

em que

- β_0, β_1 e x_i : constantes;
- ε_i : variável aleatória, com suposição de normalidade, ou seja, $\varepsilon \sim N(0, \sigma^2)$;

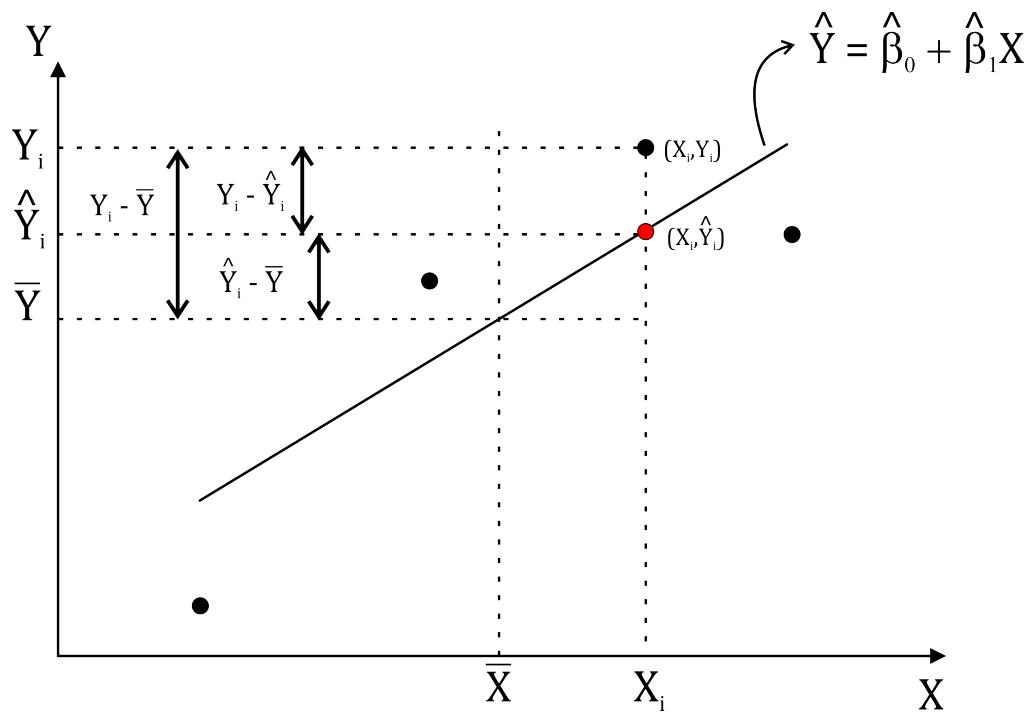


Figura 11.2: Representação gráfica dos diversos desvios envolvidos em uma análise de Regressão linear simples.

- $E[\varepsilon_i] = 0;$
- $Var[\varepsilon_i] = \sigma^2;$
- $Cov[\varepsilon_i, \varepsilon_j] = 0, i \neq j, i, j = 1, 2, \dots, n.$

Os estimadores dos parâmetros β_0 e β_1 são dados nas equações (11.8) e (11.9), respectivamente. Considerando o modelo de regressão linear simples, dado na equação (11.4), então a distribuição de probabilidade Y , correspondente ao valor pré-fixado x de X , é dada por:

$$Y \sim N(\beta_0 + \beta_1 x; \sigma^2),$$

ou seja,

$$\begin{aligned} E[Y|x] &= E[\beta_0 + \beta_1 x + \varepsilon], \\ &= E[\beta_0 + \beta_1 x] + E[\varepsilon], \\ &= \beta_0 + \beta_1 x + 0, \\ &= \beta_0 + \beta_1 x, \end{aligned}$$

e

$$\begin{aligned} Var[Y] &= Var[\beta_0 + \beta_1 x + \varepsilon], \\ &= Var[\beta_0 + \beta_1 x] + Var[\varepsilon], \\ &= 0 + \sigma^2, \\ &= \sigma^2. \end{aligned}$$

Assim, fica provado a distribuição de $Y|x$. Dessa forma, podemos afirmar que a expressão (11.15) também tem distribuição normal,

$$Y_i \sim N(\beta_0 + \beta_1 x_i; \sigma^2),$$

independentes para $i = 1, 2, \dots, n$. Observe que os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$ são combinações lineares de y_1, \dots, y_n , logo, também tem distribuição normal. Iniciando com a esperança de $\hat{\beta}_1$, temos

$$\begin{aligned} E[\hat{\beta}_0] &= E\left[\frac{SPXY}{SQX}\right], \\ &= E\left[\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right], \end{aligned}$$

observe que

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n \frac{\sum_{i=1}^n (x_i)}{n} = 0$$

e

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n y_i(x_i - \bar{x}) - \bar{y} \sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n y_i(x_i - \bar{x}),$$

então,

$$\begin{aligned} E[\hat{\beta}_1] &= E\left[\frac{\sum_{i=1}^n y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right], \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} E[y_i], \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} (\beta_0 + \beta_1 x_i), \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x}) \beta_0 + \beta_1 \sum_{i=1}^n x_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \end{aligned}$$

sabendo que $\sum_{i=1}^n (x_i - \bar{x}) = 0$ e $\sum_{i=1}^n x_i (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})$, então

$$E[\hat{\beta}_1] = \frac{\beta_1 \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta_1. \quad (11.16)$$

Para a esperança de $\hat{\beta}_0$, temos

$$\begin{aligned} E[\hat{\beta}_0] &= E[\bar{y} - \hat{\beta}_1 \bar{x}], \\ &= E\left[\frac{\sum_{i=1}^n y_i}{n}\right] - E[\hat{\beta}_1] \bar{x}, \\ &= \frac{1}{n} \sum_{i=1}^n E[y_i] - \beta_1 \bar{x}, \\ &= \frac{1}{n} \left[n\beta_0 + \beta_1 \sum_{i=1}^n x_i \right] - \beta_1 \bar{x}, \\ &= \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0. \end{aligned} \quad (11.17)$$

Calculando a variância para $\hat{\beta}_1$,

$$\begin{aligned} Var[\hat{\beta}_1] &= Var\left[\frac{\sum_{i=1}^n y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right], \\ &= \frac{1}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} \sum_{i=1}^n Var(y_i)(x_i - \bar{x})^2, \\ &= \frac{1}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} \sum_{i=1}^n \sigma^2 (x_i - \bar{x})^2, \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} \sigma^2, \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned} \quad (11.18)$$

A variância para $\hat{\beta}_0$,

$$\begin{aligned} \text{Var}[\hat{\beta}_0] &= \text{Var}[\bar{y} - \hat{\beta}_1 \bar{x}], \\ &= \text{Var}[\bar{y}] + \text{Var}[\hat{\beta}_1 \bar{x}] - 2\text{Cov}[\bar{y}, \hat{\beta}_1 \bar{x}], \end{aligned}$$

a $2\text{Cov}[\bar{y}, \hat{\beta}_1 \bar{x}] = 0$, pois

$$\begin{aligned} \text{Cov}[\bar{y}, \hat{\beta}_1 \bar{x}] &= \text{Cov}\left[\frac{\sum_{i=1}^n y_i}{n}, \frac{\bar{x} \sum_{i=1}^n y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right], \\ &= \frac{\bar{x}}{n \sum_{i=1}^n (x_i - \bar{x})^2} \text{Cov}\left[\sum_{i=1}^n y_i, \sum_{i=1}^n y_i(x_i - \bar{x})\right], \\ &= \frac{\bar{x}}{n \sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n \text{Cov}[y_i, y_i(x_i - \bar{x})], \\ &= \frac{\bar{x}}{n \sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n (x_i - \bar{x}) \text{Cov}[y_i, y_i], \\ &= \frac{\bar{x}}{n \sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n (x_i - \bar{x}) \sigma^2, \end{aligned}$$

como $\sum_{i=1}^n (x_i - \bar{x}) = 0$, então $\text{Cov}[\bar{y}, \hat{\beta}_1 \bar{x}] = 0$. Assim, a $\text{Var}[\hat{\beta}_0]$ fica da seguinte forma,

$$\begin{aligned} \text{Var}[\hat{\beta}_0] &= \text{Var}[\bar{y}] + \text{Var}[\hat{\beta}_1 \bar{x}], \\ &= \frac{\sum_{i=1}^n \text{Var}[y_i]}{n^2} + \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \bar{x}^2, \\ &= \frac{n\sigma^2}{n^2} + \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \bar{x}^2, \\ &= \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]. \end{aligned} \tag{11.19}$$

11.3.4.1 Verificação da existência do modelo de regressão linear

Quando testamos se o modelo de regressão linear existe, estamos querendo verificar a seguinte hipótese:

$$\begin{aligned} H_0 &\quad \beta_1 = 0, \\ H_1 &\quad \beta_1 \neq 0 \end{aligned}$$

. Sob a hipótese H_0 , temos:

$$\frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{(n-2)}, \tag{11.20}$$

em que $t_{(n-2)}$ é a distribuição t de Student com $v = n - 2$ graus de liberdade. Assim, para um teste de nível α , rejeitamos H_0 , se o módulo da estatística em (11.20) for maior que o quantil $t_{\alpha/2, n-2}$ da distribuição t de Student com $n - 2$ graus de liberdade.

11.3.5 Medidas de adequação do modelo

Uma medida simples da qualidade do ajuste do modelo de regressão aos dados é dado pelo coeficiente de determinação, representado por R^2 e refere-se a proporção da variação das variáveis

dependentes (Y) que é explicada pelo modelo. Assim,

$$\begin{aligned} R^2 &= \frac{SQ_{reg}}{SQ_{tot}}, \\ &= \frac{SPXY^2}{SQX \times SQY}, \end{aligned} \quad (11.21)$$

$$= 1 - \frac{SQ_{res}}{SQ_{tot}}. \quad (11.22)$$

Uma outra medida que é apresentada com o objetivo de propiciar comparações de qualidade do ajuste do modelo linear simples, com o ajuste de outro modelo linear com um número maior de parâmetros, é o coeficiente de determinação ajustado (R_{aj}^2), dado por

$$R_{aj}^2 = 1 - [(n - I)/(n - k)](1 - R^2), \quad (11.23)$$

em que n é o número de observações, I é uma variável indicador que é igual a 1 se o modelo possui intercepto, e 0, caso contrário, e k o número de parâmetros.

Características de R_{aj}^2 :

- $R_{aj}^2 < R^2$;
- pode ser negativo;
- não tem a interpretação prática de R^2 ;
- é mais justo para a comparação de modelos, isto é, quanto mais próximo de 1, melhor o modelo.

11.4 Exercícios

Exercício 11.1: Considere os dados de consumo de combustível (km/l) ao longo do tempo (mês) de um veículo, após uma regulagem eletrônica, apresentados a seguir.

Tempo (X)	1	2	3	4	5	6	7	8	9	10
Consumo (Y)	10,6	10,2	9,6	9,2	8,6	9,0	8,0	7,9	7,3	7,6

Considerando que podemos explicar o consumo ao longo do tempo, apresente o modelo de regressão linear simples estimado para essas duas variáveis.

Solução na página 235

Exercício 11.2: Tentando verificar se existe uma relação linear entre a umidade relativa (%) do ar de secagem de semente e a porcentagem de germinação (%), um estudo foi realizado, do qual se obteve uma amostra de 4 pares de observações, que segue:

Umidade relativa (X)	25	35	45	55
Germinação (Y)	91	95	93	97

Verifique:

- se é possível verificar a existência de relação entre a umidade relativa de secagem com a porcentagem de germinação de sementes, ao nível de significância de 5% de probabilidade;
- qual seria o valor da porcentagem de germinação esperado, considerando uma umidade relativa de 30%?;
- apresente o modelo de regressão linear simples, entre essas duas variáveis.

Solução na página 235

Exercício 11.3: Tentando verificar se existe uma relação linear entre a umidade relativa (%) do ar de secagem de semente e a porcentagem de germinação (%), um estudo foi realizado, do qual se obteve uma amostra de 4 pares de observações, que segue:

Umidade relativa (X)	25	35	45	55
Germinação (Y)	91	95	93	97

Verifique:

- se é possível verificar a existência de relação entre a umidade relativa de secagem com a porcentagem de germinação de sementes, ao nível de significância de 5% de probabilidade;
- qual seria o valor da porcentagem de germinação esperado, considerando uma umidade relativa de 30%?;
- apresente o modelo de regressão linear simples, entre essas duas variáveis.

Solução na página 235

Apêndice A - Introdução ao R

....

Apêndice B - Tabelas

Esse apêndice apresenta as tabelas das principais distribuições amostrais, que serão utilizadas ao longo do curso. Há também outros tipos de tabelas que poderão ser apresentados, tais como a tabelas de números aleatórios, dentre outras.

Tabela da distribuição normal padrão

$$P(Z > z) = \int_z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \quad 0 \leq z \leq 3,99$$

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,49601	0,49202	0,48803	0,48405	0,48006	0,47608	0,47210	0,46812	0,46414
0,1	0,46017	0,45620	0,45224	0,44828	0,44433	0,44038	0,43644	0,43251	0,42858	0,42465
0,2	0,42074	0,41683	0,41294	0,40905	0,40517	0,40129	0,39743	0,39358	0,38974	0,38591
0,3	0,38209	0,37828	0,37448	0,37070	0,36693	0,36317	0,35942	0,35569	0,35197	0,34827
0,4	0,34458	0,34090	0,33724	0,33360	0,32997	0,32636	0,32276	0,31918	0,31561	0,31207
0,5	0,30854	0,30503	0,30153	0,29806	0,29460	0,29116	0,28774	0,28434	0,28096	0,27760
0,6	0,27425	0,27093	0,26763	0,26435	0,26109	0,25785	0,25463	0,25143	0,24825	0,24510
0,7	0,24196	0,23885	0,23576	0,23270	0,22965	0,22663	0,22363	0,22065	0,21770	0,21476
0,8	0,21186	0,20897	0,20611	0,20327	0,20045	0,19766	0,19489	0,19215	0,18943	0,18673
0,9	0,18406	0,18141	0,17879	0,17619	0,17361	0,17106	0,16853	0,16602	0,16354	0,16109
1,0	0,15866	0,15625	0,15386	0,15151	0,14917	0,14686	0,14457	0,14231	0,14007	0,13786
1,1	0,13567	0,13350	0,13136	0,12924	0,12714	0,12507	0,12302	0,12100	0,11900	0,11702
1,2	0,11507	0,11314	0,11123	0,10935	0,10749	0,10565	0,10383	0,10204	0,10027	0,09853
1,3	0,09680	0,09510	0,09342	0,09176	0,09012	0,08851	0,08691	0,08534	0,08379	0,08226
1,4	0,08076	0,07927	0,07780	0,07636	0,07493	0,07353	0,07215	0,07078	0,06944	0,06811
1,5	0,06681	0,06552	0,06426	0,06301	0,06178	0,06057	0,05938	0,05821	0,05705	0,05592
1,6	0,05480	0,05370	0,05262	0,05155	0,05050	0,04947	0,04846	0,04746	0,04648	0,04551
1,7	0,04457	0,04363	0,04272	0,04182	0,04093	0,04006	0,03920	0,03836	0,03754	0,03673
1,8	0,03593	0,03515	0,03438	0,03362	0,03288	0,03216	0,03144	0,03074	0,03005	0,02938
1,9	0,02872	0,02807	0,02743	0,02680	0,02619	0,02559	0,02500	0,02442	0,02385	0,02330
2,0	0,02275	0,02222	0,02169	0,02118	0,02068	0,02018	0,01970	0,01923	0,01876	0,01831
2,1	0,01786	0,01743	0,01700	0,01659	0,01618	0,01578	0,01539	0,01500	0,01463	0,01426
2,2	0,01390	0,01355	0,01321	0,01287	0,01255	0,01222	0,01191	0,01160	0,01130	0,01101
2,3	0,01072	0,01044	0,01017	0,00990	0,00964	0,00939	0,00914	0,00889	0,00866	0,00842
2,4	0,00820	0,00798	0,00776	0,00755	0,00734	0,00714	0,00695	0,00676	0,00657	0,00639
2,5	0,00621	0,00604	0,00587	0,00570	0,00554	0,00539	0,00523	0,00508	0,00494	0,00480
2,6	0,00466	0,00453	0,00440	0,00427	0,00415	0,00402	0,00391	0,00379	0,00368	0,00357
2,7	0,00347	0,00336	0,00326	0,00317	0,00307	0,00298	0,00289	0,00280	0,00272	0,00264
2,8	0,00256	0,00248	0,00240	0,00233	0,00226	0,00219	0,00212	0,00205	0,00199	0,00193
2,9	0,00187	0,00181	0,00175	0,00169	0,00164	0,00159	0,00154	0,00149	0,00144	0,00139
3,0	0,00135	0,00131	0,00126	0,00122	0,00118	0,00114	0,00111	0,00107	0,00104	0,00100
3,1	0,00097	0,00094	0,00090	0,00087	0,00084	0,00082	0,00079	0,00076	0,00074	0,00071
3,2	0,00069	0,00066	0,00064	0,00062	0,00060	0,00058	0,00056	0,00054	0,00052	0,00050
3,3	0,00048	0,00047	0,00045	0,00043	0,00042	0,00040	0,00039	0,00038	0,00036	0,00035
3,4	0,00034	0,00032	0,00031	0,00030	0,00029	0,00028	0,00027	0,00026	0,00025	0,00024
3,5	0,00023	0,00022	0,00022	0,00021	0,00020	0,00019	0,00019	0,00018	0,00017	0,00017
3,6	0,00016	0,00015	0,00015	0,00014	0,00014	0,00013	0,00013	0,00012	0,00012	0,00011
3,7	0,00011	0,00010	0,00010	0,00010	0,00009	0,00009	0,00008	0,00008	0,00008	0,00008
3,8	0,00007	0,00007	0,00007	0,00006	0,00006	0,00006	0,00006	0,00005	0,00005	0,00005
3,9	0,00005	0,00005	0,00004	0,00004	0,00004	0,00004	0,00004	0,00004	0,00003	0,00003

Tabela da distribuição normal padrão

$$P(Z < z) = \int_z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \quad -\infty \leq z \leq 0$$

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,49601	0,49202	0,48803	0,48405	0,48006	0,47608	0,47210	0,46812	0,46414
-0,1	0,46017	0,45620	0,45224	0,44828	0,44433	0,44038	0,43644	0,43251	0,42858	0,42465
-0,2	0,42074	0,41683	0,41294	0,40905	0,40517	0,40129	0,39743	0,39358	0,38974	0,38591
-0,3	0,38209	0,37828	0,37448	0,37070	0,36693	0,36317	0,35942	0,35569	0,35197	0,34827
-0,4	0,34458	0,34090	0,33724	0,33360	0,32997	0,32636	0,32276	0,31918	0,31561	0,31207
-0,5	0,30854	0,30503	0,30153	0,29806	0,29460	0,29116	0,28774	0,28434	0,28096	0,27760
-0,6	0,27425	0,27093	0,26763	0,26435	0,26109	0,25785	0,25463	0,25143	0,24825	0,24510
-0,7	0,24196	0,23885	0,23576	0,23270	0,22965	0,22663	0,22363	0,22065	0,21770	0,21476
-0,8	0,21186	0,20897	0,20611	0,20327	0,20045	0,19766	0,19489	0,19215	0,18943	0,18673
-0,9	0,18406	0,18141	0,17879	0,17619	0,17361	0,17106	0,16853	0,16602	0,16354	0,16109
-1,0	0,15866	0,15625	0,15386	0,15151	0,14917	0,14686	0,14457	0,14231	0,14007	0,13786
-1,1	0,13567	0,13350	0,13136	0,12924	0,12714	0,12507	0,12302	0,12100	0,11900	0,11702
-1,2	0,11507	0,11314	0,11123	0,10935	0,10749	0,10565	0,10383	0,10204	0,10027	0,09853
-1,3	0,09680	0,09510	0,09342	0,09176	0,09012	0,08851	0,08691	0,08534	0,08379	0,08226
-1,4	0,08076	0,07927	0,07780	0,07636	0,07493	0,07353	0,07215	0,07078	0,06944	0,06811
-1,5	0,06681	0,06552	0,06426	0,06301	0,06178	0,06057	0,05938	0,05821	0,05705	0,05592
-1,6	0,05480	0,05370	0,05262	0,05155	0,05050	0,04947	0,04846	0,04746	0,04648	0,04551
-1,7	0,04457	0,04363	0,04272	0,04182	0,04093	0,04006	0,03920	0,03836	0,03754	0,03673
-1,8	0,03593	0,03515	0,03438	0,03362	0,03288	0,03216	0,03144	0,03074	0,03005	0,02938
-1,9	0,02872	0,02807	0,02743	0,02680	0,02619	0,02559	0,02500	0,02442	0,02385	0,02330
-2,0	0,02275	0,02222	0,02169	0,02118	0,02068	0,02018	0,01970	0,01923	0,01876	0,01831
-2,1	0,01786	0,01743	0,01700	0,01659	0,01618	0,01578	0,01539	0,01500	0,01463	0,01426
-2,2	0,01390	0,01355	0,01321	0,01287	0,01255	0,01222	0,01191	0,01160	0,01130	0,01101
-2,3	0,01072	0,01044	0,01017	0,00990	0,00964	0,00939	0,00914	0,00889	0,00866	0,00842
-2,4	0,00820	0,00798	0,00776	0,00755	0,00734	0,00714	0,00695	0,00676	0,00657	0,00639
-2,5	0,00621	0,00604	0,00587	0,00570	0,00554	0,00539	0,00523	0,00508	0,00494	0,00480
-2,6	0,00466	0,00453	0,00440	0,00427	0,00415	0,00402	0,00391	0,00379	0,00368	0,00357
-2,7	0,00347	0,00336	0,00326	0,00317	0,00307	0,00298	0,00289	0,00280	0,00272	0,00264
-2,8	0,00256	0,00248	0,00240	0,00233	0,00226	0,00219	0,00212	0,00205	0,00199	0,00193
-2,9	0,00187	0,00181	0,00175	0,00169	0,00164	0,00159	0,00154	0,00149	0,00144	0,00139
-3,0	0,00135	0,00131	0,00126	0,00122	0,00118	0,00114	0,00111	0,00107	0,00104	0,00100
-3,1	0,00097	0,00094	0,00090	0,00087	0,00084	0,00082	0,00079	0,00076	0,00074	0,00071
-3,2	0,00069	0,00066	0,00064	0,00062	0,00060	0,00058	0,00056	0,00054	0,00052	0,00050
-3,3	0,00048	0,00047	0,00045	0,00043	0,00042	0,00040	0,00039	0,00038	0,00036	0,00035
-3,4	0,00034	0,00032	0,00031	0,00030	0,00029	0,00028	0,00027	0,00026	0,00025	0,00024
-3,5	0,00023	0,00022	0,00022	0,00021	0,00020	0,00019	0,00019	0,00018	0,00017	0,00017
-3,6	0,00016	0,00015	0,00015	0,00014	0,00014	0,00013	0,00013	0,00012	0,00012	0,00011
-3,7	0,00011	0,00010	0,00010	0,00010	0,00009	0,00009	0,00008	0,00008	0,00008	0,00008
-3,8	0,00007	0,00007	0,00007	0,00006	0,00006	0,00006	0,00006	0,00005	0,00005	0,00005
-3,9	0,00005	0,00005	0,00004	0,00004	0,00004	0,00004	0,00004	0,00003	0,00003	0,00003

Tabela da distribuição normal padrão

$$P(0 < Z < z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \quad 0 \leq z \leq 3,99$$

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

Tabela da distribuição normal padrão

$$P(Z < z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \quad 0 \leq z \leq 3,99$$

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997

Tabela dos quantis superiores da distribuição t de Student

$$P(T > t) = \int_t^{\infty} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\nu/2)} \frac{1}{\left(1 + \frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}} dx = \alpha$$

ν	α						
	0,4	0,25	0,1	0,05	0,025	0,01	0,005
1,0	0,325	1,000	3,078	6,314	12,706	31,821	63,657
2,0	0,289	0,816	1,886	2,920	4,303	6,965	9,925
3,0	0,277	0,765	1,638	2,353	3,182	4,541	5,841
4,0	0,271	0,741	1,533	2,132	2,776	3,747	4,604
5,0	0,267	0,727	1,476	2,015	2,571	3,365	4,032
6,0	0,265	0,718	1,440	1,943	2,447	3,143	3,707
7,0	0,263	0,711	1,415	1,895	2,365	2,998	3,499
8,0	0,262	0,706	1,397	1,860	2,306	2,896	3,355
9,0	0,261	0,703	1,383	1,833	2,262	2,821	3,250
10,0	0,260	0,700	1,372	1,812	2,228	2,764	3,169
11,0	0,260	0,697	1,363	1,796	2,201	2,718	3,106
12,0	0,259	0,695	1,356	1,782	2,179	2,681	3,055
13,0	0,259	0,694	1,350	1,771	2,160	2,650	3,012
14,0	0,258	0,692	1,345	1,761	2,145	2,624	2,977
15,0	0,258	0,691	1,341	1,753	2,131	2,602	2,947
16,0	0,258	0,690	1,337	1,746	2,120	2,583	2,921
17,0	0,257	0,689	1,333	1,740	2,110	2,567	2,898
18,0	0,257	0,688	1,330	1,734	2,101	2,552	2,878
19,0	0,257	0,688	1,328	1,729	2,093	2,539	2,861
20,0	0,257	0,687	1,325	1,725	2,086	2,528	2,845
21,0	0,257	0,686	1,323	1,721	2,080	2,518	2,831
22,0	0,256	0,686	1,321	1,717	2,074	2,508	2,819
23,0	0,256	0,685	1,319	1,714	2,069	2,500	2,807
24,0	0,256	0,685	1,318	1,711	2,064	2,492	2,797
25,0	0,256	0,684	1,316	1,708	2,060	2,485	2,787
26,0	0,256	0,684	1,315	1,706	2,056	2,479	2,779
27,0	0,256	0,684	1,314	1,703	2,052	2,473	2,771
28,0	0,256	0,683	1,313	1,701	2,048	2,467	2,763
29,0	0,256	0,683	1,311	1,699	2,045	2,462	2,756
30,0	0,256	0,683	1,310	1,697	2,042	2,457	2,750
31,0	0,256	0,682	1,309	1,696	2,040	2,453	2,744
32,0	0,255	0,682	1,309	1,694	2,037	2,449	2,738
33,0	0,255	0,682	1,308	1,692	2,035	2,445	2,733
34,0	0,255	0,682	1,307	1,691	2,032	2,441	2,728
35,0	0,255	0,682	1,306	1,690	2,030	2,438	2,724

continua...

ν	α						
	0,4	0,25	0,1	0,05	0,025	0,01	0,005
36,0	0,255	0,681	1,306	1,688	2,028	2,434	2,719
37,0	0,255	0,681	1,305	1,687	2,026	2,431	2,715
38,0	0,255	0,681	1,304	1,686	2,024	2,429	2,712
39,0	0,255	0,681	1,304	1,685	2,023	2,426	2,708
40,0	0,255	0,681	1,303	1,684	2,021	2,423	2,704
41,0	0,255	0,681	1,303	1,683	2,020	2,421	2,701
42,0	0,255	0,680	1,302	1,682	2,018	2,418	2,698
43,0	0,255	0,680	1,302	1,681	2,017	2,416	2,695
44,0	0,255	0,680	1,301	1,680	2,015	2,414	2,692
45,0	0,255	0,680	1,301	1,679	2,014	2,412	2,690
46,0	0,255	0,680	1,300	1,679	2,013	2,410	2,687
47,0	0,255	0,680	1,300	1,678	2,012	2,408	2,685
48,0	0,255	0,680	1,299	1,677	2,011	2,407	2,682
49,0	0,255	0,680	1,299	1,677	2,010	2,405	2,680
50,0	0,255	0,679	1,299	1,676	2,009	2,403	2,678
51,0	0,255	0,679	1,298	1,675	2,008	2,402	2,676
52,0	0,255	0,679	1,298	1,675	2,007	2,400	2,674
53,0	0,255	0,679	1,298	1,674	2,006	2,399	2,672
54,0	0,255	0,679	1,297	1,674	2,005	2,397	2,670
55,0	0,255	0,679	1,297	1,673	2,004	2,396	2,668
56,0	0,255	0,679	1,297	1,673	2,003	2,395	2,667
57,0	0,255	0,679	1,297	1,672	2,002	2,394	2,665
58,0	0,255	0,679	1,296	1,672	2,002	2,392	2,663
59,0	0,254	0,679	1,296	1,671	2,001	2,391	2,662
60,0	0,254	0,679	1,296	1,671	2,000	2,390	2,660
61,0	0,254	0,679	1,296	1,670	2,000	2,389	2,659
62,0	0,254	0,678	1,295	1,670	1,999	2,388	2,657
63,0	0,254	0,678	1,295	1,669	1,998	2,387	2,656
64,0	0,254	0,678	1,295	1,669	1,998	2,386	2,655
65,0	0,254	0,678	1,295	1,669	1,997	2,385	2,654
66,0	0,254	0,678	1,295	1,668	1,997	2,384	2,652
67,0	0,254	0,678	1,294	1,668	1,996	2,383	2,651
68,0	0,254	0,678	1,294	1,668	1,995	2,382	2,650
69,0	0,254	0,678	1,294	1,667	1,995	2,382	2,649
70,0	0,254	0,678	1,294	1,667	1,994	2,381	2,648
71,0	0,254	0,678	1,294	1,667	1,994	2,380	2,647
72,0	0,254	0,678	1,293	1,666	1,993	2,379	2,646
73,0	0,254	0,678	1,293	1,666	1,993	2,379	2,645
74,0	0,254	0,678	1,293	1,666	1,993	2,378	2,644
75,0	0,254	0,678	1,293	1,665	1,992	2,377	2,643
76,0	0,254	0,678	1,293	1,665	1,992	2,376	2,642
77,0	0,254	0,678	1,293	1,665	1,991	2,376	2,641
78,0	0,254	0,678	1,292	1,665	1,991	2,375	2,640
79,0	0,254	0,678	1,292	1,664	1,990	2,374	2,640
80,0	0,254	0,678	1,292	1,664	1,990	2,374	2,639
81,0	0,254	0,678	1,292	1,664	1,990	2,373	2,638
82,0	0,254	0,677	1,292	1,664	1,989	2,373	2,637
83,0	0,254	0,677	1,292	1,663	1,989	2,372	2,636

continua...

ν	α						
	0,4	0,25	0,1	0,05	0,025	0,01	0,005
84,0	0,254	0,677	1,292	1,663	1,989	2,372	2,636
85,0	0,254	0,677	1,292	1,663	1,988	2,371	2,635
86,0	0,254	0,677	1,291	1,663	1,988	2,370	2,634
87,0	0,254	0,677	1,291	1,663	1,988	2,370	2,634
88,0	0,254	0,677	1,291	1,662	1,987	2,369	2,633
89,0	0,254	0,677	1,291	1,662	1,987	2,369	2,632
90,0	0,254	0,677	1,291	1,662	1,987	2,368	2,632
91,0	0,254	0,677	1,291	1,662	1,986	2,368	2,631
92,0	0,254	0,677	1,291	1,662	1,986	2,368	2,630
93,0	0,254	0,677	1,291	1,661	1,986	2,367	2,630
94,0	0,254	0,677	1,291	1,661	1,986	2,367	2,629
95,0	0,254	0,677	1,291	1,661	1,985	2,366	2,629
96,0	0,254	0,677	1,290	1,661	1,985	2,366	2,628
97,0	0,254	0,677	1,290	1,661	1,985	2,365	2,627
98,0	0,254	0,677	1,290	1,661	1,984	2,365	2,627
99,0	0,254	0,677	1,290	1,660	1,984	2,365	2,626
100,0	0,254	0,677	1,290	1,660	1,984	2,364	2,626
120,0	0,254	0,677	1,289	1,658	1,980	2,358	2,617
101,0	0,254	0,677	1,290	1,660	1,984	2,364	2,625
102,0	0,254	0,677	1,290	1,660	1,983	2,363	2,625
103,0	0,254	0,677	1,290	1,660	1,983	2,363	2,624
104,0	0,254	0,677	1,290	1,660	1,983	2,363	2,624
105,0	0,254	0,677	1,290	1,659	1,983	2,362	2,623
106,0	0,254	0,677	1,290	1,659	1,983	2,362	2,623
107,0	0,254	0,677	1,290	1,659	1,982	2,362	2,623
108,0	0,254	0,677	1,289	1,659	1,982	2,361	2,622
109,0	0,254	0,677	1,289	1,659	1,982	2,361	2,622
110,0	0,254	0,677	1,289	1,659	1,982	2,361	2,621
111,0	0,254	0,677	1,289	1,659	1,982	2,360	2,621
112,0	0,254	0,677	1,289	1,659	1,981	2,360	2,620
113,0	0,254	0,677	1,289	1,658	1,981	2,360	2,620
114,0	0,254	0,677	1,289	1,658	1,981	2,360	2,620
115,0	0,254	0,677	1,289	1,658	1,981	2,359	2,619
116,0	0,254	0,677	1,289	1,658	1,981	2,359	2,619
117,0	0,254	0,677	1,289	1,658	1,980	2,359	2,619
118,0	0,254	0,677	1,289	1,658	1,980	2,358	2,618
119,0	0,254	0,677	1,289	1,658	1,980	2,358	2,618
120,0	0,254	0,677	1,289	1,658	1,980	2,358	2,617
121,0	0,254	0,677	1,289	1,658	1,980	2,358	2,617
122,0	0,254	0,677	1,289	1,657	1,980	2,357	2,617
123,0	0,254	0,676	1,288	1,657	1,979	2,357	2,616
124,0	0,254	0,676	1,288	1,657	1,979	2,357	2,616
125,0	0,254	0,676	1,288	1,657	1,979	2,357	2,616
126,0	0,254	0,676	1,288	1,657	1,979	2,356	2,615
127,0	0,254	0,676	1,288	1,657	1,979	2,356	2,615
128,0	0,254	0,676	1,288	1,657	1,979	2,356	2,615
129,0	0,254	0,676	1,288	1,657	1,979	2,356	2,614

continua...

ν	α						
	0,4	0,25	0,1	0,05	0,025	0,01	0,005
130,0	0,254	0,676	1,288	1,657	1,978	2,355	2,614
140,0	0,254	0,676	1,288	1,656	1,977	2,353	2,611
150,0	0,254	0,676	1,287	1,655	1,976	2,351	2,609
160,0	0,254	0,676	1,287	1,654	1,975	2,350	2,607
170,0	0,254	0,676	1,287	1,654	1,974	2,348	2,605
180,0	0,254	0,676	1,286	1,653	1,973	2,347	2,603
190,0	0,254	0,676	1,286	1,653	1,973	2,346	2,602
200,0	0,254	0,676	1,286	1,653	1,972	2,345	2,601
240,0	0,254	0,676	1,285	1,651	1,970	2,342	2,596
250,0	0,254	0,675	1,285	1,651	1,969	2,341	2,596
300,0	0,254	0,675	1,284	1,650	1,968	2,339	2,592
∞	0,253	0,674	1,282	1,645	1,960	2,326	2,576

Tabela dos quantis superiores da distribuição de Qui-quadrado

$$P(X > x) = \int_x^{\infty} \frac{1/2}{\Gamma(\nu/2)} \left(\frac{1}{2}t\right)^{\nu/2-1} e^{-\frac{1}{2}t} dt = \alpha$$

ν	α											
	0,995	0,990	0,975	0,950	0,900	0,400	0,250	0,100	0,050	0,025	0,010	0,005
1,0	0,000	0,000	0,001	0,004	0,016	0,708	1,323	2,706	3,841	5,024	6,635	7,879
2,0	0,010	0,020	0,051	0,103	0,211	1,833	2,773	4,605	5,991	7,378	9,210	10,597
3,0	0,072	0,115	0,216	0,352	0,584	2,946	4,108	6,251	7,815	9,348	11,345	12,838
4,0	0,207	0,297	0,484	0,711	1,064	4,045	5,385	7,779	9,488	11,143	13,277	14,860
5,0	0,412	0,554	0,831	1,145	1,610	5,132	6,626	9,236	11,070	12,833	15,086	16,750
6,0	0,676	0,872	1,237	1,635	2,204	6,211	7,841	10,645	12,592	14,449	16,812	18,548
7,0	0,989	1,239	1,690	2,167	2,833	7,283	9,037	12,017	14,067	16,013	18,475	20,278
8,0	1,344	1,646	2,180	2,733	3,490	8,351	10,219	13,362	15,507	17,535	20,090	21,955
9,0	1,735	2,088	2,700	3,325	4,168	9,414	11,389	14,684	16,919	19,023	21,666	23,589
10,0	2,156	2,558	3,247	3,940	4,865	10,473	12,549	15,987	18,307	20,483	23,209	25,188
11,0	2,603	3,053	3,816	4,575	5,578	11,530	13,701	17,275	19,675	21,920	24,725	26,757
12,0	3,074	3,571	4,404	5,226	6,304	12,584	14,845	18,549	21,026	23,337	26,217	28,300
13,0	3,565	4,107	5,009	5,892	7,042	13,636	15,984	19,812	22,362	24,736	27,688	29,819
14,0	4,075	4,660	5,629	6,571	7,790	14,685	17,117	21,064	23,685	26,119	29,141	31,319
15,0	4,601	5,229	6,262	7,261	8,547	15,733	18,245	22,307	24,996	27,488	30,578	32,801
16,0	5,142	5,812	6,908	7,962	9,312	16,780	19,369	23,542	26,296	28,845	32,000	34,267
17,0	5,697	6,408	7,564	8,672	10,085	17,824	20,489	24,769	27,587	30,191	33,409	35,718
18,0	6,265	7,015	8,231	9,390	10,865	18,868	21,605	25,989	28,869	31,526	34,805	37,156
19,0	6,844	7,633	8,907	10,117	11,651	19,910	22,718	27,204	30,144	32,852	36,191	38,582
20,0	7,434	8,260	9,591	10,851	12,443	20,951	23,828	28,412	31,410	34,170	37,566	39,997
21,0	8,034	8,897	10,283	11,591	13,240	21,991	24,935	29,615	32,671	35,479	38,932	41,401
22,0	8,643	9,542	10,982	12,338	14,041	23,031	26,039	30,813	33,924	36,781	40,289	42,796
23,0	9,260	10,196	11,689	13,091	14,848	24,069	27,141	32,007	35,172	38,076	41,638	44,181
24,0	9,886	10,856	12,401	13,848	15,659	25,106	28,241	33,196	36,415	39,364	42,980	45,559
25,0	10,520	11,524	13,120	14,611	16,473	26,143	29,339	34,382	37,652	40,646	44,314	46,928
26,0	11,160	12,198	13,844	15,379	17,292	27,179	30,435	35,563	38,885	41,923	45,642	48,290
27,0	11,808	12,879	14,573	16,151	18,114	28,214	31,528	36,741	40,113	43,195	46,963	49,645
28,0	12,461	13,565	15,308	16,928	18,939	29,249	32,620	37,916	41,337	44,461	48,278	50,993
29,0	13,121	14,256	16,047	17,708	19,768	30,283	33,711	39,087	42,557	45,722	49,588	52,336
30,0	13,787	14,953	16,791	18,493	20,599	31,316	34,800	40,256	43,773	46,979	50,892	53,672

continua...

ν	α											
	0,995	0,990	0,975	0,950	0,900	0,400	0,250	0,100	0,050	0,025	0,010	0,005
31,0	14,458	15,655	17,539	19,281	21,434	32,349	35,887	41,422	44,985	48,232	52,191	55,003
32,0	15,134	16,362	18,291	20,072	22,271	33,381	36,973	42,585	46,194	49,480	53,486	56,328
33,0	15,815	17,074	19,047	20,867	23,110	34,413	38,058	43,745	47,400	50,725	54,776	57,648
34,0	16,501	17,789	19,806	21,664	23,952	35,444	39,141	44,903	48,602	51,966	56,061	58,964
35,0	17,192	18,509	20,569	22,465	24,797	36,475	40,223	46,059	49,802	53,203	57,342	60,275
36,0	17,887	19,233	21,336	23,269	25,643	37,505	41,304	47,212	50,998	54,437	58,619	61,581
37,0	18,586	19,960	22,106	24,075	26,492	38,535	42,383	48,363	52,192	55,668	59,893	62,883
38,0	19,289	20,691	22,878	24,884	27,343	39,564	43,462	49,513	53,384	56,896	61,162	64,181
39,0	19,996	21,426	23,654	25,695	28,196	40,593	44,539	50,660	54,572	58,120	62,428	65,476
40,0	20,707	22,164	24,433	26,509	29,051	41,622	45,616	51,805	55,758	59,342	63,691	66,766
41,0	21,421	22,906	25,215	27,326	29,907	42,651	46,692	52,949	56,942	60,561	64,950	68,053
42,0	22,138	23,650	25,999	28,144	30,765	43,679	47,766	54,090	58,124	61,777	66,206	69,336
43,0	22,859	24,398	26,785	28,965	31,625	44,706	48,840	55,230	59,304	62,990	67,459	70,616
44,0	23,584	25,148	27,575	29,787	32,487	45,734	49,913	56,369	60,481	64,201	68,710	71,893
45,0	24,311	25,901	28,366	30,612	33,350	46,761	50,985	57,505	61,656	65,410	69,957	73,166
46,0	25,041	26,657	29,160	31,439	34,215	47,787	52,056	58,641	62,830	66,617	71,201	74,437
47,0	25,775	27,416	29,956	32,268	35,081	48,814	53,127	59,774	64,001	67,821	72,443	75,704
48,0	26,511	28,177	30,755	33,098	35,949	49,840	54,196	60,907	65,171	69,023	73,683	76,969
49,0	27,249	28,941	31,555	33,930	36,818	50,866	55,265	62,038	66,339	70,222	74,919	78,231
50,0	27,991	29,707	32,357	34,764	37,689	51,892	56,334	63,167	67,505	71,420	76,154	79,490
51,0	28,735	30,475	33,162	35,600	38,560	52,917	57,401	64,295	68,669	72,616	77,386	80,747
52,0	29,481	31,246	33,968	36,437	39,433	53,942	58,468	65,422	69,832	73,810	78,616	82,001
53,0	30,230	32,018	34,776	37,276	40,308	54,967	59,534	66,548	70,993	75,002	79,843	83,253
54,0	30,981	32,793	35,586	38,116	41,183	55,992	60,600	67,673	72,153	76,192	81,069	84,502
55,0	31,735	33,570	36,398	38,958	42,060	57,016	61,665	68,796	73,311	77,380	82,292	85,749
56,0	32,490	34,350	37,212	39,801	42,937	58,040	62,729	69,919	74,468	78,567	83,513	86,994
57,0	33,248	35,131	38,027	40,646	43,816	59,064	63,793	71,040	75,624	79,752	84,733	88,236
58,0	34,008	35,913	38,844	41,492	44,696	60,088	64,857	72,160	76,778	80,936	85,950	89,477
59,0	34,770	36,698	39,662	42,339	45,577	61,111	65,919	73,279	77,931	82,117	87,166	90,715
60,0	35,534	37,485	40,482	43,188	46,459	62,135	66,981	74,397	79,082	83,298	88,379	91,952
61,0	36,301	38,273	41,303	44,038	47,342	63,158	68,043	75,514	80,232	84,476	89,591	93,186
62,0	37,068	39,063	42,126	44,889	48,226	64,181	69,104	76,630	81,381	85,654	90,802	94,419
63,0	37,838	39,855	42,950	45,741	49,111	65,204	70,165	77,745	82,529	86,830	92,010	95,649
64,0	38,610	40,649	43,776	46,595	49,996	66,226	71,225	78,860	83,675	88,004	93,217	96,878
65,0	39,383	41,444	44,603	47,450	50,883	67,249	72,285	79,973	84,821	89,177	94,422	98,105
66,0	40,158	42,240	45,431	48,305	51,770	68,271	73,344	81,085	85,965	90,349	95,626	99,330
67,0	40,935	43,038	46,261	49,162	52,659	69,293	74,403	82,197	87,108	91,519	96,828	100,554
68,0	41,713	43,838	47,092	50,020	53,548	70,315	75,461	83,308	88,250	92,689	98,028	101,776
69,0	42,494	44,639	47,924	50,879	54,438	71,337	76,519	84,418	89,391	93,856	99,228	102,996
70,0	43,275	45,442	48,758	51,739	55,329	72,358	77,577	85,527	90,531	95,023	100,425	104,215
71,0	44,058	46,246	49,592	52,600	56,221	73,380	78,634	86,635	91,670	96,189	101,621	105,432
72,0	44,843	47,051	50,428	53,462	57,113	74,401	79,690	87,743	92,808	97,353	102,816	106,648
73,0	45,629	47,858	51,265	54,325	58,006	75,422	80,747	88,850	93,945	98,516	104,010	107,862
74,0	46,417	48,666	52,103	55,189	58,900	76,443	81,803	89,956	95,081	99,678	105,202	109,074
75,0	47,206	49,475	52,942	56,054	59,795	77,464	82,858	91,061	96,217	100,839	106,393	110,286

continua...

ν	α											
	0,995	0,990	0,975	0,950	0,900	0,400	0,250	0,100	0,050	0,025	0,010	0,005
76	47,997	50,286	53,782	56,920	60,690	78,485	83,913	92,166	97,351	101,999	107,583	111,495
77	48,788	51,097	54,623	57,786	61,586	79,505	84,968	93,270	98,484	103,158	108,771	112,704
78	49,582	51,910	55,466	58,654	62,483	80,526	86,022	94,374	99,617	104,316	109,958	113,911
79	50,376	52,725	56,309	59,522	63,380	81,546	87,077	95,476	100,749	105,473	111,144	115,117
80	51,172	53,540	57,153	60,391	64,278	82,566	88,130	96,578	101,879	106,629	112,329	116,321
81	51,969	54,357	57,998	61,261	65,176	83,586	89,184	97,680	103,010	107,783	113,512	117,524
82	52,767	55,174	58,845	62,132	66,076	84,606	90,237	98,780	104,139	108,937	114,695	118,726
83	53,567	55,993	59,692	63,004	66,976	85,626	91,289	99,880	105,267	110,090	115,876	119,927
84	54,368	56,813	60,540	63,876	67,876	86,646	92,342	100,980	106,395	111,242	117,057	121,126
85	55,170	57,634	61,389	64,749	68,777	87,665	93,394	102,079	107,522	112,393	118,236	122,325
86	55,973	58,456	62,239	65,623	69,679	88,685	94,446	103,177	108,648	113,544	119,414	123,522
87	56,777	59,279	63,089	66,498	70,581	89,704	95,497	104,275	109,773	114,693	120,591	124,718
88	57,582	60,103	63,941	67,373	71,484	90,723	96,548	105,372	110,898	115,841	121,767	125,913
89	58,389	60,928	64,793	68,249	72,387	91,742	97,599	106,469	112,022	116,989	122,942	127,106
90	59,196	61,754	65,647	69,126	73,291	92,761	98,650	107,565	113,145	118,136	124,116	128,299
91	60,005	62,581	66,501	70,003	74,196	93,780	99,700	108,661	114,268	119,282	125,289	129,491
92	60,815	63,409	67,356	70,882	75,100	94,799	100,750	109,756	115,390	120,427	126,462	130,681
93	61,625	64,238	68,211	71,760	76,006	95,818	101,800	110,850	116,511	121,571	127,633	131,871
94	62,437	65,068	69,068	72,640	76,912	96,836	102,850	111,944	117,632	122,715	128,803	133,059
95	63,250	65,898	69,925	73,520	77,818	97,855	103,899	113,038	118,752	123,858	129,973	134,247
96	64,063	66,730	70,783	74,401	78,725	98,873	104,948	114,131	119,871	125,000	131,141	135,433
97	64,878	67,562	71,642	75,282	79,633	99,892	105,997	115,223	120,990	126,141	132,309	136,619
98	65,694	68,396	72,501	76,164	80,541	100,910	107,045	116,315	122,108	127,282	133,476	137,803
99	66,510	69,230	73,361	77,046	81,449	101,928	108,093	117,407	123,225	128,422	134,642	138,987
100	67,328	70,065	74,222	77,929	82,358	102,946	109,141	118,498	124,342	129,561	135,807	140,169
120	83,85	86,92	91,57	95,70	100,62	123,29	130,05	140,23	146,57	152,21	158,95	163,65
240	187,32	191,99	198,98	205,14	212,39	244,92	254,39	268,47	277,14	284,80	293,89	300,18
480	403,95	410,87	421,19	430,20	440,75	487,22	500,52	520,11	532,08	542,60	555,01	563,56
600	514,53	522,37	534,02	544,18	556,06	608,15	622,99	644,80	658,09	669,77	683,52	692,98
960	850,89	861,01	876,03	889,08	904,29	970,47	989,18	1016,57	1033,19	1047,76	1064,87	1076,62

Tabela dos quantis superiores da distribuição F ($\alpha = 0,25$)

$$P(X > x) = \int_x^{\infty} \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{t^{\frac{\nu_1 - 2}{2}}}{\left(1 + \frac{t\nu_1}{\nu_2}\right)^{\frac{\nu_1 + \nu_2}{2}}} dt = 0,25$$

ν_2	ν_1													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	5,828	7,500	8,200	8,581	8,820	8,983	9,102	9,192	9,263	9,320	9,367	9,406	9,440	9,468
2	2,571	3,000	3,153	3,232	3,280	3,312	3,335	3,353	3,366	3,377	3,386	3,393	3,400	3,405
3	2,024	2,280	2,356	2,390	2,409	2,422	2,430	2,436	2,441	2,445	2,448	2,450	2,452	2,454
4	1,807	2,000	2,047	2,064	2,072	2,077	2,079	2,080	2,081	2,082	2,082	2,083	2,083	2,083
5	1,692	1,853	1,884	1,893	1,895	1,894	1,894	1,892	1,891	1,890	1,889	1,888	1,887	1,886
6	1,621	1,762	1,784	1,787	1,785	1,782	1,779	1,776	1,773	1,771	1,769	1,767	1,765	1,764
7	1,573	1,701	1,717	1,716	1,711	1,706	1,701	1,697	1,693	1,690	1,687	1,684	1,682	1,680
8	1,538	1,657	1,668	1,664	1,658	1,651	1,645	1,640	1,635	1,631	1,627	1,624	1,622	1,619
9	1,512	1,624	1,632	1,625	1,617	1,609	1,602	1,596	1,591	1,586	1,582	1,579	1,576	1,573
10	1,491	1,598	1,603	1,595	1,585	1,576	1,569	1,562	1,556	1,551	1,547	1,543	1,540	1,537
11	1,475	1,577	1,580	1,570	1,560	1,550	1,542	1,535	1,528	1,523	1,518	1,514	1,510	1,507
12	1,461	1,560	1,561	1,550	1,539	1,529	1,520	1,512	1,505	1,500	1,495	1,490	1,486	1,483
13	1,450	1,545	1,545	1,534	1,521	1,511	1,501	1,493	1,486	1,480	1,475	1,470	1,466	1,462
14	1,440	1,533	1,532	1,519	1,507	1,495	1,485	1,477	1,470	1,463	1,458	1,453	1,449	1,445
15	1,432	1,523	1,520	1,507	1,494	1,482	1,472	1,463	1,456	1,449	1,443	1,438	1,434	1,430
20	1,404	1,487	1,481	1,465	1,450	1,437	1,425	1,415	1,407	1,399	1,393	1,387	1,382	1,378
25	1,387	1,466	1,458	1,441	1,424	1,410	1,398	1,387	1,378	1,370	1,363	1,357	1,352	1,347
30	1,376	1,452	1,443	1,424	1,407	1,392	1,380	1,369	1,359	1,351	1,343	1,337	1,331	1,326
40	1,363	1,435	1,424	1,404	1,386	1,371	1,357	1,345	1,335	1,327	1,319	1,312	1,306	1,300
50	1,355	1,425	1,413	1,393	1,374	1,358	1,344	1,332	1,321	1,312	1,304	1,297	1,291	1,285
60	1,349	1,419	1,405	1,385	1,366	1,349	1,335	1,323	1,312	1,303	1,294	1,287	1,280	1,274
70	1,346	1,414	1,400	1,379	1,360	1,343	1,329	1,316	1,305	1,296	1,287	1,280	1,273	1,267
80	1,343	1,411	1,396	1,375	1,355	1,338	1,324	1,311	1,300	1,291	1,282	1,275	1,268	1,262
90	1,341	1,408	1,393	1,372	1,352	1,335	1,320	1,307	1,296	1,287	1,278	1,270	1,263	1,257
100	1,339	1,406	1,391	1,369	1,349	1,332	1,317	1,304	1,293	1,283	1,275	1,267	1,260	1,254
200	1,331	1,396	1,380	1,358	1,337	1,319	1,304	1,291	1,279	1,269	1,260	1,252	1,245	1,238
300	1,328	1,393	1,377	1,354	1,333	1,315	1,300	1,286	1,275	1,264	1,255	1,247	1,240	1,233
∞	1,323	1,386	1,369	1,346	1,325	1,307	1,291	1,277	1,265	1,255	1,246	1,237	1,230	1,223

continua...

v_2	v_1													
	15	20	25	30	40	50	60	70	80	90	100	200	300	∞
1	9,493	9,581	9,634	9,670	9,714	9,741	9,759	9,772	9,782	9,789	9,795	9,822	9,831	9,849
2	3,410	3,426	3,436	3,443	3,451	3,456	3,459	3,462	3,464	3,465	3,466	3,471	3,473	3,476
3	2,455	2,460	2,463	2,465	2,467	2,469	2,470	2,470	2,471	2,471	2,471	2,473	2,473	2,474
4	2,083	2,083	2,083	2,082	2,082	2,082	2,082	2,082	2,081	2,081	2,081	2,081	2,081	2,081
5	1,885	1,882	1,880	1,878	1,876	1,875	1,874	1,874	1,873	1,873	1,872	1,871	1,870	1,869
6	1,762	1,757	1,753	1,751	1,748	1,746	1,744	1,743	1,742	1,742	1,741	1,739	1,738	1,737
7	1,678	1,671	1,667	1,663	1,659	1,657	1,655	1,654	1,653	1,652	1,651	1,648	1,647	1,645
8	1,617	1,609	1,603	1,600	1,595	1,591	1,589	1,588	1,586	1,586	1,585	1,581	1,580	1,578
9	1,570	1,561	1,555	1,551	1,545	1,541	1,539	1,537	1,536	1,535	1,534	1,530	1,528	1,526
10	1,534	1,523	1,517	1,512	1,506	1,502	1,499	1,497	1,495	1,494	1,493	1,489	1,487	1,484
11	1,504	1,493	1,486	1,481	1,474	1,469	1,466	1,464	1,463	1,461	1,460	1,455	1,454	1,450
12	1,480	1,468	1,460	1,454	1,447	1,443	1,439	1,437	1,435	1,434	1,433	1,428	1,426	1,422
13	1,459	1,447	1,438	1,432	1,425	1,420	1,416	1,414	1,412	1,411	1,409	1,404	1,402	1,398
14	1,441	1,428	1,420	1,414	1,405	1,400	1,397	1,394	1,392	1,391	1,389	1,383	1,381	1,377
15	1,426	1,413	1,404	1,397	1,389	1,383	1,380	1,377	1,375	1,373	1,372	1,366	1,363	1,359
20	1,374	1,358	1,348	1,340	1,330	1,324	1,319	1,316	1,313	1,311	1,310	1,302	1,300	1,294
25	1,342	1,325	1,314	1,306	1,294	1,287	1,282	1,279	1,276	1,273	1,272	1,263	1,260	1,254
30	1,321	1,303	1,291	1,282	1,270	1,263	1,257	1,253	1,250	1,247	1,245	1,236	1,233	1,226
40	1,295	1,276	1,263	1,253	1,240	1,231	1,225	1,220	1,217	1,214	1,212	1,201	1,197	1,188
50	1,280	1,259	1,245	1,235	1,221	1,212	1,205	1,200	1,196	1,193	1,190	1,178	1,174	1,164
60	1,269	1,248	1,234	1,223	1,208	1,198	1,191	1,186	1,182	1,178	1,176	1,163	1,158	1,147
70	1,262	1,240	1,225	1,214	1,199	1,189	1,181	1,176	1,171	1,168	1,165	1,151	1,146	1,135
80	1,256	1,234	1,219	1,208	1,192	1,181	1,174	1,168	1,163	1,160	1,157	1,142	1,136	1,124
90	1,252	1,229	1,214	1,202	1,186	1,176	1,168	1,162	1,157	1,153	1,150	1,135	1,129	1,116
100	1,248	1,226	1,210	1,198	1,182	1,171	1,163	1,157	1,152	1,148	1,145	1,129	1,123	1,109
200	1,232	1,209	1,192	1,179	1,162	1,149	1,140	1,133	1,128	1,123	1,120	1,100	1,093	1,074
300	1,227	1,203	1,186	1,173	1,155	1,142	1,133	1,125	1,119	1,115	1,111	1,090	1,081	1,060
∞	1,216	1,191	1,174	1,160	1,140	1,127	1,116	1,108	1,102	1,096	1,091	1,066	1,054	1,000

Tabela dos quantis superiores da distribuição F ($\alpha = 0,10$)

$$P(X > x) = \int_x^{\infty} \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{t^{\frac{\nu_1 - 2}{2}}}{\left(1 + \frac{t\nu_1}{\nu_2}\right)^{\frac{\nu_1 + \nu_2}{2}}} dt = 0,10$$

ν_2	ν_1													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,20	60,47	60,71	60,90	61,07
2	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392	9,401	9,408	9,415	9,420
3	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230	5,222	5,216	5,210	5,205
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,907	3,896	3,886	3,878
5	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297	3,282	3,268	3,257	3,247
6	3,776	3,463	3,289	3,181	3,108	3,055	3,014	2,983	2,958	2,937	2,920	2,905	2,892	2,881
7	3,589	3,257	3,074	2,961	2,883	2,827	2,785	2,752	2,725	2,703	2,684	2,668	2,654	2,643
8	3,458	3,113	2,924	2,806	2,726	2,668	2,624	2,589	2,561	2,538	2,519	2,502	2,488	2,475
9	3,360	3,006	2,813	2,693	2,611	2,551	2,505	2,469	2,440	2,416	2,396	2,379	2,364	2,351
10	3,285	2,924	2,728	2,605	2,522	2,461	2,414	2,377	2,347	2,323	2,302	2,284	2,269	2,255
11	3,225	2,860	2,660	2,536	2,451	2,389	2,342	2,304	2,274	2,248	2,227	2,209	2,193	2,179
12	3,177	2,807	2,606	2,480	2,394	2,331	2,283	2,245	2,214	2,188	2,166	2,147	2,131	2,117
13	3,136	2,763	2,560	2,434	2,347	2,283	2,234	2,195	2,164	2,138	2,116	2,097	2,080	2,066
14	3,102	2,726	2,522	2,395	2,307	2,243	2,193	2,154	2,122	2,095	2,073	2,054	2,037	2,022
15	3,073	2,695	2,490	2,361	2,273	2,208	2,158	2,119	2,086	2,059	2,037	2,017	2,000	1,985
20	2,975	2,589	2,380	2,249	2,158	2,091	2,040	1,999	1,965	1,937	1,913	1,892	1,875	1,859
25	2,918	2,528	2,317	2,184	2,092	2,024	1,971	1,929	1,895	1,866	1,841	1,820	1,802	1,785
30	2,881	2,489	2,276	2,142	2,049	1,980	1,927	1,884	1,849	1,819	1,794	1,773	1,754	1,737
40	2,835	2,440	2,226	2,091	1,997	1,927	1,873	1,829	1,793	1,763	1,737	1,715	1,695	1,678
50	2,809	2,412	2,197	2,061	1,966	1,895	1,840	1,796	1,760	1,729	1,703	1,680	1,660	1,643
60	2,791	2,393	2,177	2,041	1,946	1,875	1,819	1,775	1,738	1,707	1,680	1,657	1,637	1,619
70	2,779	2,380	2,164	2,027	1,931	1,860	1,804	1,760	1,723	1,691	1,665	1,641	1,621	1,603
80	2,769	2,370	2,154	2,016	1,921	1,849	1,793	1,748	1,711	1,680	1,653	1,629	1,609	1,590
90	2,762	2,363	2,146	2,008	1,912	1,841	1,785	1,739	1,702	1,670	1,643	1,620	1,599	1,581
100	2,756	2,356	2,139	2,002	1,906	1,834	1,778	1,732	1,695	1,663	1,636	1,612	1,592	1,573
200	2,731	2,329	2,111	1,973	1,876	1,804	1,747	1,701	1,663	1,631	1,603	1,579	1,558	1,539
300	2,722	2,320	2,102	1,964	1,867	1,794	1,737	1,691	1,652	1,620	1,592	1,568	1,546	1,527
∞	2,706	2,303	2,084	1,945	1,847	1,774	1,717	1,670	1,632	1,599	1,570	1,546	1,524	1,505

continua...

v_2	v_1													
	15	20	25	30	40	50	60	70	80	90	100	200	300	∞
1	61,22	61,74	62,06	62,27	62,53	62,69	62,79	62,87	62,93	62,97	63,01	63,17	63,22	63,33
2	9,425	9,441	9,451	9,458	9,466	9,471	9,475	9,477	9,479	9,480	9,481	9,486	9,488	9,491
3	5,200	5,184	5,175	5,168	5,160	5,155	5,151	5,149	5,147	5,145	5,144	5,139	5,137	5,134
4	3,870	3,844	3,828	3,817	3,804	3,795	3,790	3,786	3,782	3,780	3,778	3,769	3,767	3,761
5	3,238	3,207	3,187	3,174	3,157	3,147	3,140	3,135	3,132	3,129	3,126	3,116	3,112	3,105
6	2,871	2,836	2,815	2,800	2,781	2,770	2,762	2,756	2,752	2,749	2,746	2,734	2,730	2,722
7	2,632	2,595	2,571	2,555	2,535	2,523	2,514	2,508	2,504	2,500	2,497	2,484	2,480	2,471
8	2,464	2,425	2,400	2,383	2,361	2,348	2,339	2,333	2,328	2,324	2,321	2,307	2,302	2,293
9	2,340	2,298	2,272	2,255	2,232	2,218	2,208	2,202	2,196	2,192	2,189	2,174	2,169	2,159
10	2,244	2,201	2,174	2,155	2,132	2,117	2,107	2,100	2,095	2,090	2,087	2,071	2,066	2,055
11	2,167	2,123	2,095	2,076	2,052	2,036	2,026	2,019	2,013	2,009	2,005	1,989	1,983	1,972
12	2,105	2,060	2,031	2,011	1,986	1,970	1,960	1,952	1,946	1,942	1,938	1,921	1,915	1,904
13	2,053	2,007	1,978	1,958	1,931	1,915	1,904	1,896	1,890	1,886	1,882	1,864	1,858	1,846
14	2,010	1,962	1,933	1,912	1,885	1,869	1,857	1,849	1,843	1,838	1,834	1,816	1,810	1,797
15	1,972	1,924	1,894	1,873	1,845	1,828	1,817	1,808	1,802	1,797	1,793	1,774	1,768	1,755
20	1,845	1,794	1,761	1,738	1,708	1,690	1,677	1,667	1,660	1,655	1,650	1,629	1,622	1,607
25	1,771	1,718	1,683	1,659	1,627	1,607	1,593	1,583	1,576	1,569	1,565	1,542	1,534	1,518
30	1,722	1,667	1,632	1,606	1,573	1,552	1,538	1,527	1,519	1,512	1,507	1,482	1,474	1,456
40	1,662	1,605	1,568	1,541	1,506	1,483	1,467	1,455	1,447	1,439	1,434	1,406	1,397	1,377
50	1,627	1,568	1,529	1,502	1,465	1,441	1,424	1,412	1,402	1,395	1,388	1,359	1,349	1,327
60	1,603	1,543	1,504	1,476	1,437	1,413	1,395	1,382	1,372	1,364	1,358	1,326	1,315	1,291
70	1,587	1,526	1,486	1,457	1,418	1,392	1,374	1,361	1,350	1,342	1,335	1,302	1,291	1,265
80	1,574	1,513	1,472	1,443	1,403	1,377	1,358	1,344	1,334	1,325	1,318	1,284	1,271	1,245
90	1,564	1,503	1,461	1,432	1,391	1,365	1,346	1,332	1,321	1,312	1,304	1,269	1,256	1,228
100	1,557	1,494	1,453	1,423	1,382	1,355	1,336	1,321	1,310	1,301	1,293	1,257	1,244	1,214
200	1,522	1,458	1,414	1,383	1,339	1,310	1,289	1,273	1,261	1,250	1,242	1,199	1,183	1,144
300	1,510	1,445	1,401	1,369	1,325	1,295	1,273	1,256	1,243	1,233	1,224	1,178	1,160	1,115
∞	1,487	1,421	1,375	1,342	1,295	1,263	1,240	1,222	1,207	1,195	1,185	1,130	1,106	1,000

Tabela dos quantis superiores da distribuição F ($\alpha = 0,05$)

$$P(X > x) = \int_x^{\infty} \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{t^{\frac{\nu_1 - 2}{2}}}{\left(1 + \frac{t\nu_1}{\nu_2}\right)^{\frac{\nu_1 + \nu_2}{2}}} dt = 0,05$$

ν_2	ν_1													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	161,4	199,5	215,7	224,5	230,1	233,9	236,7	238,8	240,5	241,8	242,9	243,9	244,6	245,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,39	19,40	19,41	19,41	19,42	19,42
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,75	8,73	8,72
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,936	5,912	5,891	5,873
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,704	4,678	4,655	4,636
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,027	4,000	3,976	3,956
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,603	3,575	3,550	3,529
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,313	3,284	3,259	3,237
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,102	3,073	3,048	3,025
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,943	2,913	2,887	2,865
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,818	2,788	2,761	2,739
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,717	2,687	2,660	2,637
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,635	2,604	2,577	2,554
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,565	2,534	2,507	2,484
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,507	2,475	2,448	2,424
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,310	2,278	2,250	2,225
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,236	2,198	2,165	2,136	2,111
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165	2,126	2,092	2,063	2,037
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,124	2,077	2,038	2,003	1,974	1,948
50	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,073	2,026	1,986	1,952	1,921	1,895
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993	1,952	1,917	1,887	1,860
70	3,978	3,128	2,736	2,503	2,346	2,231	2,143	2,074	2,017	1,969	1,928	1,893	1,863	1,836
80	3,960	3,111	2,719	2,486	2,329	2,214	2,126	2,056	1,999	1,951	1,910	1,875	1,845	1,817
90	3,947	3,098	2,706	2,473	2,316	2,201	2,113	2,043	1,986	1,938	1,897	1,861	1,830	1,803
100	3,936	3,087	2,696	2,463	2,305	2,191	2,103	2,032	1,975	1,927	1,886	1,850	1,819	1,792
200	3,888	3,041	2,650	2,417	2,259	2,144	2,056	1,985	1,927	1,878	1,837	1,801	1,769	1,742
300	3,873	3,026	2,635	2,402	2,244	2,129	2,040	1,969	1,911	1,862	1,821	1,785	1,753	1,725
∞	3,841	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880	1,831	1,789	1,752	1,720	1,692

continua...

v_2	v_1													
	15	20	25	30	40	50	60	70	80	90	100	200	300	∞
1	245,9	248,0	249,2	250,1	251,1	251,7	252,2	252,5	252,7	252,9	253,0	253,6	253,8	254,3
2	19,43	19,44	19,45	19,46	19,47	19,47	19,47	19,48	19,48	19,48	19,48	19,49	19,49	19,49
3	8,703	8,660	8,634	8,617	8,594	8,581	8,572	8,566	8,561	8,557	8,554	8,540	8,536	8,526
4	5,858	5,803	5,769	5,746	5,717	5,699	5,688	5,679	5,673	5,668	5,664	5,646	5,640	5,628
5	4,619	4,558	4,521	4,496	4,464	4,444	4,431	4,422	4,415	4,409	4,405	4,385	4,378	4,365
6	3,938	3,874	3,835	3,808	3,774	3,754	3,740	3,730	3,722	3,716	3,712	3,690	3,683	3,669
7	3,511	3,445	3,404	3,376	3,340	3,319	3,304	3,294	3,286	3,280	3,275	3,252	3,245	3,230
8	3,218	3,150	3,108	3,079	3,043	3,020	3,005	2,994	2,986	2,980	2,975	2,951	2,943	2,928
9	3,006	2,936	2,893	2,864	2,826	2,803	2,787	2,776	2,768	2,761	2,756	2,731	2,723	2,707
10	2,845	2,774	2,730	2,700	2,661	2,637	2,621	2,610	2,601	2,594	2,588	2,563	2,555	2,538
11	2,719	2,646	2,601	2,570	2,531	2,507	2,490	2,478	2,469	2,462	2,457	2,431	2,422	2,404
12	2,617	2,544	2,498	2,466	2,426	2,401	2,384	2,372	2,363	2,356	2,350	2,323	2,314	2,296
13	2,533	2,459	2,412	2,380	2,339	2,314	2,297	2,284	2,275	2,267	2,261	2,234	2,225	2,206
14	2,463	2,388	2,341	2,308	2,266	2,241	2,223	2,210	2,201	2,193	2,187	2,159	2,150	2,131
15	2,403	2,328	2,280	2,247	2,204	2,178	2,160	2,147	2,137	2,130	2,123	2,095	2,085	2,066
20	2,203	2,124	2,074	2,039	1,994	1,966	1,946	1,932	1,922	1,913	1,907	1,875	1,865	1,843
25	2,089	2,007	1,955	1,919	1,872	1,842	1,822	1,807	1,796	1,787	1,779	1,746	1,735	1,711
30	2,015	1,932	1,878	1,841	1,792	1,761	1,740	1,724	1,712	1,703	1,695	1,660	1,647	1,622
40	1,924	1,839	1,783	1,744	1,693	1,660	1,637	1,621	1,608	1,597	1,589	1,551	1,537	1,509
50	1,871	1,784	1,727	1,687	1,634	1,599	1,576	1,558	1,544	1,534	1,525	1,484	1,469	1,438
60	1,836	1,748	1,690	1,649	1,594	1,559	1,534	1,516	1,502	1,491	1,481	1,438	1,422	1,389
70	1,812	1,722	1,664	1,622	1,566	1,530	1,505	1,486	1,471	1,459	1,450	1,404	1,388	1,353
80	1,793	1,703	1,644	1,602	1,545	1,508	1,482	1,463	1,448	1,436	1,426	1,379	1,361	1,325
90	1,779	1,688	1,629	1,586	1,528	1,491	1,465	1,445	1,429	1,417	1,407	1,358	1,340	1,302
100	1,768	1,676	1,616	1,573	1,515	1,477	1,450	1,430	1,415	1,402	1,392	1,342	1,323	1,283
200	1,717	1,623	1,561	1,516	1,455	1,415	1,386	1,364	1,346	1,332	1,321	1,263	1,240	1,189
300	1,700	1,606	1,543	1,497	1,435	1,393	1,363	1,341	1,323	1,308	1,296	1,234	1,210	1,150
∞	1,666	1,571	1,506	1,459	1,394	1,350	1,318	1,293	1,273	1,257	1,243	1,170	1,138	1,000

Tabela dos quantis superiores da distribuição F ($\alpha = 0,025$)

$$P(X > x) = \int_x^{\infty} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{t^{\frac{\nu_1-2}{2}}}{\left(1 + \frac{t\nu_1}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}}} dt = 0,025$$

ν_2	ν_1													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	647,7	799,5	864,1	899,5	921,8	937,1	948,2	956,6	963,2	968,6	973,0	976,7	979,8	982,5
2	38,50	39,00	39,16	39,24	39,29	39,33	39,35	39,37	39,38	39,39	39,40	39,41	39,42	39,42
3	17,44	16,04	15,43	15,10	14,88	14,73	14,62	14,54	14,47	14,41	14,37	14,33	14,30	14,27
4	12,21	10,64	9,979	9,605	9,364	9,197	9,074	8,980	8,905	8,844	8,794	8,751	8,715	8,684
5	10,00	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,681	6,619	6,568	6,525	6,488	6,456
6	8,813	7,260	6,599	6,227	5,988	5,820	5,695	5,600	5,523	5,461	5,410	5,366	5,329	5,297
7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823	4,761	4,709	4,666	4,628	4,596
8	7,571	6,059	5,416	5,053	4,817	4,652	4,529	4,433	4,357	4,295	4,243	4,200	4,162	4,130
9	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026	3,964	3,912	3,868	3,831	3,798
10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779	3,717	3,665	3,621	3,583	3,550
11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588	3,526	3,474	3,430	3,392	3,359
12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,436	3,374	3,321	3,277	3,239	3,206
13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312	3,250	3,197	3,153	3,115	3,082
14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209	3,147	3,095	3,050	3,012	2,979
15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123	3,060	3,008	2,963	2,925	2,891
20	5,871	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837	2,774	2,721	2,676	2,637	2,603
25	5,686	4,291	3,694	3,353	3,129	2,969	2,848	2,753	2,677	2,613	2,560	2,515	2,476	2,441
30	5,568	4,182	3,589	3,250	3,026	2,867	2,746	2,651	2,575	2,511	2,458	2,412	2,372	2,338
40	5,424	4,051	3,463	3,126	2,904	2,744	2,624	2,529	2,452	2,388	2,334	2,288	2,248	2,213
50	5,340	3,975	3,390	3,054	2,833	2,674	2,553	2,458	2,381	2,317	2,263	2,216	2,176	2,140
60	5,286	3,925	3,343	3,008	2,786	2,627	2,507	2,412	2,334	2,270	2,216	2,169	2,129	2,093
70	5,247	3,890	3,309	2,975	2,754	2,595	2,474	2,379	2,302	2,237	2,183	2,136	2,095	2,059
80	5,218	3,864	3,284	2,950	2,730	2,571	2,450	2,355	2,277	2,213	2,158	2,111	2,071	2,035
90	5,196	3,844	3,265	2,932	2,711	2,552	2,432	2,336	2,259	2,194	2,140	2,092	2,051	2,015
100	5,179	3,828	3,250	2,917	2,696	2,537	2,417	2,321	2,244	2,179	2,124	2,077	2,036	2,000
200	5,100	3,758	3,182	2,850	2,630	2,472	2,351	2,256	2,178	2,113	2,058	2,010	1,969	1,932
300	5,075	3,735	3,160	2,829	2,609	2,451	2,330	2,234	2,156	2,091	2,036	1,988	1,947	1,910
∞	5,024	3,689	3,116	2,786	2,567	2,408	2,288	2,192	2,114	2,048	1,993	1,945	1,903	1,866

continua...

v_2	v_1													
	15	20	25	30	40	50	60	70	80	90	100	200	300	∞
1	984,8	993,1	998,0	1001	1005	1008	1009	1011	1011	1012	1013	1015	1016	1018
2	39,43	39,44	39,45	39,46	39,47	39,47	39,48	39,48	39,48	39,48	39,48	39,49	39,49	39,49
3	14,25	14,16	14,11	14,08	14,03	14,01	13,99	13,97	13,97	13,96	13,95	13,92	13,92	13,90
4	8,657	8,560	8,501	8,461	8,411	8,381	8,360	8,346	8,335	8,326	8,319	8,289	8,278	8,257
5	6,428	6,329	6,268	6,227	6,175	6,144	6,123	6,107	6,096	6,087	6,080	6,048	6,037	6,015
6	5,269	5,168	5,107	5,065	5,012	4,980	4,959	4,943	4,932	4,923	4,915	4,882	4,871	4,849
7	4,568	4,467	4,405	4,362	4,309	4,276	4,254	4,239	4,227	4,218	4,210	4,176	4,165	4,142
8	4,101	3,999	3,937	3,894	3,840	3,807	3,784	3,768	3,756	3,747	3,739	3,705	3,693	3,670
9	3,769	3,667	3,604	3,560	3,505	3,472	3,449	3,433	3,421	3,411	3,403	3,368	3,357	3,333
10	3,522	3,419	3,355	3,311	3,255	3,221	3,198	3,182	3,169	3,160	3,152	3,116	3,104	3,080
11	3,330	3,226	3,162	3,118	3,061	3,027	3,004	2,987	2,974	2,964	2,956	2,920	2,908	2,883
12	3,177	3,073	3,008	2,963	2,906	2,871	2,848	2,831	2,818	2,808	2,800	2,763	2,750	2,725
13	3,053	2,948	2,882	2,837	2,780	2,744	2,720	2,703	2,690	2,680	2,671	2,634	2,621	2,595
14	2,949	2,844	2,778	2,732	2,674	2,638	2,614	2,597	2,583	2,573	2,565	2,526	2,513	2,487
15	2,862	2,756	2,689	2,644	2,585	2,549	2,524	2,506	2,493	2,482	2,474	2,435	2,422	2,395
20	2,573	2,464	2,396	2,349	2,287	2,249	2,223	2,205	2,190	2,179	2,170	2,128	2,114	2,085
25	2,411	2,300	2,230	2,182	2,118	2,079	2,052	2,032	2,017	2,005	1,996	1,952	1,936	1,906
30	2,307	2,195	2,124	2,074	2,009	1,968	1,940	1,920	1,904	1,892	1,882	1,835	1,819	1,787
40	2,182	2,068	1,994	1,943	1,875	1,832	1,803	1,781	1,764	1,751	1,741	1,691	1,673	1,637
50	2,109	1,993	1,919	1,866	1,796	1,752	1,721	1,698	1,681	1,667	1,656	1,603	1,584	1,545
60	2,061	1,944	1,869	1,815	1,744	1,699	1,667	1,643	1,625	1,611	1,599	1,543	1,524	1,482
70	2,028	1,910	1,833	1,779	1,707	1,660	1,628	1,604	1,585	1,570	1,558	1,500	1,480	1,436
80	2,003	1,884	1,807	1,752	1,679	1,632	1,599	1,574	1,555	1,540	1,527	1,467	1,446	1,400
90	1,983	1,864	1,787	1,731	1,657	1,610	1,576	1,551	1,531	1,516	1,503	1,441	1,419	1,371
100	1,968	1,849	1,770	1,715	1,640	1,592	1,558	1,532	1,512	1,496	1,483	1,420	1,397	1,347
200	1,900	1,778	1,698	1,640	1,562	1,511	1,474	1,447	1,425	1,407	1,393	1,320	1,293	1,229
300	1,877	1,755	1,674	1,616	1,536	1,484	1,446	1,417	1,395	1,377	1,361	1,285	1,255	1,182
∞	1,833	1,708	1,626	1,566	1,484	1,428	1,388	1,357	1,333	1,313	1,296	1,205	1,166	1,000

Tabela dos quantis superiores da distribuição F ($\alpha = 0,01$)

$$P(X > x) = \int_x^{\infty} \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{t^{\frac{\nu_1 - 2}{2}}}{\left(1 + \frac{t\nu_1}{\nu_2}\right)^{\frac{\nu_1 + \nu_2}{2}}} dt = 0,01$$

ν_2	ν_1													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	4052	4999	5403	5624	5763	5858	5928	5981	6022	6055	6083	6106	6125	6142
2	98,50	99,00	99,16	99,24	99,29	99,33	99,35	99,37	99,38	99,39	99,40	99,41	99,42	99,42
3	34,11	30,81	29,45	28,71	28,23	27,91	27,67	27,48	27,34	27,22	27,13	27,05	26,98	26,92
4	21,19	18,00	16,69	15,97	15,52	15,20	14,97	14,79	14,65	14,54	14,45	14,37	14,30	14,24
5	16,25	13,27	12,06	11,39	10,96	10,67	10,45	10,28	10,15	10,05	9,963	9,888	9,825	9,770
6	13,74	10,92	9,780	9,148	8,746	8,466	8,260	8,102	7,976	7,874	7,790	7,718	7,657	7,605
7	12,24	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719	6,620	6,538	6,469	6,410	6,359
8	11,25	8,649	7,591	7,006	6,632	6,371	6,178	6,029	5,911	5,814	5,734	5,667	5,609	5,559
9	10,56	8,022	6,992	6,422	6,057	5,802	5,613	5,467	5,351	5,257	5,178	5,111	5,055	5,005
10	10,04	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,942	4,849	4,772	4,706	4,650	4,601
11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,744	4,632	4,539	4,462	4,397	4,342	4,293
12	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,388	4,296	4,220	4,155	4,100	4,052
13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,191	4,100	4,025	3,960	3,905	3,857
14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	4,030	3,939	3,864	3,800	3,745	3,698
15	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,004	3,895	3,805	3,730	3,666	3,612	3,564
20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457	3,368	3,294	3,231	3,177	3,130
25	7,770	5,568	4,675	4,177	3,855	3,627	3,457	3,324	3,217	3,129	3,056	2,993	2,939	2,892
30	7,562	5,390	4,510	4,018	3,699	3,473	3,304	3,173	3,067	2,979	2,906	2,843	2,789	2,742
40	7,314	5,179	4,313	3,828	3,514	3,291	3,124	2,993	2,888	2,801	2,727	2,665	2,611	2,563
50	7,171	5,057	4,199	3,720	3,408	3,186	3,020	2,890	2,785	2,698	2,625	2,562	2,508	2,461
60	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,718	2,632	2,559	2,496	2,442	2,394
70	7,011	4,922	4,074	3,600	3,291	3,071	2,906	2,777	2,672	2,585	2,512	2,450	2,395	2,348
80	6,963	4,881	4,036	3,563	3,255	3,036	2,871	2,742	2,637	2,551	2,478	2,415	2,361	2,313
90	6,925	4,849	4,007	3,535	3,228	3,009	2,845	2,715	2,611	2,524	2,451	2,389	2,334	2,286
100	6,895	4,824	3,984	3,513	3,206	2,988	2,823	2,694	2,590	2,503	2,430	2,368	2,313	2,265
200	6,763	4,713	3,881	3,414	3,110	2,893	2,730	2,601	2,497	2,411	2,338	2,275	2,220	2,172
300	6,720	4,677	3,848	3,382	3,079	2,862	2,699	2,571	2,467	2,380	2,307	2,244	2,190	2,142
Inf	6,635	4,605	3,782	3,319	3,017	2,802	2,639	2,511	2,407	2,321	2,248	2,185	2,130	2,082

continua...

ν_2	ν_1													
	15	20	25	30	40	50	60	70	80	90	100	200	300	∞
1	6157	6208	6239	6260	6286	6302	6313	6320	6326	6330	6334	6349	6355	6365
2	99,43	99,44	99,45	99,46	99,47	99,47	99,48	99,48	99,48	99,48	99,48	99,49	99,49	99,49
3	26,87	26,69	26,57	26,50	26,41	26,35	26,31	26,28	26,26	26,25	26,24	26,18	26,16	26,12
4	14,19	14,02	13,91	13,83	13,745	13,69	13,65	13,62	13,60	13,59	13,57	13,52	13,50	13,46
5	9,722	9,553	9,449	9,379	9,291	9,238	9,202	9,176	9,157	9,142	9,130	9,075	9,057	9,020
6	7,559	7,396	7,296	7,229	7,143	7,091	7,057	7,032	7,013	6,998	6,987	6,934	6,916	6,880
7	6,314	6,155	6,058	5,992	5,908	5,858	5,824	5,799	5,781	5,766	5,755	5,702	5,685	5,650
8	5,515	5,359	5,263	5,198	5,116	5,065	5,032	5,007	4,989	4,975	4,963	4,911	4,894	4,859
9	4,962	4,808	4,713	4,649	4,567	4,517	4,483	4,459	4,441	4,426	4,415	4,363	4,346	4,311
10	4,558	4,405	4,311	4,247	4,165	4,115	4,082	4,058	4,039	4,025	4,014	3,962	3,944	3,909
11	4,251	4,099	4,005	3,941	3,860	3,810	3,776	3,752	3,734	3,719	3,708	3,656	3,638	3,602
12	4,010	3,858	3,765	3,701	3,619	3,569	3,535	3,511	3,493	3,478	3,467	3,414	3,397	3,361
13	3,815	3,665	3,571	3,507	3,425	3,375	3,341	3,317	3,298	3,284	3,272	3,219	3,202	3,165
14	3,656	3,505	3,412	3,348	3,266	3,215	3,181	3,157	3,138	3,124	3,112	3,059	3,040	3,004
15	3,522	3,372	3,278	3,214	3,132	3,081	3,047	3,022	3,004	2,989	2,977	2,923	2,905	2,868
20	3,088	2,938	2,843	2,778	2,695	2,643	2,608	2,582	2,563	2,548	2,535	2,479	2,460	2,421
25	2,850	2,699	2,604	2,538	2,453	2,400	2,364	2,337	2,317	2,302	2,289	2,230	2,210	2,169
30	2,700	2,549	2,453	2,386	2,299	2,245	2,208	2,181	2,160	2,144	2,131	2,070	2,049	2,006
40	2,522	2,369	2,271	2,203	2,114	2,058	2,019	1,991	1,969	1,952	1,938	1,874	1,851	1,805
50	2,419	2,265	2,167	2,098	2,007	1,949	1,909	1,880	1,857	1,839	1,825	1,757	1,733	1,683
60	2,352	2,198	2,098	2,028	1,936	1,877	1,836	1,806	1,783	1,764	1,749	1,678	1,653	1,601
70	2,306	2,150	2,050	1,980	1,886	1,826	1,785	1,754	1,730	1,711	1,695	1,622	1,596	1,540
80	2,271	2,115	2,015	1,944	1,849	1,788	1,746	1,714	1,690	1,671	1,655	1,579	1,552	1,494
90	2,244	2,088	1,987	1,916	1,820	1,759	1,716	1,684	1,659	1,639	1,623	1,546	1,518	1,457
100	2,223	2,067	1,965	1,893	1,797	1,735	1,692	1,659	1,634	1,614	1,598	1,518	1,490	1,427
200	2,129	1,971	1,868	1,794	1,694	1,629	1,583	1,548	1,521	1,499	1,481	1,391	1,357	1,279
300	2,099	1,940	1,836	1,761	1,660	1,594	1,547	1,511	1,483	1,460	1,441	1,346	1,309	1,220
∞	2,039	1,878	1,773	1,696	1,592	1,523	1,473	1,435	1,404	1,379	1,358	1,247	1,200	1,000

Apêndice C - Coordenadas polares

Observando a Figura 14.1, podemos partitionar um plano usando coordenadas retangulares ou cartesianas, ou usando coordenadas polares. Se p é um ponto cujas coordenadas são (x, y) e as polares (r, θ) , então

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta. \end{aligned}$$

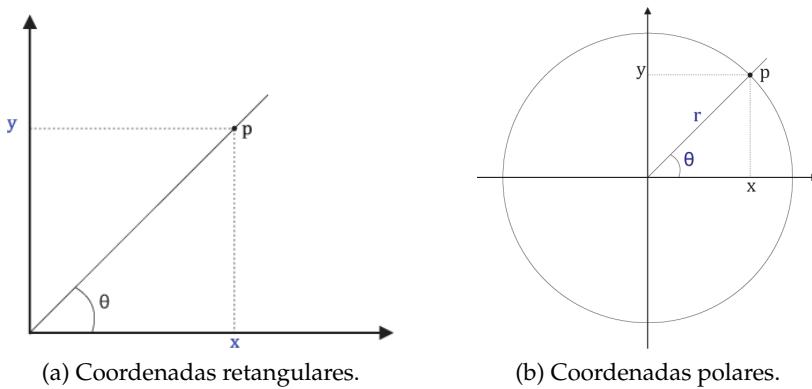


Figura 14.1: Coordenadas polares e retangulares.

A área de um círculo (A_c) de raio r como sabemos é

$$A_c = \pi r^2.$$

Se temos dois círculos concêntricos de raios r_1 e r_2 com $r_1 < r_2$, então a área da região R (chamada coroa circular) compreendida entre as duas circunferências, como visto na Figura 14.2, é:

$$A_R = \pi(r_2^2 - r_1^2).$$

A área de um setor circular de raio r e ângulo θ é definida por

$$A_S = \frac{\theta r^2}{2}.$$

Logo juntando, as Figuras 14.2a e 14.2b, teremos a área de um pedaço de abertura θ de uma coroa circular com raio menor r_1 e raio maior r_2 , como apresentado na Figura 14.3, definida por:

$$A_c = \frac{\theta(r_2^2 - r_1^2)}{2}.$$

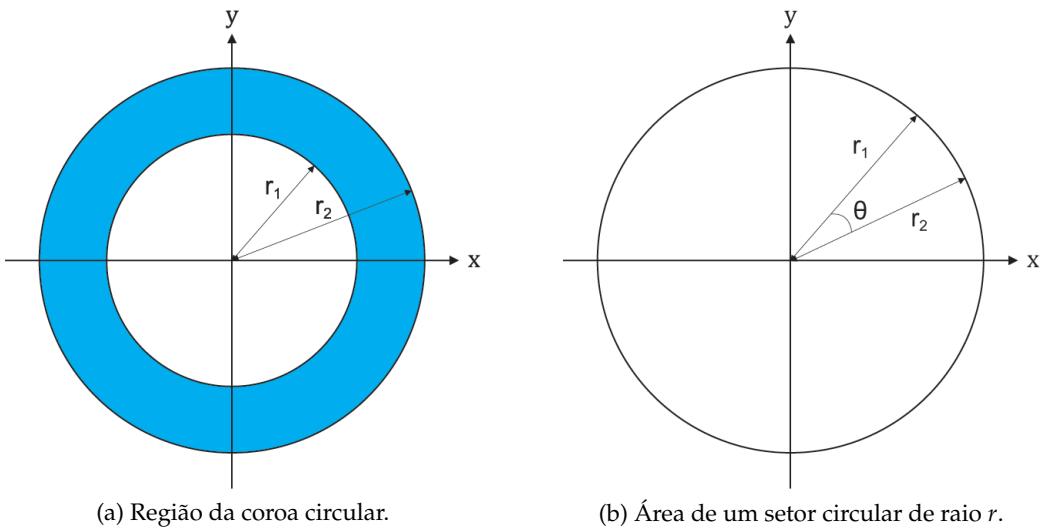
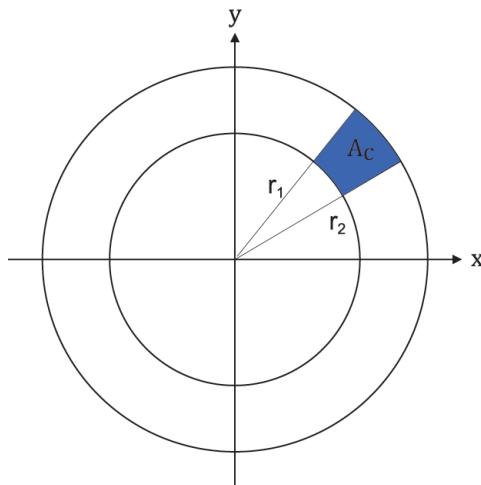
Figura 14.2: Região da coroa circular e área de um setor circular de raio r .

Figura 14.3: Área de um pedaço da coroa circular.

Dessa forma, a área da região no lado direito da Figura 14.3 é definida por:

$$\Delta A = \frac{\Delta\theta[(r_2^2 - r_1^2)]}{2}.$$

Sendo $\Delta r = (r_2 - r_1)$, em que

$$\begin{aligned} r_2^2 - r_1^2 &= (r_2 - r_1)(r_2 + r_1), \\ &= \Delta r(r_2 + r_1), \end{aligned}$$

e observando que $(r_1 + r_2)/2$ é o raio médio. Se passarmos à situação idealizada em que os dois raios são muito próximos, teremos $r_1 \approx r_2 \approx (r_1 + r_2)/2$, e assim

$$\begin{aligned} \Delta A &= \Delta\theta \left[\Delta r \frac{r_2 + r_1}{2} \right], \\ &= \Delta\theta \times \Delta r \times r. \end{aligned} \tag{14.1}$$

Assim,

$$\int \int_{\mathbb{R}} f(r, \theta) \Delta A = \lim_{\Delta \rightarrow 0} \sum \sum f(r, \theta) r \Delta\theta \Delta r.$$

Portanto, se o raio da região varia no intervalo $[a, b]$ e o ângulo $[\alpha_1, \alpha_2]$, então

$$\int \int_{\mathbb{R}} f(r, \theta) d\theta dr = \int_a^b \int_{\alpha_1}^{\alpha_2} f(r, \theta) r d\theta dr. \quad (14.2)$$

Gabarito dos Exercícios

Solução dos Exercícios do Capítulo 1

Solução do Exercício 1.1 na página 16:

Sabemos que no Brasil as barragens de mineração apresentam três tipos: barragem a montante, barragem a jusante e barragem linha de centro, dos quais esses tipos levam o custo e a segurança da barragem, porém todos eles depositam água juntamente com os rejeitos, e isto proporciona uma certa instabilidade para a barragem. Desse modo, podemos ter como problema inicial que a instabilidade das barragens podem estar relacionados aos tipos de barragens desenvolvidas no Brasil. Existe um outro tipo de barragem que pode ser desenvolvido a seco, porém pouco conhecido no Brasil, do qual o material acumulado é drenado e acumulado em pilhas, de modo a ficarem expostos à secagem do sol.

Assim, como estudo inicial sobre essa alternativa de construção de barragem, poderíamos nortear a pesquisa sob duas hipóteses a serem estudadas:

- Pesquisa 1: H_0 : Barragens do tipo depósito de rejeito a seco são mais baratas que as existentes no Brasil;
- Pesquisa 2: H_0 : Barragens do tipo depósito de rejeito a seco são mais seguras que as existentes no Brasil;

Solução do Exercício 1.2 na página 16:

As propriedades do somatório é parte inerente a quase todas as medidas estatísticas, como por exemplo, as medidas resumo, que compreende as medidas de posição e dispersão. Podemos observar nas propriedades (I), (III), (IV), do Teorema 1.1, que são fundamentais para as provas do Teorema 3.1 para as medidas de posição. As propriedades (V), (VII), (VIII) estão relacionadas as propriedades da variância, das quais algumas dessas estão no Teorema 4.2. Outras propriedades abordadas que envolvem duas variáveis no Teorema 1.1, podem ser utilizadas no Capítulo 11 quando formos estudar regressão e correlação linear. Claro, que as aplicações do somatório não se resumem apenas a esses resultados como consequências. Ao longo de todo livro, iremos abordar direta ou indiretamente essa técnica.

Solução do Exercício 1.3 na página 16:

Essas medidas não são iguais por diversos fatores, dentre elas, a imprecisão dos instrumentos de medida. Isso é o que define chamarmos essa característica de interesse no estudo de variável, pois de elemento a elemento, o valor assumido dessa característica varia. Outros aspectos dessa variação podem ser as diferentes condições ambientais referentes ao tempo de medida, dentre outros.

Solução do Exercício 1.4 na página 16:

- a) $\sum_{i=1}^4 X_i = 2 + 4 + 5 + 1 = 12;$
- b) $\sum_{i=1}^5 4 \times X_i^2 = 4 \times 2^2 + 4 \times 4^2 + \dots + 4 \times 1^2 = 200;$
- c) $\sum_{i=2}^5 X_i = 4 + 5 + 1 + 2 = 12;$
- d) $\sum_{i=1}^n X_i \times Y_i = 2 \times 1 + 4 \times 2 + \dots + 2 \times 8 = 46;$
- e) $\sum_{i=1}^n (3X_i + 2Y_i) = (3 \times 2 + 2 \times 1) + (3 \times 2 + 2 \times 2) + \dots + (3 \times 2 + 2 \times 8) = 80;$
- f) $\sum_{i=1}^n X_i Y_i + \sum_{i=1}^n Y_i^2 = (2 \times 1 + 4 \times 2 + \dots + 2 \times 8) + (1^2 + 2^2 + \dots + 8^2) = 149;$
- g) $\sum_{i=1}^n X_i = 2 + 4 + \dots + 2 = 14;$
- h) $\sum_{i=1}^n Y_i = 1 + 2 + \dots + 8 = 19;$
- i) $\sum_{i=1}^n X_i^2 = 2^2 + 4^2 + \dots + 2^2 = 50;$
- j) $\sum_{i=1}^n Y_i^2 = 1^2 + 2^2 + \dots + 8^2 = 103;$
- k) $\sum_{i=1}^n (Y_i X_i) = \sum_{i=1}^n Y_i \times X_i = 2 \times 1 + 4 \times 2 + \dots + 2 \times 8 = 46;$
- l) $(\sum_{i=1}^n Y_i)^2 = (1 + 2 + \dots + 8)^2 = 361;$
- m) $(\sum_{i=1}^n X_i)^2 = (2 + 4 + \dots + 2)^2 = 196;$
- n) Considerando que $\sum_{i=1}^n X_i / n = (2 + 4 + \dots + 2) / 5 = 2,8$ unid., então $\sum_{i=1}^n (X_i - 2,8)^2 = (2 - 2,8)^2 + (4 - 2,8)^2 + \dots + (2 - 2,8)^2 = 10,8$ unid.²;
- o) Considerando que $\sum_{i=1}^n Y_i / n = (1 + 2 + \dots + 8) / 5 = 3,8$ unid., então $\sum_{i=1}^n (Y_i - 3,8)^2 = (2 - 3,8)^2 + (4 - 3,8)^2 + \dots + (2 - 3,8)^2 = 30,8$ unid.²;
- p) $\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} = 2^2 + 4^2 + \dots + 2^2 - (2 + 4 + \dots + 2)^2 / 5 = 10,8$ unid.²;
- q) $\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} = 1^2 + 2^2 + \dots + 8^2 - (1 + 2 + \dots + 8)^2 / 5 = 30,8$ unid.²;
- r) (h) e (j), bem como (i) e (j), apresentam respostas iguais, respectivamente.

Solução do Exercício 1.5 na página 16:

- a) O Globo (Rio de Janeiro), Folha de São Paulo (São Paulo), Tribuna do Norte (Rio Grande do Norte), A crítica (Manaus), A Gazeta (São Paulo);
- b) NET, TIM, Brisanet, Vivo, Oi;
- c) Samara, Diego, Anna Albuquerque, Matheus, Giliwiline;
- d) 85, 75, 65, 42, 57.

Solução do Exercício 1.6 na página 17:

- a) A população em estudo representa todos os atendimentos a empresa telefônica A;
- b) Verificar se o tempo médio, em minutos, era superior a 10 minutos;

- c) Tempo de atendimento em minutos, e a variável é aleatória contínua;
- d) Não podemos identificar o tamanho da população com essas informações, mas o tamanho da amostra é possível, dado por $n = 10$.

Solução do Exercício 1.7 na página 17:

Pesquisa científica	Ramos da Estatística
Identificar o problema e o objetivo da pesquisa	-
Formular a hipótese estudada	Estatística Inferencial
Revisão de literatura	-
Formular um plano amostral e identificar as variáveis de interesse para a pesquisa	Amostragem (Não o classificamos com um ramo da estatística)
Coleta, crítica e tratamento dos dados	Estatística descritiva
Apresentação dos dados	Estatística descritiva
Análise e interpretação dos dados	Estatística inferencial e Probabilidade
Conclusão e derivação da conclusão que poderá rejeitar ou não a hipótese estudada, gerando assim, uma confirmação ou indagações para outros problemas	Estatística inferencial e Probabilidade
Apresentação dos dados por meio de trabalhos científicos para a propagação do conhecimento sobre o problema estudado	Estatística descritiva

Solução do Exercício 1.8 na página 17:

Devemos observar nessa tabela, a presença de algumas variáveis como: Região (Norte, Nordeste, Sudeste, Sul, Centro-Oeste e Brasil), Zona (Urbano e Rural), Estado Nutricional (Magreza, Adequado, Sobrepeso I e Sobrepeso II e III) e Número.

Vale lembrar que os valores internos da tabela representam a distribuição percentual de pessoas que se enquadram em um determinado estado nutricional, separados por região. Assim, considerando que temos todos os valores de IMC das pessoas entrevistadas (2.028 pessoas), e que esta é representada pela variável X , (a) a soma de todos os IMCs do Brasil, levando em consideração a todas as outras variáveis, pode ser representada por:

$$\sum_{i=1}^{2.028} X_i.$$

(b) Agora o total do IMC de homens do nordeste (586 pessoas), e que estes são representados pela variável Y , então

$$\sum_{j=1}^{586} Y_j \tag{15.1}$$

(c) Por fim, considerando o total do IMC de homens da zona urbana, podemos aproveitar a notação anterior, e complementar com um outro indexador, $k = 1, 2$; se $k = 1$ teremos uma pessoa da zona urbana, e se $k = 2$, teremos uma pessoa da zona rural, então como representação de soma geral teríamos

$$\sum_{j=1}^{586} Y_{jk}, \quad k = 1, 2. \tag{15.2}$$

As expressões (15.1) e (15.2) são equivalentes. Restringindo a soma apenas para homens da zona urbana, temos

$$\sum_{j=1}^{586} Y_{j1}.$$

Este último resultado significa que alguns valores de Y_{j1} serão iguais a 0, pois sabemos pela tabela que há apenas 267 homens da zona urbana.

Solução do Exercício 1.9 na página 18:

Fazendo:

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

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

$$\begin{aligned} D &= \sum_{i=1}^n (X_i - A)^2 = \sum_{i=1}^n (X_i^2 - 2AX_i + A^2) = \sum_{i=1}^n X_i^2 - \sum_{i=1}^n 2AX_i + \sum_{i=1}^n A^2 \\ \frac{dD}{dA} &= -2 \sum_{i=1}^n X_i + 2nA \end{aligned}$$

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

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

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

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

Certificando se o ponto é de máximo ou de mínimo,

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

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

Solução dos Exercícios do Capítulo 2**Solução do Exercício 2.1 na página 30:**

Como desejamos não incluir as condições limiares às indagações, podemos expressar as frequências acumuladas *abaixo de* e *acima de* como

$$F_{ac\downarrow_i} = \sum_{j=1}^{i-1} F_j, \quad i = 1, 2, \dots, k, \quad (15.3)$$

e

$$F_{ac\uparrow_i} = \sum_{j=i+1}^k F_j, \quad i = 1, 2, \dots, k, \quad (15.4)$$

respectivamente, sendo k o número de grupos ou classes.

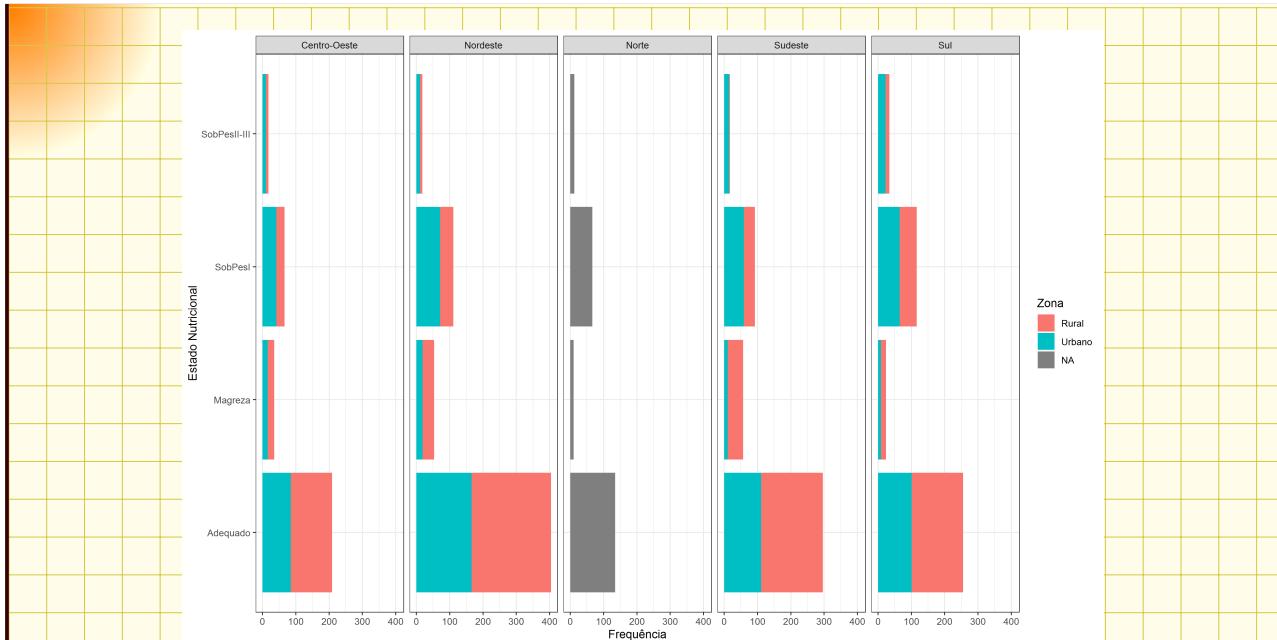
Solução do Exercício 2.2 na página 30:

Iremos apresentar os resultados para a distribuição de frequência do estado nutricional a nível de Brasil, sem especificar as regiões do país. Portanto, temos:

Zona	Estado Nutricional (X_{ij})	Frequência (F_{ij})
Urbano ($i = 1$)	Magreza ($j = 1$)	62
	Adequado ($j = 2$)	594
	Sobrepeso I ($j = 3$)	306
	Sobrepeso II e III ($j = 4$)	76
Rural ($i = 2$)	Magreza ($j = 1$)	116
	Adequado ($j = 2$)	710
	Sobrepeso I ($j = 3$)	140
	Sobrepeso II e III ($j = 4$)	24
Total		2.028

Solução do Exercício 2.3 na página 31:

Vamos representar esses dados por meio de um gráfico de barras, bem como o seu código em R, isto é,



E segue o script R:

Script:

```

1 # Pacotes
2 install.packages("ggplot2")
3
4 # Dados
5 dado <- read.table("ex2.2.csv", h = T, sep = ";")
6
7 # Anexando o pacote
8 library(ggplot2)
9
10 # Gerando o grafico
11 ggplot(dado) +
12   aes(x = estnut, fill = zona, weight = freq) +
13   geom_bar() +
14   scale_fill_hue() +
15   labs(x = "Estado_Nutricional", y = "Frequência", fill = "Zona") +
16   coord_flip() +
17   theme_bw() +
18   facet_grid(vars(), vars(regiao))

```

Solução do Exercício 2.4 na página 31:

Para tabularmos esses dados, temos que a variável é quantitativa contínua. Assim, tabularemos esses dados agrupando-os em intervalos de classe. Calculando o número de classes (k), inicialmente, temos:

$$k = \sqrt{n} = \sqrt{25} = 5 \text{ classes.}$$

Sabemos que a amplitude total é dada por A_t , e a amplitude de cada classe (c) é dada por:

$$c = \frac{A_t}{k-1} = \frac{59,4}{5-1} = 14,85.$$

Por fim, para o cálculo do limite inferior da primeira classe, temos:

$$LI_{1a} 9,6 - 14,85 / 2 = 2,175.$$

Ao final, basta seguir o algoritmo para o desenvolvimento da distribuição de frequências, abordado nesse capítulo, e assim, apresentamos a tabela de distribuição de frequências:

Classe		Fi	PM	Fr	Fac1	Fac2	Fp	Fac1p	Fac2p
1	2,17 — 17,02	6,00	9,59	0,24	6,00	25,00	24,00	24,00	100,00
2	17,02 — 31,88	10,00	24,45	0,40	16,00	19,00	40,00	64,00	76,00
3	31,88 — 46,73	6,00	39,30	0,24	22,00	9,00	24,00	88,00	36,00
4	46,73 — 61,58	2,00	54,16	0,08	24,00	3,00	8,00	96,00	12,00
5	61,58 — 76,42	1,00	69,00	0,04	25,00	1,00	4,00	100,00	4,00

Solução dos Exercícios do Capítulo 3

Solução do Exercício 3.1 na página 48:

Vamos inicialmente inserir na tabela de notas, a coluna dos pontos médios das classes, e a frequência acumulada abaixo de ($f_{ac\downarrow i}$), do qual segue:

Notas	F_i	\tilde{X}_i	$F_{ac\downarrow i}$
0 — 2	4	1	4
2 — 4	12	3	16
4 — 6	15	5	31
6 — 8	13	7	44
8 — 10	6	9	50

- Média

$$\bar{X} = \frac{1 \times 4 + 3 \times 12 + \dots + 9 \times 6}{50} = 5,2 \text{ pontos},$$

em que a notas dos estudantes se distribuem em torno desse valor;

- Mediana

$$Md(X) = 4 + \frac{25 - 16}{15} \times 2 = 5,2 \text{ pontos},$$

em que a distribuição das notas dos estudantes abaixo e acima de 5,2 pontos, representam 50% dos dados;

- Moda

$$Mo(X) = 4 + \frac{3}{3+2} \times 2 = 5,2 \text{ pontos},$$

em que esta representa a nota de maior frequência.

Solução do Exercício 3.2 na página 48:

Usando o fato de que $N = n$, então $\mu = \bar{X}$. Essa soma de quadrados é minimizada quando o desvio é em relação a media amostral. Logo a primeira soma é a menor, desde que $\bar{X} \neq \mu$.

Solução do Exercício 3.3 na página 48:

Para as duas transformações X e Y , dada uma amostra fixada, as medidas \bar{X} e S podem ser consideradas como uma constante, pois elas não se alteram. Então, usando a propriedade (II) para os Teoremas 3.1, 3.2 e 3.3, temos que respectivamente, a média, a mediana e a moda de Y , são iguais a 0. Usando esses mesmos teoremas para o cálculo da média, mediana e moda de Z , precisamos redefinir essa variável como $Z_i = X_i/S + \bar{X}/S$. Assim, para X_i/S basta considerarmos que X_i está multiplicado por uma constante $k_1 = 1/S$ e daí usamos a propriedade (I) dos teoremas citados, acrescido de outra constante $k_2 = \bar{X}/S$, do qual usamos a propriedade (II) desses mesmos teoremas, logo, a média, mediana e moda serão também iguais a 0. Para que haja um entendimento mais claro, faremos a prova apenas para a média nas duas transformações, e para as demais medidas, deixaremos como exercício para os leitores.

- Média de Y

$$\begin{aligned}\bar{Y} &= \sum_{i=1}^n \frac{Y_i}{n} \\ &= \sum_{i=1}^n \frac{X_i - \bar{X}}{n} \\ &= \sum_{i=1}^n \frac{X_i}{n} - \bar{X} \\ &= \bar{X} - \bar{X} = 0;\end{aligned}$$

- Média de Z

$$\begin{aligned}\bar{Z} &= \sum_{i=1}^n \frac{Z_i}{n} \\ &= \sum_{i=1}^n \frac{X_i \times k_1 - k_2}{n} \\ &= \sum_{i=1}^n \frac{X_i \times k_1}{n} - k_2 \\ &= k_1 \sum_{i=1}^n \frac{X_i}{n} - k_2 \\ &= k_1 \times \bar{X} - k_2 \\ &= \bar{X}/S - \bar{X}/S, \quad k_1 = 1/S \text{ e } k_2 = \bar{X}/S \\ &= 0.\end{aligned}$$

Solução do Exercício 3.4 na página 48:

Considere o salário dos funcionários representado pela variável X , cuja média salarial é dada por \bar{X} . Considere o reajuste salarial representado pela variável Y , dada por:

$$\begin{aligned}Y &= X + 0,10X \\ &= 1,1X.\end{aligned}$$

Assim, para calcularmos a média reajustada, isto é, \bar{Y} , basta usarmos o resultado da propriedade (II) do Teorema 3.1, isto é, $\bar{Y} = 1,1\bar{X} = 1,1 \times 1.500 = 1.650$ reais.

Solução do Exercício 3.5 na página 48:

Considere que a nota dos alunos podem ser representados pela variável X_i , com $i = 1, 2, \dots, n$, em que n representa o número de alunos que fizeram a prova. Considere ainda o índice H na variável X_{Hj} , representa o aluno j do sexo masculino que obteve nota X_{Hj} , tal que $j = 1, 2, \dots, n_H$, sendo n_H o número de rapazes que fizeram a prova. E de modo similar, o índice M na variável X_{Mk} , representa a aluna k do sexo feminino que obteve nota X_{Mj} , tal que $k = 1, 2, \dots, n_M$, sendo n_M o número de moças que fizeram a prova. Dessa forma a proporção

de mulheres que fizeram a prova é:

$$\sum_{i=1}^n X_i = \sum_{k=1}^{n_M} X_k + \sum_{j=1}^{n_H} X_j$$

$$n\bar{X} = n_M\bar{X}_M + n_H\bar{X}_H$$

$$5,8n = 6,3n_M + 4,3n_H \quad (n = n_H + n_M)$$

$$5,8n = 6,3n_M + 4,3(n - n_M)$$

$$5,8n - 4,3n = 6,3n_M - 4,3n_M$$

$$0,75n = n_M.$$

Portanto, há 75% dos alunos que realizaram a prova, eram mulheres.

Solução do Exercício 3.6 na página 48:

Essa questão limita a menor nota tirada nessa prova, bem como o número de alunos com a nota desejada. Dessa forma, o valor máximo para que tenhamos uma quantidade n_X de alunos que retiraram nota 90 pontos é considerar que os demais ($n - n_X$) retiraram o valor mínimo que é 60 pontos. Caso algum dos ($n - n_X$) alunos tenham retirado nota acima de 60 pontos, com certeza o número de alunos que retirou 90 pontos será menor que n_X . Assim, o valor de n_X pode ser encontrado da seguinte forma:

$$\begin{aligned} \frac{\sum_{i=1}^{30} X_i}{30} &\leq 70 \\ \frac{60(30 - n_X) + 90n_X}{30} &\leq 70 \\ -60n_X + 90n_X &\leq 2100 - 1800 \\ n_X &\leq 300/30 = 10. \end{aligned}$$

Logo, o número máximo de alunos que tirou 90 pontos foi $n_H \leq 10$.

Solução dos Exercícios do Capítulo 4**Solução do Exercício 4.1 na página 64:**

I) Considerando uma amostra X_1, X_2, \dots, X_n , assumimos que \bar{X} e S_X são a média e o desvio padrão. Dado a transformação $Y_i = X_i \pm c$, para $i = 1, 2, \dots, n$ e c uma constante, podemos afirmar, pelos Teoremas 3.1 e 4.3, respectivamente, que a média e o desvio padrão de Y_i , podem ser dadas por $\bar{Y} = \bar{X} \pm c$ e $S_Y = S_X$, respectivamente. Dessa forma, sabendo que o coeficiente de variação de X_i é dado por $CV_X = S_X/\bar{X}$, então o coeficiente de variação de Y_i é:

$$CV_Y = \frac{S_Y}{\bar{Y}}$$

Solução do Exercício 4.2 na página 64:**Solução do Exercício 4.3 na página 64:****Solução do Exercício 4.4 na página 64:****Solução do Exercício 4.5 na página 64:****Solução do Exercício 4.6 na página 65:****Solução do Exercício 4.7 na página 65:**

Solução dos Exercícios do Capítulo 5

Solução do Exercício 5.1 na página 99:

Verifique a Definição 5.1 e formule uma resposta.

Solução do Exercício 5.2 na página 99:

Verifique as Definições 5.2 e 5.3 e formule uma resposta.

Solução do Exercício 5.3 na página 99:

- Consideremos C_i e K_i , as faces superiores cara e coroa, respectivamente, para as moedas $i = 1, 2$. Assim, o espaço amostral é dado por $\Omega = \{(C_1, C_2), (C_1, K_2), (K_1, C_2), (K_1, K_2)\}$ e $\#\Omega = 4$;
- Segue o resultado para o dado:

Dado	1	2	3	4	5	6
Resultado	Ímpar	Par	Ímpar	Par	Ímpar	Par

Considerando que foi lançado duas vezes o dado, cada elemento é um par de resultados, e ainda, mesmo alguns resultados representando elementos iguais, como por exemplo, $\{(Par, Par)\} = \{(2, 2), (2, 4), (2, 6), \dots\}$, no espaço amostral só teremos o elemento (Par, Par) . Assim, o espaço amostral é dado por $\Omega = \{(Par, Par), (Par, Ímpar), (Ímpar, Par), (Ímpar, Ímpar)\}$, com $\#\Omega = 4$;

- O espaço amostral para a somas das faces superiores de dois dados lançados é $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, com cardinalidade $\#\Omega = 11$;
- O espaço amostral para esse experimento aleatório, considerando F - pessoa do sexo feminino e M - pessoa do sexo masculino, é:

$$\Omega = \{(F, F, F), (F, F, M), (F, M, F), (M, F, F), (F, M, M), \\ (M, M, F), (M, F, M), (M, M, M)\},$$

com cardinalidade $\#\Omega = 8$.

Solução do Exercício 5.4 na página 99:

Essa questão exige um conhecimento avançado de progressão aritmética. Vejamos, inicia-se o jogo com A, posteriormente, B. Enquanto ninguém ganhar o jogo prossegue. Dessa forma, as jogadas em que A pode ganhar são $1^{\text{a}}, 3^{\text{a}}, 5^{\text{a}}, 7^{\text{a}}, \dots$, e assim por diante. As jogadas em que B pode ganhar são $2^{\text{a}}, 4^{\text{a}}, 6^{\text{a}}, \dots$, e assim por diante.

- Vejamos, as chances de A ganhar:

Jogada	Situação	Probabilidade
1^{a}	3 moedas caras ou 3 coroas	$2 \times (1/2)^3 = 1/4$
3^{a}	A perder, B perder e A ganhar	$(3/4)^2 \times 1/4$
5^{a}	A perder, B perder, A perder, B perder e A ganhar	$(1/3)^4 \times 1/4$
:	:	:

Percebe-se que a sequência $1/4, 1/4[(3/4)^2]^1, 1/4[(3/4)^2]^2, 1/4[(3/4)^2]^3, \dots$, é uma progressão geométrica (PG) em que o primeiro elemento é $a_1 = 1/4$ e a razão é $q = (3/4)^2$.

Pode-se mostrar que a soma infinita de uma PG é dada por:

$$S = \frac{a_1}{1 - q}, \quad (15.5)$$

que representa a probabilidade de A ganhar, isto é,

$$P(\text{A ganhar}) = \frac{1/4}{1 - (3/4)^2} = 4/7$$

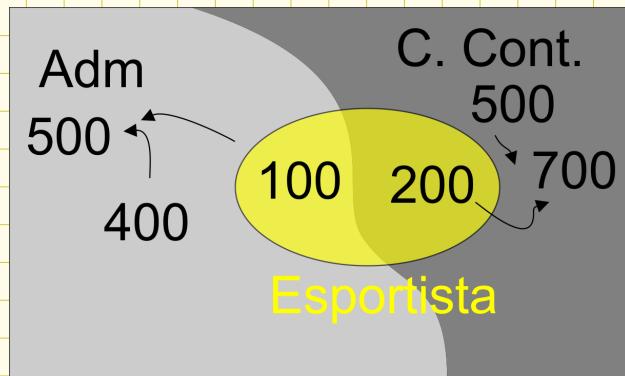
b) As chances de B ganhar são:

Jogada	Situação	Probabilidade
2 ^a	A perder e B ganhar	$(1 - 1/4) \cdot 1/4 = 3/4 \times 1/4$
4 ^a	A perder, B perder, A perder e B ganhar	$(3/4)^3 \times 1/4$
6 ^a	A perder, B perder, A perder, B perder, A perder e B ganhar	$(3/4)^5 \times 1/4$
:	:	:

A série pode ser expressa como: $3/16, 3/16 \times [(3/4)^2]^1, 3/16 \times [(3/4)^2]^2, \dots$, e assim por diante. Pelo mesmo raciocínio feito na probabilidade de A ganhar, temos que a probabilidade de B ganhar é:

$$P(\text{B ganhar}) = \frac{3/16}{1 - (3/4)^2} = \frac{3}{7}$$

Solução do Exercício 5.5 na página 99:



- a) E - Evento ser Esportista, logo $P(E) = 4.000/10.000 = 0,40$;
- b) EA - Esportista e aluno da Administração, logo $P(EA) = 100/10.000$;
- c) E - Evento ser esportista; C - Evento ser da C. Contábeis, logo, $P(E \cup C) = P(E) + P(C) - P(E \cap C) = 4.000/10.000 + 700/10.000 - 200/10000 = 4.500/10.000$.
- d) E - Evento ser esportista; A - Evento ser da administração, como $P(E \cup A) = P(E) + P(A) - P(E \cap A) = 4.000/10.000 + 500/10.000 - 100/10.000 = 4.400/10.000$, logo $P(E \cup A)^c = 1 - 4.400/10.000$.

Solução do Exercício 5.6 na página 99:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \Rightarrow P(B) = P(A \cup B) + P(A \cap B) - P(A) = 0,5 + 0,1 - 0,2 = 0,4$$

Solução do Exercício 5.7 na página 100:

- a) $P(A^c) = 1 - \frac{1}{2} = \frac{1}{2}$;
 b) $P(B^c) = 1 - \frac{1}{4} = \frac{3}{4}$;
 c) $P(A \cap B) = 0$;
 d) $P(A \cup B) = P(A) + P(B) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$;
 e) $P(A^c \cap B^c) = P[(A \cup B)^c] = 1 - \frac{3}{4} = \frac{1}{4}$.

Solução do Exercício 5.8 na página 100:

- a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{3} = \frac{1}{2}$;
 b) $P(A^c \cup B) = P(A^c) + P(B) - P(A^c \cap B) = P(A^c) + P(B) = \frac{5}{6}$, uma vez que $P(A^c \cap B) = 0$,
 pois $B \subset A$ (já que $P(A^c \cap B) = P(B)$);
 c) $P(A^c \cap B^c) = P[(A \cup B)^c] = 1 - P(A \cup B) = 1 - \frac{1}{2} = \frac{1}{2}$;

Solução do Exercício 5.9 na página 100:

Dado 1	1	2	3	4	5	6
Dado 2	1	2	3	4	5	6

- a) A - Evento: soma ser menor que 4, logo $P(A) = \frac{\#\{\text{Soma ser menor que 4}\}}{36} = \frac{3}{36}$;
 b) B - Evento: soma ser 9, logo $P(B) = \frac{\#\{\text{Soma ser 9}\}}{36} = \frac{4}{36}$
 c) C - Evento: o primeiro resultado ser maior que o segundo, logo $P(C) = \frac{\#\{\text{1º Resultado ser maior que o 2º Resultado}\}}{36} = \frac{15}{36}$.

Solução do Exercício 5.10 na página 100:

Considerando que um baralho tem 52 cartas, 4 cartas rei, e 13 cartas com naipe copas, sendo 1 rei de copas, e ainda R: carta rei, C: Carta de copas, logo

$$P(R \cup C) = P(R) + P(C) - P(R \cap C) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}.$$

Solução do Exercício 5.11 na página 100:

Evento J_1 : o jogador 1 marca o pênalti; Evento J_2 : o jogador 2 marca o pênalti; Evento J_3 : o jogador 3 marca o pênalti. Assim,

a) Todos acertarem:

$$P(J_1 \cap J_2 \cap J_3) = P(J_1) \times P(J_2) \times P(J_3) = \frac{2}{3} \times \frac{4}{5} \times \frac{7}{10} = \frac{28}{75}.$$

b) Apenas um acertar:

$$\begin{aligned} & P(J_1 \cap J_2^c \cap J_3^c) + P(J_1^c \cap J_2 \cap J_3^c) + P(J_1^c \cap J_2^c \cap J_3) \\ & \left(\frac{2}{3} \times \frac{1}{5} \times \frac{3}{10} \right) + \left(\frac{1}{3} \times \frac{4}{5} \times \frac{3}{10} \right) + \left(\frac{1}{3} \times \frac{1}{5} \times \frac{7}{10} \right) \\ & \frac{1}{25} + \frac{2}{25} + \frac{7}{150} \\ & \frac{1}{6}. \end{aligned}$$

c) Todos errarem:

$$P[J_1^c \cap J_2^c \cap J_3^c] = \frac{1}{3} \times \frac{1}{5} \times \frac{3}{10} = \frac{1}{50}.$$

Solução do Exercício 5.12 na página 100:

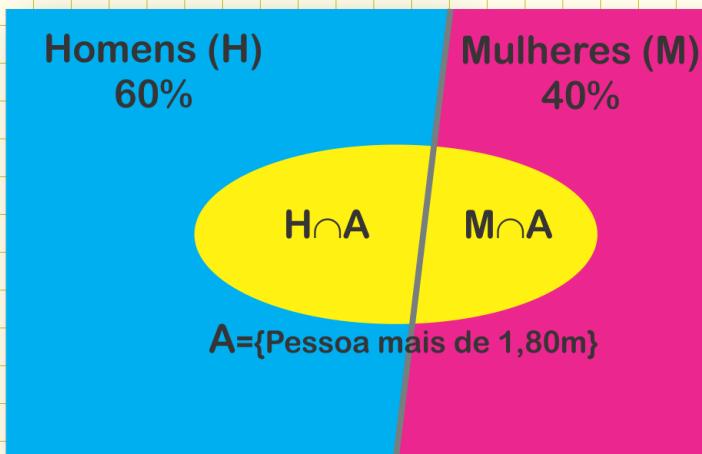
Considerando os eventos V : bola de volei; B : bola de basquete; A_1 : bola no armário 1; A_2 : bola no armário 2, assim

$$\text{a)} P(V|A_1) = \frac{P(V \cap A_1)}{P(A_1)} = \frac{3/9}{4/9} = \frac{3}{4}.$$

$$\text{b)} P(B|A_2) = \frac{P(B \cap A_2)}{P(A_2)} = \frac{2/9}{5/9} = \frac{2}{5}.$$

c) Essa questão nos remete uma atenção. O enunciado está desejando calcular a probabilidade de escolher um armário e em seguida uma de suas bolas, sendo esta de basquete. Como neste ítem não foi identificado qual armário seria o escolhido, vamos identificar o evento A_1^* que representa a escolha do armário 1, e o evento A_2^* que representa a escolha do armário 2. Estes eventos são diferentes dos eventos A_1 e A_2 , respectivamente, pois estes últimos representam os eventos de escolher as bolas nestes respectivos armários e não a escolha do armário. Logo, a chance de escolhermos um dos armários é 50%, isto é, $P(A_1^*) = P(A_2^*) = 1/2$. Dessa forma, a probabilidade de escolhermos um armário e uma bola de basquete pode ser dada por:

$$\begin{aligned} P(B) &= P(A_1^*) \times P(B|A_1) + P(A_2^*) \times P(B|A_2) \\ &= 1/2 \times 1/4 + 1/2 \times 2/5 \\ &= \frac{13}{40} = 0,3250. \end{aligned}$$

Solução do Exercício 5.13 na página 101:

Seja A o evento do estudante ter mais de 1,80m, e considere M o evento do estudante ser mulher. Temos as seguintes informações disponíveis, $P(H) = 0,60$, $P(M) = 0,40$, $P(A|H) = 0,05$ e $P(A|M) = 0,02$. Para calcularmos o evento A , temos

$$\begin{aligned} P(A) &= P(A \cap M) + P(A \cap H) \\ &= P(A|M)P(M) + P(A|H)P(H) \\ &= 0,02 \times 0,40 + 0,05 \times 0,6 \\ &= 0,038. \end{aligned}$$

Assim, para calcularmos $P(M|A)$ temos

$$\begin{aligned}P(M|A) &= \frac{P(M \cap A)}{P(A)} \\&= \frac{0,008}{0,038} = 0,2105.\end{aligned}$$

Solução do Exercício 5.14 na página 101:

Considere o evento H - homem está vivo; e o evento M - mulher está viva. Assim,

a) $P(H \cap M^c) = P(H) \times P(M^c) = \frac{3}{5} \times \frac{1}{4} = \frac{3}{20}$.

b) $P(H^c \cap M) = P(H^c) \times P(M) = \frac{2}{5} \times \frac{3}{4} = \frac{6}{20}$.

c) $P(H \cap M) = P(H) \times P(M) = \frac{3}{5} \times \frac{3}{4} = \frac{9}{20}$.

Solução do Exercício 5.15 na página 101:

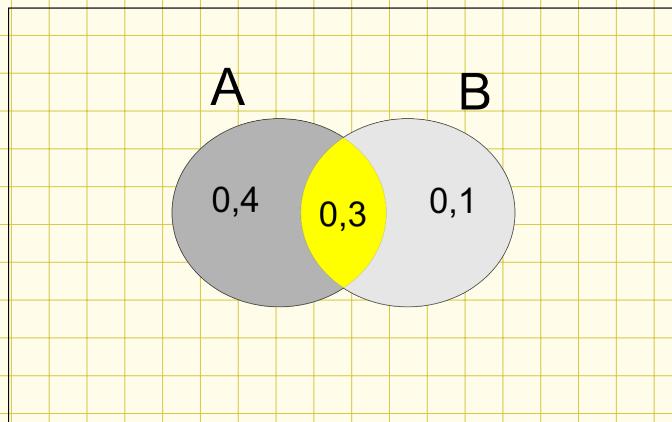
a) $P(A \cup B) = P(A) + P(B) \Rightarrow P(B) = 0,8 - 0,5 = 0,3$.

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Se A e B são independentes, então $P(A \cap B) = P(A) \times P(B)$. Logo,

$$\begin{aligned}P(A \cup B) &= P(A) + P(B) - P(A)P(B) \\&= P(A) + P(B)[1 - P(A)] \\&= P(A) + P(B)P(A^c) \\0,8 &= 0,5 + P(B) \times 0,5,\end{aligned}$$

que resulta em $P(B) = \frac{0,8-0,5}{0,5} = \frac{0,3}{0,5} = 0,60$.

Solução do Exercício 5.16 na página 101:



Considerando que $P(B^c) = 1 - P(B) = 0,6$, então

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{0,4}{0,6} = 0,6667.$$

Solução do Exercício 5.17 na página 101:

Problema resolvido pelo teorema de Bayes. Consideraremos o evento G o evento das vitórias

do São Paulo, e X o evento de chover. Assim,

$$\begin{aligned} P(X|G) &= \frac{P(G|X)P(X)}{P(G|X)P(X) + P(G|X^c)P(X^c)} \\ &= \frac{0,7 \times 0,3}{0,7 \times 0,3 + 0,8 \times 0,7} \\ &= \frac{0,21}{0,77} = 0,2727. \end{aligned}$$

Solução do Exercício 5.18 na página 101:

Considerando D o evento defeito, temos

$$\begin{aligned} P(B|D) &= \frac{P(D|B)P(B)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)} \\ &= \frac{0,05 \times 0,50}{0,03 \times 0,30 + 0,05 \times 0,50 + 0,02 \times 0,20} \\ &= \frac{0,0250}{0,038} = 0,6579. \end{aligned}$$

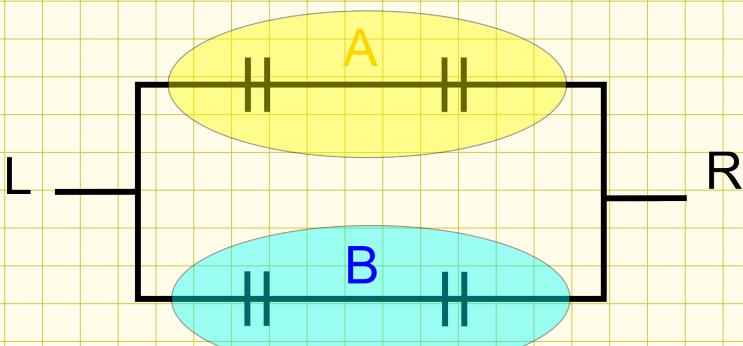
Solução do Exercício 5.19 na página 101:

Seja T o evento de gostar de teatro, e B o evento de gostar de cinema. Assim, temos que $T = (T \cap B) \cup (T \cap B^c)$.

- a) Como $P(T \cap B) = 0$, então $P(T \cap B^c) = P(T) = 1/3$.
- b) $P(T \cap B^c) = P(T) - P(T)P(B) = 1/3 - 1/3 \times 1/2 = 1/6$.
- c) Nesse caso, $T \subset B$, logo $T \cap B^c = \emptyset$. Assim, $P(T \cap B^c) = 0$.
- d) $P(T \cap B^c) = P(T) - P(T \cap B) = 1/3 - 1/8 = 5/24$.
- e) Sabemos que $P(T^c|B^c) = 3/4$, e que $P(T^c|B^c) + P(T|B^c) = 1$, então $P(T|B^c) = 1/4$. Sabemos também que

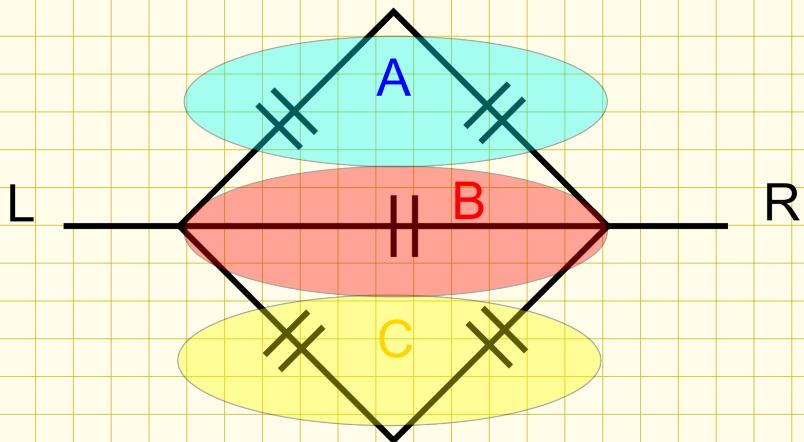
$$\begin{aligned} P(T \cap B^c) &= P(T|B^c) \times P(B^c) \\ &= P(T|B^c) \times [1 - P(B)] \\ &= 1/4(1 - 1/2) \\ &= 1/8. \end{aligned}$$

Solução do Exercício 5.20 na página 102:



$$\begin{aligned}
 P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
 &= p^2 + p^2 - p^4 \\
 &= 2p^2 - p^4.
 \end{aligned}$$

Solução do Exercício 5.21 na página 102:



$$\begin{aligned}
 P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + \\
 &\quad + P(A \cap B \cap C) \\
 &= p + p + p - p^2 - p^2 - p^2 + p^3 \\
 &= 3p - 3p^2 + p^3.
 \end{aligned}$$

Solução do Exercício 5.22 na página 102:

$$F_X(x) = \begin{cases} 0, & \text{se } x < 0, \\ \frac{157.267.837}{160.044.908}, & \text{se } 0 \leq x < 1, \\ \frac{158.093.995}{160.044.908}, & \text{se } 1 \leq x < 2, \\ 1, & \text{se } x \geq 2. \end{cases}$$

Solução do Exercício 5.23 na página 103:

Considere a função de probabilidade:

x	5	7	8	20
$p_X(x)$	0,2	0,3	0,4	0,1

Assim, temos que:

- $P(X \geq 7) = P(X = 5) + P(X = 7) = 0,2 + 0,3 = 0,5$;
- $P(X < 7) = P(X = 5) = 0,2$;
- $P(8 \leq X \leq 18) = P(X = 8) = 0,4$;

d) $P(X \geq 15) = P(X = 20) = 0,1;$

e) A esperança é calculada da seguinte forma:

$$E[X] = 5 \times 0,2 + 7 \times 0,3 + 8 \times 0,4 + 20 \times 0,1 = 8,3 \text{ unid.},$$

a mediana é um número entre 7 e 8, pois abaixo e acima desses respectivos números temos 50%, e por conveniência assumimos $\mu_d = (7+8)/2 = 7,5$ unid., isto é, o ponto médio entre os dois valores centrais, e por fim, a moda é o valor de maior chance, isto é, $\mu_0 = 8$ unid..

Solução do Exercício 5.24 na página 103:

Seja

$$P(A) = P(A \cap B) + P(A \cap B^c) \quad (15.6)$$

e seja

$$P(B) = P(A \cap B) + P(B \cap A^c) \quad (15.7)$$

Considere ainda que os eventos $A \cap B^c$ e $A^c \cap B$ são disjuntos. Assim,

$$P[(A \cap B^c) \cup (A^c \cap B)] = P(A \cap B^c) + P(A^c \cap B) \quad (15.8)$$

Substituindo (15.6) e (15.7) em (15.8), temos

$$\begin{aligned} P[(A \cap B^c) \cup (A^c \cap B)] &= [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)] \\ &= P(A) + P(B) - 2P(A \cap B), \end{aligned}$$

o que prova o resultado.

Solução dos Exercícios do Capítulo 6

Solução do Exercício 6.1 na página 135:

Verifique a Definição 5.22 e formule sua resposta.

Solução do Exercício 6.2 na página 135:

Distribuição Binomial:

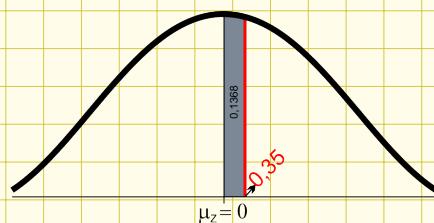
$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Considerando $p = 0,02$ e $n = 100$, então

$$\begin{aligned} P(X = 3) &= \binom{100}{3} (0,02)^3 (1 - 0,02)^{100-3} \\ &= 0,1823. \end{aligned}$$

Solução do Exercício 6.3 na página 135:

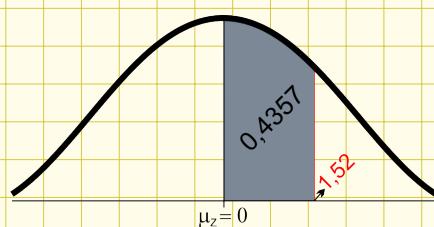
a) $P(0,00 \leq Z \leq 0,35) = 0,1368$. Graficamente, temos



Tabela

Z	...	0,05	
⋮		⋮	
0,3	...	0,1368	

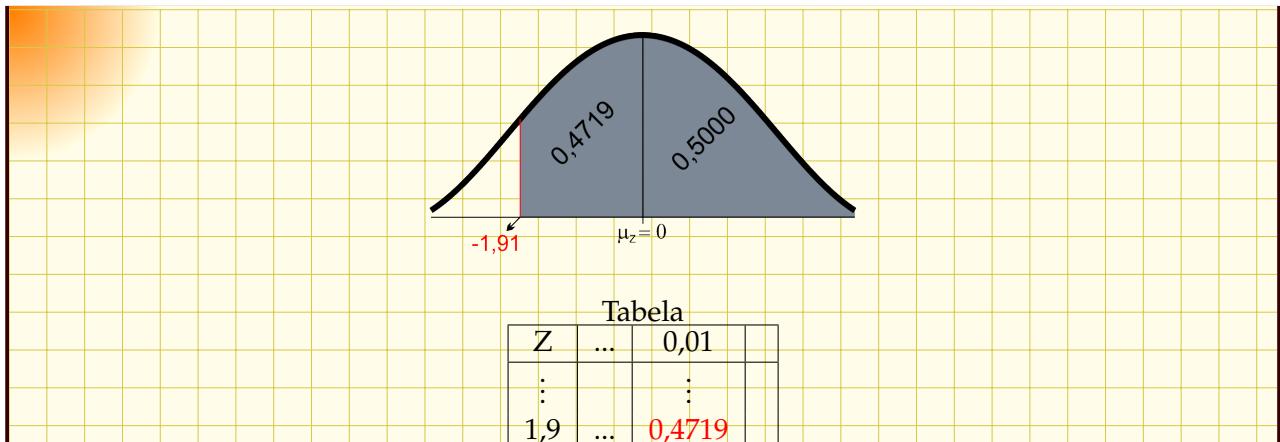
b) $P(0,00 \leq Z \leq 1,52) = 0,4357$. Graficamente, temos



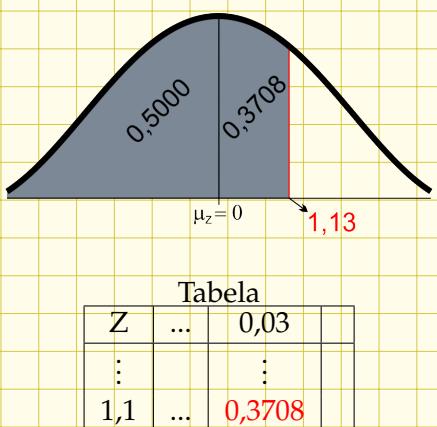
Tabela

Z	...	0,02	
⋮		⋮	
1,5	...	0,4357	

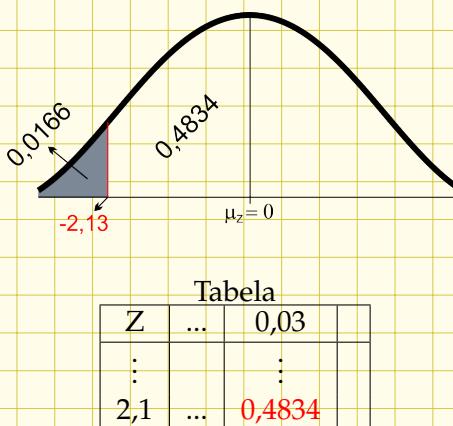
c) $P(Z \geq -1,91) = 0,4719$. Graficamente, temos



d) $P(Z \leq 1,13) = 0,5000 + 0,3708 = 0,8708$. Graficamente, temos



e) $P(Z \leq -2,13) = 0,5000 - 0,4834 = 0,0166$. Graficamente, temos



Solução do Exercício 6.4 na página 135:

- Fazendo a transformação $Z = \frac{X-\mu}{\sigma} = \frac{33-40}{6} = -1,17$, temos que $P(X \leq 33) = P(Z \leq -1,17) = 0,5000 - P(-1,17 \leq Z \leq 0) = 0,5000 - 0,3790 = 0,1210$.
- Fazendo a transformação $Z = \frac{X-\mu}{\sigma} = \frac{29-40}{6} = -1,83$, temos que $P(X \geq 29) = P(Z \geq -1,83) = P(-1,83 \leq Z \leq 0) + 0,5000 = 0,4664 + 0,5000 = 0,9664$.
- Fazendo a transformação $Z = \frac{X-\mu}{\sigma} = \frac{29-40}{6} = -1,83$, temos que $P(X \geq 29) = P(Z \geq -1,83) = P(-1,83 \leq Z \leq 0) + 0,5000 = 0,4664 + 0,5000 = 0,9664$.
- $Z = -0,20 \Rightarrow X = \mu + Z\sigma = 40 + (-0,20) \times 6 = 38,8$.

Tabela

Z	...	0,00		
⋮		⋮		
0,2	...	0,07926		

e) $Z = 1,28 \Rightarrow X = \mu + Z\sigma = 40 + 1,28 \times 6 = 47,68.$

Tabela				
Z	...	0,08		
⋮		⋮		
1,2	...	0,39973		

f) $P(Z > 0) = 0,5000.$

Solução do Exercício 6.5 na página 135:

$$Z = \frac{X - \mu}{\sigma} = \frac{5 - 6}{1,5} = -0,67.$$

$$\begin{aligned} P(X \geq 5) &= P(Z \geq -0,67) = 0,5000 + P(0 \leq Z \leq -0,67) \\ &= 0,5000 + 0,2486 = 0,7486. \end{aligned}$$

Logo, entre 120 candidatos, o número de candidatos aprovados é $120 \times 0,7486 = 89,83 \approx 90$ candidatos.

Solução do Exercício 6.6 na página 135:

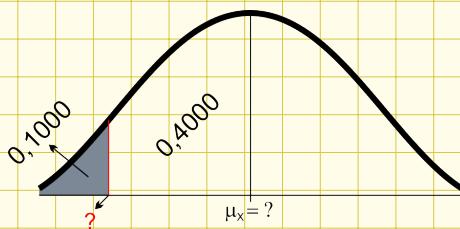


Tabela				
Z	...	0,08		
⋮		⋮		
1,2	...	0,39973		

$$Z = -1,28 \Rightarrow \mu_X = X - Z\sigma = 400 - (-1,28) \times 20 = 425,60g$$

Solução do Exercício 6.7 na página 136:

CONCEITO A: Considerando que $\mu + \sigma = 4,5 + 2,3 = 6,8$, então

$$Z_{\mu+\sigma} = \frac{6,8 - 4,5}{2,3} = 1.$$

$$\text{Logo, } P(\bar{X}_A \geq \mu + \sigma) = P(\bar{X}_A \geq 6,8) = P(Z \geq 1) = 0,5000 - 0,3413 = 0,1587;$$

CONCEITO B: Considerando que $\mu + \sigma = 4,5 + 2,3 = 6,8$, então

$$Z_{\mu+\sigma} = \frac{6,8 - 4,5}{2,3} = 1,$$

e $Z_\mu = (\mu - \mu)/\sigma = 0$. Logo, $P(2,3 \leq \bar{X}_B \leq 6,8) = P(Z_\mu \leq Z \leq Z_{\mu+\sigma}) = P(0 \leq Z \leq 1) = 0,3413$;

CONCEITO C: Considerando

$$Z_{\mu-\sigma} = \frac{2,2 - 4,5}{2,3} = -1.$$

Logo, $P(2,3 \leq \bar{X}_C \leq 4,5) = P(-1 \leq Z \leq 0) = P(-1 \leq Z \leq 0) = 0,3413$;

$P(\bar{X}_D \leq 2,2) = P(Z_{\mu-\sigma} \leq -1) = 0,1587$.

Assim, o número de alunos que receberam cada um dos conceitos são

Conceito	Cálculo	Nº Alunos
A	$0,1587 \times 100$	16
B	$0,3413 \times 100$	34
C	$0,3413 \times 100$	34
D	$0,1587 \times 100$	16

Solução do Exercício 6.8 na página 136:

$$E[X] = \sum_{i=1}^n x_i p(x_i) = 60.000 \times 0,3 + 20.000 \times 0,7 = 26.000.$$

Solução do Exercício 6.9 na página 136:

$$\text{a)} P(X = 0) = \binom{35}{0} (0,05)^0 (0,95)^{35} = 0,95^{35} = 0,1660834;$$

$$\text{b)} P(X = 1) = \binom{35}{1} (0,05)^1 (0,95)^{34} = 35 \times 0,05 \times 0,95^{34} = 0,3059431;$$

$$\text{c)} P(X \leq 1) = 0,1660834 + 0,3059431 = 0,4720265;$$

$$\text{d)} P(X > 2) = 1 - [P(X \leq 1) + P(X = 2)] = 1 - [0,4720265 + 0,2737385] = 0,254235.$$

Solução do Exercício 6.10 na página 136:

Considere que X representa o número de câmeras que não passaram no teste e que a probabilidade de sucesso é $p = 0,15$. Assim, podemos calcular o tamanho necessário de uma amostra para que a probabilidade de no mínimo uma câmera não tenha passado no teste seja de 90%, da seguinte forma:

$$P(X \geq 1) = 1 - P(X = 0) \geq 0,90.$$

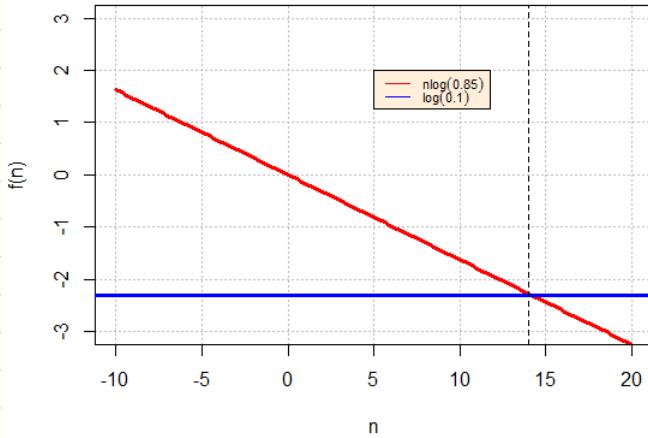
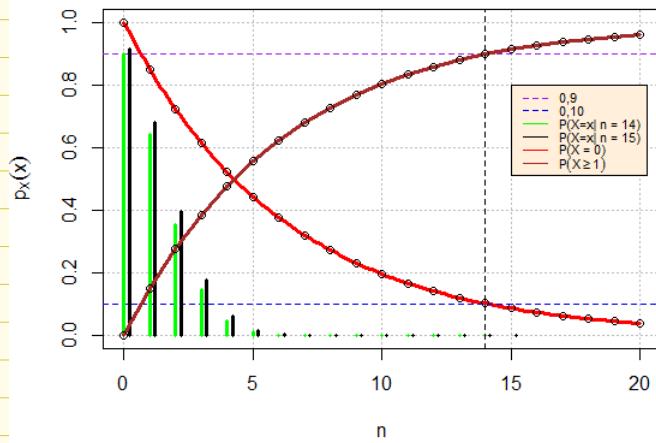
Observe que,

$$\begin{aligned} P(X = 0) &= \binom{n}{0} 0,15^0 (1 - 0,15)^n \\ &= 0,85^n. \end{aligned}$$

Assim, para que $P(X \geq 1) \geq 0,90$, temos que $P(X = 0) \leq 0,10$, logo,

$$\begin{aligned} 0,85^n &\leq 0,10 \\ n \log(0,85) &\leq \log(0,10) \\ -0,16252n &\leq -2,30259 \quad \times (-1) \\ 0,16252n &\geq 2,30259 \\ n &\geq \frac{2,30259}{0,16252} \\ n &\geq 14,1780. \end{aligned}$$

Para $n = 14$, $P(X \geq 1) = 0,8972$, e para $n = 15$, $P(X \geq 0,9126)$. Portanto $n \geq 15$. As figuras abaixo podem auxiliar no entendimento do resultado.



Solução do Exercício 6.11 na página 136:

a) $P(X = 0) = \frac{0,30^0 e^{-0,30}}{0!} = 0,7408182;$

b) Como $\lambda^* = (0,3 \times 8)/24 = 0,1$ alterações/8 horas, então $P(X > 1) = 1 - [P(X = 0) + P(X = 1)] = 1 - [0,9048374 + 0,09048374] = 0,00467884;$

c) Como $\lambda^* = (0,3 \times 2)/1 = 0,6$ alterações/2 dias, então $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0,5488116 + 0,329287 + 0,09878609 = 0,9768847$;

d) $E[X] = \lambda = 0,30$ alterações/dia, e variância $Var[X] = \lambda^2 = 0,09$ (alterações/dia)².

Solução do Exercício 6.12 na página 137:

a) $P(X = 0) = \binom{100}{0} 0,05^0 0,95^{100} = 0,005920529$;

b) $P(X > 5) = 1 - [P(X = 0) + P(X = 1) + \dots + P(X = 5)] = 1 - 0,6159991 = 0,3840009$;

c) $P(X = 15) = \binom{100}{15} 0,05^{15} 0,95^{85} = 0,00009880016$;

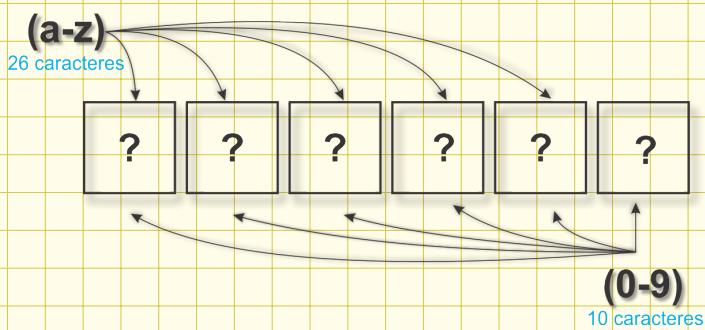
d) $E[X] = n \times p = 100 \times 0,05 = 5$ consumidores;

e) A variância é dada por $Var[X] = np(1-p) = 100 \times 0,05 \times 0,95 = 4,75$ consumidores², e o desvio padrão é dado por $\sigma_X = \sqrt{4,75} = 2,179449$ consumidores.

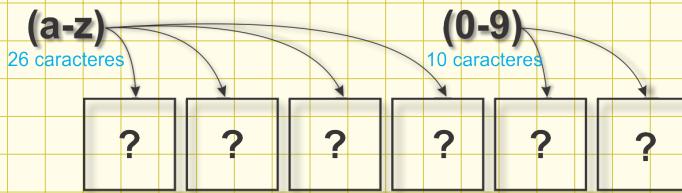
Solução do Exercício 6.13 na página 137:

Seja $N = 36^6$, $n_{(I)} = 26^5 \times 60$, $n_{(II)} = 26^5 \times 10^2$, $p_{(I)} = n_{(I)}/N$ e $p_{(II)} = n_{(II)}/N$, então:

a) $P(X = 2) = \binom{10}{2} p_{(I)}^2 (1 - p_{(I)})^{10-2} = 45 \times 0,3274937^2 \times (1 - 0,3274937)^8 = 0,2019246$.



b) $P(X = 0) = \binom{15}{0} p_{(II)}^0 (1 - p_{(II)})^{15} = 1 \times 0,5458229^2 \times (1 - 0,5458229)^{10} = 0,0000072174$.



Solução dos Exercícios do Capítulo 7

Solução dos Exercícios do Capítulo 8

Solução dos Exercícios do Capítulo 9

Solução do Exercício 9.1 na página 159:

Solução do Exercício 9.2 na página 159:

Solução do Exercício 9.3 na página 159:

Solução do Exercício 9.4 na página 159:

Solução do Exercício 9.5 na página 159:

Solução do Exercício 9.6 na página 160:

Solução dos Exercícios do Capítulo 10**Solução do Exercício 10.1 na página 167:**

- a) Vamos assumir que as questões são independentes e que a variável aleatória X denota o número de acertos entre as 10 questões. Portanto, X tem distribuição binomial com parâmetros $n = 10$ e p desconhecido. Supondo que adotamos a seguinte regra de decisão: o aluno não está advinhando se acertar 8 ou mais questões. Isto equivale a

Rejeitar H_0 se $X \geq 8$.

Dessa forma, a região crítica do teste é $C_Y = \{X : X > 8\}$. É possível também que o aluno acerte 8 ou mais questões e esteja advinhando, isto é, podemos rejeitar H_0 dado que ela é verdadeira. A probabilidade de que isso ocorra é:

$$P(X \geq 8 | p = 1/2) = \sum_{k=8}^{10} (1/2)^k (1 - 1/2)^{10-k} \approx 0,055.$$

Solução do Exercício 10.2 na página 167:**Solução do Exercício 10.3 na página 167:****Solução do Exercício 10.4 na página 167:****Solução do Exercício 10.5 na página 168:****Solução do Exercício 10.6 na página 168:**

Solução dos Exercícios do Capítulo 11

Solução do Exercício 11.1 na página 179:

Solução do Exercício 11.3 na página 179:

Solução do Exercício 11.3 na página 179:

Solução dos Exercícios do Capítulo 12

Solução dos Exercícios do Capítulo 13

Solução dos Exercícios do Capítulo 14

Referências Bibliográficas

- COMPANY, B. P. *BP statistical review of world energy*. London, 2018.
- DEVORE, J. L. *Probabilidade e Estatística para Engenharia e Ciências*. 6. ed. São Paulo: Cengage Learning, 2006. 692 p. Tradução Joaquim Pinheiro Nunes da Silva.
- FERREIRA, D. F. *Estatística Básica*. 2 revisada. ed. Lavras: Editora UFLA, 2009. 664 p.
- HITE, S. *Woman and love: A cultural revolution in progress*. New York: Knopf, 1987. 922 p.
- JAMES, B. R. *Probabilidade: um curso em nível intermediário*. 3. ed. Rio de Janeiro: IMPA, 2004. 304 p.
- LOHR, S. L. *Sampling: Design and analysis*. Boca Raton: CRC Press, 2019. 596 p.
- Magalhães, M. N.; LIMA, A. C. P. de. *Noções de Probabilidade e Estatística*. 7. ed. São Paulo: Edusp, 2015. 416 p.
- MONTGOMERY, D. C.; RUNGER, G. C. *Estatística Aplicada e Probabilidade para Engenheiros*. 6. ed. Rio de Janeiro: LTC, 2016. 629 p. Tradução Verônica Calado.
- MORETTIN, L. G. *Estatística básica: probabilidade e inferência*. Volume único. São Paulo: Pearson Prentice Hall, 2010. 390 p.
- Presidential Comission. *On the Space Shuttle Challenger Accident*. Washington, 1986. I, 256 p.
- SEKER, S.; KALKAN, T.; UZUM, G.; CELIK, C.; CEREZCI, O. Em effects of different mobile handsets on rats' brain. In: *2002 IEEE International Symposium on Electromagnetic Compatibility*. [S.l.: s.n.], 2002. v. 2, p. 667–670 vol.2.
- SILVA, J. G. C. da. *Planejamento da Pesquisa Experimental: Base conceitual e metodológica*. Pelotas: Instituto de Física e Matemática, 2007. 509 p.
- TAVARES, E. L.; ANJOS, L. A. do. Perfil antropométrico da população idosa brasileira. resultados da pesquisa nacional sobre saúde e nutrição. *Cad. Saúde Pública*, Rio de janeiro, v. 15, n. 4, p. 759–768, out-dez 1999.

Índice Remissivo

- desvio padrão
 - distribuição Bernoulli, 106
 - distribuição Binomial, 111
 - distribuição Poisson, 120
- distribuição
 - Bernoulli, 104
 - Binomial, 107
 - Normal, 122
 - Normal Padrão, 128
 - Poisson, 116
- espaço
 - amostral, 67
 - contínuo, 68
 - discreto, 68
- esperança matemática
 - distribuição Bernoulli, 106
 - distribuição Binomial, 111
 - distribuição Poisson, 120
- experimentos
 - aleatórios, 66
 - de Bernoulli, 104, 106, 116
- função de probabilidade
 - Bernoulli, 104
 - Binomial, 108
 - Poisson, 118
- função densidade de probabilidade
 - Normal, 123
 - Normal Padrão, 128
- medidas de posição, 32
 - mediana, 36
 - moda, 44
 - média, 32
- R
 - pacote stats
 - `dbinom()`, 107
 - `dpois()`, 117, 118, 121, 122
 - `pbinom()`, 107
 - `ppois()`, 117, 121

