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, 19 de junho de 2021

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]. 2021. Disponível em: <https://bendeivide.github.io/book-epaec/>

Autor correspondente e mantenedor da obra:
Ben Dêivide de Oliveira Batista
Contato: <ben.deivide@ufsj.edu.br>
Site pessoal: <http://bendeivide.github.io/>

Licença

Epígrafe

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

Sumário

Prefácio

Esse é um **livro digital** da 1ª 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, 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 nesse local ou consequências dessa fonte energética na economia do país.

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. Nessa 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 entreterimento 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 números 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. Você já se perguntou como é que lojas virtuais lhe oferta produtos sendo que nunca acessou aquele site antes? Já percebeu que o Netflix quando lhe oferece uma série a tela de entrada as 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.

[a] Ver 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 dos dados dessas 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.

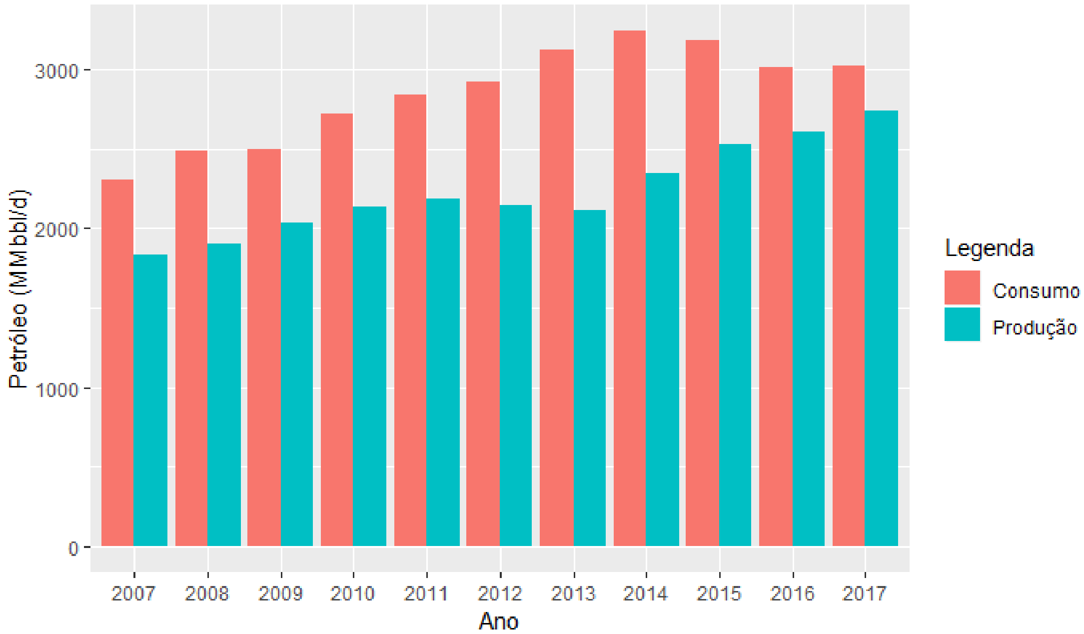


Figura 1.1: Produção e consumo de petróleo do Brasil nos períodos de 2007 a 2017.

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:

```
# BP Statistical Review 2018

# Producao e consumo de petroleo no brasil (Milhares de barris por dia)
ano  <- as.factor(c(2007:2017, 2007:2017))
prodcons <- c(1831, 1897, 2029, 2137, 2179, 2145, 2110, 2341, 2525, 2608,
    2734, 2308, 2481, 2498, 2716, 2839, 2915, 3124, 3242, 3181, 3013, 3017)
id <- c(rep("Produção", 11), rep("Consumo",11))

# Objeto que armazena as informações
dados <- data.frame(ano, Legenda = id, prodcons)

# Pacote utilizado
install.packages("ggplot2")
library(ggplot2)

# Funcao para criacao do grafico de barras
ggplot(dados, aes(x=ano, y=prodcons, fill=Legenda)) +
  geom_bar(position="dodge", stat="identity") +
  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 aproximadamete 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 inferencias 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.

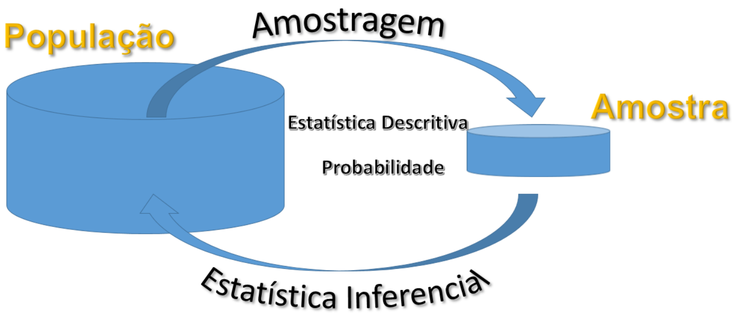


Figura 1.2: Ilustração animada sobre as disposições gerais da Estatística.

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]:

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

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:

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

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.

[a] A 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 á 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.

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 $i = 1, 2, \ldots, n$. 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 PRODES[1] (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, percebamos 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

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

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

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

[4]Essa 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>

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 é 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.

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, ..., 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,

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

sendo que m pode representar o tamanho amostral, n, ou o tamanho populacional, N. No caso da representação anterior, podemos simplicar 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,

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

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

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

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,

$$\sum_{i=1}^{n} X_i Y_i = X_1 Y_1 + X_2 Y_2 + \ldots, X_n Y_n. \tag{1.2}$$

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:

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

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 $n = 3$,

$$\begin{aligned}\sum_{i=1}^{3}\sum_{j=1}^{3} X_iY_j &= \sum_{j=1}^{3} X_1Y_j + \sum_{j=1}^{3} X_2Y_j + \sum_{j=1}^{3} X_3Y_j \\ &= (X_1Y_1 + X_1Y_2 + X_1Y_3) + (X_2Y_1 + X_2Y_2 + X_2Y_3) + (X_3Y_1 + X_3Y_2 + X_3Y_3).\end{aligned}$$

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

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} X_iY_j &= \sum_{j=1}^{n} X_1Y_j + \sum_{j=1}^{n} X_2Y_j + \ldots + \sum_{j=1}^{n} X_nY_j \\ &= (X_1Y_1 + X_1Y_2 + \ldots + X_1Y_n) + (X_2Y_1 + X_2Y_2 + \ldots + X_2Y_n) + \ldots \\ &\quad \ldots + (X_nY_1 + X_nY_2 + \ldots + X_nY_n). \end{aligned} \tag{1.4}$$

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 $\sum_{i=1}^{n}\sum_{j=1}^{n} X_iY_j$ da seguinte forma,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} X_iY_j = \sum_{\substack{i=1\\ j=1}}^{n} X_iY_j. \tag{1.5}$$

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,

$$\begin{aligned}(X_1 + X_2 + \ldots, X_n)^2 &= (X_1 + X_2 + \ldots + X_n) \times (X_1 + X_2 + \ldots + X_n) \\ &= X_1X_1 + X_1X_2 + \ldots + X_1X_n + \\ &\quad + X_2X_1 + X_2X_2 + \ldots + X_2X_n + \ldots \\ &\quad \ldots + X_nX_1 + X_nX_2 + \ldots + X_nX_n \\ &= \sum_{i=1}^{n} X_i^2 + 2\sum_{j>i=1}^{n} X_iX_j. \end{aligned} \tag{1.6}$$

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 $X_1 \times X_2 + X_2 \times X_1 = 2X_1X_2$ que o resultado será o mesmo, e simplifica a notação. Generalizando a soma para os demais casos, temos $2\sum_{j>i=1}^{n} X_iX_j$, 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 $P(.)$ é uma função que mede a chance de determinado X assumir um determinado valor, isto é, $P(.)$ 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 deteminado 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:

I) $\sum\limits_{i=1}^{n} aX_i = a \sum\limits_{i=1}^{n} X_i$

II) $\sum\limits_{i=1}^{n} X_i Y_i \leq \sum\limits_{i=1}^{n} X_i \sum\limits_{i=1}^{n} Y_i$;

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

IV) $\sum\limits_{i=1}^{n} k = nk$;

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

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

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

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

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

Prova:

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

(ii) Observando as expressões (1.2) e (1.3), claramente que $\sum\limits_{i=1}^{n} X_i Y_i < \sum\limits_{i=1}^{n} X_i \sum\limits_{i=1}^{n} Y_i$. 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,

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

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

(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,

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

Exercícios propostos

Exercício 1.1: De acordo com o Exemplo 1.1, identifique um problema para a 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 137

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

Solução na página 137

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 137

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

$$X = \{2, 4, 5, 1, 2\},$$
$$Y = \{1, 2, 3, 5, 8\}.$$

Obtenha:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

r) Qual conclusão se pode chegar sobre os itens (n) e (p), bem como (o) e (q)?

Solução na página 137

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 137

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:

a) Qual a população em estudo?

b) Qual o problema indagado?

c) Qual(is) a(s) variável(is) em estudo do trabalho, como também a natureza dessa(s) variável(is)?

d) Podemos identificar o tamanho da população e da amostra, com essas informações?

Solução na página 137

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

Solução na página 137

Exercício 1.8: Os dados retirados de Tavares e Anjos (1999), representam a distribuição percentual do estado nutricional em homens idosos brasileiros (idade $\geq$ 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	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 137

[a] A unidade de IMC em kg/m^2.

[b] A 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 137

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 está 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, ..., X_n, de tamanho n. Com o ordenamento, usaremos um parêntese no índice, $X_{(1)}$, $X_{(2)}$, ..., $X_{(n)}$, de modo que, $X_{(1)} = \min_i(X_i)$ e $X_{(n)} = \max_i(X_i)$ para $i = 1, 2, \ldots, n$. Da mesma forma, vale para a representação dos valores observados dessas respectivas variáveis, isto é, valores observados sem ordenação denotados por $x_{(1)}$, $x_{(2)}$, ..., $x_{(n)}$, 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 simplicada, 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),

$$F_{r_i} = \frac{F_i}{\sum_{i=1}^{k} F_i}, \quad i = 1, 2, \ldots, k, \tag{2.1}$$

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).

$$F_{\%_i} = F_{r_i} \times 100, \quad i = 1, 2, \ldots, k, \tag{2.2}$$

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	$0,20$	$0,20 \times 100 = 35$
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 perguntássemos, 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),

$$F_{ac\downarrow_i} = \sum_{j=1}^{i} F_j, \quad i = 1, 2, \ldots, k, \tag{2.3}$$

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 $F_{ac\downarrow}$, cujo valor calculado é denotado por $f_{ac\uparrow}$, dado pela expressão (2.4),

$$F_{ac\uparrow_i} = \sum_{j=i}^{k} F_j, \quad i = 1, 2, \ldots, k, \tag{2.4}$$

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	-	-

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

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

$$Fr_{ac_\downarrow} = \frac{F_{ac\downarrow_i}}{\sum_{i=1}^{k} F_i} \quad i = 1, 2, \ldots, k, \tag{2.6}$$

e

$$Fr_{ac_\uparrow} = \frac{F_{ac\uparrow_i}}{\sum_{i=1}^{k} F_i} \quad i = 1, 2, \ldots, k, \tag{2.7}$$

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

$$F_{ac_\downarrow\%} = Fr_{ac_\downarrow} \times 100 \quad i = 1, 2, \ldots, k, \tag{2.8}$$

e

$$F_{ac_{\uparrow}\%} = Fr_{ac_{\uparrow}} \times 100 \quad i = 1, 2, \ldots, k, \tag{2.9}$$

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 anterioremente, 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	fr_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	0,30	1	30	100
1	7	0,35	35	10	17	0,33	0,85	33	85
2	4	0,20	35	14	10	0,47	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:

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

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 esses 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),

$$A_t = \max_i(X_i) - \min_i(X_i), \tag{2.11}$$

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

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

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

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),

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

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
$Li_{1a} \mid\!\!— Ls_{1a}$

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 $Ls_{2a} = L1_{1a} + c$, em que c é a amplitude da classe. Inserindo agora a segunda classe na tabela, temos:

Classe
$Li_{1a} \mid\!\!— Ls_{1a}$
$Li_{2a} \mid\!\!— Ls_{2a}$

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
$Li_{1a} \mid\!\!— Ls_{1a}$
$Li_{2a} \mid\!\!— Ls_{2a}$
$\vdots$
$Li_{ka} \mid\!\!—\!\!\mid Ls_{ka}$

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 $Li_{ka} \mid\!\!— Ls_{ka}$. Um problema surge com os dados, porque ao serem agrupadados nas classes nós perdemos essa informação. Vejamos a representação de uma classe com a sua frequência absoluta,

Classe	F_i
Li_{1a} $\vert$— Ls_{1a}	f_1
Li_{2a} $\vert$— Ls_{2a}	f_2
$\vdots$	
Li_{ka} $\vert$—$\vert$ 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 $i = 1, 2, \ldots, k$, em cada classe é usa 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:

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

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 (ºF) 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 $x_{(1)} = 31^oF$ e $x_{(36)} = 84^oF$, logo a amplitude total é $A_t = 53^oF$. Na sequência, calculamos amplitude da classe, $c = 53/(6-1) = 10,6^oF$, por fim, o limite inferior da primeira classe, $Li_{1a} = 31 - 10,6/2 = 25,7^oF$. Assim, começaremos pela primeira classe, em que já temos o limite inferior dela e seu limite superior será $Li_{1a} = 25,7 + 10,6 = 36,3^oF$. As demais classes, segue o procedimento descrito anteriormente. Por fim, verificaremos quais o 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

2.3 Representação gráfica

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 139

[a] Entenda 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 $\geq$ 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 distribuição de frequência baseado nas informações da tabela? Apresente-a(s).

Solução na página 139

[a] A unidade de IMC em kg/m^2.

[b] A 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 2.3:

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

Solução na página 139

Capítulo 3

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 exista 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 $X_1, X_2, \ldots, X_n$, de uma população $X_1, X_2, \ldots, X_N$, de tamanhos n e N, respectivamente, definimos a média aritmética por:

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

e

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

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:

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

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, $\ldots$, 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 $i = 1, 2, \ldots, k$ e $\sum_{i=1}^{k} F_i = n$, então a média aritmética de uma amostra, é definida por:

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

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:

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

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:

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

Se calculássemos a média sem agrupamento, o valor seria $\bar{X} = 65{,}86^{\text{o}}$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 $0,18$ºF 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:

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

Observamos que as observações $x_1 = 1\ unid.$ e $x_{17} = 40\ und.$ 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:

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

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:

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

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 $X_1, X_2, \ldots, X_n$, 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 $i = 1, 2, \ldots, n$, a nova média aritmética é dada por $\bar{Y} = \bar{X} \pm c$;

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

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

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,

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

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 or próprios valores observados. Mas especificamente, o valor da mediana é o ponto central dos dados, em que abaixo desse valor, representa as 50% menores observação, 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 $X_1, X_2, \ldots, X_n$, de uma população $X_1, X_2, \ldots, X_N$, de tamanhos n e N, respectivamente, definimos a mediana por:

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

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,

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

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 $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ 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:

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

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* ($f_{ac\downarrow_i}$), 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 $X_{(4)}, X_{(5)}, \ldots, X_{(10)}$, 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 $X_{(11)}$, $X_{(12)}, \ldots, X_{(14)}$, 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.

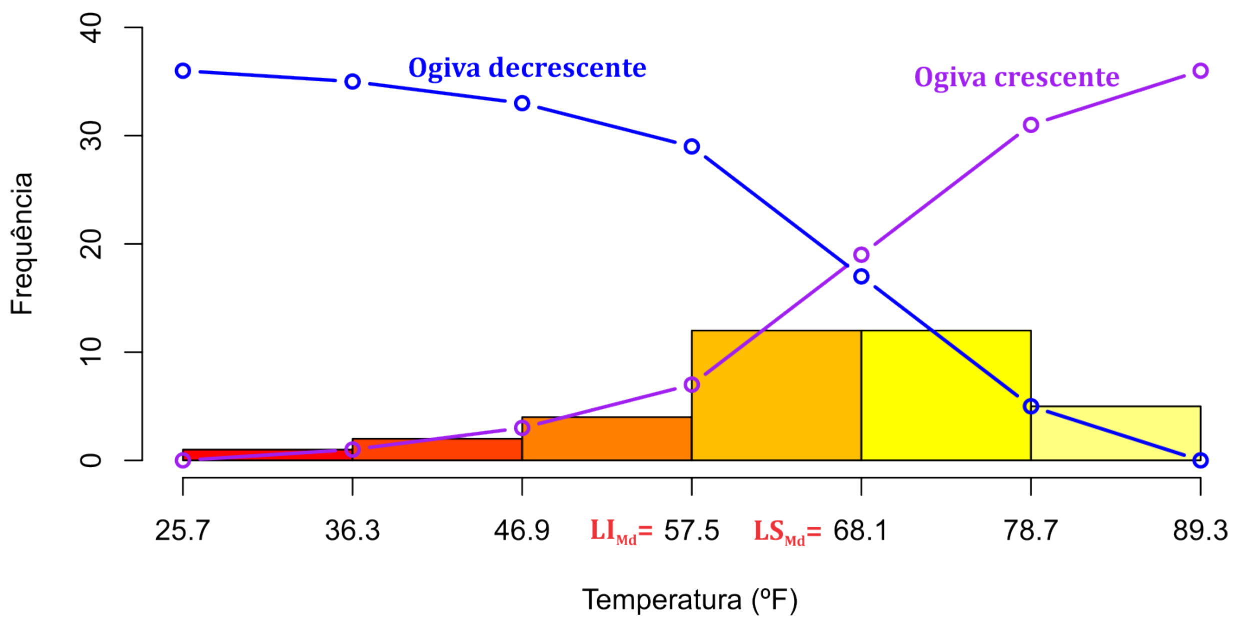


Figura 3.1: Histograma de frequência e ogivas para a dedução do cálculo da mediana.

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 $57,5 \vdash 68,1$ porque $f_{ac\downarrow_4} = 19$, isto é, abaixo de $68,1$ ºF temos as primeiras 19 observações, e nessa classe contemplamos as observações ordenadas $x_{(8)}, x_{(9)}, \ldots, x_{(19)}$, 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.

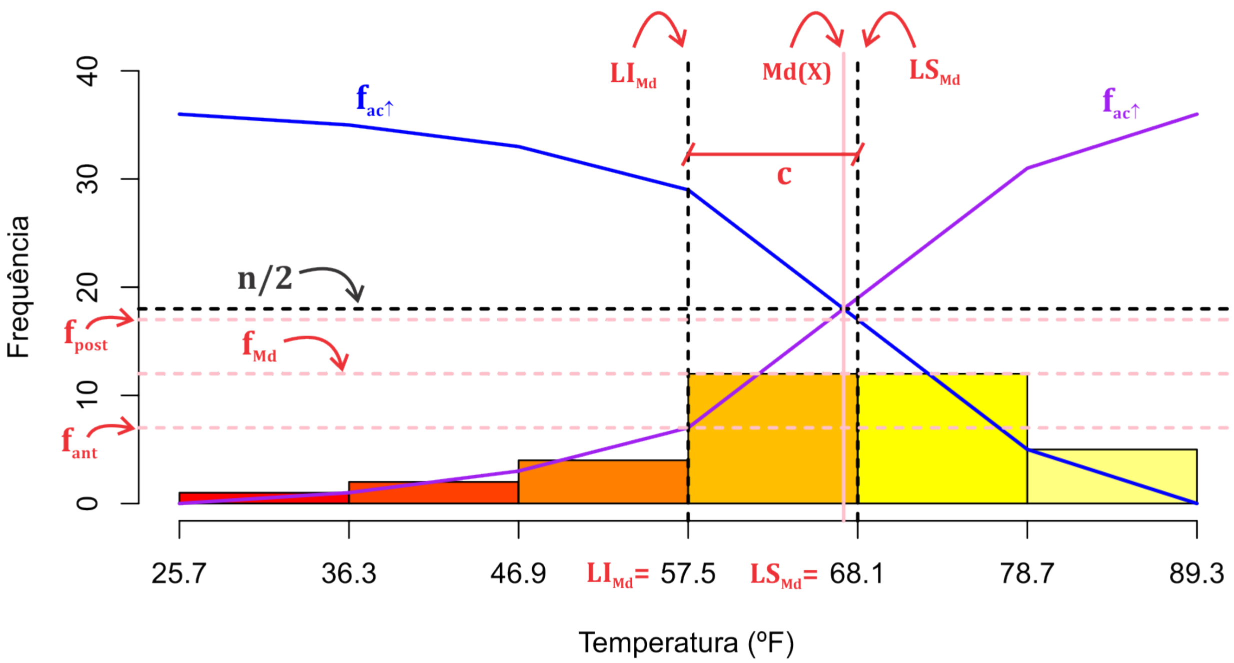


Figura 3.2: Histograma de frequência e ogivas para a dedução do cálculo da mediana com as notações.

1° Método

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

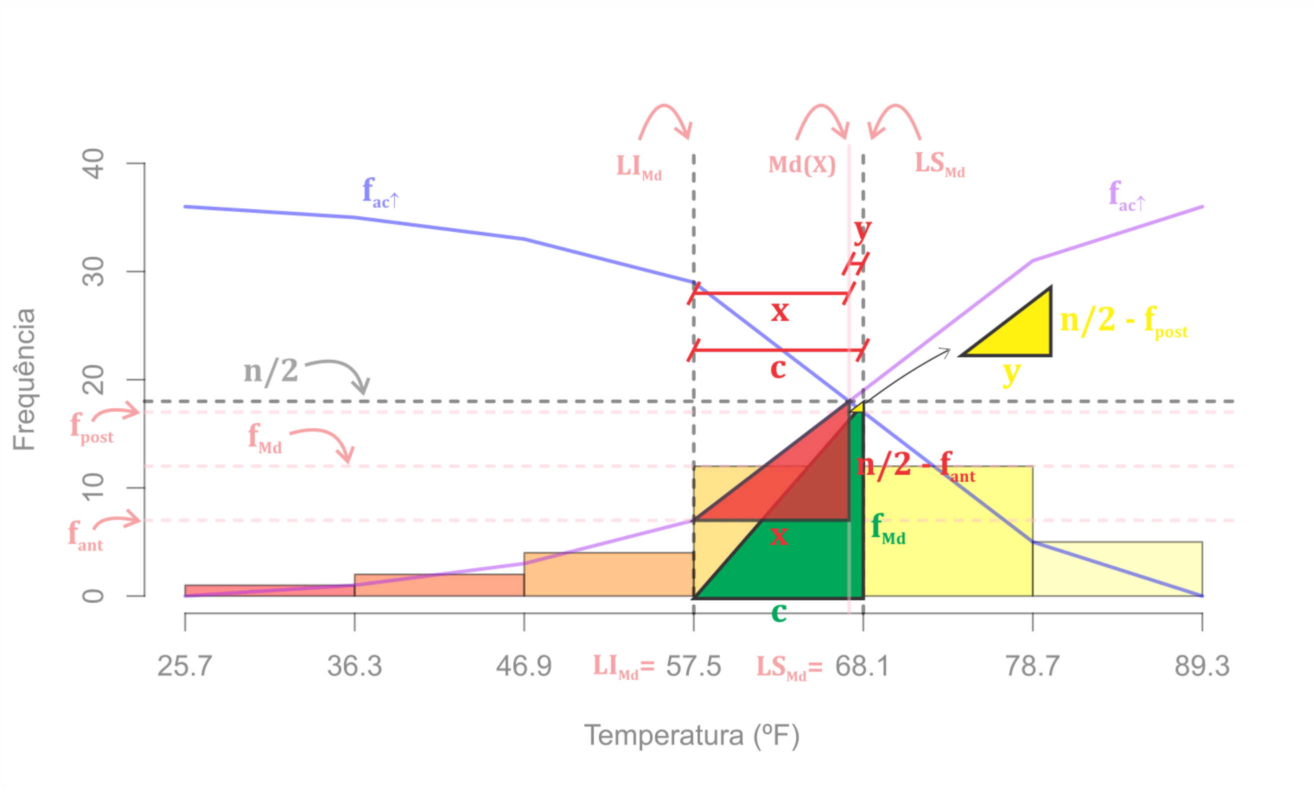


Figura 3.3: Detalhamento do Histograma de frequência e ogivas para a dedução do cálculo da mediana com as notações.

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

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

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

2° 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).

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

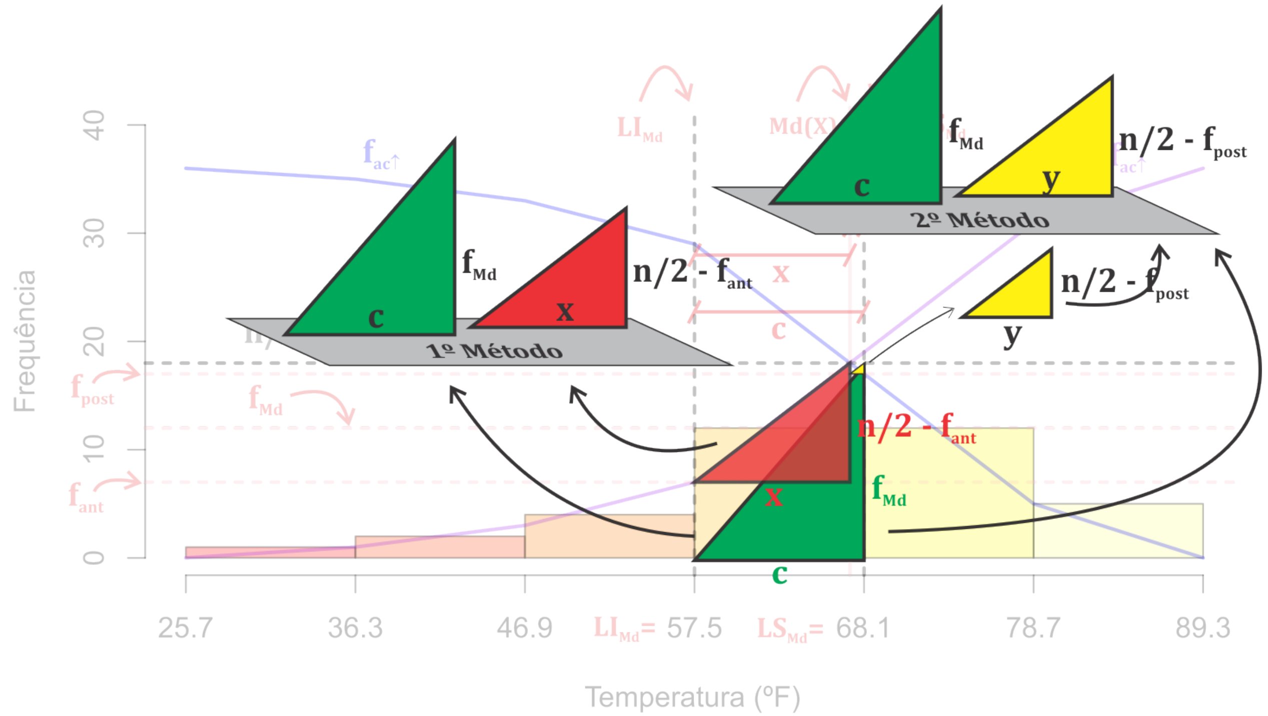


Figura 3.4: Semelhança de triângulos para a dedução do cálculo da mediana com as notações.

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

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

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

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

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,

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

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),

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

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

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

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 variavel 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 $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ 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 $i = 1, 2, \ldots, n$, a mediana aritmética é dada por $Md(Y) = Md(X) \pm c$;

II) Se para uma amostra $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ 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 $i = 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_n$, e c uma constante, e que $Y_i = X_i \pm c$, para $i = 1, 2, \ldots, n$, então a mediana de Y_i é dado por:

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

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

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

III) Considerando uma amostra $X_1, X_2, \ldots, X_n$, 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

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

que resulta em

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

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

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

se

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

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 $X_{\left(\frac{n+2}{2}\right.}$, isto é,

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

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

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

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 $X_1, X_2, \ldots, X_n$, de uma população $X_1, X_2, \ldots, X_N$, 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.

[a] Entendemos 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.

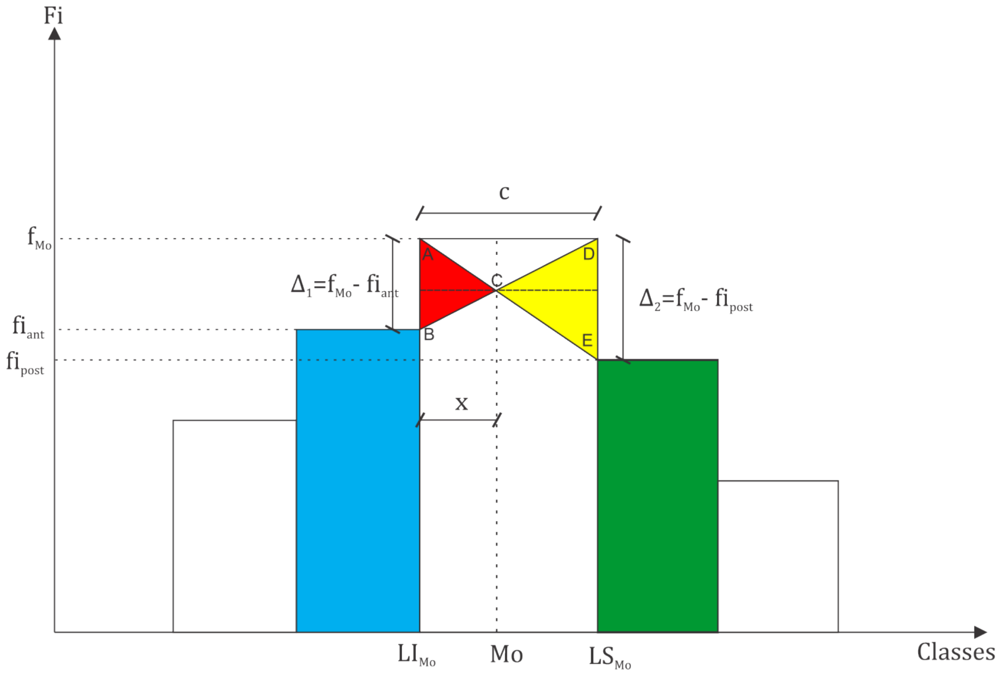


Figura 3.5: Histograma de frequência para a deduação do cálculo da moda.

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

$$Mo(X) = 3Md(X) - 2\overline{X}. \tag{3.14}$$

Formalmente, definimos

Definição 3.6: Moda para dados agrupados com intervalo de classe

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

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

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 variavel 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,

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

Vejamos a moda para a classe 5,

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

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 $X_1, X_2, \ldots, X_n$ 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 $i = 1, 2, \ldots, n$, a moda é dada por $Mo(Y) = Mo(X) \pm c$;

II) Se para uma amostra $X_1, X_2, \ldots, X_n$ 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 $i = 1, 2, \ldots, n$, a nova moda é dada por $MoY = MoX \times c$. 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 complicada ou dividida pela constante.

   Sejam
   $$Y_i = cX_i,$$

então

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

Por fim, apresentamos o Exemplo 3.10 de aes 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 140

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 140

Exercício 3.3: Considere uma amostra $X_1, X_2, \ldots, X_n$ 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 $i = 1, 2, \ldots, n$, sendo $S = \sqrt{S^2}$. Apresente as medidas de posição para essas transformações, Y e Z.

Solução na página 140

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 140

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 140

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 140

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 $\bar{X}_{\text{gA}} = 9,018686$ unid., $\bar{X}_{\text{gB}} = 9,018686$ unid. e $\bar{X}_{\text{gC}} = 9,018686$ 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:

```
# Semente de simulação
# set.seed(10)

# Dados dos grupos gerados
set.seed(10); gA <- rnorm(10, mean = 10, sd = 2)
set.seed(10); gB <- rnorm(9, mean = 10, sd = 4)
set.seed(10); gC <- rnorm(9, mean = 10, sd = 6)

# Media do grupo gA
(media.gA <- mean(gA))

# gerando gB[10] e media de gB
gB[10] <- gB10 <- length(gA) * media.gA - sum(gB)
(media.gB <- mean(gB))


# gerando gC[10] e media de gC
gC[10] <- gc10 <- length(gA) * media.gA - sum(gC)
(media.gC <- mean(gC))

# Apresentando os dados
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, simplemente 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
$\bar{X}_{\text{gA}} = 9,018686$	$\bar{X}_{\text{gB}} = 9,018686$	$\bar{X}_{\text{gC}} = 9,018686$

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

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

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,

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

Isso representa, uma variação de 155971 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 aplitude em termos amostras, como será apresentado na Definição 4.2, a seguir.

Definição 4.2: Amplitude em uma amostra

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

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

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 aplitude para dados agrupados com e sem intervalo de classe.

Definição 4.3: Amplitude em dados agrupados

Seja uma amostra $X_1, X_2, \ldots, X_n$, 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 $i = 1, 2, \ldots, k$ e $\sum_{i=1}^{k} F_i = n$, então a amplitude de uma amostra, denotada por A, é definida por:

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

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 encontrados 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:

$$A = 84 - 31 = 52 \text{ °F.}$$

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 discrepantes[1];
- 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 estimador[2] viesado[3] 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.

[1] Entendemos 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.

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

[3] Dizemos 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 $X_1, X_2, \ldots, X_n$, a amplitude é dada por $A_X = X_{(n)} - X_{(1)}$, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \ldots, n$, a nova amplitude não se altera, isto é, $A_Y = A_X$.

II) Se para uma amostra $X_1, X_2, \ldots, X_n$, a amplitude é dada por $A_X = X_{(n)} - X_{(1)}$, então para uma transformação de $Y_i = X_i \times c$, para $i = 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_n$, e c uma constante, e que $Y_i = X_i \pm c$, para $i = 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_n$, e c uma constante, e que $Y_i = X_i \times c$, para $i = 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_N$ e sua respectiva amostra $X_1, X_2, \ldots, X_n$, 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:

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

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:

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

e

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

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:

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

e

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

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

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

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 $X_1, X_2, \ldots, X_N$, 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:

$$\sigma^2 = \frac{SQ_p}{N}, \tag{4.11}$$

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

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

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,

$$\hat{\sigma}^2 = \frac{SQ}{n}. \tag{4.13}$$

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, $(\hat{\sigma}_1^2 + \hat{\sigma}_2^2 + \ldots + \hat{\sigma}_k^2)/k$, 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 $X_1, X_2, \ldots, X_n$, de tamanho n, com $\bar{X} = \sum_{i=1}^{n} X_i/n$, então a variância amostral, denotada por S^2, é definida como:

$$S^2 = \frac{SQ}{n-1}, \tag{4.14}$$

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

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

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:

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

- Variância amostral para o grupo gB:

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

- Variância amostral para o grupo gC:

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

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 $X_1, X_2, \ldots, X_n$, 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 $i = 1, 2, \ldots, k$ e $\sum_{i=1}^k F_i = n$, então a variância de uma amostra, denotada por S^2, é definida por:

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

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 $\bar{\tilde{X}} = \sum_{i=1}^k \tilde{X}_i F_i / \sum_{i=1}^k F_i$, ou se forma similar,

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

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, mas 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 $X_1, X_2, \ldots, X_n$, a variância é dada por S_X^2, expressão (4.14), então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \ldots, n$, então a nova variância não se altera, isto é, $S_Y^2 = S_X^2$.

II) Se para uma amostra $X_1, X_2, \ldots, X_n$, a variância é dada por S_X^2, expressão (4.14), então para uma transformação de $Y_i = X_i \times c$, para $i = 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_n$, e c uma constante, e que $Y_i = X_i \pm c$, para $i = 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_n$, e c uma constante, e que $Y_i = X_i \times c$, para $i = 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_n$, de uma população $X_1, X_2, \ldots, X_N$, 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:

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

em que σ é apresentado na Definição 4.4,

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

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:

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

- Desvio padrão amostral para o grupo gB:

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

- Desvio padrão amostral para o grupo gC:

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

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 $S_1^2 = 10\ unid$ para a amostra 1, e $S_2^2 = 20\ unid$ 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 $\bar{X}_1 = 100\ unid$ e para a amostra 2, seja $\bar{X}_1 = 50\ unid$. 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, mas concentrado os dados estão em torno da média, ao passo que, à medida que o desvio padrão se distancia de zero, mas 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 $X_1, X_2, \ldots, X_n$, a amplitude é dada por $A_X = X_{(n)} - X_{(1)}$, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \ldots, n$, então o novo desvio padrão não se altera, isto é, $S_Y = S_X$.

II) Se para uma amostra $X_1, X_2, \ldots, X_n$, a amplitude é dada por $A_X = X_{(n)} - X_{(1)}$, então para uma transformação de $Y_i = X_i \times c$, para $i = 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_n$, e c uma constante, e que $Y_i = X_i \pm c$, para i

$= 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_n$, e c uma constante, e que $Y_i = X_i \times c$, para $i = 1, 2, \ldots, n$, 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 $X_1, X_2, \ldots, X_n$, de uma população $X_1, X_2, \ldots, X_N$, 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:

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

e

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

Essa medida permite com que possamos comparar a dispersão com dois ou mais grupos com características completamente diferente 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 diferente 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:

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

e

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

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	$CV_{\mathrm{gA}} = \frac{1,399787}{9,018686} \times 100 = 15,52\%$
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 (ºC), uma observação igual a 0ºC 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 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 $X_1, X_2, \ldots, X_n$, 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:

$$EA_{\bar{X}} = \bar{X} - \mu. \tag{4.22}$$

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, $\ldots$, 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:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}, \tag{4.23}$$

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 informação, e que o desvio padrão populacional geralemente 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, $\ldots$, 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:

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}, \tag{4.24}$$

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 $n \to N$, isto é, à medida que n aumenta, a média amostral tende a μ, logo o $EA_{\bar{X}} \to 0$. 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:

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

- Erro padrão da média amostral para o grupo gB:

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

- Erro padrão da média amostral para o grupo gC:

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

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:

I) Se para uma amostra $X_1, X_2, \ldots, X_n$, o coeficiente de variação, Definição 4.8, então para uma transformação de $Y_i = X_i \pm c$, para $i = 1, 2, \ldots, n$, então o novo coeficiente de variação é igual a $CV_Y = S/(\bar{X} \pm c) \times 100$.

II) Se para uma amostra $X_1, X_2, \ldots, X_n$, o coeficiente de variação, Definição 4.8, então para uma transformação de $Y_i = X_i \times c$, para $i = 1, 2, \ldots, n$, 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 141

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 $\vert$— 2	4
2 $\vert$— 4	12
4 $\vert$— 6	15
6 $\vert$— 8	13
8 $\vert$— 10	6

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

Solução na página 141

Exercício 4.3: Para uma amostra $X_1, X_2, \ldots, X_n$ de tamanho n. Desejamos usar a transformação $Y_i = \beta_0 + \beta_1 X_i$, para $i = 1, 2, \ldots, n$, 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 141

Exercício 4.4: Sabemos que a conversão da temperatura de graus Celsius (ºC) para a escala de Fahrenheit (ºF) 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 70ºC e variância de $2(^oC)^2$, como poderíamos representar essas medidas na escala Fahrenheit?

Solução na página 141

Exercício 4.5: Se tivéssemos estudando as 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 141

Exercício 4.6: Considere uma amostra $X_1, X_2, \ldots, X_n$ 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^2_{n+1}, partindo do pressuposto que só sabemos das informações $\bar{X}_n$, S^2_n 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 $x_{37} = 63$ºF, use os resultados obtidos e determine $\bar{x}_{n+1}$ e s^2_{n+1} após desse novo dado as observações.

Solução na página 141

Exercício 4.7: Considere uma amostra $X_1, X_2, \ldots, X_n$ 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^2_X. Considere também as transformações $Y_i = X_i - \bar{X}$ e $Z_i = (X_i - \bar{X})/S$, para $i = 1, 2, \ldots, n$, sendo $S = \sqrt{S^2}$. Apresente as medidas de dispersão, desvio padrão e variância, para essas transformações, Y e Z.

Solução na página 141

Capítulo 5

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 Jerónimo 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.

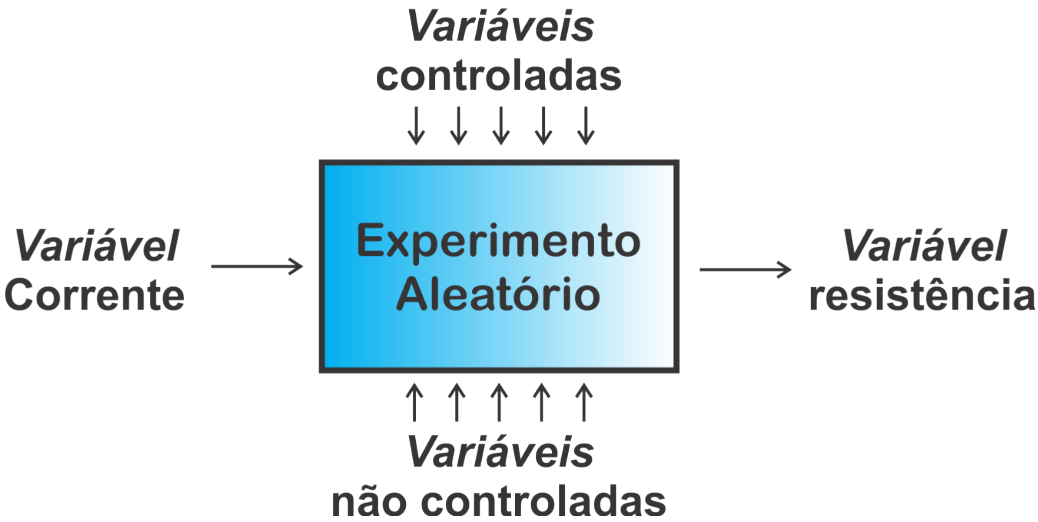


Figura 5.1: Componente aleatória de um experimento aleatório.

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:

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

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:

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

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,

$$\Omega = \{sim,\ não\}.$$

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

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

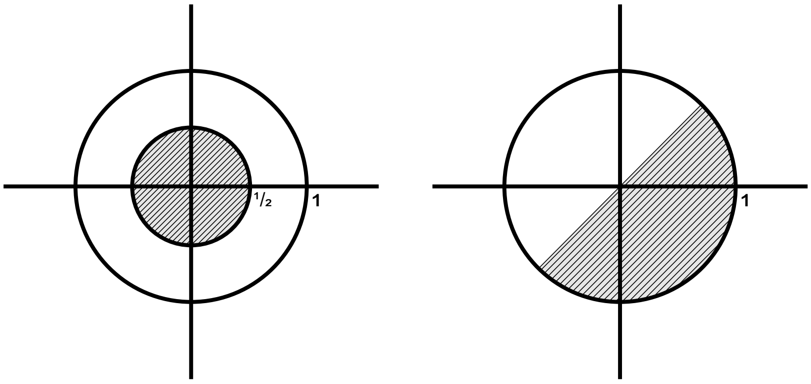
Vejamos alguns eventos para esse exemplo:

$$\begin{aligned}
A &= \text{“distância entre o ponto escolhido e a origem é”} \ \leq 1/2 \\
B &= \text{“distância entre o ponto escolhido e a origem é”} \ \geq 15 \\
C &= \text{“1ª Coordenada do ponto escolhido é maior que a 2ª.}
\end{aligned}$$

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:

$$\begin{aligned}
A &= \{(x, y) \in \Omega : \ \sqrt{x^2 + y^2} \leq 1/2\}, \\
B &= \emptyset = \text{conjunto vazio}, \\
A &= \{(x, y) \in \Omega : \ x > y\}.
\end{aligned}$$

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



(a) Evento A (b) Evento C

Figura 5.2: Escolha do ponto em um círculo unitário.

Diante, 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,

$$A = \{\omega \in \Omega : \omega \neq \omega\}. \tag{5.1}$$

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 $\emptyset \nsubseteq A$. 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,

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

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,

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

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

Definição 5.10: Eventos Disjuntos ou multuamente exclusivos

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

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 $\overline{A}$, A^c, ou $\Omega - A$, é o subconjunto dos elementos de Ω exceto os elementos do evento A, isto é,

$$A^c = \{\omega \in \Omega : \ \omega \notin A\}. \tag{5.4}$$

Exemplo 5.12:

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

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

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

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

Então o complemento de A será:

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

Definição 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,

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

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 $A_1, A_2, \ldots, A_n$, não vazios, é uma partição do evento A, se e somente se,

I) $A = \bigcup_{i=1}^{n} A_i$;

II) $A_1, A_2, \ldots, A_n$ são multuamente 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: $\omega \in (A \cup B)^c \Leftrightarrow \omega \notin A$ ou $\omega \notin B \Leftrightarrow \omega \in A^c$ 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: $\omega \in (A \cap B)^c \Leftrightarrow \omega \notin (A \cap B) \Leftrightarrow \omega \notin A$ e $\omega \notin B \Leftrightarrow \omega \in A^c$ 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 probabildade.

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) $A\varnothing = \varnothing$ (Elemento absorvente);

X) $A \cup \varnothing = A$;

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 $\exists \omega \in A \cap A^c : \omega \in A$ e $\omega \in A^c$. Mas por definição $A^c = \{\omega \in \Omega : \omega \notin A\}$, então a afirmação $\exists \omega \in A \cap A^c : \omega \in A$ e $\omega \in A^c$, é absurdo. Logo, $A \cap A^c = \emptyset$;

II) $\omega \in A \cup A^c \Leftrightarrow \omega \in A$ ou $\omega \in A^c \Leftrightarrow \omega \in A$ ou $\omega \notin A \Leftrightarrow \omega \in \Omega$;

III) Vamos apresentar a prova por contradição. Supomos que $\Omega^c \neq \emptyset$. Isso significa que $\exists \omega : \omega \in \Omega^c$ 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, $\exists \omega : \omega \in \Omega$ 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) $\omega \in A \cap \Omega \Leftrightarrow \omega \in A$ e $\omega \in \Omega \Leftrightarrow \omega \in A$;

VII) $\omega \in A \cap A \Leftrightarrow \omega \in A$ e $\omega \in A \Leftrightarrow \omega \in A$;

VIII) $\omega \in A \cup \Omega \Leftrightarrow \omega \in A$ 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, $\omega \in A$ 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) $\omega \in (A - B) \Leftrightarrow \omega \in A$ e $\omega \notin B \Leftrightarrow \omega \in A$ e $\omega \notin (A \cap B) \Leftrightarrow A - (A \cap B)$. Da mesma forma, $\omega \in (A - B) \Leftrightarrow \omega \in A$ e $\omega \notin B \Leftrightarrow \omega \in A$ 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) $\omega \in (A \cap B) \Leftrightarrow \omega \in A$ ou $\omega \in B \Leftrightarrow \omega \in A$ ou $\omega \in (B \cap A^c) \Leftrightarrow \omega \in A \cup (B \cap A^c)$;

XV) $\omega \in A \cup B \Leftrightarrow \omega \in A$ ou $\omega \in B \Leftrightarrow \omega \in (A \cap B)$ 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 $P : \Omega \to \mathbb{R}$, 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 $A_1 \cap A_2$ 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 probabilidade 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.

a) Qual é o espaço amostral?

b) Qual é a probabilidade de a peça ser proveniente da cavidade 1 ou 2?

c) 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:

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

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

Exemplo 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 $P(X = 1) = P(X = 2) = P(X = 3) = \ldots = P(X = 6) = 1/6$, 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:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{5.6}$$

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

$$P(A \cup B) = P(A) + P(B). \tag{5.7}$$

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

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n). \tag{5.8}$$

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,

$$A = \{\omega \in \Omega : \omega = \text{“lar X que não excede 240kWh em janeiro”}\},$$
$$B = \{\omega \in \Omega : \omega = \text{“lar X que não excede 240kWh em julho”}\}.$$

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

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

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:

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

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,

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

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:

$$P(A^C) = 1 - P(A). \tag{5.11}$$

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:

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

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 $Q : \Omega \to [0, 1]$ definida por

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

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

Prova:

Para verificarmos se $Q(.)$ 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

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

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

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

Prova:

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

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}$$
$$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1),$$

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

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

pela definição 5.18, temos

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

podendo ser reescrito como

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

Assim, usando a indução sucessivas vezes, chegaremos a expressão

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

Observe que por hipótese, todos os condicionamentos da expressão do lado direito, têm probabilidades positivas, pois contém $\bigcap_{i=1}^{n} A_i$, 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 $A_1, A_2, \ldots,$ 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 $A_1, A_2, \ldots, A_n$ 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 $i = 1, 2, \ldots, n$. □

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,

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

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). $(i) \to (ii)$: Se $P(AB) = P(A)P(B)$, então

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

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

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

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

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

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

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

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:

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

Que é o mesmo que entender que $P(A) \times P(B) = 1/6 = P(A \cap B)$. 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 $A_1, A_2, \ldots, A_n$ de Ω são independentes se e somente se:

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j), \quad \text{para } i \neq j; \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k), \quad \text{para } i \neq j \neq k; \\ &\vdots \\ P(\cap_{i=1}^{n} A_i) &= \prod_{i=1}^{n} P(A_i). \end{aligned} \tag{5.14}$$

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

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}, \qquad k = 1, 2, \ldots, n, \tag{5.15}$$

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

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 retirado de Morettin (2010), para elucidar essas definições.

Exemplo 5.23: Retirado de Morettin (2010)

A tabela abaixo dá a distribuição das probabilidade dos quatro tipos sanguíneos, numa certa comunidade.

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

Calcular a probabilidade de que:

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

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

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

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

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

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

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

$$\begin{aligned} P[(AB)^c \cup B^c] &= P[(AB)^c] + P(B^c) - P[(AB)^c \cap B^c] \\ &= 0,9000 + 0,9500 - P[(AB)^c \cap B^c]. \end{aligned} \tag{5.16}$$

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

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

Ao final, temos a tabela completada da seguinte forma:

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

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

Podemos ainda expressar mais 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:

$$P(A) = P(A \cap B) + P(A \cap B^c), \tag{5.18}$$
$$P(A^c) = P(A^c \cap B) + P(A^c \cap B^c), \tag{5.19}$$
$$P(B) = P(B \cap A) + P(B \cap A^c), \text{ e} \tag{5.20}$$
$$P(B^c) = P(B^c \cap A) + P(B^c \cap A^c), \tag{5.21}$$

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

$$\begin{aligned} A \cup B &= (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) \\ &= (A \cap B^c) \cup (A^c \cap B) \\ &\neq \Omega \quad \text{(Pelo fato de A e B não serem partição do espaço amostral).} \end{aligned} \tag{5.22}$$

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 $X(.)$. 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 $X : \Omega \to \mathbb{R}$ 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,

$$\Omega = \left\{ \begin{array}{ccccc} (A_1, A_2) & (A_1, E_1) & (A_1, E_2) & (A_1, E_3) & (A_2, E_1) \\ (A_2, E_2) & (A_2, E_3) & (E_1, E_2) & (E_1, E_3) & (E_2, E_3) \end{array} \right\}.$$

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.

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 dados valor x será denotado por $P(X = x)$, $p_X(x)$ ou p_i com $i = 1, 2, \ldots$, 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:

$$P(C_3) = P(\{\omega \in \Omega : \omega \in C_3\}) = P(\{(A_1, A_2)\}) = \frac{\#\{(A_1, A_2)\}}{\#\Omega} = \frac{1}{10}. \tag{5.23}$$

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,

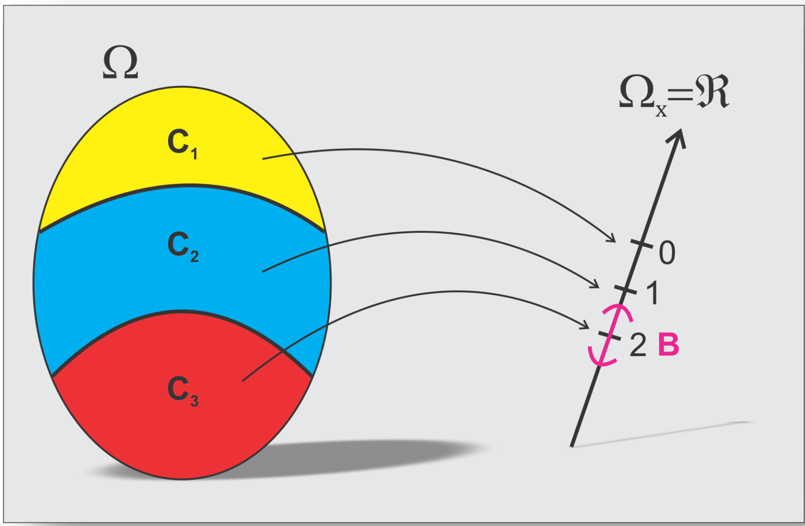


Figura 5.3: Espaço amostral e espaço amostral induzido pela variável aleatória X.

$$
\begin{aligned}
p_X(2) = P(X=2) &= P_X(B \in \Omega_X), \quad (B = \{2\}, \text{Figura 5.3}) \\
&= P_X(\{x \in \Omega_X : X(\omega) \in B,\ \omega \in \Omega\}) \\
&= P(X^{-1}(2)) \\
&= P(\{\omega \in \Omega : X(\omega) = 2\}) \\
&= P(\{\omega \in \Omega : \omega \in C_3\}) \\
&= P(\{(A_1, A_2)\}) \\
&= P(C_3) \\
&= \frac{1}{10}, \quad \text{(resultado (5.23))}. \qquad (5.24)
\end{aligned}
$$

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 $x_1, x_2, \ldots$, de uma variável aleatória discreta X, e suas respectivas probabilidades, p_1, p_2, $\ldots$, 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	$\ldots$
p_X	p_1	p_2	$\ldots$

Em notação, dizemos que a função de probabilidade é dada da seguinte forma:

$$P(X = x_i) = p_X(x_i) = p_i, \quad i = 1, 2, \ldots. \tag{5.25}$$

Uma distribuição de probabilidade satisfaz os seguintes axiomas:

1. $0 \leq p_i \leq 1$;
2. $\sum\limits_i p_i = 1$.

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 $f : \mathbb{R} \to \mathbb{R}$, 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:

$$P(a \leq X \leq b) = \int_a^b f(x)dx. \tag{5.26}$$

Portanto, como $P_X(X = x) = 0$, em se tratando de variáveis aleatórias contínuas, como consequência, as probabilidade 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 $e^{-x/128} \geq 0, \quad \forall x \geq 0$, 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,

$$\int e^{-mx}dx = \frac{-e^{-mx}}{m} + c. \tag{5.27}$$

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

$$\begin{aligned}
\int_0^{\infty} \frac{1}{128}e^{-\frac{1}{128}x}dx &= \frac{1}{128}\underbrace{\int_0^{\infty} e^{-\frac{1}{128}x}dx}_{\text{idêntica a } eq.(5.27)}, \\
&= \frac{1}{128}\left[\frac{-e^{-(1/128)x}}{1/128}\right]_0^{\infty}, \\
&= \frac{1}{128}\left[0 - \left(\frac{-e^{-(1/128)\times 0}}{1/128}\right)\right], \\
&= \frac{1}{128}\left[128e^{0}\right] \\
&= \frac{1}{128} \times 128 \times 1 = 1.
\end{aligned}$$

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 probilidade 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.

$$F(X) = \begin{cases} 0, & \text{se} \quad x < 1; \\ 0,23, & \text{se} \quad 1 \leq x < 2; \\ 0,50, & \text{se} \quad 2 \leq x < 3; \\ 0,80, & \text{se} \quad 3 \leq x < 4; \\ 0,92, & \text{se} \quad 4 \leq x < 5; \\ 1, & \text{se} \quad x \geq 5. \end{cases}$$

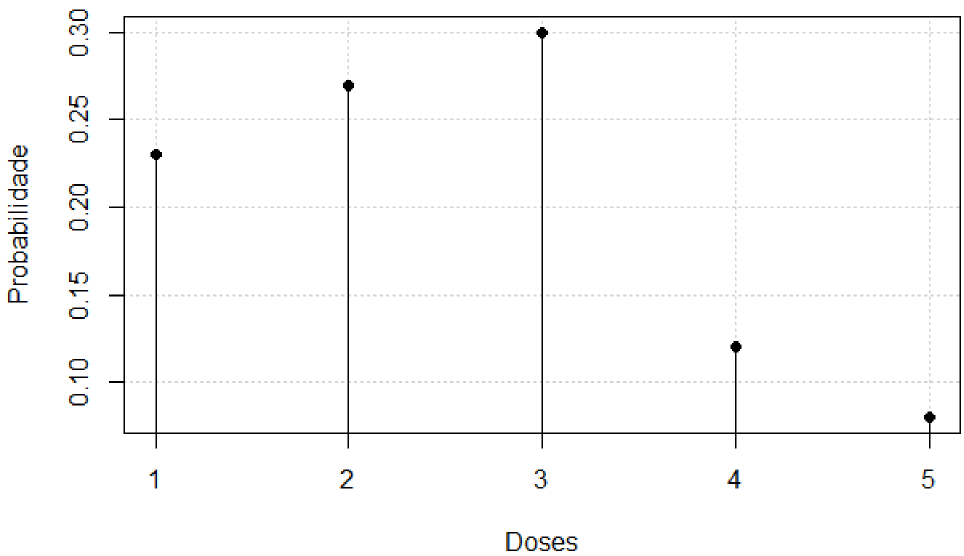


Figura 5.4: Gráfico da função de probabilidade.

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:

$$F(x) = \int_{-\infty}^{x} f(t)dt, \quad x \in \mathbb{R}. \tag{5.30}$$

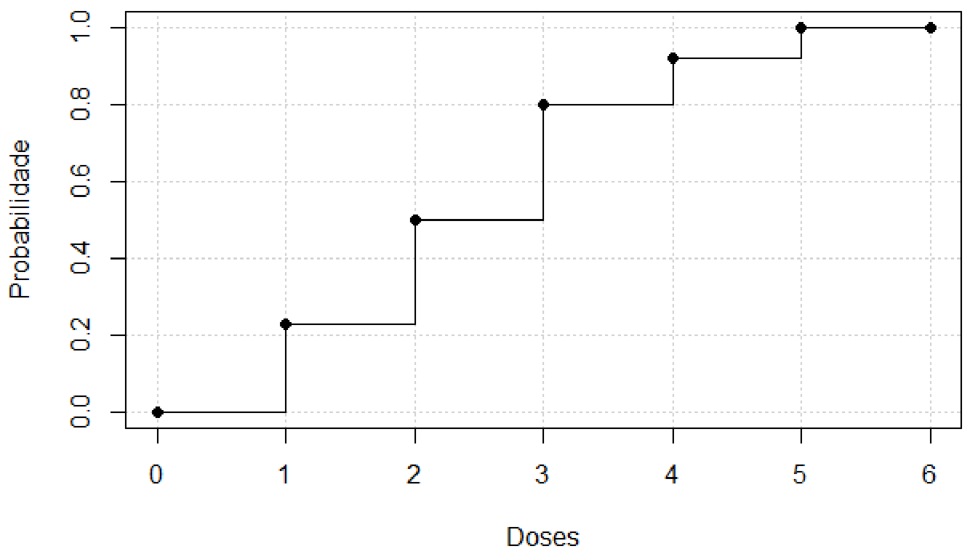


Figura 5.5: Gráfico da função de distribuição.

5.10.2.1 Propriedades

Apresentamos algumas propriedades da função de distribuição:

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

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

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

Essas propriedades podem ser melhor observadas pela Figura 5.6.

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:

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

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

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

5.11 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.

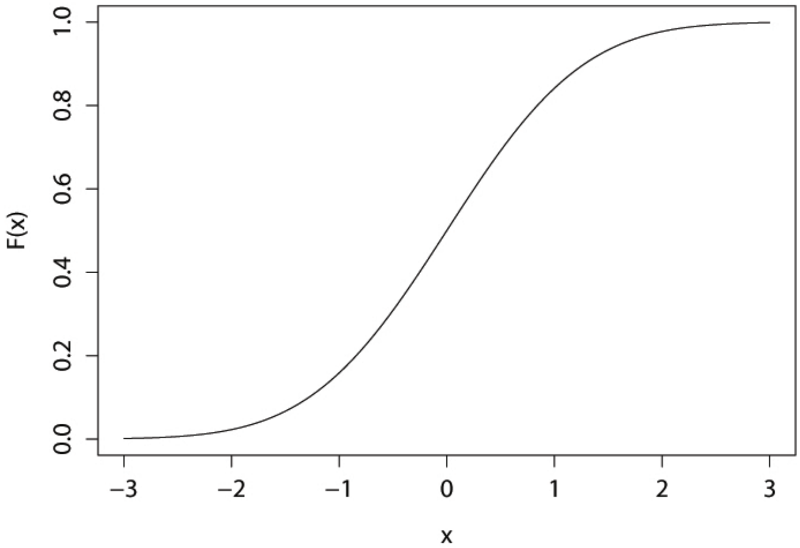


Figura 5.6: Gráfico da função de distribuição.

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,
$$E[X] = \sum_i x_i p_X(x_i), \tag{5.31}$$
para um conjunto de pontos $x_1, x_2, \ldots, x_i, \ldots$;

ii) se X for contínua,
$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx, \tag{5.32}$$
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:

$$P(X \geq \mu_d) = P(X \leq \mu_d) = 0,50, \tag{5.33}$$

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:

$$P(X = \mu_o) = \max(p_1, p_2, \ldots, p_n). \tag{5.34}$$

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,

$$E[X] = \sum_{i=1}^{n} x_i p(x_i) = (0) \times 0,3 + 5 \times 0,2 + \ldots + 20 \times 0,1 = 9 \text{ unid.}.$$

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 $E[X] = \mu$, enttão a variância de X, denotada por σ_X^2 ou $Var[X]$, é definida:

i) se X for discreta, por:

$$\sigma_X^2 = \sum_i (x_i - \mu)^2 P_X(x_i), \tag{5.35}$$

para um conjunto de pontos $x_1, x_2, \ldots, x_i, \ldots$;

ii) se X for contínua, por:

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x_i - \mu)^2 f_X(x) dx, \tag{5.36}$$

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 quadradado 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:

$$\sigma_X = +\sqrt{\sigma_X^2}, \tag{5.37}$$

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 $(\Omega, \mathcal{F}, P)$, então

$$Var[X] = E[X^2] - (E[X])^2, \tag{5.38}$$

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:

$$\begin{aligned} Var[X] &= E[X^2] + (E[X] - E[X]) - (E[X])^2 \\ &= E[X(X-1)] + E[X] - (E[X])^2. \end{aligned} \tag{5.39}$$

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

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

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

5.12 Exercícios

Exercício 5.1: Defina o que é um experimento aleatório e exemplifique.

Solução na página 142

Exercício 5.2: Defina o que é um espaço amostral e exemplifique.

Solução na página 142

Exercício 5.3: Para cada um dos casos abaixo, escreva o espaço amostral correspondente e conte seus elementos.

a) Uma moeda é lançada duas vezes e observam-se as faces obtidas.

b) Um dado é lançado duas vezes e a ocorrência de face par ou ímpar é observado.

c) Dois dados são lançados simultaneamente e estamos interessados na soma das faces observadas.

d) Em uma cidade, famílias com 3 crianças são selecionadas ao acaso, anotando-se o sexo de cada uma.

Solução na página 142

Exercício 5.4: Para cada um dos casos abaixo, escreva o espaço amostral correspondente e conte seus elementos.

a) Uma moeda é lançada duas vezes e observam-se as faces obtidas.

b) Um dado é lançado duas vezes e a ocorrência de face par ou ímpar é observado.

c) Dois dados são lançados simultaneamente e estamos interessados na soma das faces observadas.

d) Em uma cidade, famílias com 3 crianças são selecionadas ao acaso, anotando-se o sexo de cada uma.

Solução na página 142

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:

a) Ser esportista.

b) Ser esportista e aluno da Administração.

c) Ser esportista ou aluno da Ciências contábeis.

d) Não ser esportista nem aluno da Administração.

Solução na página 142

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 142

Exercício 5.7: Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{4}$, e A e B são multuamente exclusivos, calcular:

a) $P(A^c)$.

b) $P(B^c)$.

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

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

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

Solução na página 143

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 143

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 143

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 143

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 143

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 144

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 144

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 145

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 multuamente exclusivos.
2. A e B serem independentes.

Solução na página 145

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 145

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 145

Exercício 5.18: Três máquinas A, B, e C produzem, respectivamente 30%, 50% e 20% do total de peças de uma fábrica. As porcentagens e peças defeituosas nas respectivas máquinas são 3%, 5% e 2%. Uma peça é sorteada ao acaso, e verifica-se que é defeituosa. Qual a probabilidade de que a peça tenha vindo da máquina B?
Solução na página 146

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.

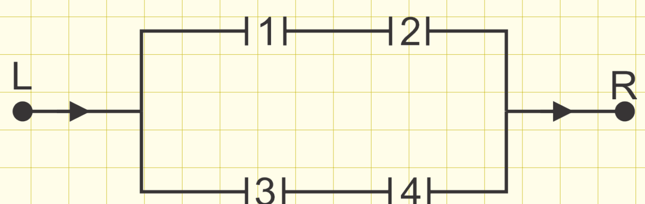
c) Todos que gostam de teatro gostam de cinema.

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 $3/4$.

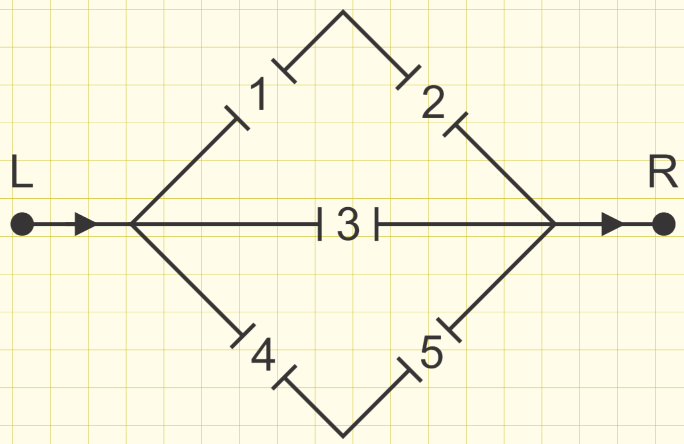
Solução na página 146

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 146

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 146

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 2.777.072 milhões de pessoas tomaram a primeira dose da vacina, 1.950.914 milhões de pessoas tomaram 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 147

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 147

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

$$P(C) = P(A) + P(A) - 2P(A \cap B).$$

Solução na página 147

Capítulo 6

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

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

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.

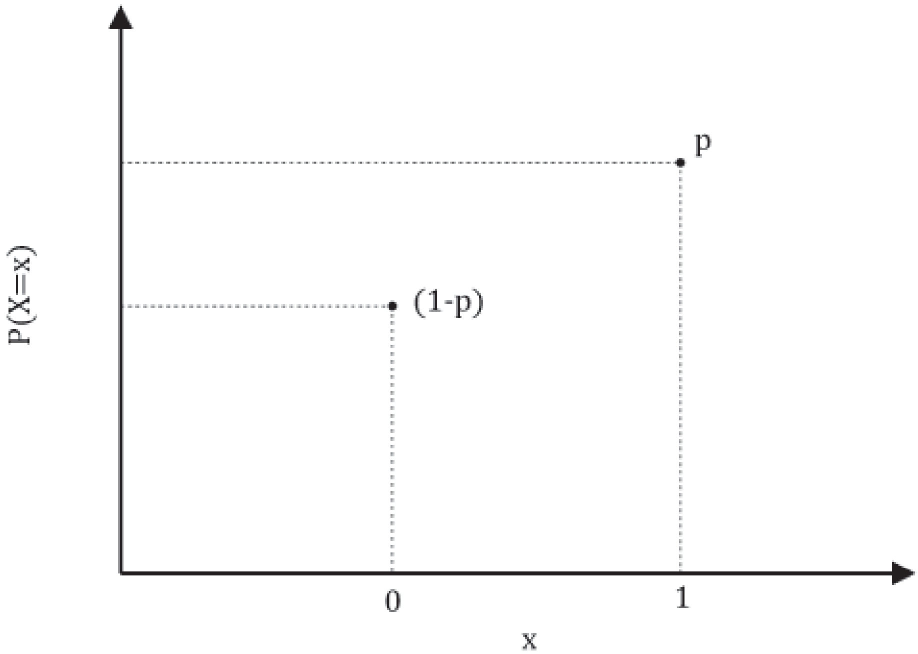


Figura 6.1: Função de probabilidade da distribuição Bernoulli com parâmetro p.

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:

$$\mu_X = p, \tag{6.2}$$

a variância é dada por:

$$\sigma_X^2 = p(1-p), \tag{6.3}$$

e o desvio padrão é dado por:

$$\sigma_X = \sqrt{p(1-p)}. \tag{6.4}$$

□

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

$$P(X = x) \quad = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{para } x = 0, 1, 2, \ldots, n, \\ 0, & \text{caso contrário,} \end{cases} \tag{6.5}$$

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, $X \sim Binomial(n, p)$ 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:

a) qual a distribuição de X?

b) 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, $Y \sim Bernoulli(p)$, 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 $X \sim Binomial(n, p)$. 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.

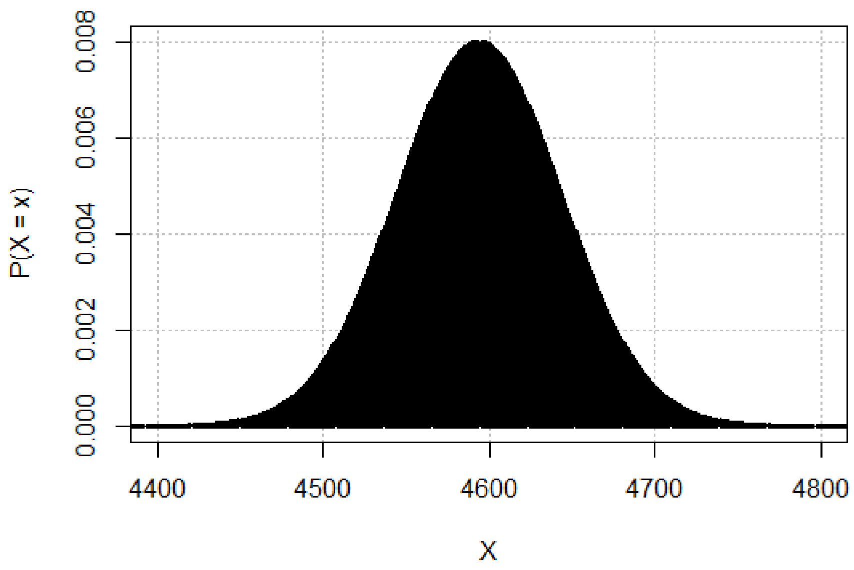


Figura 6.2: Distribuição de $X \sim \mathit{Binomial}(10{,}000,\ 10^9/36^6)$.

Código R 6.1

Script:

```
# Valores
n <- 10000; x <- 0:n; p <- 0.459394
# Probabilidade
px <- dbinom(x, size = n, prob = p)
# Gráfico
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 $\sum_{x=0}^{n} P(X = x) = 1$.

i) 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

$$\begin{aligned}
x = 0 \quad &\Rightarrow \quad \tbinom{3}{0} p^0 (1-p)^{3-0} = (1-p)^3 = q^3, \\
x = 1 \quad &\Rightarrow \quad \tbinom{3}{1} p^1 (1-p)^{3-1} = 3p(1-p)^2 = 3pq^2, \\
x = 2 \quad &\Rightarrow \quad \tbinom{3}{2} p^2 (1-p)^{3-2} = 3p^2(1-p)^1 = 3p^2 q, \\
x = 3 \quad &\Rightarrow \quad \tbinom{3}{3} p^3 (1-p)^{3-3} = p^3(1-p)^0 = p^3.
\end{aligned}$$

Agora observe,

$$(p+q)^3 = p^3 + 3p^2 q + 3pq^2 + q^3,$$

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

$$\sum_{x=0}^{n} P(X = x) = (p+q)^n. \tag{6.6}$$

Sabendo que $q = 1 - p$, então

$$\begin{aligned}
\sum_{x=0}^{n} P(X = x) &= [p + (1-p)]^n, \\
&= 1^n = 1. \qquad (6.7)
\end{aligned}$$

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:

```
#---------------------------------------------------
# Consideramos a alteracao do valor do parametro p
#---------------------------------------------------
# Grafico da distribuicao binomial para n = 30 e p = 0.5
x <- 0:30; n <- 30; p <- 0.5
px <- dbinom(x, size = n, prob = p)
plot(x, px, type = "h", xlab = "X", ylab = expression("P(X=x)"), panel.first
    = grid(col="gray"), ylim = c(0, 0.5), lwd = 2)
# n = 30 e p = 0.1
x <- 0:30; n <- 30; p <- 0.1
px <- dbinom(x, size = n, prob = p)
lines(x, px, col = "red", type = "h", lwd = 2)
# n = 30 e p = 0.9
x <- 0:30; n <- 30; p <- 0.9
px <- dbinom(x, size = n, prob = p)
lines(x, px, col = "green", type = "h", lwd = 2)
# Legendas
legend(20, 0.5, legend = c("n=30;p=0.5",
                           "n=30;p=0.1",
                           "n=30;p=0.9"),
       col = c("black", "red", "green"),
       bg = "antiquewhite1", lty = 1, cex = 0.8, lwd = 2)
```

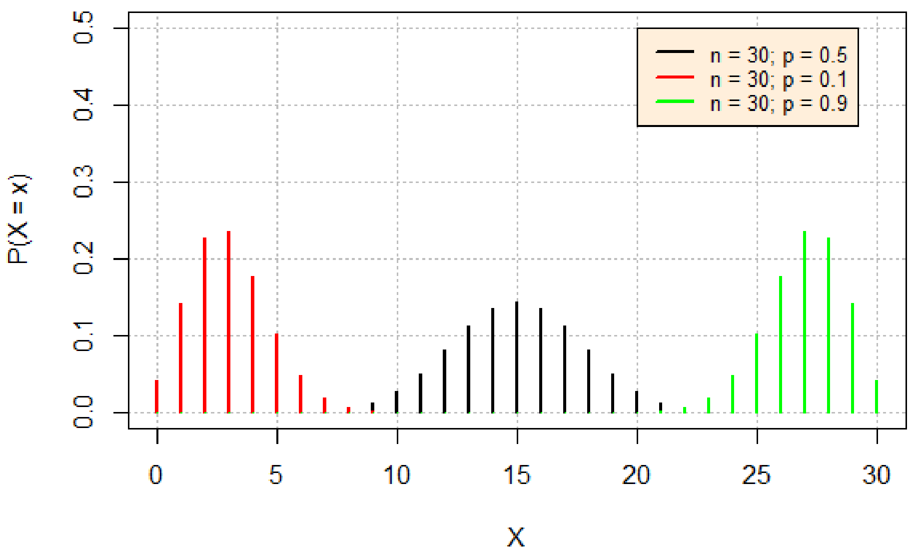


Figura 6.3: Gráfico da função de probabilidade da distribuição binomial, de parâmetros $n = 30$ e diferentes valores de p.

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:

$$E[X] = np, \tag{6.8}$$

a variância é dada por:

$$\sigma_X^2 = np(1-p). \tag{6.9}$$

e o desvio padrão é dado por:

$$\sigma_X = \sqrt{np(1-p)}. \tag{6.10}$$

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 $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes e identicamente distribuídas[a] (*iid*) com distribuição de Bernoulli de parâmetro p, então $Y = \sum_{i=1}^{n} X_i \sim Binomial(n, p)$. Logo,

$$\mu_Y = E[Y] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = nE[X] = np,$$

considerando $E[X]$ em (6.2). A variância pode ser obtida pelo mesmo raciocínio feito para obter a esperança, isto é,

$$\sigma_Y = Var[Y] = Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} Var[X_i] = nVar[X] = np(1-p),$$

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:

$$
\begin{aligned}
\mu = E[X] &= \sum_{x=0}^{n} x p_X(x) \\
&= 0 \times p_X(0) + \sum_{x=1}^{n} x p_X(x) \\
&= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!x}{x!(n-x)!} p p^{x-1} (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)!(n-x)!} p p^{x-1} (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}.
\end{aligned}
$$

Usando a transformação $y = x - 1$, então

$$
\begin{aligned}
\mu = E[X] &= np \underbrace{\sum_{y=0}^{n-1} \frac{(n-1)!}{y![(n-1)-x]!} p^y (1-p)^{(n-1)-x}}_{=1,\text{ consequência do resultado (6.7).}} \\
\mu = E[X] &= np, \qquad (6.11)
\end{aligned}
$$

como queríamos demonstrar. Para o cálculo da variância, calcularemos inicialmente $E[X^2]$, que segue:

$$
\begin{aligned}
E[X^2] &= \sum_{x=0}^{n} x^2 p_X(x) \\
&= 0 \times p_X(0) + \sum_{x=1}^{n} x^2 p_X(x) \\
&= \sum_{x=1}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!x^2}{x!(n-x)!} p p^{x-1} (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n(n-1)!x}{(x-1)!(n-x)!} p p^{x-1} (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!x}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}.
\end{aligned}
$$

Usando a transformação $y = x - 1$, então

$$
\begin{aligned}
E[X^2] &= np \sum_{y=0}^{n-1} (y+1) \frac{(n-1)!}{y![(n-1)-y]!} p^y (1-p)^{(n-1)-y} \\
&= np \left[\underbrace{\sum_{y=0}^{n-1} y \frac{(n-1)!}{y![(n-1)-y]!} p^y (1-p)^{(n-1)-y}}_{E[Y]=(n-1)p} + \underbrace{\sum_{y=0}^{n-1} \frac{(n-1)!}{y![(n-1)-y]!} p^y (1-p)^{(n-1)-y}}_{=1} \right] \\
&= np[(n-1)p + 1] \\
&= np + (np)^2 - np^2. \qquad (6.12)
\end{aligned}
$$

Usando a Definição 5.24 e substituindo os termos pelas expressões (6.11) e (6.12), então

$$
\begin{aligned}
\sigma_X^2 = Var[X] &= np + (np)^2 - np^2 - (np)^2 \\
&= np - np^2 \\
&= np(1-p), \qquad (6.13)
\end{aligned}
$$

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.

[a] O 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 que as variáveis aleatórias 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 senha 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

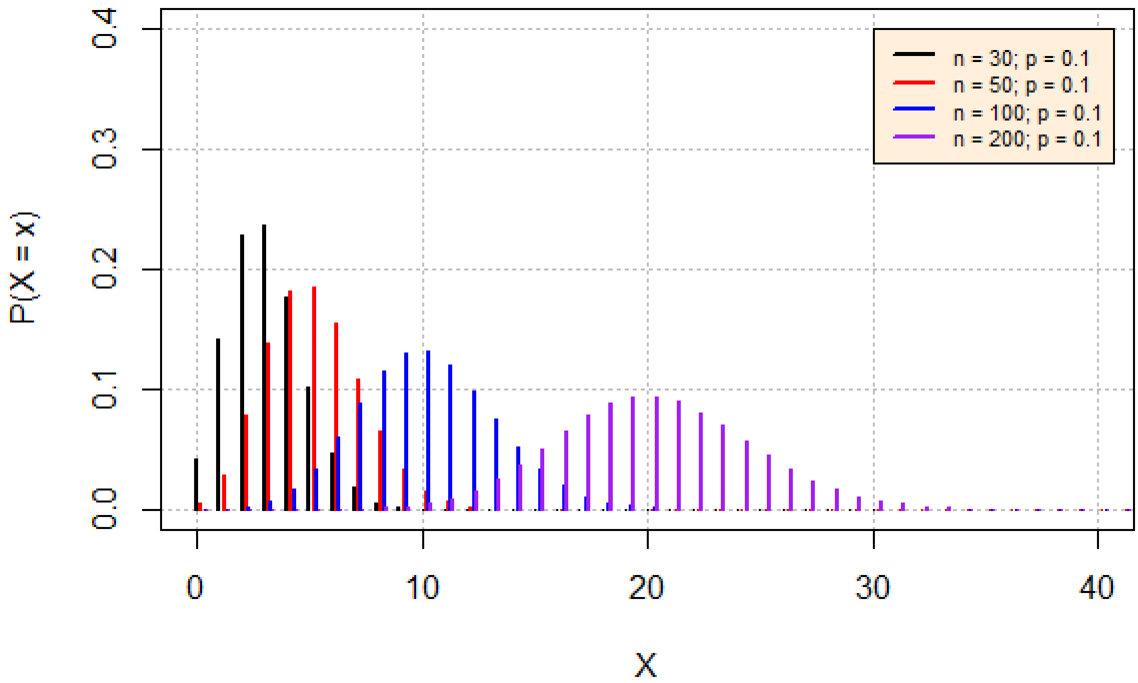


Figura 6.4: Gráfico da função de probabilidade da distribuição binomial, de parâmetros diferentes de n e $p = 0.1$.

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 o 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:

```
#-----------------------------------------------
# Consideramos a alteracao do valor do parametro n
#-----------------------------------------------
# Grafico da distribuicao binomial para n = 30 e p = 0.1
n <- 30; x <- 0:n; p <- 0.1
px <- dbinom(x, size = n, prob = p)
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)
# n = 30 e p = 0.1
n <- 50; x <- 0:n; p <- 0.1
px <- dbinom(x, size = n, prob = p)
lines(x + 0.2, px, col = "red", type = "h", lwd = 2)
# n = 100 e p = 0.1
n <- 100; x <- 0:n; p <- 0.1
px <- dbinom(x, size = n, prob = p)
lines(x + 0.3, px, col = "blue", type = "h", lwd = 2)
# n = 200 e p = 0.1
n <- 200; x <- 0:n; p <- 0.1
px <- dbinom(x, size = n, prob = p)
lines(x + 0.4, px, col = "purple", type = "h", lwd = 2)
# Legendas
legend(30, 0.4, legend = c("n=30;p=0.1",
                           "n=50;p=0.1",
                           "n=100;p=0.1",
                           "n=200;p=0.1"),
       col = c("black", "red", "blue", "purple"),
       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 $n \to \infty$ e $p \to 0$, 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 $(n)_x = n(n-1)\ldots(n-x+1)$, 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 $\left(1-\frac{\lambda}{n}\right)^{n-x} = \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}$, então

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

Considerando agora que a probabilidade $p \to 0$ com $n \to \infty$, λ passa a ser considerado como aproximadamente constante. Na aplicação do limite observe que

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

e por um resultado clássico de limite, segue também que

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

Desse modo, concluímos que

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

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 $X \sim Binomial(250 \times 4, 1/500)$. 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 $n \to \infty$ e $p \to 0$.

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

$$P(X=x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & \text{para } x = 0,1,2,\ldots, \\ 0, & \text{caso contrário,} \end{cases} \tag{6.14}$$

em que $\lambda > 0$. Em notação, $X \sim Poisson(\lambda)$ representa que X tem distribuição Poisson com parâmetro λ.

Vejamos o Exemplo 6.6, para apresentar uma aplicação dessa distribuição.

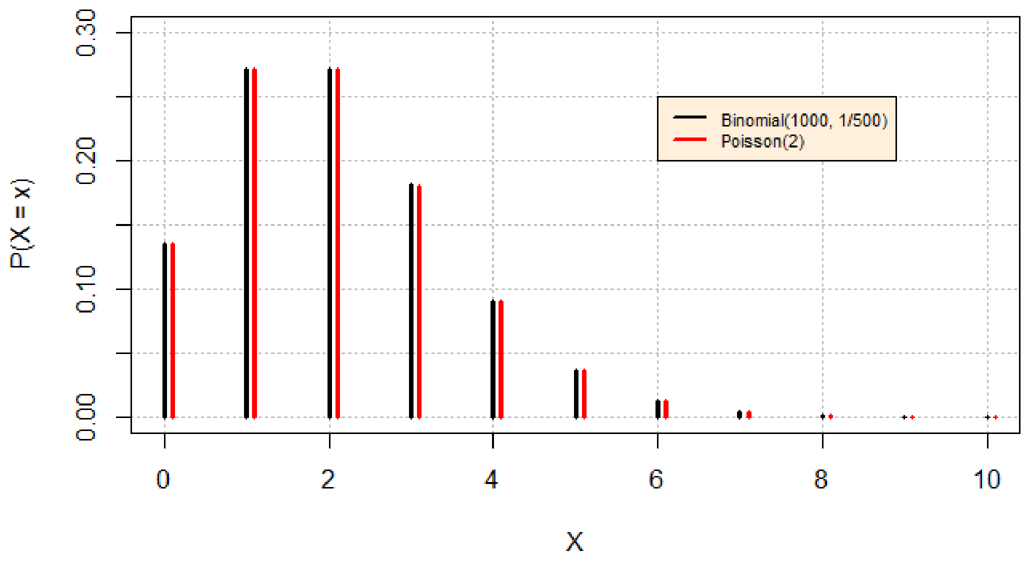


Figura 6.5: Relação entre as distribuições Poisson e Binomial para $n = 1000$, $p = 1/500$ e $lambda = np$

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.

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,

$$\begin{cases} \lambda^* & \rightarrow \quad 1,5 \\ 0,2 & \rightarrow \quad 1 \end{cases} \quad \therefore \lambda^* = 0,3 \text{ falhas por } 1,5\ m^2.$$

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 demostrar os itens:

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

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

i) Sabe-se que $\lambda > 0$, $x = 0, 1, 2, \ldots$, e a constante $e = 2,7188281...$, então

- $e^{-\lambda} > 0$,
- $\lambda^x > 0$,
- $x! > 0$,

portanto, $P(X = x) \geq 0$.

ii) Tendo,

$$
\begin{aligned}
\sum_{x=0}^{\infty} P(X = x) &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}, \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}, \\
&= e^{-\lambda} \left[\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^x}{x!} + \ldots \right], \\
&= e^{-\lambda} \left[\underbrace{1 + \lambda + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^x}{x!} + \ldots}_{\text{série de Taylor}} \right],
\end{aligned}
$$

e sabendo que $e^x = 1 + x + \frac{x^2}{2!} + \ldots$, 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

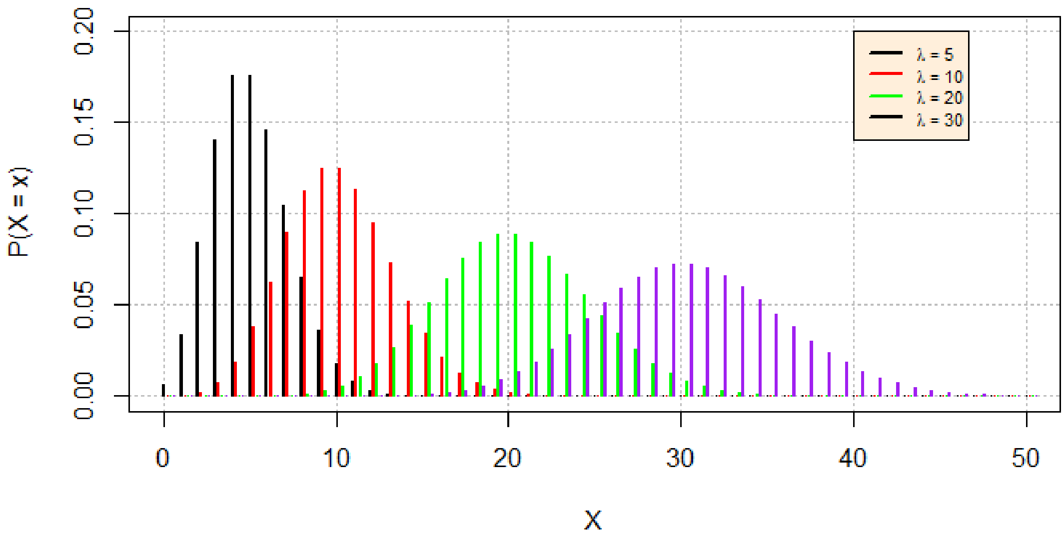


Figura 6.6: Comportamento da distribuição Poisson com relação ao parâmetro λ.

Código R 6.4

Script:

```r
#-------------------------------------------------------
# Consideramos a alteracao do valor do parametro lambda
#-------------------------------------------------------
# Grafico da distribuicao poisson para lambda=5
x <- 0:50; lambda = 5
px <- dpois(x, lambda = lambda)
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)
# n = 30 e lambda = 10
x <- 0:50; lambda = 10
px <- dpois(x, lambda = lambda)
lines(x+0.2, px, col = "red", type = "h", lwd = 2)
# n = 30 e lambda = 20
x <- 0:50; lambda = 20
px <- dpois(x, lambda = lambda)
lines(x+0.4, px, col = "green", type = "h", lwd = 2)
# n = 30 e lambda = 30
x <- 0:50; lambda = 30
px <- dpois(x, lambda = lambda)
lines(x+0.6, px, col = "purple", type = "h", lwd = 2)
# Legendas
legend(40, 0.2, legend = c(expression(lambda~"=5"),
                           expression(lambda~"=10"),
                           expression(lambda~"=20"),
                           expression(lambda~"=30")),
       col = c("black", "red", "green"),
       bg = "antiquewhite1", lty = 1, cex = 0.7, lwd = 2)
```

Para completarmos a caracterização dessa distribuição, apresentamos o Teorema 6.6 para a esperança, 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:

$$\mu_X = E[X] = \lambda, \tag{6.15}$$

a variância é dada por:

$$\sigma_X^2 = Var[X] = \lambda, \tag{6.16}$$

e o desvio padrão é dado por:

$$\sigma_X = \sqrt{\lambda}. \tag{6.17}$$

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,

$$\begin{aligned}
Var(X) &= \sum_{y=0}^{\infty}(y+1)^2\frac{e^{-\lambda}\lambda^{(y+1)}}{(y+1)!} - \lambda^2,\\
&= \sum_{y=0}^{\infty}(y+1)^2\frac{e^{-\lambda}\lambda^{(y+1)}}{(y+1)y!} - \lambda^2,\\
&= \sum_{y=0}^{\infty}(y+1)\frac{e^{-\lambda}\lambda^{(y+1)}}{y!} - \lambda^2,\\
&= \sum_{y=0}^{\infty}y\frac{e^{-\lambda}\lambda^{y}\lambda}{y!} + \sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^{y}\lambda}{y!} - \lambda^2,\\
&= \lambda\left[\underbrace{\sum_{y=0}^{\infty}y\frac{e^{-\lambda}\lambda^{y}}{y!}}_{=E[X]=\lambda} + \underbrace{\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^{y}}{y!}}_{=1} - \lambda\right],\\
&= \lambda(\cancel{\lambda} + 1 - \cancel{\lambda}) = \lambda.
\end{aligned}$$

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 papeis 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,

$$\begin{cases} \lambda^* & \rightarrow & 1,5 \\ 0,2 & \rightarrow & 1 \quad \therefore \lambda^* = 0,3 \text{ falhas por } 1,5\ m^2. \end{cases}$$

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 de 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

$$f_X(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, & \text{para} \quad -\infty < x < \infty, \\ 0, & \text{caso contrário,} \end{cases} \tag{6.18}$$

em que os parâmetros μ e σ^2 satisfazem $\mu \in \mathbb{R}$ e $\sigma^2 > 0$. Em notação, $X \sim Normal(\mu, \sigma^2)$ representa que X tem distribuição Normal com parâmetros μ e σ^2.

Código R 6.5

Script:

```r
# Triangulo pascal

# Codigo para linhas acumuladas. Por exemplo, 'linhas = 10',
# imprime as linhas de 1 a 10 do trianbulo de Pascal
tpacum <- function(linhas) {
  lapply(0:linhas, function(i) choose(i, 0:i))
}

# Codigo para imprimir uma determinada linha do triangulo
# de Pascal. Por exemplo, 'linha = 10', imprime apenas a
# linha 10
tpind <- function(linha) {
  choose(linha, 0:linha)
}

# Gerando os graficos das linhas do triangulo de pascal
# como barras do gráfico de barras. Vamos gerar os graficos
n <- c(10, 25, 50, 100, 500, 1000)
par(mfrow = c(3, 2))
for(i in n) {
  x <- tpind(i)
  plot(x, type = "h", xlab = paste("Ordem␣dos␣coeficientes␣da
␣␣␣␣␣␣␣linha␣", i), ylab = "Valor␣do␣coeficiente", panel.first =
         grid(col="gray"), lwd = 3)
}
```

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) $f_X(x) \to 0$ quando $x \to \pm\infty$;

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

A esperança de X é dada por:

$$E[X] = \mu. \tag{6.19}$$

A variância é dada por:

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

O desvio padrão é dado por:

$$DP(X) = \sqrt{\sigma^2} = \sigma. \tag{6.20}$$

6.3.1 Normal padrão

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

$$Z = \frac{X - \mu}{\sigma}, \tag{6.21}$$

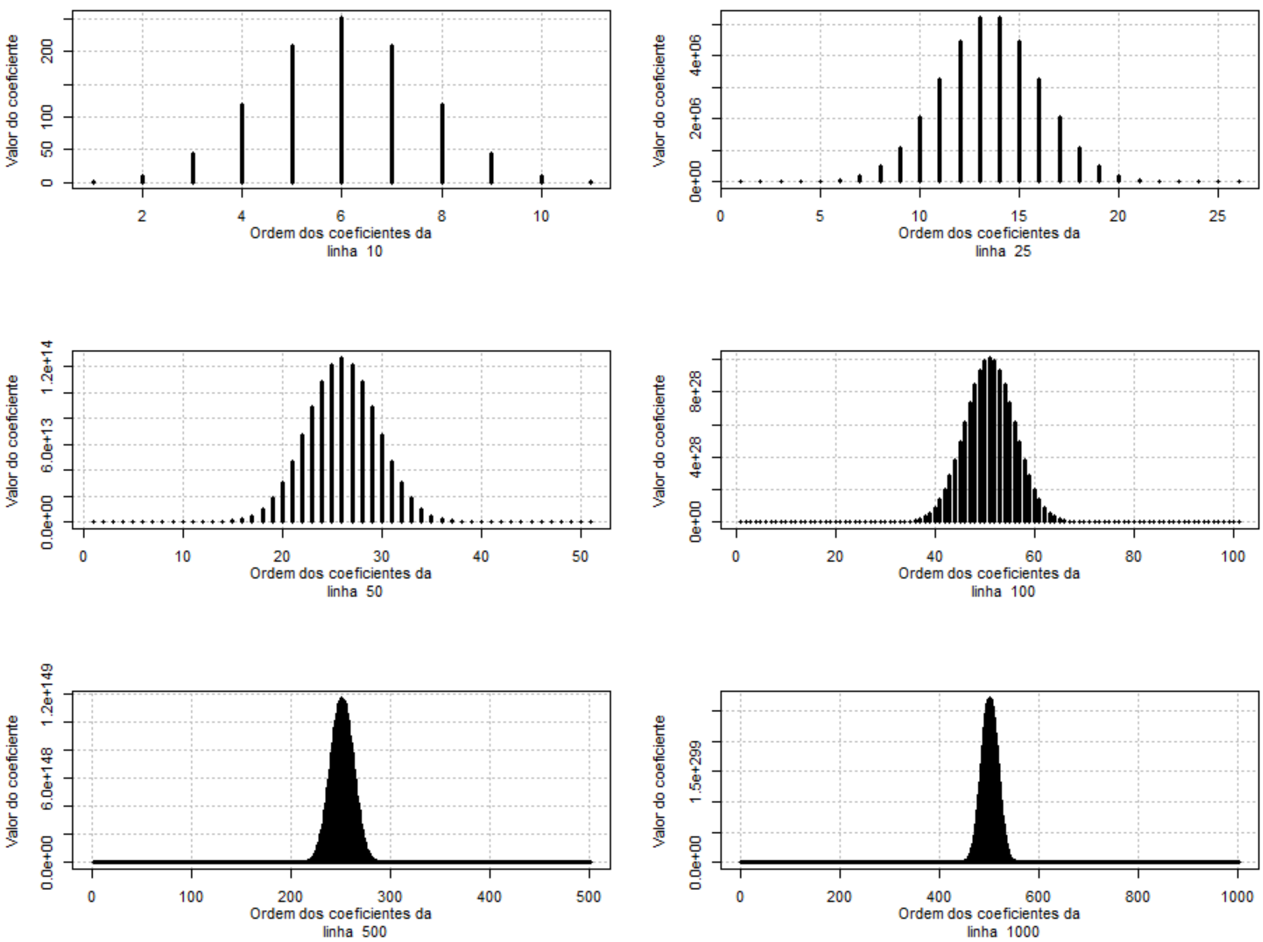


Figura 6.7: Gráfico de barras dos coeficientes para determinadas linhas do triângulo Pascal

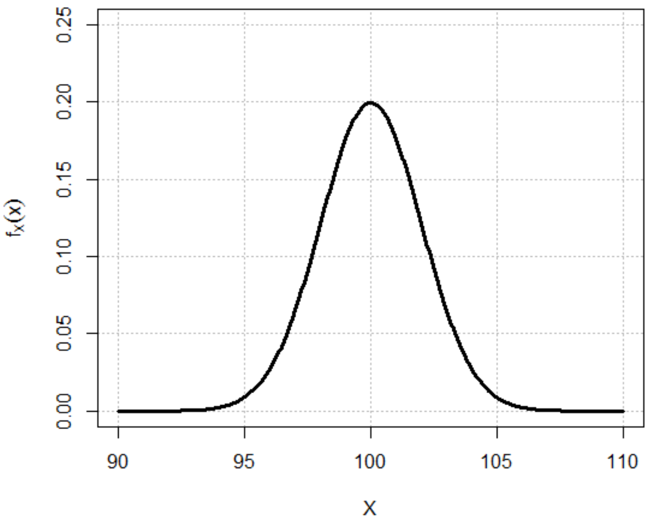


Figura 6.8: Função densidade da distribuição normal.

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

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

e

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

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

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 148

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 148

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 148

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 149

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 150

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 150

Exercício 6.7: Na 1ª prova de Estatística a média foi 4,5 e o desvio padrão 2,3. Considerando o método científico de aprovação:

- Conceito A: nota média $\geq \mu + \sigma$;
- Conceito B: $\mu \leq$ média $\leq \mu + \sigma$;
- Conceito C: $\mu - \sigma \leq$ média $\leq \mu$;
- Conceito D: média $\leq \mu - \sigma$;

a) Quantos alunos receberam cada um dos conceitos?

b) Quantos alunos foram aprovados (Conceito A, B e C);

c) Quantos alunos foram aprovados com distinção (média $\geq \mu + 2\sigma$);

d) Considere o método comum de aprovação ($X \geq 5,0$), quantos foram reprovados.

OBSERVAÇÃO: 100 alunos fizeram essa prova.

Solução na página 150

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 151

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:

a) nenhum defeito;

b) apenas um defeito;

c) não mais que um defeito;

d) acima de dois defeitos.

Solução na página 151

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 independente. 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 151

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.

1. Qual a probabilidade de que não haja alterações em um dia?
2. Qual a probabilidade de ocorra mais do que uma alteração em 8 horas?

3. Qual a probabilidade de ocorrer no máximo duas alterações em 2 dias?

4. Qual o valor esperado dessas alterações por dia? E a variabilidade dessas alterações em torno de 0,30 alterações por dia?

Solução na página 152

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 1000 consumidores apresentam comportamentos independentes quanto ao interesse pelas compras nesse site, determine:

a) nenhum consumir respondeu 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 152

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 previlegiadas 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 152

Amostragem

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.

A ideia é mostrar como os resultados de uma pesquisa pode ser desastrosas!

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.

A partir daqui, vou utilizar a notação de Fuller (2009) e Bolfarine (2005).

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!

- Apresentar os exemplos em que os elementos da população não se remete apenas a pessoas;
- falar da notação do tamanho da população N;
- falar sobre população finita ou infinita e apresentar exemplos; casos em que a população é um rio, como delimitar os elementos.
- mencionar que o estudo realizado sobre todos os elementos da população, denominamos Censo;
- Enfatizar que no Censo não precisamos realizar inferências;
- Porém, no censo, o custo, o tempo, problemáticas que surgem, impedem muitas vezes de trabalhar com a população;
- Daí abrimos mão do todo, e usamos uma amostra;
- notação para o tamanho da amostra n
- a amostra deve preservar as principais características da população. Quando isso ocorre, dizemos que essa amostra é representativa, é uma amostra aleatória. Quando isso não ocorre, temos uma amostra tendenciosa. Apresentar situações reais de amostra representativa e amostra tendenciosa;
- mostrar também que temos um conjunto de técnicas de como proceder em obter amostras, dependendo da população, e que esse conjunto de técnicas será estudado no capítulo 5.

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 conclomerados 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 estimação nada mais é do que criarmos mecanismos para estrapolar 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 comentendo 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.

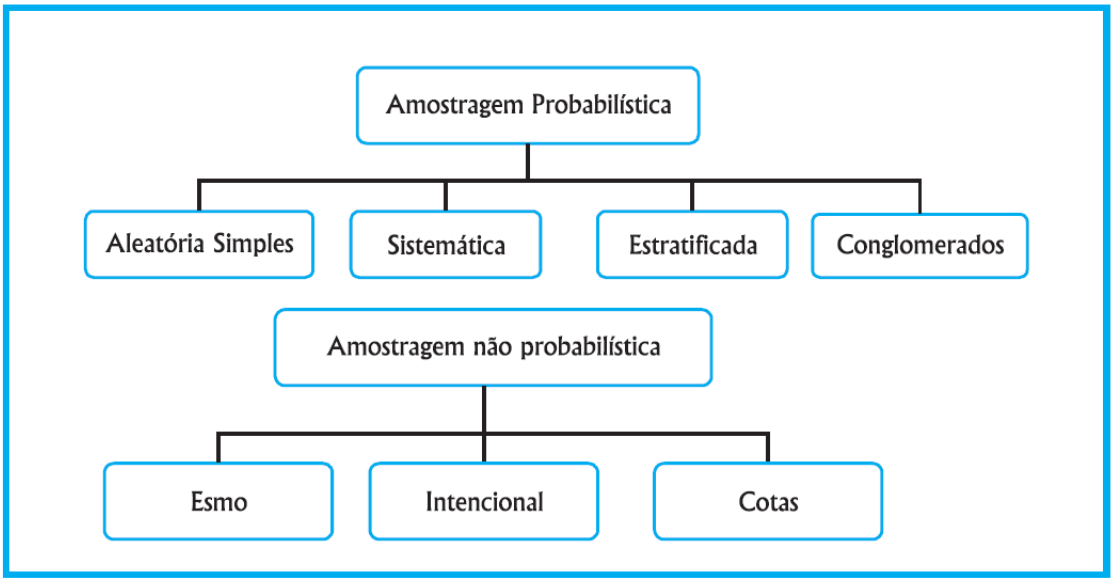


Figura 7.1: Métodos de amostragem probabilística e não-probabilística.

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

Capítulo 9

Teoria da estimação

Teoria da decisão

10.1 Introdução

10.2 Testes de hipóteses

10.3 Erros tipo I e II

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

Capítulo 11

Correlação e regressão linear simples

11.1 Introdução

11.2 Correlação linear

11.2.1 Coeficiente de correlação linear

11.2.2 Teste de hipóteses acerca do coeficiente de correlação linear

11.3 Regressão linear simples

11.3.1 Modelo

11.3.2 Estimação dos parâmetros do modelo

11.3.3 Teste de hipóteses para o modelo de regressão

11.3.4 Medidas de adequação do modelo

Apêndice A - Introdução ao R

....

Gabarito dos Exercícios

Solução dos Exercícios do Capítulo 1

Solução do Exercício 1.1 na página 16:
Não Disponível!

Solução do Exercício 1.2 na página 16:
Não Disponível!

Solução do Exercício 1.3 na página 16:
Não Disponível!

Solução do Exercício 1.4 na página 16:
Não disponível!

Solução do Exercício 1.5 na página 16:
Não Disponível!

Solução do Exercício 1.6 na página 17:
Não Disponível!

Solução do Exercício 1.7 na página 17:
Não Disponível!

Solução do Exercício 1.8 na página 17:

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:

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

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

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

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

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

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

Certificando se o ponto é de máximo ou de mínimo,

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

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

Solução dos Exercícios do Capítulo 2

Solução do Exercício 2.1 na página 27:
Como desejamos não incluir as condições limiares às indagações, podemos expressar as frequências acumuladas *abaixo de* e *acima de* como

$$F_{ac\downarrow_i} = \sum_{j=1}^{i-1} F_j, \quad i = 1, 2, \ldots, k, \tag{13.1}$$

e

$$F_{ac\uparrow_i} = \sum_{j=i+1}^{k} F_j, \quad i = 1, 2, \ldots, k, \tag{13.2}$$

respectivamente, sendo k o número de grupos ou classes.

Solução do Exercício 2.2 na página 27:

Solução do Exercício 2.3 na página 28:

Solução dos Exercícios do Capítulo 3

Solução do Exercício 3.1 na página 45:

Solução do Exercício 3.2 na página 45:

Solução do Exercício 3.3 na página 45:

Solução do Exercício 3.4 na página 45:

Solução do Exercício 3.5 na página 45:

Solução do Exercício 3.6 na página 45:

Solução dos Exercícios do Capítulo 4

Solução do Exercício 4.1 na página 61:

Solução do Exercício 4.2 na página 61:

Solução do Exercício 4.3 na página 61:

Solução do Exercício 4.4 na página 61:

Solução do Exercício 4.5 na página 61:

Solução do Exercício 4.6 na página 61:

Solução do Exercício 4.7 na página 62:

Solução dos Exercícios do Capítulo 5

Solução do Exercício 5.1 na página 96:
Ver as definições mostradas em sala de aula.

Solução do Exercício 5.2 na página 96:
Ver as definições mostradas em sala de aula.

Solução do Exercício 5.3 na página 96:
Ver as definições mostradas em sala de aula.

Solução do Exercício 5.4 na página 96:

a) $\Omega = \{CC, CK, KC, CC\}$, Cara - C; Coroa - K.

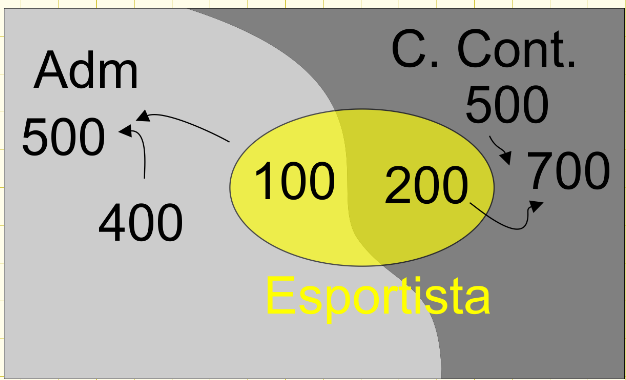
b) Ω = {(Par, Par), (Par, Ímpar), (Ímpar, Par), (Ímpar, Ímpar) }.

c) $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

d) O espaço amostral para esse experimento aleatório é:

$$\Omega = \{(F, F, F), (F, F, M), (F, M, F), (M, F, F), (F, M, M), (M, M, F), (M, F, M), (M, M, M)\}.$$

Solução do Exercício 5.5 na página 96:



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 97:
$P(A \cup B) = P(A) + P(B) - P(A \cap B) \Rightarrow P(B) = P(A \cup B) + P(A \cap B) - P(A) = 0,5 + 0,1 - 0,2 = 0,4$

Solução do Exercício 5.7 na página 97:
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 97:
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 97:

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 $P(C) = \frac{\#\{1^{\text{o}} \text{ Resultado ser maior que o } 2^{\text{o}} \text{ Resultado}\}}{36} = \frac{15}{36}$.

Solução do Exercício 5.10 na página 97:
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 97:
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:

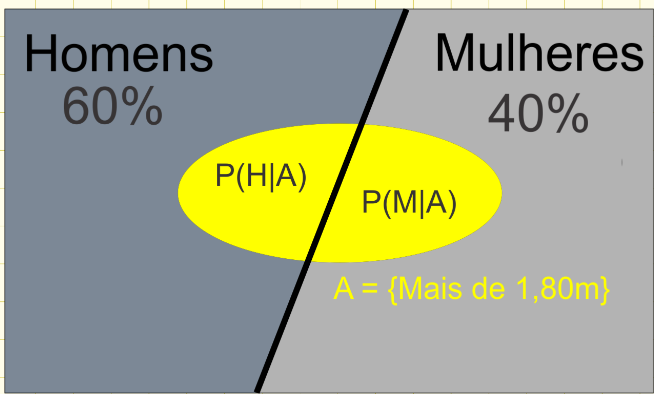
$$\begin{aligned} & P(J_1 \cap J_2^c \cap J_3^c) + P(J_1^c \cap J_2 \cap J_3^c) + P(J_1^c \cap J_2^c \cap J_3) \\ & \left(\frac{2}{3} \times \frac{1}{5} \times \frac{3}{10}\right) + \left(\frac{1}{3} \times \frac{4}{5} \times \frac{3}{10}\right) + \left(\frac{1}{3} \times \frac{1}{5} \times \frac{7}{10}\right) \\ & \frac{1}{25} + \frac{2}{25} + \frac{7}{150} \\ & \frac{1}{6}. \end{aligned}$$

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 98:
Considerando os eventos V: bola de volei; B: bola de basquete; A_1: bola no armário 1; A_2: bola no armário 2, assim
a) $P(V|A_1) = \frac{P(V \cap A_1)}{A_1} = \frac{3/9}{4/9} = \frac{3}{4}$.
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, $P(A_2^*) = P(A_2^*) = 1/2$. 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 98:



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

$$\begin{aligned} P(M|A) &= \frac{P(M \cap A)}{P(A)} \\ &= \frac{0,008}{0,038} = 0,2105. \end{aligned}$$

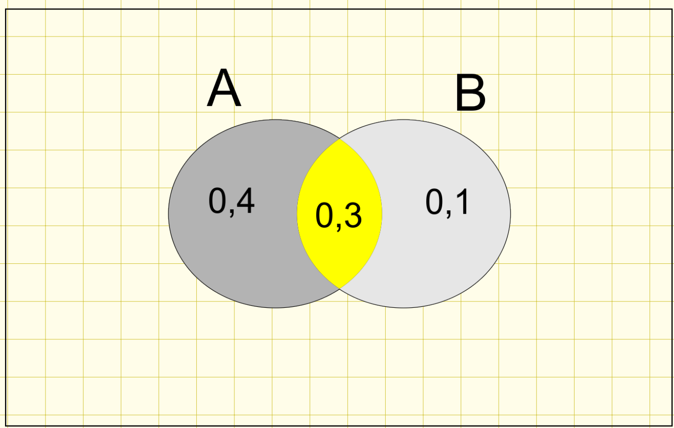
Solução do Exercício 5.14 na página 98:
Considere o evento H - homem está vivo; e o evento M - mulher está viva. Assim,
a) $P(H \cap M^c) = P(H) \times P(M^c) = \frac{3}{5} \times \frac{1}{4} = \frac{3}{20}$.
b) $P(H^c \cap M) = P(H^c) \times P(M) = \frac{2}{5} \times \frac{3}{4} = \frac{6}{20}$.
c) $P(H \cap M) = P(H) \times P(M) = \frac{3}{5} \times \frac{3}{4} = \frac{9}{20}$.

Solução do Exercício 5.15 na página 98:
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 98:



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 98:
Problema resolvido pelo teorema de Bayes. Consideremos o evento G o evento das vitórias do São Paulo, e X o evento de chover. Assim,

$$\begin{aligned} P(X|G) &= \frac{P(G|X)P(X)}{P(G|X)P(X) + P(G|X^c)P(X^c)} \\ &= \frac{0,7 \times 0,3}{0,7 \times 0,3 + 0,8 \times 0,7} \\ &= \frac{0,21}{0,77} = 0,2727. \end{aligned}$$

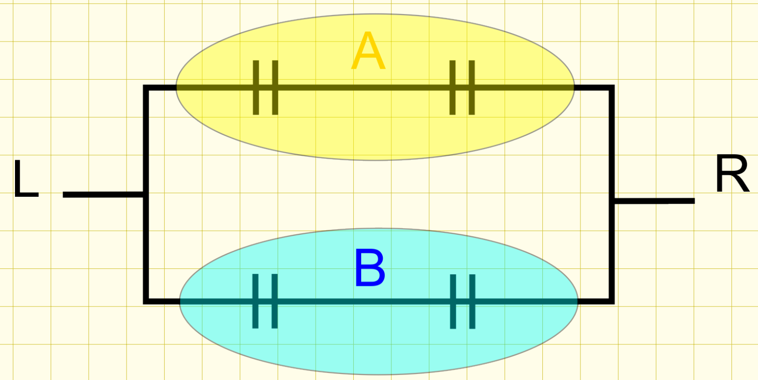
Solução do Exercício 5.18 na página 98:
Considerando D o evento defeito, temos

$$\begin{aligned} P(B|D) &= \frac{P(D|B)P(B)}{P(D|A)P(A)+P(D|B)P(B)+P(D|C)P(C)} \\ &= \frac{0,05\times 0,50}{0,03\times 0,30+0,05\times 0,50+0,02\times 0,20} \\ &= \frac{0,0250}{0,038} = 0,6579. \end{aligned}$$

Solução do Exercício 5.19 na página 99:
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

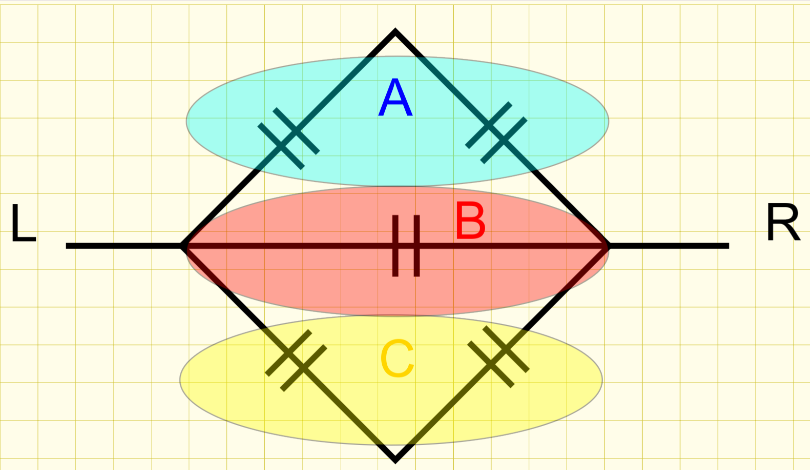
$$\begin{aligned} P(T \cap B^c) &= P(T|B^c) \times P(B^c) \\ &= P(T|B^c) \times [1 - P(B)] \\ &= 1/4(1 - 1/2) \\ &= 1/8. \end{aligned}$$

Solução do Exercício 5.20 na página 99:



$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= p^2 + p^2 - p^4 \\ &= 2p^2 - p^4. \end{aligned}$$

Solução do Exercício 5.21 na página 99:



$$
\begin{aligned}
P(A\cup B\cup C) &= P(A)+P(B)+P(C)+P(A\cap B)+P(A\cap C)+P(B\cap C)-\\
&\quad -P(A\cap B\cap C)\\
&= p+p+p+p^2+p^2+p^2-p^3\\
&= 3p+3p^2-p^3.
\end{aligned}
$$

Solução do Exercício 5.22 na página 99:

Solução do Exercício 5.23 na página 100:
Considere a função de probabilidade:

x	5	7	8	20
$p_X(x)$	0,2	0,3	0,4	0,1

Assim, seguem as respostas...

Solução do Exercício 5.24 na página 100:
Seja

$$P(A) = P(A\cap B) + P(A\cap B^c) \tag{13.3}$$

e seja

$$P(B) = P(A\cap B) + P(B\cap A^c) \tag{13.4}$$

Considere ainda que os eventos $A\cap B^c$ e $A^c\cap B$ são disjuntos. Assim,

$$P[(A\cap B^c)\cup(A^c\cap B)] = P(A\cap B^c) + P(A^c\cap B) \tag{13.5}$$

Substituindo (13.3) e (13.4) em (13.5), 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 123:
Definido e exemplificado em sala de aula.

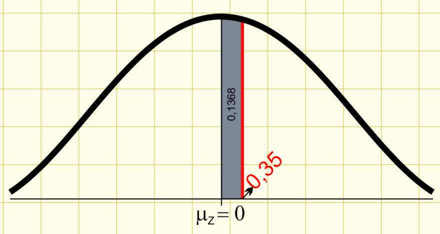
Solução do Exercício 6.2 na página 123:
Distribuição Binomial:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Considerando $p = 0,02$ e $n = 100$, então

$$\begin{aligned} P(X = 3) &= \binom{100}{3} (0,02)^3 (1 - 0,02)^{100-3} \\ &= 0,1823. \end{aligned}$$

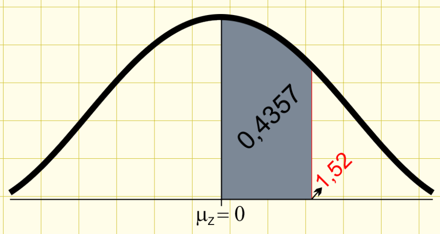
Solução do Exercício 6.3 na página 123:
a) $P(0,00 \leq Z \leq 0,35) = 0,1368$. Graficamente, temos



Tabela

Z	...	0,05	
⋮		⋮	
0,3	...	0,1368	

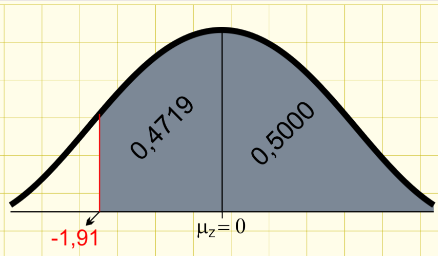
b) $P(0,00 \leq Z \leq 1,52) = 0,4357$. Graficamente, temos



Tabela

Z	...	0,02	
⋮		⋮	
1,5	...	0,4357	

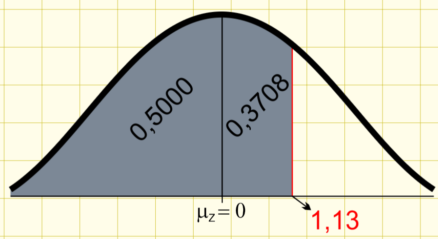
c) $P(Z \geq -1,91) = 0,4719$. Graficamente, temos



Tabela

Z	...	0,01	
⋮		⋮	
1,9	...	0,4719	

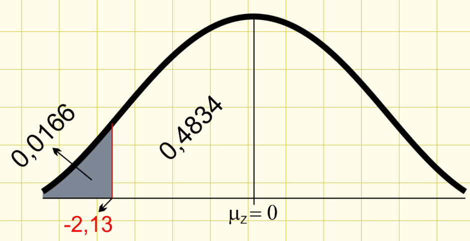
d) $P(Z \leq 1,13) = 0,5000 + 0,3708 = 0,8708$. Graficamente, temos



Tabela

Z	...	0,03	
⋮		⋮	
1,1	...	0,3708	

e) $P(Z \leq -2,13) = 0,5000 - 0,4834 = 0,0166$. Graficamente, temos



Tabela

Z	...	0,03	
⋮		⋮	
2,1	...	0,4834	

Solução do Exercício 6.4 na página 123:

a) Fazendo a transformação $Z = \frac{X-\mu}{\sigma} = \frac{33-40}{6} = -1,17$, temos que $P(X \leq 33) = P(Z \leq -1,17) = 0,5000 - P(-1,17 \leq Z \leq 0) = 0,5000 - 0,3790 = 0,1210$.

b) Fazendo a transformação $Z = \frac{X-\mu}{\sigma} = \frac{29-40}{6} = -1,83$, temos que $P(X \geq 29) = P(Z \geq -1,83) = P(-1,83 \leq Z \leq 0) + 0,5000 = 0,4664 + 0,5000 = 0,9664$.

c) Fazendo a transformação $Z = \frac{X-\mu}{\sigma} = \frac{29-40}{6} = -1,83$, temos que $P(X \geq 29) = P(Z \geq -1,83) = P(-1,83 \leq Z \leq 0) + 0,5000 = 0,4664 + 0,5000 = 0,9664$.

d) $Z = -0,20 \Rightarrow X = \mu + Z\sigma = 40 + (-0,20) \times 6 = 38,8$.

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) $P(Z > 0) = 0,5000.$

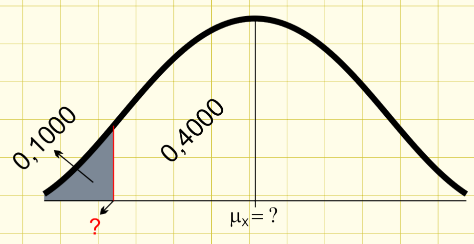
Solução do Exercício 6.5 na página 123:

$$Z = \frac{X-\mu}{\sigma} = \frac{5-6}{1,5} = -0,67.$$

$$\begin{aligned} P(X \geq 5) = P(Z \geq -0,67) &= 0,5000 + P(0 \leq Z \leq -0,67) \\ &= 0,5000 + 0,2486 = 0,7486. \end{aligned}$$

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 123:



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 124:
CONCEITO A: Considerando que $\mu + \sigma = 4,5 + 2,3 = 6,8$, então

$$Z_{\mu+\sigma} = \frac{6,8-4,5}{2,3} = 1.$$

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 124:
$E[X] = \sum_{i=1}^{n} x_i p(x_i) = 60.000 \times 0,3 + 20.000 \times 0,7 = 26.000.$

Solução do Exercício 6.9 na página 124:

Solução do Exercício 6.10 na página 124:
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,

$$\begin{aligned} P(X \geq 0) &= \binom{n}{0} 0,15^0 (1 - 0,15)^n \\ &= 0,85^n. \end{aligned}$$

Assim, para que $P(X \geq 1) = 0,90$, temos que $P(X = 0) = 0,10$, logo,

$$\begin{aligned} 0,85^n &\leq 0,10 \\ n \log(0,85) &> \log(0,10) \\ n &> \frac{\log(0,10)}{\log(0,85)} \\ n &>\approx 14,17 = 14. \end{aligned}$$

Para $n = 14$, $P(X \geq 1) = 0,8972$, e para $n = 15$, $P(X \geq 0,9126)$.

Solução do Exercício 6.11 na página 124:

Solução do Exercício 6.12 na página 125:

Solução do Exercício 6.13 na página 125:
Seja $N = 36^6$, $n_{(I)} = 26^5 \times 60$ e $n_{(II)} = 26^5 \times 10^2$, então:

a) $P(X = 2) = \binom{10}{2} \left(\frac{26^5 \times 60}{36^6}\right)^2 \left(\frac{26^5 \times 60}{36^6}\right)^{10-2}$

Solução dos Exercícios do Capítulo 7

Solução dos Exercícios do Capítulo 8

Solução dos Exercícios do Capítulo 9

Solução dos Exercícios do Capítulo 10

Solução dos Exercícios do Capítulo 11

Solução dos Exercícios do Capítulo 12

Referências Bibliográficas

aES, M. N. M.; LIMA, A. C. P. de. *Noções de Probabilidade e Estatística*. 7. ed. São Paulo: Edusp, 2015. 416 p.

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.

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 Challanger Accident*. Washington, 1986. I, 256 p.

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