1 Introdução à Inferência Estatística

Um típico problema estatístico consiste em realizar um ou uma série de experimentos aleatórios; coletar alguns dados do(s) experimento(s); extrair informações dos dados, interpretar os resultados e realizar algumas conclusões. Uma análise descritiva e exploratória de dados pode ser conduzida para se obter algumas medidas de resumo dos dados, tais como a média, mediana, desvio-padrão, etc., e alguns gráficos, tais como gráfico de barras, histogramas, boxplots, etc.. Embora este tipo de análise tenha a vantagem de ser simples e não requerer suposição alguma a respeito dos dados, ela pode ser restritiva, no sentido de não nos permitir ganhar entendimento suficiente do problema. Desta forma, focamos, neste capítulo, em um método mais sofisticado de análise de dados: a *inferência estatística*.

1.1 Tipos de Problema

O objetivo da inferência estatística é obter conclusões sobre algumas características de um conjunto de interesse, denominado *população*, com base na informação contida em um conjunto de dados disponíveis proveniente da população, denominado *amostra*. A base para o estabelecimento dessas conclusões são certos *modelos probabilísticos*, em relação aos quais as questões de interesse são especificadas. Em geral, vários modelos com diferentes níveis de complexidade podem ser propostos para um mesmo problema e a adoção de um ou de outro depende, não só do grau de conhecimento que temos a respeito da característica que está sendo investigada, como também dos objetivos do estudo.

Exemplo 1

Suponha que um grupo de pesquisadores esteja interessado em estudar a distribuição das alturas dos brasileiros adultos. Se o grupo tivesse recurso para medir as alturas de todos os brasileiros adultos, eles teriam meios de obter sua distribuição exata e, desta forma, produzir os correspondentes parâmetros de interesse como, por exemplo, a altura média dos brasileiros adultos. No entanto, o grupo carece de recursos e, por não ser possível determinar a distribuição exata, eles precisam supor um modelo probabilístico para as alturas dos brasileiros adultos. É razoável supor que tal distribuição é normal, embora as alturas não possam assumir valores negativos. No entanto, essa suposição não é suficiente para produzir os parâmetros de interesse. Neste caso, por exemplo, a altura média dos brasileiros adultos permanece desconhecida. Portanto, é necessário que o grupo colete uma amostra de indivíduos dessa população, realize a medição da altura nos indivíduos da amostra, e utilize os resultados de tais medições para realizar inferência a cerca dos parâmetros desconhecidos do modelo probabilístico fixado.

Com o Exemplo 1 podemos notar que estabelecido um modelo probabilístico para o problema em estudo, gostaríamos de deduzir algumas propriedades deste modelo com base em uma amostra da população. Neste sentido, em inferência estatística, os dados são vistos como observações ou realizações de um elemento aleatório definido sobre um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. A característica que estamos interessados em estudar é, portanto, uma variável aleatória X definida nesse espaço de probabilidade. O modelo probabilístico que estabelecemos para estudar a característica X é uma suposição a cerca da sua distribuição de probabilidade. Nestas notas, vamos supor que a distribuição de probabilidade da variável aleatória X pertence a uma família paramétrica de distribuições, i.e., fixamos a distribuição de X, tal como Bernoulli, exponencial, normal, etc., mas não conhecemos os parâmetros da distribuição. O *parâmetro* da distribuição é um valor (desconhecido) que resume a característica X na população toda e, sobre o qual, gostaríamos de realizar inferência com base em uma amostra. A *amostra* é o elemento aleatório que produz os dados e que consiste na possível observação da característica X em um subconjunto de indivíduos ou objetos da população.

Definição 1: População e Amostra

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $X : \Omega \to \mathbb{R}$ uma variável aleatória que representa uma característica observável associada a uma **população** $\mathcal{P}$ de indivíduos ou objetos. Uma **amostra** da população $\mathcal{P}$ é o vetor $(X_1, \dots, X_n)$, em que $X_i : \Omega \to \mathbb{R}$ é uma variável aleatória que representa a característica observável X medida no indivíduo ou objeto i, $i = 1, 2, \dots, n$.

Observe que o conjunto de dados $(x_1, \dots, x_n)$ que é produzido após analisarmos a característica X nos indivíduos ou objetos amostrados é uma das possíveis observações do vetor $(X_1, \dots, X_n)$.

Definição 2: Inferência estatística

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $X : \Omega \to \mathbb{R}$ uma variável aleatória que representa uma característica observável associada a uma população $\mathcal{P}$ de indivíduos ou objetos. Se X tem função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$), em que $\theta \in \mathbb{R}$ é um parâmetro desconhecido, e $(X_1, \dots, X_n)$ é uma amostra da população $\mathcal{P}$ obtida a partir da característica X, então chamamos de **inferência estatística** o problema que consiste em especificar um ou mais valores para θ baseado em um conjunto de valores observados $(x_1, \dots, x_n)$ do vetor $(X_1, \dots, X_n)$.

Inferência estatística pode ser dividida em duas grandes áreas: *estimação de parâmetros* e *teste de hipóteses*. O objetivo da estimação é procurar, segundo algum critério, valores que representem adequadamente os parâmetros desconhecidos de uma distribuição de probabilidade com base em uma amostra de observações. No caso de problemas de teste de hipóteses, o objetivo é utilizar a amostra de observações da distribuição de probabilidade para verificar a validade de afirmações sobre um valor (ou valores) do(s) parâmetro(s) desconhecido(s).

Exemplo 2

Um fabricante nos forneceu 100.000 pequenos rebites. Uma junta firmemente rebitada exige que o rebite encaixe adequadamente em seu furo, o que não ocorre se o rebite tiver rebarbas. Antes de aceitar essa remessa, desejamos ter alguma ideia sobre a magnitude de p, a proporção de rebites defeituosos, i.e., os rebites com rebarbas. Neste problema,

- a característica de interesse X é a variável que indica se o rebite é ou não defeituoso;
- o modelo probabilístico que assumimos para X é a distribuição Bernoulli, i.e., $X \sim \text{Bernoulli}(p)$, em que $p \in]0, 1[$ é desconhecido;
- o parâmetro que estamos interessados em realizar inferência é a proporção p de rebites com rebarbas.

Definido o plano amostral e o tamanho n da amostra a ser coletada, $n < 100.000$, veremos adiante que uma boa aproximação para proporção p de rebites com rebarbas é a *proporção amostral*

$$\hat{p} := \frac{1}{n} \sum_{i=1}^{n} X_i,$$

em que $(X_1, \ldots, X_n)$ é uma amostra obtida a partir da característica X medida em n objetos da população de 100.000 pequenos rebites. Suponha que, após observamos uma amostra de 5 desses rebites, obtemos o seguinte conjunto de dados

$$(x_1, x_2, x_3, x_4, x_5) = (1, 0, 0, 1, 1).$$

Neste caso, a *estimativa* baseada na proporção amostral é

$$\hat{p} = \frac{1}{5} \sum_{i=1}^{5} x_i = \frac{3}{5}.$$

Como esse tamanho da amostra é muito pequeno em relação ao tamanho da população, a estimativa pode não ser tão precisa. Neste sentido, é fundamental definir um tamanho de amostra que seja suficientemente grande antes de realizar a sua coleta.

É importante observar que neste capítulo não estamos interessados em realizar um estudo extensivo sobre critérios a cerca de como obter um tamanho de amostra suficientemente grande, muito embora trataremos brevemente desse assunto na Seção 1.2.

Exemplo 3

Em geral, uma medição tem imperfeições que dão origem a um erro no resultado da medição. Suponha que os engenheiros responsáveis pela manutenção de uma máquina empacotadora de arroz estejam interessados em verificar se o peso médio μ dos pacotes é realmente de um quilograma como foi estabelecido na última calibragem da máquina. Neste problema,

- a característica de interesse X é o peso do pacote de arroz;
- o modelo probabilístico que assumimos para X é a distribuição Normal, i.e., $X \sim$ Normal$(\mu; 0,001)$, em que $\mu \in \mathbb{R}$ é desconhecido;
- o parâmetro que estamos interessados em realizar inferência é o peso médio μ do pacote de arroz.

Definido o plano amostral e o tamanho n da amostra a ser coletada, veremos adiante como estabelecer critérios para testar a veracidade da hipótese de que a máquina está de fato calibrada, i.e., $\mu = 1\text{kg}$. Tais critérios serão estabelecidos com base em uma amostra $(X_1, \dots, X_n)$ da característica X, i.e., X_i representa o peso do i-ésimo pacote de arroz, $i = 1, 2, \dots, n$.

No Exemplo 3, o desvio-padrão do peso do pacote é conhecido e igual a $0,1\text{kg}$, o que significa que o peso de um pacote de arroz se distancia de μ por no máximo $0,1\text{kg}$.

1.2 Estatísticas e suas distribuições

Definição 3: Estatística

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $(X_1, X_2, \dots, X_n)$ uma amostra da variável aleatória $X : \Omega \to \mathbb{R}$ que representa uma característica observável associada a uma população $\mathcal{P}$ de indivíduos ou objetos. Uma **estatística** é qualquer função $T(X_1, \dots, X_n)$ da amostra que não depende de parâmetros desconhecidos da distribuição de X.

Em outras palavras, uma estatística é um valor ou conjunto de valores observáveis que resume a característica X na amostra e, que utilizaremos para realizar inferência estatística sobre o(s) parâmetro(s) desconhecido(s) da população. É claro que a própria amostra $(X_1, X_2, \dots, X_n)$ é uma estatística. Todavia, queremos utilizar estatísticas que simplifiquem a informação dos dados originais. Neste sentido, estamos interessados em estatísticas cujo conjunto imagem seja bem mais simples que o de $(X_1, X_2, \dots, X_n)$. Eis alguns exemplos.

Exemplo 4

São exemplos de estatísticas.

- Média amostral: $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$;
- Variância amostral: $S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$;
- O menor valor da amostra: $X_{(1)} := \min(X_1, \dots, X_n)$;
- O maior valor da amostra: $X_{(n)} := \max(X_1, \dots, X_n)$;
- Amplitude amostral: $W := X_{(n)} - X_{(1)}$.

Se $(X_1, X_2, \dots, X_n)$ é uma amostra da variável aleatória X com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$), em que $\theta \in \mathbb{R}$ é um parâmetro desconhecido, então a distribuição de probabilidade $p(x_1, \dots, x_n|\theta)$ (ou $f(x_1, \dots, x_n|\theta)$) do vetor $(X_1, X_2, \dots, X_n)$ também depende do parâmetro desconhecido θ. Qualquer estatística $T(X_1, \dots, X_n)$ é uma variável aleatória, logo sua distribuição de probabilidade $p(t|\theta)$ (ou $f(t|\theta)$) também possui parâmetro θ. Como as inferências que desejamos realizar em relação ao parâmetro θ são baseados em uma estatística $T(X_1, \dots, X_n)$ é natural que um dos problemas centrais da inferência estatística seja o de encontrar a forma da distribuição de $T(X_1, \dots, X_n)$.

1.2.1 Distribuições conjuntas

Nesta subseção, estamos interessados em discutir a distribuição da amostra $(X_1, \dots, X_n)$ de uma variável aleatória X com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$), em que $\theta \in \mathbb{R}$ é um parâmetro desconhecido.

Definição 4: Distribuição conjunta de probabilidade

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra da variável aleatória $X : \Omega \to \mathbb{R}$ com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$, em que $\theta \in \mathbb{R}$ é um parâmetro desconhecido. A **distribuição conjunta de probabilidade** do vetor $(X_1, X_2 \ldots, X_n)$ é uma função $p : \mathbb{R}^n \to [0, 1]$ (ou $f : \mathbb{R}^n \to [0, +\infty[$) tal que quando X é uma variável aleatória discreta

$$p(x_1, \ldots, x_n|\theta) = \begin{cases} P\left(X_1 = x_1, \ldots, X_n = x_n\right), & \text{se } (x_1, \ldots, x_n) \in Im(X_1, \ldots, X_n), \\ 0, & \text{caso contrário.} \end{cases}$$

e quando X é uma variável aleatória contínua $f(x_1, \ldots x_n|\theta)$ é tal que

$$P\left(a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n\right) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n|\theta) dx_1 \ldots dx_n,$$

em que $[a_1, b_1]$, …, $[a_n, b_n]$ são intervalos reais.

Exemplo 5

Arthur é professor de inglês, uma de suas turmas é composta por 5 alunas. O conjunto $\{1, 3, 5, 5, 7\}$ representa as notas que elas obtiveram na última prova. Seja X uma variável aleatória que representa a nota da aluna e considere que uma amostra (X_1, X_2) seja retirada dessa população **com reposição**. Qual é a distribuição conjunta do vetor (X_1, X_2)?

Solução. *Considerando todas as possíveis amostras de tamanho $n = 2$, obtidas com reposição dessa população, teremos um total de 25 amostras. Assim, podemos escrever a distribuição conjunta do vetor (X_1, X_2) através da tabela:*

	1	3	5	7	$p(x_1)$
1	1/25	1/25	2/25	1/25	1/5
3	1/25	1/25	2/25	1/25	1/5
5	1/25	1/25	4/25	2/25	2/5
7	1/25	1/25	2/25	1/25	1/5
$p(x_2)$	1/5	1/5	2/5	1/5	1

Na tabela do Exemplo 5 obtivemos a distribuição conjunta do vetor (X_1, X_2), bem como as distribuições unidimensionas. Assim, a primeira e a última linha da tabela fornece a distribuição de X_2, isto é, $(x_2, p(x_2))$, enquanto a primeira a última coluna estabelece a distribuição de X_1, ou seja, $(x_1, p(x_1))$. Essas distribuições são chamadas *distribuições marginais*.

Definição 5: Distribuições marginais

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra da variável aleatória $X : \Omega \to \mathbb{R}$ com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$, em que $\theta \in \mathbb{R}$ é um parâmetro desconhecido. A **distribuição marginal de probabilidade** da variável aleatória X_i, $i = 1, 2, \ldots, n$, é uma função $p : \mathbb{R} \to [0, 1]$ (ou $f : \mathbb{R} \to [0, +\infty[$) tal que quando X é uma variável aleatória discreta

$$p(x_i|\theta) := \sum_{x_1 \in \mathrm{Im}(X_1)} \cdots \sum_{x_n \in \mathrm{Im}(X_n)} p(x_1 \ldots, x_i, \ldots, x_n|\theta)$$

e quando X é uma variável aleatória contínua $f(x_i|\theta)$ é tal que

$$P\left(a \leq X_i \leq b\right) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_i, \ldots x_n|\theta) dx_1 \ldots dx_n,$$

em que $[a, b]$ é um intervalo real e $p(x_1, \ldots, x_n|\theta)$ ou $f(x_1, \ldots, x_n|\theta)$ é a função de distribuição conjunta de probabilidade do vetor $(X_1, \ldots, X_n)$.

Encontrar a distribuição conjunta de um vetor $(X_1, X_2, \ldots, X_n)$ pode ser uma tarefa bem árdua quando nenhuma suposição é feita em relação a sua natureza. Para iniciar os estudos de inferência estatística, a primeira suposição que faremos é que $(X_1, X_2, \ldots, X_n)$ é uma sequência de variáveis aleatórias *independentes* com a mesma distribuição $p(x|\theta)$ (ou $f(x|\theta)$) da variável aleatória X, i.e.,

$$p(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta),$$

ou

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta),$$

em que $\theta \in \mathbb{R}$ é um parâmetro desconhecido e $p(x_1, \ldots, x_n|\theta)$ ou $f(x_1, \ldots, x_n|\theta)$ é a função de distribuição conjunta de probabilidade do vetor $(X_1, \ldots, X_n)$.

Definição 6: Amostra aleatória

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra da variável aleatória $X : \Omega \to \mathbb{R}$ com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$, em que $\theta \in \mathbb{R}$ é um parâmetro desconhecido. Dizemos que $(X_1, X_2, \ldots, X_n)$ é uma **amostra aleatória** da variável aleatória X quando $(X_1, X_2, \ldots, X_n)$ é uma sequência de variáveis aleatórias independentes e com a mesma distribuição $p(x|\theta)$ (ou $f(x|\theta)$) da variável aleatórias X.

Exemplo 6

Uma caixa contém lâmpadas cujo tempo de vida X segue uma distribuição exponencial com média $E(X) = 1/\theta$, em que $\theta > 0$ é um parâmetro desconhecido. Suponha que uma amostra aleatória (X_1, X_2) é retirada da caixa. Qual é a distribuição conjunta do vetor (X_1, X_2)?

Solução. *Como* (X_1, X_2) *é uma amostra aleatória de* X*, segue que*

$$\begin{aligned} f(x_1, x_2|\theta) &= f(x_1|\theta)f(x_2|\theta) \\ &= \theta \exp(-\theta x_1)\theta \exp(-\theta x_2) \\ &= \theta^2 \exp(-\theta(x_1 + x_2)), \end{aligned}$$

para $(x_1, x_2) \in \mathrm{Im}(X_1, X_2)$ *e* $f(x_1, x_2|\theta) = 0$*, caso contrário.*

Exemplo 7

No Exemplo 5, podemos afirmar que (X_1, X_2) é uma amostra aleatória de X?

Solução. Sim, pois

$$p(x_1, x_2|\theta) = p(x_1|\theta)p(x_2|\theta),$$

para todo $(x_1, x_2) \in \mathbb{R}^2$.

1.2.2 Distribuições amostrais

Nosso objetivo é realizar inferências sobre o(s) parâmetro(s) θ desconhecido(s) de uma população $\mathcal{P}$ através de uma amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma variável aleatória X. Nossas decisões serão baseadas em uma estatística $T(X_1, \ldots, X_n)$. Em particular, tais decisões são baseadas na distribuição de $T(X_1, \ldots, X_n)$, que é denominada *distribuição amostral*.

Exemplo 8

No Exemplo 5, qual é a distribuição amostral das estatísticas $\bar{X}$, S^2 e W?

Solução. *Podemos escrever a distribuição amostral de* $\bar{X}$, S^2 *e* W *através das tabelas:*

- *Distribuição da média amostral* $\bar{X}$*:*

$\bar{x}$	1	2	3	4	5	6	7	Σ
$p(\bar{x})$	1/25	2/25	5/25	6/25	6/25	4/25	1/25	1

- *Distribuição da variância amostral* S^2*:*

s^2	0	2	8	18	Σ
$p(s^2)$	7/25	10/25	6/25	2/25	1

- *Distribuição da amplitude amostral* W*:*

w	0	2	4	6	Σ
$p(w)$	7/25	10/25	6/25	2/25	1

A distribuição amostral de uma estatística depende da distribuição da população, do tamanho da amostra e do método de seleção da amostra. Apresentamos agora um rápido estudo a respeito de talvez a mais importante distribuição amostral: a distribuição da média amostral. As distribuições amostrais de outras estatísticas serão apresentadas ao longo do texto conforme necessitarmos delas.

Lema 1

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Se $(X_1, X_2, \ldots, X_n)$ é uma sequência de variáveis aleatórias independentes definidas sobre $(\Omega, \mathcal{F}, P)$, então as seguintes afirmações são verdadeiras

(i)

$$E\left(a\sum_{i=1}^{n} X_i + b\right) = a\sum_{i=1}^{n} E(X_i) + b \quad \text{e} \quad E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i),$$

(ii)

$$\mathrm{Var}\left(a\sum_{i=1}^{n} X_i + b\right) = a^2\sum_{i=1}^{n} \mathrm{Var}(X_i),$$

quaisquer que sejam os números reais a e b.

Lema 2

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Se $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória da variável aleatória $X : \Omega \to \mathbb{R}$ tal que $E(X) = \mu$ e $Var(X) = \sigma^2$, então

$$E(\bar{X}) = \mu \quad \text{e} \quad Var(\bar{X}) = \frac{\sigma^2}{n},$$

em que $\bar{X}$ é a média amostra.

Demonstração. *Utilizando o Lema 1, segue que*

(i)

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

(ii)

$$\begin{aligned} Var(\bar{X}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 \\ &= \frac{1}{n^2} n\sigma^2 \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

□

Com o Lema 2 conseguimos determinar a esperança e a variância da média amostral $\bar{X}$ a partir da esperança e variância de X. No entanto, ainda não encontramos a distribuição amostral da estatística $\bar{X}$. Em geral, é difícil encontrar a exata forma da distribuição de $\bar{X}$. Podemos, no entanto, utilizar o *Teorema do Limite Central* para aproximar a distribuição de $\bar{X}$. Esse teorema afirma que para uma amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma variável aleatória X associada a uma população com média μ e variância σ^2, a distribuição de $\bar{X}$ é aproximadamente Normal com média μ e variância $\frac{\sigma^2}{n}$, se o tamanho da amostra n for suficientemente grande.

Teorema 1: Teorema do Limite Central (T.L.C.)

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Se $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória da variável aleatória $X : \Omega \to \mathbb{R}$ tal que $E(X) = \mu$ e $Var(X) = \sigma^2$, então

$$\bar{X} \simeq \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right),$$

quando $n \to +\infty$.

A demonstração completa desse teorema exige recursos dos quais não dispomos, portanto não será fornecida. No entanto, o leitor ou leitora interessado(a) pode consultar Meyer (1965) para mais detalhes.

Corolário 1

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Se $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória da variável aleatória $X : \Omega \to \mathbb{R}$ tal que $E(X) = \mu$ e $Var(X) = \sigma^2$, então

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \simeq \text{Normal}(0, 1),$$

quando $n \to +\infty$.

A aproximação normal para a distribuição de $\bar{X}$ depende do tamanho da amostra n. A Figura 1 mostra a mudança que ocorre na distribuição de probabilidade das pontuações médias obtidas para o arremesso de dados honesto, conforme aumenta-se o número de dados a serem lançados. Note que, embora a distribuição da população (pontuação obtida com um dado honesto) esteja relativamente longe da normal, a distribuição das médias é aproximada razoavelmente bem pela distribuição normal conforme o número de dados lançados aumenta.

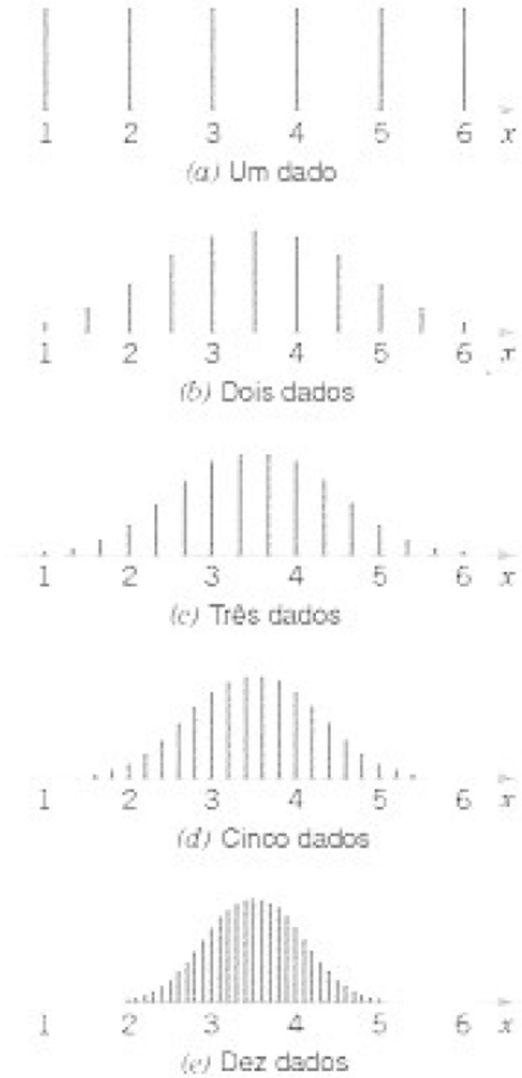

Figura 1: Distribuições das pontuações médias obtidas a partir do arremesso de dados honesto (Montgomery e Runger, 2009).

Em geral, para $n \geq 30$ a aproximação normal será satisfatória. Para $n < 30$, o T.L.C. funcionará se a distribuição da população não for muito diferente da normal. Quando $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória de uma variável aleatória $X \sim \mathsf{Normal}(\mu, \sigma^2)$, então $\bar{X} \sim \mathsf{Normal}(\mu, \sigma^2/n)$ para qualquer tamanho de amostra n. Em outras palavras, se a variável X atrelada à população possui distribuição normal, então $\bar{X}$ tem distribuição *exata* normal. Esta é uma consequência imediata do seguinte fato

$$a \sum_{i=1}^{n} X_i + b \sim \mathsf{Normal}\left(an\mu + b, a^2 n\sigma^2\right),$$

quaisquer que sejam os número reais a e b.

Exemplo 9

A capacidade máxima de um elevador é de 500kg. Se a distribuição dos pesos dos usuários do elevador é normal sendo o peso médio de $\mu = 70\text{kg}$ e o desvio-padrão dos pesos de $\sigma = 10\text{kg}$, qual é a probabilidade de uma amostra aleatória de 7 passageiros ultrapassar essa capacidade limite?

Solução. *Seja* $(X_1, X_2, \ldots, X_7)$ *uma amostra aleatória da variável aleatória* X *que representa o peso de um usuário do elevador. Como* $X \sim \text{Normal}(70, 100)$*, segue que a soma dos pesos dos usuários é tal que*

$$S := \sum_{i=1}^{7} X_i \sim \text{Normal}\,(490, 700)\,.$$

Logo,

$$\begin{aligned} P(S > 500) &= P\left(\frac{S - 490}{\sqrt{700}} > \frac{500 - 490}{\sqrt{700}}\right) \\ &= P(Z > 0,38) \\ &= 1 - P(Z \leq 0,38) \\ &= 1 - 0,64803 \\ &= 0,35197, \end{aligned}$$

em que $Z \sim \text{Normal}(0, 1)$.

Exemplo 10

No Exemplo 6, qual é a distribuição amostral aproximada da estatística $\bar{X}$?

Solução. *Como* $X \sim \text{Exponencial}(\theta)$, $\theta > 0$, *segue que*

$$E(X) = \frac{1}{\theta} \quad e \quad \text{Var}(X) = \frac{1}{\theta^2}.$$

Pelo Lema 2, segue que

$$E(\bar{X}) = \frac{1}{\theta} \quad e \quad \text{Var}(\bar{X}) = \frac{1}{2\theta^2}.$$

Portanto, pelo T.L.C.,

$$\bar{X} \simeq \text{Normal}\left(\frac{1}{\theta}, \frac{1}{2\theta^2}\right).$$

Vamos considerar uma população em que a proporção de indivíduos ou objetos portadores de uma determinada característica é p, $p \in]0,1[$. Assim, podemos definir a variável aleatória X tal que

$$X = \begin{cases} 1, & \text{se o elemento é portador da característica,} \\ 0, & \text{caso contrário.} \end{cases}$$

Logo,

$$E(X) = p \quad e \quad Var(X) = p(1-p).$$

Retirada uma amostra aleatória de tamanho n dessa população e indicando por S o total de indivíduos na amostra que são portadores da característica, temos

$$S := \sum_{i=1}^{n} X_i \sim Binomial(n, p).$$

Definindo como $\hat{p}$ a proporção de elementos portadores da característica na amostra, i.e.,

$$\hat{p} := \frac{S}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

segue que

$$Im(\hat{p}) = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}$$

e, a sua distribuição de probabilidade é

$$P\left(\hat{p} = \frac{k}{n}\right) := P\left(\frac{S}{n} = \frac{k}{n}\right) = P(S = k),$$

para todo $k \in \{0, 1, \ldots, n\}$. Em outras palavras, a distribuição da proporção amostral $\hat{p}$ é obtida a partir da distribuição de S.

Exemplo 11

Sabe-se que 20% das peças de um lote são defeituosas. Retira-se desse lote uma amostra aleatória de 8 peças e calcula-se a proporção amostral $\hat{p}$ de peças defeituosas na amostra. Qual a distribuição *exata* de $\hat{p}$?

Solução. Como $p = 0,2$, segue que a distribuição *exata* de $\hat{p}$ é

$$P\left(\hat{p} = \frac{k}{8}\right) = P(S = k) = \binom{8}{k}(0,2)^k(0,8)^{8-k}, \quad \forall k \in \{0, 1, \ldots, 8\}.$$

Neste caso,

$$Im(\hat{p}) = \left\{0, \frac{1}{8}, \frac{1}{4}, \ldots, \frac{7}{8}, 1\right\}.$$

Utilizando o T.L.C. podemos aproximar a distribuição da proporção amostral $\hat{p}$ pela distribuição Normal. Neste caso,

$$E(\hat{p}) = p \quad e \quad Var(\hat{p}) = \frac{p(1-p)}{n}.$$

Logo, para n suficientemente grande, a distribuição *aproximada* de $\hat{p}$ é

$$\hat{p} \simeq \text{Normal}\left(p, \frac{p(1-p)}{n}\right).$$

Exemplo 12

Um procedimento de controle de qualidade foi planejado para garantir uma máximo de 10% de itens defeituosos na produção. A cada 15 minutos sorteia-se uma amostra aleatória de 20 peças e, havendo mais de 15% de defeituosas, para-se a produção para verificação. Qual é a probabilidade de uma parada desnecessária?

Solução. *Seja* $(X_1, X_2, \ldots, X_{20})$ *uma amostra aleatória da variável aleatória* X *que indica se uma peça dessa produção é ou não defeituosa. Como* $X \sim \text{Bernoulli}(0,1)$, *segue que a proporção amostral de itens defeituosos é tal que*

$$\hat{p} := \frac{1}{20}\sum_{i=1}^{20} X_i \simeq \text{Normal}\left(0,10; \frac{0,10 \times 0,90}{20}\right).$$

Logo,

$$\begin{aligned} P(\hat{p} \leq 0,15) &= P\left(\frac{\hat{p} - 0,10}{\sqrt{\frac{0,10 \times 0,90}{20}}} \leq \frac{0,15 - 0,10}{\frac{0,10 \times 0,90}{20}}\right) \\ &= P(Z \leq 0,74) \\ &= 0,77035, \end{aligned}$$

em que $Z \sim \text{Normal}(0,1)$.

1.2.3 Exercícios

1. Uma caixa contém dados honestos. Dois desses dados são selecionados e lançados simultaneamente. Verifica-se a soma das faces voltadas para cima. Pergunta-se:

   (a) Qual é a população em estudo?

   (b) Qual é a característica de interesse?

   (c) Qual é a distribuição da média amostral? Represente-a em um gráfico de linhas.

2. A distribuição do número de filhos, por família, de uma zona rural está na tabela abaixo abaixo.

Número de filhos	Porcentagem
0	10
1	20
2	30
3	25
4	15
Total	100

   (a) Suponha que duas famílias sejam sorteadas com reposição dessa população. Seja X_1 o número de filhos da família que foi selecionada primeiro e X_2 o número de filhos da família selecionada em seguida. Determine a distribuição conjunta da amostra (X_1, X_2) e as respectivas distribuições marginais.

   (b) Determine $E(X_i)$ e $Var(X_i)$ para $i = 1, 2$.

   (c) Construa as distribuições amostrais das estatísticas $\bar{X}$ e S^2.

   (d) Determine a esperança e a variância de $\bar{X}$ e S^2.

3. O peso X dos adultos brasileiros segue uma distribuição Normal com média μkg e desvio-padrão $1kg$. Suponha que uma amostra aleatória (X_1, X_2) de X seja retirada dessa população. Determine:

   (a) a distribuição conjunta de (X_1, X_2) e as respectivas distribuições marginais;

   (b) $E(X_i)$ e $Var(X_i)$ para $i = 1, 2$.

4. Uma fibra sintética usada na fabricação de carpete tem uma resistência à tração que é normalmente distribuída, com média de $75,5psi$ desvio-padrão de $3,5psi$.

   (a) Qual é a probabilidade de uma amostra aleatória de 6 corpos-de-prova de fibra ter resistência média amostral à tração que exceda $75,75psi$.

   (b) Como varia o desvio-padrão da média amostral quando o tamanho da amostra aumenta de $n = 6$ para $n = 49$?

5. A resistência do concreto à compressão é normalmente distribuída com média $2.500psi$ e desvio-padrão de $50psi$.

   (a) Qual é a probabilidade de uma amostra aleatória de 5 corpos-de-prova ter resistência média amostral que caia no intervalo $[2.499psi, 2.510psi]$?

   (b) Quão grande deve ser o tamanho n da amostra aleatória para que o desvio-padrão da média amostral seja igual a $5psi$?

6. Suponha que determinada característica de uma população seja modelada através de uma variável aleatória X que possui distribuição $\mathtt{Uniforme}(0,1)$. Uma amostra aleatória de tamanho 12 é retirada dessa população.

   (a) Qual é a distribuição de probabilidade aproximada de $\bar{X}$?

   (b) Determine $E(\bar{X})$ e $Var(\bar{X})$.

   (c) Calcule $P(0,4 \leq \bar{X} \leq 0,6)$.

7. A quantidade de tempo que um consumidor gasta esperando no balcão de *check-in* de um aeroporto é uma variável aleatória com média de $8,2$ minutos e desvio-padrão de $1,5$ minuto. Suponha que uma amostra aleatória de 49 consumidores seja observada. Encontre a probabilidade do tempo médio amostral de espera na fila por esses consumidores

   (a) ser menor do que 10 minutos;

   (b) estar entre 5 e 10 minutos;

   (c) ser maior do que 6 minutos.

8. Uma máquina de empacotar determinado produto o faz segundo uma distribuição normal, com média μ e desvio-padrão de $10g$.

   (a) Em quanto deve ser regulado o peso médio μ para que apenas 10% dos pacotes tenham menos do que $500g$?

   (b) Com a máquina assim regulada, qual é a probabilidade que o peso total de 4 pacotes escolhido ao acaso seja inferior a $2kg$?

9. Um distribuidor de sementes determina, por meio de testes, que 5% das sementes não germinam. Ele vende pacotes com 200 sementes com garantia de 90% de germinação. Qual é a probabilidade de que um pacote não satisfaça a garantia?

10. Um professor dá um teste rápido, constando de 20 questões do tipo certo-errado. Para testar a hipótese de o estudante estar adivinhando a resposta, ele adota a seguinte regra de decisão: “se 13 ou mais questões tiverem corretas, ele não está adivinhando”. Qual é a probabilidade de rejeitarmos a hipótese, sendo que na realidade ela é verdadeira, i.e., a proporção de acertos é $1/2$?

11. Um indivíduo realiza 20 jogadas independentes, cuja probabilidade de vitória é $p = 0,1$, em cada uma. Seja X uma variável aleatória que representa o número de vitórias.

   (a) Encontre $P(5 \leq X \leq 8)$ de maneira exata.

   (b) Encontre $P(5 \leq X \leq 8)$ de maneira aproximada.

1.3 Estimadores pontuais

O problema de estimar um parâmetro desconhecido relacionado a uma população desconhecida é introduzido no Exemplo 1. Como vimos, o parâmetro θ de uma população é um valor real que resume uma certa característica X em toda população. Como não temos acesso a toda população, consequentemente, desconhecemos o real valor de θ. Neste sentido, desejamos utilizar uma amostra $(X_1, \ldots, X_n)$ de X para inferir a respeito deste valor. Tais inferências são baseadas em uma estatística $T(X_1, \ldots, X_n)$.

Definição 7: Espaço paramétrico

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $X : \Omega \to \mathbb{R}$ uma variável aleatória que representa uma característica observável associada a uma população $\mathcal{P}$ de indivíduos ou objetos. Suponha que X tenha função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$), em que θ é um parâmetro desconhecido. O conjunto Θ, $\Theta \subset \mathbb{R}$, em que θ toma valores é denominado **espaço paramétrico**.

Exemplo 13

Considere um estudo que esteja interessado em inferir a respeito da proporção de indivíduos de uma população que já realizaram uma viagem internacional. Neste caso,

- a característica de interesse X é a variável que indica se o indivíduo realizou ou não uma viagem internacional;
- o modelo probabilístico que assumimos para X é a distribuição Bernoulli, i.e., $X \sim \mathtt{Bernoulli}(p)$, em que $p \in]0, 1[$ é desconhecido;
- o parâmetro que estamos interessados em realizar inferência é a proporção p de indivíduos que já realizaram uma viagem internacional.

Então, espaço paramétrico é $\Theta =]0, 1[$.

Exemplo 14

Considere um estudo que esteja interessado em inferir a respeito da resistência média do concreto à compressão. Neste caso,

- a característica de interesse X é a variável que representa a resistência à compressão do concreto;
- o modelo probabilístico que assumimos para X é a distribuição Normal, i.e., $X \sim \mathtt{Normal}(\mu, 50)$, em que $\mu \in \mathbb{R}$ é desconhecido;
- o parâmetro que estamos interessados em realizar inferência é a resistência média μ do concreto à compressão.

Então, o espaço paramétrico é $\Theta = \mathbb{R}$.

Definição 8: Estimador pontual e estimativa pontual

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória $X : \Omega \to \mathbb{R}$ com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$), em que $\theta \in \Theta$ é um parâmetro desconhecido. Qualquer estatística $\hat{\theta} := T(X_1, \ldots, X_n)$ que assume valores em Θ é um **estimador** para θ. Quando $(x_1, \ldots, x_n)$ é uma observação da amostra $(X_1, \ldots, X_n)$, dizemos que $\hat{\theta} := T(x_1, \ldots, x_n)$ é uma **estimativa** para θ.

Note que usamos tanto a notação $\hat{\theta}$ para representar um estimador para θ, i.e., uma variável aleatória, quanto para representar uma estimativa para θ, ou seja, um valor numérico. A distinção de um ou de outro se fará através do contexto. Todavia, sempre que possível lembraremos o leitor e/ou a leitora a respeito da natureza do objeto $\hat{\theta}$.

Exemplo 15

No Exemplo 13, definido um plano amostral e o tamanho n da amostra a ser coletada, um estimador comumente utilizado para o parâmetro p é a proporção amostral, i.e.,

$$\hat{p} := \frac{1}{n} \sum_{i=1}^{n} X_i,$$

em que $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória obtida a partir da característica X medida em n indivíduos da população.

Exemplo 16

No Exemplo 14, definido um plano amostral e o tamanho n da amostra a ser coletada, um estimador comumente utilizado para o parâmetro μ é a média amostral, i.e.,

$$\hat{\mu} := \frac{1}{n}\sum_{i=1}^{n} X_i,$$

em que $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória obtida a partir da característica X medida em n indivíduos da população.

Um dos grandes problemas da estatística é o de encontrar um estimador razoável para o parâmetro desconhecido θ. Intuitivamente, gostaríamos que um estimador $\hat{\theta}$ estivesse, em um certo sentido, próximo do valor do parâmetro θ. Todavia, θ é desconhecido, logo essa noção de proximidade deve se dar a partir da distribuição amostral de $\hat{\theta}$.

Definição 9: Viés de um estimador

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória $X : \Omega \to \mathbb{R}$ com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$), em que $\theta \in \Theta$ é um parâmetro desconhecido. O **viés** de um estimador $\hat{\theta}$ para o parâmetro θ é uma função $B_{\hat{\theta}} : \Theta \to \mathbb{R}$ tal que

$$B_{\hat{\theta}}(\theta) := E(\hat{\theta}) - \theta.$$

Quando

- $B_{\hat{\theta}}(\theta) = 0$, dizemos que $\hat{\theta}$ é um **estimador não-viesado** para θ;
- $B_{\hat{\theta}}(\theta) > 0$, $\forall \theta \in \Theta$, dizemos que $\hat{\theta}$ é um **estimador viesado positivamente** para θ;
- $B_{\hat{\theta}}(\theta) < 0$, $\forall \theta \in \Theta$, dizemos que $\hat{\theta}$ é um **estimador viesado negativamente** para θ.

Exemplo 17

Seja $(X_1, \ldots, X_n)$ uma amostra aleatória da variável aleatória X que representa alguma característica observável de uma população $\mathcal{P}$ de forma que $E(X) = \mu$ e $Var(X) = \sigma^2$. Os parâmetros $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ são desconhecidos. Suponha que os seguintes estimadores são propostos, respectivamente, para μ e σ^2:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{e} \quad \hat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Estude o viés desses estimadores.

Solução. *Já vimos, no Lema 2, que* $E(\bar{X}) = \mu$. *Logo,*

$$B_{\bar{X}}(\mu) = E(\bar{X}) - \mu = \mu - \mu = 0.$$

Portanto, $\bar{X}$ *é um* ***estimador não-viesado*** *para* μ. *Por outro lado,*

$$\begin{aligned}
E(\hat{\sigma}^2) &= E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)\\
&= \frac{1}{n}\sum_{i=1}^{n} E\left[(X_i - \bar{X})^2\right]\\
&= \frac{1}{n}\sum_{i=1}^{n} E\left[\left(X_i - \mu + \mu - \bar{X}\right)^2\right]\\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{E\left[(X_i - \mu)^2\right] + 2E\left[(X_i - \mu)\left(\mu - \bar{X}\right)\right] + E\left[\left(\bar{X} - \mu\right)^2\right]\right\}.
\end{aligned}$$

Por definição, $E\left[(X_i - \mu)^2\right] = Var(X_i) = \sigma^2$ *e, pelo Lema 2, sabemos que* $Var(\bar{X}) := E\left[(\bar{X} - \mu)^2\right] = \dfrac{\sigma^2}{n}$. *Como* $E\left[(X_i - \mu)\left(\mu - \bar{X}\right)\right] = 0$ *(verifique!), segue que*

$$E(\hat{\sigma}^2) = \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 + \frac{\sigma^2}{n}\right) = \frac{n-1}{n}\sigma^2.$$

Logo,

$$B_{\hat{\sigma}^2}\left(\sigma^2\right) = E(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2.$$

Portanto, $B_{\hat{\sigma}^2}(\sigma^2) < 0$, $\forall \sigma^2 > 0$. *Em outras palavras,* $\hat{\sigma}^2$ *é um* ***estimador viesado negativamente*** *para* σ^2.

Observe que embora o estimador $\hat{\sigma}^2$ seja viesado para σ^2, quando o tamanho n da amostra cresce, o viés de $\hat{\sigma}^2$ tende a zero. Neste caso, dizemos que o estimador $\hat{\sigma}^2$ é **assintoticamente não-viesado** para σ^2.

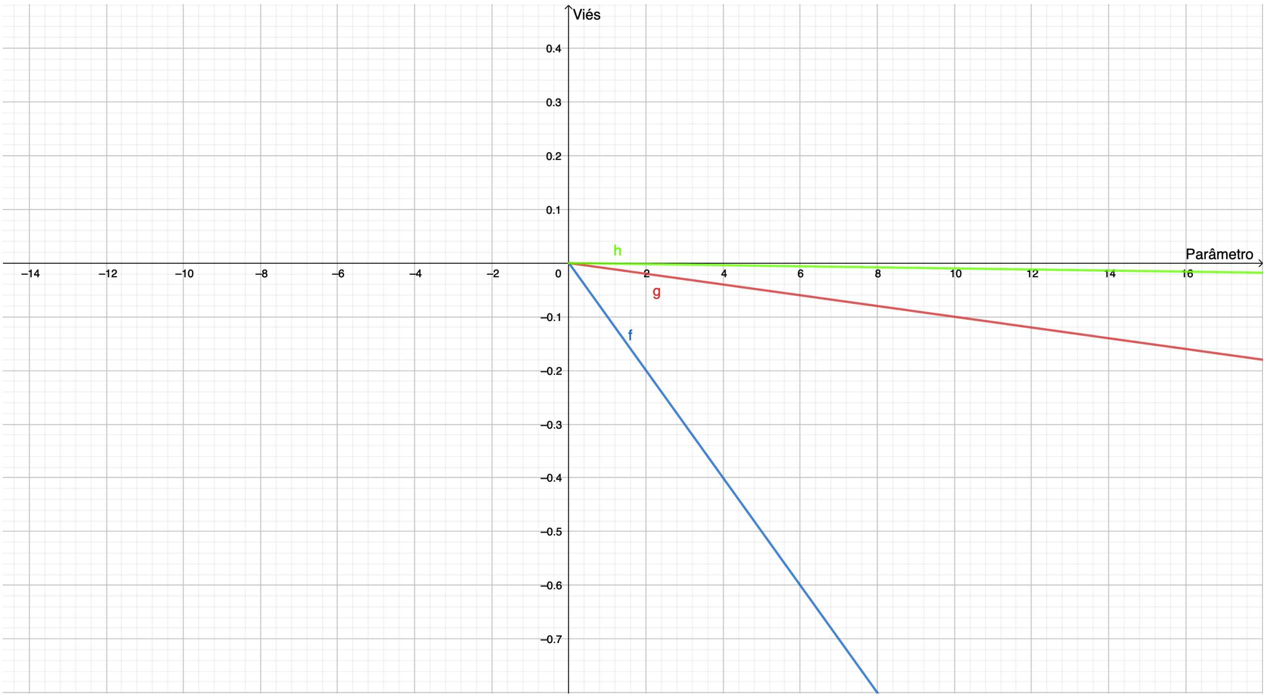

Figura 2: Viés $B_{\hat{\sigma}^2}$ do estimador $\hat{\sigma}^2$ do Exemplo 17 em função dos possíveis valores do parâmetro σ^2 considerando três tamanhos de amostra: $n = 10$ (reta azul), $n = 100$ (reta vermelha) e $n = 1.000$ (reta verde).

Exemplo 18

Seja (X_1, X_2, X_3) uma amostra aleatória da variável aleatória X que representa alguma característica observável de uma população $\mathcal{P}$ de forma que $E(X) = \theta$ e $Var(X) = 1$. Suponha que os seguintes estimadores são propostos para θ:

$$\hat{\theta}_1 := \bar{X} = \frac{X_1 + X_2 + X_3}{3} \quad \text{e} \quad \hat{\theta}_2 := \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3.$$

Estude o viés desses estimadores.

Solução. *Já vimos, no Lema 2, que* $E(\hat{\theta}_1) = E(\bar{X}) = \theta$. *Logo,* $B_{\hat{\theta}_1} = 0$, *i.e.,* $\hat{\theta}_1$ *é um **estimador não-viesado** para* θ. *Por outro lado,*

$$E(\hat{\theta}_2) = E\left(\frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3\right) = \frac{1}{2}E(X_1) + \frac{1}{4}E(X_2) + \frac{1}{4}E(X_3) = \frac{\theta}{2} + \frac{\theta}{4} + \frac{\theta}{4} = \theta.$$

Portanto, $B_{\hat{\theta}_2} = 0$, *ou seja,* $\hat{\theta}_2$ *é um **estimador não-viesado** para* θ.

No Exemplo 18, tanto $\hat{\theta}_1$ quanto $\hat{\theta}_2$ são estimadores não-viesados para θ. Isso indica que a distribuição amostral de cada estimador está centralizada no verdadeiro valor de θ. Entretanto, as variâncias dessas distribuições podem ser diferentes, que é o que ocorre com os estimadores do Exemplo 18.

Exemplo 19

Seja (X_1, X_2, X_3) uma amostra aleatória da variável aleatória X que representa alguma característica observável de uma população $\mathcal{P}$ de forma que $E(X) = \theta$ e $Var(X) = 1$. Suponha que os seguintes estimadores são propostos para θ:

$$\hat{\theta}_1 := \bar{X} = \frac{X_1 + X_2 + X_3}{3} \quad \text{e} \quad \hat{\theta}_2 := \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3.$$

Determine $Var(\hat{\theta}_1)$ e $Var(\hat{\theta}_2)$.

Solução. *Pelo Lema 2,* $Var(\hat{\theta}_1) = Var(\bar{X}) = \frac{1}{3}$. *Por outro lado,*

$$Var(\hat{\theta}_2) = Var\left(\frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3\right) = \frac{1}{4}Var(X_1) + \frac{1}{16}Var(X_2) + \frac{1}{16}Var(X_3) = \frac{6}{16} = \frac{3}{4}.$$

Portanto, $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$.

Um vez que $\hat{\theta}_1$ tem uma variância menor do que $\hat{\theta}_2$ é mais provável que o estimador $\hat{\theta}_1$ produza uma estimativa mais próxima do verdadeiro valor do parâmetro θ.

Uma decisão razoável a respeito de qual estimador utilizar dentre todos os estimadores não-viesados para um parâmetro θ desconhecido é escolher aquele que tiver a menor variância.

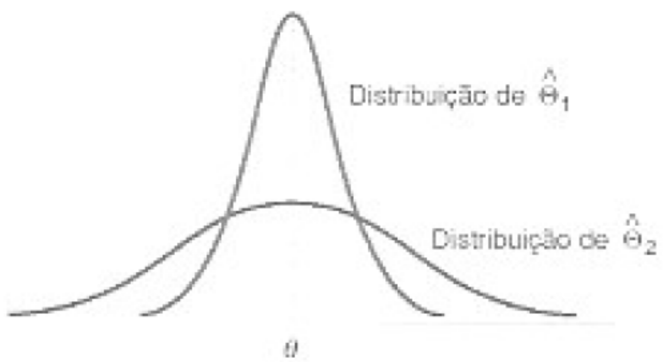

Figura 3: Distribuições amostrais de dois estimadores não-viesados para um parâmetro desconhecido qualquer (Montgomery e Runger, 2009).

Definição 10: Estimadores não-viesados de variância mínima (ENVVM)

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória $X : \Omega \to \mathbb{R}$ com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$), em que $\theta \in \Theta$ é um parâmetro desconhecido. Defina $\mathcal{U} := \{\hat{\theta} \in \Theta : B_{\hat{\theta}}(\theta) = 0\}$ como a classe de todos os estimadores não-viesados para o parâmetro θ. O estimador $\hat{\theta}^* \in \mathcal{U}$ é dito um **estimador não-viesado de variância mínima (ENVVM)** para θ quando

$$\mathsf{Var}(\hat{\theta}^*) \leq \mathsf{Var}(\hat{\theta}),$$

para todo $\theta \in \Theta$ e para todo $\hat{\theta} \in \mathcal{U}$ tal que $\hat{\theta} \neq \hat{\theta}^*$, com $\leq$ substituído por $<$ para pelo menos um valor de θ.

De certo modo, o ENVVM é o mais provável, entre todos os estimadores não-viesados para produzir uma estimativa que seja próxima do verdadeiro valor do parâmetro. Existe uma metodologia para identificar, caso exista, o ENVVM em muitas situações práticas. Todavia, essa metodologia está além do escopo destas notas.

Algumas vezes é necessário utilizar um estimador viesado. Em tais situações um dos procedimentos comumente utilizados para se avaliar o desempenho do estimador é analisar o seu *erro quadrático médio* (EQM).

Definição 11: Erro quadrático médio (EQM)

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória $X : \Omega \to \mathbb{R}$ com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$), em que $\theta \in \Theta$ é um parâmetro desconhecido. O **erro quadrático médio** de um estimador $\hat{\theta}$ para o parâmetro θ é uma função $\mathsf{EQM}_{\hat{\theta}} : \Theta \to \mathbb{R}_+$ tal que

$$\mathsf{EQM}_{\hat{\theta}}(\theta) := \mathsf{E}\left[(\hat{\theta} - \theta)^2\right].$$

É fácil ver que para qualquer estimador $\hat{\theta}$ para θ podemos escrever

$$\mathsf{EQM}_{\hat{\theta}}(\theta) = \mathsf{Var}(\hat{\theta}) + B^2_{\hat{\theta}}(\theta).$$

Assim, quando $\hat{\theta}$ é um estimador não-viesado para θ segue que

$$\mathsf{EQM}_{\hat{\theta}}(\theta) = \mathsf{Var}(\hat{\theta}).$$

Exemplo 20

Seja $(X_1, \dots, X_n)$ uma amostra aleatória da variável aleatória X que representa alguma característica observável de uma população $\mathcal{P}$ de forma que $E(X) = \mu$ e $Var(X) = \sigma_0^2$. O parâmetro $\mu \in \mathbb{R}$ é desconhecido e o parâmetro $\sigma_0^2 > 0$ é conhecido. Suponha que o seguinte estimador é proposto para μ:

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

Estude o erro quadrático médio desse estimador.

Solução. *Pelo Lema 2, temos*

$$E(\bar{X}) = \mu,$$
$$Var(\bar{X}) = \frac{\sigma_0^2}{n}.$$

Logo,

$$EQM_{\bar{X}}(\mu) = Var(\bar{X}) + B_{\bar{X}}^2(\mu) = Var(\bar{X}) + \left(E(\bar{X}) - \mu\right)^2 = \frac{\sigma_0^2}{n},$$

qualquer que seja $\mu \in \mathbb{R}$.

No Exemplo 20, qualquer que seja o valor real do parâmetro μ o erro quadrático médio será constante a igual $\frac{\sigma_0^2}{n}$. Logo, quando o tamanho n da amostra cresce, o erro quadrático médio do estimador $\bar{X}$ tende a zero.

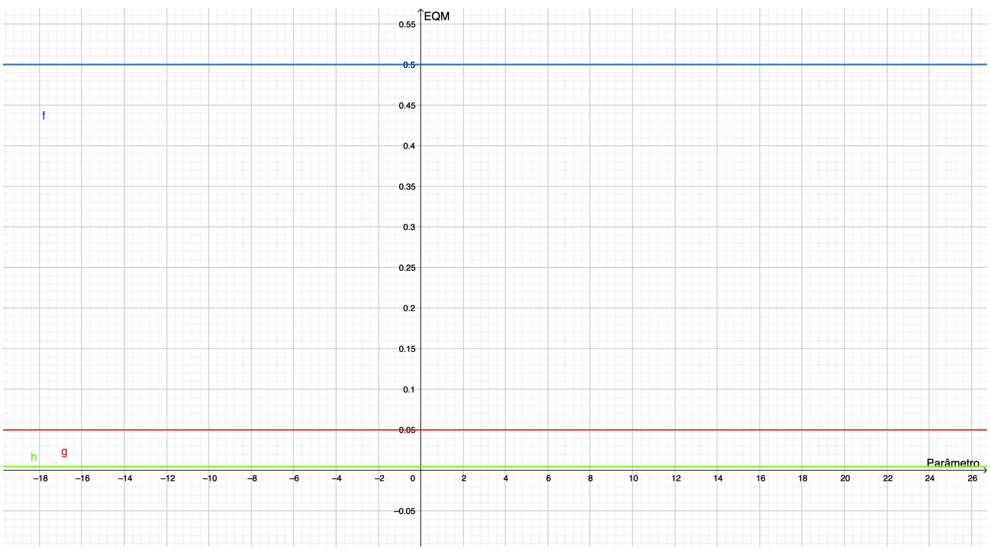

Figura 4: Erro quadrático médio $EQM_{\bar{X}}$ do estimador $\bar{X}$ do Exemplo 20 em função dos possíveis valores do parâmetro μ considerando $\sigma_0^2 = 5$ e três tamanhos de amostra: $n = 10$ (reta azul), $n = 100$ (reta vermelha) e $n = 1.000$ (reta verde).

O erro quadrático médio é um critério importante para comparar a performance de dois estimadores.

Definição 12: Melhor estimador

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ associado a um experimento aleatório. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória $X : \Omega \to \mathbb{R}$ com função de distribuição (ou densidade) de probabilidade $p(x|\theta)$ (ou $f(x|\theta)$), em que $\theta \in \Theta$ é um parâmetro desconhecido. O estimador $\hat{\theta}_1$ para θ é dito ser **melhor** que o estimador $\hat{\theta}_2$ quando

$$\mathsf{EQM}_{\hat{\theta}_1}(\theta) \leq \mathsf{EQM}_{\hat{\theta}_2}(\theta),$$

para todo $\theta \in \Theta$, com $\leq$ substituído por $<$ pelo menos para um valor de θ.

Note que quando estamos em busca do melhor estimador $\hat{\theta}^*$ para um parâmetro $\theta \in \Theta$ pertencente à classe $\mathcal{U} = \left\{\hat{\theta} \in \Theta : B_{\hat{\theta}}(\theta) = 0\right\}$ dos estimadores não-viesados para θ, então devemos ter

$$\mathsf{EQM}_{\hat{\theta}^*}(\theta) \leq \mathsf{EQM}_{\hat{\theta}}(\theta),$$

para todo $\theta \in \Theta$ e para todo $\hat{\theta} \in \mathcal{U}$ tal que $\hat{\theta} \neq \hat{\theta}^*$, com $\leq$ substituído por $<$ pelo menos para um valor de θ. Neste caso, $\hat{\theta}^*$ é também o ENVVM para θ, pois, desde que $\hat{\theta}^*$ e $\hat{\theta}$ pertencem à classe $\mathcal{U}$, $\mathsf{EQM}_{\hat{\theta}^*}(\theta) = \mathsf{Var}(\hat{\theta}^*)$ e $\mathsf{EQM}_{\hat{\theta}}(\theta) = \mathsf{Var}(\hat{\theta})$.

Exemplo 21

Seja (X_1, X_2, X_3) uma amostra aleatória da variável aleatória X que representa alguma característica observável de uma população $\mathcal{P}$ de forma que $\mathsf{E}(X) = \theta$ e $\mathsf{Var}(X) = 1$. Suponha que os seguintes estimadores são propostos para θ:

$$\hat{\theta}_1 := \bar{X} = \frac{X_1 + X_2 + X_3}{3} \quad \text{e} \quad \hat{\theta}_2 := \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3.$$

Decida entre $\hat{\theta}_1$ e $\hat{\theta}_2$ qual é o melhor estimador para θ.

Solução. *Por um lado, no Exemplo 18 encontramos que* $\hat{\theta}_1$ *e* $\hat{\theta}_2$ *são estimadores não-viesado para* θ, *i.e.,*$B_{\hat{\theta}_1} = B_{\hat{\theta}_2} = 0$ *Por outro lado, no Exemplo 19, verificamos que* $\mathsf{Var}(\hat{\theta}_1) < \mathsf{Var}(\hat{\theta}_2)$, *em particular,* $\mathsf{Var}(\hat{\theta}_1) = 1/3$ *e* $\mathsf{Var}(\hat{\theta}_2) = 3/4$. *Assim,* $\mathsf{EQM}_{\hat{\theta}_1}(\theta) < \mathsf{EQM}_{\hat{\theta}_2}(\theta)$ *para todo* $\theta \in \Theta$. *Portanto,* $\hat{\theta}_1$ *é um estimador melhor do que* $\hat{\theta}_2$ *para* θ.

1.3.1 Exercícios

1. Seja $(X_1, X_2, \ldots, X_7)$ uma amostra aleatória da variável aleatória X que representa uma característica observável de uma população $\mathcal{P}$ de forma que $E(X) = \mu$ e $Var(X) = \sigma^2$, $\mu \in \mathbb{R}$ e $\sigma^2 > 0$. Suponha que os seguintes estimadores são proposto para μ:

$$\hat{\mu}_1 := \bar{X} = \frac{1}{7}\sum_{i=1}^{7} X_i \quad \text{e} \quad \hat{\mu}_2 := \frac{2X_1 - X_6 + X_4}{2}.$$

   (a) Determine o viés dos estimadores $\hat{\mu}_1$ e $\hat{\mu}_2$ para o parâmetro μ.

   (b) Calcule $Var(\hat{\mu}_1)$ e $Var(\hat{\mu}_2)$.

   (c) Decida entre $\hat{\mu}_1$ e $\hat{\mu}_2$ qual é o melhor estimador para o parâmetro μ de acordo com o erro quadrático médio desses estimadores.

   (d) Represente em um mesmo plano cartesiano os gráficos dos erros quadrático médio de $\hat{\mu}_1$ e $\hat{\mu}_2$ em função do parâmetro μ.

2. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória X que representa uma característica observável de uma população $\mathcal{P}$ de forma que $X \sim Bernoulli(p)$, $p \in]0, 1[$. Suponha que os seguintes estimadores são proposto para p:

$$\hat{p}_1 := \frac{S}{n} = \frac{1}{n}\sum_{i=1}^{7} X_i \quad \text{e} \quad \hat{p}_2 := \begin{cases} 1, & \text{se a primeira observação consistir sucesso,} \\ 0, & \text{caso contrário.} \end{cases}$$

   (a) Determine o viés dos estimadores $\hat{p}_1$ e $\hat{p}_2$ para o parâmetro p.

   (b) Calcule $Var(\hat{p}_1)$ e $Var(\hat{p}_2)$.

   (c) Decida entre $\hat{p}_1$ e $\hat{p}_2$ qual é o melhor estimador para o parâmetro p de acordo com o erro quadrático médio desses estimadores.

   (d) Represente em um mesmo plano cartesiano os gráficos dos erros quadrático médio de $\hat{p}_1$ e $\hat{p}_2$ em função do parâmetro p.

3. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória X que representa uma característica observável de uma população $\mathcal{P}$ de forma que $E(X) = \mu$ e $Var(X) = \sigma^2$, $\mu \in \mathbb{R}$ e $\sigma^2 > 0$. Suponha que o estimador $\bar{X}^2$ tenha sido proposto para μ^2:

   (a) Determine o viés do estimador $\bar{X}^2$ para o parâmetro μ^2.

   (b) O que acontece com a viés do estimador conforme aumentamos o tamanho n da amostra?

   (c) Represente no plano cartesiano o gráfico do erro quadrático médio de $\bar{X}^2$ em função do parâmetro μ^2.

4. Dados sobre a espessura de óxido de semicondutores são os seguintes: 425, 431, 416, 419, 421, 436, 418, 410, 431, 433, 423, 426, 410, 435, 436, 428, 411, 426, 409, 437, 422, 428, 413 e 416.

   (a) Calcule a estimativa pontual da espessura média do óxido para todas os semicondutores da população.

   (b) Calcule a estimativa pontual do desvio-padrão do óxido para todas os semicondutores da população.

   (c) Calcule a estimativa pontual da proporção de semicondutores na população que tem espessura do óxido maior do que 430 angström.

   Nos itens (a), (b) e (c), considere como estimador, respectivamente, a média amostral, o desvio-padrão amostral e a proporção amostral.

5. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória X que representa uma característica observável de uma população $\mathcal{P}$ de forma que $X \sim \mathrm{Bernoulli}(p)$, $p \in]0,1[$. Suponha que os seguintes estimadores são proposto para p:

$$\hat{p}_1 := \frac{S}{n} = \frac{1}{n}\sum_{i=1}^{7} X_i \quad \text{e} \quad \hat{p}_2 := \frac{S + \sqrt{n}/2}{n + \sqrt{n}}.$$

   (a) Determine o viés dos estimadores $\hat{p}_1$ e $\hat{p}_2$ para o parâmetro p.

   (b) Calcule $\mathrm{Var}(\hat{p}_1)$ e $\mathrm{Var}(\hat{p}_2)$.

   (c) Decida entre $\hat{p}_1$ e $\hat{p}_2$ qual é o melhor estimador para o parâmetro p de acordo com o erro quadrático médio desses estimadores.

   (d) Represente em um mesmo plano cartesiano os gráficos dos erros quadrático médio de $\hat{p}_1$ e $\hat{p}_2$ em função do parâmetro p.

6. Em uma pequena vila no norte do Brasil foi conduzido um levantamento estatístico a fim de se estudar a expectativa de vida das pessoas que vivem naquela região. Após acessar os documentos do cartório local obteve-se os seguintes tempos de vida associados a cada elemento da população: 54, 61, 58, 49 e 62. Para todas as possíveis amostras de tamanho 2 que podem ser retiradas dessa população com reposição,

   (a) calcule a estimativa pontual do tempo de vida médio para todas as pessoas dessa população,

   (b) calcule a estimativa pontual do desvio-padrão do tempo de vida médio para todas as pessoas dessa população.

   Nos itens (a) e (b), considere como estimador, respectivamente, a média amostral e o desvio-padrão amostral.