7

Distribuições Amostrais e Estimação Pontual de Parâmetros

OBJETIVOS DE APRENDIZAGEM

Depois de um cuidadoso estudo deste capítulo, você deve ser capaz de:

1. Explicar os conceitos gerais de estimação de parâmetros de uma população ou de uma distribuição de probabilidades
2. Explicar o papel importante da distribuição normal como uma distribuição amostral
3. Entender o Teorema do Limite Central
4. Explicar propriedades importantes dos estimadores pontuais, incluindo tendenciosidade, variância e erro quadrático médio.

7-1 Introdução

- O campo da inferência estatística consiste naqueles métodos usados para tomar decisões ou tirar conclusões acerca de uma **população.**

- Para tirar conclusões, esses métodos utilizam a informação contida em uma **amostra** proveniente da população.

- Inferência Estatística pode ser dividida em duas grandes áreas:

    - Estimação de Parâmetros

    - Teste de Hipóteses

7-1 Introdução

Suponha que queiramos obter uma estimativa pontual de um parâmetro de uma população. Sabemos que antes dos dados serem coletados, as observações são consideradas variáveis aleatórias, isto é, $X_1, X_2, \ldots, X_n$. Logo, qualquer função da observação, ou qualquer **estatística**, é também uma variável aleatória. Por exemplo, a média da amostra $\overline{X}$ e a variância da amostra S^2 são estatísticas e são também variáveis aleatórias.

Desde que uma estatística seja uma variável aleatória, ela terá uma distribuição de probabilidades. Chamamos a distribuição de probabilidades de uma estatística de uma **distribuição amostral**. A noção de uma distribuição amostral é muito importante e será discutida e ilustrada mais adiante neste capítulo.

Definição

Uma **estimativa pontual** de algum parâmetro de uma população θ é um único valor númerico $\hat{\theta}$ de uma estatística $\hat{\Theta}$. A estatística $\hat{\Theta}$ é chamada de **estimador pontual**.

7-1 Introdução

Problemas de estimação ocorrem frequentemente em engenharia. Geralmente necessitamos estimar:

- A média μ de uma única população
- A variância σ^2 (ou desvio-padrão σ) de uma única população
- A proporção p de itens em uma população que pertence a uma classe de interesse
- A diferença nas médias de duas populações, $\mu_1 - \mu_2$
- A diferença nas proporções de duas populações, $p_1 - p_2$

7-1 Introdução

Estimativas pontuais razoáveis desses parâmetros são dadas a seguir:

- Para μ, a estimativa é $\hat{\mu} = \bar{x}$ a média amostral
- Para σ2, a estimativa é $\hat{\sigma}^2 = s^2$, variância amostral
- Para p, a estimativa é $\hat{p} = x/n$, a proporção da amostra, sendo x o número de itens em uma amostra aleatória de tamanho n que pertence à classe de interesse.
- Para $\mu_1 - \mu_2$, a estimativa é $\hat{\mu}_1 - \hat{\mu}_2 = \bar{x}_1 - \bar{x}_2$, a diferença entre as médias de duas amostras aleatórias independentes.
- Para $p_1 - p_2$, a estimativa é $\hat{p}_1 - \hat{p}_2$, a diferença entre duas proporções amostrais, calculadas a partir de duas amostras aleatórias independentes.

7-2 Distribuições Amostrais e Teorema do Limite Central

A Inferência Estatística lida em tomar decições acerca de uma população, baseando-se na informação contida em uma amostra aleatória proveniente daquela população.

Definições:

As variáveis aleatórias $X_1, X_2, \ldots, X_n$ são uma **amostra aleatória** de tamanho n, se (a) os X_i's forem variáveis aleatórias independentes e (b) cada X_i tiver a mesma distribuição de probabilidades.

Uma **estatística** é qualquer função das observações em uma amostra aleatória.

A distribuição de probabilidade de uma estatística é chamada de uma **distribuição amostral**.

7-2 Distribuições Amostrais e Teorema do Limite Central

Se estivermos amostrando de uma população que tenha uma distribuição desconhecida de probabilidades, a distribuição amostral média da média da amostra será aproximadamente normal, com média μ e variância σ^2/n, se o tamanho n da amostra for grande. Esse é um dos mais utéis teoremas em estatística, o chamado **Teorema do Limite Central**. O enunciado é dado a seguir:

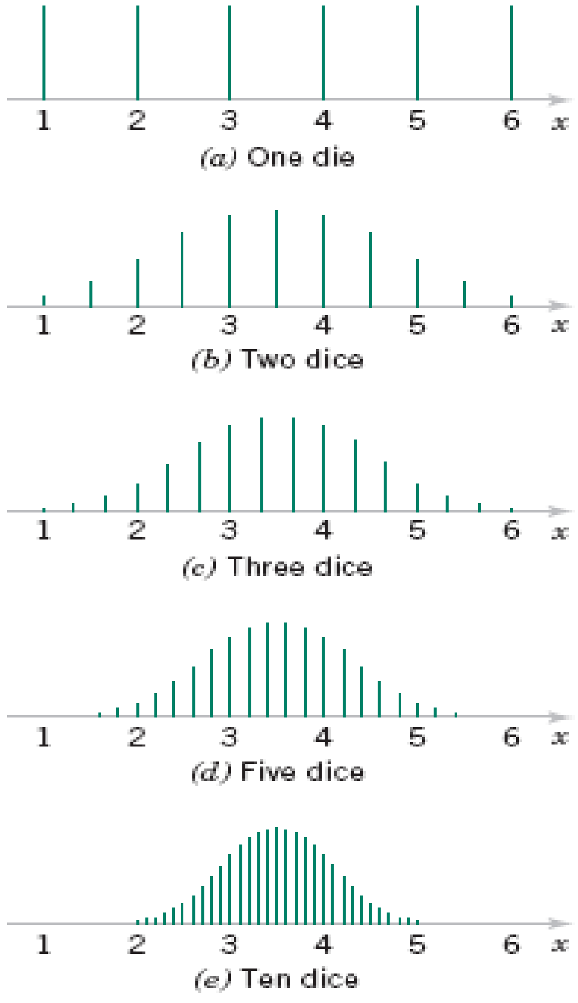
Se $X_1, X_2, \ldots, X_n$ for uma amostra aleatória de tamanho n, retirada de uma população (finita ou infinita), com média μ e variância finita σ^2, e se $\overline{X}$ for a média da amostra, então a forma limite da distribuição de:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \qquad (7\text{-}1)$$

quando $n \to \infty$ é a distribuição normal padrão.

7-2 Distribuições Amostrais e Teorema do Limite Central

Figura 7-1 Distribuições das pontuações médias obtidas quando do arremesso de dados. [Adaptado, com permissão de Box, Hunter, and Hunter (1978).]



7-2 Distribuições Amostrais e Teorema do Limite Central

Exemplo 7-1

Uma companhia eletrônica fabrica resistores que têm uma resistência média de 100 ohms e um desvio-padrão de 10 ohms. A distribuição de resistências é normal. Encontre a probabilidade de uma amostra aleatória de n= 25 resistores ter uma resistência média menor que 95 ohms.

Note que a distribuição amostral de $\overline{X}$ é normal, com média $\mu_{\overline{X}} = 100$ ohms e um desvio padrão de:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Consequentemente, a probabilidade desejada corresponde à área sombreada na Fig. 7-1. Padronizando o ponto $\overline{X} = 95$ na Fig 7.2, encontramos que

$$z = \frac{95 - 100}{2} = -2.5$$

E desse modo,

7-2 Distribuições Amostrais e Teorema do Limite Central

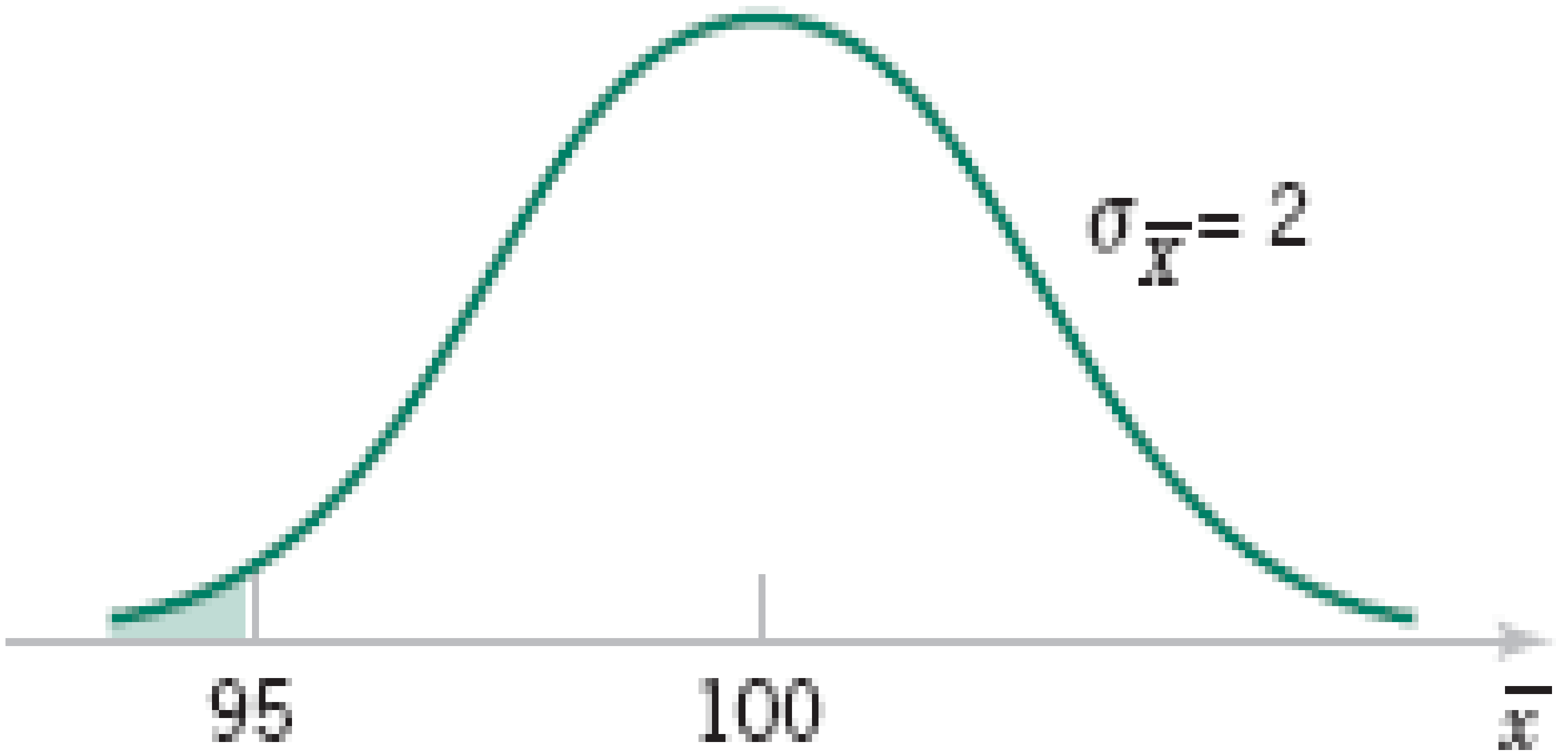


Figure 7-2 Probabilidade do Exemplo 7-1

7-2 Distribuições Amostrais e Teorema do Limite Central

Distribuição Amostral Aproximada de uma Diferença nas Médias de Amostras

Se tivermos duas populações independentes, com médias μ_1 e μ_2 e variâncias σ_1^2 e σ_2^2, e se $\overline{X}_1$ e $\overline{X}_2$ forem as médias de duas amostras aleatórias independentes de tamnho n1 e n2 dessas populações então a distribuição amostral de

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \qquad (7\text{-}4)$$

Será aproximadamente normal padrão, se as condições do teorema do limite central se aplicarem. Se as duas populações forem normais, então a distribuição amostral de Z será exatamente normal padrão.

7-3 Conceitos Gerais de Estimação Pontual

7-3.1 Estimadores Não-tendenciosos

Definição

O Estimador pontual $\hat{\Theta}$ é um **estimador não-tendencioso** para o parâmetro θ, se

$$E(\hat{\Theta}) = \theta$$

Se o estimador for tendencioso, então a diferença

$$E(\hat{\Theta}) - \theta$$

É chamada de **tendenciosidade** do estimator $\hat{\Theta}$.

7-3 Conceitos Gerais de Estimação Pontual

Exemplo 7-4

Suponha que X seja uma variável aleatória com média μ e variância σ^2. Seja X1, X2, … , Xn uma amostra aleatória de tamanho n, proveniente de uma população representada por X. Mostre que a média da amostra $\overline{X}$ e a variância da amostra S^2 são estimadores não-tendenciosos de μ e σ^2 respectivamente.

Considere primeiro a média da amostra. Na Seção 5.5 do Capítulo 5, mostramos que E($\overline{X}$) = μ. Consequentemente a média da amostra $\overline{X}$ é um estimador não-tendencioso da média da população μ. Considere agora a variância da amostra. Temos:

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right] = \frac{1}{n-1} E \sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$= \frac{1}{n-1} E \sum_{i=1}^{n}(X_i^2 + \overline{X}^2 - 2\overline{X}X_i) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right]$$

7-3 Conceitos Gerais de Estimação Pontual

Exemplo 7-4 (continuação)

A última igualdade vem da equação para a média de uma função linear no Capítulo 5. Entretanto, uma vez que $E(X_i^2) = \mu^2 + \sigma^2$ e $E(\overline{X}^2) = \mu^2 + \sigma^2/n$, temos

$$
\begin{aligned}
E(S^2) &= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2+\sigma^2) - n(\mu^2+\sigma^2/n)\right] \\
&= \frac{1}{n-1}(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2) \\
&= \sigma^2
\end{aligned}
$$

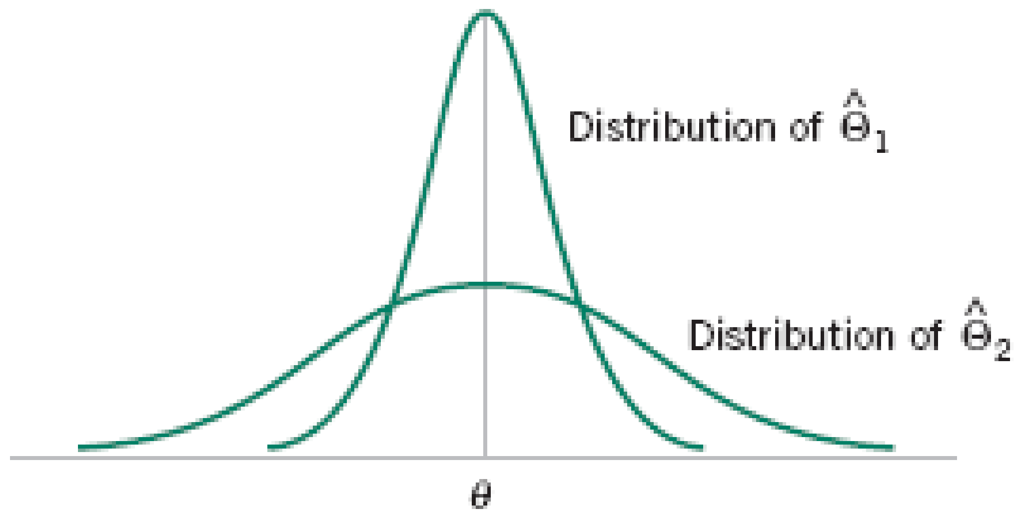
7-3 Conceitos Gerais de Estimação Pontual

7-3.2 Variância de um Estimador Pontual

Definição

Se considerarmos todos os estimadores não-tendenciosos de θ, aquele com a menor variância será chamado de estimador **não-tendencioso de variância mínima (ENTVM).**

Figure 7-5 As distribuições amostrais de dois estimadores não-tendenciosos $\hat{\Theta}_1$ e $\hat{\Theta}_2$.



7-3 Conceitos Gerais de Estimação Pontual

7-3.2 Variância de um Estimador Pontual

Se $X_1, X_2, \ldots, X_n$ for uma amostra aleatória de tamanho n, proveniente de uma distribuição normal com média μ e variância σ^2, então a média da amostra $\overline{X}$ será o ENTVM (Estimador Não-Tendencioso de Variância Mínima) para μ.

7-3 Conceitos Gerais de Estimação Pontual

7-3.3 Erro-Padrão de um Estimador

Definição

O erro-padrão de um estimador $\hat{\Theta}$ é o seu desvio-padrão, dado por $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. Se o erro padrão envolver parâmetros desconhecidos que possam ser estimados, então a substituição daqueles valores em $\sigma_{\hat{\Theta}}$ produz um erro-padrão estimado denotado por $\hat{\sigma}_{\hat{\Theta}}$.

7-3 Conceitos Gerais de Estimação Pontual

7-3.3 Erro-Padrão de Estimador

Suponha que estejamos amostrando a partir de uma distribuição normal, com média μ e variância σ^2. Agora, a distribuição de $\overline{X}$ é normal, com média μ e variância σ^2/n; assim o erro-padrão de $\overline{X}$ é:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Se não conhecêssemos σ, mas substituirmos o desvio-padrão S da amostra na equação anterior, então o erro-padrão etimado de $\overline{X}$ seria:

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$$