

# Estimação: (A) Propriedades e Distribuições Amostrais

Wagner H. Bonat  
Fernando P. Mayer  
Elias T. Krainski

Universidade Federal do Paraná  
Departamento de Estatística  
Laboratório de Estatística e Geoinformação

01/05/2018



# Sumário

## 1 Introdução

## 2 Estimação pontual

- Parâmetros, estimadores e estimativas
- Propriedades dos estimadores

## 3 Distribuições amostrais

- Introdução
- Distribuição amostral da média
- Distribuição amostral da proporção

## 4 Exercícios

# Inferência estatística

Seja  $X$  uma variável aleatória com função densidade (ou de probabilidade) denotada por  $f(x, \theta)$ , em que  $\theta$  é um parâmetro desconhecido. Chamamos de **inferência estatística** o problema que consiste em especificar um ou mais valores para  $\theta$ , baseado em um conjunto de valores  $X$ .

A inferência pode ser feita de duas formas:

- estimativa pontual
- estimativa intervalar

# Redução de dados

Um experimentador usa as informações em uma amostra aleatória  $X_1, \dots, X_n$  para se fazer inferências sobre  $\theta$ .

Normalmente  $n$  é grande e fica inviável tirar conclusões baseadas em uma longa **lista** de números.

Por isso, um dos objetivos da inferência estatística é **resumir** as informações de uma amostra, da maneira mais **compacta** possível, mas que ao mesmo tempo seja também **informativa**.

Normalmente esse resumo é feito por meio de **estatísticas**, por exemplo, a média amostral e a variância amostral.

# População e amostra

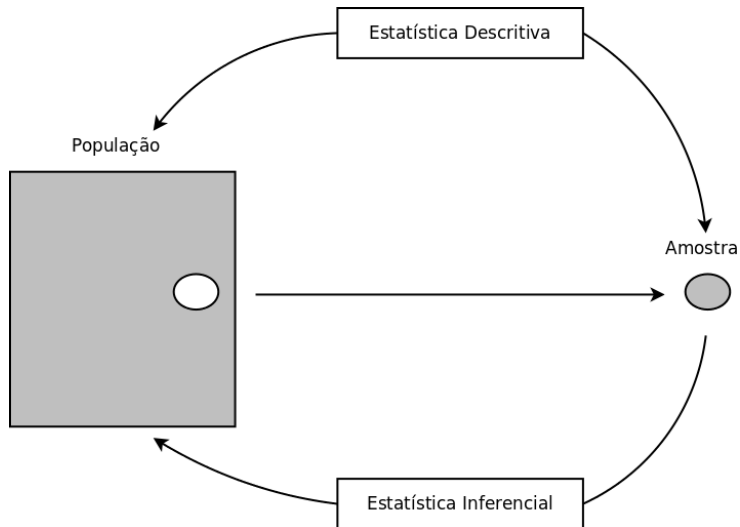
O conjunto de valores de uma característica associada a uma coleção de indivíduos ou objetos de interesse é dito ser uma população.

Uma sequência  $X_1, \dots, X_n$  de  $n$  variáveis aleatórias independentes e identicamente distribuídas (iid) com função densidade (ou de probabilidade)  $f(x, \theta)$  é dita ser uma amostra aleatória de tamanho  $n$  da distribuição de  $X$ .

Como normalmente  $n > 1$ , então temos que a fdp ou fp conjunta será

$$f(\mathbf{x}, \theta) = f(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta).$$

# População e amostra



# Sumário

- 1 Introdução
- 2 Estimação pontual
  - Parâmetros, estimadores e estimativas
  - Propriedades dos estimadores
- 3 Distribuições amostrais
  - Introdução
  - Distribuição amostral da média
  - Distribuição amostral da proporção
- 4 Exercícios

# Parâmetro e Estatística

## População → censo → parâmetro

*Uma medida numérica que descreve alguma característica da **população**, usualmente representada por letras gregas:  $\theta$ ,  $\mu$ ,  $\sigma$ , ...*

Exemplo: média populacional =  $\mu$

---

## População → amostra → estatística

*Uma medida numérica que descreve alguma característica da **amostra**, usualmente denotada pela letra grega do respectivo parâmetro com um acento circunflexo:  $\hat{\theta}$ ,  $\hat{\mu}$ ,  $\hat{\sigma}$ , ..., ou por letras do alfabeto comum:  $\bar{x}$ ,  $s$ , ...*

Exemplo: média amostral =  $\bar{x}$



# Parâmetros

É importante notar que um parâmetro não é restrito aos modelos de probabilidade. Por exemplo:

- $X \sim N(\mu, \sigma^2) \Rightarrow$  parâmetros:  $\mu, \sigma^2$
- $Y \sim \text{Poisson}(\lambda) \Rightarrow$  parâmetro:  $\lambda$
- $Y = \beta_0 + \beta_1 X \Rightarrow$  parâmetros:  $\beta_0, \beta_1$
- $L_t = L_\infty[1 - e^{-k(t-t_0)}] \Rightarrow$  parâmetros:  $L_\infty, k, t_0$

# Estatística

Qualquer função da amostra que não depende de parâmetros desconhecidos é denominada uma estatística, denotada por  $T(\mathbf{X}) = T(X_1, X_2, \dots, X_n)$

Exemplos:

- $T_1(\mathbf{X}) = \sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n$
- $T_2(\mathbf{X}) = \prod_{i=1}^n X_i = X_1 \cdot X_2 \cdot \dots \cdot X_n$
- $T_3(\mathbf{X}) = X_{(1)}$
- $T_4(\mathbf{X}) = \sum_{i=1}^n (X_i - \mu)^2$

# Estatística

Qualquer função da amostra que não depende de parâmetros desconhecidos é denominada uma estatística, denotada por  $T(\mathbf{X}) = T(X_1, X_2, \dots, X_n)$

Exemplos:

- $T_1(\mathbf{X}) = \sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n$
- $T_2(\mathbf{X}) = \prod_{i=1}^n X_i = X_1 \cdot X_2 \cdot \dots \cdot X_n$
- $T_3(\mathbf{X}) = X_{(1)}$
- $T_4(\mathbf{X}) = \sum_{i=1}^n (X_i - \mu)^2$

Verificamos que  $T_1$ ,  $T_2$ ,  $T_3$  são estatísticas, mas  $T_4$  não.

Como é uma função da amostra, então uma estatística também é uma **variável aleatória** → distribuições amostrais

# Estimador

## Espaço paramétrico

O conjunto  $\Theta$  em que  $\theta$  pode assumir seus valores é chamado de **espaço paramétrico**

## Estimador

Qualquer estatística que assume valores em  $\Theta$  é um estimador para  $\theta$ .

## Estimador pontual

Dessa forma, um **estimador pontual** para  $\theta$  é qualquer estatística que possa ser usada para estimar esse parâmetro, ou seja,

$$\hat{\theta} = T(\mathbf{X})$$

# Estimador

## Observações:

1. Todo estimador é uma estatística, mas nem toda estatística é um estimador.
2. O valor assumido pelo estimador pontual é chamado de **estimativa pontual**,

$$\hat{\theta} = T(\mathbf{X}) = T(X_1, \dots, X_n) = t$$

ou seja, o estimador é uma **função** da amostra, e a estimativa é o **valor observado** de um estimador (um número) de uma amostra particular.

# Estimação pontual

A ideia geral por trás da estimação pontual é muito simples:

Quando a amostragem é feita a partir de uma população descrita por uma função  $f(x, \theta)$ , o conhecimento de  $\theta$  a partir da amostra, gera todo o conhecimento para a população.

Dessa forma, é natural que se procure um **método** para se achar um **bom** estimador para  $\theta$ .

Existem algumas **propriedades** que definem o que é um bom estimador, ou o “**melhor**” estimador entre uma série de candidatos.

# Estimação pontual

**Localização do problema:** Considere  $X_1, \dots, X_n$  uma amostra aleatória de uma variável aleatória  $X$  com fdp ou fp  $f(x, \theta)$ ,  $\theta \in \Theta$ . Sejam:

$$\hat{\theta}_1 = T_1(X_1, \dots, X_n) \quad \hat{\theta}_2 = T_2(X_1, \dots, X_n)$$

Qual dos dois estimadores pontuais é **melhor** para  $\theta$ ?

Como não conhecemos  $\theta$ , não podemos afirmar que  $\hat{\theta}_1$  é melhor do que  $\hat{\theta}_2$  e vice-versa.

O problema da estimação pontual é então escolher um estimador  $\hat{\theta}$  que se aproxime de  $\theta$  segundo algumas **propriedades**.

# Estimação pontual

**Exemplo 1:** Considere uma amostra aleatória  $(X_1, \dots, X_n)$  de uma variável aleatória  $X \sim N(\mu = 3, \sigma^2 = 1)$  e os estimadores pontuais para  $\mu$

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{e} \quad \hat{\theta}_2 = \frac{X_{(1)} + X_{(n)}}{2}$$

Qual dos dois estimadores pode ser considerado como o **melhor** para estimar o verdadeiro valor de  $\mu$ ?

Considere os seguintes pseudo-códigos para um estudo de simulação do comportamento destes dois estimadores:



# Estimação pontual

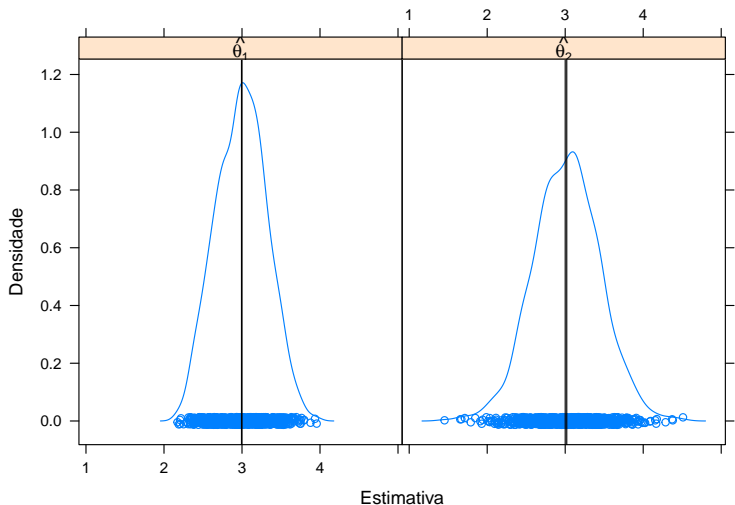
## Pseudo-código 1

- Simule uma amostra de tamanho  $n = 10$  da distribuição considerada
- Para essa amostra, calcule a média ( $\hat{\theta}_1$ ) e o ponto médio ( $\hat{\theta}_2$ )
- Repita os passos (1) e (2) acima  $m = 1000$  vezes
- Faça um gráfico da densidade das  $m = 1000$  estimativas de  $\hat{\theta}_1$  e  $\hat{\theta}_2$  e verifique seu comportamento

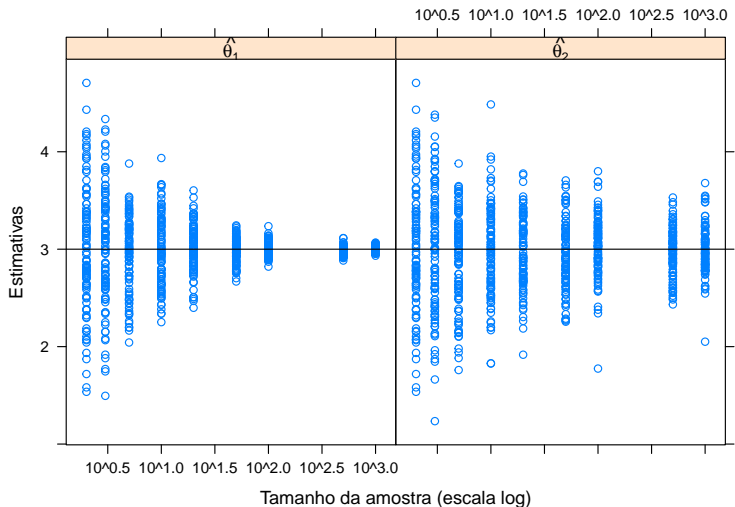
## Pseudo-código 2

- Simule amostras de tamanhos ( $n$ ) 2, 3, 5, 10, 20, 50, 100, 500, 1000 da distribuição considerada
- Para cada amostra de tamanho  $n$ , calcule a média ( $\hat{\theta}_1$ ) e o ponto médio ( $\hat{\theta}_2$ )
- Repita os passos (1) e (2) acima  $m = 100$  vezes
- Faça um gráfico das  $m = 100$  estimativas de  $\hat{\theta}_1$  e  $\hat{\theta}_2$  para cada tamanho de amostra  $n$  e verifique seu comportamento

# Estimação pontual: Pseudo-código 1 - $X \sim N(3, 1)$



# Estimação pontual: Pseudo-código 2 - $X \sim N(3, 1)$

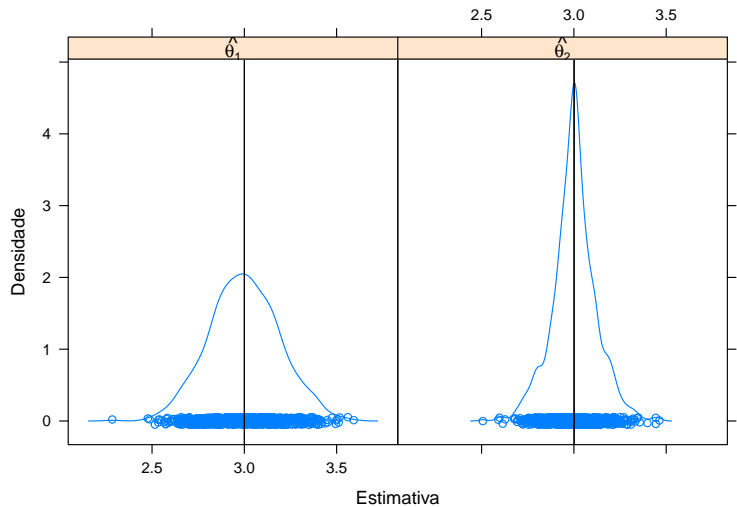


# Estimação pontual

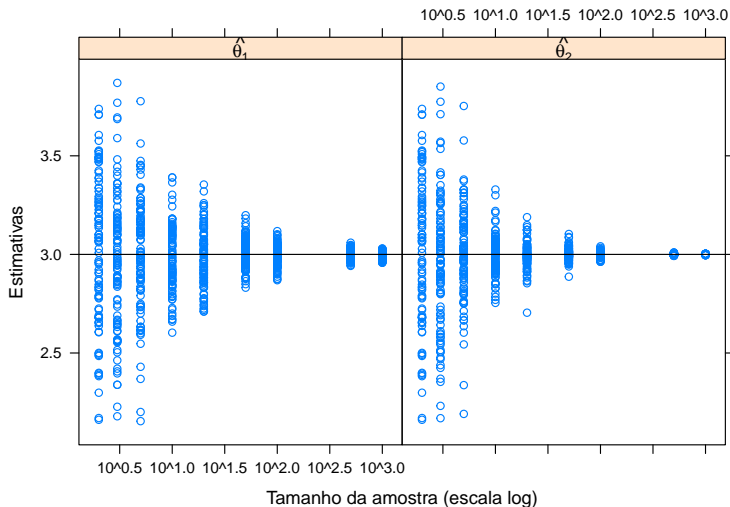
**Exemplo 2:** Considere uma amostra aleatória  $(X_1, \dots, X_n)$  de uma variável aleatória  $Y \sim U(\min = 2, \max = 4)$  (distribuição uniforme no intervalo  $[2, 4]$ ) e os estimadores pontuais para  $\mu$

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{e} \quad \hat{\theta}_2 = \frac{X_{(1)} + X_{(n)}}{2}$$

Qual dos dois estimadores pode ser considerado como o **melhor** para estimar a média de  $Y$ ?

Estimação pontual: Pseudo-código 1 -  $Y \sim U(2, 4)$ 

# Estimação pontual: Pseudo-código 2 - $Y \sim U(2, 4)$



# Propriedades dos estimadores

De modo geral, um “**bom**” estimador deve ser:

1. Não viciado
2. Consistente
3. Eficiente

# Propriedades dos estimadores: 1. Vício

## Erro quadrático médio (EQM)

O Erro Quadrático Médio (EQM) de um estimador  $\hat{\theta}$  de  $\theta$  é dado por

$$\begin{aligned}\text{EQM}[\hat{\theta}] &= E[(\hat{\theta} - \theta)^2] \\ &= \text{Var}[\hat{\theta}] + B[\hat{\theta}]^2\end{aligned}$$

onde

$$B[\hat{\theta}] = E[\hat{\theta}] - \theta$$

é denominado de **vício** do estimador  $\hat{\theta}$ . Portanto, dizemos que um estimador é **não viciado** para  $\theta$  quando

$$B[\hat{\theta}] = 0 \quad \Rightarrow \quad E[\hat{\theta}] = \theta$$



# Propriedades dos estimadores: 1. Vício

## Estimador não viciado

Seja  $(X_1, \dots, X_n)$ , uma amostra aleatória de uma variável aleatória com fdp ou fp  $f(x, \theta)$ ,  $\theta \in \Theta$ , dizemos que o estimador  $\hat{\theta} = T(\mathbf{X})$  é não viciado para  $\theta$  se

$$E[\hat{\theta}] = E[T(\mathbf{X})] = \theta \quad \forall \theta \in \Theta$$

Um estimador  $\hat{\theta}$  é dito **assintoticamente não viciado** se

$$\lim_{n \rightarrow \infty} E[\hat{\theta}] = \theta$$

Ou seja, para grandes amostras,  $\hat{\theta}$  passa a ser imparcial.

## Propriedades dos estimadores: 2. Consistência

### Estimador consistente

Seja  $(X_1, \dots, X_n)$ , uma amostra aleatória de uma variável aleatória com fdp ou fp  $f(x, \theta)$ ,  $\theta \in \Theta$ , o estimador  $\hat{\theta} = T(\mathbf{X})$  é consistente para  $\theta$  se satisfaz simultaneamente

$$\lim_{n \rightarrow \infty} E[\hat{\theta}] = \theta$$

e

$$\lim_{n \rightarrow \infty} \text{Var}[\hat{\theta}] = 0$$

## Propriedades dos estimadores: 3. Eficiência

### Eficiência relativa

Sejam  $\hat{\theta}_1 = T_1(\mathbf{X})$  e  $\hat{\theta}_2 = T_2(\mathbf{X})$  dois estimadores pontuais **não viciados** para  $\theta$ . A eficiência relativa de  $\hat{\theta}_1$  em relação a  $\hat{\theta}_2$  é

$$ER[\hat{\theta}_1, \hat{\theta}_2] = \frac{\text{Var}[\hat{\theta}_2]}{\text{Var}[\hat{\theta}_1]}$$

Se:

- $ER[\hat{\theta}_1, \hat{\theta}_2] > 1 \Rightarrow \hat{\theta}_2$  é mais eficiente
- $ER[\hat{\theta}_1, \hat{\theta}_2] < 1 \Rightarrow \hat{\theta}_1$  é mais eficiente

# Propriedades dos estimadores

**Exemplo:** média amostral  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  como estimador da média populacional  $\mu$ :

$$E(\bar{x}) = E \left[ \frac{1}{n} \sum_{i=1}^n x_i \right] = \mu$$

$$\text{Var}(\bar{x}) = \text{Var} \left[ \frac{1}{n} \sum_{i=1}^n x_i \right] = \frac{\sigma^2}{n}$$

Portanto  $\bar{x}$  é um estimador **não viciado** e **consistente** para  $\mu$ .

# Propriedades dos estimadores

**Exemplo:** variância amostral  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  como estimador da variância populacional  $\sigma^2$ :

$$E(\hat{\sigma}^2) = E \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right] = \left( \frac{n-1}{n} \right) \sigma^2$$

Portanto  $\hat{\sigma}^2$  é um estimador **viciado** para  $\sigma^2$ . (Embora seja um estimador **assintoticamente** não viciado).

Para eliminar esse vício, podemos definir então um novo estimador:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \text{ e}$$

$$E(S^2) = E \left[ \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right] = \sigma^2$$

que é então um estimador **não viciado** para  $\sigma^2$ .

## Exemplo 7.11

Uma amostra  $(X_1, \dots, X_n)$  é retirada de uma população com  $X \sim N(\mu, \sigma^2)$ , e dois estimadores são propostos para  $\mu$ :

$$\hat{\mu}_1 = \bar{X} \quad \text{e} \quad \hat{\mu}_2 = \text{mediana}(X_1, \dots, X_n)$$

Qual dos dois é melhor para  $\mu$ ?

## Exemplo 7.11

Podemos notar que

$$\begin{aligned}E(\hat{\mu}_1) &= E(\bar{X}) = \mu \\ \text{Var}(\hat{\mu}_1) &= \text{Var}(\bar{X}) = \sigma^2/n\end{aligned}$$

$$\begin{aligned}E(\hat{\mu}_2) &= E(\text{mediana}(X_1, \dots, X_n)) = \mu \\ \text{Var}(\hat{\mu}_2) &= \text{Var}(\text{mediana}(X_1, \dots, X_n)) = (\pi/2)(\sigma^2/n)\end{aligned}$$

Portanto, ambos são estimadores não viciados e consistentes. Mas:

$$ER[\hat{\mu}_1, \hat{\mu}_2] = \frac{\text{Var}[\hat{\mu}_1]}{\text{Var}[\hat{\mu}_2]} = \frac{\sigma^2/n}{(\pi/2)(\sigma^2/n)} = \frac{2}{\pi} = 0,63$$

Como  $ER[\hat{\mu}_1, \hat{\mu}_2] < 1$  então  $\hat{\mu}_1 = \bar{X}$  é mais **eficiente**.

# Propriedades dos estimadores

O **erro padrão** de um estimador dá uma ideia da **precisão** da estimativa.

O erro padrão (EP) de um estimador é o seu desvio-padrão (raíz quadrada da variância), ou seja,

$$\text{EP}(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$$

**Exemplo:** Sabemos que a distribuição de  $\bar{X}$  tem média  $\mu$  e variância  $\sigma^2/n$ . Então o erro padrão de  $\bar{X}$  é

$$\text{EP}(\bar{X}) = \sqrt{\text{Var}(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$



# Sumário

- 1 Introdução
- 2 Estimação pontual
  - Parâmetros, estimadores e estimativas
  - Propriedades dos estimadores
- 3 Distribuições amostrais
  - Introdução
  - Distribuição amostral da média
  - Distribuição amostral da proporção
- 4 Exercícios

# Distribuições amostrais

De maneira geral, uma amostra de tamanho  $n$  será descrita pelos valores  $x_1, x_2, \dots, x_n$  das variáveis aleatórias  $X_1, X_2, \dots, X_n \Rightarrow$  **Amostra Aleatória**

No caso de uma Amostragem Aleatória Simples (AAS) **com reposição**,  $X_1, X_2, \dots, X_n$  serão variáveis aleatórias **independentes e identicamente distribuídas** (iid) com função de probabilidade (fp) ou função densidade de probabilidade (fdp) conjunta dada por

$$f(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdots f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

Onde o mesmo valor do parâmetro  $\theta$  é utilizado em cada um dos termos no produto.

# Distribuições amostrais

Quando uma amostra  $X_1, X_2, \dots, X_n$  é obtida, geralmente estamos interessados em um resumo destes valores, que pode ser expresso matematicamente pela estatística  $T(x_1, x_2, \dots, x_n)$ .

Dessa forma,  $Y = T(x_1, x_2, \dots, x_n)$  é **também uma variável aleatória**. Se  $Y$  é uma VA, então ela possui uma **distribuição de probabilidade**.

Uma vez que a distribuição de  $Y$  é derivada da amostra  $X_1, X_2, \dots, X_n$ , vamos denominá-la de **distribuição amostral** de  $Y$ .

# Distribuições amostrais

## Distribuições amostrais

A distribuição de probabilidade de uma estatística  $Y = T(x_1, x_2, \dots, x_n)$  é denominada de **distribuição amostral** de  $Y$ . Assim, uma estatística também é uma variável aleatória, pois seus valores mudam conforme a amostra aleatória.

**Exemplo:** duas estatísticas comumente utilizadas para o resumo de uma amostra aleatória são a **média amostral**  $\bar{x}$ , e a **proporção amostral**  $\hat{p}$ . Cada uma delas também possui uma distribuição amostral.

# Distribuição amostral da média

Para estudarmos a distribuição amostral da estatística  $\bar{X}$ , considere uma população identificada pela VA  $X$ , com parâmetros

$$E(X) = \mu = \text{média}$$

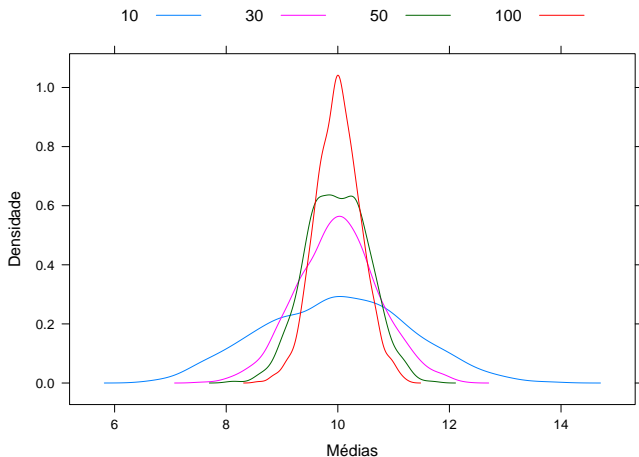
$$\text{Var}(X) = \sigma^2 = \text{variância}$$

supostamente conhecidos. Em seguida, realizamos os seguintes passos:

1. Retiramos  $m$  amostras aleatórias (AAS com reposição) de tamanho  $n$  dessa população
2. Para cada uma das  $m$  amostras, calculamos a média amostral  $\bar{x}$
3. Verificamos a distribuição das  $m$  médias amostrais e estudamos suas propriedades

## Exemplo 7.13

Seja  $X \sim N(10, 16)$ , como se comporta  $\bar{X}$  para  $n = 10, 30, 50, 100$ ?



# Distribuição amostral da média

Através do estudo da distribuição da média amostral chegamos em um dos resultados mais importantes da inferência estatística.

## Distribuição amostral da média

- $E(\bar{X}) = \mu_{\bar{X}} = \mu$
- $\text{Var}(\bar{X}) = \sigma_{\bar{X}}^2 = \sigma^2/n$

Portanto, se

$$X \sim N(\mu, \sigma^2) \quad \text{então} \quad \bar{X} \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}}^2)$$

mas, como

$$\mu_{\bar{X}} = \mu \quad \text{e} \quad \sigma_{\bar{X}}^2 = \sigma^2/n$$

então, a **distribuição amostral** da média amostral  $\bar{X}$  é

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

# Distribuição amostral da média

Pode-se mostrar que, para amostras suficientemente grandes, a **média amostral  $\bar{X}$  converge para o verdadeiro valor da média populacional  $\mu$**  (é um **estimador não viesado** de  $\mu$ ).

Além disso, a variância das médias amostrais  $\sigma_{\bar{X}}^2$  tende a diminuir conforme  $n \rightarrow \infty$  (é um estimador **consistente**).

Estes resultados sugerem que, quando o tamanho da amostra aumenta,

independente do formato da distribuição da população original,

a **distribuição amostral de  $\bar{X}$  aproxima-se cada vez mais de uma distribuição Normal**, um resultado fundamental na teoria de probabilidade conhecido como **Teorema Central do Limite**.



# Distribuição amostral da média

## Teorema Central do Limite (TCL)

Para amostras aleatórias simples  $(X_1, X_2, \dots, X_n)$ , retiradas de uma população com média  $\mu$  e variância  $\sigma^2$ , a distribuição amostral da média  $\bar{X}$ , terá forma dada por

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

no limite quando  $n \rightarrow \infty$ , onde  $Z \sim N(0, 1)$ .

- Se a população for normal, então  $\bar{X}$  terá distribuição *exata* normal.
- A rapidez da convergência para a normal depende da distribuição da população da qual as amostras foram geradas.

# Distribuição amostral da média

Ver figura [dist\\_amostrais.pdf](#)

# Distribuição amostral da média

Em palavras, o teorema garante que para  $n$  grande, a distribuição da média amostral, devidamente padronizada, **se comporta segundo um modelo normal** com média 0 e variância 1.

Pelo teorema, temos que quanto maior o tamanho da amostra, **melhor é a aproximação**.

Estudos envolvendo simulações mostram que, em muitos casos, **valores de  $n$  ao redor de 30** fornecem aproximações bastante boas para as aplicações práticas.

# Distribuição amostral da média

Quando calculamos a probabilidade de um valor estar em um determinado intervalo de valores, podemos usar o modelo Normal, como vimos anteriormente.

No entanto, quando temos uma **amostra**, e queremos calcular probabilidades associadas à **média amostral** (a probabilidade da média amostral estar em um determinado intervalo de valores), precisamos necessariamente usar os resultados do TCL.

# Distribuição amostral da média

**Exemplo:** Uma máquina de empacotamento que abastece pacotes de feijão apresenta distribuição normal com média de 500 g e desvio-padrão de 22 g. De acordo com as normas de defesa do consumidor, os pacotes de feijão não podem ter peso inferior a 2% do estabelecido na embalagem.

- a. Determine a probabilidade de **um pacote** selecionado aleatoriamente ter a peso inferior a 490 g.
- b. Determine a probabilidade de **20 pacotes** selecionados aleatoriamente terem peso médio inferior a 490 g.
- c. Como podemos interpretar os resultados dos itens anteriores? O que é mais indicado para se tomar uma decisão sobre o funcionamento da máquina: selecionar um pacote ou uma amostra de pacotes?

# Distribuição amostral da proporção

Muitas vezes, o interesse é conhecer uma **proporção**, e não a média de uma população.

Suponha que uma amostra de tamanho  $n$  foi obtida de uma população, e que  $x \leq n$  observações nessa amostra pertençam a uma classe de interesse (ex.: pessoas do sexo masculino).

Dessa forma, a proporção amostral

$$\hat{p} = \frac{x}{n} = \frac{\text{número de sucessos}}{\text{total de tentativas}}$$

é o “melhor estimador” para a proporção populacional  $p$ .

Note que  $n$  e  $p$  são os parâmetros de uma **distribuição binomial**.

## Distribuição amostral da proporção

**Exemplo:** em 5 lançamentos de uma moeda considere que o evento “cara” (C) seja o sucesso (“sucesso” = 1; “fracasso” = 0). Um possível resultado seria o conjunto {C, C, R, R, C}. A proporção amostral seria

$$\hat{p} = \frac{x}{n} = \frac{\text{número de sucessos}}{\text{total de tentativas}} = \frac{3}{5} = 0,6$$

**Exemplo:** em uma amostra de 2500 eleitores de uma cidade, 1784 deles eram favoráveis à reeleição do atual prefeito. A proporção amostral é então

$$\hat{p} = \frac{x}{n} = \frac{\text{número de sucessos}}{\text{total de tentativas}} = \frac{1784}{2500} = 0,7136$$

# Distribuição amostral de uma proporção

A distribuição amostral de uma **proporção** é a distribuição das proporções de todas as possíveis amostras de tamanho  $n$  retiradas de uma população.

## Exemplo:

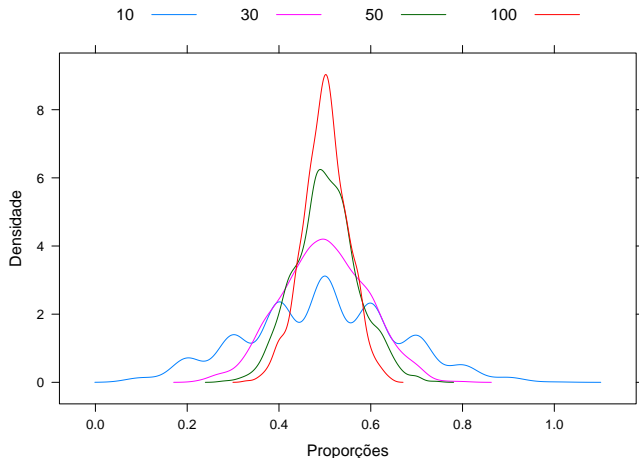
- Uma moeda é lançada  $n = 10, 30, 50, 100$  vezes, e a proporção de caras é registrada
- Esse processo é repetido  $m = 1000$  vezes

Com isso, concluímos que:

- A média das proporções para  $n \rightarrow \infty$  tende para a verdadeira proporção populacional  $p = 0,5$
- A **distribuição amostral** das proporções segue aproximadamente uma **distribuição normal**



# Distribuição amostral de uma proporção



# Distribuição amostral de uma proporção

Através do estudo da distribuição amostral da proporção, chegamos aos seguintes resultados

- $E(\hat{p}) = \mu_{\hat{p}} = p$
- $\text{Var}(\hat{p}) = \sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$

Ou seja,  $\hat{p}$  é um estimador **não viciado** e **consistente** para  $p$ .

Assim, a **distribuição amostral** de  $\hat{p}$  será

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

# Distribuição amostral de uma proporção

Note que o **erro padrão** de  $\hat{p}$  será

$$EP(\hat{p}) = \sqrt{\text{Var}(\hat{p})} = \sqrt{\frac{p(1-p)}{n}}$$

Assim, usando o TCL, podemos mostrar que a quantidade

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

segue uma distribuição **normal padrão** com média 0 e variância 1.

Quando não conhecemos  $p$ , usamos  $\hat{p} = x/n$  como estimativa para calcular o erro padrão.

## Exemplo 7.17

Suponha que a proporção de peças fora da especificação em um lote é de 40%. Uma amostra de 30 peças foi selecionada. Qual é a probabilidade da proporção de peças defeituosas ser menor do que 0,5?

## Exemplo 7.17

Suponha que a proporção de peças fora da especificação em um lote é de 40%. Uma amostra de 30 peças foi selecionada. Qual é a probabilidade da proporção de peças defeituosas ser menor do que 0,5?

$X \sim \text{Bin}(30, 0.4)$ . Assim:

$$\begin{aligned} P(\hat{p} < 0.5) &= P(X/30 < 0.5) = P(X < 15) \\ &= \sum_{x=0}^{14} \binom{30}{x} 0.4^x 0.6^{30-x} \end{aligned}$$

Usando o R:

```
pbinom(14, size = 30, prob = 0.4)
```

```
# [1] 0.8246309
```

## Exemplo 7.17

Considerando a aproximação pela Normal, temos

$$\hat{p} \sim N\left(0.4, \frac{0.4(1-0.4)}{30}\right)$$

Assim,

$$\begin{aligned} P(\hat{p} < 0.5) &\approx P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.5 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{30}}}\right) \\ &= P(Z < 1.12) = 0.8686 \end{aligned}$$

Usando o R:

```
pnorm(1.12)
```

```
# [1] 0.8686431
```

# Sumário

- 1 Introdução
- 2 Estimação pontual
  - Parâmetros, estimadores e estimativas
  - Propriedades dos estimadores
- 3 Distribuições amostrais
  - Introdução
  - Distribuição amostral da média
  - Distribuição amostral da proporção
- 4 Exercícios

# Exercícios recomendados

- Seção 7.1 - 1 a 3
- Seção 7.2 - 1 a 5
- Seção 7.3 - 1 a 7
- Seção 7.5 - 1 a 5, 9 a 12