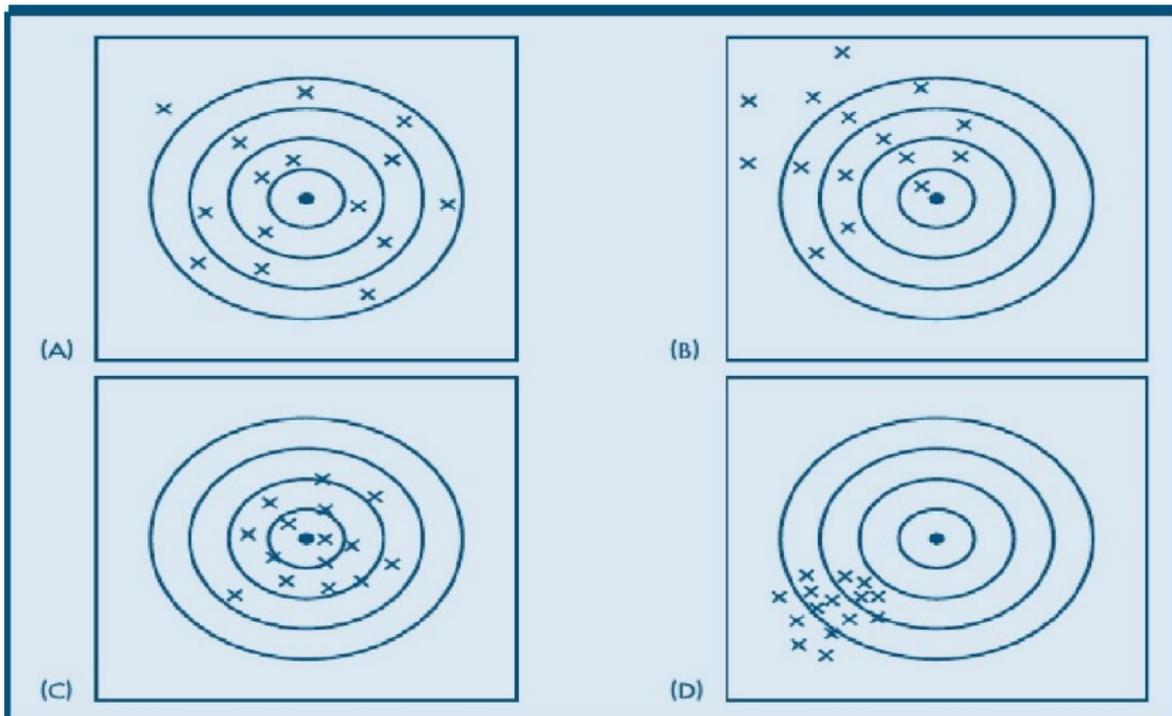
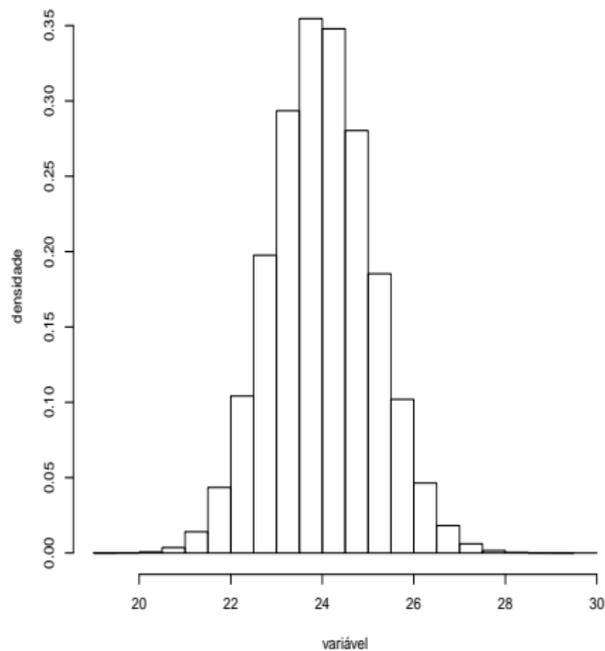


Estimação (de Bussab & Morettin (2010))

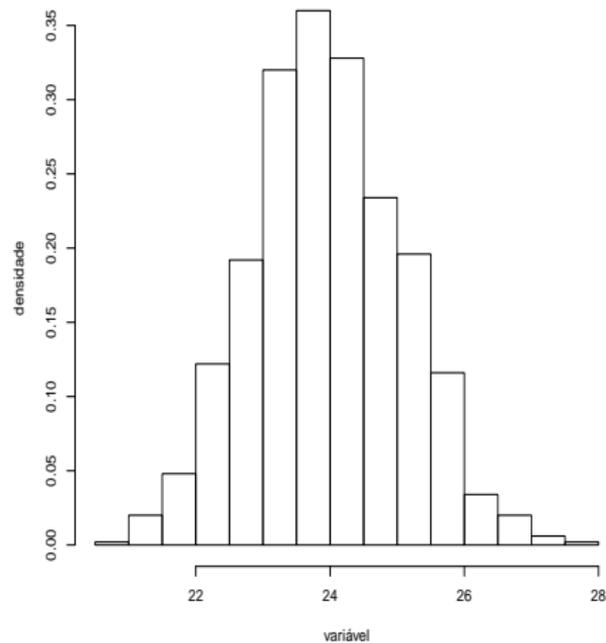


População e amostra

Histograma: população



Histograma: amostra



Estatística

- Uma estatística é uma característica da amostra. Ou seja, se X_1, \dots, X_n é uma amostra, $T = \text{função}(X_1, \dots, X_n)$ é uma estatística.
- Exemplos
 - $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + \dots + X_n}{n}$: a média amostral é uma estatística.
 - $X_{(1)} = \min\{X_1, \dots, X_n\}$.
 - $X_{(n)} = \max\{X_1, \dots, X_n\}$.
 - $X_{(i)}$ é o i -ésimo valor da amostra ordenada.
- Note que uma estatística é uma função que em uma determinada amostra, assume um valor numérico.

Estatística

- Para que serve uma estatística? Para “estimar” os valores de uma distribuição, ou características de uma população.
- População:
 - média $_P$
 - variância $_P$
- Amostra:
 - média $_A = \sum_{i=1}^n \frac{x_i}{n}$ “estima” a média $_P$
 - variância $_A = \sum_{i=1}^n \frac{(x_i - \text{média}_A)^2}{n}$ “estima” a variância $_P$

Estatística

- Exemplo de uso de uma estatística para estimar uma quantidade de interesse: Seja θ a proporção, na cidade de Campinas, de consumidores de um determinado produto.
 - Planejamento do experimento: planeja-se obter uma amostra aleatória simples de tamanho $n = 100$, sem reposição.
 - Cada X_i , $i = 1, \dots, 100$, vai assumir o valor 1 se a pessoa i consome o produto, e 0 se não.
 - Estatística: $T = \frac{X_1 + \dots + X_{100}}{100}$
 - Experimento: uma vez que o experimento foi implementado, T assume um valor t_0 , que estima θ , ou seja, $\hat{\theta} = t_0$

Parâmetros

- Cada quantidade de interesse (como θ no exemplo anterior) é chamado de parâmetro da população.
- Para apresentar uma estimativa de um parâmetro (como $\hat{\theta}$), devemos escolher uma estatística (como T).
- Note que da maneira que o experimento é planejado, a estatística T é uma variável aleatória, uma vez que o experimento pode apresentar resultados diversos, se repetido diversas vezes.
- Portanto, a estatística T (não necessariamente $T = \frac{X_1 + \dots + X_{100}}{100}$) possui uma distribuição, que será a **distribuição amostral de T**.

Aplicação e Exemplos de Distribuição Amostral

- Planejamento: pretendemos avaliar a honestidade de uma moeda. Para isso, planeja-se lançar uma moeda 50 vezes, sendo cada lançamento independente dos demais. Definimos:

$$X_i = \begin{cases} 1, & \text{quando cara} \\ 0, & \text{quando coroa} \end{cases}$$

- Assim, a variável $Y = \sum_{i=1}^{50} X_i$ = número de caras nos 50 lançamentos.
- Y é uma v.a. que pode assumir os valores $\{0, 1, 2, \dots, 50\}$.

Aplicação e Exemplos de Distribuição Amostral

- $Y = \sum_{i=1}^{50} X_i \sim \text{binomial}(50, p)$, pois é uma soma de Bernoullis iid.
- $p = P(X_i = 1) = P(\text{obter cara})$.
- Y é a estatística usada para avaliar a honestidade da moeda.
- Se a moeda for honesta $p = \frac{1}{2}$.
- Definimos por fim $T = \frac{Y}{n}$, e a partir de T , estimamos p .
- Supondo que foram observadas 30 caras nos 50 lançamentos, temos que $T = \frac{30}{50} = 0,6 = \hat{p}$.

Aplicação e Exemplos de Distribuição Amostral

- Qual a importância de saber a distribuição de Y ? Para avaliar se a ocorrência de 30 caras em 50 lançamentos nos traz evidências se a moeda é honesta ou não.
- Assumindo que a moeda é honesta, dado $p = \frac{1}{2}$:

$$P(Y = 30 | n = 50, p = 0.5) = \binom{50}{30} (0,5)^{30} (0,5)^{20} = 0,042.$$

- Então $P(Y \geq 30 | n = 50, p = 0,5) = 0,08$.

Aplicação e Exemplos de Distribuição Amostral

- Assim, se a moeda for honesta, a probabilidade de ocorrer mais de 30 caras em 50 lançamentos é de aproximadamente 0,08.
- Essa probabilidade é evidência suficiente contra a honestidade da moeda?

Distribuição Amostral

- **Resultado:** Seja X uma v.a. com média μ e variância σ^2 . Seja X_1, \dots, X_n uma amostra casual simples de X .
 - $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$
 - $E(\bar{X}_n) = \mu$
 - $Var(\bar{X}_n) = \frac{\sigma^2}{n}$
- Exemplo: X_1, X_2, X_3 ensaios Bernoulli(0.3) independentes
 - $E(X_i) = 0,3 \Rightarrow E(\bar{X}_3) = 0,3$
 - $Var(X_i) = 0,3 \times 0,7 = 0,21 \Rightarrow Var(\bar{X}_3) = \frac{0,21}{3} = 0,07$

Teorema Central do Limite

- **Resultado (T.C.L.):** Para amostras casuais simples X_1, \dots, X_n colhidas de uma população com média μ e variância σ^2 , a distribuição amostral de \bar{X}_n aproxima-se de uma distribuição Normal de média μ e variância $\frac{\sigma^2}{n}$, quando n for suficientemente grande.

Teorema Central do Limite

- Exemplo: X_1, \dots, X_n uma amostra aleatória

- $\exp(2)$: $f_{X_i}(x) = 2e^{-2x}\mathbb{I}_{(x>0)}$

- $E(X_i) = \frac{1}{2}$

- $Var(X_i) = \frac{1}{4}$

Suponha que X_i modela o tempo de vida de um transistor em horas.

Os tempos de 100 transistores são coletados. Desejamos estudar a variável aleatória \bar{X}_{100} , e pelo T.C.L., temos que:

- $E(\bar{X}_{100}) = \frac{1}{2}$

- $Var(\bar{X}_{100}) = \frac{1/4}{100} = \frac{1}{400}$

- $\bar{X}_{100} \sim N\left(\frac{1}{2}, \frac{1}{400}\right)$

Teorema Central do Limite

- **Resultado do T.C.L.:** Se X_1, \dots, X_n é uma amostra casual simples com média μ e variância σ^2 , e definimos $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$, quando n for suficientemente grande:

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

- Retomando o exemplo dos transistores: $\frac{\bar{X}_{100} - (1/2)}{(1/2)/\sqrt{100}} \sim N(0, 1)$

Teorema Central do Limite

- Utilidade do Resultado:

$$\begin{aligned}P(\bar{X}_{100} \leq x) &= P\left(\frac{\bar{X}_{100} - (1/2)}{(1/2)/\sqrt{100}} \leq \frac{x - (1/2)}{(1/2)/\sqrt{100}}\right) \\ &= P(Z \leq 10(2x - 10))\end{aligned}$$

$$\begin{aligned}P(\bar{X}_{100} \geq x) &= 1 - P(\bar{X}_{100} \leq x) \\ &= 1 - P\left(\frac{\bar{X}_{100} - (1/2)}{(1/2)/\sqrt{100}} \leq \frac{x - (1/2)}{(1/2)/\sqrt{100}}\right) \\ &= 1 - P(Z \leq 10(2x - 10))\end{aligned}$$