Estatística Econômica II

Lupércio França Bessegato
Departamento de Estatística/UFJF

Roteiro Geral

1. Introdução
2. Modelos probabilísticos
3. Amostragem
4. **Distribuições amostrais e estimação**
5. Testes de significância
6. Comparações de médias
7. Tópicos especiais
8. Referências

Estatística Econômica II - 2018
2

Distribuições Amostrais e Estimação

Roteiro do Módulo

4. Distribuições amostrais e estimação:
   a) Distribuição da média amostral
   b) Erro padrão
   c) Distribuições amostrais
   d) Estimativas
   e) Distribuição t
   f) Intervalos de confiança

Estatística Econômica II - 2018
4

Distribuição da Média Amostral

Distribuição Amostral

- Estatística de uma amostra aleatória:
  - √ Assume diferentes valores de amostra para amostra da mesma população
  - √ Estatística amostrais são variáveis aleatórias
    - – Variam de acordo a um padrão

Média Amostral

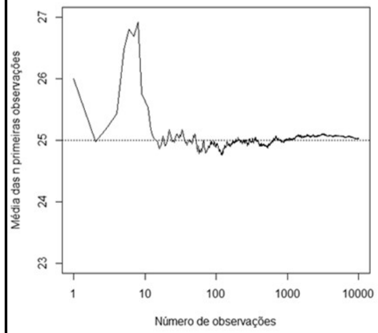
- $\bar{X}$ varia de amostra para amostra
  - √ Apesar de sua variação, essa medida é uma medida razoável da média populacional μ.
  - √ Para amostras cada vez maiores, é garantido que a estatística $\bar{X}$ fica cada vez mais próxima do parâmetro μ.

Lei dos Grandes Números

- Sejam observações aleatórias e independentes de uma mesma população, com uma média $\mu < \infty$.
  - √ À medida que o número de observações aumenta, a média $\bar{X}$ dos valores observados aproxima-se cada vez mais da média μ da população.
  - √ Os resultados médios de muitas observações são estáveis e previsíveis

Exemplo

- Média acumulada de 10.000 observações
  - √ Simulação: Normal($\mu = 25, \sigma = 3$)



- Média das observações torna-se previsível a longo prazo.
- Cassinos podem contar com regularidade a longo prazo
  - √ Ganhos médios da banca em dezenas de milhares de jogadas estarão bastante próximos da média da distribuição dessas jogadas.

9

Estatística Econômica II - 2018

Lei dos Grandes Números

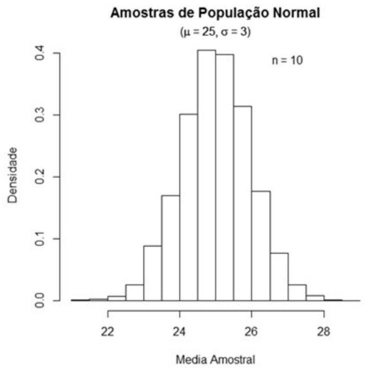
- Exemplo:
  - √ Demanda média
    - Possível predizer, apesar de os clientes tomarem decisões independentes
- Quão grande é um número grande para assegurar que a média $\bar{X}$ dos resultados vá, de fato, ficar próxima da média μ da distribuição?
  - √ Quanto maior a variabilidade dos resultados, serão necessárias mais tentativas.

10

Estatística Econômica II - 2018

Verificação Empírica

- E se o tamanho da amostra não for grande?
- Simulação:
  - √ 10.000 amostras de tamanho 10.



A que distância da média verdadeira (desconhecida) ficaram as médias de amostras de tamanho ?

11

Estatística Econômica II - 2018

Média de $\bar{X}$

$\bar{X}$: média de amostra aleatória de tamanho n

- √ População grande com média μ e desvio-padrão σ.

Qual o padrão de comportamento de $\bar{X}$?

- Média da distribuição amostral de $\bar{X}$:

$$\mu_{\bar{X}} = \mu$$

  - √ Distribuição amostral de $\bar{X}$ é centrada em μ
    (estimador não enviesado do parâmetro μ)

12

Estatística Econômica II - 2018

Desvio-padrão da distribuição amostral de $\bar{X}$:

√ Quão perto o estimador ficará do parâmetro μ na maioria das amostras?

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

√ Médias variam menos que as observações individuais

√ Menor quanto maior for a amostra

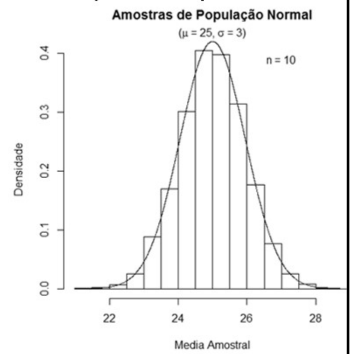
– Resultados de grandes amostras são menos variáveis do que os resultados de pequenas amostras

(mais precisos!)

Distribuição Amostral de $\bar{X}$

- População com distribuição N(μ,σ):

√ Média amostral $\bar{X}$ de n observações independentes tem distribuição $N\left(\mu, \frac{\sigma}{n}\right)$



Exemplo – Simulação

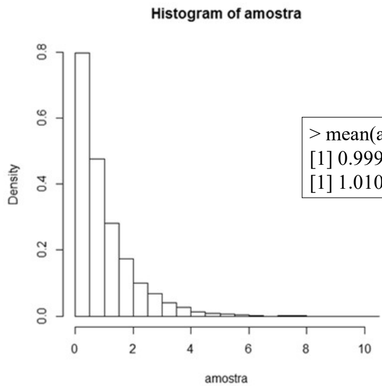
- População exponencial com média 1:

√ $\lambda = 1$

√ Geração de 10.000 valores dessa população

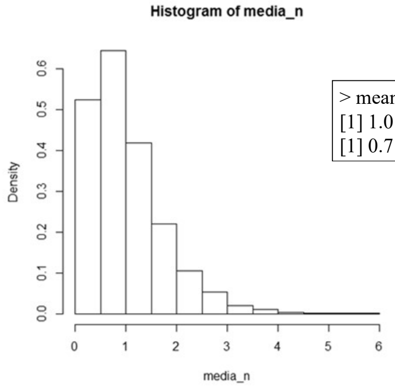
√ Amostra de tamanho 1 (n = 1)

Amostra n =1



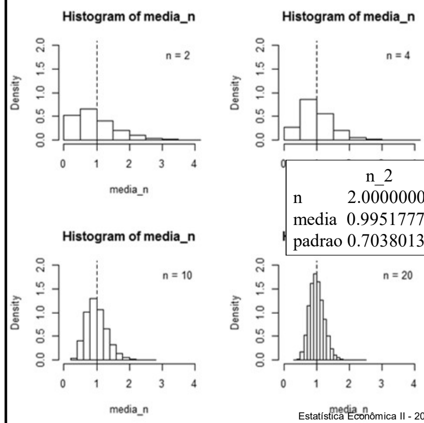

```
> mean(amostra); sd(amostra)
[1] 0.9990838
[1] 1.010478
```

Amostra n = 2



> mean(media_n); sd(media_n)
[1] 1.012711
[1] 0.7129089

Amostras de tamanhos 2, 4, 10 e 20

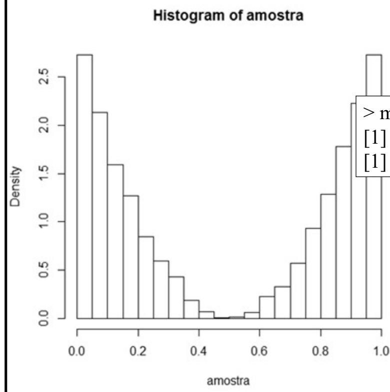


	n_2	n_4	n_10	n_20
n	2.0000000	4.0000000	10.0000000	20.0000000
media	0.9951777	1.0004503	0.9963200	0.9997876
padrao	0.7038013	0.5018562	0.3190797	0.2223573

Exemplo – Simulação

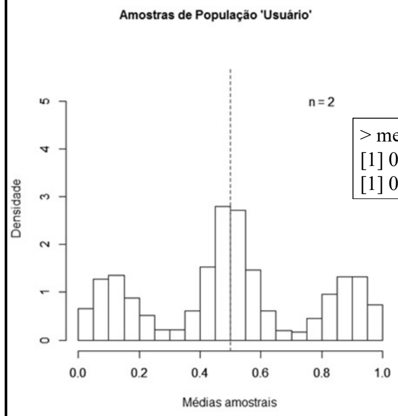
- População com densidade em U:
  - √ $f(x) = 12\,(x - 0{,}5)^2$
  - √ Geração de 10.000 valores dessa população
  - √ Amostra de tamanho 1 (n = 1)

Amostra n =1



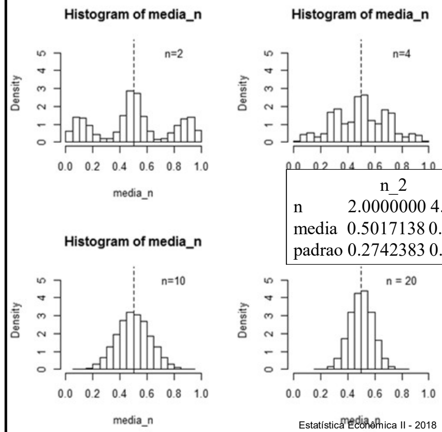
> mean(amostra); sd(amostra)
[1] 0.5061657
[1] 0.3877259

Amostra n = 2




```
> mean(media_n); sd(media_n)
[1] 0,501576
[1] 0.2735612
```


Amostras de tamanhos 2, 4, 10 e 20



	n_2	n_4	n_10	n_20
n	2.0000000	4.0000000	10.0000000	20.00000000
media	0.5017138	0.4986490	0.4994313	0.49993781
padrao	0.2742383	0.1957755	0.1221512	0.08599018

Teorema Central do Limite

- Amostra aleatória de tamanho n de uma população qualquer, com média μ e desvio-padrão σ:
  - √ Se *n* é suficientemente grande, a distribuição da média amostral $\bar{X}$ é aproximadamente normal.

Exemplo

- X: tempo para execução de manutenção preventiva de unidade de ar condicionado
  - √ Distribuído exponencialmente
  - √ Média: $\mu = 1$ hora
  - √ Desvio padrão: $\sigma = 1$ hora
- Na empresa, há 70 dessas unidades.
- Qual a probabilidade de que o tempo médio exceda 50 minutos?

- Aproximação pelo TCL (n = 70)
  - √ Média amostral ($\bar{X}$) tem distribuição aproximadamente normal com média $\mu = 1$ e desvio padrão: $\dfrac{\sigma}{\sqrt{70}} = \dfrac{1}{\sqrt{70}} = 0,1195$ hora
  - √ Cálculo da probabilidade de exceder 50 min:
    - 50 min = $\frac{50}{60} = 0{,}8333$ hora

$$\text{P}\left\{\bar{X} > 0,83\right\} = \text{P}\left\{Z > \frac{0,8333 - 1}{0,1195} = -1,395\right\}$$
$$= 1 - \Phi(-1,40) = 1 - 0,081$$
$$= 0,919$$

25

Estatística Econômica II - 2018

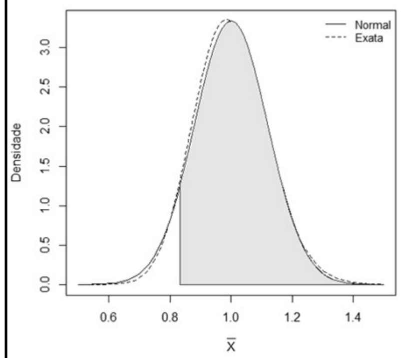
- Mudança de unidade para minutos.
  - √ Média amostral ($\bar{X}$) tem distribuição aproximadamente normal com média $\mu = 1$ e desvio padrão: $\dfrac{\sigma}{\sqrt{70}} = \dfrac{60}{\sqrt{70}} = 7,1714$ hora
  - √ Cálculo da probabilidade de exceder 50 min:

$$\text{P}\left\{\bar{X} > 0,83\right\} = \text{P}\left\{Z > \frac{50 - 60}{7,1714} = -1,394\right\}$$
$$= 1 - \Phi(-1,39)$$
$$= 1 - 0,081 = 0,918$$

26

Estatística Econômica II - 2018

- Distribuição exata e aproximada



- Curva Normal é uma boa aproximação
- Probabilidade
  - √ Aproximada: 0,9222
  - √ Exata: 0,9251
  - √ Diferença: 0,0029

27

Estatística Econômica II - 2018

Comentários

- √ A aproximação normal para a média amostral depende do tamanho da amostra
- √ Com população contínua, unimodal e simétrica, na maioria dos casos, o TCL trabalha bem para pequenas amostras (n = 4, 5).
- √ Em muitos casos de interesse prático, a aproximação normal será satisfatória para n ≥ 30
- √ Se n < 30, o TCL funcionará se a distribuição da população não for muito diferente da normal

28

Estatística Econômica II - 2018

Erro Padrão

Erro Padrão

- Desvio padrão de $\bar{X}$: $\mathrm{dp}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
  - √ Notação: $\mathrm{dp}(\bar{X})$
  - √ Ela não pode ser usada na maioria das vezes (em geral, σ não é conhecido)
- Erro padrão de $\bar{X}$: $\mathrm{ep}(\bar{X}) = \frac{S}{\sqrt{n}}$
  - √ Notação: $\mathrm{ep}(\bar{X})$
  - √ Substitui σ por S (desvio padrão da amostra)

Estatística Econômica II - 2018 30

Erro Padrão da Média Amostral

$$\mathrm{ep}(\bar{x}) = \frac{s}{\sqrt{n}}$$

- Medida de precisão de $\bar{X}$.
- Distribuição amostral para amostras grandes

$$\frac{\bar{X} - \mu}{\mathrm{dp}(\bar{X})} \overset{\text{as.}}{\sim} \mathrm{N}(0,1)$$

Distribuição exata se a população for normal

$$\frac{\bar{X} - \mu}{\mathrm{ep}(\bar{X})} \overset{\text{as.}}{\sim} \mathrm{N}(0,1)$$

Estatística Econômica II - 2018 31

Intervalo de 2 Erros Padrão

- Valores observados na amostra

$$\bar{x} \pm 2\,\mathrm{ep}(\bar{x})$$

  - √ Em uma grande amostra, é muito raro uma média amostral estar distante da média populacional (desconhecida) mais que dois erros padrão
  - √ Em média, isso ocorre em 5% das vezes (para amostras > 20)

Estatística Econômica II - 2018 32

- Consequência importante:
  - √ Suspeitaremos de qualquer valor hipotético para μ (teórico) que esteja afastado do valor de $\bar{x}$ (observado na amostra) mais que 2 ep($\bar{X}$).
  - √ Suspeitaremos ainda mais de um valor hipotético de μ que estiver afastado da nossa estimativa de $\bar{x}$ mais do que 3 ep($\bar{X}$).

33

Estatística Econômica II - 2018

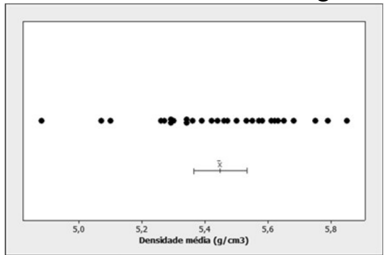
Exemplo

- Densidade média da Terra
  - √ No século XVII Cavendish executou um grande número de experimentos para determinar a densidade da terra
  - √ Amostra:
    - – 29 determinações de Cavendish da densidade média da Terra

34

Estatística Econômica II - 2018

- Determinações de Cavendish
  - √ $\bar{x} = 5{,}447931$ g/cm³
  - √ $s = 0{,}2209457$ g/cm³



$$\text{ep}(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{0,2209457}{\sqrt{29}} = 0,04102858$$

  - √ Intervalo de 2 erros padrão:

$$\bar{x} \pm 2\,\text{ep}(\bar{x}) = 5,447931 \pm 2 \times 0,04102858 = [5,37; 5,53]$$

35

Estatística Econômica II - 2018

- μ = 5,7 é um valor teórico plausível para a densidade média da terra
- Distância da estimativa ao valor teórico, medida em desvios padrão

$$\frac{\bar{x} - \mu}{\text{ep}(\bar{x})} = \frac{5,447931 - 5,7}{0,04102858} = -6,14$$

  - √ Se acreditamos na confiabilidade dos procedimentos experimentais de Cavendish
    - – Parece seguro apostar que valor teórico estaria errado
  - √ Se acreditamos no valor teórico
    - – Parece seguro apostar que os procedimentos experimentais de Cavendish estariam incorretos

36

Estatística Econômica II - 2018

Distribuições Amostrais

Exemplo

- Proporção de adolescentes que testemunharam a venda de droga na escola:
  - √ National Center on Addction and Substance Abuse
  - √ Entrevista por telefone com 1000 adolescentes (12 a 17 anos) da Universidade de Columbia sobre questões relacionadas ao estilo de vida.
  - √ 27% afirmaram ter presenciado venda de drogas ilegais na escola

38

Estatística Econômica II - 2018

- Proporção amostral ($\hat{p}$):
  - √ Na amostra, $\hat{p} = 0{,}27$.
- $\hat{p}$ é uma variável aleatória
  - √ Em outras amostras com 1.000 adolescentes:
    - Incluídas pessoas diferentes nas amostras
    - Seriam obtidas estimativas diferentes da proporção amostral
- A proporção amostral ($\hat{p}$) estima a proporção populacional (p)

39

Estatística Econômica II - 2018

Modelo Probabilístico

- Y: quantidade de indivíduos na amostra com característica de interesse

$$Y \sim \text{binomial}(n, p)$$

n: tamanho da amostra
p: proporção verdadeira

  - √ Assumindo que n < 10% da população

40

Estatística Econômica II - 2018

Exemplo

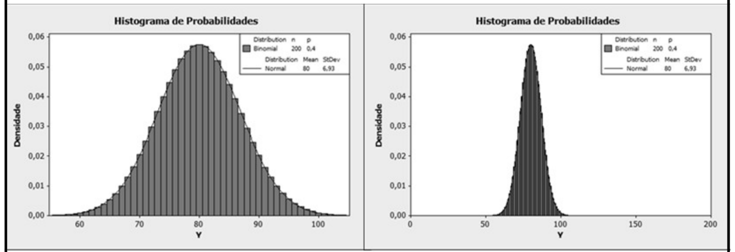
$$Y \sim \text{binomial}(n = 200, p = 0,4)$$

√ Média de Y: $\mu_Y = np = (200)(0,4) = 80$

√ Desvio padrão de Y:

$$\sigma_Y = \sqrt{n\,p\,(1-p)} = \sqrt{(200)(0,4)(0,6)} = 6,93$$

- Histograma de probabilidades:



√ Para grandes valores de n, a distribuição binomial pode ser aproximada pela normal com média np e desvio padrão $\sqrt{np(1-p)}$.

Proporção Amostral

$$\hat{P} = \frac{Y}{n}$$

- Consequência:
  - √ Para grandes amostras, a distribuição de $\hat{P}$ é aproximadamente normal com média np e desvio padrão $\sqrt{\frac{p(1-p)}{n}}$.
- Erro padrão de $\hat{P}$: $\text{ep}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p}}{n}}$
  - √ Podemos usar $\hat{p}$ e seu erro padrão para informar sobre o verdadeiro valor de p.

Exemplo – Continuação

- Adolescentes testemunhas de venda de droga:
  - √ Erro padrão da estimativa

$$\text{ep}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p}}{n}} = \sqrt{\frac{0,27 \times 0,73}{1000}}$$
$$= 0,014039$$

  - √ Intervalo de 2 erros padrão

$$\hat{p} \pm 2\,\text{ep}(\hat{p}) = 0,27 \pm 2 \times 0,014039$$
$$= [0,24; 0,30]$$

  - √ É bastante seguro apostar que o verdadeiro valor de p pertence ao intervalo (entre 24% e 30%)

Pesquisa de Opinião

- Modelo binomial é aproximação adequada para amostragem sem reposição de população real
  - √ Quando tamanho da amostra (n) for menor que 10% do tamanho da população (N)
  - √ Condição satisfeita por quase todas as amostragens reais

- Consequência:
  - √ Tamanho amostral (n) é o que é importante para precisão da proporção amostral
- Importante:
  - √ Amostra deve ser aleatória e todos elementos da população deve ter a mesma chance de ser escolhido
  - √ Em geral, pesquisas de opinião não são AAS
- Tendência em citar uma única medida de erro
  - √ $\text{ep}(\hat{p})$ varia com $\hat{p}$.

Estimativas

Estimativas

- Estimativas pontuais:
  - √ Resultados de estimadores que são um único número
  - √ Ex.: média amostral, proporção amostral
- Estimativas por intervalos:
  - √ Resultados de estimadores que são um intervalo de valores
  - √ Ex.: intervalo de 2 desvios padrão

Propriedades de Bons Estimadores

- Distribuição amostral do estimador ($\hat{\theta}$) seja a mais concentrada possível em torno do valor verdadeiro (e desconhecido) do parâmetro (θ).
  - √ Quase toda vez que for extraída uma amostra, a estimativa resultante ($\hat{\theta}$) estará próxima do verdadeiro valor (θ).

Vício do Estimador

- Distância média entre o centro da distribuição amostral do estimador e o verdadeiro valor do parâmetro.
  - √ $E(\hat{\theta}) = \theta$

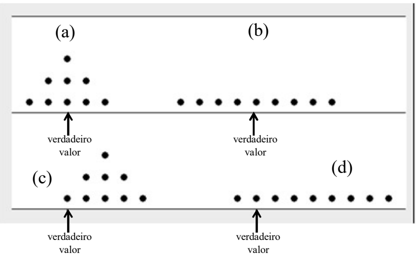
$\mathrm{E}(\bar{X}) = \mu$

$\mathrm{E}(\hat{P}) = p$

Estimadores não viciados

Precisão do Estimador

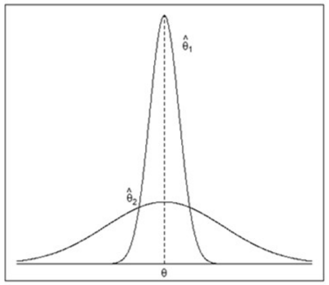
- Quão variável é o estimador em amostragens repetidas?



a) Nenhum vício, alta precisão
b) Nenhum vício, baixa precisão
c) Viciada, alta precisão
d) Viciada, baixa precisão

Acurácia (Eficiência)

- Estimador fornece estimativas próximas do parâmetro estimado, com alta precisão



Erro Padrão

- $\mathrm{ep}(\hat{\theta})$: erro padrão de qualquer estimativa
  - √ Estima a variabilidade dos valores de $\hat{\theta}$ em amostragens repetidas
  - √ Medida de precisão de $\hat{\theta}$.

Estimativas por Intervalo

- Para grande amostras, grande parte dos estimadores têm distribuição amostral

$$\frac{\hat{\theta}-\mu}{\mathrm{dp}(\hat{\theta})} \overset{\text{as.}}{\sim} \mathrm{N}(0,1)$$

- Probabilidades das estimativas:
  - √ Rara a amostra em que estimativa esteja afastada do verdadeiro valor mais que $3\mathrm{ep}(\hat{\theta})$
    - 0,3% das vezes (≈ 1 amostra em 300)

$$\mathrm{P}\left\{|\hat{\theta}-\theta| \leq 2\,\mathrm{ep}\left(\hat{\theta}\right)\right\} \approx 0,954$$

$$\mathrm{P}\left\{|\hat{\theta}-\theta| \leq 3\,\mathrm{ep}\left(\hat{\theta}\right)\right\} \approx 0,997$$

- Consequências:
  - √ É razoavelmente seguro apostar que o verdadeiro valor do parâmetro (θ) está em algum lugar da estimativa ± 2 erros padrão.

$$\hat{\theta} \pm 2\,\mathrm{ep}(\hat{\theta})$$

  - √ Os dados fornecem evidência contra qualquer valor hipotético (teórico) do valor do parâmetro que esteja mais 2 erros padrão da estimativa.
    - Evidência é ainda mais forte se o valor hipotético for afastado da estimativa mais de 3 erros padrão

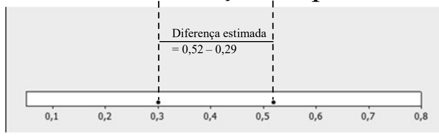
Exemplo

- Percepções acerca do desrespeito da polícia por minorias
  - √ Pesquisa em New Jersey, em 1998 sobre se a polícia abordava de acordo com perfil étnico
  - √ Amostra com 139 pessoas do grupo “negro ou hispânicos” e 378 pessoas do grupo “brancos”
  - √ Resultados amostrais para a resposta SIM:
    - Grupo 1: 52%
    - Grupo 2: 29%
  - √ Qual é a diferença verdadeira entre as proporções entre os dois grupos

57

Estatística Econômica II - 2018

- Proporção daqueles que acreditam que a polícia usou construção de perfil



  - √ Qual a verdadeira diferença (deseja-se estimar $p_1 - p_2$)
  - √ Estimativa pontual da diferença pelos dados

$$\hat{p}_1 - \hat{p}_2 = 0,52 - 0,29 = 0,23$$

  - √ Qual um intervalo de valores prováveis?
    - Qual o erro padrão para a estimativa da diferença?

58

Estatística Econômica II - 2018

Erro Padrão de uma Diferença – Amostras Independentes

- √ Desvio padrão de uma diferença entre variáveis aleatórias independentes:

$$\mathrm{dp}(X_1 - X_2) = \sqrt{\mathrm{dp}(X_1)^2 + \mathrm{dp}(X_2)^2}\ .$$

- √ Erro padrão de uma diferença entre estimativas diferentes

$$\mathrm{ep}(\mathrm{estimativa}_1 - \mathrm{estimativa}_2) = \sqrt{\mathrm{ep}(\mathrm{estimativa}_1)^2 + \mathrm{ep}(\mathrm{estimativa}_2)^2}$$

$$\mathrm{ep}(\hat{\theta}_1 - \hat{\theta}_2) = \sqrt{\mathrm{ep}(\hat{\theta}_1)^2 + \mathrm{ep}(\hat{\theta}_2)^2}\ .$$

Estimativas independentes ⇒ Estimativas não relacionadas de populações diferentes

59

Estatística Econômica II - 2018

Erro Padrão de Diferença de Proporções – Amostras Independentes

- √ $\hat{p}_1$: proporção amostral de amostra de tamanho n_1
- √ $\hat{p}_2$: proporção amostral de amostra de tamanho n_2

- Erro padrão para uma única proporção:

$$\mathrm{ep}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p}}{n}}$$

- Erro padrão de diferença de proporções amostrais independentes

$$\mathrm{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\mathrm{ep}(\hat{p}_1)^2 + \mathrm{ep}(\hat{p}_2)^2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

60

Estatística Econômica II - 2018

Exemplo – Continuação

- Erro padrão da estimativa da diferença:

$$\text{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{0,52 \times 0,48}{139} + \frac{0,29 \times 0,71}{378}} = 0,0483776$$

- Distância estimativa até zero

$$\frac{\hat{p}_1 - \hat{p}_2}{\text{ep}(\hat{p}_1 - \hat{p}_2)} = \frac{0,23}{0,0483776} = 4,7543$$

√ Zero está afastado da estimativa aproximadamente 4, 75 erros padrão

– Parece seguro apostar que as proporções verdadeiras são diferentes

61

Estatística Econômica II - 2018

- Intervalo de dois erros padrão:

$$\begin{aligned}(\hat{p}_1 - \hat{p}_2) \pm 2\,\text{ep}(\hat{p}_1 - \hat{p}_2) &= 0,23 \pm 2 \times 0,0483776 \\ &= [0,13; 0,33]\end{aligned}$$

√ É seguro apostar que a verdadeira diferença de proporção é algo entre 13% e 33%

√ Na verdade, não sabemos onde a diferença se localiza neste intervalo

62

Estatística Econômica II - 2018

Erro Padrão de Diferença de Médias – Amostras Independentes

√ $\bar{x}_1$: média de amostra de tamanho n_1

√ s_1: desvio padrão de amostra de tamanho n_1

√ $\bar{x}_2$: média de amostra de tamanho n_2

√ s_2: desvio padrão de amostra de tamanho n_2

- Erro padrão para uma única média amostral:

$$\text{ep}(\bar{x}) = \frac{s}{\sqrt{n}}$$

63

Estatística Econômica II - 2018

- Erro padrão de estimativa de diferença de médias amostrais independentes

$$\text{ep}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\ .$$

64

Estatística Econômica II - 2018

Exemplo – Continuação

- Preço de produto agrícola
  - √ Pesquisa por amostragem com poucos produtores
  - √ Valores obtidos em amostras aleatórias independentes

Mês	n	$\bar{x}$	s
Setembro	45	\$3,61	\$0,19
Julho	90	\$2,95	\$0,22

  - √ Estimar a diferença das médias nacionais de produção, nos meses de Julho e Setembro

- Estimativa da diferença de médias:

$$\bar{x}_1 - \bar{x}_2 = 3,61 - 2,95 = 0,66\ .$$

- Erro padrão da diferença:

$$\mathrm{ep}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0,19^2}{45} + \frac{0,22^2}{90}}$$
$$= 0,0366\ .$$

- Intervalo de 2 erros padrão:

$$(\bar{x}_1 - \bar{x}_2) \pm 2\,\mathrm{ep}(\bar{x}_1 - \bar{x}_2) = 0,66 \pm 2 \times 0,0366$$
$$= [0,59; 0,73]\ .$$

  - √ É seguro apostar que a verdadeira diferença dos preços médios nacionais é algo entre \$0,59 e \$0,73.

Interpretação de Intervalo para a Diferença

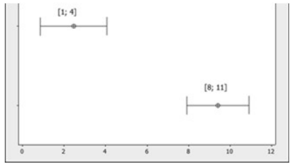
[a, b]: intervalo para $\theta_1 - \theta_2$ (verdadeira):

- √ a e b > 0:
  - Seguro apostar que $\theta_1 > \theta_2$.
- √ a e b < 0:
  - Seguro apostar que $\theta_1 < \theta_2$.
- √ a e b tem sinais opostos:
  - Inclui a possibilidade de que $\theta_1 = \theta_2$ (ou seja $\theta_1 - \theta_2 = 0$).

- Exemplo:
  - √ $-0,1 \le \theta_1 - \theta_2 \le 6$
    - θ_1 pode ser maior que θ_2 por até 6 unidades
    - Pouco provável que seja apreciavelmente menor

Intervalos Individuais e Diferenças

- Intervalos individuais não se sobrepõem



√ Podemos concluir que $\hat{\theta}_1$ e $\hat{\theta}_2$ estão afastados entre si no mínimo a semiamplitude do intervalo (2 erros padrão)

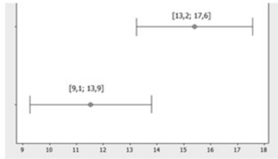
√ Podemos apostar com bastante segurança que os verdadeiros valores de θ_1 e θ_2 são diferentes.

- Intervalos individuais se sobrepõem

$$\hat{\theta}_1 = 15,4 \qquad \hat{\theta}_2 = 11,5$$

$$\text{ep}(\hat{\theta}_1) = 1,1 \qquad \text{ep}(\hat{\theta}_2) = 1,2$$

$$13,2 \leq \theta_1 \leq 17,6\ . \quad 9,1 \leq \theta_1 \leq 13,9\ .$$



- θ_1 e θ_2 são iguais?
- Intervalo da diferença

$$\hat{\theta}_1 - \hat{\theta}_2 = 3,9$$

$$\text{ep}(\hat{\theta}_1 - \hat{\theta}_2) = \sqrt{1,1^2 + 1,2^2} = 1,63 \qquad 0 \notin [0,64; 6,36]\ .$$

$$0,64 \leq \theta_1 - \theta_2 \leq 6,36\ .$$

√ Podemos apostar que os verdadeiros valores de θ_1 e θ_2 são diferentes!

- Distância entre $\hat{\theta}_1$ e $\hat{\theta}_2$:

$$\hat{\theta}_1 - \hat{\theta}_2 = 3,9$$

$$\text{ep}(\hat{\theta}_1 - \hat{\theta}_2) = \sqrt{1,1^2 + 1,2^2} = 1,63$$

$$0,64 \leq \theta_1 - \theta_2 \leq 6,36\ .$$

√ $\hat{\theta}_1$ e $\hat{\theta}_2$ estão afastados entre si 2,39 erros padrão da diferença

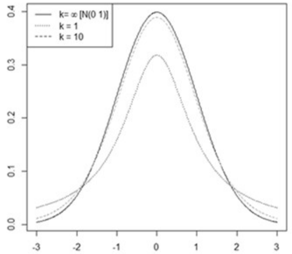
- Importante:

√ Variabilidade de uma diferença é menor do que a soma da variabilidade de suas componentes

$$\text{dp}(X_1 - X_2) = \text{dp}(X_1 + X_2) < \text{dp}(X_1) + \text{dp}(X_2)\ .$$

Distribuição t

Distribuição t de Student



√ Parâmetro:
- Graus de liberdade (k)

√ Aumentando os graus de liberdade, a curva da densidade t se aproxima da normal

√ Diminuindo os graus de liberdade, as caudas da curva de densidade ficam mais grossas (pesadas).

Para dados selecionados em população normal:

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

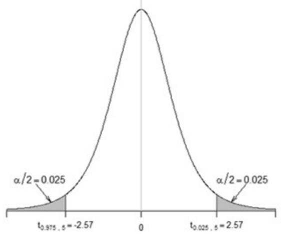
$$T = \frac{\bar{X} - \mu}{\mathrm{ep}(\bar{X})} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim \mathrm{t}_{n-1} \ .$$

Mede a separação entre X e μ, em erros padrão

- S é uma nova fonte de variabilidade
  - √ Variabilidade adicional é refletida nas caudas mais pesadas que as da normal

- Intervalos (0 ± 2):
  - √ $\mathrm{P}\{-2 \ \leq Z \ \leq 2\} = 0{,}9545.$
  - √ $\mathrm{P}\{-2 \ \leq T_{29} \leq 2\} = 0{,}9450.$
  - √ $\mathrm{P}\{-2 \ \leq T_{7} \leq 2\} = 0{,}9144.$
  - √ $\mathrm{P}\{-2 \ \leq T_{3} \leq 2\} = 0{,}8607.$
- Para tamanhos amostrais pequenos distribuição t pode ter variabilidade bem maior que a normal!

- Intervalos simétricos de 95%:
  - √ $\mathrm{P}\{-1{,}96 \ \leq Z \ \leq 1{,}96\} = 0{,}95.$
  - √ $\mathrm{P}\{-2{,}05 \ \leq T_{29} \leq 2{,}05\} = 0{,}95.$
  - √ $\mathrm{P}\{-2{,}36 \ \leq T_{7} \leq 2{,}36\} = 0{,}95.$
  - √ $\mathrm{P}\{-3{,}18 \ \leq T_{3} \leq 3{,}18\} = 0{,}95.$



$\mathrm{P}\{-2{,}36 \ \leq T_{7} \leq 2{,}36\} = 0{,}95.$

Comandos Excel

- Distribuição normal padrão:
  - √ Função de distribuição acumulada ($P\{Z \leq q\} = x$):
    - `DISTNORMP(q)`
  - √ Quantil ($P\{Z \leq x\} = p$):
    - `INVNORMALP(p)`
- Distribuição t com gl graus de liberdade:
  - √ Função de distribuição acumulada
    - `DIST.T(q; g_liberdade; VERDADEIRO)`
  - √ Quantil:
    - `INV.T(p; g_liberdade)`

77

Estatística Econômica II - 2018

- Intervalos simétricos de 99%?
  - √ $P\{-2{,}58 \leq Z \leq 2{,}58\} = 0{,}99$.
  - √ $P\{-2{,}76 \leq T_{29} \leq 2{,}76\} = 0{,}99$.
  - √ $P\{-3{,}50 \leq T_7 \leq 3{,}50\} = 0{,}99$
  - √ $P\{-5{,}84 \leq T_3 \leq 5{,}84\} = 0{,}99$.

78

Estatística Econômica II - 2018

- Intervalos simétricos de 90%:
  - √ $P\{-1{,}64 \leq Z \leq 1{,}64\} = 0{,}90$.
  - √ $P\{-1{,}70 \leq T_{29} \leq 1{,}70\} = 0{,}90$.
  - √ $P\{-1{,}89 \leq T_7 \leq 1{,}89\} = 0{,}90$.
  - √ $P\{-2{,}35 \leq T_3 \leq 2{,}35\} = 0{,}90$.

79

Estatística Econômica II - 2018

Exercício

- Determine:
  - √ $P\{T_8 \geq 1{,}06\} = x$.
  - √ $P\{T_{15} \geq 0{,}9\} = x$.
  - √ $P\{T_{20} \leq 2{,}7\} = x$.
  - √ $P\{T_{40} \leq 5{,}4\} = x$.
  - √ $P\{T_6 \geq x\} = 0{,}001$.
  - √ $P\{T_{13} \leq x\} = 0{,}2$.

81

Estatística Econômica II - 2018

Usos da Distribuição t

- Aplicação da estatística T se aplica apenas a dados amostrados de uma população normal.
  - √ Para grandes amostras a distribuição t é aproximadamente normal

- Para amostras pequenas, é necessário um número ligeiramente maior de erros padrão de ambos os lados de $\bar{X}$ para preservar cobertura do intervalo
- Intervalos para a proporção:
  - √ São necessárias outras abordagens para lidar com dados de pequenas amostras

Suposição de Normalidade

- Normalidade é suposição razoável:
  - √ Muitas populações encontradas na prática são bem aproximadas pela normal
  - √ Desvios moderados de normalidade têm pequeno efeito nas conclusões da inferência
- Se normalidade não é suposição razoável:
  - √ Alternativa
    - – Usar procedimentos não-paramétricos para construir IC
    - – Testes de hipóteses não-paramétricos

Intervalos de Confiança

Intervalos de Confiança

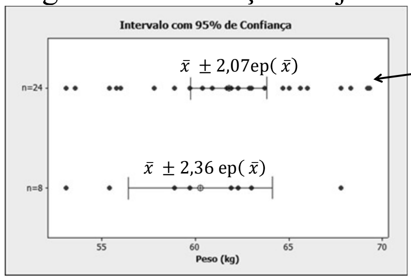
- Intervalo em que o comprimento dependerá do grau de confiança desejado:
  - √ Intervalo de dois erros padrão para a média:
    - – Cobertura de 95,4%
    - (em geral, apenas para grandes amostras)

Exemplo

- Roupas para atletas
  - √ Venda de roupas e equipamentos pela internet
  - √ Características físicas de diferentes tipos de clientes do sexo masculino
  - √ Amostra aleatória de tamanho 24 corredores
  - √ Dados:

67,8	61,9	63	53,1	62,3	59,7	55,4	58,9	60,9	69,2	63,7	68,3
64,7	65,6	56	57,8	66	62,9	53,6	65	55,8	60,4	69,3	61,7

- Comprimento do intervalo depende do grau de confiança desejado



Mais dados, mais precisão

IC de 2 ep (grandes amostras) cobertura de 95,4%

	$\bar{x}$	s_X	ep($\bar{x}$)	t	IC 95%	IC 2 ep	Cobertura 2 ep
n = 8	60,26	4,602	1,627	2,362	[56,41; 64,11]	[56,41; 64,11]	91,4%
n = 24	61,79	4,808	0,981	2,069	[59,76; 63,82]	[59,76; 63,82]	94,3%

Intervalo com 95% de Confiança para um Parâmetro

- Contém o verdadeiro valor do parâmetro para 95% das amostras extraídas:

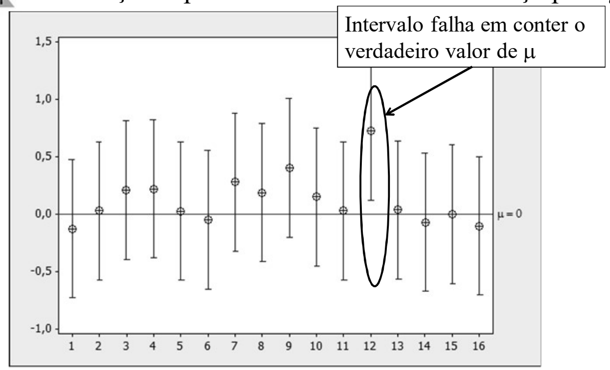
$$\hat{\theta} \pm t_{gl}\,\mathrm{ep}(\hat{\theta})\ .$$

  - √ gl: graus de liberdade da t.
  - √ t_{gl}: multiplicador para um IC de 95%
- Intervalo de confiança para o verdadeiro valor do parâmetro:
  - √ `estimativa ± (t)(erros padrão)`
  - √ `estimativa ± margem de erro`

• Valor do multiplicador t para um IC de 95%

gl	7	8	9	10	11	12	13	14	15	16	17
t	2,365	2,306	2,262	2,228	2,201	2,179	2,160	2,145	2,131	2,120	2,110
gl	18	19	20	25	30	35	40	45	50	60	∞
t	2,101	2,093	2,086	2,060	2,042	2,030	2,021	2,014	2,009	2,000	1,960

√ Construção repetida de intervalo de confiança para μ



√ Confiança no intervalo é por ele de ter sido produzido por método que funciona 95% das vezes

Exemplo

- Continuação do exemplo primeiro
  - √ IC de 95% para amostra completa
  - √ Nossa confiança no intervalo [59,76; 63,82] provém do fato de ter sido produzido por um método que funciona 95% das vezes

Cenários de Operação com Intervalos de Confiança

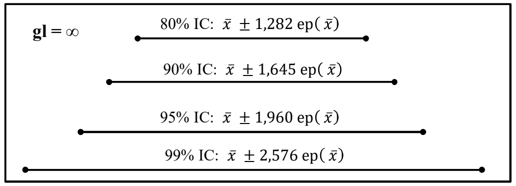
- Amostragem repetida:
  - √ Supondo a longo prazo:
    - Extração de amostras repetidamente
    - Construção de intervalo a partir de cada amostra
    - Em média, 95% das amostras fornecerão intervalos que conterão verdadeiro valor do parâmetro

- Muitos estudos
  - √ Grande número de pesquisadores executa independentemente estudos:
    - 95% dos pesquisadores englobarão o verdadeiro valor do parâmetro em seus intervalos
    - 5% deles deixarão de fazê-lo

- Planejamento de um estudo:
  - √ Planeja-se fazer um estudo (futuro) em que será construído um intervalo a partir de amostra aleatória que será extraída
    - Há uma probabilidade de 95% de este intervalo (que será construído) englobar o verdadeiro valor do parâmetro
    - Há uma chance de 5% de não conseguir fazê-lo

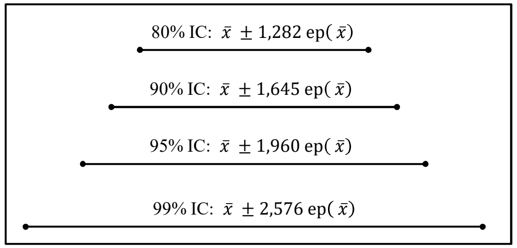
Ajustamento do Nível de Confiança

Taxa de sucesso em estimativas por intervalo	Coeficiente de confiança	Nível de confiança	Frequência de cobertura



√ Quantidade de erros padrão para obter intervalo com nível de confiança desejado

Nível de confiança	80%	90%	95%	99%
gl (∞)	1,282	1,645	1,960	2,576



- √ Quanto maior o nível de confiança, maior o comprimento do intervalo
  - A afirmação será menos precisa

- Se queremos fazer uma afirmação na qual confiamos mais

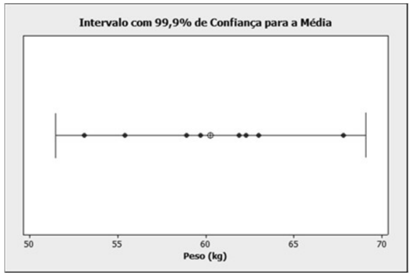
  (ou seja, para corrermos um risco menor de estar errado)
  - Precisamos fazer uma afirmação menos precisa
- Informações muito imprecisas são pouco úteis

Exemplo - Continuação

- Intervalo com 99,9% de confiança para a amostra com n = 8:
  - √ Média amostral:
    - $\bar{x} = 60{,}26$
  - √ Erro padrão da média amostral:
    - $\text{ep}(\bar{x}) = 1{,}627$
  - √ Multiplicador:
    - $t = 5{,}408$
  - √ Intervalo:

$$\bar{x} \pm t_{gl} \times \text{ep}(\bar{x})$$
$$60,26 \pm 5,408 \times 1,627$$
$$[51,46; 69,06] \; .$$

- Gráfico do IC com 99,9% de confiança:



  - √ Limite inferior é menor do qualquer das medidas observadas!

- Como podemos fazer uma afirmação precisa com confiança?
  - √ Erros padrão de algumas estimativas

$$\text{ep}(\bar{x}) = \frac{s_X}{\sqrt{n}} \; . \quad \text{ep}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \; .$$

  - √ Erros padrão aproximadamente proporcionais a $\frac{1}{\sqrt{n}}$
    - Erro padrão diminui quando n aumenta
    - São necessárias 4 vezes mais observações para duplicar a precisão (IC com metade do comprimento)

Intervalo de Confiança para a Média Populacional

- Intervalo de confiança para a média verdadeira da população (μ):
  - √ média amostral ± (t)(erros padrão)

$$\bar{x} \pm t_{gl} \times \text{ep}(\bar{x}) \ . \quad \text{ep}(\bar{x}) = \frac{s_X}{\sqrt{n}} \ . \qquad \text{gl} = n - 1 \ .$$

- Hipóteses para construção destes IC's:
  i. Amostra aleatória de população com média μ (observações independentes)
     - Esta hipótese é crucial
  ii. População com distribuição normal
     - Esta hipótese é menos importante
     - Intervalos funcionam bem na prática, exceto quando a população apresenta severos desvios da normal

- Comportamento dos dados que fariam duvidar da validade da construção desses ICs:
  - √ Presença de *outliers*
    - Valores atípicos
  - √ Presença de *clusters*

Exemplo

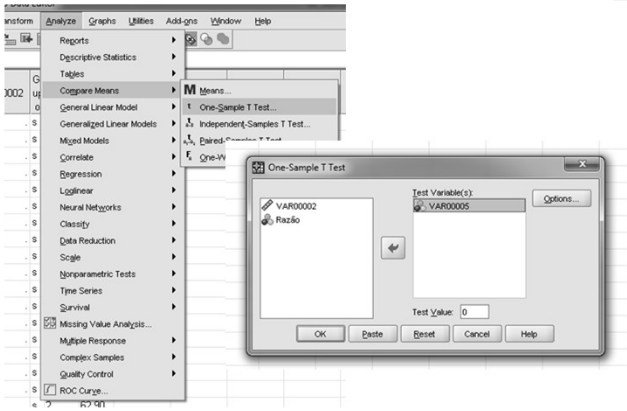
- Roupas para atletas
  - √ Venda de roupas e equipamentos pela internet
  - √ Amostra aleatória de tamanho 24 corredores
  - √ Dados:

n	$\bar{x}$	s_X	t_{23}
24	61,79	4,808	2,069

  - √ Intervalo com 95% de confiança para a média:

$$\bar{x} \pm t_{gl} \times \frac{s_X}{\sqrt{n}}$$
$$61,79 \pm 2,069 \times \frac{4,808}{\sqrt{n}}$$
$$61,79 \pm 2,030 = [59,76; 63,82] \ .$$

- Comandos – SPSS:



- Saída – SPSS:
  - √ Estatísticas amostrais:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
VAR00005	24	61,7917	4,80778	,98138

  - √ Intervalo com 95% de confiança

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
VAR00005	62,964	23	,000	61,7916	59,7615	63,8218

Intervalo de Confiança para a Proporção Populacional

- Intervalo de confiança para a proporção verdadeira da população (p):
  - √ `proporção amostral ± (z)(erros padrão)`

$$\hat{p} \pm z \times \mathrm{ep}(\hat{p})\ . \qquad \mathrm{ep}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ .$$

  - √ Aplicado quando *n* é grande
  - √ Valores de z:

Cobertura do IC	t
95%	1,960
90%	1,645
99%	2,576

Tamanho da Amostra

- Tamanho da amostra para que os intervalos funcionem adequadamente depende do valor da proporção
- Tamanho amostral mínimo vs. valor de $\hat{p}$:
  - √ Valor mínimo exigido para assegurar qye a soma dos erros absolutos nas extremidades constitua menos do que 10% do comprimento do intervalo

Valor de $\hat{p}$	0,05	0,1	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
n mínimo	960	400	220	125	76	47	23	13	11	10
Valor de $\hat{p}$	0,95	0,90	0,85	0,80	0,75	0,70	0,65	0,60	0,55	0,50

Samuels, M. L.; Lu, T.-F. C. Sample size requirements for the back-of-the-Envelope binomial confidence interval. *The American Statistician*, v. 46, n.3, p. 228-231, 1992.

Exemplo

- Relacionamentos no ambiente de trabalho
  - √ Pesquisa feita com 200 CEOs de empresas americanas (*Fortune*, 1994)
  - √ Pergunta:
    - "Você aprova ou desaprova romances no ambiente de trabalho entre funcionários não-casados? Ou você diria que a empresa não deveria interferir nessa questão?"
  - √ Resultado:
    - 70% responderam que a empresa não deveria interferir
  - √ Suposição:
    - Amostra aleatória das principais empresas

111

Estatística Econômica II - 2018

- Tamanho amostral é adequado para construção de IC aproximado?

Valor de $\hat{p}$	0,05	0,1	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
n mínimo	960	400	220	125	76	47	23	13	11	10
Valor de $\hat{p}$	0,95	0,90	0,85	0,80	0,75	0,70	0,65	0,60	0,55	0,50

  - √ Tamanho da amostra (n = 200) é maior que o mínimo permitido
    - Pode ser usada IC aproximado pela normal
- Se amostra é menor que o mínimo prescrito
  - √ Usar intervalo exato (Clopper-Pearson)
  - √ Métodos aproximados mais completos (Bohning, Newcomb)

112

Estatística Econômica II - 2018

- Intervalo com 95% de confiança para a verdadeira proporção dos executivos que acreditam que a empresa não deve interferir
- Estimativa da pesquisa:
  - $\hat{p} = 0{,}7$

$$\hat{p} \pm z \text{ ep}(\hat{p}) = \hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$= 0,7 \pm 1,96 \times \sqrt{\frac{0,7 \times 0,3}{200}}$$
$$= [0,64; 0,77] \ .$$

113

Estatística Econômica II - 2018

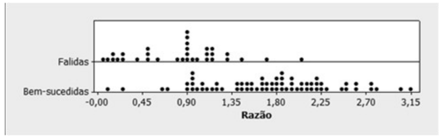
Exemplo

- Empresas bem-sucedidas vs. empresas falidas
  - √ De que maneira empresas que vão à falência diferem daquelas que continuam a operar?
- Estudo comparativo de diversas características:
  - √ Grupo bem-sucedidas:
    - 68 empresas
  - √ Grupo falidas:
    - 33 empresas
- Variável de interesse:
  - √ Razão entre patrimônio e dívidas

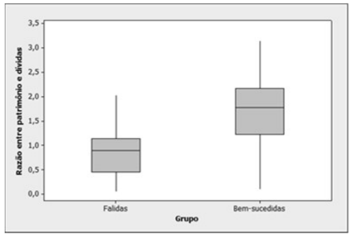
114

Estatística Econômica II - 2018

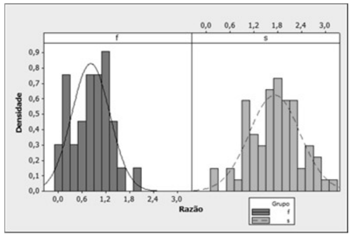
- Análise gráfica:
  - √ Gráfico de pontos:



  - √ Box-plot:



√ Histograma:



√ Empresas bem-sucedidas parecem se aproximar mais da normalidade

√ Parece haver dois subgrupos entre as empresas que faliram

√ Não existem outliers ou outros fortes desvios da normalidade

- Resumo das estatísticas:

	n	$\bar{x}$	s_X	$ep(\bar{x})$
Bem-sucedidas	68	1,7256	0,6393	1,627
Falidas	33	0,8236	0,4811	0,981

- Há diferença entre as razões médias patrimônio/dívidas dos dois grupos?

Comparação de Duas Médias

- Médias desconhecidas:
  - √ μ_1: média da 1ª população (grupo).
  - √ μ_2: média da 2ª população (grupo).
- Parâmetro:
  - √ Diferença entre as médias populacionais: $\mu_1 - \mu_2$
- Estimativa:
  - √ Diferença entre as médias amostrais: $\bar{x}_1 - \bar{x}_2$

- Intervalo de confiança para diferença entre médias populacionais ($\mu_1 - \mu_2$)

diferença entre médias amostrais $\pm\, t$ erros padrão da diferença .

$$\bar{x}_1 - \bar{x}_2 \pm t \times \text{ep}(\bar{x}_1 - \bar{x}_2).$$

√ Duas amostras independentes:

$$\text{ep}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

√ Para cálculo manual:

$$\text{gl} = \min\{(n_1 - 1), (n_2 - 1)\}.$$

– Aproximação conservadora

119

Estatística Econômica II - 2018

Exemplo – Continuação

- Qual a diferença da razão média patrimônio/dívidas entre as firmas bem-sucedidas e aquelas que faliram?

√ Resumo das estatísticas:

	n	$\bar{x}$	s_X
Bem-sucedidas	68	1,7256	0,6393
Falidas	33	0,8236	0,4811

√ Graus de liberdade: min(68 – 1, 33 – 1) = 32

√ $t_{32}(0{,}025) = 2{,}037$ (bilateral)

120

Estatística Econômica II - 2018

- Intervalo com 95% de confiança para a diferença de razões médias

$$\bar{x}_1 - \bar{x}_2 \pm t_{\text{gl}}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(1,7256 - 0,8236) \pm t_{32}\sqrt{\frac{(0,6393)^2}{68} + \frac{(0,4811)^2}{33}}$$

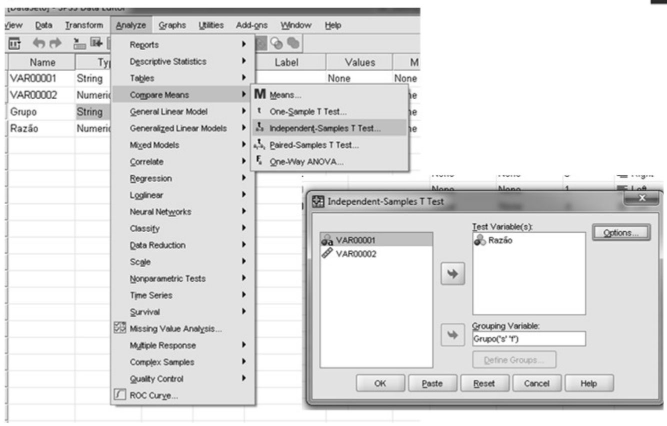
$$0,902 \pm 2,037 \times 0,114 = [0,670; 1,134].$$

√ Com 95% de confiança, a verdadeira razão patrimônio/dívidas das empresas bem-sucedidas excede aquela para as empresas falidas entre 0,67 e 1,13.

121

Estatística Econômica II - 2018

- Comandos – SPSS:



122

Estatística Econômica II - 2018

- Saída – SPSS:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Razão	Equal variances assumed	2,721	,102	7,172	99	,000	,902	,126	,652	1,151
	Equal variances not assumed			7,904	81,692	,000	,902	,114	,675	1,129

√ Construção manual de IC é mais conservativa
- Intervalo ligeiramente maior: [0,670; 1,134]

Comparação de Duas Proporções

- Três situações amostrais:
  - √ Duas proporções de amostras independentes
  - √ Comparação de proporções de mesma amostra, com várias categorias de resposta
  - √ Comparação de muitos itens binomiais em mesma amostra

Exemplo – Amostras Independentes

- Relação entre o fumo e desempenho escolar
  - √ Pesquisa com adolescentes (1998)
  - √ 1.000 adolescentes com idade entre 12 e 17 anos entrevistados por telefone
  - √ 68% dos 870 adolescentes que não fumavam obtiveram bons conceitos (A's e B's)
  - √ 41% dos 130 dos adolescentes fumantes obtiveram bons conceitos
  - √ Há diferença de desempenho escolar entre os dois grupos de adolescentes?

Comparação de Duas Proporções

- Proporções desconhecidas:
  - √ p_1: proporção de interesse da 1ª população.
  - √ p_2: proporção de interesse da 2ª população.
- Parâmetro:
  - √ Diferença entre as proporções populacionais: $p_1 - p_2$
- Estimativa:
  - √ Diferença entre as proporções amostrais: $\hat{p}_1 - \hat{p}_2$

- Intervalo de confiança para diferença entre proporções populacionais ($p_1 - p_2$)

diferença entre proporções amostrais $\pm\ z$ erros padrão da diferença .

$$\hat{p}_1 - \hat{p}_2 \pm z \times \text{ep}(\hat{p}_1 - \hat{p}_2).$$

√ Baseados em aproximações normais para amostras grandes.
  - n_i maior que o mínimo para o p_i correspondente

√ Intervalo com 95% de confiança:
  - z = 1,96

√ Erro padrão para amostras independentes:

$$\text{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

127

Estatística Econômica II - 2018

Exemplo – Continuação

- Qual a diferença da proporção de alunos com bons conceitos entre alunos fumamtes e não fumantes?

√ Resumo das estatísticas:

	n	$\hat{p}$
Não fumantes	870	0,68
Fumantes	130	0,41

√ Erro padrão para a diferença de proporções:

$$\text{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{0,68 \times 0,32}{870} + \frac{0,41 \times 0,59}{130}} = 0,0459.$$

128

Estatística Econômica II - 2018

- Intervalo com 95% de confiança para a diferença de proporções populacionais:

$$\hat{p}_1 - \hat{p}_2 \pm z \times \text{ep}(\hat{p}_1 - \hat{p}_2)$$
$$(0,68 - 0,41) \pm 1,96 \times 0,0459$$
$$0,27 \pm 0,090 = [0,18; 0,36].$$

√ Não fumantes têm uma probabilidade maior de obter bons conceitos do que os fumantes.

- Tamanhos amostrais mínimos exigidos:

√ Grupo fumantes: 47

√ Grupo não-fumantes: 13

- Intervalo de confiança exato para a diferença:

√ [0,181632; 0,361604]

129

Estatística Econômica II - 2018

- Importante:

√ Não se pode concluir que o ato de fumar seja a causa de fumantes obterem conceitos inferiores
  - É um estudo observacional

- O que poderia explicar o fato de pessoa que fuma (na pesquisa) é a que mais provavelmente obtém conceito ruim?
- Diferenças bem pequenas na formulação das questões da pesquisa podem afetar os resultados obtidos!

130

Estatística Econômica II - 2018

Comparação de Proporções

- Situação 1 – Amostras independentes:
  - √ Erro padrão de $p_1 - p_2$:

$$\mathrm{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

  - √ Aplicável apenas quando as duas proporções amostrais provêm de 2 amostras independentes.
  - √ Exemplo anterior:
    - São duas amostras separadas de indivíduos: Fumantes e não-fumantes
    - Resposta binomial (S/N)
    - Comparação da proporção de interesse em cada grupo

132

Estatística Econômica II - 2018

Exemplo – Eleição

- Eleição nos EUA, em 1996

Estado	n	Pesquisas de opinião pré-eleitorais				Resultados da eleição		
		Clinton	Dole	Perot	Outros/ Indecisos	Clinton	Dole	Perot
Nova Jérsei	1.000	51	33	8	8	53	36	9
Nova Iorque	1.000	59	25	7	9	59	31	8
Connecticut	1.000	51	29	11	9	52	35	10

  - √ Situação de independência:
    - Comparar apoio ao candidato Clinton em dois estados

133

Estatística Econômica II - 2018

Comparações de Proporções

- Situação 2:
  - √ Em uma única amostra, indivíduo é enquadrado em uma categoria
- Exemplo:
- Comparação para saber se um candidato está à frente de outro

134

Estatística Econômica II - 2018

Exemplo – Eleições

Estado	n	Pesquisas de opinião pré-eleitorais				Resultados da eleição		
		Clinton	Dole	Perot	Outros/ Indecisos	Clinton	Dole	Perot
Nova Jérsei	1.000	51	33	8	8	53	36	9

  - √ Proporções são relativas ao mesmo conjunto de pessoas
  - √ Não são independentes
  - √ É incorreto usar a fórmula do erro padrão para o caso de independência
    - Fornece resultados muito pequenos

135

Estatística Econômica II - 2018

Exemplo – Propaganda

- Por que estações de rádio veiculam propagandas de serviços públicos?

Razão	%
A mensagem é relevante para a estação ou necessidade dos ouvintes	52
Organização sem fins lucrativos é amplamente conhecida ou representa uma causa válida	25
Tamanho do anúncio se encaixa nos intervalos disponíveis	15
Porta-voz de celebridade oferece credibilidade	5
Qualidade de produção, chance, formato correto	3

√ Comparação entre qualquer par de porcentagens não é independente

136

Estatística Econômica II - 2018

Comparações de Proporções

- Situação 3:
  - √ Proporções que provêm da mesma amostra, mas para questões diferentes
  - √ Não são independentes

137

Estatística Econômica II - 2018

Exemplo – Sistema de Saúde

- Reação da população ao Sistema de Saúde

Entrada na tabela é a % que concorda	Austrália	Canadá	N.Z.	R. Unido	EUA
Dificuldade na obtenção de cuidados necessários	15	20	18	15	28
Mudanças recentes prejudicarão qualidade	28	46	38	12	18
Sistema deveria ser reestruturado	30	23	32	14	33
Nº de contas não cobertas pelo seguro	7	27	12	44	8
Tamanho da amostra	1.000	1.000	1.000	1.000	1.000
Gastos com cuidados médicos (US$/pessoa)[a]	1.805	2.095	1.352	1.347	4.090

[a]Adaptado para diferentes custos de vida

- √ Situação 1: Independente
  - Comparação % da mesma afirmação em 2 países diferentes
- √ Situação 3:
  - Comparação % de 2 afirmações diferentes no mesmo país.

138

Estatística Econômica II - 2018

Exemplo – Fumo/Álcool

- Características por condição

Entrada na tabela é a % que diz sim	Fumante[a]	Não-fumante	Que bebe[b]	Que não bebe
Obtém geralmente A ou B?	41	68		
Lê uma ou mais horas por dia?	54	72	56	75
Embriaga-se ao menos uma vez por mês?	63	10		
Já fumou maconha?	79	14	52	12
Tamanho da amostra	130	870	260	740

[a]Fumou cigarros nos últimos 30 dias

[b]Consumiu mais do que alguns goles de álcool nos últimos 30 dias

- √ Situação 1: Independente
  - Comparação entre fumantes e não-fumantes no mesmo item
- √ Situação 3:
  - Comparação % de fumantes que se embriagam (63%) com os que já usaram maconha (79%).

139

Estatística Econômica II - 2018

Erros Padrão – Diferença de Proporções

1. Proporções de amostras independentes:

$$\text{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

2. Várias categorias de resposta, em amostra única

$$\text{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1 + \hat{p}_2 - (\hat{p}_1 - \hat{p}_2)^2}{n}}.$$

3. Muitos itens sim/não, em amostra única

$$\text{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\min\{\hat{p}_1 + \hat{p}_2, \hat{q}_1 + \hat{q}_2\} - (\hat{p}_1 - \hat{p}_2)^2}{n}}. \qquad \hat{q}_1 = 1 - \hat{p}_1 \quad \hat{q}_2 = 1 - \hat{p}_2.$$

√ Use a regra de 10% para cada proporção

- Intervalo de confiança para a diferença de proporções

$$\hat{p}_1 - \hat{p}_2 \pm z \times \text{ep}(\hat{p}_1 - \hat{p}_2).$$

Exemplo – Eleições

- Diferenças nos níveis de apoio a Clinton e Dole

Estado	n	Pesquisas de opinião pré-eleitorais				Resultados da eleição		
		Clinton	Dole	Perot	Outros/ Indecisos	Clinton	Dole	Perot
Nova Jérsei	1.000	51	33	8	8	53	36	9

√ Clinton com 18% à frente ($\hat{p}_1 - \hat{p}_2 = 0{,}51 - 0{,}33$)

√ Erro padrão da diferença:

$$\text{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{0,51 + 0,33 - (0,51 - 0,33)^2}{1.000}} = 0,0284183.$$

√ IC de 95% de confiança para a verdadeira diferença

$$\hat{p}_1 - \hat{p}_2 \pm z \times \text{ep}(\hat{p}_1 - \hat{p}_2)$$

$$0,18 \pm 1,96 \times 0,0284183 = [0,12; 0,24].$$

√ Diferença na eleição: 17%

Exemplo – Características Estudantes

- Comparação da % de fumantes que usam maconha e dos que se embriagam

Entrada na tabela é a % que diz sim	Fumante[a]	Não-fumante	Que bebe[b]	Que não bebe
Obtém geralmente A ou B?	41	68		
Lê uma ou mais horas por dia?	54	72	56	75
Embriaga-se ao menos uma vez por mês?	63	10		
Já fumou maconha?	79	14	52	12
Tamanho da amostra	130	870	260	740

√ Situação de amostragem – 3

√ % experimentaram maconha é maior que a dos que se embriagaram ($\hat{p}_1 - \hat{p}_2 = 0{,}16$)

√ n =130 é suficientemente grande

√ Erro padrão da diferença:

$$\hat{q}_1 = 1 - \hat{p}_1 = 1 - 0,79 = 0,21$$
$$\hat{q}_2 = 1 - \hat{p}_2 = 1 - 0,63 = 0,37.$$
$$\text{ep}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\min\{\hat{p}_1 + \hat{p}_2, \hat{q}_1 + \hat{q}_2\} - (\hat{p}_1 - \hat{p}_2)^2}{n}}$$
$$= \sqrt{\frac{\min\{0,79 + 0,63; 0,21 + 0,37\} - (0,70 - 0,63)^2}{130}}$$
$$= \sqrt{\frac{0,58 - 0,16^2}{130}} = 0,06530402.$$

√ IC de 95% de confiança para a diferença

$$\hat{p}_1 - \hat{p}_2 \pm z \times \text{ep}(\hat{p}_1 - \hat{p}_2)$$
$$0,16 \pm 1,96 \times 0,06530402 = [0,03; 0,29].$$

√ Verdadeira diferença poderia ser tão pequena quanto 3% e tão grande quanto 29%.

Tamanho Amostral

- Qual tamanho deve ter minha amostra?
  - √ Resp.: Tão grande quanto possível
- Situações:
  - √ Proporção
  - √ Média

Tamanho de Amostra – Uma Proporção

- Margem de erro (m):
  - √ Expressa como um valor entre 0 e 1.

$$m = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

- Tamanho da amostra (n)

$$n = \left(\frac{z}{m}\right)^2 \times p^*(1-p^*).$$

  - √ p^*: valor 'adivinhado' (prévio à amostragem)
  - √ $\hat{p}$: valor conhecido após amostragem
  - √ z: multiplicador para o nível de confiança
  - √ Margem de erro não maior que m

Exemplo – Relacionamento em Empresas

- IC de 95% para a proporção de executivos que julgam que empresa não deve interferir
  - √ IC: [0,64; 0,77]
  - √ m = 0,065
  - √ n = 200
- Tamanho da amostra para $m \leq 0,03$
  - ▪ $p^* = 0,7$

$$n = \left(\frac{z}{m}\right)^2 \times p^*(1-p^*) = \left(\frac{1,96}{0,03}\right)^2 \times 0,7 \times 0,3 = 897.$$

$$n = \left(\frac{z}{m}\right)^2 \times p^*(1-p^*) = \left(\frac{1,96}{0,03}\right)^2 \times 0,5 \times 0,5 = 1.068.$$

N máximo

Intervalo com 95% de Confiança para uma Proporção

- Tamanhos amostrais mínimos:

	Margem de erro (m)			
p*(a)	0,10	0,05	0,03	0,01
0,5	96	384	1.067	9.604
0,3 ou 0,7	81	323	896	8.067
0,1 ou 0,9	35	138	384	3.457

(a)p* é o valor 'adivinhado para a proporção

Tamanho de Amostra – Uma Média

- Margem de erro (m):

$$m = t \times \frac{s_X}{\sqrt{n}}.$$

√ s_X : estimativa amostral de σ.

√ t: multiplicador para o nível de confiança

- Tamanho da amostra (n)

$$n = \left(\frac{z\sigma^*}{m}\right)^2.$$

√ σ^*: valor 'adivinhado' (prévio à amostragem)

√ Quando n é grande $t \approx z$

Obtenção de σ^*

- Para se obter σ^*:
  1. Desvio padrão de dados "semelhantes" coletados no passado
  2. Estudo piloto para planejar o estudo principal
  3. Usando a expressão:

$$\sigma^* = \frac{\text{máximo imaginável} - \text{mínimo imaginável}}{6}.$$

- 3 σ para cada lado da média

Exemplo

- Roupas para atletas

√ Venda de roupas e equipamentos pela internet

√ Características físicas de diferentes tipos de clientes do sexo masculino

√ Amostra aleatória de tamanho 24 corredores

√ Dados e resultados anteriores:

n	$\bar{x}$	s_X	t	IC 95%	m
24	61,79	4,808	2,069	[59,76; 63,82]	2,030

√ Tamanho da amostra para dobrar a precisão:

– $m = 1{,}015$

- Tamanho da amostra (n):
  - √ $\sigma^* = 4{,}808$

$$n = \left(\frac{z\sigma^*}{m}\right)^2 = \left(\frac{1,96 \times 4,808}{1,015}\right)^2$$
$$= 86,2.$$

  - √ Precisaríamos tomar cerca de 87 observações

153

Estatística Econômica II - 2018

Comentários

- Tamanhos de amostras obtidos a partir dessas expressões não são muito confiáveis
  - √ O ele fraco é a estimativa σ^*
- Desvios padrões amostrais de pequenos estudos piloto são notadamente variáveis
- Suposição de mesma variabilidade com dados passados é frequentemente não confiável
- Método 3 é ainda pior

154

Estatística Econômica II - 2018

Referências

Bibliografia Recomendada

- AGRESTI, A.; FINLAY, B. *Métodos estatísticos para as ciências sociais.* Penso, 2012.
- MOORE, D. S.; MCCABE, G. P.; DUCKWORTH, W. M.; SLOVE, S. L. *A prática da estatística empresarial: como usar dados para tomar decisões*. LTC, 2006.
- WILD, J. W.; SEBER, G. A. F. *Encontros com o acaso: um primeiro curso de análise de dados e inferência*. LTC, 2004.

156

Estatística Econômica II - 2018