Estatística II

Licenciatura em Gestão do Desporto
2.º Ano/2.º Semestre
2024/2025

Aulas Teórico-Práticas N.º 14 e 15 (Semana 8)

Docente: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

https://doity.com.br/estatistica-aplicada-a-nutricao

https://basiccode.com.br/produto/informatica-basica/

Conteúdos Programáticos

Aulas Teórico-Práticas (Semanas 1 a 3)	Aulas Teórico-Práticas (Semanas 4 a 7)	Aulas Teórico-Práticas (Semanas 7 a 9)	Aulas Teórico-Práticas (Semanas 10 a 13)
• **Capítulo 1**: Revisões e Distribuições de Amostragem	• **Capítulo 2**: Estimação	• **Capítulo 3**: Testes de Hipóteses	• **Capítulo 4**: Modelo de Regressão Linear Múltipla

Material didático: Exercícios do Livro Murteira et al (2015), Formulário e Tabelas Estatísticas

Bibliografia: B. Murteira, C. Silva Ribeiro, J. Andrade e Silva, C. Pimenta e F. Pimenta; *Introdução à Estatística*, 2ª ed., Escolar Editora, 2015.

Intervalo de Confiança para uma proporção p

1

Intervalo de Confiança para p

Estimador de p

$$\boxed{\bar{P} = \hat{P}}$$

Variável Fulcral

$$Z = \frac{\bar{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\circ}{\sim} N(0;1).$$

Portanto, quando a amostra é grande, o **I. C. para p com $100(1-\alpha)\%$ de confiança** é:

$$\left]\bar{P} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{P}(1-\bar{P})}{n}}\,;\ \bar{P} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{P}(1-\bar{P})}{n}}\right[.$$

ProbabilidadesEstatistica_2019 (uevora.pt)

IC para p: Formulário

Considerar em Testes de Hipóteses

Considerar em Intervalos de Confiança

População de Bernoulli

Proporção	$\dfrac{\overline{X}-\theta}{\sqrt{\dfrac{\theta(1-\theta)}{n}}} \overset{a}{\sim} N(0,1)$	$\dfrac{\overline{X}-\theta}{\sqrt{\dfrac{\overline{X}(1-\overline{X})}{n}}} \overset{a}{\sim} N(0,1)$
Diferença de proporções	$\dfrac{\overline{X}_1-\overline{X}_2-(\theta_1-\theta_2)}{\sqrt{\dfrac{\theta_1(1-\theta_1)}{m}+\dfrac{\theta_2(1-\theta_2)}{n}}} \overset{a}{\sim} N(0,1)$	$\dfrac{\overline{X}_1-\overline{X}_2-(\theta_1-\theta_2)}{\sqrt{\dfrac{\overline{X}_1(1-\overline{X}_1)}{m}+\dfrac{\overline{X}_2(1-\overline{X}_2)}{n}}} \overset{a}{\sim} N(0,1)$
Igualdade de proporções	$\dfrac{\overline{X}_1-\overline{X}_2-(\theta_1-\theta_2)}{\sqrt{\left(\dfrac{1}{m}+\dfrac{1}{n}\right)\hat{\theta}(1-\hat{\theta})}} \overset{a}{\sim} N(0,1)$ onde $\hat{\theta}=\dfrac{m\overline{X}_1+n\overline{X}_2}{m+n}$	

Intervalo de Confiança para uma proporção p: Exercícios

2

Exercício Suplementar que não consta do livro Murteira et al (2015)

Numa certa cidade A recolheu-se uma amostra aleatória de 150 homens tendo 54 afirmado que viam o telejornal todos os dias.

a) Com 90% de confiança, será que se pode considerar que a proporção de homens, daquela cidade, que veem o telejornal todos os dias é de 40%.

b) Mantendo-se o resto constante, qual deveria ser a dimensão da amostra de forma a que o erro de estimativa do intervalo de confiança não ultrapasse 5%?

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (a): IC para p

Sejam:

- X_i a v. a. que designa se o i-ésimo homem afirmou ver o telejornal,
- $\overline{P}$ a v. a. que representa a proporção de homens que afirmaram ver o telejornal, em n homens.

$n = 150$ e $\overline{p} = \frac{54}{150} = 0{,}36.$

a) Afirmação: $p = 0{,}4$.

I. C. a 90% para p é dado por:

$$\left]\overline{P} - z_{0,95}\sqrt{\frac{\overline{P}(1-\overline{P})}{n}};\ \overline{P} + z_{0,95}\sqrt{\frac{\overline{P}(1-\overline{P})}{n}}\right[.$$

Substituindo pelos valores conhecidos, com $z_{0,95} = 1{,}645$, obtém-se

$$\left]0{,}36 - 1{,}645\sqrt{\frac{0{,}36(1-0{,}36)}{150}};0{,}36 + 1{,}645\sqrt{\frac{0{,}36(1-0{,}36)}{150}}\right[=]0{,}3464; 0{,}5077[.$$

Deste modo, face aos resultados obtidos (0,4 está contido do I. C. a 90%) não é de rejeitar a hipótese de que a proporção de homens que vê o telejornal todos os dias é de 40%, pois com 90% de confiança a percentagem de homens que vê diariamente o telejornal situa-se entre 34,64% e 50,77%.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (b): IC para p

Margem de erro ou Erro de estimativa é metade da amplitude do IC

b) Erro de estimativa $\leq 0{,}05$, então $n = ?$

O erro de estimativa associado ao I. C. a 90% para p é:

$$z_{0,95}\sqrt{\frac{\overline{P}(1-\overline{P})}{n}},$$

que se pretende que seja inferior ou igual a 0,5. Portanto,

$$z_{0,95}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \leq 0{,}05 \Leftrightarrow 1{,}645\sqrt{\frac{0{,}36(1-0{,}36)}{n}} \leq 0{,}05$$

$$\Leftrightarrow \sqrt{n} \geq \frac{1{,}645\sqrt{0{,}36(1-0{,}36)}}{0{,}05} \Leftrightarrow \sqrt{n} \geq 15{,}792$$

$$\Rightarrow n \geq 15{,}792^2 \Leftrightarrow n \geq 249{,}4 \Rightarrow n \geq 250$$

Desta forma, a dimensão mínima da amostra que garante que o erro de estimativa do I. C. a 90% é no máximo de 5% é de 250 homens.

ProbabilidadesEstatistica_2019 (uevora.pt)

Murteira et al (2015)
Capítulo 7

40. Uma cadeia de televisão fixou como objectivo ter uma audiência, para a telenovela XYZ, de pelo menos 55% dos telespectadores. Consultada uma empresa de audiometria, ficou estabelecido que semanalmente a empresa interrogaria por telefone, dentro do horário da novela, uma amostra casual de 200 telespectadores e registaria quantos estavam a ver a referida novela. Em determinada semana obtiveram-se 90 respostas positivas. Construa um intervalo de confiança a 95% para a percentagem de telespectadores da telenovela XYZ nessa semana. Com base neste intervalo, diga o que pode concluir sobre o cumprimento do objectivo fixado.

Exercício 40

Pop. de Bernoulli: $X \sim Be(\theta)$

$$X = \begin{cases} 1 & \text{(se o telespectador assiste à novela)} \\ 0 & \text{(caso contrário} \end{cases}$$

$X \sim Be(\theta)$ objetivo: $\theta \geqslant 0.55$

Amostra: $n = 200$ telespecadores

$$\sum_{i=1}^{200} x_i = 90 \qquad \bar{x} = \frac{90}{200} = 0.45$$

$$\frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} N(0,1) \qquad z_{0.025} = 1.96$$

$$P\left(-z_{0.025} < \frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} < z_{0.025}\right) = 0.95$$

Exercício 40

$$IC_{95\%}(\theta) = \bar{x} \pm z_{0.025}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} =$$

$$= 0.45 \pm 1.96\sqrt{\frac{0.45(1-0.45)}{200}} =$$

$$= (0.381, 0.519)$$

$\theta = 0.55 \notin IC_{95\%}(\theta)$ e $\forall \theta \in IC_{95\%}(\theta)$ temos $\theta < 0.55$

O intervalo de confiança obtido sugere que o objetivo fixado não foi cumprido.

Murteira et al (2015)
Capítulo 7

44. Pretende saber-se qual a proporção de consumidores do produto A. Para tal foram recolhidas amostras de 200 pessoas nas cidades F e G, tendo-se observado, respectivamente, 140 e 160 consumidores desse produto.

   a) Determine, com um grau de confiança de 95%, a percentagem de consumidores do produto A na cidade F.

   b) Pode afirmar-se, com um grau de confiança de 99%, que o consumo do produto A é superior na cidade F?

Exercício 44 a)

Duas Populações de Bernoulli independentes

$X_F \sim B(1, \theta_F)$ $\qquad$ $X_G \sim B(1, \theta_G)$

θ_F = Proporção de consumidores do produto A na cidade F.

θ_G = Proporção de consumidores do produto A na cidade G.

$n_F = n_G = 200$ (amostras grandes)

$\sum_{i=1}^{200} x_{F_i} = 140$ $\qquad$ $\sum_{i=1}^{200} x_{G_i} = 160$

$\bar{x}_F = \frac{140}{200} = 0.7$ $\qquad$ $\bar{x}_G = \frac{160}{200} = 0.8$

Exercício 44 a)

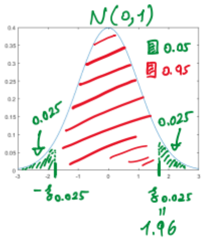
a) I.C. para θ_F com $1 - \alpha = 0.95$

$\alpha = 0.05 \quad \frac{\alpha}{2} = 0.025$

Variável fulcral: $Z = \dfrac{\bar{X}_F - \theta_F}{\sqrt{\dfrac{\bar{X}_F(1-\bar{X}_F)}{n_F}}} \overset{a}{\sim} N(0,1)$

$P\left(-z_{0.025} < Z < z_{0.025}\right) = 0.95 \Leftrightarrow$

$\Leftrightarrow P\left(-z_{0.025} < \dfrac{\bar{X}_F - \theta_F}{\sqrt{\dfrac{\bar{X}_F(1-\bar{X}_F)}{n_F}}} < z_{0.025}\right) =$



$= P\left(\bar{X}_F - z_{0.025}\sqrt{\dfrac{\bar{X}_F(1-\bar{X}_F)}{n_F}} < \theta_F < \bar{X}_F + z_{0.025}\sqrt{\dfrac{\bar{X}_F(1-\bar{X}_F)}{n_F}}\right) = 0.95$

I.A.$_{95\%}$ $\bar{X}_F \pm z_{0.025}\sqrt{\dfrac{\bar{X}_F(1-\bar{X}_F)}{n_F}}$

I.C.$_{95\%}$: $\bar{x}_F \pm \underset{\text{Tabela 5}}{z_{0.025}} \cdot \sqrt{\dfrac{\bar{x}_F(1-\bar{x}_F)}{n_F}} =$

$= 0.7 \pm 1.96\sqrt{\dfrac{0.7(1-0.7)}{200}} =$

$= (0.6365, 0.7635)$

ou I.C.$_{95\%}$ $= (63.65\%, 76.35\%)$

Intervalo de Confiança para a Diferença de Proporções p_1-p_2

3

Intervalo de Confiança para p_1-p_2

Estimadores de p1 e p2

$$\overline{P1} = \hat{P}_1$$

$$\overline{P2} = \hat{P}_2$$

Variável Fulcral

$$Z = \frac{(\overline{P}_1 - \overline{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \overset{\circ}{\sim} N(0; 1).$$

Portanto, quando as amostras são grandes, o **I. C. para $p_1 - p_2$ com $100(1-\alpha)\%$ de confiança** é dado por:

$$\left]\overline{P}_1 - \overline{P}_2 - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{P}_1(1-\overline{P}_1)}{n_1} + \frac{\overline{P}_2(1-\overline{P}_2)}{n_2}}; \overline{P}_1 - \overline{P}_2 + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{P}_1(1-\overline{P}_1)}{n_1} + \frac{\overline{P}_2(1-\overline{P}_2)}{n_2}}\right[.$$

ProbabilidadesEstatistica_2019 (uevora.pt)

IC para p_1-p_2 : Formulário

População de Bernoulli

Proporção	$\dfrac{\bar{X}-\theta}{\sqrt{\dfrac{\theta(1-\theta)}{n}}} \overset{a}{\sim} N(0,1)$	$\dfrac{\bar{X}-\theta}{\sqrt{\dfrac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} N(0,1)$
Diferença de proporções	$\dfrac{\bar{X}_1-\bar{X}_2-(\theta_1-\theta_2)}{\sqrt{\dfrac{\theta_1(1-\theta_1)}{m}+\dfrac{\theta_2(1-\theta_2)}{n}}} \overset{a}{\sim} N(0,1)$	$\dfrac{\bar{X}_1-\bar{X}_2-(\theta_1-\theta_2)}{\sqrt{\dfrac{\bar{X}_1(1-\bar{X}_1)}{m}+\dfrac{\bar{X}_2(1-\bar{X}_2)}{n}}} \overset{a}{\sim} N(0,1)$
Igualdade de proporções	$\sqrt{\left(\frac{}{m}+\frac{}{n}\right)\theta(1-\theta)}$	$\hat{\theta} = \dfrac{m\bar{X}_1 \ldots}{m \ldots}$

Considerar em Testes de Hipóteses

Considerar em Intervalos de Confiança

Intervalo de Confiança para a Diferença de Proporções p1-p2: Exercícios

4

Exercício Suplementar que não consta do livro Murteira et al (2015)

Numa certa cidade A recolheu-se uma amostra aleatória de 150 homens tendo 54 afirmado que viam o telejornal todos os dias. Numa outra cidade do país, cidade B, 80 dos 200 homens selecionados aleatoriamente responderam afirmativamente.

Com 95% de confiança, será de admitir que a proporção de homens que vê o telejornal todos os dias é igual nas duas cidades?

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício: IC para p_1-p_2

Sejam:

- X_{1i} a v. a. que designa se o i-ésimo homem, da cidade A, afirmou ver o telejornal, $i = 1, \dots, n_1$,
- X_{2i} a v. a. que designa se o i-ésimo homem, da cidade B, afirmou ver o telejornal, $i = 1, \dots, n_2$,
- $\overline{P}_1$ a v. a. que representa a proporção de homens, da cidade A, que afirmaram ver o telejornal, em n_1 homens,
- $\overline{P}_2$ a v. a. que representa a proporção de homens, da cidade B, que afirmaram ver o telejornal, em n_2 homens.

$$n_1 = 150;\ \ \overline{p}_1 = \frac{54}{150} = 0{,}36;\ \ n_2 = 200 \text{ e } \overline{p}_2 = \frac{80}{200} = 0{,}4.$$

O I. C. a 95% para $p_1 - p_2$ é dado por:

$$\left]\overline{P}_1 - \overline{P}_2 - z_{0,975}\sqrt{\frac{\overline{P}_1(1-\overline{P}_1)}{n_1} + \frac{\overline{P}_2(1-\overline{P}_2)}{n_2}}; \overline{P}_1 - \overline{P}_2 + z_{0,975}\sqrt{\frac{\overline{P}_1(1-\overline{P}_1)}{n_1} + \frac{\overline{P}_2(1-\overline{P}_2)}{n_2}}\right[.$$

Substituindo pelos valores conhecidos, sendo $z_{0,975} = 1{,}96$, obtém-se:

$$\left](0{,}36 - 0{,}4) \pm 1{,}96\sqrt{\frac{0{,}36(1-0{,}36)}{150} + \frac{0{,}4(1-0{,}4)}{200}}\right[=]-0{,}143; 0{,}063[.$$

Portanto, com 95% de probabilidade a diferença entre a percentagem de homens da cidade A e cidade B que veem o telejornal diariamente está entre -14,3% e 6,3%. Como o 0 está contido no intervalo não é de excluir a hipótese de que a percentagem de homens é idêntica nas duas cidades, com a referida confiança.

ProbabilidadesEstatistica_2019 (uevora.pt)

Murteira et al (2015)
Capítulo 7

44. Pretende saber-se qual a proporção de consumidores do produto A. Para tal foram recolhidas amostras de 200 pessoas nas cidades F e G, tendo-se observado, respectivamente, 140 e 160 consumidores desse produto.

a) Determine, com um grau de confiança de 95%, a percentagem de consumidores do produto A na cidade F.

b) Pode afirmar-se, com um grau de confiança de 99%, que o consumo do produto A é superior na cidade F?

Exercício 44 b)

b) I.C. a 99% para $\theta_F - \theta_G$.

$1-\alpha = 0.99 \quad \alpha = 0.01 \quad \frac{\alpha}{2} = 0.005$

Variável fulcral: $$Z = \frac{\overline{X}_F - \overline{X}_G - (\theta_F - \theta_G)}{\sqrt{\frac{\overline{X}_F(1-\overline{X}_F)}{m_F} + \frac{\overline{X}_G(1-\overline{X}_G)}{m_G}}} \overset{a}{\sim} N(0,1)$$

$$P(-z_{0.005} < Z < z_{0.005}) = 0.99 = \cdots =$$

$$= P\left(\overline{X}_F - \overline{X}_G - z_{0.005}\sqrt{\frac{\overline{X}_F(1-\overline{X}_F)}{m_F} + \frac{\overline{X}_G(1-\overline{X}_G)}{m_G}} < \theta_F - \theta_G < X_F - \overline{X}_G + z_{0.005}\sqrt{\frac{\overline{X}_F(1-\overline{X}_F)}{m_F} + \frac{\overline{X}_G(1-\overline{X}_G)}{m_G}}\right)$$

$$IC_{95\%}: \bar{x}_F - \bar{x}_G \pm z_{0.005}\sqrt{\frac{\bar{x}_F(1-\bar{x}_F)}{m_F} + \frac{\bar{x}_G(1-\bar{x}_G)}{m_G}} =$$

Tabela 5

$$= 0.7 - 0.8 \pm 2.576\sqrt{\frac{0.7(1-0.7)}{200} + \frac{0.8(1-0.8)}{200}} =$$

$$= (-0.2108, \ 0.0108)$$

Resposta: Não, porque o valor $\theta_F - \theta_G = 0$ pertence ao $I.C._{99\%}(\theta_F - \theta_G)$

Murteira et al (2015)
Capítulo 7

49. Com o objectivo de avaliar o efeito da dimensão da turma sobre o aproveitamento dos alunos recolheu-se a seguinte informação (amostras casuais de alunos da mesma disciplina):

Classificação (0-20)	**< 8**	**[8, 10)**	**[10, 12)**	**[12, 14)**	**≥ 14**	**Total**
Turmas grandes	12	42	30	11	5	100
Turmas pequenas	4	30	40	21	5	100

a) Construa um intervalo de confiança a 90% para a nota média dos alunos de turmas pequenas.

b) Construa um intervalo de confiança a 90% para a diferença na classificação média das duas populações e comente o resultado obtido.

c) Com base num intervalo de confiança a 90%, diga se pode afirmar que a proporção de alunos com nota positiva é maior nas turmas pequenas.

Exercício 49 c)

c) Duas populações de Bernoulli

Turmas grandes:

$$Y_G \equiv \begin{cases} 1 & (\text{se o aluno tem positiva}) \\ 0 & (\text{caso contrário}) \end{cases} \sim Ber(\theta_G)$$

Turmas pequenas:

$$Y_p \equiv \begin{cases} 1 & (\text{se o aluno tem positiva}) \\ 0 & (\text{caso contrário}) \end{cases} \sim Ber(\theta_P)$$

$$\bar{y}_G = \frac{30+11+5}{100} = \frac{46}{100} = 0.46$$

$$\bar{y}_P = \frac{40+21+5}{100} = \frac{66}{100} = 0.66$$

$$\frac{\bar{Y}_G - \bar{Y}_P - (\theta_G - \theta_P)}{\sqrt{\frac{\bar{Y}_G(1-\bar{Y}_G)}{m_G} + \frac{\bar{Y}_P(1-\bar{Y}_P)}{m_P}}} \overset{a}{\sim} N(0,1)$$

Exercício 49 c)

$$IC_{90\%}(\theta_G - \theta_P) = (\bar{y}_G - \bar{y}_P) \pm z_{0.05}\sqrt{\frac{\bar{y}_G(1-\bar{y}_G)}{n_G} + \frac{\bar{y}_P(1-\bar{y}_P)}{n_P}} =$$

$$= (0.46 - 0.66) \pm 1.645\sqrt{\frac{0.46(1-0.46)}{100} + \frac{0.66(1-0.66)}{100}} =$$

$$= (-0.3131 \,;\, -0.087)$$

Como todos os valores de $\theta_g - \theta_p$ pertencentes ao IC são negativos, com base neste IC conclui-se que a proporção de positivas deve ser maior nas turmas pequenas $\theta_p > \theta_g$.

Obrigada!

Questões?