Estatística II

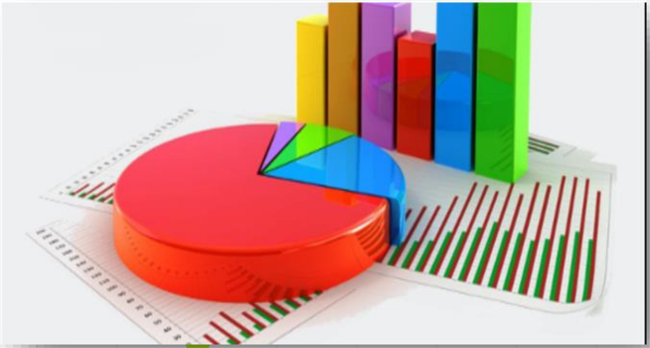
Licenciatura em Gestão do Desporto

2.º Ano/2.º Semestre

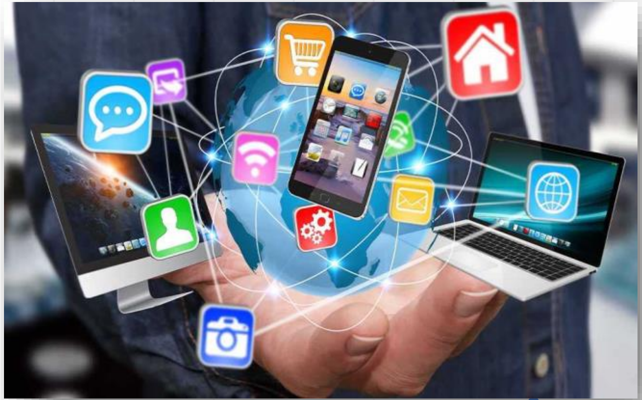
2023/2024

Aulas Teórico-Práticas N.º 4 e 5 (Semana 3)

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 5)	Aulas Teórico-Práticas (Semanas 5 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.

Diferença de Médias Amostrais

Distribuições de Amostragem

1

Diferença de Médias Amostrais

Sejam $X_{11}, X_{12}, \dots, X_{1n_1}$ e $X_{21}, X_{22}, \dots, X_{2n_2}$ duas a. a. independentes, de dimensão n_1 e n_2 retiradas de duas populações com médias μ_1 e μ_2 e desvios padrão σ_1 e σ_2, respetivamente, e

$$\overline{X}_1 = \sum_{i=1}^{n_1} \frac{X_{1i}}{n_1} \quad \text{e} \quad \overline{X}_2 = \sum_{i=1}^{n_2} \frac{X_{2i}}{n_2}.$$

ProbabilidadesEstatistica2019.pdf

Diferença de Médias: Variâncias Conhecidas

Se as populações forem Normais, sendo $X_1 \sim N(\mu_1; \sigma_1)$ e $X_2 \sim N(\mu_2; \sigma_2)$, com σ_1 e σ_2 conhecidos, então, pelo Teorema da aditividade da distribuição Normal, tem-se que:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0; 1).$$

Se as populações não forem Normais, mas as amostras forem de grande dimensão então, pelo corolário do T. L. C., vem $Z \overset{\circ}{\sim} N(0; 1)$, considerando o já anteriormente exposto em situação análoga.

ProbabilidadesEstatistica2019.pdf

Diferença de Médias: Variâncias Desconhecidas e Iguais

Se as populações forem Normais, sendo $X_1 \sim N(\mu_1; \sigma_1)$ e $X_2 \sim N(\mu_2; \sigma_2)$, com σ_1 e σ_2 desconhecidos mas iguais ($\sigma_1 = \sigma_2$), então demonstra-se que:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}.$$

Se as populações não forem Normais, mas as amostras forem de grande dimensão então, por extensão do T. L. C., a expressão anterior segue aproximadamente uma $N(0; 1)$.

ProbabilidadesEstatistica2019.pdf

Diferença de Médias: Variâncias Desconhecidas e Diferentes

Se as populações forem Normais, sendo $X_1 \sim N(\mu_1; \sigma_1)$ e $X_2 \sim N(\mu_2; \sigma_2)$, com σ_1 e σ_2 desconhecidos e diferentes ($\sigma_1 \neq \sigma_2$), então, pela aproximação de Welch tem-se que (Murteira *et al.*, 2007):

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \mathrel{\overset{\circ}{\sim}} t_v, \text{ onde } v = \left[\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{S_2^2}{n_2}\right)^2} \right].$$

sendo $[r]$ a parte inteira de r, ou seja, arredonda-se por defeito o valor obtido.

Também neste caso, se as populações não forem Normais, mas as amostras forem de grande dimensão então, por extensão do T. L. C., a expressão anterior segue aproximadamente uma $N(0; 1)$, sendo válidas as observações anteriores.

ProbabilidadesEstatistica2019.pdf

Diferença de Médias Amostrais: Exercícios

Distribuições de Amostragem

2

Uma dada empresa farmacêutica lançou no mercado um novo medicamento, para dormir, que tem estado a ser utilizado nos hospitais. Constatou-se que os doentes não sujeitos a este medicamento em média dormiam 7,5 horas, com desvio padrão de 1,4 horas, ao passo que os doentes aos quais se administrou este medicamento dormiam em média 8 horas com desvio padrão de 2 horas.

Num determinado hospital observaram-se 31 doentes não sujeitos ao referido medicamento e 61 sob a referida medicação. Qual a probabilidade de os doentes do primeiro grupo observado dormirem em média mais do que os do segundo grupo? Assuma a normalidade das distribuições.

ProbabilidadesEstatistica2019.pdf

Exercício: Diferença de Médias Amostrais e Variâncias Conhecidas

Sejam:

- X_1 a v.a. que representa o número de horas de sono dos doentes não sujeitos ao medicamento,
- X_2 a v.a. que representa o número de horas de sono dos doentes sujeitos ao medicamento,

com

$$X_1 \sim N(\mu_1 = 7{,}5; \sigma_1 = 1{,}4) \text{ e } X_2 \sim N(\mu_2 = 8; \sigma_2 = 2).$$

$$\left.\begin{array}{l} X_1 \text{ e } X_2 \text{ dist. Normal} \\ \sigma_1\ (= 1{,}4) \text{ e } \sigma_2\ (= 2) \text{ conhecidos} \end{array}\right| \Rightarrow Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0; 1).$$

$$P(\overline{X}_1 > \overline{X}_2) = P(\overline{X}_1 - \overline{X}_2 > 0) = 1 - P(\overline{X}_1 - \overline{X}_2 \leq 0) = 1 - P\left(Z \leq \frac{0 - (7{,}5 - 8)}{\sqrt{\frac{1{,}4^2}{31} + \frac{2^2}{61}}}\right) = 1 - \Phi(1{,}39)$$

$$= 1 - 0{,}9177 = 0{,}0823.$$

ProbabilidadesEstatistica2019.pdf

Uma dada empresa farmacêutica lançou no mercado um novo medicamento, para dormir, que tem estado a ser utilizado nos hospitais. Constatou-se que os doentes não sujeitos a este medicamento em média dormiam 7,5 horas, enquanto os doentes aos quais se administrou este medicamento dormiam em média 8 horas.

Num determinado hospital observaram-se n_1 doentes não sujeitos ao referido medicamento e n_2 sob a referida medicação tendo-se obtido, respetivamente, os seguintes desvios-padrão: 1,4 horas e 2 horas. Determine a probabilidade de os doentes do primeiro grupo dormirem em média menos do que os do segundo grupo, quando $n_1 = 20$ e $n_2 = 27$, quando se verifica a normalidade das distribuições e se considera:

a) A igualdade das variâncias populacionais.

b) A desigualdade das variâncias populacionais.

ProbabilidadesEstatistica2019.pdf

Exercício: Diferença de Médias Amostrais e Variâncias Desconhecidas e Iguais

Sejam:

- X_1 a v.a. que representa o número de horas de sono dos doentes não sujeitos ao medicamento,
- X_2 a v.a. que representa o número de horas de sono dos doentes sujeitos ao medicamento,

Com $X_1 \sim N(\mu_1 = 7{,}5; \sigma_1 =?)$ e $X_2 \sim N(\mu_2 = 8; \sigma_2 =?)$.

$n_1 = 20;\ \ s_1 = 20;\ \ n_2 = 27 \text{ e } s_2 = 2.$

a)

X_1 e X_2 dist. Normal

σ_1 e σ_2 desconhecidos, mas iguais

$$\Rightarrow T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t_{n_1+n_2-2=45}.$$

$$P(\overline{X}_1 < \overline{X}_2) = P(\overline{X}_1 - \overline{X}_2 < 0) = P\left(T < \frac{0-(7{,}5-8)}{\sqrt{\dfrac{(20-1)1,4^2 + (27-1)2^2}{20+27-2}}\sqrt{\dfrac{1}{20}+\dfrac{1}{27}}}\right) = P(T < 0{,}957)$$

$$\approx 0{,}83.$$

ProbabilidadesEstatistica2019.pdf

Exercício: Diferença de Médias Amostrais e Variâncias Desconhecidas e Diferentes

b)

X_1 e X_2 dist. Normal

σ_1 e σ_2 desconhecidos e diferentes

$$\Rightarrow T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{\circ}{\sim} t_{v=44},$$

pois

$$v = \left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} \right] = \left[\frac{\left(\frac{1,4^2}{20} + \frac{2^2}{27}\right)^2}{\frac{1}{20 - 1}\left(\frac{1,4^2}{20}\right)^2 + \frac{1}{27 - 1}\left(\frac{2^2}{27}\right)^2} \right] = [44,9] = 44.$$

$$P(\overline{X}_1 < \overline{X}_2) = P(\overline{X}_1 - \overline{X}_2 < 0) = P\left(T < \frac{0 - (7,5 - 8)}{\sqrt{\frac{1,4^2}{20} + \frac{2^2}{27}}} \right) = P(T < 1,008) \approx 0,84.$$

ProbabilidadesEstatistica2019.pdf

Proporção Amostral

Distribuições de Amostragem

3

Proporção Amostral

Seja $X_1, X_2, \ldots, X_n$, uma a. a. duma população Bernoulli com probabilidade de sucesso p, em que X_i toma o valor 1 se for um sucesso e o valor 0 se for um insucesso. A **proporção amostral** de sucessos é dada por:

$$\overline{P} = \sum_{i=1}^{n} \frac{X_i}{n}.$$

Propriedades:

- $\mu_{\overline{P}} = E(\overline{P}) = p.$
- $\sigma_{\overline{P}}^2 = Var(\overline{P}) = \frac{p(1-p)}{n}.$

ProbabilidadesEstatistica2019.pdf

Proporção Amostral

Se a dimensão da amostra for pequena então sabe-se que,

$$P\left(\sum_{i=1}^{n} X_i = x\right) = {}^{n}C_x p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n, \text{com } 0 < p < 1,$$

pelo que

$$P(\overline{P} = a) = P(X = na) = {}^{n}C_{na} p^{na} (1-p)^{n-na}, \qquad a = 0, \frac{1}{n}, \frac{2}{n} \ldots, \frac{n}{n}.$$

Se a dimensão da amostra for grande então, pelo Teorema de Moivre-Laplace (i.e., variante do T. L. C. para a distribuição Binomial),

$$\overline{P} \overset{\circ}{\sim} N\left(p; \sqrt{\frac{p(1-p)}{n}}\right), \text{ ou seja, } Z = \frac{\overline{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\circ}{\sim} N(0; 1).$$

Observação: Esta aproximação tem sido muitas vezes criticada, pois tem mostrado resultados consideravelmente insatisfatórios e, apenas, continua a ser usada dada a sua simplicidade e implementação nos diversos programas estatísticos. Recomenda-se especial atenção se a proporção for tendencialmente pequena ou grande, i.e., valores próximos de 0 ou de 1 (Pires e Amado, 2008).

ProbabilidadesEstatistica2019.pdf

Proporção Amostral: Exercícios

Distribuições de Amostragem

4

Numa dada Repartição de Finanças, sabe-se que 70% dos contribuintes pagam o IUC (Imposto Único de Circulação) dentro do prazo.

a) Qual a probabilidade de numa amostra de 35 IUC, pelo menos 65% terem sido pagos dentro do prazo?

b) Para a probabilidade da alínea anterior ser no mínimo 80%, qual deve ser a dimensão mínima da amostra a recolher (admita que a amostra será de grande dimensão)?

ProbabilidadesEstatistica2019.pdf

Exercício: Proporção Amostral

Sejam:

- X_i a v. a. que designa se o contribuinte i paga o IUC dentro do prazo, $i = 1, \dots, n$,
- $\overline{P}$ a v.a. que representa a proporção de contribuintes que pagam o IUC dentro do prazo, em n contribuintes.

$p = 0{,}7$.

a)

$$\left.\begin{array}{l} X \text{ distribuição Bernoulli} \\ n\ (= 35) \text{ grande} \end{array}\right| \Rightarrow Z = \frac{\overline{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\circ}{\sim} N(0;1).$$

$$P(\overline{P} \geq 0{,}65) = 1 - P(\overline{P} < 0{,}65) = 1 - P\left(Z < \frac{0{,}65 - 0{,}7}{\sqrt{\frac{0{,}7 \times 0{,}3}{35}}}\right) = 1 - \Phi(-0{,}65) = \Phi(0{,}65) = 0{,}7422.$$

ProbabilidadesEstatistica2019.pdf

Exercício: Proporção Amostral

b) $n = ?$

$$P(\overline{P} \geq 0{,}65) \geq 0{,}8 \Leftrightarrow 1 - P(\overline{P} < 0{,}65) \geq 0{,}8 \Leftrightarrow P\left(Z < \frac{0{,}65 - 0{,}7}{\sqrt{\frac{0{,}7 \times 0{,}3}{n}}}\right) \leq 0{,}2 \Leftrightarrow \Phi(-0{,}11\sqrt{n}) \leq 0{,}2$$

$$\Leftrightarrow \Phi(0{,}11\sqrt{n}) \geq 0{,}8$$

Como $\Phi(0{,}84) \approx 0{,}8$

$$\Leftrightarrow 0{,}11\sqrt{n} \geq 0{,}84 \Leftrightarrow \sqrt{n} \geq 7{,}6364 \Rightarrow n \geq 58{,}3 \Rightarrow n \geq 59.$$

ProbabilidadesEstatistica2019.pdf

Diferença de Proporções Amostrais

Distribuições de Amostragem

5

Diferença de Proporções Amostrais

Sejam $X_{11}, X_{12}, \ldots, X_{1n_1}$ e $X_{21}, X_{22}, \ldots, X_{2n_2}$ duas a. a. independentes, de dimensão n_1 e n_2 retiradas de duas populações Bernoulli, com parâmetros p_1 e p_2, respetivamente, em que X_{ij} toma o valor 1 se for um sucesso e o valor 0 se for um insucesso, e

$$\overline{P}_1 = \sum_{i=1}^{n_1} \frac{X_{1i}}{n_1} \quad \text{e} \quad \overline{P}_2 = \sum_{i=1}^{n_2} \frac{X_{2i}}{n_2}.$$

Se as dimensões das amostras forem grandes, então pelo Teorema de Moivre Laplace tem-se:

$$\overline{P}_1 \overset{\circ}{\sim} N\left(p_1; \sqrt{\frac{p_1(1-p_1)}{n_1}}\right) \quad \text{e} \quad \overline{P}_2 \overset{\circ}{\sim} N\left(p_2; \sqrt{\frac{p_2(1-p_2)}{n_2}}\right),$$

ProbabilidadesEstatistica2019.pdf

Diferença de Proporções Amostrais

donde pelo Teorema da aditividade da distribuição Normal vem:

$$\overline{P}_1 - \overline{P}_2 \overset{\circ}{\sim} N\left(p_1 - p_2; \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right),$$

ou seja,

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

ProbabilidadesEstatistica2019.pdf

Diferença de Proporções Amostrais: Exercícios

6

Distribuições de Amostragem

A proporção de clientes que optaram pela marca de telemóveis *Noko* na loja *TeleMN* foi 0,35 e na loja *Optcel* foi 0,29. Calcule a probabilidade de, recolhendo uma amostra de 200 clientes na primeira loja e de 150 clientes na segunda, a proporção amostral de clientes que optaram pela marca *Noko* na loja *TeleMN* ser superior à da loja *Optcel*.

ProbabilidadesEstatistica2019.pdf

Exercício: Diferença de Proporções Amostrais

Sejam:

- X_{1i} a v. a. que designa se o cliente i optou pela marca *Noko* na loja *TeleMN*, $i = 1, \dots, n_1$.
- X_{2i} a v. a. que designa se o cliente i optou pela marca *Noko* na loja *Optcel*, , $i = 1, \dots, n_2$.
- $\overline{P}_1$ a v. a. que representa a proporção de clientes que optaram pela marca *Noko* na loja *TeleMN*, em n_1 clientes
- $\overline{P}_2$ a v. a. que representa a proporção de clientes que optaram pela marca *Noko* na loja *Optcel*, em n_2 clientes.

$p_1 = 0{,}35$ e $p_2 = 0{,}29$.

X_1 e X_2 dist. Bernoulli
n_1 $(= 200)$ e n_2 $(= 150)$ grandes

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

$$P\left(\overline{P}_1 > \overline{P}_2\right) = P\left(\overline{P}_1 - \overline{P}_2 > 0\right) = 1 - P\left(\overline{P}_1 - \overline{P}_2 \leq 0\right) = 1 - P\left(Z \leq \frac{0 - (0{,}35 - 0{,}29)}{\sqrt{\frac{0{,}35(1-0{,}35)}{200} + \frac{0{,}29(1-0{,}29)}{150}}}\right)$$

$$= 1 - \Phi(-1{,}2) = 1 - \left(1 - \Phi(1{,}2)\right) = \Phi(1{,}2) = 0{,}8849.$$

ProbabilidadesEstatistica2019.pdf

Variância Amostral

Distribuições de Amostragem

7

Variância Amostral

Seja $X_1, X_2, \ldots, X_n$, uma a. a. duma dada população com variância σ^2. A **variância amostral** é definida por:

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}.$$

Propriedades:

- $\mu_{S^2} = E(S^2) = \sigma^2$.
- Se a distribuição da população for Normal então $\sigma_{S^2}^2 = Var(S^2) = \frac{2\sigma^4}{n-1}$.

Se a distribuição da população for Normal, com variância σ^2, então:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

ProbabilidadesEstatistica2019.pdf

Variância Amostral: Exercícios

Distribuições de Amostragem

8

Uma determinada empresa farmacêutica lançou no mercado um novo medicamento, para dormir, que tem estado a ser utilizado nos hospitais. Constatou-se que os doentes sujeitos a este medicamento em média dormiam 8 horas, sendo o desvio padrão de 2 horas, e que a distribuição do número de horas de sono podia ser considerada Normal.

Qual a probabilidade de, numa amostra aleatória de 31 doentes sujeitos ao referido medicamento:

a) A variância amostral ser superior a 5 horas?

b) A variância amostral ser inferior a 2,25 horas?

ProbabilidadesEstatistica2019.pdf

Exercício: Variância Amostral

Seja X a v.a. que representa o número de horas de sono dos doentes sujeitos ao medicamento, com $X \sim N(\mu = 8;\ \sigma = 2)$.

$$\left.\begin{array}{l} X \text{ dist. Normal} \\ n = 31 \end{array}\right| \Rightarrow \chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1=30}.$$

a) $P(S^2 > 5) = 1 - P(S^2 \leq 5) = 1 - P\left(\chi^2 \leq \frac{(31-1)\times 5}{2^2}\right) = 1 - P(\chi^2 \leq 37{,}5) \approx 1 - 0{,}84 = 0{,}16.$

b) $P(S^2 < 2{,}25) = 1 - P\left(\chi^2 < \frac{(31-1)\times 2{,}25}{2^2}\right) = P(\chi^2 \leq 16{,}875) \approx 0{,}026.$

ProbabilidadesEstatistica2019.pdf

Obrigada!

Questões?