Estatística II

Licenciatura em Gestão do Desporto

2.º Ano/2.º Semestre

2024/2025

Aulas Teórico-Práticas N.º 5 e 6 (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 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.

Proporção Amostral

Distribuições de Amostragem

1

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:

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

ProbabilidadesEstatistica2019.pdf

Proporção Amostral

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

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

pelo que

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

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

2

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, $i = 1, \ldots, n$,
- $\overline{P}$ a v.a. que representa a proporção de contribuintes que pagam o IUC dentro do prazo, em n contribuintes.

$p = 0{,}7.$

a)

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

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

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

Murteira et al (2015)
Capítulo 6

19. Da experiência passada apurou-se que em 5% das declarações de IRS entregues constam deduções ilegais. Para efeitos de controlo foram examinadas 1000 declarações escolhidas casualmente de entre todas as entregues. Supondo que se mantém o padrão de anos anteriores, calcule a probabilidade de pelo menos 60 terem esse tipo de ilegalidade.

Exercício 19

$$X = \begin{cases} 1, & \text{se IRS tem deduções ilegais} \\ 0, & \text{caso contrário} \end{cases} \Rightarrow X \sim B(1,\theta), \text{ onde } \theta = 0.05$$

$X \sim B(1, 0.05)$

Amostra casual: $(X_1, X_2, \dots, X_{1000})$, $n = 1000$

Quer-se: $P\left(\sum_{i=1}^{1000} X_i \geq 60\right)$, onde $\sum_{i=1}^{1000} X_i \sim B(1000, 0.05)$

Resultado exacto

$$P\left(\sum_{i=1}^{1000} X_i \geq 60\right) = 1 - P\left(\sum_{i=1}^{1000} X_i \leq 59\right) \approx 0.08673$$

Exercício 19

Resultado Aproximado pelo TLC: $Z = \dfrac{\overline{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \overset{\circ}{\sim} N(0;1).$

$$P(\bar{P} \geq 60/1000) = P\left(Z \geq \frac{0{,}06 - 0{,}05}{\sqrt{\frac{0{,}05 \times (1 - 0{,}05)}{1000}}}\right) = 1 - \Phi(1{,}451) \sim 1 - 0{,}9265 = 0{,}0735$$

Murteira et al (2015)
Capítulo 6

24. Num clube desportivo, a proporção de adeptos com opinião favorável sobre a direcção é de 75%.

a) Em 1000 adeptos seleccionados casualmente qual é, aproximadamente, a probabilidade de se observarem menos de 720 com opinião favorável à direcção?

b) Qual deverá ser a dimensão mínima de uma amostra casual de adeptos para que o desvio entre a frequência relativa da amostra e a verdadeira proporção de adeptos favoráveis à direcção não atinja 0.02 em pelo menos 95% dos casos?

Exercício 24 a)

Resultado Aproximado pelo TLC: $Z = \dfrac{\overline{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \overset{\circ}{\sim} N(0;1).$

$$\mathrm{P}(\bar{P} < 720/1000) = \mathrm{P}\left(\mathrm{Z} < \frac{0{,}72 - 0{,}75}{\sqrt{\frac{0{,}75 \times (1-0{,}75)}{1000}}}\right) = \Phi(-2{,}191)$$

$$= 1 - \Phi(2{,}191) \sim 1 - 0{,}9857 = 0{,}0143$$

Exercício 24 (b)

Quer-se o menor n tal que $P(|\bar{X}-\theta| < 0.02) \geq 0.95$

Pelo TLC: $Z = \dfrac{\bar{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \overset{a}{\sim} N(0,1)$

Logo,

$P(|\bar{X}-\theta| < 0.02) \geq 0.95 \Rightarrow P(-0.02 < \bar{X}-\theta < 0.02) \geq 0.95 \Rightarrow$

Exercício 24 (b)

$$\Rightarrow P\left(-\frac{0.02}{\sqrt{\frac{0.75\times0.25}{n'}}} < \frac{\bar{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n'}}} < \frac{0.02}{\sqrt{\frac{0.75\times0.25}{n'}}}\right) \geq 0.95 \Rightarrow$$

$$\Rightarrow P\left(-\frac{0.02\sqrt{n'}}{\sqrt{0.1875}} < Z < \frac{0.02\sqrt{n'}}{\sqrt{0.1875}}\right) \geq 0.95 \Rightarrow$$

$$\Rightarrow \Phi\left(\frac{0.02\sqrt{n'}}{\sqrt{0.1875}}\right) - \Phi\left(-\frac{0.02\sqrt{n'}}{\sqrt{0.1875}}\right) \geq 0.95 \Rightarrow$$

$$\Rightarrow \Phi\left(\frac{0.02\sqrt{n'}}{\sqrt{0.1875}}\right) - \left[1 - \Phi\left(\frac{0.02\sqrt{n'}}{\sqrt{0.1875}}\right)\right] \geq 0.95 \Rightarrow$$

$$\Rightarrow 2\,\Phi\left(\frac{0.02\sqrt{n'}}{\sqrt{0.1875}}\right) - 1 \geq 0.95 \Rightarrow \Phi\left(\frac{0.02\sqrt{n'}}{\sqrt{0.1875}}\right) \geq 0.975 \Rightarrow$$

$$\Rightarrow \frac{0.02\sqrt{n'}}{\sqrt{0.1875}} \geq 1.96 \Rightarrow \sqrt{n'} \geq \frac{1.96\sqrt{0.1875}}{0.02} \Rightarrow$$

$$\Rightarrow n' \geq \left(\frac{1.96\sqrt{0.1875}}{0.02}\right)^2 = 1800.75 \longrightarrow \boxed{n' = 1801}$$

Diferença de Proporções Amostrais

Distribuições de Amostragem

3

Diferença de Proporções 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 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,

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

ProbabilidadesEstatistica2019.pdf

Diferença de Proporções Amostrais: Exercícios

Distribuições de Amostragem

4

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*, $i = 1, \ldots, n_1$.
- X_{2i} a v. a. que designa se o cliente i optou pela marca *Noko* na loja *Optcel*, , $i = 1, \ldots, n_2$.
- $\overline{P}_1$ 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

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

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

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

ProbabilidadesEstatistica2019.pdf

Murteira et al (2015)
Capítulo 6

27. Depois de uma vigorosa campanha publicitária a quota de mercado das batatas fritas "As Estaladiças" passou de 8% para 10%. Suponha que se tinham realizados 2 inquéritos por amostragem, um antes de se iniciar a campanha (amostra de dimensão 100) e outro duas semanas depois do final da campanha (amostra de dimensão 300).

   a) Qual a probabilidade de se concluir, recorrendo aos referidos inquéritos, que o ganho de quota de mercado tinha sido superior a 5 pontos percentuais?

   b) Qual a probabilidade de os inquéritos concluírem por uma perda de quota de mercado?

Exercício 27 a)

Antes da campanha : $X_1 \sim B(1, 0.08) \longrightarrow$ Amostra : $m = 100$

Depois da campanha : $X_2 \sim B(1, 0.1) \longrightarrow$ Amostra : $n = 300$

Exercício 27 a)

(a)

quotas mercado nas amostras

Quer-se $P(\bar{X}_2 - \bar{X}_1 > 0.05) = P(\bar{X}_1 - \bar{X}_2 < -0.05)$

Sabe-se que: $Z = \dfrac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\dfrac{\theta_1(1-\theta_1)}{m} + \dfrac{\theta_2(1-\theta_2)}{n}}} \overset{a}{\sim} N(0,1)$, onde $\theta_1 = 0.08$, $\theta_2 = 0.1$, $m = 100$, $n = 300$

Logo,

$$P(\bar{X}_1 - \bar{X}_2 < -0.05) = P\left(\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{m} + \frac{\theta_2(1-\theta_2)}{n}}} < \frac{-0.05 - (0.08 - 0.1)}{\sqrt{\frac{0.08 \times 0.92}{100} + \frac{0.1 \times 0.9}{300}}}\right) =$$

$$= P\left(Z < \frac{-0.03}{\sqrt{0.001036}}\right) \approx \Phi(-0.93) = 1 - \Phi(0.93) \approx 1 - 0.8238 = 0.1762$$

Exercício 27 b)

$$P(\bar{X}_1 > \bar{X}_2) = P(\bar{X}_1 - \bar{X}_2 > 0) = 1 - P\left(\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{m} + \frac{\theta_2(1-\theta_2)}{n}}} \leq \frac{0 - (0.08 - 0.1)}{\sqrt{\frac{0.08 \times 0.92}{100} + \frac{0.1 \times 0.9}{300}}} \right) =$$

$$= \boxed{1 - P(\bar{X}_1 - \bar{X}_2 \leq 0)}$$

$$= 1 - P\left(Z \leq \frac{0.02}{\sqrt{0.001036}} \right) \approx 1 - \Phi(0.62) = 1 - 0.7324 = 0.2676$$

Distribuições do Mínimo e Máximo

Distribuições de Amostragem

5

Estatísticas de Ordem: Distribuição do Mínimo e Distribuição do Máximo

$$P(X_{(1)} \leq x) = 1- P(X_{(1)} > x) = 1- [P(X_1 > x)\times\ldots\times P(X_n > x)] = 1- [1- P(X < x)^n] = 1- [1-F(x)]^n$$

$$P(X_{(n)} \leq x) = P(X_1 \leq x)\times\ldots\times P(X_n \leq x) = P(X < x)^n = F(x)^n$$

Distribuições de Amostragem

Formulário

- **DISTRIBUIÇÃO DO MÍNIMO E DO MÁXIMO**

$G_1(x) = 1 - [1 - F(x)]^n \quad ; \quad G_n(x) = [F(x)]^n$

ProbabilidadesEstatistica2019.pdf

Distribuições do Mínimo e Máximo: Exercícios

Distribuições de Amostragem

6

Murteira et al (2015)
Capítulo 6

13. O tempo que um aluno despende por dia no *messenger* tem distribuição exponencial com média 2 horas. Seleccionaram-se 5 dias ao acaso tendo-se observado o tempo despendido no *messenger* em cada um deles.

a) Calcule a probabilidade do tempo médio despendido no *messenger*, por dia, ser superior a 4 horas?

b) Qual a probabilidade de o tempo máximo despendido por dia, não ultrapassar 6 horas?

Exercício 13 a)

X - v.a. tempo diário gasto no facebook , em horas

$X \sim Ex(1/2) = Ex(0.5)$, $\lambda = 0.5$

Amostra: $(X_1, X_2, X_3, X_4, X_5)$, $n = 5$, onde $X_i \sim Ex(0.5)$ $(i = 1, 2, 3, 4, 5)$

Tempo médio na amostra: $\dfrac{\sum_{i=1}^{5} X_i}{5} = \bar{X}$

Quer-se $P(\bar{X} > 4) = P\left(\sum_{i=1}^{5} X_i > 20\right)$

Como $X_i \sim Ex(0.5) \Rightarrow \sum_{i=1}^{5} X_i \sim G(5, 0.5) \Rightarrow \cancel{\underbrace{\dfrac{\sum_{i=1}^{5} X_i}{5}}_{\bar{X}} \sim G\left(5, \dfrac{0.5}{1/5}\right) \Leftrightarrow \bar{X} \sim G(5, \overset{\lambda^*}{\overset{\uparrow}{2.5}})} \not\Rightarrow$

$\cancel{\Rightarrow \underbrace{2\lambda^* \bar{X}}_{Q} \sim \chi^2(2 \times 5) \Leftrightarrow Q \sim \chi^2(10)}$ $\Rightarrow 2\lambda \sum_{i=1}^{5} X_i \overset{\cdot}{\sim} \chi^2(2 \times 5) \Leftrightarrow Q \overset{\cdot}{\sim} \chi^2(10)$

Assim, $P\left(\sum_{i=1}^{5} X_i > 20\right) = P\left(2\lambda \sum_{i=1}^{5} X_i > 2 \times 0.5 \times 20\right) = P(Q > 20) \approx 0.02925$

Distribuições: Exponencial, Gama e Qui-Quadrado

Formulário

- **EXPONENCIAL** $X \sim \mathrm{Ex}(\lambda)$, $(\lambda > 0)$; $X \sim \mathrm{Ex}(\lambda) \Leftrightarrow X \sim \mathrm{G}(1, \lambda)$

  $f(x \mid \lambda) = \lambda e^{-\lambda x}$, $x > 0$ $\quad E(X) = \dfrac{1}{\lambda}$; $\mathrm{Var}(X) = \dfrac{1}{\lambda^2}$; $\Im(\lambda) = 1/\lambda^2$

  Propriedades:

  - $X_i \sim \mathrm{Ex}(\lambda)$ ($i = 1, 2, \ldots, k$) independentes $\Rightarrow \sum_{i=1}^{k} X_i \sim G(k, \lambda)$ e $\min_i X_i \sim \mathrm{Ex}(k\lambda)$

- **GAMA** $X \sim G(\alpha, \lambda)$, $(\lambda > 0, \alpha > 0)$

  $f(x \mid \alpha, \lambda) = \dfrac{\lambda^{\alpha} e^{-\lambda x} x^{\alpha - 1}}{\Gamma(\alpha)}$, $\quad x > 0$; $E(X) = \dfrac{\alpha}{\lambda}$; $\mathrm{Var}(X) = \dfrac{\alpha}{\lambda^2}$; $\Im(\lambda) = \alpha / \lambda^2$, λ^2 conhecido

  Propriedades:

  - $X_i \sim G(\alpha_i; \lambda)$, ($i = 1, 2, \ldots, k$) independentes $\Rightarrow \sum_{i=1}^{k} X_i \sim \mathrm{G}\left(\sum_{i=1}^{k} \alpha_i ; \lambda\right)$
  - $X \sim G(\alpha, \lambda)$ então $Y = cX \sim G(\alpha, \dfrac{\lambda}{c})$ onde c constante positiva

- **QUI-QUADRADO** $X \sim \chi^2(n)$, (n inteiro positivo).

  $X \sim \chi^2(n) \Leftrightarrow X \sim \mathrm{G}(n/2; 1/2)$; $E(X) = n$; $\mathrm{Var}(X) = 2n$

  Propriedades:

  - $X_i \sim \chi^2_{(n_i)}$ ($i = 1, 2, \ldots, k$) independentes $\Rightarrow \sum_{i=1}^{k} X_i \sim \chi^2_{(n)}$ com $n = \sum_{i=1}^{k} n_i$
  - $X \sim \mathrm{G}(n; \lambda) \Leftrightarrow 2\lambda X \sim \chi^2(2n)$
  - $X_i \sim N(0,1)$, ($i = 1,2,\ldots,n$) independentes $\Rightarrow \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$

Exercício 13 b)

Quer-se $P[\max(X_i) \leq 6]$

<u>Distribuição do máximo</u>

$G_n(x) = P[\max(X_i) \leq x] = [F_X(x)]^n$, onde $n = 5$. Logo,

$P[\max(X_i) \leq 6] = G_5(6) = [F_X(6)]^5$

• $F_X(6) = P(X \leq 6) = \int_0^6 0.5\, e^{-0.5x}\, dx = \left[-e^{-0.5x}\right]_0^6 = 1 - e^{-3}$

Assim,

FMax(6) = (1- e^-3)^5 = (0,95)^5 = 0,774

Murteira et al (2015)
Capítulo 6

14. O tempo decorrido desde a participação da avaria até à reparação (tempo de reparação) de um certo tipo de máquina é uma variável aleatória com distribuição exponencial de média 4 horas.

a) Em dez avarias participadas qual a probabilidade de o menor dos tempos de reparação ser superior a 2 horas? E do maior dos tempos de reparação não ultrapassar as 8 horas?

b) Numa amostra casual de 40 avarias qual a probabilidade da média dos tempos de reparação ser inferior a 5.1 horas?

Exercício 14 a)

X - tempo de reparação, em horas $\rightarrow$ $X \sim Ex(1/4)$

$(X_1, X_2, \dots, X_{10})$

$P(X_{(1)} > 2)$

- $X_{(1)} = Y \sim Ex\left(10 \times \frac{1}{4}\right) = Ex(2.5)$

$$P(X_{(1)} > 2) = P(Y > 2) = \int_2^{+\infty} 2.5\, e^{-2.5y}\, dy = \left[-e^{-2.5y}\right]_2^{+\infty} = e^{-5} \approx 0.0067$$

Outra forma de resolver, muito melhor:

- $G_1(x) = 1 - [1 - F(x)]^n$

$$P(X_{(1)} > 2) = 1 - P(X_{(1)} \leq 2) = 1 - G_1(2) = 1 - \left\{1 - [1 - F_X(2)]^{10}\right\} = [1 - F_X(2)]^{10}$$

$$\rightarrow F_X(2) = P(X \leq 2) = \int_0^2 \frac{1}{4} e^{-\frac{1}{4}x} dx = \left[-e^{-\frac{1}{4}x}\right]_0^2 = 1 - e^{-\frac{1}{2}}$$

Logo, $P(X_{(1)} > 2) = \left[1 - (1 - e^{-1/2})\right]^{10} = \left(e^{-\frac{1}{2}}\right)^{10} = e^{-5} \approx 0.0067$

Exercício 14 a)

$$P(X_{(10)} \leq 8) = G_{10}(8) = \left[F_x(8)\right]^{10}$$

$$\rightarrow F_x(8) = \int_0^8 \frac{1}{4} e^{-\frac{1}{4}} dx = \left[-e^{-\frac{1}{4}}\right]_0^8 = 1 - e^{-2}$$

Logo,

$$P(X_{(10)} \leq 8) = (1 - e^{-2})^{10} \approx 0.2336$$

Exercício 14 b)

Amostra : $(X_1, X_2, \ldots, X_{40})$

$$P(\bar{X} < 5.1) = P\left(\sum_{i=1}^{40} X_i < 5.1 \times 40\right) = P\left(\sum_{i=1}^{40} X_i < 204\right) =$$

$$= P\left(\underbrace{2\lambda \sum_{i=1}^{40} X_i}_{\chi^2(2\times 40)} < 2 \times \frac{1}{4} \times 204\right) = P\left(\chi^2(80) < 102\right) \approx 0.9508$$

Distribuições de Amostragem: Resumo

7

Distribuições de Amostragem

Formulário

- **POPULAÇÕES NORMAIS**

Média	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\overline{X}-\mu}{S'/\sqrt{n}} \sim t(n-1)$
Diferença de médias	$\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \sim N(0,1)$	$Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1'^2}{m}+\frac{S_2'^2}{n}}} \overset{a}{\sim} t(\nu)$
	$T=\frac{\frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{1}{m}+\frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2+(n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	onde ν é o maior inteiro contido em r, $r=\frac{\left(\frac{s_1'^2}{m}+\frac{s_2'^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{s_1'^2}{m}\right)^2+\frac{1}{n-1}\left(\frac{s_2'^2}{n}\right)^2}$
Variância	$\frac{nS^2}{\sigma^2}=\frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	
Relação de variâncias	$\frac{S_1'^2}{S_2'^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1,n-1)$	

Distribuições de Amostragem

Formulário

- **GRANDES AMOSTRAS**

Caso geral

Média	$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0,1)$	$\dfrac{\overline{X}-\mu}{S'/\sqrt{n}} \overset{a}{\sim} N(0,1)$
Diferença de médias	$\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} \overset{a}{\sim} N(0,1)$	$\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1'^2}{m}+\dfrac{S_2'^2}{n}}} \overset{a}{\sim} N(0,1)$

População de Bernoulli

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

8 Estimação Pontual

Estimadores e Estimativas

Estatística

i) **ESTATÍSTICA**: uma estatística é uma função (qualquer) que depende da amostra e não depende de nenhum parâmetro desconhecido.

Exemplo: $X \sim \text{Poisson}(\lambda)$ $\qquad \lambda = ?$

a.a. $(X_1, X_2, \ldots, X_{10})$

$$T_1 = \bar{X} = \frac{1}{10} \sum_{i=1}^{n} X_i$$

$$T_2 = 3$$

$$T_3 = \frac{1}{10} \sum_{i=1}^{10} (X_i - \lambda)^2$$

Quais destas funções são estatísticas?

Slides Professora Claúdia Nunes

Estatística

i) ESTATÍSTICA: uma estatística é uma função (qualquer) que depende da amostra e não depende de nenhum parâmetro desconhecido.

Exemplo: $X \sim Poisson(\lambda)$ $\qquad \lambda = ?$

a.a. $(X_1, X_2, \ldots, X_{10})$

$T_1 = \bar{X} = \frac{1}{10} \sum_{i=1}^{n} X_i$: É UMA ESTATÍSTICA

$T_2 = 3$: É UMA ESTATÍSTICA

$T_3 = \frac{1}{10} \sum_{i=1}^{10} (X_i - \lambda)^2$ NÃO É ESTATÍSTICA!!

Slides Professora Claúdia Nunes

Estatística

DEF. 4: ESTATÍSTICA:

Uma estatística é uma função das variáveis observáveis, e é por si própria uma variável observável que não contém qualquer parâmetro desconhecido.

EXEMPLO 5: Seja a a.a. $[X_1, X_2, \ldots, X_n]$. A média amostral $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ é uma estatística.

EXEMPLO 6: Seja a v.a $X \sim N(\mu, \sigma^2)$ com μ e σ^2 desconhecidos, então $x - \mu$ não é estatística porque não é observável, é função do parâmetro desconhecido μ.

EXEMPLO 6: Seja uma v.a. com distribuição $N(\mu;\sigma^2)$ Quais são Estatísticas?

a) $X^2-\mu$

b) $\dfrac{X}{\sigma^2}$

c) X^2-3

d) $X-4$

e) $X - \log X^3$

Quais destas funções são estatísticas?

apostila.PDF (ufpr.br)

Estimador

ii) ESTIMADOR : é uma estatística que toma valores no espaço do parâmetro desconhecido

Exemplo: $X \sim Ber(p)$ $\quad p = ?$ $\quad p \in [0,1]$

a.a. $(X_1, \ldots, X_{20})$

$$T_1 = \sum_{i=1}^{20} X_i$$

$$T_2 = \frac{1}{20} \sum_{i=1}^{20} X_i$$

$$T_3 = \frac{\bar{X} - p}{1 - p}$$

Quais destas funções são estimadores do parâmetro p?

Slides Professora Claúdia Nunes

Estimador

ii) <u>ESTIMADOR</u> : é uma estatística que toma valores no espaço do parâmetro desconhecido

Exemplo: $X \sim Ber(p)$ $\quad p = ?$ $\quad p \in [0,1]$

a.a. $(X_1, \ldots, X_{20})$

$$T_1 = \cancel{\sum_{i=1}^{20}} X_i \in \{0,1,\ldots,20\} \not\subset [0,1]$$

$$T_2 = \frac{1}{20} \sum_{i=1}^{20} X_i = \bar{X} \in [0,1]$$

$$T_3 = \cancel{\frac{\bar{X} - p}{1-p}}$$

não é estimador porque nem sequer é estatística

Slides Professora Claúdia Nunes

Estimador vs. Estimativa

Dada uma amostra aleatória $(X_1, X_2, ..., X_n)$ chama-se **estatística** a uma função da amostra aleatória que não envolve parâmetros desconhecidos.

Definição

Chama-se **estimador** de $\underline{\theta}$ a uma estatística que a cada amostra observada faz corresponder um valor que estima $\underline{\theta}$, a que chamamos uma **estimativa** de $\underline{\theta}$.

$$\Theta^*(X_1, X_2, ..., X_n) \quad \text{é um estimador}$$
$$\theta^*(x_1, x_2, ..., x_n) \quad \text{é uma estimativa}$$

Não confundir

– estimador– variável aleatória

– estimativa–valor aproximado do parâmetro, obtido pelo estimador usando uma amostra particular

moduloI_aula3_4_Estimação.pdf

Definição: Um estimador dum parâmetro da população é uma variável aleatória (v. a.) que depende da informação amostral e cujas realizações fornecem aproximações para o parâmetro desconhecido. A um valor específico assumido por este estimador para uma amostra em concreto chama-se **estimativa**.

Estimador vs. Estimativa

Exemplo: São estimadores

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

Observada, por exemplo, a amostra $(1, 2, 0, 3, 1, 5)$

As estimativas associadas àqueles estimadores são:

i) $\overline{x} = \frac{1}{6}\sum_{i=1}^{6} x_i = 2$ e ii) $s^2 = \frac{1}{5}\sum_{i=1}^{6}(x_i - \overline{x})^2 = 3.2$

Consideremos o caso de termos uma dada população com distribuição $F(x|\theta)$, com um parâmetro desconhecido θ.

Como se podem definir vários estimadores de um parâmetro, põe-se o problema de escolher, se possível "o melhor". Há então que considerar certas propriedades que um estimador deve verificar.

modulol_aula3_4_Estimação.pdf

Conceitos Básicos

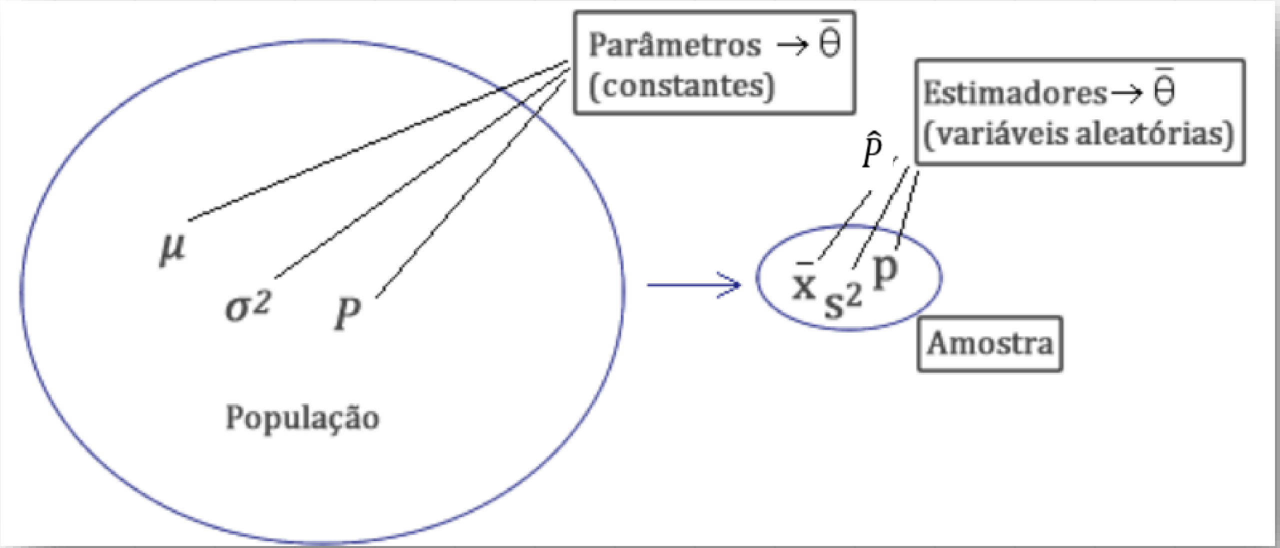
Parâmetro é uma quantidade numérica, em geral desconhecida, que descreve uma característica da população. Normalmente é representado por letras gregas como μ e σ, entre outras.

Estimador é uma função dos valores da amostra que utilizamos para estimar um parâmetro populacional. Os estimadores, em geral, são representados por letras gregas com acento circunflexo: $\hat{\mu}$, $\hat{\sigma}$, etc.

Estimativa é o valor numérico obtido através do estimador.

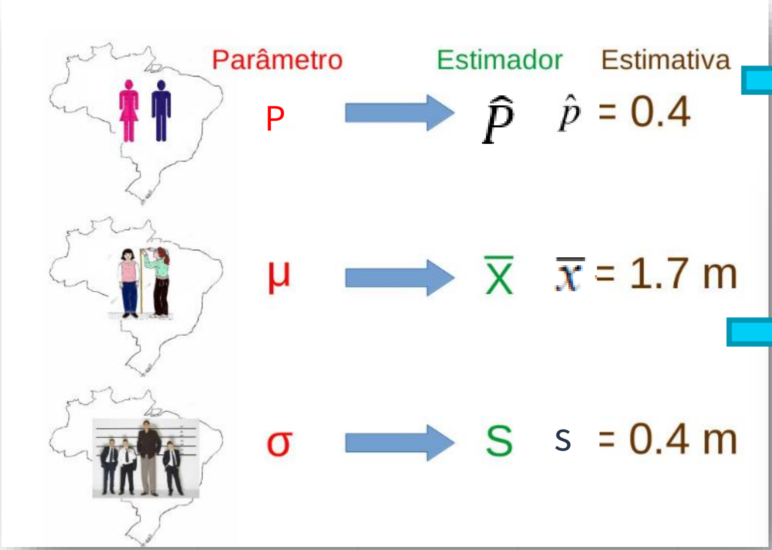
https://slideplayer.com.br/slide/2870987

Estimadores vs. Estimativas



http://tics.ifsul.edu.br/matriz/conteudo/disciplinas/est/uk/1/3.html

Estimadores vs. Estimativas



https://me323-unicamp.github.io/aulas/slides/parte10/parte10.html

$$\hat{p} = \frac{x}{n}$$

Sendo X o número de elementos da amostra (n) que apresenta a característica de estudo;

A estimativa do parâmetro da proporção populacional (P) é dada por esta fórmula; e o respetivo estimador é X/n.

Parâmetro Desconhecido	Estatística Estimador	Estimativa Pontual
μ	$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$	$\bar{x}$
σ^2	$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$	s^2

https://sites.google.com/site/estatisticabasicacc/conteudo

Variância corrigida (s^2)	Variância não corrigida ($s^{2`}$)
$\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$	$\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$

Resumo dos Principais Estimadores e Estimativas em estudo...

Parâmetro a estimar	Estimador	Estimativa
μ	$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
σ^2	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$
p	$\widehat{P} = \frac{X^{(a)}}{n}$	$\hat{p} = \frac{x^{(b)}}{n}$
$\mu_1 - \mu_2$	$\overline{X_1} - \overline{X_2}$	$\overline{x_1} - \overline{x_2}$
σ_1^2 / σ_2^2	S_1^2 / S_2^2	s_1^2 / s_2^2
$p_1 - p_2$	$\widehat{P_1} - \widehat{P_2}$	$\hat{p}_1 - \hat{p}_2$

$^{(a)}$ X - v.a. que conta ... e ˜ $^{(b)}$ x - número observado de sucessos na amostra de dimensão n.

moduloI_aula3_4_Estimação.pdf

9

Método dos Momentos

Estimadores

Métodos de Estimação

Vamos então falar dos principais métodos de estimação paramétrica.

Dos **métodos de estimação paramétrica** vamos referir:

o Método dos momentos e
o Método da Máxima verosimilhança

modulol_aula3_4_Estimação.pdf

Métodos dos Momentos

Introduzido por Karl Pearson no início do século XX, foi o primeiro método de estimação a ser apresentado e que tem uma filosofia muito simples.

O método consiste em:

– considerar como estimadores dos parâmetros desconhecidos as soluções das equações que se obtêm igualando os momentos teóricos aos momentos empíricos.

Momentos empíricos ou Momentos amostrais

Momentos teóricos ou Momentos populacionais

É um método de aplicação geral, tendo como única condição que a distribuição tenha um número suficiente de momentos (teóricos).

moduloI_aula3_4_Estimação.pdf

Métodos dos Momentos

Sejam $\theta_1, ..., \theta_k$ parâmetros desconhecidos de uma v.a. X.
O método dos momentos consiste em igualar momentos teóricos e momentos empíricos, i.e.,

$$
\begin{aligned}
&E[X] = m_1' && \text{c/} && m_1' = \tfrac{1}{n}\textstyle\sum_{i=1}^n x_i \\
&E[X^2] = m_2' && \text{c/} && m_2' = \tfrac{1}{n}\textstyle\sum_{i=1}^n x_i^2 \\
&\vdots \\
&E[X^k] = m_k' && \text{c/} && m_k' = \tfrac{1}{n}\textstyle\sum_{i=1}^n x_i^k
\end{aligned}
$$

$m_k' = \frac{1}{n}\sum_{i=1}^n x_i^k$ são os chamados momentos empíricos, calculados à custa da amostra $(x_1, ..., x_n)$.

Aquelas igualdades dão-nos estimativas que são concretização dos estimadores com as expressões correspondentes.

moduloI_aula3_4_Estimação.pdf

Métodos dos Momentos: Exemplo

Considere-se X ∩ N(μ, σ^2). Quais são os estimadores dos momentos de μ e σ^2?

Métodos dos Momentos: Exemplo

Exemplo

Consideremos $X \frown \mathcal{N}(\mu, \sigma)$. Quais os estimadores de momentos de μ e σ^2?

Tem-se :

$$\begin{array}{lcl} E[X] = \mu & \text{e} & M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \\ E[X^2] = \sigma^2 + \mu^2 & \text{e} & M_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{array}$$

$$\begin{cases} \mu = \frac{1}{n}\sum_{i=1}^{n} x_i \\ \sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \end{cases} \Leftrightarrow \begin{cases} \mu = \overline{x} \\ \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \end{cases}.$$

donde

$$\mu^* = \overline{X}$$

$$(\sigma^2)^* = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mu^{*2} \Rightarrow (\sigma^2)^* = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

moduloI_aula3_4_Estimação.pdf

Portanto, os estimadores são:

$$\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \left(\frac{n-1}{n}\right) S^2 \end{cases}.$$

Variância não corrigida (s^2)

Variância corrigida (s^2)

$$\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}$$

Métodos dos Momentos

Os cálculos não são complicados, mas no cálculo dos momentos empíricos aparecem potências de expoente elevado quando há muitos parâmetros, conduzindo a estimativas instáveis.
Por isso, como regra prática deve evitar-se recorrer ao método dos momentos para mais de quatro parâmetros.

Nota: Os estimadores obtidos pelo método dos momentos são menos eficientes do que os estimadores de máxima verosimilhança, que passamos já a apresentar.

moduloI_aula3_4_Estimação.pdf

10

Método dos Momentos: Exercícios

Estimadores

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

Exercício 1: Suponha que $X1, X2, ., Xn$ é uma a.a. proveniente de uma distribuição exponencial com o parâmetro λ. Determine o estimador dos momentos de λ.

Exercício 1: Estimador dos Momentos

$$f(x) = \lambda e^{-\lambda x}, 0 \leq x < \infty$$

$$\mu = E(X) = \frac{1}{\lambda}$$

1º Momento da Amostra: $\frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$

1º Momento da População: $E[x] = \frac{1}{\lambda}$

Método dos Momentos: $E[x] = \bar{x}$

$$\frac{1}{\hat{\lambda}_{MM}} = \bar{x}$$

$$\hat{\lambda}_{MM} = \frac{1}{\bar{x}}$$

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

Exercício 2: Seja X1, X2, ., Xn uma a.a. que segue uma v.a. com função densidade de probabilidade dada por

$$f(x|\alpha) = (\alpha + 1) \cdot x^{\alpha} \, I_{(0,1)}(x) \quad \alpha > 0.$$

Encontre um estimador para α utilizando o método dos momentos.

Exercício 2: Estimador dos Momentos

solução pelo método dos momentos: $\mu_1' = m_n$

$$\mu_1' = E(X) = \int_0^1 x \cdot f(x|\alpha)\,dx = (\alpha+1)\cdot\int_0^1 x^{\alpha+1}\,dx = (\alpha+1)\cdot\left.\frac{x^{\alpha+2}}{\alpha+2}\right|_0^1$$

Logo, $\mu_2' = \boxed{\dfrac{\alpha+1}{\alpha+2}}$

Tabela 1.1: Tabela de Primitivas Elementares

f	$P f = F$
$c, c \in IR$	$c\,x$
$x^\alpha \ (\alpha \neq -1)$	$\dfrac{x^{\alpha+1}}{\alpha+1}$
$\dfrac{1}{x}$	$\log\lvert x\rvert$
e^x	e^x
$\cos x$	$\sin x$
$\sin x$	$-\cos x$
$\sec^2 x$	$\mathrm{tg}x$
$\mathrm{cosec}^2 x$	$-\mathrm{cotg}x$
$\dfrac{1}{1+x^2}$	$\mathrm{arctg}x$
$\dfrac{1}{\sqrt{1-x^2}}$	$\arcsin x$
$\cosh x$	$\sinh x$
$\sinh x$	$\cosh x$

[Tópico_08_.pdf (usp.br)]

Exercício 2: Estimador dos Momentos

Assim,

$$m_1 = \mu_1' \implies \frac{\widehat{\alpha}+1}{\widehat{\alpha}+2} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

Segue que,

$$\begin{aligned}
\widehat{\alpha}+1 &= (\widehat{\alpha}+2)\cdot\overline{X} \\
\widehat{\alpha}+1 &= \widehat{\alpha}\,\overline{X}+2\,\overline{X} \\
\widehat{\alpha}\,(1-\overline{X}) &= 2\,\overline{X}-1
\end{aligned}$$

Portanto,

$$\widehat{\alpha} = \frac{2\,\overline{X}-1}{1-\overline{X}}$$

é o estimador para α pelo métodos dos momentos.

Obrigada!

Questões?