Estatística II

Licenciatura em Gestão do Desporto

2.º Ano/2.º Semestre

2023/2024

Aulas Teórico-Práticas N.º 8 e 9 (Semana 5)

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.

Distribuições de Amostragem e TLC: Exercícios

Murteira et al (2015)

1

2. O número de gralhas por página, em certo tipo de publicações, é uma variável aleatória com distribuição de Poisson cuja média está estimada em 0.3. Supõe-se que existe independência entre o número de gralhas em páginas diferentes.
   a) Numa amostra de cinco páginas, qual a probabilidade de as duas primeiras terem uma gralha cada, e de as restantes não terem gralhas?
   b) Se a amostra casual for de 20 páginas, calcule a probabilidade de o número total de gralhas encontrado ser de pelo menos 8.
   c) Para uma amostra de 50 páginas, obtenha o valor esperado e a variância da média de gralhas nessa amostra.
   d) Voltando às amostras da alínea a), calcule e interprete $P\{\max(X_i) \leq 1\}$.
   e) Numa publicação do tipo apresentado com 100 páginas, qual a probabilidade de pelo menos 80 delas não terem qualquer gralha?

Exercício 2 a)

X - v.a. nº gralhas por página $\rightarrow X \sim Po(0.3)$, $\lambda = 0.3$

(a)

Amostra : $n=5$, $(X_1, X_2, X_3, X_4, X_5)$, onde $X_i \sim Po(0.3)$ $(i=1,2,3,4,5)$

$P(X_1=1, X_2=1, X_3=0, X_4=0, X_5=0)$

$$P = P(X_1=1)\, P(X_2=1)\, P(X_3=0)\, P(X_4=0)\, P(X_5=0) = [P(X=1)]^2\, [P(X=0)]^3 \approx 0.02008$$

Exercício 2 b)

(b)

Amostra: $n = 20$, $(X_1, X_2, \dots, X_{20})$, onde $X_i \sim Po(0.3)$ $(i = 1, 2, \dots, 20)$.

Nº total de gralhas na amostra (20 páginas): $\sum_{i=1}^{20} X_i \sim Po(20 \times 0.3) = Po(6)$

$$P\left(\sum_{i=1}^{20} X_i \geq 8\right) = 1 - P\left(\sum_{i=1}^{20} X_i < 8\right) = 1 - P\left(\sum_{i=1}^{20} X_i \leq 7\right) \approx 1 - 0.74398 = 0.25602$$

Exercício 2 c)

Amostra: $n = 50$, $(X_1, X_2, \dots, X_{50})$, onde $X_i \sim Po(0.3)$ $(i = 1, 2, \dots, 50)$

Média de gralhas na amostra : $\dfrac{\sum_{i=1}^{50} X_i}{50} = \bar{X}$

- $E(\bar{X}) = E\left(\dfrac{\sum_{i=1}^{50} X_i}{50}\right) = \dfrac{1}{50} E\left(\sum_{i=1}^{50} X_i\right) = \dfrac{1}{50} \sum_{i=1}^{50} \overbrace{E(X_i)}^{\lambda = 0.3} = \dfrac{1}{50} \sum_{i=1}^{50} 0.3 = \dfrac{1}{50} \times 50 \times 0.3 = 0.$

- $Var(\bar{X}) = Var\left(\dfrac{\sum_{i=1}^{50} X_i}{50}\right) = \left(\dfrac{1}{50}\right)^2 Var\left(\sum_{i=1}^{50} X_i\right) \overset{X_i \text{ iid}}{=} \dfrac{1}{50^2} \sum_{i=1}^{50} \overbrace{Var(X_i)}^{\lambda = 0.3} = \dfrac{1}{50^2} \sum_{i=1}^{50} 0.3 =$

$= \dfrac{1}{50^2} \times 50 \times 0.3 = \dfrac{0.3}{50} = 0.006$

Exercício 2 d)

Amostra: $(X_1, X_2, X_3, X_4, X_5)$, $n = 5$

$P[\underbrace{\max(X_i)}_{X_{(n)}} \leq 1] = P(X_{(5)} \leq 1) \rightarrow$ Probabilidade de a página com mais gralhas na amostra não ter mais de 1 gralha

Distribuição do máximo da amostra

Sendo $G_n(x) = P[\max(X_i) \leq x]$ a função de distribuição do máximo, tem-se que:

$G_n(x) = [F(x)]^n$, onde $n = 5$. Então:

$$P[\max(X_i) \leq 1] = G_5(1) = [F_X(1)]^5 = [P(\underset{\downarrow \atop P_o(0.3)}{X} \leq 1)]^5 \approx 0.8285$$

Logo, na amostra, a probabilidade de a página com mais gralhas não ter mais que uma é cerca de 0.83.

Exercício 2 e)

Y - v.a. nº páginas sem gralhas num conjunto de 100

$Y \sim B(100, \theta)$, onde $\theta = P(X=0) \approx 0.74082 \longrightarrow Y \sim B(100, 0.74082)$

Quer-se: $P(Y \geq 80) = 1 - P(Y < 80) = 1 - P(Y \leq 79) \approx 1 - 0.89398 = 0.10602$

ou TLC ... $\quad Z = \dfrac{Y - n\theta}{\sqrt{n\theta(1-\theta)}} \overset{a}{\sim} N(0,1)$

$\ldots = 1 - P(Y \leq 79) = 1 - P\left(\dfrac{Y - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \dfrac{79 - 100 \times 0.74082 + \frac{1}{2}}{\sqrt{100 \times 0.74082 \times 0.25918}}\right) =$

$= 1 - P(Z \leq 1.24) = 1 - \Phi(1.24) = 1 - 0.8925 = 0.1075$

9. O tempo, em minutos, que uma pessoa demora a servir-se na cantina da faculdade é uma variável aleatória com função densidade

$$f(x) = e^{-x} \quad (x > 0).$$

Se às 14:00 se encontram 40 pessoas na fila do almoço dessa cantina, calcule a probabilidade de que nenhuma delas esteja por servir às 14:30.

Exercício 9

X - v.a. tempo atendimento por pessoa, em minutos

$f(x) = e^{-x} \quad (x > 0) \quad \rightarrow \quad X \sim Ex(1) \quad , \lambda = 1$

Amostra casual : $(X_1, X_2, \ldots, X_{40})$, $n = 40$, onde $X_i \sim Ex(1)$ $(i = 1, 2, \ldots, 40)$

Quer-se $P\left(\sum_{i=1}^{40} X_i \leq 30\right)$, onde $\sum_{i=1}^{40} X_i$ é o tempo total de atendimento das 40 pessoas

Como $X_i \sim Ex(1) \Rightarrow \sum_{i=1}^{40} X_i \sim G(40, 1) \Rightarrow \underbrace{2\lambda \sum_{i=1}^{40} X_i}_{Q} \sim \chi^2(2 \times 40) \Leftrightarrow Q \sim \chi^2(80)$

Assim,

$$P\left(\sum_{i=1}^{40} X_i \leq 30\right) = P\left(2\lambda \sum_{i=1}^{40} X_i \leq 2 \times 1 \times 30\right) = P(Q \leq 60) \approx 0.04625$$

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 $X_i \sim Ex(0.5) \Rightarrow \sum_{i=1}^{5} X_i \sim G(5, 0.5) \Rightarrow$ $\cancel{\underbrace{\frac{\sum_{i=1}^{5} X_i}{5}}_{\bar{X}} \sim G\left(5, \frac{0.5}{1/5}\right) \Leftrightarrow \bar{X} \sim G(5, \overset{\lambda^*}{2.5}) \Rightarrow}$

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

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

Exercício 13 b)

Quer-se $P[\max(X_i) \le 6]$ [illegible] ~~(X é v.a. contínua)~~

Distribuição do máximo

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

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

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

Assim,

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 $\rightarrow$ $X \sim Ex(1/4)$

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

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

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

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

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

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

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

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)

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

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

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

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

Exercício 14 b)

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

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

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

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 (Poisson)

$$\sum_{i=1}^{1000} X_i \overset{a}{\sim} Po(n\theta), \text{ onde } n = 1000 \text{ e } \theta = 0.05 \Rightarrow \sum_{i=1}^{1000} X_i \overset{a}{\sim} Po(50)$$

$$P\left(\sum_{i=1}^{1000} X_i \geq 60\right) = 1 - P\left(\sum_{i=1}^{1000} X_i \leq 59\right) \approx 0.09226$$

Abordando a distribuição binomial com a distribuição de Poisson

A distribuição de Poisson é uma boa aproximação à distribuição binomial desde que:

-O tamanho da amostra é grande: *n ≥ 100*

-A probabilidade *p é* pequena: *p ≤ 0,1*

– *μ* estar na ordem de: *np ≤ 10*

Resultado aproximado (TLC)

$$Z = \frac{\sum_{i=1}^{n} X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \overset{a}{\sim} N(0,1), \text{ onde } \theta = 0.05 \text{ e } n = 1000$$

- $X \sim Bin(n,p)$ pode ser aproximada por $\tilde{X} \sim N(np, np(1-p))$, a aproximação tem menor erro quando $np > 5$ e $n(1-p) > 5$
- $X \sim Poisson(\lambda)$ pode ser aproximada por $\tilde{X} \sim N(\lambda, \lambda)$, a aproximação tem menor erro quando $\lambda > 5$

$$P\left(\sum_{i=1}^{1000} X_i \geq 60\right) = P\left(\frac{\sum_{i=1}^{1000} X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \geq \frac{60 - 1000 \times 0.05 - 1/2}{\sqrt{1000 \times 0.05 \times 0.95}}\right) = 1 - \Phi\left(\frac{9.5}{\sqrt{47.5}}\right) \approx 0.08404$$

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)

$$X = \begin{cases} 1, & \text{se adepto é favorável} \\ 0, & \text{caso contrário} \end{cases} \quad \rightarrow \quad X \sim B(1, 0.75)$$

(a)

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

Quer-se $P\left(\sum_{i=1}^{1000} X_i < 720\right)$, onde $\sum_{i=1}^{1000} X_i \sim B(1000, 0.75)$

Resultado exacto (Binomial)

$$P\left(\sum_{i=1}^{1000} X_i < 720\right) = P\left(\sum_{i=1}^{1000} X_i \leq 719\right) \approx 0.0137 \quad \rightarrow \quad \text{binomcdf}(1000, 0.75, 719)$$

Exercício 24 a)

Resultado aproximado (TLC)

$$Z = \frac{\sum_{i=1}^{n} X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \overset{a}{\sim} N(0,1) \quad , \quad \text{onde } \theta = 0.75 \text{ e } n = 1000$$

$$P\left(\sum_{i=1}^{1000} X_i < 720\right) = P\left(\sum_{i=1}^{1000} X_i \leq 719\right) = P\left(\frac{\sum_{i=1}^{1000} X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{719 - 1000 \times 0.75 + 1/2}{\sqrt{1000 \times 0.75 \times 0.25}}\right) =$$

$$= P(Z \leq -2.23) = \Phi(-2.23) = 1 - \Phi(2.23) = 1 - 0.9871 = 0.0129$$

Exercício 24 b)

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

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

Logo,

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

Exercício 24 b)

$$\Rightarrow P\left(-\frac{0.02}{\sqrt{\frac{0.75\times 0.25}{n'}}} < \frac{\bar{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n'}}} < \frac{0.02}{\sqrt{\frac{0.75\times 0.25}{n'}}}\right) \geqslant 0.95 \Rightarrow$$

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

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

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

Exercício 24 b)

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

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

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

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 : $n = 100$

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)

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

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

30. De uma população normal de média 8 e desvio padrão 4 tomou-se uma amostra casual de dimensão 100.

   a) Qual a probabilidade de a média da amostra diferir da média da população em mais de 0.5?

   b) Se a população não fosse normal, qual seria essa probabilidade?

Exercício 30 a)

$X \sim N(8, 4^2) \longrightarrow$ Amostra casual : $n = 100$

(a)

$$P(|\bar{X} - \mu| > 0.5) = 1 - P(|\bar{X} - \mu| \leq 0.5) = 1 - P(-0.5 \leq \bar{X} - \mu \leq 0.5)$$

Sabe-se que : $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$. Logo,

$$1 - P(-0.5 \leq \bar{X} - \mu \leq 0.5) = 1 - P\left(-\frac{0.5}{4/\sqrt{100}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.5}{4/\sqrt{100}}\right) = 1 - \left[\Phi(1.25) - \Phi(-1.25)\right] =$$

$$= 1 - \left[2\,\Phi(1.25) - 1\right] = 2 - 2\,\Phi(1.25) \approx 0.2113$$

Exercício 30 b)

Se a população não fosse normal, recorrendo ao TLC, utilizar-se-ia o resultado:

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0,1).$$

Pelo que a probabilidade seria <u>aproximadamente</u> igual à calculada em (a).

37. De uma população normal de média e variância desconhecidas retirou-se uma amostra casual.

a) Determine a percentagem de amostras em que a sua média difere da média da população, por valores superiores ao desvio padrão da população, considerando que a amostra tem 4 observações.

b) Determine a percentagem de amostras de 5 observações em que as suas médias diferem da média da população, por valores superiores ao do desvio padrão da amostra.

Exercício 37 a)

$X \sim N(\mu, \sigma^2)$

(a)

Amostra casual : $n = 4$

Quer-se $P(|\bar{X} - \mu| > \sigma) = 1 - P(|\bar{X} - \mu| \leq \sigma) = 1 - P(-\sigma \leq \bar{X} - \mu \leq \sigma)$

Sabe-se que : $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

Logo,

$$1 - P(-\sigma \leq \bar{X} - \mu \leq \sigma) = 1 - P\left(-\frac{\sigma}{\sigma/\sqrt{4}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\sigma}{\sigma/\sqrt{4}}\right) = \underbrace{1 - P(-2 \leq Z \leq 2)}_{1 - [\Phi(2) - \Phi(-2)]} \approx$$

$$\approx 1 - 0.9545 = 0.0455$$

Exercício 37 b)

Amostra casual: $n = 5$

Quer-se $P\left(|\bar{X}-\mu| > S\right) = 1 - P\left(|\bar{X}-\mu| \leq S\right) = 1 - P\left(-S \leq \bar{X}-\mu \leq S\right)$

Sabe-se que: $T = \dfrac{\bar{X}-\mu}{S'/\sqrt{n}} = \dfrac{\bar{X}-\mu}{S/\sqrt{n-1}} \sim t(n-1) = t(4)$ $\quad \left(n S^2 = (n-1) S'^2\right)$

Logo,

$$1 - P\left(-S \leq \bar{X}-\mu \leq S\right) = 1 - P\left(-\frac{S}{S/\sqrt{4}} \leq \frac{\bar{X}-\mu}{S/\sqrt{n-1}} \leq \frac{S}{S/\sqrt{4}}\right) = 1 - P(-2 < T < 2)$$

$$\approx 1 - 0.88388 = 0.11612$$

Exercício 37 b)

Amostra: $n = 4$

$$P(|\bar{X} - \mu| > S') = 1 - P(|\bar{X} - \mu| \leq S') = 1 - P(-S' \leq \bar{X} - \mu \leq S')$$

Sabe-se que: $T = \dfrac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1) = t(3)$

Logo,

$$1 - P(-S' \leq \bar{X} - \mu \leq S') = 1 - P\left(-\frac{S'}{S'/\sqrt{4}} \leq \frac{\bar{X} - \mu}{S'/\sqrt{n}} \leq \frac{S'}{S'/\sqrt{4}}\right) = 1 - P(-2 < T < 2) \approx$$

$$\approx 1 - 0.86067 = 0.13933$$

39. De uma população normal de variância 64, tomou-se uma amostra de dimensão 3.

a) Qual a probabilidade de a variância da amostra exceder 78?

b) Responda à mesma pergunta para uma amostra de dimensão 16.

Exercício 39 a)

$X \sim N(\mu, 64) \longrightarrow$ Amostra : $n = 3$

(a)

Quer-se $P(S^2 > 78)$

Sabe-se que : $Q = \dfrac{n S^2}{\sigma^2} \sim \chi^2(n-1) = \chi^2(2)$, onde $n = 3$ e $\sigma^2 = 64$

Logo,

$$P(S^2 > 78) = P\left(\frac{n S^2}{\sigma^2} > \frac{3 \times 78}{64}\right) = P(Q > 3.65625) \approx 0.1607$$

Exercício 39 b)

Amostra : $n = 16$

$$Q = \frac{n s^2}{\sigma^2} \sim \chi^2_{(n-1)} = \chi^2_{(15)}$$

Logo,

$$P(s^2 > 78) = P\left(\frac{n s^2}{\sigma^2} > \frac{16 \times 78}{64}\right) = P(Q > 19.5) \approx 0.192$$

41. Um centro emissor de cartões de segurança tem dois equipamentos de personalização, de funcionamento independente. O tempo de processamento, em segundos, para cada um deles tem comportamento normal com o mesmo tempo médio, sendo o desvio padrão do primeiro 10, e do segundo 15. Considerando amostras de 16 cartões personalizados de cada um dos equipamentos,

a) Calcule a probabilidade de a diferença entre as médias das duas amostras, em valor absoluto, ser superior a 5.

b) Qual a probabilidade do desvio padrão dos tempos de processamento dos cartões da amostra do primeiro equipamento ser superior ao da amostra do segundo equipamento?

Exercício 41 a)

X_1 - v.a. tempo processamento do equipamento 1, em segundos

X_2 - v.a. tempo processamento do equipamento 2, em segundos

$X_1 \sim N(\mu_1, 10^2) \longrightarrow$ Amostra casual : $n = 16$

, onde $\mu_1 = \mu_2$

$X_2 \sim N(\mu_2, 15^2) \longrightarrow$ Amostra casual : $n = 16$

Exercício 41 a)

Quer-se: $P\left(|\bar{X}_1-\bar{X}_2|>5\right) = 1-P\left(|\bar{X}_1-\bar{X}_2|\leq 5\right) = 1-P\left(-5\leq \bar{X}_1-\bar{X}_2\leq 5\right)$

Sabe-se que: $Z = \dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \sim N(0,1)$. Logo,

$$1-P\left(-5\leq \bar{X}_1-\bar{X}_2\leq 5\right) = 1-P\left(\frac{-5-0}{\sqrt{\frac{10^2}{16}+\frac{15^2}{16}}}\leq \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\leq \frac{5-0}{\sqrt{\frac{10^2}{16}+\frac{15^2}{16}}}\right) =$$

$$= 1-P\Bigg(\underbrace{-\frac{5}{\sqrt{20.3125}}}_{(-1.11)}\leq Z\leq \underbrace{\frac{5}{\sqrt{20.3125}}}_{(1.11)}\Bigg) = 1-\left[\Phi(1.11)-\Phi(-1.11)\right] =$$

$$= 1-\left[2\Phi(1.11)-1\right] = 2-2\Phi(1.11) = 2-2\times 0.8665 = 0.267$$

Exercício 41 b)

Quer-se : $P(S_1 > S_2) = P(S_1^2 > S_2^2)$ $\longrightarrow$ Relação de variâncias

- $S_1^2 = \dfrac{(m-1)S_1'^2}{m} = \dfrac{15}{16} S_1'^2$

- $S_2^2 = \dfrac{(n-1)\, S_2'^2}{n} = \dfrac{15}{16} S_2'^2$

Exercício 41 b)

Então,

$$P(S_1^2 > S_2^2) = P\left(\frac{15}{16} S_1'^2 > \frac{15}{16} S_2'^2\right) = P(S_1'^2 > S_2'^2) = P\left(\frac{S_1'^2}{S_2'^2} > 1\right)$$

Sabe-se que: $F = \frac{S_1'^2}{S_2'^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$, onde: $\sigma_1^2 = 10^2 = 100$; $\sigma_2^2 = 15^2 = 225$;
$m = 16$ e $n = 16$

Logo,

$$P\left(\frac{S_1'^2}{S_2'^2} > 1\right) = P\left(\underbrace{\frac{S_1'^2}{S_2'^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}}_{F(15,15)} > 1 \cdot \frac{225}{100}\right) = P(F > 2.25) \approx 0.064$$

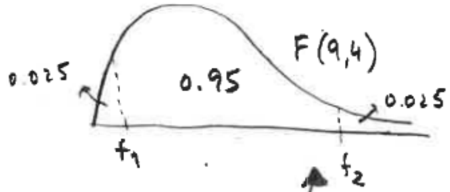
45. De duas populações normais independentes, com variâncias iguais, foram extraídas duas amostras casuais com dimensões 10 e 5, respectivamente. Determine os valores tais que entre eles esteja, com 95% de probabilidade, a razão das variâncias corrigidas das amostras.

Exercício 45

$X_1 \sim N(\mu_1, \sigma_1^2) \longrightarrow$ Amostra casual : $m = 10$

$X_2 \sim N(\mu_2, \sigma_2^2) \longrightarrow$ Amostra casual : $n = 5$

$\sigma_1^2 = \sigma_2^2$



Quer-se encontrar os valores f_1 e f_2, tais que: $P\left(f_1 \leq \frac{S_1'^2}{S_2'^2} \leq f_2\right) = 0.95$

Sabe-se que $F = \frac{S_1'^2}{S_2'^2} \cdot \underbrace{\frac{\sigma_2^2}{\sigma_1^2}}_{1} \sim F\,(m-1, n-1) = F\,(9,4)$ . . Logo,

$$P\left(f_1 < \frac{S_1'^2}{S_2'^2} < f_2\right) = 0.95 \Rightarrow P\left(f_1 \times 1 \leq \frac{S_1'^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leq f_2 \times 1\right) = 0.95 \Rightarrow P(f_1 < F < f_2) = 0.95$$

Exercício 45

$$\Rightarrow \begin{cases} P(F < f_1) = 0.025 \\ P(F > f_2) = 0.025 \end{cases} \Rightarrow \begin{cases} P\left(\frac{1}{F} > \frac{1}{f_1}\right) = 0.025 \text{ , onde } \frac{1}{F} \sim F(4,9) \\ f_2 = 8.90 \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \frac{1}{f_1} = 4.72 \\ '' \end{cases} \Rightarrow \begin{cases} f_1 = \frac{1}{4.72} \\ '' \end{cases} \Rightarrow \begin{cases} f_1 = 0.212 \\ f_2 = 8.90 \end{cases}$$

46. Admita a existência de duas populações nas quais são definidas as variáveis aleatórias X e Y, com distribuições normais, e tais que

$$E(X) = 20, \ E(Y) = 22 \ \text{e} \ \operatorname{Var}(X) = \operatorname{Var}(Y) = 16 .$$

Considere que se obtêm duas amostras casuais independentes, uma de cada população, com 9 e 16 elementos, respectivamente.

a) Qual a probabilidade de a média da segunda amostra exceder a média da primeira em mais de três unidades?

b) Qual a probabilidade de o desvio padrão corrigido da primeira amostra ultrapassar o dobro do desvio padrão corrigido da segunda amostra?

Exercício 46 a)

$X_1 \sim N(20, 16) \rightarrow$ Amostra casual : $m = 9$

$X_2 \sim N(22, 16) \rightarrow$ Amostra casual : $n = 16$

(a)

Quer-se $P(\overline{X}_2 > \overline{X}_1 + 3) = P(\overline{X}_2 - \overline{X}_1 > 3) = P(\overline{X}_1 - \overline{X}_2 < -3)$

Sabe-se que : $Z = \dfrac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} \sim N(0,1)$

Logo,

$$P(\overline{X}_1 - \overline{X}_2 < -3) = P\left(\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < \frac{-3 - (20 - 22)}{\sqrt{\frac{16}{9} + \frac{16}{16}}}\right) = P(Z < -0.6) \approx$$

≈ 0.27425

Exercício 46 b)

Quer-se $P\left(S'_1 > 2S'_2\right) = P\left(\frac{S'_1}{S'_2} > 2\right)$

Sabe-se que : $F = \frac{S'^2_1}{S'^2_2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1) = F(8,15)$

Logo,

$$P\left(\frac{S'_1}{S'_2} > 2\right) = P\left(\frac{S'^2_1}{S'^2_2} > 4\right) = P\left(\frac{S'^2_1}{S'^2_2} \frac{\sigma_2^2}{\sigma_1^2} > 4 \times \frac{16}{16}\right) = P(F > 4) \approx 0.01$$

2

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: $X \sim Poisson(\lambda)$ $\quad \lambda = ?$

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

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 $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ 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) $\frac{X}{\sigma^2}$

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

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

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

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

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

Quais destas funções são estimadores do parâmetro p?

Slides Professora Claúdia Nunes

Estimador

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

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

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

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

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

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

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

modulol_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

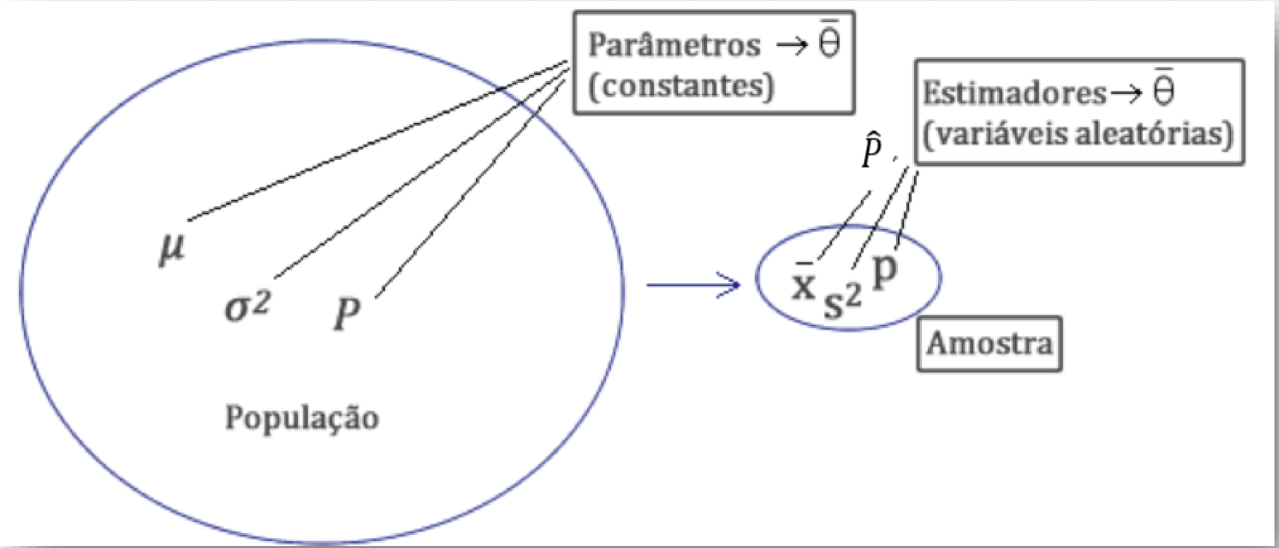
Parâmetro é uma quantidade numérica, em geral desconhecida, que descreve uma característica da população. Normalmente é representado por letras gregas como μ e σ, entre outras.

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.

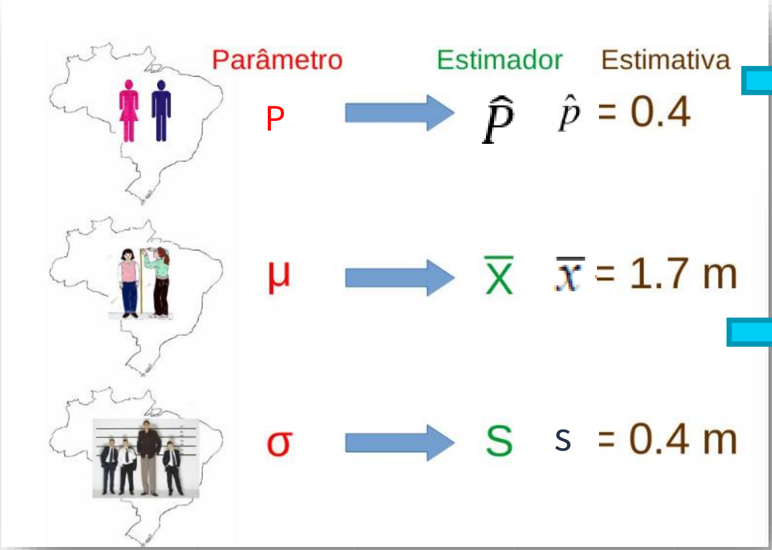
https://slideplayer.com.br/slide/2870987

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

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

Variância corrigida (s^2)

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

Variância não corrigida ($s^{2`}$)

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

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

Obrigada!

Questões?