Universidade do Minho
Escola de Engenharia
Departamento de Produção e Sistemas
Prof. Ana Cristina Braga

ESTATÍSTICA APLICADA

FORMULÁRIO

1º semestre 2020/2021

MIEI (2º ano)

MEDIDAS DE LOCALIZAÇÃO	MEDIDAS DE DISPERSÃO
Média $\bar{x} = \frac{1}{n}\sum_i x_i \qquad \bar{x} \approx \sum_k f_{r_k} M_k = \frac{1}{n}\sum_k f_k M_k$	**Erro Quadrático Médio** $EQM = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Mediana nº de elementos $Med = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$ com n par $Med = x_{\left(\frac{n+1}{2}\right)}$ com n ímpar limite inferior amplitude $Med = LI + \frac{0.5 - F_{r_A}^{-}}{F_{r_{Med}} - F_{r_A}^{-}}\Delta$	**Variância** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $s^2 \approx \frac{n}{n-1}\sum_k f_{r_k}(M_k - \bar{x})^2 = \frac{1}{n-1}\sum_k f_k (M_k - \bar{x})^2$ valores agregados **Desvio padrão** $s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$
Moda $Mod = LI + \frac{d_1}{d_1 + d_2}\Delta$ com $d_1 = f_{Mod} - f_A^{-}$ $\quad d_2 = f_{Mod} - f_D^{+}$	**Amplitude** $A = X_{(n)} - X_{(1)}$

PROBABILIDADES

$0 \le P(A) \le 1$; $P(S) = 1$; $P(A) + P(\bar{A}) = 1$; $P(\varnothing) = 0$; $P(\bar{A} \cap B) = P(B) - P(A \cap B)$	
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	$P(A \cap B) = P(A \mid B) \times P(B)$
$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$	$P(B_r \mid A) = \frac{P(B_r).P(A \mid B_r)}{\sum_{i=1}^{k} P(B_i).P(A \mid B_i)}$

DISTRIBUIÇÕES DE PROBABILIDADE

DISCRETO	CONTÍNUO
• propriedades da f.p.: 1) $f(x) \ge 0$ 2) $\sum_x f(x) = 1$	• propriedades da f.d.p.: 1) $f(x) \ge 0$ 2) $\int_x f(x)dx = 1$
• função probabilidade acumulada $F(x) = P(X \le x) = \sum_{-\infty}^{x} f(x)$	• função probabilidade acumulada $F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx$
• valor esperado $E(x) = \sum_x x \times f(x)$	• valor esperado $E(x) = \int_x x \times f(x)dx$
• variância σ^2 $V(x) = E(x^2) - [E(x)]^2$ $V(x) = \sum_x (x - \mu)^2 f(x)$	• variância $V(x) = E(x^2) - [E(x)]^2$ $V(x) = \int_x (x - \mu)^2 f(x)dx$

FUNÇÕES DE PROBABILIDADE

DISCRETAS

Distribuição de Bernoulli

$$f(x)=\pi^x(1-\pi)^{1-x} \qquad x=0 \text{ ou } 1$$

$$\mu=\pi \quad \sigma^2=\pi(1-\pi)$$

π -> probabilidade de sucesso

Distribuição Binomial *B(n,π)*

$$f(x)=C^n_x\,\pi^x(1-\pi)^{n-x} \qquad x=0,1,2,\ldots,n$$

$$\mu=n\pi \quad \sigma^2=n\pi(1-\pi)$$

π -> probabilidade de sucesso

Distribuição Poisson *P(λ)*

$$f(x)=\frac{\lambda^x e^{-\lambda}}{x!} \qquad x=0,1,2,\ldots$$

$$\mu=\lambda \quad \sigma^2=\lambda$$

Aproximação da Binomial à Poisson

n grande e π muito pequeno

$$\lambda=n\pi$$

$n \geq 20$ e $\pi \leq 0{,}05$

$n \geq 100$ e $n\pi < 10$

CONTÍNUAS

Distribuição Uniforme U*(α,β)*

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \alpha<x<\beta\\ 0 & \text{outros valores}\end{cases}$$

$$\mu=\frac{\beta+\alpha}{2} \qquad \sigma^2=\frac{1}{12}(\beta-\alpha)^2$$

Distribuição Exponencial *EN(1/θ)*

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & outros\end{cases}$$

$$\mu=\theta \quad \sigma^2=\theta^2$$

Distribuição Normal *N(μ,σ²)*

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mu=\mu \quad \sigma^2=\sigma^2$$

$$Z=\frac{X-\mu}{\sigma}$$

Aproximação da Binomial à Normal

Condições $\begin{cases} n\pi>5\\ n(1-\pi)>5\end{cases}$

$$\mu=n\pi$$

$$\sigma^2=n\pi(1-\pi)$$

<u>Correcção de Yates</u>

$$P(X\leq x)\approx P(X<x+0.5)$$

$$P(Y\geq y)\approx P(Y>y-0.5)$$

INTERVALOS DE CONFIANÇA E TESTES DE HIPÓTESES PARA UMA AMOSTRA

Parâmetro a estimar	Tipo de População	Dimensão da amostra	Conhece σ ?	E.T ~ Distribuição	Intervalo de Confiança	Notas
Média μ	Normal	Qualquer	Sim	$Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\bar{x}-z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}$	$z_{(1-\alpha/2)}$: quantil da tabela acumulada da Normal padrão à esquerda (tabela 5)
	Qualquer	$n\geq 30$	Não	$Z=\frac{\bar{x}-\mu}{s/\sqrt{n}}\sim N(0,1)$	$\bar{x}-z_{(1-\alpha/2)}\frac{s}{\sqrt{n}}<\mu<\bar{x}+z_{(1-\alpha/2)}\frac{s}{\sqrt{n}}$	Estimador do desvio padrão: $\sigma\approx s$ **(1)**
	Normal	$n<30$	Não	$T=\frac{\bar{x}-\mu}{s/\sqrt{n}}\sim t_{n-1}$	$\bar{x}-t_{(\alpha/2),n-1}\frac{s}{\sqrt{n}}<\mu<\bar{x}+t_{(\alpha/2),n-1}\frac{s}{\sqrt{n}}$	
Proporção binomial π	Binomial	$n>30$ **(2)**	-	$Z=\frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}\sim N(0,1)$	$p-z_{(1-\alpha/2)}\sqrt{\frac{p(1-p)}{n}}<\pi<p+z_{(1-\alpha/2)}\sqrt{\frac{p(1-p)}{n}}$	Estimador da proporção binomial $\pi\approx p=\frac{x}{n}$
Variância σ^2	População Normal			$Q=\frac{(n-1)s^2}{\sigma^2}\sim\chi^2_{n-1}$	$\frac{(n-1)s^2}{\chi^2_{(\alpha/2),n-1}}<\sigma^2<\frac{(n-1)s^2}{\chi^2_{(1-\alpha/2),n-1}}$	

INTERVALOS DE CONFIANÇA E TESTES DE HIPÓTESES PARA DUAS AMOSTRAS

Parâmetro a estimar	Tipo de População	Dimensão da amostra	Conhece σ ?	E.T ~ Distribuição	Intervalo de Confiança	Notas
Diferença entre as médias $\mu_1-\mu_2$	Normais	Quaisquer	σ_1 e σ_2 Sim	$Z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$(\bar{x}_1-\bar{x}_2)\pm z_{(1-\alpha/2)}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$	
	Quaisquer	$n_1\geq 30$ e $n_2\geq 30$	σ_1 e σ_2 Não	$Z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}\sim N(0,1)$	$(\bar{x}_1-\bar{x}_2)\pm z_{(1-\alpha/2)}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$	Estimadores dos desvios padrão: $\sigma_1\approx s_1$, $\sigma_2\approx s_2$
	Normais	$n_1<30$ e $n_2<30$	σ_1 e σ_2 Não e $\sigma_1^2=\sigma_2^2$	$T=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{GL}$	$(\bar{x}_1-\bar{x}_2)\pm t_{(\alpha/2),GL}\,s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	$GL=n_1+n_2-2$ $s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$
	Normais Amostras dependentes	$n_1<30$ e $n_2<30$	σ_1 e σ_2 Não	$T=\frac{\bar{D}_i-(\mu_1-\mu_2)}{\frac{s_{D_i}}{\sqrt{n}}}\sim t_{(n-1)}$	$\bar{D}_i-t_{(n-1),\alpha/2}\cdot\frac{s_{D_i}}{\sqrt{n}}<\mu_1-\mu_2<\bar{D}_i+t_{(n-1),\alpha/2}\cdot\frac{s_{D_i}}{\sqrt{n}}$	$s_{D_i}=s_{n-1}$ para $D_i=X_{1i}-X_{2i}$
Diferença de proporções $\pi_1-\pi_2$	Binomial	$n_1\geq 30$ e $n_2\geq 30$	-	$Z=\frac{(p_1-p_2)-(\pi_1-\pi_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}\sim N(0,1)$ **(3)**	$(p_1-p_2)\pm z_{(1-\alpha/2)}\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$	Estimadores das proporções binomiais **(4)** $p_1=\frac{x_1}{n_1}$ e $p_2=\frac{x_2}{n_2}$
Razão de variâncias σ_1^2/σ_2^2	Normais	Quaisquer	-	$F=\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}\sim F_{\nu_1,\nu_2}$	$\frac{s_1^2}{s_2^2}\frac{1}{F_{(\alpha/2),\nu_1,\nu_2}}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{s_1^2}{s_2^2}\frac{1}{F_{(1-\alpha/2),\nu_1,\nu_2}}$	$\nu_1=n_1-1$ e $\nu_2=n_2-1$ $\frac{1}{F_{(1-\alpha/2),\nu_1,\nu_2}}=F_{(\alpha/2),\nu_2,\nu_1}$

(1) O desvio padrão σ , sendo desconhecido, é estimado através de $s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$; **(2)** Proporção para amostras de pequena dimensão necessário recorrer à solução exata através da distribuição binomial; **(3)** e **(4)** No teste à diferença de proporções se $H_0:\pi_1-\pi_2=0$, a E.T. passa a ser:

$$Z=\frac{(p_1-p_2)}{\sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\sim N(0,1),\ \text{com}\ p=\frac{x_1+x_2}{n_1+n_2}.$$

PLANEAMENTO COMPLETAMENTE ALEATÓRIO

H_0: Não existem diferenças significativas na variável resposta devido aos diferentes tratamentos

$\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$ ou $\alpha_j = 0$ com $j = 1,2, ..., k$

H_1: Pelo menos 2 tratamentos são diferentes ($\alpha_j \neq 0$ para pelo menos um valor de j).

R.R: $F > c$

Tabela ANOVA

Fonte de variação	Soma dos Quadrados	Graus de liberdade	Média dos Quadrados	Estatística de teste, F
Tratamentos (Entre grupos)	SQT	$k-1$	MQT	F = MQT/MQR
Resíduos (dentro dos grupos)	SQR	$\Sigma n_j - k$	MQR	F = MQT/MQR
Total	STQ	$N - 1$		

$$SQT = \sum_{j=1}^{k} \frac{T_{.j}^2}{n_j} - \frac{T_{..}^2}{N} \qquad STQ = \sum_{j=1}^{k} \sum_{i=1}^{n_j} y_{ij}^2 - \frac{T_{..}^2}{N} \qquad STQ = SQT + SQR \qquad N = \sum_{j=1}^{k} n_j$$

Intervalos de confiança para diferenças entre pares de médias de tratamentos i e j com i $i \neq j = 1, 2, .., k$

$$(\bar{y}_i - \bar{y}_j) - t_{(N-k, \alpha/2)} \cdot \sqrt{MQR} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \leq \mu_i - \mu_j \leq (\bar{y}_i - \bar{y}_j) + t_{(N-k, \alpha/2)} \cdot \sqrt{MQR} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

PLANEAMENTO COM BLOCOS ALEATÓRIOS

H_{01}: Não existem diferenças significativas na variável resposta devido aos diferentes tratamentos

$\mu_{1.} = \mu_{2.} = \mu_{3.} = ... = \mu_{k.}$ ou $\alpha_j = 0$ com $j = 1,2, ..., k$

H_{11}: Pelo menos 2 tratamentos são diferentes ($\alpha_j \neq 0$ para pelo menos um valor de j).

R.R: $F_1 > c_1$

H_{02}: Não existem diferenças significativas na variável resposta devido aos diferentes blocos

$\mu_{.1} = \mu_{.2} = \mu_{.3} = ... = \mu_{.b}$ ou $\beta_i = 0$ com $i = 1,2, ..., b$

H_{12}: Pelo menos 2 blocos são diferentes ($\beta_i \neq 0$ para pelo menos um valor de i).

R.R: $F_2 > c_2$

Tabela ANOVA

Fonte de variação	Soma dos Quadrados	Graus de liberdade	Média dos Quadrados	Estatística de teste, F
Tratamentos (colunas)	SQT	$k-1$	MQT	F_1 = MQT/MQR
Blocos (linhas)	SQB	$b-1$	MQB	F_1 = MQT/MQR
Resíduos (dentro dos grupos)	SQR	$(k-1)(b-1)$	MQR	F_2 = MQB/MQR
Total	STQ	$N - 1$		

$$SQT = \frac{1}{b} \sum_{j=1}^{k} T_{.j}^2 - \frac{T_{..}^2}{N} \qquad SQB = \frac{1}{k} \sum_{j=1}^{b} T_{i.}^2 - \frac{T_{..}^2}{N} \qquad STQ = \sum_{j=1}^{k} \sum_{i=1}^{b} y_{ij}^2 - \frac{T_{..}^2}{N} \qquad STQ = SQT + SQB + SQR \qquad N = k.b$$

TESTES DO “BOM AJUSTE” DO QUI-QUADRADO

Probabilidades completamente especificadas na hipótese nula

H_0: $p_1 = p_{10}$, $p_2 = p_{20}$, ..., $p_k = p_{k0}$ e $p_{10} + p_{20} + ... + p_{k0} = 1$

R.R: Q > c com $c = \chi^2_{k-1,\alpha}$

Probabilidades não estão completamente especificadas na hipótese nula

H_0: As probabilidades correspondentes das classes provêm de uma distribuição da família.....

R.R: Q > c com $c = \chi^2_{g.l.,\alpha}$

g.l. = nº de celas – 1 – nº de parâmetros estimados

$$Q = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \quad \text{com } e_i = n.p_i$$