UNIVERSIDADE DE ÉVORA
ESCOLA DE CIÊNCIAS E TECNOLOGIA

Formulário de Introdução à Probabilidade e Estatística

Engenharia Civil, Engenharia das Energias Renováveis, Engenharia Geológica, Engenharia Informática e Engenharia Mecatrónica

Estatística Descritiva

Localização	Dados Não Agrupados	Dados Agrupados
Média	$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$	$\overline{X}=\frac{1}{n}\sum_{i=1}^{k}n_iX_i'$ k=nº de classes ou categorias

Localização	Dados Não Agrupados e Dados Agrupados Discretos	Dados Agrupados Contínuos
Moda	M_o=Valor ou categoria que se repete mais vezes	$M_0=l_i+A_i\frac{\Delta_1}{\Delta_1+\Delta_2}$ $\Delta_1=n_i-n_{i-1}$ e $\Delta_2=n_i-n_{i+1}$
Mediana	$M_e=\begin{cases}\frac{1}{2}(X_{([\frac{n}{2}])}+X_{([\frac{n}{2}+1])}), & n \text{ par}\\ X_{([\frac{n+1}{2}])}, & n \text{ ímpar}\end{cases}$	$M_e=l_i+A_i\frac{\frac{n}{2}-N_{i-1}}{n_i}$
Quantis	$Q_p=\begin{cases}\frac{1}{2}(X_{(np)}+X_{(np+1)}), & np \text{ inteiro}\\ X_{([np]+1)}, & np \text{ não inteiro}\end{cases}$	$Q_p=l_i+A_i\frac{np-N_{i-1}}{n_i}$

Dispersão	Dados Não Agrupados	Dados Agrupados
Variância	$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}\sum_{i=1}^{n}X_i^2-\frac{n}{n-1}\overline{X}^2$	$S^2=\frac{1}{n-1}\sum_{i=1}^{k}n_i(X_i'-\overline{X})^2=\frac{1}{n-1}\sum_{i=1}^{k}n_iX_i'^2-\frac{n}{n-1}\overline{X}^2$
Momento Empírico de ordem m	$M_m'=\frac{1}{n}\sum_{i=1}^{n}X_i^m$	$M_m'=\frac{1}{n}\sum_{i=1}^{k}n_iX_i'^m$
Momento Empírico Centrado de ordem m	$M_m=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^m$	$M_m=\frac{1}{n}\sum_{i=1}^{k}n_i(X_i'-\overline{X})^m$

Outras Medidas de Dispersão (Dados agrupados e não agrupados):

Amplitude (*Range*) $\Delta=X_{(n)}-X_{(1)}$; **Dispersão Quartal** $Q=Q_{0,75}-Q_{0,25}$

Intervalo de Variação $Q'=Q_{0,90}-Q_{0,10}$; **Coeficiente de Variação** $CV=\frac{S}{\overline{X}}\times 100\%$

Construção de Classes: $A_i=\frac{\Delta}{k}$; $k=\left[\frac{\ln(n)}{\ln(2)}\right]+1$ **(Regra de Sturges:)**

Medidas de Assimetria (*Skewness*) (Dados agrupados e não agrupados):

Coeficiente de Assimetria de Fisher $\beta_1 = \frac{M_3}{S^3} = \begin{cases} < 0, & \text{Dist. Assimétrica Negativa} \quad (\overline{X} < M_e) \\ = 0, & \text{Dist. Simétrica} \quad (\overline{X} = M_e) \\ > 0, & \text{Dist. Assimétrica Positiva} \quad (\overline{X} > M_e) \end{cases}$

Grau de Assimetria de Pearson $G_P = \frac{\overline{X} - M_o}{S} = \begin{cases} < 0, & \text{Dist. Assimétrica Negativa} \\ = 0, & \text{Dist. Simétrica} \\ > 0, & \text{Dist. Assimétrica Positiva} \end{cases}$

Coeficiente de Assimetria de Bowley $G_B = \dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} = \begin{cases} < 0, & \text{Dist. Ass. Negativa} \\ = 0, & \text{Dist. Simétrica} \\ > 0, & \text{Dist. Ass. Positiva} \end{cases}$

Medidas de Achatamento (*Kurtosis*) (Dados agrupados e não agrupados):

Coeficiente de Achatamento $\beta_2 = \frac{M_4}{S^4} = \begin{cases} < 3, & \text{Dist. Platicúrtica} \\ = 3, & \text{Dist. Mesocúrtica} \\ > 3, & \text{Dist. Leptocúrtica} \end{cases}$

Coeficiente Percentil de Achatamento $K = \dfrac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \begin{cases} > 0,263, & \text{Dist. Platicúrtica} \\ = 0,263, & \text{Dist. Mesocúrtica} \\ < 0,263, & \text{Dist. Leptocúrtica} \end{cases}$

Medidas de Associação Amostral:

Covariância amostral de (X, Y)**:** $S_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) = \frac{1}{n-1} \sum_{i=1}^{n} X_i Y_i - \frac{n}{n-1} \overline{X}\,\overline{Y}$

Correlação amostral de (X, Y)**:** $r = \frac{S_{XY}}{S_X\, S_Y}$

Distribuições de Probabilidades

Univariadas			
Distribuição	$\mathbf{P(X = x)}$	$\mathbf{E[X]}$	$\mathrm{Var}[\mathbf{X}]$
Binomial $X \frown B(n, p)$	${}^{n}C_x p^x (1-p)^{n-x}, \ x = 0, \ldots, n$	np	$np(1-p)$
Hipergeométrica $X \frown H(N, n, p)$	$\dfrac{{}^{Np}C_x \, {}^{Nq}C_{n-x}}{{}^{N}C_n}, \ x = 0, \ldots, n$	np	$np(1-p)\frac{N-n}{N-1}$
Poisson $X \frown P(\lambda)$	$\dfrac{e^{-\lambda}\lambda^x}{x!}, \ x = 0, 1, \ldots$	λ	λ
Distribuição	$\mathbf{f(x)}$	$\mathbf{E[X]}$	$\mathrm{Var}[\mathbf{X}]$
Normal $X \frown N(\mu, \sigma)$	$\dfrac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ x \in \mathbb{R}$	μ	σ^2
Exponencial $X \frown \mathrm{Exp}(\lambda)$	$\lambda e^{-\lambda x}, \ x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Aproximações:

- Se $X \frown B(n, p)$, $n \geq 20$ e $p \leq 0,05$, então $X \dot{\frown} P(np)$
- Se $X \frown B(n, p)$, $n > 50$ e $0,1 < p < 0,9$, então $X \dot{\frown} N(np, \sigma^2 = np(1-p))$
- Se $X \frown P(\lambda)$, $\lambda > 20$, então $X \dot{\frown} N(\lambda, \sigma^2 = \lambda)$

Intervalos de Confiança

Parâmetro: μ		
$\boldsymbol{\sigma^2}$ conhecido?	**Condições**	**IC a** $\mathbf{100(1-\alpha)\%}$
Sim	População Normal e n qualquer ou População qualquer e $n > 30$	$\left]\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right[$
Não	População Normal	$\left]\overline{X} - t_{n-1;1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}; \overline{X} + t_{n-1;1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right[$
Não	População qualquer e $n > 30$	$\left]\overline{X} - Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right[$

Parâmetro: p	
População	**IC a** $\mathbf{100(1-\alpha)\%}$
Bernoulli $n > 30$	$\left]\overline{P} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{P}(1-\overline{P})}{n}}; \overline{P} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{P}(1-\overline{P})}{n}}\right[$

Parâmetro: σ^2	
População	**IC a** $\mathbf{100(1-\alpha)\%}$
Normal	$\left]\frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}; \frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right[$

Parâmetro: $\mu_1 - \mu_2$		
$\boldsymbol{\sigma_1^2}$ **e** $\boldsymbol{\sigma_2^2}$ conhecidos?	**Populações**	**IC a** $\mathbf{100(1-\alpha)\%}$
Sim	Normais ou Quaisquer e $n_1 > 30$ e $n_2 > 30$	$\left]\overline{X_1} - \overline{X_2} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; \overline{X_1} - \overline{X_2} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right[$
Não $(\sigma_1^2 = \sigma_2^2)$	Normais	$\left]\overline{X_1} - \overline{X_2} - t_{n_1+n_2-2;1-\frac{\alpha}{2}} S^*; \overline{X_1} - \overline{X_2} + t_{n_1+n_2-2;1-\frac{\alpha}{2}} S^*\right[$ $S^* = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
Não $(\sigma_1^2 = \sigma_2^2)$	Quaisquer e $n_1 > 30$ e $n_2 > 30$	$\left]\overline{X_1} - \overline{X_2} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}; \overline{X_1} - \overline{X_2} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right[$

Parâmetro: $p_1 - p_2$	
Populações	**IC a 100(1 − α)%**
Bernoulli $n_1 > 30$ e $n_2 > 30$	$\left]\overline{P_1} - \overline{P_2} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{P_1}(1-\overline{P_1})}{n_1} + \frac{\overline{P_2}(1-\overline{P_2})}{n_2}}; \overline{P_1} - \overline{P_2} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{P_1}(1-\overline{P_1})}{n_1} + \frac{\overline{P_2}(1-\overline{P_2})}{n_2}}\right[$

Parâmetro: σ_1^2/σ_2^2	
Populações	**IC a 100(1 − α)%**
Normais	$\left]\frac{1}{F_{n_1-1,n_2-1;1-\frac{\alpha}{2}}}\frac{S_1^2}{S_2^2}; \frac{1}{F_{n_1-1,n_2-1;\frac{\alpha}{2}}}\frac{S_1^2}{S_2^2}\right[$

Testes de Hipóteses

$H_0: \mu = \mu_0$ **ou** $H_0: \mu \leq \mu_0$ **ou** $H_0: \mu \geq \mu_0$		
$\boldsymbol{\sigma^2}$ conhecido?	**População**	**Estatística de Teste**
Sim	Normal	$Z = \sqrt{n}\,\frac{\overline{X} - \mu_0}{\sigma} \frown N(0,1)$
Sim	Qualquer e $n > 30$	$Z = \sqrt{n}\,\frac{\overline{X} - \mu_0}{\sigma} \dot{\frown} N(0,1)$
Não	Normal	$T = \sqrt{n}\,\frac{\overline{X} - \mu_0}{S} \frown t_{(n-1)}$
Não	Qualquer e $n > 30$	$Z = \sqrt{n}\,\frac{\overline{X} - \mu_0}{S} \dot{\frown} N(0,1)$

$H_0: p = p_0$ **ou** $H_0: p \leq p_0$ **ou** $H_0: p \geq p_0$	
População	**Estatística de Teste**
Bernoulli $n > 30$	$Z = \frac{\overline{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \dot{\frown} N(0,1)$

$H_0: \sigma^2 = \sigma_0^2$ **ou** $H_0: \sigma^2 \leq \sigma_0^2$ **ou** $H_0: \sigma^2 \geq \sigma_0^2$	
População	**Estatística de Teste**
Normal	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \frown \chi^2_{(n-1)}$

$H_0: \mu_1-\mu_2=\mu_0$ **ou** $H_0: \mu_1-\mu_2\leq\mu_0$ **ou** $H_0: \mu_1-\mu_2\geq\mu_0$		
$\boldsymbol{\sigma_1^2}$ **e** $\boldsymbol{\sigma_2^2}$ conhecidos?	**Populações**	**Estatística de Teste**
Sim	Normais	$Z=\dfrac{\overline{X_1}-\overline{X_2}-\mu_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \frown N(0,1)$
Sim	Quaisquer e $n_1>30$ e $n_2>30$	$Z=\dfrac{\overline{X_1}-\overline{X_2}-\mu_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \dot{\frown} N(0,1)$
Não ($\sigma_1^2=\sigma_2^2$)	Normais	$T=\dfrac{\overline{X_1}-\overline{X_2}-\mu_0}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}},\ \frown t_{(n_1+n_2-2)}$
Não ($\sigma_1^2=\sigma_2^2$)	Quaisquer e $n_1>30$ e $n_2>30$	$Z=\dfrac{\overline{X_1}-\overline{X_2}-\mu_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \dot{\frown} N(0,1)$

$H_0: p_1-p_2=p_0$ **ou** $H_0: p_1-p_2\leq p_0$ **ou** $H_0: p_1-p_2\geq p_0$	
Populações	**Estatística de Teste**
Bernoulli $n_1>30$ e $n_2>30$	$\dfrac{\overline{P_1}-\overline{P_2}-p_0}{\sqrt{\overline{P}^*(1-\overline{P}^*)(\frac{1}{n_1}+\frac{1}{n_2})}} \dot{\frown} N(0,1)$ $\overline{P}^*=\frac{n_1\overline{P_1}+n_2\overline{P_2}}{n_1+n_2}$

$H_0: \frac{\sigma_1^2}{\sigma_2^2}=\sigma_0^2$ **ou** $H_0: \frac{\sigma_1^2}{\sigma_2^2}\leq\sigma_0^2$ **ou** $H_0: \frac{\sigma_1^2}{\sigma_2^2}\geq\sigma_0^2$	
Populações	**Estatística de Teste**
Normais	$\dfrac{1}{\sigma_0^2}\dfrac{S_1^2}{S_2^2} \frown F_{(n_1-1;n_2-1)}$

Regressão Linear Simples: $Y_i=\beta_0+\beta_1X_i+\epsilon_i$

- $\widehat{\beta}_0=\overline{Y}-\widehat{\beta}_1\overline{X}$
- $\widehat{\beta}_1=\dfrac{\sum_{i=1}^n(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^n(X_i-\overline{X})^2}=\dfrac{S_{XY}}{S_X^2}=r\,\dfrac{S_Y}{S_X}$
- $\widehat{Var}[\widehat{\beta}_0]=\dfrac{\widehat{\sigma}^2\sum_{i=1}^n X_i^2}{n^2\,S_X^2}$
- $\widehat{Var}[\widehat{\beta}_1]=\dfrac{\widehat{\sigma}^2}{n\,S_X^2}$
- $\widehat{\sigma}^2=\displaystyle\sum_{i=1}^n\frac{e_i^2}{n-2}=\sum_{i=1}^n\frac{(Y_i-\hat{Y}_i)^2}{n-2}=\frac{n-1}{n-2}(S_Y^2-\widehat{\beta}_1S_{XY})$

IC a $(1-\alpha)\,100\%$	
Parâmetro	**Intervalo de Confiança**
β_0	$\left]\widehat{\beta}_0 - t_{n-2;1-\frac{\alpha}{2}}\sqrt{\widehat{Var}[\widehat{\beta}_0]};\ \widehat{\beta}_0 + t_{n-2;1-\frac{\alpha}{2}}\sqrt{\widehat{Var}[\widehat{\beta}_0]}\right[$
β_1	$\left]\widehat{\beta}_1 - t_{n-2;1-\frac{\alpha}{2}}\sqrt{\widehat{Var}[\widehat{\beta}_1]};\ \widehat{\beta}_1 + t_{n-2;1-\frac{\alpha}{2}}\sqrt{\widehat{Var}[\widehat{\beta}_1]}\right[$
Y_S (Predição)	$\left]\hat{Y}_S - t_{n-2;1-\frac{\alpha}{2}}\sqrt{\widehat{\sigma}^2\left(\frac{1}{n}+\frac{(X_S-\overline{X})^2}{S_X}\right)};\ \hat{Y}_S + t_{n-2;1-\frac{\alpha}{2}}\sqrt{\widehat{\sigma}^2\left(\frac{1}{n}+\frac{(X_S-\overline{X})^2}{S_X}\right)}\right[$

$H_0:\beta_0=\beta_{0,0}$ **ou** $H_0:\beta_0\leq\beta_{0,0}$ **ou** $H_0:\beta_0\geq\beta_{0,0}$	$H_0:\beta_1=\beta_{1,0}$ **ou** $H_0:\beta_1\leq\beta_{1,0}$ **ou** $H_0:\beta_1\geq\beta_{1,0}$
Estatística de Teste	**Estatística de Teste**
$T=\dfrac{\hat{\beta}_0-\beta_0}{\sqrt{Var[\hat{\beta}_0]}} \frown t(n-2)$	$T=\dfrac{\hat{\beta}_1-\beta_1}{\sqrt{Var[\hat{\beta}_1]}} \frown t(n-2)$

Testes Não Paramétricos

Teste	**Estatística de Teste**	**Decisão de Rejeição de H_0**
Ajustamento	$\chi^2=\sum_{i=1}^{k}\dfrac{(O_i-E_i)^2}{E_i} \frown \chi^2_{(k-p-1)}$	$\chi^2_{Obs} > \chi^2_{k-p-1;1-\alpha}$
Independência	$\chi^2=\sum_{i=1}^{L}\sum_{j=1}^{C}\dfrac{(O_{ij}-E_{ij})^2}{E_{ij}} \frown \chi^2_{(L-1)(C-1)}$	$\chi^2_{Obs} > \chi^2_{(L-1)(C-1);1-\alpha}$
Independência tabelas 2×2 (Correcção de Yates)	$\chi^2=\dfrac{n(\lvert O_{11}O_{22}-O_{12}O_{21}\rvert-0,5n)^2}{O_{1.}O_{2.}O_{.1}O_{.2}} \frown \chi^2_{(L-1)(C-1)}$	$\chi^2_{Obs} > \chi^2_{(L-1)(C-1);1-\alpha}$