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 (variável contínua)	Dados Agrupados (variável discreta)
Média $\bar{X}$ (mean)	$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ ou $\frac{\text{sum}}{N}$	$\overline{X}=\frac{1}{n}\sum_{i=1}^{k}n_iX'_i$ 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 $\Delta_2=n_i-n_{i+1}$
(median) $\tilde{X}$ 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	$Q_p=\begin{cases}\frac{1}{2}(X_{(np)}+X_{(np+1)}), & np \text{ inteiro}\\ X_{([np]+1)}, & np \text{ não inteiro}\end{cases}$	$Q_p=l_i+A_i\frac{np-N_{i-1}}{n_i}$

$Q_1 = P_{25}$
$Q_2 = \tilde{X} = P_{50}$
$Q_3 = P_{75}$

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

Desvio-padrão (std. Deviation)
$S=\sqrt{S^2}$

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

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

$a = \max - \min$

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

(Homogeneidade)
+ homogéneo CV <

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

- CV > 50%, baixa representatividade da média
- CV < 50%, a média será tanto mais representativa dos valores da amostra quanto menor for o valor deste coeficiente.

Amplitude interquartil
$IQ = Q_3 - Q_1$

Outlier moderado
$x_i < Q_1 - 1,5\, IQ$
$x_i > Q_3 + 1,5\, IQ$

Outlier severo
$x_i < Q_1 - 3\, IQ$
$x_i > Q_3 + 3\, IQ$

Função de probabilidade marginal

$f_x(x) = P(X=x, -\infty < y < +\infty) = \sum_{y \in D_y} f(x,y)$ x fixo

$f_y(y) = P(-\infty < x < +\infty, Y=y) = \sum_{x \in D_x} f(x,y)$ y fixo

⚠ Dada uma variável aleatória bidimensional diz-se que X e Y são independentes se e só se

$f(x,y) = f_x(x) \times f_y(y)$

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

Coeficiente de assimetria do SPSS:

$$\gamma_{SPSS} = \frac{\text{skewness}}{\text{std. error of skewness}}$$

$\gamma_{SPSS} < -1,96$ (assimetria negativa)

$|\gamma_{SPSS}| < 1,96$ (simétrica)

$\gamma_{SPSS} > 1,96$ (assimetria positiva)

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

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 de achatamento do SPSS:

$$k_{SPSS} = \frac{\text{kurtosis}}{\text{std. error of kurtosis}}$$

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

$k_{SPSS} < -1,96$ (platicúrtica)

$|k_{SPSS}| < 1,96$ (mesocúrtica)

$k_{SPSS} > 1,96$ (leptocúrtica)

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): $r = \frac{S_{XY}}{S_X S_Y}$

ou Correlação linear de Pearson

desvio padrão - s_x e s_y

Distribuições de Probabilidades

n - provas

p - probabilidade de sucesso

utilizada para modelar contagens em intervalos de tempo

Univariadas			
Distribuição	**P(X = x)**	**E[X]**	**Var[X]**
① Binomial $X \frown B(n,p)$	${}^nC_x p^x (1-p)^{n-x}, \; x = 0, \ldots, n$	np	$np(1-p)$
Hipergeométrica $X \frown H(N,n,p)$	$\frac{{}^{Np}C_x \, {}^{Nq}C_{n-x}}{{}^NC_n}, \; x = 0, \ldots, n$	np	$np(1-p)\frac{N-n}{N-1}$
② Poisson $X \frown P(\lambda)$	$\frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0, 1, \ldots$	λ	λ
Distribuição	**f(x)**	**E[X]**	**Var[X]**
③ Normal $X \frown N(\mu,\sigma)$	$\frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in \mathbb{R}$	μ	σ^2
Exponencial $X \frown \mathrm{Exp}(\lambda)$	$\lambda e^{-\lambda x}, \; x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

λ - nº médio de sucessos que ocorrem no intervalo de tempo

Distribuição normal reduzida ou normal padrão

Se $X \sim N(\mu, \sigma)$, então

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

(nós podemos ver na tabela)

Propriedades

- $P(Z \le -z) = P(Z \ge z) = 1 - P(Z < z) = 1 - \phi(z)$
- $P(a < X \le b) = P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$
- $P(Z \ge -z) = P(Z \le z)$
- $P(Z > z) = 1 - P(Z \le z)$

Teorema da Aditividade

① $X_1 + X_2 + \ldots + X_k = \sum_{i=1}^{k} X_i \sim B\left(\sum_{i=1}^{k} n_i \,;\, p\right)$

② $X_1 + X_2 + \ldots + X_k = \sum_{i=1}^{k} X_i \sim P(k\lambda)$

③ ver folha 6

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 $X \dot{\frown} N(\lambda, \sigma^2 = \lambda)$

Parâmetros de variáveis aleatória

$$\mu = \underset{E[X]}{E(X)} = \sum_{x_i \in D_x} x_i P(X = x_i) = \sum_{x_i \in D_x} x_i f(x_i) = \text{ex: } (k-1)3p + (1-k)p + \ldots$$

valor esperado ou média

Do tipo:

x	$k-1$	$1-k$	k
$f(x)$	$3p$	p	p

$$\sigma^2 = \underset{V[X]}{\mathrm{Var}(X)} = E(X^2) - E^2(X) = \left(\sum_{x_i \in D_x} x_i^2 f(x_i)\right) - \mu^2$$

variância

$\sigma = \sqrt{\mathrm{Var}(X)}$

desvio-padrão

se aumentar a confiança o intervalo terá maior amplitude.
Pelo contrário se diminuir a confiança ter-se-á um IC com menor amplitude.

Intervalos de Confiança

IC para a diferença de duas médias com variâncias conhecidas

$$IC_{100(1-\alpha)\%}(\mu_X - \mu_Y) = \left]\bar{x} - \bar{y} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} ; \bar{x} - \bar{y} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right[$$

Parâmetro: μ média		
σ^2 conhecido?	Condições	IC a $100(1-\alpha)\%$
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 percentagem	
População	IC a $100(1-\alpha)\%$
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 variância	
População	IC a $100(1-\alpha)\%$
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: $\mu_1 - \mu_2$ diferença de médias		
σ_1^2 e σ_2^2 conhecidos?	Populações	IC a $100(1-\alpha)\%$
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 $n_2 > 30$	$\left]\overline{X_1} - \overline{X_2} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}; \overline{X_1} - \overline{X_2} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right[$

Função de probabilidade condicionada

- Função de probabilidade de X condicionada a $\{Y = y\}$

$$f_{X|Y=y}(x) = P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{f(x,y)}{f_Y(y)}$$

- Função de probabilidade de Y condicionada a $\{X = x\}$

$$f_{X=x|Y}(y) = P(Y=y \mid X=x) = \frac{P(X=x, Y=y)}{P(X=x)} = \frac{f(x,y)}{f_X(x)}$$

Função (massa de) probabilidade

$$f(x) = \begin{cases} P(X=x) & \text{se } x \in D \\ 0 & \text{se } x \notin D \end{cases}$$

⚠ construir tabela

x	...
$f(x)$	...

Função de distribuição

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

(função em escada)

↳ Propriedades

$$P(X > x_1) = 1 - P(X \le x_1) = 1 - F(x_1)$$

$$P(x_1 \le X \le x_2) = P(X \le x_2) - P(X \le x_1) = F(x_2) - F(x_1)$$

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

Populações	Parâmetro: σ_1^2/σ_2^2 IC a $100(1-\alpha)\%$
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 → teste significância

Níveis de significância (α)
↳ mais usados: 0,01 ; 0,05 e 0,1

Teste de hipótese para a média

$H_0: \mu = \mu_0$ ou $H_0: \mu \leq \mu_0$ ou $H_0: \mu \geq \mu_0$		
σ^2 conhecido?	População	Estatística de Teste
Sim	Normal	$\underbrace{Z = \sqrt{n}\frac{\overline{X} - \mu_0}{\sigma} \frown N(0,1)}_{t_{obs}}$
Sim	Qualquer e $n > 30$	$Z = \sqrt{n}\frac{\overline{X} - \mu_0}{\sigma} \frown N(0,1)$
Não	Normal	T-student média o que tiver na pergunta $T = \sqrt{n}\frac{\overline{X} - \mu_0}{S} \frown t_{(n-1)}$ df? std deviation
Não	Qualquer e df? $n > 30$	$Z = \sqrt{n}\frac{\overline{X} - \mu_0}{S} \frown N(0,1)$

p-value (sig)

std error mean $\frac{S}{\sqrt{n}}$

<u>One sample test</u>
- Test value ⇔ μ_0
- Mean Difference ⇔ $\overline{x} - \mu_0$
- t ⇔ t_{obs} ← $n_1 + n_2 - 2$
- df ⇔ graus de liberdade $(n-1)$

Independent samples test
- Lower = 2× Mean Difference − Upper

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

$H_0: \sigma^2 = \sigma_0^2$ ou $H_0: \sigma^2 \leq \sigma_0^2$ ou $H_0: \sigma^2 \geq \sigma_0^2$	
População	Estatística de Teste
Normal	df? std deviation $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \frown \chi^2_{(n-1)}$ Qui-quadrado variância

? <u>Cálculo do p-value</u>

Exemplo:

$t_{19;\,0.995} \simeq 3.865$

vamos à tabela procurar o valor 3,865 para descobrir

$0.995 = \underbrace{1 - \frac{\alpha}{2}}$ (teste bilateral)

⇔ $\alpha = 0,001$

↓
p-value

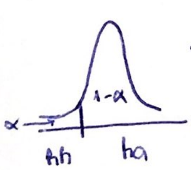
$1 - \alpha$ para testes unilaterais

(ver página 6)

Teste de hipótese para a média com <u>σ^2 desconhecido</u>

<u>Teste unilateral esquerdo</u>

$H_0: \mu \geq \mu_0$ vs $H_1: \mu < \mu_0$



rejeita-se H_0 se $t_{obs} \leq -t_{n-1;1-\alpha}$

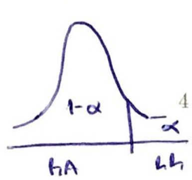
cálculo do p-value

p-value = $P(T \leq t_{obs})$

(ver página 6)

<u>Teste unilateral direito</u>

$H_0: \mu \leq \mu_0$ vs $H_1: \mu > \mu_0$



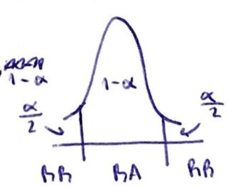
rejeita-se H_0 se $t_{obs} \geq t_{n-1;1-\alpha}$

cálculo do p-value

p-value = $P(T \geq t_{obs})$
$= 1 - F(t_{obs})$

<u>Teste bilateral</u>

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$



rejeita-se H_0 se $|t_{obs}| \geq t_{n-1;1-\frac{\alpha}{2}}$

cálculo do p-value

p-value = $2 \times P(T \geq |t_{obs}|)$

Cálculo do p-value

Ex:

$z_{obs} = t_{1-\alpha}$

$2,59 = z_{1-\alpha}$ (procurar na tabela)

$= z_{0,97}$

$1-\alpha = 0,97$

$\alpha = 1 - 0,97$

$\alpha = 0,03$

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
$T = \dfrac{\hat{\beta}_0 - \beta_0}{\sqrt{\widehat{Var}[\widehat{\beta}_0]}} \frown t(n-2)$	$T = \dfrac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{Var}[\widehat{\beta}_1]}} \frown t(n-2)$

Testes Não Paramétricos

$n = k - 1$

k - classes

p - ?

só uma variável

2×3

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

teste unilateral direito

Teste de hipótese para a média com σ^2 conhecido

- $H_0: \mu \geq \mu_0$ vs $H_1: \mu < \mu_0$ | $H_0: \mu_X - \mu_Y \geq \mu_0$ ...
  rejeitar H_0 se $z_{obs} \leq -z_{1-\alpha}$ | $t_{obs} \leq -t_{1-\alpha}$
  p-value $= P(Z \leq z_{obs}) = \Phi(z_{obs})$
- $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ | $H_0: \mu_X - \mu_Y = \mu_0$ vs ...
  rejeitar H_0 se $|z_{obs}| \geq z_{1-\frac{\alpha}{2}}$ | $|z_{obs}| \geq z_{1-\frac{\alpha}{2}}$
  p-value $= 2 \times P(Z \geq |z_{obs}|) = 2\left[1 - \Phi(z_{obs})\right]$
- $H_0: \mu \leq \mu_0$ vs $H_1: \mu > \mu_0$ | $H_0: \mu_X - \mu_Y \leq \mu_0$ ...
  rejeitar H_0 se $z_{obs} \geq z_{1-\alpha}$ | $z_{obs} \geq z_{1-\alpha}$
  p-value $= P(Z \geq z_{obs}) = 1 - \Phi(z_{obs})$

Potência teste (médias)

- $\mu \leq \mu_0$
  $\mu_c: \mu_0 + z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$
  $\beta(\mu_1) = P\left(Z > \frac{\mu_c - \mu_1}{\sigma} \times \sqrt{n}\right)$
- $\mu \geq \mu_0$
  faço -x sinal

Potência teste (percentagem)

$H_0: p \leq p_0$

valor crítico → $p_c = p_0 + z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}$

p_1 - valor real

$\pi(p_1) = P\left(Z > \frac{p_c - p_1}{\sqrt{p_1(1-p_1)}} \times \sqrt{n}\right)$

$p \geq p_0$

...

Erro de estimativa $= z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

P-value (variância)

- bilateral
  $2\min\left(P(X^2_k < X^2_{obs})\right)$
- unilateral
  $P(X^2_k > X^2_{obs})$

$H_0: \mu_1-\mu_2=\mu_0$ ou $H_0: \mu_1-\mu_2\le\mu_0$ ou $H_0: \mu_1-\mu_2\ge\mu_0$		
σ_1^2 e σ_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 $n_2>30$	$Z=\dfrac{\overline{X_1}-\overline{X_2}-\mu_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \overset{.}{\frown} N(0,1)$
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 $n_2>30$	$Z=\dfrac{\overline{X_1}-\overline{X_2}-\mu_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \overset{.}{\frown} N(0,1)$

n - df
$\overline{X}$ - mean
μ_0 - quando não está no enunciado $\mu_0 = 0$
S - std. deviation

	Y_1	Y_2	Y_3	
X_1	O_{11} E_{11}	O_{12} E_{12}	O_{13} E_{13}	$O_{1\cdot}$
X_2	O_{21} E_{21}	O_{22} E_{22}	O_{23} E_{23}	$O_{2\cdot}$
	$O_{\cdot 1}$	$O_{\cdot 2}$	$O_{\cdot 3}$	n

$$E_{12} = \frac{O_{1\cdot} \times O_{\cdot 2}}{n}$$

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

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

$A \cup A = A$ $\quad A \cap A = A$
$A \cup \overline{A} = \Omega$ $\quad A \cap \overline{A} = \emptyset$
$A \cup \emptyset = A$ $\quad A \cap \emptyset = \emptyset$
$A \cup \Omega = \Omega$ $\quad A \cap \Omega = A$

Distribuição Normal
Teorema da Aditividade

$Y = \sum_{i=1}^{n} a_i X_i \sim N(\mu, \sigma)$

com $\mu = \sum_{i=1}^{n} a_i \mu_i$ e

$\sigma = \sqrt{\sum_{i=1}^{n} a_i^2 \sigma_i^2}$

Canónicas
variáveis independentes

1. $\sum_{i=1}^{n} X_i \sim N(\mu; \sqrt{n}\,\sigma)$
2. $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$

Regressão Linear Simples: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

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

mutuamente exclusivos ou disjuntos ou incompatíveis

$A \cap B = \emptyset$
$\hookrightarrow P(A \cup B) = P(A) + P(B)$

$P(\Omega) = 1$
$P(\emptyset) = 0$
$P(\overline{A}) = 1 - P(A)$
$P(B - A) = P(B) - P(A \cap B)$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probabilidade condicionada

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ se } P(B) > 0$$

Acontecimentos independentes

$P(A \cap B) = P(A) \times P(B)$

com $P(A) \geq 0$ e $P(B) \geq 0$

Assim:

$P(A|B) = P(A)$
$P(B|A) = P(B)$