Universidade do Minho
Escola de Engenharia

FORMULÁRIO

1º semestre (2011 /2012)

Índice

FORMULÁRIO DE ESTATÍSTICA DESCRITIVA

1. MEDIDAS DE LOCALIZAÇÃO

1.1. MÉDIA

$\boxed{\bar{x} = \frac{1}{n}\sum_i x_i}$ dados não-agrupados

$\boxed{\bar{x} \approx \sum_k fr_k M_k = \frac{1}{n}\sum_k f_k M_k}$ dados agrupados

1.2. MEDIANA

$\boxed{Med = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}}$ n par

$\boxed{Med = x_{\left(\frac{n+1}{2}\right)}}$ n ímpar

$\boxed{Med = LI + \frac{0.5 - F_{rA}^{\;-}}{F_{rMed}^{\;+} - F_{rA}^{\;-}}\Delta}$ representação histograma

1.3. MODA

$$\boxed{\begin{aligned} &Mod = LI + \frac{d_1}{d_1 + d_2}\Delta \\ &d_1 = f_{Mod} - f_A^{\;-} \\ &d_2 = f_{Mod} - f_D^{\;+} \end{aligned}}$$ representação histograma

2. MEDIDAS DE DISPERSÃO

2.1. QUARTIL

$\boxed{Q_i = x_{\left(n\times\frac{i}{4}\right)}}$

Nota: Q_2=Med

2.2. PERCENTIL

$\boxed{P_i = x_{\left(n\times\frac{i}{100}\right)}}$

Nota: P_{50}=Med; P_{25}=Q_1; P_{75}=Q_3

2.3. VARIÂNCIA

$\boxed{s^2 = \frac{1}{n-1}\sum_i \left(x_i - \bar{x}\right)^2}$ dados não-agrupados

$\boxed{s^2 \approx \frac{n}{n-1}\sum_k fr_k \left(M_k - \bar{x}\right)^2 = \frac{1}{n-1}\sum_k f_k \left(M_k - \bar{x}\right)^2}$ dados agrupados

2.4. AMPLITUDE

$\boxed{A = x_{i_{\text{máx}}} - x_{i_{min}}}$

FORMULÁRIO DE PROBABILIDADES

AXIOMA 1: $0 \le P(A) \le 1$

AXIOMA 2: $P(S) = 1$

AXIOMA 3: se A e B mutuamente exclusivos então $P(A \cup B) = P(A) + P(B)$

Propriedades

1. $P(A) + P(\overline{A}) = 1$
2. $P(\varnothing) = 0$
3. $P(\overline{A} \cap B) = P(B) - P(A \cap B)$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probabilidade Condicional

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

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

Acontecimentos Independentes

$$P(A \mid B) = P(A)$$

$$P(B \mid A) = P(B)$$

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

Teorema de Bayes

$$P(A_N \mid B) = \frac{P(B \mid A_N) \times P(A_N)}{\sum_{I=1}^{N} P(B \mid A_I) \times P(A_I)}$$

FORMULÁRIO DE DISTRIBUIÇÕES DE PROBABILIDADE

DISCRETO

propriedades da função probabilidade:

1) $f(x) \geq 0$

2) $\sum_{x} f(x) = 1$

função probabilidade acumulada

$$F(x) = P(X \leq x) = \sum_{-\infty}^{x} f(x)$$

valor esperado

$$E(x) = \sum_{x} x \times f(x)$$

variância

$$V(x) = E(x^2) - [E(x)]^2$$

$$V(x) = \sum_{x} (x - \mu)^2 f(x)$$

CONTÍNUO

propriedades da função densidade de probabilidade:

1) $f(x) \geq 0$

2) $\int_{x} f(x) dx = 1$

função probabilidade acumulada

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x) dx$$

valor esperado

$$E(x) = \int_{x} x \times f(x) dx$$

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

$$f(x) = \theta^x (1-\theta)^{1-x} \qquad x = 0,1$$

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

Distribuição Binomial [**B(n,p)**]

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,2,\ldots,n$$

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

Distribuição Poisson [**P(λ)**]

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

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

Aproximação da Binomial à Poisson

N grande e p muito pequeno

$\lambda = np$

Distribuição Uniforme (discreta)

$$f(x) = \frac{1}{k} \qquad x = x_1, x_2, \ldots, x_k$$

$$\mu = \sum_i \frac{x_i}{k} \quad \sigma^2 = \sum_i \frac{(x_i - \mu)^2}{k}$$

CONTÍNUAS

Distribuição Uniforme [**U(a,b)**]

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & outros \end{cases}$$

$$\mu = \frac{a+b}{2} \quad \sigma^2 = \frac{(b-a)^2}{12}$$

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 $\begin{cases} np > 5 \\ nq > 5 \end{cases}$

$\mu = np$

$\sigma^2 = npq$

<u>Correcção de Yates</u>

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

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

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
	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 p	Bernoulli	$n > 30$ **(2)**	-	$Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$	$\hat{p} - z_{(1-\alpha/2)} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{(1-\alpha/2)} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	Estimador da proporção binomial $p \approx \hat{p} = \frac{x}{n}$
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 $p_1 - p_2$	Bernoulli	$n_1 \geq 30$ e $n_2 \geq 30$	-	$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0,1)$ **(3)**	$(\hat{p}_1 - \hat{p}_2) \pm z_{(1-\alpha/2)} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	Estimadores das proporções binomiais **(4)** $\hat{p}_1 = \frac{x_1}{n_1}$ e $\hat{p}_2 = \frac{x_2}{n_2}$
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 exacta através da distribuição binomial; **(3)** e **(4)** No teste à diferença de proporções se $H_0 : p_1 - p_2 = 0$, a E.T. passa a ser

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1), \text{ com } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

Teste do "bom ajuste" do Qui-Quadrado para grandes amostras

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

- Probabilidades não totalmente especificadas na hipótese nula

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

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

graus de liberdade = n^o de celas -1- n^o de parâmetros estimados

$Q = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$ com a frequência esperada dada por $e_i = n.p_i$

Tabelas de Contingência

		Característica B					
		B_1	B_2	B_3		B_b	$n_{i.}$
	A_1	f_{11}	f_{12}	f_{13}	$\cdots$	f_{1b}	
Característica	A_2	f_{21}	f_{22}	f_{23}	$\cdots$	f_{2b}	
A	A_3	f_{31}	f_{32}	f_{33}	$\cdots$	f_{3b}	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	
	A_a	f_{a1}	f_{a2}	f_{a3}	$\cdots$	f_{ab}	
	$n_{.j}$						n

1. . Teste de independência

Hipótese nula:

$H_0 : p_{ij} = p_{i.}p_{.j}$ (as variáveis são independentes), $i = 1, ..., a$ e $j = 1, ..., b$

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

2 . Teste de homogeneidade

Hipótese nula:

$H_0 : w_{1j} = w_{2j} = ... = w_{aj}$ (as subpopulações A são equivalentes) para $j = 1, ..., b.$

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

$$Q = \sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$ com a frequência esperadada dada por $e_{ij} = \frac{n_{i.}n_{.j}}{n}$ $\quad i = 1, ..., a \ e \ j = 1, ..., b$

Planeamento completamente aleatório

$$SQT = \sum_{j=1}^{k} n_j(\overline{y}_{.j} - \overline{Y})^2 \qquad SQT = \sum_{j=1}^{k} \frac{T_{.j}^2}{n_j} - \frac{1}{N}T_{..}^2$$
$$STQ = \sum_{j=1}^{k} \sum_{i=1}^{n_j}(y_{ij} - \overline{Y})^2 \qquad STQ = \sum_{j=1}^{k} \sum_{i=1}^{n_j} y_{ij}^2 - \frac{1}{N}T_{..}^2$$
$$SQR = \sum_{j=1}^{k} \sum_{i=1}^{n_j}(y_{ij} - \overline{y}_{.j})^2 \qquad SQR = STQ - SQT \quad \text{com} \quad N = \sum_{j=1}^{k} n_j$$

$T_{.j}$ é o total dos valores obtidos para o tratamento j ; $T_{..}$ é o grande total

Tendo-se $STQ = SQT + SQR$

Modelo populacional: $y_{ij} = \mu + \alpha_j + e_{ij}$

com $i = 1, ..., n_j$ e $j = 1, ..., k$ $\quad e_{ij} \sim N(0, \sigma^2)$

Teste às diferenças entre os tratamentos

$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k = 0$

(não existem diferenças entre as médias das k populações).

H_1 : os efeitos da aplicação dos tratamentos são significativos

(ou existem diferenças entre os tratamentos).

R.R : $F > c$ em que c (Fisher) é determinado por forma a $\alpha = P\left[F > c; H_0\right]$

Tabela ANOVA

Fonte de variação	Soma dos quadrados	graus de liberdade	Média dos quadrados	v.a. F
Tratamentos	SQT	k-1	MQT=SQT/(k-1)	
Resíduos	SQR	$\Sigma n_j - k$	MQR=SQR/$(\Sigma n_j - k)$	F=$\frac{MQT}{MQR}$
Total	STQ	$\Sigma n_j - 1$		

- Intervalos de confiança para diferenças entre pares de médias de tratamentos i e j com

$i \neq j = 1, 2, ..., k$

$$T = \frac{(\overline{y}_{.i} - \overline{y}_{.j}) - (\mu_i - \mu_j)}{\sqrt{\frac{SQR}{N-k}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \sim t_{N-k}$$

Planeamento com blocos aleatórios

$$SQT = b\sum_{j=1}^{k}(\overline{y}_{.j} - \overline{Y})^2 \qquad SQT = \frac{1}{b}\sum_{j=1}^{k} T_{.j}^2 - \frac{1}{kb}T_{..}^2$$
$$SQB = k\sum_{i=1}^{b}(\overline{y}_{i.} - \overline{Y})^2 \qquad SQB = \frac{1}{k}\sum_{i=1}^{b} T_{i.}^2 - \frac{1}{kb}T_{..}^2$$
$$STQ = \sum_{i=1}^{b} \sum_{j=1}^{k}(y_{ij} - \overline{Y})^2 \qquad STQ = \sum_{i=1}^{b} \sum_{j=1}^{k} y_{ij}^2 - \frac{1}{kb}T_{..}^2$$
$$SQR = \sum_{i=1}^{b} \sum_{j=1}^{k}(y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{Y})^2 \qquad SQR = STQ - SQT - SQB$$

$T_{i.}$ é o total dos valores obtidos para o bloco i ; $T_{.j}$ é o total dos valores obtidos para o tratamento j

Modelo populacional: $y_{ij} = \mu + \alpha_j + \beta_i + e_{ij}$

para $i = 1, ..., b$ e $j = 1, ..., k$ $\quad e_{ij} \sim N(0, \sigma^2)$

Teste às diferenças entre os tratamentos

$H_{01} : \alpha_1 = \alpha_2 = ... = \alpha_k = 0$

(não existem diferenças significativas entre os tratamentos).

H_{11} : os efeitos da aplicação dos tratamentos são significativos

(ou existem diferenças entre os tratamentos).

R.R : $F_1 > c$ em que c (Fisher) é determinado a partir de

$\alpha = P\left[F_{1,((k-1),(b-1)(k-1))} > c; H_{01}\right] \quad e \quad F_1 = \frac{MQT}{MQR}$

Teste às diferenças entre os blocos

$H_{02} : \beta_1 = \beta_2 = ... = \beta_b = 0$

(não existem diferenças significativas entre os efeitos dos blocos)

H_{12}: existem diferenças entre os efeitos dos blocos.

R.R : $F_2 > c$ em que c (Fisher) é determinado a partir de

$\alpha = P\left[F_{2,((b-1),(b-1)(k-1))} > c; H_{02}\right] \quad e \quad F_2 = \frac{MQB}{MQR}$

Tabela ANOVA

F. de variação	S. dos quadrados	graus de liberdade	Média dos quadrados	v.a. F
Tratamentos	SQT	k-1	MQT=SQT/(k-1)	
Blocos	SQB	b-1	MQB=SQB/(b-1)	$F_1 = \frac{MQT}{MQR}$
Resíduos	SQR	$(k-1).(b-1)$	MQR=SQR/$(k-1).(b-1)$	$F_2 = \frac{MQB}{MQR}$
Total	STQ	$k.b-1$		

Intervalos de confiança para diferenças entre pares de médias de tratamentos:

$$T = \frac{(\overline{y}_{j1} - \overline{y}_{j2}) - (\mu_{j1} - \mu_{j2})}{\sqrt{MQR(\frac{2}{b})}} \quad \sim \quad t_{(b-1)(k-1)}$$

Planeamento factorial com replicações

$$SQF_A = qr \sum_{i=1}^{p} (\overline{y}_{i..} - \overline{Y})^2$$
$$SQF_B = pr \sum_{j=1}^{q} (\overline{y}_{.j.} - \overline{Y})^2$$
$$SQR = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} (y_{ijk} - \overline{y}_{ij.})^2$$
$$STQ = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} (y_{ijk} - \overline{Y})^2$$

$$SQF_A = \frac{\sum_{i=1}^{p} T_{i.}^2}{rq} - \frac{T_{..}^2}{pqr}$$
$$SQF_B = \frac{\sum_{j=1}^{q} T_{.j}^2}{rp} - \frac{T_{..}^2}{pqr}$$
$$SQR = \sum_{ijk} y_{ijk}^2 - \frac{\sum_{ij} T_{ij}^2}{r}$$
$$STQ = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} y_{ijk}^2 - \frac{T_{..}^2}{pqr}$$
$$SQI_{AB} = STQ - SQF_A - SQF_B - SQR$$

T_{ij} é a soma das observações da célula (i, j)

Modelo populacional

$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$

para $i = 1, ..., p$, $j = 1, ..., q$. e $k = 1, ..., r$ e $e_{ijk} \sim N(0, \sigma^2)$

Testes de hipóteses

- Factor A

$H_{01} : \alpha_1 = \alpha_2 = ... = \alpha_p = 0$

H_{11}: existem diferenças significativas entre os níveis de A

R.R : $F_1 > c$ com c (Fisher) determinado de

$\alpha = Pr[F_{1_{(p-1),pq(r-1)}} > c; H_{01}]$ e $F_1 = \frac{MQF_A}{MQR}$

- Factor B

$H_{02} : \beta_1 = \beta_2 = ... = \beta_q = 0$

H_{12} : existem diferenças significativas entre os níveis de B

R.R : $F_2 > c$ com c (Fisher) determinado de

$\alpha = Pr[F_{2_{(q-1),pq(r-1)}} > c; H_{02}]$ e $F_2 = \frac{MQF_B}{MQR}$

- Interacção AB

$H_{03} : \gamma_{11} = \gamma_{12} = ... = \gamma_{21} = ... = \gamma_{pq} = 0$

H_{13}: existem diferenças significativas devido a interacção

R.R : $F_3 > c$ com c (Fisher) determinado de

$\alpha = Pr[F_{3_{(p-1)(q-1),pq(r-1)}} > c; H_{03}]$ e $F_3 = \frac{MQI_{AB}}{MQR}$

Tabela ANOVA

Fonte de variação	Soma dos Quadrados	graus de lib.	Média dos Quadrados	v.a F
Factor A	SQF$_A$	p-1	MQF$_A$	
Factor B	SQF$_B$	q-1	MQF$_B$	F$_1 = \frac{MQF_A}{MQR}$
Interacção AxB	SQI$_{AB}$	(p-1).(q-1)	MQI$_{AB}$	F$_2 = \frac{MQF_B}{MQR}$
Resíduos	SQR	p.q.(r-1)	MQR	F$_3 = \frac{MQI_{AB}}{MQR}$
Total	STQ	pqr-1		

Testes não paramétricos

Teste de Kruskal-Wallis

H_0: Não existem diferenças significativas entre os efeitos dos tratamentos ou as médias das distribuições das k populações são idênticas

H_1: Nem todas as k distribuições têm médias idênticas.

R.R: $H \geq c$ onde c é determinado de $\alpha = Pr[H \geq c; H_0]$ e

$$H = \frac{12}{n(n+1)}[\frac{W_1^2}{n_1} + \frac{W_2^2}{n_2} + ... + \frac{W_k^2}{n_k}] - 3(n+1) \text{com } W_i = \sum_{j=1}^{n_i} R_{ij} \quad i = 1, 2, ..., k$$

Para $k > 3$ ou $n_1, n_2, ...$ e/ou $n_i > 5$, a distribuição assintótica de H é a χ^2 com $k-1$ graus de liberdade.

A 'estatística' ajustada é

$$H^{'} = \frac{H}{1 - \frac{\sum_{j=1}^{l} q_j(q_j^2 - 1)}{n(n^2-1)}}$$

em que l é o número de conjuntos com observações repetidas existente e q_j é o número de elementos nesse conjunto j $(j = 1, ..., l)$. A 'estatística' $H^{'}$ tem ainda uma distribuição assintótica χ^2_{k-1}.

Planeamento com blocos. Teste de Quade

Os dados consistem num conjunto de b variáveis aleatórias independentes a k dimensões $(y_{i1}, y_{i2}, ..., y_{ik}),\ i = 1, ..., b$, chamadas blocos.

Os cálculos para este teste devem estar assim ordenados:

Amplitude do bloco: A_i	Graduações do bloco de acordo com a sua amplitude: $R(A_i)$	Matriz S_{ij}
$A_i = max_j(y_{ij}) - min_j(y_{ij})$		$S_{ij} = R(A_i)[R(y_{ij}) - \frac{k+1}{2}]$

$R(y_{ij})$ - graduações das observações $y_{ij},\ \ (j = 1, ..., k)$

$S_j = \sum_{i=1}^{b} S_{ij};\quad SQT = \frac{1}{b}\sum_{j=1}^{k} S_j^2;\quad STQ = \sum_{i=1}^{b}\sum_{j=1}^{k} S_{ij}^2$

Se não existirem observações repetidas, STQ reduz-se a

$b(b+1)(2b+1)k(k+1)(k-1)/72.$

Teste de hipóteses

H_0 : Não existem diferenças significativas entre os tratamentos

(ou, os efeitos dos tratamentos são idênticos)

H_1 : Pelo menos um dos tratamentos tende a conseguir valores observados maiores do que um outro tratamento.

R.R : $T > c$ onde c é um ponto crítico da distribuição F que corresponde ao nível de significância α,

com $(k-1)$ e $(b-1)(k-1)$ graus de liberdade

$$T = \frac{(b-1)SQT}{STQ - SQT}$$

Comparações dois a dois

Os tratamentos i e j são considerados significativamente diferentes se

$$|S_i - S_j| > c\sqrt{\frac{2b(STQ - SQT)}{(b-1)(k-1)}}$$

sendo c o ponto crítico da distribuição t-Student, com $(b-1)(k-1)$ graus de liberdade que corresponde a uma região de rejeição de tamanho α (nível de significância)

Testes de ajuste de distribuições

Testes do tipo de Kolmogorov para pequenas amostras

$S(x)$ é a função de distribuição empírica que é definida como fracção dos X_i's (elementos da amostra) que são menores ou iguais a X, para cada X $(-\infty < X < +\infty)$

Dados: $F^*(x)$ é uma função distribuição completamente especificada.

A. Teste bilateral	B. Teste unilateral	C. Teste unilateral
$H_0 : F(x) = F^*(x)\ \forall_x$	$H_0 : F(x) \geq F^*(x)$	$H_0 : F(x) \leq F^*(x)$
$H_1 : F(x) \neq F^*(x)$	$H_1 : F(x) < F^*(x)$	$H_1 : F(x) > F^*(x)$
$T = sup_x \vert\ F^*(x) - S(x)\ \vert$	$T^+ = sup_x[\ F^*(x) - S(x)\ ]$	$T^- = sup_x[\ S(x) - F^*(x)\ ]$

R.R: $T\ \ (T^+\ ou\ T^-) > c$ com c calculado de $\alpha = Prob(Rej\ \ H_0;\ H_0) = Prob(T > c;\ H_0\ \text{ de } \mathbf{A.})$.

Os pontos críticos da distribuição de T (T^+ ou T^-) correspondem a $p = 1 - \alpha$

Teste de Lilliefors para a Normal

DADOS: Os dados consistem numa amostra aleatória $X_1, X_2, ..., X_n$ de tamanho n associada com alguma função distribuição desconhecida $F(x)$.

H$_0$: A amostra aleatória foi retirada de uma distribuição normal, com média e/ou variância não especificadas.

H$_1$: A função distribuição dos $X_i's$ não é normal.

R.R: $T_1 > c$ sendo c o ponto crítico da distribuição de T_1 que corresponde a $p = 1 - \alpha$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{e} \quad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

$$Z_i = \frac{X_i - \overline{X}}{s}, \quad i = 1, 2, ..., n \quad T_1 = sup_z |\ F^*(z) - S(z)\ |$$

Teste de Lilliefors para a exponencial

DADOS: Os dados consistem numa amostra aleatória $X_1, X_2, ..., X_n$ de tamanho n associada com alguma função distribuição desconhecida $F(x)$.

H_0 : A amostra aleatória segue a distribuição exponencial:

$$F(x) = F^*(x) = \begin{cases} 1 - e^{-x/\beta}, & \text{para} \quad x > 0 \\ 0 & \text{para} \quad x < 0 \end{cases}$$

em que β é um parâmetro desconhecido.

H_1: A distribuição dos $X_i's$ não é exponencial.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad Z_i = \frac{X_i}{\overline{X}}, \quad i = 1, 2, ..., n$$

$$F^*(z) = \begin{cases} 1 - e^{-z}, & \text{para} \quad z > 0 \\ 0 & \text{para} \quad z < 0 \end{cases}$$

$$T_2 = sup_z |\ F^*(z) - S(z)\ |$$

R.R: $T_2 > c$ sendo c o ponto crítico da distribuição de T_2 que corresponde a $p = 1 - \alpha$.

Teste a duas distribuições. Amostras independentes.

Teste de Kolmogorov - Smirnov

DADOS: Os dados consistem em duas amostras aleatórias independentes, uma de tamanho n, $X_1, X_2, ..., X_n$ e outra de tamanho m, $Y_1, Y_2, ..., Y_m$ retiradas de duas populações com distribuições $F(x)$ e $G(y)$ (ou $G(x)$) respectivamente.

A. Teste bilateral	B. Teste unilateral	C. Teste unilateral
$H_0 : F(x) = G(x)\ \forall_x$	$H_0 : F(x) \leq G(x)$	$H_0 : F(x) \geq G(x)$
$H_1 : F(x) \neq G(x)$	$H_1 : F(x) > G(x)$	$H_1 : F(x) < G(x)$
$T_1 = sup_x |\ S_1(x) - S_2(x)\ |$	$T_1^+ = sup_x[\ S_1(x) - S_2(x)\ ]$	$T_1^- = sup_x[\ S_2(x) - S_1(x)\ ]$

com $S_1(x)$ a função empírica baseada na amostra $X_1, X_2, ..., X_n$ e $S_2(x)$ a função empírica baseada em $Y_1, Y_2, ..., Y_m$

R.R: T_1 (T_1^+ ou T_1^-) $> c$ sendo c o ponto crítico da distribuição da estatística que corresponde a um nível de significância α.

Teste a k distribuições. Amostras independentes.Teste unilateral de Smirnov

DADOS: k amostras aleatórias de tamanho iguais a n. Se as distribuições empíricas forem, respectivamente, $S_1(x), S_2(x), ..., S_k(x)$, e as funções distribuição $F_1(x)$, $F_2(x), ..., F_k(x)$ representam as k populações, desconhecidas,

$H_0 : F_1(x) \leq F_2(x) \leq ... \leq F_k(x)$ para todo o x

$H_1 : F_i(x) > F_j(x)$ para algum $i < j$ e algum x

R.R: $T_2 > c$ sendo c o ponto crítico, que corresponde a $p = 1 - \alpha$, ao nível de significância α.

$$T_2 = sup_{x,i<k}[\ S_i(x) - S_{i+1}(x)\ ] \quad i = 1, ..., k-1$$

Regressão linear e simples

$Y_i \sim N(\alpha + \beta x_i, \sigma^2)$

$Y_i = \alpha + \beta x_i + e_i$ com $x_i = X_i - \overline{X}, \quad i = 1, .., n$

Os estimadores de máxima verosimilhança, para os parâmetros α, β e σ^2 são

$\tilde{\alpha} = \frac{\sum_{i=1}^n Y_i}{n} = \overline{Y} \quad ; \quad \tilde{\beta} = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2} = \frac{\sum_{i=1}^n (X_i - \overline{X}) Y_i}{\sum_{i=1}^n (X_i - \overline{X})^2};$

$\tilde{\sigma}^2 = \frac{1}{(n-2)} \sum_{i=1}^n [Y_i - \tilde{\alpha} - \tilde{\beta}(X_i - \overline{X})]^2$

Testes de hipóteses

$$T_1 = \frac{\tilde{\alpha} - \alpha}{\sqrt{\frac{\tilde{\sigma}^2}{n}}} \sim t_{n-2} \quad ; \qquad T_2 = \frac{\tilde{\beta} - \beta}{\sqrt{\frac{\tilde{\sigma}^2}{\sum_1^n (X_i - \overline{X})^2}}} \sim t_{n-2}$$

$H_0 : \beta = 0$

$H_1 : \beta > 0 \ \ (ou \quad \beta \neq 0)$

R.R: $T_2 \geq c$ com $c = t_{n-2,\alpha}$ (ou $\alpha/2$)

Do mesmo modo, a 'estatística' T_1 pode ser usada para calcular intervalos de confiança e testes de hipóteses relacionados com o parâmetro α

Média e variância de um valor estimado de Y:

$$E[Y_0] = E[\tilde{\alpha}] + (X_0 - \overline{X}) E[\tilde{\beta}] = \alpha + \beta(X_0 - \overline{X})$$

$$var[Y_0] = \sigma^2 \left(\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\right)$$

Regressão linear e múltipla

$$Y_i = \alpha + \beta x_i + \gamma z_i + e_i$$

em que $x_i = X_i - \overline{X}, z_i = Z_i - \overline{Z}$ e e_i é o erro aleatório de observação,normalmente distribuído com média zero e variância comum σ^2 $(i = 1, ..., n)$.

$E[Y] = \alpha + \beta(X - \overline{X}) + \gamma(Z - \overline{Z})$.

$\tilde{\alpha} = \frac{\sum_{i=1}^{n} Y_i}{n} = \overline{Y} \quad ; \quad \tilde{\sigma}^2 = \frac{1}{(n-3)} \sum_{i=1}^{n} (Y_i - \tilde{\alpha} - \tilde{\beta} x_i - \tilde{\gamma} z_i)^2$

$$\begin{cases} \sum_{i=1}^{n} x_i Y_i = \tilde{\beta} \sum_{i=1}^{n} x_i^2 + \tilde{\gamma} \sum_{i=1}^{n} x_i z_i \\ \sum_{i=1}^{n} z_i Y_i = \tilde{\beta} \sum_{i=1}^{n} x_i z_i + \tilde{\gamma} \sum_{i=1}^{n} z_i^2 \end{cases}$$

As 'estatísticas' T_1, T_2 e T_3 para testes de hipóteses e intervalos de confiança, em relação, respectivamente, aos parâmetros α, β e γ, são:

$T_1 = \frac{\tilde{\alpha} - \alpha}{\sqrt{\frac{\tilde{\sigma}^2}{n}}}; \qquad T_2 = \frac{\tilde{\beta} - \beta}{\sqrt{\frac{\tilde{\sigma}^2}{\sum x_i^2 - \frac{(\sum x_i z_i)^2}{\sum z_i^2}}}}; \qquad T_3 = \frac{\tilde{\gamma} - \gamma}{\sqrt{\frac{\tilde{\sigma}^2}{\sum z_i^2 - \frac{(\sum x_i z_i)^2}{\sum x_i^2}}}}$ e seguem a distribuição t-Student com $n-3$ graus de liberdade.

Regressão não-linear

i) $E[Y_i] = \alpha + \beta X_i^2$

O modelo matemático geral, é: $Y_i = \alpha + \beta w_i + \gamma w_i^2 + e_i$ com $w_i = W_i - \overline{W}$ e o $e_i \sim N(0, \sigma^2)$ $(i = 1, ..., n)$. Define-se $X = W$ e $Z = W^2$, o que reduz este caso à regressão múltipla e linear.

ii) $E[Y_i] = X_i^{\beta}$

O modelo matemático mais geral e comum é: $Y_i = \alpha e^{\beta X_i} u_i$. Os erros aleatórios u_i $(i = 1, ..., n)$ têm agora uma distribuição, em geral não simétrica e centrada em 1.

Lineariza-se o modelo passando-se a ter: $lnY_i = ln\alpha + \beta X_i + lnu_i$ e aplica-se a análise de regressão linear e simples.

Testes de independência estocástica

• Coeficiente de correlação da amostra. Teste de Pearson

$X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

R.R: $|R| \geq c$. O valor de c é determinado de $\alpha = P_r[|R| \geq c; H_0]$

$$R = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sqrt{(\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n})(\sum_{i=1}^{n} Y_i^2 - \frac{(\sum_{i=1}^{n} Y_i)^2}{n})}}$$

e que é o coeficiente de correlação da amostra de Pearson.

A variável $T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \sim t_{n-2}$ e o teste resume-se a, rejeitar H_0 se $|T| \geq c$ com c determinado de $\alpha = P_r[|T| \geq c; H_0]$.

- Correlação de Spearman

$(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ amostra aleatória bivariada, de tamanho n

$R(X_i)$ graduação do valor de X_i

$R(Y_i)$ graduação do valor de Y_i $(i = 1, 2, ..., n)$

A medida de correlação de Spearman R_S é definida por

$$R_S = \frac{\sum_{i=1}^{n} [R(X_i) - \frac{n+1}{2}][R(Y_i) - \frac{n+1}{2}]}{\frac{n(n^2-1)}{12}}$$

ou

$$R_S = 1 - \frac{6T}{n(n^2-1)} \quad \text{com} \quad T = \sum_{i=1}^{n} [R(X_i) - R(Y_i)]^2$$

caso não existam observações repetidas. Existindo repetições deve usar-se a expressão:

$$R_S = \frac{\sum_{i=1}^{n} R(X_i)R(Y_i) - n(\frac{n+1}{2})^2}{\sqrt{\sum_{i=1}^{n} R(X_i)^2 - n(\frac{n+1}{2})^2} . \sqrt{\sum_{i=1}^{n} R(Y_i)^2 - n(\frac{n+1}{2})^2}}$$

A. Teste bilateral

H_0: As variáveis X e Y são independentes.

H_1: (a) Existe uma tendência para os maiores valores de X formarem pares com os maiores valores de Y, ou

(b) Existe uma tendência para os menores valores de X formarem pares com os maiores valores de Y.

R.R: $R_S > c_1$ ou $R_S < c_2$, sendo c_1 o ponto crítico que corresponde a $1 - \frac{\alpha}{2}$ e c_2 o ponto crítico que corresponde a $\frac{\alpha}{2}$

B. Teste unilateral para correlação positiva

H_0: As variáveis X e Y são independentes.

H_1: Existe uma tendência para os maiores valores de X e de Y formarem pares.

R.R: $R_S > c$, em que c é o ponto crítico que corresponde a $1 - \alpha$

C. Teste unilateral para correlação negativa

H_0: As variáveis X e Y são independentes.

H_1: Existe uma tendência para os menores valores de X formarem pares com os maiores valores de Y e vice-versa.

R.R: $R_S < c$ sendo c o ponto crítico que corresponde a α.