INSTITUTO SUPERIOR POLITÉCNICO DE VISEU
ESCOLA SUPERIOR DE TECNOLOGIA
Departamento de Matemática
ESTATÍSTICA APLICADA

FORMULÁRIO

k - nº de amostras; ***n_j*** - nº de observações na amostra *j*; **N**$=\sum_{j=1}^{k} n_j$ - nº total de observações

Estatística de teste do **TESTE DE BARTLETT**: $B=\frac{1}{C}\left[(N-k)\ln\left(S_p^2\right)-\sum_{j=1}^{k}(n_j-1)\ln\left(S_j^2\right)\right] \underset{\text{Sob } H_0}{\sim} \chi^2_{k-1}$

$$S_j^2=\frac{1}{n_j-1}\sum_{i=1}^{n_j}(X_{ij}-\overline{X}_j)^2\,,\quad S_p^2=\frac{1}{N-k}\sum_{j=1}^{k}(n_j-1)\,S_j^2 \quad \text{e} \quad C=1+\frac{1}{3(k-1)}\left[\sum_{j=1}^{k}\frac{1}{n_j-1}-\frac{1}{N-k}\right].$$

Estatística de teste do **TESTE DE LEVENE**: $W=\frac{N-k}{k-1}\cdot\frac{\sum_{j=1}^{k} n_j(\overline{Z}_j-\overline{Z})^2}{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Z_{ij}-\overline{Z}_j)^2} \underset{\text{Sob } H_0}{\sim} F^{k-1}_{N-k}$

$Z_{ij}=\left|X_{ij}-\overline{X}_j\right|$, $\quad \overline{Z}_j=\frac{1}{n_j}\sum_{i=1}^{n_j} Z_{ij}$ - média dos Z_{ij} na amostra *j*

e $\quad \overline{Z}=\frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{n_j} Z_{ij}$ - média global dos Z_{ij} .

ANOVA COM UM FACTOR

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Variância (Soma Média de Quadrados)	Razão F
Entre grupos	**SS**$_A$=$\sum_{j=1}^{k} n_j\left(\overline{x}_j-\overline{\overline{x}}\right)^2$	k-1	$MS_A=\frac{SS_A}{k-1}$	$\frac{MS_A}{MS_E}$
Dentro dos grupos	**SS**$_E$=$\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\overline{x}_j\right)^2$	N-k	$MS_E=\frac{SS_E}{N-k}$	
Total	**SS**$_T$=$\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\overline{\overline{x}}\right)^2$	N-1		

$$\overline{\overline{\mathbf{x}}}=\frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}=\frac{1}{N}\sum_{j=1}^{k} n_j\overline{x}_j$$

Teste HSD de Tuckey

A hipótese H$_0$: $\mu_i=\mu_j$ é rejeitada se $\left|\overline{x}_i-\overline{x}_j\right|\geq S_{T(1-\alpha)}\cdot\sqrt{\frac{MS_E}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$.

onde $S_{T(1-\alpha)}$ é tal que $P\left(W\leq S_{T(1-\alpha)}\right)=1-\alpha$ com $W\sim S_{T(k,N\text{-}k)}$.
Note que se as amostras tiverem todas o mesmo tamanho *n*, fica:

A hipótese H$_0$: $\mu_i=\mu_j$ é rejeitada se $\left|\overline{x}_i-\overline{x}_j\right|\geq S_{T(1-\alpha)}\cdot\sqrt{\frac{MS_E}{n}}$.

onde $S_{T(1-\alpha)}$ é tal que $P\left(W\leq S_{T(1-\alpha)}\right)=1-\alpha$ com $W\sim S_{T(k,N\text{-}k)}$ e *N*=*nk*.

Teste de Scheffé

A hipótese nula H_0: $\mu_i = \mu_j$ é rejeitada se

$$\left|\bar{x}_i - \bar{x}_j\right| \geq \sqrt{(k-1)F_{(1-\alpha)}} \cdot \sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

onde, $F_{(1-\alpha)}$ é o quantil de probabilidade (1-α) da distribuição F_{N-k}^{k-1}, isto é, $P\left(F_{N-k}^{k-1} \leq F_{(1-\alpha)}\right) = 1-\alpha$.

ANOVA COM DOIS FACTORES

Modelo sem interacção (uma observação por célula)

a - nº de níveis do factor A; ***b*** - nº de níveis do factor B.

Fonte de Variação	Soma dos Quadrados (SS)	Graus de Liberdade	Variância (Soma Média dos Quadrados)	Razões F
Entre Grupos **Factor A**	$\mathbf{SS_A}=\mathbf{b}\sum_{i=1}^{a}(\bar{x}_{i\bullet}-\bar{\bar{x}})^2$	**a-1**	$\mathbf{MS_A}=\frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
Entre Grupos **Factor B**	$\mathbf{SS_B}=\mathbf{a}\sum_{j=1}^{b}(\bar{x}_{\bullet j}-\bar{\bar{x}})^2$	**b-1**	$\mathbf{MS_B}=\frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
Residual	$\mathbf{SS_E}=\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij}-\bar{x}_{i\bullet}-\bar{x}_{\bullet j}+\bar{\bar{x}})^2$	**(a-1)(b-1)**	$\mathbf{MS_E}=\frac{SS_E}{(a-1)(b-1)}$	
Total	$\mathbf{SS_T}=\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij}-\bar{\bar{x}})^2$	**ab-1**		

$$\bar{x}_{\bullet j}=\frac{1}{a}\sum_{i=1}^{a}x_{ij}, \quad \bar{x}_{i\bullet}=\frac{1}{b}\sum_{j=1}^{b}x_{ij} \quad \textbf{e} \quad \bar{\bar{x}}=\frac{1}{a+b}\left(\sum_{j=1}^{b}\bar{x}_{\bullet j}+\sum_{i=1}^{a}\bar{x}_{i\bullet}\right)=\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}$$

Teste HSD de Tuckey

Para o Factor A:

A hipótese nula H_0: $\mu_r = \mu_s$ (os grupos *r* e *s* do factor A têm médias iguais) é rejeitada se

$$\left|\bar{x}_{r\bullet}-\bar{x}_{s\bullet}\right| \geq S_{T(1-\alpha)}\sqrt{\frac{MS_E}{b}}$$

onde $S_{T(1-\alpha)}$ é o quantil de probabilidade (1-α) da distribuição da "Studentized Range" com (*a*, *(a-1)(b-1)*) graus de liberdade;

Para o Factor B:

A hipótese nula H_0: $\mu_r = \mu_s$ (os grupos *r* e *s* do factor B têm médias iguais) é rejeitada se

$$\left|\bar{x}_{\bullet r}-\bar{x}_{\bullet s}\right| \geq S_{T(1-\alpha)}\sqrt{\frac{MS_E}{a}}$$

onde $S_{T(1-\alpha)}$ é o quantil de probabilidade (1-α) da distribuição da "Studentized Range" com (*b*, *(a-1)(b-1)*) graus de liberdade;

departamento de matemática

Modelo com interacção (n observações por célula)

n - nº de observações em cada célula; ***a*** - nº de níveis do factor A; ***b*** - nº de níveis do factor B.

Fonte de Variação	Soma dos Quadrados (SS)	Graus de Liberdade	Variância (Soma Média dos Quadrados)	Razões F
Factor A	$SS_A = nb\sum_{i=1}^{a}(\bar{x}_{i\bullet\bullet}-\bar{\bar{x}})^2$	**a-1**	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
Factor B	$SS_B = na\sum_{j=1}^{b}(\bar{x}_{\bullet j\bullet}-\bar{\bar{x}})^2$	**b-1**	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
Interacção	$SS_I = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij\bullet}-\bar{x}_{i\bullet\bullet}-\bar{x}_{\bullet j\bullet}+\bar{\bar{x}})^2$	**(a-1)(b-1)**	$MS_I = \frac{SS_I}{(a-1)(b-1)}$	$\frac{MS_I}{MS_E}$
Residual	$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{ij\bullet})^2$	**ab(n-1)**	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\bar{\bar{x}})^2$,	**abn-1**		

$$\bar{x}_{\bullet j\bullet} = \frac{1}{na}\sum_{i=1}^{a}\sum_{k=1}^{n}x_{ijk} \quad , \quad \bar{x}_{i\bullet\bullet} = \frac{1}{nb}\sum_{j=1}^{b}\sum_{k=1}^{n}x_{ijk} \quad , \quad \bar{x}_{ij\bullet} = \frac{1}{n}\sum_{k=1}^{n}x_{ijk} \quad e \quad \bar{\bar{x}} = \frac{1}{nab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}x_{ijk} .$$

Teste HSD de Tuckey

Para o Factor A:

A hipótese nula H_0: $\mu_r = \mu_s$ (os grupos r e s do factor A têm médias iguais) é rejeitada se

$$\left|\bar{x}_{r\bullet\bullet} - \bar{x}_{s\bullet\bullet}\right| \geq S_{T(1-\alpha)}\sqrt{\frac{MS_E}{nb}}$$

onde, $S_{T(1-\alpha)}$ é o quantil de probabilidade $(1-\alpha)$ da distribuição da "Studentized Range" com (a, $ab(n-1)$) graus de liberdade.

Para o Factor B:

A hipótese nula H_0: $\mu_r = \mu_s$ (os grupos r e s do factor B têm médias iguais) é rejeitada se

$$\left|\bar{x}_{\bullet r\bullet} - \bar{x}_{\bullet s\bullet}\right| \geq S_{T(1-\alpha)}\sqrt{\frac{MS_E}{na}}$$

onde, $S_{T(1-\alpha)}$ é o quantil de probabilidade $(1-\alpha)$ da distribuição da "Studentized Range" com (b, $ab(n-1)$) graus de liberdade.

ANÁLISE DE REGRESSÃO E CORRELAÇÃO

$$b = [b_0 \quad b_1 \quad \cdots \quad b_k]^t = \left(X^t X\right)^{-1} X^t y$$

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} d_i^{\,2} \qquad SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \qquad r^2 = \frac{SSR}{SST}$$

departamento de matemática

<u>*Teste de Durbin-Watson*</u>: $d=\dfrac{\sum_{i=1}^{n-1}\left(d_{i+1}-d_i\right)^2}{\sum_{i=1}^{n} d_i^{\,2}}$

<u>*Testes sobre os coeficientes de regressão*</u>: $\dfrac{\hat{\beta}_i-\beta_i^0}{S_{\hat{\beta}_i}} \underset{\text{Sob } H_0}{\sim} t_{n-k-1}$, com $S_{\hat{\beta}_i}=S\sqrt{c_{ii}}$ onde c_{ii} é o elemento diagonal da linha i+1 da matriz $\left(X^tX\right)^{-1}$ e $S^2=\dfrac{SSE}{n-k-1}$

<u>*Teste F*</u>: $F=\dfrac{SSR/k}{SSE/(n-k-1)}=\dfrac{SSR/k}{S^2}=\dfrac{n-k-1}{k}\times\dfrac{R^2}{1-R^2}\underset{\text{Sob } H_0}{\sim} F^k_{n-k-1}$

<u>***Na Regressão Simples***</u>:

$$b_1=\frac{\sum_{i=1}^{n}x_iy_i-n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n}x_i^{\,2}-n\,\bar{x}^{\,2}};\quad b_0=\bar{y}-b_1\bar{x}\ ;\ r^2=\frac{b_0\sum_{i=1}^{n}y_i+b_1\sum_{i=1}^{n}y_ix_i-n\bar{y}^2}{\sum_{i=1}^{n}y_i^{\,2}-n\bar{y}^2}$$

$$S_{\hat{\beta}_0}^{\ 2}=S^2\frac{\sum_{i=1}^{n}x_i^{\,2}}{n\sum_{i=1}^{n}x_i^{\,2}-n^2\bar{x}^2}\qquad S_{\hat{\beta}_1}^{\ 2}=S^2\frac{1}{\sum_{i=1}^{n}x_i^{\,2}-n\bar{x}^2};$$

MEDIDAS DE ASSOCIAÇÃO

Coeficiente de Contingência de Pearson: $C=\sqrt{\dfrac{\chi^2}{\chi^2+n}}$; $0\le C\le\sqrt{\dfrac{q-1}{q}}$ onde q=min{r,s}

Coeficiente de Tshuprow: $T=\sqrt{\dfrac{\chi^2}{n\sqrt{(r-1)(s-1)}}}$

Coeficiente V de Cramer: $V=\sqrt{\dfrac{\chi^2}{n(q-1)}}$ onde q=min{r,s}

TESTE DE KRUSKAL-WALLIS

Quando **não há empates** ou o nº de empates é muito pequeno, a estatística de teste é:

$$H=\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}-3(N+1).$$

Se $\mathbf{n_i\geq 5}$, i=1,...,k então, sob H_0, H tem aproximadamente distribuição χ^2_{k-1}.

Quando **há muitos empates** a estatística de teste a usar deve ser:

$$H'=\frac{1}{S^2}\left[\sum_{i=1}^{k}\frac{R_i^2}{n_i}-\frac{N(N+1)^2}{4}\right],\text{ onde}\qquad S^2=\frac{1}{N-1}\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}R_{ij}^2-\frac{N(N+1)^2}{4}\right].$$

Se $\mathbf{n_i\geq 5}$, i=1,...,k, então, sob H_0, H' tem aproximadamente distribuição χ^2_{k-1}.