FORMULARIO ANALISIS DE DATOS CURSO 2003-04

Anova

Anova de una vía

$$SCT = \sum\sum(Y_{ij} - \overline{Y})^2 \qquad MCT = \frac{SCT}{N-1}$$

$$SCE = \sum n_j(\overline{Y}_j - \overline{Y})^2 \qquad MCE = \frac{SCE}{k-1} \qquad F = \frac{MCE}{MCR} \sim F_{k-1,N-k}$$

$$SCR = \sum\sum(Y_{ij} - \overline{Y}_j)^2 \qquad MCR = \frac{SCR}{N-k}$$

Test de Bartlet

$$s_j^2 = \frac{\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_j)^2}{n_j} \qquad C = 2.3026\left[(N-k)\ \mathrm{Log}\left(\frac{\sum_{j=1}^{k} n_j s_j^2}{N-k}\right) - \sum_{j=1}^{k}(n_j-1)\mathrm{Log}\ \frac{(n_j-1)}{n_j}s_j^2\right]$$

$$A = 1 + \frac{1}{3(k-1)}\left[\sum_{j=1}^{k}\frac{1}{n_j-1} - \frac{1}{\sum_{j=1}^{k}(n_j-1)}\right]$$

El estadístico de contraste es : $B = \frac{C}{A} \sim \chi^2_{k-1}$

Test de comparaciones múltiples de Sheffé

$$S = \frac{\left|\overline{Y}_j - \overline{Y}_i\right|}{\sqrt{MCR\left(\frac{1}{n_j} + \frac{1}{n_i}\right)}} \qquad \frac{S^2}{k-1} \sim F_{k-1,N-k}$$

Anova de dos víasy una observación por celda

$$SCT = \sum\sum(Y_{ij} - \overline{Y})^2 \qquad MCT = \frac{SCT}{ab-1}$$

$$SCA = \sum b(\overline{A}_j - \overline{Y})^2 \qquad MCA = \frac{SCA}{a-1} \qquad F = \frac{MCA}{MCR} \sim F_{a-1,(a-1)(b-1)}$$

$$SCB = \sum a(\overline{B}_i - \overline{Y})^2 \qquad MCB = \frac{SCB}{b-1} \qquad F = \frac{MCB}{MCR} \sim F_{b-1,(a-1)(b-1)}$$

$$SCR = \sum\sum(Y_{ij} - \overline{A}_j - \overline{B}_i)^2 \qquad MCR = \frac{SCR}{(b-1)(a-1)}$$

Test de comparciones múltiples de Scheffé

a) $S = \frac{\left|\overline{A}_j - \overline{A}_l\right|}{\sqrt{MCR\left(\frac{2}{b}\right)}}$ b) $S = \frac{\left|\overline{B}_j - \overline{B}_l\right|}{\sqrt{MCR\left(\frac{2}{a}\right)}}$

---

a) $\frac{S^2}{a-1} \sim F_{a-1,(a-1)(b-1)}$ b) $\frac{S^2}{b-1} \sim F_{b-1,(a-1)(b-1)}$

Anova de dos vías con r observaciones por celda efectos fijos

$$SCT = \sum\sum\sum(Y_{ijl} - \overline{Y})^2 \qquad MCT = \frac{SCT}{rab-1}$$

$$SCA = \sum rb(\overline{A}_j - \overline{Y})^2 \qquad MCA = \frac{SCA}{a-1} \qquad F = \frac{MCA}{MCR} \sim F_{a-1,ab(r-1)}$$

$$SCB = \sum ra(\overline{B}_i - \overline{Y})^2 \qquad MCB = \frac{SCB}{b-1} \qquad F = \frac{MCB}{MCR} \sim F_{b-1,ab(r-1)}$$

$$\text{SCinter} = \sum\sum r(\overline{Y}_{il} - \overline{A}_j - \overline{B}_i + \overline{Y})^2 \qquad MC\text{inter} = \frac{SC\text{inter}}{(a-1)(b-1)} \qquad F = \frac{MC\text{inter}}{MCR} \sim F_{(a-1)(b-1),ab(r-1)}$$

$$SCR = \sum\sum\sum(Y_{ijl} - \overline{Y}_{ij})^2 \qquad MCR = \frac{SCR}{ab(r-1)}$$

Test de comparaciones múltiples de Scheffé

$$S = \frac{\left|\overline{A}_j - \overline{A}_i\right|}{\sqrt{MCR\frac{2}{rb}}} \qquad \frac{S^2}{a-1} \sim F_{a-1,ab(r-1)}$$

$$S = \frac{\left|\overline{B}_j - \overline{B}_i\right|}{\sqrt{MCR\frac{2}{rb}}} \qquad \frac{S^2}{b-1} \sim F_{b-1,ab(r-1)}$$

$$S = \frac{\left|\overline{Y}_{ij} - \overline{Y}_{kl}\right|}{\sqrt{MCR\frac{2}{r}}} \qquad \frac{S^2}{ab-1} \sim F_{ab-1,ab(r-1)}$$

Efectos aleatorios en los dos factores:

$$F = \frac{MCA}{MC\text{inter}} \sim F_{a-1,(a-1)(b-1))} \qquad F = \frac{MCB}{MC\text{inter}} \sim F_{b-1,(a-1)(b-1)} \qquad F = \frac{MC\text{inter}}{MCR} \sim F_{(a-1)(b-1),ab(r-1)}$$

Efectos aleatorios en el factor B:

$$F = \frac{MCA}{MC\text{inter}} \sim F_{a-1,(a-1)(b-1))} \qquad F = \frac{MCB}{MCR} \sim F_{b-1,ab(r-1)} \qquad F = \frac{MC\text{inter}}{MCR} \sim F_{(a-1)(b-1),ab(r-1)}$$

Regresión

Estimadores puntuales de los parámetros en regresión simple

$$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\sigma}^2 = s_r^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

Coeficiente de determinación

$$R^2 = 1 - \frac{(n-2)s_r^2}{ns_y^2} = r_{xy}^2$$

Distribuciones muestrales

$$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{ns_x^2}) \qquad \hat{\beta}_0 \sim N(\beta_0, \frac{\sigma^2(\bar{x}^2 + s_x^2)}{ns_x^2}) \qquad \frac{(\text{n - 2})\text{s}_\text{r}^2}{\sigma^2} \sim \chi_{n-2}^2$$

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{s_r^2}{ns_x^2}}} \sim t_{n-2} \qquad \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{s_r^2(\bar{x}^2 + s_x^2)}{ns_x^2}}} \sim t_{n-2}$$

Intervalos de confianza para β_1 y β_0 a nivel 1-α

$$\left[\hat{\beta}_1 -_{1-\alpha/2} t_{n-2}\sqrt{\frac{s_r^2}{ns_x^2}}, \hat{\beta}_1 +_{1-\alpha/2} t_{n-2}\sqrt{\frac{s_r^2}{ns_x^2}}\right] \quad \left[\hat{\beta}_0 -_{1-\alpha/2} t_{n-2}\sqrt{\frac{s_r^2(\bar{x}^2 + s_x^2)}{ns_x^2}}, \hat{\beta}_0 +_{1-\alpha/2} t_{n-2}\sqrt{\frac{s_r^2(\bar{x}^2 + s_x^2)}{ns_x^2}}\right]$$

Regiones de rechazo para el contraste $H_0:\beta_1=b_1$ y $H_0: \beta_0=b_0$ a nivel α

$$R_1 = \{\ \hat{\beta}_1 : \left|\hat{\beta}_1 - \text{b}_1\right| >_{1-\alpha/2} t_{n-2}\sqrt{\frac{s_r^2}{ns_x^2}}\ \} \qquad R_1 = \{\ \hat{\beta}_0 : \left|\hat{\beta}_0 - \text{b}_0\right| >_{1-\alpha/2} t_{n-2}\sqrt{\frac{s_r^2(\bar{x}^2 + s_x^2)}{ns_x^2}}\ \}$$

ANOVA para el contraste sobre bondad de ajuste lineal

$$SCT = \sum (y_i - \bar{y})^2 \qquad MCT = \frac{SCT}{n-1}$$

$$SCReg = \sum (\hat{y}_i - \bar{y})^2 \qquad MCReg = \frac{SCReg}{1} \qquad F = \frac{MCReg}{MCR} \sim F_{1,n-2}$$

$$SCR = \sum (y_i - \hat{y}_i)^2 \qquad MCR = \frac{SCR}{n-2}$$

Intervalos de confianza para la media y para la predicción en $x=x_0$

$$\left[\hat{y}_0 -_{1-\alpha/2} t_{n-2}\sqrt{s_r^2(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{ns_x^2})}, \hat{y}_0 +_{1-\alpha/2} t_{n-2}\sqrt{s_r^2(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{ns_x^2})}\right]$$

$$\left[\hat{y}_0 -_{1-\alpha/2} t_{n-2}\sqrt{s_r^2(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{ns_x^2})}, \hat{y}_0 +_{1-\alpha/2} t_{n-2}\sqrt{s_r^2(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{ns_x^2})}\right]$$

Regresión Múltiple

Estimadores puntuales de los parámetros en regresión múltiple

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \hat{\sigma}^2 = s_r^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-(p+1)}$$

Distribuciones muestrales

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2((X'X)^{-1})_{ii}) \qquad \frac{(n-p-1)s_r^2}{\sigma^2} \sim \chi^2_{n-p-1}$$

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{s_r^2((X'X)^{-1})_{ii}}} \sim t_{n-p-1}$$

Coeficiente de determinación y coeficiente de determinación corregido

$$R^2 = 1 - \frac{(n-p-1)s_r^2}{ns_y^2} \qquad \overline{R}^2 = 1 - \frac{s_r^2}{\frac{ns_y^2}{n-1}}$$

Intervalos de confianza para los parámetros a nivel 1-α

$$\left[\hat{\beta}_i - {}_{1-\alpha/2}t_{n-p-1}\sqrt{s_r^2((X'X)^{-1})_{ii}},\ \hat{\beta}_i + {}_{1-\alpha/2}t_{n-p-1}\sqrt{s_r^2((X'X)^{-1})_{ii}}\right]$$

Regiones de rechazo para el contraste H_0:β_ι=b_i a nivel α

$$R_1 = \left\{ \hat{\beta}_i : \left|\hat{\beta}_i - b_i\right| > {}_{1-\alpha/2}t_{n-p-1}\sqrt{s_r^2((X'X)^{-1})_{ii}} \right\}$$

ANOVA para el contraste sobre bondad de ajuste lineal

$$SCT = \sum (y_i - \bar{y})^2 \qquad MCT = \frac{SCT}{n-1}$$

$$SCReg = \sum (\hat{y}_i - \bar{y})^2 \qquad MCReg = \frac{SCReg}{p} \qquad F = \frac{MCReg}{MCR} \sim F_{p,n-p-1}$$

$$SCR = \sum (y_i - \hat{y}_i)^2 \qquad MCR = \frac{SCR}{n-p-1}$$

Intervalos de confianza a nivel 1-α para la media y para la predicción en x=x_0

$$\left[\hat{y}_0 - {}_{1-\alpha/2}t_{n-p-1}\sqrt{s_r^2 x'_0 (X'X)^{-1}x_0},\ \hat{y}_0 + {}_{1-\alpha/2}t_{n-p-1}\sqrt{s_r^2 x'_0 (X'X)^{-1}x_0}\right]$$

$$\left[\hat{y}_0 - {}_{1-\alpha/2}t_{n-p-1}\sqrt{s_r^2(1 + x'_0 (X'X)^{-1}x_0)},\ \hat{y}_0 + {}_{1-\alpha/2}t_{n-p-1}\sqrt{s_r^2(1 + x'_0 (X'X)^{-1}x_0)}\right]$$

Análisis discriminante

x_{il} vector de observaciones del individuo i que está en la población l. $i = 1,\ldots,n_l$ $l = 1,\ldots,k$

Método de Fisher:

$$D_l = (x - \overline{x}^{(l)})'\hat{\Sigma}^{-1}(x - \overline{x}^{(l)}) \qquad \text{con} \qquad \hat{\Sigma} = \frac{\sum_{l=1}^{k} n_l S_l}{n-k}$$

Matrices de varianzas-covarianzas dentro (W) y entre (B) grupos

$$\mathrm{W} = \sum_{l=1}^{\mathrm{k}} \sum_{i=1}^{n_l} (x_{il} - \overline{x}^{(l)})(x_{il} - \overline{x}^{(l)})' = \sum_{l=1}^{k} n_l S_l \qquad \mathrm{B} = \sum_{l=1}^{k} n_l (\overline{x}^{(l)} - \overline{x})(\overline{x}^{(l)} - \overline{x})'$$

Transformación a variedades canónicas

$$Z_j = \sqrt{n-k}\,\gamma_j' X - \sqrt{n-k}\,\gamma_j' \overline{X} \qquad j = 1,\ldots,s \quad s = min(p, k\text{-}1)$$

con γ_j' el vector propio de $\mathrm{W}^{-1}\mathrm{B}$ ordenados de mayor a menor valor propio asociado

Estadístico Lambda de Wilks

$$\Lambda = \prod_{j=i}^{s} \frac{1}{1+\hat{\lambda}_j} \qquad -(\mathrm{n} - \frac{\mathrm{p}+\mathrm{k}}{2} - 1) Ln\Lambda \sim \chi^2_{(p-i-1)(k-i-2)}$$