Diseño de experimentos (un factor)

Modelo: $Y_i = \mu_i + U$; $U \sim N(0;\sigma)$ $i = 1, ..., I$

Muestra: $Y_{ij} \sim N(\mu_i;\sigma^2)$ independientes; $i = 1, ..., I$; $j = 1., ..., n_i$; $\sum_i n_i = n$.

$$\bar{y}_{i.} = \frac{1}{n_i}\sum_j y_{ij} \ ; \qquad \bar{y}_{..} = \frac{1}{n}\sum_i n_i \bar{y}_{i.}$$

$$\hat{\mu}_i = \bar{y}_{i.} = \frac{1}{n_i}\sum_j y_{ij} \ , \ i = 1, ..., I \qquad \hat{\sigma}^2 = S_R^2 = \frac{1}{n-I}\sum_i\sum_j (y_{ij} - \bar{y}_{i.})^2$$

$$IC_{1-\alpha}(\mu_i) = \left(\bar{y}_{i\cdot} \pm t_{n-I;\alpha/2}\ S_R\sqrt{\tfrac{1}{n_i}}\right) \qquad IC_{1-\alpha}(\sigma^2) = \left(\tfrac{(n-I)}{\chi^2_{n-I;\alpha/2}} S_R^2 \ ; \tfrac{(n-I)}{\chi^2_{n-I;1-\alpha/2}} S_R^2\right)$$

Tabla ANOVA

Suma de cuadrados	g.l.	Varianza	Estadístico
$SCE = \sum_i n_i(\bar{y}_{i.} - \bar{y}_{..})^2$	$I-1$	$\frac{SCE}{I-1}$	$F = \frac{SCE/(I-1)}{SCR/(n-I)}$
$SCR = \sum_i\sum_j (y_{ij} - \bar{y}_{i.})^2$	$n-I$	$S_R^2 = \frac{SCR}{n-I}$	
$SCT = \sum_i\sum_j (y_{ij} - \bar{y}_{..})^2$	$n-1$		

$$IC_{1-\alpha}(\mu_i - \mu_j) = \left(\bar{y}_{i.} - \bar{y}_{j.} \pm t_{n-I;\alpha/2}\ S_R\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right) \quad ; \quad S_R^2 = \frac{\sum_i (n_i - 1)s_i^2}{n-I}$$

Diseño de experimentos (dos factores sin interacción)

(el <u>diseño por bloques</u> utiliza estas fórmulas con $K = 1$)

Modelo: $Y_{ij} = \mu + \alpha_i + \beta_j + U$; $U \sim N(0;\sigma)$; $i = 1, \ldots, I$; $j = 1, \ldots, J$

Muestra: $Y_{ijk} \sim N(\mu + \alpha_i + \beta_j;\sigma^2)$ independientes; $i = 1, ..., I$; $j = 1., ..., J$; $k = 1, ..., K$

$$\begin{aligned}
\hat{\mu} &= \bar{y}_{...} = \frac{1}{IJK}\sum_i\sum_j\sum_k y_{ijk} \\
\hat{\alpha}_i &= \bar{y}_{i..} - \bar{y}_{...} = \left(\frac{1}{JK}\sum_j\sum_k y_{ijk}\right) - \bar{y}_{...} \\
\hat{\beta}_j &= \bar{y}_{.j.} - \bar{y}_{...} = \left(\frac{1}{IK}\sum_i\sum_k y_{ijk}\right) - \bar{y}_{...} \\
\hat{\sigma}^2 &= S_R^2 = \frac{1}{IJK - I - J + 1}\sum_i\sum_j\sum_k (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 = \frac{1}{IJK - I - J + 1}\sum_i\sum_j\sum_k \hat{e}_{ijk}^2
\end{aligned}$$

Observación: $\hat{e}_{ijk} = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} = y_{ijk} - (\hat{\alpha}_i + \hat{\beta}_j + \hat{\mu})$

Tabla ANOVA

Suma de cuadrados	g.l.	Varianza	Estadístico
$SCE(\alpha) = JK\sum_i \hat{\alpha}_i^2$	$I-1$	$\frac{SCE(\alpha)}{I-1}$	$F(\alpha)$
$SCE(\beta) = IK\sum_j \hat{\beta}_j^2$	$J-1$	$\frac{SCE(\beta)}{J-1}$	$F(\beta)$
$SCR = \sum_i\sum_j\sum_k \hat{e}_{ijk}^2$	$IJK - I - J + 1$	$\frac{SCR}{IJK-I-J+1}$	
$SCT = \sum_i\sum_j\sum_k (y_{ijk} - \bar{y}_{...})^2$	$IJK - 1$		

Estadísticos F: $F(\alpha) = \dfrac{SCE(\alpha)/(I-1)}{SCR/(IJK-I-J+1)}$; $F(\beta) = \dfrac{SCE(\beta)/(J-1)}{SCR/(IJK-I-J+1)}$

$$IC_{1-\alpha}(\alpha_i - \alpha_j) = \left(\bar{y}_{i.} - \bar{y}_{j.} \pm t_{(I-1)(J-1)K;\alpha/2}\ S_R \sqrt{\frac{1}{JK} + \frac{1}{JK}} \right)$$

$$IC_{1-\alpha}(\beta_i - \beta_j) = \left(\bar{y}_{.i} - \bar{y}_{.j} \pm t_{(I-1)(J-1)K;\alpha/2}\ S_R \sqrt{\frac{1}{IK} + \frac{1}{IK}} \right)$$

Diseño de experimentos (dos factores con interacción)

Modelo: $Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + U$; $U \sim N(0;\sigma)$ $i = 1, ..., I$; $j = 1., ..., J$

Muestra:

$Y_{ijk} \sim N(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}; \sigma^2)$, independientes; $i = 1, ..., I$; $j = 1., ..., J$; $k = 1, ..., K$

$$\begin{aligned}
\hat{\mu} &= \bar{y}_{...} = \frac{1}{IJK} \sum_i \sum_j \sum_k y_{ijk} \\
\hat{\alpha}_i &= \bar{y}_{i..} - \bar{y}_{...} = \frac{1}{JK} \sum_j \sum_k y_{ijk} - \bar{y}_{...} \\
\hat{\beta}_j &= \bar{y}_{.j.} - \bar{y}_{...} = \frac{1}{IK} \sum_i \sum_k y_{ijk} - \bar{y}_{...} \\
(\hat{\alpha\beta})_{ij} &= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} = \frac{1}{K} \sum_k y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \\
\hat{\sigma}^2 &= S_R^2 = \frac{1}{IJ(K-1)} \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2 = \frac{1}{IJ(K-1)} \sum_i \sum_j \sum_k \hat{e}_{ij}^2
\end{aligned}$$

Tabla anova

Suma de cuadrados	G.l.	Varianza	F
$SCE(\alpha) = JK \sum_i \hat{\alpha}_i^2$	$I-1$	$\frac{SCE(\alpha)}{I-1}$	$F(\alpha)$
$SCE(\beta) = IK \sum_j \hat{\beta}_j^2$	$J-1$	$\frac{SCE(\beta)}{J-1}$	$F(\beta)$
$SCE(\alpha\beta) = K \sum_i \sum_j (\hat{\alpha\beta})_{ij}^2$	$(I-1)(J-1)$	$\frac{SCE(\alpha\beta)}{(I-1)(J-1)}$	$F(\alpha\beta)$
$SCR = \sum_i \sum_j \sum_k \hat{e}_{ij}^2$	$IJ(K-1)$	$\frac{SCR}{IJ(K-1)}$	
$SCT = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2$	$IJK-1$		

Estadísticos F:

$$F(\alpha) = \frac{SCE(\alpha)/(I-1)}{SCR/IJ(K-1)}\ ; \quad F(\beta) = \frac{SCE(\beta)/(J-1)}{SCR/IJ(K-1)}\ ; \quad F(\alpha\beta) = \frac{SCE(\alpha\beta)/(I-1)(J-1)}{SCR/IJ(K-1)}$$

$$IC_{1-\alpha}(\alpha_i - \alpha_j) = \left(\bar{y}_{i.} - \bar{y}_{j.} \pm t_{IJ(K-1);\alpha/2}\ S_R \sqrt{\frac{1}{JK} + \frac{1}{JK}} \right)$$

$$IC_{1-\alpha}(\beta_i - \beta_j) = \left(\bar{y}_{.i} - \bar{y}_{.j} \pm t_{IJ(K-1);\alpha/2}\ S_R \sqrt{\frac{1}{IK} + \frac{1}{IK}} \right)$$

Regresión lineal simple

Modelo: $Y\,|_{X=x} = \beta_0 + \beta_1 x + U \qquad U \sim N(0;\sigma)$

$$\hat\beta_1 = \frac{cov_{xy}}{v_x}\,; \qquad \hat\beta_0 = \bar y - \frac{cov_{xy}}{v_x}\bar x = \bar y - \hat\beta_1 \bar x$$

$$\hat\sigma^2 = S_R^2 = \frac{1}{n-2}\sum_i (y_i - \hat y_i)^2 = \frac{1}{n-2}\sum_i (y_i - \hat\beta_0 - \hat\beta_1 x_i)^2$$

$$\begin{aligned}
IC_{1-\alpha}(\beta_0) &= \left(\hat\beta_0 \pm t_{n-2;\alpha/2}\ S_R\sqrt{\frac{1}{n}+\frac{\bar x^2}{nv_x}}\right)\\
IC_{1-\alpha}(\beta_1) &= \left(\hat\beta_1 \pm t_{n-2;\alpha/2}\ S_R\sqrt{\frac{1}{nv_x}}\right)\\
IC_{1-\alpha}(\sigma^2) &= \left(\frac{(n-2)S_R^2}{\chi^2_{n-2;\alpha/2}}\,;\frac{(n-2)S_R^2}{\chi^2_{n-2;1-\alpha/2}}\right)
\end{aligned}$$

Tabla anova

Suma de cuadrados	G.l.	Varianza	Estadístico
$SCE = \sum_i (\hat y_i - \bar y)^2$	1	$\frac{SCE}{1}$	$F = \frac{SCE/1}{SCR/(n-2)}$
$SCR = \sum_i (y_i - \hat y_i)^2$	$n-2$	$\frac{SCR}{n-2}$	
$SCT = \sum_i (y_i - \bar y)^2$	$n-1$		

$$SCT = nv_y\,; \qquad SCR = nv_y(1-r^2)\,; \qquad \text{donde } r = \frac{cov_{xy}}{\sqrt{v_x v_y}}$$

Para $X = x_0$; $\hat y_0 = \hat\beta_0 + \hat\beta_1 x_0$:

$$IC_{1-\alpha}(\text{valor medio de } Y\,|_{X=x_0}) = \left(\hat y_0 \pm t_{n-2;\alpha/2}\ S_R\sqrt{\frac{1}{n}+\frac{(x_0-\bar x)^2}{nv_x}}\right)$$

$$IC_{1-\alpha}(\text{valor de } Y\,|_{X=x_0}) = \left(\hat y_0 \pm t_{n-2;\alpha/2}\ S_R\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar x)^2}{nv_x}}\right)$$

Regresión lineal múltiple

Modelo: $Y\Big|_{X_1=x_1,\dots,X_k=x_k} = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + U \qquad U \sim N(0;\sigma)$

$$\hat\sigma^2 = S_R^2 = \frac{1}{n-k-1}\sum_i (y_i - \hat y_i)^2 = \frac{1}{n-k-1}\sum_i (y_i - \hat\beta_0 - \hat\beta_1 x_{1i} - \cdots - \hat\beta_k x_{ki})^2$$

$$IC_{1-\alpha}(\beta_i) = \left(\hat\beta_i \pm t_{n-k-1;\alpha/2}\cdot(\text{error típico de } \hat\beta_i)\right),\ i = 0,\dots,k$$

Tabla anova

Suma de cuadrados	G.l.	Varianza	Estadístico
$SCE = \sum_i (\hat y_i - \bar y)^2$	k	$\frac{SCE}{k}$	$F = \frac{SCE/k}{SCR/(n-k-1)}$
$SCR = \sum_i (y_i - \hat y_i)^2$	$n-k-1$	$\frac{SCR}{n-k-1}$	
$SCT = \sum_i (y_i - \bar y)^2$	$n-1$		

$$R^2 = \frac{SCE}{SCT} = \frac{SCT - SCR}{SCT}\,; \qquad F = \frac{R^2}{1-R^2}\,\frac{n-k-1}{k}$$

Modelo: $\text{Prob}(Y=1|\, X_1=x_1,\ldots,X_k=x_k) = \dfrac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k)}}$

$\text{IC}_{1-\alpha}(\beta_j) = \left(\hat{\beta}_i \pm z_{\alpha/2} \cdot (\text{error típico en } \hat{\beta}_j)\right)$

Regla de clasificación de los individuos:

- Si $\hat{\beta}_0+\hat{\beta}_1 x_1+\cdots+\hat{\beta}_k x_k > 0$, se clasifica $Y=1$.
- Si $\hat{\beta}_0+\hat{\beta}_1 x_1+\cdots+\hat{\beta}_k x_k < 0$, se clasifica $Y=0$.