

## ANÁLISIS DE LA VARIANZA MULTIFACTORIAL

Estudia la influencia de dos o más factores (variables explicativas) sobre la media de una variable aleatoria (variable respuesta).

- Definición de la variable a explicar
- Definición de los distintos factores que pueden influir en la respuesta y, en cada uno de ellos, sus distintos niveles o grupos.

Se analizan tres casos:

1. Dos factores (diseño por bloques)
2. Dos factores con interacción
3. Tres factores (Cuadros latinos)

## ANÁLISIS DE LA VARIANZA CON DOS FACTORES 'DISEÑO POR BLOQUES'

El Modelo:  $\{ Y_{ij} = \mu + \alpha_i + \beta_j + U \}_{(i=1, 2, \dots, k), (j=1, 2, \dots, m)}$

$Y_{ij}$  es la respuesta de la variable en el i-ésimo nivel del factor 1 ( $\alpha$ ) y en el j-ésimo nivel del factor 2 ( $\beta$ )

$\mu_{ij} = E(Y_{ij}) = \mu + \alpha_i + \beta_j$        $\left\{ \begin{array}{l} \alpha_i \equiv \text{representa el efecto que sobre la media global } \mu \text{ tiene el nivel } i \text{ del factor 1} \\ \beta_j \equiv \text{representa el efecto que sobre la media global } \mu \text{ tiene el nivel } j \text{ del factor 2} \\ U \equiv \text{representa la variación aleatoria de las } Y_{ij} \text{ (igual en distribución para todas ellas)} \end{array} \right.$

Suponemos que  $U$  sigue una distribución normal  $N(0, \sigma)$  lo que implica que  $Y_{ij}$  sigue una distribución normal  $N(\mu_{ij}, \sigma)$  y que no hay interacción entre los factores.

$$\sum_{i=1}^k \alpha_i = \sum_{j=1}^m \beta_j = 0$$

## MUESTRA ALEATORIA - UNA OBSERVACIÓN POR CASILLA -

$$Y_{ij} \approx N\left(\mu + \alpha_i + \beta_j, \sigma^2\right) \text{ independientes}$$

Factor 1 (α)	Factor 2 (β)					Medias por filas
	Niveles	1	2	..	m	
	1	y <sub>11</sub>	y <sub>12</sub>	..	y <sub>1m</sub>	$\bar{y}_{1\bullet}$
	2	y <sub>21</sub>	y <sub>22</sub>	..	y <sub>2m</sub>	$\bar{y}_{2\bullet}$
	.	..	..	..	..	
	.	..	..	..	..	
	k	y <sub>k1</sub>	y <sub>k2</sub>	..	y <sub>km</sub>	$\bar{y}_{k\bullet}$
Medias por columnas		$\bar{y}_{\bullet 1}$	$\bar{y}_{\bullet 2}$	..	$\bar{y}_{\bullet m}$	$\bar{y}_{\bullet\bullet}$

$$(y_{ij} - \bar{y}_{\bullet\bullet}) = (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet}) + (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}) + (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})$$

$$\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{\bullet\bullet})^2 = \sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2 + \sum_{i=1}^k \sum_{j=1}^m (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 + \sum_{i=1}^k \sum_{j=1}^m (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2$$

$$\underbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{\bullet\bullet})^2}_{\begin{array}{l} \text{SCT} \\ \text{Suma Cuadrados Total} \end{array}} = \underbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2}_{\begin{array}{l} \text{SCR} \\ \text{Suma Cuadrados Residual} \end{array}} + \underbrace{m \sum_{i=1}^k (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2}_{\begin{array}{l} \text{SCE}(\alpha) \\ \text{Suma Cuadrados Explicada Factor}(\alpha) \end{array}} + \underbrace{k \sum_{j=1}^m (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2}_{\begin{array}{l} \text{SCE}(\beta) \\ \text{Suma Cuadrados Explicada Factor}(\beta) \end{array}}$$

$$\underbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{..})^2}_{\text{SCT Suma Cuadrados Total}} = \underbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{..})^2}_{\text{SCR Suma Cuadrados Residual}} + \underbrace{m \sum_{i=1}^k (\bar{y}_{i\bullet} - \bar{y}_{..})^2}_{\text{SCE } (\alpha) \text{ Suma Cuadrados Explicada Factor } (\alpha)} + \underbrace{k \sum_{j=1}^m (\bar{y}_{\bullet j} - \bar{y}_{..})^2}_{\text{SCE } (\beta) \text{ Suma Cuadrados Explicada Factor } (\beta)}$$

$$\hat{S}_y^2 = \frac{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{..})^2}{k \cdot m - 1} \quad \hat{S}_r^2 = \frac{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{..})^2}{(k-1)(m-1)} \quad \hat{S}_\alpha^2 = \frac{m \sum_{i=1}^k (\bar{y}_{i\bullet} - \bar{y}_{..})^2}{(k-1)} \quad \hat{S}_\beta^2 = \frac{k \sum_{j=1}^m (\bar{y}_{\bullet j} - \bar{y}_{..})^2}{(m-1)}$$



$$\hat{S}_y^2 = \frac{\overbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{..})^2}^{\text{SCT}}}{k \cdot m - 1}$$

$$\hat{S}_\alpha^2 = \frac{\overbrace{m \sum_{i=1}^k (\bar{y}_{i\bullet} - \bar{y}_{..})^2}^{\text{SCE } (\alpha)}}{(k-1)}$$

$$\hat{S}_\beta^2 = \frac{\overbrace{k \sum_{j=1}^m (\bar{y}_{\bullet j} - \bar{y}_{..})^2}^{\text{SCE } (\beta)}}{(m-1)}$$



$$\overbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2}^{\text{SCR}} = \overbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{\bullet\bullet})^2}^{\text{SCT}} - m \sum_{i=1}^k (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 - k \sum_{j=1}^m (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2$$



$$\overbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2}^{\text{SCR}} = \overbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{\bullet\bullet})^2}^{\text{SCT}} - m \sum_{i=1}^k (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 - k \sum_{j=1}^m (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2$$

$$\hat{s}_r^2 = \frac{\text{SCR}}{(k-1)(m-1)}$$

$$\hat{s}_y^2 = \frac{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{\bullet\bullet})^2}{k \cdot m - 1} \quad \mapsto \quad \text{SCT} = (k \cdot m - 1) \sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{\bullet\bullet})^2$$

$$\text{SCE}(\alpha) = m \sum_{i=1}^k (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$$

$$\text{SCE}(\beta) = k \sum_{j=1}^m (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2$$

En la práctica ....

$$\text{SCR} = \text{SCT} - \text{SCE}(\alpha) - \text{SCE}(\beta)$$



TABLA ANOVA II

Fuente de variación	Suma de cuadrados	gr. libertad	Varianza	Test F Fisher-Snedecor	Coeficiente Determinación $R^2$
Factor( $\alpha$ )	SCE( $\alpha$ )	$k - 1$	$\hat{S}_{\alpha}^2 = \frac{\text{SCE}(\alpha)}{k - 1}$	$F_{\alpha} = \frac{\hat{S}_{\alpha}^2}{\hat{S}_r^2}$	$R_{\alpha}^2 = \frac{\text{SCE}(\alpha)}{\text{SCT}} \cdot 100$
Factor ( $\beta$ )	SCE( $\beta$ )	$m - 1$	$\hat{S}_{\beta}^2 = \frac{\text{SCE}(\beta)}{m - 1}$	$F_{\beta} = \frac{\hat{S}_{\beta}^2}{\hat{S}_r^2}$	$R_{\beta}^2 = \frac{\text{SCE}(\beta)}{\text{SCT}} \cdot 100$
Residual	SCR	$(k - 1)(m - 1)$	$\hat{S}_r^2 = \frac{\text{SCR}}{(k - 1)(m - 1)}$		
Total	SCT	$k \cdot m - 1$			$R^2 = \frac{\text{SCE}(\alpha) + \text{SCE}(\beta)}{\text{SCT}} \cdot 100$

Análisis Estadístico: ANOVA II  
- Contrastos del efecto de cada factor -

		Estadístico de Contraste
$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$	El Factor 1 no influye	$F_{\alpha} = \frac{\hat{S}_{\alpha}^2}{\hat{S}_r^2}$
$H_a : \text{Algún } \alpha_i \neq 0$		
$H_0 : \beta_1 = \beta_2 = \dots = \beta_m = 0$	El Factor 2 no influye	$F_{\beta} = \frac{\hat{S}_{\beta}^2}{\hat{S}_r^2}$
$H_a : \text{Algún } \beta_j \neq 0$		



Estadístico de contraste:  $F_\alpha = \frac{\hat{S}_\alpha^2}{\hat{S}_r^2}$

$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$  El Factor 1 no influye

$H_a : \text{Algun } \alpha_i \neq 0$

Se acepta la hipótesis nula  $H_0$  si

$$F_\alpha = \frac{\hat{S}_\alpha^2}{\hat{S}_r^2} \leq F_{\alpha, (k-1), (k-1)(m-1)}$$

En caso contrario, se acepta la hipótesis alternativa

Estadístico de contraste:  $F_\beta = \frac{\hat{S}_\beta^2}{\hat{S}_r^2}$

$H_0 : \beta_1 = \beta_2 = \dots = \beta_m = 0$  El Factor 2 no influye

$H_a : \text{Algun } \beta_j \neq 0$

Se acepta la hipótesis nula  $H_0$  si

$$F_\beta = \frac{\hat{S}_\beta^2}{\hat{S}_r^2} \leq F_{\beta, (m-1), (k-1)(m-1)}$$

En caso contrario, se acepta la hipótesis alternativa





## Coefficiente Determinación

El Coeficiente de Determinación explicada por el modelo viene dada por la expresión:

$$R^2 = \frac{SCE(\alpha) + SCE(\beta)}{SCT} \cdot 100$$

En particular, para cada uno de los factores, se expresa:

$$R_\alpha^2 = \frac{SCE(\alpha)}{SCT} \cdot 100 \quad R_\beta^2 = \frac{SCE(\beta)}{SCT} \cdot 100$$

Cuando se rechaza la hipótesis nula  $H_0$ , se puede realizar una análisis posterior



Cuando se rechaza la hipótesis nula, se puede construir un Intervalo de Confianza para la diferencia de medias, con un nivel de significación  $\alpha$ :

FACTOR 1

$$IC(\alpha_i - \alpha_j) = \left[ \underbrace{\overline{y}_{i\bullet} - \overline{y}_{j\bullet}}_{\text{diferencia medias}} \pm \underbrace{t_{\alpha/2; (k-1)(m-1)} \hat{S}_r \sqrt{\frac{1}{m} + \frac{1}{m}}}_{\text{Error típico}} \right]$$

FACTOR 2:  $IC(\beta_i - \beta_j) = \left[ \underbrace{\overline{y}_{\bullet i} - \overline{y}_{\bullet j}}_{\text{diferencia medias}} \pm \underbrace{t_{\alpha/2; (k-1)(m-1)} \hat{S}_r \sqrt{\frac{1}{k} + \frac{1}{k}}}_{\text{Error Típico}} \right]$

Uno de las opciones más utilizadas en las Pruebas Post hoc es el Test de Bonferroni



Se realizan comparaciones simultáneas entre todas las posibles parejas de niveles en cada factor. Para ello se construyen Intervalos de Confianza:

**FACTOR 1**

$$IC(\alpha_i - \alpha_j) = \left[ \overbrace{(\bar{y}_{i\bullet} - \bar{y}_{j\bullet})}^{\text{Diferencia medias}} \pm t_{(\alpha/2c); (k-1)(m-1)} \hat{s}_r \sqrt{\frac{1}{m} + \frac{1}{m}} \right]$$

**FACTOR 2**

$$IC(\beta_i - \beta_j) = \left[ \overbrace{(\bar{y}_{\bullet i} - \bar{y}_{\bullet j})}^{\text{Diferencia medias}} \pm t_{(\alpha/2c); (k-1)(m-1)} \hat{s}_r \sqrt{\frac{1}{k} + \frac{1}{k}} \right]$$

donde 'c' es el número de contrastes, esto es, el número de tests que hay que realizar para comparar todos los pares de medias  $c = \binom{k}{2}$



Cuando el CERO no se encuentra en el Intervalo de Confianza, se afirma que las medias no son iguales.



Se desea estudiar la eficiencia (en cuanto la menor emisión de  $CO_2$ ) de cinco máquinas desaladoras. Se piensa que la cantidad de sal en el agua puede influir en dicha eficiencia.

Factor 2: Nivel de Sal

Factor 1: Máquinas	Poca Sal	Bastante Sal	Mucha Sal
	Máquina I	24	26
Máquina II	27	30	32
Máquina III	26	27	30
Máquina IV	25	28	28
Máquina V	28	29	31



media total y  
medias marginales

	Poca Sal	Bastante Sal	Mucha Sal	$\bar{y}_{i\bullet}$
Máquina I	24	26	29	$\bar{y}_{1\bullet} = 26,33$
Máquina II	27	30	32	$\bar{y}_{2\bullet} = 29,67$
Máquina III	26	27	30	$\bar{y}_{3\bullet} = 27,67$
Máquina IV	25	28	28	$\bar{y}_{4\bullet} = 27$
Máquina V	28	29	31	$\bar{y}_{5\bullet} = 29,33$
$\bar{y}_{\bullet j}$	$\bar{y}_{\bullet 1} = 26$	$\bar{y}_{\bullet 2} = 28$	$\bar{y}_{\bullet 3} = 30$	$\bar{y}_{\bullet\bullet} = 28$



Analizando los  
datos en conjunto:

$$\bar{y}_{\bullet\bullet} = \frac{\sum_{i=1}^5 \sum_{j=1}^3 y_{ij}}{n} = \frac{420}{15} = 28$$

$$\hat{S}_r^2 = \frac{\sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - \bar{y}_{\bullet\bullet})^2}{n-1} = \frac{70}{14} = 5 \Rightarrow \boxed{SCT = \sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - \bar{y}_{\bullet\bullet})^2 = 70 \quad \hat{S}_r^2 = 5}$$



Pensamos en la descomposición de la variable de respuesta  $y_{ij}$   
a efectos de analizar el efecto que sobre la media global tiene  
los dos factores:

$$(y_{ij} - \bar{y}_{\bullet\bullet}) = (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet}) + (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}) + (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})$$

$$\underbrace{\sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - \bar{y}_{\bullet\bullet})^2}_{SCT} = \underbrace{\sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2}_{SCR} + \underbrace{3 \sum_{i=1}^5 (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2}_{SCE(\alpha)} + \underbrace{5 \sum_{j=1}^3 (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2}_{SCE(\beta)}$$



**Analizando el FACTOR 1**

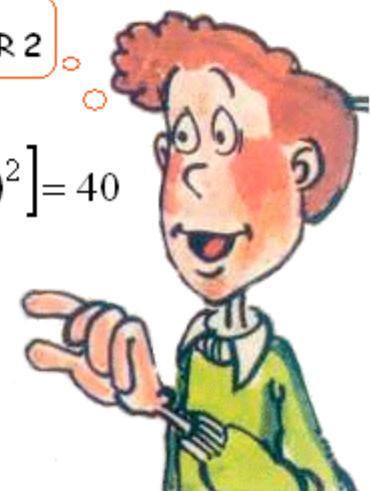
$$\text{SCE}(\alpha) = 3 \sum_{i=1}^5 (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 = \\ = 3[(26,33 - 28)^2 + (29,67 - 28)^2 + (27,67 - 28)^2 + (27 - 28)^2 + (29,33 - 28)^2] = 25,36$$

$$\hat{s}_\alpha^2 = \frac{\text{SCE}(\alpha)}{(5-1)} = \frac{3 \sum_{i=1}^5 (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2}{(5-1)} = \frac{25,36}{4} = 6,34$$

**Analizando el FACTOR 2**

$$\text{SCE}(\beta) = 5 \sum_{j=1}^3 (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2 = 5[(26 - 28)^2 + (28 - 28)^2 + (30 - 28)^2] = 40$$

$$\hat{s}_\beta^2 = \frac{5 \sum_{j=1}^3 (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2}{(3-1)} = \frac{40}{2} = 20$$



**La Variación Residual**

$$\text{SCR} = \sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2 = \text{SCT} - \text{SCE}(\alpha) - \text{SCE}(\beta) = 70 - 40 - 25,36 = 4,64$$

$$\hat{s}_r^2 = \frac{\sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2}{(5-1).(3-1)} = \frac{4,64}{8} = 0,58$$

Fuente Variación	Suma Cuadrados	grados libertad	Varianza	Test F
Factor ( $\alpha$ )	$SCE(\alpha) = 25,36$	$k - 1 = 4$	$\hat{S}_{\alpha}^2 = \frac{SCE(\alpha)}{k - 1} = 6,34$	$F_{\alpha} = \frac{\hat{S}_{\alpha}^2}{\hat{S}_r^2} = \frac{6,34}{0,58} = 10,93$
Factor ( $\beta$ )	$SCE(\beta) = 40$	$m - 1 = 2$	$\hat{S}_{\beta}^2 = \frac{SCE(\beta)}{m - 1} = 20$	$F_{\beta} = \frac{\hat{S}_{\beta}^2}{\hat{S}_r^2} = \frac{20}{0,58} = 34,48$
Residual	$SCR = 4,64$	$(k - 1)(m - 1) = 8$	$\hat{S}_r^2 = \frac{SCR}{(k - 1)(m - 1)} = 0,58$	
Total	$SCT = 70$	$km - 1 = 14$		



Hemos de contrastar las siguientes hipótesis:

TABLA  
ANOVA II

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$  « Los resultados medios obtenidos no dependen de la máquina elegida »

$H_a : \text{Algún } \alpha_i \neq 0$  « Los resultados medios obtenidos dependen de la máquina elegida »

$H_0 : \beta_1 = \beta_2 = \beta_3$  « Los resultados medios obtenidos no dependen de la cantidad de sal »

$H_a : \text{Algún } \beta_j \neq 0$  « Los resultados medios obtenidos dependen de la cantidad de sal »

Se acepta  $H_0$  si  $F_{\alpha} = \frac{\hat{S}_{\alpha}^2}{\hat{S}_r^2} \leq F_{\alpha; (k-1), (k-1)(m-1)}$

En caso contrario, se rechaza la hipótesis nula

Se acepta  $H_0$  si  $F_{\beta} = \frac{\hat{S}_{\beta}^2}{\hat{S}_r^2} \leq F_{\beta; (m-1), (k-1)(m-1)}$

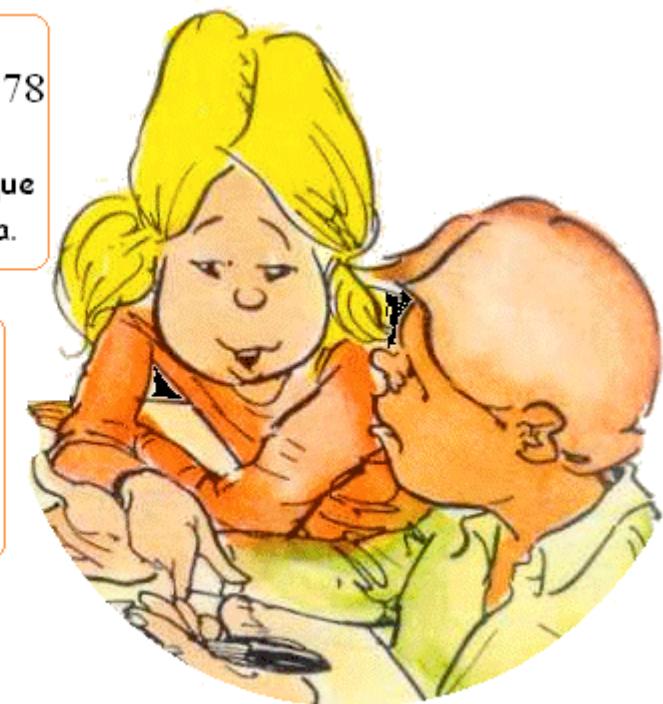
En caso contrario, se rechaza la hipótesis nula

$$F_{\alpha} = \frac{\hat{S}_{\alpha}^2}{\hat{S}_r^2} = \frac{6,34}{0,58} = 10,93 > F_{\alpha; (k-1), (k-1)(m-1)} = F_{0,05; 4, 8} = 3,8378$$

En consecuencia, se rechaza la hipótesis nula, y por tanto, se concluye que los resultados medios obtenidos de  $CO_2$  dependen de la máquina elegida.

$$F_{\beta} = \frac{\hat{S}_{\beta}^2}{\hat{S}_r^2} = \frac{20}{0,58} = 34,48 > F_{\beta; (m-1), (k-1)(m-1)} = F_{0,05; 2, 8} = 4,4590$$

En consecuencia, se rechaza la hipótesis nula, y por tanto, se concluye que los resultados medios obtenidos de  $CO_2$  dependen del factor cantidad de sal.



Adviértase que si no hubiéramos tenido en cuenta la cantidad de sal, nos situamos ante el análisis de la varianza de un sólo factor (ANOVA I). Entonces, se plantea el estudio:

	Poca Sal	Bastante Sal	Mucha Sal	$\bar{y}_{i\bullet}$	$s_{i\bullet}^2$
Máquina I	24	26	29	$\bar{y}_{1\bullet} = 26,33$	$s_{1\bullet}^2 = 6,33$
Máquina II	27	30	32	$\bar{y}_{2\bullet} = 29,67$	$s_{2\bullet}^2 = 6,33$
Máquina III	26	27	30	$\bar{y}_{3\bullet} = 27,67$	$s_{3\bullet}^2 = 4,33$
Máquina IV	25	28	28	$\bar{y}_{4\bullet} = 27$	$s_{4\bullet}^2 = 3$
Máquina V	28	29	31	$\bar{y}_{5\bullet} = 29,33$	$s_{5\bullet}^2 = 2,33$
				$\bar{y}_{\bullet\bullet} = 28$	



Nos encontramos en una situación de ANOVA I .....

$$(y_{ij} - \bar{y}_{..}) = (y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..}) \mapsto$$

$$\underbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{..})^2}_{\text{SCT}} = \underbrace{\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{i.})^2}_{\text{SCR}} + \underbrace{\sum_{i=1}^k m (\bar{y}_{i.} - \bar{y}_{..})^2}_{\text{SCE}}$$

$$\bar{y}_{..} = \frac{\sum_{i=1}^5 \sum_{j=1}^3 y_{ij}}{n} = \frac{420}{15} = 28$$

La variable en conjunto ya se encuentra analizada, resultando

$$\hat{s}_y^2 = \frac{\sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - \bar{y}_{..})^2}{n-1} = \frac{70}{14} = 5 \Rightarrow \boxed{\text{SCT} = \sum_{i=1}^5 \sum_{j=1}^3 (y_{ij} - \bar{y}_{..})^2 = 70 \quad \hat{s}_y^2 = 5}$$



Al analizarlo por GRUPOS

$$s_{i.}^2 = \frac{1}{m-1} \sum_{j=1}^m (y_{ij} - \bar{y}_{i.})^2 \quad \hat{s}_r^2 = \frac{1}{(n-k)} \sum_{i=1}^k (m-1)s_{i.}^2 = \frac{1}{(n-k)} \sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_{i.})^2 = \frac{\text{SCR}}{(n-k)}$$





Dentro de los GRUPOS ....

$$\hat{S}_r^2 = \frac{1}{n-5} \sum_{i=1}^5 (m-1)s_{i\bullet}^2 = \frac{SCR}{15-5} = \\ = \frac{1}{15-5} [(3-1)(6,33) + (3-1)(6,33) + (3-1)(4,33) + (3-1)(3) + (3-1)(2,33)] = \frac{44,64}{10} = 4,464$$

**SCR = 44,64**

Finalmente, la variación explicada ENTRE GRUPOS:  $SCE = SCT - SCR = 70 - 44,64 = 25,36$

Fuente Variación	Suma Cuadrados	grados libertad	Varianza	Test F
Explicada (Máquinas)	$SCE = 25,36$	$k-1 = 4$	$\hat{S}_e^2 = \frac{SCE}{k-1} = 6,34$	$F = \frac{\hat{S}_e^2}{\hat{S}_r^2} = \frac{6,34}{4,46} = 1,42$
Residual	$SCR = 44,64$	$n-k = 10$	$\hat{S}_r^2 = \frac{SCR}{n-k} = 4,46$	
Total	$SCT = 70$	$n-1 = 14$		

$$F = \frac{\hat{S}_e^2}{\hat{S}_r^2} = 1,42 \leq F_{\alpha, (k-1), (n-k)} = F_{0,05, 4, 10} = 3,4780$$

Por tanto, se acepta la hipótesis nula de que la emisión media de CO<sub>2</sub> no depende del tipo de máquina.





Para Introducir los datos en el SPSS  
y calcular el ANOVA I

maquinas-sal.sav - SPSS Editor de datos

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

22 :

	Maquinas	Sal
1	1,00	24,00
2	1,00	26,00
3	1,00	29,00
4	2,00	27,00
5	2,00	30,00
6	2,00	32,00
7	3,00	26,00
8	3,00	27,00
9	3,00	30,00
10	4,00	25,00
11	4,00	28,00
12	4,00	28,00
13	5,00	28,00
14	5,00	29,00
15	5,00	31,00
16		

Vista de datos Vista de variables

Analizar → Comparar medias → ANOVA de un factor...

ANOVA de un factor

Dependientes: Sal

Factor: Maquinas

Contrastes Post hoc... Opciones...

Aceptar Pegar Restablecer Cancelar Ayuda

ANOVA de un factor

SPSS El procesador está preparado

**Resultados6 - Visor SPSS**

Archivo Edición Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ventana ?

Resultados  
ANOVA de un factor  
Título  
Notas  
ANOVA

→ **ANOVA de un factor**

**ANOVA**

Sal

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	25,333	4	6,333	1,418	,297
Intra-grupos	44,667	10	4,467		
Total	70,000	14			

SPSS El procesador está preparado

Si hubiéramos realizado en el SPSS el Análisis de la varianza con dos factores, se tendría:

	Poca Sal	Bastante Sal	Mucha Sal
Máquina I	24	26	29
Máquina II	27	30	32
Máquina III	26	27	30
Máquina IV	25	28	28
Máquina V	28	29	31





SPSS CO2.sav - SPSS Editor de datos

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 : CO 24

	CO	Maquina	Sal
1	24,00	1,00	1,00
2	26,00	1,00	2,00
3	29,00	1,00	3,00
4	27,00	2,00	1,00
5	30,00	2,00	2,00
6	32,00	2,00	3,00
7	26,00	3,00	1,00
8	27,00	3,00	2,00
9	30,00	3,00	3,00
10	25,00	4,00	1,00
11	28,00	4,00	2,00
12	28,00	4,00	3,00
13	28,00	5,00	1,00
14	29,00	5,00	2,00
15	31,00	5,00	3,00
16			

Vista de datos Vista de variables /

SPSS El procesador está pr

Analizar ▾

- Informes
- Estadísticos descriptivos
- Tablas
- Comparar medias
- Modelo lineal general
- Modelos mixtos
- Correlaciones
- Regresión
- Loglineal
- Clasificar
- Reducción de datos
- Escalas
- Pruebas no paramétricas
- Series temporales
- Supervivencia
- Respuesta múltiple
- Análisis de valores perdidos...
- Muestras complejas

Univariente... Multivariante... Medidas repetidas... Componentes de la varianza...

■ Univariate

Dependiente: # CO

Factores fijos: # Maquina # Sal

Factores aleatorios:

Covariables:

Ponderación MCP:

Aceptar Pegar Restablecer Cancelar Ayuda

## Resultados2 - Visor SPSS

Archivo Edición Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ventana ?



- Resultados
  - Análisis de varianza univariante
    - Título
    - Notas
    - Factores inter-sujetos
    - Pruebas de los efectos inter-sujetos

### ► Análisis de varianza univariante

Variable dependiente: CO

Fuente	Suma de cuadrados tipo I	gl	Media cuadrática	F	Significación
Modelo corregido	65,333 <sup>a</sup>	6	10,889	18,667	,000
Intersección	11760,000	1	11760,000	20160,000	,000
Maquina	25,333	4	6,333	10,857	,003
Sal	40,000	2	20,000	34,286	,000
Error	4,667	8	,583		
Total	11830,000	15			
Total corregida	70,000	14			

a. R cuadrado = ,933 (R cuadrado corregida = ,883)

SPSS El procesador está preparado





	Poca Sal	Bastante Sal	Mucha Sal
Máquina I	24	26	29
Máquina II	27	30	32
Máquina III	26	27	30
Máquina IV	25	28	28
Máquina V	28	29	31

Recapitulando, en el análisis de la varianza con dos factores, con una observación por casilla, los datos obtenidos eran:

Fuente Variación	Suma Cuadrados	Varianza	Test F
Factor ( $\alpha$ )	$SCE(\alpha) = 25,36$	$\hat{S}_{\alpha}^2 = \frac{SCE(\alpha)}{k-1} = 6,34$	$F_{\alpha} = \frac{\hat{S}_{\alpha}^2}{\hat{S}_r^2} = \frac{6,34}{0,58} = 10,93$
Factor ( $\beta$ )	$SCE(\beta) = 40$	$\hat{S}_{\beta}^2 = \frac{SCE(\beta)}{m-1} = 20$	$F_{\beta} = \frac{\hat{S}_{\beta}^2}{\hat{S}_r^2} = \frac{20}{0,58} = 34,48$
Residual	$SCR = 4,64$	$\hat{S}_r^2 = \frac{SCR}{(k-1)(m-1)} = 0,58$	
Total	$SCT = 70$		

El porcentaje de variabilidad explicada viene dada por el coeficiente de determinación :

$$R^2 = \frac{\underbrace{SCE}_{\text{explicada}}}{\underbrace{SCT}_{\text{explicada}}} = \frac{SCE(\alpha) + SCE(\beta)}{SCT} = \frac{\underbrace{SCE(\alpha)}_{\substack{\text{explicada} \\ \text{factor } \alpha}}}{\underbrace{SCT}_{\text{explicada}}} + \frac{\underbrace{SCE(\beta)}_{\substack{\text{explicada} \\ \text{factor } \beta}}}{\underbrace{SCT}_{\text{explicada}}}$$

$$R^2 = \frac{\underbrace{SCE}_{\text{explicada}}}{\underbrace{SCT}_{\text{explicada}}} = \frac{65,36}{70} = 0,93$$

$$R_\alpha^2 = \frac{\underbrace{SCE(\alpha)}_{\substack{\text{explicada} \\ \text{factor } \alpha}}}{\underbrace{SCT}_{\text{explicada}}} = \frac{25,36}{70} = 0,37 \quad R_\beta^2 = \frac{\underbrace{SCE(\beta)}_{\substack{\text{explicada} \\ \text{factor } \beta}}}{\underbrace{SCT}_{\text{explicada}}} = \frac{40}{70} = 0,57$$



$$F_\alpha = \frac{\hat{S}_\alpha^2}{\hat{S}_r^2} = \frac{6,34}{0,58} = 10,93 > F_{\alpha; (k-1), (k-1)(m-1)} = F_{0,05; 4, 8} = 3,8378$$

En consecuencia, se rechaza la hipótesis nula, y por tanto, se concluye que los resultados medios obtenidos de  $CO_2$  dependen de la máquina elegida.



$$F_\beta = \frac{\hat{S}_\beta^2}{\hat{S}_r^2} = \frac{20}{0,58} = 34,48 > F_{\beta; (m-1), (k-1)(m-1)} = F_{0,05; 2, 8} = 4,4590$$

En consecuencia, se rechaza la hipótesis nula, y por tanto, se concluye que los resultados medios obtenidos de  $CO_2$  dependen del factor cantidad de sal.



Cuando se rechaza la hipótesis nula se pueden realizar pruebas simultáneas 'Pruebas Post hoc', construyendo un intervalo de confianza para la diferencia de medias entre todas las posibles parejas de niveles en cada factor - Test de Bonferroni.

$$\text{FACTOR 1} \quad IC(\alpha_i - \alpha_j) = \left[ \overbrace{(\bar{y}_{i\bullet} - \bar{y}_{j\bullet})}^{\text{Diferencia medias}} \pm \overbrace{t_{(\alpha/2c), (k-1)(m-1)} \hat{S}_r \sqrt{\frac{1}{m} + \frac{1}{m}}}^{\text{Error Típico}} \right]$$

$$\text{FACTOR 2} \quad IC(\beta_i - \beta_j) = \left[ \overbrace{(\bar{y}_{\bullet i} - \bar{y}_{\bullet j})}^{\text{Diferencia medias}} \pm \overbrace{t_{(\alpha/2c), (k-1)(m-1)} \hat{S}_r \sqrt{\frac{1}{k} + \frac{1}{k}}}^{\text{Error Típico}} \right]$$

$t_{0,025; 8} = 2,306$	$\bar{y}_{\bullet i} - \bar{y}_{\bullet j}$	Error Típico	Intervalo de confianza 85%	
			Límite Inferior	Límite Superior
Poca sal	Poca sal	-----	-----	-----
	Bastante sal	-2	0,4817	- 3,11 - 0,89
	Mucha sal	-4	0,4817	- 5,11 - 2,89
Bastante sal	Poca sal	2	0,4817	0,89 3,11
	Bastante sal	-----	-----	-----
	Mucha sal	-2	0,4817	- 3,11 - 0,89
Mucha sal	Poca sal	4	0,4817	2,89 5,11
	Bastante sal	2	0,4817	0,89 3,11
	Mucha sal	-----	-----	-----

$t_{0,0025; 8} = 3,817$		$\bar{y}_{i\bullet} - \bar{y}_{j\bullet}$	Error Típico	Intervalo de confianza 95%	
				Límite Inferior	Límite Superior
Máquina I	Máquina I	-----	-----	-----	-----
	Máquina II	- 3,34	0,6219	- 5,71	- 0,97
	Máquina III	- 1,34	0,6219	- 3,71	1,03
	Máquina IV	- 0,67	0,6219	- 3,04	1,7
	Máquina V	- 3	0,6219	- 5,37	- 0,63
Máquina II	Máquina I	3,34	0,6219	0,97	5,71
	Máquina II	-----	-----	-----	-----
	Máquina III	2	0,6219	- 0,37	4,37
	Máquina IV	2,67	0,6219	0,3	5,04
	Máquina V	0,34	0,6219	- 2,03	2,71
Máquina III	Máquina I	1,34	0,6219	- 1,03	3,71
	Máquina II	- 2	0,6219	- 4,37	0,37
	Máquina III	-----	-----	-----	-----
	Máquina IV	0,67	0,6219	- 1,7	3,04
	Máquina V	- 1,66	0,6219	- 4,03	0,71
Máquina IV	Máquina I	0,67	0,6219	- 1,7	3,04
	Máquina II	- 2,67	0,6219	- 5,04	- 0,3
	Máquina III	- 0,67	0,6219	- 3,04	1,7
	Máquina IV	-----	-----	-----	-----
	Máquina V	- 2,33	0,6219	- 4,7	0,04
Máquina V	Máquina I	3	0,6219	0,63	5,37
	Máquina II	- 0,34	0,6219	- 2,71	2,03
	Máquina III	1,66	0,6219	- 0,71	4,03
	Máquina IV	2,33	0,6219	- 0,04	4,7
	Máquina V	-----	-----	-----	-----

Si en el INTERVALO no se encuentra el CERO, rechazamos que las medias son iguales.

## MAQUINAS-SAL-ANOVA2.sav - SPSS Editor de datos



Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?



Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
Emisiones	Numérico	8	0		Ninguno	Ninguno	8	Centrado	Ordinal
Maquinas	Cadena	14	0		Ninguno	Ninguno	9	Izquierda	Nominal
Sal	Cadena	20	0		Ninguno	Ninguno	12	Izquierda	Nominal

Vista de datos Vista de variables /

## MAQUINAS-SAL-ANOVA2.sav - SPSS Editor de datos



Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?



19 : Sal

	Emisiones	Maquinas	Sal
1	24	Máquina I	Poca Sal
2	26	Máquina I	Bastante Sal
3	29	Máquina I	Mucha Sal
4	27	Máquina II	Poca Sal
5	30	Máquina II	Bastante Sal
6	32	Máquina II	Mucha Sal
7	26	Máquina III	Poca Sal
8	27	Máquina III	Bastante Sal
9	30	Máquina III	Mucha Sal
10	25	Máquina IV	Poca Sal
11	28	Máquina IV	Bastante Sal
12	28	Máquina IV	Mucha Sal
13	28	Máquina V	Poca Sal
14	29	Máquina V	Bastante Sal
15	31	Máquina V	Mucha Sal

Vista de datos Vista de variables /

## Univariate: Opciones

## Medias marginales estimadas

Factores e interacciones de los factores: Mostrar las medias para:

(GLOBAL)  
Maquinas  
Sal

Maquinas  
Sal

Comparar los efectos principales

Ajuste del intervalo de confianza:

Bonferroni

## Mostrar

- |   |  |
|---|--|
| <input checked="" type="checkbox"/> Estadísticos descriptivos | <input type="checkbox"/> Pruebas de homogeneidad         |
| <input type="checkbox"/> Estimaciones del tamaño del efecto   | <input type="checkbox"/> Diagramas de dispersión x nivel |
| <input type="checkbox"/> Potencia observada                   | <input type="checkbox"/> Gráfico de los residuos         |
| <input type="checkbox"/> Estimaciones de los parámetros       | <input type="checkbox"/> Falta de ajuste                 |
| <input type="checkbox"/> Matriz de coeficientes de contraste  | <input type="checkbox"/> Función estimable general       |

Nivel de significación: ,05

Los intervalos de confianza son del 95%

Continuar

Cancelar

Ayuda

Dependiente:

Emisiones

Factores fijos:

Maquinas

Sal

Factores aleatorios:

Covariables:

Ponderación MCP:

Modelo...

Contrastes...

Gráficos...

Post\_hoc...

Guardar...

Opciones...

Cancelar

Ayuda

Variable dependiente: Emisiones

(I) Sal	(J) Sal	Diferencia entre medias (I-J)	Error típ.	Significación <sup>a</sup>	Intervalo de confianza al 95 % para diferencia <sup>a</sup>	
					Límite inferior	Límite superior
Bastante Sal	Mucha Sal	-2,000*	,483	,010	-3,457	-,543
	Poca Sal	2,000*	,483	,010	,543	3,457
Mucha Sal	Bastante Sal	2,000*	,483	,010	,543	3,457
	Poca Sal	4,000*	,483	,000	2,543	5,457
Poca Sal	Bastante Sal	-2,000*	,483	,010	-3,457	-,543
	Mucha Sal	-4,000*	,483	,000	-5,457	-2,543

Comparaciones por pares



Variable dependiente: Emisiones

(I) Maquinas	(J) Maquinas	Diferencia entre medias (I-J)	Error típ.	Significación <sup>a</sup>	Intervalo de confianza al 95 % para diferencia <sup>a</sup>	
					Límite inferior	Límite superior
Máquina I	Máquina II	-3,333*	,624	,007	-5,723	-,943
	Máquina III	-1,333	,624	,650	-3,723	1,057
	Máquina IV	-,667	,624	1,000	-3,057	1,723
	Máquina V	-3,000*	,624	,013	-5,390	-,610
Máquina II	Máquina I	3,333*	,624	,007	,943	5,723
	Máquina III	2,000	,624	,125	-,390	4,390
	Máquina IV	2,667*	,624	,027	,277	5,057
	Máquina V	,333	,624	1,000	-2,057	2,723
Máquina III	Máquina I	1,333	,624	,650	-1,057	3,723
	Máquina II	-2,000	,624	,125	-4,390	,390
	Máquina IV	-,667	,624	1,000	-1,723	3,057
	Máquina V	-1,667	,624	,282	-4,057	,723
Máquina IV	Máquina I	-,667	,624	1,000	-1,723	3,057
	Máquina II	-2,667*	,624	,027	-5,057	-,277
	Máquina III	-,667	,624	1,000	-3,057	1,723
	Máquina V	-2,333	,624	,057	-4,723	,057
Máquina V	Máquina I	3,000*	,624	,013	,610	5,390
	Máquina II	-,333	,624	1,000	-2,723	2,057
	Máquina III	1,667	,624	,282	-,723	4,057
	Máquina IV	2,333	,624	,057	-,057	4,723

\*. La diferencia de las medias es significativa al nivel ,05.

a. Ajuste para comparaciones múltiples: Bonferroni.