

ANÁLISIS DE LA VARIANZA

MODELO DOS CRITERIOS DE CLASIFICACIÓN CON INTERACCIONES

El modelo de Análisis de la Varianza para dos criterios de clasificación de efectos fijos o Modelo I, supone que los elementos observados se asignan a celdas en donde se combinan dos factores y se dispone de un grupo de observaciones en cada combinación, a diferencia del diseño en bloques al azar que solamente cuenta con una observación por celda. Este modelo considera además que los efectos de ambos factores interactúan.

Por ejemplo se cuenta con 24 parcelas y 6 combinaciones de dos factores. El primer factor se refiere a dos tipos de semilla y el segundo factor a tres niveles de aplicación de fertilizante. En cada celda o combinación quedan los tratamientos y se asignan 4 parcelas al azar. La variable observada es el rendimiento.

Experimento con dos factores de variación y el mismo número de observaciones por casilla. Se dispone de un conjunto de observaciones homogéneas y se asigna los tratamientos (combinación de niveles de los dos factores) aleatoriamente a las unidades.

Los datos resultantes serían los siguientes

Factor A		A_1	...	A_r
Factor B	B_1	$\begin{bmatrix} x_{111} \\ \vdots \\ x_{11t} \end{bmatrix}$...	$\begin{bmatrix} x_{1r1} \\ \vdots \\ x_{1rt} \end{bmatrix}$
	\vdots	\vdots	\ddots	\vdots
	B_s	$\begin{bmatrix} x_{s11} \\ \vdots \\ x_{s1t} \end{bmatrix}$...	$\begin{bmatrix} x_{rs1} \\ \vdots \\ x_{rst} \end{bmatrix}$

El modelo de dos criterios de clasificación tiene la siguiente forma.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad i=1,2,\dots,a \quad j=1,2,\dots,b \quad k=1,\dots,n_{ij}$$

Y_{ijk}	Valor de la observación en el tratamiento i, bloque j
μ	Media general de la población
α_i	Efecto del tratamiento $i=1,2,\dots,a$
β_j	Efecto del bloque $j=1,2,\dots,b$
$(\alpha\beta)_{ij}$	Interacción del nivel i del primer factor y nivel j del segundo
ε_{ijk}	Error aleatorio independiente $\varepsilon_{ijk} \rightarrow N(0, \sigma^2)$

Los niveles $i=1,2,\dots,a$ del primer α factor y los niveles $j=1,2,\dots,b$ del segundo factor β se identifican como efectos principales. Si los efectos principales son estrictamente aditivos, todas las interacciones son cero, pero si no es el caso, entonces éstas adoptan diversos valores. Las $a \times b$ combinaciones dan lugar a las interacciones.

$$\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{i=1}^a (\alpha\beta)_{\bar{i}} = \sum_{j=1}^b (\alpha\beta)_{\bar{j}} = 0$$

DESCOMPOSICIÓN DE LASUMA DE CUADRADOS TOTAL.

Se calculan medias por efecto principal y por las combinaciones de ambos efectos.

$$\bar{Y}_{ij.} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk} \quad \bar{Y}_{.j.} = \frac{1}{a} \sum_{i=1}^a \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk} \quad \bar{Y}_{i..} = \frac{1}{b} \sum_{j=1}^b \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$$

$$\bar{Y}_{...} = \frac{1}{a} \sum_{i=1}^a \frac{1}{b} \sum_{j=1}^b \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$$

Las desviaciones de cada observación respecto de la media total se expresan alternativamente como diversas desviaciones para aislar los efectos de los factores principales las interacciones y la parte residual.

$$Y_{ijk} - \bar{Y} = (\bar{Y}_i - \bar{Y}) + (\bar{Y}_j - \bar{Y}) + (\bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}) + (Y_{ijk} - \bar{Y}_{ij.})$$

La suma de cuadrados total de desviaciones respecto de la media global se expresa en forma análoga por las sumas de cuadrados de efectos principales, interacciones y la parte residual.

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y})^2 &= bn \sum_{i=1}^a (\bar{Y}_{i..} - \bar{Y}_{...})^2 + an \sum_{j=1}^b (\bar{Y}_{.j.} - \bar{Y}_{...})^2 + n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \\ &+ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2 \end{aligned}$$

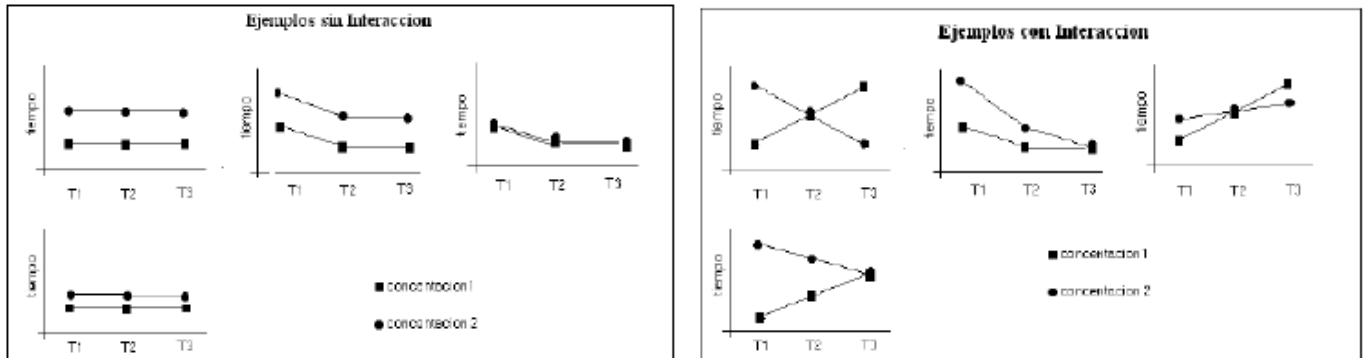
Las sumas de cuadrados junto con sus respectivos grados de libertad se combinan para obtener los cuadrados medios y finalmente las estadísticas de prueba F para efectos principales e interacciones.

ANOVA CON DOS FACTORES E CON INTERACCIONES

Fuente de Varianza	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	F Calculada
Factor A	$bn \sum_{i=1}^a (\bar{Y}_{i..} - \bar{Y}_{...})^2$	(a-1)	$bn \sum_{i=1}^a (\bar{Y}_{i..} - \bar{Y}_{...})^2 / (a-1)$ Msa	Msa/Se ²
Factor B	$an \sum_{j=1}^b (\bar{Y}_{.j.} - \bar{Y}_{...})^2$	(b-1)	$an \sum_{j=1}^b (\bar{Y}_{.j.} - \bar{Y}_{...})^2 / (b-1)$ Msb	Msb/Se ²
Interacción AB	$n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$	(a-1)(b-1)	$n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 / ((a-1)(b-1))$ Msab	Msab/Se ²
Residual	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2$	ab(n-1)	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2 / (ab(n-1))$ Se ²	
Total	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y})^2$	abn-1		

Cuando el número de observaciones dentro de cada casillero es constante decimos que el **diseño es balanceado**

Para cada observación, podríamos considerar un modelo que involucre una **media general**, el **efecto del tratamiento** y el **efecto de la interacción**



Sin embargo, podría ocurrir que el efecto de cierto tratamiento no sea el mismo para los distintos niveles de concentración. En este caso diríamos que hay interacción. Como vemos en el segundo cuadro

Se deben de cumplir los siguientes requisitos

- 1) **Normalidad.** Los datos obtenidos en cada nivel de los factores se ajustan razonablemente a una distribución Normal Y_{ij} sigue una distribución $N(\mu_{ij}, \sigma)$ para cada ij
- 2) **Homocedasticidad** la variabilidad de los datos en cada nivel de los factor es similar (contraste de varianzas) $\sigma^2 = \text{Var}(Y_{ij})$ igual para todo i, j
- 3) **Linealidad los residuos** (diferencia de los datos a su media, en cada nivel de los factores) se distribuyen alrededor del cero $E(\mu) = 0$
- 4) **Independencia** las observaciones se realizan de forma independiente unas de otras.

Ejemplo 1

Considérese un experimento que consiste en aplicar 2 tipos de semilla combinados con 3 niveles de fertilizante a 24 parcelas. Se asignan 4 parcelas a cada una de las 6 combinaciones o tratamientos resultantes. La variable respuesta es el rendimiento (Kg/M^2)

Cálculos Fundamentales.

Se calcularán medias por combinación de tipo de semilla y nivel de fertilizante y las medias marginales por semilla y fertilizante. La combinación que produce el mayor rendimiento corresponde al Fertilizante en nivel Alto y a las Semilla 1, con $18.225 \text{ Kg}/\text{M}^2$

Semilla	Nivel de Fertilizante			Media
	Bajo	Medio	Alto	
1	14.30	18.10	17.60	
	14.50	17.60	18.20	
	11.50	17.10	18.90	
	13.60	17.60	18.20	
Media	13.475	17.600	18.225	16.433
2	12.60	10.50	15.70	
	11.20	12.80	17.50	
	11.00	8.30	16.70	
	12.10	9.10	16.60	
Media	11.725	10.175	16.625	12.842
Media	12.600	13.888	17.425	14.638
Semilla	Bajo	Medio	Alto	Media
Semilla 1	13.475	17.600	18.225	16.433
Semilla 2	11.725	10.175	16.625	12.842
Media	12.600	13.888	17.425	14.638

Se procede a calcular las sumas de cuadrados que se requieren en la tabla de ANOVA2, para ello se utilizan las tablas auxiliares siguientes:

Sumas de Cuadrados de Efectos Principales

Semilla	Bajo	Medio	Alto	Media
Semilla 1	13.475	17.600	18.225	16.433
Semilla 2	11.725	10.175	16.625	12.842
Media	12.600	13.888	17.425	14.638

Diferencias	Cuadrado	Suma de C Semillas
1.796	3.22501736	
-1.796	3.22501736	a*n* Suma
Suma	6.45003472	77.4004167

Diferencias	-2.038	-0.750	2.788	Suma
Cuadrado	4.151	0.562	7.770	12.484
	Suma de C Fertilizante	bn Suma		99.8725

Suma de cuadrados de residuales

Semilla	Nivel de Fertilizante		
	Bajo	Medio	Alto
1	14.30	18.10	17.60
	14.50	17.60	18.20
	11.50	17.10	18.90
	13.60	17.60	18.20
Media	13.475	17.600	18.225
2	12.60	10.50	15.70
	11.20	12.80	17.50
	11.00	8.30	16.70
	12.10	9.10	16.60
Media	11.725	10.175	16.625

Diferencias Respecto a Media de su Grupo

0.8250	0.5000	-0.6250
1.0250	0.0000	-0.0250
-1.9750	-0.5000	0.6750
0.1250	0.0000	-0.0250
0.8750	0.3250	-0.9250
-0.5250	2.6250	0.8750
-0.7250	-1.8750	0.0750
0.3750	-1.0750	-0.0250

Cuadrado de las Diferencias			
0.6806	0.2500	0.3906	
1.0506	0.0000	0.0006	
3.9006	0.2500	0.4556	
0.0156	0.0000	0.0006	6.9950
0.7656	0.1056	0.8556	
0.2756	6.8906	0.7656	
0.5256	3.5156	0.0056	
0.1406	1.1556	0.0006	15.0025

Suma de Cuadrados Residual	21.9975
----------------------------	---------

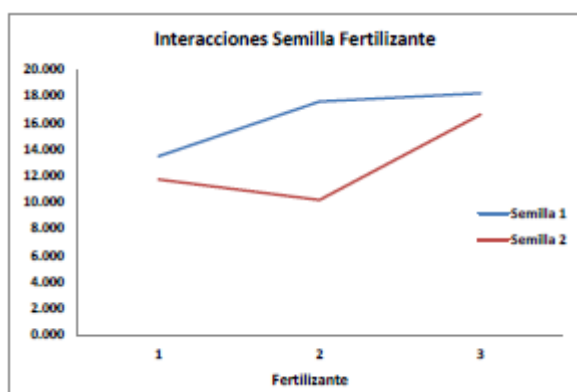
Datos originales		
14.30	18.10	17.60
14.50	17.60	18.20
11.50	17.10	18.90
13.60	17.60	18.20
12.60	10.50	15.70
11.20	12.80	17.50
11.00	8.30	16.70
12.10	9.10	16.60

n =	24
Suma de Cuadrados Total	243.37625

sumas de cuadrados se trasladan una Tabla de Análisis de

Estas sumas de cuadrados se trasladan una Tabla de Análisis de la Varianza

Anova dos criterios de clasificación con interacciones					
Fuente de Variación	Suma de Cuadrados	GL	Cuadrados Medios	F	P VALUE
Fertilizante	99.87	2	49.9363	40.8616	2.03371E-07
Semilla	77.40	1	77.4004	63.3348	2.63979E-07
Interaccion F*S	44.11	2	22.0529	18.0453	5.00437E-05
Residual	22.00	18	1.2221		
Total	243.38	23			



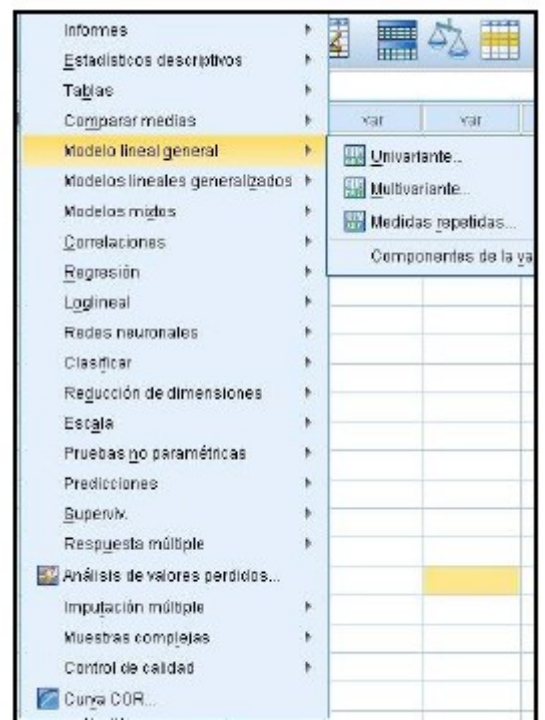
Las estadísticas F asociadas a los efectos principales y la interacción resultan significativas. La siguiente gráfica muestra las diferencias entre los efectos principales y la interacción. Como se observa en la tabla que muestra las medias de las combinaciones por tratamiento.

Cálculos Mediante SPSS.

Los datos para SPSS se ordenan en columnas que identifican los efectos principales y la variable respuesta

	SEMILLA	FERTILIZANTE	RENDIMIENTO
1	1	1	14.30
2	1	1	14.50
3	1	1	11.50
4	1	1	13.60
5	1	2	18.10
6	1	2	17.60
7	1	2	17.10
8	1	2	17.60
9	1	3	17.60
10	1	3	16.20
11	1	3	18.90
12	1	3	16.20
13	2	1	12.60
14	2	1	11.20
15	2	1	11.00
16	2	1	12.10
17	2	2	10.50
18	2	2	12.80
19	2	2	8.30
20	2	2	9.10
21	2	3	16.70
22	2	3	17.50
23	2	3	16.70

Se procede a seleccionar el menú de **Análisis** la opción **Modelo lineal general** y a continuación **Univariante**.



La variable dependiente es RENDIMIENTO y los Factores serán SEMILLA y FERTILIZANTE. En la opción modelo se selecciona Modelo I (Efectos Fijos)

En opciones se seleccionan medias para cada factor y la interacción. Se procede a aceptar el procedimiento y se obtienen los siguientes reportes:

Pruebas de los efectos inter-sujetos

Variable dependiente: RENDIMIENTO

Origen	Suma de cuadrados tipo I	gl	Media cuadrática	F	Sig.	Parámetro de no centralidad Parámetro	Potencia observada ^b
Modelo corregido	221.379 ^a	5	44.276	36.230	.000	181.149	1.000
Intersección	5142.154	1	5142.154	4207.695	.000	4207.695	1.000
SEMILLA	77.400	1	77.400	63.335	.000	63.335	1.000
FERTILIZANTE	99.872	2	49.936	40.862	.000	81.723	1.000
SEMILLA * FERTILIZANTE	44.106	2	22.053	18.045	.000	36.091	.999
Error	21.998	18	1.222				
Total	5385.530	24					
Total corregida	243.376	23					

a. R cuadrado = .910 (R cuadrado corregida = .885)

b. Calculado con alfa = .05

Estadísticos descriptivos

Variable dependiente: RENDIMIENTO

SEMILLA	FERTILIZANTE	Media	Desviación típica	N
1	1	13.4750	1.37204	4
	2	17.6000	.40825	4
	3	18.2250	.53151	4
	Total	16.4333	2.34107	12
2	1	11.7250	.75443	4
	2	10.1750	1.97210	4
	3	16.6250	.73655	4
	Total	12.8417	3.09969	12
Total	1	12.6000	1.38770	8
	2	13.8875	4.18208	8
	3	17.4250	1.04163	8
	Total	14.6375	3.25293	24

1. SEMILLA

Variable dependiente: RENDIMIENTO

SEMILLA	Media	Error típ.	Intervalo de confianza 95%	
			Límite inferior	Límite superior
1	16.433	.319	15.763	17.104
2	12.842	.319	12.171	13.512

2. FERTILIZANTE

Variable dependiente: RENDIMIENTO

FERTILIZANTE	Media	Error típ.	Intervalo de confianza 95%	
			Límite inferior	Límite superior
1	12.600	.391	11.779	13.421
2	13.887	.391	13.066	14.709
3	17.425	.391	16.604	18.246

Contraste de Levene sobre la igualdad de las varianzas error^a

Variable dependiente: RENDIMIENTO

F	gl1	gl2	Sig.
2.483	5	18	.071

Contrasta la hipótesis nula de que la varianza error de la variable dependiente es igual a lo largo de todos los grupos.

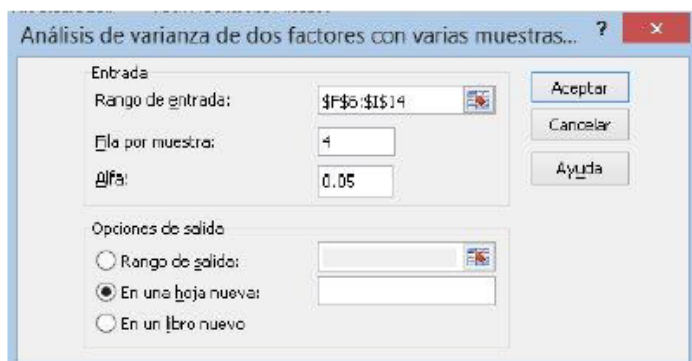
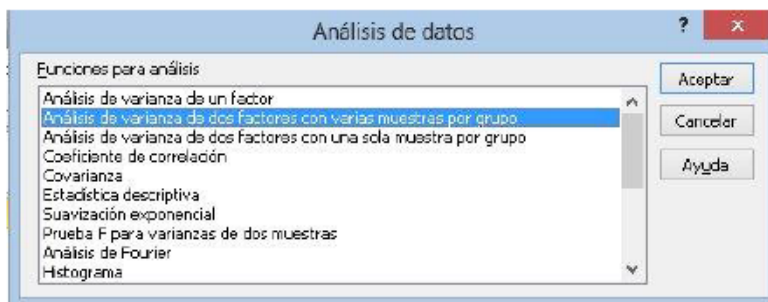
a. Diseño: Intersección + SEMILLA + FERTILIZANTE + SEMILLA * FERTILIZANTE

Cálculos Mediante EXCEL

Los datos en Excel se deben estructurar en una tabla como la siguiente

	FER1	FER2	FER3
SEM1	14.30	18.10	17.60
SEM1	14.50	17.60	18.20
SEM1	11.50	17.10	18.90
SEM1	13.60	17.60	18.20
SEM2	12.60	10.50	15.70
SEM2	11.20	12.80	17.50
SEM2	11.00	8.30	16.70
SEM2	12.10	9.10	16.60

Excel dispone entre sus Complementos la opción **Análisis de Datos**. Al seleccionar esta opción dentro del título del encabezado principal Datos, se abre la siguiente ventana, en la cual se selecciona **Análisis de varianza de dos factores con varias muestras por grupo**.



En la ventana que se abre a continuación, se define el rango incluyendo los títulos de columnas. Se declaran 4 filas por muestra que corresponden al número de observaciones por celda.

El reporte se presenta en una página nueva del mismo libro de Excel donde se opera.

RESUMEN	FER1	FER2	FER3	Total
SEM1				
Cuenta	4	4	4	12
Suma	53.9	70.4	72.9	197.2
Promedio	13.475	17.6	18.225	16.43333333
Varianza	1.8825	0.16666667	0.2825	5.48060606
SEM2				
Cuenta	4	4	4	12
Suma	46.9	40.7	66.5	154.1
Promedio	11.725	10.175	16.625	12.84166667
Varianza	0.56916667	3.88916667	0.5425	9.60810606
Total				
Cuenta	8	8	8	
Suma	100.8	111.1	139.4	
Promedio	12.6	13.8875	17.425	
Varianza	1.92571429	17.4898214	1.085	

Finalmente la Tabla de Análisis de la Varianza se presenta en la siguiente forma:

ANÁLISIS DE VARIANZA

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Muestra	77.4004	1	77.4004	63.3348	2.63979E-07	4.4138734
Columnas	99.8725	2	49.9363	40.8616	2.03371E-07	3.5545571
Interacción	44.1058	2	22.0529	18.0453	5.00437E-05	3.5545571
Dentro del grupo	21.9975	18	1.2221			
Total	243.37625	23				

Ejemplo 2

En un estudio sobre memoria verbal se seleccionaron al azar 50 personas mayores y 50 jóvenes (factor 1: edad) Dentro de cada grupo se asignaron al azar, 10 personas a 5 distintos grupos se le dieron las siguientes instrucciones (factor 2; método).

Grupo 1 (contar) se les pidió que contarán el número de letras de cada palabra

Grupo 2 (rimar) se les pidió que rimasen cada palabra con otra

Grupo 3 (adjetivar) se les pidió que a cada palabra le asignaran un adjetivo

Grupo 4 (imaginar) se les pidió que a cada palabra le asignaran una imagen

Grupo 5 (recordar) se les pidió que memorizaran las palabras

A los 4 primeros grupos no se les dijo que deberían recordar las palabras. Finalmente, tras revisar la lista 3 veces, se recogió el número de palabras recordadas por cada grupo (variable de respuesta)

		Factor 2 Método				
		Contar	Rimar	Adjetivar	Imaginar	Recordar
F a c t o r 1	M a y o r e s	9	7	11	12	10
		8	9	13	11	19
		6	6	8	16	14
		8	6	6	11	5
		10	6	14	9	10
		4	11	11	23	11
		6	6	13	12	14
		5	3	13	10	15
		7	8	10	19	11
		7	7	11	11	11
J ó v e n e s	J ó v e n e s	8	10	14	20	21
		6	7	11	16	19
		4	8	18	16	17
		6	10	14	15	15
		7	4	13	18	22
		6	7	22	16	16
		5	10	17	20	22
		7	6	16	22	22
		9	7	12	14	18
		7	7	11	19	21

Realice los cálculos correspondientes al análisis de la varianza con 2 factores e interacciones.