ANÁLISIS DE LA VARIANZA
MODELO COMPLETO DE BLOQUES AL AZAR

EL MODELO

En el modelo de Análisis de la Varianza para un criterio de clasificación de efectos fijos o Modelo I, los elementos observados se supone que fueron asignados a cada grupo en forma completamente aleatorizada. Hay situaciones en las que no se tiene una aleatorización completa, sino que se presentan ciertos agrupamientos de elementos por algún otro criterio, que no es el objetivo principal del análisis, pero que permite agrupar los elementos. Por ejemplo, en un productor de ganado porcino pretende probar 2 nuevas dietas de alimentación y compararlas con la dieta actual, que se considerará como control. Dispone de 15 cerdos que a su vez proceden de 5 camadas diferentes y decide asignar 5 cerdos a cada dieta, pero de diferente camada cada cerdo.

Lo que el productor registra es la ganancia en Kg. de cada cerdo, después de un período determinado.

Bloque Camada	Dieta Control	Dieta 1	Dieta 2
1	3.74	4.47	5.65
2	4.58	6.78	7.00
3	4.58	5.19	6.08
4	4.47	5.19	5.74
5	4.79	6.85	7.55

El productor desea probar la hipótesis nula $Ho: \mu_1 = \mu_2 = \mu_3$

De manera equivalente $Ho: \alpha_1 = \alpha_2 = \alpha_3 = 0$

La hipótesis alternativa será entonces $H_1: \mu_i \neq \mu_j \; para\ cualquier\ \ i \neq j$

En forma equivalente $H_1: \alpha_i \neq 0 \; para\ alguna\ i$

El modelo de bloques completamente al azar tiene la siguiente forma. Este modelo supone aditividad en los efectos de tratamientos y bloques y sin interacción.

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad \text{i=1,2…a} \quad \text{j=1,2,….b}$$

Y_{ij} Valor de la observación en el tratamiento i, bloque j

μ Media general de la población

α_i Efecto del tratamiento i

β_j Efecto de bloque j

ε_{ij} Error aleatorio independiente y normalmente distribuido $\varepsilon_{ij} \rightarrow N(0, \sigma^2)$

También se cumplen las siguientes restricciones

$$\sum_{i=1}^{a} \alpha_i = 0 \qquad\qquad \sum_{j=1}^{b} \beta_j = 0$$

La estructura de los datos queda en forma de una matriz axb:

Bloque	Tratamiento			
	1	2	……	a
1	Y_{11}	Y_{21}	…..	Y_{a1}
2	Y_{12}	Y_{22}	…..	Y_{a2}
….	…..	…..	…..	…..
b	Y_{1b}	Y_{2b}	…..	Y_{ab}

Se definen las siguientes medias general. Por tratamiento y por bloque.

$$\bar{Y}_{...} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij} \qquad \bar{Y}_{i.} = \frac{1}{b} \sum_{j=1}^{b} Y_{ij} \qquad \bar{Y}_{.j} = \frac{1}{a} \sum_{i=1}^{a} Y_{ij}$$

Partiendo del modelo, la siguiente diferencia se estima en términos de las medias empíricas de la siguiente forma:

$$Y_{ij} - \mu = \alpha_i + \beta_j + \varepsilon_{ij}$$

$$\left(Y_{ij} - \bar{Y}_{..}\right) = \left(\bar{Y}_{i.} - \bar{Y}_{..}\right) + \left(\bar{Y}_{.j} - \bar{Y}_{..}\right) + \left(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}\right)$$

La variabilidad de toda la muestra se resume en la suma de cuadrados total (SCT).

$$SCT = \sum_{i=1}^{a} \sum_{j=1}^{b} \left(Y_{ij} - \bar{Y}_{..}\right)^2$$

La suma de cuadrados total, a su vez se descompone en las siguientes sumas de cuadrados, los productos cruzados y dobles productos se anulan debido a primeros momentos respecto de la media. La primera suma del miembro derecho es la suma de cuadrados entre tratamientos, la segunda es la suma de cuadrados entre bloques y la tercera es la suma de cuadrados residual, después de remover las sumas de cuadrados entre tratamientos y entre bloques.

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \left(Y_{ij} - \bar{Y}_{..}\right)^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\bar{Y}_{.j} - \bar{Y}_{..}\right)^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} \left(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}\right)^2$$

La suma de cuadrados total tiene asociados ab-1 grados de libertad. Se tienen originalmente axb observaciones, para calcular la suma de cuadrados se pierde un grado de libertad por el cálculo previo de la media total. En forma similar se parte de los a tratamientos, cuya suma de cuadrados pierde un grado de libertad por la media total y queda con a-1. De manera similar la suma de cuadrados de bloques tiene b-1 grados de libertad y la suma de cuadrados de los residuos queda con (a-1)(b-1) grados de libertad. El cociente de las sumas de cuadrados entre sus grados de libertad da lugar a los cuadrados medios.

Se cumple la aditividad de los grados de libertad de las sumas de cuadrados anteriores:

(ab-1)=(a-1)+(b-1)+(a-1)(b-1)

Tanto el error cuadrático medio de los tratamientos, como el error cuadrático medio entre bloques son estimadores de la varianza con un sesgo que depende de sus efectos. Se puede ver que si las hipótesis de nulidad se cumplen, entonces se tienen estimadores insesgados de la varianza. Por otra parte el cuadrado medio del residual es insesgado independientemente de que se cumplan o no las nulidades de los efectos de tratamientos. Es lógico considerar el cociente del cuadrado medio de tratamientos entre el cuadrado medio del error para probar la nulidad de efecto de los tratamientos. A medida que el cociente es mayor a 1 se tiene más evidencia de que la hipótesis de nulidad no es sostenible.

$$E(ECMEntreTratamientos) = \sigma^2 + \sum_{i=1}^{a} \frac{\alpha_i^2}{a-1}$$

Y_{ab}

$$E(ECMEntreBloques) = \sigma^2 + \sum_{j=1}^{b} \frac{\beta_j^2}{b-1}$$

$$E(ECMdeResiduales) = \sigma^2$$

El supuesto de normalidad tiene como consecuencia que las sumas de cuadrados se distribuyan como Ji cuadrada con sus correspondientes grados de libertad en forma independiente. El cociente de dos variables que se distribuyen Ji cuadrada, se distribuye F con los grados de libertad de numerador y denominador. Entonces el cociente del cuadrado medio de tratamientos entre el cuadrado medio del error se distribuye F con a-1 y (a-1)(b-1) grados de libertad.

$$F_a = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} \left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 /(a-1)}{\sum_{i=1}^{a} \sum_{j=1}^{b} \left(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}\right)^2 /[(a-1)(b-1)]} \to F_{a-1,(a-1)(b-1)}$$

$$F_b = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{.j}-\bar{Y}_{..}\right)^2/(b-1)}{\sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..}\right)^2/\left[(a-1)(b-1)\right]} \rightarrow F_{b-1,(a-1)(b-1)}$$

Los cálculos se arreglan en una tabla de análisis de la varianza de la siguiente forma:

La hipótesis de nulidad de bloques, formalmente es incorrecto probarla, pues los bloques tienen la restricción de aleatoriedad. En términos simples, los bloques intencionalmente se construyen diferentes y no tiene sentido probar que son diferentes.

TABLA ANOVA BLOQUES AL AZAR

Fuente de Variación	Sumas de Cuadrados	Grados de Libertad	Cuadrados Medios	$F_{(a-1),(a-1)(b-1)}$
Tratamientos	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)^2$	a - 1	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)^2\frac{1}{(a-1)}$ (I)	(I)/(III)
Bloques	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{.j}-\bar{Y}_{..}\right)^2$	b -1	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{.j}-\bar{Y}_{..}\right)^2\frac{1}{(b-1)}$ (II)	(Ii)/(III)
Residual	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..}\right)^2$	(a- 1)(b -1)	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..}\right)^2\frac{1}{(a-1)(b-1)}$ (III)	
Total	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij}-\bar{Y}_{..}\right)^2$	ab - 1	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij}-\bar{Y}_{..}\right)\frac{1}{(n-1)}^2$ (IV)	

Ejemplo

En el ejemplo de las dietas se calculan las medias por tratamiento, por bloque y total,

Bloque	Control	Tra 1	Trat 2	Media Bloques
1	3.74	4.47	5.65	4.620
2	4.58	6.78	7.00	6.120
3	4.58	5.19	6.08	5.283
4	4.47	5.19	5.74	5.133
5	4.79	6.85	7.55	6.397
Media Tratamiento	4.432	5.696	6.404	5.511

Las sumas de cuadrados indicadas en la tabla de ANOVA se presentan a continuación:

ANOVA bloques al azar

Fuente de Variación	Sumas de Cuadrados	Grados de Libertad	Cuadrados Medios	F	P value
Tratamientos	9.979573333	2	4.98978667	26.1548918	0.00030961
Bloques	6.430893333	4	1.60772333	8.42717989	0.00572898
Residual	1.526226667	8	0.19077833		
Total	17.93669333	14	1.28119238		

Como la probabilidad asociada a la igualdad de varianzas es pequeña, la hipótesis se rechaza y se procede a continuación con las pruebas de comparación múltiple de tratamientos mediante Scheffé.

La hipótesis que se prueba en cada caso es que las diferencias de las medias son iguales a cero

$$Ho: \mu_A - \mu_B = 0 \quad vs \quad H_1: \mu_A - \mu_B \neq 0$$

La estadística de prueba de Scheffé se calcula como el cociente de la diferencia entre la medias aritmética mayor menos la menor, dividida entre el Error Estándar del Cuadrado Medio del Residual (CMR) que se extrae de la tabla de análisis de la varianza.

$$E_S = \frac{\bar{Y}_A - \bar{Y}_B}{EE(E_S)} \qquad CMR = \sum_{i=1}^{a} \sum_{j=1}^{b} \left(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}\right)^2 \frac{1}{(a-1)(b-1)}$$

El error estándar de la diferencia se obtiene mediante la raíz cuadrada del producto del CMDG multiplicado por la suma de los recíprocos de los tamaños de muestra de las medias de los grupos involucrados.

$$EE(E_S) = \sqrt{CMR\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

El discriminante de la prueba (Umbral de Scheffé) se obtiene mediante la raíz cuadrada del número total de tratamientos menos 1, multiplicada por el percentil de la distribución F correspondiente al nivel de significancia (α) de la prueba con a-1 y (a-1)(b-1) grados de libertad.

$$\sqrt{(a-1)F(\alpha, a-1, (a-1)(b-1)}$$

Grupo A	Grupo B	Media A	Media B	Diferencia M1-M2	n_A	n_B	Error Est. Diferencia	Estadística Sheffé	$\sqrt{(a-1)F(\alpha,a-1,(a-1)(b-1)}$ Umbral de Sheffé	Conclusión
Trat 2	Control	6.404	4.432	1.972	5	5	0.27624506	7.1386	2.9863	Rechazo
Trat 2	Trat 1	6.404	5.696	0.708	5	5	0.27624506	2.5629	2.9863	No Rechaza
Trat 1	Control	5.696	4.432	1.264	5	5	0.27624506	4.5756	2.9863	Rechazo

Las diferencias entre control y tratamientos resultan significativas, pero entre los dos tratamientos la diferencia no resulta significativa.

Intervalos de confianza para las diferencias de medias en prueba de Scheffé.

Se obtiene al restar y sumar el error estándar de la diferencia, multiplicado por el umbral.

$$(\bar{Y}_A - \bar{Y}_B) \pm EE(E_s)\sqrt{(a-1)F(\alpha, a-1, (a-1)(b-1)}$$

Grupo A	Grupo B	Diferencia M1-M2	Error Est. Diferencia	$\sqrt{(a-1)F(\alpha,a-1,(a-1)(b-1)}$ Umbral de Sheffé	$EE(E_s)\sqrt{(a-1)F(\alpha,a-1,(a-1)(b-1)}$	Límite Inferior	Límite Superior
Trat 2	Control	1.972	0.2762451	2.986	0.8249	1.1471	2.7969
Trat 2	Trat 1	0.708	0.2762451	2.986	0.8249	-0.1169	1.5329
Trat 1	Control	1.264	0.2762451	2.986	0.8249	0.4391	2.0889

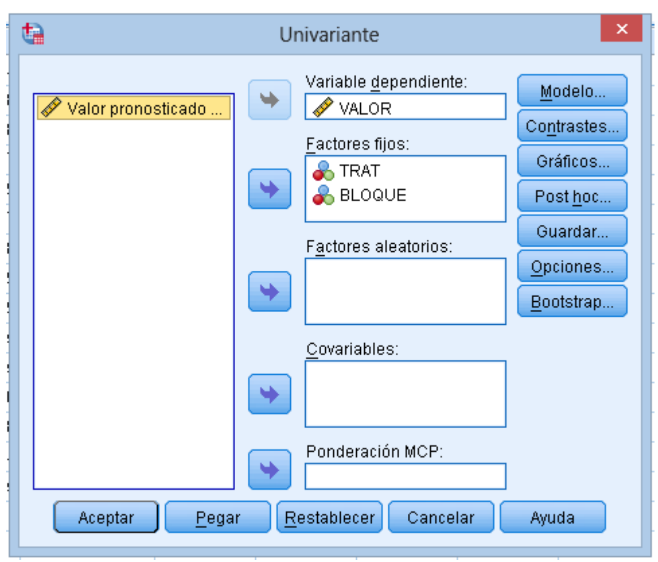
Observe que el intervalo para la diferencia entre tratamiento 1 y tratamiento 2 cubre al origen.

SOLUCIÓN MEDIANTE SPSS

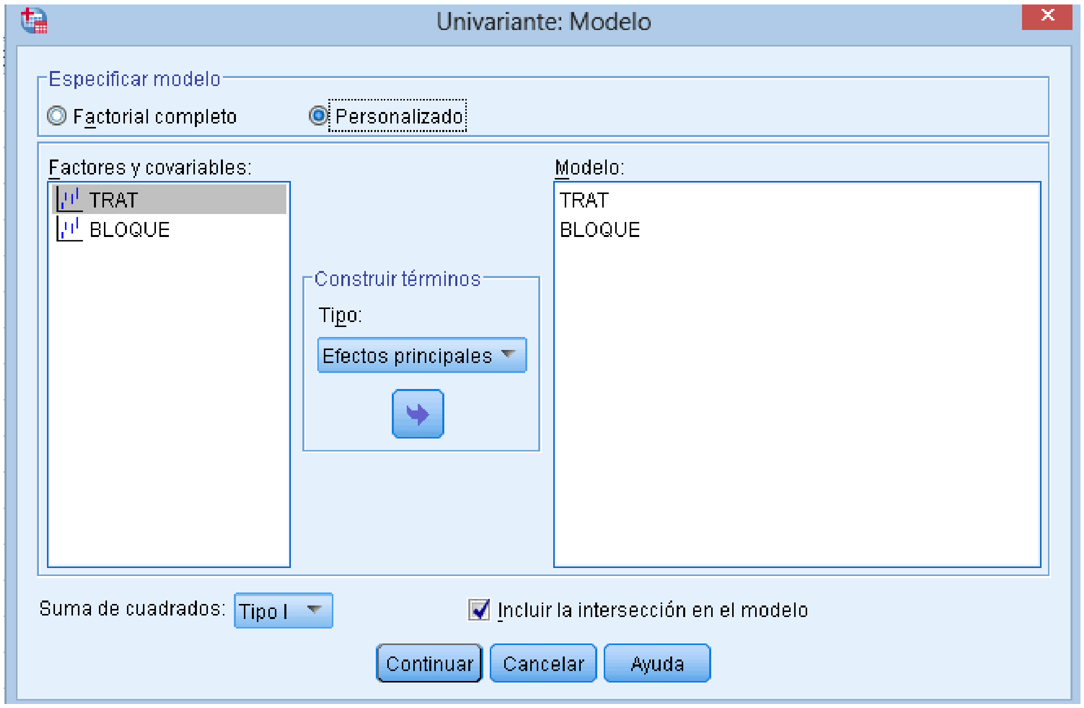
La estructura de datos se integra por tres columnas, la primera columna con la identificación del tratamiento, la segunda con el identificador de bloque y la tercera con el valor de la variable.

	TRAT	BLOQUE	VALOR
1	1	1	3.74
2	1	2	4.58
3	1	3	4.58
4	1	4	4.47
5	1	5	4.79
6	2	1	4.47
7	2	2	6.78
8	2	3	5.19
9	2	4	5.19
10	2	5	6.85
11	3	1	5.65
12	3	2	7.00
13	3	3	6.08
14	3	4	5.74
15	3	5	7.55

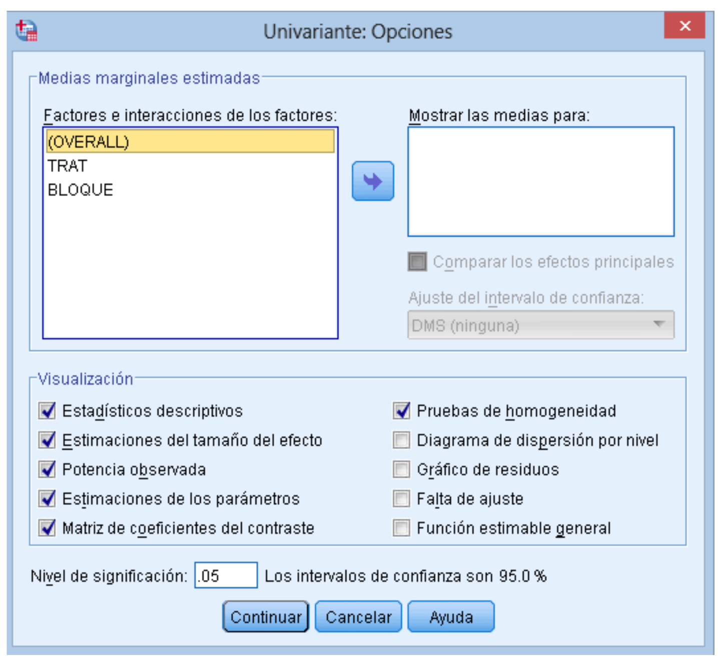
A continuación, en el menú de Analizar de SPSS se selecciona Modelo Lineal General y dentro de este Univariante. En la ventana que se abre, se trasladan las variables de tratamiento y bloques a Factores Fijos y en la Variable Dependiente se traslada el valor de la variable de análisis.



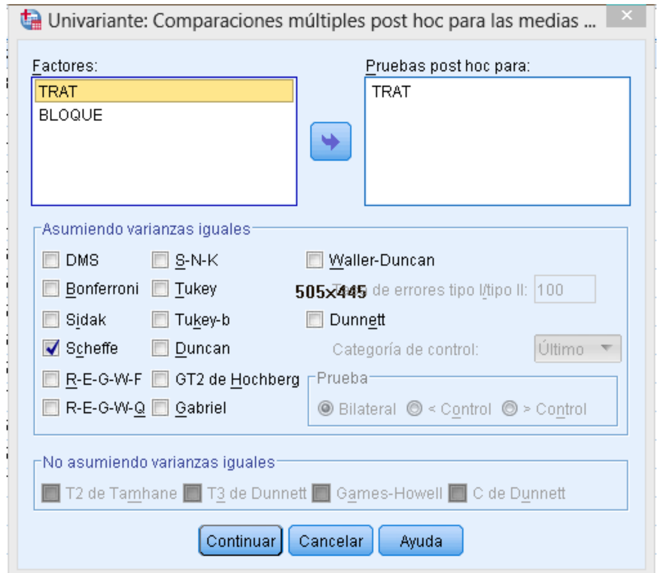
El siguiente paso consiste en la especificación del Modelo. Se abre la ventana siguiente, en la cual se selecciona Personalizado y tanto tratamientos como bloques se trasladan a modelo. De Tipo se selecciona Efectos Principales. La suma de cuadrados se selecciona para Tipo I y se deja la inclusión de la intersección en el modelo.



Se procede a seleccionar en las Opciones los reportes adicionales que se deseen, como Estadísticas Descriptivas, estimación de parámetros, Pruebas de Homogeneidad de Varianzas, etc.



Finalmente se seleccionan las pruebas para comparaciones múltiples. En este caso se selecciona la prueba de Scheffé por su robustez.



Reportes de SPSS de Análisis de la Varianza y Estimaciones de coeficientes del modelo para tratamientos y bloques prueba de Levene para homogeneidad de varianzas.

Contraste de Levene sobre la igualdad de las varianzas error[a]

Variable dependiente: VALOR

F	gl1	gl2	Sig.
	14	0	

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

a. Diseño: Intersección + TRAT + BLOQUE

Pruebas de los efectos inter-sujetos

Variable dependiente: VALOR

Origen	Suma de cuadrados tipo I	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial	Parámetro de no centralidad Parámetro	Potencia observada[b]
Modelo correg	16.410[a]	6	2.735	14.336	.000683	.915	86.019	.999
Intersección	455.512	1	455.512	2387.649	.000000	.997	2387.649	1.000
TRAT	**9.980**	**2**	**4.98979**	**26.15489**	**.000310**	**.867**	**52.310**	**1.000**
BLOQUE	**6.431**	**4**	**1.60772**	**8.42718**	**.005729**	**.808**	**33.709**	**.948**
Error	**1.526**	**8**	**.190778**					
Total	473.448	15						
Total corregid	17.937	14						

a. R cuadrado = .915 (R cuadrado corregida = .851)

b. Calculado con alfa = .05

Estimaciones de los parámetros

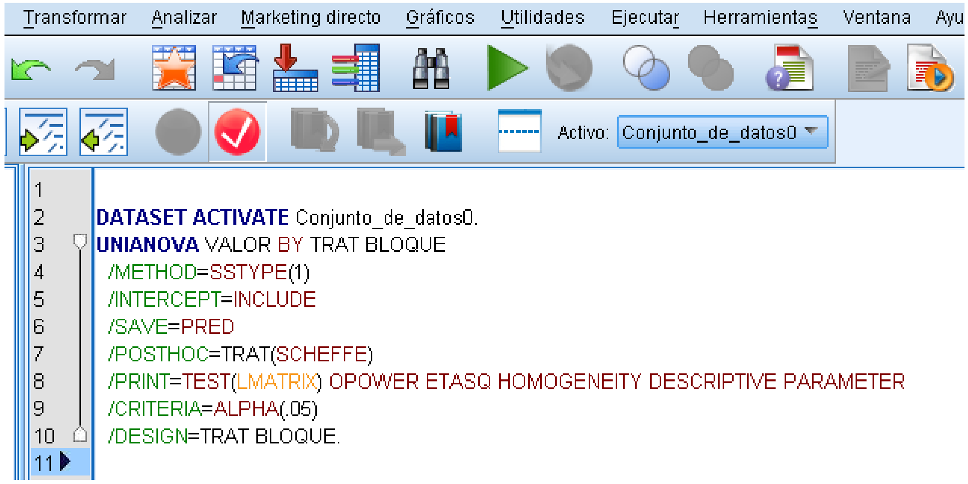
Variable dependiente: VALOR

Parámetro	B	Error típ.	t	Sig.	Intervalo de confianza 95% Límite inferior	Límite superior	l cuadrado pa	e no centralida	encia observa
Intersección	7.290	.298	24.432	.000	6.602	7.978	.987	24.432	1.000
[TRAT=1]	-1.972	.276	-7.139	.000	-2.609	-1.335	.864	7.139	1.000
[TRAT=2]	-.708	.276	-2.563	.033	-1.345	-.071	.451	2.563	.614
[TRAT=3]	0[a]								
[BLOQUE=1]	-1.777	.357	-4.982	.001	-2.599	-.954	.756	4.982	.991
[BLOQUE=2]	-.277	.357	-.776	.460	-1.099	.546	.070	.776	.106
[BLOQUE=3]	-1.113	.357	-3.122	.014	-1.936	-.291	.549	3.122	.780
[BLOQUE=4]	-1.263	.357	-3.542	.008	-2.086	-.441	.611	3.542	.872
[BLOQUE=5]	0[a]								

a. Al parámetro se le ha asignado el valor cero porque es redundante.

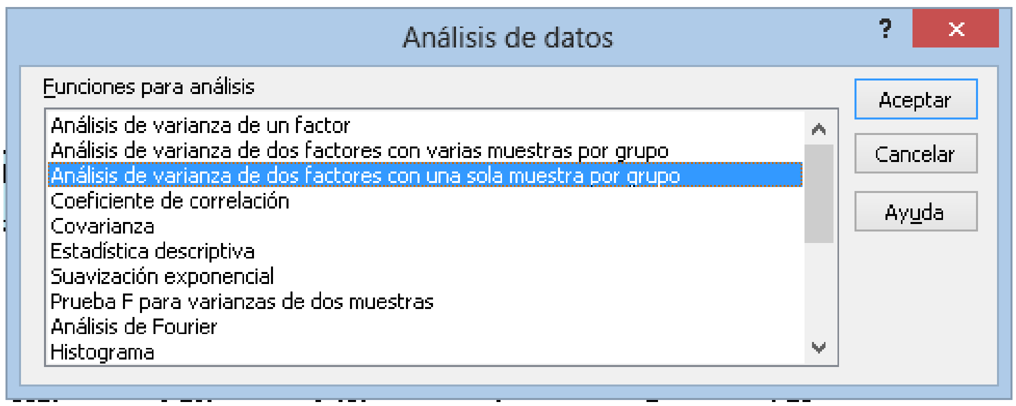
b. Calculado con alfa = .05

Código de SPSS



SOLUCIÓN MEDIANTE EXCEL

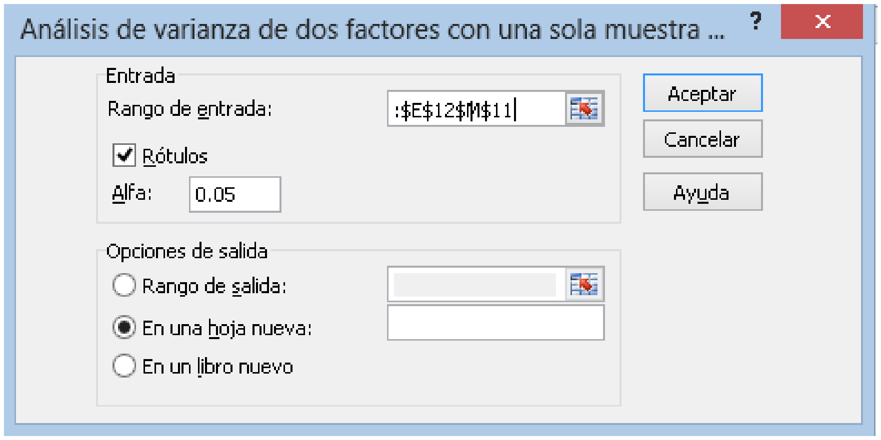
Los cálculos para el análisis de bloques al azar con una observación por celda se pueden efectuar en Excel de forma muy sencilla, siguiendo los siguientes pasos.



En el menú principal se debe seleccionar la opción **Datos** y dentro de ésta la opción **Análisis de Datos**, que debe estar previamente activada a partir de los complementos. Se abre la siguiente ventana de diálogo. En la cual se selecciona el **Análisis de la varianza de dos factores con una sola muestra por grupo.**

A continuación se abre la ventana específica, en la cual se define la zona de la página de Excel que contiene los datos, incluidos los títulos o rótulos de renglones y columnas.

Bloque	Control	Tra 1	Trat 2
1	3.74	4.47	5.65
2	4.58	6.78	7.00
3	4.58	5.19	6.08
4	4.47	5.19	5.74
5	4.79	6.85	7.55



El reporte que se genera inmediatamente en una página nueva es el siguiente. La estadística F asociada a tratamientos es la que se identifica como columnas en la tabla de análisis de la varianza.

Análisis de varianza de dos factores con una sola muestra por grupo

RESUMEN	*Cuenta*	*Suma*	*Promedio*	*Varianza*
1	3	13.8600	4.6200	0.9289
2	3	18.3600	6.1200	1.7908
3	3	15.8500	5.2833	0.5690
4	3	15.4000	5.1333	0.4056
5	3	19.1900	6.3967	2.0585
Control	5	22.1600	4.4320	0.16307
Tra 1	5	28.4800	5.6960	1.13048
Trat 2	5	32.0200	6.4040	0.69573

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Filas	6.43089333	4	1.60772333	8.4271799	0.005728984	3.837853355
Columnas	9.97957333	2	4.98978667	26.1548918	0.000309606	4.458970108
Error	1.52622667	8	0.19077833			
Total	17.9366933	14				

Ejercicio 2

Debido a la proliferación de los campos de golf y a la gran cantidad de agua que necesitan, un grupo de científicos estudia la calidad de varios tipos de césped para implantarlo en invierno en los campos de golf. Para ello, miden la distancia recorrida por una pelota de golf, en el campo, después de bajar por una rampa (para proporcionar a la pelota una velocidad inicial constante). El terreno del que disponen tiene mayor pendiente en la dirección norte-sur, por lo que se aconseja dividir el terreno en cinco bloques de manera que las pendientes de las parcelas individuales dentro de cada bloque sean las mismas. Se utilizó el mismo método para la siembra y las mismas cantidades de semilla. Las mediciones son las distancias desde la base de la rampa al punto donde se pararon las pelotas. En el estudio se incluyeron las variedades: Agrostis Tenuis (Césped muy fino y denso, de hojas cortas y larga duración), Agrostis Canina (Hoja muy fina, estolonífera. Forma una cubierta muy tupida), Paspalum Notatum (Hojas gruesas, bastas y con rizomas. Forma una cubierta poco densa), Paspalum Vaginatum (Césped fino, perenne, con rizomas y estolones).

	Variedad de Césped			
Bloques	Agrostis Tenuis	Agrostis Canina	Paspalum Notatum	Paspalum Vaginatum
1	1.3	2.2	1.8	3.9
2	1.6	2.4	1.7	4.4
3	0.5	0.4	0.6	2.0
4	1.2	2.0	1.5	4.1
5	1.1	1.8	1.3	3.4

1. Identificar los elementos del estudio (factores, unidades experimentales, variable respuesta, etc.) y plantear detalladamente el modelo matemático utilizado en el experimento.
2. ¿Son los bloques fuente de variación?
3. Existen diferencias reales entre las distancias medias recorridas por una pelota de golf en los distintos tipos de césped?
4. Utilizando el método de Scheffé, ¿qué tipo de cesped ofrece menor resistencia al recorrido de las pelotas?