ANALISIS DE LA VARIANZA
MODELO CON UN CRITERIO DE CLASIFICACION ANOVA 1
COMPARACIONES MÚLTIPLES POSTERIORES

Los métodos para verificar hipótesis empleados hasta ahora, involucran uno o dos grupos de muestras extraídos de otras tantas poblaciones. Sin embargo, una generalización obvia es involucrar tres o más muestras. Si se trata de 3 muestras independientes, puede resultar lógico que si se trata de verificar su igualdad de medias $Ho: \mu_1 = \mu_2 = \mu_3$ se recurriera a tomar las combinaciones de los 3 grupos tomando 2 a la vez $Ho: \mu_1 = \mu_2$, y $Ho: \mu_2 = \mu_3$ la hipótesis $Ho: \mu_1 = \mu_3$ esta alternativa, sin duda razonable, tiene el inconveniente de incrementar el error tipo I.

Para probar la hipótesis de igualdad de k grupos $Ho: \mu_1 = \mu_2 = \mu_3$ se recurre al análisis de la varianza. Esta técnica agrupa una amplia gama de tópicos del análisis estadístico. Su desarrollo inicial se debe a R.A. Fisher. Debido a ello es que la estadística de prueba y la distribución asociada a ésta se designa como prueba F.

EL MODELO.

Es muy conveniente adoptar un modelo lineal para describir la estructura de los datos vinculada a k grupos independientes determinados por los que se llamarán k tratamientos.

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$ Para i=1,2…k j=1,2,…n_i

Y_{ij} Es el valor de la observación j del grupo i

μ Es la media global de la población.

τ_i Es el efecto del tratamiento i

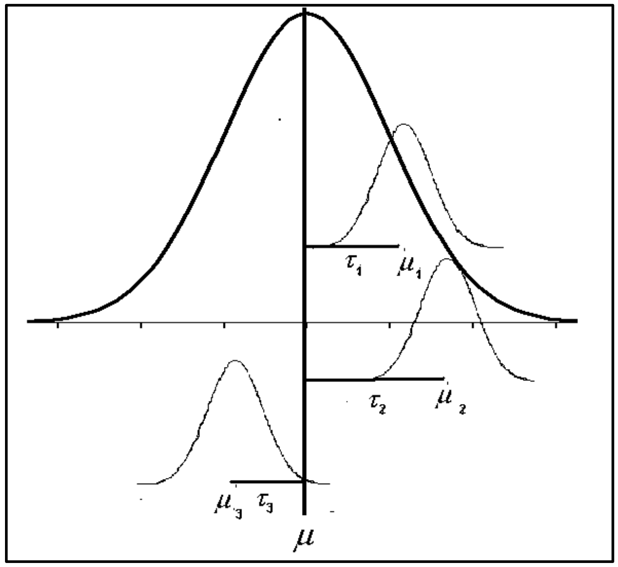
ε_{ij} Es el error aleatorio normalmente distribuido, independiente y con la misma varianza para los k tratamientos

$$\varepsilon_{ij} \rightarrow N\left(0, \sigma_\varepsilon^2\right)$$

Los tratamientos se supone que han sido determinados específicamente por el investigador, esto es, no se seleccionaron al azar. A este modelo se le identifica por ello como Modelo de Efectos Fijos o Modelo I. La intención es probar la igualdad de medias de los k tratamientos y no se puede hacer extensiva a tratamientos no considerados. Se pretenderá estimar los parámetros

$$\left(\mu, \tau_i . \sigma_\varepsilon^2\right)$$

El siguiente diagrama ilustra las relaciones entre la media de la superpoblación, los efectos de tratamientos y las medias por tratamiento.



En este modelo los efectos de tratamiento τ_i se definen como desviaciones respecto de la media de la superpoblación y se adopta la siguiente restricción cuya justificación se presenta adelante.

$$\sum_{i=1}^{k} \tau_i = 0$$

DESCOMPOSICIÓN DE LAS SUMA DE CUADRADOS TOTAL.

En primer término, se establece la estructura de datos en muestra que conviene considerar para que todos los conceptos queden adecuadamente definidos. Se consideran K muestras de tamaños ni no necesariamente iguales.

$$\begin{array}{cccc} G_1 & G_2 & \ldots\ldots & G_K \\ Y_{11} & Y_{12} & \ldots & Y_{1K} \\ \ldots. & \ldots. & \ldots. & \ldots. \\ Y_{1n_1} & Y_{1n_2} & \ldots. & Y_{Kn_k} \end{array}$$

Se definen las medias por grupo y la media general de la siguiente forma:

$$Y_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \qquad n = \sum_{i=1}^{k} n_i \qquad Y_{...} = \frac{1}{n}\sum_{i=1}^{K}\ \sum_{j=1}^{n_i} Y_{ij}$$

La variabilidad total de la muestra se resume en la suma de cuadrados de desviaciones de las observaciones respecto de la media general.

$$SCT = \sum_{i=1}^{K}\ \sum_{j=1}^{n_i}\left(Y_{ij} - Y_{..}\right)^2$$

Con la introducción de la suma y resta de las medias por grupo dentro del paréntesis de la expresión anterior se logra una forma equivalente.

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

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

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

Pero se puede verificar fácilmente que el doble producto se anula. Entonces la suma de cuadrados total SCT se descompone en las dos sumas de cuadrados que corresponden a la suma de cuadrados dentro de grupos (tratamientos) y la suma de cuadrados entre tratamientos.

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

Una expresión alternativa de la suma de cuadrados entre grupos que resulta útil para efectos de cálculo es la siguiente

$$\sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 = n_i \sum_{j=1}^{n_i} \left(\bar{Y}_{ii.} - \bar{Y}_{..}\right)^2$$

Los grados de libertad asociados a estas sumas de cuadrados se pueden entender en la forma siguiente. En el cálculo de la suma de cuadrados totales intervienen las n observaciones y se requiere el cálculo de la media general para construir las diferencias, ello provoca que se pierda un grado de libertad. En consecuencia la SCT tiene n-1 grados de libertad.

En el cálculo de la suma de cuadrados dentro de grupos intervienen n observaciones, pero para calcular las diferencias, antes se deben calcular las K medias correspondientes a los grupos y por ello pierde K grados de libertad. Como consecuencia la SCDentro tiene n-K grados de libertad.

Finalmente, en la suma de cuadrados entre de grupos intervienen las K medias y la media general que hace perder un grado de libertad a las K observaciones correspondientes a las medias y entonces se tienen K-1 grados de libertad asociados a la SCEntre.

La estadística de prueba F se calcula como el cociente de la suma de cuadrados medios entre grupos dividida entre la suma de cuadrados medios dentro de grupos. Ambas están asociadas a distribuciones Ji cuadrada y su cociente a una distribución F de Fisher. Cuyos grados de libertad son heredados de ambas sumas de cuadrados.

Las sumas de cuadrados medios entre grupos y dentro de grupos se obtienen al dividir las sumas de cuadrados entre sus respectivos grados de libertad.

Ambas sumas de cuadrados medios son estimadores de la varianza. Pero la suma de cuadrados medios dentro de grupos es un estimador insesgado, en tanto que la suma de cuadrados medios entre grupos tiene una componente de sesgo positivo que se anula solamente que todos los efectos de tratamiento τ_i sean nulos. Esto es equivalente a que la hipótesis de igualdad de medias se cumpla estrictamente $Ho: \mu_1 = \mu_2 = \ldots = \mu_k$ A medida que la hipótesis no se cumpla, el sesgo se incrementa y el cociente resulta mayor

$$E(ECMDentro) = \sigma^2 \qquad E(ECMEntre) = \sigma^2 + \frac{n\sum_{i=1}^{k}\tau_i^2}{k-1}$$

Al construir la estadística de prueba como el cociente de ambas sumas de cuadrados medios, se espera que el cociente sea mayor o igual a 1. En la medida que el cociente sea mayor de 1 quiere decir que hay mayor evidencia para rechazar la hipótesis de igualdad de medias. La estadística de prueba se distribuye F de Fisher con K-1 y n-K grados de libertad

$$F = \frac{(SCMEntre)}{(SCMDentro)}$$

TABLA DE ANOVA1

Los cálculos correspondientes se presentan en una Tabla de Análisis de la Varianza con objeto de que el investigador pueda valorar la forma como se comportan sus datos y la parte que aportan a la varianza sus orígenes.

Fuente de variación	Sumas de Cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre Grupos	$\sum_{i=1}^{K}\sum_{j=1}^{n_i}(Y_{i.} - Y_{..})^2$	k-1	$\sum_{i=1}^{K}\sum_{j=1}^{n_i}(Y_{i.} - Y_{..})^2 / (k-1)$ (a)	(a)/(b)
Dentro de Grupos	$\sum_{i=1}^{K}\sum_{j=1}^{n_i}(Y_{ij} - Y_{.i.})^2$	n - k	$\sum_{i=1}^{K}\sum_{j=1}^{n_i}(Y_{ij} - Y_{.i.})^2 / (n-k)$ (b)	
Total	$\sum_{i=1}^{K}\sum_{j=1}^{n_i}(Y_{ij} - Y_{..})^2$	n - 1		

ESTIMADORES DE MÍNIMOS CUADRADOS.

Supóngase que se tienen n observaciones y que se desea contar con estimaciones de los k+1 parámetros $\mu, \tau_1, \tau_2, \ldots, \tau_k$ mediante mínimos cuadrados.

$$Y_{ij} = \hat{\mu} + \hat{\tau}_i + e_{ij}$$

Se parte de la suma de cuadrados de los residuales.

$$\sum_{i=1}^{K} \sum_{j=1}^{n_i} e_i^{\,2} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(Y_{ij} - \hat{\mu} - \hat{\tau}_i\right)^2$$

La suma de cuadrados de los residuales, que se deriva respecto de los parámetros de interés y se igualan a cero.

$$\frac{d}{d\hat{\mu}} \sum_{i=1}^{K} \sum_{j=1}^{n_i} e_i^{\,2} = -2 \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(Y_{ij} - \hat{\mu} - \hat{\tau}_i\right) = 0$$

$$\frac{d}{d\tau_i} \sum_{i=1}^{K} \sum_{j=1}^{n_i} e_i^{\,2} = -2 \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(Y_{ij} - \hat{\mu} - \hat{\tau}_i\right) = 0$$

De donde se obtienen las ecuaciones:

$$Y_{..} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} Y_{ij} = n\hat{\mu} + \sum_{i=1}^{K} n_i \hat{\tau}_i$$

$$Y_{i.} = \sum_{j=1}^{n_i} Y_{ij} = n_i \hat{\mu} + \sum_{j=1}^{n_i} \hat{\tau}_i$$

Estas k+1 ecuaciones en forma desagregada se presentan a continuación.

$$Y_{..} = n\hat{\mu} + n_1\hat{\tau}_1 + n_2\hat{\tau}_2 + \ldots\ldots\ldots + n_k\hat{\tau}_k$$

$$Y_{1.} = n_1\hat{\mu} + n_1\hat{\tau}_1$$

$$Y_{2.} = n_2\hat{\mu} + n_2\hat{\tau}_2$$

……………………………………

$$Y_{k.} = n_k\hat{\mu} + n_k\hat{\tau}_k$$

Observe que si se suman las ecuaciones, miembro a miembro a partir de la segunda, se obtiene la primera. Por lo tanto no son independientes y en consecuencia no se tiene una solución única para $\mu, \tau_1, \tau_2, \ldots, \tau_k$ La solución es adoptar la restricción.

$$\sum_{i=1}^{k} \tau_i = 0$$

Entonces se llega a los estimadores:

$$\hat{\mu} = \bar{Y}_{..} \qquad\qquad \hat{\tau}_i = \bar{Y}_{i.} - \bar{Y}_{..}$$

Ejemplo. Un grupo de 19 cerdos fueron asignados al azar, poco tiempo después de su nacimiento a 4 dietas diferentes. Después de 5 meses fueron pesados. Se pretende probar la igualdad de las medias de los pesos en kilogramos alcanzados por los cerdos con las 4 dietas.

$$Ho: \mu_1 = \mu_2 = \mu_3 = \mu_4 \qquad H_1: \mu_i \neq \mu_j \text{ } para\ cualquier\ i \neq j$$

DIETA 1	DIETA 2	DIETA 3	DIETA 4
60.7	68.6	102.5	87.8
56.8	67.6	101.9	84.1
64.9	73.8	100	82.9
58.5	66.1	96.3	85.5
61.6	69.6		90.1

Se calculan las estadísticas básicas de las 4 dietas.

Estadística	DIETA 1	DIETA 2	DIETA 3	DIETA 4
ni	5	5	4	5
Media por Grupo	60.50	69.14	100.18	86.08
Media Global	77.86			
Desv Est	3.09	2.91	2.79	2.89

La tabla anterior muestra una diferencia de casi 40 kg. en el promedio de los cerdos de la dieta 1 y la dieta 3 que corresponden a los valores mínimo y máximo de las medias observadas. Intuitivamente se puede afirmar que hay una clara diferencia entre las 4 dietas.

A continuación, se procede al cálculo de la tabla de análisis de la varianza. El elevado valor de la estadística F=163.378 y su probabilidad asociada, permiten confirmar holgadamente que la hipótesis de igualdad de medias se rechaza. La aportación de la variación entre grupos es abrumadora 4,216.718. Por otra parte los valores observados dentro de cada grupo presentan homogeneidad y por tanto la aportación de la suma de cuadrados dentro de grupos es mínima (129.0475)

Fuente de variación	Sumas de Cuadrados	Grados de Libertad	Cuadrados Medios	F	Probabilidad
Entre Grupos	4,216.7188	3	1,405.5729	163.3786	0.000000
Dentro de Grupos	129.0475	15	8.6032		
Total	4,345.7663	18			

PRUEBA DE LEVENE PARA HOMOGENEIDAD DE VARIANZAS

Para verificar el supuesto de homogeneidad de varianzas, también conocida como homoscedasticidad, existen varias pruebas, una de las más conocidas es la prueba de Bartlett, sin embargo, esta prueba suele ser más sensible a la falta de normalidad. La prueba que se utiliza con mayor frecuencia en la actualidad es la prueba de Howard Levene (1960).

La hipótesis nula que se prueba: $Ho: \sigma_1^2 = \sigma_2^2 = \ldots\ldots = \sigma_k^2 \ vs \ H_1: \sigma_1^2 \neq \sigma_2^2 \ para\ i \neq j$

La estadística de prueba

$$W = \frac{(n-k)}{(k-1)} \frac{\sum_{i=1}^{k} n_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} (Z_{iji} - Z_{i.})^2}$$

.

Donde

$$Z_{ij} = |Y_{ij} - Y_{i.}| \qquad Z_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij} \qquad Z_{..} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij}$$

La estadística W se distribuye como F con k-1 y n-k grados de libertad. Existe una versión modificada de la prueba que utiliza la mediana en lugar de la media y que se conoce como Brown y Forsythe

A continuación se aplica la prueba a los datos del ejemplo 1 las columnas permiten el cálculo ordenado

PRUEBA DE LEVENE PARA HOMOGENEIDAD DE VARIANZAS

Grupo	Y_{ij}	$\bar{Y}_{i.}$	Z_{ij}	$\bar{Z}_{i.}$	ni	$n_i(\bar{Z}_{i.}-\bar{Z}_{..})^2$	$(Z_{iji}-\bar{Z}_{i.})^2$
1	60.7		0.20				4.32640
1	56.8		3.70				2.01640
1	64.9		4.40				4.49440
1	58.5		2.00				0.07840
1	61.6	60.5	1.10	2.28	5	0.061080	1.39240
2	68.6		0.54				2.27406
2	67.6		1.54				0.25806
2	73.8		4.66				6.82254
2	66.1		3.04				0.98406
2	69.6	69.14	0.46	2.05	5	0.073779	2.52174
3	102.5		2.33				0.09000
3	101.9		1.73				0.09000
3	100.0		0.17				3.42250
3	96.3	100.2	3.88	2.03	4	0.083491	3.42250
4	87.8		1.72				0.33178
4	84.1		1.98				0.09986
4	82.9		3.18				0.78146
4	85.5		0.58				2.94466
4	90.1	86.08	4.02	2.30	5	0.080045	2.97218
		77.86	2.17	2.17	19	0.298395	39.3234

	Muestra	Grados Lib
K =	4	3
N=	19	15

Estadística de Prueba	
W =	0.03794112
Prob F (3,15)	0.98971314

El valor de W = 0.0379412, da lugar a la probabilidad asociada a una distribución F con 3 y 15 grados de libertad P= 0.9897314. Esto es la probabilidad de igualdad de varianzas es muy alta, por tanto la hipótesis nula de homoscedasticidad no se rechaza.

COMPARACIONES MÚLTIPLES MEDIANTE PRUEBA DE SHEFFEÉ

La hipótesis que se prueba con el modelo ANOVA1 es $Ho: \mu_1 = \mu_2 = \mu_3$ El rechazo de esta hipótesis no implica que las K medias sean diferentes, sino que al menos una es diferente. El investigador obviamente deseará saber cuales grupos motivaron el rechazo de la hipótesis. En primera instancia el investigador pretendería efectuar pruebas comparando las diferentes combinaciones 2 a 2 de los K grupos mediante pruebas de t para dos muestras independientes, pero este procedimiento es inadecuado, porque eleva artificialmente la probabilidad del error tipo I. En lugar de ello se debe aplicar alguno de los métodos conocidos como de Comparaciones Múltiples, Entre éstos se pueden mencionar Tukey, Duncan, Dunnett, Sheffeé, etc. En nuestro caso utilizaremos la prueba de Sheffeé, por ser una prueba conservadora y que brinda excelentes resultados.

Usualmente el procedimiento de la prueba consiste en ordenar las medias aritméticas de menor a mayor e iniciar por el contraste de las medias extremas. Si el par extremo se rechaza, entonces se procede a contrastar las extremas restantes.

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

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

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

$$E_S = \frac{\bar{Y}_A - \bar{Y}_B}{EE(E_S)} \qquad\qquad CMDG = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i.}\right)^2 /(n-k)$$

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.

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

El discriminante de la prueba (umbral de Sheffeé) se obtiene mediante la raíz cuadrada del número total de grupos menos 1, multiplicada por el percentil de la distribución F correspondiente al nivel de significancia (α) de la prueba con k-1 y n-k grados de libertad.

$$\sqrt{(k-1)F(\alpha, k-1, n-k)}$$

A continuación se presentan los contrastes para los grupos del ejemplo de las dietas. En todos los casos la igualdad de medias se rechazó, pues la E_s superó el umbral.

CM Dentro de G = 8.6031667

Grupo A	Grupo B	Media A	Media B	Diferencia M1-M2	n_A	n_B	Error Est. Diferencia	Estadística Sheffé	$\sqrt{(k-1)F(\alpha,k-1,n-k)}$ Umbral de Sheffé	Conclusión
3	1	100.175	60.500	39.675	4	5	1.967593708	20.1642	3.1404	Rechazo
3	2	100.175	69.140	31.035	4	5	1.967593708	15.7731	3.1404	Rechazo
3	4	100.175	86.080	14.095	4	5	1.967593708	7.1636	3.1404	Rechazo
4	1	86.080	60.500	25.580	5	5	1.855065138	13.7893	3.1404	Rechazo
4	2	86.080	69.140	16.940	5	5	1.855065138	9.1318	3.1404	Rechazo
2	1	69.140	60.500	8.640	5	5	1.855065138	4.6575	3.1404	Rechazo

Intervalos de confianza para las diferencias de medias en prueba de Sheffeé.
Se obtiene al restar y sumar el error estándar de la diferencia, multiplicado por el umbral.

$$\left(Y_A - Y_B\right) \pm EE(E_s)\sqrt{(k-1)F(\alpha, k-1, n-k)}$$

Grupo A	Grupo B	Diferencia M1-M2	Error Est. Diferencia	$\sqrt{(k-1)F(\alpha,k-1,n-k)}$ Umbral de Sheffé	$EE\sqrt{(k-1)F(\alpha,k-1,n-k)}$	Límite Inferior	Límite Superior
3	1	39.675	1.9675937	3.140	6.1790	33.4960	45.8540
3	2	31.035	1.9675937	3.140	6.1790	24.8560	37.2140
3	4	14.095	1.9675937	3.140	6.1790	7.9160	20.2740
4	1	25.580	1.8550651	3.140	5.8257	19.7543	31.4057
4	2	16.940	1.8550651	3.140	5.8257	11.1143	22.7657
2	1	8.640	1.8550651	3.140	5.8257	2.8143	14.4657

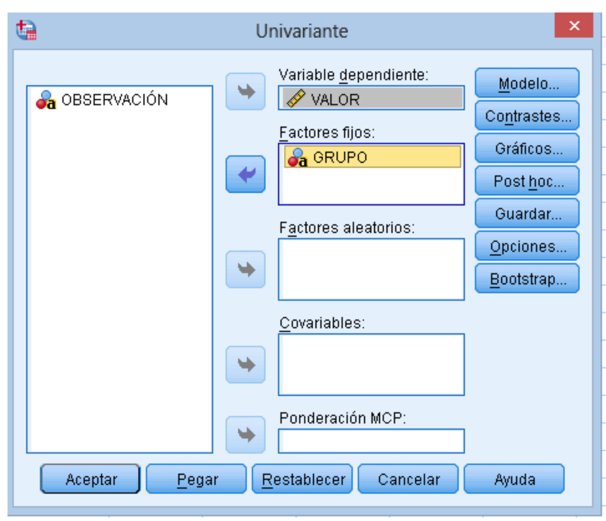
SOLUCIÓN MEDIANTE SPSS

Estructura de los Datos.

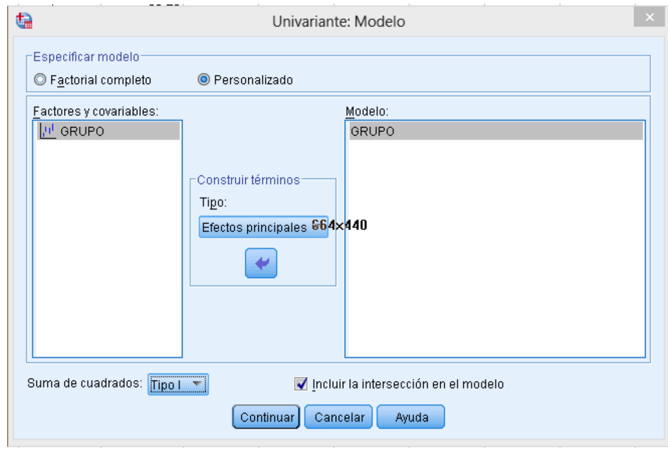
Los datos se presentan en tres columnas: la columna que identifica a la observación la columna que identifica al grupo y el valor de la observación:

	OBSERVACIÓN	GRUPO	VALOR
1	1	1	60.70
2	2	1	56.80
3	3	1	64.90
4	4	1	58.50
5	5	1	61.60
6	6	2	68.60
7	7	2	67.60
8	8	2	73.80
9	9	2	66.10
10	10	2	69.60
11	11	3	102.50
12	12	3	101.90
13	13	3	100.00
14	14	3	96.30
15	15	4	87.80
16	16	4	84.10
17	17	4	82.90
18	18	4	85.50
19	19	4	90.10

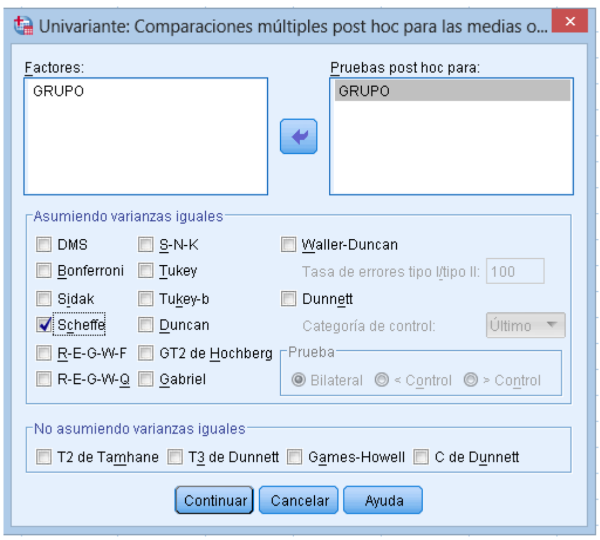
A continuación se procede a la opción Analizar y se selecciona Modelo Lineal General y dentro de éste Univariante. La variable VALOR se traslada a la celda de variable dependiente. El GRUPO se traslada a la celda de factores fijos. Posteriormente se selecciona la pestaña Modelo. Dentro de la ventana que se presenta, se selecciona personalizar. En Tipo se selecciona Efectos principales. En Modelo se asigna GRUPO. En Suma de cuadrados se selecciona la opción Tipo I (Efectos Fijos) y se conserva la opción de inclusión del cálculo de la intersección. Se oprime el botón de Continuar.

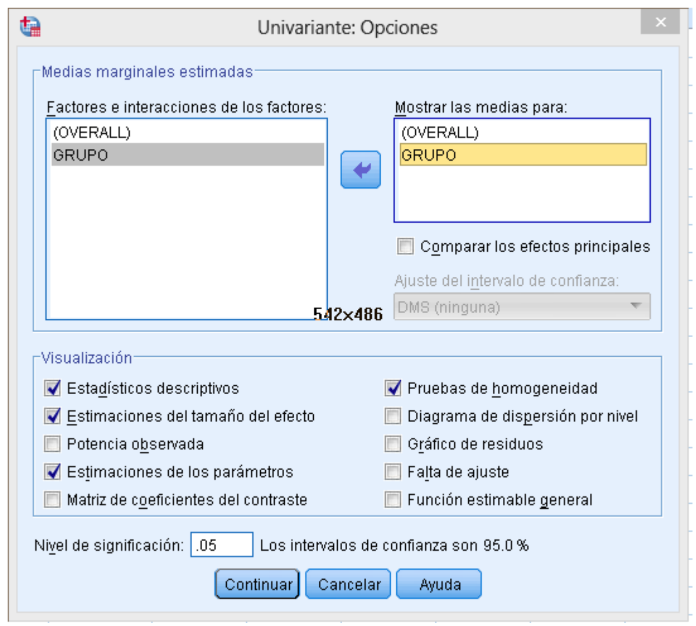


Se seleccionan opciones complementarias. En guardar se selecciona Valores Pronosticados No tipificados y su error típico (estándar). Posteriormente se selecciona la opción Post Hoc que corresponde a pruebas de comparación posteriores. Aquí se selecciona Scheffeé.



En el botón de opciones se marcan las estadísticas básicas, estimaciones de parámetros y pruebas de homogeneidad de varianzas (Levene).





Los reportes emitidos por SPSS son los siguientes. Compare las estadísticas de estos reportes con los calculados manualmente.

Estadísticos descriptivos
Variable dependiente: VALOR

GRUPO	Media	Desviación típica	N
1	60.5000	3.09435	5
2	69.1400	2.90826	5
3	100.1750	2.79449	4
4	86.0800	2.89344	5
Total	77.8579	15.53806	19

Pruebas de los efectos inter-sujetos

Variable dependiente: VALOR

Origen	Suma de cuadrados tipo I	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
Modelo corregido	4216.719[a]	3	1405.573	163.379	.000000	.970305
Intersección	**115175.184**	**1**	**115175.184**	**13387.534**	**.000000**	**.998881**
GRUPO	**4216.719**	**3**	**1405.573**	**163.379**	**.000000**	**.970305**
Error	**129.048**	**15**	**8.603**			
Total	**119520.950**	**19**				
Total corregida	4345.766	18				

a. R cuadrado = .970 (R cuadrado corregida = .964)

2. GRUPO

Variable dependiente: VALOR

GRUPO	Media	Error típ.	Intervalo de confianza 95%	
			Límite inferior	Límite superior
1	60.500	1.312	57.704	63.296
2	69.140	1.312	66.344	71.936
3	100.175	1.467	97.049	103.301
4	86.080	1.312	83.284	88.876

Contraste de Levene sobre la igualdad de las varianzas err

Variable dependiente: VALOR

F	gl1	gl2	Sig.
.038	**3**	**15**	**.990**

Contrasta la hipótesis nula de que la varianza error de la va

a. Diseño: Intersección + GRUPO

Comparaciones múltiples

Variable dependiente: VALOR Scheffe

(I)GRUPO		Diferencia de medias (I-J)	Error típ.	Sig.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
1	2	-8.6400*	1.85507	.003	-14.4657	-2.8143
	3	-39.6750*	1.96759	.000	-45.8540	-33.4960
	4	-25.5800*	1.85507	.000	-31.4057	-19.7543
2	1	8.6400*	1.85507	.003	2.8143	14.4657
	3	-31.0350*	1.96759	.000	-37.2140	-24.8560
	4	-16.9400*	1.85507	.000	-22.7657	-11.1143
3	1	39.6750*	1.96759	.000	33.4960	45.8540
	2	31.0350*	1.96759	.000	24.8560	37.2140
	4	14.0950*	1.96759	.000	7.9160	20.2740
4	1	25.5800*	1.85507	.000	19.7543	31.4057
	2	16.9400*	1.85507	.000	11.1143	22.7657
	3	-14.0950*	1.96759	.000	-20.2740	-7.9160

Basadas en las medias observadas. El término de error es la media cuadrática(Error) = 8.603.

*. La diferencia de medias es significativa al nivel .05.

Ejemplo 2.

Un ingeniero de desarrollo de productos está interesado en maximizar la resistencia a la tensión de una fibra sintética que se empleará en la manufactura de tela para camisas de hombre. El ingeniero sabe que la resistencia esta, en función del porcentaje de algodón que tiene la fibra. Y sospecha qué si incrementa el algodón, se incrementa la resistencia. Y decide probar muestras a cinco niveles de porcentaje de algodón: 15%, 20%, 25%, 30% y 35%.
Tenemos un experimento con un criterio de clasificación de 5 niveles del factor y 5 repeticiones.

Resistencia a la tensión observada

Num	**Porcentaje de algodón (lb/plg2)**				
	15	**20**	**25**	**30**	**35**
1	7	12	14	19	7
2	7	17	18	25	10
3	15	12	18	22	11
4	11	18	19	19	15
5	9	18	19	23	11