SOLUCIONES A EJERCICIOS DE ADEVA SOBRE DISEÑOS EQUILIBRADOS DE DOS FACTORES CON INTERACCIONES

DE LOS MODELOS FIJO, ALEATORIO Y MIXTO

CON INDICACIONES SOBRE LA APLICACIÓN DEL PROGRAMA STATGRAPHICS CENTURIÓN A SU RESOLUCIÓN Y AL ANÁLISIS POSTERIOR AL ADEVA

DISEÑO DE EXPERIMENTOS CON DOS FACTORES

Este diseño sirve para analizar de forma simultánea los efectos que pueden producir los diferentes niveles de dos factores sobre una magnitud que se estudia en los individuos de una población y las interacciones que pueden darse entre los niveles de esos dos factores.

Variable de respuesta:
Representa el conjunto de valores que puede tomar una magnitud en los individuos de una población.

Factor:
Es cualquier condición o situación que pueda influir en la variable de respuesta. Llamaremos A y B a los dos factores.

Tratamiento:
Es cada una de las combinaciones entre un nivel de un factor y un nivel del otro que se analizan en el diseño.

Nivel de un factor:
Es cada una de las distintas facetas, formas, grados o valores que puede presentar ese factor en el diseño. Cada nivel puede producir un efecto distinto sobre la variable de respuesta.

En el diseño de experimentos se toma una muestra de *n* individuos de la población, se divide al azar en tantos grupos como tratamientos se hayan formado, y cada grupo se somete a un tratamiento distinto.

Unidad muestral: *Es cada individuo sobre el que se realiza el experimento.*

Réplicas de un tratamiento: *son las unidades muestrales que reciben ese tratamiento*.

Notaremos I al número de niveles del factor A (fila); J al número de niveles del factor B (columna) y K al número de réplicas de cada tratamiento.

EJEMPLO I: Variable de respuesta: longitudes de los tallos de plantas herbáceas (en cm.)

NIVELES de A	NIVELES DE B: Grados de humedad			
Suelos	10%	20%	25%	30%
arenoso	20 15 10	25 14 15	17 20 23	15 14 18
calcáreo	20 25 18	25 24 15	22 25 28	18 13 17
arcilloso	22 25 16	22 31 19	28 30 23	20 14 16

En este ejemplo , el tipo de suelo es el factor fila, con 3 niveles, ***(I = 3)*** *y el grado de humedad es el factor columna, con 4 niveles* ***(J = 4).*** *En el diseño se tomaron 36 plantas* ***(n=36)*** *y se repartieron en 12 grupos* ***(3 x 4 = 12*** *tratamientos* ***)*** *con 3 réplicas en cada uno* ***(K = 3) .***

EJEMPLO II: Variable de respuesta: Aumento de peso, en gr. de las colonias de algas sometidas a los distintos tratamientos

NIVELES de A	NIVELES DE B: Salinidad			
Temperatura	1	2	3	4
15^0	16 14	20 18	13 12	11 10
20^0	8 11	17 14	15 12	10 9
25^0	7 7	17 18	14 16	6 10
30^0	1 3	16 18	8 6	3 5

En este ejemplo , la temperatura es el factor fila, con 4 niveles, ***(I = 4)*** *y la salinidad es el factor columna, con 4 niveles* ***(J = 4).*** *En el diseño se tomaron 36 plantas* ***(n=32)*** *y se repartieron en 16 grupos* ***(4 x 4 = 16*** *tratamientos* ***)*** *con 2 réplicas en cada uno* ***(K = 2) .***

ADEVA DE DOS CRITERIOS PARA EL DISEÑO FACTORIAL CON INTERACCIONES en un Diseño factorial equilibrado

EJEMPLO I

Grados de humedad

Suelos	10%	20%	25%	30%
arenoso	20 15 10	25 14 15	17 20 23	15 14 18
calcáreo	20 25 18	25 24 15	22 25 28	18 13 17
arcilloso	22 25 16	22 31 19	28 30 23	20 14 16

Para introducir estos datos, llamaremos SUELO a la columna del primer factor, donde escribiremos las letras a (arenoso), **c** (calcáreo) **y r** (arcilloso) **doce veces cada una.**

Llamaremos HUMED a la columna del segundo factor, donde escribiremos los números 1 (10%), **2 (**20%), **3 (**25%) **y 4** (30%), **tres veces cada uno, a la derecha de cada letra de la columna suelo.**

Llamaremos TALLO a la columna de la variable de respuesta, donde escribiremos cada dato de la tabla en la fila donde están los dos niveles que le corresponden.

	SUELO	HUMED	TALLO
1	a	1	20
2	a	1	15
3	a	1	10
4	a	2	25
5	a	2	14
6	a	2	15
7	a	3	17
8	a	3	20
9	a	3	23
10	a	4	15
11	a	4	14
12	a	4	18
13	c	1	20
14	c	1	25
15	c	1	18
16	c	2	25
17	c	2	24
18	c	2	15
19	c	3	22
20	c	3	25
21	c	3	28
22	c	4	18
23	c	4	13
24	c	4	17
25	r	1	22
26	r	1	25
27	r	1	16
28	r	2	22
29	r	2	31
30	r	2	19
31	r	3	28
32	r	3	30
33	r	3	23
34	r	4	20
35	r	4	14
36	r	4	16

ANOVA DEL DISEÑO CRUZADO CON DOS FACTORES Y K RÉPLICAS
o ANOVA DE DOS CRITERIOS CON INTERACCIONES

El objetivo de este ANOVA *es utilizar los datos de un diseño cruzado para contrastar, a la vez, tres hipótesis nulas sobre los niveles de los dos factores:*

La ausencia de efectos de los niveles del primer factor.

La ausencia de efectos de los niveles del segundo factor.

La ausencia de interacciones de niveles del primer factor con niveles del segundo

En este diseño, las variables aleatorias que representan la variabilidad propia de las unidades muestrales se designan con la notación ε_{ijk}

SUPUESTOS DE VALIDEZ DEL ANOVA DE DOS CRITERIOS

1 – **INDEPENDENCIA:**
Las variables aleatorias ε_{ijk} ***son independientes entre sí.***

2– **NORMALIDAD:**
Todas estas variables aleatorias siguen leyes Normales $N(0, \sigma_e^2)$

3– **HOMOGENEIDAD DE LAS VARIANZAS:**
Todas estas leyes normales tienen la misma varianza, σ_e^2 ***que se llama "varianza de los errores" o "varianza residual"***

MODELOS DEL DISEÑO DE DOS FACTORES CON K RÉPLICAS.

Modelo I o de efectos FIJOS *(Los niveles de cada factor se eligen a voluntad.)*

Los efectos α_i *son parámetros desconocidos,* que verifican $\sum_{i=1}^{I} \alpha_i = 0$

Para cada índice i, el parámetro $\mu_{i\bullet} = \mu + \alpha_i$ se llama ***media del nivel*** A_i.

Los efectos β_j *son parámetros desconocidos,* que verifican $\sum_{j=1}^{J} \beta_j = 0$

Para cada índice j, el parámetro $\mu_{\bullet j} = \mu + \beta_j$ se llama ***media del nivel*** B_j.

Los efectos $(\alpha\beta)_{ij}$ *son parámetros desconocidos,* que verifican

$$\sum_{i=1}^{I} (\alpha\beta)_{ij} = 0 \quad \text{y} \quad \sum_{j=1}^{J} (\alpha\beta)_{ij} = 0$$

Para cada grupo *(i, j)*, el parámetro $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ se llama ***media del tratamiento*** $A_i B_j$.

Los objetivos del ADEVA en un modelo I son:

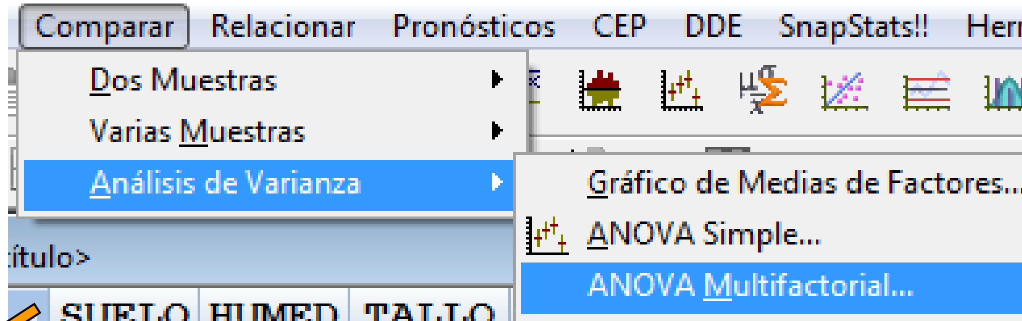
- *Decidir si existen interacciones entre los niveles de ambos factores.*
- *Decidir si las medias de los diferentes niveles de cada factor son distintas.*
- *Estimar los parámetros* $\mu_{i\bullet}$ *y* α_i ; $\mu_{\bullet j}$ *y* β_j ; μ_{ij} *y* $(\alpha\beta)_{ij}$

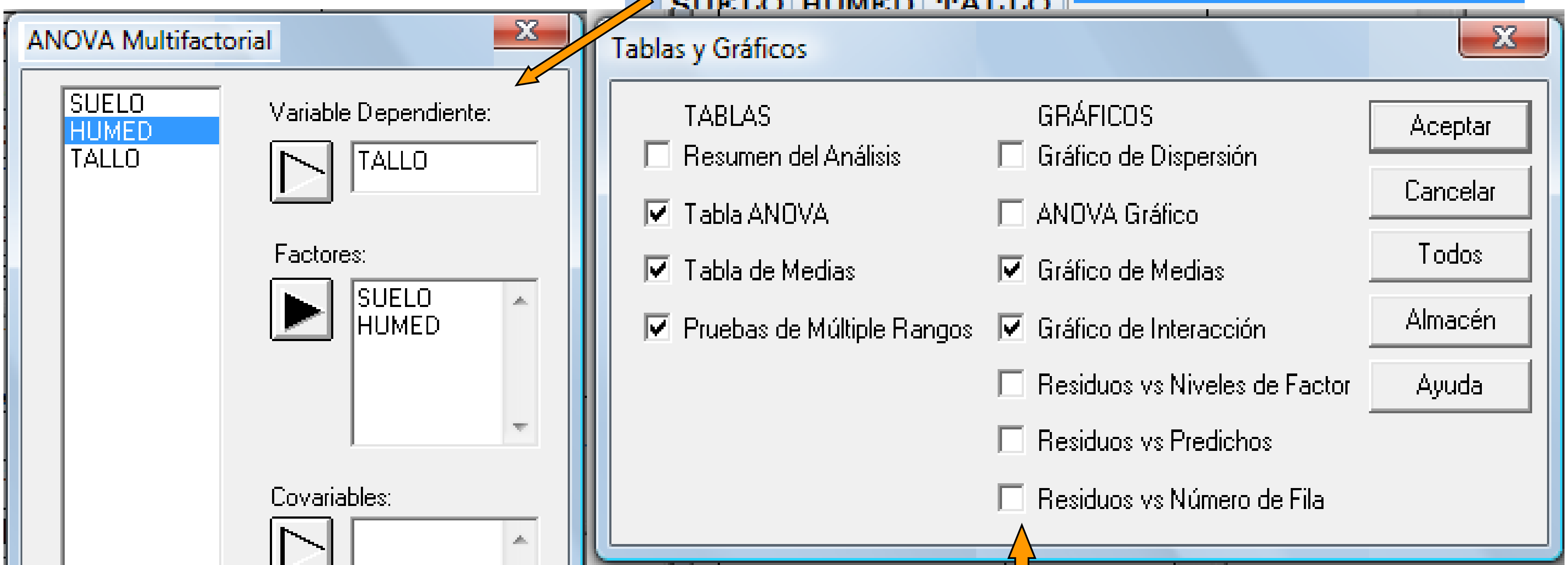
Hipótesis que se contrastan

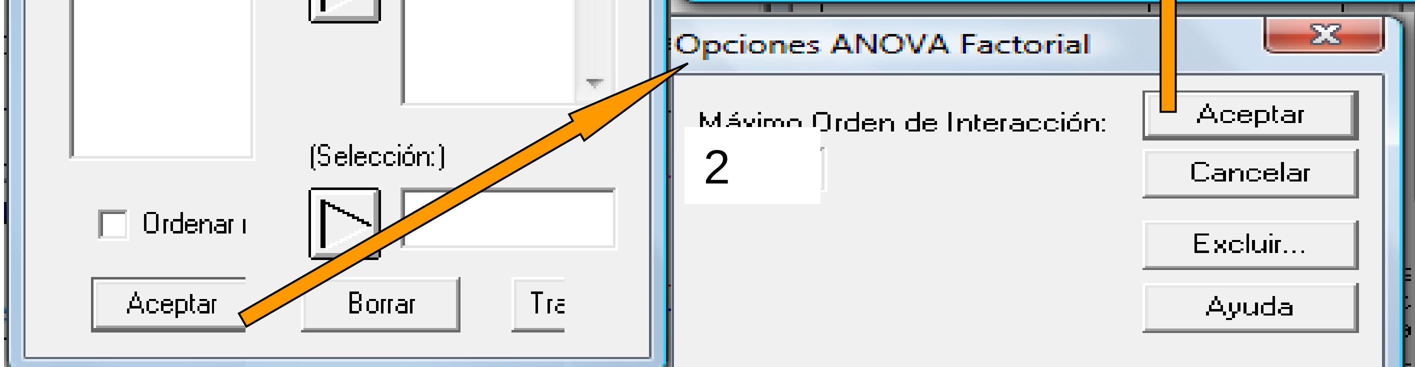
hipótesis nula.	hipótesis alternativa
H_0: *Las medias de los niveles* A_i *son iguales:* $\mu_{i.} = \mu \; \forall i$	H_1: $\exists\, \mu_{i.} \neq \mu$
H'_0: *Las medias de los niveles* B_j *son iguales:* $\mu_{.j} = \mu \; \forall j$	H'_1: $\exists\, \mu_{.j} \neq \mu$
H''_0: *No hay interacciones entre niveles de A y niveles de B:* $(\alpha\beta)_{ij} = 0 \;\; \forall i, \forall j$	H''_1: $\exists\, (\alpha\beta)_{ij} \neq 0$

Supondremos que el ejemplo 1 corresponde a un diseño de efectos fijos.

Las instrucciones para que el programa realice el ANOVA son las siguientes:







En "opciones ANOVA" se escribe un 2 si se quiere contrastar la interacción, y1 si se supone que no existe interacción.

RESULTADOS DEL ANOVA ☑ Tabla ANOVA

TABLA DEL ANOVA Y P- VALOR DE CADA UNO DE LOS CONTRASTES

Análisis de Varianza para TALLO - Suma de Cuadrados Tipo III

Fuente	*Suma de Cuadrados*	*Gl*	*Cuadrado Medio*	*Razón-F*	*Valor-P*
EFECTOS PRINCIPALES					
A:SUELO	160,889 ***SCA***	***I -1*** 2	80,4444 ***MCA***	4,45 $\frac{MCA}{MCD}$	0,0227
B:HUMED	300,111 ***SCB***	***J -1*** 3	100,037 ***MCB***	5,53 $\frac{MCB}{MCD}$	0,0049
INTERACCIONES					
AB	44,8889 ***SCI***	***(I-1)(J-1)*** 6	7,48148 ***MCI***	0,41 $\frac{MCI}{MCD}$	0,8626
RESIDUOS	434,0 ***SCD***	***IJ(K-1)*** 24	18,0833 ***MCD***		
TOTAL (CORREGIDO)	939,889 ***SCT***	***IJK - 1*** 35			

Todas las razones-F se basan en el cuadrado medio del error residual ***MCD***

Los cocientes o razones F se basan todos en el MCD si el diseño es de efectos fijos

La interpretación de esta tabla es la siguiente:

Se rechaza la hipótesis nula de igualdad de las medias de los niveles del suelo a cualquier nivel de significación mayor que 0.0227.

Se rechaza la hipótesis nula de igualdad de las medias de los niveles del grado de humedad a cualquier nivel de significación mayor que 0.0049

Se acepta la hipótesis nula de que NO hay interacción entre los dos factores a cualquier nivel de significación inferior a 0.8626.

En particular, a un nivel de significación de 0.10 se rechazan las dos primeras hipótesis nulas y se acepta la tercera; los valores críticos de los tres contrastes a este nivel son, respectivamente,

$$F_{2,24;0.9} = 2.54, \qquad F_{3,24;0.9} = 2.33, \qquad F_{6,24;0.9} = 2.04$$

Interpretación de las decisiones de los contrastes en el modelo I:

a) *Si en el ADEVA no se rechaza ninguna hipótesis nula, entonces sólo deben estimarse la varianza de los errores,* σ_e^2 *, y la media general,* μ.

b) *Si se rechaza la hipótesis de falta de interacción,* se admite que existen interacciones entre sus distintos niveles, lo que indica una dependencia entre los dos factores, por lo que no tiene sentido estudiar por separado los efectos de cada factor.

En este caso deben estimarse:

- La varianza de los errores, σ_e^2
- Las medias de los tratamientos, μ_{ij}
- Los efectos de las interacciones, $(\alpha\beta)_{ij}$.

c) *Si se acepta la hipótesis de falta de interacción,* se acepta la independencia entre los dos factores, por lo que deben estimarse las medias y los efectos de los niveles de cada factor cuya hipótesis nula se haya rechazado.

Si se rechaza H_0*,* deben estimarse las medias de los niveles del factor A y compararlas mediante pruebas a posteriori para decidir entre cuáles existen diferencias significativas.

Si se rechaza H'_0*,* deben estimarse las medias de los niveles del factor B y compararlas mediante pruebas a posteriori para decidir entre cuáles existen diferencias significativas.

Las pruebas a posteriori que se utilizan para estos contrastes son análogas a las que se utilizan en el ADEVA doble con una réplica.

Prueba de mínima diferencia significativa de Fisher (LSD).
Prueba de Bonferroni. Prueba de Scheffé. Prueba de Tukey

CONCLUSIONES QUE SE DEDUCEN DE UN ADEVA SIGNIFICATIVO en el modelo de efectos fijos:

Como se han rechazado las hipótesis nulas de igualdad de las medias de los niveles de ambos factores, hemos de estimar estas medias y realizar pruebas a posteriori a un nivel de significación dado, (5%) para decidir qué medias de cada nivel son las que difieren entre sí al 5%.

En “Tabla de Medias” se encuentran las medias de los niveles de ambos factores y las medias de todos los tratamientos, junto con intervalos de confianza, cuyo nivel (95%) se obtiene en “Opciones de Ventana”.

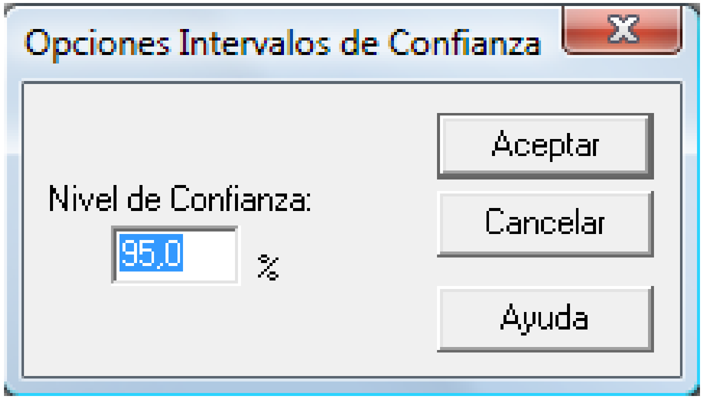
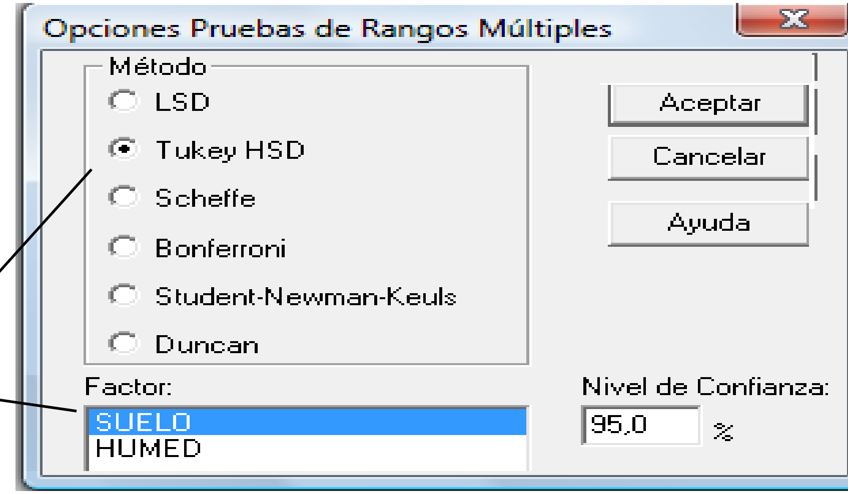


Tabla de Medias por Mínimos Cuadrados para TALLO con intervalos de confianza del 95,0%

			Error	*Límite*	*Límite*
Nivel	*Casos*	*Media*	*Est.*	*Inferior*	*Superior*
MEDIA GLOBAL	36	20,0556			
SUELO					
a	12	17,1667	1,22758	14,6331	19,7003
c	12	20,8333	1,22758	18,2997	23,3669
r	12	22,1667	1,22758	19,6331	24,7003
HUMED					
1	9	19,0	1,41748	16,0745	21,9255
2	9	21,1111	1,41748	18,1856	24,0367
3	9	24,0	1,41748	21,0745	26,9255
4	9	16,1111	1,41748	13,1856	19,0367
SUELO por HUMED					
a,1	3	15,0	2,45515	9,9328	20,0672
a,2	3	18,0	2,45515	12,9328	23,0672
a,3	3	20,0	2,45515	14,9328	25,0672
a,4	3	15,6667	2,45515	10,5995	20,7339
c,1	3	21,0	2,45515	15,9328	26,0672
c,2	3	21,3333	2,45515	16,2661	26,4005
c,3	3	25,0	2,45515	19,9328	30,0672
c,4	3	16,0	2,45515	10,9328	21,0672
r,1	3	21,0	2,45515	15,9328	26,0672
r,2	3	24,0	2,45515	18,9328	29,0672
r,3	3	27,0	2,45515	21,9328	32,0672
r,4	3	16,6667	2,45515	11,5995	21,7339

En “Pruebas de Múltiple Rangos” se pueden realizar varios contrastes a posteriori sobre la igualdad entre las medias de los niveles de un mismo factor, eligiendo tanto el factor como los contrastes en el cuadro adjunto.



Pruebas de Múltiple Rangos para TALLO por SUELO

Método: 95,0 porcentaje Tukey HSD

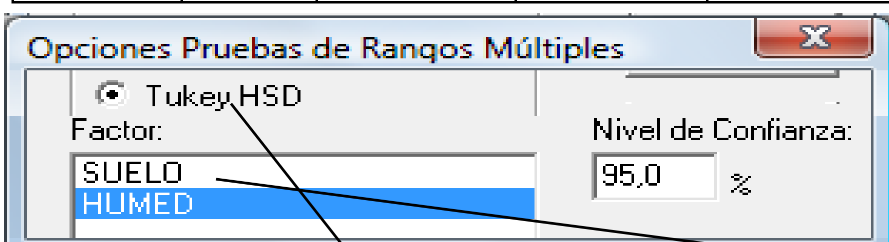
SUELO	*Casos*	*Media LS*	*Sigma LS*	*Grupos Homogéneos*
a	12	17,1667	1,22758	X
c	12	20,8333	1,22758	XX
r	12	22,1667	1,22758	X

Contraste	*Sig.*	*Diferencia*	*+/- Límites*
a - c		-3,66667	4,33681
a - r	*	-5,0	4,33681
c - r		-1,33333	4,33681

* indica una diferencia significativa.

Son diferentes las medias de los suelos a y r

Son diferentes las medias de los grados 3 y 4



Pruebas de Múltiple Rangos para TALLO por HUMED

Método: 95,0 porcentaje Tukey HSD

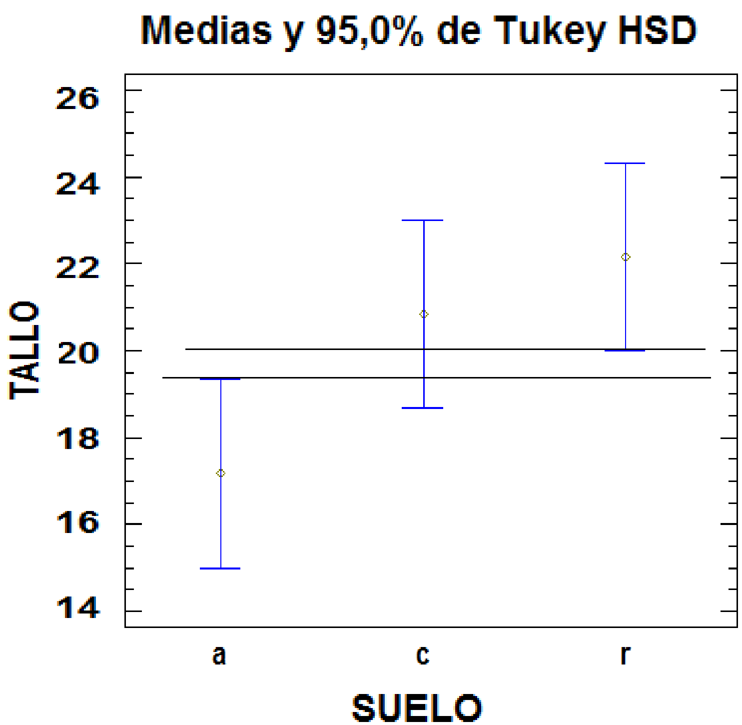
HUMED	*Casos*	*Media LS*	*Sigma LS*	*Grupos Homogéneos*
4	9	16,1111	1,41748	X
1	9	19,0	1,41748	XX
2	9	21,1111	1,41748	XX
3	9	24,0	1,41748	X

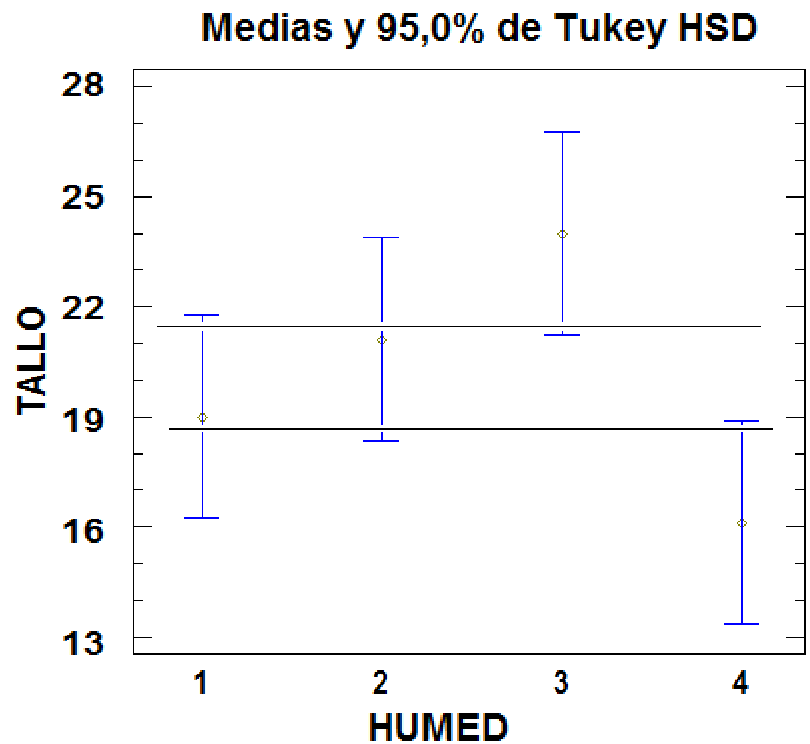
Contraste	*Sig.*	*Diferencia*	*+/- Límites*
1 - 2		-2,11111	5,53139
1 - 3		-5,0	5,53139
1 - 4		2,88889	5,53139
2 - 3		-2,88889	5,53139
2 - 4		5,0	5,53139
3 - 4	*	7,88889	5,53139

* indica una diferencia significativa.

Gráfico de Medias

En “Gráfico de medias” se representan las diferencias entre ellas mediante intervalos de confianza obtenidos por el mismo método y nivel con los que se ha realizado el contraste.
Si los intervalos de dos medias con el nivel de confianza dado se solapan, eso indica que esas dos medias no son diferentes al nivel de significación complementario del nivel de confianza.





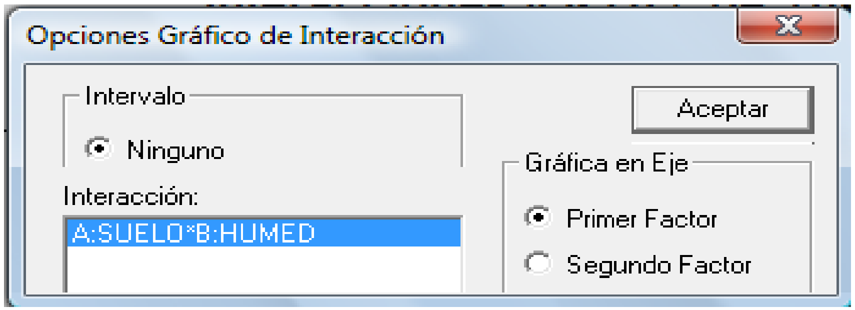
En este gráfico vemos como el intervalo de la media del suelo c se solapa con los intervalos de las otras dos, por tanto la media del suelo c no es distinta de ninguna de las otras dos al nivel del 5%, como dice la tabla del contraste.

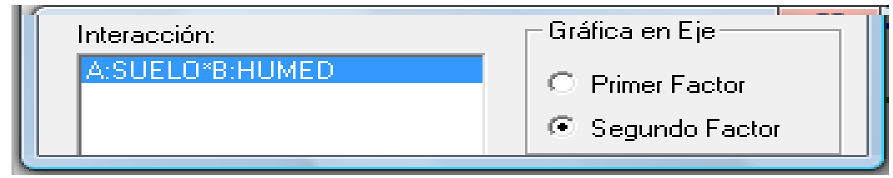
En este gráfico vemos como los únicos intervalos que no se solapan entre sí son el de la media del grado 3 y el de la media del grado 4, por tanto estas son las únicas medias que difieren entre sí al nivel del 5%, como indica la tabla del contraste.

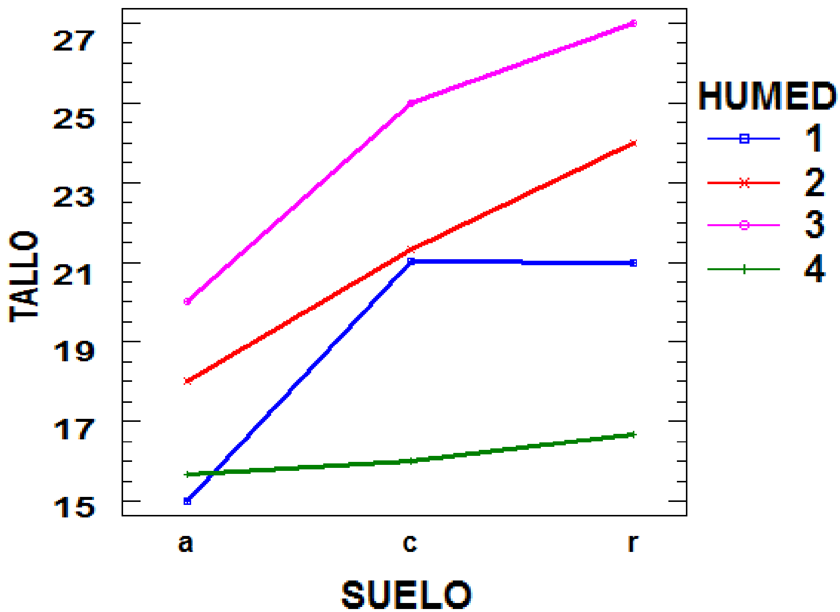
El factor que se sitúa en el eje de abscisas se elige en “opciones de ventana”

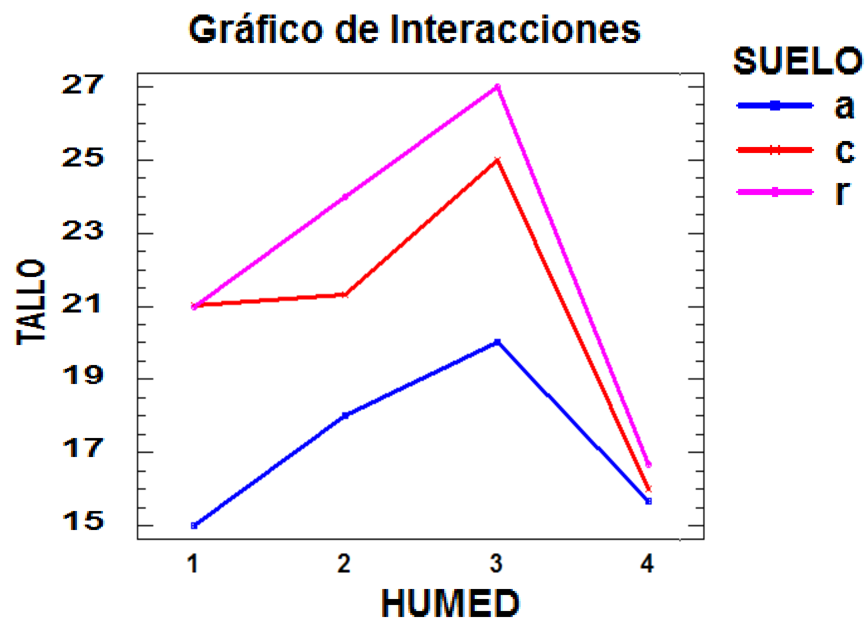
En este ANOVA se ha aceptado la hipótesis nula de que NO hay interacción entre los dos factores al nivel de significación del 5%, lo que quiere decir que los efectos de los niveles de cada factor no dependen del nivel del otro factor.

Esto se refleja en el gráfico de interacción:









En este gráfico, las líneas que representan los efectos de todos los grados de humedad suben siempre del suelo a al r, por tanto los efectos de la humedad no dependen de los tipos de suelo.

En este gráfico, las líneas que representan los efectos de los tipos de suelo suben del grado 1 al 3 y luego bajan al 4, pero todos se comportan igual, por tanto los tipos de suelo no dependen de los grados de humedad .

Modelo II o de efectos ALEATORIOS:

Los niveles de cada factor se eligen al azar, entre un amplio conjunto de posibles niveles.

Los efectos α_i son variables aleatorias independientes

siguen una misma ley normal $N(0,\sigma_A^2)$, donde σ_A^2 *es la varianza de los efectos del factor A.*

Los efectos β_j son variables aleatorias independientes

siguen una misma ley normal $N(0,\sigma_B^2)$, donde σ_B^2 *es la varianza de los efectos del factor B.*

Los efectos $(\alpha\beta)_{ij}$ *son variables aleatorias independientes*

que siguen una misma ley normal $N(0,\sigma_I^2)$, donde σ_I^2 es *la varianza de las interac ciones entre todos los pares* (A_i, B_j).

Se llama *Varianza total* al parámetro $\sigma_T^2 = \sigma_A^2 + \sigma_B^2 + \sigma_I^2 + \sigma_e^2$, que indica la variabilidad total de la población.

Las varianzas $\sigma_A^2, \sigma_B^2, \sigma_I^2$ y σ_e^2 se llaman componentes de la varianza total, porque indican la influencia relativa de cada factor, de las interacciones entre ellos y del azar en la variabilidad total de la población.

Los objetivos del ADEVA en un modelo II son

- *Contrastar si existe variabilidad e interacción entre los niveles.*
- *Estimar los parámetros componentes de la varianza total.*

Hipótesis que se contrastan

hipótesis nulas

H_0: *Los niveles del factor A no producen variabilidad:* $\sigma_A^2 = 0$

H'_0: *Los niveles del factor B no producen variabilidad:* $\sigma_B^2 = 0$

H''_0: *No existen interacciones entre niveles de A y de B:* $\sigma_I^2 = 0$

hipótesis alternativas

H_1: *Los niveles del factor A producen variabilidad:* $\sigma_A^2 > 0$

H'_1: *Los niveles del factor B producen variabilidad:* $\sigma_B^2 > 0$

H''_1: *Existen interacciones entre niveles de A y de B:* $\sigma_I^2 > 0$

EJEMPLO II:

Temperatura	Salinidad 1	2	3	4
15^0	16 14	20 18	13 12	11 10
20^0	8 11	17 14	15 12	10 9
25^0	7 7	17 18	14 16	6 10
30^0	1 3	16 18	8 6	3 5

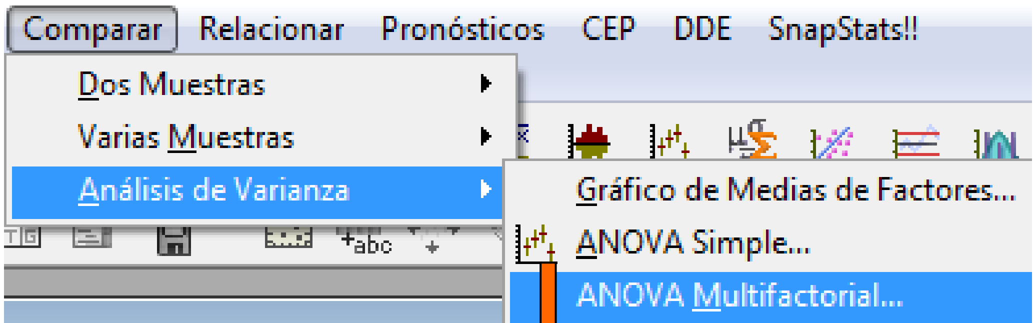
Supondremos que este ejemplo corresponde a un diseño de efectos aleatorios.

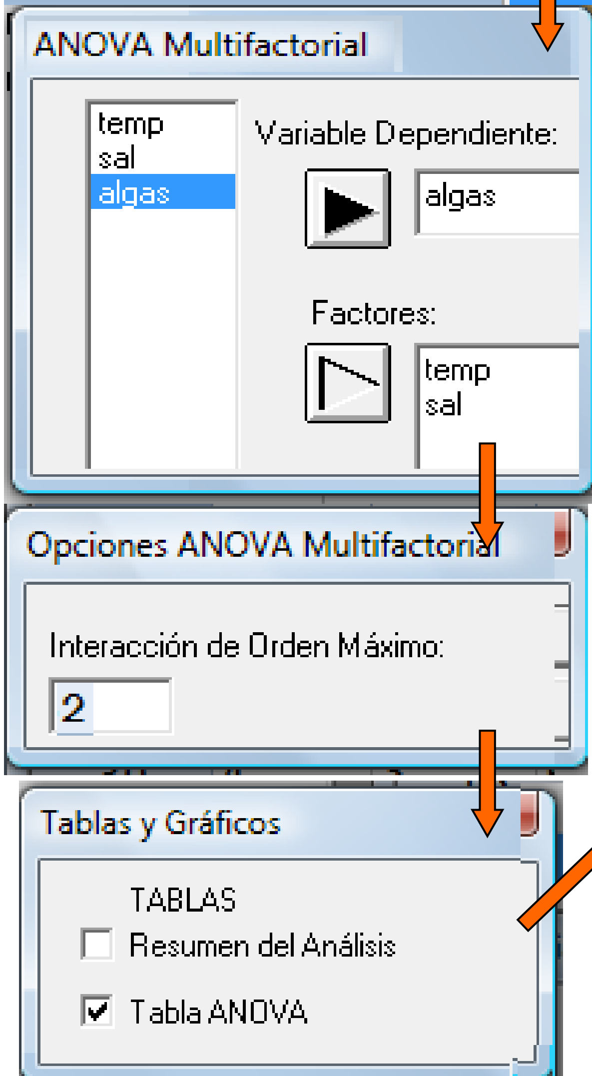
Para introducir estos datos, llamaremos GRADOS a la columna del primer factor, donde escribiremos los números 1 (15%), **2 (**20%), **3 (**25%) **y 4** (30%), **ocho veces cada uno.**

Llamaremos SAL a la columna del segundo factor, donde escribiremos los números 1, 2, 3 y 4 dos veces cada uno, a la derecha de cada número de la columna grados.

Llamaremos ALGAS a la columna de la variable de respuesta, donde escribiremos cada dato de la tabla en la fila donde están los dos niveles que le corresponden.

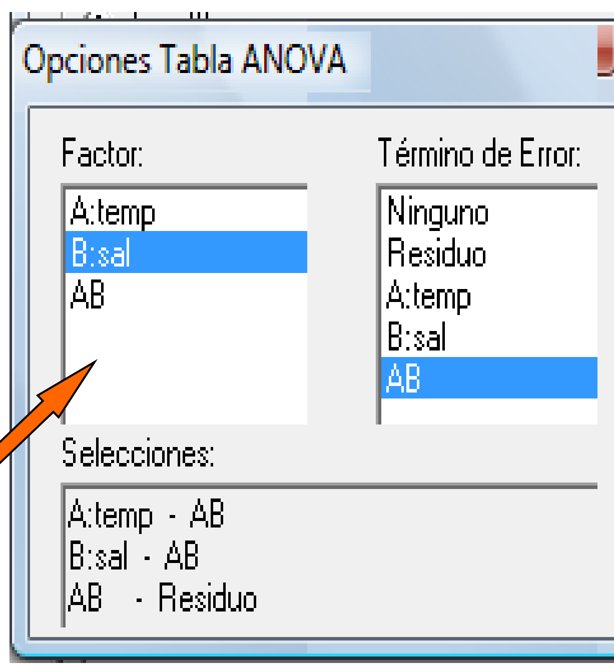
	temp	sal	algas
1	1	1	16
2	1	1	14
3	1	2	20
4	1	2	18
5	1	3	13
6	1	3	12
7	1	4	11
8	1	4	10
9	2	1	8
10	2	1	11
11	2	2	17
12	2	2	14
13	2	3	15
14	2	3	12
15	2	4	10
16	2	4	9
17	3	1	7
18	3	1	7
19	3	2	17
20	3	2	18
21	3	3	14
22	3	3	16
23	3	4	6
24	3	4	10
25	4	1	1
26	4	1	3
27	4	2	16
28	4	2	18
29	4	3	8
30	4	3	6
31	4	4	3
32	4	4	5



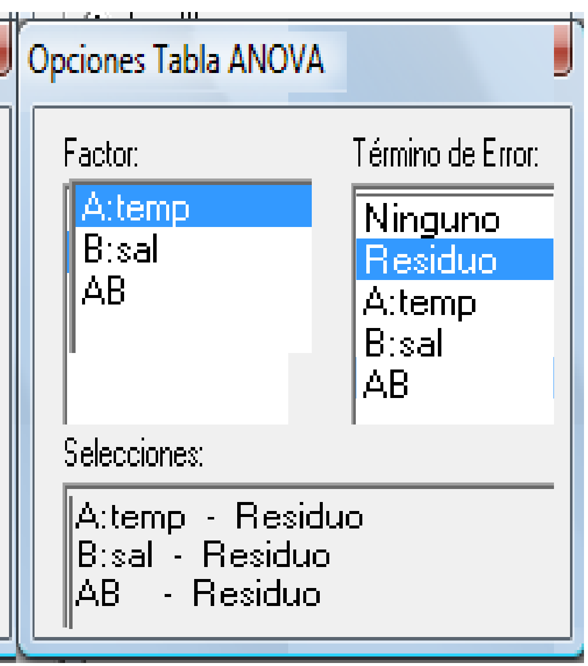


En el diseño de efectos aleatorios sólo necesitamos la tabla de ANOVA, en la que deben modificarse los cocientes entre medias de cuadrados que serán los estadísticos de contraste.
Esto se consigue pulsando opciones de ventana y eligiendo para los dos factores la interacción (AB) como término de error.

Cocientes F para efectos aleatorios



Cocientes F para efectos fijos (viene por defecto)



☑ Tabla ANOVA

TABLA DEL ANOVA Y P- VALOR DE CADA UNO DE LOS CONTRASTES

Análisis de Varianza para algas - Suma de Cuadrados Tipo III

Fuente	*Suma de Cuadrados*	*Gl*	*Cuadrado Medio*	*Razón-F*	*Valor-P*	
EFECTOS PRINCIPALES						
A:temp	191,344 ***SCA***	3	63,7813 ***MCA***	4,84	0,0284	(1)
B:sal	442,344 ***SCB***	3	147,448 ***MCB***	11,20	0,0022	(1)
INTERACCIONES						
AB	118,531 ***SCI***	9	13,1701 ***MCI***	5,62	0,0014	(0)
RESIDUOS	37,5 ***SCD***	16	2,34375 ***MCD***			
TOTAL (CORREGIDO)	789,719 ***SCT***	31				

Las Razones-F se basan en los siguientes cuadrados medios: (0) Residuo (1) AB

Los estadísticos F de los factores son MCA / MCI y MCB / MCI y el de la interacción es MCI / MCD

La interpretación de esta tabla es la siguiente:

Se rechaza la hipótesis nula de varianza nula de los grados de temperatura a cualquier nivel de significación mayor que 0.0284

Se rechaza la hipótesis nula de varianza nula de los niveles de salinidad a cualquier nivel de significación mayor que 0.0022

Se rechaza la hipótesis nula de varianza nula de la interacción entre los dos factores a cualquier nivel de significación mayor que 0.0014

En particular, a un nivel de significación de 0.05 se rechazan las tres hipótesis nulas *pero a un nivel de significación de 0.025 sólo se rechazan las dos últimas.*

Cuando se rechaza alguna de las hipótesis nulas del ANOVA de efectos aleatorios debe estimarse la varianza correspondiente a esa hipótesis:

Conclusiones del Ejemplo II

Las estimaciones de las varianzas no nulas son:

$$\hat{\sigma}_A^2 = \frac{63.781 - 13.170}{4 \times 2} = 6.326; \quad \hat{\sigma}_B^2 = \frac{147.448 - 13.170}{4 \times 2} = 16.785;$$

$$\hat{\sigma}_{AB}^2 = \frac{13.170 - 2.344}{2} = 5.413 \quad \hat{\sigma}_e^2 = 2.344; \quad \hat{\sigma}_T^2 = 30.868$$

Los porcentajes de las varianzas parciales en la varianza total son:

$$\hat{\sigma}_A^2 \longrightarrow 20.5\% \quad \hat{\sigma}_B^2 \longrightarrow 54.4\%; \quad \hat{\sigma}_{AB}^2 \longrightarrow 17.5\%; \quad \hat{\sigma}_e^2 \longrightarrow 7.6\%$$

Estas estimaciones no se calculan en el programa, sino que deben obtenerse en la calculadora.

Modelo III o de efectos MIXTOS, con el factor A de efectos fijos

(Los niveles del factor A se eligen a voluntad y los del factor B se eligen al azar.)

Los efectos α_i *son parámetros desconocidos,* que verifican $\sum_{i=1}^{I} \alpha_i = 0$

Para cada índice i**, el parámetro** $\mu_{i\bullet} = \mu + \alpha_i$ **se llama** *media del nivel* A_i.

Los efectos β_j *son variables aleatorias independientes entre sí y de los errores,* **que siguen una misma ley normal** $N(0, \sigma_B^2)$, **donde** σ_B^2 ***es la varianza de los efectos de todos los posibles niveles del factor B.***

Los efectos $(\alpha\beta)_{ij}$ *son variables aleatorias independientes entre sí, de las* β_j *y de los errores,* **que siguen una misma ley normal** $N(0, \sigma_I^2)$, **donde** σ_I^2 ***es la varianza de las interacciones entre todos los posibles pares*** (A_i, B_j).

Se llama *Varianza total* **al parámetro** $\sigma_T^2 = \sigma_B^2 + \sigma_I^2 + \sigma_e^2$, **que indica la variabilidad total de la población.**

Las varianzas σ_B^2, σ_I^2 y σ_e^2 *se llaman* ***componentes de la varianza total,*** *porque indican la influencia relativa del factor B, de las interacciones entre los factores y del azar en la variabilidad total de la población.*

Los objetivos del ADEVA en un modelo mixto son:

- *Decidir si las medias de los niveles del factor A son distintas entre sí.*
- *Contrastar si existe variabilidad entre los niveles del factor B.*
- *Decidir si existen interacciones entre los niveles de ambos factores.*

Hipótesis que se contrastan

hipótesis nulas

H_0: *Las medias de los niveles A_i son iguales:* $\mu_{i.} = \mu \ \forall i$

H'_0: *Los niveles del factor B no producen variabilidad:* $\sigma_B^2 = 0$

H''_0: *No existen interacciones entre niveles de A y de B:* $\sigma_I^2 = 0$

hipótesis alternativas

H_1: *Las medias de los niveles A_i no son iguales:* $\exists \ \mu_{i.} \neq \mu$

H_1': *Los niveles del factor B producen variabilidad:* $\sigma_B^2 > 0$

H_1'': *Existen interacciones entre niveles de A y de B:* $\sigma_I^2 > 0$

Interpretación de las decisiones de los contrastes en los modelos aleatorio y mixto:

Si el ADEVA de un modelo de efectos aleatorios es significativo, ***sólo deben estimarse las varianzas de los efectos cuya hipótesis nula se haya rechazado.***

Los estimadores de las varianzas no nulas son:

$$\hat{\sigma}_A^2 = \frac{MCA - MCI}{JK}; \quad \hat{\sigma}_B^2 = \frac{MCB - MCI}{IK}; \quad \hat{\sigma}_I^2 = \frac{MCI - MCD}{K}$$

Si el ADEVA de un modelo mixto, con factor de efectos fijos A, es significativo, ***deben estimarse las medias de los niveles A_i, si se rechaza la hipótesis nula de igualdad y las varianzas de los efectos de los niveles B_i o de las interacciones, si se rechaza la hipótesis nula respectiva.***

Sus estimadores son: $\hat{\sigma}_B^2 = \dfrac{MCB - MCD}{IK}; \qquad \hat{\sigma}_I^2 = \dfrac{MCI - MCD}{K}$

En ambos modelos, deben estimarse la varianza de los errores y la varianza total de la población, que es la suma de todas las varianzas que sean positivas porque se ha rechazado su hipótesis nula correspondiente.

En el modelo II: $\hat{\sigma}_T^2 = \hat{\sigma}_A^2 + \hat{\sigma}_B^2 + \hat{\sigma}_I^2 + \hat{\sigma}_e^2$

En el modelo III: $\hat{\sigma}_T^2 = \hat{\sigma}_B^2 + \hat{\sigma}_I^2 + \hat{\sigma}_e^2$

Por último, deben calcularse los porcentajes con los que cada varianza contribuye a la total.

EJEMPLO 3

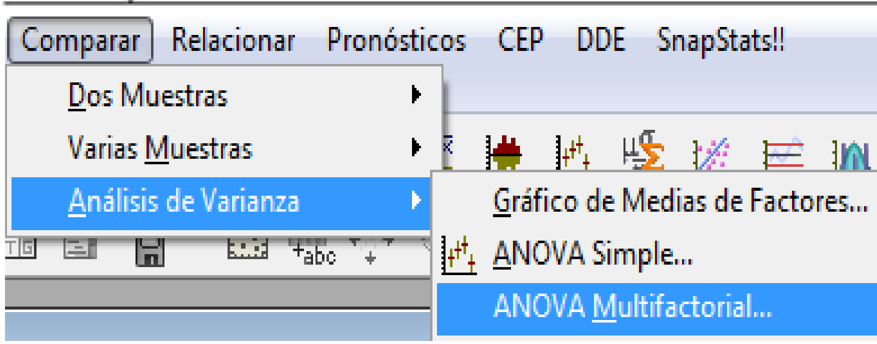
3 En la siguiente tabla se dan los resultados de un diseño factorial en el que se analiza la mortalidad de las lapas de una determinada zona costera frente a cuatro substancias contaminantes, (A1, A2, A3, A4), en tres niveles de cobertura de algas, elegidos al azar (bajo, medio, alto). El diseño consta de 3 réplicas para cada tratamiento; los datos son los índices de mortalidad de las lapas en las unidades experimentales después de sufrir la acción de los contaminantes durante un mes.

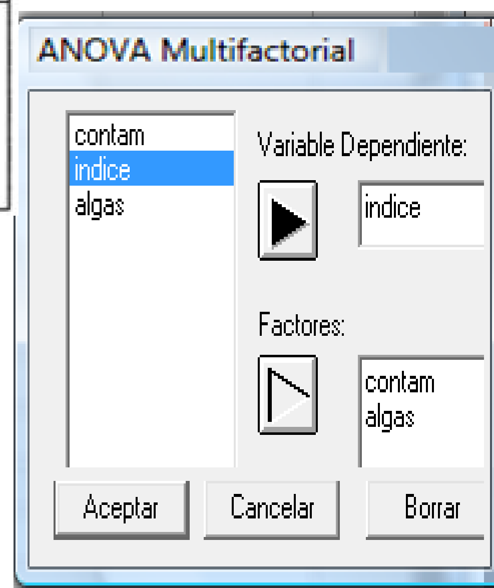
a) En el supuesto de que se verifican las condiciones de un ANOVA, efectúe el contraste de las hipótesis que corresponden a este modelo de diseño a un nivel de significación del 10 %, explique las conclusiones que se deducen de estos contrastes.

b) Calcule estimaciones de las varianzas que contribuyen a la varianza total de la población y halle el porcentaje con que contribuye cada una.

c) Determine, si es posible, cuál de estos contaminantes es más mortífero para estas lapas.

	Cobertura de algas								
Contam.	bajo			medio			alto		
A1	.28	.43	.57	.41	.54	.67	.71	.60	.63
A2	.26	.45	.53	.44	.55	.69	.69	.62	.65
A3	.41	.57	.69	.53	.65	.77	.81	.61	.72
A4	.44	.58	.71	.51	.67	.74	.84	.58	.71

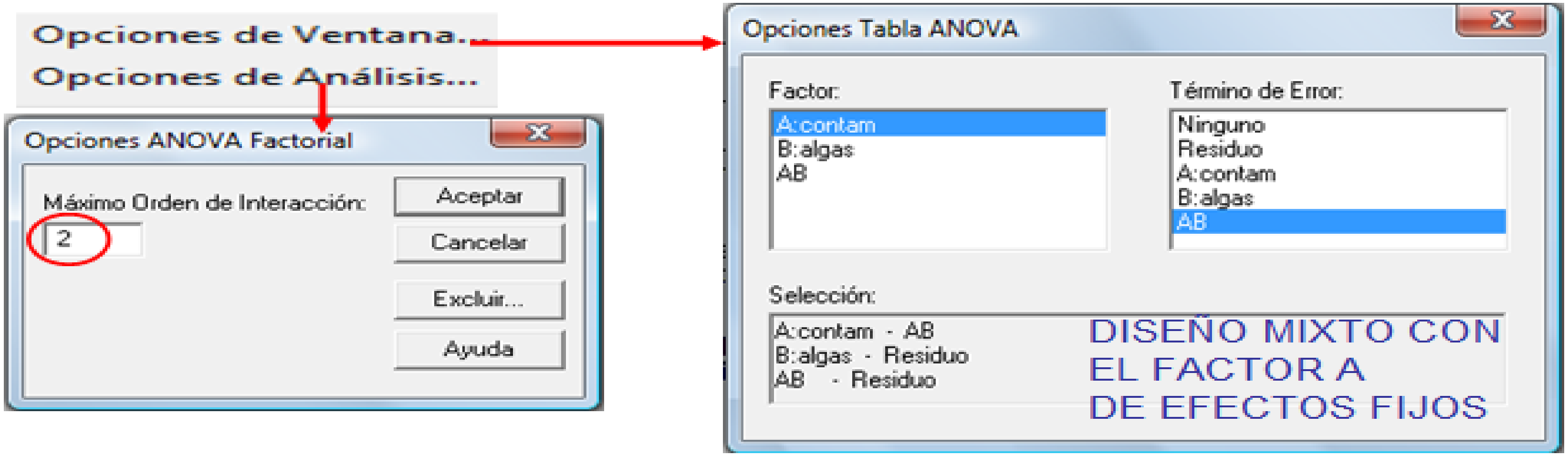




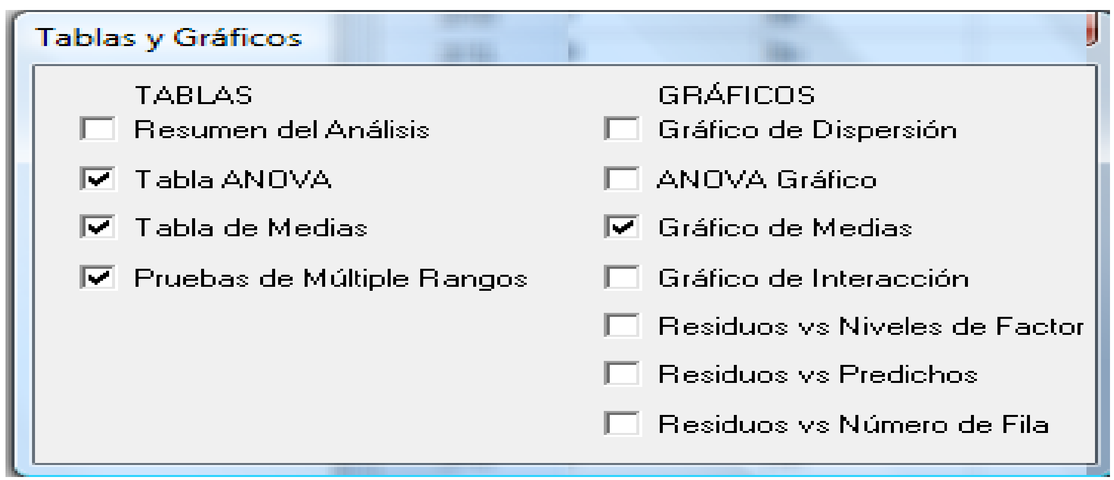
	conta	indice	algas
1	A1	0,28	b
2	A1	0,43	b
3	A1	0,57	b
4	A1	0,41	m
5	A1	0,54	m
6	A1	0,67	m
7	A1	0,71	a
8	A1	0,6	a
9	A1	0,63	a
10	A2	0,26	b
11	A2	0,45	b
12	A2	0,53	b
13	A2	0,44	m
14	A2	0,55	m
15	A2	0,69	m
16	A2	0,69	a
17	A2	0,62	a
18	A2	0,65	a
19	A3	0,41	b
20	A3	0,57	b
21	A3	0,69	b
22	A3	0,53	m
23	A3	0,65	m
24	A3	0,77	m
25	A3	0,81	a
26	A3	0,61	a
27	A3	0,72	a
28	A4	0,44	b
29	A4	0,58	b
30	A4	0,71	b
31	A4	0,51	m
32	A4	0,67	m
33	A4	0,74	m
34	A4	0,84	a
35	A4	0,58	a
36	A4	0,71	a

En el diseño de efectos mixtos, con el factor A de efectos fijos, deben modificarse en la tabla de ANOVA los cocientes entre medias de cuadrados que serán los estadísticos de contraste. Esto se consigue pulsando opciones de ventana y eligiendo como términos de error la interacción (AB) para el factor A y el residuo para el factor B.

Cocientes F para efectos mixtos (A fijo)



También necesitaremos la tabla de medias, la prueba de rangos y el gráfico de medias, pero sólo para el factor A



☑ Tabla ANOVA

TABLA DEL ANOVA Y P- VALOR DE CADA UNO DE LOS CONTRASTES

Análisis de Varianza para indice - Suma de Cuadrados Tipo III

Fuente	*Suma de Cuadrados*	*Gl*	*Cuadrado Medio*	*Razón-F*	*Valor-P*	
EFECTOS PRINCIPALES						
A:conta	0,0921222 ***SCA***	3	0,0307074 ***MCA***	14,63	0,0036	(1)
B:algas	0,211806 ***SCB***	2	0,105903 ***MCB***	7,45	0,0030	(0)
INTERACCIONES						
AB	0,0125944 ***SCI***	6	0,00209907 ***MCI***	0,15	0,9878	(0)
RESIDUOS	0,341067 ***SCD***	24	0,0142111 ***MCD***			
TOTAL (CORREGIDO)	0,657589 ***SCT***	35				

Las Razones-F se basan en los siguientes cuadrados medios: (0) Residuo (1) AB

El estadístico F del factor A es MCA / MCI, el del factor B es MCB / MCD y el de la interacción es MCI / MCD

La interpretación de esta tabla es la siguiente:

Se rechaza la hipótesis nula de igualdad de las medias de los contaminantes a cualquier nivel de significación mayor que 0.0036

Se rechaza la hipótesis nula de varianza nula de los niveles de cobertura de algas a cualquier nivel de significación mayor que 0.0030

Se acepta la hipótesis nula de varianza nula de la interacción entre los dos factores a cualquier nivel de significación menor que 0.9878

En particular, a un nivel de significación de 0.10 se rechazan las dos primeras hipótesis nulas *pero se acepta la tercera*

CONCLUSIONES QUE SE DEDUCEN DE ESTE ADEVA

Como se ha rechazado la hipótesis nula de igualdad de las medias de los contaminantes, hemos de estimar estas medias y realizar pruebas a posteriori a un nivel del 10 %, para decidir qué medias difieren entre sí a ese nivel.

Tabla de Medias por Mínimos Cuadrados para indice con intervalos de confianza del 90,0%

			Error	*Límite*	*Límite*
Nivel	*Casos*	*Media*	*Est.*	*Inferior*	*Superior*
MEDIA GLOBAL	36	0,590556			
conta					
1	9	0,537778	0,0152719	0,508102	0,567454
2	9	0,542222	0,0152719	0,512546	0,571898
3	9	0,64	0,0152719	0,610324	0,669676
4	9	0,642222	0,0152719	0,612546	0,671898

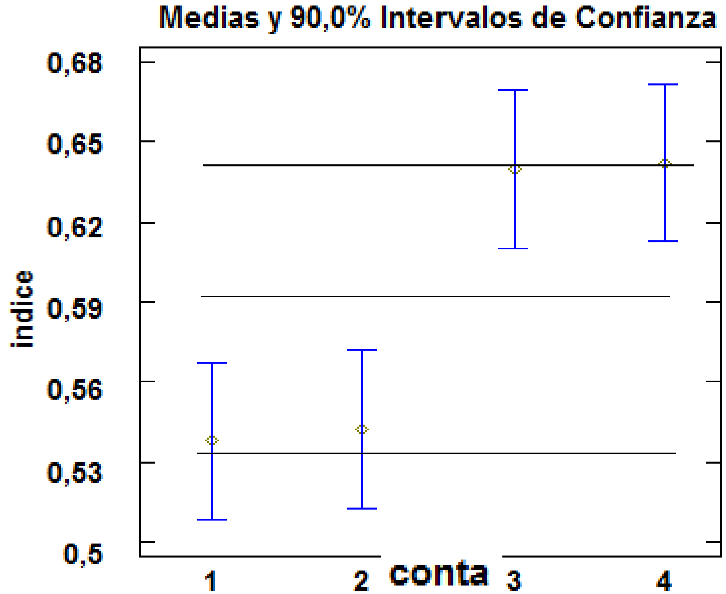
Pruebas de Múltiple Rangos para indice por conta

Método: 90,0 porcentaje Scheffe

conta	*Casos*	*Media LS*	*Sigma LS*	*Grupos Homogéneos*
1	9	0,537778	0,0152719	X
2	9	0,542222	0,0152719	X
3	9	0,64	0,0152719	X
4	9	0,642222	0,0152719	X

Contraste	*Sig.*	*Diferencia*	*+/- Límites*
1 - 2		-0,00444444	0,0678398
1 - 3	*	-0,102222	0,0678398
1 - 4	*	-0,104444	0,0678398
2 - 3	*	-0,0977778	0,0678398
2 - 4	*	-0,1	0,0678398
3 - 4		-0,00222222	0,0678398

* indica una diferencia significativa.



En este gráfico se solapan entre sí los intervalos de los contaminantes 1 y 2 y los de los contaminantes 3 y 4, pero no se solapan los de los dos primeros con los dos últimos, luego las medias 1 y 2 son distintas de las 3 y 4, al nivel del 10%, como indica la tabla del contraste.

Además, como se ha rechazado que la varianza de los niveles de cobertura es nula, deben estimarse esa varianza, la varianza residual, la varianza total y los porcentajes de cada componente en la total.

ESTIMACIÓN DE LA VARIANZA DE LOS EFECTOS DE LA COBERTURA DE ALGAS:

$\hat{\sigma}^2_B$ = (MCB - MCD)/IK = (0,105903 - 0,0142111)/(4x3) = 0,0916919/12 = 0,007641

ESTIMACIÓN DE LA VARIANZA DE LOS ERRORES: $\hat{\sigma}^2_e$ = MCD = 0,014211

ESTIMACIÓN DE LA VARIANZA TOTAL: $\hat{\sigma}^2_T$ = 0,007641 + 0,014211 = 0,021852

EL PORCENTAJE CON QUE CONTRIBUYE $\hat{\sigma}^2_B$ ES 0,007641 x 100/0,021852 = 34,97 %;

EL PORCENTAJE CON QUE CONTRIBUYE $\hat{\sigma}^2_e$ ES 0,014211 x 100/0,021852 = 65,03 %

Estas estimaciones no se calculan en el programa, sino que deben obtenerse en la calculadora.

RESUMEN DE LAS HIPÓTESIS QUE SE CONTRASTAN EN LOS TRES MODELOS DE ANOVA

En el modelo I

hipótesis nula.	hipótesis alternativa
H_0: *Las medias de los niveles A_i son iguales:* $\mu_{i.} = \mu \; \forall i$	H_1: $\exists \, \mu_{i.} \neq \mu$
H'_0: *Las medias de los niveles B_j son iguales:* $\mu_{.j} = \mu \; \forall j$	H'_1: $\exists \, \mu_{.j} \neq \mu$
H''_0: *No hay interacciones entre niveles de A y niveles de B:* $(\alpha\beta)_{ij} = 0 \; \forall i, \forall j$	H''_1: $\exists \, (\alpha\beta)_{ij} \neq 0$

En el modelo II

hipótesis nula.	hipótesis alternativa
H_0: *Los niveles del factor A no producen variabilidad:* $\sigma_A^2 = 0$	H_1: $\sigma_A^2 > 0$
H'_0: *Los niveles del factor B no producen variabilidad:* $\sigma_B^2 = 0$	H'_1: $\sigma_B^2 > 0$
H''_0: *No existen interacciones entre niveles de A y de B:* $\sigma_I^2 = 0$	H''_1: $\sigma_I^2 > 0$

En el modelo III *con A fijo*

hipótesis nula.	hipótesis alternativa
H_0: *Las medias de los niveles A_i son iguales:* $\mu_{i.} = \mu \; \forall i$	H_1: $\exists \, \mu_{i.} \neq \mu$
H'_0: *Los niveles del factor B no producen variabilidad:* $\sigma_B^2 = 0$	H'_1: $\sigma_B^2 > 0$
H''_0: *No existen interacciones entre niveles de A y de B:* $\sigma_I^2 = 0$	H''_1: $\sigma_I^2 > 0$