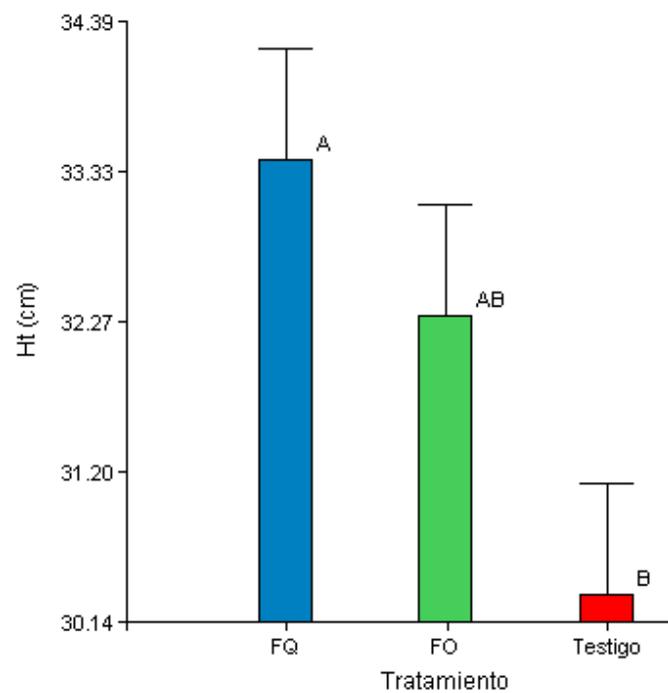


ANÁLISIS DE VARIANZA

Comparando tres o más medias



Jorge Fallas
2012

Contenido

1. Introducción	1
1.1. Terminología	2
2. Análisis de varianza de una vía o completamente aleatorio (modelo I, efectos fijos)	4
2.1 Lógica del análisis de una vía	5
2.2 Estimaciones de las varianzas	8
2.3 La prueba F (igualdad de varianzas)	11
2.4 Pruebas a posteriori	13
A. Diferencia mínima significativa irrestricta LSD	14
B. Diferencia mínima significativa restringida de Fisher	16
C. Prueba de rangos múltiples de Duncan	17
D. Prueba de Student-Newman-Keuls (S.N.K)	18
E. Tukey	20
F. Scheffé	21
G. Dunnett (comparación con un control)	22
H. Intervalo de Bonferroni o corrección de Bonferroni	23
2.4.1. Limitaciones de las pruebas a posteriori	24
2.5. Pruebas a priori o planeadas	25
3. Diseño de bloques completos al azar	28
3.1. ANDEVA de dos vías	30
3.1.1. ANDEVA de dos vías con una observación por celda	31
A. Tabla de análisis de varianza	34
3.2 Diseño factorial con “n” observaciones por celda	40
3.2.1 ¿Qué es la interacción?	44
3.2.2 Tabla de análisis de varianza para un diseño factorial	44
3.2.3. Ventajas del diseño de dos vías	46
4. Transformaciones	46
5. Comentario final	48
6. Bibliografía	48

7. Ejercicios

El presente documento se distribuye bajo licencia CC BY-NC-SA de “Creative Commons” “reconocimiento-No comercial-Compartir bajo la misma licencia”; la cual permite a otros entremezclar, ajustar y construir con base en su trabajo para fines no comerciales, siempre y cuando se de crédito y licencia de sus nuevas creaciones, en los términos idénticos.

1. Introducción

En las secciones previas del curso hemos estudiado el tema de prueba de hipótesis para dos medias independientes (μ_1 y μ_2). La pregunta que nos hacíamos era: ¿Existe suficiente evidencia estadística en el set de datos como para creer que las muestras provienen de dos poblaciones diferentes?; o por el contrario ¿Indica la evidencia que se trata de dos muestras de la misma población?.

En el presente capítulo extendemos esta pregunta a tres o más muestras y lo denominamos *análisis de varianza- ANDEVA (ANOVA en inglés)*. En este caso comparamos las estimaciones de la varianza entre muestras y al interior de cada muestra para determinar mediante una prueba F si se trata de una misma población (H_0 : las muestras provienen de la misma población) o si por el contrario, provienen de diferentes poblaciones. Una vez probado que al menos un par de medias es diferente se procede a aplicar una prueba de comparaciones múltiples de medias para determinar cuáles de ellos son estadísticamente diferentes. Un método de análisis alternativo es realizar comparaciones planeadas; para las cuales no es necesario realizar primero una prueba F (Fig. 1)



Figura 1: Flujograma de un análisis de varianza.

El análisis de varianza permite analizar el efecto de una o más variables o categorías en un conjunto de datos. Cada “tratamiento” puede tener varias observaciones (e.g. 20 plántulas por tratamiento) o por el contrario tener una *única* observación por tratamiento (e.g número de semillas germinadas por lote de cien semillas). Veamos algunos ejemplos:

Diseño experimental

En este caso el investigador(a) tiene control sobre los grupos a investigar (tratamientos) y asigna cada tratamiento a los sujetos experimentales de manera aleatoria.

A. Un tratamiento y diferentes niveles del mismo

Usted desea determinar el efecto de tres métodos de plantación (bolsa, raíz desnuda y plantón) en la sobrevivencia de árboles de laurel. El tratamiento es *método de plantación* y estamos interesados en determinar si al menos uno de los métodos utilizados es estadísticamente superior a los otros; o si por el contrario el tratamiento no tiene efecto y por lo tanto podemos asumir que todos brindan el mismo resultado.

B. Dos tratamientos y diferentes niveles de cada uno

Ahora, supongamos que utilizamos plántulas tres especies arbóreas (A, B y C) y que además utilizamos dos métodos de plantación (bolsa y raíz desnuda). En este caso tenemos seis tratamientos: *dos métodos de plantación y tres especies de árboles* y por esta razón a este análisis de varianza se le denomina de dos vías.

Diseño no experimental

En este caso el investigador(a) no tiene control sobre los grupos a investigar y solo puede asignarlos a una de varias posibles categorías.

A. Un tratamiento y diferentes niveles del mismo

Usted desea saber si ENOS (El Niño Oscilación Sur) tiene un efecto en la cantidad de lluvia anual en su área de estudio. Para el estudio usted divide los años con datos de precipitación en Neutro, Niña y Niño y luego realiza un análisis de varianza de una vía.

B. Dos tratamientos y diferentes niveles de cada uno

Usted desea saber cómo afecta ENOS (El Niño Oscilación Sur) y la elevación la precipitación anual a nivel nacional. Para el estudio usted divide los años con datos de precipitación en Neutro, Niña y Niño y además los clasifica por rangos de elevación y luego realiza un análisis de varianza de dos vías.

El tema de análisis de varianza es complejo y por lo tanto en el presente capítulo sólo se tratarán dos de los modelos más simples de ANDEVA: *el análisis de varianza simple o de una vía* y *el análisis de dos vías*.

Antes de proceder a los métodos de análisis, se definirán algunos términos de uso frecuente en esta área de la estadística inferencial.

1.1. Terminología

Población: Es el total o universo al cual se desea aplicar la inferencia o conclusión del estudio.

Muestra: Es una parte o porción de la realidad bajo estudio.

Deducción: A partir del todo (población) se deriva una afirmación que aplica a una condición particular (muestra).

Inducción: A partir de una porción de la realidad (muestra) se hace una afirmación sobre el todo (población).

Unidad experimental: Individuo, objeto, grupo o conjunto de sujetos experimentales a los cuales se les aplica un determinado tratamiento. Por ejemplo, la unidad experimental puede ser una parcela en una plantación, un grupo de semillas, una persona entrevistada o un árbol que se mide. En algunos textos se le denomina a la unidad experimental *caso*.

Tratamientos o variables: Procesos o acciones cuyos efectos serán medidos en el material experimental y posteriormente comparados entre sí para determinar si existen diferencias estadísticamente significativas. Los tratamientos pueden ser cualitativos (e.g. ENOS) ó cuantitativos (e.g. dosis de fertilizantes).

Observación: Es la medición realizada en una unidad experimental.

Testigo: Unidad experimental al cual no se le aplica el tratamiento (sirve de referencia) que es utilizado para determinar si los tratamientos tienen un efecto estadísticamente discernible sobre el material experimental.

Variable respuesta: Es aquella propiedad o cualidad de la unidad experimental que se cuantifica. Para mayor detalle sobre niveles de medición ver capítulo uno.

Repetición: Réplica estadísticamente independiente de un tratamiento. En este caso el tratamiento es aplicado dos o más veces a diferentes unidades experimentales; cada aplicación brinda una estimación independiente de la respuesta del sujeto experimental al tratamiento. Cuantas más réplicas se tenga mejor será la estimación del error experimental. En la mayoría de los casos se recomienda un mínimo de tres observaciones independientes por tratamiento.

Seudo replicación: Es el resultado de muestrear dos o más veces la misma condición (muestras no independientes). Por ejemplo, al evaluar la densidad de peces en dos ríos; uno contaminado y otro no, si se muestrean cinco sitios al azar en cada uno de ellos, dichas muestras no representan réplicas ya que se está muestreando el mismo río. En el sentido estadístico para que se consideren réplicas debería de elegirse al azar dos o más ríos por condición (contaminado-no contaminado) y luego obtener muestras independientes de cada uno de ellos. Esto permitiría estimar la variabilidad natural de cada uno de los sistemas acuáticos en los cuales viven los peces que se muestrean. Aun cuando el análisis de los datos presupone la existencia de réplicas independientes, en la mayoría de los estudios en el área de recursos naturales no es posible cumplir con este supuesto.

Medición: Proceso de asignar un valor (nominal, ordinal, razón, intervalo) a un fenómeno, proceso u objeto

Significancia estadística: Esta es una regla que permite afirmar que la diferencia observada entre dos o más tratamientos o grupos es el resultado del efecto del tratamiento o de la variable de clasificación y no del azar. Con frecuencia se declaran como significativas aquellas diferencias que tienen una probabilidad inferior a 0,1 (o sea 10%) de ocurrir en forma aleatoria.

En algunos textos de estadística se recomienda utilizar un asterisco (*) para designar las diferencias significativas a un 5% ($P < 0.05$), dos asteriscos (**) para designar diferencias significativas al 1% ($P < 0.01$) y tres asteriscos (***) para designar diferencias significativas al 0,1% ($P < 0.001$). Sin embargo, dado que los paquetes estadísticos le brindan el valor de P, se recomienda reportarlo conjuntamente con el tamaño de la muestra.

Consistencia: Un método de análisis estadístico es consistente cuando la significancia de la prueba depende exclusivamente de: 1) la diferencia entre las dos medias, 2) el error estándar de las diferencias, 3) el número de grados de libertad del error, y 4) el nivel de significancia al cual se hace la prueba.

Aleatorización: Asignación aleatoria de los tratamientos a los sujetos o unidades experimentales. Esto elimina cualquier sesgo conocido o desconocido en la asignación de los tratamientos.

Error experimental: Variación natural del material experimental no controlado por el investigador(a). Este no es un error adrede o derivado de la aplicación errónea de técnicas de medición sino simplemente un componente propio del material experimental.

Cuasi o pseudo experimento: Estudio en el cual se utilizan los principios propuestos para el diseño de experimentos; sin embargo no es posible asignar los tratamientos en forma aleatoria. Este tipo de estudios es común en el área de ecología y en general en estudios de tipo observacional.

2. Análisis de varianza de una vía o completamente aleatorio (modelo I, efectos fijos)

Este es el diseño experimental más sencillo y es similar al muestreo simple al azar. Los tratamientos se asignan al azar a una serie de unidades experimentales seleccionadas previamente. En general, este no es el diseño más eficiente en el área de experimentación sin embargo es flexible y le permite al investigador(a) someter a prueba cualquier número de “tratamientos”. Es deseable que se asigne el mismo número de unidades experimentales por tratamiento. Otra ventaja del diseño es que determina el error experimental utilizando el máximo número posible de grados de libertad.

El diseño es apropiado para condiciones de laboratorio sin embargo es poco utilizado en experimentos de campo ya que otros diseños como el de bloques al azar brinda una mayor precisión en la estimación del error. El número mínimo de unidades experimentales requeridas para el experimento será igual al número de tratamientos por el número de repeticiones.

Recuerde que los tratamientos se deben asignar al azar a cada una de las unidades experimentales; por ejemplo, si se tienen tres tratamientos y tres repeticiones (nueves opciones para asignar el tratamiento) se puede utilizar una tabla de números al azar para seleccionar valores de 1 a 9, los primeros tres números se asignarán al tratamiento uno, los segundos tres al tratamiento dos y los últimos tres al tratamiento tres. En síntesis, este diseño es apropiado cuando se desea someter a prueba pocos tratamientos con material experimental homogéneo y cuando existe la posibilidad de que algunas de las unidades experimentales se pierdan.

Bajo este modelo de ANDEVA los sujetos son asignados en forma aleatoria a “n” tratamientos. La premisa es que si los tratamientos no tienen ningún efecto sobre los diferentes grupos experimentales entonces sus promedios serán estadísticamente iguales. La hipótesis nula a probar es la siguiente:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots$$

La hipótesis alternativa es que la media de al menos dos grupos es diferente.

Ha: al menos un par de medias es diferente

A diferencia de la prueba “t” o “Z” presentada en el capítulo sobre prueba de hipótesis, en este caso no aplica el concepto de direccionalidad en la prueba; ya que H_0 puede ser falsa por diversas razones. Por ejemplo, dados tres tratamientos, μ_1 puede ser diferente de μ_2 pero a su vez igual a μ_3 .

2.1 Lógica del análisis de una vía; Error! Marcador no definido.

Supongamos que sesenta plántulas de pochote son elegidas al azar de un vivero y que luego son asignadas al azar a cada uno de tres tratamientos de fertilización (no fertilización-testigo; abono orgánico y abono químico)¹. Los estadísticos descriptivos se muestran en el cuadro 1 y la figura 2 presenta el comportamiento de los datos por tratamiento.

Algunas observaciones sobre el set de datos:

1. Si no existieran diferencias en cuanto a la respuesta de las plántulas a los tratamientos, las medias de los tres tratamientos deberían ser iguales. Por el contrario, de existir un efecto atribuible a alguno de los tratamientos, al menos un par de medias deberían ser diferentes. La media de los tratamientos es mayor que la media del grupo control, esto es un primer indicio de que posiblemente las plantas responden de manera diferenciada a los tratamientos.
2. La diferencia en altura entre el fertilizante orgánico y el químico es de 1,1 cm. El valor parece pequeño y por lo tanto no es fácil confirmar o desechar un efecto de tratamiento. Otra forma de valorar esta diferencia es desde una perspectiva práctica: ¿es para usted como responsable del vivero importante la diferencia? Otra pregunta que podría hacerse es ¿cuál es el costo de cada tratamiento? ¿Estaría usted dispuesto a sacrificar crecimiento en beneficio del ambiente?

¹ Los datos se encuentran en el archivo anova_cuadro1.xlsx-

3. Las gráficas indican que no existen valores extremos y que las observaciones tienden a agruparse de manera más homogénea en el fertilizante químico que en el orgánico.

Cuadro 1: Estadísticos descriptivos para sesenta plántulas de pochote sometidas a tres tratamientos de fertilización. Altura total (cm) después de 6 meses.

Tratamiento	Media (cm)	Tamaño muestra	Desviación estándar (cm)	Coef. Variación (%)
No fertilizante (k=1)	$\bar{X}_1 = 30,34$	20	3,04	10,0
Fertilizante orgánico (k=2)	$\bar{X}_2 = 32,31$	20	3,37	10,4
Fertilizante químico (k=3)	$\bar{X}_3 = 33,42$	20	3,99	11,9
Media general	$\bar{X}_G = 31,02$	60	3,66	11,8

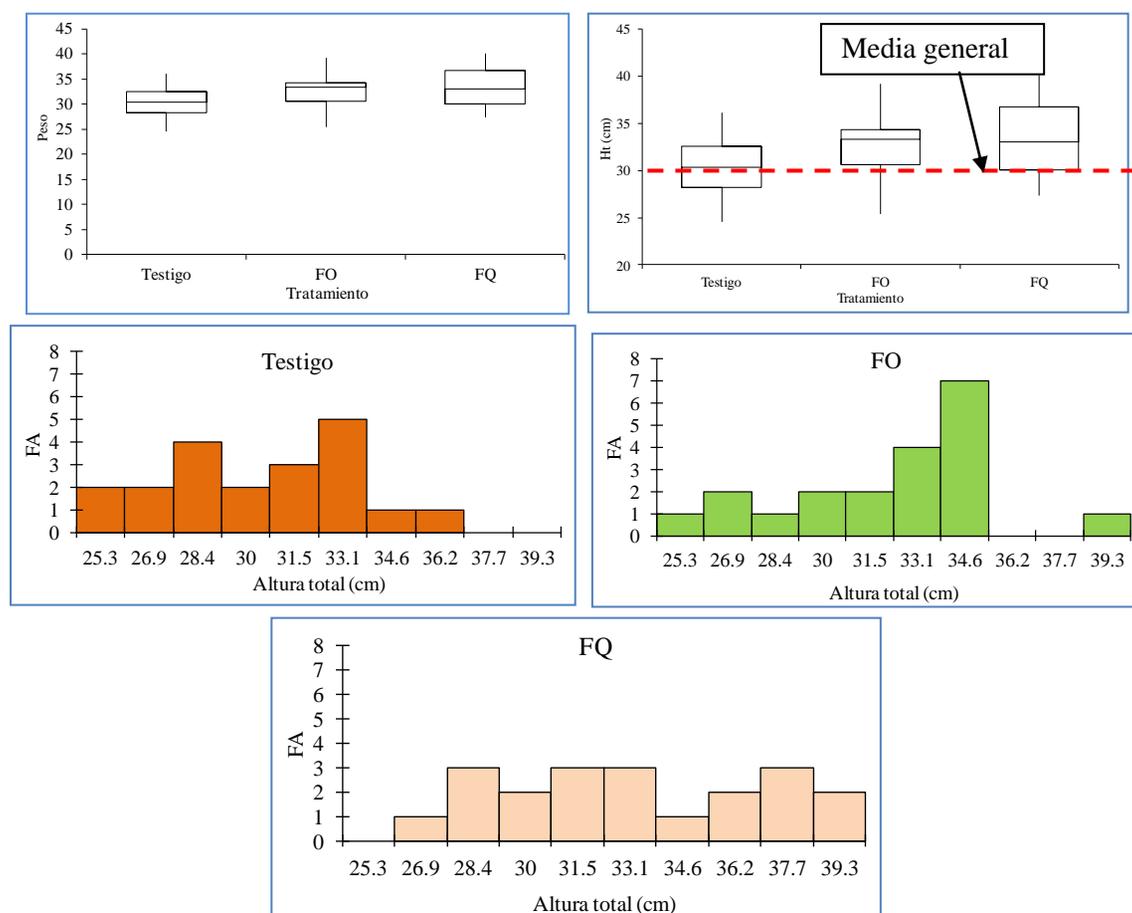


Figura 2: Diagrama de Box-Whisker e histogramas por tratamiento.

Bajo los supuestos del diseño irrestricto al azar el valor esperado de cualquier observación X_{ij} puede estimarse mediante el modelo:

$$X_{ij} = \mu + \beta_i + \varepsilon_{ij} \quad (1)$$

con $i = 1, \dots, n; j = 1, \dots, n$ y $\varepsilon_{ij} = (0, \sigma^2)$

en donde:

μ : representa la media de la población

β_i : representa el efecto del tratamiento i

ε_{ij} : representa el error con una distribución normal $(0, \sigma^2)$

La estimación de máximo verosimilitud de los parámetros anteriores son los siguientes estadísticos:

$$\hat{\mu} = \bar{X} \text{ Estimador de } \mu, \text{ media general} \quad (2)$$

$$\hat{\beta}_j = (\bar{X}_j - \bar{X}) \text{ estimador de } \beta_j, \text{ variación explicada por el tratamiento} \quad (3)$$

$$\hat{\varepsilon}_{ij} = X_{ij} - \bar{X}_i \text{ estimador de } \varepsilon_{ij}, \text{ error experimental} \quad (4)$$

El error asociado a cada observación es el resultado de la variabilidad no explicada por el diseño experimental y es el que se debe minimizar a través de la elección de un diseño estadístico apropiado. La implicación de un error grande es que no se detectarán diferencias significativas entre los tratamientos.

$$\hat{\varepsilon}_{ij} = X_{ij} - \hat{\beta}_j - \hat{\mu} \quad (5)$$

A continuación se muestra cómo se calculan los diferentes componentes del análisis de varianza completamente al azar.

Variación al interior de los tratamientos

El cuadro 2 presenta los estadísticos descriptivos por tratamiento. Observe que el coeficiente de variación oscila entre 10,0 % (testigo) y 11.9% (FQ); esta variabilidad es inherente o propia del grupo experimental y no puede atribuirse a los tratamientos. La respuesta observada en el presente caso es típica de experimentos con organismos vivos; ya que cada sujeto responde en forma diferente a un mismo estímulo o tratamiento.

La variación al interior de los grupos refleja la variabilidad inherente de los individuos que componen el grupo y es independiente del tratamiento que recibe dicho grupo.

Cuadro 2: Estadísticos descriptivos por tratamiento para la variable altura total (cm) de 60 plantas de pochote a la edad de seis meses.

Tratamiento	Todos	Testigo	FO	FQ
Number	60	20	20	20
Media	32,02	30,34	32,31	33,42
S	3,66	3,04	3,37	3,99
CV %	11,8	10,0	10,4	11,9
Asimetría	0,13	-0,08	-0,48	0,17
Mínimo	24,54	24,54	25,42	27,34
Q ₁	29,09	28,20	30,61	30,09
Mediana	32,22	30,37	33,32	33,02
Q ₃	34,31	32,56	34,31	36,75
Máximo	40,08	36,07	39,10	40,08

Variación entre tratamientos

Si observamos la figura 2 es evidente que la respuesta de las plántulas asignadas a cada uno de los tratamientos no es homogénea. La variabilidad entre grupos refleja tanto la influencia del tratamiento como la variabilidad natural o inherente de los individuos que conforman cada uno de los grupos experimentales. Sin embargo, para que la hipótesis nula sea falsa el efecto del tratamiento debe ser tal que cause una respuesta lo suficientemente diferente en los individuos de cada uno de los grupos. En otras palabras, la variabilidad entre grupos debe ser mayor que la variabilidad dentro de grupos.

La variación entre grupos refleja tanto la variabilidad propia de los individuos que componen el grupo más la variabilidad atribuible al efecto diferencial del tratamiento que recibe cada grupo.

Cuando la hipótesis nula es verdadera (*no hay diferencia entre tratamientos*) las dos estimaciones de varianza (*dentro de grupos y entre grupos*) medirían lo mismo: variabilidad inherente; y por lo tanto su valor debería ser muy cercano a uno. Por el contrario, cuando H_0 es falsa, la estimación de la *variabilidad entre grupos* debería ser mayor que la *variabilidad dentro de grupos*; ya que la primera mide tanto la variabilidad inherente más la variabilidad asociada a los tratamientos.

2.2 Estimaciones de las varianzas

El análisis de varianza se basa en la partición de la variabilidad total asociada al experimento (SC_{total}) en dos estimaciones independientes de varianza: la varianza *dentro de los grupos experimentales* y la *varianza entre los grupos experimentales*. Dado que H_0 sea verdadera, ambas estimaciones de varianza son una estimación de la varianza poblacional σ^2 y por lo tanto la razón de varianzas debe ser cercana a uno. En el capítulo dos se definió la varianza como la desviación cuadrática de las observaciones con respecto a la media dividido entre $n-1$. En el análisis de

varianza de una vía se utiliza el mismo concepto; con la diferencia de que existen las siguientes desviaciones cuadráticas:

- ✓ *Variación dentro de grupos:* En este caso se calculan las desviaciones cuadráticas con respecto a la media en cada uno de los grupos experimentales. Esta estimación de varianza representa la variación inherente de los individuos en cada uno de los tratamientos.

$$SCD = \sum_{j=1}^r \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = \sum_{j=1}^r \sum_{i=1}^{n_j} x_{ij}^2 - \sum_{j=1}^r n_j \bar{x}_j^2 \quad (6)$$

- ✓ *Variación entre grupos:* La estimación de σ^2 entre los grupos experimentales mide tanto la variabilidad inherente de cada grupo experimental como la variabilidad asociada al tratamiento asignado a cada grupo. Para su cálculo se utiliza la media de cada tratamiento y la media general. En otras palabras, esta estimación mide cuán variable es la media de cada tratamiento con respecto a la media general. Cuando cada tratamiento tiene el mismo número de casos, la media general es una estimación de la media de cada uno de los grupos experimentales.

$$SCE = \sum_{j=1}^r n_j (\bar{x}_j - \bar{x}_{..})^2 \quad (7)$$

- ✓ *Variación con respecto a la media general o variación total:* Esta es una estimación de la variación total asociada al experimento sin considerar los grupos experimentales o tratamientos. La variación total es igual a la suma de la variación al interior de los grupos y entre grupos.

$$SCT = \sum_{j=1}^r \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2 \quad (8)$$

Las fórmulas utilizadas para estimar las varianzas anteriores son:

Varianza dentro de grupos:

$$SC_{dentro} = SC_w / gl \quad (9)$$

- en donde: SC_{dentro} : suma de cuadrados al interior de cada grupo
- gl : grados de libertad ($N - k$); N : total observaciones; k : No. de tratamientos.

Varianza entre grupos:

$$SC_{entre} = SC_k / gl \quad (10)$$

- en donde: SC_{entre} : suma de cuadrados entre grupos
- gl : grados de libertad ($k - 1$); k : número de tratamientos

Varianza total:

$$SC_{total} = SC_{dentro} + SC_{entre} \quad (11)$$

La tabla de análisis de varianza de una vía contiene los siguientes elementos para el modelo de efectos fijos (modelo tipo I):

Cuadro 3: Tabla de análisis de varianza

Fuente de variación	Gl	SC	CM	F	CM esperado
Entre (K)	k-1	SC _k	SC _k / k-1	CM _k / CM _{dentro}	$\sigma^2 + n \sigma_k^2$
Dentro (error)	N-k	SC _{dentro}	SC _{dentro} / N-k		σ^2
Total	N-1	SC _T			

Los resultados de aplicar las fórmulas anteriores se resumen en una tabla de análisis de varianza como la que se muestra en el cuadro 4.

Cuadro 4: Tabla de análisis de varianza de una vía.

H ₀ : All population means (of Ht) are equal
H ₁ : Not all population means (of Peso) are equal
p-value = 0,02362

ANOVA Table				
Fuente	DF	SS	MS	F
Tratamientos	2	97,4977	48,7489	4,00273
Error	57	694,197	12,1789	
Total	59	791,695		

Software: XLStatistics

Simbología:

DF: grados de libertad SS: suma de cuadrados

MS: cuadrado medio F: Estadístico F

p-value: Probabilidad de que Ho sea verdadera

Comentarios:

1. El cuadrado medio de los tratamientos 48,75 cm² representa la varianza esperada para los tratamientos más el error experimental (σ^2).
2. El *cuadrado medio del error* representa la varianza no explicada esperada para el experimento (σ^2); en este caso 12,18 cm² (recuerde que este número representa una varianza y por lo tanto tiene unidades al cuadrado). Si se desea disminuir el error experimental debe aumentarse el número de repeticiones; algo que no siempre es económico o posible. El valor del cuadrado medio del errores igual a: $\sigma^2 = (S^2_1 + S^2_2 + S^2_3) / T$; en donde T representa el número de tratamientos.
3. El valor de P es 0,02362 y por lo tanto las diferencias se declaran como estadísticamente significativas a un nivel de significancia de 0,05.

4. La prueba F indica la relación entre la *variabilidad explicada* por el experimento (CM_k) y la *variabilidad no explicada o error* (CM_{dentro}). Cuanto mayor sea el valor de F mayor será la probabilidad de que el experimento se declare como significativo; o sea que al menos dos pares de medias sean diferentes.
5. El análisis de varianza presentado asume que existe el mismo número de observaciones por tratamiento ($n=20$ en este caso; diseño balanceado). Cuando no se cuente con el mismo número de observaciones por tratamiento deben hacerse los ajustes necesarios en el cálculo de las medias y de las sumas de cuadrados. Algunos paquetes estadísticos hacen el ajuste en forma automática.

Un método alternativo de realizar una inferencia sobre los tratamientos es calcular sus intervalos de confianza como se muestra a continuación:

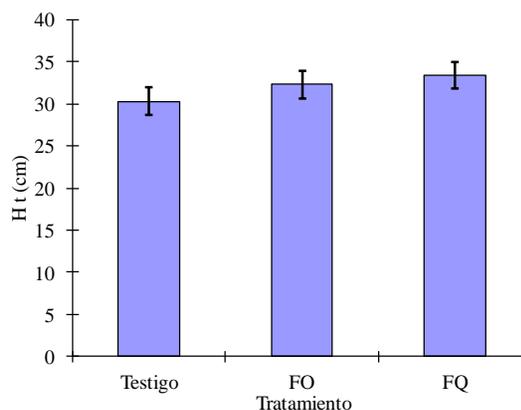
A. IC 95% considerando las varianzas iguales

Category	ME	Lower	Upper
Testigo	1.63	28.70	31.97
FO	1.63	30.67	33.94
FQ	1.63	31.78	35.05

A. IC 95% considerando las varianzas diferentes

Categoría	ME	Lower	Upper
Testigo	1.42	28.91	31.76
FO	1.58	30.73	33.88
FQ	1.87	31.55	35.29

ME: cuadrado medio de error



Note que la elección del método de cálculo tiene un pequeño efecto en la amplitud de los intervalos de confianza. En este caso, lo correcto es utilizar varianzas iguales basados en el resultado de la prueba de Hartley (F_{max}) (ver cuadro 5).

2.3 La prueba F (igualdad de varianzas)

La prueba F fue introducida en el capítulo sobre *prueba de hipótesis* y se utilizó para probar la igualdad entre varianzas obtenidas de dos muestras independientes. El estadístico “F” se denominó así en honor de R. A. Fisher, quien creó el método de análisis de varianza en los años 20s. La razón de F se forma a partir de la comparación de dos estimaciones independientes de la varianza poblacional (σ^2).

$$F = CM_{entre} / CM_{dentro}$$

Cuanto mayor sea el valor de CM_{entre} mayor será el valor de F y menor será el valor del p calculado.

Dado que H_0 sea verdadera, el estadístico F seguirá la distribución teórica de F con $N-k$, $k-1$ grados de libertad; en donde N es el tamaño de la muestra y k el número de tratamientos. Recuerde que la distribución F es positiva; ya que no existen varianzas negativas.

Interpretación de F

Los resultados de la tabla de análisis de varianza (cuadro 4) indican que el valor de F es 4,00273 y que el valor de P (p-value) equivalente para dicho valor de F es 0,02362. Si se utiliza un alfa de 0,05 se debe rechazar H_0 ; ya que el “ p ” calculado es inferior al p crítico. Note sin embargo que si hubiésemos seleccionado previamente un nivel de significancia de 1% ($\alpha=0,01$) no se debería rechazar H_0 . Es refuerza la importancia de definir el nivel de significancia antes de realiza cualquier prueba estadística.

¿Qué sabemos hasta ahora? Al rechazar H_0 sabemos que al menos dos pares de medias son diferentes; sin embargo el ANOVA no nos indica cuáles son. Para responder a esta interrogante debemos hacer una prueba de medias a posteriori como se indica en la sección 2.4.

Los *supuestos de la prueba F* para un análisis de varianza de una vía son los mismos que para una prueba “ t ” de dos muestras independientes; a saber:

- ✓ Muestras independientes y muestreo aleatorio. Esto asegura la validéz de los resultados.
- ✓ Homogeneidad de varianzas ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots$). La prueba F es robusta a la violación de este supuesto siempre y cuando el número de observaciones sea igual por tratamiento. La prueba de Bartlett puede utilizarse para probar por la homogeneidad de las varianzas.
- ✓ Los errores son independientes al interior de los tratamientos y entre tratamientos. Un buen diseño experimental asegura que este supuesto se cumpla.
- ✓ La distribución de errores es normal al interior de cada tratamiento. La prueba F es robusta a la violación de este supuesto siempre y cuando los datos muestren el mismo patrón de asimetría y curtosis. Graficando los datos por tratamiento se puede evaluar este supuesto.

El complemento XLXTatiscs ofrece las gráficas que se muestran en la figura 3 para evaluar los supuestos del análisis de varianza.

Comentarios:

- 1- La gráfica de probabilidad normal y el histograma de residuos indican que los datos no se apartan del supuesto de normalidad.
- 2- Los graficos de residuo para los valores ajustados y por tratamiento indican que la varianza es homogénea para el set de datos. La prueba de igualdad de varianzas de Hartley (F_{max}) (cuadro 5) confirma esta apreciación visual (el p calculado 0,1966 es mayor que p critico 0,05).

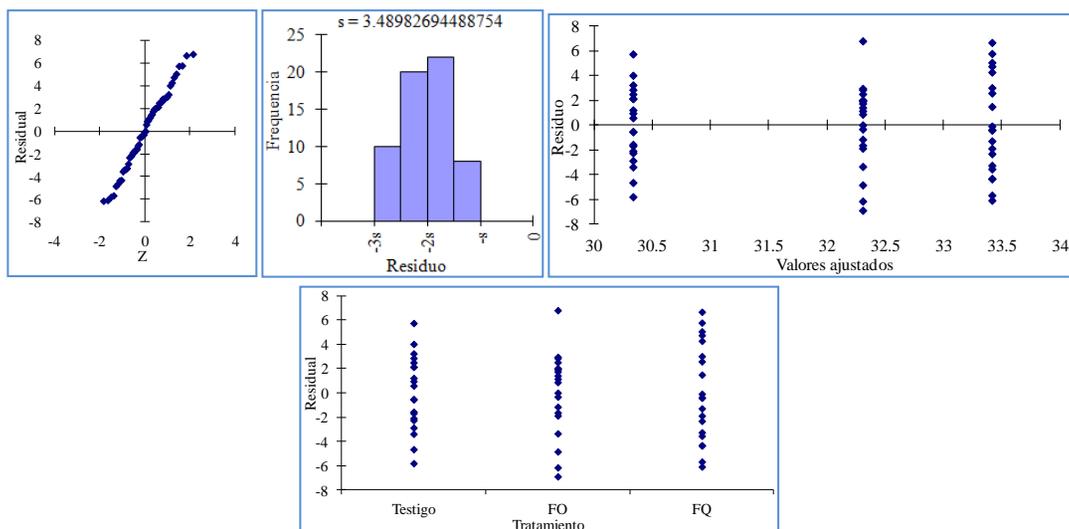


Figura 3: Análisis gráfico de los residuos.

Cuadro 5: Prueba de Barlett para igualdad de varianzas.

Data	
Number of Groups	3
Total Sample Size	60
S_{\max}^2	15,93
S_{\min}^2	9,257

H ₀ : Population variances of all categories equal	
H ₁ : Population variances of all categories not equal	
F _{max}	1,7208
DF ₁	3
DF ₂	19
p-value =	0,1966

Los supuestos de normalidad e igualdad de varianzas son críticos para realizar las pruebas de hipótesis; sin embargo el análisis de varianza es robusto a dichas violaciones; especialmente conforme aumenta el tamaño de la muestra. En la práctica, se pueden aceptar violaciones a dichos supuestos sin invalidar los resultados. También es posible transformar los datos (e.g. raíz cuadrada, logaritmos, recíproco, angular) para hacerlos compatibles con los supuestos del análisis de varianza; sin embargo dichas transformaciones afectará la magnitud de las interacciones y en general se recomiendan con la última opción. Otra alternativa es utilizar pruebas no paramétricas incluido el remuestreo.

2.4 Pruebas a posteriori

Dado que el análisis de varianza indica que al menos uno de las medias es diferente de las otras, procedemos a realizar pruebas a posteriori. A continuación se listan algunas de las pruebas a posteriori más comunes; sin embargo en un caso real el investigador deberá seleccionar aquella que mejor se ajuste a los objetivos de su investigación.

El (la) investigador(a) debe tener claro que cuando se realizan pruebas a posteriori el error tipo I (rechazar H_0 cuando sea verdadera) puede ser muy superior al nominal indicado por el valor de alfa utilizado en las mismas. Las pruebas a posteriori son apropiadas para comparar set de tratamientos que no provienen de un diseño experiemntal (e.g.efecto de ENOS en precipitación media anual) y en especial cuando el objetivo del estudio es realizar un análisis preliminar o cuando se desea seleccionar el mejor tratamiento.

En general, estas pruebas no son apropiadas para ensayos que incluyan tratamientos estructurados tales como concentraciones de un factor (e.g fertilizante, humedad, luz) o diseños factoriales para los cuales la variable respuesta es evaluada para todos las combinaciones de los diferentes niveles de dos o más factores. En estos casos los contrastes planeados son preferidos. Cuando no sea posible aplicar comparaciones planeadas ortogonales (independientes), el intervalo de Bonferroni es el apropiado ya que reparte el error tipo I por igual entre todas las comparaciones.

A. Diferencia mínima significativa irrestricta LSD

Según Saville (1990) esta es la mejor opción para realizar un análisis explotario de las diferencias entre tratamientos. Entre sus ventajas están: simpleza, es concistente, se puede utilizar en diseños no balanceados y con desigualdad de varianzas; su potencia es excelente y su error tipo I es conocido y constante. Cuando se trabaja con diseños con igual numero de observaciones por tratamiento (balanceados), esta prueba es equivalente a la prueba de la diferencia mínima significativa de Fisher para toda comparación de medias de efectos principales.

Se denomina irrestricto porque no requiere de la significancia previa de una prueba F (análisis de varianza). Entre las desventajas se citan: muchas de las comparaciones están correlacionadas o sea son no independientes (sin embargo esto aplica también a los otros métodos) y existe una mayor probabilidad de declarar H_0 como falsa siendo en realidad verdadera (prueba es más liberal, error tipo I es superior al indicado por el valor de alfa).

Otros autores (Little y Hills,1976; Lowry, 1992) recomiendan que esta prueba se utilice sólo para comparaciones independientes y con promedios adyacentes. Es estadístico de prueba es el siguiente:

$$\mathbf{Diferencia: } t_0 = |(y_1 - y_2)| / (2 * S_p^2 / r)^{0.5} = \text{diferencia/error estándar de las diferencias} \quad (13)$$

En donde: \bar{y}_1 : media 1; \bar{y}_2 : media 2

r: número de unidades experimentales que recibe el tratamiento

S_{2p} : Varianza ponderada de las dos medias

$$S_{2p} = [(r_1 - 1) * S_{21} + (r_2 - 1) * S_{22}] / (r_1 + r_2) - 2 \quad (14)$$

El valor absoluto de la expresión 13 se compara con el valor de “t” crítico correspondiente y se declaran como significativas todas aquellas diferencias superiores a dicho valor ($t_{\alpha/2, gl}$).

A continuación se presenta el resultado del análisis realizado con BioEstat² para los datos del archivo cuadro1_bioestat.csv.

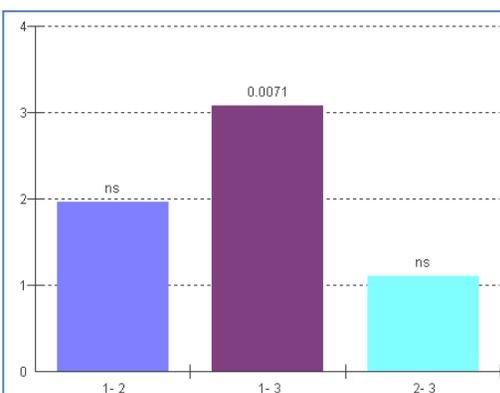
Interface grafica del programa BioEstat.

Columna 1: Testigo (no fertilizante)

Columna 2: Fertilizante orgánico

Columna 3: Fertilizante químico

FONTES DE VARIAÇÃO	GL	SQ	QM
Tratamentos	2	97.498	48.749
Erro	57	694.197	12.179
F =	4.0027	---	---
(p) =	0.0230	---	---
Média (Coluna 1) =	30.3350	---	---
Média (Coluna 2) =	32.3075	---	---
Média (Coluna 3) =	33.4175	---	---
Teste t:	Diferença	t	(p)
Médias (1 e 2) =	1.9725	1.7874	ns
Médias (1 e 3) =	3.0825	2.7932	0.0071
Médias (2 e 3) =	1.1100	1.0058	ns



Comentarios:

1. El programa realiza el análisis de varianza y además calcula un valor de “t” de Estudiante para cada uno de los contrastes o comparaciones. Los datos del análisis de varianza son los mismos obtenidos con XLStatistics.
2. La prueba LSD irrestricta indica que sola la diferencia entre testigo (media 1) y fertilizante químico (media 3) es estadísticamente significativa ($p < 0,01$).

Contraste	Valor (cm)	t de Estudiante	p
Testigo - FO	$ 30.34 - 32.31 = 1.97$	1,7874	NS
Testigo - FQ	$ 30.34 - 33,42 = 3,08$	2,7932	0,0071 (S)
FO - FQ	$ 32.31 - 33.42 = 1,11$	1,0058	NS

NS: no significativo

S: significativo

B. Diferencia mínima significativa restringida de Fisher

La prueba LSD de Fisher es una de las más antiguas (conocida como LSD restringida) y se aplica en dos pasos. Primero, se debe declarar el análisis de varianza para el experimento como significativo y luego se aplica la prueba LSD. Sin embargo según Saville (1990) esta prueba brinda resultados inconsistentes. Esto significa que su veredicto cambiará de experimento a experimento aún cuando no cambien las diferencias entre tratamientos, los grados de libertad ni el cuadrado medio del error. Por lo anterior Saville (1990) recomienda utilizar la *LSD irrestricta* con un alfa de 0.01 para resguardar al investigador contra los errores tipo I. El estadístico de prueba es el siguiente:

$$\text{Diferencia: } t_0 = (\bar{y}_i - \bar{y}_j) / (2\text{CME} / r)^{0.5} \quad (15)$$

En donde: \bar{y}_i : media i; \bar{y}_j : media j

r: número de repeticiones por tratamiento

CME: Cuadrado medio del error obtenido de la tabla de análisis de varianza

El valor absoluto de la expresión 15 se compara con el “t” crítico correspondiente y se declaran como significativas todas aquellas diferencias superiores a dicho valor ($t_{\alpha/2, gl}$). Algunas veces el análisis de varianza será significativo sin embargo el LSD no mostrará diferencias significativas entre tratamientos.

A continuación se presenta el resultado del análisis realizado con InfoStat³ para los datos del archivo anova_cuadro.xls.

Análisis de la varianza

Variable	N	R ²	R ² Aj	CV
Ht (cm)	60	0.12	0.09	10.90

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo.	97.50	2	48.75	4.00	0.0236
Tratamiento	97.50	2	48.75	4.00	0.0236
Error	694.20	57	12.18		
Total	791.69	59			

Test:LSD Fisher Alfa=0.01 DMS=2.94090

Error: 12.1789 gl: 57

Tratamiento Medias n E.E.

FQ	33.42	20	0.78	A
FO	32.31	20	0.78	A B
Testigo	30.34	20	0.78	B

Medias con una letra común no son significativamente diferentes ($p < 0.01$)

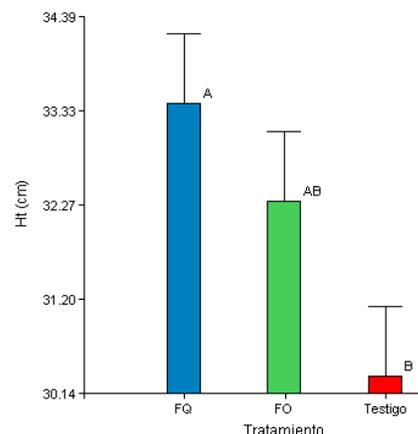
Test:LSD Fisher Alfa=0.05 DMS=2.20988

Error: 12.1789 gl: 57

Tratamiento Medias n E.E.

FQ	33.42	20	0.78	A
FO	32.31	20	0.78	A B
Testigo	30.34	20	0.78	B

Medias con una letra común no son significativamente diferentes ($p < 0.05$)



³ <http://www.infostat.com.ar/>

Comentarios:

1. Se presentan los resultados de LSD para dos niveles de significancia (0,05 y 0,01). Note que al cambiar el valor de alfa, cambia el valor de DMS (diferencia mínima significativa). Para un alfa de 0,01 se requiere una diferencia superior a 2,94 cm entre dos tratamientos para declarar el contraste como significativo, en tanto que para un alfa de 0,05 solo se requiere de una diferencia de 2,12cm.
2. Observe que los datos del análisis de varianza son los mismos obtenidos con XLStatistics y BioEstat.
3. El error estándar (EE) es igual a $(CME / n)^{0,5} = (12,18 / 20)^{0,5} = 0,78$
4. La prueba LSD indica que la media del fertilizante químico (FQ) y del orgánico (FO) es igual y que la media del fertilizante orgánico es igual a la media del testigo. Observe que solo el testigo es diferente del fertilizante químico. Las diferencias son las siguientes:

Comparación	Valor (cm)	DMS (0,05) (cm)	DMS (0,01) (cm)
Testigo - FO	$ 30.34 - 32.31 = 1,97$ NS	2,21	2,94
Testigo - FQ	$ 30.34 - 33,42 = 3,08$ S	2,21	2,94
FO - FQ	$ 32.31 - 33.42 = 1,11$ NS	2,21	2,94

NS: no significativo

S: significativo

5. El resultado de la prueba LSD restringida es consistente con el de la prueba LSD irrestricta.

C. Prueba de rangos múltiples de Duncan

En este caso las comparaciones entre las medias no se hacen utilizando una única diferencia crítica como se hace por ejemplo en una prueba t de Estudiante o en la LSD de Fischer. La prueba de Duncan ajusta la diferencia crítica considerando si los dos promedios que se comparan son adyacentes o sí por el contrario existe uno o más medias entre las medias que se están comparando.

La prueba de Duncan tiene una consistencia moderada; sin embargo no debe utilizarse cuando las varianzas son heterogéneas. Sin embargo Lehmann y Shaffer (1977) argumenta que dicha prueba no debe utilizarse porque la derivación de las tasas de error dependen de una condición de monotocidad en los datos (siempre crecen o decrecen; pero no oscilan); la cual no se presenta en datos reales. La fórmula para la prueba de Duncan es:

$$\text{Diferencia crítica: } K_r (CME \text{ dentro de grupos} / n)^{0,5} \quad (16)$$

El valor de K_r se obtiene de la tabla de comparaciones de rangos múltiples de Duncan y n corresponde al tamaño medio de las muestras. A continuación se presenta el resultado del análisis realizado con InfoStat para los datos del archivo anova_cuadro1.cvs.

Test:Duncan Alfa=0.05

Error: 12.1789 gl: 57

Tratamiento	Medias	n	E.E.	
FQ	33.42	20	0.78	A
FO	32.31	20	0.78	A B
Testigo	30.34	20	0.78	B

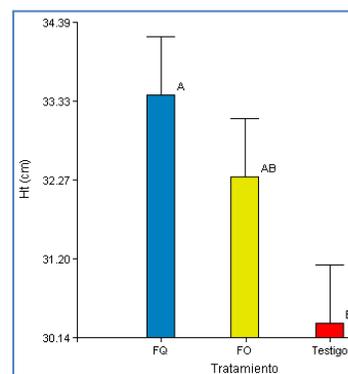
Medias con una letra común no son significativamente diferentes ($p \leq 0.05$)

Test:Duncan Alfa=0.01

Error: 12.1789 gl: 57

Tratamiento	Medias	n	E.E.	
FQ	33.42	20	0.78	A
FO	32.31	20	0.78	A B
Testigo	30.34	20	0.78	B

Medias con una letra común no son significativamente diferentes ($p \leq 0.01$)



Comentarios:

1. Se presentan los resultados de la prueba de Duncan para dos niveles de significancia (0,05 y 0,01). La significancia de los contrastes no cambia con el nivel de significancia.
2. El error estándar (EE) es igual a $(CME / n)^{0,5} = (12,18 / 20)^{0,5} = 0,78$
3. La prueba de Duncan indica que la media del fertilizante químico (FQ) y del orgánico (FO) es igual y que la media del fertilizante orgánico es igual a la media del testigo. Observe que solo el testigo es diferente del fertilizante químico. Las diferencias son las siguientes:

Comparación	Valor (cm)	Resultado
Testigo - FO	$ 30.34 - 32.31 = 1,97$	NS
Testigo - FQ	$ 30.34 - 33,42 = 3,08$	S
FO - FQ	$ 32.31 - 33.42 = 1,11$	NS

NS: no significativo

S: significativo

D. Prueba de Student-Newman-Keuls (S.N.K)

La prueba de Student-Newman-Keuls utiliza una filosofía similar a la de Duncan para calcular el valor de la diferencia crítica; sin embargo utiliza otra tabla para obtener los valores de los rangos críticos. En opinión de los estadísticos la tabla de Newman-Keul y de rangos múltiples de Tukey son más defendibles desde el punto de vista matemático que la de Duncan y por lo tanto los resultados obtenidos son más confiables. Esta prueba no debe utilizarse cuando las varianzas no sean homogéneas. La prueba opera de la siguiente manera:

1. Prueba por la homogeneidad de las k medias a un nivel de significación α_k . Si se rechaza H_0 (las medias son homogéneas) se procede con la siguiente prueba, de lo contrario se detiene el proceso.
2. El paso dos es probar por la homogeneidad de cada subconjunto de $k-1$ medias, usando un nivel de significación α_{k-1} , si no se rechaza H_0 , el procedimiento se detiene.

3. El procedimiento se repite con $k-2$ medias y así sucesivamente hasta que no sea posible rechazar H_0 (o se las medias son homogéneas).

El nivel de significación en la etapa i -ésima es: $\alpha_i = 1 - (1 - \alpha)^{i-1}$; el método controla tasa de error por comparación sin embargo no la tasa de error por experimento. El incremento en la tasa de error puede conducir a una disminución del error de tipo II, razón por la cual algunos autores afirman que esta prueba es más potente.

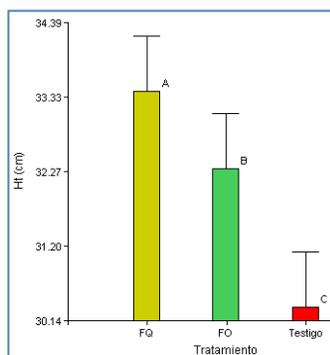
La fórmula para la prueba de Newman-Keuls es:

$$\text{Diferencia crítica: } q_r \left(\frac{\text{CME dentro de grupos}}{n} \right)^{0,5} \quad (17)$$

El valor de “q” se obtiene de la tabla de comparaciones múltiples de Newman-Keul y rangos múltiples de Tukey. A continuación se presenta el resultado del análisis realizado con InfoStat para los datos del archivo anova_cuadro1.cvs.

```
Test:SNK Alfa=0.05
Error: 12.1789 gl: 57
Tratamiento Medias n E.E.
FQ          33.42 20 0.78 A
FO          32.31 20 0.78 B
Testigo     30.34 20 0.78 C
Medias con una letra común no son significativamente diferentes (p<= 0.05)
```

```
Test:SNK Alfa=0.01
Error: 12.1789 gl: 57
Tratamiento Medias n E.E.
FQ          33.42 20 0.78 A
FO          32.31 20 0.78 B
Testigo     30.34 20 0.78 C
Medias con una letra común no son significativamente diferentes (p<= 0.01)
```



Comentarios:

1. Se presentan los resultados de la prueba de Duncan para dos niveles de significancia (0,05 y 0,01). La significancia de los contrastes no cambia con el nivel de significancia.
2. El error estándar (EE) es igual a $(\text{CME} / n)^{0,5} = (12,18 / 20)^{0,5} = 0,78$.
3. Para un alfa de 0,05 y 0,01, la prueba de Student-Newman-Keuls declara como significativos todos los contrastes.

Comparación	Valor (cm)	Resultado
Testigo - FO	$ 30.34 - 32.31 = 1,97$	S
Testigo - FQ	$ 30.34 - 33,42 = 3,08$	S
FO - FQ	$ 32.31 - 33.42 = 1,11$	S

NS: no significativo

S: significativo

Con fines ilustrativos, a continuación se muestra el resultado de realizar tres pruebas de hipótesis por separado utilizando XLStatistics. Para un alfa de 0,05 solo los contrastes testigo Vs fertilizantes son significativos. En otras palabras, esta prueba no detecta una diferencia significativa entre la aplicación de fertilizante orgánico y químico.

Hypothesis Tests		Hypothesis Tests		Hypothesis Tests	
H ₀ : $\mu_1 - \mu_2 = 0$		H ₀ : $\mu_1 - \mu_2 = 0$		H ₀ : $\mu_1 - \mu_2 = 0$	
Alternative <input type="radio"/> ≠ <input checked="" type="radio"/> > <input type="radio"/> <		Alternative <input type="radio"/> ≠ <input checked="" type="radio"/> > <input type="radio"/> <		Alternative <input type="radio"/> ≠ <input checked="" type="radio"/> > <input type="radio"/> <	
H ₁ : $\mu_1 - \mu_2 > 0$		H ₁ : $\mu_1 - \mu_2 > 0$		H ₁ : $\mu_1 - \mu_2 > 0$	
T	1.943237	T	2.746833	T	0.950425
DF	38	DF	38	DF	38
p-value	= 0.029711	p-value	= 0.004574	p-value	= 0.173951
F. orgánico > testigo		F. químico > testigo		F. químico > F. Orgánico	

E. Tukey

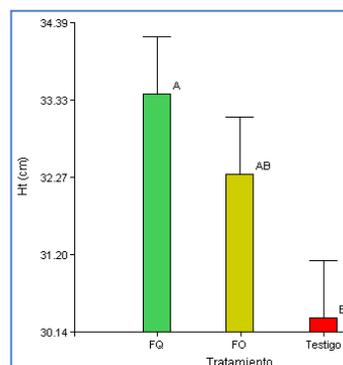
La prueba de Tukey es similar a una prueba t de Estudiante en cuanto a que se calcula una única diferencia crítica para realizar todas las comparaciones entre las medias; sin embargo es también similar a la prueba de Duncan y de Newman-Keuls en cuanto a que el valor de esta diferencia crítica depende del número de comparaciones que se haga.

Una desventaja de la prueba de Tukey es que es la más inconsistente de entre las pruebas para realizar comparaciones múltiples. Esto significa que su veredicto cambiará de experimento a experimento aún cuando no cambien las diferencias entre tratamientos, los grados de libertad ó el cuadrado medio del error (Saville, 1990). El valor de q_r se obtiene de la tabla de Newman-Keul y de rangos múltiples de Tukey. La fórmula para la prueba de Tukey es:

$$\text{Diferencia honestamente significativa: HSD} = q_r (\text{CME dentro de grupos} / n)^{0,5} \quad (18)$$

Esta prueba no es apropiada cuando existen grandes diferencias en el tamaño de la muestra por tratamiento. A continuación se presenta el resultado del análisis realizado con InfoStat para los datos del archivo anova_cuadro1.csv.

Test: Tukey Alfa=0.05 DMS=2.65568			
Error: 12.1789 gl: 57			
Tratamiento	Medias	n	E.E.
FQ	33.42	20	0.78 A
FO	32.31	20	0.78 A B
Testigo	30.34	20	0.78 B
Medias con una letra común no son significativamente diferentes (p<= 0.05)			
Test: Tukey Alfa=0.01 DMS=3.34849			
Error: 12.1789 gl: 57			
Tratamiento	Medias	n	E.E.
FQ	33.42	20	0.78 A
FO	32.31	20	0.78 A
Testigo	30.34	20	0.78 A
Medias con una letra común no son significativamente diferentes (p<= 0.01)			



Comentarios:

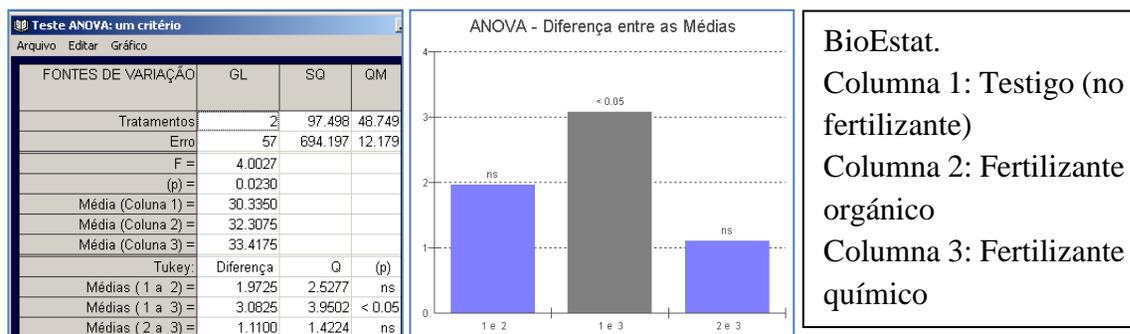
1. Se presentan los resultados de la prueba de Duncan para dos niveles de significancia (0,05 y 0,01). La significancia de los contrastes cambia con el nivel de significancia.
2. El error estándar (EE) es igual a $(CME / n)^{0,5} = (12,18 / 20)^{0,5} = 0,78$
3. Para un alfa de 0,05 la prueba de Tukey declara como significativa únicamente la diferencia entre la media del testigo y del fertilizante químico.

Comparación	Valor (cm)	Resultado
Testigo - FO	$ 30.34 - 32.31 = 1,97$	NS
Testigo - FQ	$ 30.34 - 33,42 = 3,08$	S
FO - FQ	$ 32.31 - 33.42 = 1,11$	NS

NS: no significativo S: significativo

4. Para un alfa de 0,01 la prueba de Tukey declara como no significativos ninguno de los contrastes (Ver intervalos de Bonferroni).

Con fines ilustrativos, a continuación se muestra el resultado de la prueba realizada con el programa BioEstat para un alfa de 0,05. La conclusión es que solo el contraste testigo Vs fertilizante químico alcanzó un valor de p inferior a 0,05.



F. Scheffé

La prueba de Scheffé es similar a la prueba de Tukey pues se calcula un único valor crítico para realizar las comparaciones entre las medias. La principal diferencia entre Scheffé y los otros métodos de comparaciones múltiples es que utiliza la tabla F y no las tablas de rangos estandarizados según la distribución “t” de estudiante de las otras pruebas.

La prueba de Scheffé es la más rigurosa (conservadora) y por lo tanto la probabilidad del error tipo I es menor (rechazar H_0 cuando es verdadera); sin embargo es poco utilizado en la vida real (Saville, 1990). Una ventaja de Scheffé es que puede utilizarse para realizar comparaciones que involucran más de dos medias; por ejemplo, testigo versus abono1 y abono2. La fórmula para la prueba de Scheffé es:

$$\text{Diferencia crítica: } ((K-1) * F_{df N-1, df error})^{0,5} (2 * \text{CME dentro de grupos} / n)^{0,5} \quad (19)$$

Donde, K es el número de grupos o tratamientos y F es el valor de la tabla F con los respectivos grados de libertad. Esta prueba no es apropiada cuando existen grandes diferencias en el tamaño de la muestra por tratamiento. Ni Infostat ni BioEstat permiten calcular esta prueba.

G. Dunnett (comparación con un control)

La prueba debe utilizarse cuando se quiere comparar la media del grupo control (o sea aquel grupo al que no se le aplica tratamiento) con las medias de los grupos experimentales (grupos que reciben tratamiento). El método de cálculo de la prueba de Dunnett es similar a una prueba *t* y a la prueba de Scheffe. Dunnett utiliza una única diferencia crítica para realizar las comparaciones múltiples. La prueba de Dunnett no requiere que la prueba F del análisis de varianza sea significativa para aplicarla. La fórmula para la prueba de Dunnett es:

$$\text{Diferencia crítica: } d_r (\text{CME dentro de grupos} / n)^{0,5} \quad (20)$$

El valor de d_r se obtienen de la tabla de Dunnett y corresponde a : $t_{D_{\alpha/2; k, v}}$; endo donde K es el número de tratamientos (incluido el control) y “v” los grados de libertad del CME.

A continuación se muestra el resultado de la prueba realizada con el programa BioEstat para un alfa de 0,05.

Configuración de la ventana de diálogo. Asegúrese que la primera columna contiene los datos del testigo.

Seleccione una prueba unilateral (de una cola). Se seleccionó un alfa de 0,05 para la prueba.

Teste ANOVA: um critério			
Arquivo Editar Gráfico			
FONTES DE VARIAÇÃO	GL	SQ	QM
Tratamentos	2	97.498	48.749
Erro	57	694.197	12.179
F =	4.0027		
(p) =	0.0230		
Média (Coluna 1) =	30.3350		
Média (Coluna 2) =	32.3075		
Média (Coluna 3) =	33.4175		
Dunnett: Diferença	SE	Q	V Crítico
Colunas 1 e 2	1.9725	1.1036	1.6915
Colunas 1 e 3	3.0825	1.1036	2.7932
			Conclusão
			p < 0.05
			p < 0.05

Comentarios:

1. El error estándar (EE) es igual a $(\text{CME} / n)^{0,5} = (12,18 / 20)^{0,5} = 0,78$.

2. Para un alfa de 0,05 la prueba de Dunnett declara como significativas las diferencias entre testigo y fertilizante orgánico y testigo y fertilizante químico.

Comparación	Valor (cm)	Resultado
Testigo - FO	$ 30.34 - 32.31 = 1.97$	$P < 0,05$ S
Testigo - FQ	$ 30.34 - 33,42 = 3,08$	$P < 0,05$ S

NS: no significativo

S: significativo

H. Intervalo de Bonferroni o corrección de Bonferroni

Como ya se ha mencionado, la principal limitante de las pruebas múltiples es controlar el error tipo I. La corrección de Bonferroni se sustenta en la idea de que si usted somete a prueba "n" hipótesis dependientes o independientes (pareadas) para un mismo set de datos, la forma de mantener la tasa de error constante para el estudio es probar cada hipótesis individual a un nivel de significancia $1/n$ veces más de lo que sería si sólo se sometiera a prueba una hipótesis.

El objetivo es mantener igual el nivel de significancia (α) para toda la familia de pruebas y por lo tanto la corrección de Bonferroni somete a prueba cada uno de los contrastes a un nivel de significancia de α/n . Por ejemplo, si se realizan tres contrastes simultáneos y se desea mantener el nivel de significancia en 0,05; el alfa elegido para la prueba debe ser $0,05/3 = 0,0166$.

La prueba permite controlar el error tipo I de los contrastes de manera simultánea utilizando la siguiente expresión: $c=k(k-1)/2$; donde k es el número de tratamientos a comparar. El nivel de significancia de cada contraste es ajustado de acuerdo al número de comparaciones realizadas.

A continuación se presenta el resultado del análisis realizado con InfoStat para los datos del archivo anova_cuadro1.cvs.

```

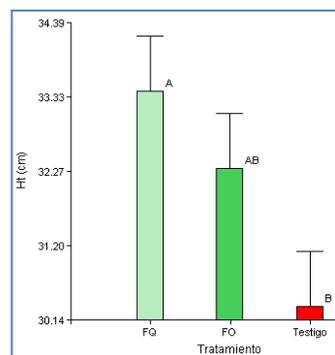
Test:Bonferroni Alfa=0.05 DMS=2.72219
Error: 12.1789 gl: 57
Tratamiento Medias n E.E.
FQ          33.42 20 0.78 A
FO          32.31 20 0.78 A B
Testigo     30.34 20 0.78 B
Medias con una letra común no son significativamente diferentes (p<= 0.05)

```

```

Test:Bonferroni Alfa=0.01 DMS=3.38131
Error: 12.1789 gl: 57
Tratamiento Medias n E.E.
FQ          33.42 20 0.78 A
FO          32.31 20 0.78 A
Testigo     30.34 20 0.78 A
Medias con una letra común no son significativamente diferentes (p<= 0.01)

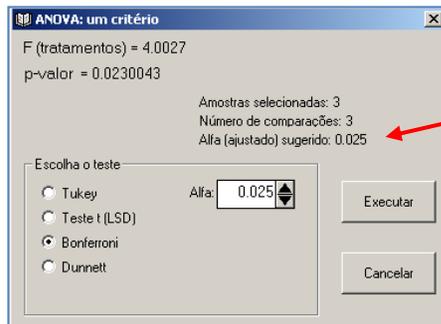
```



Comentarios:

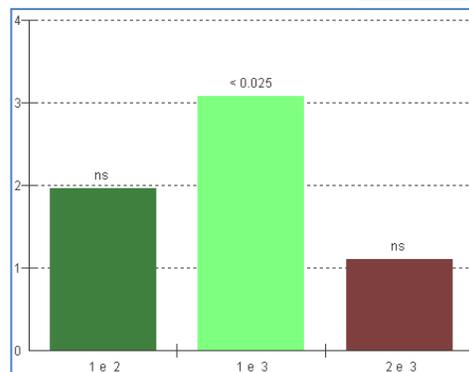
1. Para un alfa de 0,05, la corrección de Bonferroni declara como significativa solo la diferencia entre el testigo y fertilizante químico. Para el alfa 0,01 no declaró ningún contraste como significativo.

Con fines ilustrativos, a continuación se muestra el resultado de la prueba realizada con el programa BioEstat para un alfa de 0,025. La conclusión es que solo el contraste testigo Vs fertilizante químico alcanzó un valor de p inferior a 0,025.



Para tres comparaciones, el programa sugiere utilizar un alfa de 0,025.

FONTES DE VARIAÇÃO	GL	SQ	QM
Tratamentos	2	97.498	48.749
Erro	57	694.197	12.179
F =	4.0027		
(p) =	0.0230		
Média (Coluna 1) =	30.3350		
Média (Coluna 2) =	32.3075		
Média (Coluna 3) =	33.4175		
Bonferroni:	Diferença	E	(p)
Médias (1 e 2) =	1.9725	1.7874	ns
Médias (1 e 3) =	3.0825	2.7932	< 0.025
Médias (2 e 3) =	1.1100	1.0058	ns



Resumen

Como ya se mencionó, la filosofía y rigurosidad estadística de cada uno de los métodos para realizar comparaciones múltiples es diferente; sin embargo, para el ejemplo analizado, la mayoría de los métodos solo detectó diferencias significativas entre la media del testigo y el fertilizante químico ($p < 0,05$). Por lo tanto, se puede concluir que no existe una diferencia significativa ($p > 0,05$) entre los métodos de fertilización. Solo la prueba de Dunnett declaró como significativa ($p < 0,05$) dicha diferencia.

Conclusiones:

- Los resultados del análisis de varianza indican que no existe diferencia en el crecimiento de las plántulas a las cuales se les aplicó el abono orgánico con respecto a aquellas a las cuales se les aplicó el abono químico a un nivel de significancia de 0,05%.
- El crecimiento de los árboles con abono químico o orgánico fue superior al de los árboles sin abono a un nivel de significancia de 0,05.
- Dados los resultados, no es posible saber cuál de los dos abonos es el mejor.

2.4.1. Limitaciones de las pruebas a posteriori

Las pruebas a posteriori deben aplicarse sólo cuando la prueba F del análisis de varianza indica que existen diferencias entre algunas de las medias de los tratamientos. Esto minimiza la probabilidad de incurrir en un error tipo I (rechazar H_0 cuando es verdadera). Las diferencias

críticas o significativas en las pruebas a posteriori son mayores comparadas con las pruebas planeadas con el objeto de evitar rechazar Ho cuando esta es verdadera. En este sentido, estas pruebas tienden a ser conservadoras.

Las comparaciones planeadas son aquellas que se definen antes de realizar el experimento y por lo tanto no requieren de un valor de F significativo previo a su aplicación. Para un experimento con “c” comparaciones independientes entre medias, la probabilidad de obtener al menos una comparación significativa por efecto de azar está dada por:

$$1 - (1 - \alpha)^c \tag{21}$$

Lo que es aproximadamente igual a $C \cdot \alpha$ para valores pequeños de α . Si comparamos la media menor y mayor de un experimento (usualmente el testigo versus el mejor tratamiento) se esperarían los siguientes resultados en función del número de comparaciones:

No. de tratamientos	alfa nominal	% de veces en que el valor tabulado será excedido y por lo tanto Ho rechazada (error tipo I)
3	0.05	13
6	0.05	40
10	0.05	60
20	0.05	90

Fuente: Kirk, 1968.

Para evitar cometer errores tipo I má allá de lo deseado por el investigador (a), siempre que sea posible se debe optar por las pruebas planeadas. En opinión de algunos autores el método de Duncan muestra un buen control del error tipo I; su potencia es superior a Newman-Keuls, Tukey y Scheffe (Snedecor y Cochran, 1980); además es un método con una consistencia moderada y en la práctica se muestra mucho más consistente que las otras pruebas múltiples (Saville, 1990). A continuación se ordenan las pruebas para comparaciones múltiples de acuerdo a su tendencia a cometer errores tipo I (rechazar Ho cuando es verdadera).

<p>Liberales</p> <p>LSD-----Duncan-----NK-----Tukey-----Scheffé</p> <p>Liberales</p> <p>Se detecta más diferencias y diferencias pequeñas son declaradas como significativas.</p>	<p>Conservadoras</p> <p>Conservadoras</p> <p>Se detecta menos diferencias y diferencias pequeñas no son declaradas como significativas</p>
--	---

2.5. Pruebas a priori o planeadas

En la sección anterior se mencionó que se recomienda utilizar las pruebas a posteriori con fines exploratorios; o sea en aquellos estudios en los que no es posible definir las hipótesis que se desean someter a prueba antes de realizar el experimento.

Se espera que en la mayoría de los experimentos el o la investigadora tenga una idea clara de cuáles comparaciones son de interés y cuáles no lo son. Una de las propiedades de los contrastes planeados es que son independientes y por lo tanto no padecen de las limitaciones mencionadas para las pruebas a posteriori (error tipo I).

Dado un experimento no "T" tratamientos existes "T-1" contrastes o comparaciones independientes (ortogonales). Por ejemplo, en el caso anterior con tres tratamientos (tipos de fertilizantes) solo sería posible hacer dos comparaciones; una de ellas puede ser el efecto del fertilizante orgánico con respecto al químico (equivalente a una prueba t de dos poblaciones) y la otra, la respuesta del testigo con respecto al fertilizante químico o al orgánico. Note sin embargo que este caso la comparación de interés debe ser entre los fertilizantes (orgánico vs químico).

Un contraste es una combinación lineal de medias en el cual la suma de los coeficientes de cada una de las medias suma cero.

$$\sum C_i \mu_i \text{ en donde } \sum C_i = 0 \quad (22)$$

Al diseñar una tabla de coeficientes de comparación, utilice los siguientes criterios como una guía:

1. Recuerde que siempre se realizará una comparación entre dos grupos o medias. Si ambos grupos tienen el mismo número de tratamientos; asigne el valor 1 a uno de los grupos y el valor -1 al otro grupo. Por ejemplo, si se desea comparar el efecto del fertilizante orgánico versus el efecto del fertilizante químico, asigne un 1 a los tratamientos que incluyen el fertilizante químico y un -1 a los tratamientos que incluyen el fertilizante orgánico; note que la suma de los coeficientes es igual a cero.
2. Cuando los grupos que se comparan no tienen el mismo número de tratamientos (e.g. tres tratamientos con factor A Vs dos tratamientos con factor B) asígnele al primer grupo un coeficiente igual al número de tratamientos del segundo grupo y al segundo grupo un valor igual al número de tratamientos del primer grupo pero de signo opuesto. Por ejemplo, si tenemos tres tratamientos (A,B,C) y deseamos comparar los primeros dos (A y B) versus el tercero (C) debemos asignar los coeficientes así: +1,+1, -2; note que la suma de los coeficientes es igual a cero.
3. Aplique un divisor común a todos los coeficientes de tal forma que su valor se reduzca al entero más pequeño posible. Por ejemplo, dado un experimento factorial de tres niveles del factor A y dos niveles del factor B, tenemos seis tratamientos (A, B, C, D, E, F). Si se desea comparar el efecto de los tratamientos A y B versus los tratamientos C, D, E, F; los coeficientes deberían ser: +4 +4 -2 -2 -2 -2; sin embargo estos valores pueden dividirse entre dos y reducir los coeficientes a: +2 +2 -1 -1 -1 -1; note que la suma de los coeficientes es igual a cero. A continuación se muestra otro ejemplo:

$$\begin{array}{cccccc} & \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \mathbf{E} & \mathbf{F} \\ \text{A versus B, C, D, E, F} & 5 & -1 & -1 & -1 & -1 & -1 \end{array} \quad \Sigma C_i = 5 + (-1) + (-1) + (-1) + (-1) + (-1) = 0$$

$$\begin{array}{cccccc} & \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \mathbf{E} & \mathbf{F} \\ \text{A y B versus C, D, E} & -3 & -3 & 2 & 2 & 2 & 0 \end{array} \quad \Sigma C_i = -3 + (-3) + 2 + 2 + 2 = 0$$

Nota: observe que F es cero porque no se incluye en la comparación.

$$\begin{array}{cccccc} & \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \mathbf{E} & \mathbf{F} \\ \text{C y D versus E, F} & 0 & 0 & -2 & -2 & 2 & 2 \end{array} \quad \Sigma C_i = -2 + (-2) + 2 + 2 = 0$$

Nota: observe que A y B son cero porque no se incluye en la comparación.

4. Si desea probar por el efecto de la interacción entre cualquier par de comparaciones simplemente multiplique los coeficientes correspondientes de los efectos principales. Para que un conjunto de comparaciones planeadas sean independientes u ortogonales es necesario que cumplan con las siguientes condiciones:
- La suma de los coeficientes de las comparaciones debe ser igual a cero. Observe que la matriz de coeficientes planteada en el numeral tres cumple con esta condición.
 - La suma de los productos para cualquier par de comparaciones debe ser igual a cero. Observe que la matriz de coeficientes planteada en el numeral tres para A versus B, C, D, E, F no cumple con esta condición.

Ejemplo: A continuación se muestra el resultado de dos contrastes utilizando InfoStat:

Análisis de la varianza

Variable	N	R ²	R ² Aj	CV
Ht (cm)	60	0.12	0.09	10.90

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo.	97.50	2	48.75	4.00	0.0236
Tratamiento	97.50	2	48.75	4.00	0.0236
Error	694.20	57	12.18		
Total	791.69	59			

Contrastes

Tratamiento	Contraste	SC	gl	CM	F	p-valor	
Contrastel		1.97	38.91	1	38.91	3.19	0.0792
Total		38.91	1	38.91	3.19	0.0792	

Coefficientes de los contrastes

Tratamiento	Ct.1
FO	1.00
FQ	0.00
Testigo	-1.00

Análisis de la varianza

Variable	N	R ²	R ² Aj	CV
Ht (cm)	60	0.12	0.09	10.90

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo.	97.50	2	48.75	4.00	0.0236
Tratamiento	97.50	2	48.75	4.00	0.0236
Error	694.20	57	12.18		
Total	791.69	59			

Contrastes

Tratamiento	Contraste	SC	gl	CM	F	p-valor	
Contrastel		3.08	95.02	1	95.02	7.80	0.0071
Total		95.02	1	95.02	7.80	0.0071	

Coefficientes de los contrastes

Tratamiento	Ct.1
FO	0.00
FQ	1.00
Testigo	-1.00

Testigo versus fertilizante orgánico

Matriz de contraste

NF	FO	FQ
-1	1	0

Para un alfa de 0,05 no se rechaza

Ho.

Testigo versus fertilizante químico

Matriz de contraste

NF	FO	FQ
-1	0	1

Para un alfa de 0,01 se rechaza Ho.

El valor de p para la prueba de hipótesis de dos medias independientes es 0,0045.

Si usted necesita hacer un número pequeño de contrastes y algunos de ellos no son independientes, puede utilizar el intervalo de Bonferroni; el cual depende del número de comparaciones que se desea hacer, del nivel de significancia seleccionado y de los grados de libertad del cuadrado medio del error.

La prueba de Bonferroni utiliza una prueba “t” de Estudiante para realizar comparaciones múltiples; sin embargo a diferencia de los otros métodos presentados previamente, permite controlar la tasa de error general del estudio mediante la asignación de un error a cada comparación igual al error general del experimento dividido entre el número de comparaciones. De esta forma el nivel de significancia es ajustado según el número de comparaciones realizadas.

3. Diseño de bloques completos al azar

Este diseño experimental es equivalente a un muestreo estratificado en donde los tratamientos se asignan al azar a un conjunto de unidades experimentales denominado bloques; en algunos textos de estadística también se les denomina repeticiones (e.g. Steel y Torrie, 1980); sin embargo el primer término es más apropiado para evitar la confusión con las repeticiones de un diseño completamente al azar.

Los adjetivos *completo* y *al azar* indican que cada bloque contiene todos los tratamientos y que éstos son asignados al azar al interior del bloque. También se puede tener un diseño de *bloques balanceados incompletos al azar*; en este caso uno de los tratamientos es omitido en cada bloque; sin embargo en todos los casos el efecto de los tratamientos y sus diferencias son estimados con la misma precisión; ya que todos los bloques tienen igual número de repeticiones. Por ejemplo, si se tienen cuatro tratamientos (a, b, c y d) se puede omitir uno por bloque como se muestra a continuación:

Bloque	Tratamientos	tratamiento omitido
1	b a c	d
2	a b d	c
3	a c d	b
4	b c d	a

El objetivo del diseño es seleccionar los bloques de tal forma que su variabilidad interna sea mínima, a la vez que se maximiza la variabilidad entre los bloques. En otras palabras, las unidades experimentales de un bloque deben ser lo más similares posibles entre sí pero a la vez lo más disimilares con respecto a los otros bloques.

El éxito del diseño depende de la capacidad del investigador(a) de formar bloques homogéneos y a la vez de aplicar los tratamientos por bloque en forma uniforme. Uno de los supuestos del diseño de bloques al azar es que el efecto de los bloques es independiente del efecto de los otros factores (tratamientos) y por lo tanto en este tipo de diseño no interesa la interacción entre tratamientos y bloques; aun cuando el investigador(a) puede calcularlo si lo desea. El diseño podría analizarse como un experimento de dos factores sin repeticiones (Manly, 1992).

En el caso de experimentos de campo que involucren parcelas (e.g. plantaciones, respuesta a raleos, respuesta a fertilización) es recomendable ubicar las parcelas de un bloque lo más cercanas una de las otras; ya que esto tiende a reducir la variabilidad del bloque. En caso de presentarse una gradiente en el área de estudio (e.g. pendiente, pluviosidad, humedad), los bloques deben ubicarse a lo largo de la gradiente y las parcelas por bloque deben ubicarse paralelas al gradiente (Fig. 4).

Dado el espacio físico requerido por este tipo de diseño con frecuencia se incluyen pocos tratamientos y una única repetición. Aunque no se requiere que cada bloque tenga la misma forma sí es una característica deseable para evitar incluir una fuente de variabilidad adicional en el diseño.

La asignación de tratamientos por bloque debe ser aleatoria; por ejemplo, si se tienen tres tratamientos por bloque usted puede enumerar sus tratamientos de uno a cuatro y luego utilizar una tabla de números al azar para asignar los tratamientos por bloque como se ilustra a continuación:

1	2	3	4	5
C	A	C	B	B
D	C	D	C	D
A	D	B	A	C
B	B	A	D	A

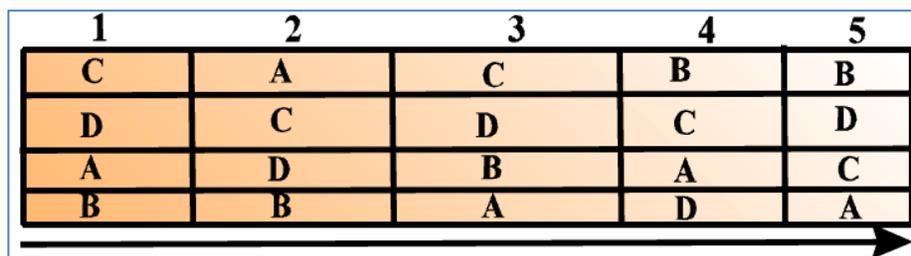


Figura 4: Ubicación de bloques y unidades experimentales en un terreno con una gradiente. La flecha indica el sentido de la gradiente. Los tratamientos A, B, C, D fueron asignados en forma aleatoria en cada uno de los bloques (1 a 5).

En síntesis, bajo condiciones de campo este diseño experimental permite obtener estimaciones más precisas que un diseño completamente al azar. El diseño permite incluir cualquier número de tratamientos y repeticiones; también es posible incluir dos o más repeticiones de un tratamiento por bloque.

En presencia de datos perdidos puede utilizarse la técnica de la parcela perdida desarrollada por Yates para realizar el análisis. Este diseño no es apropiado cuando se espera un alto número de datos perdidos en el estudio; bajo estas condiciones es mejor utilizar un diseño completamente al azar.

A continuación se muestra cómo se calculan los diferentes componentes del análisis de varianza x dos vías.

3.1. ANDEVA de dos vías

La ANDEVA de dos vías permite evaluar la respuesta del grupo experimental al efecto combinado de dos factores. Normalmente cada factor provee dos o más tratamientos. Por ejemplo, podemos evaluar la respuesta en crecimiento de plántulas de pochote a los siguientes factores:

- ✓ Factor 1: dos formulaciones de fertilizante (1 y 2). Cada formulación es un tratamiento.
- ✓ Factor 2: tres tipos de envases (A,B,C). Cada tipo de envase es un tratamiento.

El efecto de los tratamientos anteriores podría analizarse realizando dos experimentos por separado y utilizando un diseño de varianza de una vía (irrestricto al azar) en cada caso. Sin embargo, esto no permitiría saber si existe una interacción entre los métodos de fertilización y los tipos de envases.

Al igual que en el caso del ANDEVA de una vía, en el análisis de dos vías se divide la variabilidad total del experimento en diversas fuentes de variación (estimaciones independientes de varianza) y luego se realizan las pruebas F para determinar la significancia de los tratamientos principales (e.g. fertilizante y tipo de envase) y la interacción (fertilizante * tipo de envase).

En el análisis de varianza de dos vías el investigador(a) debe tener claridad del modelo de ANDEVA que está utilizando. Por ejemplo, ¿representan los niveles de los factores una muestra de “n” posibles niveles? o por el contrario ¿son los únicos niveles de importancia para el experimento? La respuesta a esta pregunta permite definir los factores o efectos principales como *fijos* o *aleatorios*.

Factor fijo: Los niveles incluidos en el experimento son los únicos relevantes para la investigación. Por ejemplo, extracción con bueyes, chapulín y tractor de orugas; medición de diámetros con forcípula y cinta diamétrica; medición de pH con dos instrumentos.

Factor aleatorio: Los niveles del factor incluidos en el experimento representan una muestra aleatoria de “n” posibles niveles. Por ejemplo, tres niveles de N (nitrógeno) en un ensayo de fertilizante ó dos procedencias de semilla de “n” posibles sitios.

La combinación de los dos tipos de factores brindan los siguientes modelos de ANDEVA:

Niveles de Factor 1	Niveles de Factor2	Designación	Comentario
Fijo	Fijo	Modelo I, Fijo	Uso muy frecuente
Aleatorio	Aleatorio	Modelo II, Aleatorio	Uso poco frecuente
Fijo	Aleatorio	Modelo III, Mixto	Uso frecuente
Aleatorio	Fijo	Modelo III, Mixto	Uso frecuente

En cuanto al diseño experimental, este puede implementarse mediante tratamientos con una sola observación o con múltiples observaciones por tratamiento como se ilustra a continuación.

3.1.1. ANDEVA de dos vías con una observación por celda

La tabla de análisis de varianza de dos vías con una observación por celda contiene los siguientes elementos (*Modelo de efectos fijos*):

Fuente de variación	Gl	SC	CM	F	CM Esperado
Total	FC-1	SC _T	-----	-----	-----
Filas (F)	F-1	SC _F	SC _F / F-1	CM _F / CM _E	$\sigma^2 + C \sigma^2_F$
Columnas (C)	C-1	SC _C	SC _C / C-1	CM _C / CM _E	$\sigma^2 + F \sigma^2_C$
Error	(R-1)(C-1)	SC _E	SC _E / (R-1)(C-1)		σ^2

Al igual que en el análisis de varianza de una vía, el objetivo del diseño es repartir la varianza total en tres componentes: efecto de filas (tratamiento 1), columnas (tratamiento 2) y error. Los supuestos del análisis son los mismos que para el diseño de una vía, a saber: errores independientes; con una distribución normal, varianzas homogéneas y aditividad en los efectos principales. El modelo lineal que describe la relación entre los componentes del diseño es:

$$X_{ij} = \mu + \beta_j + \pi_i + \varepsilon_{ij} \quad \text{con } i = 1, \dots, n; j = 1, \dots, n; \varepsilon_{ij} = (0, \sigma^2) \quad (23)$$

En donde:

- μ : representa la media de la población, constante para todas las observaciones.
- β_j : representa el efecto del tratamiento j, constante para todas las observaciones en tratamiento j.
- π_i : representa el efecto del tratamiento i, constante para todas las observaciones en tratamiento i.
- ε_{ij} : representa el error experimental, independiente y con una distribución normal para cada tratamiento $(0, \sigma^2)$.

El error asociado a cada observación es el resultado de la variabilidad no explicada por el diseño experimental y es el que se debe minimizar a través de la elección de un diseño estadístico apropiado. La implicación de un error experimental grande es que no se detectarán diferencias significativas entre los tratamientos.

A continuación se muestra cómo se calculan los diferentes componentes del análisis de varianza para un *diseño de bloques completos al azar* con una observación por tratamiento. El experimento consiste en determinar el efecto de cuatro tratamientos pregerminativos en semillas de teca (*Tectona grandis*). La investigación previa realizada por el (la) responsable del ensayo indica que los métodos más utilizados han sido:

1. Someter las semillas a remojo (agua)
2. Asolear las semillas previo a su siembra (sol)
3. Escarificar las semillas (abración)

Además y con fines de *control* se decidió incluir un lote de semillas sin ningún tratamiento. La variable respuesta fue el número de semillas germinadas por lote de cien semillas al cabo de diez días.

En este caso se asume que los tres métodos pregerminativos representan los únicos tratamientos de interés para el experimento y por lo tanto el diseño se define como de *efecto fijo*. Para aumentar el poder del experimento se decidió dividir el lote de semillas en ocho bloques y aplicar a cada bloque los tratamientos en forma aleatoria (*efecto aleatorio*). En este tipo de experimento la respuesta de interés es la tasa de germinación y no la variabilidad entre bloques.

Un diseño alternativo sería instalar cada uno de los bloques en viveros diferentes; nuevamente en este caso no estaríamos interesados en el efecto del vivero sobre la germinación y por lo tanto no es una prueba de hipótesis que nos interese someter a prueba.

El cuadro 6 resume los resultados del ensayo al cabo de las diez semanas y la figura 5 presenta el porcentaje de semillas germinadas por tratamiento y por bloque.

Cuadro 6: Respuesta hipotética de semillas de *Tectona grandis* a tres tratamientos pregerminativos. Los valores representan germinación por lotes de cien semillas al cabo de diez días del tratamiento.

Tratamientos	Bloques									% Germinación
	1	2	3	4	5	6	7	8	Total	
T1: control	40	70	40	40	20	30	30	30	300	37,5
T2: agua	80	90	80	70	60	70	60	70	580	72,5
T3: Sol	50	60	50	40	30	40	50	40	360	45,0
T4: Abrasión	80	90	100	90	100	100	100	100	760	95,0
Total	250	310	270	240	210	240	240	240	2000	62,5
% Germinación	62,55	77,5	67,5	60,0	52,5	60,0	60,0	60,0	62,5	62,5

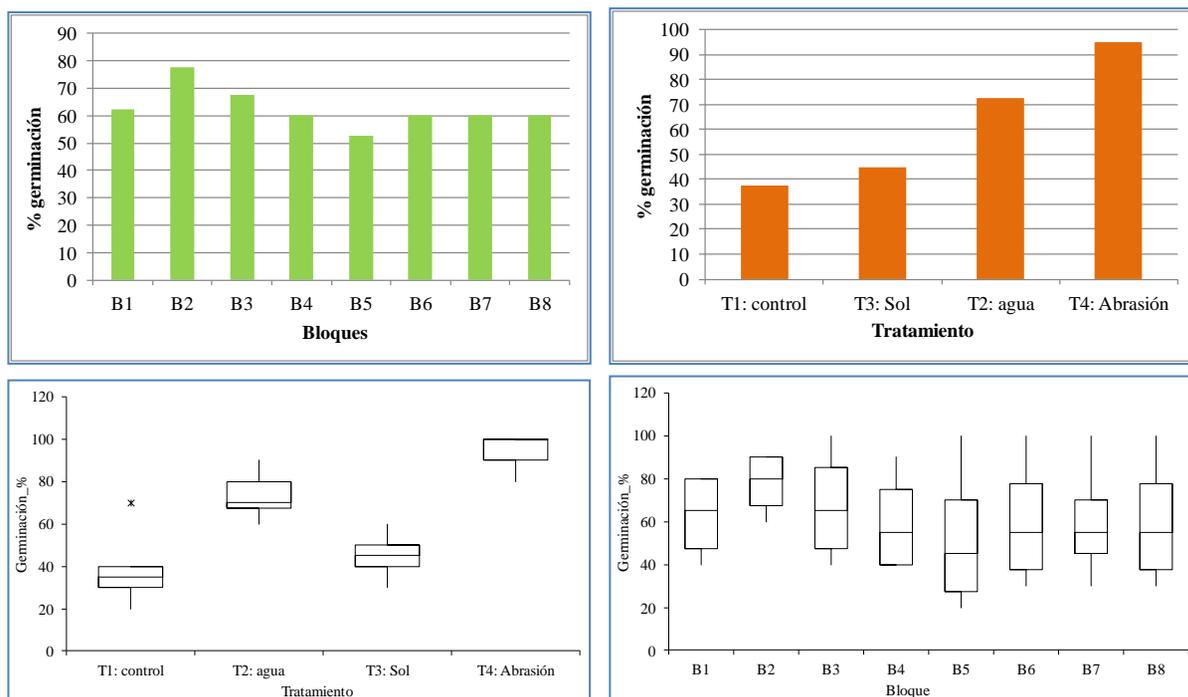


Figura 5: Porcentaje de germinación por lotes de 100 semillas. El testigo representa la germinación esperada en ausencia de un tratamiento.

A continuación se brindan los estadísticos descriptivos por tratamiento y por bloque.

Cuadro 7: Estadísticos descriptivos por tratamiento y bloque para el porcentaje de germinación de semillas de *Tectona grandis* sometidas a tres tratamientos pregerminativos. Los valores representan germinación por lotes de cien semillas al cabo de diez días del tratamiento.

Tratamiento	Total	T1: control	T2: agua	T3: Sol	T4: Abrasión
Number	32	8	8	8	8
Mean	62,5	37,5	72,5	45	95
St Dev	25,4	14,88	10,35	9,26	7,56
Skew	0,116	1,604	0,386	5,3E-17	-1,323
Min	20	20	60	30	80
Q ₁	40	30	67,5	40	90
Median	60	35	70	45	100
Q ₃	82,5	40	80	50	100
Max	100	70	90	60	100

Bloque	All	B1	B2	B3	B4	B5	B6	B7	B8
Number	32	4	4	4	4	4	4	4	4
Mean	62,5	62,5	77,5	67,5	60	52,5	60	60	60
St Dev	25,4	20,62	15,00	27,54	24,50	35,94	31,62	29,44	31,62
Skew	0,117	-0,200	-0,370	0,323	0,544	0,889	0,632	0,941	0,632
Min	20	40	60	40	40	20	30	30	30
Q ₁	40	47,5	67,5	47,5	40	27,5	37,5	45	37,5
Median	60	65	80	65	55	45	55	55	55
Q ₃	82,5	80	90	85	75	70	77,5	70	77,5
Max	100	80	90	100	90	100	100	100	100

COMENTARIOS:

1. Los bloques proveen una fuente adicional de variación conocida y por tanto esta variación puede sustraerse de la suma total de cuadrados, lo que a su vez hace que se reduzca el error no explicado por el experimento. El *cuadrado medio del error (MS)* representa la variabilidad que permanece como no explicada una vez que se ha eliminado la variabilidad debido a los bloques y a los tratamientos.

En este caso la reducción debido a los bloques es pequeña y por ende su incidencia en la potencia del ensayo es también pequeña. Cuanto más pequeño sea el valor del *cuadrado medio del error (MS)* mayor será la posibilidad de rechazar la hipótesis nula y por ende declarar a los tratamientos como significativos. El gráfico 5 permite apreciar que la mayor variabilidad en el número de semillas germinadas (variable respuesta) se presenta entre los tratamientos; esto es lo deseable en un diseño de bloques completos al azar; ya que las columnas representan repeticiones.

2. El testigo y el tratamiento de “sol” presentaron la menor germinación: 37.5% y 45%, respectivamente.
3. El método de abrasión es el que muestra la mayor germinación (95%).

Este análisis de tipo exploratorio nos brinda una primera impresión sobre cuáles tratamientos podrían ser iguales y cuáles diferentes. A continuación se analiza el resultado de la tabla de análisis de varianza.

A. Tabla de análisis de varianza

La tabla de análisis de varianza de dos vías permite, al igual que para el análisis de una vía, dividir la variabilidad total asociada al experimento en tres fuentes de variación; a saber:

- ✓ Variación inherente de cada tratamiento. Estimación de σ^2 . En los paquetes estadísticos a esta valor se le denomina “error”.

- ✓ Efecto principal del factor de las columnas (repeticiones, en este caso). Estimación de σ^2 más el efecto principal de los bloques.
- ✓ Efecto principal del factor de las filas (métodos pregerminativos, en este caso). Estimación de σ^2 más el efecto principal de los tratamientos.

A continuación se presenta el resultado del análisis de varianza obtenidos con el complemento para Excel XLStatistics (1Num2Cat).

Cuadro 8: Tabla de análisis de varianza de dos vías.

ANOVA Table					
Source	DF	SS	MS	F	p-value
Tratamiento	3	16700	5566,667	64,944	8,569E-11
Bloque	7	1500	214,286	2,5	0,0490957
Error	21	1800	85,714		
Total	31	20000			

Hypothesis Tests

Note: Tests for main effects (due to Tratamiento and Bloque separately) may not be appropriate if interaction is present and serious

H ₀ : Germinación_% does not depend on Tratamiento	
H ₁ : Germinación_% depends on Tratamiento	
p-value =	8.569E-11
H ₀ : There is no variability in Germinación_% due to Bloque	
H ₁ : There is variability in Germinación_% due to Bloque	
p-value =	0.0490957

Fuente de variación	Valor de F	Valor de P	Ho:
Tratamientos	64,944 (5566,667/85,714)	≤0.001	$\mu_1 = \mu_2 = \mu_3 = \mu_4$
Bloques	2,50 (214.29/85.71)	0.05	$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$

COMENTARIOS:

- ✓ La prueba “F” para los tratamientos indica que al menos la media de un par de tratamientos es estadísticamente diferente ($P \leq 0,001$). En cuanto a los bloques, las diferencias no son significativas a un alfa de 0,01; sin embargo sí lo serían a un alfa de 0,05. En este tipo de diseño experimental los bloques tienen como objetivo reducir la estimación del error estándar de los tratamientos y por tanto no tienen interés como un tratamiento en sí mismos.
- ✓ Observe que el valor del cuadrado medio de los bloques (214,29) es 2,5 veces el tamaño del cuadrado medio del error (85,71); esto indica que el diseño de bloques al azar mejoró la exactitud del experimento.
- ✓ La eficiencia del diseño en *bloques al azar (BA)* con respecto al diseño *irrestringido o completamente al azar (CA)* puede evaluarse mediante la siguiente expresión (Snedecor and Cochran, 1980):

$$\sigma^2_{CA} / \sigma^2_{BA} \quad (24)$$

En donde,

$$S^2 = \sigma^2_{CA} = \frac{(J-1) S_b^2 + J(I-1) S^2}{(IJ-1)} \quad (25)$$

En donde,

I: tratamientos

J: Bloques o repeticiones

S_b^2 : Cuadrado medio del error de los bloques o repeticiones

Para nuestro ejemplo tenemos:

I: 4

J: 8

S_b^2 : 214,29

S^2 : 85,71

$$\sigma^2_{CA} / \sigma^2_{BA} = \{(8-1)*214,29 + 8(4-1)* 85,71\} / 85,71*\{4*8\}-1 = 1,33$$

Los datos indican que si se hubiese utilizado un diseño completamente al azar se habrían requerido once réplicas ($1,33*8$) para lograr el mismo error estándar para los tratamientos del diseño de bloques al azar. Los cálculos anteriores no consideran el hecho de que el diseño completamente al azar proveerá un mayor número de grados de libertad para el error comparado con el diseño de bloques al azar (en este caso 40 Vs 21) y por lo tanto Fisher (1935-1951, citado por Snedecor and Cochran, 1980) sugiere modificar la expresión anterior para considerar el efecto de los grados de libertad:

$$(gl_{BA}) * (gl_{CA}) (\sigma^2_{CA} / \sigma^2_{BA}) / (gl_{BA} + 3) * (gl_{CA} + 1)$$

$$\{(21+1) (40+3) * 1.33\} / (21+3)*(40+1) = 1,27$$

Observe que aun con la corrección de Fisher, todavía se requerirían unas diez réplicas para lograr el mismo nivel de error que con el diseño de bloques completos al azar.

El éxito del diseño de bloques completos al azar está en elegir material muy similar entre sí para luego aplicar los tratamientos. Por ejemplo, dos sitios con condiciones de suelo, clima, pendiente y uso previo muy similares son candidatos para servir como réplicas en un experimento de manejo o restauración forestal.

Una vez detectadas diferencias entre los tratamientos se procede a aplicar alguna de las pruebas a posteriori o contrastes planeados para determinar cuáles medias son diferentes. A continuación se ilustra cómo utilizar BioEstat para realizar un análisis de varianza de dos vías con la prueba a posteriori de Tukey.

1. Importar el archivo cuadro_6_bioestat.csv a BioEstat

cuadro_6_bioestat.csv - Notepad

```
File Edit Format View Help
40,80,50,80
70,90,60,90
40,80,50,100
40,70,40,90
20,60,30,100
30,70,40,100
30,60,50,100
30,70,40,100
```

➔

cuadro_6_bioestat

	- 1 - Testigo	- 2 - Agua	- 3 - Sol	- 4 - Abrasión
1	40.000	80.000	50.000	80.000
2	70.000	90.000	60.000	90.000
3	40.000	80.000	50.000	100.000
4	40.000	70.000	40.000	90.000
5	20.000	60.000	30.000	100.000
6	30.000	70.000	40.000	100.000
7	30.000	60.000	50.000	100.000
8	30.000	70.000	40.000	100.000

Para asignar la etiqueta a cada columna, selecciónela y haga un clic sobre F2. Luego digite el texto.

2. Seleccione de menú de Estadísticas, Análise da Variância, ANOVA: dois criterios

BioEstat 5.3

Arquivo Editar Estatísticas Gráficos Sugestões Configurar Ajuda

Último Teste Ctrl+U Escolha um teste

Dados 1

- Amostragem
- Análise Multivariada
- Análise de Sobrevida
- Análise da Variância**
 - ANOVA: um critério
 - ANOVA: dois critérios**
 - ANOVA: fatorial a x b
- Bootstrap - Reamostragem
- Correlação

3. Seleccionar columnas 1 a 4

Seleção de amostras para Teste ANOVA: dois critérios

Colunas disponíveis = 0

Colunas selecionadas = 4

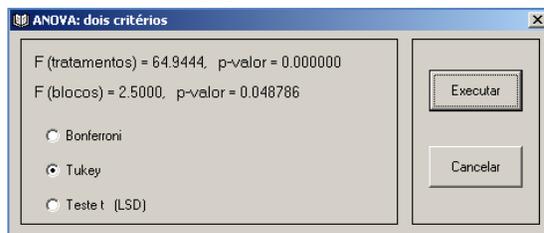
- # 1 - Testigo
- # 2 - Agua
- # 3 - Sol
- # 4 - Abrasão

Duplo Clique para excluir a coluna

Executar Estatística Cancelar

5. Executar Estatística

6. Marcar Tukey y Executar



7. Editar y copiar resultados

FO	GL	SQ	QM
Tratamentos	3	16700,0000	5566,667
Blocos	7	1500,0000	214,286
Erro	21	1800,0000	85,714

8. Pegar resultados en documento de Word

FONTES DE VARIAÇÃO	GL	SQ	QM
Tratamentos	3	16700,0	5566,667
Blocos	7	1500,0	214,286
Erro	21	1800,0	85,714

Análisis de varianza

F (tratamentos) = 64,9444
 p (tratamentos) = < 0,0001
 F (blocos) = 2,5000
 p (blocos) = 0,0488

Valor de F para tratamientos
 Valor de p calculado para tratamientos
 Valor de F para bloques
 Valor de p calculado para bloques

Médias (tratamentos):

Média (Coluna 1) = 37,5
 Média (Coluna 2) = 72,5
 Média (Coluna 3) = 45,0
 Média (Coluna 4) = 95,0

Medias por tratamento

Tukey	Q	(p)
Médias (1 a 2) =	10,6927	< 0,01
Médias (1 a 3) =	2,2913	ns
Médias (1 a 4) =	17,5665	< 0,01
Médias (2 a 3) =	8,4014	< 0,01
Médias (2 a 4) =	6,8739	< 0,01
Médias (3 a 4) =	15,2753	< 0,01

Resultado de comparaciones múltiples de Tukey.

ns: No significativo.

Médias (blocos):

Média (Linha 1) =	62,5
Média (Linha 2) =	77,5
Média (Linha 3) =	67,5
Média (Linha 4) =	60,0
Média (Linha 5) =	52,5
Média (Linha 6) =	60,0
Média (Linha 7) =	60,0
Média (Linha 8) =	60,0

Medias por bloque

Tukey	Q	(p)
Médias (1 a 2) =	3,2404	ns
Médias (1 a 3) =	1,0801	ns
Médias (1 a 4) =	0,5401	ns
Médias (1 a 5) =	2,1602	ns
Médias (1 a 6) =	0,5401	ns
Médias (1 a 7) =	0,5401	ns
Médias (1 a 8) =	0,5401	ns
Médias (2 a 3) =	2,1602	ns
Médias (2 a 4) =	3,7804	ns
Médias (2 a 5) =	5,4006	< 0,05
Médias (2 a 6) =	3,7804	ns
Médias (2 a 7) =	3,7804	ns
Médias (2 a 8) =	3,7804	ns
Médias (3 a 4) =	1,6202	ns
Médias (3 a 5) =	3,2404	ns
Médias (3 a 6) =	1,6202	ns
Médias (3 a 7) =	1,6202	ns
Médias (3 a 8) =	1,6202	ns
Médias (4 a 5) =	1,6202	ns
Médias (4 a 6) =	0,0000	ns
Médias (4 a 7) =	0,0000	ns
Médias (4 a 8) =	0,0000	ns
Médias (5 a 6) =	1,6202	ns
Médias (5 a 7) =	1,6202	ns
Médias (5 a 8) =	1,6202	ns
Médias (6 a 7) =	0,0000	ns
Médias (6 a 8) =	0,0000	ns
Médias (7 a 8) =	0,0000	ns

Resultado de comparaciones múltiples de Tukey para diferencias entre bloques.

ns: No significativo.

La prueba F_{\max} de Hartley indica que las varianzas son iguales entre tratamientos ($p > 0,05$); la razón entre la varianza máxima y mínima es 3,87. Al evaluar los resultados de esta prueba recuerde que es sumamente sensible a las desviaciones de *normalidad* en el set de datos y por lo tanto rechazar H_0 puede significar no-normalidad en vez de desigualdad de varianzas. Por lo anterior, muchos estadísticos no recomiendan esta prueba (Sokal y Rohlf, 1969).

Hartley's F_{\max} Test	
Data	
Number of Groups	4
Total Sample Size	32
S_{\max}^2	221.43
S_{\min}^2	57.143
H ₀ : Population variances of all categories equal	
H ₁ : Population variances of all categories not equal	
F_{\max}	3.875
DF ₁	4
DF ₂	7
p-value	0.057292819

prueba F_{\max} de Hartley (XLStatistics)

COMENTARIOS: Las comparaciones múltiples de Tukey declaran como significativas ($p < 0,01$) todas las comparaciones con la excepción del testigo) versus sol y por lo tanto se recomendaría el tratamiento No. 4 (abración) como el más efectivo. Una ventaja del tratamiento 4 es que también presentó la menor variabilidad (7,6%).

3.2 Diseño factorial con “n” observaciones por celda

Un experimento factorial es aquel en el cual dos o más factores se prueban en todas las combinaciones posibles. Cada uno de los factores debe estar representado por dos o más niveles. Por ejemplo, en un experimento factorial de 2×2 existen 4 tratamientos; en uno de 3×3 existen 9 combinaciones o tratamientos y así sucesivamente. Este tipo de experimento permite evaluar la interacción entre los tratamientos; o sea, el efecto diferencial de un factor sobre el otro. Algunos ejemplos de diseños factoriales son:

- ✓ Medir la respuesta de un cultivo a dos niveles de fertilización y tres niveles de pH.
- ✓ Medir la erosión asociada a dos métodos de preparación de terreno en dos tipos de suelos.
- ✓ Comparar la eficiencia de dos métodos de divulgación en tres grupos de edad.

Para ilustrar el análisis de varianza de dos vías con mediciones múltiples por tratamiento utilizaremos también un set de datos hipotético. Suponga que un(a) viverista está interesado(a) en determinar si el método de inoculación de micorriza en roble y el método de trasplante tienen un efecto en el crecimiento de las plántulas. Su diseño experimental consiste de dos métodos de inoculación y dos métodos de trasplante con veinte plántulas por tratamiento (cuadro 9).

Cuadro 9: Asignación de plántulas por tratamiento. Análisis de varianza factorial de 2×2 . Diseño completamente al azar.

Tratamiento	Método Inoculación 1	Método Inoculación 2	Subtotal
Mét. Trasplante 1	$n(I_1 T_1) = 20$	$n(I_2 T_1) = 20$	$n T_1 I_{1 y 2} = 40$
Mét. Trasplante 2	$n(I_1 T_2) = 20$	$n(I_2 T_2) = 20$	$n T_2 I_{1 y 2} = 40$
Subtotal	$n(I_1 y T_{1 y 2}) = 40$	$n(I_2 y T_{1 y 2}) = 40$	N total = 80

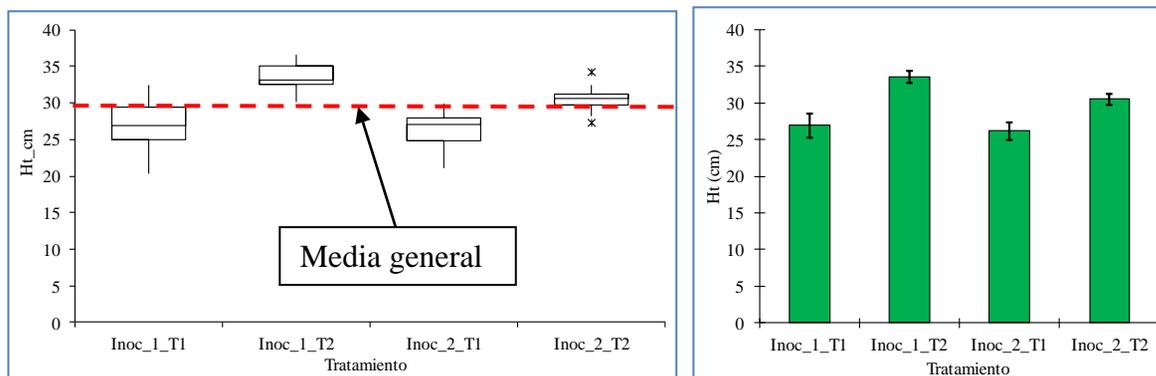
En este caso, todos los tratamientos tienen el mismo número de plantas y por lo tanto el diseño se denomina *balanceado u ortogonal*. Observe que el diseño considera todos los niveles del factor I y todos los niveles del factor J y por lo tanto se denomina factorial. El diseño es completamente al azar con “n” réplicas por tratamiento. Los supuestos expuestos para los diseños de varianza anteriores también aplican al presente.

El cuadro 10 y las figuras 6 y 7 muestran la respuesta en altura total (cm) de las árboles de roble al cabo de 6 meses. Para diseños balanceados la media general es igual a la medio de todos los datos, así como también a la media de columnas y filas (ver cuadro 10).

Cuadro 10: Estadísticos para la altura media de las plántulas (cm) al cabo de seis meses.

Tratamiento	Todos	Inoc_1_T1	Inoc_1_T2	Inoc_2_T1	Inoc_2_T2
Number	80	20	20	20	20
Mean	29.28	26.91	33.54	26.17	30.51
St Dev	3.81	3.45	1.66	2.56	1.58
Skew	-0.37	-0.20	-0.20	-0.78	0.22
Min	20.34	20.34	30.22	21.18	27.35
Q ₁	27.02	24.93	32.53	24.81	29.73
Median	29.79	26.88	33.17	27.11	30.63
Q ₃	32.17	29.41	35.05	27.95	31.14
Max	36.46	32.37	36.46	29.90	34.28

Software: XLStatistics



Categoría	ME	L. Inf.	L. Sup.
Inoc_1_T1	1,613	25,3	28,5
Inoc_1_T2	0,775	32,8	34,3
Inoc_2_T1	1,200	25,0	27,4
Inoc_2_T2	0,738	29,8	31,2

Figura 6: Altura total (cm) al cabo de seis meses para plántulas de roble sometidas a dos métodos de inoculación y dos de trasplante. Tamaño de muestra veinte plantas por tratamiento. A. Diagrama de cajas. B. Diagrama de barras con intervalo de confianza al 95%.

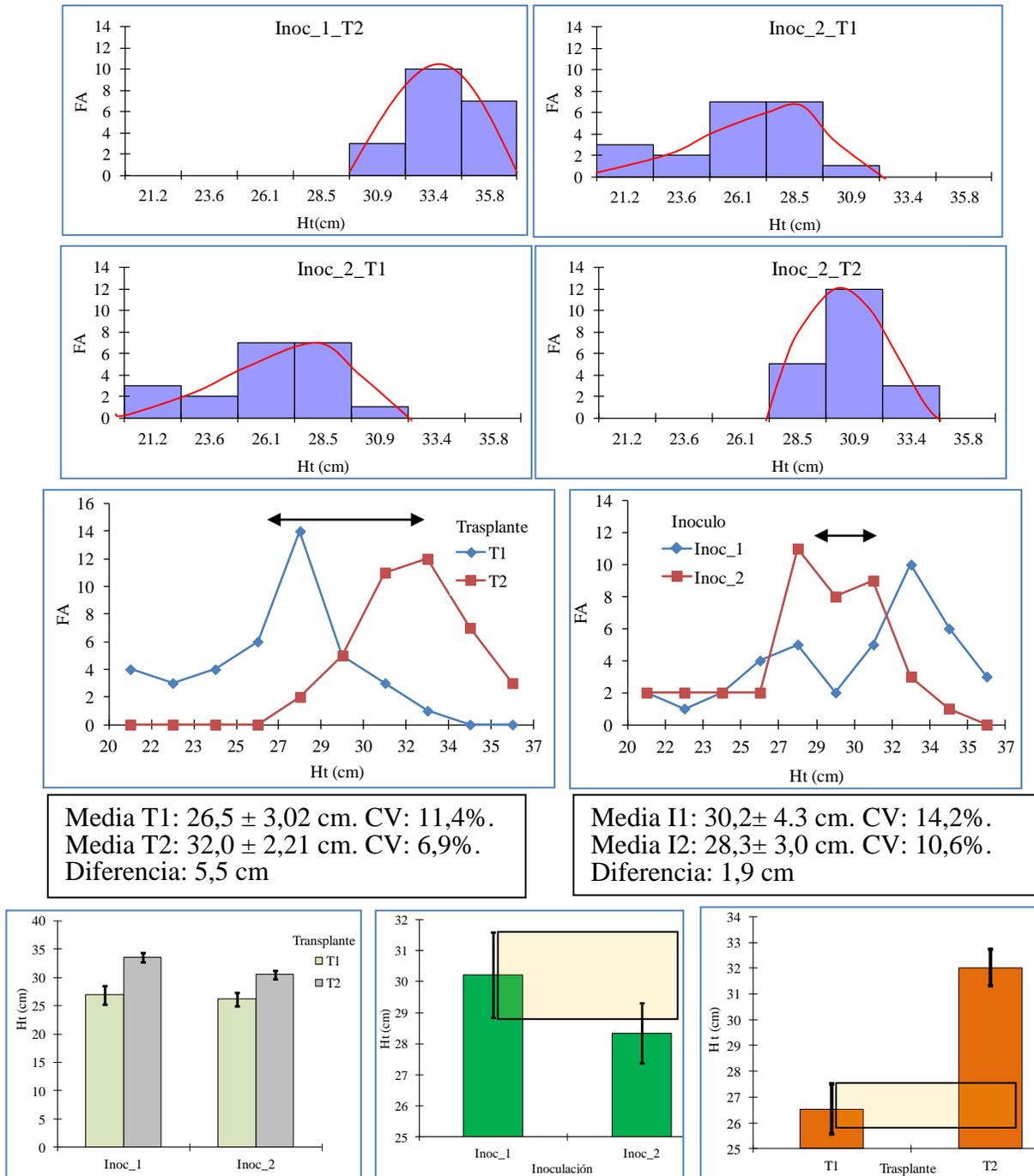


Figura 7: Altura total (cm) de plántulas de roble al cabo de seis meses por tratamiento. Tamaño de muestra veinte plantas por tratamiento. Barras indican intervalo de confianza al 95%.

COMENTARIOS:

1. Los datos indican que las plántulas roble de responden de manera diferente a los métodos inoculación y de trasplante (figs. 6 y 7) y que el método de trasplante parece ejercer una mayor influencia en el crecimiento de las plantas.
2. La mayor variabilidad entre los tratamientos corresponde a los métodos de inoculación 10,6 a 14,2%.

- Los intervalos de confianza (95%) para las medias por tratamiento indican que probablemente ambos sean estadísticamente significativos a un alfa de 0,05.
- La prueba Fmax de Hartly indica que las varianzas no son homogéneas entre los tratamientos ($p < 0,01$).

Data		H_0 : Population variances of all categories equal	
Number of Groups	4	H_1 : Population variances of all categories not equal	
Total Sample Size	80	F_{max}	4.77783002
S_{max}^2	11.879	DF_1	4
S_{min}^2	2.4862	DF_2	19
		p-value =	0.007742136

Software: XLStatistics.

- El análisis de residuos evidencia que la distribución de los datos es normal sin embargo la varianza entre tratamientos no es homogénea.

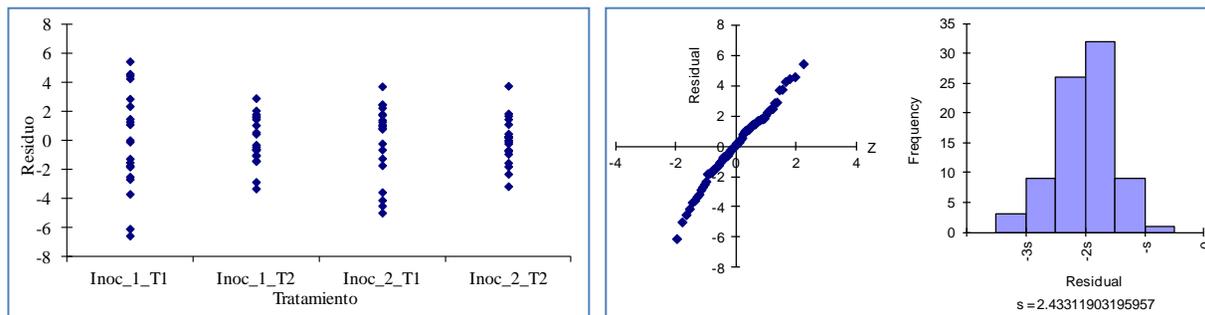


Figura 8: Análisis de residuos para la respuesta en altura total (cm) de plántulas sometidas a dos métodos de inoculación y dos de trasplante. Altura total a los seis meses.

El análisis anterior es de tipo exploratorio y por lo tanto debe confirmarse mediante el análisis de varianza; el cual permite responder a las siguientes preguntas:

- ¿Cuál es el efecto de los métodos de inoculación? $H_0 : \mu_1 = \mu_2$
- ¿Cuál es el efecto de los métodos de trasplante? $H_0 : \mu_1 = \mu_2$
- ¿Cuál es el efecto de combinar el método de inoculación con el método de trasplante?
 $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

Las hipótesis anteriores prueban por los **efectos principales** (métodos de inoculación y métodos de trasplante); o sea se realiza una variante de un análisis de varianza de una vía para las columnas ($H_0 : \mu_1 = \mu_2$) y otro para las filas ($H_0 : \mu_1 = \mu_2$). La diferencia media entre métodos de trasplante y de inoculación es la siguiente:

Media T1: $26,5 \pm 3,02$ cm. CV: 11,4%.

Media T2: $32,0 \pm 2,21$ cm. CV: 6,9%.

Diferencia: 5,5 cm

Media I1: $30,2 \pm 4,3$ cm. CV: 14,2%.

Media I2: $28,3 \pm 3,0$ cm. CV: 10,6%.

Diferencia: 1,9 cm

Al interpretar el valor de los efectos principales se debe recordar que representan medias de un factor asumiendo que la respuesta de los individuos no es afectada por los niveles del otro factor; o sea bajo el supuesto de que no existe interacción ó que ésta es mínima. El concepto de interacción se discute a continuación.

3.2.1 ¿Qué es la interacción?

El diseño factorial con múltiples observaciones por tratamiento permite probar por la interacción entre los diferentes niveles de cada factor. Por ejemplo, ¿es la respuesta de las plántulas de roble independiente de los métodos de trasplante o por el contrario, su crecimiento se debe parcialmente al efecto del método de trasplante? En caso de existir un efecto conjunto (positivo o negativo) el mismo recibe el nombre de *interacción* y su significancia puede determinarse a través de una prueba F.

Si la prueba F en la tabla de análisis de varianza es significativa entonces se puede concluir que la respuesta de las plántulas no es independiente de los métodos de trasplante. La forma más simple de apreciar si existe interacción en el set de datos es graficando las medias de las celdas del cuadro 10 como se ilustra en la figura 9A. En ausencia de interacción, las líneas deben ser paralelas; aún cuando las diferencias entre tratamientos para un factor no sean exactamente iguales debido a la variabilidad muestral. La presencia de interacción se ilustra en la figura 9B.

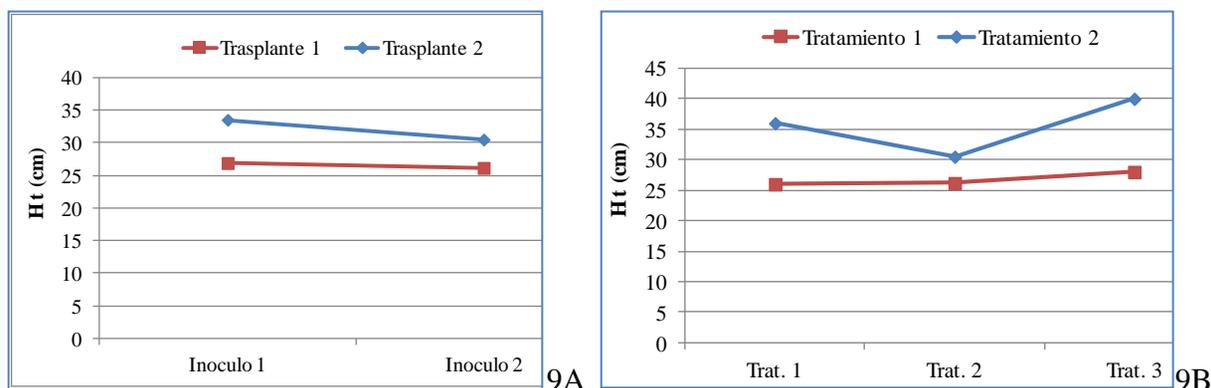


Figura 9: A. Ausencia de interacción entre factores. B. Interacción fuerte entre factores.

Cuando la presencia de un factor incrementa o afecta positivamente la respuesta del material experimental del otro factor se dice que existe un efecto de *sinergia*; en tanto que cuando se produce un efecto adverso se dice que los factores ejercen una acción de *interferencia*.

3.2.2 Tabla de análisis de varianza para un diseño factorial

La tabla de análisis de varianza factorial permite, al igual que para el análisis de una vía, dividir la variabilidad total asociada al experimento en cuatro fuentes de variación:

- ✓ Variación inherente de cada celda (tratamiento). Estimación de σ^2 . En algunos paquetes estadísticos a este valor se le denomina “error” o “within cells”.
- ✓ Efecto principal del factor de las columnas (método de inoculación, en este caso). Estimación de σ^2 más el efecto principal de las columnas.
- ✓ Efecto principal del factor de las filas (método de trasplante, en este caso). Estimación de σ^2 más el efecto principal de las filas.
- ✓ Interacción (método de inoculación y método de trasplante). Estimación de σ^2 más el efecto principal de la interacción columnas*filas.

A continuación se presenta el resultado del análisis de varianza obtenidos con el programa XLSTATistics.

ANOVA Table					
Source	DF	SS	MS	F	p-value
Inóculo	1	71,234	71,234	12,033	0,0009
Trasplante	1	601,979	601,979	101,684	1,152E-15
Inóculo*Trasplante	1	26,163	26,163	4,419	0,0388
Error	76	449,925	5,920		
Total	79	1149,301			

H ₀ :	Ht_cm does not depend on Inoculo
H ₁ :	Ht_cm depends on Inoculo
p-value =	0,000864

Prueba por el efecto principal del factor uno: método de inoculación.

H ₀ :	Ht_cm does not depend on Trasplante
H ₁ :	Ht_cm depends on Trasplante
p-value =	1,152E-15

Prueba por el efecto principal del factor dos: método de trasplante.

Hypothesis Tests	
Note: Tests for main effects (due to Inoculo and Trasplante separately) may not be appropriate if interaction is present and serious	
H ₀ :	There is no interaction between Inoculo and Trasplante in their effect on Ht_cm
H ₁ :	There is interaction between Inoculo and Trasplante in their effect on Ht_cm
p-value =	0,0388

Prueba por la interacción entre método de inoculación y de trasplante

Comentarios:

1. Las pruebas “F” indican, al igual que en el análisis de varianza de una vía, la razón entre estimaciones independientes de la varianza. Si el valor “p” asociado a cada razón es menor que el “p” crítico, entonces se declara la misma como significativa. O sea, al menos un par de medias asociadas a dicha prueba es estadísticamente diferente.

Fuente de variación	Valor de F	Valor de P	Ho:
A (Métodos Inococ.)	$71,234 / 5,920 = 12,033$	0,0009	$\mu_1 = \mu_2$
B (Métodos Transp.)	$601,979 / 5,920 = 101,684$	1,152E-15	$\mu_1 = \mu_2$
AB (interacción)	$26,163 / 5,920 = 4,419$	0,0388	$\mu_1 = \mu_2 = \mu_3 = \mu_4$

En el presente caso y para un alfa de 0,01 tanto los métodos de inoculación (A) como los métodos de trasplante (B) son estadísticamente significativos. La interacción entre los dos factores no es significativa ($P > 0.01$).

3.2.3. Ventajas del diseño de dos vías

El diseño de dos vías con múltiples observaciones por tratamiento tiene las siguientes ventajas con respecto al de una vía:

- ✓ Permite obtener conclusiones sobre un ámbito mayor de situaciones con un solo experimento.
- ✓ Se puede dar respuesta a dos preguntas con la misma inversión. Por ejemplo, se pueden utilizar las mismas 80 plántulas de roble para evaluar el efecto de los métodos de inoculación y de trasplante.
- ✓ El diseño de dos vías es más eficiente que el de una vía. La eficiencia se mide en términos del tamaño de muestra requerido para lograr un determinado error.

4. Transformaciones

La validez de las conclusiones del análisis de varianza depende de que los datos cumplan con los supuestos del análisis. Con frecuencia y como se observó en la sección anterior, dichos supuestos no se cumplen. Ya se mencionó, que cuando del tamaño de muestra es el mismo para todos los tratamientos la prueba F es robusta a violaciones en el supuesto de homogeneidad de varianzas y que el supuesto de normalidad puede obtenerse con muestras de tamaño mediano (30 o más observaciones).

Para remediar la violación de los supuestos del análisis de varianza se aplican transformaciones a los datos originales y el análisis es realizado con los datos transformados. Una desventaja de transformar los datos es que las conclusiones aplican a los datos transformados y no a los originales. Las transformaciones más frecuentes son (Snedecor y Cochran, 1980):

Transformación logarítmica

Esta transformación es útil cuando la desviación estándar de cada tratamiento tiende a ser proporcional a su respectiva media; o sea cuando el coeficiente de variación tiende a ser constante (Manly, 1992). Otro criterio para elegir esta transformación es que los efectos principales son multiplicativos y no aditivos. Esta transformación presenta las siguientes limitaciones: no se puede aplicar a valores negativos ni a ceros. Esta limitación puede eliminarse sumando 1 o cualquier otra constante al valor de cada tratamiento antes de aplicarla.

Se puede utilizar cualquier logaritmo; sin embargo con frecuencia se utiliza el de base 10. La transformación logarítmica tiende además a reducir el grado de asimetría positiva del set de datos. Esta transformación es muy utilizada para datos sobre crecimiento de organismos. Una vez aplicada una transformación logarítmica a los datos las conclusiones deben expresarse en términos de tasas de cambio (proporciones o porcentajes) y no en términos absolutos

Transformación de raíz cuadrada (conteos)

Esta transformación se aplica normalmente a datos obtenidos a partir de conteos y para muestras pequeñas ($n < 20$); en general estos datos siguen la distribución conocida como Poisson y se aplica cuando la probabilidad de presencia de un evento (Ej. individuo con una característica dada) es muy baja.

Cuando la muestra es mayor que veinte, la distribución de Poisson y la Normal son muy similares y por tanto es probable que no se requiera de una transformación (Fry, 1996). Algunos ejemplos de conteos son número de semillas germinadas por lotes de 100 semillas, número de árboles infestados con una bacteria por parcela de 1000 m² ó número de animales silvestres atropellados por autos en carreteras. En estos casos, la varianza tiende a ser proporcional a la media. Cuando el set de datos incluye ceros se recomienda codificarlos antes de aplicar la transformación; para ello simplemente adicione 0,5 al valor de cada observación.

Transformación angular o de arcoseno

Esta transformación aplica a datos que expresan porcentajes o proporciones obtenidas a partir de un conteo o experimento. Por ejemplo, porcentaje de semillas que germinan por lote de 100 semillas o fracción de árboles bifurcados por parcela de 500 m². La transformación aplicada es arcoseno $(x)^{0.5}$ (nota: x debe expresarse como una fracción).

La transformación es poco efectiva cuando los valores se encuentran en el ámbito 40 a 100% (0,4 a 1); y en general se recomienda para aquellos casos en que la mayoría de las observaciones se encuentran entre 0 y 0,3 y 0,7 a 1,0.

Para tamaños de muestra (n) inferiores a 50, los valores de 0 y 1 deben reemplazarse por $1/4n$ y $(n-1/4)/n$; respectivamente. Al utilizar un paquete estadístico para realizar la transformación debe asegurarse que el cálculo se haga en grados y no en radianes; en caso de utilizar radianes el resultado debe multiplicarse por 57,296 para convertirlos a grados decimales.

Recuerde que al utilizar transformaciones se modifica la escala original de medición y por lo tanto los intervalos de confianza deben reportarse para los datos transformados y no para los originales. Estos nuevos intervalos serán asimétricos con respecto a la media.

Ley de potencia de Taylor

Esta transformación utiliza un análisis de regresión lineal entre el \log_{10} de la varianza de los grupos y el \log_{10} de las medias de los grupos. El objetivo del análisis de regresión es determinar la transformación óptima que se debe aplicar a los datos.

La transformación es $X*P$ en donde $P= 1- (b/2)$; en donde “b” es la pendiente de una regresión lineal simple. El método es apropiado para aquellos datos que muestran un relación lineal entre \log_{10} de la varianza y \log_{10} de las medias por tratamiento o grupo (Fry, 1996).

5. Comentario final

El tema de análisis de varianza es uno de los más interesantes y útiles en el área de la estadística inferencial; sin embargo, es también uno de los más complejos. En el presente capítulo se han presentado los elementos básicos relacionados con el tema y por lo tanto, se recomienda al estudiante que busque asesoramiento cuando requiera diseñar un experimento o cuando se enfrente al reto de analizar datos obtenidos en ambientes no controlados.

Cuando se utilice un paquete estadístico debe leerse cuidadosamente el manual con el objeto de determinar la forma correcta en que deben estructurarse los datos. También se debe entender la terminología utilizada por el autor del manual o de cualquier libro de estadística; ya que diferentes autores expresan el mismo concepto con diferentes nombres y símbolos matemáticos.

6. Bibliografía

Bernhardson, C. S. 1975. Type I error rates when multiple comparisons procedures follow a significant F test ANOVA. *Biometrics* 31:229.

Carmer, S. G. and Swanson, M. R. 1971. Detection of differences between means: A Monte Carlo study of five pairwise multiple comparison procedure. *Agronomy Journal*, 63:940-945.

Carmer, S. G. and Swanson, M. R. 1973. An evaluation of ten pairwise multiple comparisons procedures by Monte Carlo methods. *Journal of The American Statistician Association*, 74:122-124.

Cochran William y Cox Gertrude. 1974. Diseños experimentales. Trillas. México. Tercera reimpresión. 661p.

Duncan, D. B. 1955. Multiple range and multiple F tests. *Biometrics*, 11:1-42.

Dunnnett, C. W. 1955. A multiple comparisons procedure for comparing several treatments with a control. *Journal American Statistical Association*, 50:1096.

Dunnnett, C. W. 1964. New tables for multiple comparisons with a control. *Biometrics* 20: 482.

Gill, J. L. 1973. Current Status of multiple comparisons of means in designed experiments. *Journal Dairy Science*, 56:973.

Lehmann, E. L. and Shaffer, J. P. 1977. On a fundamental theorem in multiple comparisons. *Journal American Statistical Association*, 72:576.

Lowry, S. R. 1992. Use and misuse of multiple comparisons in animal experiments. *Journal Animal Science*, 70:1971-1977.

Manly, B. F.J. 1992. *The design and analysis of research studies*. Cambridge University Press. Great Britain. 353p.

Mize, C. W. and Schultz, R. C. 1985. Comparing treatment means correctly and appropriately. *Canadian Journal of Forestry Research* 15(6): 1142.

Saville, D. J. 1990. Multiple comparison procedures: the practical solution. *The American Statistician*, Vol 44(2): 174-180.

Scheffé, H. 1953. A method for judging all contrasts in the analysis of variance. *Biometrika*, 40: 87-104.

Snedecor, G.W. and Cochran, W.G. 1980. *Statistical methods*. Seventh Ed. Iowa, The Iowa State University Press. 507p.

Sokal, R. R. and Rohlf, 1969. *Biometry*. W. H. Freeman and Company. USA.

Steel, R.G.D. y Torrie, J.H. 1980. *Principles and procedures of Statistics: A Biometrical Approach* (2nd Ed.). McGraw-Hill, New York. 629p.

Tukey, J. W. 1949. Comparing individual means in the analysis of variance. *Biometrics*, 5:99-114.

7. Ejercicios

1. Defina los siguientes términos: ANDEVA, Fcrítico, CM, Df, varianza inherente, varianza total, varianza entre grupos, varianza dentro de grupos, significancia, tabla de análisis de varianza, pruebas a posteriori, Tukey, Dunnett, Scheffe, Newman-Keuls, comparaciones planeadas, normalidad, ANDEVA de dos vías, interacción, efectos principales, diferencia “práctica”, error.
2. Dados seis tratamientos (A,B,C,D,E,F) determine el número de pruebas “t” requeridas para probar todas las posibles combinaciones de tratamientos.
3. Una investigadora desea someter a prueba cuatro métodos de escarificación de semillas para una especie del bosque seco tropical. Para ello selecciona al azar cuatro grupos de 100 semillas cada una y les asigna a cada grupo un tratamiento al azar. Plantee la hipótesis nula y alternativa. Desde un punto de vista teórico ¿por qué sería incorrecto plantear H_a como: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$?
4. Realice un análisis de varianza para los datos del archivo ejercicio_p4.xlsx. El Alcalde de Liberia desea saber si existe alguna diferencia estadística entre la disponibilidad de pago asociada a los entrevistadores.
5. Realice un análisis de varianza para los datos del archivo ejercicio_p5.xlsx. Si el análisis de varianza es significativo, utilice una prueba a posteriori para determinar cuáles tratamientos son diferentes. ¿Cuál tratamiento recomendaría usted? ¿Por qué?
6. Un investigador desea saber si la altura total de la especie “X” está asociada al tipo de suelo donde crece. Para ello selecciona un parche de bosque no alterado en cuatro tipos de suelos y en cada uno de ellos establece una parcela de 1000 m² y mide la altura de los árboles. Los resultados se presentan en el archivo “ejercicio_p6.xlsx”. Realice un análisis de varianza para someter a prueba la hipótesis de que el crecimiento en altura total es independiente del tipo de suelo. Incluya el análisis exploratorio (grafico y numérico).

Sugerencias:

- ✓ Grafique los datos (Box-Whisker). Observe la tendencia de los datos. ¿Qué es evidente en la gráfica?
- ✓ Calcule estadísticos descriptivos por sitio. Analice la variabilidad al interior de cada sitio. ¿Qué le indica esto?
- ✓ Realice el análisis de varianza.
- ✓ Seleccione uno o dos métodos de comparaciones múltiples.
- ✓ Redacte su conclusión estadística y práctica. ¿Qué sugerencias haría usted en cuanto a la muestra?

7. A continuación se brindan valores de compactación de suelo (Kg/cm^2) para tres sitios en la Estación Experimental Horizontes, Guanacaste y para un sitio en Pital de San Carlos. Cada sitio representa las condiciones de suelo después de iniciado el proceso de sucesión natural. Los datos provienen de Alfaro, R. Ernesto. 1998. Efecto del desarrollo del bosque secundario en diferentes etapas sucesionales en las propiedades físicas y químicas del suelo, Estación Experimental Horizontes, Guanacaste. Costa Rica. Anteproyecto de tesis de Ing. Forestal. Esc. Ciencias Ambientales. UNA.

La pregunta a responder es: ¿Muestran los datos evidencia estadística para suponer que conforme se avanza en la sucesión la compactación del suelo se reduce?. Datos en archivo ejercicio_p7.xlsx.

Sugerencias:

- ✓ Grafique los datos (Box-Whisker). Observe la tendencia de los datos. ¿Qué es evidente en la gráfica?
- ✓ Calcule estadísticos descriptivos por sitio. Analice la variabilidad al interior de cada sitio. ¿Qué le indica esto?
- ✓ Realice el análisis de varianza.
- ✓ Seleccione uno o dos métodos de comparaciones múltiples.
- ✓ Redacte su conclusión estadística y práctica. ¿Qué sugerencias haría usted en cuanto a la muestra?

8. El archivo ejercicio_p7.xlsx contiene 240 mediciones de altura (cm) para dos especies forestales a las cuales se le aplicaron los siguientes tratamientos:

Testigo: no recibió tierra micorrizada

Tierra 25%: Plántulas se trasplantaron en bolsas que contenían un 25% de tierra micorrizada.

Tierra 50%: Plántulas se trasplantaron en bolsas que contenían un 50% de tierra micorrizada.

Tierra 75%: Plántulas se trasplantaron en bolsas que contenían un 75% de tierra micorrizada.

Realice un análisis de varianza para dicho experimento. Recomendaría usted alguno de los tratamientos. ¿Por qué?

Sugerencias:

- ✓ Grafique los datos (Box-Whisker). Observe la tendencia de los datos. ¿Qué es evidente en la gráfica?
- ✓ Calcule estadísticos descriptivos por sitio. Analice la variabilidad al interior de cada sitio. ¿Qué le indica esto?
- ✓ Realice el análisis de varianza.
- ✓ Seleccione uno o dos métodos de comparaciones múltiples.
- ✓ Redacte su conclusión estadística y práctica.