

LO QUE LOS BIÓLOGOS PUEDEN USAR PARA ANALIZAR
SUS DATOS EXPERIMENTALES: LECCIÓN 3: ¿COMO USAR
ANÁLISIS DE VARIANZA PARA ALGUNOS CASOS COMUNES?

ANDEVA PARA DISEÑO COMPLETAMENTE AL AZAR (DCA)

TOOLS BIOLOGISTS MAY USE TO ANALYZE THEIR EXPERIMENTAL DATA: LESSON 3.
? HOW TO USE ANALYSIS OF VARIANCE IN SOME COMMON CASES? ANOVA FOR A
COMPLETELY RANDOM DESIGN (CRD)

Alexander Bustos B.Sc¹

Daniel Rodríguez Caicedo M.Sc.^{1,2}

Fernando Cantor Ph.D.¹

Foto Flickr

RESUMEN

El análisis de datos experimentales que provienen de diferentes tratamientos es una tarea usual en el trabajo de los biólogos tanto en campo como en el laboratorio. Sin embargo, la forma en que se analizan e interpretan esos datos, puede algunas veces omitir supuestos estadísticos para la aplicación de las pruebas o utilizar pruebas que no son las más adecuadas para el análisis de los datos. En esta lección presentamos el uso adecuado del análisis de varianza para comparar más de dos tratamientos en un experimento y elementos básicos para su interpretación utilizando datos experimentales de campo.

¹ Docente, Programa de Biología Aplicada, Facultad de Ciencias, Universidad Militar "Nueva Granada"

² Autor para correspondencia: daniel.rodriguez@umng.edu.co

SUMMARY

The analysis of experimental data coming from different treatments is usual for the work of biologists in both field and in the laboratory. However, the way they are analyzed and interpreted these data can sometimes omit statistical assumptions for the application of the tests or use tests that are not optimal for data analysis. In this lesson we present the proper use of the analysis of variance to compare more than two treatments in an experiment and basic elements for its interpreting using data of field experiment.

INTRODUCCIÓN

El desarrollo de experimentos en biología y en general en investigación tiene por objeto conocer y medir el efecto de algún tratamiento diseñado por el investigador sobre alguna variable de respuesta. De la cantidad de tratamientos a evaluar y sus características depende el análisis estadístico que se va a efectuar a los datos.

Rodríguez y otros (2006) presentan la prueba de t para comparar promedios de dos poblaciones. Sin embargo, cuando se desea comparar un número mayor de promedios, el uso de esa prueba no es recomendable debido a que se incrementa la probabilidad de cometer error tipo I, es decir rechazar la hipótesis nula de no diferencia entre los promedios aun que ésta sea cierta (Daniel, 1977).

De esa manera, si se desean efectuar todas las comparaciones posibles entre por ejemplo los promedios provenientes de un experimento en el cual se evaluaron cuatro tratamientos, sería necesario efectuar cuatro pruebas de t. Si

para cada una de ellas se asume una confianza $(1-\alpha) = 0,95$ la probabilidad de que las seis pruebas sean correctas es de $0,95^6 = 0,735$.

Por lo anterior lo mejor es utilizar una prueba que permita el análisis de todos los tratamientos al mismo tiempo y para esto se usa el análisis de varianza.

El análisis de varianza (ANDEVA)

El análisis de varianza, ANDEVA se utiliza para analizar resultados de experimentos en los que se han utilizado diseños como el Diseño Completamente al Azar (DCA), en experimentos factoriales con diseño DCA y en el diseño de Bloques Completos al Azar (BCA).

Cuando se diseñan experimentos bien sea en el campo o en el laboratorio es preciso, antes de iniciar el experimento, identificar las fuentes de variación que puedan afectar las mediciones de las variables pero que no son el interés principal del investigador y que se denominan como factores "nuisance" (NIST/SEMATECH 2006), con el fin de discriminar el efecto de tales fuentes y separarlo del efecto de los factores de interés. De esta manera es posible determinar cual diseño experimental se puede utilizar para medir el efecto de las fuentes de variación sobre las variables evaluadas en el ensayo.

El análisis de varianza permite estimar y evaluar hipótesis acerca de las medias de las poblaciones (Daniel, 1977). Sin embargo, es preciso recordar que las hipótesis a evaluar deben formularse antes de coleccionar los datos y no después. El plantear hipótesis después de coleccionar los datos puede invalidar la prueba estadística utilizada (Zar, 1999).

Al igual que para comparar los promedios de dos poblaciones con la prueba de t (ver Rodríguez et al, 2006), en el ANAVA también es

importante conocer los supuestos que deben cumplirse para usar correctamente esta prueba. Es decir para evaluar la hipótesis nula de no diferencia entre las medias de las poblaciones ($H_0: \mu_1 = \mu_2 = \mu_3 = \mu_n$) se asumen como supuestos que: 1). Las muestras fueron tomadas al azar, 2). Existe homogeneidad de varianzas, 3). Hay independencia en los errores, 4) Los errores se ajustan a una distribución normal y 5). Los efectos de los tratamientos son aditivos (Crawley, 1993)

Con el fin de determinar la cantidad nitrógeno total presente en la materia seca de plantas de papa, en un experimento bajo invernadero, se sembraron en macetas plásticas tres plantas de cada una de cinco variedades de papa (*Solanum phureja*).

Utilicemos los datos de porcentaje de nitrógeno total (NITRTOT) respecto a la materia seca encontrados en tubérculos de papa de cinco genotipos diferentes, para ilustrar mejor estos supuestos. El diseño corresponde a un Diseño Completamente al Azar (DCA) y lo primero que debemos confirmar es que hay homogeneidad de varianzas.

Se utilizó, en un número anterior, la prueba de F para evaluar homogeneidad de varianzas de dos poblaciones (ver Rodríguez et al, 2006). Sin embargo, al igual que no se recomienda el uso de la prueba de t para comparar promedios de más de dos poblaciones tampoco debe usarse la prueba de F para evaluar la homogeneidad de varianzas de más de dos poblaciones, en tal caso se usa la prueba de Barlet (Zar 1999).

A continuación se citan los comandos para importar datos, la explicación de cada uno se hizo en la lección 1 de esta serie.

```
> ejemploDCA<-read.table("c:/datos/DCA.txt",header=TRUE)
```

```
> attach(ejemploDCA)
> ejemploDCA
```

Ahora podemos utilizar la prueba de Bartlett para evaluar la hipótesis nula

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_5^2$$

Que plantea que existe homogeneidad de varianzas en las cinco poblaciones (cinco genotipos) de este ejemplo, así:

```
> bartlett.test(NITRTOT~GEN)
```

El comando "bartlett.test" permite realizar la prueba con los argumentos entre paréntesis (Figura 1) así: el primero es la variable de respuesta (NITRTOT) y el segundo separado por coma es la variable categórica (GEN).

El valor obtenido de $p = 0,3143$ es mayor a 0,05 es decir se acepta la hipótesis nula, entonces se concluye que existe homogeneidad de varianzas. También, se hace importante mencionar que esta prueba es muy sensible a violar el supuesto de normalidad.

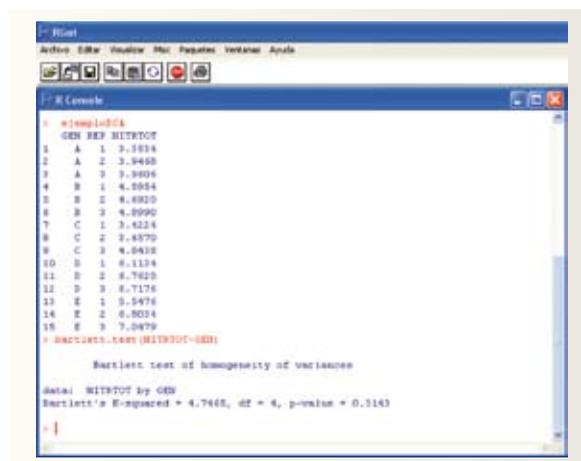


Figura 1. Prueba de Bartlett para comparar varianzas entre poblaciones.

Para evaluar la normalidad de los datos se puede utilizar la prueba de Shapiro, pero como esta prueba se realiza con los residuales del análisis de varianza, en el R debemos primero realizar el análisis de varianza en el que usamos el comando `aov()` que permite hacer inferencia acerca de los promedios de los tratamientos.

Entonces deseamos conocer si el contenido de nitrógeno total es igual en todos los genotipos es decir evaluar la hipótesis nula de no diferencia entre las medias de las poblaciones.

$$(H_0: \mu_1 = \mu_2 = \dots = \mu_5)$$

El comando que permite hacer el anova es: `aov()`. En el paréntesis los argumentos son: primero la variable de respuesta seguida del símbolo `~` (virgulilla) y luego la variable que categoriza los tratamientos en este caso genotipo (GEN). Utilizando punto y coma(;) podemos escribir otra línea de comandos en este caso usar la función `summary()` para obtener el resumen del `andeva`. Sin embargo, en este caso como el `summary` es de la prueba de anova y no de los datos originales, debemos poner en el paréntesis el nombre del objeto donde se guardó el anova, es decir ANDEVA para este ejemplo (Figura 2). La línea de comando queda así:

```
ANDEVA<-aov(NITRTOT~GEN);
summary(ANDEVA)
```

De esta manera obtenemos un valor de $p = 0,0001532$ *** mucho menor que 0,05 (los asteriscos indican que es altamente significativa la diferencia), en este caso, puesto que el valor de $p < 0,05$ se rechaza la hipótesis nula y por tanto puede decirse que por lo menos uno de los promedios

de una de las poblaciones es diferente de los demás. Los grados de libertad de GEN son iguales al número de categorías - 1 en este caso cinco genotipos - 1. Grados de Libertad = 4

Como ya tenemos el ANOVA podemos crear un objeto llamado RESIDUALES para incluir los residuales del ANOVA y con los residuales realizar la prueba de normalidad de Shapiro que permite evaluar la hipótesis nula de que los residuales siguen una distribución normal.

```
RESIDUALES<- residuals(ANDEVA)
> shapiro.test(RESIDUALES)
```

Shapiro-Wilk normality test

```
data: RESIDUALES
W = 0.9743, p-value = 0.9157
```

En este caso obtenemos un valor de W que es el estadístico teórico de prueba y debe compararse con valores tabulados en tablas y si el valor calculado es menor al tabulado se rechaza la hipótesis nula de normalidad. Sin embargo, en el R obtenemos directamente un valor de probabilidad que si es mayor a un valor teórico escogido, en nuestro caso 0,05, se acepta la hipótesis nula de normalidad de los residuales. Como obtuvimos normalidad en los residuales podemos entonces continuar nuestro análisis paramétrico para comparar los promedios de tratamientos.

Ahora sabemos que sí se presentaron diferencias pero no sabemos entre cuales tratamientos existen esas diferencias. Para averiguarlo podemos utilizar una prueba de comparación múltiple como la de Tukey para comparar por parejas de tratamientos pero utilizando el análisis de varianza.

PruebaTuk<-TukeyHSD(ANDEVA,
"GEN"); PruebaTuk

En el objeto PruebaTuk vamos a almacenar la prueba de Tukey cuyo comando en R es TukeyHSD(). Los argumentos en el paréntesis son el análisis de varianza que esta almacenado en el objeto ANDEVA y la variable o factor que define los tratamientos (GEN). Luego digitamos el nombre del objeto (PruebaTuk) para ver los resultados de la prueba (Figura 2). En la tabla se obtienen las probabilidades, en este caso igual que en el anova si el $p < 0,05$ existe diferencia significativa entre las dos poblaciones evaluadas.

Por ejemplo, al comparar Genotipo A con el D (A-D) se obtiene un valor de $p = 0,0007448$ es decir existe diferencia entre estas dos poblaciones. Por el contrario al comparar A con B (A-B) el valor de $p = 0,2623$ indica que entre estas dos poblaciones no existe diferencia estadísticamente significativa.

De la misma manera podemos ver gráficamente esas diferencias utilizando el comando Plot (PruebaTuk) para graficar la prueba (Figura 2).

Las líneas en la figura 2 muestran los intervalos de confianza al 95% para las diferencias de los promedios. La línea punteada representa el cero, de manera que aquellos intervalos que incluyen en su interior dicha línea, no se consideran estadísticamente diferentes de cero, o lo que es equivalente, no existe diferencia estadísticamente significativa entre la pareja de promedios comparados.

ANDEVA para diseños en bloques completos al azar (BCA)

Usemos los datos experimentales de parcelas de campo en los que se montaron bloques perpendiculares a un gradiente de humedad. Se evaluó la producción total de tubérculos en gramos por planta de papa, con tres dosis de un fertilizante, bajo diseño de bloques completos al azar. En este ejemplo es importante resaltar que la presencia de un gradiente de

humedad en las parcelas fue un factor que debía considerarse antes de iniciar el experimento pero que no interesaba evaluar su efecto en los tratamientos, factor "nuisance". Por tal razón se usaron bloques ubicados de manera perpendicular al gradiente de humedad (Figura 3)

De esta manera tenemos tres bloques y cada bloque tiene tres dosis de un fertilizante, es decir tenemos tres tratamientos con tres repeticiones, cada bloque tiene una repetición de cada tratamiento por esto son bloques

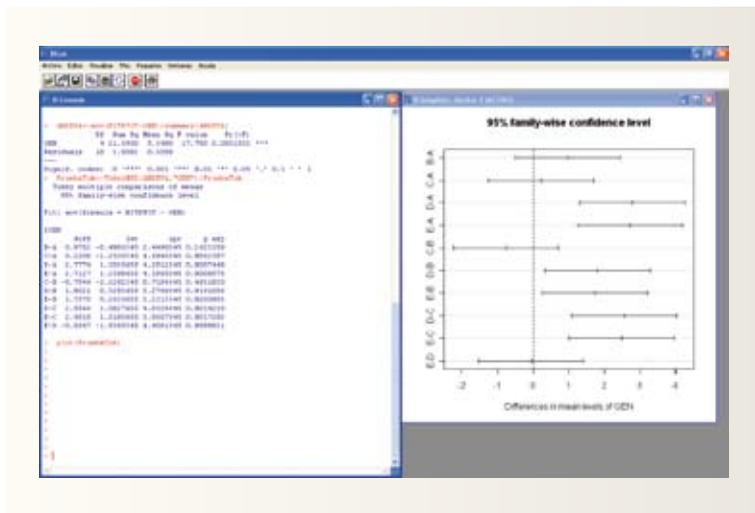
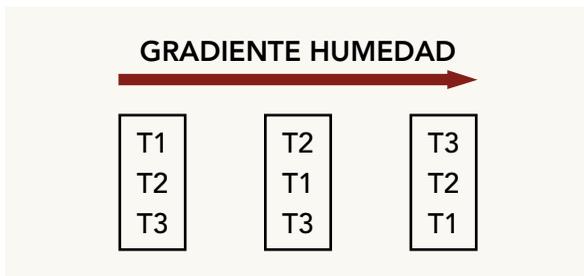


Figura 2. Análisis de varianza, prueba de comparación múltiple de Tukey y grafico de Tukey para los cinco genotipos del estudio.

	Bloque	Dosis	Gramos.planta
1	1	1	8100
2	1	2	15875
3	1	3	6800
4	2	1	13470
5	2	2	11150
6	2	3	7600
7	3	1	4550
8	3	2	9600
9	3	3	12225

completos y como dentro de cada bloque se asignaron los tratamientos de manera aleatoria el diseño se conoce como Diseño en Bloques Completos al Azar.



La pregunta que se quiere resolver es si existen diferencias en los promedios de los tratamientos evaluados respecto al peso de tubérculos de papa sometidos a tres tipos de fertilizante. En términos estadísticos la hipótesis nula es:

$$(H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5)$$

Los comandos en el R son:
`bloque<-as.factor(bloque)`
`dosis<-as.factor(dosis)`

Este comando permite tomar los valores numéricos de los bloques y dosis como factores

para que el programa (R) los tome como factores en el análisis.

```
andeva<-aov(gramos.planta~bloque
+dosis) > summary(andeva)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloque	2	6179117	3089558	0.170	0.8494
dosis	2	23401117	11700558	0.644	0.5722
Residuals	4	72677817	18169454		

La prueba de andeva en este caso arroja un $p > 0,05$ y por tanto se acepta la hipótesis nula de no diferencia entre los promedios de tratamientos, por esto no aparecen los códigos de significancia en la consola del R.

BIBLIOGRAFÍA

- 1- Crawley M. 1993. GLIM for ecologists Blackwell Scientific Publications. pp. 379
- 2- Daniel W. 1977. Base para el análisis de las ciencias de la salud. Editorial Limusa S. A. pp 193 - 237
- 3- Rodríguez D, Cantor F, Bustos A. 2006. Lo que los biólogos deberían conocer para analizar sus datos experimentales. Lección 1: Selección del software estadístico. Revista Facultad de Ciencias Básicas 3 (1)189 – 209.
- 3- ZarJ. 1999. Biostatistical analysis. Prentice Hall. Fourth edition pp 65 – 89.
- 4- NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, 30 Septiembre de 2008.