

Nota Técnica

MÉTODOS APROPIADOS DE ANÁLISIS ESTADÍSTICO SUBSIGUIENTES AL ANÁLISIS DE VARIANZA (ANDEVA)¹

Fabio A. Blanco*

Palabras clave: Análisis estadístico, varianza, ANDEVA

RESUMEN

Se presentan las bases que permiten elegir los métodos apropiados de análisis estadístico subsiguientes al análisis de varianza (ANDEVA) cuando los tratamientos son de efecto fijo. Se consideran 2 aspectos esenciales: 1) si el propósito es estimar parámetros o probar hipótesis con respecto a ellos y, 2) la estructura del conjunto de tratamientos experimentales. Se describen y dan ejemplos de conjuntos de tratamientos con estructura anidada (o grupal), de regresión (o gradiente), sin estructura y con estructura factorial. Además, se explica la relación entre cada caso de estructura de tratamientos y la técnica apropiada de análisis que le corresponde.

ABSTRACT

Appropriate statistical analyses methods subsequent to analysis of variance (ANOVA). The bases that allow to choose the appropriate statistical analyses following ANOVA are presented, for the case fixed effects treatments. Two essential aspects are taken into account: 1) whether the purpose of the researcher is to estimate parameters or to test hypothesis regarding them and, 2) the structure type of the set of experimental treatments. Treatment sets with either nested (also called group) or regression (also called gradient), without structure or with factorial structure are described, and examples of them are given. In addition, the relationship between each structure type and the corresponding appropriate statistical analysis is explained.

INTRODUCCION

En todas las ciencias se pueden utilizar los métodos estadísticos en las fases de planeamiento de la investigación, análisis, interpretación y presentación de los resultados. Chew (1976) fue uno de los primeros en alertar sobre la alta frecuencia con que los métodos estadísticos han sido

usados erróneamente en las publicaciones científicas. Frecuencias elevadas de uso inapropiado de los métodos estadísticos y deficiencias en los diseños experimentales fueron encontradas en publicaciones periódicas de prestigio como *Agronomy Journal* (Petersen 1977), *Phytopathology* (Johnson y Berger 1982), *Journal of Animal Science* (Lowry 1992), para mencionar sólo algunas.

1/ Recibido para publicación el 31 de julio del 2000.

* Escuela de Ciencias Agrarias, Universidad Nacional. Apartado postal 86-3000, Heredia, Costa Rica. Correo electrónico: fblanco@una.ac.cr

El problema del mal uso de los métodos estadísticos ha sido presentado aún con títulos atractivos como: "If Galileo published in HortScience" (Si Galileo publicara en HortScience) (Little 1978) o discutido en artículos satíricos como: "Baby Bear's dilemma: a statistical tale" (El dilema del Bebé Oso: un cuento estadístico) (Carmar y Walker 1982), y ha sido ampliamente debatido en publicaciones periódicas (ej. Journal of Economic Entomology (Perry 1986), Environmental Entomology (Jones 1984), Plant Disease (Swallow 1984)). Es posible que las deficiencias señaladas entonces en esas revistas sean menos frecuentes en la actualidad.

Una categoría de errores comunes en el uso de la estadística, consiste en la mala elección de técnicas de análisis posteriores al análisis univariado de varianza (ANDEVA en español, ANOVA en inglés). El ANDEVA es la técnica de análisis más conocida entre los investigadores agrícolas de los países de África, el Caribe y el Pacífico (Riley 1997) y posiblemente lo sea en América Latina. Consiste en partir la variación total en partes atribuibles a fuentes reconocidas (como tratamientos y bloques) y al error experimental. Pero, ¿qué hacer después del ANDEVA?

Por supuesto que el uso inapropiado de los métodos estadísticos tiene consecuencias prácticas. En muchos casos, simplemente impide que los investigadores alcancen los objetivos propuestos, complica la interpretación de los resultados, evita llegar a conclusiones correctas y resta habilidad para comunicar con sencillez los resultados y conclusiones (Perry 1986). Sobra decir que al lector también se le aumenta innecesariamente la complejidad de la información.

El propósito de este artículo es presentar los métodos apropiados de análisis estadístico subsiguientes al análisis de varianza univariado; no se considera el análisis multivariado.

Relación entre objetivos y métodos estadísticos

Los objetivos de una investigación inductiva se relacionan con los métodos estadísticos en 2 aspectos: primero, que los objetivos indican si el propósito es estimar parámetros (valores verdaderos) de poblaciones o probar hipótesis res-

pecto de ellas y segundo en que de los objetivos se deriva la estructura de los tratamientos y de estos los métodos estadísticos que deben emplearse para el análisis de los datos.

¿Estimación o prueba de hipótesis?

Los objetivos de una investigación pueden traducirse a preguntas que deben ser contestadas. Hay 2 categorías generales de preguntas (Pearce 1988):

Las que demandan una respuesta de "sí" o "no", por ejemplo: ¿es el herbicida A más efectivo que el B?

Las que demandan una respuesta cuantitativa, por ejemplo: ¿Cuánto más efectivo es el herbicida A que el B? Otro ejemplo: ¿Cuánto cambio habrá en el rendimiento de cierto cultivo por cada incremento de 1 kg/ha de fertilizante aplicado?

En el lenguaje estadístico, se responde a preguntas de la primera clase mediante *pruebas de hipótesis*, y a las de la segunda clase mediante *estimaciones*. Las *pruebas de hipótesis* más comunes son sobre parámetros de poblaciones. Las *estimaciones* también son de parámetros. Tanto estimación como prueba de hipótesis son procesos inductivos, es decir, se analizan muestras aleatorias y se concluye con respecto a poblaciones. La inferencia se hace respaldada con probabilidades.

En la fase de planeamiento de un experimento, todo investigador debería detenerse a contestar si su objetivo es *probar o estimar*, o ambos. A juicio de Chew (1976) a los investigadores se les ha inculcado en demasía la idea de *probar*, a expensas de la de *estimar*. En su opinión, hacer un experimento para *probar* si 2 o más tratamientos experimentales causan diferentes resultados (lo cual implica comparar los tratamientos), puede ser superfluo pues, casi seguramente, todos los tratamientos incluidos en un experimento difieren entre sí, aunque sea en el milésimo lugar decimal. "Así como 2 guisantes de una misma vaina no son idénticos, tampoco lo serán las medias de 2 tratamientos" (Chew 1976). Además, el hecho de que una prueba estadística, a un nivel de significancia dado, declare o no diferentes las medias de 2 o más tratamientos, es

tan sólo un asunto de precisión, pues una diferencia entre medias verdaderas de tratamientos, por pequeña que sea, puede llegarse a declarar significativa con solo aumentar el tamaño de muestra (número de repeticiones) tanto como sea necesario.

La falta de claridad en cuanto a si se desea *estimar o probar*, puede tener consecuencias nocivas al menos en 2 sentidos: primero en el diseño experimental, puesto que la *estimación* comúnmente demanda más repeticiones que la *comparación* (Pearce 1988) y, segundo, en que a veces no se efectúa la estimación, sino algún tipo de comparación, aunque el objetivo declarado haya sido *estimar*. En el primer caso la investigación adolecería de falta de precisión, mientras que en el segundo simplemente no se lograría el objetivo.

La estructura del diseño de tratamientos y el análisis estadístico

Las ideas expresadas de aquí en adelante tienen como contexto sólo a los conjuntos de tratamientos de efecto fijo, que son los que el investigador escoge específicamente por que le interesa estimar sus efectos. Estos casos constituyen cerca del 100% de las investigaciones donde se usa ANDEVA. No se aplican a los tratamientos de efecto aleatorio, que son aquellos que constituyen una muestra aleatoria de una población mayor de tratamientos, en cuyo caso el interés radica en estimar, no sus efectos particulares, sino la varianza entre estos.

Hecha la aclaración anterior, se puede afirmar que las preguntas a responder con una investigación determinan el conjunto de tratamientos y, automáticamente, el diseño de tratamientos. Urquhart (1981) estableció muy didácticamente la diferencia entre el diseño de tratamientos y el diseño experimental, desagregando de este a aquel; sus conceptos han sido ampliamente aceptados y difundidos. El diseño de tratamientos tiene que ver con la estructura de relaciones presente entre los tratamientos mientras que el diseño experimental especifica como se relacionan los tratamientos con las unidades experimentales. El análisis post-andeva depende del diseño de tratamientos. En el Cuadro 1 se cla-

sifican los diseños de tratamientos en *estructurados y no estructurados*, se dan ejemplos de cada clase de estructura y se indica la técnicas post-andeva correspondientes a cada caso.

No discernir las diferentes estructuras de tratamientos trae como consecuencia la ausencia de criterio para elegir las técnicas estadísticas apropiadas, lo cual resulta en errores frecuentes.

A continuación se describen las estructuras de los diseños y se dan ejemplos y pautas para el análisis estadístico posterior al ANDEVA.

Estructura grupal o anidada. La importancia de distinguir la estructura de grupos radica en que las preguntas que originaron el experimento se contestan comparando los grupos o subgrupos o estimando las diferencias entre ellos.

Como primer ejemplo considere el que se presenta en el Cuadro 1. Los 6 tratamientos pueden dividirse en 2 grupos: de un lado los primeros 5, con los cuales supuestamente se mejora la nutrición de las plantas y del otro el testigo. En el primer grupo se distinguen 2 subgrupos: los fertilizantes de un lado, y el inoculante del otro; este ejerce su acción por medio de microorganismos mientras que aquellos lo hacen directamente, mediante elementos químicos. Entre los 4 fertilizantes también se puede establecer una división de acuerdo con su naturaleza química: los orgánicos y los inorgánicos. Por último, dentro de los fertilizantes orgánicos hay 2 tratamientos, lo mismo que entre los inorgánicos.

Como un segundo ejemplo, considere un experimento con los tratamientos siguientes:

Testigo, sin herbicida.
Herbicida A, a 2 L/ha.
Herbicida A, a 3 L/ha.
Herbicida B, a 1 L/ha.
Herbicida B, a 2 L/ha.

Obviamente, los tratamientos 2-5 forman un grupo y tienen una relación entre sí, diferente a la que tienen con el tratamiento 1, que constituye otro grupo. Los tratamientos 2 al 5 se caracterizan porque todos consisten en aplicar herbicida, mientras que el tratamiento 1 consiste en no aplicarlo. A su vez, los tratamientos 2 y 3 tienen,

Cuadro 1. Clasificación de los diseños de tratamiento, ejemplos de diseños y técnicas apropiadas de análisis posterior al ANDEVA, con base en Urquhart (1981).

Diseño de tratamientos	Ejemplo	Técnicas apropiadas de análisis posterior al ANDEVA
Estructurado		
a. Con estructura grupal o anidada.	Seis tratamientos: 2 fertilizantes orgánicos, 2 fertilizantes inorgánicos, 1 inoculante 1 testigo.	Estimación o prueba de contrastes o comparaciones
b. Con estructura de gradiente o regresión.	Cuatro tratamientos: 0, 50, 100 y 200 mg/L de un regulador del crecimiento.	Análisis de regresión: Pruebas de modelos y/o estimación de modelos y/o parámetros
c. Con estructura factorial.	Seis tratamientos de 2 factores: a) 3 fertilizantes b) con o sin encalado del suelo.	Una o más de las técnicas de análisis citadas en esta columna, según sea la estructura de cada factor.
No Estructurado*		
(Tratamientos de efecto fijo)	Cinco tratamientos: 5 cultivares de frijol.	Separación de medias ** de tratamientos en grupos (pruebas como Tukey, Duncan, etc.) o estimación de diferencias.

* Se omite el diseño no estructurado con tratamientos de efecto aleatorio.

** Subsiste controversia sobre el uso de algunas técnicas de separación de medias y aún sobre la pertinencia de tal separación.

entre sí, un nexo específico, pues ambos son del herbicida A. Su nexo es distinto al que tienen con los tratamientos 4 y 5, que son del herbicida B. Aquellos forman un subgrupo y estos otro. Como las dosis del herbicida A no son las mismas que las del herbicida B se dice que las dosis están anidadas dentro de los herbicidas.

En los 2 experimentos citados, las preguntas de investigación que motivaron la escogencia de los tratamientos también determinaron la estructura de grupos. Esas preguntas se contestan comparando los grupos o subgrupos o estimando las diferencias entre ellos. En estos casos, la técnica estadística post-ANDEVA apropiada es el análisis de contrastes o comparaciones. Cualquiera otra técnica atentaría contra los objetivos de la investigación. No es requisito una prueba de F significativa en el ANDEVA; de hecho no hay necesidad de realizarla (Chew 1976). El ANDEVA, en estos casos, sólo

es útil para obtener el cuadrado medio del error que es la varianza estimada s^2 , la cual es útil en los análisis posteriores.

Estructura de gradiente o de regresión.

Considere ahora un conjunto hipotético de tratamientos que consiste de un solo herbicida aplicado en dosis crecientes, por ejemplo: 0, 1.0, 2.0 y 3.0 L/ha, o las 4 dosis de reguladores de crecimiento del Cuadro 1. Las dosis pueden o no estar igualmente espaciadas aunque, para facilidad del análisis estadístico, es deseable que lo estén. Estos diseños de tratamientos tienen *estructura de gradiente o de regresión*. Son de carácter cuantitativo. Otros ejemplos de esta clase son niveles crecientes de: temperatura, distancia de siembra, dosis de fertilizante, etc.

Salvo excepciones muy calificadas, los experimentos con esta estructura de tratamientos tienen objetivos como:

- a. Determinar si, dentro de cierto rango de valores de X (niveles de tratamiento, ej. dosis), la tendencia de la variable dependiente Y (variable de respuesta) es de forma lineal o curvilínea.
- b. Encontrar un modelo de ecuación matemática que razonablemente caracterice la relación entre la variable independiente X y la variable dependiente Y, y estimar los parámetros de la ecuación.
- c. Determinar el valor de X que causa cierto grado de efecto, como el máximo beneficio biológico o económico o el LD₅₀, que es la dosis que causa un cierto efecto (ej. la muerte) en el 50% de la población de individuos.

Cuando la estructura de tratamientos es de gradiente o regresión, los objetivos sólo se alcanzan mediante técnicas de regresión. El uso de comparaciones entre medias de tratamientos no tiene cabida en este contexto; hacerlo implicaría desviarse de los objetivos de la investigación. Lo afirmado para la estructura grupal se repite para la de regresión: no se necesita una prueba de F significativa en el ANDEVA, y ni siquiera hay necesidad de hacer la prueba; el ANDEVA, en estos casos, sólo es útil para obtener el cuadrado medio del error que es la varianza estimada s^2 .

Repitiendo la argumentación de Chew (1976), si los tratamientos fueran, por ejemplo, 10, 20, 30 y 40 mg/L de un producto químico y con los resultados se estimara la ecuación $Y=17.5 + 26.2 X$, "...eso significa que se espera un incremento en Y de 26.2 unidades por cada unidad de aumento de X. Así, no solo 10 y 20 mg/L dan resultados diferentes, sino también 10 y 10.1 mg/L".

No obstante lo anterior, debe reconocerse que en diseños de tratamientos con estructura de regresión, a veces es útil considerar una estructura de 2 grupos básicos: 1) el testigo (comúnmente el nivel cero) y, 2) el resto de los niveles. Si la estructura se mira de esa forma, se impone una comparación entre ambos grupos o una estimación de su diferencia y posteriormente un análisis de regresión excluyendo al testigo.

Conjuntos de tratamientos sin estructura. Un conjunto de tratamientos puede carecer de estructura. Un ejemplo lo constituye un grupo de genotipos de una especie que un investigador recibe para que los evalúe. Si en este caso no hay nexos particulares entre 2 o más tratamientos, que permitan establecer relaciones entre los tratamientos, entonces se dice que el conjunto de tratamientos es *no estructurado*. Teniendo presente que la estructura proviene de las preguntas que dan origen al conjunto de tratamientos, se concluye que tales preguntas no existen, sino que se desea poner los tratamientos a competir bajo las mismas condiciones con el propósito de escoger los mejores. Estos experimentos son de carácter preliminar; dan lugar a la formulación de preguntas e hipótesis, que a su vez originan diseños estructurados de tratamientos.

El análisis estadístico posterior al ANDEVA de experimentos cuyos tratamientos no tienen estructura, se restringe al caso en que la hipótesis general $H_0: m_1 = m_2 = \dots m_t$ ha sido rechazada durante el ANDEVA. A diferencia de los casos anteriores, este es un requisito. "Cuando la prueba del ANDEVA falla en detectar diferencias significativas entre medias de tratamientos, el experimento simplemente no provee suficiente evidencia para afirmar que existen diferencias entre los tratamientos. Por tanto, sin información adicional una prueba no significativa deja al investigador con la conclusión general de que las poblaciones tienen medias esencialmente iguales. Desde un punto de vista práctico, no se pueden hacer inferencias adicionales acerca de las medias de tratamientos" (Lentner y Bishop 1993).

Aparte de lo anterior, el análisis estadístico post-ANDEVA de experimentos sin estructura de tratamientos ha generado mucha controversia. Aunque la costumbre de comparar todas las medias entre sí mediante pruebas múltiples (de rango único como T o Tukey o de rango múltiple como Duncan o SNK) todavía está muy difundida, ciertamente ha perdido mucho terreno, y en su lugar han emergido alternativas no objetadas basadas en el análisis de conglomerados; un ejemplo de estas es la prueba de Bautista et al. (1997).

Estructura factorial. Un factorial puede estar constituido de p factores ($p= 2, 3, \dots$), cada uno a 2 o más niveles. En el ejemplo del Cuadro 1, el factor A (fertilizante) tiene 3 niveles, que pueden denotarse a_1, a_2 y a_3 , que son los 3 fertilizantes; el factor B tiene 2 niveles, que pueden denotarse b_1 y b_2 , que son respectivamente: con y sin encalado. Los tratamientos se forman obteniendo todas las combinaciones posibles entre los niveles de los factores, así: $a_1b_1, a_1b_2, a_2b_1, a_2b_2, a_3b_1, a_3b_2$. Se dice que este es un factorial 3×2 porque A tiene 3 niveles y B tiene 2. El resultado es 6 tratamientos factoriales.

Considere ahora los tratamientos 2 al 5 del segundo ejemplo de estructura grupal mencionado antes. Si las 2 dosis fueran las mismas para ambos herbicidas (por ejemplo 1 y 3 L/ha), entonces los tratamientos 2 al 5 tendrían una *estructura factorial* 2×2 ó 2^2 ; es decir, 2 factores (herbicidas y dosis) con 2 niveles (las 2 dosis) cada uno. Si se añade el tratamiento 1 (testigo), los 5 tratamientos serían un híbrido de las estructuras factorial y grupal, al cual se le llama *factorial aumentado* o *factorial con tratamientos adicionales*.

Cada factor puede ser de naturaleza cualitativa (herbicidas, formas de realizar una operación, variedades de un cultivo, sexo, etc.) o cuantitativa (cantidades de herbicida o de fertilizante aplicado, distancias de siembra, etc.) y tiene su propia estructura (grupal, de gradiente) o carece de ella.

Los experimentos factoriales deben analizarse como tales (Little 1978) para que el ANDEVA conteste si existe o no interacción entre los factores, y si los niveles de cada factor tienen o no efectos diferentes. Bajo ningún concepto se justifica que las medias de todos los tratamientos factoriales se comparen entre sí mediante alguna prueba de comparaciones múltiples (Duncan, Tukey, etc.). Esto significaría ignorar la estructura factorial y renunciar a las ventajas que ella ofrece.

Si por medio del ANDEVA se concluye que no hay interacción, eso significa que los efectos y las diferencias entre las medias (pero no necesariamente las medias) de los niveles a_1, a_2 , etc., del factor A se mantienen iguales, cualquiera que sean los niveles de B o en general de los otros factores. Por el contrario, si la interacción

existe, los efectos de a_1, a_2 , etc., y las diferencias entre las medias de estos niveles, dependen de cual sea el nivel de B o de los niveles de los otros factores.

El análisis post-ANDEVA de un experimento factorial depende de si la interacción es significativa o no y de cual sea la estructura de los niveles de cada factor. En presencia de interacción, los niveles de un factor particular deben examinarse por separado en cada nivel o combinaciones de niveles de los otros factores, mientras que en ausencia de interacción lo apropiado es proseguir con el examen de cada factor en forma independiente de los otros factores. En cualquiera de los casos, si un factor tiene estructura grupal se prueban los contrastes apropiados, si tiene estructura de gradiente se efectúa un análisis de regresión, sin importar, en ambos casos, lo que el ANDEVA indique con respecto a si las medias de sus niveles difieren o no. Por el contrario, si un factor es carente de estructura, antes de proseguir el análisis es indispensable constatar en el ANDEVA que las medias de los niveles difieren.

CONCLUSION

Debe reconocerse que existe una relación entre los objetivos de la investigación y los métodos estadísticos que deben emplearse. Por tanto, nadie debe tener por norma aplicar siempre el mismo método de análisis estadístico posterior al análisis de varianza (ANDEVA). De esta forma la herramienta estadística responderá las expectativas del investigador si las hipótesis de la experiencia están bien planteadas y el método experimental –incluyendo el estadístico– está bien planificado. Con lo que las consecuencias de los errores de juicio serán mínimas.

LITERATURA CITADA

- BAUTISTA M.G., SMITH D.W., STEINER R.L. 1997. A cluster-based approach to means separation. *Journal of Agricultural Biological and Environmental Statistics* 2(2):179-197.

- CARMER S.G., WALKER W.M. 1982. Baby Bear's Dilemma: A statistical tale. *Agronomy Journal* 74:122-124.
- CHEW V. 1976. Comparing treatment means: A compendium. *HortScience* 11(4):348-356.
- JOHNSON S.B., BERGER R.D. 1982. On the status of statistics in phytopathology. *Phytopathology* 72(8):1014-1015.
- JONES D. 1984. Use, misuse and role of multiple-comparison procedures an ecological and agricultural entomology. *Environmental Entomology* 13(3):635-649.
- LITTLE T.M. 1978. If Galileo published in Hort Science. *HortScience* 13(5):504-506.
- LENTNER M., BISHOPT. 1993. Experimental design and analysis. Valley Book. Company. Blacksburg, VA. 585 p.
- LOWRY S.R. 1992. Use and misuse of multiple comparisons in animal experiments. *Journal of Animal Science*. 70:1971-1977.
- PEARCE S.C. 1988. Analysis of data from agricultural experiments. *Tropical Agriculture (Trinidad)* 65(1):1-5.
- PERRY J.N. 1986. Multiple comparison procedures: A dissenting view. *Journal of Economic Entomology* 79(5):1149-1155.
- PETERSEN R.G. 1977. Use and misuse of multiple comparison procedures. *Agronomy Journal* 69:205-208.
- RILEY J. 1997. Utilización de la estadística en las ciencias biológicas: principales problemas y soluciones. *In: Enseñanza de la Biometría: problemas, experiencias y soluciones*. Ed. by J. Camacho. International Biometric Society Network for Central America and the Caribbean. p. 1-18.
- SWALLOW W.H. 1984. Those overworked and often misused mean separation procedures-Duncan's, LSD, etc. *Plant Disease* 68(10):919-921.
- URQUHART N.S. 1981. The anatomy of a study. *HortScience* 16(5):621-627.