

## Procedimientos de comparación múltiple; su uso indiscriminado en la investigación agropecuaria<sup>1</sup>

Luis Barrales V.<sup>2</sup> y Fernando Silva F.<sup>3</sup>

### INTRODUCCION

El análisis estadístico de la información obtenida de experimentos agronómicos, proporciona conocimientos acerca de la forma en que las unidades experimentales respondieron a la aplicación de tratamientos. En general, una primera etapa común es realizar un análisis de varianza, lo que permite dimitir la hipótesis de igualdad de las medias poblacionales de los tratamientos ensayados. Si la dicitima de F rechaza la hipótesis nula planteada, será necesario considerar el empleo de otras metodologías estadísticas que permitan explicar la naturaleza de la respuesta en más detalle. Petersen (1977) señala que de las técnicas estadísticas disponibles para este fin, son los Procedimientos de Comparación Múltiple (PCM) los usados con mayor frecuencia, y también más a menudo mal utilizados. Al respecto, señala que en el volumen 67 de *Agronomy Journal* (1975), el 40% de los autores usaron algunos de estos procedimientos. De ellos, el 40% los usó en forma enteramente inapropiada; el 30% los empleó no exactamente en una forma incorrecta, sino que los utilizó en experimentos para los cuales se dispone de metodologías más adecuadas, y sólo en el 30% de los artículos los experimentos presentaron las características para los cuales estos procedi-

mientos fueron concebidos. En relación a esto, Petersen (1977) enfatiza que los PCM son apropiados para aquellas situaciones en que los tratamientos son de un carácter cualitativo y no estructurados.

Little (1978) manifiesta que los PCM son aplicables en experimentos donde los tratamientos consisten de un conjunto de material no relacionado, tales como variedades o productos químicos.

Tal como se ha señalado, los PCM son una herramienta estadística que puede ser utilizada en situaciones muy específicas, dependiendo éstas de las características de los tratamientos. Sin embargo, los investigadores en el campo agronómico han utilizado estos procedimientos en forma indiscriminada. Esta situación no ha estado ausente de la investigación agropecuaria realizada en nuestro país.

El presente trabajo tiene por objeto puntualizar y evaluar las deficiencias en el uso de los PCM, para lo cual se efectuó una revisión de los artículos de investigación publicados en la revista *AGRICULTURA TÉCNICA*, desde 1965 a 1978, de donde se seleccionaron algunos ejemplos para ilustrar los casos más comunes de uso inapropiado de estas técnicas.

Los artículos seleccionados no serán citados por volumen, página o autores, ya que la intención de este trabajo no es de crítica a algunos autores por el empleo inadecuado de procedimientos que, también, han sido utilizados inapropiadamente por un gran número de investigadores extranjeros, sino evidenciar que ello produce una pérdida de eficiencia en la interpretación de resultados

<sup>1</sup>Recepción originales: 1º de octubre de 1979.

<sup>2</sup>Ing. Agr., M.S., Programa Biometría, Estación Experimental La Platina, Instituto de Investigaciones Agropecuarias (INIA), Casilla 5427, Santiago, Chile.

<sup>3</sup>Estadístico, Programa Biometría, Estación Experimental La Platina, Instituto de Investigaciones Agropecuarias (INIA), Casilla 5427, Santiago, Chile.

experimentales, y llamar la atención para la mejor elección de procedimientos estadísticos.

### DISCUSION

Las objeciones al uso de los PCM, no serán hechas con un enfoque estadístico matemático, sino que se objetará su empleo en aquellos casos donde metodologías más adecuadas son aplicables. Ante esta situación resulta irrelevante discutir los méritos relativos de los diversos PCM, los que ya han sido señalados por Keuls (1952), Tukey (1949), Duncan (1955) y Scheffé (1953), y revisados por O'Neill y Wetheril (1971).

Varios autores, entre ellos Chew (1976) y Petersen (1977), coinciden en señalar que los PCM casi nunca son apropiados para experimentos en los cuales:

1. Los tratamientos son niveles graduados de una variable cuantitativa.
2. Para tratamientos cualitativos, los cuales presentan cierta estructura.
3. Para tratamientos provenientes de la combinación factorial de dos o más factores a dos o más niveles.

Para cada una de estas situaciones existen técnicas estadísticas más apropiadas, que permiten una mayor eficiencia en la interpretación de los resultados experimentales.

En la revisión de los artículos publicados en Agricultura Técnica se encontró que un 31% de los autores utilizó algún tipo de comparación múltiple (generalmente Duncan). En el 50% de estos artículos los procedimientos fueron inapropiadamente usados.

#### *Uso apropiado de los Procedimientos de Comparación Múltiple*

Los tipos de experimentos para los cuales los PCM son apropiados, son aquellos cuyo objetivo es determinar el o los mejores tratamientos desde un conjunto cualitativo de ellos. En este tipo de experimentos el uso más común de estos procedimientos, es efectuar comparaciones de cada una de las medias con cada una de las otras medias de tratamientos, con el objeto de detectar posibles grupos entre un conjunto de tratamientos no estructurados.

A modo de ejemplo, se seleccionó un estudio acerca del comportamiento varietal en rendimiento de maíz dulce. Un extracto de

la información original se presenta en el Cuadro 1. Este ejemplo ilustra el uso adecuado de una prueba de comparación múltiple (Duncan), la que separa las variedades en cuatro grupos, dentro de los cuales los rendimientos no son significativamente diferentes.

**Cuadro 1 — Rendimiento en toneladas de mazorcas por hectárea, con 73% de humedad.**

<i>Variedad</i>	<i>Rendimiento</i>
1	33,6 a*
2	33,0 a
3	30,3 ab
4	29,5 ab
5	28,5 abc
6	26,7 bc
7	26,2 bc
8	25,4 bc
9	23,9 c
10	18,3 d
11	14,0 d

\*Valores seguidos por una misma letra no son significativamente diferentes para  $P = 0,05$ , según prueba de rango múltiple de Duncan.

Se pueden detectar los siguientes 4 grupos:

(1, 2, 3, 4, 5); (3, 4, 5, 6, 7, 8); (5, 6, 7, 8, 9); (10, 11)

#### *Usos inapropiados de los Procedimientos de Comparación Múltiple*

De los artículos que utilizaron un PCM en forma inapropiada, en el 59% de los casos fue aplicado a tratamientos consistentes en niveles graduados de un factor o factores cuantitativos (cantidad de fertilizantes, densidad de plantas, temperaturas, etc.).

El procedimiento estadístico más apropiado para esta situación, es el empleo de la metodología de superficie de respuesta que, en términos generales, consiste en ajustar, mediante técnicas de regresión, una función entre la respuesta (Y) y el o los factores en estudio (X).

Un ejemplo que permite ilustrar el problema es un estudio acerca del efecto de la aplicación de distintas dosis de nitrato sódico, sobre el rendimiento en raíces de remolacha azucarera.

Los Cuadros 2 y 3 muestran los análisis presentados por el autor en la publicación

original, sobre los cuales basa su discusión: "Los resultados indican la existencia de diferencias significativas entre los tratamientos

con nitrato de sodio y el testigo. Sin embargo, no se observan diferencias significativas entre los tratamientos que incluían diversas cantidades de nitrato".

Cuadro 2 — Análisis de varianza de los rendimientos (Ton/ha) de raíces de remolacha azucarera.

F. V.	Gl	SC	CM	F
Bloques	4	131,2	32,8	
Tratamientos	4	760,0	190,0	6,244**
Error Exp.	16	486,8	30,4	
Total	24	1.378,0		

\*\*Valor de F. significativo (P = 0,01).

Cuadro 3 — Efecto de los tratamientos sobre el rendimiento de raíces de remolacha azucarera (Ton/ha).

Tratamientos Kg/ha de nitrato sódico	Rendimiento Ton/ha
0	56,6 a
300	64,2 b
600	71,8 b
900	70,0 b
1200	69,4 b

El objetivo de este experimento fue relacionar las dosis de nitrato sódico con el rendimiento de raíces de remolacha azucarera. El usar un PCM no fue una forma adecuada de visualizar la relación, ya que estos procedimientos sólo permiten separar medias en grupos dentro de los cuales éstas no son significativamente diferentes.

Un análisis más adecuado es examinar los datos, a través de una relación funcional entre el rendimiento de raíces y las dosis de nitrato sódico. La ecuación resultante proporciona una descripción del cambio de rendimiento con las distintas dosis de nitrato sódico, no tan sólo con aquéllas que se incluyeron en el experimento, sino también con aquéllas dosis no ensayadas que están dentro del rango estudiado.

En el Cuadro 4 se presenta un análisis de varianza adecuado a esta situación, en el cual se descompone la suma de cuadrados de tratamientos en componentes, debido a la parte explicada por las contribuciones adicionales de las funciones polinómicas que se ajustan a los datos. De este análisis se concluye que una función de respuesta de tipo cuadrático sería la que mejor explica la respuesta del rendimiento, en relación a los niveles de nitrato sódico. La ecuación encontrada es:

Cuadro 4 — Análisis de varianza de los rendimientos de raíces de remolacha azucarera.

F. V.	Gl	SC	CM	F
Bloques	4	131,2	32,8	
Tratamientos	4	760,0	190,0	
efecto lineal	1	493,0		16,20*
efecto cuadrático	1	0,7		7,81*
efecto cúbico	1	237,8		0,02 n/s
desviación	1	28,5		0,94 n/s
Error Exp.	16	486,8	30,4	
Total	24	1.378,0		

\*Valores de F. Significativo (P = 0,05).

Ecuación de pronóstico basada en el análisis de varianza.

$$\hat{Y} = 56,43428 + 0,035938 (X) - 0,00002 (X^2)$$

$$\hat{Y} = 56,43428 + 0,035038 (X) - 0,00002 (X^2), \text{ con un } R^2 = 0,961.$$

El conocimiento cuantitativo de la relación, permite estimar el rendimiento promedio de raíces que podría haberse observado para dosis de nitrato sódico no incluidas en el experimento.

Es interesante destacar que en algunos casos particulares, dependiendo de las características del modelo estimado, es posible encontrar el valor óptimo de la variable independiente que determina el valor máximo de la variable respuesta.

Nada de la información que se ha descrito es proporcionado por el empleo de un PCM; sin embargo, el conocimiento obtenido de la aplicación de estos procedimientos, es también proporcionado por la función de respuesta. Al respecto, Chew (1976) señala que al comparar los efectos de los niveles graduados de un factor cuantitativo, y si la ecuación de respuesta es significativa, todos los niveles son significativamente diferentes

en sus efectos, sin importar la magnitud de la diferencia entre los niveles.

Otro de los problemas más frecuentes en el uso de los PCM, es el de aplicarlos a tratamientos cualitativos que presentan cierta estructura; este problema lo presentó el 14% de los autores que emplearon inapropiadamente estos procedimientos.

Cuando los tratamientos consisten de factores cualitativos (métodos de preparación de suelo, siembra, formas de fertilización, etc.), frecuentemente es posible formular con anterioridad a la realización del experimento, importantes comparaciones entre tratamientos o entre grupo de tratamientos, con el propósito de responderse ciertas preguntas específicas. Estas comparaciones son estimadas y probadas por medio de la metodología de contrastes ortogonales.

El experimento seleccionado para ilustrar esta situación es un ensayo en raps, cuyo objetivo fue estudiar el efecto de diferentes fertilizantes nitrogenados y el efecto de épocas de aplicación del nitrógeno. El Cuadro 5 pre-

Cuadro 5 — Rendimiento de grano del raps de primavera con diferentes fuentes y épocas de aplicación de N.

Tratamientos	Rendimiento qq/ha
T <sub>1</sub> Urea todo a la siembra	31,0 a*
T <sub>2</sub> Salitre todo a la siembra	28,2 a
T <sub>3</sub> Urea 1/2 a la siembra, salitre 1/2 a la florescencia	31,3 a
T <sub>4</sub> Salitre 1/2 a la siembra, salitre 1/2 a la florescencia	31,8 a
T <sub>5</sub> Urea 1/2 a la siembra, urea 1/2 a la florescencia	31,4 a
T <sub>6</sub> Sin nitrógeno	24,9 b

Contrastes (Ci) que debieron haber sido estudiados de acuerdo a los objetivos:

$$C_1 : \frac{T_1 + T_5}{\cdot} \text{ v/s } \frac{T_2 + T_4}{\cdot}$$

Urea v/s salitre.

$$C_2 : \frac{T_1 + T_3}{\cdot} \text{ v/s } \frac{T_4 + T_5}{\cdot}$$

Aplicación de Nitrógeno de una vez v/s parcializado.

$$C_3 : \frac{T_1 + T_2 + T_3 + T_4 + T_5}{\cdot} \text{ v/s } T_6$$

Aplicación de Nitrógeno v/s no aplicación de Nitrógeno.

\*Valores seguidos por una misma letra no son significativamente diferentes para P = 0.05 según prueba de rango múltiple de Duncan.

senta los tratamientos y el rendimiento en raps para cada uno de ellos. Se utilizó el procedimiento de Rango Múltiple de Duncan, lo que permitió al autor separar las medias de los tratamientos en dos grupos y, de acuerdo a esto, obtener las siguientes conclusiones:

- a) "No hubo diferencias estadísticamente significativas entre la úrea y el salitre".
- b) "No hubo diferencia estadísticamente significativa entre aplicar todo el nitrógeno con la siembra o en dos parcialidades".
- c) "Los tratamientos con nitrógeno fueron significativamente superiores al testigo sin nitrógeno".

El conocimiento del material experimental tal vez permitió al autor obtener estas conclusiones, pero no pudo lograrlas tomando como base el PCM empleado. El autor pudo haber obtenido estas conclusiones, pero usando otro procedimiento, como es el de establecer contrastes, como los presentados en el Cuadro 5; sin embargo, él no menciona en su artículo el empleo de esta metodología. Desgraciadamente tampoco presenta el análisis de varianza o el coeficiente de variación de su experimento, lo que imposibilitó obtener una estimación del error experimental y probar los contrastes de dicho cuadro, que sí le habrían permitido obtener las conclusiones por él mencionadas en la medida en que fueran no significativos los dos primeros y significativo el último.

Otro uso inapropiado común de los PCM, es aplicarlos a los tratamientos provenientes de una estructura factorial; esta situación representó el 27% de los artículos de la revista Agricultura Técnica, que presentaban algún problema en cuanto al uso de estos procedimientos. Este relativamente alto porcentaje se debe, en gran medida, a que muchas de las investigaciones en el campo agronómico son conducidas como experimentos factoriales, en los cuales los tratamientos están formados por todas las posibles combinaciones de los niveles de los factores, ya sean éstos cualitativos o cuantitativos.

El experimento factorial tiene por objetivo proporcionar conocimiento acerca de la forma en que actúan los factores en estudio sobre una variable de respuesta, pero no es su objetivo analizar el comportamiento de los tratamientos en su conjunto.

La primera etapa, en el análisis de un ex-

perimento factorial, es particionar la suma de cuadrados de tratamientos en componentes asociados con los efectos principales e interacciones de los distintos factores y, luego, probar estadísticamente estos contrastes. Una adecuada interpretación de los datos depende del resultado de estas pruebas de significación. Si las interacciones no son significativas, toda la información está contenida en los efectos principales; por lo tanto, es posible comparar los niveles de un factor, llamémoslo A, promediados sobre los niveles de los otros factores. La forma de comparar estos niveles dependerá de la característica del factor. En caso de un factor A cualitativo, los PCM son una técnica adecuada para estudiar el comportamiento de los niveles de este factor A. Si los niveles presentaran ciertas comparaciones de interés, una técnica más adecuada que la anterior sería la de contrastes ortogonales. En el caso que el factor A sea de un carácter cuantitativo, la estimación de funciones de respuesta sería el procedimiento más informativo; esto permitiría asociar la variación entre las medias de la variable de respuesta correspondiente a los niveles del factor A, con los cambios de nivel de la variable cuantitativa.

Si las interacciones son significativas, entonces ninguna información puede ser obtenida de los factores individuales en la forma descrita anteriormente, dado que ahora los factores no son independientes, lo que implica que los cambios en la respuesta, al compararse los niveles de un factor, dependerán del nivel en que se encuentra el factor interactuante; por lo tanto, no tiene sentido comparar los niveles de un factor A promediados sobre los niveles de un factor B, pero sí debieran compararse los niveles de A (utilizando la técnica que corresponda), para cada nivel de B separadamente, y viceversa.

Cabe hacer notar que, en lo expuesto, ninguno de los procedimientos descritos se ha aplicado a las medias de los tratamientos en su conjunto. De lo que se ha mencionado se desprende un gran número de situaciones a analizar, como también diversos procedimientos alternativos de enfocar el análisis, lo que hace impracticable el estudio particular de estas situaciones en un artículo como éste. Sin embargo, se seleccionó de la revista Agricultura Técnica un trabajo que presenta algunas características interesantes de describir.

El experimento elegido tiene por objetivo: "Especificar el efecto de las diferentes épocas y dosis de aplicación del ácido giberélico (AG) en vid sobre la producción por

plantas"; el autor, además, señala que se estudiaron cuatro dosis de AG aplicadas en cinco épocas diferentes. Los datos originales se presentan en el Cuadro 6.

Cuadro 6 — Efecto de las diferentes épocas y dosis de ácido giberélico (A.G.), en la producción por planta de vid (Kg), cultivar Moscatel Rosada.

Épocas de Aplicación	Dosis de (A G)			
	0 ppm	15 ppm	25 ppm	35 ppm
Inicio floración	13,13 def *	10,83 def	8,49 ef	9,78 ef
50% de floración	15,00 f	19,96 bcd	12,20 def	10,13 ef
75% de floración	11,00 def	22,69 bc	23,64 bc	34,35 a
3 días después de plena flor	12,73 def	25,08 bc	20,19 bcd	27,24 ab
1 semana después de plena flor	7,25 ef	16,43 cde	16,11 cde	23,13 bc

\*Valores seguidos por una misma letra no son significativamente diferentes para  $P = 0,05$  según prueba de rango múltiple.

Considerando los objetivos planteados por el investigador, él debió mencionar que se trataba de un experimento factorial, y como tal, realizar el análisis e interpretación estadística. Sin embargo, sólo se limita a emplear inapropiadamente el procedimiento de Rango Múltiple de Duncan para separar los promedios de tratamientos, con lo cual no cumple con los objetivos planteados. Una adecuada interpretación de los efectos de los factores basada en esta masa de letras resulta virtualmente imposible. Un ejemplo de esto es la forma en que el autor discute sus resultados: "En la variable producción por planta se aprecia que las mayores produccio-

nes se obtuvieron de aplicar 35 ppm de AG en 75% de floración y tres días después de plena flor".

Un apropiado análisis de esta situación experimental estaría basado en un análisis factorial, cuyos factores serían: épocas de aplicación, con cinco niveles, y dosis de AG estudiadas, a sólo tres niveles, ya que la combinación del nivel 0 ppm de AG con cada uno de los niveles del factor época de aplicación, no corresponde a una estructura factorial. Esto, debido a que la aplicación de 0 ppm de AG involucra la no existencia del factor época de aplicación. Un análisis de varianza apropiado se presenta en el Cuadro 7.

Cuadro 7 — Análisis de varianza adecuado a la situación experimental.

F.V.	Gl	SC	CM
Tratamientos	15	2.916,47	194,43
Testigo v/s trat. con AG	1	529,94	529,94
Épocas	4	1.793,36	448,34
Dosis	2	175,03	87,51
Épocas x Dosis	8	418,10	52,26

Promedios de producción por planta por épocas de aplicación.

Inicio floración	9,66
50% de floración	14,10
75% de floración	26,89
3 días después de plena flor	24,17
1 semana después de plena flor	18,16

Sin una estimación del error experimental, no es posible probar la significación de estos contrastes; sin embargo, juzgando la magnitud de los cuadrados medios, pareciera que la aplicación de AG incide en la producción por planta al comparar este promedio con aquél proveniente de las unidades experimentales que no recibieron AG. También pareciera significativo el efecto de la época en que se aplica el AG, pero no existiría una diferencia entre aplicar 15, 25 ó 35 ppm; además, este análisis indicaría que los factores en estudio son independientes. Por lo tanto, la información queda resumida en el factor época de aplicación del AG, que es un factor cualitativo y, en consecuencia, es apropiada la aplicación de un PCM que permita determinar la o las mejores épocas de aplicación. Observando las magnitudes de estos promedios puede señalarse que las mejores épocas de aplicación correspondieron a 75% de floración y tres días después de plena flor, independiente de la dosis de AG aplicada. Con este ejemplo se manifiesta claramente que los PCM son de poca ayuda en la interpretación de resultados experimentales, cuando se aplican inapropiadamente.

## CONCLUSIONES

Considerando el alto porcentaje de publicaciones que usaron inapropiadamente los Procedimientos de Comparación Múltiple, es evidente que existe una deficiencia en el empleo de éstas técnicas en la investigación agropecuaria realizada en nuestro país, siendo el grado de esta deficiencia similar al que presentan las publicaciones extranjeras.

La mayor deficiencia en el empleo de estas técnicas se detectó en aquellos trabajos en que los tratamientos consistieron en los niveles graduados de un factor cuantitativo. Esta situación era de esperar, ya que existe una errónea tendencia por parte de los investigadores, a utilizar metodologías estadísticas que les permitan discernir cualitativamente si un tratamiento es diferente a otro.

El uso indiscriminado de los Métodos de Comparaciones Múltiples, puede ocasionar pérdida de información y disminución de la eficiencia en la interpretación de los resultados experimentales.

## R E S U M E N

Se realizó una revisión completa de los artículos de investigación publicados en la revista AGRICULTURA TÉCNICA, desde 1965 a 1978, con el objeto de puntualizar y evaluar el uso indiscriminado de los Procedimientos de Comparación Múltiple.

El 31% de los autores utilizó algunos de estos procedimientos; en el 50% de estos artículos las metodologías fueron inadecuadamente usadas.

Ejemplos típicos de mal uso fueron seleccionados y se indicaron los análisis adecuados para cada situación. Estas ejemplarizaciones permitieron evidenciar que el uso indiscriminado de los Procedimientos de Comparación Múltiple produce una pérdida de la eficiencia en la interpretación de resultados experimentales.

## S U M M A R Y

### MULTIPLE COMPARISON PROCEDURES; ITS INDISCRIMINATE USE IN AGRICULTURAL RESEARCH.

A complete review of the contributions to the Chilean scientific journal AGRICULTURA TÉCNICA, from 1965 to 1978, was carried out to detect and evaluate the indiscriminate use of Multiple Comparison Procedures.

In 31% of the papers the authors used one of these tests. In 50% of these, the multiple range tests were inappropriate for the type of data involved.

Typical examples of misuse were selected and a more appropriate analysis was indicated for each case. These examples gave evidence that the indiscriminate use of multiple comparison tests may result in a reduced efficiency in the interpretation of the experimental data.

#### LITERATURA CITADA

- CHEW, V.** 1976. Comparing treatment means: a compendium. *Hort Science*, 11 (4): 348-357.
- DUNCAN, D. B.** 1955. Multiple range and multiple F tests. *Biometrics*, 11: 1-42.
- KEULS, M.** 1952. The use of the studentized range in connection with the analysis of variance. *Euphytica* 1: 112-122.
- LITTLE, T. M.** 1978. If Galileo published in *Hort Science*. *Hort Science*, 13 (5): 504-506.
- O'NEILL, R.** and **WETHERILL, G. B.** 1971. The present state of multiple comparison methods. *J. Roy. Stat. Soc., B*, 36: 218-250.
- PETERSEN, R. G.** 1977. Use and misuse of multiple comparison procedures. *Agronomy Journal*, 69: 205-208.
- SCHEFFÉ, H.** 1953. A method for judging all contrast in the analysis of variance. *Biometrika* 40: 87-104.
- TUKEY, J. W.** 1949. Comparing individual means in the analysis of variance. *Biometrics* 5: 99-114.