COMPARACIÓN DE LOS PROCEDIMIENTOS DE TUKEY, DUNCAN, DUNNETT, HSU Y BECHHOFER PARA SELECCIÓN DE MEDIAS

A COMPARISON OF TUKEY, DUNCAN, DUNNETT, HSU AND BECHHOFER PROCEDURES FOR SELECTION OF MEANS

Jesús A. **García-Villalpando**[1], Alberto **Castillo-Morales**[1,2], Martha Elva **Ramírez-Guzmán**[1], Gilberto **Rendón-Sánchez**[1] y Mario U. **Larqué-Saavedra**[3]

[1]Especialidad de Postgrado en Estadística. ISEI. Colegio de Postgraduados. 56230, Montecillo, Estado de México. Actualmente en [2]Departamento de Matemáticas. Universidad Autónoma Metropolitana. Unidad Iztapalapa. México, D. F. (acm@xanum.uam.mx). [3]Departamento de Sistemas. Universidad Autónoma Metropolitana. Unidad Azcapotzalco. México, D. F.

Resumen

Para seleccionar a los tratamientos con las mejores medias, los investigadores agrícolas en México utilizan el análisis de varianza y la prueba de Tukey o de Duncan. Estos dos métodos generan comparaciones múltiples de parejas de medias y no es claro si su uso es adecuado para seleccionar medias. El objetivo del presente trabajo fue determinar la habilidad para seleccionar medias de los procedimientos de Tukey y Duncan y del método de Dunnett (comparaciones múltiples con un testigo) en comparación con los métodos de Bechhofer y Hsu que son específicos para la selección de medias. Las comparaciones se hicieron en un modelo balanceado con un criterio de clasificación, utilizándose como referencia el método de Bechhofer. Los resultados mostraron que el método de Bechhofer fue el más adecuado para la selección de medias. El método de Dunnett coincidió en sus resultados numéricos con el de Hsu (comparaciones contra el mejor); ambos superaron a los métodos de Tukey y Duncan, por lo cual el método de Dunnett fue una buena alternativa. Entre los métodos de Tukey y Duncan, Duncan fue mejor, ya que el de Tukey fue demasiado conservador.

Palabras clave: Comparaciones múltiples, estadística aplicada, selección de tratamientos, separación de medias, zona de indiferencia.

Introducción

El análisis estadístico de datos derivados de un experimento tiene como propósito proveer información referente a la manera en que las unidades experimentales responden a los tratamientos aplicados. El primer paso consiste en someter los datos a un análisis de varianza para establecer si hay diferencias significativas entre las medias de los tratamientos. El rechazo de la hipótesis nula de igualdad de medias en el análisis de varianza, conduce a preguntar cuáles diferencias entre las medias muestrales son las responsables del rechazo.

Recibido: Marzo, 1999. Aprobado: Junio, 2000.
Publicado como ARTÍCULO en Agrociencia 35: 79-86. 2001.

Abstract

To select the treatments with the best means, agricultural researchers in Mexico use analysis of variance and the Tukey or Duncan tests. These two methods generate multiple comparisons of pairs of means, and it is not clear if they can adequately be used to select means. The objective of this study was to compare the ability of the Tukey and Duncan procedures and the Dunnett method (multiple comparisons versus a control) to select means with the methods of Bechhofer and Hsu, which are specifically designed to this end. Comparisons were done in a one-way balanced model using the Bechhofer method as a reference. Results showed that Bechhofer was the most suitable method for selection of means. The Dunnett method coincided in numerical results with that of Hsu (comparisons against the best), and both surpassed Tukey's and Duncan's methods. Hence, Dunnett was a good alternative. Between the Tukey and Duncan methods, Duncan was better since Tukey was too conservative.

Key words: Multiple comparisons, applied statistics, selection of treatments, separation of means, indifference zone.

Introduction

The statistical analysis of data derived from an experiment has the purpose of providing information on the way in which experimental units respond to the treatments applied. The first step is to subject the data to an analysis of variance to determine whether there are significant differences among the treatment means. Rejection of the null hypothesis of equality of means in the analysis of variance leads us to ask which differences between the sample means are responsible for the rejection. To answer this question, there are procedures of multiple comparisons between all the pairs of means, that are used even when the objective is to select the treatments with the best means. This objective is a different one, as is the solution. Tukey's, Duncan's and Dunnett's procedures for multiple comparisons are widely used in agricultural research and have been

Para ello existen procedimientos de comparaciones múltiples entre todas las parejas de medias, que se usan aun cuando el objetivo sea seleccionar a los tratamientos con las mejores medias. Este objetivo es diferente y su resolución también. Los procedimientos de comparaciones múltiples de Tukey, Duncan y Dunnett son ampliamente usados en investigación agrícola y están descritos en numerosos libros de metodología estadística (Miller, 1966).

Cuando el investigador efectúa un trabajo experimental para seleccionar el mejor de *k* tratamientos, esto es, el tratamiento con mayor (o menor) media, o seleccionar un subconjunto, lo más pequeño posible de entre los *k* tratamientos, que tenga una alta probabilidad de incluir al mejor o a los mejores tratamientos, no es adecuado que compare todas las parejas posibles de los tratamientos, sino que debe usar un método de selección de medias. En este caso, el análisis de varianza aporta la estimación de la varianza del error con sus grados de libertad y la información básica sobre si existen o no diferencias significativas en las medias de los tratamientos. Dos métodos específicos para la selección de medias son el de Bechhofer (1954) y el de Hsu (1981, 1984a,b).

Gupta y Sobel (1958), con el estadístico de Dunnett (1955), presentaron un procedimiento para seleccionar todos los tratamientos tan buenos o mejores que el tratamiento testigo. Gupta (1965) generó un procedimiento que no necesita formular una zona de indiferencia, pero, a cambio, el subconjunto seleccionado tiene un tamaño aleatorio.

Los procedimientos de Bechhofer (1954) y de Gupta (1965) están relacionados debido a que usan el mismo modelo, y por tanto, las expresiones para la probabilidad y las tablas correspondientes son similares (Lund, 1991).

Los procedimientos de comparación múltiple más comúnmente utilizados fueron analizados por Einot y Gabriel (1975) usando la potencia; ellos concluyeron que los métodos de Tukey y Scheffé tienen menor potencia, Duncan se ubica en un término medio y el de Ryan presenta la mayor potencia. Ellos recomiendan a Tukey por su simplicidad y por disponer de límites de confianza. Ramírez y Castillo (1985), al estudiar las zonas de rechazo de la prueba de F del análisis de varianza y cuatro pruebas de comparación múltiple, para el caso de tres tratamientos, ponen énfasis en los eventos en que la hipótesis de igualdad de tratamientos fue rechazada por el análisis de varianza pero no por la prueba de comparación, y viceversa. Ellos encontraron que las pruebas de Tukey y Dunnett tienen menor probabilidad de no coincidir con la prueba de F del análisis de varianza.

Lorenz *et al.* (1982) aplicaron los métodos de Tukey y de Hsu y encontraron que el método de Hsu seleccionó un subconjunto con menos tratamientos que el de Tukey.

described in numerous books on statistical methodology (Miller, 1966).

When the researcher carries out an experiment to select the best of *k* treatments, that is, the treatment with the highest (or lowest) mean, or to select the smallest possible subset among *k* treatments that has a high probability of containing the best treatment or treatments, it is not suitable to compare all of the possible pairs of treatments. Instead, he should use a method for selecting means. In this case, the analysis of variance provides an estimation of the variance of the error with its degrees of freedom and the basic information on whether or not significant differences exist among the means of the treatments. Two specific methods for selection of means are those of Bechhofer (1954) and Hsu (1981, 1984a,b).

Gupta and Sobel (1958), using the statistic of Dunnett (1955), presented a procedure to select all the treatments that are equal to or better than the control. Gupta (1965) created a procedure that does not need to formulate a zone of indifference but, in which, the selected subset has a random size.

The procedures of Bechhofer (1954) and Gupta (1965) are related because they use the same model, and so the expressions of probability and the corresponding tables are similar (Lund, 1991).

The most commonly used procedures for multiple comparisons were contrasted by Einot and Gabriel (1975) using the power. They concluded that Tukey's and Scheffé's methods were less powerful, Duncan's is intermediate and Ryan's is the most powerful. They recommend Tukey for its simplicity and because it allows to obtain confidence limits. Ramírez and Castillo (1985), in the study of zones of rejection of the F test in the analysis of variance and four tests of multiple comparisons for the case of three treatments, emphasize the events in which the hypothesis of treatments equality was rejected by the analysis of variance but not by the test of comparisons, and *vice versa.* They found that Tukey's and Dunnett's tests have a lower probability of not coinciding with the F test in the analysis of variance.

Lorenz *et al*. (1982) applied Tukey's and Hsu's methods and found that Hsu's method selected a subset with fewer treatments than did the Tukey method.

The previous considerations suggests the hypothesis that methods specifically designed for selection of means should be better for this purpose. The possibility remains that the methods of comparison of means commonly used by researchers make conservative or erroneous selections.

The objective of this study was to determine the ability to select means of the Tukey and Duncan procedures of multiple comparison and the Dunnett method (multiple comparison with a control) in regard to the methods of Bechhofer and Hsu, which are specific for selection of means.

Lo anterior sugiere la hipótesis de que los métodos específicos para la selección de medias deben ser mejores para este fin. Queda la posibilidad de que los métodos de comparaciones de medias, que son conocidos por los investigadores y se usan para seleccionar, hagan selecciones conservadoras o erróneas.

El objetivo del presente trabajo fue determinar la habilidad para seleccionar medias de los procedimientos de comparación múltiple de Tukey y Duncan y del método de Dunnett (comparaciones múltiples con un testigo), en comparación con los métodos de Bechhofer y Hsu, que son específicos para la selección de medias.

METODOLOGÍA

Procedimiento de Bechhofer

Este procedimiento es específico para seleccionar medias. No se plantea prueba de hipótesis, sino la probabilidad, P*, de que el tratamiento con la mayor media (parámetro) sea el que produzca la mayor media muestral.

En general se tienen k poblaciones ($k \geq 2$) denotadas por π_i (i=1,2,...,k), y se desconoce cuál población tiene la mayor media. Se denota por [1],[2],...,[k] a los subíndices desconocidos tales que $\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[k-1]} \leq \mu_{[k]}$, de manera que [] en el subíndice indica ordenamiento de acuerdo con los valores de los parámetros, pero se desconoce a qué subíndice i en la lista de poblaciones $\pi_1, \pi_2, \ldots, \pi_k$ o de medias poblacionales $\mu_1, \mu_2, \ldots, \mu_k$ corresponde el subíndice [k]. Las k medias muestrales también se pueden ordenar, obteniendo $\bar{Y}_{(1)} \leq \bar{Y}_{(2)} \leq \ldots \leq \bar{Y}_{(k-1)} \leq \bar{Y}_{(k)}$; debe notarse que el orden que las medias muestrales dan a las poblaciones, π_i, no es el mismo que el orden que producen las medias poblacionales μ_i, esto es, (i) puede ser diferente de [i]. Si se desea seleccionar una población, el procedimiento escoge la de mayor media muestral; esto es, a la población $\pi_{(k)}$. Se hace una selección correcta si [k]=(k). Para seleccionar t poblaciones, el procedimiento selecciona a las que tienen las t mayores medias muestrales, esto es, a $\pi_{[k]}, \pi_{[k-1]}, \ldots, \pi_{[k-t+1]}$.

Para identificar a la mejor media, Bechhofer (1954) requiere que $\mu_{[k]}$ difiera de $\mu_{[k-1]}$ por una cantidad δ>0. Delta (δ) genera una zona de indiferencia ya que si $\mu_{[k]} - \mu_{[k-1]} < \delta$, el procedimiento selecciona a una de las dos medias sin que entre ellas se distinga diferencia bajo la probabilidad establecida. Para asegurar la probabilidad de selección correcta, Bechhofer obtiene el tamaño de muestra para una configuración de las medias poblacionales conocida como la menos favorable, ya que es la que dificulta más separar a la mayor media de las demás. Esta configuración ocurre cuando la mayor media, la de la población [k], difiere por δ unidades de las demás medias, todas ellas iguales, esto es

$$\mu_{[k]} - \mu_{[k-1]} = \delta, \text{ y } \mu_{[1]} = \mu_{[2]} = \ldots = \mu_{[k-1]} \quad (1)$$

Bechhofer resuelve para el número de repeticiones el sistema de ecuaciones que resulta de

$$\text{P[selección correcta / } \mu_{[k]} - \mu_{[k-1]} \geq \delta \text{ y } \mu_{[1]} = \mu_{[2]} = \ldots = \mu_{[k-1]}] \geq \text{P*} \quad (2)$$

METHODOLGY

Bechhofer procedure

This procedure is specific for the selection of means. It does not propose to test hypothesis, but the probability, P*, that the treatment with the highest mean (parameter) is the one which produces the highest sample mean.

In general we have, k populations ($k \geq 2$), denoted by π_i (i=1,2,...,k), and it is unknown which population has the highest mean. [1],[2],...,[k], are unknown subindexes such that $\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[k-1]} \leq \mu_{[k]}$, and therefore [] in the subindex indicates an ordering according to the values of the parameters, but we ignore to which subindex i in the populations $\pi_1, \pi_2, \ldots, \pi_k$ or population means $\mu_1, \mu_2, \ldots, \mu_k$, the subindex [$k$] belongs. The k sample means can also be ordered, as $\bar{Y}_{(1)} \leq \bar{Y}_{(2)} \leq \ldots \leq \bar{Y}_{(k-1)} \leq \bar{Y}_{(k)}$. It should be noted that the order given to the populations, π_i, by the sample means, is not the same as that produced by the population means, μ_i. That is, (i) can be different from [i]. If a population is to be selected, the procedure chooses the one with the highest sample mean; that is, population $\pi_{(k)}$. A correct selection is made if [k]=(k). To select t populations, the procedure selects those that have the t highest sample means, that is, $\pi_{[k]}, \pi_{[k-1]}, \ldots, \pi_{[k-t+1]}$.

To identify the best mean, Bechhofer (1954) requires that $\mu_{[k]}$ differ from $\mu_{[k-1]}$ by a quantity δ>0. Delta (δ) generates a zone of indifference since if $\mu_{[k]} - \mu_{[k-1]} < \delta$, the procedure selects one of the two means without distinguishing a difference between them under the established probability. To assure the probability of correct selection, Bechhofer obtains the sample size for a configuration of population means known as the least favorable, since it is the configuration that makes more difficult to separate the highest mean from the rest. This configuration occurs when the highest mean, that of population [k], differs by d units from the rest of the means, all of them equal. That is

$$\mu_{[k]} - \mu_{[k-1]} = \delta, \text{ y } \mu_{[1]} = \mu_{[2]} = \ldots = \mu_{[k-1]} \quad (1)$$

Bechhofer solves the system of equations for the number of repetitions, which results from

$$\text{P[selección correcta / } \mu_{[k]} - \mu_{[k-1]} \geq \delta \text{ y } \mu_{[1]} = \mu_{[2]} = \ldots = \mu_{[k-1]}] \geq \text{P*} \quad (2)$$

He obtains

$$n = \left[\frac{\sigma d}{\delta} \right]^2 \quad (3)$$

where δ is the specified constant in (2) and d is a tabulated value that guarantees that [k]=(k), with probability P*. Bechhofer (1954) gives values of $d = \sqrt{n}\delta / \sigma$ associated with probabilities 0.05 (0.05)0.80, 0.82(0.02)0.90, 0.91(0.01)0.99, 0.995, 0.999 and 0.9995 for correct selection P*, for 2(1)14 treatments k, and to select 1(1) [k/2] t means. The nomenclature x(y)z means that from "x" value, progress is with steps of size "y" until arriving at "z" value; thus, 0.82(0.02)0.90 means values 0.82, 0.84, 0.86, 0.88, 0.90.

Obtiene

$$n = \left[\frac{\sigma d}{\delta}\right]^2 \qquad (3)$$

donde δ es la constante especificada en (2) y d es un valor tabulado que garantiza que $[k]=(k)$, con probabilidad P*. Bechhofer (1954) da valores de $d = \sqrt{n}\delta / \sigma$ asociados con probabilidades 0.05 (0.05) 0.80, 0.82 (0.02) 0.90, 0.91 (0.01) 0.99, 0.995, 0.999 y 0.9995 de selección correcta P*, para 2(1)14 tratamientos k, y para seleccionar 1(1)[k/2] medias t. La nomenclatura x(y)z quiere decir que a partir del valor "x" se avanza con pasos de tamaño "y" hasta llegar al valor "z", así, 0.82 (0.02) 0.90 quiere decir los valores 0.82, 0.84, 0.86, 0.88, 0.90.

Procedimiento de Hsu

Hsu (1981, 1984a,b) propone el método de comparaciones múltiples con el mejor (desconocido). Compara cada tratamiento con el mejor de los demás tratamientos. Las comparaciones de interés entre los parámetros son $\mu_i - \text{máx}_{j\neq i}\{\mu_j\}$ para $1\leq i\leq k$. Construyó intervalos de confianza simultáneos al nivel $1-\alpha$=P* para $\mu_i - \text{máx}_{j\neq i}\{\mu_j\}$, conocidos como intervalos de comparaciones múltiples con el mejor, los cuales están dados por

$$\mu_i - \text{máx}_{j\neq i}\,\mu_j \in \left[\text{mín}\left\{\left(\overline{Y}_i - \text{máx}_{j\neq i}\overline{Y}_j - DE\right),0\right\},\ \text{máx}\left\{\left(\overline{Y}_i - \text{máx}_{j\neq i}\overline{Y}_j + DE\right),0\right\}\right] (1\leq i\leq k) \qquad (4)$$

donde $DE = D(\alpha,k,v,1/2)S\sqrt{2/n}$; con S, la desviación estándar del error en el análisis de varianza; n, el número de repeticiones de cada tratamiento; y $D(\alpha,k,v,1/2)$ se distribuye como el máximo de k variables que siguen la distribución t multivariada con correlaciones iguales a 1/2; y $\alpha=1-$P*, k y v son nivel de significancia, número de tratamientos, y grados de libertad de S, respectivamente. Los valores tabulados para D se presentan en Dunnett (1955) para $k-1=1(1)9$. García (2000)[4] amplía las tablas para k=2(1)50 y k=55(5)100 tratamientos.

Gupta (1965) demostró que si se selecciona un subconjunto **S** de tratamientos de acuerdo con la regla

$$\mathbf{S} = \left\{i{:}\overline{Y}_i > \text{máx}_{j\neq i}\ \overline{Y}_j - DE\right\} \qquad (5)$$

entonces **S** contiene el mejor tratamiento con una probabilidad de al menos P*.

Comparación de procedimientos

En este caso se utilizó el procedimiento de Bechhofer como referencia para luego aplicar los procedimientos de separación de medias de Duncan, Tukey, Dunnett y Hsu.

Hsu's procedure

Hsu (1981, 1984a,b) proposes the method of multiple comparison with the best (unknown). He compares each treatment with the best of the rest of the treatments. The comparisons of interest among the parameters are $\mu_i - \text{max}_{j\neq i}\{\mu_j\}$ for $1\leq i\leq k$. He constructed simultaneous intervals of confidence at the level $1-\alpha$=P* para $\mu_i - \text{máx}_{j\neq i}\{\mu_j\}$, known as intervals of multiple comparisons with the best, which are given by

$$\mu_i - \text{máx}_{j\neq i}\,\mu_j \in \left[\text{mín}\left\{\left(\overline{Y}_i - \text{máx}_{j\neq i}\overline{Y}_j - DE\right),0\right\},\ \text{máx}\left\{\left(\overline{Y}_i - \text{máx}_{j\neq i}\overline{Y}_j + DE\right),0\right\}\right] (1\leq i\leq k) \qquad (4)$$

where $DE = D(\alpha,k,v,1/2)S\sqrt{2/n}$; with $S,$ the standard deviation of error in the analysis of variance, $n,$ the number of repetitions of each treatment, and $D(\alpha,k,v,1/2)$ is distributed as the maximum of k variables that follow the multivariate t distribution with correlations equal to 1/2; and $\alpha=1-$P*, k and v are level of significance, number of treatments, and degrees of freedom of S, respectively. Tabulated values for D are presented in Dunnett (1955) for $k-1=1(1)9$. García (2000)[4] expands the tables for k=2(1)50 and k=55(5)100 treatments.

Gupta (1965) demonstrated that if a subset **S** of treatments is selected following the rule

$$\mathbf{S} = \left\{i{:}\overline{Y}_i > \text{máx}_{j\neq i}\ \overline{Y}_j - DE\right\} \qquad (5)$$

then **S** contains the best treatment with a probability of at least P*.

Comparison of procedures

In this case the Bechhofer procedure was used as a reference to compare with the Duncan, Tukey, Dunnett and Hsu procedures for separation of means.

Experiments were simulated in order to separate the best treatment or treatments with the criteria of Duncan, Dunnett, Hsu and Tukey. The statistical package SAS (1988), version 6.03 with MS-DOS and SAS (1996), version 6.11 in Windows 95 were used in a PC with a Pentium processor at 200 Mhz and 16 Mb in RAM.

To carry out the simulations, the model $Y_{ia}=\mu_i+\varepsilon_{ia}(1\leq i\leq k,\ 1\leq a\leq n)$; was used; thus, for t treatments to have equal means and of greater magnitude, the first tn random observations were generated with the model:

$$Y_{ia} = \delta + \text{rannor}\ (.);\ i = 1,2,...,t,\ a = 1,2,...,n \qquad (6)$$

and the following k-t treatments that have smaller means equal to each other were generated by:

$$Y_{ia} = \text{rannor}\ (\cdot);\ i = t+1,\ t+2,...,k,\ a = 1,2,...,n \qquad (7)$$

[4] García, V. J. A. 2000. Comparación de los procedimientos de Tukey, Duncan, Dunnett, Hsu y Bechhofer para selección de medias. Tesis de Maestría en Ciencias. Colegio de Postgraduados, Montecillo, México. pp: 68-87.

Se efectuó una simulación de experimentos, con el fin de separar al mejor o mejores tratamientos, con los criterios de Duncan, Dunnett, Hsu y Tukey. Se utilizó el paquete estadístico SAS (1988), versión 6.03 bajo MS-DOS y SAS (1996), versión 6.11 bajo Windows 95, en una microcomputadora con procesador Pentium a 200 Mhz y 16 Mb en memoria RAM.

Para efectuar las simulaciones se consideró el modelo $Y_{ia}=\mu_i+\varepsilon_{ia}$ ($1\leq i\leq k$, $1\leq a\leq n$); así, para que t tratamientos tuvieran media igual y de mayor magnitud, las primeras tn observaciones aleatorias se generaron con el modelo:

$$Y_{ia}= \delta + \text{rannor}\ (.);\ i = 1,2,...,t,\ a = 1,2,...,n \qquad (6)$$

y los siguientes $k-t$ tratamientos que tienen menor media e igual entre ellos, se generaron con:

$$Y_{ia}= \text{rannor}\ (\cdot);\ i = t + 1,\ t + 2,...,k,\ a = 1,2,...,n \qquad (7)$$

donde rannor ($\cdot$) es la función de SAS que genera números aleatorios de una distribución normal estándar. Delta (δ) fue derivada de (3), como $\delta = d\sigma / \sqrt{n}$, con σ=1; donde n fue el número de repeticiones y d el valor de tablas de Bechhofer, de manera que el argumento probabilístico P* quedó satisfecho.

Lo anterior se hizo así para asegurar que los primeros t tratamientos fueran los de mayor magnitud de media. En este caso, el procedimiento de Bechhofer seleccionaría a los t tratamientos con mayores medias.

Para cada uno de los valores t=1 y t=2, con n=5 y k=10 se generaron 1000 series diferentes de datos aleatorios, 1000 experimentos, sujetos a lo descrito en (6) y (7). Posteriormente, para cada experimento, se obtuvieron las medias de cada tratamiento y el ordenamiento de las mismas en forma descendente. A partir del ordenamiento de las medias, y de acuerdo con cada uno de los cuatro procedimientos, se identificó al subconjunto de los tratamientos con mayor media e iguales entre sí.

Los datos requeridos para efectuar la simulación con t=1 (seleccionar una media) fueron:

a) $\delta = d\sigma / \sqrt{n}$: con d=3.4182 tomado de las tablas de Bechhofer (1954) para P*=0.95 (probabilidad de seleccionar correctamente la mejor media), k=10 tratamientos y t=1, con n=5 repeticiones y σ=1, por lo que, δ=3.4182(1)/$\sqrt{5}$ =1.52866.
b) Las medias muestrales ordenadas en forma descendente.
c) El cuadrado medio del error.
d) Los grados de libertad, en este caso ν=N$-k$=50$-$10=40, donde N es el número total de observaciones.
e) Un vector de valores críticos. En los casos de Dunnett, Hsu y Tukey fue un solo valor. Para el procedimiento de Duncan, por su naturaleza secuencial, los rangos estudentizados fueron extraídos para subconjuntos de tamaño p=2, 3,...,10.

Para seleccionar dos medias (t=2) se utilizó $\delta = d\sigma / \sqrt{n}$; con d=3.720 tomado de las mismas tablas. Para P*=0.95, k=10, t=2, con where rannor ($\cdot$) is the function of SAS that generates random numbers of a standard normal distribution. Delta (δ) was derived from (3), as $\delta = d\sigma / \sqrt{n}$, with σ=1; where n was the number of repetitions and d the value from the Bechhofer tables, so that the probabilistic argument P* is satisfied.

The aforementioned steps were carried out in this way to make sure that the first t treatments were those with the highest means. In this case, the Bechhofer procedure would select the t treatments with the highest means.

For each of the values t=1 and t=2, with n=5 and k=10, 1000 different series of random data, 1000 experiments, subject to that described in (6) and (7) were generated. Later, for each experiment, the means of each treatment were obtained and ordered from highest to lowest. With the means ordered, and according to each of the four procedures, the subset of treatments equal to each other and with the highest was identified.

The data required to carry out the simulation with t=1 (select one mean) were:

a) $\delta = d\sigma / \sqrt{n}$: with d=3.4182, taken from the tables in Bechhofer (1954) for P*=0.95 (probability of correctly selecting the best mean), k=10 treatments and t=1, with n=5 repetitions and σ=1, so that d =3.4182(1)/$\sqrt{5}$ =1.52866.
b) The sample means ordered from highest to lowest.
c) The error mean square.
d) Degrees of freedom, in this case ν=N$-k$=50$-$10=40, where N is the total number of observations.
e) A vector of critical values. In the cases of Dunnett, Hsu and Tukey, it was a single value. For the Duncan procedure, because of its sequential nature, studentized ranges were extracted for subsets of size p=2,3,...,10.

To select two means, (t=2), $\delta = d\sigma / \sqrt{n}$ was used with d=3.720 taken from the same tables; for P*=0.95, k=10, t=2, with n=5 and s=1, which gives δ=3.720(1)/$\sqrt{5}$ =1.66354. As in the case t=1; additionally the steps in paragraphs b), c), d) and e) were required.

Results and Discussion

In each type of experiment (t=1 and t=2), frequency in terms of cardinality of the selected subsets was determined for each method (Duncan, Dunnett, Hsu and Tukey).

For t=1, the procedures had to separate one treatment (Treatment 1). If they separated more treatments, the procedure was more conservative. The Dunnett and Hsu procedures produced the same results; thus, from this point on, only Dunnett's will be referred to, since it is the better known in agricultural research. Bechhofer always separated only one treatment and made the correct selection with the given probability (0.95). As can be seen in Table 1, Duncan correctly separated a single treatment, the best, in 21.4% of the cases; Dunnett

n=5 y σ=1, dando δ=3.720 (1)/ $\sqrt{5}$ =1.66354. Al igual que para t=1 se requirieron los incisos b), c), d) y e).

Resultados y Discusión

En cada tipo de experimento (t=1 y t=2) se determinó la frecuencia en cuanto a la cardinalidad de los subconjuntos seleccionados por cada método (Duncan, Dunnett, Hsu y Tukey).

Para t=1, los procedimientos debieron separar un tratamiento (Tratamiento 1); si separaron más tratamientos, el procedimiento resultó más conservador. Los procedimientos de Dunnett y de Hsu dieron los mismos resultados, así que en adelante sólo se mencionará a Dunnett, por ser más conocido en la investigación agrícola. Bechhofer separó siempre un tratamiento e hizo la selección correcta con la probabilidad prevista (0.95). De acuerdo con el Cuadro 1, Duncan separó correctamente un solo tratamiento, el mejor, en 21.4% de los casos, Dunnett lo hizo en 9.8% y Tukey 1.3%, este último resulta demasiado conservador.

Ninguno de los procedimientos cometió el error de no incluir en su selección a la mejor media. El error de separar más medias que las que la situación requiere (una, en este caso) se cometió en 78.6% de las veces por Duncan, 90.2% por Dunnett y 98.7% por Tukey.

Para t=2, los procedimientos debieron separar dos tratamientos (Tratamiento 1 y Tratamiento 2, simultáneamente). En el Cuadro 2 se observa que Duncan separó a los dos mejores tratamientos 41.2% de las veces; Dunnett, 26.2%; mientras que Tukey lo hizo 4.8%. Al igual que para t=1, Tukey resultó muy conservador.

El error de seleccionar más de dos medias tuvo porcentajes de 53.7, 72.4 y 95.2% para Duncan, Dunnett y Tukey, respectivamente.

Duncan cometió el error de no incluir las dos mejores medias de los tratamientos en 5.1% de los casos; Dunnett, en 1.4%; y Tukey no cometió este error. El 5.1% de error en Duncan es similar al 5% que comete Bechhofer. Por su parte, Hsu (1984a) indica que su procedimiento comete menos errores que Bechhofer, lo mismo que pasó con Dunnett, que produjo los mismos resultados que el método de Hsu.

La comparación de los métodos de Dunnett ó Hsu contra Tukey (Cuadros 1 y 2) confirma lo indicado por Lorenz *et al*. (1982), en donde Hsu fue menos conservador, ya que separó subconjuntos con menos tratamientos que el de Tukey. Esto enfatiza la inconveniencia de utilizar el método de Tukey para seleccionar medias.

Cuando se generaron datos de uno (t=1) y dos (t=2) promedios mayores, se verificó la probabilidad de error del procedimiento de Bechhofer. Así, cuando t=1 y t=2 se contabilizaron 47 y 50 casos, respectivamente, en cada uno de los 1000 experimentos, en los que no se separó un

did so in 9.8% and Tukey in 1.3%, Tukey being too conservative.

None of the procedures committed the error of not including the best mean in its selection. The error of separating more means than the situation requires (one, in this case) was committed 78.6% of the times by Duncan, 90.2% by Dunnett and 98.7% by Tukey.

For t=2, the procedures had to separate two treatments (Treatment 1 and Treatment 2 simultaneously). In Table 2, it can be observed that Duncan separated the two best treatments 41.2% of the times, Dunnett 26.2%, while Tukey did so 4.8%. As in t=1, Tukey turned out to be too conservative.

The error of selecting more than two means had percentages of 53.7, 72.4 and 95.2% for Duncan, Dunnett and Tukey, respectively.

Duncan committed the error of not including the two best means of the treatments in 5.1% of the cases, Dunnett in 1.4%, and Tukey did not commit this error. The 5.1% error in Duncan is similar to the 5% error that Bechhofer commits. Hsu (1984a) reports that his procedure commits fewer errors than Bechhofer, as did Dunnett, whose method produced the same results as Hsu.

The comparison of the Dunnett or Hsu methods with Tukey (Tables 1 and 2) confirm what Lorenz *et al*. (1982) reported, that Hsu was less conservative since it separated subsets with fewer treatments than Tukey. This stresses the inconvenience of using the Tukey method for selecting means.

When the data for one (t=1) and two (t=2) largest averages are generated, the probability of error of the Bechhofer procedure was verified. Thus, when t=1 and t=2, 47 and 50 cases, respectively, were counted in each

Cuadro 1. Porcentajes para el número de tratamientos seleccionados como los mejores por tres procedimientos de separación de medias en 1000 experimentos simulados con k=10 tratamientos, para t=1 tratamientos a seleccionar.
Table 1. Percentages for the number of treatments selected as the best by three procedures of separation of means in 1000 experiments simulated with k=10 treatments for t=1 treatments to be selected.

Número de tratamientos seleccionados como el (los) mejor(es)	Duncan	Dunnett %	Tukey
1	21.4	9.8	1.3
2	11.2	9.0	1.6
3	9.2	8.9	2.6
4	9.0	9.4	4.2
5	8.4	10.5	4.7
6	7.2	9.3	6.4
7	8.3	10.3	7.8
8	8.1	10.2	11.3
9	9.6	11.9	16.9
10	7.6	10.7	43.2
Total	100.0	100.0	100.0

Cuadro 2. Porcentajes para el número de tratamientos seleccionados como los mejores por tres procedimientos de separación de medias en 1000 experimentos simulados con *k*=10 tratamientos, para *t*=2 tratamientos a seleccionar.

Table 2. Percentages for the number of treatments selected as the best by three procedures of separation of means in 1000 experiments simulated with *k*=10 treatments for *t*=2 treatments to be selected.

Número de tratamientos seleccionados como el (los) mejor(es)	Duncan	Dunnett	Tukey
		%	
1	5.1	1.4	0.0
2	41.2	26.2	4.8
3	17.0	18.7	6.9
4	11.4	14.4	8.1
5	8.4	12.0	11.1
6	5.4	9.0	10.5
7	4.6	7.4	11.1
8	3.8	5.7	13.7
9	2.0	3.3	16.5
10	1.1	1.9	17.3
Total	100.0	100.0	100.0

conjunto que incluyera al mejor tratamiento (Tratamiento 1) o a los dos mejores tratamientos (Tratamiento 1 y Tratamiento 2). El Cuadro 3 presenta los porcentajes del número de tratamientos seleccionados de estos casos para t=1 y el Cuadro 4 para t=2.

Cuando ocurre que la media del mejor tratamiento no es la mayor, los procedimientos de Duncan, Dunnett y Tukey aumentan el número de medias seleccionadas, con incrementos mayores en Tukey (Cuadro 3).

of the 1000 experiments in which a subset including the best treatment (Treatment 1) or the two best treatments (Treatment 1 and Treatment 2) was not separated. Table 3 presents the percentages of the number of treatments selected of these cases for t=1 and Table 4 does so for t=2.

When it happens that the mean of the best treatment is not the highest, Duncan, Dunnett and Tukey increase the number of selected means, Tukey having the highest increments (Table 3).

The Bechhofer method is exact in its probability of selection error. However, it can be impractical since it produces large sample sizes. In practice, when there are conditions that impede using more than a certain number of repetitions, once the value of correct selection probability is given, d is retrieved from the tables to solve for d in (3) and thus the width of the zone of indifference for the experiment being planned is determined. This can also be applied to find the selection capacity of the experiments already carried out.

Conclusions

To facilitate the use of the Bechhofer procedure, it is advisable to obtain the value of d from the number of repetitions and the probability of correct selection the researcher chooses. Among the Tukey and Duncan methods, the latter was better since it selects subsets with fewer treatments; Tukey was too conservative. The Dunnett and Hsu procedures coincided in separating

Cuadro 3. Porcentajes para el número de tratamientos seleccionados como los mejores por tres procedimientos de separación de medias en 47 experimentos simulados con *k*=10 tratamientos para seleccionar *t*=1 tratamiento cuando, por efectos de la simulación aleatoria, el Tratamiento 1 no tuvo la mayor media.

Table 3. Percentages for the number of treatments selected as the best by three procedures of separation of means in 47 experiments simulated with *k*=10 treatments to select *t*=1 treatment when, because of effects of random simulation, Treatment 1 did not have the highest mean.

Número de tratamientos seleccionados como el (los) mejor(es)	Duncan	Dunnett	Tukey
		%	
1	0.00	0.00	0.00
2	4.26	2.13	0.00
3	0.00	2.13	0.00
4	0.00	0.00	0.00
5	6.38	2.13	0.00
6	4.26	2.13	0.00
7	8.51	8.51	0.00
8	17.02	12.77	4.26
9	27.66	23.40	10.63
10	31.91	46.80	85.11
Total	100.00	100.00	100.00

Cuadro 4. Porcentajes para el número de tratamientos seleccionados como los mejores por tres procedimientos de separación de medias en 50 experimentos simulados con *k*=10 tratamientos para seleccionar *t*=2 tratamientos cuando, por efectos de la simulación aleatoria, los Tratamientos 1 y 2 no tuvieron las dos mayores medias.

Table 4. Percentages for the number of treatments selected as the best by three procedures of separation of means in 50 experiments simulated with *k*=10 treatments to select *t*=2 treatments when, because of effects of random simulation, Treatments 1 and 2 did not have the two highest means.

Número de tratamientos seleccionados como el (los) mejor(es)	Duncan	Dunnett	Tukey
		%	
1	14	4	0
2	2	6	0
3	22	10	4
4	18	20	4
5	10	16	12
6	8	12	10
7	8	12	6
8	6	8	14
9	8	6	20
10	4	6	30
Total	100	100	100

El método de Bechhofer es exacto en su probabilidad de error de selección, sin embargo, puede resultar poco práctico, ya que produce tamaños de muestra grandes. Cuando en la práctica se tengan condiciones que impiden usar más de un cierto número de repeticiones, una vez que se fija el valor de la probabilidad de selección correcta, de las tablas se obtiene d para despejar δ en (3) y así conocer la amplitud de la zona de indiferencia que se tiene para el experimento planeado; esto también se puede aplicar para conocer la capacidad de selección de experimentos realizados.

Conclusiones

Para facilitar el uso del procedimiento de Bechhofer conviene obtener el valor de δ a partir del número de repeticiones y la probabilidad de selección correcta deseada por el investigador. Entre los métodos de Tukey y Duncan, el segundo fue mejor pues selecciona conjuntos con menos tratamientos; el de Tukey fue demasiado conservador. Los procedimientos de Dunnett y Hsu coincidieron en la separación de medias, cuando la comparación en Dunnett se hizo tomando como testigo al tratamiento con mayor media. Para propósitos de selección de medias con los valores de t y k usados, la prueba de Duncan, que selecciona al grupo de medias iguales que el tratamiento con mayor media, tuvo la ventaja, ante Dunnett, de seleccionar grupos más reducidos de medias, pero cometió el error de no incluir en la selección al mejor tratamiento, con mayor frecuencia que Dunnett. Cuando se desea seleccionar un número establecido de tratamientos el método adecuado es el de Bechhofer. Si se desea seleccionar a las mejores medias desconociendo el número de ellas, cuidando que se mantengan las probabilidades de error conocidas, puede emplearse el método de Hsu o el de Dunnett, que seleccionan al conjunto de tratamientos con media igual al control, este último el de mayor media. El método de Duncan separó conjuntos de menos tratamientos al seleccionar los tratamientos con igual media que el mayor, mientras que el método de Tukey separa conjuntos con mayor número de tratamientos.

Agradecimiento

Los autores agradecen las atinadas sugerencias del editor, ya que permitieron enriquecer la calidad del trabajo.

Literatura citada

Bechhofer, R. E. 1954. A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Stat. 25: 16-39.

Dunnett, C. W. 1955. A multiple comparison procedure for comparing several treatments with a control. J. Am. Stat. Assoc. 70: 574-583.

means when comparison in Dunnett was done using the treatment with the highest mean as the control. For purposes of selection of means with the values t and k used, the Duncan test, which selects the groups of means equal to the treatment with the highest mean, had the advantage of selecting smaller groups of means than those selected by Dunnett. But it committed the error of not including the best treatment in the selection more frequently than Dunnett. When the researcher wants to select a set number of treatments, the suitable method is that of Bechhofer. If he wants to select the best means, regardless of their number but being careful of maintaining known probabilities of error, the Hsu or Dunnett methods can be used. These select the set of treatments with the mean that is equal to that of the control, which has the highest mean. The Duncan method separated sets of fewer treatments, selecting the treatments with means equal to the highest, while the Tukey method separates sets with the largest number of treatments.

—End of the English version—

Einot, I., and K. R. Gabriel. 1975. A study of the powers of several methods of multiple comparisons. J. Am. Stat. Assoc. 70: 574-583.

Gupta, S. S., and M. Sobel. 1958. On selecting a subset which contains all populations better than a standard. Ann. Math. Stat. 29: 235-244.

Gupta, S. S. 1965. On some multiple decision (selection and ranking) rules. Technometrics 7: 225-245.

Hsu, J. C. 1981. Simultaneous confidence intervals for all distances from the best. Ann. Stat. 9: 1026-1034.

Hsu, J. C. 1984a. Ranking and selection and multiple comparisons with the best. *In*: Design of Experiments: Ranking and Selection. Santner, T. J., and A. C. Tamhane (eds.). Marcel Dekker. New York. pp: 23-33.

Hsu, J. C. 1984b. Constrained simultaneous confidence intervals for multiple comparisons with the best. Ann. Stat. 12: 1136-1144.

Lorenz, R. C., J. C. Hsu, and O. H. Tuovinen. 1982. Performance variability, ranking, and selection analysis of membrane filters for enumerating coliform bacteria in river water. J. Am. Water Works Assoc. 74: 429-437.

Lund, R. E. 1991. Algorithm AS267. Probabilities and standardized differences for selecting subsets containing the best population. J. Royal Stat. Soc. App. Stat. 40: 495-502.

Miller, R. G. 1966. Simultaneous Statistical Inference. McGraw-Hill. New York. 272 p.

Ramírez V., B. y A. Castillo M. 1985. Estudio de las zonas de rechazo del análisis de varianza y algunas pruebas de comparaciones múltiples, para el caso de tres medias. Agrociencia Núm. 61: 65-78.

SAS Institute. 1988. SAS Language Guide for Personal Computers. Release 6.03 ed. Cary, NC. SAS Institute Inc. 558 p.

SAS Institute. 1996. SAS/STAT Software: Changes and Enhancements. Through Release 6.11. Cary, NC. SAS Institute Inc. 1104 p.