UNIVERSIDAD DE CARABOBO

FACULTAD DE INGENIERÍA

ESTUDIOS BÁSICOS DE INGENIERÍA

DEPARTAMENTO DE MATEMÁTICA

MEDICIÓN DE LA POTENCIA DE PRUEBAS DE COMPARACIÓN MÚLTIPLE DE MEDIAS PARA EXPERIMENTOS DESBALANCEADOS CON EL USO DE SIMULACIÓN DE MUESTRAS

Trabajo presentado ante el Ilustre Consejo de Facultad de Ingeniería de la Universidad de Carabobo como credencial de mérito para el ascenso dentro del escalafón del Personal Docente y de Investigación a la categoría de Profesor Asociado

Autores:
Ing. Eduardo E. Vargas C.
Ing. Edwin E. Vargas C.

Bárbula, 19 de Diciembre de 2013

INDICE GENERAL

INTRODUCCIÓN

El presente trabajo de investigación titulado "**Medición de la Potencia de Pruebas de Comparación múltiple de medias para experimentos desbalanceados con el uso de simulación de muestras**", se propone suministrar a los investigadores de prácticamente todos los campos de estudio que llevan a cabo experimentos, por lo general para descubrir algo acerca de un proceso o un sistema, una información para disminuir el esfuerzo por seleccionar un procedimiento estadístico que más se ajuste al análisis de los datos y su interpretación, ya que proporciona información concerniente al comportamiento de los procedimiento de comparación múltiple de medias de Mínima Diferencia Significativa, Tukey y Student-Newman-Keuls.

El contenido de la investigación, detalla desde el punto de vista teórico, 12 pruebas de comparación múltiple de medias, haciendo mención de los estadísticos de pruebas, desviaciones estándar y las reglas de decisión.

La estructura del informe es la siguiente:

1. Planteamiento del problema.
2. Descripción teórica de los procedimientos de comparación múltiple de medias.
3. Descripción de la metodología empleada para llevar a cabo la investigación.
4. Discusión y Análisis de resultados.
5. Conclusiones y recomendaciones.

CAPITULO I

EL PROBLEMA

1.1. Planteamiento del Problema.

En una perspectiva formal, Montgomery (2004) define experimento como una prueba o serie de pruebas en las que se hacen cambios deliberados en las variables de entrada de un proceso o sistema para observar e identificar las razones de los cambios que pudieran observarse en la respuesta de salida. Por otra parte, un experimento comparativo es el que utilizan los investigadores en áreas como la biología, medicina, agricultura, ingeniería, psicología y otras ciencias experimentales (Kuehl, 2000). El adjetivo comparativo implica que se establezca más de un conjunto de circunstancias en el experimento y que se comparen entre sí las respuestas a las diferentes circunstancias.

Ahora bien, los tratamientos son el conjunto de circunstancias creados para el experimento, en respuesta a la hipótesis de investigación y son el centro de la misma. Con el objeto de probar diferencias significativas entre tratamientos, en una forma elemental se usa la prueba *F* basándose en un Análisis de la Varianza (Vargas, 2009). Un análisis de este tipo resume el conocimiento acerca de la variabilidad en las observaciones del experimento. Cuando no se rechaza la hipótesis nula parecería innecesario plantearse más preguntas, ya que se concluye a partir de esta situación que no existe evidencia como para pensar que algún par de medias dentro del conjunto estudiado pudiera ser diferente.

Cuando se realiza el Análisis de Varianza y se rechaza la hipótesis nula para un determinado efecto fijo, se llega a la conclusión de que por lo menos una de las medias de los grupos involucrados en esa fuente de variación difiere del resto y no se puede especificar cual de ellas es la que presenta diferencia. Por ello es de gran utilidad efectuar comparaciones adicionales entre las medias de los grupos o tratamientos.

Según Kuehl (2000). los tipos de constrastes que consideran los investigadores con mayor frecuencia son:

- Contrastes planeados
- Constrastes polinomiales ortogonales
- Comparaciones múltiples con el mejor tratamiento
- Comparaciones múltiples con el tratamiento control
- Todas las comparaciones por pares

Pero más que efectuar comparaciones adicionales, *la importancia de la investigación radica en el hecho de dar a conocer cual es la herramienta de comparación más idónea a ser utilizada*. Según Atil y Unver (2001), se ha apreciado el uso inadecuado de procedimientos de comparación múltiple más significante para el conjunto de datos a analizar.

Vargas (2009) como recomendación, en su trabajo de ascenso plantea que además de trabajar con tratamientos balanceados, sería de especial interés incorporar como opción de ***estudio tratamientos desbalanceados*** ya que esta situación es común en el desarrollo de experimentos en la práctica, ***sobre todo cuando se pierden datos y se pierden unidades experimentales***.

Por ello, resulta de gran importancia, ***medir la potencia de pruebas*** de comparación múltiple de medias que permiten hacer comparaciones en pares cuando ***se tienen replicas desiguales*** en los diseños totalmente aleatorizados.

1.2. Objetivos de la Investigación

1.2.1. Objetivo General

Medir la Potencia de Pruebas de Comparación múltiple de medias que permiten hacer comparaciones entre todos los pares de medias, para experimentos desbalanceados con el uso de simulación de muestras.

1.2.2. Objetivos Específicos

1. Realizar una revisión y/o documentación sobre los diferentes procedimientos estadísticos utilizados para las comparaciones múltiples de medias de tratamientos que permiten hacer comparaciones entre todos los pares de medias en experimentos desbalanceados.
2. Seleccionar los procedimientos de comparaciones múltiples de medias a ser estudiados, según las limitaciones de la simulación.
3. Medir la potencia muestral de los diferentes procedimientos de comparaciones múltiples de medias seleccionados, haciendo uso de la simulación de muestras.
4. Señalar el comportamiento de los procedimientos de comparaciones múltiples de medias seleccionados, para los diferentes casos estudiados.

1.3. Justificación.

Este trabajo se encuentra enmarcado dentro de la Línea de Investigación "Aplicaciones de la Matemática a la Ingeniería y otras áreas", adscrita al departamento de Matemática de la Facultad de Ingeniería de la Universidad de Carabobo.

Uno de los objetivos del Análisis de Varianza, es comparar las medias de la variable respuesta, según los distintos tratamientos utilizados por el investigador. Al comparar más de dos medias, el análisis de varianza indicaría si existe suficiente evidencia estadística en la muestra como para decir que las medias son significativamente diferentes entre si, pero este análisis no indica qué medias difieren de otras. Generalmente, no es suficiente mostrar que las medias de tratamientos son diferentes usando la prueba F en el análisis de varianza. Así que, *los investigadores quieren comparar las medias de los tratamientos dependiendo de las propiedades de los tratamientos*.

Hay diferentes métodos para comparar las medias de tratamientos. Esta investigación suministra a los Investigadores, una información para disminuir el esfuerzo por seleccionar un procedimiento estadístico que más se ajuste al análisis de los datos y su interpretación, ya que *proporciona información concerniente al mejor procedimiento de comparaciones múltiples de medias, según las comparaciones de interés*, en lugar de normalmente seguir meramente el procedimiento usado por algunos otros científicos en su campo.

1.4. Alcance.

Esta investigación sólo estableció criterios de comparación, en términos de los errores de tipo I y tipo II para las siguientes pruebas de comparaciones múltiples de medias:

- Mínimas Diferencias Significativas
- Prueba de Tukey
- Prueba de Student – Neuman – Keuls

Una vez señalado el comportamiento de los diferentes procedimientos a estudiar, se procedió a verificar los resultados en un caso con data real obtenida de la bibliografía disponible.

1.5. Limitaciones.

Como en toda investigación, el tiempo, la necesidad de obtención de resultados lo más pronto posible, los recursos financieros para el desarrollo adecuado de la actividad de investigar, y una plataforma de cálculo/simulación disponible, capaz de almacenar la información proveniente de la simulación, se hicieron presentes. Pero la limitante que más impacto tuvo en la investigación fue la dificultad para poder incluir dentro de la sintaxis de programación, las pruebas de Contrastes, entre ellas las de Contrastes Ortogonales.

CAPITULO II

MATERIALES Y MÉTODOS

2.1. Clasificación de las pruebas de Rango Múltiple.

Hay diversas clasificaciones de las pruebas de rango múltiple. A continuación se ofrecen dos clasificaciones a saber según Montgomery (1998):

De acuerdo a su naturaleza, que a su vez se clasifican en:

1. *Pruebas Predeterminadas*: su aplicación esta restringida a las situaciones en las que las comparaciones a realizar son planificadas previamente. Se les da nombre a los tratamientos de antemano. Estas pruebas pueden conducir a resultados sin validez cuando son aplicadas en comparaciones sugeridas por los datos.

   En esta categoría se incluyen las pruebas Mínima Diferencia Significativa, Contrastes Ortogonales y Dunnett.

2. *Pruebas de Efectos Sugeridos por los Datos*: se aplican en situaciones en las que se conoce poco acerca de la naturaleza de los tratamientos y se vacila en proponer comparaciones con sentido. Se incluyen las pruebas de Scheffé, SNK, W de Tuckey, amplitudes Múltiples de Duncan, t de razón k Bayesiana de Waller-Duncan, de Gabriel y Bonferroni.

De acuerdo a las comparaciones de Interés, a su vez se clasifican en:

1. *Pruebas Contra un Testigo*: su objetivo es comparar todas las medias de los tratamientos contra un testigo preestablecido. Se incluye en este grupo la prueba de Dunnett.

2. *Pruebas que Involucran solo algunas comparaciones*: son pruebas predeterminadas, con las cuales es posible efectuar solo algunas comparaciones, dentro de las que puede incluirse o no las comparaciones contra un testigo. Esta restricción depende de las características de cada prueba. Se incluyen aquí la prueba de Mínima Diferencia Significativa y Contrastes Ortogonales.

3. *Pruebas que involucran la comparación de todos los pares posibles de medias*: es posible comparar las medias de todos los tratamientos, bien sea que se trate de efectos sugeridos por los datos o que se considere un efecto planeado. Se incluyen en esta categoría las pruebas de Scheffé, W de Tuckey, Amplitudes Múltiples de Duncan, t de razón k Bayesiana de Waller-Duncan, de Gabriel y Bonferroni.

2.2. Pruebas de Comparaciones Múltiples de Medias

A continuación, se detallan los procedimientos de comparación múltiple de medias recopilados luego de la revisión bibliográfica:

2.2.1. Contrastes.

Un *Contraste* conlleva a una combinación lineal de los totales de tratamientos con el objeto de hacer comparaciones planeadas. Las comparaciones o contrastes entre las medias de los tratamientos se pueden construir de manera que se respondan a las preguntas específicas sobre el experimento (Kuehl, 2001).

En la comparación se emplea la *prueba **F***, a fin de responder preguntas pertinentes, conllevando esto a la partición de los grados de libertad y la suma de cuadrados para los tratamientos en componentes de comparación. Las comparaciones deben ser independientes y ortogonales.

De este modo una Comparación o Contraste es un conjunto de funciones lineales de la forma:

$$C = \sum_{i=1}^{t} C_i U_i = C_1 U_1 + C_2 U_2 + ... + C_t U_t \tag{2.1}$$

donde $\sum_{i=1}^{t} C_i = 0$

U_i : pueden ser totales o medias de tratamientos.

La suma de cuadrados de un Contraste es igual:

$$SCC = \frac{\left(\sum_{i=1}^{t} C_i U_i \right)^2}{\sum_{i=1}^{t} \frac{C_i^2}{n_i}} \tag{2.2}$$

donde n_i es el número de observaciones de cada tratamiento. Si todas las n_i son iguales, e iguales a n, se tiene que:

$$SCC = n \cdot \frac{\left(\sum_{i=1}^{t} C_i U_i \right)^2}{\sum_{i=1}^{t} C_i^2} \tag{2.3}$$

Una comparación tiene un grado de libertad de modo que la prueba **T** es la apropiada, sin embargo es conveniente usar la prueba **F** con los totales del tratamiento como numerador y como denominador suma de cuadrados del error experimental con 1 y $n - t$ grados de libertad para tratamientos y error.

Para probar la Hipótesis $Ho: C = 0$ **vs.** $Ha: C \neq 0$ a un nivel de significación α, se utiliza:

$$Fo = \frac{CMC}{CMEE} \qquad (2.4)$$

Donde: $CMC = SCC$
$SCEE$: Suma de Cuadrados del Error Experimental

Bajo la hipótesis nula, Fo tiene una distribución F con 1 y $n-t$ grados de libertad.

Así, la hipótesis nula se rechaza si: $F_0 > F_{\alpha,1,n-t}$.

O de manera equivalente se calcula: $To = \frac{C}{SC}$.

Este valor se compara con $t_{\alpha,n-t}$, y se rechaza H_0 si se cumple que $T_0 > t_{\alpha,n-t}$ tomando en cuenta que:

$$SC = \frac{CMEE}{r}\left(C_1^2 + C_2^2 + ... + C_t^2\right) \qquad (2.5)$$

2.2.2. Contrastes Ortogonales:

Cierto tipo de Comparaciones, conocidas como Contrastes Ortogonales, tienen propiedades especiales con respecto a la partición de la suma de cuadrados en el Análisis de Varianza y en cuanto a sus interrelaciones. La ortogonalidad implica que un Contraste no contiene información sobre otro (Kuehl, 2001).

Supóngase que existen dos contrastes, C y D, donde:

$$C = \sum_{i=1}^{t} c_i \bar{y}_i \quad \text{y} \quad D = \sum_{i=1}^{t} d_i \bar{y}_i \tag{2.6}$$

Para que sean contrastes ortogonales se debe cumplir:

$$\sum_{i=1}^{t} \frac{c_i d_i}{r_i} = \frac{c_1 d_1}{r_1} + \frac{c_2 d_2}{r_2} + \ldots + \frac{c_t d_t}{r_t} = 0 \tag{2.7}$$

Existen $t-1$ contrastes mutuamente ortogonales entre t medias de tratamientos. Otra característica especial de los *Contrastes Ortogonales* es que la sumatoria de cuadrados de los $t-1$ Contrastes Ortogonales es igual a la suma de cuadrados del Análisis de Varianza.

Los coeficientes de los Contrastes deben ser elegidos antes de realizar el experimento y analizar los datos. Para elegir los coeficientes de los Contrastes existen diferentes maneras. Las siguientes reglas son de utilidad en ese sentido:

a. Si se van a comparar dos grupos de igual tamaño, simplemente asígnense coeficientes +1 a los miembros de un grupo y -1 a los integrantes del otro grupo.
b. En la comparación de grupos con diferente número de tratamientos, asígnense al primer grupo tantos coeficientes como número de tratamientos tenga el segundo grupo, y al último tantos coeficientes del signo opuesto como número de tratamientos tenga el primer grupo.
c. Redúzcanse los coeficientes a los enteros más pequeños posibles. Por ejemplo, en la comparación de un grupo de dos tratamientos con un grupo de cuatro, en virtud de la regla (b) se tienen los coeficientes: +4,+4,-2,-2,-2,-2, pero estos pueden reducirse a +2,+2,-1, -1,-1,-1.
d. Los coeficientes de interacción siempre pueden determinarse mediante la multiplicación de los coeficientes correspondientes a los efectos principales.

2.2.3. .- Prueba de Scheffé.

En diferentes situaciones no se saben de antemano los *Contrastes* que se desean comparar, o se desea realizar $t-1$ comparaciones.

Scheffé (1953), propuso un método para probar todas las comparaciones posibles o construir Intervalos de Confianza para las correspondientes funciones lineales de parámetros. Este método proporciona la protección señalada del error con respecto al experimento para cualquier número de comparaciones; en consecuencia, es bastante conservador y, en general, se usa para comparaciones no planeadas o sugeridas por los datos.

Considere cualquier comparación $C = \sum_{i=1}^{t} C_i \bar{y}_{i.}$, entre t medias de tratamientos con un error estándar:

$$S_C = \sqrt{CMEE \cdot \sum_{i=1}^{t} \frac{C_i^2}{r_i}} \qquad (2.8)$$

La hipótesis nula para la comparación $Ho: C = 0$, se rechaza si:

$$|C| > S(\alpha u) \qquad (2.9)$$

donde $S(\alpha u)$ es el estadístico de Scheffé, que se calcula como:

$$S(\alpha u) = S_C \sqrt{(t-1) F_{t-1,(\alpha u)}^{n-t}} \qquad (2.10)$$

Los Intervalos de Confianza simultáneos de 100(1-α) % para todas las comparaciones posibles con el estadístico de Scheffé se construyen usando: $C \pm S(\alpha u)$

Scheffe (1953,1959) propuso otro método para el control del error experimental para un conjunto de contrastes u otras pruebas de hipótesis lineales en el análisis de modelos lineales, incluyendo comparaciones pareadas. Dos medias son declaradas significativamente diferentes si:

$$\left|\bar{y}_i - \bar{y}_j\right| \geq s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\sqrt{(k-1)F_{v\ \alpha}^{k-1}} \qquad (2.11)$$

En el caso de tener igual repeticiones por tratamiento, se tiene:

$$\left|\bar{y}_i - \bar{y}_j\right| \geq s\sqrt{\frac{2(k-1)}{n}F_{v\ \alpha}^{k-1}} \qquad (2.12)$$

Donde $F_{v\ \alpha}^{k-1}$ es el valor crítico a un nivel α de una distribución F con k-1 grados de libertad para el numerador y v grados de libertad para el denominador.

La prueba de Scheffe es compatible con la prueba F del ANOVA completo en que el Método de Scheffe nunca declara un contraste significativo si la Prueba F completa es no significativa. Es de destacar que otros métodos de comparaciones múltiples pueden encontrar contrastes significativos cuando la prueba F completa es no significativa y por esta razón sufre una pérdida de potencia cuando es usada con una Prueba F preliminar.

El método de Scheffe puede ser más potente que los Métodos de Bonferroni o Sidak si el número de comparaciones es relativamente grande respecto al número de medias a comparar. Para comparaciones pareadas, la prueba de Sidak es generalmente más potente.

2.2.4. Prueba de las Mínimas Diferencias Significativas (MDS).

Es un procedimiento usado para comparar un conjunto de medias y también para comparar cada una de las medias de un conjunto con un tratamiento control. Este procedimiento fue traído a discusión por *Fisher*.

Cada hipótesis $Ho : U_i = U_j$ se puede probar con el estadístico T-Student:

$$to = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{CMEE\left(\frac{1}{r_i} + \frac{1}{r_j}\right)}} \tag{2.13}$$

Cuando se establece la probabilidad de *error tipo I* en algún valor α y la varianza S^2 tiene $n-t$ grados de libertad, la hipótesis nula se rechaza para cualquier valor observado de $\left|\bar{y}_i - \bar{y}_j\right|$ si $|t_0| > t_{\alpha,n-t}$. La *mínima diferencia significativa* es un método abreviado para realizar todas las pruebas T por pares. Simbólicamente se escribe:

$$MDS(\alpha) = t_{(\alpha/2,glEE)} \cdot S(\bar{d}) \tag{2.14}$$

Donde:

- $MDS(\alpha)$: mínima diferencia significativa al nivel de significación α
- $t_{(\alpha,glEE)}$: valor de la distribución T-Student al nivel α con los grados de libertad del error.
- $S(\bar{d})$: error típico de la diferencia de 2 medias.

A su vez:

$$S\left(\bar{d}\right)=\sqrt{\frac{2\cdot CMEE}{r}} \qquad (2.15)$$

donde:

- $CMEE$ **:** Cuadrado Medio del Error Experimental.
- r : Número de replicaciones.

La hipótesis nula $Ho: U_i = U_j$, se rechaza si:

$$\left|\bar{y}_i - \bar{y}_j\right| > MDS(\alpha) \qquad (2.16)$$

(Todas las diferencias entre medias son comparadas con el $MDS(\alpha)$ calculado).

En total se pueden generar $t(t-1)\frac{1}{2}$ comparaciones. Si la diferencia excede el $MDS(\alpha)$, se dice que las medias provienen de poblaciones distintas.

2.2.5. Prueba de Student-Newman-Keuls o SNK.

Student (1927), sugirió el uso de rangos como procedimientos para rechazar muestras que difieren en una serie de análisis. Newman hizo uso de las ideas de Student para dividir un grupo de medias de tratamientos ordenados en subgrupos homogéneos y presentó tablas para ciertos valores de q, donde:

$$q = \frac{y_{\max} - y_{\min}}{S(\bar{y})} \qquad (17)$$

donde:

- $y_{\max}$: media más alta.
- $y_{\min}$: media más baja.
- $S(\bar{y})$: estimado de la desviación estándar de cada media si el número de réplicas es igual para los tratamientos y se calcula como: $S(\bar{y}) = \sqrt{\frac{CMEE}{r}}$ (18)

El estadístico SNK es:

$$W(\alpha,t) = q_{(\alpha,t,glEE)} \cdot S(\bar{y}) \tag{2.19}$$

Donde:

- $q(\alpha,t,glEE)$ **:** es el intervalo estadístico de Student.
- t : Número de medias en el conjunto.
- $glEE$: Grados de libertad del error experimental.

La hipótesis nula $Ho: U_i = U_j$, para las medias mayor y menor, digamos $y_{\max}, y_{\min}$, se rechaza si:

$$|y_{\max} - y_{\min}| > W_{(\alpha,t)} \tag{2.20}$$

Al aplicar la prueba SNK se debe seguir los siguientes pasos:

a. Escoger el nivel de significación α.
b. Calcular $S(\bar{y})$ y los valores:

   $W_t = q(\alpha,t,glEE)$

   $W_{t-1} = q(\alpha,t-1,glEE)S(\bar{y})$

   $W_2 = q(\alpha,t-2,glEE)S(\bar{y})$
c. Ordenar las medias de tratamiento desde la más alta a la más baja. Es decir : $\bar{y}_{\max}, \bar{y}_{t-1}, ..., \bar{y}_2, \bar{y}_{\min}$
d. Comparar los rangos de los t tratamientos $\bar{y}_{\max} - \bar{y}_{\min}$ con los W_t. Si $\bar{y}_{\max} - \bar{y}_{\min} > W_t$ se subdividen las medias en 2 grupos de $t-1$ medias cada uno y se comparan los rangos $\bar{y}_{\max} - \bar{y}_2$ con $\bar{y}_{t-1} - \bar{y}_{\min}$ y así sucesivamente.

Es importante acotar que la prueba no se realiza si existe un conjunto de medias que contiene a $y_{\max}, y_{\min}$ con mayor tamaño t, que no es significativo según SNK .

2.2.6. Prueba de Rangos Múltiples de Duncan.

Duncan (1955), desarrolló una nueva prueba de amplitud múltiple. Dicha prueba es parecida a la prueba *SNK* en cuanto usa amplitudes múltiples y esta guiada por resultados. Se aparta del procedimiento *SNK* en cuanto al nivel de significación, constante en *SNK* y variable en Duncan dependiendo del número de medias que entran en una etapa. La idea es que a medida que el número de medias aumenta, menor es la probabilidad que se asemejen.

Si el rango es de 2 y 3 medias se usa α = 5 %. Para 4 medias úsese α = 14%, y para t medias: $\alpha' = 1 - (1 - \alpha)^{t-1}$

Con esto se construye la tabla de amplitud estudentizada.

El estadístico de Duncan es:

$$R_{(\alpha,t,glEE)} = W_{(\alpha,t,glEE)} \cdot S(\bar{y}). \qquad (21)$$

donde:

- $W_{(\alpha,t,glEE)}$: estadístico t de amplitud estudentizada al nivel de significación α con t tratamientos.
- $glEE$: grados de libertad del error experimental.

El procedimiento a seguir para probar la hipótesis nula $Ho: U_i = U_j$ contra la alternativa $Ha: U_i \neq U_j$ es similar a SNK, pero utilizando $R(\alpha, t, glEE)$ en lugar de $W(\alpha, t, glEE)$.

Si $\left|\bar{y}_i - \bar{y}_j\right| > R_{(\alpha,t,glEE)}$, se declara significativa la diferencia.

2.2.7. Prueba de Tukey.

Tukey (1953), propuso un procedimiento de comparación múltiple que esta basado en los intervalos y es aplicable a los pares de medias; necesita de un solo valor para juzgar la significancia de todas las diferencias. Ya que solo se hacen comparaciones por pares, el valor crítico es menor que el exigido por el método de Scheffé, todos los pares constituyen una familia y la tasa de error es familiar.

Tukey (1952, 1953) propone un prueba diseñada específicamente para comparaciones pareadas basadas en un rango estudentizado, algunas veces llamada “prueba de diferencias significativas honesta”, que controla la tasa de error máximo experimental cuando los tamaños de muestra son iguales. Tukey (1953) y Kramer (1956) de manera independiente proponen una modificación para tamaños de muestras desiguales. Hayter (1984), hizo pruebas al método de Tukey o de Tukey-Kramer y comprobó el procedimiento de control que estas pruebas tienen sobre la tasa máxima de error experimental. El Método de Tukey-Kramer es más potente que los métodos de Bonferroni, Sidak o Scheffe para comparaciones pareadas.

El procedimiento consiste en el cálculo de un valor crítico mediante su aplicación a diferencias entre todos los pares de medias. El valor crítico de Tuckey es:

$$W_\alpha = q_{\alpha(t,glEE)}\sqrt{\frac{MS_E}{n}} \qquad (2.22)$$

donde:

$q_{\alpha(t,glEE)}$: es el punto porcentual y se puede hallar tabla de puntos porcentuales superiores de la amplitud estudentizada, en la que t es el número de tratamientos y,
$glEE$: corresponde a los grados de libertad del error.
MS_E: el cuadrado medio del error experimental y,
n: el número de muestras por tratamiento.

Para que un par de medias se declare significativamente diferente debe suceder que:

$$\left|\bar{y}_i - \bar{y}_j\right| > W_\alpha \qquad (2.23)$$

La prueba de Tuckey tiene menos poder que la prueba SNK y Duncan (Carmer y Swanson, 1973).

Esta prueba también puede usarse para cálculos de un conjunto de intervalos de confianza para las diferencias mediante: $\left|\bar{y}_i - \bar{y}_j\right| \pm W_\alpha$

2.2.8. Prueba *t* de Razón Bayesiana de Waller – Duncan:

Duncan (1975) citado por Chacín (2000), en su afán de perfeccionamiento de los procedimientos para realizar comparaciones múltiples de medias con la colaboración de Waller propuso la siguiente prueba de comparaciones múltiples de medias.

La prueba se caracteriza por que no interviene un nivel de significancia, en su lugar se escoge una gravedad o peso del error tipo I o tipo II. Se aconseja razones de 50:1, 100:1 500:1 en lugar de α = 10 %, 5% y 1%.

En esta prueba se utiliza una tabla que contiene valores t de riesgo mínimo para K=100 (la que se utilizará en los ejemplos de este trabajo). La tabla se consulta en base al valor para el contraste F de tratamiento por lo cual se calcula el valor de b como:

$$b = \left|\frac{F}{F-1}\right| \qquad (2.24)$$

Con el valor $t-1$ y $glEE$ se va a la tabla y se interpola en caso de ser necesario el t de riesgo mínimo. Luego se calcula el valor crítico:

$$DMS = t \cdot S(\bar{d}) \qquad (2.25)$$

donde:

- t : valor de riesgo mínimo.
- $S(\bar{d})$: desviación estándar de 2 medias.

Para probar la hipótesis nula $Ho: U_i = U_j$ contra la alternativa $Ha: U_i \neq U_j$, este valor se compara con cada diferencia de medias: $\left|\bar{y}_i - \bar{y}_j\right|$. Si $\left|\bar{y}_i - \bar{y}_j\right| > DMS$ se declara diferencia significativa.

2.2.9. Prueba de Dunnett.

Probablemente uno de los casos mas frecuentemente encontrados es la comparación de un control con cada tratamiento. Por ejemplo, en Medicina como también en las industrias se podría desear probar varias drogas nuevas y compararlo con una droga estándar.

El procedimiento de Dunnett requiere de un solo valor para juzgar la significancia de las diferencias observadas entre cada tratamiento y el control. Se pueden efectuar comparaciones con alternativas Unilaterales y Bilaterales. La tasa de error es familiar y pueden construirse intervalos de confianza.

El valor crítico de Dunnett esta dado por:

$$d' = t_{(\alpha,t)} \cdot S(\bar{d}). \qquad (2.26)$$

donde:

- $t(\alpha,t)$**:** estadístico de Dunnett al nivel de significación α y t tratamientos para el conjunto apropiado de alternativas de una y dos colas.
- $S(\bar{d})$: es la desviación estándar expresada como :

$$S(\bar{d}) = \sqrt{\frac{2 \cdot CMEE}{r}} \qquad (2.27)$$

Las estimaciones de los intervalos de confianza bilaterales para las medias de los tratamientos individuales y la media del control para: $Ho: \mu_i - \mu_c = 0$ son: $\bar{y}_i - \bar{y}_c \pm d'$. Los límites superiores del intervalo de un lado, si se manifiesta superioridad de la media de tratamiento por ser mayor que la media control, son:
$\bar{y}_i - \bar{y}_c - d'$. En caso de ser menor que la media control, son: $\bar{y}_i - \bar{y}_c + d'$.
Para la prueba $Ho: \mu_i - \mu_c = 0$ contra la alternativa $Ha: \mu_i - \mu_c \neq 0$, se rechaza si: $|\bar{y}_i - \bar{y}_c| > d'$. Para la alternativa de un lado $Ho: \mu \leq \mu_j$ contra la alternativa $Ha: \mu \geq \mu_j$ se rechaza si: $\bar{y}_i - \bar{y}_c \geq d'$. Para la alternativa de un lado $Ho: \mu \geq \mu_j$ contra la alternativa $Ha: \mu \leq \mu_j$, se rechaza si: $\bar{y}_i - \bar{y}_c < -d'$.

2.2.10. Prueba de Gabriel.

Esta prueba fue desarrollada por K. R. Gabriel, en la Universidad Hebrea de Jerusalén, Israel. El procedimiento fue designado por las siglas STP que significa procedimiento de pruebas simultáneas.

El procedimiento se propone probar la homogeneidad de todos los conjuntos de medias involucradas en el *Análisis de Varianza*. Las decisiones de significación están basadas en las sumas de cuadrados entre medias que están en un conjunto, se usa el mismo valor crítico como el usado para el *test F*.

Se dice que las decisiones son transitivas, en el sentido de que cada subconjunto que contenga otro subconjunto significativo es por sí significativo. Una decisión sería incompleta en el conjunto significativo si resulta ningún subconjunto significativo.

Se asume que los siguientes datos están disponibles para K poblaciones: $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ son medias normales e independientemente distribuidas con varianzas $\frac{\sigma^2}{n_1}, \frac{\sigma^2}{n_2}, ..., \frac{\sigma^2}{n_k}$, siendo

$n_1, n_2, ..., n_k$, conocidas pero σ^2 desconocida. Se asume S^2 un estimador independiente e insesgado de σ^2. Cada conjunto p de todas o algunas de las k poblaciones se juzga como de varianzas significativamente diferentes, si y solo si:

$$S_p^2 > S_{k-1}^2 \cdot F_{(\alpha,k-1,n_0)} \qquad (2.28)$$

donde:

- $F_{(\alpha,k-1,n_0)}$: es el punto superior de α de la distribución F con k – 1 y n_0 grados de libertad.
- S_p^2: es la suma de cuadrados para la hipótesis de homogeneidad de las varianzas de las poblaciones en p calculada como:

$$S_p^2 = \sum_i^p n_i\left(\bar{x}_i - \bar{x}_p\right)^2 \qquad \bar{x}_p = \sum_i^p \frac{n_i \bar{x}_i}{n_p} \qquad n_p = \sum_i^p n_i$$

- $\sum_i^p$ denota la sumatoria sobre todas las muestras de la población en " p ".

2.2.11. Prueba de Bonferroni

Kuehl (2000) indica que los valores tabulados para el estadístico t de Student en relación con la t de Bonferroni se expresa como: $t_{\alpha_e/2,k,glEE}$, para las pruebas de 2 colas, K número de comparaciones y glEE grados de libertad del error.

Los intervalos de confianza simultáneos usan el estadístico t de Bonferroni en lugar de la t Student y vienen dado por:

$$C \pm t_{\alpha_e/2,k,glEE} SC \qquad (2.29)$$

Para probar la hipótesis nula: $Ho:\mu_i - \mu_j = 0$ equivalente a $Ho: C = 0$ se calcula la estadística: $to = \frac{C}{SC}$, donde C es la diferencia de medias $\bar{y}_i - \bar{y}_j$ y SC es la desviación estándar correspondiente a la diferencia de medias. Se rechaza la hipótesis nula si $|to| \geq t_{\alpha_e/2,k,glEE}$ por lo que se declara diferencia significativa entre las medias de tratamiento.

2.2.12. Prueba de Sidak.

La prueba de Sidak (Citada por Games, 1977), considera significativa una diferencia si:

$$\left|\bar{y}_i - \bar{y}_j\right| \geq t_{(\varepsilon,\nu)} \cdot s \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \qquad (2.30)$$

Donde $\varepsilon = 1 - (1-\alpha)^{1/(k(k-1)/2)}$ para comparaciones entre k medias. Si los tamaños de muestras son iguales, la prueba se simplifica a:

$$\left|\bar{y}_i - \bar{y}_j\right| \geq t_{(\varepsilon,\nu)} \cdot s \cdot \sqrt{\frac{2}{n}} \qquad (2.31)$$

Se puede usar la desigualdad aditiva de Bonferroni y la desigualdad multiplicativa de Sidak para controlar la Tasa de Error Máximo Experimental para un conjunto de contrastes u otras pruebas de hipótesis, no solo para comparaciones pareadas. La desigualdad de Bonferroni puede arrojar inferencias simultáneas en alguna aplicación estadística con pruebas de más de una hipótesis. Otros métodos discutidos después para comparaciones de par en par pueden también ser adaptados para contrastes generales (Miller, 1981).

2.2.13. Prueba de Hochberg.

Hochberg (1974) propone un método similar al de Tukey, pero éste usa los módulos máximos estudentizados en vez de los rangos estudentizados y emplea la desigualdad no correlacionada de Sidak. Este método es generalmente menos potente que el método de

Tukey-Kramer y siempre menos potente que la prueba de Tukey para igual tamaño de muestras. Dos medias se declaran significativamente diferentes si:

$$\left|\bar{y}_i - \bar{y}_j\right| \geq m_{(\alpha;c,v)} \cdot s \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \qquad (2.32)$$

Donde $m_{(\alpha;c,v)}$ es el valor crítico de nivel α del módulo máximo estudentizado con distribución de c variables aleatorias normales independientes con v grados de libertad y $c = k(k-1)/2$. Para igual número de muestras, la prueba se simplifica a:

$$\left|\bar{y}_i - \bar{y}_j\right| \geq m_{(\alpha;c,v)} \cdot s \cdot \sqrt{\frac{2}{n}} \qquad (2.33)$$

2.3. Metodología

Al momento de desarrollar un proyecto es conveniente contar con una metodología para definir los aspectos que deben considerarse en la ejecución de la investigación, así como también considerar la herramienta de simulación para tratar de conocer el comportamiento de las diferentes pruebas de comparaciones múltiples de medias.

2.3.1. Planificación.

Esta etapa consideró toda la planificación del proyecto, el estudio de tópicos esenciales tales como diseños de experimentos, análisis de varianza, pruebas de comparaciones de medias, ratas experimentales de errores tipo I, II y potencia de prueba.

2.3.2. Recopilación y Análisis de la Información.

Se hizo una revisión bibliográfica minuciosa con el propósito de recopilar la mayor cantidad de información posible entorno a la descripción teórica como a la efectividad de pruebas de comparaciones múltiples de medias, tanto a nivel teórico como a nivel de aplicaciones en áreas como la medicina, ciencias computacionales y ciencias sociales entre otras, encontrándose aspectos importantes que ayudaron a conducir las conclusiones de esta investigación.

Una vez culminada la fase anterior se procedió a analizar en detalle toda la información obtenida, logrando identificar las necesidades de investigación. Esto es, que lenguaje de programación podría ser el más apropiado y que procedimientos se ajustarían a dicho lenguaje.

2.3.3. Selección de Herramienta de Simulación.

Selección de la herramienta computacional, considerando las necesidades que se desean cubrir, y que los procedimientos a los cuales se les desean conocer su comportamiento.

II.3.4. Simulación de los Procedimientos.

Una vez que se ha seleccionado el software de simulación y los procedimientos de comparaciones múltiples de medias de interés para la investigación, se efectuó la simulación para señalar el comportamiento de estos últimos, según las diferentes situaciones estudiadas.

2.3.5. Obtención de Resultados.

Realizado el código que permite obtener muestras de diseños de experimentos completamente aleatorizados y balanceados, se hicieron las corridas según los esquemas planteados para la simulación de muestras.

2.3.5. Conclusiones y Recomendaciones.

Finalmente resultó conveniente puntualizar un conjunto de conclusiones y recomendaciones a fin de mostrar las implicaciones de la investigación y futuras líneas de investigación en el tema de estudio.

CAPITULO III

RESULTADOS

En esta sección se muestran los resultados obtenidos una vez seleccionada la herramienta de simulación y seleccionadas las pruebas de comparación de medias a ser estudiadas.

3.1. Herramienta de simulación.

Luego de una revisión de herramientas de cómputo para la ejecución de la simulación, se seleccionó el software SCILAB, por su disponibilidad y facilidad de codificación de las distintas rutinas que permiten entre otras cosas, generar muestras aleatorias con distribución normal y evaluar los distintos procedimientos de comparación de medias para las muestras generadas.

3.2. Procedimientos a simular.

De acuerdo con la revisión bibliográfica, en el presente documento se muestra la descripción de los siguientes procedimientos para las pruebas de comparación múltiple de medias:

1.- Mínima Diferencia Significativa.
2.- Contrastes Ortogonales.
3.- Prueba de Dunnett.
4.- Prueba de Scheffé.
5.- Prueba de Student, Newman y Keuls, (SNK).
6.- Prueba W de Tuckey-Kramer.
7.- Prueba de Amplitudes Múltiples de Duncan.
8.- Prueba t de Razón K Bayesiana de Waller- Duncan.

9.- Prueba de Bonferronni.

10.- Prueba de Gabriel.

11.- Prueba t de Sidak.

12. Método de Hochberg.

El objetivo de la simulación, es determinar los errores tipo I y tipo II muestrales (mas que el error tipo II, es determinar la potencia como el complemento del error tipo II), para las diferentes pruebas de comparaciones múltiples de medias. Sin embargo, no todos los procedimientos listados pudieron ser considerados dentro de la simulación, ya que se aplicaron los siguientes criterios de selección:

1. Posibilidad de efectuar la mayor cantidad de comparaciones de medias.

   Las pruebas de comparación múltiple de medias, de acuerdo a las comparaciones de interés, se clasifican en *pruebas contra un testigo, pruebas que involucran sólo algunas comparaciones y pruebas que involucran la comparación de todos los pares posibles de medias*. Debido a que las muestras son generadas aleatoriamente, no se sabe de ante mano, donde se encuentran las diferencias reales una vez que el Análisis de Varianza resulte significativo, por ello es necesario utilizar procedimientos que permitan realizar todas las comparaciones posibles sin afectar las tasas de error tipo I.
2. Disponibilidad de valores de riesgo mínimo para las condiciones de interés de la simulación.

   Las situaciones simuladas dependieron de variables como número de tratamientos y número de muestras por tratamiento. Estas variables afectan directamente los grados de libertad del error experimental en el Análisis de Varianza, y si además consideramos el nivel de significación, la gran variedad de situaciones amerita tener disponible los valores de riesgo mínimo para cada situación. Esta fue una limitante, ya que para algunas pruebas no se encontraron dichos valores oportunamente.
3. Facilidad para la codificación en el software utilizado para la simulación.

   Este criterio se explica por si solo.

4. Uso histórico de la prueba por parte de investigadores.
5. Posibilidad de plantear contrastes sin necesidad de conocer la naturaleza de los tratamientos.

De acuerdo con una revisión de al menos 200 títulos y resúmenes de investigaciones cuyas palabras claves involucraban el uso de procedimientos múltiples de medias, se pudo medir cuáles de los procedimientos resultaban comúnmente usados por los investigadores en sus trabajos.

Con esto se presenta el cuadro 3.1 que muestra el cumplimiento o no de cada uno de los criterios por parte de las pruebas listadas anteriormente.

Cuadro 3.1. Cumplimiento, de los diferentes procedimientos o pruebas de comparación múltiple de medias, de los criterios de selección para la simulación.

Prueba o procedimiento	Cumplimiento del Criterio				
	(1)	(2)	(3)	(4)	(5)
Mínima Diferencia Significativa	No	Si	Si	Si	Si
Contrastes Ortogonales	No	Si	No	No	Si
Prueba de Dunnett	No	No	Si	Si	Si
Prueba de Scheffé	Si	Si	Si	Si	No
Prueba de Student, Newman y Keuls (SNK)	No	Si	Si	Si	No
Prueba W de Tuckey-Kramer	Si	Si	Si	Si	No
Prueba de Amplitudes Múltiples de Duncan	Si	Si	Si	Si	No
Prueba t de Razón K Bayesiana de Waller- Duncan	Si	No	Si	No	No
Prueba de Bonferronni	Si	Si	Si	Si	No
Prueba de Gabriel	Si	No	Si	No	No
Prueba t de Sidak	No	No	Si	No	Si
Método de Hochberg	No	No	Si	No	Si

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Por ello, se incluyen en este trabajo de investigación los siguientes procedimientos para la simulación:

- Mínimas Diferencias Significativas

- Prueba de Tukey
- Prueba de Student – Neuman – Keuls

Es importante señalar que aunque las pruebas de Minima Diferencia Significativa y Student – Newman – Keuls no cumplen con el primer criterio, fue necesario incluirlas por el marcado uso que hacen los investigadores de estos procedimientos. Esto implica observar con detenimiento los valores muestrales de las tasas de error tipo I para estas pruebas, que pueden verse afectadas por el hecho de efectuar con ellas la mayor cantidad de comparaciones posibles de medias.

3.3. Simulación de muestras.

El experimento simulado es el Diseño Experimental Completamente al Azar. Las diferentes situaciones simuladas dependieron de las siguientes variables:

i. Número de tratamientos. En concordancia con la literatura consultada, resulta frecuente encontrar experimentos cuyo número de tratamientos oscila entre 3 y 6, en el 99% de los casos, por ende, los experimentos simulados tendrán 3, 4, 5, y 6 tratamientos. Además se incluyeron corridas para probar el comportamiento de las pruebas de comparaciones múltiples de medias para 10, 12, 14, 16, 18 y 20 tratamientos.
ii. Número de muestras por tratamiento. Por tratarse de experimentos balanceados, todos los tratamientos independientemente del número de éstos en el diseño, tendrán el mismo tamaño muestral. En concordancia con la literatura consultada, se observa en ésta que el número de observaciones por tratamiento oscila entre 3 y 6, en el 99% de los casos. Por tal razón, los tamaños de muestra de cada tratamiento para los diferentes casos de estudio serán de 3, 4, 5 o 6. Para comparar los resultados usando muestras grandes, se incluyeron corridas con 10, 20, 30, 40 y 50 muestras por tratamiento.

iii. Pérdida de Muestras para generar el desbalance. Se considera la pérdida de muestras y/o de unidades experimentales en las cantidades de 1 hasta 5 por experimento simulado.
iv. Medias de tratamientos. Para esta variable, se consideraron dos casos: medias iguales y medias diferentes, generadas de manera aleatoria, para ser lo mas general posible.
v. Nivel de significación. Basándonos en el común de las distintas investigaciones, se utilizaron para la significación estadística de las pruebas, niveles teóricos de error tipo I en el orden de 1% y 5%.
vi. Coeficiente de Variación. Esta variable permite determinar la varianza común de todos los tratamientos. Es de importancia estadística por cuanto permite medir la homogeneidad de las muestras simuladas. Para esta variable, se tomaron los siguientes valores: 1%, 5%, 10%, 20%, 50%, 100%.

Este esquema de simulación arroja un total de 11280 casos, lo cual constituye una base sólida para generar conclusiones de interés en el campo estadístico, señalando además que cada caso se repitió 10000 veces. Además, en estas 11280 situaciones, se considera la variabilidad determinada con el coeficiente de variación, lo cual expande aún más el abanico de análisis y discusión de los resultados de la investigación.

3.4. Resultados de la Simulación.

Para realizar un análisis exhaustivo de la información obtenida a través de la simulación, los resultados se muestran del macro al micro, haciendo un adecuado desglose de las características de comparación a ser estudiadas, como lo son el nivel de significación y la tasa de error tipo I obtenida con las muestras y el comportamiento de la potencia en cada uno de estos casos; también dentro de las opciones de análisis, se muestran los anteriores resultados según el numero de tratamientos y muestras de tratamientos usadas con relación al número de muestras perdidas por experimento.

3.4.1. Resultados generales.

La combinación de número de tratamientos y muestras por tratamiento arrojó como consecuencia cuatro grandes grupos o modelos de experimentos según la siguiente clasificación:

- Grupo 1: Entre 3 y 6 tratamientos. Entre 3 y 6 muestras por tratamiento.
- Grupo 2: Entre 10 y 20 tratamientos. Entre 3 y 6 muestras por tratamiento.
- Grupo 3: Entre 3 y 6 tratamientos. Entre 10 y 50 muestras por tratamiento.
- Grupo 4: Entre 10 y 20 tratamientos. Entre 10 y 50 muestras por tratamiento.

A manera de interpretación de los resultados, cuando los tratamientos se encuentren entre 3 y 6 se considera número de tratamientos pequeño, cuando los tratamientos se encuentren entre 10 y 20, se considera número de tratamientos grandes. Así, cuando las muestras por tratamiento se encuentren entre 3 y 6, se consideran muestras pequeñas y cuando las muestras por tratamiento se encuentren entre 10 y 50, se consideran muestras grandes.

Estos resultados generales, son independientes del número de tratamientos, muestras por tratamiento y muestras perdidas que tenga el experimento, ni se especifican la cantidad de muestras que se pierden en el experimento, ni el coeficiente de variación que existe entre las muestras del experimento.

La tabla 1, muestra el Error Tipo 1 (muestral) obtenido para el ANOVA y cada una de las tres pruebas de comparación múltiple de medias estudiadas, como lo son *Mínima Diferencia Significativa (MDS), Prueba de Tuckey y Prueba Student-Neuman-Keuls (SNK)*, para niveles de significación teóricos del 1% y 5%.

Tabla 1.
Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico y el Grupo de Experimentos.

Nivel_Significación	Modelo	Promedio de %Anova	Promedio de %MDS	Promedio de %Tuckey	Promedio de %SNK
1%	Grupo_1	2,6114	0,9414	3,0495	3,0495
	Grupo_2	1,6782	1,1265	1,6629	1,6629
	Grupo_3	1,1988	0,0000	1,2688	1,2688
	Grupo_4	1,0918	0,0000	4,5402	4,5402
Total 1%		**1,4662**	**0,3807**	**2,8648**	**2,8648**
Nivel_Significación	**Modelo**	**Promedio de %Anova**	**Promedio de %MDS**	**Promedio de %Tukey**	**Promedio de %SNK**
5%	Grupo_1	8,4516	4,4188	8,9919	8,9919
	Grupo_2	6,8110	10,0244	6,7084	6,7084
	Grupo_3	5,5886	0,0000	5,4680	5,4680
	Grupo_4	5,3166	0,0000	8,7917	8,7917
Total 5%		**6,1639**	**2,8148**	**7,5200**	**7,5200**

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Tanto para el nivel de significación del 1% como para el del 5%, se observa que el ANOVA es la prueba que mejor ajusta su promedio de error tipo I muestral con el teórico.

En relación con la potencia muestral obtenida para ANOVA y las pruebas de comparaciones múltiples de media, se muestra la Tabla 2, donde se visualiza los promedios de potencia muestral según el nivel de significación teórico y los grupos de experimentos simulados.

Tabla 2.
Promedios de Potencia muestral obtenidas según el nivel de significación teórico y el Grupo de Experimentos.

Nivel_Significación	Modelo	Promedio de %Anova	Promedio de %MDS	Promedio de %Tukey	Promedio de %SNK
1%	Grupo_1	40,6197	69,9081	79,9113	79,9113
	Grupo_2	62,1447	79,3384	89,1939	89,1939
	Grupo_3	65,3768	54,1458	95,8739	95,8739
	Grupo_4	78,6899	59,1021	99,3373	99,3373
Total 1%		**66,1921**	**63,7918**	**93,4411**	**93,4411**

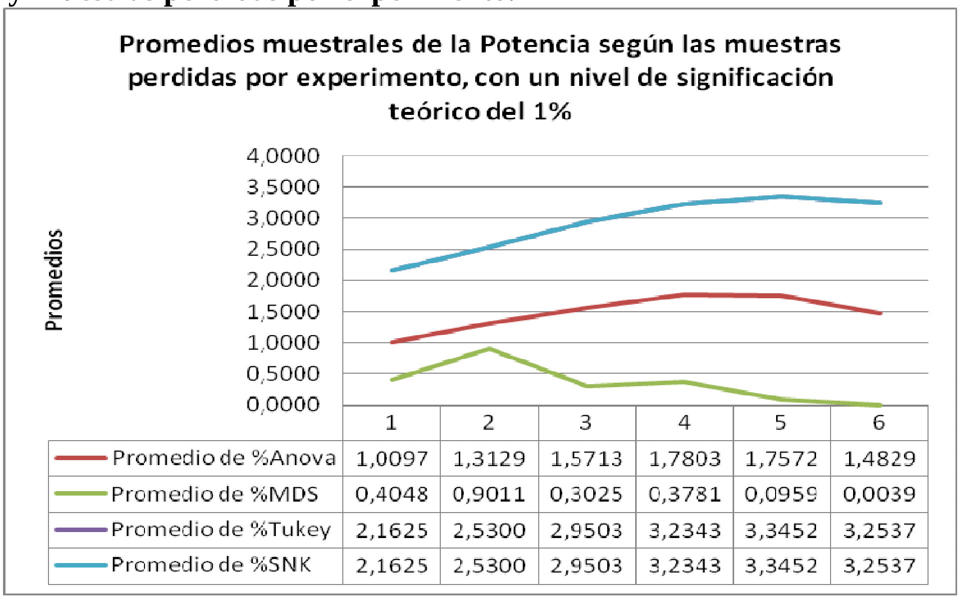
Nivel_Significación	Modelo	Promedio de %Anova	Promedio de %MDS	Promedio de %Tukey	Promedio de %SNK
5%	Grupo_1	46,2724	78,7640	85,4869	85,4869
	Grupo_2	65,5883	90,4861	93,7519	93,7519
	Grupo_3	66,8418	60,1858	97,1339	97,1339
	Grupo_4	78,8986	67,5846	99,7210	99,7210
Total 5%		**68,2113**	**72,2831**	**95,7056**	**95,7056**

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Se observa en los resultados anteriores, que las Pruebas de Tukey y SNK tienen la mayor probabilidad muestral de detectar diferencias significativas entre medias de tratamientos independientemente de la cantidad de tratamientos y muestras por tratamiento que tenga el experimento. Señalando primeramente que los resultados son generales en relación a la diversidad de situaciones simuladas.

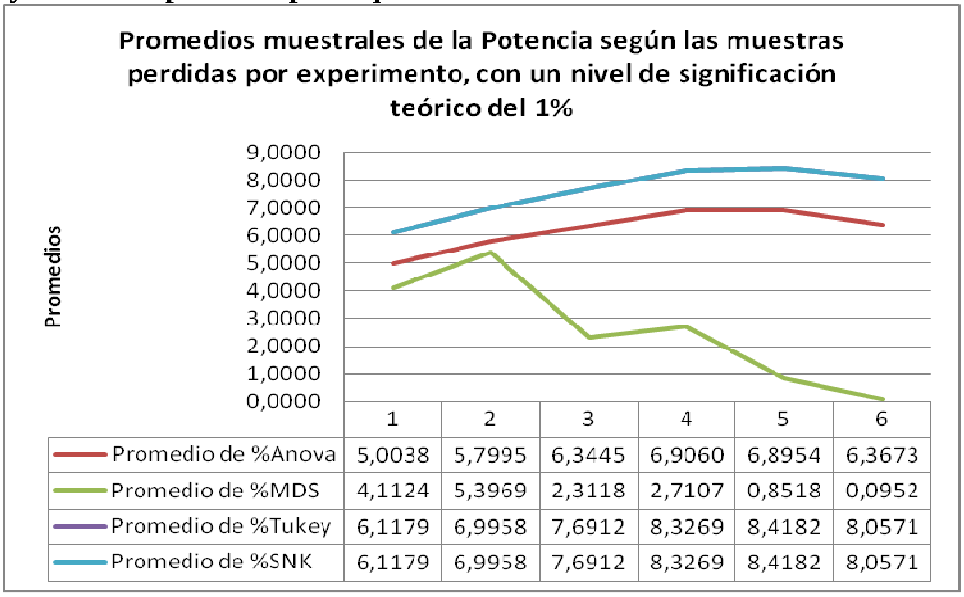
Los Gráficos 1 y 2, muestran los resultados anteriores desglosados según el número de muestras que se pierden por experimento.

Gráfico 1.
Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 1% y Muestras perdidas por experimento.



Fuente: Eduardo Vargas y Edwin Vargas (2014)

Gráfico 2.
Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 5% y Muestras perdidas por experimento.

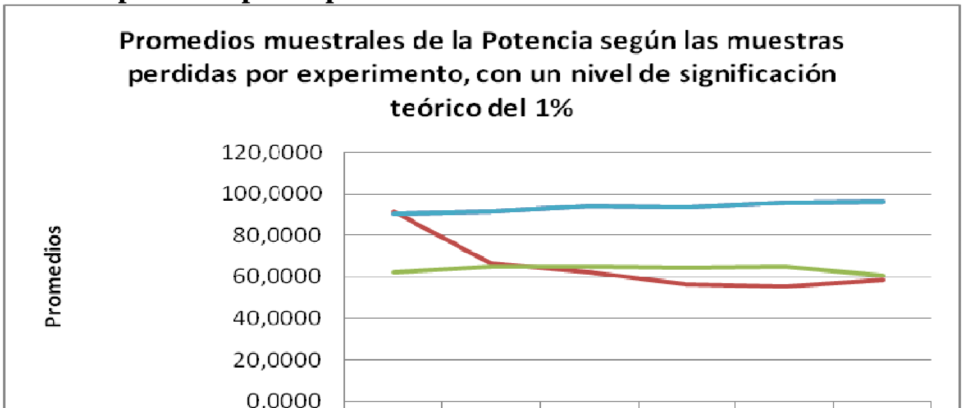


Fuente: Eduardo Vargas y Edwin Vargas (2014)

En ambos gráficos, se observa que la prueba de Mínima Diferencia Significativa es la que mejor se ajusta a los niveles de significación teóricos, mientras que el resto de las pruebas muestra un comportamiento creciente del promedio de error tipo I muestral a medida que aumenta el número de muestras perdidas por experimento. Cabe señalar que después de 5 muestras perdidas por experimento, los promedios de error tipo I para las pruebas estudiadas experimentan un descenso en su valor.

En los gráficos 3 y 4, se observa que indiferentemente del nivel de significación teórico (1% y 5%) con el que se simularon las muestras, la Potencia Muestral de las Pruebas de Tukey y SNK obtienen los mayores promedios, seguidas en orden descendente por la Prueba de MDS, para observar que la prueba de ANOVA es la que muestra los menores valores de Potencia muestral, disminuyendo cada vez más su valor a medida que el número de muestras perdidas por experimento aumenta.

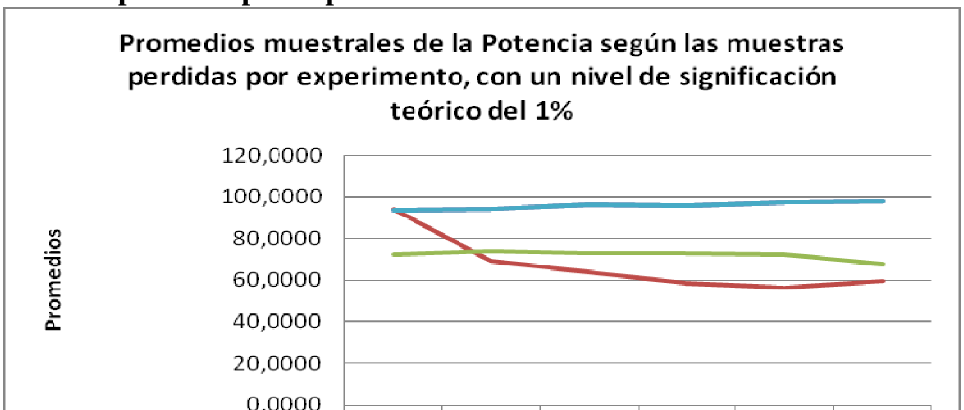
Gráfico 3.
Promedios de Potencia muestral obtenidos, según el nivel de significación teórico del 1% y Muestras perdidas por experimento.



	1	2	3	4	5	6
Promedio de %Anova	91,6477	66,6748	62,1617	56,4787	55,2274	58,4020
Promedio de %MDS	62,2271	64,7820	65,0406	64,5935	64,9018	60,6243
Promedio de %Tukey	90,5419	91,5378	94,2323	93,6629	95,9334	96,3868
Promedio de %SNK	90,5419	91,5378	94,2323	93,6629	95,9334	96,3868

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Gráfico 4.
Promedios de Potencia muestral obtenidos, según el nivel de significación teórico del 5% y Muestras perdidas por experimento.



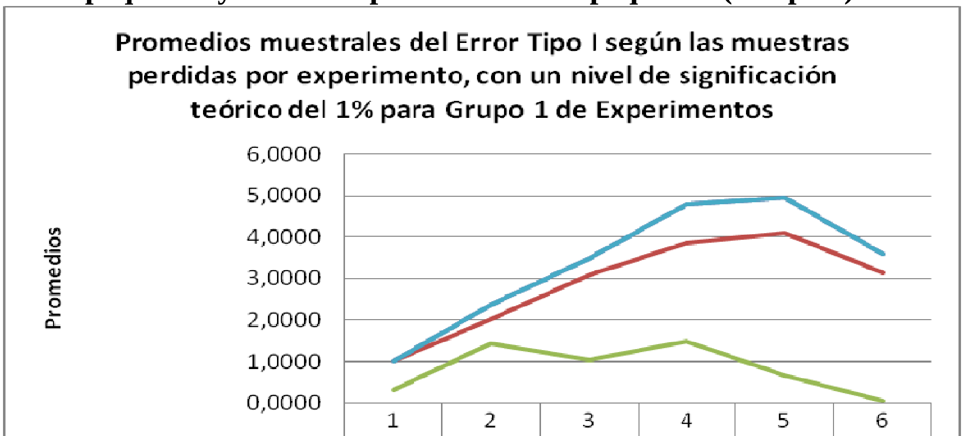
	1	2	3	4	5	6
Promedio de %Anova	94,3054	69,1977	64,0023	58,5754	56,5977	59,5992
Promedio de %MDS	72,4904	74,0519	73,0676	72,8489	72,2444	67,5638
Promedio de %Tukey	93,5752	94,3932	96,2945	95,9699	97,3820	97,7767
Promedio de %SNK	93,5752	94,3932	96,2945	95,9699	97,3820	97,7767

Fuente: Eduardo Vargas y Edwin Vargas (2014)

3.4.2. Resultados para Experimentos con número de tratamientos pequeños y muestras por tratamiento pequeñas (Grupo 1)

En los Gráficos 5 y 6 se muestran los resultados obtenidos para el Error Tipo I muestral en función de las muestras perdidas por experimento para los niveles de significación teóricos del 1% y 5% respectivamente, donde se observa que la Prueba de MDS es la que mejor se ajusta a los niveles teóricos de probabilidad de cometer error tipo I. Sin embargo, para todas las pruebas estudiadas, se observa un descenso en los valores promedios muestras del Error tipo I a partir de 5 muestras perdidas en el experimento.

Gráfico 5.
Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 1% y Muestras perdidas por experimento, para experimentos con número de tratamientos pequeños y muestras por tratamiento pequeñas (Grupo 1)

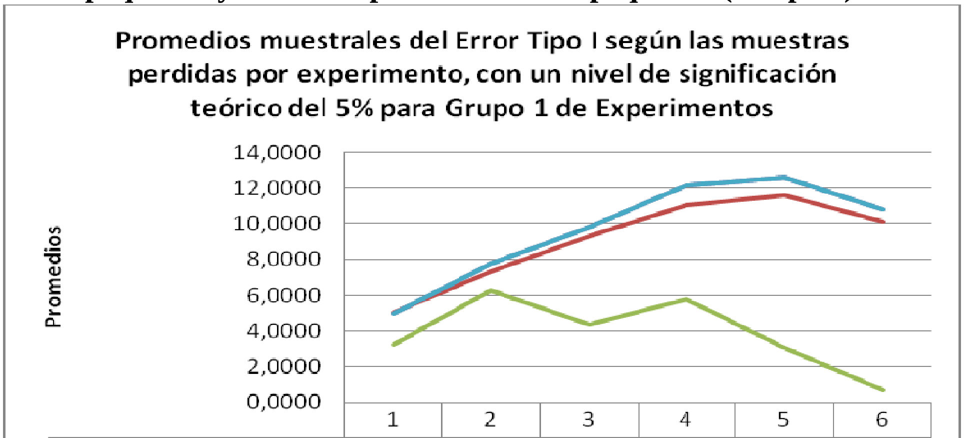


	1	2	3	4	5	6
Promedio de %Anova	1,0230	2,0283	3,0906	3,8539	4,1075	3,1404
Promedio de %MDS	0,3219	1,4408	1,0346	1,4897	0,6633	0,0538
Promedio de %Tukey	1,0220	2,3671	3,4776	4,7869	4,9460	3,5996
Promedio de %SNK	1,0220	2,3671	3,4776	4,7869	4,9460	3,5996

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Gráfico 6.

Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 5% y Muestras perdidas por experimento, para experimentos con número de tratamientos pequeños y muestras por tratamiento pequeñas (Grupo 1)



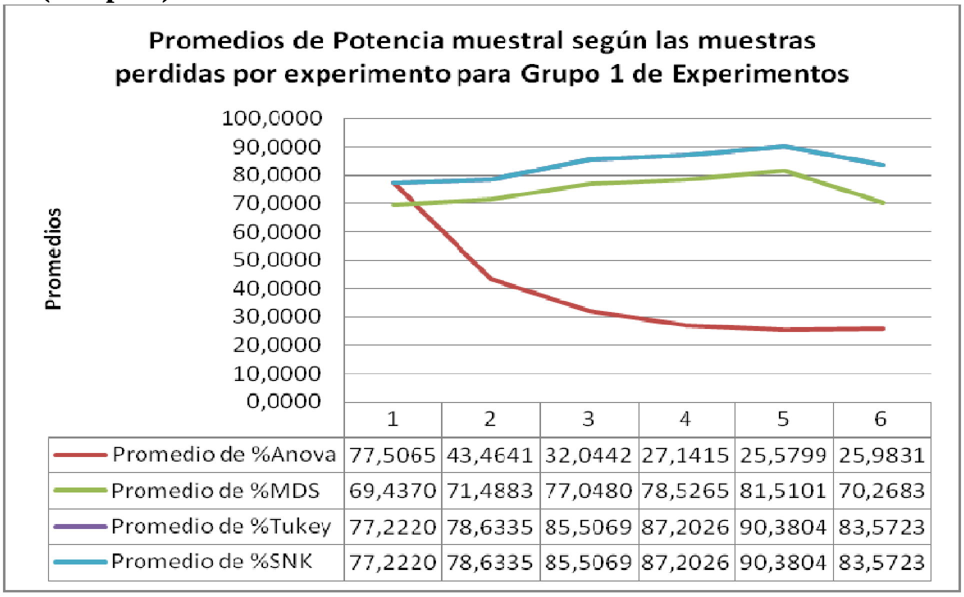
	1	2	3	4	5	6
Promedio de %Anova	5,0173	7,3060	9,3211	11,0415	11,6021	10,0913
Promedio de %MDS	3,2178	6,2732	4,3708	5,7603	3,0471	0,6688
Promedio de %Tukey	4,9896	7,7577	9,7985	12,1656	12,6046	10,7725
Promedio de %SNK	4,9896	7,7577	9,7985	12,1656	12,6046	10,7725

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Asimismo, se determinó la Potencia muestral para este grupo de experimentos, dejando en evidencia que el ANOVA es la prueba que pierda más potencia al aumentar el número de muestras perdidas por experimento cuando los experimentos tienen numero de tratamientos pequeño y muestras por tratamiento pequeñas. Esta información se puede visualizar en el gráfico 7.

Ese de notar en este gráfico 6 que las Pruebas de Tukey y SNK muestran mayor probabilidad muestral de detectar diferencia entre las medias de los tratamientos que la Prueba de MDS.

Gráfico 7.
Promedios de Potencia muestral obtenidos, según muestras perdidas por experimento, para experimentos con número de tratamientos pequeños y muestras por tratamiento pequeñas (Grupo 1)



	1	2	3	4	5	6
Promedio de %Anova	77,5065	43,4641	32,0442	27,1415	25,5799	25,9831
Promedio de %MDS	69,4370	71,4883	77,0480	78,5265	81,5101	70,2683
Promedio de %Tukey	77,2220	78,6335	85,5069	87,2026	90,3804	83,5723
Promedio de %SNK	77,2220	78,6335	85,5069	87,2026	90,3804	83,5723

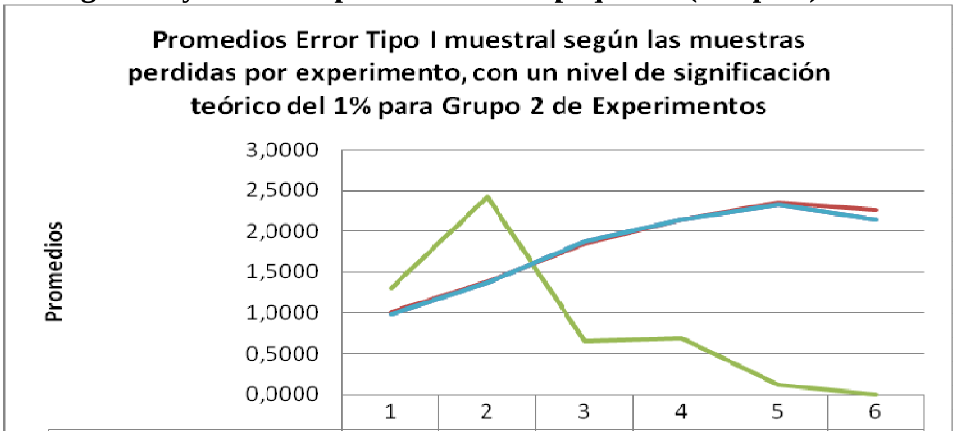
Fuente: Eduardo Vargas y Edwin Vargas (2014)

3.4.3. Resultados para Experimentos con número de tratamientos grandes y muestras por tratamiento pequeñas (Grupo 2)

Los Gráficos 8 y 9 muestran los resultados obtenidos para el Error Tipo I muestral en función de las muestras perdidas por experimento para los niveles de significación teóricos del 1% y 5% respectivamente, donde se observa que la Prueba de MDS muestra un comportamiento con tendencia a la baja a medida que aumenta el número de muestras perdidas en el experimento, lo cual es apreciable desde el punto de vista estadístico, a diferencia del ANOVA y las pruebas de Tukey y MDS, las cuales, aunque muestran un crecimiento a medida que aumenta el número de muestras perdidas por experimento, se mantienen alrededor del valor teórico del nivel de significación.

Gráfico 8.

Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 1% y Muestras perdidas por experimento, para experimentos con número de tratamientos grande y muestras por tratamiento pequeñas (Grupo 2)

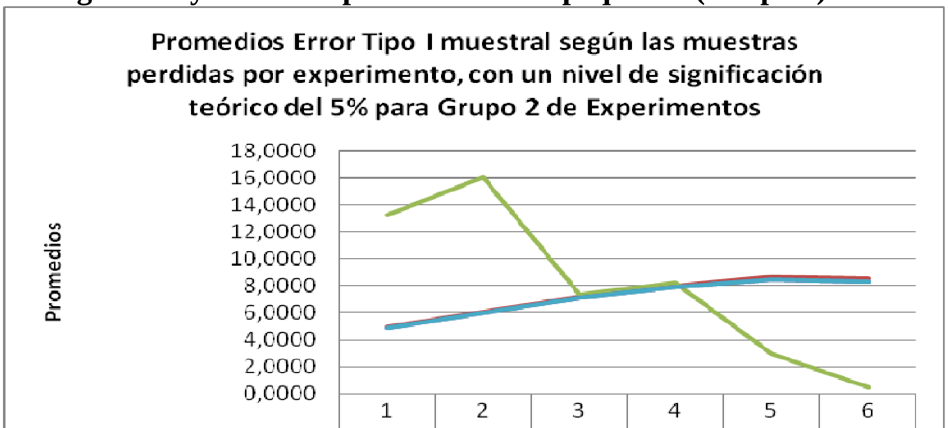


	1	2	3	4	5	6
Promedio de %Anova	1,0074	1,3888	1,8455	2,1419	2,3600	2,2633
Promedio de %MDS	1,3035	2,4185	0,6545	0,6871	0,1169	0,0033
Promedio de %Tukey	0,9797	1,3743	1,8770	2,1425	2,3288	2,1367
Promedio de %SNK	0,9797	1,3743	1,8770	2,1425	2,3288	2,1367

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Gráfico 9.

Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 5% y Muestras perdidas por experimento, para experimentos con número de tratamientos grandes y muestras por tratamiento pequeñas (Grupo 2)



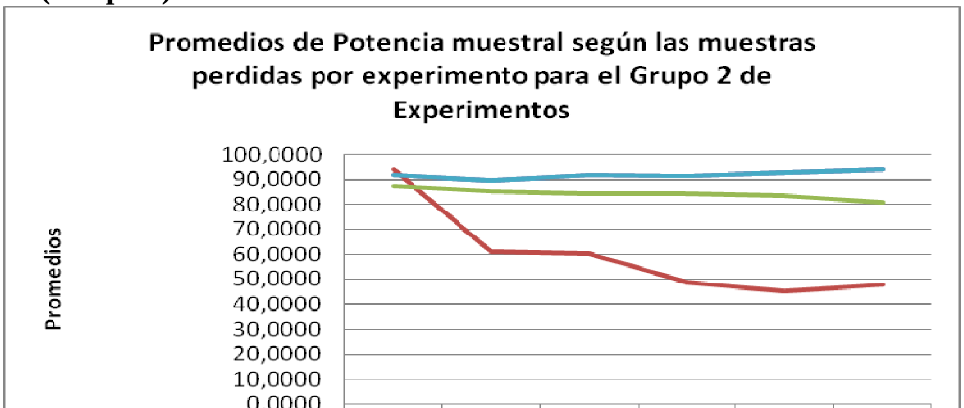
	1	2	3	4	5	6
Promedio de %Anova	4,9988	6,0610	7,1939	8,0092	8,6775	8,5833
Promedio de %MDS	13,2764	16,0560	7,3606	8,2073	2,9375	0,5058
Promedio de %Tukey	4,9142	5,9928	7,1265	7,9261	8,4828	8,2919
Promedio de %SNK	4,9142	5,9928	7,1265	7,9261	8,4828	8,2919

Fuente: Eduardo Vargas y Edwin Vargas (2014)

De igual manera, se determinó la Potencia muestral para este grupo de experimentos, mostrando nuevamente que el ANOVA es la prueba que pierde más potencia al aumentar el número de muestras perdidas por experimento cuando los experimentos tienen numero de tratamientos grandes y muestras por tratamiento pequeñas. Esta información se visualiza en el gráfico 10.

Ese de notar en este gráfico 10 que las Pruebas de Tukey y SNK muestran mayor probabilidad muestral de detectar diferencia entre las medias de los tratamientos que la Prueba de MDS.

Gráfico 10.
Promedios de Potencia muestral obtenidos, según muestras perdidas por experimento, para experimentos con número de tratamientos grandes y muestras por tratamiento pequeñas (Grupo 2)



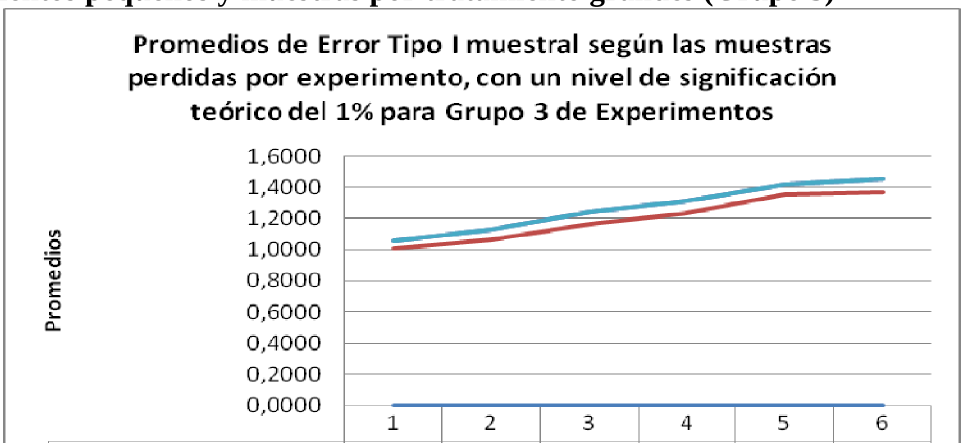
	1	2	3	4	5	6
Promedio de %Anova	94,0605	61,0137	60,3206	48,6508	45,2053	48,1092
Promedio de %MDS	87,4818	85,1949	84,2312	84,2711	83,2440	80,8069
Promedio de %Tukey	91,6177	89,6411	91,9994	91,5293	92,7247	93,9678
Promedio de %SNK	91,6177	89,6411	91,9994	91,5293	92,7247	93,9678

Fuente: Eduardo Vargas y Edwin Vargas (2014)

3.4.4. Resultados para Experimentos con número de tratamientos pequeños y muestras por tratamiento grandes (Grupo 3)

Los Gráficos 11 y 12 muestran los resultados obtenidos para el Error Tipo I muestral en función de las muestras perdidas por experimento para los niveles de significación teóricos del 1% y 5% respectivamente, donde se observa que la Prueba de MDS arrojó 0% de error tipo I muestral para todos los valores de número de muestras perdidas por experimento lo cual es apreciable desde el punto de vista estadístico. La prueba ANOVA y las pruebas de Tukey y MDS en cambio, aunque muestran un crecimiento a medida que aumenta el número de muestras perdidas por experimento, se mantienen alrededor del valor teórico del nivel de significación.

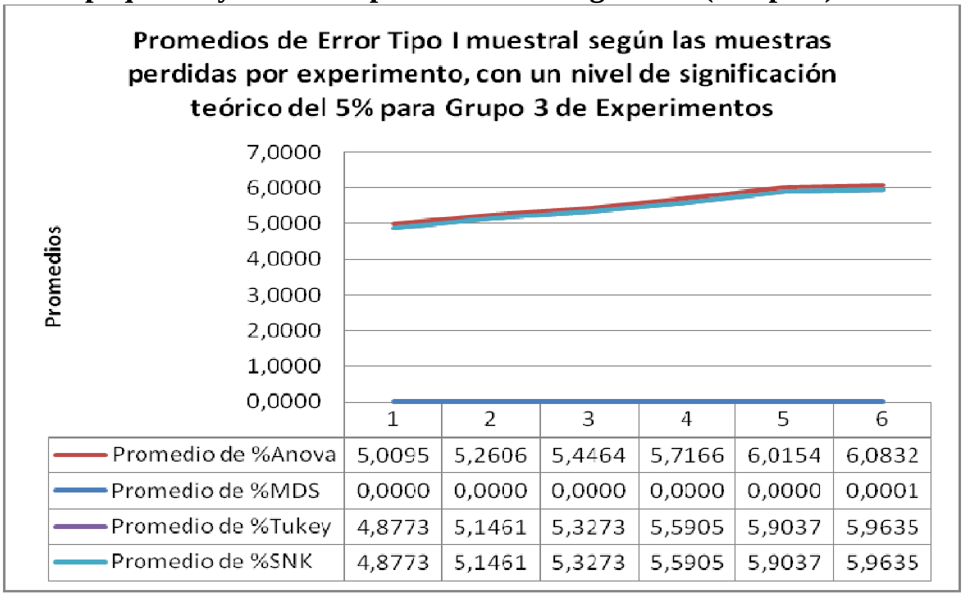
Gráfico 11.
Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 1% y Muestras perdidas por experimento, para experimentos con número de tratamientos pequeños y muestras por tratamiento grandes (Grupo 3)



	1	2	3	4	5	6
Promedio de %Anova	1,0052	1,0658	1,1641	1,2332	1,3528	1,3720
Promedio de %MDS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Promedio de %Tukey	1,0598	1,1271	1,2385	1,3098	1,4208	1,4570
Promedio de %SNK	1,0598	1,1271	1,2385	1,3098	1,4208	1,4570

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Gráfico 12.
Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 1% y Muestras perdidas por experimento, para experimentos con número de tratamientos pequeños y muestras por tratamiento grandes (Grupo 3)

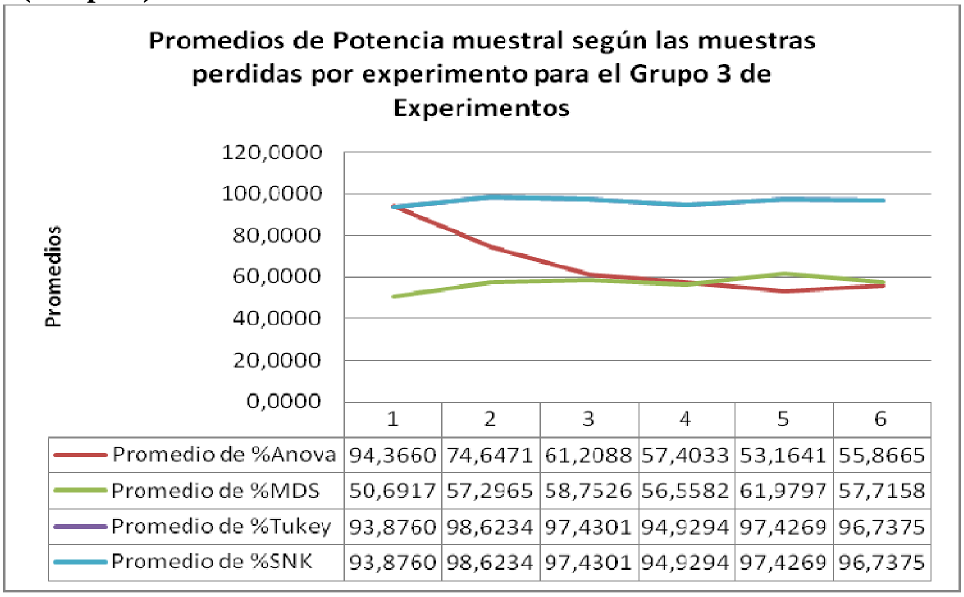


Fuente: Eduardo Vargas y Edwin Vargas (2014)

De igual manera, se determinó la Potencia muestral para este grupo de experimentos, mostrando que el ANOVA y la Prueba de MDS a medida que aumenta el número de muestras perdidas por experimento tienden a estabilizarse alrededor del 60% mientras que las Pruebas de Tukey y SNK se estabilizan alrededor del 96%. Esta información se visualiza en el gráfico 13.

Ese de notar en este gráfico 13 que las Pruebas de Tukey y SNK muestran mayor probabilidad muestral de detectar diferencia entre las medias de los tratamientos que el ANOVA y la Prueba de MDS.

Gráfico 13.
Promedios de Potencia muestral obtenidos, según muestras perdidas por experimento, para experimentos con número de tratamientos pequeños y muestras por tratamiento grandes (Grupo 3)



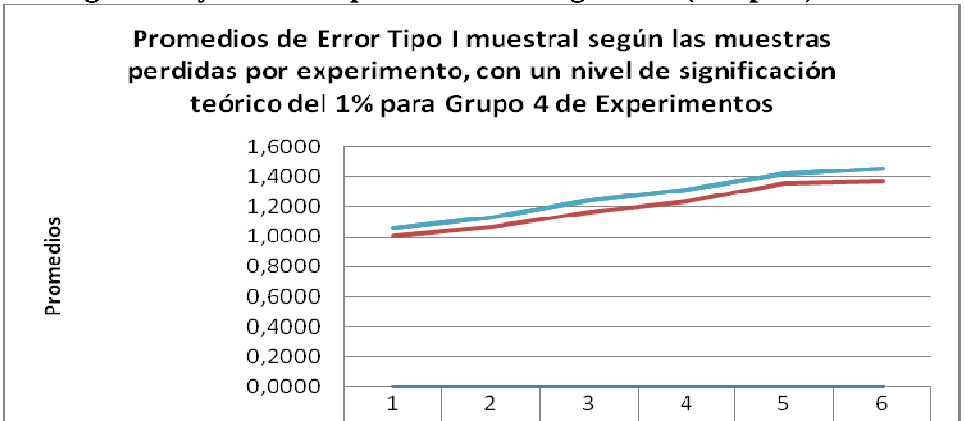
	1	2	3	4	5	6
Promedio de %Anova	94,3660	74,6471	61,2088	57,4033	53,1641	55,8665
Promedio de %MDS	50,6917	57,2965	58,7526	56,5582	61,9797	57,7158
Promedio de %Tukey	93,8760	98,6234	97,4301	94,9294	97,4269	96,7375
Promedio de %SNK	93,8760	98,6234	97,4301	94,9294	97,4269	96,7375

Fuente: Eduardo Vargas y Edwin Vargas (2014)

3.4.5. Resultados para Experimentos con número de tratamientos grandes y muestras por tratamiento grandes (Grupo 4)

Los Gráficos 14 y 15 muestran los resultados obtenidos para el Error Tipo I muestral en función de las muestras perdidas por experimento para los niveles de significación teóricos del 1% y 5% respectivamente, donde nuevamente se observa que la Prueba de MDS arrojó 0% de error tipo I muestral para todos los valores de número de muestras perdidas por experimento, lo cual es apreciable desde el punto de vista estadístico. La prueba ANOVA y las pruebas de Tukey y MDS en cambio, aunque muestran un crecimiento a medida que aumenta el número de muestras perdidas por experimento, se mantienen alrededor del valor teórico del nivel de significación.

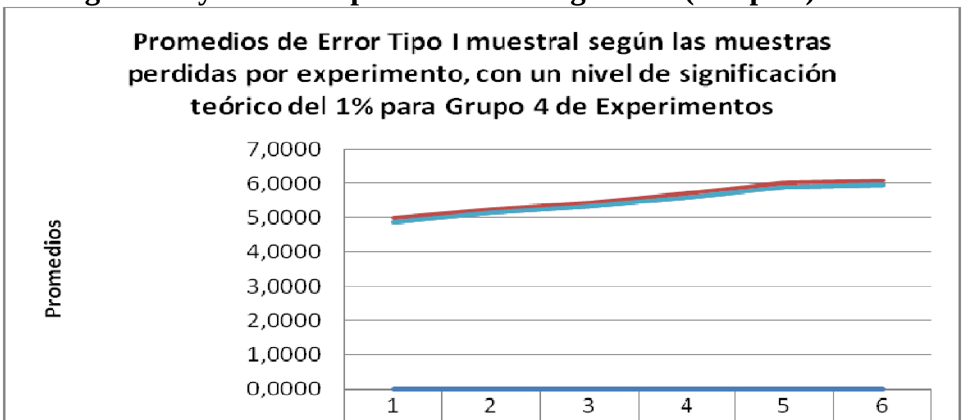
Gráfico 14.
Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 1% y Muestras perdidas por experimento, para experimentos con número de tratamientos grandes y muestras por tratamiento grandes (Grupo 4)



	1	2	3	4	5	6
Promedio de %Anova	1,0052	1,0658	1,1641	1,2332	1,3528	1,3720
Promedio de %MDS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Promedio de %Tukey	1,0598	1,1271	1,2385	1,3098	1,4208	1,4570
Promedio de %SNK	1,0598	1,1271	1,2385	1,3098	1,4208	1,4570

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Gráfico 15.
Promedios de Error tipo I muestral obtenidos, según el nivel de significación teórico del 1% y Muestras perdidas por experimento, para experimentos con número de tratamientos grandes y muestras por tratamiento grandes (Grupo 4)



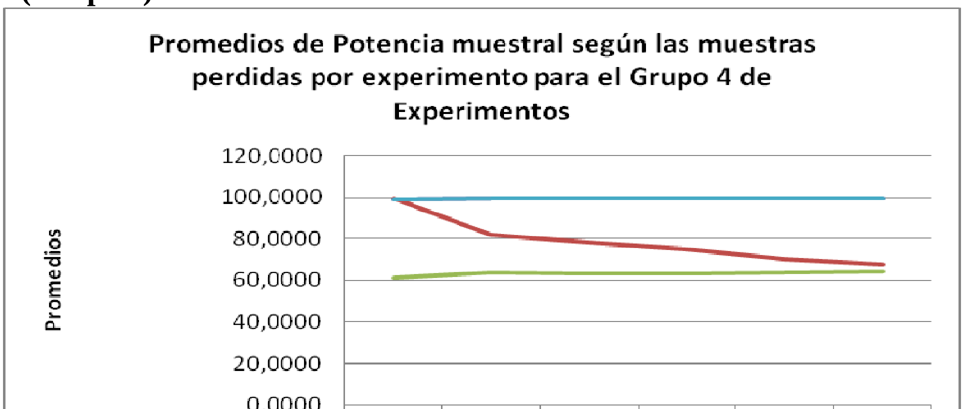
	1	2	3	4	5	6
Promedio de %Anova	5,0095	5,2606	5,4464	5,7166	6,0154	6,0832
Promedio de %MDS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
Promedio de %Tukey	4,8773	5,1461	5,3273	5,5905	5,9037	5,9635
Promedio de %SNK	4,8773	5,1461	5,3273	5,5905	5,9037	5,9635

Fuente: Eduardo Vargas y Edwin Vargas (2014)

De igual manera, se determinó la Potencia muestral para este grupo de experimentos, mostrando que el ANOVA y la Prueba de MDS a medida que aumenta el número de muestras perdidas por experimento tienden a estabilizarse alrededor del 60% mientras que las Pruebas de Tukey y SNK se estabilizan asintóticamente alrededor del 100%. Esta información se visualiza en el gráfico 16.

Ese de notar en este gráfico 16 que las Pruebas de Tukey y SNK muestran mayor probabilidad muestral de detectar diferencia entre las medias de los tratamientos que el ANOVA y la Prueba de MDS.

Gráfico 16.
Promedios de Potencia muestral obtenidos, según muestras perdidas por experimento, para experimentos con número de tratamientos grandes y muestras por tratamiento grandes (Grupo 4)



	1	2	3	4	5	6
Promedio de %Anova	99,4339	82,0523	78,4027	75,0895	70,1165	67,6706
Promedio de %MDS	61,2633	63,7702	63,6180	63,5777	63,6505	64,1803
Promedio de %Tukey	99,1123	99,4968	99,6798	99,7588	99,3921	99,7353
Promedio de %SNK	99,1123	99,4968	99,6798	99,7588	99,3921	99,7353

Fuente: Eduardo Vargas y Edwin Vargas (2014)

III.4.6. Potencia de las Pruebas según el Número de Muestras perdidas por Experimento.

Los resultados indican que las pruebas de Tukey y SNK son las que en términos generales arrojaron los mayores valores de potencia muestral en situaciones donde hay tratamientos desbalanceados (perdida de muestras por experimento). Para éstas pruebas, la potencia muestral aumenta a medida que aumenta el número de muestras perdidas por experimento.

En cambio, el ANOVA es sensible a la pérdida de muestras, disminuyendo el valor de la potencia muestral a medida que aumenta el número de muestras perdidas por experimento. Aunque ésta prueba presenta una leve mejoría en el valor de la potencia muestral cuando se tienen 5 muestras perdidas en el experimento, esta situación desde el punto de vista estadístico no es favorable.

La prueba de MDS tiene valores promedio de potencia muestral entre las pruebas de Tukey y SNK, y la prueba de ANOVA, sin embargo su comportamiento es poco deseable desde el punto de vista estadístico ya que disminuye el valor de la potencia muestral a medida que aumenta el número de muestras perdidas por experimento. La tabla 3 muestra un resumen, en cifras, de los argumentos planteados en los párrafos anteriores.

Tabla 3.
Promedios de Potencia muestral obtenidas para cada Prueba, desglosados según el número de muestras perdidas por experimento.

Muestras_Perdidas	Promedio de %Anova	Promedio de %MDS	Promedio de %Tukey	Promedio de %SNK
1	67,9363	69,4170	92,9655	92,9655
2	63,0820	69,0541	95,2634	95,2634
3	57,5271	68,7212	94,8164	94,8164
4	55,9126	68,5731	96,6577	96,6577
5	59,0006	64,0940	97,0818	97,0818
Total general	**61,0971**	**68,1982**	**95,1690**	**95,1690**

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Un mayor análisis al total general mostrado en la tabla anterior, se hace cuando para cada grupo de experimentos (los cuales se clasificaron según el número de tratamientos y el número de muestras por tratamiento) se determina la potencia muestral en función de número de muestras perdidas por experimento (recordando que Grupo 1: entre 3 y 6 tratamientos, y entre 3 y 6 muestras por tratamiento; Grupo 2: entre 10 y 20 tratamientos, y entre 3 y 6 muestras por tratamiento; Grupo 3: entre 3 y 6 tratamientos y entre 10 y 50 muestras por tratamiento; Grupo 4: entre 10 y 20 tratamientos, y entre 10 y 50 muestras por tratamiento).

La Tabla 4 muestra entre otros aspectos, como para experimentos con número de tratamientos pequeños y número muestras por tratamiento pequeños (Grupo 1) la potencia muestral es inferior a cuando se tienen experimentos con número de tratamiento grandes y número de muestras por tratamiento grandes (Grupo 4).

Tabla 4.
Promedios de Potencia muestral obtenidas para cada Prueba, desglosados según el Grupo y Número de muestras perdidas por experimento.

Modelo	Muestras_Perdidas	Promedio de %Anova	Promedio de %MDS	Promedio de %Tukey	Promedio de %SNK
Grupo_1	1	43,4641	71,4883	78,6335	78,6335
	2	32,0442	77,0480	85,5069	85,5069
	3	27,1415	78,5265	87,2026	87,2026
	4	25,5799	81,5101	90,3804	90,3804
	5	25,9831	70,2683	83,5723	83,5723
Total Grupo_1		**32,9659**	**75,8435**	**84,3843**	**84,3843**
Grupo_2	1	61,0137	85,1949	89,6411	89,6411
	2	60,3206	84,2312	91,9994	91,9994
	3	48,6508	84,2711	91,5293	91,5293
	4	45,2053	83,2440	92,7247	92,7247
	5	48,1092	80,8069	93,9678	93,9678
Total Grupo_2		**54,5761**	**84,1217**	**91,4283**	**91,4283**
Grupo_3	1	74,6471	57,2965	98,6234	98,6234
	2	61,2088	58,7526	97,4301	97,4301
	3	57,4033	56,5582	94,9294	94,9294
	4	53,1641	61,9797	97,4269	97,4269
	5	55,8665	57,7158	96,7375	96,7375
Total Grupo_3		**60,4580**	**58,4606**	**97,0295**	**97,0295**
Grupo_4	1	82,0523	63,7702	99,4968	99,4968
	2	78,4027	63,6180	99,6798	99,6798
	3	75,0895	63,5777	99,7588	99,7588
	4	70,1165	63,6505	99,3921	99,3921
	5	67,6706	64,1803	99,7353	99,7353
Total Grupo_4		**74,6663**	**63,7594**	**99,6126**	**99,6126**
Total general		**61,0971**	**68,1982**	**95,1690**	**95,1690**

Fuente: Eduardo Vargas y Edwin Vargas (2014)

3.4.7. Potencia de las Pruebas según el Coeficiente de Variación y el Número de Muestras perdidas por Experimento.

Iniciaremos este aparte mostrando los resultados globales de la Potencia muestral de cada prueba simulada, para cada Coeficiente de Variación, tal como se muestra en la tabla 5.

Tabla 5.
Promedios de Potencia muestral obtenidas según el Coeficiente de Variación.

	Coeficientes de Variación						
Prueba	1%	5%	10%	20%	50%	100%	Total
ANOVA	4,6325	33,9076	64,0827	89,0907	94,5584	80,3107	61,0971
MDS	99,9999	99,2190	96,8976	77,2596	27,9393	7,8737	68,1982
Tukey	100,0000	99,8906	99,7703	99,1652	95,2984	76,8894	95,1690
SNK	100,0000	99,8906	99,7703	99,1652	95,2984	76,8894	95,1690

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Es de recordar, que los valores anteriores no se discriminan según el número de muestras perdidas por experimento el cual va desde 1 hasta 5 muestras. Se observa que la Prueba ANOVA es la más sensible a la pérdida de información o al desbalance en el número de muestras por tratamiento con una tendencia a aumentar su potencia cuando aumenta el coeficiente de variación, mientras que las Tukey y SNK muestran un comportamiento más acorde al concepto de Coeficiente de Variación, ya que a medida que las muestras son más heterogéneas, la Potencia muestral de éstas pruebas va disminuyendo. La Prueba de MDS tiene un comportamiento creciente hasta un 10% de Coeficiente de Variación, luego disminuye su valor de potencia a niveles inferiores a un 10%, lo cual desde el punto de vista estadístico no es apreciable.

La tabla 6 muestra los resultados de la simulación desglosando para cada Coeficiente de Variación, los valores de Potencia muestral según el número de muestras perdidas por experimento.

Tabla 6.
Promedios de Potencia muestral obtenidas para cada Coeficiente de Variación, desglosados según el número de muestras perdidas por experimento.

	Coeficiente de Variación		1%		
	Muestras perdidas por Experimento				
Prueba	1	2	3	4	5
ANOVA	8,7987	9,6060	1,2500	1,4286	0,0000
MDS	100,0000	100,0000	99,9997	100,0000	100,0000
Tukey	100,0000	100,0000	100,0000	100,0000	100,0000
SNK	100,0000	100,0000	100,0000	100,0000	100,0000
	Coeficiente de Variación		5%		
	Muestras perdidas por Experimento				
Prueba	1	2	3	4	5
ANOVA	55,7427	36,5659	29,1688	22,0964	17,7088
MDS	99,5244	100,0000	98,7500	98,5714	99,1004
Tukey	99,5382	100,0000	100,0000	100,0000	100,0000
SNK	99,5382	100,0000	100,0000	100,0000	100,0000
	Coeficiente de Variación		10%		
	Muestras perdidas por Experimento				
Prueba	1	2	3	4	5
ANOVA	80,9572	69,1190	58,0459	46,2109	60,9556
MDS	96,6313	99,6050	96,0272	98,2696	93,2470
Tukey	99,1056	99,9996	99,9715	100,0000	99,9252
SNK	99,1056	99,9996	99,9715	100,0000	99,9252
	Coeficiente de Variación		20%		
	Muestras perdidas por Experimento				
Prueba	1	2	3	4	5
ANOVA	96,2294	88,3389	84,7183	85,1363	89,8283
MDS	77,6303	74,9716	78,1534	80,1236	75,2209
Tukey	98,1192	99,6624	99,1848	99,9723	99,1038
SNK	98,1192	99,6624	99,1848	99,9723	99,1038
	Coeficiente de Variación		50%		
	Muestras perdidas por Experimento				
Prueba	1	2	3	4	5
ANOVA	92,2058	94,4139	94,8275	94,4476	98,0503
MDS	31,6704	31,1186	30,5933	27,4754	15,1059
Tukey	91,5304	95,6516	95,8183	96,7944	98,0413
SNK	91,5304	95,6516	95,8183	96,7944	98,0413
	Coeficiente de Variación		100%		
	Muestras perdidas por Experimento				
Prueba	1	2	3	4	5
ANOVA	73,6839	80,4483	77,1518	86,1556	87,4604
MDS	11,0453	8,6297	8,8034	6,9986	1,8898
Tukey	69,4995	76,2666	73,9238	83,1796	85,4203
SNK	69,4995	76,2666	73,9238	83,1796	85,4203

Fuente: Eduardo Vargas y Edwin Vargas (2014)

Es de notarse en la tabla 6 como la Prueba ANOVA tiene un comportamiento contrario a lo deseable cuando hay muestras perdidas en un experimento, es decir, cuando hay desbalance en el número de muestras por tratamiento, llegando inclusive a arrojar un resultado de 0% de potencia muestral, cuando se tiene un coeficiente de variación de 1% y 5 muestras perdidas en un experimento. También se observa que las pruebas de Tukey y SNK mantienen valores de potencia muestral por encima de 90% para experimentos con coeficiente de variación entre 1% y 50%, lo cual desde el punto de vista estadístico es deseable. También es importante señalar que las dos pruebas citadas anteriormente mejoran su potencia a medida que aumenta el desbalance en los tratamientos; sólo se observan valores de potencia muestral por debajo de 90% cuando el Coeficiente de Variación es de 100%. Por otra parte, la Prueba de MDS es sensible al aumento del Coeficiente de Variación y al aumento del desbalance en el número de muestras por tratamiento, pasando de tener 100% de potencia muestral en experimentos con 1% de Coeficiente de Variación y una muestra perdida por experimento, a tener 1.88% de potencia muestral en experimentos con 100% de Coeficiente de Variación y 5 muestras perdidas por experimento.

CAPITULO IV

CONCLUSIONES Y RECOMENDACIONES

IV.1. CONCLUSIONES

La relevancia de la investigación se fundamentó en el hecho de dar a conocer el comportamiento de las pruebas de Análisis de Varianza (ANOVA), Mínima Diferencia Significativa (MDS), Tukey, y Student-Newman-Keuls (SNK), en la situación de desbalance en el número de muestras por tratamiento, lo cual se logró simulando perdida de muestras en los experimentos diseñados para este fin. Por ello, el presente trabajo de investigación arrojó las siguientes conclusiones:

- La Prueba de Análisis de Varianza (ANOVA) pierde robustez estadística a medida que los experimentos pierden muestras en sus tratamientos.
- Las Pruebas de Tukey y SNK mantienen niveles de potencia muestral por encima del 90% en términos generales, lo cual desde el punto de vista estadístico es deseable.
- La Prueba de MDS es más uniforme en su comportamiento con relación a la potencia muestral observada, ya que para los distintos valores simulados de número de muestras perdidas por experimento, el valor de la potencia muestral oscila entre 64% y 70%
- La Prueba de ANOVA se ve favorecida en cuanto a su potencia muestral cuando el número de tratamientos es grande y las muestras por tratamiento también son grandes.
- En todos los casos simulados, es deseable desde el punto de vista estadístico, utilizar las Pruebas de Tukey o SNK para detectar diferencias significativas entre medias de tratamiento, cuando se tenga desbalance en el número de muestras por tratamiento, ya que fueron las pruebas que arrojaron la mayor potencia muestral en el conjunto de experimentos simulados.

- La Prueba de MDS sólo logra, en términos generales, obtener una potencia muestral mayor al 80%, cuando el número de tratamientos es grande y las muestras por tratamiento son pequeñas.

IV.2. RECOMENDACIONES

El análisis del comportamiento en términos de error tipo I y potencia de los estadísticos que se usan en la detección de diferencias significativas en las medias de tratamientos para experimentos balaceados resulta de especial interés para los investigadores tanto del área de Cálculo Probabilístico-Estadístico como para los investigadores de otras áreas del quehacer científico que utilizan estas herramientas dentro de sus estudios. Por ello, luego de analizar las pruebas de Análisis de Varianza (ANOVA), Mínima Diferencia Significativa (MDS), Tukey, y Student-Newman-Keuls (SNK), en las diferentes situaciones de esta investigación, se recomienda lo siguiente:

- Simular situaciones con un mayor número de muestras perdidas por tratamiento, para establecer comportamientos asintóticos de la potencia muestral observada para cada prueba estudiada.
- Incorporar el resto de las Pruebas de Comparaciones Múltiples de Medias a analizar en concordancia con las limitaciones de la investigación, para ofrecer a los investigadores un panorama más amplio de orientación al momento de seleccionar el procedimiento de comparación de medias de tratamiento que mejor se adapte a su situación estudiada.
- Para los resultados obtenidos de este trabajo, aumentar su análisis tomando en cuenta elementos con número de tratamientos y cantidad de muestras por tratamiento.

BIBLIOGRAFÍA

Atil, H. and Unver, Y.: "***Multiple Comparisons*"**. Journal of Biological Sciences 1 (8): 723-727, 2001.

Bellorín, L., y Rivas, J: "***Técnicas de Documentación e Investigación I***". Universidad Nacional Abierta, Caracas, Sexta Edición, 2000.

García, R.: "***Inferencia Estadística y Diseños de Experimentos***". Editorial Universitaria de Buenos Aires, Ciudad de Buenos Aires, Primera Edición, 2004.

Kuehl, R.: ***"Diseño de Experimentos"***. Editorial Thomson. Mexico, 2000.

Li, Ch.: "***Introducción a la Estadística Experimental***". Ediciones Omegam Barcelona. Primera Edición (traducida), 1977.

Little, T., y Hills, F.: "***Métodos Estadísticos para la Investigación en la Agronomía***". Editorial Trillas, México, Segunda Edición, 1998.

Montgomery, D.: "***Probabilidad y Estadística Aplicadas a la Ingeniería***". Editorial McGraw Hill, México, Primera Edición, 1998.

---------------------.: "***Diseño y Análisis de Experimentos***". Editorial Limusa, México, Segunda Edición, 2004.

Steel, R., y Torrie, J.: "***Bioestadística: Principios y Procedimientos***". Editorial McGraw Hill

Vargas, E. "***Determinación del procedimiento de comparación múltiple de medias de tratamientos para experimentos balanceados con la mayor potencia, usando muestras simuladas"***. Trabajo de Ascenso para ascender a la categoría de Profesor Agregado.