

Comparación de los Métodos de Mínima Diferencia Significativa (LSD) y Dunnett para experimentos balanceados

Analysis of multiple comparison methods of mean least significant difference (LSD) and Dunnett for balanced experiments using simulated samples

Eduardo Vargas Cano, Edwin Vargas Cano

Palabras clave: Mínima Diferencia Significativa, Prueba de Dunnett, Simulación de muestras

Key words: Least Significant Difference, Dunnett test, simulation samples

RESUMEN

La investigación está centrada en analizar los métodos de comparación múltiple de medias de tratamientos de Mínima Diferencia Significativa (LSD) y Dunnett para experimentos balanceados, según su potencia muestral observada, usando muestras simuladas. Es de vital importancia una vez realizado el experimento, definir si existe o no una diferencia significativa en las medias de los tratamientos, y contar con un procedimiento estadístico confiable que permita observar dicha diferencia. Para ello, se simularon 1920 situaciones experimentales, a las cuales se le aplicó el ANOVA y cada una de las dos pruebas antes mencionadas para determinar diferencias significativas entre las medias de tratamientos, determinando un error tipo I o tipo II muestral según sea el caso. En general, se obtuvo que en términos de la potencia, la prueba de Mínima Diferencia Significativa resulta ser más potente que la prueba de Dunnett, evidenciando tener una mayor probabilidad muestral de detectar diferencias significativas entre las medias de tratamientos, cuando éstas realmente existen.

ABSTRACT

The research is focused on analyzing the methods of multiple comparison of means of treatments of Minimum Significant Difference (LSD) and Dunnett for balanced experiments, according to their observed sample power, using simulated samples. It is of vital importance once the experiment is done, to define whether or not there is a significant difference in the means of the treatments, and to have a reliable statistical procedure that allows to observe this difference. For this, 1920 experimental situations were simulated, to which the ANOVA was applied and each of the two tests mentioned above to determine significant differences between the treatment means, determining a type I error or sample type II, as the case may be. In general, it was found that, in terms of power, the Minimum Significant Difference test turns out to be more powerful than the Dunnett test, evidencing having a greater probability of sampling to detect significant differences between treatment means, when these actually exist.

INTRODUCCIÓN

El diseño de experimentos está referido a los métodos de muestreo para reducir la variación en función de amplificar la señal de la naturaleza y así adquirir una cantidad especificada de información a través de éstos. Este muestreo aleatorio tiene dos propósitos. Primero, evitar la posibilidad de sesgo, y segundo, suministrar una base probabilística para la selección de la muestra (Mendenhall, Beaver & Beaver, 2015). A la selección de muestras aleatorias independientes de k poblaciones se le denomina *Diseño Completamente aleatorizado*.

Investigadores de prácticamente todos los campos de estudio realizan experimentos, por lo general para descubrir algo acerca de un proceso o un sistema, donde el experimento se puede definir como una prueba o serie de pruebas en las que se hacen cambios deliberados en las variables de entrada de un proceso o sistema (diseño de tratamientos) para observar e identificar las razones de los cambios en la respuesta (Montgomery, 2004). Los tratamientos deberán relacionarse con los objetivos de investigación y el diseño de tratamientos requiere la identificación de los mismos, en cuanto al papel que desempeñan, para la evaluación de la hipótesis. Cuando se realiza un experimento para contestar a preguntas específicas, los tratamientos se seleccionan de manera que las comparaciones entre ellos contesten esas preguntas, aquí juega un papel importante la etapa de diseño de experimentos.

Los investigadores de las áreas como la biología, medicina, agricultura, ingeniería psicología y otras ciencias, utilizan experimentos del tipo comparativo para probar sus hipótesis. Kuehl (2001) declara que 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 (tratamientos).

Para probar diferencias significativas entre tratamientos, en una forma elemental se usa la prueba F basándose en un ANOVA (Análisis de la Varianza). 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. Si se rechaza la hipótesis nula cuando se usa la prueba F , entonces surge la siguiente interrogante: ¿Dónde están las diferencias reales? Steel & Torrie (1985) plantean que si tal pregunta no se formula, entonces es razonablemente claro pensar que la máxima diferencia observada puede declararse significativa ya que H_0 se ha rechazado. Para responder a la interrogante planteada, se sugiere la utilización de procedimientos de comparación múltiple permiten detectar las diferencias en las medias de tratamiento.

Se debe tener claro cuál es la herramienta de comparación más idónea a ser utilizada para una determinada situación de análisis.

Según Atil & Unver (2001), en diferentes situaciones se ha apreciado el uso inadecuado de procedimientos de comparación múltiple más significativa para el conjunto de datos a analizar.

Por ello, resulta conveniente el señalamiento del comportamiento de los diferentes procedimientos de comparaciones múltiples de medias, según la naturaleza de la data muestral o simulada, tal como lo afirma el trabajo realizado por Rafter, Abell & Braselton (2002), donde se incluye además, sugerencias sobre los mejores métodos. Sauder & DeMars (2019) hacen una recomendación para el uso de procedimientos de comparaciones múltiples de medias en estudios del área de la psicología, con el uso de simulación usando el software SPSS. Basu, Cai, Das & Sun (2018) desarrollaron procedimientos orientados a datos que tienen como objetivo maximizar el número esperado de positivos verdaderos sujetos a una restricción en la tasa ponderada de descubrimientos falsos

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 sus propiedades con el objeto de identificar cuál afecta con mayor o menor valor, a la media de las observaciones.

Aspectos Teóricos

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 (Citado por Montgomery, 2004; Kuehl, 2001).

Cada hipótesis $H_0: U_i = U_j$ se puede probar con el estadístico T-Student:

$$t_0 = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{CMEE \left(\frac{1}{r_i} + \frac{1}{r_j} \right)}} \quad (1)$$

donde:

- \bar{y}_i : Media muestral de la muestra "i".
- \bar{y}_j : Media muestral de la muestra "j".
- CMEE: Cuadrado Medio del Error Experimental.
- r_i, r_j : Número de replicas.

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 $|\bar{y}_i - \bar{y}_j|$ 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:

$$MDS(\alpha) = t_{\frac{\alpha}{2}, gLEE} \cdot S(\bar{d}) \quad (2)$$

donde:

- $MDS(\alpha)$: mínima diferencia significativa al nivel de significación α
- $t_{\frac{\alpha}{2}, gLEE}$: 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:

$$S(\bar{d}) = \sqrt{\frac{2 \cdot CMEE}{r}} \quad (3)$$

donde:

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

La hipótesis nula $H_0: U_i = U_j$, se rechaza si:

$$|\bar{y}_i - \bar{y}_j| \text{ si } |t_0| > MDS(\alpha) \quad (4)$$

(Todas las diferencias entre medias son comparadas con el MDS (α) calculado).

En total se pueden generar $t(t-1)\frac{1}{2}$ comparaciones. Si la diferencia excede el MDS (α), se dice que las medias provienen de poblaciones distintas (Montgomery, 2004; Kuehl, 2001).

Prueba de Dunnett

Probablemente uno de los casos más 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 está dado por (Montgomery, 2004; Kuehl, 2001):

$$d' = t(\alpha, t) \cdot S(\bar{d}) \quad (5)$$

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 :

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

Las estimaciones de los intervalos de confianza bilaterales para las medias de los tratamientos individuales y la media del control para: $H_0: \mu_i - \mu_c = 0$ son:

$$\bar{y}_i - \bar{y}_c \pm d' \quad (6)$$

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:

$$\bar{y}_i - \bar{y}_c - d' \quad (7)$$

En caso de ser menor que la media control, son:

$$\bar{y}_i - \bar{y}_c + d' \quad (8)$$

Para la prueba $H_0: \mu_i - \mu_c = 0$ contra la alternativa $H_a: \mu_i - \mu_c \neq 0$, se rechaza si:

$$|\bar{y}_i - \bar{y}_c| > d' \quad (9)$$

Para la alternativa de un lado $H_0: \mu \leq \mu_j$ contra la alternativa $H_a: \mu \geq \mu_j$ se rechaza si:

$$\bar{y}_i - \bar{y}_c \geq d' \quad (10)$$

Para la alternativa de un lado $H_0: \mu \geq \mu_j$ contra la alternativa $H_a: \mu \leq \mu_j$, se rechaza si:

$$\bar{y}_i - \bar{y}_c < -d' \quad (11)$$

METODOLOGÍA

Simulación de muestras

Giri & Miguel (2018) sostienen que un modelo de simulación permite crear conocimiento en forma de explicaciones causales, tomando en cuenta la validez de exportar resultados del modelo al mundo modelado en virtud de la similitud entre modelo y mundo, analizable en términos de identidad parcial de estructura para eliminar la similitud superficial que repita los resultados empíricos al ajustar datos por calibración.

El experimento simulado es el Diseño Experimental Completamente al Azar y 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.
- 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 bibliografía 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.

- iii. *Medias de tratamientos*. Para esta variable, se consideraron dos casos: medias iguales y medias diferentes. Intentando abarcar la mayor cantidad de casos, se simularon experimentos con medias en el orden hipotético de 0.1, 1.0, 10.0 y 100.0 unidades de medida.
- iv. *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%.
- v. *Coefficiente de Variación*. Esta variable permite determinar la varianza común de todos los tratamientos. Es de importancia estadística ya que mide 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 128 casos, lo cual constituye una base para generar conclusiones de interés en el campo estadístico. Además, cada una de estas 128 situaciones, será probada con 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, y bajo las condiciones de medias iguales para medir el error tipo I muestral, y medias diferentes

para medir el error tipo II, o mejor aún, la potencia muestral simulada de la prueba. Para cada situación se generaron 10000 muestras, a las cuales se le aplicó el ANOVA y cada una de las dos pruebas antes mencionadas para determinar diferencias significativas entre las medias de tratamientos, determinando un error tipo I (si se simulan medias iguales) o tipo II muestral según sea el caso.

Las situaciones simuladas se agrupan en dos grupos:

- Medias iguales, diferentes varianzas, diferentes niveles de significancia
- Medias diferentes, diferentes varianzas, diferentes niveles de significancia.

RESULTADOS

La probabilidad (muestral) de ocurrencia del error de tipo I para las pruebas de

ANOVA, Mínima Diferencia Significativa (MDS) y Dunnett, partiendo de niveles de significación de 1% 5%, se observan en la tabla 1.

Tabla 1. Probabilidad de Error Tipo I, obtenido en la simulación

		Prueba	Coeficiente de Variación					
			1%	5%	10%	20%	50%	100%
Nivel de Significación	0,01	ANOVA	0,9763	0,9563	0,9525	1,0538	0,9775	0,9450
		MDS	10,5600	10,4063	10,3994	10,3013	10,2538	10,1413
		DUNNETT	1,0263	1,0063	0,9850	1,0825	0,9850	0,9613
Nivel de Significación	0,05	ANOVA	4,9950	4,8125	5,1075	5,0325	5,0150	5,4775
		MDS	37,8250	37,6000	37,7175	40,1250	38,0150	38,1238
		DUNNETT	5,7750	4,8388	5,1138	5,0913	5,0700	4,9400

En la tabla 1 se puede observar que:

- Los porcentajes de las veces que las diferentes pruebas detectaron una diferencia significativa entre las medias de los tratamientos. Estos porcentajes pueden considerarse son la tasa de error tipo I muestral, ya que originalmente las medias de los tratamientos se consideraron como iguales. Las tasas de error tipo I teórica para estas corridas fueron

de 0.01 y 0.05, por tanto es de esperar que los porcentajes obtenidos en la simulación estén alrededor del 1% y 5% según sea el caso.

- Para un nivel de significación de 0,01, que tanto para el ANOVA como para la prueba de Dunnett, el porcentaje de error tipo I aunque sobre pasa el 1%, no supera el 2%.
- Para un nivel de significación de 0.05 en el ANOVA y la prueba de

Dunnett, el porcentaje de error tipo I aunque en algunos valores del coeficiente de variación supera el 5%, no sobre pasa el 6%. Además, también se determinó la potencia (muestral) obtenida para las pruebas de Mínima Diferencia Significativa y Dunnett, partiendo de niveles de significación de 1%

y 5%, la cual se observa en la tabla 2, y se midió como el porcentaje de las veces que teniendo experimentos con medias de tratamientos diferentes, las pruebas de Mínima Diferencia Significativa y Dunnett detectaron diferencias estadísticamente significativas entre éstas.

Tabla 2. Potencia muestral observada, obtenida en la simulación

		Prueba	Coeficiente de Variación					
			1%	5%	10%	20%	50%	100%
Nivel de Significación	0,01	ANOVA	100	99,4088	94,7988	85,8875	78,2425	59,5175
		MDS	100	99,8888	97,6713	91,6300	83,7350	73,2575
		DUNNETT	100	99,5650	95,4463	86,4963	79,0263	61,9125
	0,05	ANOVA	100	99,9075	97,4925	90,3163	81,9613	71,2100
		MDS	100	99,9988	99,3100	96,1500	90,4075	85,7700
		DUNNETT	100	99,9450	97,9038	90,7488	82,3338	72,8850

De la tabla 2 se observa que:

- Al considerar medias diferentes y realizar el conteo de las veces que las pruebas detectan diferencias significativas, se debe esperar que las pruebas mantengan sus porcentajes de detección de diferencias significativas en números cercanos al 100%. Tal situación se refleja hasta un coeficiente de variación del 20%, donde a partir de este valor, se observa que la prueba que sufre una disminución considerable es la de Dunnett.
- La prueba que logra mantener en mayor porcentaje altos porcentajes de detección de diferencias significativas es la de Mínima Diferencia Significativa.

- Es de entender que a mayor varianza, las distribuciones de los tratamientos tienden a solaparse unas con otras, haciendo más compleja la detección de diferencias significativas entre las medias de los tratamientos.

Otras de las estadísticas muestrales de interés determinadas con la data simulada, fueron la probabilidad condicionada muestral de que siendo significativa la prueba de ANOVA las pruebas de MDS y Dunnett resulten significativas, y la probabilidad de que siendo no significativo el ANOVA las pruebas de MDS y Dunnett resulten significativas, todas éstas, según la naturaleza de las medias y el nivel de significación. Tales valores se observan en las tablas 3 y 4, respectivamente.

Tabla 3. Probabilidad condicional muestral de que siendo ANOVA significativo, las pruebas de MDS y Dunnett resulten significativas

		Prueba	Naturaleza de las Medias	
			Medias Iguales	Medias Diferentes
Nivel de Significación	0,01	MDS	100,00	100,00
		DUNNETT	59,78	97,13
	0,05	MDS	100,00	100,00
		DUNNETT	65,35	97,93

De la tabla 3 se puede observar que:

- Independiente de la naturaleza de las medias y del nivel de significación, la prueba de Mínima Diferencia Significativa detecta diferencias significativas entre las medias de tratamiento el 100% de las veces que el ANOVA detecta tal situación.
- La prueba de Dunnett resulta ser más conservadora en este aspecto,

ya que siendo las medias de tratamiento iguales, aunque ANOVA detecte por error significancia estadística en los tratamientos, Dunnett sólo comete este error un 62,57% de la veces, en promedio. Y siendo las medias diferentes, solo deja de detectar diferencias significativas luego de que ANOVA detecta, un 2,47% de las veces, en promedio.

Tabla 4. Probabilidad condicional muestral de que siendo ANOVA no significativo, las pruebas de MDS y Dunnett resulten significativas

		Prueba	Naturaleza de las Medias	
			Medias Iguales	Medias Diferentes
Nivel de Significación	0,01	MDS	9,45	61,62
		DUNNETT	0,43	21,70
	0,05	MDS	34,67	75,49
		DUNNETT	1,88	32,26

De la tabla 4 se observa que:

- La prueba de Mínima Diferencia Significativa resultó significativa estadísticamente en un 68,18% de las

veces que ANOVA no detectó tal significación en el caso de medias de tratamientos diferentes.

CONCLUSIONES

La importancia de la investigación se fundamentó en el hecho de dar a conocer cual es la herramienta de comparación más idónea a ser utilizada de entre las estudiadas (Mínima Diferencia Significativa y Dunnett). Por ello, el presente trabajo de investigación arrojó las siguientes conclusiones:

- El coeficiente de variación resulta ser una variable determinante tanto en el error tipo I como en la potencia muestral obtenida de la simulación.
- El Análisis de Varianza (ANOVA) muestra robustez al comparar los valores teóricos de significación estadística con los valores obtenidos muestralmente. Incluso logra mantener niveles de potencia por encima de 75% para coeficientes de variación de hasta 50%.
- La prueba de Mínima Diferencia Significativa (MDS) se comporta como la mejor en comparación con la prueba de Dunnett, en términos de la potencia muestral observada. Sin embargo, cuando se observa el comportamiento de su tasa de error tipo I muestral, la prueba de Dunnett logra mantenerse con una desviación media cuadrática de 5,5% con respecto al valor teórico fijado.
- En términos de la potencia de la prueba, la prueba de Mínima Diferencia Significativa es más potente que la prueba de Dunnett, teniendo entonces una mayor

probabilidad de detectar diferencias significativas entre las medias de tratamientos, cuando realmente existan estas diferencias.

Como comentario final, se verifica la efectividad de trabajar con coeficientes de variación bajos, es decir, menores de 25%, donde en la totalidad de las situaciones simuladas, las pruebas mantienen un comportamiento aceptable de detección o no detección de diferencias de medias significativas según sea el caso.

Obtenidos y analizados los resultados, resulta conveniente indicar que al aumentar el número de pruebas a analizar en concordancia con las limitaciones de la investigación, se puede ofrecer a los investigadores de las distintas ramas de la ciencia 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.

Es recomendable para los resultados obtenidos de este trabajo, aumentar su análisis tomando en cuenta elementos como número de tratamientos y cantidad de muestras por tratamiento. Asimismo, además de trabajar con experimentos balanceados, sería de especial interés incorporar la variable desbalance de tratamientos en los resultados, ya que esta situación es común en el desarrollo de experimentos en la práctica.

REFERENCIAS

- Atil, H. & Unver, Y. (2001). Multiple Comparisons. *Journal of Biological Sciences*, 1 (8), 723-727, 2001. <https://doi.org/10.3923/jbs.2001.723.727>
- Basu, P., Cai, T., Das, K. & Sun, W. (2018). Weighted False Discovery Rate Control in Large-Scale Multiple Testing. *Journal of the American Statistical Association*, 113 (23), 1172-1183. <https://doi.org/10.1080/01621459.2017.1336443>
- Giri, L. & Miguel, H. (2018). El modelo de simulación como generador de explicaciones causales. *Theoria: An International Journal for Theory, History and Foundations of Science*, 33 (1), 111-128. <https://www.jstor.org/stable/26355572>
- Kuehl, R. (2001). *Diseño de Experimentos*, 2da ed. México: Thompson Editores.
- Mendenhall, W., Beaver, R., & Beaver B. (2015). *Introducción a la Probabilidad y Estadística*. 14ª ed. México: Cengage Learning Editores.
- Montgomery, D. (2004). *Diseño & Análisis de Experimentos*, 2da ed. México: Editorial Limusa.
- Rafter, J., Abell, M. & Braselton, J. (2002). Multiple Comparison Methods for Means. *Society for Industrial and Applied Mathematics*, 44 (2), 259–278. <https://doi.org/10.1137/S0036144501357233>
- Sauder, D. & DeMars, C. (2019). An Updated Recommendation for Multiple Comparisons. *Advances in Methods and Practices in Psychological Science*, 2 (1), 26-44. <https://doi.org/10.1177/2515245918808784>
- Steel, R., & Torrie, J. (1985). *Bioestadística: Principios y Procedimientos*, 2da ed. Bogotá: Editorial McGraw Hill Latinoamericana.

Autores

Eduardo Vargas Cano. Ingeniero Industrial, Docente-Investigador del Departamento de Matemática, Facultad de Ingeniería, Universidad de Carabobo, Venezuela.

ORCID: <https://orcid.org/0000-0003-1293-1550>

Email: eevargas1@uc.edu.ve

Edwin Vargas Cano. Ingeniero Eléctrico, Docente-Investigador del Departamento de Matemática, Facultad de Ingeniería, Universidad de Carabobo, Venezuela.

ORCID: <https://orcid.org/0000-0002-1886-9156>

Email: vargase@uc.edu.ve

Recibido: 12-11-2021

Aceptado: 27-02-2022