

# **Evaluación mediante simulación de la potencia de prueba en diseños experimentales de medidas repetidas y de parcelas divididas, variando la magnitud del experimento y el error residual**

Camila Aguilar Gómez<sup>5</sup>, Jason Chavarría Quesada<sup>5</sup>, Francis Rodríguez Durán<sup>5</sup>

camila2099@gmail.com, jasonch541@gmail.com, rodriguezduanfrancis@gmail.com

## **RESUMEN**

Este estudio busca analizar la potencia de la prueba en dos diseños experimentales: medidas repetidas y parcelas divididas. Específicamente se centra en el estudio de la potencia para la interacción entre un factor y el tiempo, y en caso de parcelas divididas se analiza para el factor de sub- parcela. Se crean diferentes escenarios en donde se varían las condiciones iniciales de experimento, tales como: el número de sujetos, cantidad de repeticiones por sujeto, reducción del error residual, aumentar el número de niveles de un factor, aumentar el número parcelas, balancear el número de parcelas en cada nivel del factor de parcela y una reducción en el error de sub- parcela. Para ambos casos, se logra llegar a escenarios idóneos en donde la potencia de la prueba es superior a 0,80.

**PALABRAS CLAVE:** factor de sub- parcela, interacción, potencia de la prueba, simulación.

## **INTRODUCCIÓN**

En el campo experimental, es común observar investigaciones en las cuales las unidades experimentales sean personas, animales, plantas u objetos. Tales unidades, al aplicárseles cierto tratamiento, pueden reaccionar de distintas maneras, conllevando a cambios en la variable respuesta. En algunas ocasiones, estas diferencias son explicadas por las condiciones previas al experimento, o la formación de cada sujeto, y aun cuando es posible intentar controlar las condiciones preexperimentales, estas agregan variabilidad, aumentando el cuadrado medio del error. Por esta razón, Montgomery (2004) establece que una posible solución a este problema corresponde al diseño de medidas repetidas; el cual es fundamental para los análisis experimentales, ya que permite controlar la variabilidad previamente mencionada.

En los diseños de medidas repetidas, se observa a los sujetos a través del tiempo, para verificar el impacto sobre la variable respuesta, según las diferentes condiciones, lo cual se conoce como tendencias en el tiempo. Según Kuehl (2001), estas tendencias nos permiten conocer dos aspectos importantes: el primero, la velocidad de respuesta de las unidades a cada tratamiento; el segundo, durante cuánto tiempo los efectos del tratamiento impactan en las unidades de estudio.

---

<sup>5</sup> Estudiantes de Estadística de la Universidad de Costa Rica

Por otro lado, otro de los tipos de diseños experimentales, corresponde al diseño de parcelas divididas, útil cuando se tiene un experimento con múltiples factores, y uno de ellos complica o imposibilita realizar corridas aleatorias. Esto sucede cuando alguno de estos factores se tiene que aplicar sobre unidades experimentales muy grandes; por lo que, una posible solución corresponde a tomar a esta unidad experimental grande como una parcela. De igual modo, los otros factores pueden aplicarse a unidades más pequeñas, que serían divisiones de la parcela (Tapahuasco Zúñiga, 2010).

Aunado a lo anterior, una de las mediciones clave en los diseños experimentales mencionados, corresponde a la potencia de la prueba. Esta puede ser utilizada para indicar qué tan válido o creíble es el estudio, ya que puede entenderse como la probabilidad de detectar diferencias cuando realmente las hay. Ligado a esto se encuentra el tamaño de la muestra, que influye directamente en la potencia, dado que es de esperar que, a mayor tamaño de muestra, se obtenga un incremento de la potencia, y viceversa (Cárdenas Castro y Arancibia Martini, 2014).

Por la razón anteriormente mencionada, la potencia de la prueba y el tamaño de la muestra son importantes en todos los experimentos, y tal como lo menciona García-García, et al. (2013), si se omiten o no se consideran, puede llevarse al experimentador a tomar decisiones erróneas, debido al desconocimiento de una parte clave de la información.

Ahora bien, concretamente en el presente trabajo, para medidas repetidas, se busca variar el número de sujetos, número mediciones por sujeto y el error residual con el fin de determinar el comportamiento de la potencia de la prueba para la interacción entre un factor (dieta) y tiempo.

Para el caso de parcelas divididas, el interés principal se centra en identificar cómo se ve afectada la potencia de la prueba del factor de sub- parcela (horario), al cambiar el número de bloques (distritos), balancear el número de parcelas (días) en cada nivel del factor de parcela (tipo de día y una disminución en el error de sub- parcela).

## **METODOLOGÍA**

En la primera parte de la investigación, se trabajó con un ejemplo de medidas repetidas estudiado en el curso de Aplicaciones de Diseños Experimentales. El interés de este estudio era observar cómo cambia, en el tiempo, el pulso al correr según el tipo de dieta; una basada en bajo contenido graso y la otra en una dieta normal. Para esto, se le realizó tres mediciones del pulso a cada sujeto en tres tiempos diferentes: al minuto, a los 15 y a los 30 minutos. Cabe destacar que, al ser datos proporcionados en un curso, no se cuenta con el detalle de la recolección de los datos ya que son utilizados solo con fines académicos.

Dado que se quiere estudiar la potencia de la prueba asociada a la interacción entre tiempo y dieta, el modelo se expresa de la siguiente manera:

$$y_{ijk} = \beta_0 + \tau_i + \beta_1 T + \tau_j * T + \beta_{0,i} + \varepsilon_{ijk}$$

Donde  $\beta_0$  es el intercepto general fijo,  $\tau_i$  efecto de la  $i$ -ésima dieta como factor fijo,  $\beta_1 T$  coeficiente de tiempo,  $\tau_j T$  interacción entre dieta y tiempo y  $\varepsilon_{ijk}$  representa el error. Por otro lado,  $\beta_{0,i} \sim N(\mu_0, \sigma_0^2)$ ;  $\beta_{1,i} \sim N(\mu_1, \sigma_1^2)$ ;  $y \sim N(\mu, \sigma^2)$ . Con las restricciones del modelo de tratamiento de referencia:  $\gamma_1 = \text{baja en grasa} = 0$ .

Además, cada persona es aleatoria, y, como los interceptos y pendientes están asociadas a las personas, son componentes aleatorios en el modelo. Para facilitar el análisis no se tomó en cuenta la correlación entre las pendientes e interceptos.

Al realizar la estimación del modelo, se obtiene que: el coeficiente asociado a la interacción entre dieta y tiempo es 0.83; el error residual es 8.40 y las otras estimaciones se pueden observar en la tabla 1.

**Tabla 1**

*Interceptos y pendientes obtenidos del modelo de medidas repetidas*

Intercepto		Pendientes	
Baja en grasa	Normal	Baja en grasa	Normal
96.4	99.9	4.08	0.58

Para el proceso de simulación, se generan tres escenarios principales:

1. Se fija el número de sujetos en 6, tomando en cuenta 3 y 5 mediciones por sujeto, respectivamente.
2. Se fija el número de sujetos en 8, tomando en cuenta 3 y 5 mediciones por sujeto, respectivamente.
3. Se fija el número de sujetos en 10 (número inicial), tomando en cuenta 3 y 5 mediciones por sujeto, respectivamente.

Para cada uno de los escenarios planteados, se consideraron 3 errores residuales: 7, 8.39 (error inicial) y 9.

En la segunda parte, se consideró un diseño de parcelas divididas, donde el factor de distrito se compone de tres niveles: Alajuela, Curridabat y Desamparados; se tienen como parcela, la combinación entre distrito y día de la semana (21 parcelas en total); el factor de parcela es el tipo de día con dos niveles: entre semana y fin de semana; el factor de subparcela es el horario que cuenta con los niveles de mañana, tarde y noche; y la variable respuesta es el número promedio de dispositivos de seguridad, utilizados por motociclistas. Para restar complejidad a la simulación, se decide trabajar con un modelo sin ninguna interacción.

Para este diseño se emplearon datos de un estudio observacional realizado por el Observatorio del Desarrollo (OdD) en el año 2016 dirigido por el investigador Agustín Gómez Meléndez, que buscaba determinar los factores de riesgo en accidentes donde están involucradas las motocicletas. Para seleccionar la muestra se utilizó muestreo polietápico, las primeras unidades establecidas fueron los cantones, posteriormente se seleccionó un distrito en cada cantón y por último los tramos de carretera. De este modo, se emplearon dos instrumentos de recolección; uno para caracterizar los tramos de carretera y el segundo para obtener las características de los motociclistas y sus acompañantes.

Por lo que el modelo se puede escribir como:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \varepsilon_{ijl}^{(1)} + \gamma_k + \varepsilon_{ijks}^{(2)}$$

Donde  $\alpha_i$  es el efecto del i-ésimo tipo de día,  $\beta_j$  efecto del j-ésimo distrito,  $\varepsilon_{ijl}^{(1)}$  error de parcela: tipo de día dentro de distrito ( $\alpha\beta$ )<sub>ij</sub>,  $\gamma_k$  efecto del k-ésimo horario, y  $\varepsilon_{ijks}^{(2)}$  error de sub- parcela: tipo de día- horario dentro de distrito ( $\alpha\beta\gamma$ )<sub>ijk</sub> y  $y \sim N(\mu, \sigma^2)$ . Con las restricciones del modelo de tratamiento de referencia:  $\alpha_1 = \text{entre semana} = 0$ ,  $\beta_1 = \text{Alajuela} = 0$ , y  $\gamma_1 = \text{mañana} = 0$ .

Se procedió a estimar el modelo, en donde se obtuvo que la media general era de 2,02, mientras que el error de parcela fue de 0.12 y el de sub- parcela de 0.11. Por otro lado, los efectos correspondientes se detallan en la tabla 2.

**Tabla 2**

*Efectos obtenidos a partir del modelo de parcelas divididas*

	Distrito		Tipo día		Horario	
Curridabat	Desamparados		Fin de semana		Tarde	Noche
0.23	0.28	0.04		-0.15	-0.16	

Se procede a simular 3 escenarios base en donde se varían las condiciones iniciales del experimento.

Estos escenarios son:

1. Balancear el número de días en cada nivel del factor tipo de día.
2. Aumentar el número de distritos y que estos sean dentro y fuera de la GAM.
3. Disminuir el error de sub- parcela.

A partir de estos tres escenarios iniciales, se realizan combinaciones de estos, para determinar con cuáles variaciones, el factor de sub- parcela (horario), logra una potencia de la

prueba más alta. Otro aspecto para considerar es que, al realizar estos cambios, la potencia de los otros factores también se ve afectada y para efectos de visualización, se decide graficar los resultados para los 3 factores, pero solo se analizan los resultados para el factor horario.

Definidos los escenarios, se procedió a elaborar una función para cada diseño. Estas funciones facilitaron la evaluación de los diferentes escenarios para realizar las comparaciones necesarias. Dentro de cada función, se realizó un proceso que contó con 1000 iteraciones.

Para la función de medidas repetidas, se define los siguientes parámetros: el número de observaciones como  $n = n_m * n_s$ , donde  $n_m$  representa el número de mediciones por sujetos y  $n_s$  es el número de sujetos; el intervalo de tiempo para realizar cada medición el cual se establece en 30 minutos; los tipos de dieta (baja en grasa y normal); los interceptos ( $\beta_0$ ) y pendientes ( $\beta_1$ ) de las dietas; los interceptos ( $\beta_{0,i}$ ) y pendientes ( $\beta_{1,i}$ ) de las observaciones; el error residual (e.res.) donde  $e.res. \sim N(0, \sigma^2)$  y por último se generó la variable respuesta ( $y$ ) a partir de los parámetros anteriores. Después de definir los parámetros, se estiman dos modelos: uno con interacción y el otro sin interacción para poder realizar la prueba de no interacción. Luego, para obtener la potencia se contabilizó el número de veces que se rechazó la hipótesis de no interacción, es decir cuando la probabilidad asociada a la prueba mencionada anteriormente fue menor a 0.05.

Ahora, en el diseño de medidas repetidas, se contó con los siguientes parámetros: el número de observaciones como  $n = n_{dist.} * n_d * n_h$ , donde  $n_{dist.}$  es el número de distritos,  $n_d$  es el número de días que a su vez se define como  $n_d = n_{d.entre} + n_{d.fin}$  y  $n_h$  como el número de horarios; los errores de parcela ( $\epsilon_{ijl}^{(1)}$ ) y sub- parcela ( $\epsilon_{ijks}^{(2)}$ ) que siguen una distribución normal con media 0 y cierta variancia; un intercepto que se estableció como la media general  $\mu$  y una variable respuesta ( $y$ ) a partir de los parámetros anteriores. Seguido a esto, se estimó un modelo y se realizó un análisis de variancia (ANOVA, por sus siglas en inglés) de donde se extrajo la probabilidad asociada correspondiente al factor de sub- parcela (horario). En este caso, para obtener la potencia de la prueba se contó las veces en que se rechazó la hipótesis de que los efectos de los diferentes horarios (mañana, tarde y noche) eran iguales.

Para el análisis y la creación de las simulaciones, se usó el lenguaje *R* (*R Core Team, 2020*) versión 4.0.3. Además, se usó la librería *agricolae* (Mendiburu, 2020), *ggplot2* (Wickham, 2016) y *lme4* (Bates et al., 2020).

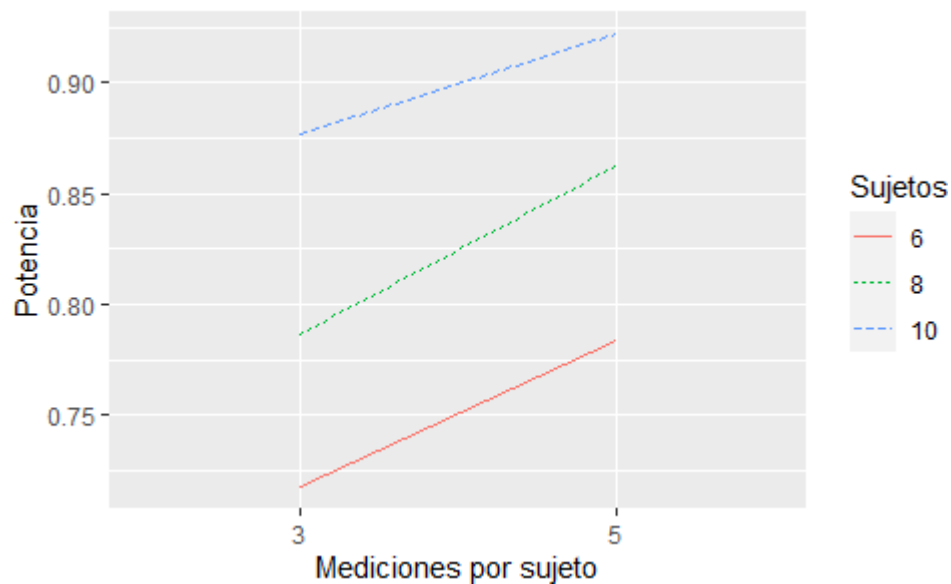
## RESULTADOS

Para el caso del diseño de medidas repetidas, cabe destacar que la potencia inicial de interacción entre el tipo de dieta y el tiempo fue de 0.877. Al ser ya una potencia buena, los escenarios manejados buscaban la manera de reducir el número de sujetos o de mediciones por sujeto, con el propósito de descubrir en cuáles condiciones se obtenía una potencia igual o mejor que la inicial, pero teniendo menos tamaño de muestra.

De este modo, se realizaron las iteraciones con base en el escenario 1. En el caso en que solo se toma en cuenta una reducción de sujetos con las mediciones iniciales (3 por sujeto), la potencia fue de 0.718; mientras que, si se aumentan las mediciones a 5 por sujeto, la potencia sería de 0.784 (ver Figura 1). A su vez, si se plantean estas mismas posibilidades, pero asumiendo un aumento del error, la potencia correspondería a 0.663 y 0.745, con 3 y 5 mediciones respectivamente (ver Figura 5 en anexos).

**Figura 1**

*Potencia de la prueba al reducir el número de sujetos y aumentar el número de mediciones*



Ahora, enfocándonos en el escenario 2, se tiene que, con 8 sujetos, la potencia de la prueba fue de 0.786; mientras que, si a esos 8 sujetos se les hicieran 5 mediciones, entonces la potencia sería de 0.862 (ver Figura 1). En este escenario, también se consideró la posibilidad de un aumento del error, y las potencias fueron de 0.741 (3 mediciones) y 0.815 (5 mediciones) (ver Figura 5 en anexos).

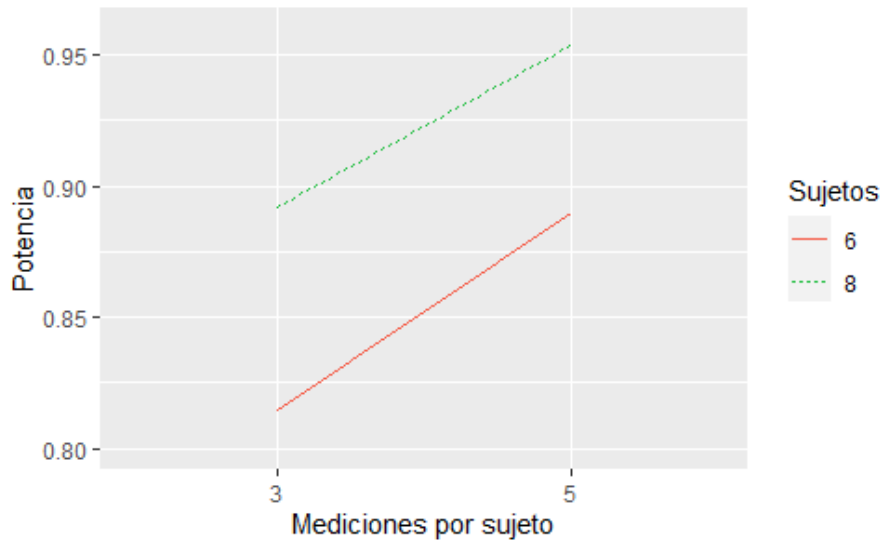
Posteriormente, analizando el escenario 3, manteniendo la misma cantidad de sujetos, pero aumentando el número de mediciones a 5, la potencia de interacción llegaba a 0.922 (ver Figura 1). Si se tomaba en cuenta que, al tener más mediciones por sujeto, se podría dar una reducción del error, por lo que la potencia queda en 0.975 (ver Figura 2).

Más adelante se procedió a comparar el escenario 1 y 2, para ver cuál presentaba condiciones más favorables. Si se parte del hecho de que se está reduciendo el número de sujetos, debería compensarse con realizar más mediciones por sujeto, por lo que se decide centrarse en 5 mediciones por sujeto. En este punto, se presentarían dos situaciones: el error se reduce o aumenta. Para el escenario 1, se tiene que la potencia al disminuir el error es de 0.89 y con un

aumento sería de 0.745, mientras que en el escenario 2 se tiene que las potencias serían 0.954 y 0.815 respectivamente (ver Figura 2).

**Figura 2**

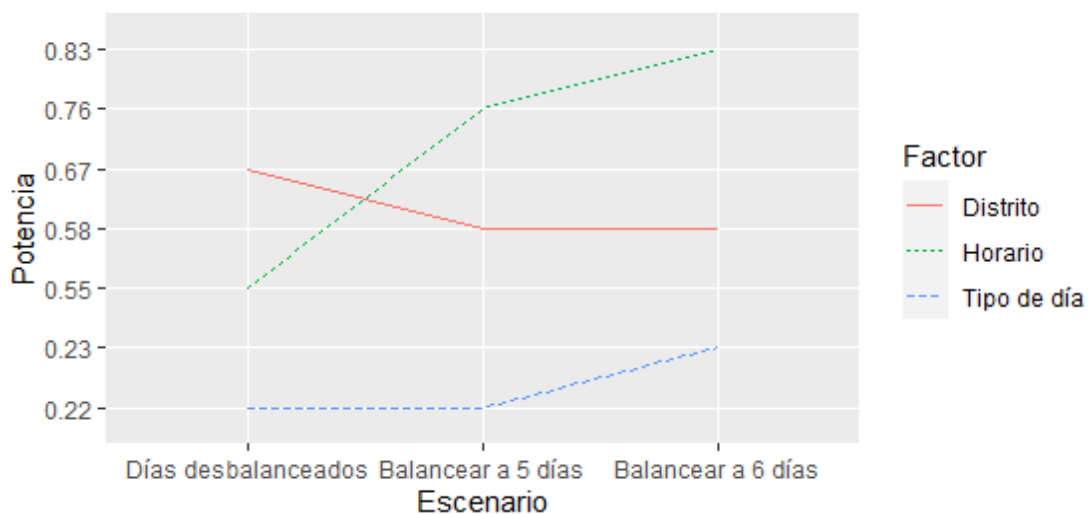
*Potencia de la prueba asumiendo una disminución del error al contar con menos sujetos*



Refiriéndonos al diseño de parcelas divididas, se tienen dos tipos de día: entre y fin de semana; con 5 y 2 repeticiones, respectivamente. Es por esto por lo que se planteó un escenario 1 donde se simula tener 5 repeticiones en cada tipo de día. El factor horario presenta una mejoría en su potencia, ya que, al balancear los días, se pasa de 0.297 a 0.423 (ver Figura 6 en anexos).

**Figura 3**

*Potencia de la prueba al aumentar el número de parcelas en cada tipo de día (entre y fin de semana), dado que hay distritos fuera y dentro de la GAM*



Aquí es importante considerar, que no necesariamente el balancear los días, es lo que está contribuyendo a la potencia del factor horario, sino que al aumentar el número de días se cuenta con más observaciones. Por esto, se decidió incrementar el número de repeticiones a 12 días por cada tipo, y se obtuvo que la potencia de la prueba fue superior a 0.80 (ver Figura 6 en anexos); de manera tal que la potencia de horario se ve afectada principalmente por un aumento en el número de observaciones que se tengan.

En el escenario 2, se decidió pasar de 3 a 6 distritos. Se simuló qué pasaría si los distritos agregados, se localizaran fuera de la GAM, y estos a su vez contarán con un efecto negativo (contrario a los otros tres que son de la GAM) sobre la respuesta promedio. Al realizar las iteraciones correspondientes, se obtuvo que la potencia pasaba de 0.297 a 0.552 (ver Figura 7 en anexos).

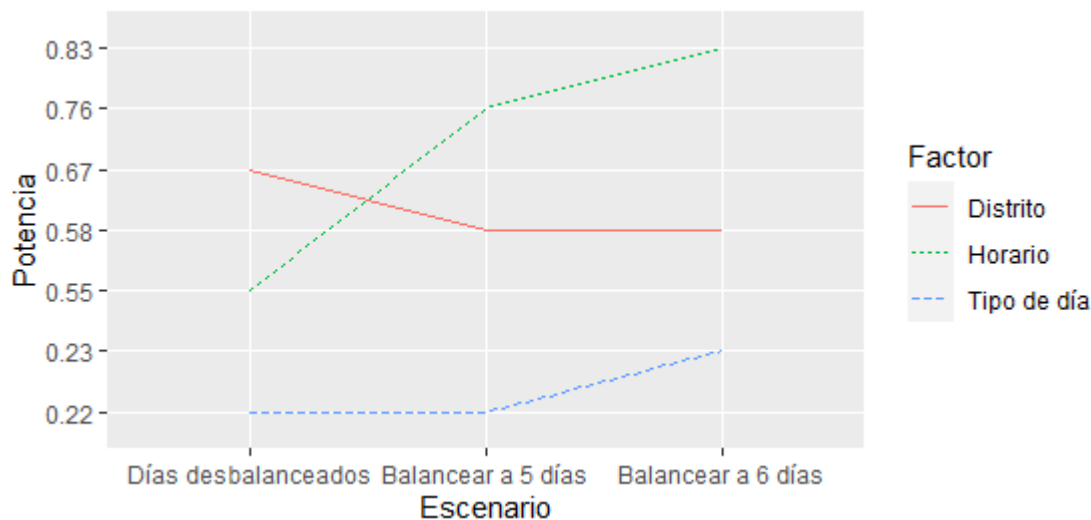
Por otro lado, en el escenario 3, se consideró la posibilidad de una reducción en el error de parcela, es decir, pasar de un error de 0.11 a 0.08. Como era de esperarse, al contar un error más pequeño, la estimación mejoró y se logró una potencia de 0.408.

Posteriormente, se plantearon otros escenarios donde se realizaban varios cambios a la vez, y no solo de uno en uno, para tratar de llegar a un escenario ideal, pero creíble. Se consideró el escenario donde se aumentaba el número de distritos y los días se balanceaban a 5 repeticiones en cada tipo de día. Los resultados arrojaron que la potencia de horario fue de 0.763 (ver Figura 3).

Seguido a esto, se elaboró un escenario, similar al anterior, pero aumentando el número de repeticiones por tipo de día a 6. Con este cambio, se consiguió mejorar, con respecto al escenario anterior, la potencia de horario la cual quedó en 0.834 (ver Figura 3).

**Figura 3**

*Potencia de la prueba al aumentar el número de parcelas en cada tipo de día (entre y fin de semana), dado que hay distritos fuera y dentro de la GAM*

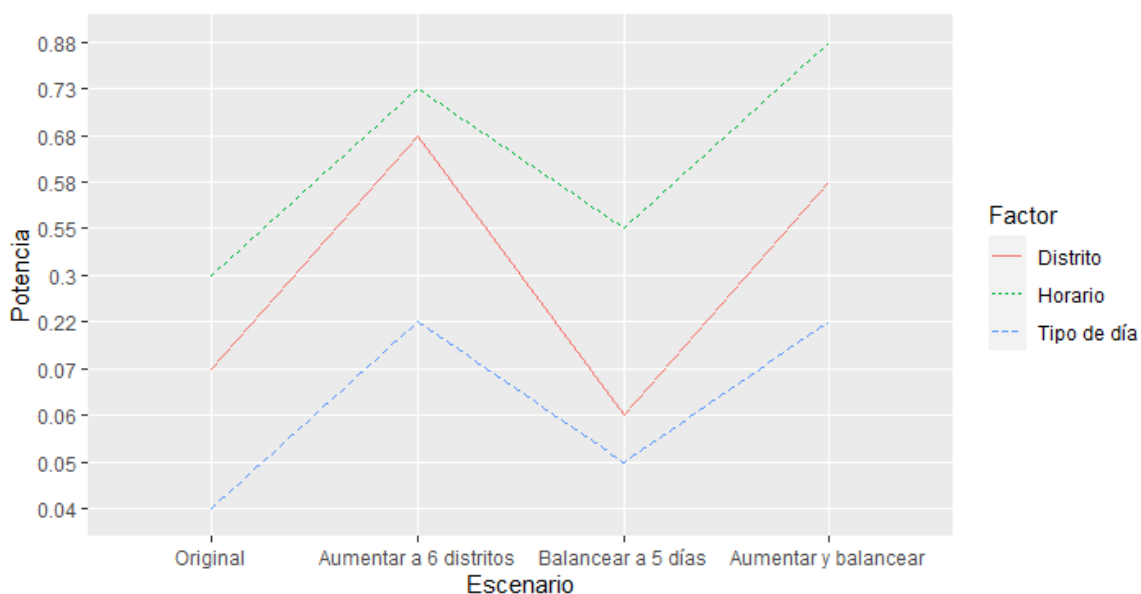




Finalmente, se elaboró un escenario en donde se trabajó con 6 distritos (3 de la GAM y 3 fuera de la GAM), así como con 10 días de la semana en total (5 entre y 5 fin), y con una reducción del error de sub- parcela. En este escenario, al considerar los aspectos antes mencionados se logró obtener una potencia de 0.881 (ver Figura 4), que se puede considerar como una muy buena potencia para el factor de sub- parcela (horario).

**Figura 4**

*Potencia de la prueba al aumentar la cantidad de distritos, balancear el número de días y asumiendo que los cambios suponen una reducción en el error de sub- parcela*



## CONCLUSIONES

En este estudio, nos centramos en dos diseños experimentales: medidas repetidas y parcelas divididas, concretamente se analizó, por medio de simulaciones, el comportamiento de la potencia de la prueba en escenarios específicos.

Para el caso de medidas repetidas, en donde se indagó cómo se veía afectada la potencia de interacción al cambiar el número de sujetos. Dado que las estimaciones originales (sin modificar nada), arrojaban una potencia superior a 0.80; se decidió buscar reducir el tamaño de la muestra. Se obtuvo que al contar con 8 sujetos y aumentar las mediciones por sujeto de 3 a 5, la potencia fue de 0.862.

Para este estudio en específico, se tiene que los resultados de la potencia son bastante estables y, por ende, la mayor modificación que podría sufrir el diseño es contar con menos sujetos. Lo cual para investigaciones en contextos parecidos puede ser de gran utilidad. No olvidando que la medición que se realiza es tomar el pulso, es bastante factible poder realizar más mediciones a los sujetos, que contar con más sujetos que tomen una dieta balanceada. Esto último pensando en que

las personas participantes del estudio deben ser capacitadas para tomar una dieta balanceada y esto representaría una mayor inversión de recursos por parte de los investigadores.

Por otro lado, se tiene el diseño de parcelas divididas, en el que se buscó establecer cuáles variaciones en el diseño del experimento, podían reflejarse en una mejora en la potencia del factor de sub- parcela. Se decidió trabajar con el factor de sub- parcela (horario), debido a que es el principal factor de interés en el estudio. Inicialmente el horario contaba con una potencia de 0.297, la cual es bastante deficiente.

Se tomaron en consideración tres variaciones: aumentar el número de distritos, balancear los días y reducir el error de parcela. Refiriéndonos específicamente al aumento de distritos, se pensó porque el investigador, antes de obtener los datos, estaba interesado en aplicar el experimento en distritos fuera y dentro de la GAM. Al tomar esta situación en consideración, se logra mejorar la potencia (0.552) pero sigue siendo deficiente.

Al contar con el mismo número de días en cada tipo de día, sí se logra mejorar la potencia del factor de sub- parcela, pero se deben tomar como mínimo 5 días en cada tipo de día. Esto nos hace cuestionarnos si realmente si el balancear la cantidad de días ayuda a esta mejoría, ya que, entre más días, mayor tamaño de muestra y esto a su vez, arrojará una mejor potencia.

A modo de cierre, se concluye que la simulación presenta un panorama amplio y podría ser utilizada por el investigador, para en un futuro replicar el estudio y obtener mejores resultados.

## **BIBLIOGRAFÍA**

- Bates, D., Maechler, M., Bolker, B., y Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. <http://10.18637/jss.v067.i01>
- Cárdenas Castro, M., y Arancibia Martini, H. (2014). Potencia estadística y cálculo del tamaño del efecto en G\*Power: complementos a las pruebas de significación estadística y su aplicación en psicología. *Salud & Sociedad*. 5(2), 210-244.
- García-García, J. A., Reding-Bernal, A., y López Alvarenga, J. C. (2013). Cálculo del tamaño de la muestra en investigación en educación médica. *Investigación en Educación Médica*, 2(8), 217-224.
- Kuehl, R.O. (2001). *Diseño de experimentos: Principios estadísticos de diseño y análisis de investigación*. International Thomson Editores S.A.
- Mendiburu, F. (2020). agricolae: Statistical Procedures for Agricultural Research. R package version 1.3-3. <https://CRAN.R-project.org/package=agricolae>
- Montgomery, D.C. (2004). Mediciones repetidas. *Diseño y análisis de experimentos*. (pp. 624-626). Editorial Limusa S.A.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

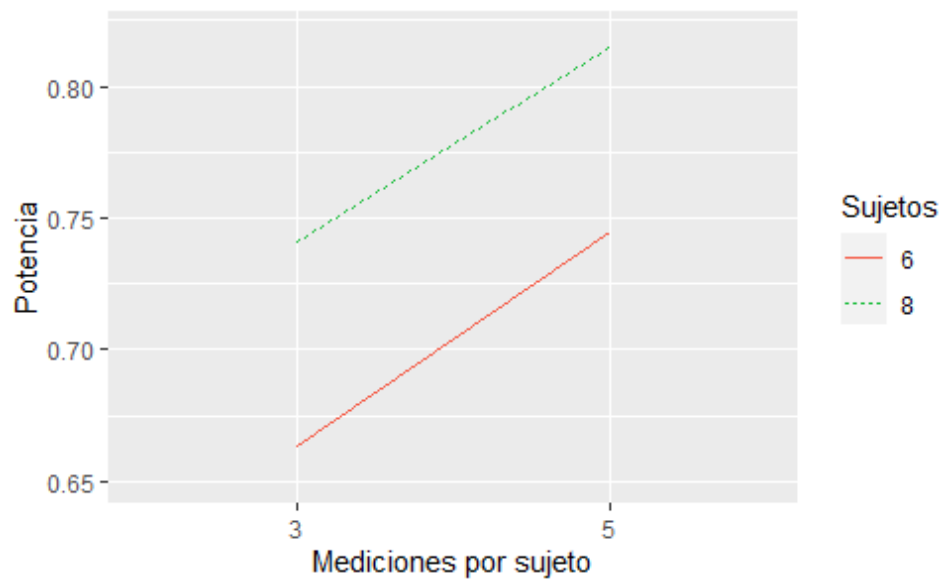
Tapahuasco Zúñiga, E. (2010). *Diseño en parcelas divididas*. Universidad Nacional de San Cristóbal de Huamanga. [http://estadistica.260mb.com/Efrain\\_DPD.pdf](http://estadistica.260mb.com/Efrain_DPD.pdf)

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York

## ANEXOS

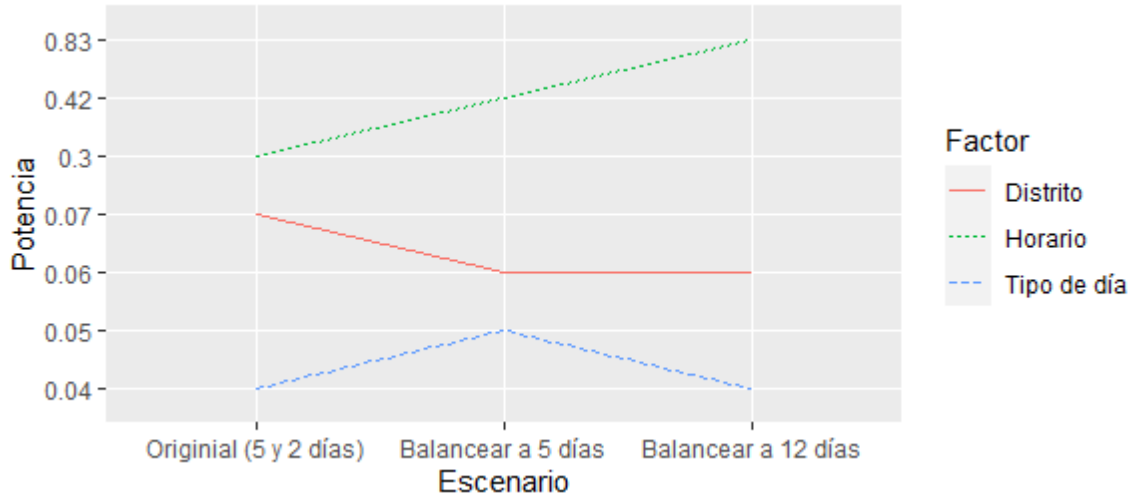
**Figura 5**

*Potencia de la prueba asumiendo un aumento del error al contar con menos sujetos*



**Figura 6**

*Potencia de la prueba al variar el número de parcelas en cada tipo de día (entre y fin de semana)*



**Figura 7**

*Potencia de la prueba al aumentar el número de niveles en el factor de parcela (distritos fuera y dentro de la GAM)*

