

SERENGUETI

Revista de Estadística

Volumen 1
Número 1
Febrero 2019

“Una nueva propuesta para el aprendizaje”



Foto: Ricardo Alvarado B.

Universidad de Costa Rica, Escuela de Estadística
Ericka Valerio, José Oviedo, Juan José Montero, Susana García

Comité editorial

El comité está integrado por estudiantes de bachillerato de la carrera de Estadística de la Universidad de Costa Rica, los cuales participaron en la elaboración del primer volumen de la revista.

El comité cumple la función de invocar a estudiantes activos que deseen participar en la divulgación de los artículos, previamente recomendados por profesores de la carrera. También se verifica que estos artículos cumplan con los lineamientos establecidos para su inclusión en la revista. Así mismo el comité tiene entre sus funciones difundir y promocionar la revista en plataformas o medios de interés para la escuela de estadística y sus estudiantes.

Ericka Valerio Salas, Universidad de Costa Rica, c.e. *erickavs189@gmail.com*

José Oviedo Jiménez, Universidad de Costa Rica, c.e. *froysur@gmail.com*

Juan José Montero Solórzano, Universidad de Costa Rica, c.e. *jmontero639@gmail.com*

Susana García Calvo, Universidad de Costa Rica, c.e. *asgc3093@gmail.com*

DEDICATORIA

A nuestro profesor, Ricardo Alvarado Barrantes, por ser una de las personas con más vocación, porque resalta el amor con que enseña, porque apoya a sus estudiantes para que salgan adelante, porque es justo y su meta final es formar verdaderos profesionales.

Gracias profe una vez más por creer en nosotros, por su apoyo incondicional, por su tiempo, por su guía y gracias por hacernos saber que aún quedan profesores con pasión por enseñar.

Contenido

INTRODUCCIÓN	5
Re-Aprendiendo: Una historia sobre la obtención de conocimiento	5
I. MODELOS LINEALES GENERALIZADOS (MLG)	8
Comparación de los modelos lineal clásico y Poisson desde la perspectiva de la aproximación Poisson a la Normal	9
II. MODELOS DE SUPERFICIES DE RESPUESTA	25
Optimización del pH y la turbidez en una bebida de pitahaya utilizando un modelo de superficies de respuesta	26
Influencia de la temperatura y el tiempo en la eficiencia de cocción de la carne de res	33
III. MODELOS MIXTOS	40
Análisis de la evolución del desempeño en una prueba de velocidad a lo largo de tres meses, dentro de un contexto de entrenamiento futbolístico, mediante un modelo mixto de medidas repetidas	41
Comportamiento de la potencia de la prueba en un diseño anidado de dos etapas en el contexto de capacitación de vendedores en una cadena de almacenes	50
Análisis del efecto de la variabilidad de los elementos de estudio en la potencia mediante la simulación de datos para un diseño de medidas repetidas	58
REFERENCIAS BIBLIOGRÁFICAS	67
AGRADECIMIENTOS	68

INTRODUCCIÓN

El diseño experimental nos permite identificar y cuantificar las causas de un efecto dentro de un estudio experimental, en esta técnica estadística se manejan una o más variables vinculadas a las causas para observar los efectos que tienen sobre una variable de interés en específico. Con base en estos diseños se desarrollaron los artículos que forman parte de la revista, elaborados por estudiantes que trabajaron en conjunto, por medio de una técnica no convencional de aprendizaje, donde se mostró que este no necesariamente se da solo por métodos tradicionales de enseñanza. Esta revista servirá como modelo para poder desarrollar en un futuro propuestas de enseñanza que permitan brindar diferentes formas y técnicas de aprendizaje.

Re-Aprendiendo: Una historia sobre la obtención de conocimiento

Muchísimos de nosotros, personas nacidas en el siglo pasado, aprendimos lo que sabemos actualmente mediante las mismas técnicas pedagógicas repetidas por muchos docentes a través de los años. Nos asustamos profundamente ante los exámenes o evaluaciones, sudamos frente al docente mal encarado que no permitía el desconocimiento en sus clases y llamamos con miles de preguntas por hacer, con tal de no ser ridiculizados frente a todos.

Ahora bien, esa educación de la que hablamos estuvo basada en una pedagogía llamada Tradicional, que surgió hace algunos siglos; *“La pedagogía tradicional comienza a gestarse en el siglo XVIII con el surgimiento de la escuela como institución y alcanza su apogeo con el advenimiento de la pedagogía como ciencia en el siglo XIX...”* (Rodríguez, 2013, p.39) En la pedagogía tradicional, lo que se “aprende” son contenidos, el maestro deposita en el estudiante las supuestas verdades que le han sido transmitidas a él, no importa el contexto del alumno, el docente expone y el alumno guarda silencio, la evaluación tiene que ver únicamente con el resultado y no con el proceso; en fin, el alumno está ahí, en el sistema educativo, solo como receptor de conocimiento. (Rodríguez, 2013, p.39)

En la Escuela Tradicional, *“El maestro es la base y condición del éxito de la educación.”* (Ceballos, 2013, p.1) Según esto, cabe preguntarnos por el alumno; ¿qué papel juega este en la educación?, ¿puede o no desarrollar él proyectos, conocimientos, investigaciones, etc., en el sistema educativo?, ¿qué sucede con las intervenciones del alumno en la consecución de su propio éxito o es que él no tiene la capacidad para propiciarlo? Vemos entonces, un tradicionalismo que no estuvo en función de nosotros, pasamos por las aulas recibiendo información, pero nunca se nos formó como forjadores de nuestro propio éxito.

Ahora bien, con esto no se tiene la intención de mostrar a la Escuela Tradicional como un monstruo del que ninguno de nosotros salió librado o que recibimos una pésima educación, pero sí nos permite reflexionar sobre cómo no estuvimos dentro de un sistema educativo que propiciara nuestra participación ni la creatividad y criticidad.

Si bien el modelo educativo tradicional nos otorgó conocimientos durante nuestros años de estudio en la primaria, secundaria, e incluso en la universidad, es claro que no puede decirse que este haya sido el mejor para hacerlo; más aún, cuando es posible contraponerlo con otros métodos de enseñanza, tales como el propio de la *pedagogía Humanista* o la *Escuela Activa*.

Para explicar de manera más clara, por qué existe en este artículo una historia sobre un proceso de re-aprendizaje, vale señalar qué se entiende por *pedagogía Humanista* y qué por la *Escuela Activa* y cuáles son sus características.

El paradigma humanista de la educación presenta un cuestionamiento frente a la Escuela Tradicional y, así mismo, permite rupturas necesarias. Así, en contraste, la educación humanista se define como “...de tipo indirecto pues en ella el docente permite que los alumnos aprendan mientras impulsa y promueve todas las exploraciones, experiencias y proyectos que éstos preferentemente inicien o decidan emprender a fin de conseguir aprendizajes vivenciales con sentido.” (Rodríguez, 2013, p.41)

Según lo anterior, el humanismo como pedagogía nos muestra distintas características a las del Tradicional. El docente sí permite y suscita en el estudiante el conocimiento, esto porque es posible el intercambio de ideas, la exploración, la investigación, el salirse de la teoría y ser partícipe en la construcción de proyectos. Además, hay un factor sumamente importante a mencionar y es el poder dar sentido a lo que se hace y aprende; sin contextualizar lo hecho, sin permitir al estudiante el cuestionamiento por lo hecho y el para qué, no existe una real incidencia.

De la mano con lo anterior, tenemos a la Escuela Activa. Actualmente, es interesante escuchar centros educativos que se rigen bajo este modelo educativo y, muchas técnicas, basadas en este mismo. Pero, qué es la Escuela Activa y cuáles son sus características. La Escuela Activa “...es un enfoque pedagógico integral que promueve la instrucción personalizada y la creación de vínculos fuertes entre la escuela y la comunidad para asegurar que los niños y niñas aprendan competencias que les sirvan para la vida.” (Mogollón & Solano, 2011, p.4) De nuevo, se rescata aquí la idea de que la educación no debe ser sosa para el estudiante, sino que la misma debe de representarle algo y debe permitir su interés y real aprendizaje.

Según los pedagogos Oscar Mogollón y Marina Solano, con la Escuela Activa se deja de lado la idea del docente como la condición del éxito del estudiante:

El alumno tiene poder de intervención, puede discutir, decidir, evaluar con otros, trabajar en grupos grandes o pequeños, solo o en parejas. En esta propuesta los alumnos creen en ellos mismos, en sus ideas, en sus capacidades, tienen altas expectativas que pueden aprender, se convencen de que serán exitosos en lo que se propongan, son reconocidos como personas diferentes a los adultos, y se les hace sentir únicos e importantes. En este sentido, la Escuela Activa propone una nueva forma de enseñar y aprender, re-definiendo lo que se considera un aprendizaje significativo y pertinente. (2011, p.9)

Con lo anterior, es posible comprender que esta pedagogía modifica lo que hemos vivido en las aulas, la Escuela Activa se muestra como aquello a extender en las instituciones educativas, ya que realmente intenta brindar al estudiante las herramientas para ser él quien construya su propio éxito junto con el docente, no como sujeto pasivo y recipiente de la información que le depositan, sino como un actor consciente de aquello que va conociendo, en un proceso en el que

se encuentra inmerso y en el que colabora con todos, además, el alumno tendrá la posibilidad de decir al docente sus ideas y darle, juntos, sentido según su realidad.

Ahora, retomando lo señalado a lo largo del artículo y, mencionando la idea que su propio título contiene, “re-aprendiendo” fue como avanzamos en este curso, ya que no obtuvimos tanto conocimiento siendo sujetos pasivos, sino que fuimos actores en este proceso. Por ejemplo, tuvimos voz para señalar al docente nuestra opinión, logramos conversar sobre nuestra condición de estudiantes y trabajadores, sobre los horarios y dificultades para llevar lo tradicional e incluso, conversamos sobre lo que las evaluaciones tradicionales nos hacían sentir.

Así, sin necesidad de asistir a las lecciones tradicionales y ser parte de sus evaluaciones, logramos aprender; pero más que eso, re-aprendimos que este proceso de enseñanza-aprendizaje puede ser distinto, puede dejarnos realmente conocimiento sin que sea tortuoso, puede significarnos algo desde nuestra propia realidad. En este curso, una enseñanza alternativa nos ayudó con nuestro trabajo y con las limitaciones de muchos para gestionar su tiempo al asistir a las clases.

Para concluir, se obtuvo conocimiento no por el método tradicional, sino por utilizar ideas distintas, como parte de este re-aprendizaje necesario ya para toda la sociedad. El conocimiento obtenido no fue por las clases presenciales, sino por los trabajos, las investigaciones, consultas al docente al que no se le temió ni se le vio como un sujeto aislado, las consultas entre nosotros los estudiantes, ya que también existió un apoyo desde todas las direcciones. Así, cabe señalar, que hubo una real obtención del conocimiento gracias a los nuevos métodos y técnicas y, gracias al docente que lo permite y que pretende comprender al alumno.



I. MODELOS LINEALES GENERALIZADOS (MLG)

Los modelos lineales generalizados están formados por dos componentes: componente aleatorio (identifica la variable respuesta y la distribución de probabilidad) y el componente sistemático (especifica las variables explicativas utilizadas en la función predictora lineal). Fue introducida por primera vez por Nelder y Wedderburn y está constituida por modelos de regresión lineal de error normal y los modelos de regresión exponencial, logística y de Poisson no lineal, entre otros tipos de modelos, como los modelos log-lineales para datos categóricos (Michael H. Kutner, 2004).



Comparación de los modelos lineal clásico y Poisson desde la perspectiva de la aproximación Poisson a la Normal

José Oviedo Jiménez¹

froysur@gmsil.com

RESUMEN

La aproximación Poisson a la Normal es una de las alternativas existentes cuando se desea aproximar una distribución a otra. Al aplicar dicha aproximación en los modelos, se parte del hecho de que, si se tienen datos de una variable respuesta que corresponde a conteos “grandes”, se puede utilizar como opción un modelo lineal clásico, ya que este sería tan bueno como un modelo Poisson. Se presenta de manera general el tema de la aproximación Poisson a la Normal como base teórica. Un estudio mediante simulaciones procura ver qué tan conveniente es aplicar la aproximación en modelos. Se le da énfasis a la expresión gráfica de los resultados para evaluar si la sugerencia de utilizar el modelo lineal clásico es generalizable o si depende de ciertas condiciones. Al final se llega a un análisis de pros y contras de utilizar un modelo u otro. Entre los hallazgos más relevantes, se encontró que los intervalos de confianza de dos colas del modelo Poisson resultan consistentes con el nivel de confianza teórico. Mientras tanto, ante un problema de heteroscedasticidad no detectado, los intervalos de confianza asociados al modelo lineal clásico incumplen el nivel de confianza teórico.

Palabras clave: Poisson, aproximación, Normal, modelos lineales

ABSTRACT

The Poisson approximation to the Normal distribution is one of many existing alternatives when distribution approximation is needed. When this approximation is applied to a model, given to the fact that a response variable data which corresponds to wider counting are the ones which are available, one could possibly use and apply a classic lineal model, since this one would be as good and recommendable as a Poisson model. The theme of Poisson approximation to the Normal one as a theoretical framework or base, is presented in a general form. A study conducted through simulations, aims to observe how convenient is to execute or apply the approximation within models. The graphic expression of the results is being emphasized to assess and evaluate if the suggestion of using the classic Lineal model is generalizable, or if it depends on certain conditions. In the end, an analysis of using one model or the other is reached with their pros and cons. Among the most relevant findings, it was found that the two-sided confidence intervals for the Poisson model were consistent with the theoretical confidence model. Meanwhile, facing a heteroscedasticity problem non-detected, the confidence intervals associated to a classic lineal model did not present consistency about the theory.

Keywords: Poisson, approximation, Normal, Lineal Models

¹ Estudiante de Estadística de la Universidad de Costa Rica



INTRODUCCIÓN

Existe mucha información relativa a aproximaciones que tienen que ver con las distribuciones Normal, Binomial, Hipergeométrica y Poisson. Se hacen relaciones entre varias de ellas, cada una con su respectivo peso teórico de por medio. Hoy en día además de las justificaciones teóricas, gracias a los recursos computacionales es fácil demostrar y darle sentido a aplicaciones relativas a la teoría existente acerca de las aproximaciones. En el presente artículo toma relevancia la aproximación de la distribución Poisson a la Normal. Dicha relación se puede resumir de la siguiente manera:

$$Poisson(\lambda) \rightarrow Normal(\lambda, \lambda)$$

La expresión anterior significa que la distribución de Poisson con parámetro λ tiende a una distribución Normal con parámetros de media y variancia iguales, o sea λ ambos. Eso se cumple cuando λ es “grande” y el tamaño de muestra tiende a infinito. Se aclara el significado de λ , que se define como un promedio de conteos discreto en un intervalo continuo. Además, la distribución de Poisson es del tipo discreta, mientras que la Normal es continua (Mendenhall & Reinmuth, 1978).

Ahora, desde la perspectiva del Teorema del Límite Central:

$$Y = \frac{X - \lambda}{\sqrt{\lambda}}$$

Sea X una variable aleatoria que se distribuye Poisson (λ). Según el Teorema del Límite Central, la variable aleatoria Y converge a una distribución Normal (0,1). Lo anterior adquiere validez si además el parámetro λ es “grande”. Es posible demostrar gráficamente (mediante simulaciones) cómo un valor de λ mayor a 12 es suficiente para llamarse grande, esto debido a que valores menores de 12 crean una distribución asimétrica, en cambio ya mayores a ese valor, hacen lucir a la distribución como una Normal (λ, λ).

Una aplicación de la teoría de la aproximación Poisson a la Normal es el uso de modelos en el contexto de diseños experimentales. Existe el modelo lineal clásico, el cual tiene variable respuesta de tipo Normal. También, un tipo de modelo lineal generalizado, este con variable respuesta Poisson. El objetivo del presente artículo es comparar mediante simulaciones si el modelo lineal clásico es tan bueno como un modelo Poisson, esto cuando la variable respuesta corresponde a conteos bien distribuidos Poisson con cierto λ como parámetro. Dicha comparación va dirigida esencialmente a intervalos de confianza, influencia de las pruebas de hipótesis y decisiones sobre supuestos.

METODOLOGÍA

El primer punto a resaltar es que se parte de la idea de suponer un diseño de un factor con dos niveles (tratamientos). Así, dentro de las simulaciones programadas se generan números aleatorios provenientes de distribuciones Poisson, esto para la variable respuesta en cada uno de los dos tratamientos. El parámetro utilizado es conocido con la letra griega λ . El método para generar valores respuesta dentro de los

tratamientos, de forma que provengan de distribuciones Poisson diferentes, se basa en los parámetros genéricos que se muestran a continuación:

Tabla N° 1

GRUPOS	PARÁMETRO
Tratamiento 1	λ
Tratamiento 2	$\lambda + \lambda * k$

Como se ilustra en Tabla N°1, siempre el valor del parámetro para el grupo 2 es mayor o igual al correspondiente al grupo 1. El rango de posibles valores utilizados en la simulación para el valor numérico llamado “k” va de 0 a 1. Lo anterior permite hacer interpretaciones simples en términos de porcentajes. Por ejemplo, si $k = 0.5$, entonces se puede decir que el parámetro de la distribución Poisson para la variable respuesta en el grupo 2 es 50 % más grande que el del grupo 1. Lo anterior es solo un ejemplo para ilustrar la importancia que adquiere en esta simulación el valor llamado k. Este valor numérico define poblacionalmente qué tan diferentes son las medias entre tratamientos y permite expresar dicha magnitud porcentualmente. Pero no solo expresa porcentualmente qué tan diferente es la media del grupo 2 comparada con la del grupo 1, de hecho, cumple la misma función para el caso de las variancias. Ese detalle es debido a la particularidad conocida de la distribución Poisson, de que su valor esperado y variancia coinciden. Se aclara todo esto porque ese control que se puede tener sobre qué tan diferentes son las variancias entre tratamientos, más adelante tomará mucha importancia al analizar lo que pasa en el caso lineal clásico.

Dentro de cada una de las mil simulaciones programadas, se ajusta de manera simultánea un modelo Poisson y uno lineal clásico, ambos con los mismos datos generados para la respuesta (eso es esencial para compararlos). Para el caso Poisson, se hace una prueba de razón de verosimilitudes (LRT) para evaluar si existe suficiente evidencia estadística para pensar que las medias dentro de cada tratamiento difieren entre sí. Luego, independientemente del resultado de la prueba se calcula un intervalo de confianza de dos colas para el cociente resultante entre las medias, siempre posicionando la media del grupo 2 en el numerador. La intención es calcular la proporción de veces (de un total de 1000), en las que el valor poblacional está correctamente estimado por el intervalo de confianza. Por razones propias de este diseño del estudio de simulaciones, se puede generalizar que el valor poblacional siempre será:

$$\frac{\lambda + \lambda * k}{\lambda} = \frac{\lambda * (1 + k)}{\lambda} = 1 + k$$

Para entender mejor, a manera de ejemplo supóngase una “realidad artificialmente creada”, con un valor de k igual a 1. Dado ese valor, es posible interpretar que poblacionalmente la media dentro del grupo 2 es dos veces la del grupo 1.

El interés se centra en la proporción de veces donde se cumple la siguiente expresión lógica para cada intervalo bilateral de 95 % de confianza:

$$\text{Límite inferior} \leq 1 + k \leq \text{Límite Superior}$$

Ahora, en materia correspondiente al modelo lineal clásico, cabe recordar que en casos de la vida real se deben realizar diagnósticos para evaluar supuestos básicos sobre normalidad dentro de cada grupo, así como para poner a prueba la igualdad de variancias entre grupos. En este estudio en particular, se le da una posición

protagónica a la evaluación de la homoscedasticidad. En cuanto a la normalidad, está garantizada de antemano, los números aleatorios se generan con parámetros λ lo suficientemente grandes (mayores a 1000). Eso quiere decir que posibles problemas de no normalidad estarían asociados a un tema de tamaño de muestra, no a un valor de λ pequeño. Se aclara que valores de λ menores a 12 se consideran “pequeños”, pero no son de interés en este estudio.

Una vez ajustado el modelo, se hace la prueba Levene de igualdad de variancias, también la prueba F del ANOVA para igualdad de medias. Sea cual sea el resultado de las pruebas anteriores, se procede a la construcción del intervalo de confianza para la diferencia de medias y se ignora cualquier incumplimiento del supuesto de homoscedasticidad. La razón es porque se desean observar casos en los que la muestra obtenida engaña en alta proporción a la prueba de Levene; también casos en los que se ignora a propósito un evidente problema de heteroscedasticidad. Se quiere observar el comportamiento de los intervalos y de la prueba de Levene cuando se sabe de antemano que sí existe cierto porcentaje de heterogeneidad de variancias.

Como el intervalo de confianza para la diferencia de medias de este modelo no es comparable con el cociente de medias del modelo Poisson, se le hace una transformación al intervalo de la diferencia de medias obtenido. Los pasos para la transformación son los siguientes:

1. Primero se calcula una equivalencia poblacional para el valor k . Hay que recordar que k es en esta simulación un valor numérico que se conoce y que se controla. Mientras que en la vida real es un valor poblacional desconocido que se intenta estimar.

$$DIFERENCIA\ POBLACIONAL = (\lambda + \lambda * k) - \lambda = \lambda * k$$

$$k = \frac{DIFERENCIA\ POBLACIONAL}{\lambda}$$

2. Ahora, una analogía básica entre población y muestra permite deducir a partir de ese valor poblacional anterior, dos valores muestrales que serían los extremos del nuevo intervalo transformado. Ambos valores muestrales llevan en el denominador al valor muestral “media del grupo 1”, debido a que de forma análoga λ es definido en este estudio como el valor poblacional del grupo 1 (Ver Tabla N° 1). Los numeradores son valores muestrales, corresponden a los límites obtenidos de cada intervalo de confianza que estima la **DIFERENCIA POBLACIONAL**. Se aclara que la **DIFERENCIA POBLACIONAL** es un valor conocido y controlado en esta simulación.

$$\frac{\text{Límite inferior de la diferencia de medias}}{\text{media del grupo 1}}$$

$$\frac{\text{Límite superior de la diferencia de medias}}{\text{media del grupo 1}}$$

3. Igual que como en el caso para Poisson, interesa evaluar qué proporción de veces se cumple la siguiente expresión lógica para cada intervalo de 95 % de confianza:

$$\frac{\text{Límite inferior Dif. de medias}}{\text{media del grupo 1}} \leq k \leq \frac{\text{Límite superior de Dif. de medias}}{\text{media del grupo 1}}$$

$$1 + \frac{\text{Límite inferior Dif. de medias}}{\text{media del grupo 1}} \leq 1 + k \leq 1 + \frac{\text{Límite superior Dif. de medias}}{\text{media del grupo 1}}$$

El último paso de agregar un 1 a toda la desigualdad no afecta el cálculo de la proporción, pero para efectos de análisis gráfico es necesario para convertirse en un intervalo totalmente comparable con el intervalo Poisson.

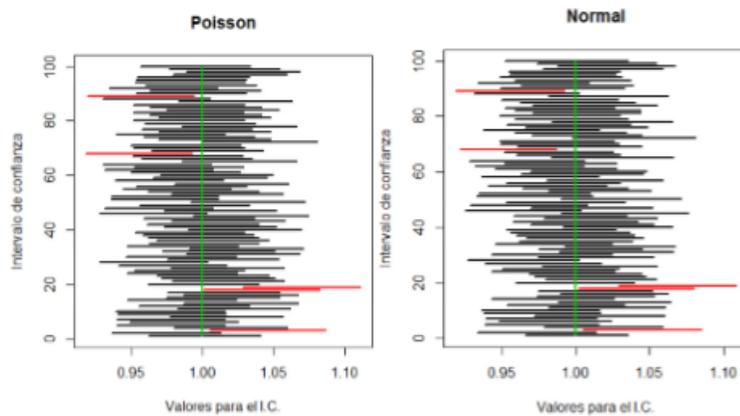
Para enriquecer el análisis, se hace un estudio comparativo con gráficos. Incluye las amplitudes de los intervalos, esto variando el tamaño de muestra y el porcentaje de heterogeneidad de las variancias entre grupos. Además, se utiliza la misma metodología para comparar las respectivas pruebas de igualdad de medias de cada modelo. Todo el proceso se hace mediante el paquete estadístico R, versión 3.5.2 (R Core Team, 2018).

RESULTADOS

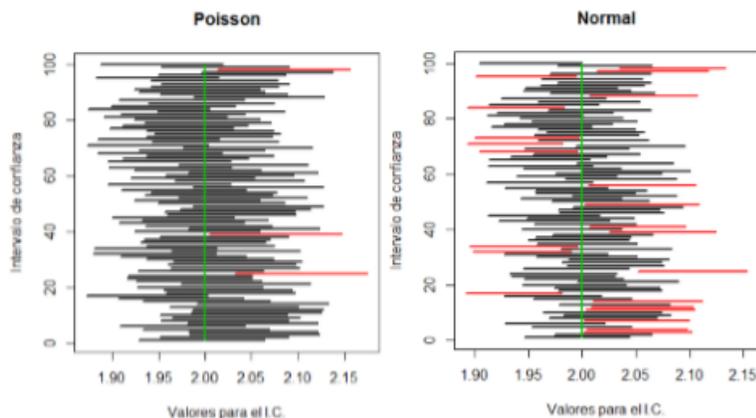
Esta sección brinda un análisis gráfico centrado en observar el comportamiento de los intervalos de confianza de dos colas. Aquí es fácil evaluar en los intervalos dos características esenciales: amplitud y capacidad para contener al valor poblacional. El análisis se cruza con factores que podrían influir directamente en los resultados. El tamaño de muestra y el porcentaje de desigualdad de medias son esos elementos que enriquecen el análisis. Pero al que se le debe prestar mayor atención es a la desigualdad de medias pues dicho indicador coincide con el porcentaje de desigualdad de variancias entre tratamientos.

Comparación de intervalos de 95% de confianza tomando en cuenta variancias entre tratamientos

Igualdad de variancias entre tratamientos



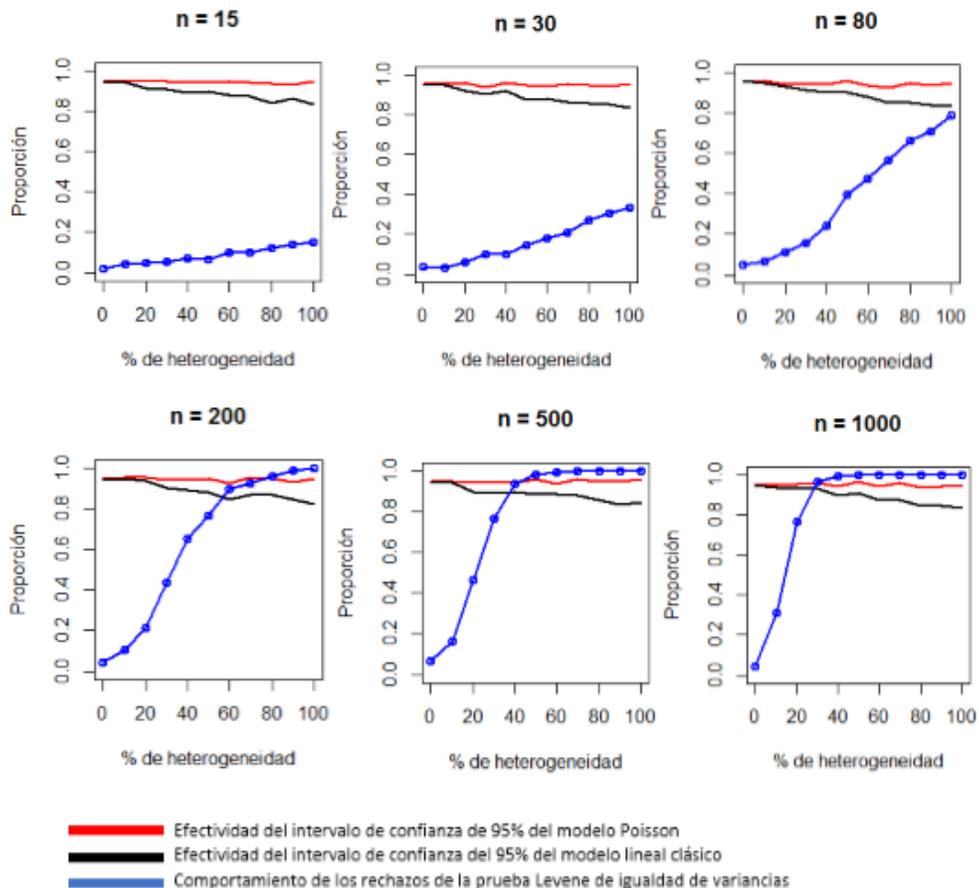
Heteroscedasticidad alta



Aunque los cálculos en este estudio se hicieron con base en mil muestras, para fines de interpretación visual en el anterior gráfico se presenta con base en cien, siempre teniendo el cuidado de mantener consistencia de acuerdo a la base mil. El valor poblacional es representado por la línea vertical de color verde. Cada recta horizontal corresponde a un intervalo de confianza de 95%. El eje x aquí corresponde a los posibles valores dentro de cada intervalo.

En el gráfico se observa que, para el escenario de igualdad de variancias entre tratamientos, el caso Poisson y lineal clásico son casi idénticos, tanto en amplitud como en contener o no al parámetro. Se cumple satisfactoriamente en ambos la característica del 95% de confianza. Obsérvese cómo en ambos casos hay cinco intervalos color rojo de los cien ilustrados. Precisamente esas líneas en color rojo son representaciones de intervalos de confianza que no contienen al valor poblacional. Por otra parte, el escenario donde hay heteroscedasticidad alta brinda un panorama muy distinto. El caso Poisson cumple con la confianza del 95% pues se aprecian cuatro líneas rojas de las cien posibles, en cambio el caso lineal clásico presenta aproximadamente veinte líneas rojas. Además, las rectas son más amplias en el caso Poisson. Quizás surja de entrada la interrogante de si aumentar el tamaño de muestra podría mejorar el comportamiento de los intervalos en el caso Normal. Es una duda razonable, el siguiente gráfico evalúa el efecto de dicho aumento.

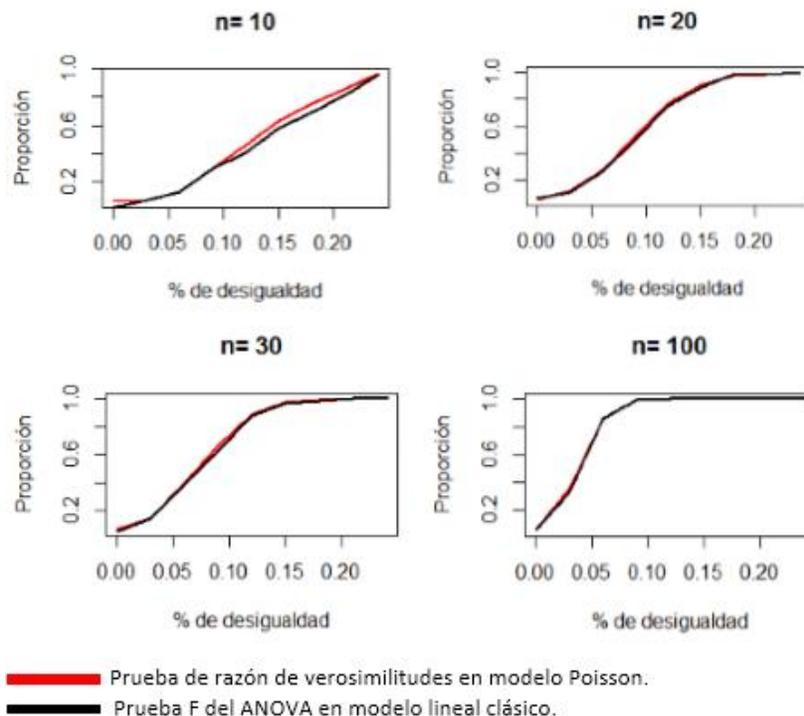
Impacto de la heterogeneidad de variancias en la efectividad de los intervalos de confianza y en la prueba de Levene.



El gráfico anterior evidencia que los aumentos en el tamaño de muestra no mejoran la capacidad que pueda tener el intervalo para contener al parámetro. El caso Poisson (en rojo) y lineal clásico (en negro) no cambian su tendencia al aumentar el tamaño de muestra. Obsérvese cómo la línea roja mantiene su tendencia constante sobre el valor de proporción igual a 0.95, esto sin importar aumentos de muestra ni aumentos en el porcentaje de heterogeneidad de variancias entre tratamientos. En cambio, dichos porcentajes de heteroscedasticidad altos (mayores a 50%) inciden negativamente en el caso Lineal clásico. La proporción de efectividad para contener al parámetro es aproximadamente de 0.8.

La línea azul representa la prueba Levene de igualdad de variancias. En presencia de una heteroscedasticidad alta y tamaño de muestra menor a 80, la prueba se ve severamente engañada. Se rechaza en muy baja proporción cuando en realidad sí debería rechazarse con contundencia. Eso implica que, en un escenario similar, si se recurre a un modelo lineal clásico y se hace el diagnóstico de la homoscedasticidad, es probable que no detecte algún problema. Lo que llevaría a un cálculo del intervalo que se supone debería ser confiable en un 95%, pero la realidad es que sería de menos confianza que la teóricamente definida.

Proporción de rechazo en las pruebas de igualdad de medias en los modelos Poisson y Lineal clásico.

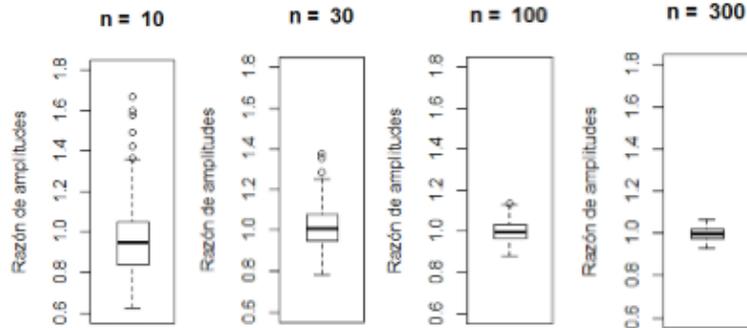


El gráfico ilustra la proporción de veces en que se rechaza cada prueba según su respectivo modelo. Cada eje x representa el porcentaje de desigualdad que se programó para la población deseada, mientras que el eje y es la proporción de veces en que las muestras de cierto tamaño conducen al rechazo de la hipótesis de que los promedios dentro de cada tratamiento son iguales. Se puede observar que las pruebas de ambos modelos se comportan muy similar tanto al aumentar el tamaño de muestra, como al aumentar el porcentaje de

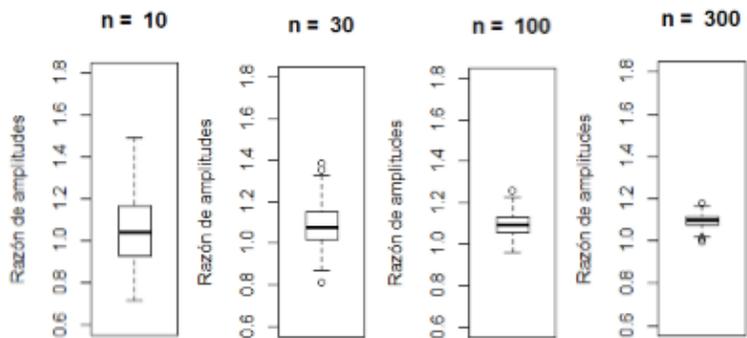
desigualdad entre las medias. Un valor de desigualdad entre medias de 20 % parece ser lo suficientemente contundente como para conducir a un alto porcentaje de rechazo de ambas pruebas, aún en tamaños de muestra pequeños. Lo anterior hace pensar que indiferente al modelo que se utilice, cualquiera de los dos cumple de manera satisfactoria la evaluación de si los promedios dentro de cada nivel del factor de estudio son iguales.

Razón de amplitudes de intervalos: I.C. Poisson / I.C. Normal

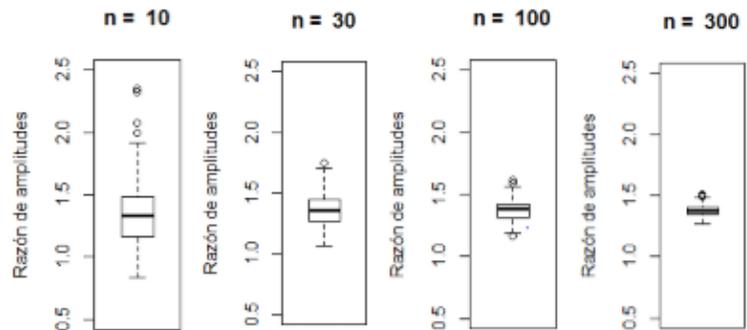
Variaciones iguales



20% de heterogeneidad de variancias



90% de heterogeneidad de variancias



El gráfico anterior toma el valor de la amplitud de cada par de intervalos (Poisson y Normal) y hace un cociente de amplitudes. Se evidencia que en un escenario de igualdad de variancias, las amplitudes de ambos son prácticamente iguales. Ya en escenarios con heterogeneidad de variancias e independientemente del

tamaño de muestra, siempre el intervalo Poisson es más ancho. En un caso de heteroscedasticidad muy alta (90%) se ve cómo el intervalo Poisson es aproximadamente 35% más amplio que el del lineal clásico.

CONCLUSIONES

La teoría de la aproximación Poisson a la Normal es sencilla y fácil de demostrar. Al llevar dicha base teórica a los modelos Poisson y lineal clásico, es mejor tener cuidado. El estudio de simulaciones a gran rasgo demuestra que, en este tema, la variancia es centro de la polémica. En situaciones en las que se tienen conteos grandes dentro de cada tratamiento, antes de aventurarse a pensar que un modelo lineal clásico es tan bueno como un Poisson, es mejor tener en cuenta algunas consideraciones:

- a) Si lo que se desea es solo saber si la media de un tratamiento es estadísticamente diferente a la de otro tratamiento, se podría decir que ambos modelos son igualmente eficientes. Las pruebas de igualdad de medias correspondientes a cada modelo brindan resultados casi iguales.
- b) Si se desea un intervalo para cuantificar qué tan diferentes son las medias, el modelo Poisson cumple con el nivel de confianza que garantiza. Mientras que el lineal clásico en igualdad de condiciones, promete un nivel de confianza y en la realidad da uno menor. Se vio por ejemplo cómo en un escenario de alta heteroscedasticidad, el intervalo de un modelo lineal clásico garantizaba un 95% y en el peor de los casos no bajó de 80%. Una situación así, dependiendo de la percepción del investigador, quizás no sería algo tan negativo y podría atreverse a tomar el riesgo.
- c) Un problema grave de heteroscedasticidad no detectado por las pruebas puede llevar a estimaciones alejadas de la realidad cuando se utiliza un modelo lineal clásico. En este sentido, vía revisión bibliográfica u opinión de experto (aunque a veces se vuelva difícil o imposible encontrarlos), quizás se pueda conjeturar acerca de la variabilidad de la población y así tratar de contemplar una “equivocación” de la prueba de hipótesis.
- d) El modelo Poisson brinda intervalos de confianza más amplios que el modelo lineal clásico. Algunos lo podrían ver como algo no muy bueno, pero esto sucede para poder garantizar la confianza deseada.

BIBLIOGRAFÍA

Mendenhall, W., & Reinmuth, J. (1978). *Estadística para administración y economía*. Editorial Iberoamericana.

Martínez, Mónica. *La distribución Poisson*. Universidad Politécnica de Valencia.

Morettini, Mariano. (2013). *Aproximaciones de distribuciones de probabilidad*. Portal de promoción y difusión pública del conocimiento académico y científico.

Reyes, Hortencia; Almendra, Félix; Morales, Marco; Tajonar, Francisco. (2015). *Problemas al usar la aproximación normal en intervalos de confianza suponiendo datos Bernoulli*.

Wackerly, D; Mendenhall, W; Scheaffer, R. (2010). *Estadística matemática con aplicaciones*. Editorial Cengage learning.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Comparación de tres modelos de conteos desde la perspectiva de la potencia de las pruebas para detectar diferencias entre las pruebas en distintos escenarios de dispersión

Juan José Montero Solórzano²

jmontero639@gmail.com

RESUMEN

Para el análisis de datos de conteo se utiliza generalmente el modelo de Poisson, donde se asume que la media y la varianza son iguales, en algunos casos no se cumple este supuesto por lo que se recurre a otras alternativas para poder analizar los datos con los que se trabaja.

Se desea modelar una situación en la que hay conteos (como variable respuesta) de dos tratamientos, por lo que se trabaja con un solo factor de diseño. La simulación está orientada a comparar el resultado de pruebas para detectar diferencias entre tratamientos. Cada prueba depende del modelo que se utiliza, por lo que se trabaja con tres modelos: Poisson, Binomial negativa y Quasi Poisson. Para todas las pruebas se establece un nivel de significancia de 0.05. La prueba de razón de verosimilitudes es utilizada para los casos Poisson y Binomial negativa. En cuanto al caso Quasi Poisson se utiliza la prueba F.

En los escenarios con subdispersión no se encuentran diferencias entre los tres modelos, con un tamaño de muestra mayor a 15 los modelos se comportan de la misma forma, desde la perspectiva de la potencia de las pruebas. Para los escenarios de equidispersión con un tamaño de muestra menor que 20 la prueba para la distribución Poisson siempre es mayor.

Palabras clave: Poisson, Binomial Negativa, Quasi Poisson, Prueba F

ABSTRACT

For the analysis of data with counting variables, the Poisson model is generally used, where it is assumed that the mean and the variance are the same, in some cases it is not possible to count with this assumption being true; and for that reason, some other alternatives emerge and are taken for making possible the analysis of data which is under work.

The aim is to modelate a situation in which there are countings (as a response variable) within two different treatments, so that the entire work remains with just one single design factor. The simulation is oriented for comparing the result of the tests to detect differences among treatments. Each test depends on the model that is being utilized, for that reason there are three models to work with: Poisson, negative Binomial and Quasi Poisson. For all tests, the significance level of 0.05 is established. The likelihood ratio test is applied for the Poisson and Negative Binominal cases. In regards of the Quasi Poisson, an F-test is utilized.

In the cases which there are sub-dispersion, differences between the three models were not found. With a sample size bigger than 15, the models tend to present the same behavior, from a power test perspective.

² Estudiante de estadística de la Universidad de Costa Rica



Poisson distribution which equi-dispersion scenarios were established, turned up to show a bigger probability associated to the test with a sample size lower than 20.

Key words: Poisson, Negative Binomial, Quasi Poisson, F-test

INTRODUCCIÓN

Para analizar datos de conteo se utiliza generalmente el modelo de Poisson, en el cual se asume que la media y la varianza son iguales, se denomina equidispersión, ($E(Y|X) = \text{Var}(Y|X) = \lambda$) se recurre a otras alternativas de análisis cuando no contamos con este supuesto. Al trabajar con sobredispersión en la distribución de Poisson vamos a tener el problema de que se afectará el error tipo I. Al realizar una prueba estadística se puede correr el riesgo de rechazar la hipótesis nula. La probabilidad de rechazar la hipótesis nula cuando en realidad es verdadera (error de tipo I). (Pita Fernández S., 2001)

El objetivo de la investigación es comparar el uso práctico de los tres modelos para conteos desde la perspectiva de la potencia de las pruebas, esperando detectar diferencias entre estas en distintos escenarios de dispersión. Cuando se presentan estos casos surgen dudas de cual modelo es el que responde mejor a cada uno de los escenarios con los que se está trabajando. Por lo que se busca determinar en cual caso es mejor cada uno de los modelos. Los modelos que se van a comparar son Poisson, Quasi Poisson y Binomial Negativa.

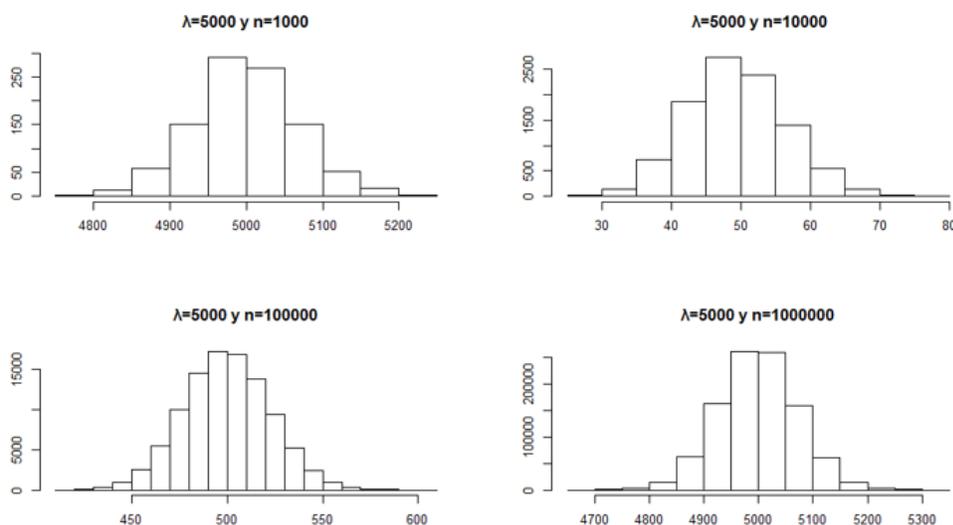
El modelo de Poisson es el modelo base para el análisis de datos que se forman mediante conteos de una variable de interés, en algún intervalo determinado. Cuando se presentan problemas de dispersión en la distribución esta se vuelve deficiente, ya que no está cumpliendo su supuesto principal. (Colorado, 2017)

METODOLOGÍA

La distribución de Poisson (λ) es una de las distribuciones más importantes de variables discretas, se utiliza en situaciones donde nos interesa determinar conteos de una variable en específico. Esta distribución tiende a la distribución normal (λ, λ), cuando el parámetro λ es grande y la muestra tiende a infinito.

Se puede observar gráficamente lo mencionado anteriormente con los siguientes gráficos:

Figura N° 1. Distribución de Poisson (5000) aproximándose a una distribución Normal (5000, 5000)



Consistente con lo visto en la figura anterior, la teoría indica que con un λ “grande” y un tamaño de n que tienda a infinito, la distribución Poisson toma la forma de una distribución normal. Esta base teórica es de gran utilidad en dicho estudio de simulaciones.

El presente estudio de simulaciones se encuentra dentro del contexto de un diseño experimental. Se desea modelar una situación en la que hay conteos (como variable respuesta) dentro de dos tratamientos, por lo que se trabaja con un solo factor de diseño. La simulación está orientada a comparar el resultado de pruebas para detectar diferencias entre tratamientos. Cada prueba depende del modelo que se utiliza, por lo que se trabaja con tres modelos: Poisson, Binomial negativa y Quasi Poisson. Para todas las pruebas se establece un nivel de significancia de 0.05. La prueba de razón de verosimilitudes es utilizada para los casos Poisson y Binomial negativa. En cuanto al caso Quasi-Poisson se utiliza la prueba F.

Para el desarrollo de la simulación se programa una función donde se analiza la potencia de la prueba, esto para ver si existen diferencias entre las medias de los dos tratamientos. Se compara dicha prueba al utilizar los tres modelos mencionados anteriormente. Esta función devuelve como resultado la cantidad de veces en las que una prueba específica se rechaza, esto según escenarios particulares. Se simula 1000 veces dentro de cada escenario. El primer escenario se programa de manera que exista subdispersión, el segundo con equidispersión y el tercero con sobredispersión. En los tres anteriores existe diferencia entre las medias de los tratamientos. Un cuarto escenario contempla de nuevo sobredispersión pero hay igualdad de medias.

La función “potencia” programada, cada vez que se ejecuta en un escenario específico devuelve un vector fila con cuatro entradas. Este vector va a contar con los siguientes elementos: n , potencia en Poisson, potencia en Binomial Negativa, potencia en Quasi Poisson. Cada una de esas potencias es el resultado de la suma de la cantidad de rechazos y luego esa suma dividida entre mil. Se generan los vectores y se agrupan en una matriz que tiene las siguientes dimensiones: filas (cantidad de valores en el vector “ n ” dado), columnas (4 para que coincida con la dimensión de la salida de la función “potencia”).

Para poder tener escenarios en función a la dispersión, se aprovecha la aproximación a la normal. Así, los números aleatorios para cada variable respuesta provienen de distribuciones normales con parámetros de media mayores a 3000. Para que esos números luzcan como conteos se redondean a enteros positivos. El parámetro de variancia se controla a conveniencia, esto es clave para poder obtener los escenarios que dependen de la dispersión. Cabe resaltar que cada vez que se generan números aleatorios, esos mismos números se analizan con cada uno de los tres modelos, eso es lo que garantiza la correcta comparabilidad. El modelo utilizado para el análisis en las tres pruebas es el siguiente:

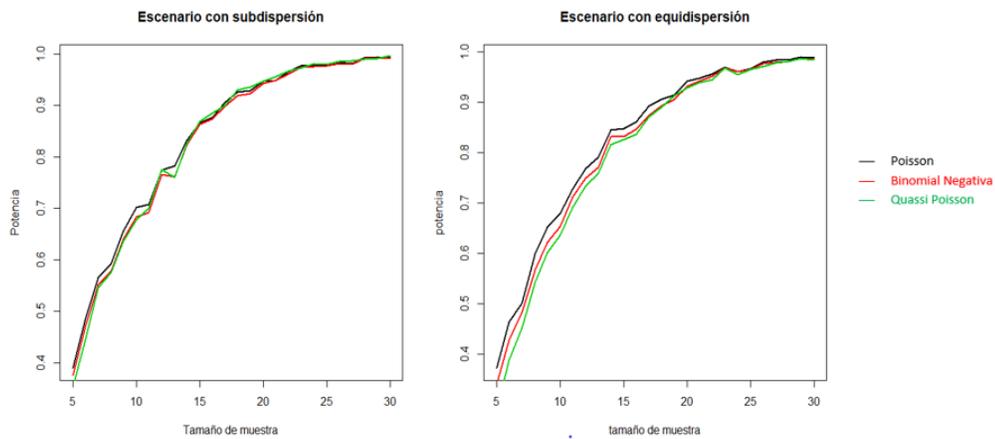
$$\log(\mu_i) = \beta_0 + \epsilon_i$$

RESULTADOS

A continuación, se muestran los principales resultados al correr la función programada. Los datos fueron analizados con el software estadístico RStudio versión 3.4.2 (RStudio Team, 2016), utilizando como principales paquetes los siguientes: car (John Fox and Sanford Weisberg, 2011), MASS (Venables, W. N. & Ripley, B. D., 2002) y AER (Christian Kleiber and Achim Zeileis, 2008). Básicamente se ilustran las curvas de potencia al ir aumentando el tamaño de muestra, todo siempre dentro de algún escenario particular. La intención principal es comparar

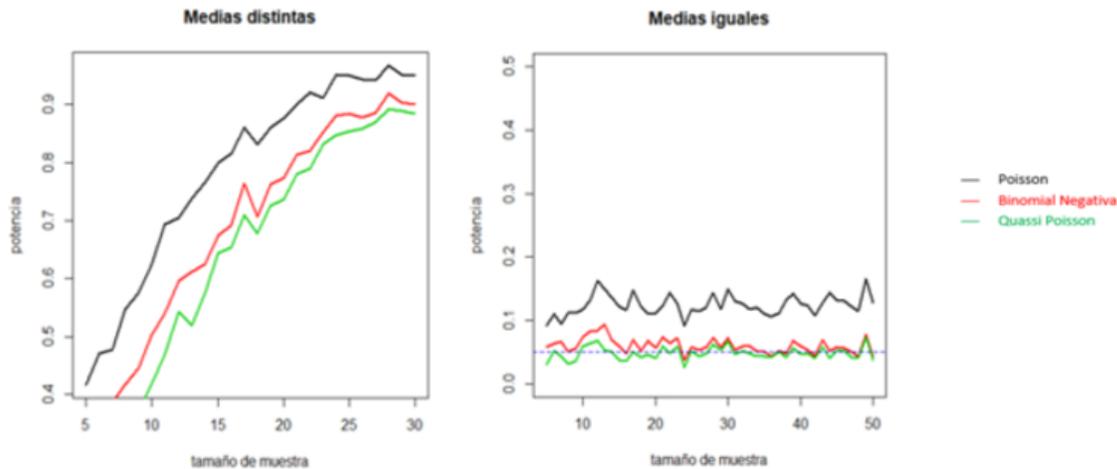
dichas curvas de potencia, esto para determinar si el uso de un modelo es más apropiado que otro u otros. Los resultados están resumidos en tres figuras. La primera agrupa dos escenarios que muestran resultados muy similares. Por otro lado, la segunda figura permite el análisis de una situación digna de profundizar un poco más. La tercera es solo para descartar que exista una situación similar a la presentada en la segunda figura.

Figura N° 2. Comparación de potencias de cada prueba para observar diferencias de medias
(Ambos escenarios con medias distintas)



En la Figura N° 2 lado izquierdo, se representa el escenario en el cual hay presencia de subdispersión. Se puede observar que las tres curvas están muy cercanas y siguen la misma tendencia. Esto quiere decir que, en dicho escenario, realmente no hay diferencia al utilizar cualquiera de los tres modelos propuestos. De hecho, a partir de un tamaño de muestra de 15, los tres modelos son exactamente iguales (desde la perspectiva de potencia de las pruebas). En el lado derecho se encuentra el escenario correspondiente a equidispersión, en él se observa que prácticamente para cualquier tamaño de muestra menor a 20, siempre la potencia es mayor en el caso de la prueba relacionada al modelo Poisson. Mientras tanto, las pruebas de los otros dos modelos son levemente menos potentes que en el caso Poisson. Ya después de un tamaño de muestra de 25, parece que las pruebas de los tres modelos son igualmente potentes. Hay que recordar que, en el caso de estos escenarios dentro de esta figura, se garantiza de antemano que las medias de cada tratamiento son realmente diferentes. Por esto, teóricamente cada prueba debería rechazarse siempre. Dicho lo anterior, aquí entra en juego la importancia de tener cierto tamaño de muestra mínimo que produzca una potencia alta. Es decir, que se rechace cuando efectivamente tenga que rechazarse alguna prueba.

**Figura N°3. Comparación de potencia de cada prueba para observar diferencias en las medias
(Escenario con sobredispersión)**



La figura N° 3 resume el caso del escenario en el cual existe sobredispersión. En primera instancia se corrió solo la función para el escenario que representa sobredispersión cuando hay medias diferentes (lado izquierdo de figura N°3). Aquí, llama la atención que la curva de potencia del caso de la prueba correspondiente al caso Poisson está realmente muy por encima de las otras dos. Es evidente que esto conduciría a pensar que en un escenario de sobredispersión es mejor el modelo Poisson que cualquiera de los otros dos. Lo curioso es que la literatura dice todo lo contrario. La teoría expone que un escenario con sobredispersión es el más problemático de todos los posibles, y que utilizar un modelo Poisson no es una buena opción.

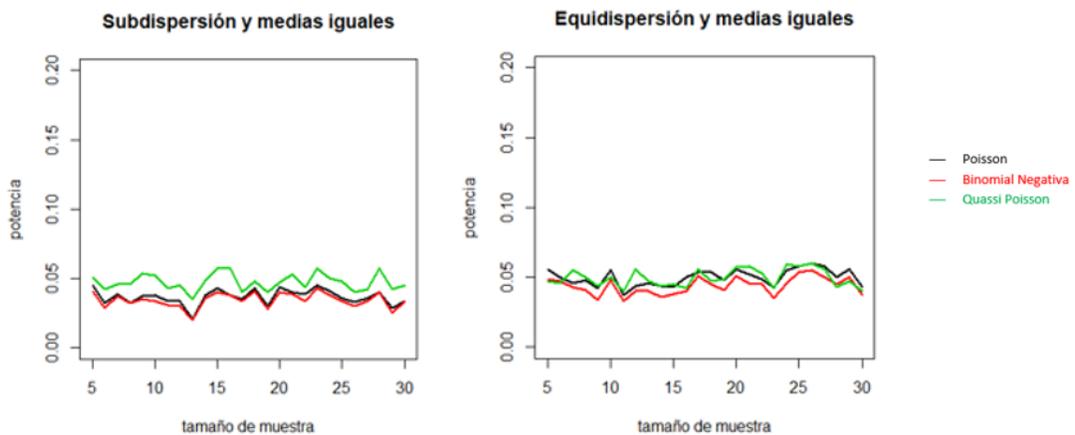
Para poder profundizar más acerca de lo que sucede en este escenario “problemático”, se creó en segunda instancia un escenario adicional que además de estar programado con sobredispersión, también lo está para que las medias sean iguales. Dicho escenario adicional está resumido en la parte derecha de la figura N°3. En este escenario lo correcto es rechazar las pruebas en una proporción muy baja, pues se trata de medias iguales. Sin embargo, se observa cómo la curva negra (caso Poisson) está muy por encima de las otras dos. Las curvas correspondientes a los modelos Quasi Poisson y Binomial negativa se posicionan sobre la línea punteada de color azul. Dicha línea punteada corresponde al valor 0.05.

Lo sucedido en ambos casos de la figura N° 3 no es casualidad ni producto del azar, la explicación se puede encontrar en los conceptos estadísticos de significancia y error tipo I. Cada una de las pruebas utilizadas, metodológicamente se establecieron con una significancia de 0.05. Además, teóricamente se sabe que cuando hay medias iguales y se hace una prueba de hipótesis para detectar diferencias entre medias, existe un error tipo I asociado. Ese error en otras palabras, corresponde a rechazar una hipótesis nula cuando en realidad no debía rechazarse. Según lo visto en la parte derecha de la figura N° 3, ese error tipo I para los casos quasi-Poisson y Binomial negativa coincide con el nivel de significancia establecido para las pruebas. Por otro lado, el caso Poisson infla al doble el error tipo I (aproximadamente está en 0.10).

También lo sucedido en la parte izquierda de la figura N°3 está íntimamente relacionado a lo anterior. Una cadena de eventos explican por qué en el caso de la prueba asociada al modelo Poisson, en un escenario con sobredispersión y diferentes medias, conduce a una potencia mucho más alta que la obtenida en las pruebas de los otros modelos. Al aumentar la probabilidad de error tipo I, disminuye la probabilidad de error tipo II; y al disminuir la probabilidad de error tipo II, aumenta la potencia. Por lo tanto, todo esto sugiere que en presencia de sobredispersión, no es aconsejable utilizar un modelo Poisson. Es preferible utilizar uno que sea Quasi-Poisson o Binomial negativa.

Para descartar que una situación similar a la anterior suceda en escenarios con subdispersión o equidispersión, se muestra a continuación la figura N°4.

Figura N°4. Comparación de la potencia de la prueba para observar en subdispersión y equidispersión (Medias iguales)



En este escenario para el caso de subdispersión lo correcto es rechazar las pruebas en una proporción muy baja, pues se trata de medias iguales. Sin embargo, se observa cómo la curva verde (caso Quasi Poisson) está por encima de las otras dos. Las curvas correspondientes a los modelos Poisson y Binomial negativa se posicionan de forma muy similar. De igual forma para el caso de equidispersión el comportamiento de las curvas para los tres modelos se comporta de forma muy similar.

CONCLUSIONES

Cuando nos encontramos en presencia de subdispersión no se observan diferencias en los tres modelos ya que como se pudo observar las líneas se encontraban muy cercanas, a partir de un tamaño de muestra mayor a 15 los modelos se comportan exactamente igual desde la perspectiva de la potencia de las pruebas.

Para el escenario de equidispersión, para tamaños de muestra menores a 20 se observa que la potencia de la prueba para la distribución de Poisson siempre es mayor, las pruebas de los otros dos modelos son levemente menos potentes que en el caso Poisson. Para un tamaño de muestra mayor a 25 las pruebas de los

tres modelos son igualmente potentes. Se garantizó de antemano que las medias de cada tratamiento fueran realmente diferentes. Por esto, teóricamente cada prueba debería rechazarse siempre, de aquí surge la importancia de tener cierto tamaño de muestra mínimo que produzca una potencia alta. Es decir, que se rechace cuando efectivamente tenga que rechazarse alguna prueba.

Para el caso de la sobredispersión se evidencia que la potencia de la prueba de la distribución Poisson es más alta que las otras, lo que llevaría a pensar que es mejor este modelo en el caso de sobredispersión. La literatura nos indica todo lo contrario, ya que un escenario con sobredispersión es el más problemático de todos los posibles, y utilizar un modelo Poisson no es una buena opción.

BIBLIOGRAFÍA

Colorado, A. A. (2017). *bdigital Repositorio Institucional UN*. Obtenido de <http://bdigital.unal.edu.co/62333/1/71294023.2017.pdf>

Delgado, M. A. (Junio de 2015). Obtenido de <https://idus.us.es/xmlui/bitstream/handle/11441/40815/Alcaide%20Delgado%20Mario%20TFG.pdf?sequence=1>

Díaz, M. d. (2009). *aulavirtual.agro.unlp.edu.ar*. Obtenido de http://aulavirtual.agro.unlp.edu.ar/pluginfile.php/5477/mod_resource/content/0/Apunte_Introduccion_a_MLG.pdf

Pita Fernández S., P. D. (19 de 9 de 2001). Obtenido de https://www.fisterra.com/mbe/investiga/signi_estadi/signi_estadisti2.pdf

Christian Kleiber and Achim Zeileis (2008). *Applied Econometrics with R*. New York: Springer-Verlag. ISBN 978-0-387-77316-2. URL <https://CRAN.R-project.org/package=AER>

Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

John Fox and Sanford Weisberg (2011). *An {R} Companion to Applied Regression*, Second Edition. Thousand Oaks CA: Sage . URL: <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>

II. MODELOS DE SUPERFICIES DE RESPUESTA

El análisis de datos mediante superficies de respuesta corresponde a técnicas y estadísticas que se derivan del modelado de datos donde se requiere conocer un punto óptimo y donde la relación entre las variables independientes y la variable respuesta es desconocida.

Por lo general, se plantea un modelo de primer orden, es decir, una función con términos lineales solamente, si los datos no presentan un buen ajuste, se asume que la relación entre las variables independientes y la respuesta presentan una curvatura por lo que el modelo se eleva a uno de segundo orden (uno con sus términos lineales y al cuadrado) y se prosigue con el análisis. De caso contrario, si presentan un buen ajuste, se asume que los valores escogidos para los niveles de variables predictoras no han llegado a su punto óptimo y en busca de éste se procede a cambiar de vecindad, es decir aumentar o disminuir estos valores de ser posible (Montgomery, 2002).



Optimización del pH y la turbidez en una bebida de pitahaya utilizando un modelo de superficies de respuesta

Ericka Valerio Salas³, Susana García Calvo³
erickavs189@gmail.com, asgc3093@gmail.com

RESUMEN

En este estudio interesa optimizar las variables pH y turbidez de una bebida donde se varían las cantidades de pitahaya y limón, frutas que se eligieron por sus importantes propiedades nutricionales, y en el caso de la pitahaya por su particular color que funciona como colorante natural. Para la respuesta turbidez se encontró que tiene un buen ajuste con un modelo lineal. Para obtener la turbidez recomendada (600 NTU), se fijó la cantidad de pitahaya en 6.9 g para disminuir costos y mejorar la apariencia de la bebida y despejando la ecuación obtenida del modelo, se evidenció que es necesario usar 13.9 ml de limón por cada 250 ml de bebida, por lo que no fue necesario realizar una segunda prueba ya que se encontraba entre las cantidades consideradas. Para la variable pH se obtuvo un buen ajuste con un modelo de segundo orden, ya que con el modelo de primer orden se obtiene una probabilidad asociada a la prueba de falta de ajuste de 0.01 y con un R^2 de 0.8. Una vez que se agregó la parte cuadrática al modelo de pH, analizando los signos de los valores propios se concluyó que es un punto silla, y al despejar la ecuación del modelo fijando la cantidad recomendada de pH en 3.2 y la cantidad de pitahaya en 6.9 g, se encontró que la cantidad recomendada de limón es de 9.8 ml por cada 250 ml de bebida, para mantener una consistencia entre la apariencia y el sabor.

Palabras clave: pH, turbidez, Modelo de Superficies de Respuesta, pitahaya, limón.

ABSTRACT

The main purpose of this study is to optimize the pH and turbidity of a drink where the amounts of pitahaya and lemon, fruits which were chosen due to their great and important nutritional properties and, in the case of the pitahaya, because of its particular color that works as a natural colorant. For the turbidity response, it was noticeable that it has a good adjustment with a lineal model. In order to obtain the recommended turbidity (600 NTU), the amount of pitahaya was fixed in 6.9 g in order to reduce the cost and to improve the drink's appearance, at the same time, clearing the obtained equation from the model, it was brought into evidence that it is necessary to use 13.9 of lemon for each 250ml of drink, and for that reason it was not necessary to perform a second test since it was found among the considered amounts patterns. For the pH variable a well adjustment was obtained with a second order model, given to the fact that with the first order model a probability associated to the lack of fit test was about 0.1 and a 0.8 for the R^2 associated probability. Once the quadratic terms was included to the model with the pH response variable, analyzing the eigenvalues, it came to conclusion that a saddle point was obtained and, resolving the model equation fixing the amount recommended of pH in 3.2 and 6.9 g of pitahaya, it was found that the recommended levels of lemon was 9.8 ml each 250 ml of drink to keep a consistency between the appearance and taste.

Keywords: pH, turbidity, Response Surface Model, pitahaya, lemon

³ Estudiante de estadística de la Universidad de Costa Rica



INTRODUCCIÓN

Dentro de la complejidad del funcionamiento del cuerpo humano existen grupos de átomos llamados radicales libres, su nombre se debe a que estos grupos tienen un electrón libre o desapareado, lo que genera que el átomo se vuelva sumamente reactivo con el fin de captar el electrón faltante de alguna otra molécula completa y de esta manera lograr la estabilidad electroquímica que necesita. Pero el proceso no termina con ahí, por el contrario, se vuelve un ejercicio reiterativo, ya que la molécula que le cedió el electrón al primer átomo, de igual manera queda inestable y se vuelve a repetir la búsqueda de algún electrón para estar completa. Al generarse esta actividad, los radicales libres están constantemente interactuando con todo su entorno y se produce un importante daño de moléculas, membranas celulares y tejidos; por esta razón es importante asegurarse de tener un control de estos grupos. Una forma de lograrlo es neutralizando la acción oxidante de los radicales libres con sustancias antioxidantes, cuya función es liberar electrones en la sangre para que los átomos incompletos puedan estabilizarse. Los antioxidantes se pueden encontrar en una buena alimentación diaria donde normalmente se encuentran vitaminas como la E y la C, esta última se caracteriza por ser el antioxidante hidrosoluble más abundante en la sangre (como se cita en Avello y Suwalsky, 2006).

Los cítricos, en términos nutricionales, son una muy buena fuente de vitamina C, dados sus compuestos fenólicos las frutas ácidas como el limón se caracterizan por su función antioxidante (Londoño *et al.*, 2006). Por otro lado, dos de los compuestos más sobresalientes del limón son los fenoles y los flavonoides, ya que se ha demostrado que estos pueden reducir el riesgo de enfermedades cardiovasculares y también neutralizar los radicales libres que generan el envejecimiento y el estrés oxidativo en las células (como se cita en Álvarez, Jorral & Genta, 2005).

En este caso, se pretende realizar una bebida compuesta por dos frutas, una de ellas exótica, la pitahaya, y otra más común, el limón, pero ambas con la particularidad de contener, principalmente, importantes propiedades antioxidantes que son sumamente necesarias en la dieta diaria de cualquier persona

En el caso de la pitahaya, es una cactácea columnar originaria de México, que principalmente se caracteriza por ser jugosa y por tener coloraciones bastante particulares, desde colores naranja hasta púrpura, el cual se debe principalmente a su alto contenido de betalaínas, que son pigmentos considerados como una alternativa al uso de colorantes artificiales en alimentos (Esquivel y Araya, 2012). Además, se ha observado que estos pigmentos cuentan con importantes propiedades antioxidantes que realizan actividades biológicas como inducir la enzima quinona reductasa que genera la detoxificación en la quimioprevención del cáncer (como se cita en García *et al.*, 2012).

Hablando de las variables respuesta, pH y turbidez, interesa determinar en qué grado las cantidades de pitahaya y limón que se le añaden a la bebida las afecta directamente, ya que se quiere encontrar un pH óptimo para evitar las múltiples consecuencias en la salud que podría acarrear producir una bebida con el pH desequilibrado. Estudios han demostrado que al consumir bebidas con altos niveles de ácido se pueden desencadenar enfermedades como caries, erosión dental y sobrepeso (como se cita en Novoa, 2017).

Con respecto a la turbidez, como se mencionó anteriormente, se quiere aprovechar al máximo las ventajas que las frutas elegidas proveen, en el caso de la pitahaya un colorante natural exótico que llama mucho la atención podría resultar atractivo en una bebida con un color tan particular.

METODOLOGÍA

Los modelos de superficies de respuesta (MSR) son comúnmente utilizados en el campo de química y en algunas ciencias agro, con el propósito de determinar si los factores de estudio están causando un efecto sobre la variable respuesta, y de ser así, se determinan los niveles de los factores que generan que la respuesta sea óptima (como se cita en Torres *et al.*, 2003).

La mayoría de situaciones donde se aplica esta metodología, tiene una relación desconocida entre la variable dependiente, o respuesta, y las variables independientes, por lo tanto, el primer objetivo es determinar una aproximación adecuada de la verdadera relación entre estas variables. Usualmente se empieza analizando los datos con un modelo de primer orden, y si fuera necesario se prosigue con uno de segundo orden (Montgomery, 2002).

En este experimento, los datos resultantes se analizan con un modelo de superficies de respuesta, ya que el objetivo del estudio es determinar en qué grado las variables independientes, en este caso la cantidad de limón y pitahaya, afectan el pH y la turbidez (NTU) de la bebida. El limón está medido en mililitros (ml) y la pitahaya en gramos (g), las cuales se disolvieron en agua hasta completar los 250 ml.

Para definir los niveles de ambas variables se realizó una pequeña prueba piloto con cuatro diferentes combinaciones de estas frutas, manteniendo una cantidad de azúcar constante con el fin de determinar la cantidad máxima y mínima a usar de limón y pitahaya de acuerdo al sabor y a la turbidez. Los valores para las combinaciones son 11.2, 22.5, 33.7 y 40 ml de limón y 31.9, 53.2, 74.4 y 95.7 g de pitahaya, se midió el pH y la turbidez de las diferentes muestras generadas.

De acuerdo a las mediciones obtenidas de la turbidez (ver Anexo 1), se decidió disminuir la cantidad de pitahaya y de limón utilizada en la bebida por recomendación del Lic. Arturo Gutiérrez, Asistente del Programa de Apoyo Tecnológico a la Industria (PATI) de la Universidad de Costa Rica, con el objetivo de disminuir los costos de producción y darle una mejor apariencia y sabor a la bebida.

Por lo anteriormente descrito, para la muestra inicial del experimento se decidió trabajar con niveles de 0 a 30 ml de limón, y de 2.3 a 11.5 g de pitahaya por cada 250 ml (ver Anexo 2). La muestra final consta de 10 combinaciones de cantidades de las frutas incluyendo los puntos axiales del modelo por recomendación del experto. Todas las mediciones se realizaron en un laboratorio de la Escuela de Tecnología de Alimentos de la Universidad de Costa Rica, y los datos fueron analizados con el software estadístico RStudio versión 3.4.2 (RStudio Team, 2016), utilizando como principales paquetes los siguientes: *lattice* (Sarkar and Deepayan, 2008), *rgl* (D. Adler et al., 2018) y *rsm* (Russell V., 2009). La significancia utilizada para las pruebas de hipótesis fue de 0.05.

Los modelos planteados son los siguientes:

1. Sea Y: cantidad de pH en la bebida.

$$Y_i = \beta_0 + \beta_1 \text{limon} + \beta_2 \text{pitahaya} + \beta_3 (\text{limon} * \text{pitahaya}) + \beta_{11} \text{limon}^2 + \beta_{22} \text{pitahaya}^2 + \varepsilon_i$$

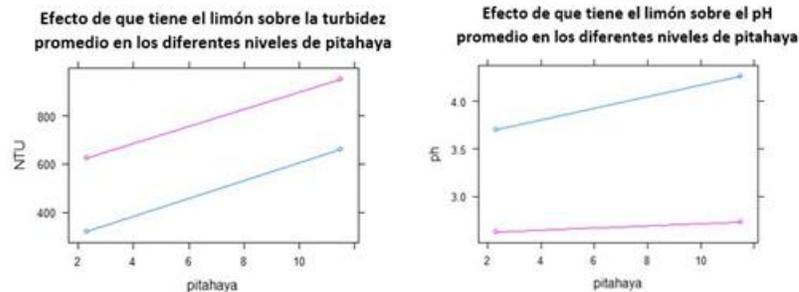
2. Sea Y: cantidad de turbidez en la bebida.

$$Y_i = \beta_0 + \beta_1 \text{limon} + \beta_2 \text{pitahaya} + \beta_3 (\text{limon} * \text{pitahaya}) + \beta_{11} \text{limon}^2 + \beta_{22} \text{pitahaya}^2 + \varepsilon_i$$

RESULTADOS

Como primer paso, una vez planteados los modelos, se realizó un análisis de varianza para estudiar si el efecto que tiene la cantidad de limón sobre el pH y turbidez promedio depende o no de la cantidad de pitahaya que se aplique, es decir, si hay o no interacción. Del análisis se concluye que la interacción entre limón y pitahaya no es significativa para el modelo cuya variable respuesta es el pH ($p\text{-value}=0.65$), al igual que para el modelo con respuesta turbidez ($p\text{-value}=0.92$), lo cual se puede observar en la Figura 1, por lo anterior, las interacciones fueron eliminadas de ambos modelos.

Figura 1.



Con los modelos reducidos, se analizó si estos presentaban un buen ajuste a un modelo de primer orden, es decir, un modelo lineal, y la hipótesis nula utilizada plantea que no hay falta de ajuste. Para la variable turbidez se obtuvo un buen ajuste a un modelo lineal con una probabilidad asociada de 0.25; en el caso del pH se presentó, por el contrario, que no existe un buen ajuste con un modelo de primer orden, con una probabilidad de 0.01, pues no parece haber una relación lineal entre las cantidades de fruta y el pH. Ambos modelos presentan coeficientes de determinación superiores a 0.8 lo cual es consistente para la variable respuesta turbidez, pero en el caso del pH resultó un R^2 alto pues ambas frutas son ácidas y aportan pH a la bebida, es por esto que, aunque el modelo no presente un buen ajuste a un modelo lineal, podría ser que la fuerte relación que hay entre las variables predictoras y la respuesta sea la que garantice que de repetirse el experimento se van a obtener resultados muy similares.

El modelo de primer orden con la variable respuesta turbidez presentó un buen ajuste, es decir, se ajusta a un modelo lineal. Previamente, en la prueba piloto se notó que al aumentar la cantidad de pitahaya el atractivo color natural que esta aporta era cada vez más oscuro y por ende menos notable, entonces, tomando en consideración el alto costo de la pitahaya, y además por la sugerencia del experto de no explorar con mayores cantidades para analizar la existencia de un punto óptimo, se decidió despejar la fórmula obtenida con el fin de analizar los datos con un modelo lineal, fijando la turbidez en una cantidad deseada y una cantidad de pitahaya escogida a conveniencia (según precio y apariencia).

La fórmula obtenida fue la siguiente:

$$\widehat{\text{turbidez}} = 210.04 + 10.02 * \text{limon} + 36.36 * \text{pitaya}$$

Se fijaron los valores de turbidez en 600 NTU y la cantidad de pitahaya en 6.9 g, teniendo en cuenta el color atractivo que aporta y el costo que implica usar una mayor cantidad por ser la pitahaya una fruta exótica. Despejando la fórmula anterior, se obtuvo que con 13.9 ml de limón y 6.9 g de pitahaya por cada 250 ml de bebida se lograría alcanzar el nivel de turbidez recomendado, este valor se encuentra entre los valores seleccionados en primera instancia para la muestra, por lo que no es necesario realizar una segunda muestra.

La falta de ajuste con un modelo lineal en el caso de la respuesta pH permite escalar el modelo a uno de segundo orden, además, se observa que el efecto que tiene la pitahaya sobre el pH de la bebida no es significativamente diferente de cero. Por lo que se procedió a plantear el siguiente modelo:

Sea Y : cantidad de pH en la bebida.

$$Y_i = \beta_0 + \beta_1 \text{limon} + \beta_2 \text{pitahaya} + \beta_{11} \text{limón}^2 + \beta_{22} \text{pitahaya}^2 + \beta_{12} (\text{limón} * \text{pitahaya}) + \varepsilon_i$$

El análisis de varianza no muestra evidencia estadística para considerar significativa la interacción entre las dos variables predictoras, por lo que el modelo se reduce a uno sin interacción. El coeficiente de terminación de este nuevo modelo es de 0.96 y el análisis de varianzas correspondiente evidencia que los términos cuadráticos resultan significativos en conjunto, con una probabilidad asociada de 0.004 con esto. Se concluyó que el modelo sí presenta un buen ajuste con uno de segundo orden. La cantidad de pitahaya parece no contribuir mucho con la curva, es decir, su término cuadrático no resultó ser significativo ($p\text{-value}=0.4$), esto puede ser por la cantidad reducida que se usó para lograr una turbidez y color deseados.

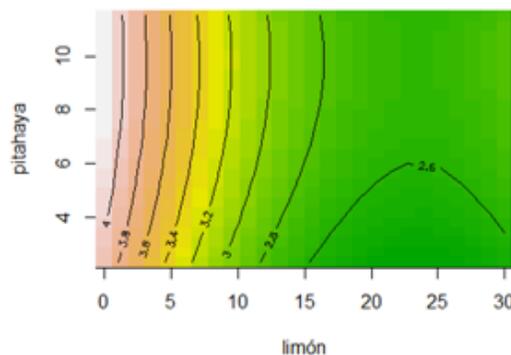
Los valores propios son 0.003 y -0.004 al ser signos opuestos se encontró un caso de punto silla y este punto se alcanza cuando la cantidad usada de limón es de 23.5 ml y 9.8 g de pitahaya.

Para buscar un punto donde se optimice el pH, se despeja la siguiente fórmula fijando la respuesta en 3.2 por recomendación del experto, ya que se conoce que el nivel de pH máximo para empacar las bebidas en caliente es de 3.2, además se fijó la cantidad de pitahaya en 6.9 g obteniendo la siguiente fórmula:

$$\widehat{3.2} = 3.76 - (0.13 * \text{limón}) + (0.09 * 6.9) + (0.003 * \text{limón}^2) - (0.004 * 6.9^2)$$

Las dos soluciones obtenidas fueron 9.8 ml y 37.2 ml, pero debido a que la segunda no se encuentra dentro del intervalo contemplado en los niveles de limón utilizados se decidió usar 9.8 ml de limón y 6.9 g de pitahaya para logra alcanzar el pH igual a 3.2 lo cual también es consistente con la turbidez deseada pues se obtienen cantidades muy parecidas de limón a utilizar. Esto se puede apreciar en la Gráfico 1, donde la cantidad de limón ronda entre los 7.5 y 10 ml fijando los valores de pH y pitahaya como se mencionó anteriormente.

Gráfico 1. Curvas de contorno para la respuesta pH según la cantidad de limón y pitahaya



CONCLUSIONES

A pesar de no utilizar una gran cantidad de pitahaya en la bebida, las propiedades antioxidantes y su función como colorante artificial siguen presentes. En el caso del limón, también se aprovecharon todos sus beneficios contenidos en el jugo, tales como disminuir el riesgo de enfermedades cardiovasculares y evitar el envejecimiento y oxidación celular gracias a su alto contenido de vitamina C.

Con respecto a las variables respuesta, la turbidez presenta una relación lineal con pendiente positiva con las variables independientes, por lo tanto, esta respuesta no debe analizarse con un MSR ya que es una línea de regresión. En el caso opuesto, el pH sí tiene una relación cuadrática con las cantidades de limón y pitahaya, además, el coeficiente de este modelo que determina qué tanto explican estas dos últimas variables de la variabilidad del pH es de 0.9, lo que significa que en el futuro se pueden predecir resultados si se realiza un experimento similar a este. Cabe resaltar que la cantidad de pitahaya no resultó significativa para la respuesta pH, lo que indica que la variable que más explica la variabilidad de esta respuesta es la cantidad de limón.

Según el interés principal de este estudio, se obtuvo que las cantidades de fruta donde se llega a alcanzar un pH óptimo son 9.8 ml de limón y 6.9 g de pitahaya por cada 250 ml de bebida.

BIBLIOGRAFÍA

Avello, M., & Suwalsky, M. (2006). Radicales libres, antioxidantes naturales y mecanismos de protección. *Atenea (Concepción)*, (494), 161-172.

Daniel Adler, Duncan Murdoch and others (2018). rgl: 3D Visualization Using OpenGL. R package version 0.99.16. <https://CRAN.R-project.org/package=rgl>

Esquivel, P., & Araya, Q. Y. (2012). Características del fruto de la pitahaya (*Hylocereus* sp.) y su potencial de uso en la industria alimentaria. *Revista Venezolana de Ciencia y Tecnología de Alimentos*, 3(1), 113-129.

García-Cruz, L., Salinas-Moreno, Y., & Valle-Guadarrama, S. (2012). Betalaínas, compuestos fenólicos y actividad antioxidante en pitahaya de mayo (*Stenocereus griseus* H.). *Revista fitotecnia mexicana*, 35(SPE5), 01-05

Londoño, J., Montoya, G., Guerrero, K., Aristizabal, L., & Arango, G. J. (2006). Los jugos de cítricos inhiben la oxidación de lipoproteínas de baja densidad: relación entre actividad captadora de radicales libres y movilidad electroforética. *Revista chilena de nutrición*, 33(3), 544-551.

Montgomery, D. (2004). Diseño y análisis de experimentos. Limusa Wiley, Segunda Edición, México.

Novoa Padilla, F. X. (2017). *Determinación del pH salival de varias bebidas no alcohólicas: y su relación con la erosión y caries dental en estudiantes de la facultad de odontología de clínica cuatro y cinco de la Universidad de Las Américas Quito* (Bachelor's thesis, Quito: Universidad de las Américas, 2017.).

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.

Russell V. Lenth (2009). Response-Surface Methods in R, Using rsm. *Journal of Statistical Software*, 32(7), 1-17. URL <http://www.jstatsoft.org/v32/i07/>.

Sarkar, Deepayan (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5

Torres, R., Chacín, F., García, J., & Ascanio, M. (2003). Optimización en modelos de superficies de respuesta. *Revista Facultad Agronomía (Maracay)*, 29, 69-83.

ANEXOS

1)

Cuadro 1
Valores de las respuestas pH y turbidez
según cantidad de fruta en la prueba piloto

Combinación	pH	turbidez
11.2 ml limón/ 31.9 gr pitahaya	3.22	600
22.5 ml limón/ 53.2 gr pitahaya	3.14	3000
33.7 ml limón/ 74.4 gr pitahaya	3.07	12000
40 ml limón/ 95.7 gr pitahaya	3.04	22500

2)

Cuadro 2
Resultado de las mediciones después de definir
los límites de las variables cantidad de pitahaya y limón

# medición	Pitahaya	Limón	pH	Turbidez
1	2.3	0	3.70	318
2	2.3	15	2.75	536
3	2.3	30	2.63	626
4	11.5	0	4.26	660
5	11.5	15	2.79	656
6	11.5	30	2.73	953
7	6.9	0	4.28	543
8	6.9	30	2.68	894
9	6.9	15	2.78	573
10	6.9	15	2.79	5

Influencia de la temperatura y el tiempo en la eficiencia de cocción de la carne de res

José Oviedo⁴, Juan José Montero⁴
froysur@gmail.com, jmontero639@gmail.com

RESUMEN

En el campo de la Tecnología de alimentos, la temperatura interna ideal de la carne depende esencialmente de dos factores: temperatura y tiempo de cocción. Dicha temperatura interna debe garantizar que la carne sea apta para el consumo humano, pero adicionalmente se quiere hacer un uso adecuado de los recursos (temperatura y tiempo). Se determina que, para la carne roja, una temperatura interna deseable es 169° Fahrenheit. Se establecen posibles rangos de combinación temperatura-tiempo donde la temperatura interna podría ser más eficiente. Partiendo de eso, se diseña un experimento en el que se utiliza la metodología de Superficies de Respuesta, mediante un Diseño Central Compuesto. Entre los principales hallazgos, se logra determinar que en efecto el tiempo y temperatura son factores significativamente importantes en este estudio. También, que la combinación 340 grados y 17 minutos produce una eficiencia máxima. Además, se determina que existe más de una combinación para lograr niveles de eficiencia deseables.

Palabras clave: Superficies de respuesta, cocción de la carne, temperatura, tiempo

ABSTRACT

The family of Food Technology indicates that the internal temperature of the meat depends on two factors: temperature and cooking time. This internal temperature must guarantee that the meat is suitable for human consumption. Additionally, it is necessary to make an adequate use of the resources (temperature and time). The desirable internal temperature of red meat is 169 ° Fahrenheit. First, possible ranges of temperature-time combination are established where the internal temperature could be more efficient. Then, an experimental design is made with the Response Surface methodology, using a Central Composite Design. Among the main findings, it is possible to determine that cooking time and temperature are statistically significant factors. Also, the combination 340 ° Fahrenheit and 17 minutes produces maximum efficiency. In addition, it is determined that there is more than one combination to achieve desirable efficiency levels.

Keywords: Response Surface Model, cooking of meat, temperature, time

⁴ Estudiante de estadística de la Universidad de Costa Rica



INTRODUCCIÓN

En la actualidad, gracias al avance científico existe conocimiento acerca de datos numéricos relacionados al consumo y preparación de los alimentos. Algunas medidas surgen con el fin de eliminar los riesgos de contaminación de ciertas comidas. Dependiendo del tipo de alimento, este puede contener microorganismos dañinos para la salud de los seres humanos. Las canales, despojos, sangre, carnes picadas, especias, frutas y verduras albergan fuertes cargas microbianas capaces de desarrollarse. La prevención al respecto, se basa especialmente en el binomio tiempo-temperatura.

El Departamento de Agricultura de los Estados Unidos de América (USDA) en publicación del año 2015 detalla que se debe cocinar la carne molida de res, cerdo, cordero y ternero hasta alcanzar una temperatura interna de 160° F (71.1° C) y que se debe medir con un termómetro para alimentos. Además, indica que toda ave debe alcanzar una temperatura interna mínima adecuada de 165° F (73.9° C). Dicho departamento recomienda usar el termómetro para alimentos para asegurar que los alimentos cocidos se mantengan a una temperatura adecuada hasta el momento de ser servidos. En Costa Rica, el profesional en Tecnología de alimentos José Alfredo Ureña Haug, indica que la temperatura a la que los diferentes patógenos son destruidos varía, así como la cocción de los diferentes alimentos. Por esto, él recomienda para el corte de carne de res, una temperatura interna de 169° F (76° C).

El presente estudio plantea el objetivo de encontrar una combinación de temperatura y tiempo idónea para la carne. De manera que sea posible mejorar los tiempos de entrega en establecimientos y así brindar un mejor servicio. Pero también, preocupándose por la seguridad de los consumidores, esto desde el punto de vista de la salubridad. Este objetivo se establece después de conversar con el Tecnólogo de alimentos José Alfredo Ureña Haug, que nos indicó los parámetros necesarios para que la cocción de la carne fuera apta para seres humanos.

METODOLOGÍA

Se hace un diseño experimental llamado central compuesto, el cual consiste en buscar un punto óptimo que determine cuál será la mejor combinación de tiempo y temperatura de cocción de la carne, además de otras combinaciones que pudiesen resultar eficientes en este contexto. Durante el proceso de diseño se contó con la colaboración del Tecnólogo de alimentos José Alfredo Ureña Haug. Él propuso los parámetros necesarios para que la cocción de la carne fuera apta para seres humanos. Además, brindó asesoría para garantizar que no se cometieran errores en el proceso de medición.

Se parte del hecho de que existe una temperatura interna ideal de la carne de res. En este caso, al decir temperatura interna ideal, se hace referencia a que sea una temperatura suficiente para garantizar que pueda ser consumida por humanos. El experto pone como parámetro para este tipo

de carne los 169 grados Fahrenheit (76 grados Celsius). Él considera que esa temperatura interna de la carne brinda una eficiencia de 100%.

Según el Tecnólogo de Alimentos, una eficiencia del 95% es aceptable. Se encuentra cerca de la temperatura interna ideal para este tipo de carne. Esa temperatura interna se alcanza mediante alguna combinación de tiempo y temperatura de horneado. Así se logra que el conteo de microorganismos perjudiciales para la salud esté dentro de lo aceptable. Más de 169 grados Fahrenheit de temperatura interna de la carne significa una pérdida de eficiencia. Esto debido a que se utilizan más recursos (tiempo, temperatura o ambos) inadecuadamente. Menos de 169 grados Fahrenheit también está asociado a pérdida de eficiencia, pues se expone la salud al aumentar el conteo de microorganismos.

Según sugerencia del Tecnólogo, el experimento se puede iniciar con temperaturas entre los 300 y 400 grados Fahrenheit. Tiempos entre los 10 y 20 minutos. Él considera que entre esos rangos se encuentra el máximo de eficiencia. Se utiliza un horno (temperaturas en grados Fahrenheit) y un termómetro especial para carne. El termómetro está en escala de grados Celsius. Se toman dos mediciones de cada combinación de tiempo-temperatura, según los cuatro valores anteriormente establecidos. Además, se hacen 6 repeticiones en lo que se llama valor central (tiempo=15, temperatura=350). A todas las corridas de combinaciones y valores centrales se les registra la temperatura interna. Dicha temperatura interna (en grados Celsius) de la carne es convertida por el experto en una medida de eficiencia que puede tomar valores de 0 a 100. La conversión consiste en que por cada grado Celsius menor o mayor a 76, se pierden 5 puntos de eficiencia. Por ejemplo, si la temperatura interna medida fuera 73 o 79, entonces el puntaje de eficiencia sería 85 para cualquiera de los dos valores. La eficiencia es utilizada como variable respuesta, y como factores se utilizan el tiempo (en minutos) y temperatura de horneado (en grados Fahrenheit). Hasta este punto lo que se hizo fue utilizar un diseño factorial con puntos centrales.

Se analizan esos datos utilizando las librerías rsm (Russell, 2009) y rgl (Adler, 2019), del paquete estadístico R versión 3.5.2 (R Core Team, 2018). En todas las pruebas se trabaja con un nivel de significancia de 0.05. Al avanzar en el análisis, mediante el ajuste de un modelo de primer orden, se logra determinar que el vecindario (límites de temperatura y tiempo) sugerido por el experto es el correcto. Se conservan los datos recolectados y se procede a agregar lo que se denomina puntos axiales. Para este caso, los axiales utilizados se redondearon. A continuación, los puntos:

Cuadro N°1
Puntos axiales

Minutos	Temperatura
15	280
15	420
8	350
22	350

Luego, se le mide el valor de temperatura interna a cada combinación axial y se convierte a medida de eficiencia dicho valor. Se construye una nueva base con los datos antiguos y los puntos axiales. Se analizan los datos de dicha base (aún se utiliza la librería "rsm"), pero se ajusta un modelo de segundo orden (que incluye términos cuadráticos), esto ayuda a darle forma a la superficie. Así se completa el diseño central compuesto.

RESULTADOS

El criterio del experto en Tecnología de Alimentos fue esencial para identificar una región donde fuera posible encontrar un máximo, pero el análisis formal empieza con el ajuste de un modelo de regresión lineal de primer orden. El modelo propuesto es el siguiente:

$$\mu|_{x_1, x_2} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2$$

Con la siguiente descripción de cada elemento:

- $\mu|_{x_1, x_2}$: Promedio de eficiencia, dado cierto tiempo y temperatura.
- β_0 : Intercepto.
- β_1 : Coeficiente para tiempo.
- X_1 : Variable llamada tiempo.
- β_2 : Coeficiente para temperatura.
- X_2 : Variable llamada temperatura.
- β_{12} : Coeficiente para la interacción tiempo-temperatura.

Al ajustar el modelo, la prueba F del ANOVA muestra para el término de interacción una probabilidad asociada de 0.17. Por lo tanto, se puede asumir que no existe interacción entre tiempo y temperatura. Se procede ahora a ajustar otro modelo de primer orden, pero sin el término de interacción.

Al ajustar el modelo nuevo, se procede a poner a prueba la hipótesis nula que dice que el modelo de primer orden presenta un buen ajuste. La probabilidad asociada a dicha prueba es de 0.0001, por lo que se rechaza la hipótesis nula y se puede pensar que existe mal ajuste. Como segundo indicador está el coeficiente de determinación asociado. En este caso dicho coeficiente tiene un valor de 0.037. Eso significa que el modelo apenas explica un 3.7% de la variabilidad total. Tal evidencia de falta de ajuste, tanto en la prueba como en el R^2 , permite verificar que el vecindario donde se encuentra el máximo de eficiencia corresponde a los límites propuestos por el experto.

El paso a seguir es ajustar un modelo de segundo orden, con los mismos datos anteriores, pero agregando puntos axiales a la base. La intención de dichos puntos es completar un diseño central compuesto, así poder modelar mejor la superficie de respuesta. El modelo utilizado es el siguiente:

$$\mu|_{x_1, x_2} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1)^2 + \beta_4 (X_2)^2 + \beta_{12} X_1 X_2$$

Con la siguiente descripción de cada elemento:

- $\mu|_{x_1, x_2}$: Promedio de eficiencia, dado cierto tiempo y temperatura
- β_0 : Intercepto.
- β_1 : Coeficiente para tiempo.
- X_1 : Variable llamada tiempo.
- β_2 : Coeficiente para temperatura.
- X_2 : Variable llamada temperatura.
- β_{12} : Coeficiente para la interacción tiempo-temperatura.
- β_3 : Coeficiente para el tiempo al cuadrado.
- β_4 : Coeficiente para la temperatura al cuadrado.

El modelo anterior es el definitivo. Se probaron los supuestos de normalidad y homoscedasticidad, y no se encontraron problemas de incumplimiento de dichos supuestos (ver anexos, Figura N° 1). En cuanto a las pruebas de hipótesis relacionadas a cada término, según las bajas probabilidades asociadas (muy cercanas a cero), debe conservarse tanto la interacción como cada uno de los términos restantes. Para este modelo, el R^2 es ahora de 0.769 y la prueba de ajuste no se rechaza. Todo parece indicar que este modelo de segundo orden es adecuado para obtener buenos resultados.

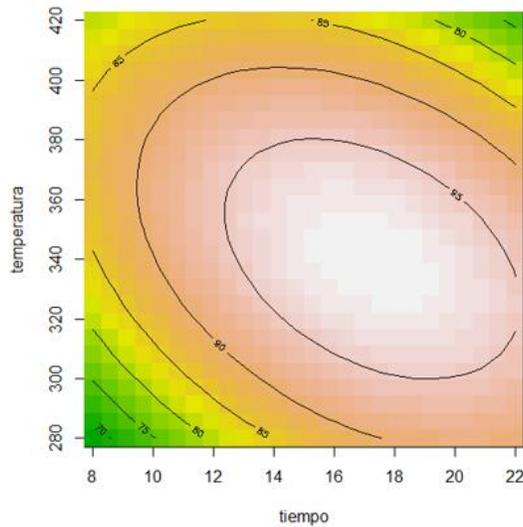
La ecuación final queda representada de la siguiente forma:

$$\hat{y} = -308.94 + 10.57X_1 + 1.85X_2 - 0.158(X_1)^2 - 0.0023(X_2)^2 - 0.015X_1X_2$$

Los valores propios asociados al ajuste del modelo son -0.001 y -0.158. Por lo anterior, el punto estacionario en este caso es un máximo. Además, la combinación de valores en dicho punto estacionario (redondeados a números enteros), corresponde a 17 minutos y 340 grados Fahrenheit. Al sustituir dicha combinación en la ecuación, da como resultado una eficiencia promedio de 98%. Además, el intervalo de confianza de 95% con esos mismos valores de temperatura y tiempo, indica que la eficiencia promedio asociada está entre 96% y 100%.

La combinación de tiempo y temperatura buscada para alcanzar el 95% de eficiencia sugerido por el experto no necesariamente corresponde a una única combinación. Para ilustrar, suponga que, si una persona quiere más rápido la carne, se puede aumentar la temperatura y disminuir el tiempo razonablemente. Eso no tendría por qué afectar la eficiencia de la temperatura interna de la carne, siempre que se utilice una combinación que funcione. A continuación, el gráfico N° 2 puede terminar de aclarar mejor esta situación.

Gráfico N° 2
Curvas de contorno para la eficiencia



Se puede observar la consistencia con el resultado del punto estacionario obtenido anteriormente. Véase que la combinación de 340 grados y 17 minutos en efecto está cercana a lo que sería la eficiencia del 100%. Por otra parte, las curvas de nivel representadas en el Gráfico N° 2 evidencian que para tener una eficiencia de 95% no es necesario tener una sola combinación de tiempo y temperatura. Por ejemplo, véase que una temperatura de 360 grados y tiempo de 13 minutos da la misma eficiencia que 310 grados y 16 minutos. Es también posible observar que una combinación de temperatura alta (420) y tiempo alto (22) incide negativamente en la eficiencia. Si bien es cierto, lo anterior logra el objetivo de entrar en un rango donde el conteo de microorganismos es aceptable. Pero como vimos antes, hay combinaciones que utilizan menos recursos y logran el mismo objetivo, por lo que se consideran más eficientes.

CONCLUSIONES

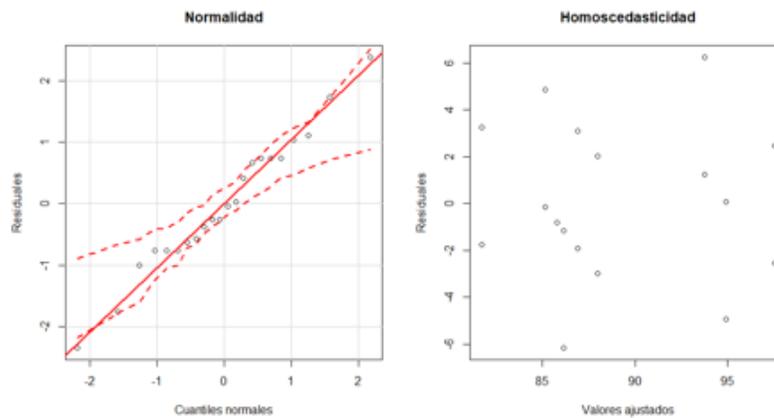
Las referencias previas y los resultados de este estudio coinciden en el hecho de que, para el caso de la carne de res, es realmente importante un equilibrio adecuado entre el tiempo y la temperatura de cocción. Dicha combinación está asociada a cierta temperatura interna de la carne. Es indispensable obtener una temperatura interna que permita que la carne sea apta para el consumo humano. Pero también, simultáneamente es deseable que no exista un desperdicio de recursos para lograr lo anterior. Tal y como se observó en los resultados, existen combinaciones temperatura-tiempo que no son eficientes. Esto a pesar que dichas combinaciones cumplan con garantizar que la carne sea apta para consumo humano. La razón es porque se hace un uso no adecuado del tiempo, de la temperatura, o de ambos.

Refiriéndose siempre a las combinaciones, se determinó que el máximo de eficiencia es posible obtenerlo con 340 grados Fahrenheit y 17 minutos. Sin embargo, otras combinaciones producen niveles de eficiencia deseables. Esto significa que el equilibrio entre la salubridad y el uso de recursos

es posible en más de un escenario. Así, se genera mayor flexibilidad, dependiendo de lo que desee cada persona en algún contexto específico. Por ejemplo, si un negocio de comidas quisiera mejorar sus condiciones de operación, podría echar mano de los resultados aquí obtenidos. Es decir, si se necesitara que la carne esté lo antes posible, bastaría con bajar el tiempo y aumentar la temperatura de manera conveniente, pero apegándose a límites establecidos por el modelo propuesto en los resultados.

ANEXOS

Figura N° 1
Verificación de supuestos



BIBLIOGRAFÍA

Agriculture, U. S. (2015). Obtenido de <https://www.fsis.usda.gov/wps/wcm/connect/fsis-content/internet/informational/en-espanol/hojasinformativas/aparatos-electrodomesticos-termometros/termometros-de-cocina/termometros-de-cocina>

Daniel Adler, Duncan Murdoch and others (2019). rgl: 3D Visualization Using OpenGL. R package version 0.100.19. <https://CRAN.R-project.org/package=rgl>

Preciado, G. F. (Diciembre de 2003). Obtenido de https://s3.amazonaws.com/academia.edu.documents/35939304/MSR.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1542572519&Signature=04D7SIjxQ2gDFLkiOzC5R9bZ9DE%3D&response-content-disposition=inline%3B%20filename%3DOPTIMIZACION_DE_UNA_SUPERFICIE_DE_RESPUE.p

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

III. MODELOS MIXTOS

Los modelos de efectos mixtos extienden el modelo lineal de forma que, el objetivo se encuentra relacionado con los factores y los factores por medio de una función de enlace específica. Pueden presentar una distribución no normal, puede existir correlación entre las observaciones. Tienen la característica de que algunas variables son tratadas como aleatorias y otras son consideradas como fijas. Cubren una amplia variedad de modelos, desde modelos de regresión lineal simple hasta modelos multinivel complejos para datos longitudinales no normales.

Dentro de los modelos mixtos encontramos el modelo de medidas repetidas, es considerado una extensión del diseño de bloques, en que el sujeto sustituye al bloque y actúa de control propio. Por medio de esto se obtiene una disminución en la variancia del error, esto se debe a que las diferencias individuales son eliminadas del error. Los sujetos de la muestra van a recibir todos los tratamientos y se generan repeticiones de la respuesta. (IBM, s.f.)



Análisis de la evolución del desempeño en una prueba de velocidad a lo largo de tres meses, dentro de un contexto de entrenamiento futbolístico, mediante un modelo mixto de medidas repetidas

José Oviedo Jiménez⁵
froysur@gmail.com

RESUMEN

La explosión de popularidad y profesionalismo que el fútbol implica en la actualidad conlleva a que cualquier equipo, ya sea chico o grande adopte las nuevas tendencias de entrenamiento para poder ser competitivo. Las mediciones de rendimiento en aspectos meramente físicos son un recurso cada vez más utilizado por los entrenadores. Por tal razón, el presente artículo toma datos acerca de una prueba de velocidad realizada como parte del proceso de entrenamiento de un equipo de fútbol. Se hace un análisis mediante un modelo mixto de medidas repetidas, esto para ver si con el paso de los meses, se logra ir mejorando el desempeño en dicha prueba. Al final se hace una comparación entre el criterio del experto y los resultados obtenidos. El principal resultado obtenido indica que desde la primera repetición y hasta la última, en general hubo una mejora en el tiempo que ronda entre los 12 y 21 segundos. Por otro lado, la mejora fue mucho mayor para los jugadores que en su primera repetición tenían tiempos muy altos. Mientras que los jugadores con tiempo inicial bajo mantenían una tendencia constante.

Palabras clave: fútbol, acondicionamiento físico, modelo mixto, medidas repetidas

ABSTRACT

The explosion of the popularity and professionalism that football soccer implies currently, lead to any kind of team, whether a small one or a big one; to adopt the new training tendencies in order to be competitive. The performance measurements in merely physical aspects are resources that are each time more utilized by soccer coaches and trainers. For such reason, this article takes data from a speed test as a part of a soccer team training process. An analysis carried through a mixed model of repeated measures is conducted in order to be able to observe if as the months pass by, an improvement in performance in the mentioned test is doable and reachable. In the end, there is a comparison between the expert criteria and the obtained results. The main obtained result shows that from the first repetition to the last one, there was a time improvement that goes between the 12 and 21 seconds. On the other hand, the improvement of the time of the players was a lot higher when they showed better timing from the first repetition while the other ones with lower timing kept a constant tendency since the beginning.

Keywords: Soccer, Physical training, mixed model, repeated measures

⁵ Estudiante de estadística de la Universidad de Costa Rica



INTRODUCCIÓN

El fútbol es un deporte que nació en Inglaterra a finales del siglo XIX, a lo largo del siglo XX tomó fuerza, más aún cuando se popularizó en el cono sur de América. Después de su auge europeo y latinoamericano, se extendió globalmente al punto de ser el deporte rey a nivel mundial. Tal disciplina deportiva ha evolucionado tanto, que hoy en día gracias a los avances tecnológicos su profesionalización alcanza niveles que sus pioneros jamás hubiesen imaginado. Sumas alucinantes de millones de dólares están día a día en juego debido al impacto del fútbol. En los clubes que conforman la élite futbolística es común ver cómo se invierte una enorme cantidad de recursos para poder estar en la cima de la máxima competencia.

Entre los recursos tecnológicos utilizados para el mejoramiento de ciertas condiciones del desarrollo futbolístico se encuentra la opción de analizar datos que contribuyan a mejorar el entrenamiento físico. Hoy en día, por ejemplo, es posible el uso de sensores capaces de monitorear y capturar información valiosa del jugador. Además, los procesos de globalización han permitido que el conocimiento desarrollado referente al contexto futbolístico esté al alcance de todos. En la actualidad cualquier persona puede tener acceso a información y documentación relativa a los métodos de entrenamiento más efectivos, incluso a los de los mejores equipos del orbe (Bangsbo, 2008).

Una parte muy importante del desarrollo futbolístico es la condición física. Nunca antes en la historia del fútbol se le había dado tanto énfasis a la parte física como en los últimos años. En el pasado bastaba con que el jugador tuviera buena técnica. Ahora el roce físico es mayor, esto demanda jugadores más potentes en cuanto a biotipo físico y a la vez rápidos (Bangsbo, 2008). Debido a la demanda de jugadores cada vez mejor formados en su constitución física, ya desde la adolescencia se incluyen programas de entrenamiento donde además de la técnica y la táctica, se realizan pruebas y rutinas para mejorar el acondicionamiento físico (Lindquist & Bangsbo, 1992).

El presente artículo toma como referencia el caso de un proyecto deportivo costarricense. El técnico y sus colaboradores utilizan métodos de entrenamiento que contemplan la actualidad futbolística, de manera que el jugador pueda tener el nivel exigido por el alto rendimiento internacional. El objetivo del artículo es analizar si los jugadores al repetir cierta prueba física por varios meses, pueden llegar a mejorar el desempeño en dicha prueba, esto mediante un tipo de entrenamiento llamado periodización táctica.

METODOLOGÍA

En el año 2015 el entrenador de fútbol costarricense Stefano Albertazzi inició un proyecto deportivo en el cantón central de San José. Se hizo una convocatoria de hombres de entre 15 y 18 años interesados en participar en torneos internacionales en la ciudad de Miami en los Estados Unidos. El gran incentivo es llegar a ser tomados en cuenta en algún club. Como parte del proceso de entrenamiento, lo primero que busca el entrenador es hacer un diagnóstico del rendimiento físico y

así empezar con el proceso de acondicionamiento. Entre los diagnósticos se encuentra una prueba muy similar a correr una milla, de ahí se mide el tiempo en segundos hasta finalizar dicha prueba. Esta fue repetida a lo largo de 3 meses (4 repeticiones). La visión del entrenador es que con el paso del tiempo y el entrenamiento periódico, se logra que los jugadores mejoren el tiempo de ejecución de la prueba.

Se hace un análisis mediante un modelo mixto de medidas repetidas, donde la variable respuesta es el tiempo en segundos. Hay una variable continua que es el mes, en principio está con valores de 1 a 4, pero para efecto de gráficos va de 0 a 3. El componente aleatorio del modelo se representa por la variable llamada sujeto, es el número de identificación de 1 a 47 de cada uno de los jugadores. Para dicho análisis se utiliza el paquete estadístico R versión 3.5.2 (R Core Team, 2018). Las librerías requeridas para el análisis son lme4 (Bates et al., 2015) y lattice (Sarkar, 2008).

El modelo final utilizado para hacer las estimaciones es:

$$Y_i = \beta_0 + \beta_1 mes + \beta_{i,0} + \beta_{i,1} mes + \epsilon_i$$

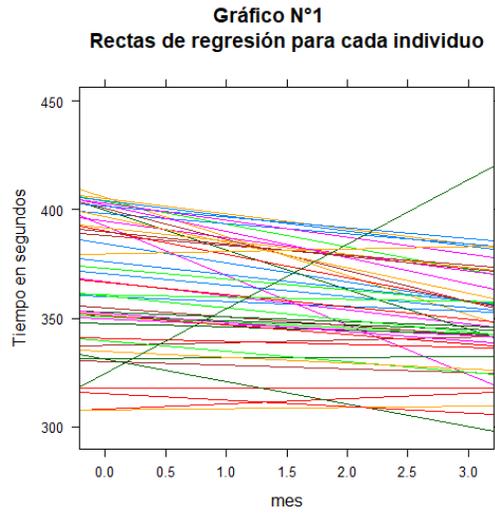
- β_0 = intercepto de la recta general
- β_1 = pendiente de la recta general
- $\beta_{i,0}$ = intercepto aleatorio
- $\beta_{i,1}$ = pendiente aleatoria
- ϵ_i = término de error

El proceso de análisis empieza con una exploración gráfica, se intenta observar la tendencia general del tiempo en segundos al pasar los meses. También, se hace un gráfico que ilustra cada línea de regresión por individuo. Luego de toda la exploración, se procede a ajustar dos modelos: uno que contempla correlación intercepto-pendiente y otro en el que no. Se comparan ambos modelos mediante una prueba de razón de verosimilitudes y se escoge uno de ellos. El modelo “ganador” se compara contra uno que no tenga el término correspondiente a la pendiente aleatoria, para ello se utiliza de nuevo la prueba de razón de verosimilitudes. El nuevo modelo “ganador” se somete a la verificación de los supuestos de normalidad y homoscedasticidad. Luego de los supuestos, se ajusta ese modelo final y se extraen las principales estimaciones puntuales y por intervalo, tanto de la parte fija como de la parte aleatoria. Todos los intervalos de confianza se hacen con un 95% de confianza.

RESULTADOS

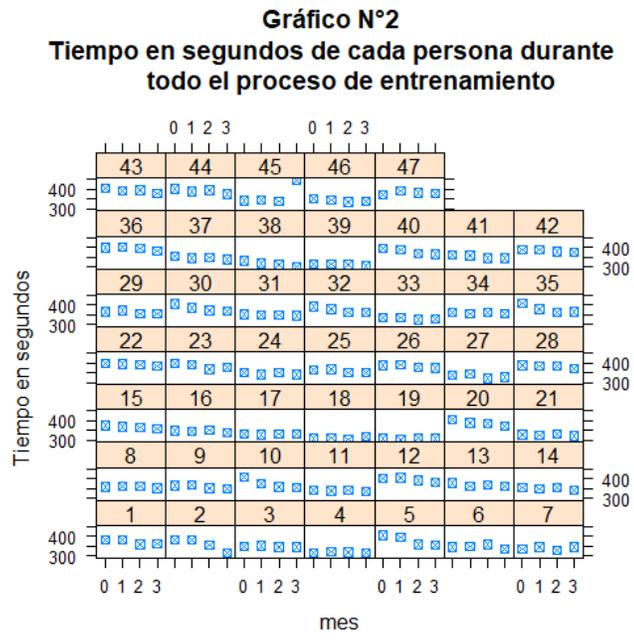
El entrenador en su momento, como parte de su proceso de toma de decisiones, observó en los datos recolectados que efectivamente los individuos expuestos a su entrenamiento mejoraron de manera general el desempeño en la prueba. Él se basó únicamente en los números que fue observando desde la primera prueba y hasta la última. No realizó algún tipo de análisis gráfico, mucho menos procedimientos estadísticos complejos. Por esta razón, en la siguiente sección se analiza si existe consistencia entre lo visto por el entrenador y la evidencia estadística resultante del análisis

gráfico exploratorio y los resultados producidos por la aplicación de un modelo mixto de medidas repetidas.



El gráfico N° 1 aporta evidencia para sospechar que existe una correlación entre las pendientes y los interceptos de las rectas correspondientes a cada individuo. Obsérvese cómo los interceptos más altos presentan pendientes decrecientes más pronunciadas. Además, conforme baja el valor del intercepto pareciera que las pendientes tienden a acercarse a un comportamiento constante. Esto sugiere que personas con tiempos iniciales muy altos tienden a mejorar su rendimiento más rápido que los que inician con tiempos muy bajos. Y es que es de esperar que personas que poseen desde el inicio rendimientos muy buenos, tiendan a mantenerse con un rendimiento similar al inicial.

Por otro lado, es posible apreciar que una de las rectas presenta un comportamiento muy distinto al de todas las demás. Si bien es cierto, algunas rectas tienen una pendiente creciente, pero ninguna es tan exageradamente pronunciada como el caso que se presenta en el gráfico N° 1. Antes de eliminar dicho caso del análisis es mejor evaluar qué puede estar sucediendo. El gráfico siguiente puede aportar más información al respecto.

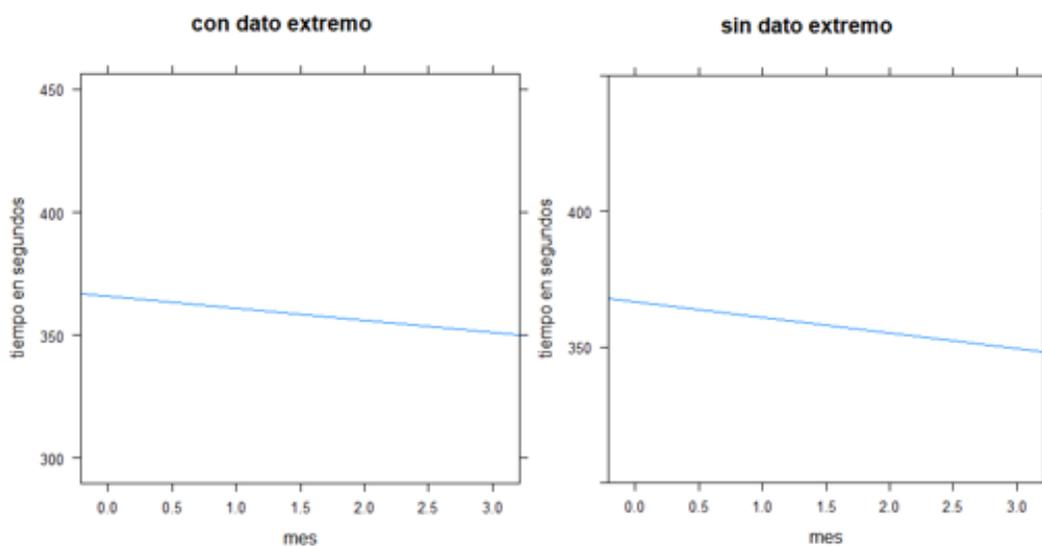


El gráfico N° 2 ilustra de forma puntual las repeticiones hechas por cada sujeto. Los sujetos están representados desde el número 1 hasta el 47, las figuras celestes denotan el tiempo en segundos en cada repetición mensual. De forma general es posible apreciar que prácticamente todos los sujetos presentan una tendencia a disminuir o mantener sus tiempos en segundos. Únicamente el sujeto 45 lleva una leve tendencia a la baja en las primeras tres repeticiones y de repente en la cuarta aumenta en exceso su tiempo. Evidentemente es un caso extremo.

A pesar que se lleva a cabo un análisis donde los sujetos son un componente aleatorio, para efectos del resto de los resultados se toma la decisión de excluir el caso extremo. Al realizar las pruebas de hipótesis propias de la metodología de las medidas repetidas, no hay diferencia entre dejar o quitar el caso extremo. Es decir, no influencia el rechazar o no una prueba, pero sí llega a afectar los intervalos de confianza que más adelante veremos. También afecta en algo a la recta de tendencia general. (Ver gráfico N° 3)

Gráfico N° 3

Rectas de tendencia general



En el gráfico N° 3 se aprecia un leve aumento del intercepto y del decrecimiento de la recta, esto al quitar el dato extremo. Ese leve cambio puede significar bastante para efectos de conclusiones y también para toma de decisiones por parte del entrenador. En este contexto una mejora de uno o dos segundos adicionales es nada despreciable. Por otro lado, al menos de una manera descriptiva, el gráfico de tendencia sugiere que el proceso de entrenamiento sí tiene un impacto en la mejoría de los tiempos de las pruebas de los individuos. Véase cómo al avanzar los meses, el tiempo en segundos decrece.

Modelo aplicado

Como se ha venido diciendo, el análisis estadístico se lleva a cabo apoyándose en la aplicación de un modelo mixto de medidas repetidas. En la sección de metodología se describe el modelo con todos los detalles. Aquí de forma general se recuerda que el modelo final utilizado es:

$$Y_i = \beta_0 + \beta_1 mes + \beta_{i,0} + \beta_{i,1} mes + \varepsilon_i$$

Lo primero que se hace es analizar si pudiese existir una correlación considerable entre pendientes e interceptos de las rectas de cada individuo. De entrada, el análisis gráfico sugiere que la correlación es fuerte y negativa. Al ajustar modelos lineales sencillos para cada sujeto y procediendo a extraer el valor correspondiente a cada intercepto y pendiente, se obtiene un coeficiente de correlación intercepto-pendiente de -0.92. Lo anterior coincide con la observación gráfica.

La razón por la cual se analiza si existe correlación intercepto-pendiente es porque se debe escoger entre dos modelos mixtos: uno en el cual la correlación esté presente y otro en la que no. Luego de ajustar los dos modelos, se comparan mediante una prueba de razón de verosimilitud. La hipótesis nula se encuentra a favor de que no existe una correlación intercepto-pendiente. Para este caso, al comparar las deviancias de los modelos, la diferencia entre estas es de 29.3 con un grado de libertad. Entonces la prueba tiene una probabilidad asociada muy cercana a cero, por lo que se rechaza la hipótesis nula y se continúa el análisis con un modelo que sí contempla correlación intercepto-pendiente.

El siguiente paso es verificar si es posible asumir que las pendientes correspondientes a las rectas de cada individuo son iguales. El procedimiento a seguir es ajustar un modelo donde se elimina la pendiente aleatoria. Después, comparar ese modelo contra el modelo que sí contempla correlación intercepto-pendiente. Se plantea la hipótesis siguiente:

$$H_0: \sigma_{\beta_1}^2 = 0$$

La hipótesis nula anterior significa que las pendientes de todos los sujetos son iguales. Al comparar las deviancias de los dos modelos se obtiene una diferencia entre ellas de 36.9 y dos grados de libertad. La probabilidad asociada es casi cero, por lo que se rechaza la hipótesis nula y se puede decir que las pendientes de los sujetos no son iguales. Se continúa el análisis con el modelo que contempla correlación intercepto-pendiente y que sí tiene el término correspondiente a pendiente aleatoria.

Evaluación de supuestos del modelo

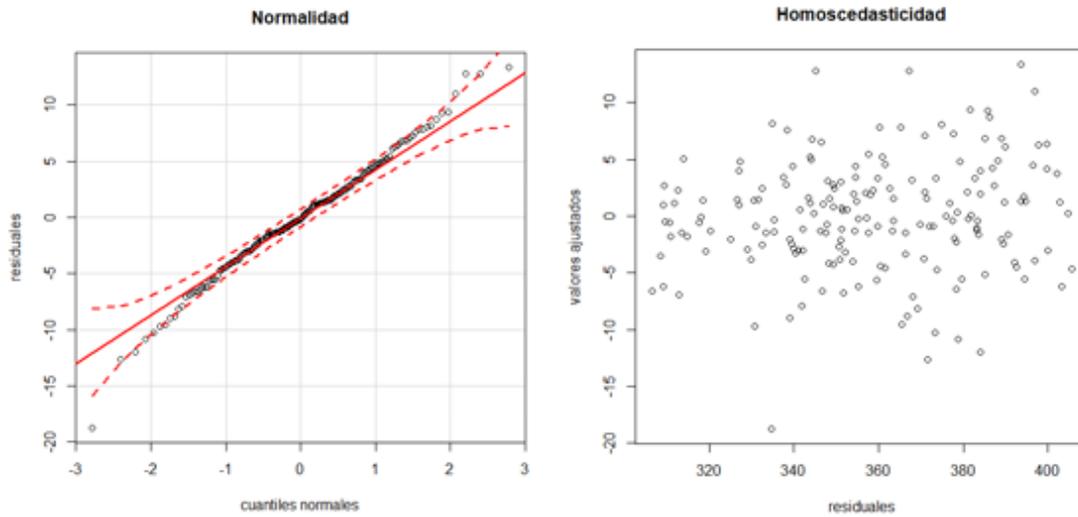
Después del proceso de selección del modelo mixto más adecuado, sigue la parte en la que se evalúan para ese modelo los supuestos de normalidad y homoscedasticidad.

Obsérvese en la figura N° 1 el lado izquierdo, los puntos correspondientes a los residuales contra los cuantiles normales tienden a estar dentro las líneas rojas discontinuas. Son muy pocos los casos donde dichos puntos se alejan del rango determinado por dichas líneas. Por esto, es posible asumir normalidad. Por otra parte, en el lado derecho los puntos correspondientes a los valores ajustados

contra los residuales (ambos del modelo), presentan una nube aleatoria, no se forma ninguna tendencia. Por lo anterior, se asume homoscedasticidad.

Figura N° 1

Evaluación de supuestos



Estimaciones

Hay que recordar que se eliminó el dato extremo de todo análisis en la parte de selección del modelo. Sin embargo, a continuación, se brinda una comparación de resultados de los intervalos de confianza de 95% cuando se incluye y se excluye el dato extremo.

Cuadro N° 1
Estimaciones por intervalo de los resultados
del modelo de medidas repetidas.

Estimaciones	Con dato extremo		Sin dato extremo	
	Inferior	Superior	Inferior	Superior
Desviación estándar intercepto aleatorio	26,35	41,41	26,03	40,24
Correlación intercepto-pendiente	-0,86	-0,52	-0,91	-0,59
Desviación estándar pendiente aleatoria	4,51	8,17	3,21	5,59
Error	7,64	10,18	5,54	7,41
Intercepto de la parte fija	360,63	380,73	362,61	382,04
Pendiente de la parte fija	-7,06	-2,82	-7,23	-4,13

Las primeras cuatro líneas del cuadro N° 1 corresponden a la parte aleatoria, mientras que las dos últimas a la parte fija. Obsérvese la correlación intercepto-pendiente, su intervalo presenta valores de correlación fuerte e inversa, en efecto esto es consistente con la decisión de haber utilizado un modelo donde la correlación fuese contemplada. En cuanto a la parte de la desviación estándar de las pendientes aleatorias de los individuos, se puede observar que al quitar el dato extremo, baja el

valor de la desviación. A pesar de esa baja, esos valores aportados por el intervalo de confianza confirman la decisión que se tomó durante el proceso de selección del modelo: dejar el término aleatorio de pendiente de individuo, debido a que existen diferencias entre esas pendientes. La cuarta fila llamada error hace referencia a la raíz cuadrada del cuadrado medio residual, obsérvese cómo dicha estimación se reduce considerablemente con tan solo haber quitado un valor extremo.

Ahora, refiriéndose a la parte fija del modelo, véase para el caso de la pendiente el cambio al quitar el dato extremo. En presencia del dato extremo, por cada aumento de un mes se da una reducción en el tiempo de la prueba física que se encuentra entre los 3 y 7 segundos. Mientras que, al quitar el dato extremo, la reducción del tiempo está entre los 4 y 7 segundos. El caso del intercepto denota que al realizar la primera prueba (o sea, en el mes cero) el tiempo en segundos se encuentra entre los 362 y 382 segundos.

CONCLUSIONES

Para el experto en el campo del fútbol es sumamente importante que aptitudes físicas como la potencia y velocidad de sus jugadores se encuentren en niveles aceptables, más aún que dichos jugadores pasan por un proceso que pretende llevarlos a competir en el exterior. El acondicionamiento físico aplicado por el entrenador -desde su propia óptica- da resultados positivos que indican que desde la primera repetición y hasta la cuarta, se mejora en promedio al menos 10 segundos. Lo anterior está apoyado solo en su propio criterio, sin ningún tipo de análisis estadístico. Esto no significa que se le deba restar importancia a la forma en la que este experto toma sus decisiones, pero agregar un poco de análisis estadístico puede adicionarle valor a su experiencia y método.

Después de haber analizado los datos aportados por el entrenador se puede decir que su ojo para calcular mejoras promedio en segundos es bastante acertado. Se vio como resultado de la aplicación del modelo de medidas repetidas que al pasar un mes, el tiempo disminuye en promedio entre 4 y 7 segundos. Eso es una medida de tendencia general, por lo tanto, no toma en cuenta algún individuo en particular. Todo lo anterior se puede traducir a que la mejora promedio de sus jugadores desde la primera y hasta la última repetición se encuentra entre los 12 y 21 segundos. Pareciera que efectivamente su metodología de entrenamiento tiene un impacto positivo en términos de rendimiento físico.

Otro punto importante que se puede tocar es el hecho de que en el análisis se detectara una correlación intercepto-pendiente tan alta (entre -0.91 y -0.59). Eso en términos estadísticos significa que el tiempo en segundos en la primera repetición y la pendiente (en este caso decreciente) de mes tienen una relación fuerte e inversa. Gracias a esa interpretación estadística y al análisis gráfico, es posible ver un detalle más profundo acerca de ciertas tendencias propias de los jugadores. Se evidencia que jugadores con tiempos iniciales (en la primera repetición) muy altos, presentan al pasar los meses mejoras enormes en sus tiempos (las pendientes decrecientes son muy pronunciadas). Por el contrario, los jugadores con los tiempos más bajos en la primera repetición tienden a mantener

tiempos muy similares en el resto de repeticiones. Un dato como este podría servir como apoyo al entrenador para así hacer trabajos físicos diferenciados.

BIBLIOGRAFÍA

Bangsbo, J. (2008). Entrenamiento de la condición física en el fútbol. Editorial Paidotribo.

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Lindquist, F., & Bangsbo, J. (1992). Comparison of various exercise tests with endurance performance during soccer in professional players.

Martin Dantas, E., García-Manso, J., Salum de Godoy, E., Sposito-Araujo, C., & Gomes, A. (2011). Aplicabilidad de los modelos de periodización del entrenamiento deportivo. Una revisión sistemática. *RICYDE. Revista Internacional De Ciencias Del Deporte*. Doi:10.5232/Ricyde, 6(20), 231-241. Recuperado de <https://www.cafyd.com/REVISTA/ojs/index.php/ricyde/article/view/316>

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Sarkar, Deepayan (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5

Comportamiento de la potencia de la prueba en un diseño anidado de dos etapas en el contexto de capacitación de vendedores en una cadena de almacenes

Susana García Calvo⁶
asgc3093@gmail.com

RESUMEN

En el presente estudio se realiza un análisis de potencia en un diseño anidado mediante un ejemplo contextualizado en recursos humanos de una cadena de almacenes. Para la situación de estudio se asume que se cuenta con tres almacenes (factor fijo). En cada uno de ellos labora una gran cantidad de capacitadores (factor aleatorio) entre los cuales se toma una muestra de forma aleatoria para que forme parte de un experimento, en este se desea determinar si los capacitadores tienen un efecto en el volumen de ventas que los agentes de venta entrenados logran al final del mes. Se definen cuatro escenarios extremos e intermedios, en los cuales se modifica la varianza de los efectos de los capacitadores sobre las ventas: (a) $\sigma^2_{\beta}=0$, (b) $\sigma^2_{\beta}=5$, (c) $\sigma^2_{\beta}=10$ y (d) $\sigma^2_{\beta}=20$. En cada uno de los escenarios se determina la proporción de veces en la que se detecta que el capacitador, como factor aleatorio, tiene un efecto en las ventas realizadas, variando en cada caso tanto el número de capacitadores como el número de agentes de venta por capacitador. Como principales hallazgos se tiene que en el escenario (a) cuanto mayor sea la cantidad de agentes de venta por capacitador mayor es la probabilidad de encontrar que la variabilidad es diferente de cero cuando realmente no es así, es decir, se incurre con mayor frecuencia en el error tipo I. Para el caso (b) y (c) se obtiene una potencia baja, especialmente cuando se tienen menos de 6 capacitadores por almacén con 5 agentes por capacitador. Por último, en el escenario (d), a diferencia de los dos escenarios anteriores, con solo el hecho de contar con más de cinco capacitadores por almacén y con más de tres agentes por capacitador, la potencia es superior a 0.75.

Palabras clave: potencia, diseño anidado, factor aleatorio, factor fijo, escenarios

ABSTRACT

In the current study a power analysis is conducted through a contextualized example within the Human Resources department of a storages network in a nested design. For the matter and subject of study it is recognized that there are three different storages (fixed factor). In each of the storages there are a great number of trainers (random factor), and a random sample was taken from it in order to make it part of an experiment; in such experiment, the aim is to determine if the trainers (mentors) have an effect or impact in the amount of sales that the trained sales agents report by the end of the month. Four extreme and intermediate scenarios are defined in which the variance of the effects of the mentors over the sales is modified: (a) $\sigma^2_{\beta}=0$, (b) $\sigma^2_{\beta}=5$, (c) $\sigma^2_{\beta}=10$ y (d) $\sigma^2_{\beta}=20$. In each of the scenarios, the proportion of times in which the mentor, as a random factor, has an effect over the reported sales; and in each of the cases, varying the number of mentors as much as the number of sales agents per each trainer. Within the main findings, it was found that scenario (a) as the bigger number of sales agents per trainer it has, the bigger probability of finding that the variability is different from zero when actually that is not correct, in other words; type I Error

⁶ Estudiante de estadística de la Universidad de Costa Rica



is incurred with greater frequency. For case (b) and (c) a lower power is obtained, especially in the cases where there are less than 6 trainers per storage with 5 agents per trainer. Lastly, scenario (d), opposite from the previous two scenarios, only with the fact of counting with more than five trainers per storage and with more than three agents per trainer, the rate is superior to 0.75.

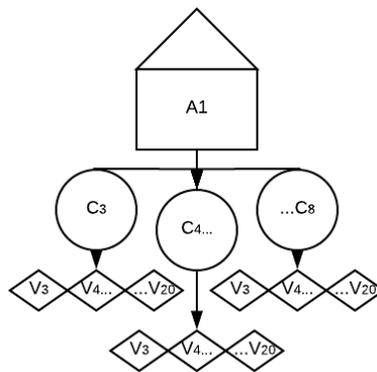
Keywords: power, nested design, random factor, fixed factor, type I Error

INTRODUCCIÓN

En un experimento, cuando se tienen dos o más factores, “A” y “B” y los niveles de uno de ellos están agrupados dentro de los niveles del otro factor, se define como un diseño anidado o jerárquico (Montgomery, 2002). Hay que tener en cuenta que si los niveles del factor “B”, por ejemplo, siempre son diferentes para los niveles de “A”, siendo “A” el nivel superior, entonces se trata de un diseño jerárquico (ver Figura 1) y no debe confundirse con un diseño factorial, donde los niveles de “B” son mismos para “A”.

Cuando se hace un análisis para este tipo de diseño, se tiene una prueba de hipótesis para cada uno de los factores, los cuales pueden ser fijos o aleatorios. Al tomar todos los valores posibles de un factor, se dice que este es fijo, y las conclusiones se dan específicamente para esos niveles. La hipótesis nula de interés para los factores fijos es que *los promedios de cada uno de sus niveles son iguales*, o lo que es equivalente, que *los efectos de todos los niveles sobre la variable de interés son cero*.

Figura 1
Estructura diseño anidado



Por otro lado, cuando se tiene un factor aleatorio se asume que los niveles que se están estudiando son solamente una muestra tomada aleatoriamente de un gran número de valores posibles, por tanto, las conclusiones sí podrían darse en términos generales del factor. Para los casos de factores aleatorios la hipótesis nula (H_0) que se prueba es la que establece *una variabilidad igual a cero entre los efectos de los niveles del factor respectivo sobre la variable de interés*.

En la mayoría de los casos, los investigadores realizan estudios previos al experimento que les permita tener una noción de los resultados que se podrían obtener bajo diferentes escenarios, por ejemplo, estudios de simulación, donde el interés es medir qué tan probable es rechazar la hipótesis bajo el escenario de que H_0 es falsa, este término se define como **potencia de la prueba** o como la probabilidad de no cometer el Error Tipo II. También se puede detectar la probabilidad de cometer el Error Tipo I, es decir rechazar H_0 cuando ésta es verdadera (Quezada, 2007; Quesada y Figuerola, 2010).

El concepto de potencia de la prueba también conocido como “poder estadístico”, surgió entre los años 1926 y 1933 atribuyendo su creación a Neyman y Pearson, sin embargo, no se dio una aplicación como tal del término hasta los años 60, cuando Jacob Cohen estableció primeramente la importancia de la potencia estadística dentro de la investigación experimental, y luego, dieron a conocer una serie de pasos a seguir para llevar a cabo un análisis de potencia (Bono, 1995).

Muchos investigadores se han dado cuenta que los análisis de potencia son de suma importancia ya que se podría evitar hacer una inversión en vano tanto económica como de tiempo, o, por el contrario, podría más bien inducir con una alta probabilidad de éxito de lograr un resultado deseado del experimento (Bono, 1995). También, es importante recalcar que según la potencia que tenga la prueba aumenta proporcionalmente conforme aumenta el tamaño de la muestra para el estudio, entonces, un investigador puede controlar el poder del estudio ajustando el tamaño de la muestra y viceversa (Suresh y Chandrashekara, 2012).

Por otro lado, un punto importante por el cual se decide abordar el tema de almacenes y sus capacitaciones en ventas, es porque hoy en día los departamentos de recursos humanos de importantes cadenas de almacenes, tienen como objetivo clave el capacitar en habilidades blandas y técnicas de venta a los empleados, de manera que el nivel de ganancias de la compañía sea cada vez mayor (Martínez, 2010). Bajo esta situación pueden surgir muchas inquietudes sobre quien es el profesional capacitador más adecuado para dejar en sus manos el entrenamiento de los vendedores, no obstante, esos cuestionamientos se pueden responder mediante diseños experimentales.

Dado lo anterior en este estudio se parte de un experimento hipotético que realiza el departamento de recursos humanos de una cadena de almacenes, la cual posee tres almacenes (factor fijo) donde cada seis meses se toma una muestra aleatoria de capacitadores en cada almacén para entrenar a un grupo agentes de venta con el fin de aumentar el volumen de ventas. El objetivo del jefe de recursos humanos es determinar si el capacitador como factor aleatorio tiene un efecto en las ventas que puedan lograr los agentes después del entrenamiento, para ello realiza un experimento con un diseño anidado, pero dado que no quiere incurrir en muchos gastos, desea saber cómo debe distribuir su muestra para responder a su pregunta de investigación correctamente.

Se decide desarrollar la simulación con un diseño jerárquico con el objetivo de analizar el comportamiento de la potencia de la prueba bajo la hipótesis nula de que el capacitador, como factor aleatorio, no tiene un efecto en las ventas promedio realizadas por los agentes de venta. Además, ya que las simulaciones pueden aportar información valiosa en la toma de decisiones para hacer un experimento de una u otra manera, y al no contarse con un amplio banco de ejemplos o trabajos realizados con estas características, también es motivación para realizar el análisis presente en este trabajo.

METODOLOGÍA

Los datos que se simulan tienen la siguiente estructura: almacenes, etapa superior y con efecto fijo; capacitadores, etapa intermedia y con efecto aleatorio (ver Figura 1).

Para la simulación se parte de que se tienen solamente tres almacenes de electrodomésticos en diferentes partes del país (A_1, A_2, A_3). La muestra de capacitadores se varía entre 3 y 8 (C_3, C_2, \dots, C_8), mientras que el número de agentes de ventas varía entre 3 y 20 (V_3, V_2, \dots, V_{20}). El rendimiento de cada capacitador se ve reflejado en el promedio de las ventas en colones realizadas por los vendedores de su grupo respectivo a final de mes.

Cabe resaltar que este diseño anidado que se plantea es balanceado, es decir, cada almacén tiene la misma cantidad de capacitadores y cada capacitador tiene la misma cantidad de agentes de venta a cargo.

En el software estadístico RStudio versión 3.4.2 (RStudio Team, 2016), utilizando como principales paquetes los siguientes: lme4 (Bates D. et al., 2015), nlme (Pinheiro J. et al., 2018), gridExtra (Baptiste A., 2017), dplyr (Hadley W. et al., 2019), se construye una simulación donde se generan quinientas muestras aleatorias dados los parámetros definidos en la ecuación 1. Los efectos de los capacitadores no se especifican en la ecuación 1 debido a que dependen del número de capacitadores que se definan, lo cual se va a variar, así como el número de réplicas (agentes) por capacitador. Además, dado que el capacitador es un factor aleatorio, los efectos se generan a partir de una distribución normal con media 10 y varianzas 0, 5, 10 y 20, las cuales definen los diferentes escenarios.

$$y_{ijk} = 70\,000 + 23\,000A_2 - 54\,300A_3 + \beta_{j(i)}C_{j(i)} + \varepsilon_{(ij)} \quad (1)$$

Donde:

- $i = 1, 2, 3$
- $j = 1, \dots, V_k$

De esta manera se reproduce 500 veces la prueba $H_0: \sigma^2_{\beta} = 0$ para los cuatro escenarios, para así calcular la proporción de rechazo de H_0 y determinar la potencia que tiene la prueba de los efectos de los capacitadores.

Para generar los datos primeramente se creó la matriz de estructura correspondiente al diseño anidado, luego se genera la variable respuesta “ventas” a partir de números aleatorios provenientes de una distribución normal con la media que define el modelo de la ecuación (1) y una variancia (σ^2) de 5000.

Para contrastar H_0 : *el capacitador como un factor aleatorio no tiene un efecto sobre el promedio de ventas alcanzado por los agentes*, se genera un estadístico de prueba F, el cual se construye a partir de los cuadrados medios esperados tomando como base la siguiente tabla:

Tabla 1.
Cuadrados medios esperados en un diseño anidado de tres etapas

Fuente	A fijo, B y C aleatorios
Factor A	$\sigma^2 + bcn \frac{\sum \alpha_i^2}{a-1} + cn\sigma_{\beta}^2 + n\sigma_{\alpha}^2$
Factor B (dentro de A)	$\sigma^2 + cn\sigma_{\beta}^2 + n\sigma_{\alpha}^2$
Factor C (dentro de B)	$\sigma^2 + n\sigma_{\alpha}^2$
Residual	σ^2

Fuente: se tomó como base la tabla 13.1 del Montgomery (2002) p.559

La fórmula para el cálculo de la F es la siguiente:

$$F_{B(A)} = \frac{CMB(A)}{CMC(B)}$$

Los cuadrados medio fueron extraídos del ANOVA del modelo, y con ello se calcula la probabilidad asociada, la cual se contrasta con una significancia de 0.05 para determinar si se rechaza o no la hipótesis nula. Una vez realizados los cálculos anteriores se calcula la potencia para analizar su comportamiento.

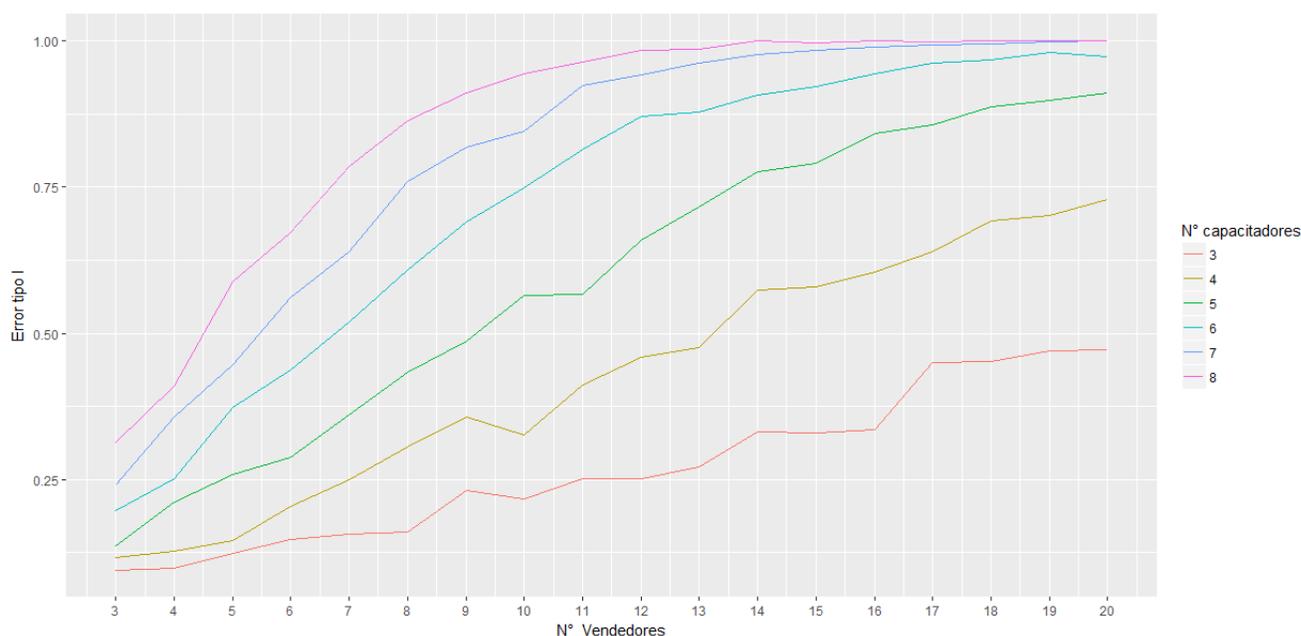
RESULTADOS

En el primer escenario se asume que el capacitador como factor aleatorio no tiene un efecto en las ventas que puedan lograr los vendedores, es decir que la variabilidad de los efectos del instructor en las ventas logradas dentro de cada almacén es cero.

Lo correcto en este caso sería no rechazar la hipótesis nula, sin embargo, como se muestra en el Gráfico 1, al aumentar el número de vendedores por capacitador (número réplicas) hay una tendencia creciente de la probabilidad de rechazar la hipótesis nula a pesar de que ésta es verdadera (Error Tipo I).

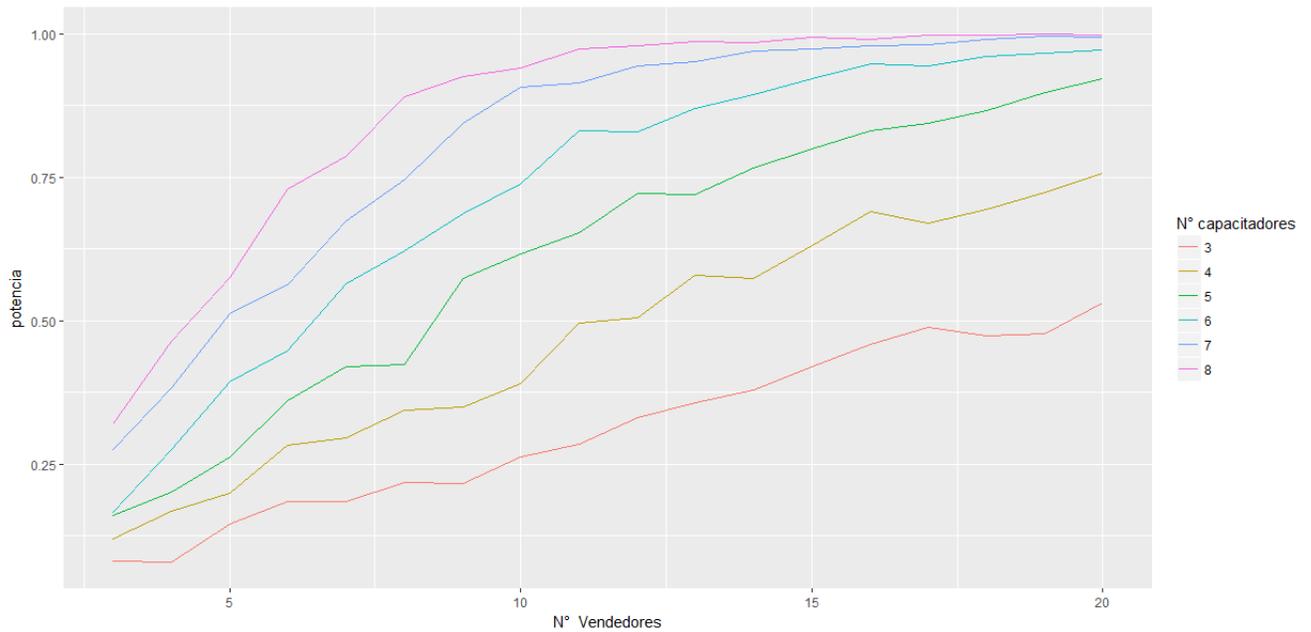
El comportamiento descrito es el mismo sin importar el número de capacitadores que se seleccionen, sin embargo, es claro que conforme mayor sea el número de capacitadores que se elijan, la probabilidad de cometer el Error Tipo I es mayor (ver Gráfico 1).

Gráfico 1. Error tipo I según número de capacitadores y vendedores, asumiendo que $\sigma_p^2 = 0$



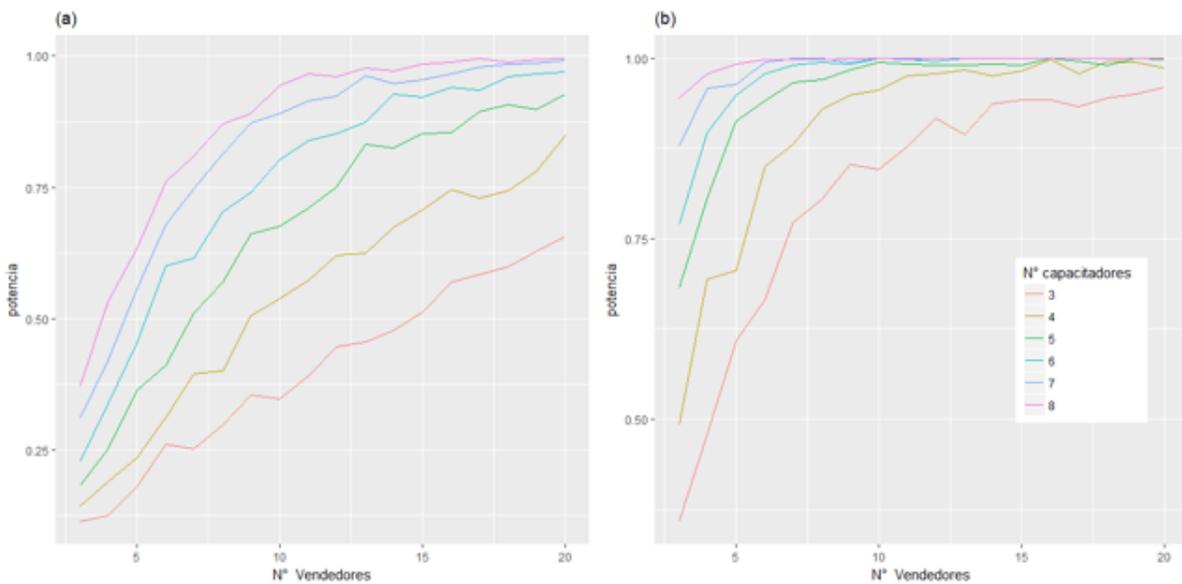
Por otro lado, si se asume poca variabilidad entre los efectos de los capacitadores sobre las ventas, como por ejemplo cinco, se observa un comportamiento similar al gráfico anterior. También se puede observar en el gráfico 2 que la potencia de la prueba es baja cuando el número de vendedores por almacén es inferior a 5, especialmente cuando se cuenta con menos de 6 capacitadores.

Gráfico 2. Potencia según número de capacitadores y vendedores, asumiendo que $\sigma_{\beta}^2 = 5$



En la Figura 3 se puede observar que, al aumentar la variabilidad de los efectos de los capacitadores sobre las ventas, la potencia de la prueba cuando la varianza es 20 crece a un ritmo más acelerado en comparación al caso donde la varianza es 10. De hecho, sin importar el número de vendedores, cuando se tienen más de cinco capacitadores por almacén, se tiene una potencia superior al 0.75, a es decir que en el 75% de los casos se tomó la decisión correcta de rechazar la hipótesis nula.

Figura 2. Potencia según número de capacitadores y vendedores, asumiendo que (a) $\sigma_{\beta}^2 = 10$ y (b) $\sigma_{\beta}^2 = 20$



Entonces, por ejemplo, si un almacén cuenta con 5 capacitadores y 10 vendedores la probabilidad de que el capacitador influya sobre las ventas logradas por los vendedores bajo su cargo, cuando realmente esta influencia sí existe, es de aproximadamente el 75%, es decir, 75 veces de 100 se tomó una decisión correcta.

CONCLUSIONES

En la práctica se debe ser precavido cuando se trabaja con tamaños de muestra grande, pues hay una tendencia a que las pruebas de hipótesis resulten significativas realmente no lo sean, es decir se puede llegar a decir que los capacitadores sí tienen un efecto en las ventas realizadas cuando realmente no es así, tal y como se mostró en el primer escenario.

Cabe resaltar que la conclusión anterior aplica también si se parte de que los capacitadores son factores fijos, por ende, si este fuera el caso, con tamaños de muestra grandes sería mejor verificar si los intervalos de confianza de las diferencias de medias de las ventas de los capacitadores incluyen al *delta*, es decir la diferencia a partir de la cual se considera que las discrepancias son relevantes, esto para cada uno de los almacenes.

Tener una alta probabilidad de cometer el Error Tipo I, cuando se cuentan con muchos vendedores por capacitador, puede incidir en una mala toma de decisiones, especialmente si se toman a los capacitadores como factores fijos, ya que puede que se recompense con dinero a ciertos capacitadores considerados como mejores, cuando realmente no es así.

Por ende, si a partir de alguna experiencia previa se sabe que los capacitadores no influyen en las ventas que puedan alcanzar los vendedores, se recomienda trabajar con una muestra máxima de 11 vendedores por almacén, distribuidos equitativamente en 3 capacitadores, para que la probabilidad de cometer el Error Tipo I no supere el 20% (Ver Anexo 1).

Si se conoce que los capacitadores sí tienen un efecto sobre las ventas de los vendedores, y se sabe que la variabilidad es de una magnitud importante, no es necesario tener muchas réplicas por almacén, siempre y cuando se tengan más de 4 capacitadores, ya que bajo estas condiciones se tendría una potencia superior a 0.75. Lo anterior es importante ya que para el experimento los almacenes se podrían ahorrar costos, pues en lugar de tener a una gran cantidad de vendedores en capacitación, estos estarían generando ganancias.

BIBLIOGRAFÍA

Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. <https://CRAN.R-project.org/package=gridExtra>

Bono, R. (1995). Consideraciones generales en torno a los estudios de potencia. *Anales de psicología*, 11(2), 193.

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2019). dplyr: A Grammar of Data Manipulation. R package versión 0.8.0.1. <https://CRAN.R-project.org/package=dplyr>

- Martínez, G. D. J. (2010). Capacitación de vendedores. *Contribuciones a la Economía*, (2010-11).
- Montgomery, D. (2002). Diseño y análisis de experimentos. *Limusa Wiley, Segunda Edición, México*.
- Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2018). *_nlme: Linear and Nonlinear Mixed Effects Models_*. R package version 3.1-137, <URL: <https://CRAN.R-project.org/package=nlme>>.
- Quesada, J., & Figuerola, J. (2010). Potencia de una prueba estadística: aplicación e interpretación en ecología del comportamiento. *Temas actuales de Etología*, 22, 19-37.
- Quezada, C. (2007). Potencia estadística, sensibilidad y tamaño de efecto: ¿un nuevo canon para la investigación? *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile*, (16), 159-170.
- Suresh, K. P., & Chandrashekhara, S. (2012). Sample size estimation and power analysis for clinical research studies. *Journal of human reproductive sciences*, 5(1), 7.

ANEXOS

Anexo 1. Combinaciones del número de capacitadores y vendedores para que la probabilidad de error tipo I sea inferior o igual a 0.2.

N° Capacitadores	N° Vendedores	Error Tipo I
5	3	0.09
4	3	0.10
3	3	0.11
3	5	0.11
3	4	0.12
3	8	0.12
3	9	0.12
4	4	0.15
4	5	0.15
6	3	0.15
3	11	0.19
4	6	0.19
7	3	0.20

Análisis del efecto de la variabilidad de los elementos de estudio en la potencia mediante la simulación de datos para un diseño de medidas repetidas

Ericka Valerio Salas⁷
erickavs189@gmail.com

RESUMEN

Con este estudio se pretende analizar el efecto de la variabilidad presente en las unidades de estudio y el tamaño de muestra respecto a la potencia mediante la simulación de datos y el planteamiento de los diferentes escenarios. Así mismo busca establecer criterios de decisión para una potencia deseada en caso de que se desee realizar un nuevo experimento similar al que se ha realizado previamente. Para esto se utilizó el material facilitado en la clase de Diseños Experimentales Aplicados impartido por la Escuela de Estadística de la Universidad de Costa Rica. Se utilizaron diferentes variancias para el error del modelo (0.1, 1, 1.5, 2, 2.5) y diferentes variancias para las pendientes de los sujetos de cada sexo (0.1, 1, 5). Así mismo se crearon bases con diferente tamaño de muestra (15, 30 y 100) y se utilizaron diferentes cantidades de repeticiones para las mediciones (4, 6, 8 y 10) para analizar el impacto que esto podría tener en la potencia. Entre los principales resultados se obtuvo que, para este caso específico, la cantidad de sujetos tomados en cuenta en la realización del experimento resultó no afectar la potencia para ninguno de los escenarios. Al aumentar la variabilidad entre sujetos del mismo sexo aumenta la potencia, esta relación se pudo apreciar de manera inversa con variabilidad de los errores o la variabilidad general entre sujetos donde se aprecia que al tener una menor variabilidad la potencia era mayor. Se encontró que a mayor número de repeticiones para un mismo período de tiempo aumenta la probabilidad de no rechazar la hipótesis nula cuando ésta es cierta.

Palabras clave: Medidas repetidas, simulaciones, potencia.

ABSTRACT

This study pretends to analyze the effect of the variability presented in the study units and the sample size regards to the power within de data simulation on the setting of different scenarios. Also, it looks for the establishing of decision criteria for a wanted power in case of a new similar experiment has to be reproduced. For this, data provided for the Statistics School's Applied Experimental Design class was used. The setting of diverse error variances (0.1, 1, 1.5, 2, 2.5) and the subject's incline variances (0.1, 1, 5) was used for the model in the analysis. Likewise, for the creation of data frames, different sample size (15, 30, 100) and repeated measures on the model (4, 6, 8, 10) was taken on consideration to the purpose of analyzing the impact in the power with the multiple scenarios. Within the main

⁷ Estudiante de estadística de la Universidad de Costa Rica



results, for each case, the number of subjects taking in consideration for the experiment didn't show a different power tendency. If the variance between the same sex subjects is raised, the power rose as well. Within the error variance, an opposite relation was observed, lower variances shown a higher power. The bigger amount of repeated measures in a same period, the bigger probability of no rejection of the null hypothesis, when this is true, was found.

Keywords: Repeated measures, simulations, power

INTRODUCCIÓN

El proceso de estudiar un fenómeno de interés usualmente incluye tener que hacer una serie de supuestos acerca de cómo trabaja, los cuales toman la forma de relaciones lógicas o matemáticas, que constituyen un modelo o una simulación del funcionamiento del fenómeno de interés (Law & Kelton, 1999). Gracias a la capacidad de las computadoras modernas, ahora se pueden explorar escenarios de comportamientos de datos. Controlando las características generales de poblaciones y de muestras, se pueden simular diferentes aproximaciones a un mismo problema.

Dentro de la estadística, para la generación de modelos que puedan representar el comportamiento de una población, las simulaciones entran a ser de particular interés ya que las fuentes de aleatoriedad se pueden representar con distribuciones de probabilidad (Law & McComas, 1999). Así ayudan a evaluar cómo las muestras que se pueden encontrar o medir, pueden afectar en las estimaciones de parámetros poblacionales. Esto se hace posible ya que se pueden extraer grandes cantidades de muestras derivadas de distribuciones con parámetros definidos por un experto, lo que lleva a la posibilidad de tener un ambiente controlado. Debido a que se conoce el escenario real, o el resultado al que se quiere llegar con la estimación, se puede observar cómo, al hacer variaciones en los parámetros que simulan las posibles muestras, las respuestas al problema planteado pueden tener diferentes resultados.

Entre los escenarios posibles, se puede hacer un ambiente que siga las características de un Diseño de Medidas Repetidas. Este diseño se enfoca en observar secuencialmente la misma variable respuesta en el paso del tiempo. Esta estrategia se sigue con el fin de reducir la variabilidad del error, ya que los efectos de los tratamientos se analizan a través de la respuesta media de los sujetos en los diferentes tratamientos (Arnau, 2007). La importancia de utilizar este diseño radica en que al realizar mediciones repetidas se pueden describir, pronosticar y explicar los procesos (ya sean biológicos, psicológicos, sociales u otros) que se producen como consecuencia del paso del tiempo (Fernández et al., 2007).

En el análisis de datos con una estructura de medidas repetidas, se plantea una hipótesis nula donde se afirma que el efecto que tiene la variable independiente sobre la respuesta promedio es el mismo para todos los sujetos, es decir, que la pendiente es la misma en todos los casos. Si se define una característica categórica de interés, por ejemplo, el sexo, que tiene dos categorías, se puede estudiar un ambiente que siga el diseño de medidas repetidas.

Estas pruebas permiten calcular la potencia, la cual se define como la probabilidad de aceptación de la hipótesis alternativa cuando es verdadera. La estimación de la potencia para una hipótesis dada depende de varios factores, entre ellos el tamaño de muestra, ya que según la teoría, a mayor tamaño de muestra, mayor será la potencia, el nivel de significancia pues al aumentar la probabilidad de cometer el error Tipo I y el tamaño del efecto del fenómeno de interés, pues al definir un tamaño del efecto más grande también resultará en una potencia mayor, (Bono, 1995).

Esta investigación se concentra en la modelación para un caso de diseño de medidas repetidas, con el fin de evaluar de qué forma el tamaño del efecto, la variancia de las pendientes, la variancia del error, la cantidad de sujetos y la cantidad de repeticiones en un mismo período, podrían afectar la potencia para la prueba de la hipótesis sobre la interacción.

METODOLOGÍA

Para la reproducción de datos, se utilizó a modo de prueba piloto los datos correspondientes al experimento denominado “Ortodoncia”, laboratorio impartido en la clase de Diseños Experimentales Aplicados, de la Escuela de Estadística de la Universidad de Costa Rica a cargo del profesor Ricardo Alvarado. En dicho ejercicio se cuenta con una base de datos recolectados por la Escuela de Odontología de la Universidad de Carolina del Norte, esta cuenta con observaciones de 27 niños a los cuales se les dio seguimiento desde los 8 a los 14 años de edad, midiendo la distancia que existe entre la pituitaria y la escotadura pterigomaxilar mediante rayos X con el fin de analizar si la tasa de crecimiento varía entre sexos.

Esta base cuenta con 4 variables:

- **Distancia:** la distancia entre la pituitaria y la escotadura pterigomaxilar, medida en milímetros.
- **Sujeto:** el número de sujeto al que corresponde la medida.
- **Sexo:** El sexo del sujeto al que corresponde la medida.
- **Año:** Año en el que fue medida la distancia.

Debido a los objetivos de la simulación, el análisis de la potencia debe ser limitado a una única prueba, se decidió analizar la significancia de la interacción entre las variables sexo y edad. La hipótesis nula de la prueba, que corresponde a la ausencia de interacción entre sexo y tiempo, se realizará mediante un análisis de variancia comparando un modelo con interacción con uno sin esta, y mediante la probabilidad asociada a una Chi-Cuadrado en el estadístico asociado a la razón de verosimilitud de ambos modelos. Cabe destacar que para el planteamiento del modelo se asume la no correlación entre las pendientes y los interceptos, lo cual quiere decir que se asume que no hay relación entre la distancia a los ocho años y la pendiente de cada sujeto. Además, se asume que las pendientes no son iguales entre niños de un mismo sexo, esto debido a que se desea analizar escenarios en los cuales exista mucha variabilidad entre sujetos del mismo sexo.

El modelo planteado es el siguiente:

$$y_{ij} = \beta_0 + \beta_1 Edad + \tau_j + \tau_j^* * Edad + \beta_{0,i} + \beta_{1,i} Edad + \varepsilon_{ij}$$

3. Donde i es el sujeto y j es el sexo correspondiente.

Mediante el planteamiento del modelo anterior y utilizando los datos brindados en el laboratorio, se obtuvieron los siguientes coeficientes del modelo:

Tabla 1. Coeficientes del modelo con los datos “Ortodoncia”

Coficiente de Intercepto (β_0)	16.857
Coficiente de Edad (β_1)	0.632
Coficiente de interacción entre sexo y edad ($\tau_j^* * Edad$)	-0.152
Variancia del intercepto del sujeto (efecto aleatorio)	2.249

Los datos presentados anteriormente fueron utilizados para la generación de las muestras mediante simulación, donde los coeficientes del efecto aleatorio (sujetos) serían generados aleatoriamente; para el intercepto del sujeto se fijó la variancia en 2.249, para el intercepto de la edad para cada sujeto y se utilizaron diferentes variancias (0.1, 1 y 5) y se generaron valores provenientes de una distribución normal con media 0 y alguna de las variancias establecidas. Esto permitió modificar entre los interceptos de los sujetos del mismo sexo. También fueron generados aleatoriamente valores para hacer variar las pendientes, con una distribución normal con media 0 y variancias en el conjunto 0.5, 1, 1.5, 2 y 2.5. Esto agrega variabilidad a las pendientes de los sujetos indiferentemente del sexo. Se utilizó una cantidad de repeticiones de 4, 6, 8 y 10 en el período de 8 a 14 años y una cantidad de sujetos de cada sexo de 15, 30 y 100.

Para definir los efectos de la variable sexo, se optó por plantear dos escenarios: uno donde la magnitud del efecto del sexo no fuera muy grande (0.5) y otro donde la magnitud fuera mayor (5), a esta distancia entre el promedio de distancias para sexo y edad se le llamó “ δ ”.

Una vez obtenidos los datos simulados se procedió a comparar un modelo con interacción entre sexo y edad y otro sin esta y, mediante un análisis de variancia se pudo determinar si dicha interacción era significativa, utilizando la probabilidad asociada a la hipótesis de no interacción. Este proceso se repitió 500 veces, y con estas probabilidades asociadas se procedió a calcular la potencia al determinar la proporción de veces en que se rechaza la hipótesis nula, usando un nivel de significancia de 0.05.

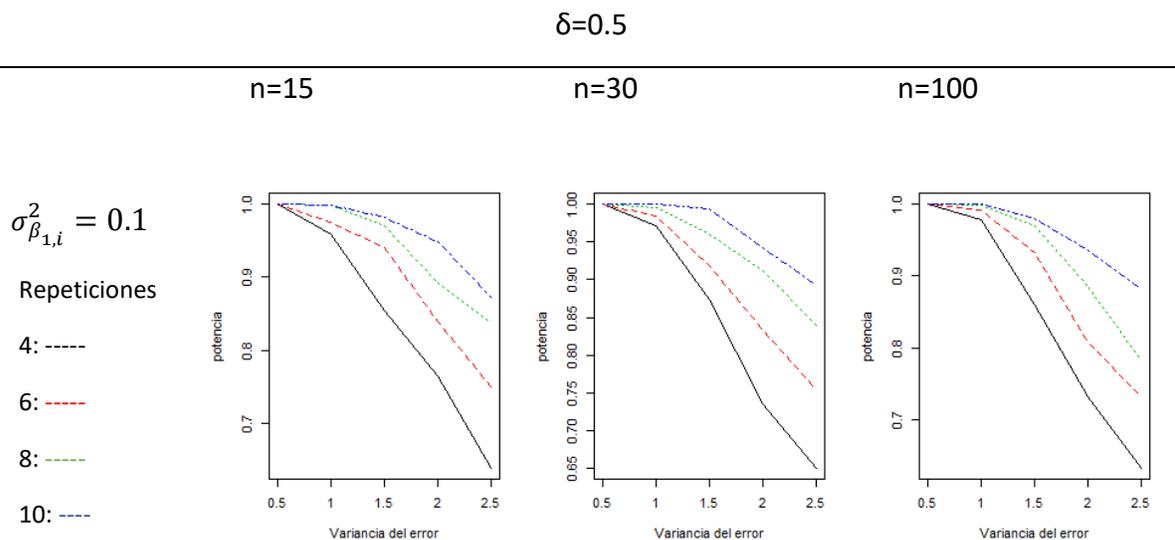
Se utilizó el software estadístico R (versión 3.5.2) (R Core Team, 2018) y fue necesaria la utilización de las librerías lme4 (Bates et al., 2015), lattice (Sarkar, 2008) y ggplot2 (Wickham, 2009)

RESULTADOS

Asumiendo un δ igual a 0.5, sería suficiente para determinar significativa la distancia entre promedios de la pendiente general de cada sexo, es decir, determinar que sí hay un efecto del sexo en la tasa de crecimiento, para el primer escenario del Figura 1., se puede apreciar que asumiendo que existe una baja variabilidad (0.1) entre las pendientes de los sujetos entre sexos, al aumentar la variancia del error, la potencia disminuye en diferente magnitud al aumentar la cantidad de mediciones realizadas a los niños y niñas de los 8 a los 14 años.

Las tendencias de la potencia para las diferentes cantidades de sujetos, dentro de cada variancia de la pendiente utilizada es muy similar indiferentemente del número de mediciones que se tomen en el periodo de 8 a 14 años, por lo cual se podría decir que es indiferente si se utilizan 15 niños o 100 niños de cada sexo para el experimento. Consistentemente la potencia se ve afectada al aumentar la variabilidad del error debido a que dicha variabilidad podría generar confusión en los resultados. (ver Figura 1.)

Figura 1. Potencia según variancia del error y cantidad de repeticiones realizadas.



$$\sigma_{\beta_{1,i}}^2 = 1$$

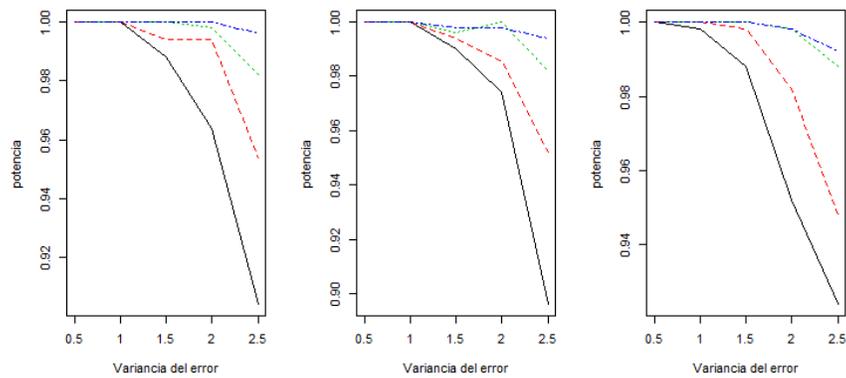
Repeticiones

4: ----

6: -.-.-

8: -.-.-

10: -.-.-



$$\sigma_{\beta_{1,i}}^2 = 5$$

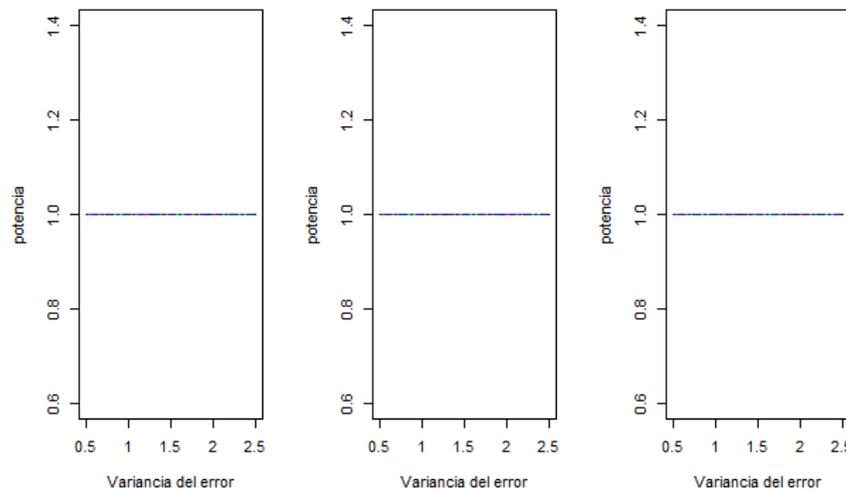
Repeticiones

4: ----

6: -.-.-

8: -.-.-

10: -.-.-

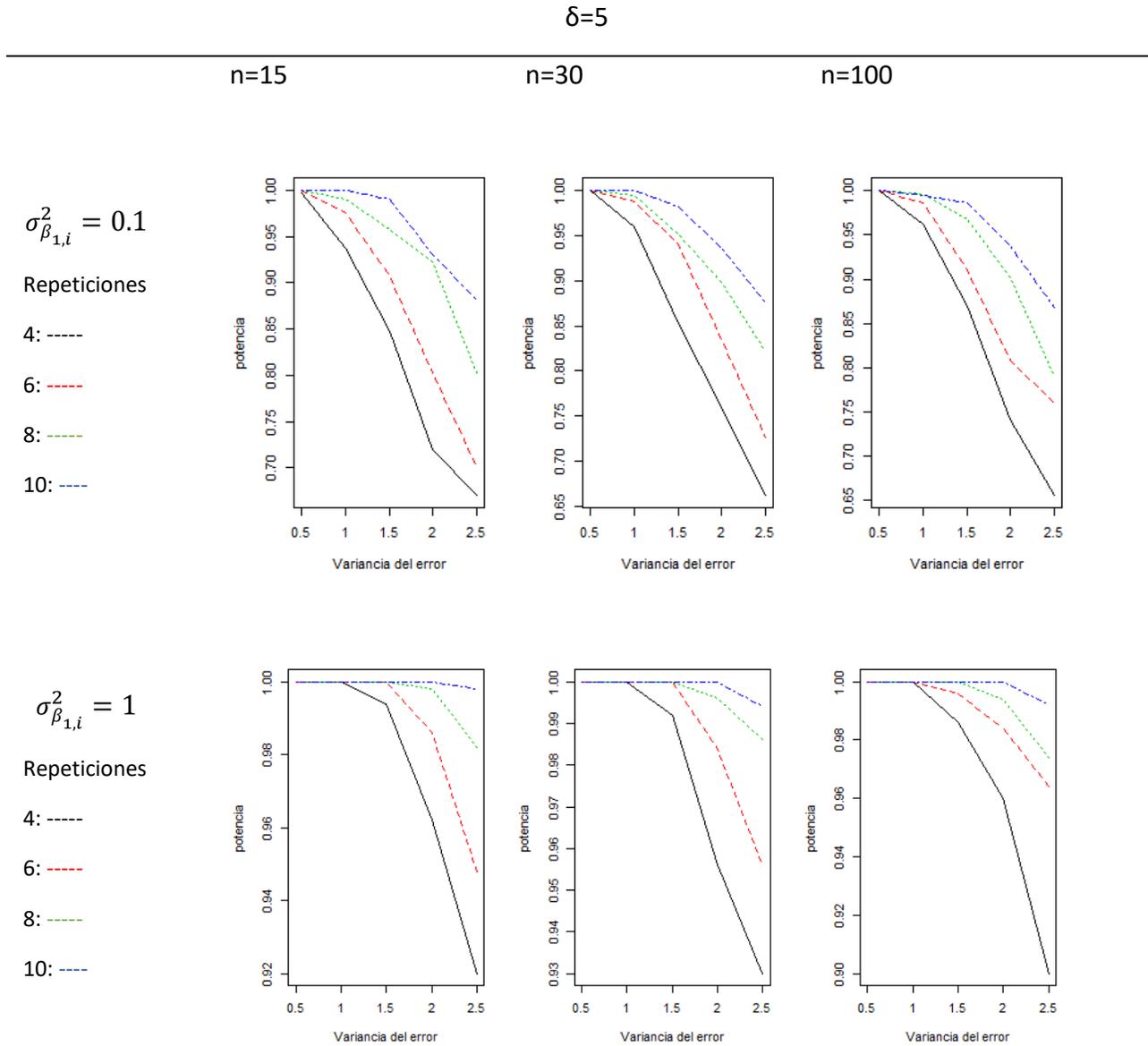


Así mismo, en la figura anterior se puede apreciar que, al aumentar la variabilidad de las pendientes entre niños del mismo sexo, la curvatura correspondiente a la potencia converge a 1, es decir que la probabilidad de rechazar la hipótesis nula cuando ésta es falsa tiende a 1 al aumentar la variabilidad de la pendiente, lo cual tiene sentido pues al haber mucha variabilidad entre sujetos del mismo sexo, la probabilidad asociada a la prueba respecto a la interacción del sexo con el sujeto debería ser casi nula. Esto porque esta variabilidad no permite observar diferencias entre sexos. Las tendencias entre la cantidad de sujetos utilizados en el experimento mantienen un patrón similar entre las variancias de las pendientes al igual que las diferencias entre potencias obtenidas para cada repetición simulada.

En la Figura 2, donde se determina que el δ debe ser mayor (5), se puede apreciar que en estos escenarios se presentan las mismas semejanzas entre tendencias respecto a la cantidad de sujetos utilizados, es decir que establecer un delta mayor no varió en gran magnitud las probabilidades de no

rechazar la interacción cuando ésta es significativa y para cada caso específico se alcanza una potencia similar indiferentemente del delta planteado.

Figura 2. Potencia según variancia del error y cantidad de repeticiones realizadas.



$$\sigma_{\beta_{1,i}}^2 = 5$$

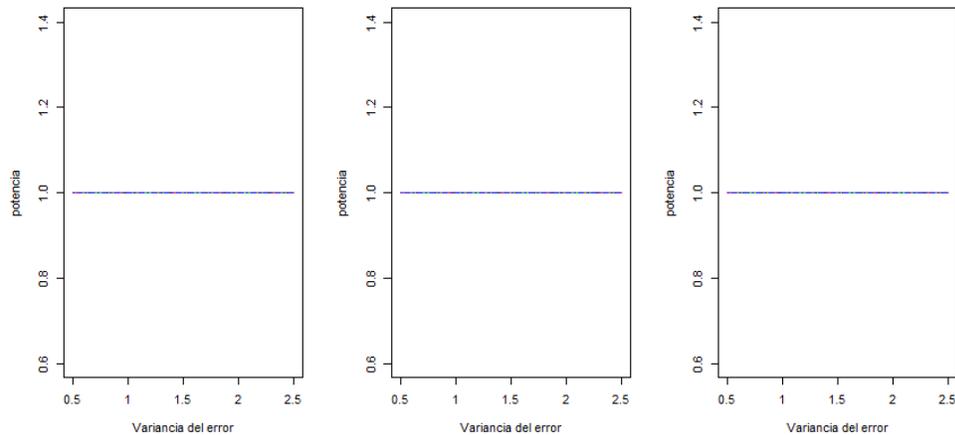
Repeticiones

4: - - - -

6: - - - -

8: - - - -

10: - - - -



CONCLUSIONES

De acuerdo a lo observado en este ejercicio de simulación, si un especialista decidiera realizar este experimento, se podría utilizar la menor cantidad de sujetos posible lo cual ahorraría costos en la recolección de la información, además de que si se presenta mucha variabilidad entre sujetos indiferentemente del sexo se espera lograr una potencia de 1, es decir que la probabilidad de no rechazar la hipótesis nula cuando ésta es falsa, es decir Error Tipo II es casi nula.

Así mismo, si los datos presentan una variabilidad de los residuos muy pequeña, es de esperarse que la potencia tienda a 1 indiferentemente de la variabilidad de las pendientes de los sujetos entre un mismo sexo, ya que esto genera confusión en los resultados de la interacción. Si la variabilidad de los residuos es alta se podría tomar en consideración la variabilidad existente entre los sujetos del mismo sexo si se tiene como meta una potencia específica.

Queda en evidencia que cantidad de sujetos no es tan influyente en la potencia como la cantidad de repeticiones utilizadas en este caso de Ortodoncia, ya que a mayor cantidad de repeticiones se observa un aumento en la potencia obtenida, por lo cual si se desea maximizar la potencia se recomienda considerar el presupuesto y los costos de lo implica que realizar más mediciones longitudinales.

Cuando existe mucha variabilidad entre sujetos es de considerar que esto podría afectar en los resultados en cuanto a rechazo o no rechazo de la prueba pues puede generar confusión, y podría no detectarse un efecto del sexo en la tasa de crecimiento cuando sí lo hubiese.

BIBLIOGRAFÍA

- Arnau-Gras, J. (2007). Estudios longitudinales de medidas repetidas. Modelos de diseño y de análisis. *Avances en medición*, 5, 9-26.
- Bono, R. (1995). Consideraciones generales en torno a los estudios de potencia. *Anales de psicología*, 11(2), 193.
- Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Fernández, P., Livacic-Rojas, P., & Vallejo, G. (2007). Cómo elegir la mejor prueba estadística para analizar un diseño de medidas repetidas. *International Journal of Clinical and Health Psychology*, 7(1).
- H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.
- Law, A & Kelton, D. (1999). *Simulation Modeling and Analysis*. 3ra edición. New York: McGraw-Hill
- Law, A & McComas, M. (1999). Expertfit: Total support for simulation input modeling. Recuperado de <http://delivery.acm.org/10.1145/330000/324224/p261-law.pdf?ip=163.178.19.201&id=324224&acc=ACTIVE%20SERVICE&key=842BC1E250410AEB%2EF2CED886D8E8C30C%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&acm=1552499565f2651c7feda70048e944552b065d6a15>
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Sarkar, Deepayan (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5

REFERENCIAS BIBLIOGRÁFICAS

Ceballos, Á. (2004) La escuela tradicional. Revisado el día 20 de diciembre del año 2018, del sitio web <https://profesorailianartiles.files.wordpress.com/2013/03/escuela-pedagc3b3gica-tradicional-vs-nueva.pdf>

IBM, K. C. (s.f.). Modelos lineales mixtos generalizados. Obtenido de IBM Knowledge Center: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/glimm/idh_glimm.html

Michael H. Kutner, C. J. (2004). Applied Linear Statistical Models.

Mogollón, O. y Solano, M. (2011) Escuela Activas: Apuestas para mejorar la calidad de la educación. Revisado el día 9 de enero del año 2018, del sitio web https://www.epdc.org/sites/default/files/documents/Active_Schools_Spanish.pdf

Montgomery, D. (2004). Diseño y análisis de experimentos. Limusa Wiley, Segunda Edición, México.

Rodríguez, J. (2013) Una mirada a la pedagogía tradicional y humanista. Presencia Universitaria. (No. 5), p. 39. Revisado el día 6 de diciembre del año 2018, del sitio web http://eprints.uanl.mx/3681/1/Una_mirada_a_la_pedagog%C3%ADa_tradicional_y_humanista.pdf

AGRADECIMIENTOS

Los integrantes del proyecto Revista Serengueti quieren agradecer de manera muy especial a todas las personas que de una u otra manera colaboración para llevar a cabo esta revista, tanto en la parte de asesoría como en el diseño.

Asesorías: Ricardo Alvarado B. (Profesor de Estadística), José Ureña H. (Tecnólogo de alimentos), Stefano Albertazzi (Entrenador de fútbol), Arturo Gutiérrez P. (Tecnólogo de alimentos), Ana Lucía Valerio Salas (Licenciada en Enseñanza del Inglés).

Diseño: José Aguilar U. (estudiante de Estadística y Comunicación Colectiva).

Sin el apoyo de todos ellos no hubiera sido posible terminar con éxito este proyecto tan importante.

