

Efecto en la potencia de la prueba e intervalos de confianza cuando se incumple el supuesto de normalidad

Angie María Rojas Venegas⁴, Jorge Arturo Solórzano Fallas⁴, Joseph Andrey Porras Rivas⁴
angierojas893@gmail.com, arturosolo12@gmail.com, joseph.porras.rivas@gmail.com

RESUMEN

El presente artículo estudia qué ocurre con la potencia de la prueba e intervalos de confianza en un diseño factorial con tres tratamientos, ante el incumplimiento del supuesto de normalidad en distribuciones como: Poisson, Uniforme, Binomial Negativa y la distribución real del ingreso por hogar de Costa Rica. Se plantea una simulación donde se evalúa el supuesto de normalidad, que obtenga los intervalos de confianza y la potencia de la prueba para las diferentes distribuciones planteadas. En los resultados se aprecia que conforme aumenta el tamaño de la muestra los intervalos de confianza de la normal se mantienen más estrechos en comparación con las no normales si se mantiene una varianza pequeña y constante. Por otro lado, la potencia de la prueba para la distribución normal es más alta en comparación con las no normales, no obstante, conforme aumenta el tamaño de muestra la potencia de la prueba es mayor en especial para la normal.

PALABRAS CLAVE: inferencia estadística, potencia, intervalos de confianza, datos reales, tratamientos, tamaño de muestra.

ABSTRACT

This article studies what happens with the power of the test and confidence intervals in a factorial design with three treatments, in the event of non-compliance with the assumption of normality in distributions such as: Poisson, Uniform, Negative Binomial and the real distribution of income per house of Costa Rica. A simulation is proposed where the assumption of normality is evaluated, which obtains the confidence intervals and the power of the test for the different distributions proposed.

The results show that as the sample size increases, the confidence intervals of the normal distribution remain narrower compared to the non-normal distributions if a small and constant variance is maintained.

On the other hand, the power of the test for the normal distribution it is higher compared to the non-normal distributions, nevertheless, as the sample size increases, the power of the test is higher, especially for the normal distribution.

KEYWORDS: statistical inference, power, confidence intervals, real data, treatments, sample size.

⁴Estudiante de la Universidad de Costa Rica

INTRODUCCIÓN

La estadística es un campo muy extenso en el cual se hacen investigaciones, donde se tiene como finalidad realizar inferencias estadísticas que ayuden a contestar distintas interrogantes que surgen en la vida diaria; según Fernández et al. (2016) la inferencia se entiende como una parte de la estadística, que estudia grandes grupos a partir de una pequeña parte de la población de estudio; la colección de individuos a analizar se llama población, y la parte que apoya este análisis se llama muestra. Técnicamente, el razonamiento implica proyectar las conclusiones obtenidas a todo el grupo una vez estudiada la muestra.

Esto debido a que muchos investigadores como Carl Friedrich Gauss crearon diferentes técnicas para hacer inferencias a la población a partir del análisis de datos, en especial Friedrich, quien formuló la ecuación de la curva, a lo que también se le conoce como la campana de Gauss y popularmente denominada como la distribución normal. Con respecto a lo anterior, Llinás (2017) menciona que la distribución normal es la distribución continua más utilizada en la aplicación de la teoría de la probabilidad. Constituye la base para el desarrollo de muchas teorías y métodos estadísticos.

Esta distribución se caracteriza por la simetría alrededor de la media, que es consistente con la mediana y otras características. La gran ventaja de la distribución normal es que nos permite calcular la probabilidad que tienen los datos en la distribución, por lo que podemos inferir los datos de la muestra a la población sabiendo que se distribuyen normalmente y con esto realizar un análisis fiable. Sin embargo, existen distribuciones que no son simétricas, según Martínez (2018) la asimetría se define como la falta o ausencia de simetría de la distribución con respecto a la media.

Según lo expuesto, este artículo revisa los intervalos de confianza y potencia de la prueba si se viola el supuesto de normalidad de los residuos. Al calcular estos intervalos de confianza para la estimación del parámetro desconocido en las distribuciones, se quiere aproximar un rango de valores en los cuales se pueda encontrar este parámetro.

Ahora bien, Puente (2018) sugiere que la estimación de intervalos tampoco permite conocer el valor exacto del parámetro desconocido, pero define un intervalo de confianza en el que ya existe un cierto nivel de confianza o el mismo valor, es decir, el parámetro está en el intervalo. Estimar por intervalo significa conocer el rango del parámetro desconocido de la población, pero siempre existe la probabilidad de que el intervalo no contenga el parámetro. El supuesto de probabilidad asume un nivel de incertidumbre llamado significancia.

Por otro lado, cuando se hacen análisis de datos generalmente el investigador se plantea una hipótesis nula y una hipótesis alternativa acerca de lo que desea indagar; no obstante, el investigador conoce que puede incurrir en dos tipos de error: por ejemplo, el tipo I, que es la probabilidad de rechazar la hipótesis nula siendo cierta o el tipo II, que consiste en no rechazar la hipótesis nula siendo falsa; dado que es importante conocer en cuál error se está incurriendo, existe una manera de conocer la probabilidad de que nuestra prueba detecte una diferencia si realmente existe, a esta prueba se le conoce con el nombre de potencia de la prueba y según Llinás (2018) la potencia de la prueba tiene las siguientes propiedades:

⁴Estudiante de la Universidad de Costa Rica

- a) Mucho más lejos se encuentre la media verdadera de la media planteada, mayor es la potencia de la prueba si los datos se mantienen iguales.
- b) A mayor nivel de significancia, menor potencia de la prueba, si los datos permanecen iguales.
- c) A mayor varianza de la población, menor es la potencia, si los datos se mantienen constantes.
- d) A mayor tamaño de muestra, mayor es la potencia de la prueba si los datos permanecen constantes.

A su vez el propósito de este estudio es exploratorio, por ello, se tiene como objetivo general analizar el efecto de la potencia de la prueba y los intervalos de confianza ante el incumplimiento del supuesto de normalidad, más específicamente, examinar el impacto que tienen las distribuciones no normales en los intervalos de confianza en comparación con las distribuciones normales, así como, comparar el efecto de la no normalidad en la potencia de la prueba ante distribuciones normales o no normales, y por último, contrastar la potencia de la prueba ante diferentes tamaños de muestra en las distintas distribuciones.

METODOLOGÍA

La aplicación del presente trabajo inicia con la generación de números aleatorios por medio de diferentes distribuciones conocidas, como lo son: la distribución Normal, la distribución Poisson, la distribución Binomial Negativa y la distribución Uniforme. Además, se genera una muestra de una distribución de datos reales, tomados de un archivo que contiene la encuesta de Ingresos y Gastos de Costa Rica, proporcionado por el M.Sc. Johnny Madrigal Pana, profesor del curso de Principios y Técnicas de Muestreo en el I Ciclo lectivo del año 2021; estos datos son utilizados en la creación de un modelo factorial simple de un factor con tres niveles, el cual se puede escribir de la siguiente forma:

$$\mu_j = \mu + \tau_j$$

Siendo:

- μ : corresponde a la media general
- μ_j : es la media del tratamiento j
- τ_j : es el efecto del j-ésimo tratamiento.

Cabe mencionar que la distribución de datos reales trae su propia media y varianza, por ende, a la hora de generar la muestra de datos, estos adoptan de manera análoga, una media y varianza específica para esta muestra recolectada. Para lograr que esta muestra se pueda adaptar a la media y varianza que se requiere en la simulación se realiza una transformación de la forma:

$$W = \frac{x - \mu_x}{\sigma_x} \cdot \sigma_y + \mu_y$$

Donde:

- x: Dato generado en la muestra.
- μ_x : Promedio de la muestra generada.
- σ_x : Desviación estándar de la muestra generada.
- σ_y : Desviación estándar deseada.
- μ_y : Promedio deseado.

Las distribuciones antes mencionadas tienen mucha importancia en el mundo estadístico y por eso fueron seleccionadas; estas tienen diferentes usos, como por ejemplo la Poisson, la cual es utilizada en conteos, por ejemplo, según Wackerly et al. (2008) la distribución de Poisson proporciona un buen modelo para la probabilidad del número Y de accidentes automovilísticos, industriales y otros tipos en una unidad de tiempo determinada; por otro lado, la Binomial Negativa no es tan conocida como la Poisson o la Normal, pero esta es muy utilizada por las aseguradoras para el cálculo de riesgos, según Martínez (2017) la distribución Binomial Negativa supone un modelo donde los asegurados no tienen el mismo riesgo de tener un siniestro, entonces se logra justificar el hecho de que a algunos clientes se les dé un descuento y a otros un recargo ya que esto está directamente relacionado a su riesgo. Por otro lado, la uniforme es una distribución sencilla pero muy utilizada en algunos campos de la vida cotidiana, como por ejemplo el lanzamiento de un dado o en juegos de azar.

El modelo factorial simple se crea con el fin de verificar la normalidad de cada una de las distribuciones, a su vez, se plantean los datos en función de un factor creado que contiene tres niveles, y se les extrae sus residuales, estos son ingresados al gráfico (QQPlots) para verificar la normalidad de cada una de las distribuciones generadas. La elaboración de estos gráficos se realiza por medio de la librería `car: An {R} Companion to Applied Regression` (Fox y Weisberg, 2019).

Por otro lado, se crean intervalos de confianza para las diferentes distribuciones, esto para analizar la amplitud de los intervalos para cada una de las distintas distribuciones planteadas, y observar si los intervalos contienen o no la diferencia real de las medias y el cero. Este análisis se realiza por medio de un gráfico generado con la librería `ggplot2: Elegant Graphics for Data Analysis` (Wickham, 2016), el cual permite visualizar la amplitud de los intervalos de confianza para cada una de las comparaciones de las medias.

Cabe destacar que dichos intervalos de confianza fueron creados por medio de la generación de números aleatorios, los cuales, se toman como variable respuesta para cada uno de los tratamientos creados; tanto la variable respuesta como el tratamiento simulado son ingresados a un modelo lineal para estimar el Cuadrado Medio Residual (CMRes) y los grados de libertad residuales.

Con respecto a lo anterior, el CMRes es utilizado para la estimación del error estándar y los grados de libertad para estimar el cuantil necesario para realizar el cálculo de los intervalos de confianza.

Una vez obtenidos estos valores se procede a la creación de los intervalos, los cuales tienen la forma:

$$LIM = dif \pm qt \cdot ee$$

Siendo:

LIM: La estimación del intervalo de confianza.

dif: La diferencia entre cada par de medias.

qt: El cuantil utilizado para la estimación del intervalo de confianza.

ee: El error estándar estimado para esas diferencias.

De manera análoga, se realiza la potencia de la prueba para cada una de las diferentes distribuciones; esta prueba se realiza con el fin de verificar por medio de la simulación, cuáles

distribuciones obtienen una potencia de la prueba mayor en igualdad de condiciones, es decir, igual cantidad de repeticiones y mismos promedios hipotéticos. Cabe mencionar, que se tiene un nivel de significancia constante, fijado en 0.05.

Dicha potencia es estimada por medio de un modelo lineal, tomando como variable dependiente los números generados aleatoriamente con cada una de las diferentes distribuciones y como variable predictora el tratamiento simulado. A este modelo se le realiza un análisis de varianza para lograr determinar si existen diferencias entre los tratamientos simulados, y por medio de la extracción y almacenamiento de la probabilidad asociada en cada una de las iteraciones del ciclo, se realiza el cálculo del promedio de las probabilidades asociadas que sean menores a la significancia planteada, el cual corresponde a la potencia de la prueba. Es importante recalcar que el usuario puede manipular las medias de los tres tratamientos planteados, así como la cantidad de réplicas por tratamiento, esto por medio de los cuatro deslizadores que aparecen en la parte izquierda de la aplicación. Cabe destacar que dicha simulación solo inicia en el momento en que se presione el botón “Generar” (ver Figura 3 en anexos).

Para la realización del presente artículo se requirió la utilización del programa estadístico R Core Team, 2021) en su versión 4.1.0, el cual, por medio del paquete Shiny: Web Application Framework for R (Chang et al. 2021), permite realizar aplicaciones interactivas. A su vez, la aplicación utiliza la librería shinythemes (Chang, 2021) para poder cambiar el aspecto de la interfaz de la aplicación. Además, se parte de la hipótesis de que los intervalos de confianza y la potencia de la prueba si se ven afectados por el incumplimiento del supuesto de normalidad.

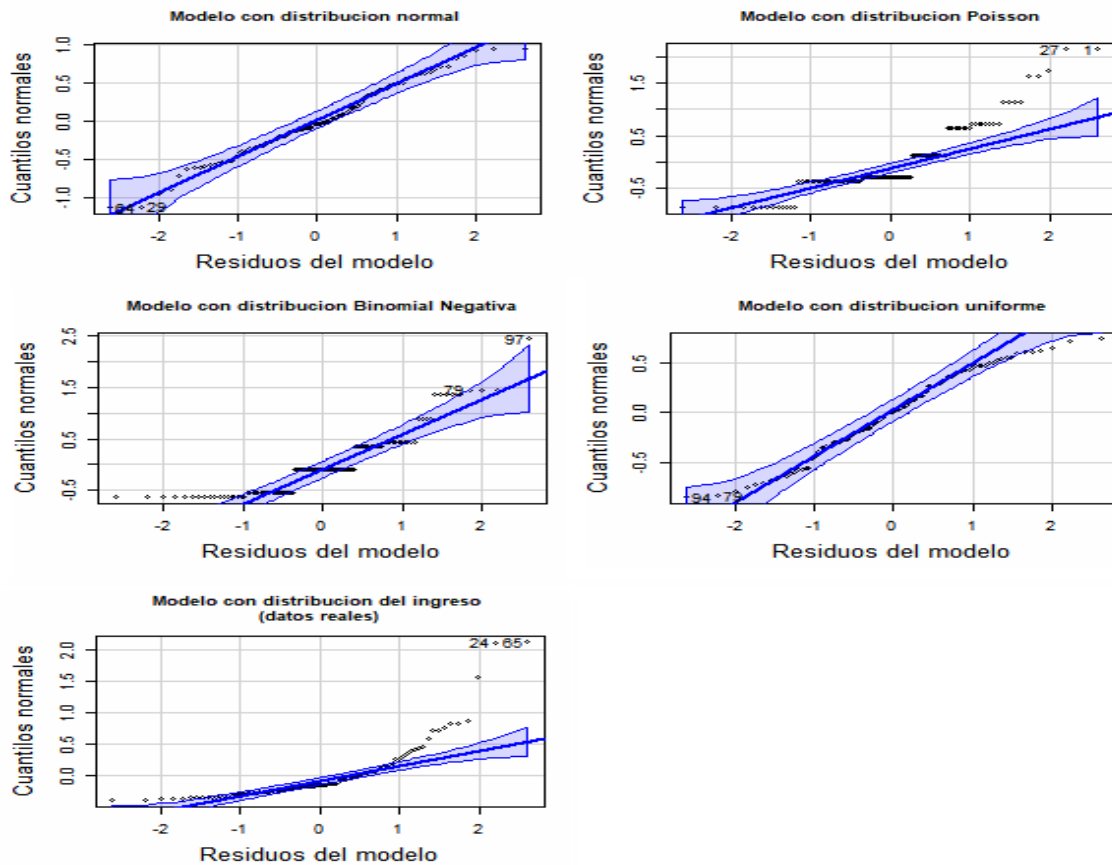
RESULTADOS

A continuación, se muestran los resultados de la simulación, como se puede observar en la Figura 1, al realizar el análisis gráfico para la prueba de hipótesis de normalidad de los residuos, se obtiene que para ciertas distribuciones los datos se alejan de las bandas de confianza, ejemplo de ello, Poisson, Uniforme, Binomial Negativa y la distribución de los datos reales del ingreso de Costa Rica.

La Figura 1 es una corrida de la simulación con 36 réplicas por tratamiento, se observa que con esta cantidad de réplicas la única distribución que mantiene el supuesto de normalidad de los residuos es la normal, con las demás distribuciones es evidente el incumplimiento de este supuesto.

Figura 1.

Verificación del supuesto de normalidad residual en modelos con diferentes distribuciones ante un mismo tamaño de réplicas por tratamiento.

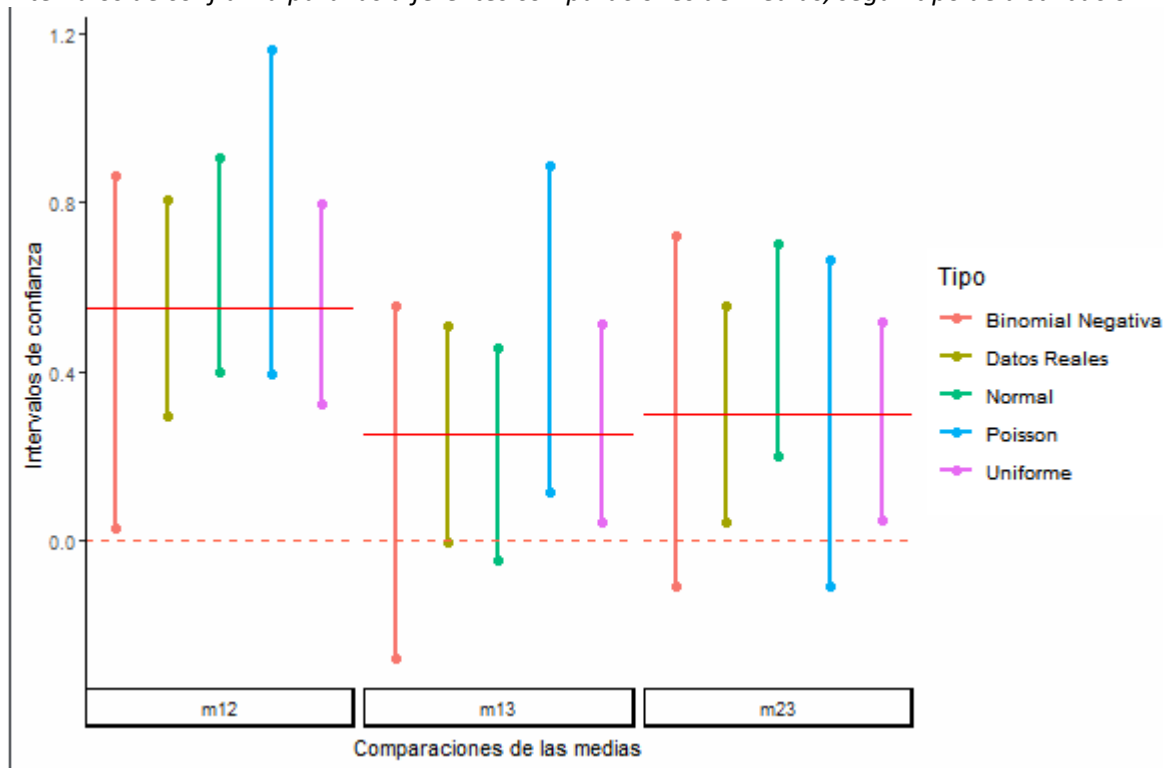


Además, al observar los intervalos de confianza en la Figura 2, los cuales fueron generados con las mismas 36 réplicas presentadas anteriormente, pero realizando la corrida con distintas medias hipotéticas, se tiene un valor de 0.75 en la primera media, 0.2 en la segunda media y 0.5 en la tercera media; se puede apreciar que todas las distribuciones presentan intervalos de confianza diferentes, en otras palabras, la amplitud del intervalo para todas las distribuciones es distinta. La línea roja continua representa la diferencia de medias reales para cada comparación y la línea roja punteada representa la posición del cero; además, en el eje de las abscisas se tienen las distintas comparaciones entre las tres medias, siendo m12 la comparación entre la media del tratamiento 1 y el tratamiento 2, m13 la comparación entre la media del tratamiento 1 y el tratamiento 3, y por último, m23 la comparación entre la media del tratamiento 2 y el tratamiento 3; también se puede diferenciar por color la distribución utilizada para la creación de cada intervalo de confianza.

⁴Estudiante de la Universidad de Costa Rica

Figura 2.

Intervalos de confianza para las diferentes comparaciones de medias, según tipo de distribución.



Por otro lado, la normal, la Uniforme y la distribución de datos reales del Ingreso producen intervalos de confianza similares, los cuales son más estrechos que los producidos por la Poisson o la Binomial Negativa, siendo esta última la que tiene los intervalos de confianza más amplios. Además, en todos los casos se puede observar que estos intervalos incluyen la diferencia real que existe entre el contraste de las tres medias y también que en algunos casos los intervalos de confianza incluyen el cero.

Al examinar la potencia de la prueba donde se mantiene una significancia del 0.05 constante, se puede apreciar en la Tabla 1 como la potencia de la prueba cambia por el incumplimiento del supuesto de normalidad en las distintas distribuciones.

Tabla 1.

Potencia de la prueba ante diferentes distribuciones, varianza constante, según cantidad de réplicas por tratamiento y promedios hipotéticos.

Significancia	Normal	Poisson	Binomial negativa	Uniforme	Datos reales
0.05	1.00	0.88	0.88	1.00	1.00

⁴Estudiante de la Universidad de Costa Rica

Las distribuciones que tienen la potencia más baja son la Poisson y la Binomial Negativa, mientras que la Normal, la Uniforme y la distribución del Ingreso en Costa Rica tienen la potencia más alta.

También se puede observar que, aunque la uniforme no cumple con el supuesto de normalidad, la potencia de esta es muy similar a la potencia de la normal, esto debido a que estas distribuciones son similares entre sí y la uniforme solo incumple el supuesto en las colas de la distribución.

CONCLUSIONES

En primer lugar, al aumentar la cantidad de réplicas por tratamiento un número considerable de veces, el supuesto de normalidad solo se cumple para la distribución normal y se incumple para las distribuciones Poisson, Binomial Negativa, Uniforme y la distribución real del Ingreso de Costa Rica; esta última es explicada porque Costa Rica tiene una distribución del ingreso asimétrica debido a que: el ingreso de los hogares no es repartido proporcionalmente, la mayor fuente proviene de los salarios, seguidamente del ingreso autónomo, las pensiones, remesas, el ingreso por renta de la propiedad, los subsidios estatales y las becas

El violar el supuesto de normalidad hace que las diferentes distribuciones tengan un impacto en los intervalos de confianza. En cuanto a las distribuciones que no cumplen con el supuesto de normalidad, los intervalos de confianza son más amplios en comparación con la distribución que cumple el supuesto, no obstante, todos los intervalos contienen la diferencia real entre las medias; esto es un indicador de que, si se repite muchas veces la estimación de los intervalos el 95% de las repeticiones, el intervalo va a contener el parámetro real. Con respecto a los intervalos que incluyen el cero, se puede decir que no se puede afirmar que exista una diferencia en la comparación de ese par de medias.

La potencia de la prueba sí se ve afectada por el incumplimiento del supuesto de normalidad; sin embargo, este factor no es lo único que influye, debido a que la distribución de datos también va a afectar este resultado, como lo es el caso de la distribución real del Ingreso de Costa Rica, en la cual, la potencia de la prueba es de uno o cero al cambiar las medias, por lo que se puede decir, que se van a encontrar diferencias siempre que estas existan; esto se puede evidenciar, ya que, una diferencia en los ingresos de Costa Rica se puede detectar de manera más sencilla y rápida.

Ahora bien, al aumentar la cantidad de réplicas por tratamiento, se observa que la potencia de las distribuciones utilizadas aumenta independientemente de si cumple el supuesto de normalidad o no, en el caso de que encuentre diferencias reales entre las medias; este resultado concuerda con la teoría donde Llinás (2018) menciona que la potencia de la prueba aumenta conforme aumenta el tamaño de muestra; pero en el caso en que las medias sean iguales, se van a obtener potencias bajas puesto que es más difícil para las pruebas detectar diferencias cuando estas no existen.

Por otro lado, no es recomendable utilizar distribuciones que no cumplan el supuesto de normalidad cuando se tiene un tamaño de muestra pequeño, dado que estas distribuciones tienden a tener intervalos de confianza muy amplios y potencias de la prueba más bajas; esto se puede

⁴Estudiante de la Universidad de Costa Rica

remediar por medio del aumento del tamaño de muestra, lo cual hace que estos intervalos de confianza se estrechen, y también logra aumentar la potencia de la prueba. Ahora bien, existen medidas remediales que se pueden realizar ante el incumplimiento de un supuesto, como lo es el supuesto de normalidad, pero se presenta la limitación que el estudio de estas medidas se sale del alcance de los objetivos planteados.

Además, se presentan limitaciones como: el tiempo para ampliar el conocimiento en el programa de Shiny y sus paquetes, por otro lado, al realizar muchas iteraciones el programa se vuelve más lento para generar los datos, además, el profesional debe de mantener una actualización constante de los paquetes y del programa, por último, este programa impide que la simulación pueda usarse para crear una aplicación web.

BIBLIOGRAFÍA

Chang, W. (2021). shinythemes: Themes for Shiny. R package version 1.2.0.

<https://CRAN.R-project.org/package=shinythemes>

Chang, W. Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A. and Borges, B. (2021). shiny: Web Application Framework for R. R package version 1.6.0.

<https://CRAN.R-project.org/package=shiny>

Fernández Palacín, F. Espejo Miranda, I. y López Sánchez, M. (2016). *Inferencia estadística: teoría y problemas (2a. ed.)*. Cádiz, Spain: Servicio de Publicaciones de la Universidad de Cádiz.

<https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/ereader/sibdi/33882?page=38>.

Fox, J. y Weisberg, S (2019). *car: An {R} Companion to Applied Regression*, Third Edition. Thousand Oaks CA: Sage. Recuperado de: <https://socialsciences.mcmaster.ca/jfox/Books/Companion>

Llinás Solano, H. (2017). *Estadística descriptiva y distribuciones de probabilidad*. Barranquilla, Colombia: Universidad del Norte. <https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/ereader/sibdi/70059?page=301>.

Llinás Solano, H. (2018). *Estadística inferencial*. Universidad del Norte. <https://www-digitaliapublishing-com.ezproxy.sibdi.ucr.ac.cr/a/54947?page=206>.

Martínez, A. (2017). *Análisis de las Primas de Riesgo en Seguros de Automóviles:*

Una Aplicación de los Modelos Lineales Generalizados. Universidad Autónoma Metropolitana.

http://mat.izt.uam.mx/mcmai/documentos/tesis/Gen.13-O/Martinez_Gutierrez_Adrian.pdf

Martínez, A. (2018, 4 de abril). Tipos de distribución. RStudio. Disponible en: https://rstudio-pubs-static.s3.amazonaws.com/377130_08851253a31b41d18c25fd08fad316c3.html

Puente Viedma, C. D. L. (2018). *Estadística descriptiva e inferencial*. Madrid, Spain: Ediciones IDT. <https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/ereader/sibdi/59931?page=161>.

⁴Estudiante de la Universidad de Costa Rica

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Wackerly, D., Mendenhall, W. y Scheaffer, R. (2008) Mathematical Statistics with Applications. 7th ed. Belmont, CA: Thomson Brooks/Cole, 2008. ISBN: 978-0-495-11081-1.

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

ANEXOS

Figura 3.

Parámetros de la simulación (deslizadores)

Normalidad

Generar

Cantidad de replicas por tratamiento:

6 36 45

6 12 18 24 30 36 42 45

Media 1

0 0.75 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Media 2

0 0.2 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Media 3

0 0.5 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1