

# EL USO DE LA HERRAMIENTA “ANÁLISIS DE DATOS” DE EXCEL COMO COMPLEMENTO PARA EL APRENDIZAJE DE LA ESTADÍSTICA EN EL AULA

Álvaro Toledo, Inés Vicencio

Departamento de Matemáticas y Física.  
Facultad de Ingeniería, Ciencia y Tecnología.  
Universidad Bernardo O’Higgins, Chile.  
alvaro.toledo@ubo.cl , ines.vicencio@ubo.cl

| RESUMEN   | ABSTRACT   |
|---|--|
| <p>Este estudio presenta una propuesta sobre el uso de TIC para la enseñanza de la estadística en el aula, en específico, se muestra la utilización de la herramienta Análisis de Datos disponible en Excel con la que es posible realizar completos resúmenes descriptivos y gráficos, así como también, el generar datos aleatorios que se pueden utilizar para comprobar empíricamente conceptos habitualmente presentados de forma teórica, tales como el Teorema Central de Límite y el concepto de nivel de Confianza, fundamental para la construcción de un intervalo de confianza.</p> | <p>This study presents a proposal about the use of ICT for the teaching of statistics in the classroom, specifically, it shows the use of the Data Analysis tool available in Excel with which it is possible to make complete descriptive and graphic summaries, as well as, generate random data wich can be used to empirically verify concepts usually presented in a theoretical way, such as the Central Limit Theorem and the concept of confidence level, fundamental for the construction of a confidence interval.</p> |
| PALABRAS CLAVE:   | KEYWORDS:  |
| <p>TIC – Excel<br/>Complemento Análisis de Datos.</p>   | <p>ICT – Excel – Data Analysis Complement.</p>   |

## 1. INTRODUCCIÓN

El uso de tecnologías de información y comunicación (TIC) es un complemento necesario para el desarrollo de una clase de Estadística, específicamente el uso de software se ha hecho fundamental para ejemplificar la teoría del análisis descriptivos de datos, la simulación de datos y su utilidad como complemento al concepto de variable aleatoria, la comprobación de teoremas relevantes, la aplicación en inferencia, entre otros. En este sentido, son muchos los autores que consideran muy importante la actitud hacia el uso de software y el alcance que dan a estos los alumnos y docentes.

Por una parte, el profesorado ha enfrentado problemas para incorporar TIC en su práctica docente. Estos problemas se ven reflejados en una articulación inadecuada de estas tecnologías con las disciplinas que imparte, pobre contextualización, y en el mejor de los casos, un aprovechamiento insuficiente de su potencial. Por otra parte, el estudiantado es un factor aún más complejo para descifrar. En efecto, muchos estudiantes actuales no necesitan adaptarse a las nuevas tecnologías porque nacieron con ellas, lo que supone una problemática aún mayor para autoridades educativas y profesores (Hernández, 2013, p. 166)

En esta misma línea, autores como Batanero (2009) destacan cómo la tecnología ha influido en la estadística y su enseñanza, siendo reconocida por la Internacional Association for Statistical Education (IASE) en los sucesivos Congresos Internacionales sobre la Enseñanza de la Estadística, donde se discute sobre el software disponible para la enseñanza, los cambios implicados en el contenido y la metodología, y el efecto en el aprendizaje y las actitudes de los alumnos. “Asimismo, se destaca cómo la tecnología ha reducido el tiempo de cálculo, permitiendo trabajar con aplicaciones reales en clase”. (Espinoza y Fernández, 2014, p. 90)

Se plantea entonces el uso del software Microsoft Excel, el porqué de su elección es simple, está disponible en la mayoría de los laboratorios de computación en colegios o Universidades. “Microsoft Excel es la hoja de cálculo más popular que se utiliza para almacenar información en columnas y filas, que luego pueden ser organizadas y/o procesadas” (Giles, 2002, p.2). Muchos autores consideran Microsoft Excel como una excelente herramienta para la enseñanza de la estadística. “Una ventaja importante de la hoja de cálculo Excel es que se ha convertido en un software estándar en los entornos de enseñanza, profesionales y familiares. Excel tiene una interfaz amigable y es fácil usar” (Cao y Naya, 2010, p.2).

Ahora, si bien Excel no es un programa propiamente estadístico y presenta ciertas limitantes respecto a los alcances de los métodos estadísticos y de cálculo (Heiser, 2006, Simonoff, 2008; Mac-Cullough y Heiser, 2008), es suficiente para abordar, por ejemplo, un primer curso de Estadística de nivel universitario (Estadística Descriptiva) e inclusive parte de un segundo curso (Inferencia) solo haciendo uso de las funciones directas ejecutables en la planilla Excel o bien mediante las diferentes opciones que proporciona la herramienta de Análisis de Datos (Levine, Stephan y Szabat, 2010; Brenes y Vanegas, 2014)

Artículos sobre el uso de gráficos interactivos en Excel para la enseñanza de la estadística pueden encontrarse en Coll y Blasco (2010), elementos de simulación para la comprensión de teoremas de probabilidad en Brenes y Vanegas (2014) y elementos de contraste de hipótesis con apoyo de Excel en Batanero y Díaz (2015).

## **2. EL COMPLEMENTO ANÁLISIS DE DATOS DE EXCEL Y EL APRENDIZAJE DE LA ESTADÍSTICA**

“La estadística no es sólo una colección de conceptos y técnicas, sino sobre todo una forma de razonar. Es necesaria en la mayoría de los ámbitos, en cualquier ciencia o cualquier trabajo” (Martín, Cabero y De Paz, 2008)

El uso de Excel puede favorecer el proceso de aprendizaje de la Estadística al: (a) trasladar el contenido teórico estudiado al ámbito de la aplicación práctica sin necesidad de tener que realizar,

cuando se trabaja con gran cantidad de datos, los engorrosos cálculos necesarios para obtener las medidas estadísticas requeridas; y (b) promover el tratamiento de forma autónoma y crítica de información (Coll Serrano y Blasco Blasco, 2010, p.31)

Excel cuenta con un complemento denominado “Análisis de datos” tal herramienta permite trabajar datos de forma descriptiva, simular datos e incluso realizar inferencia estadística en base a intervalos de confianza y test de hipótesis entre otras opciones.

El paso inicial, es activar tal complemento desde la planilla de Excel. Esta opción está disponible en: Archivo Opciones Complementos Herramienta para análisis Ir (Esta descripción es para la versión Ms Excel 2013, para versiones anteriores la opción es similar). La visualización del complemento aparecerá en la opción “Datos”.

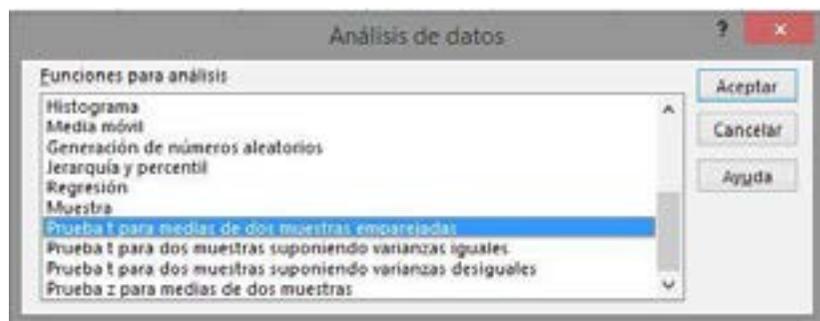


Figura 1.  
Ventana de herramientas.  
Análisis de datos.

Las opciones disponibles en la herramienta “Análisis de Datos” son múltiples, por ejemplo, posee herramientas relacionadas con el resumen descriptivo de datos, opciones: Estadística Descriptiva e Histograma. Herramientas de simulación: opción Generación de números aleatorios, donde es posible generar un conjunto de datos asociados a una distribución de probabilidad (Bernoulli, Binomial, Poisson, Normal, entre otras). Herramientas de Inferencia Estadística: opciones: Prueba t para medias de dos poblaciones emparejadas, Prueba t para dos muestras suponiendo varianzas iguales, Prueba z para medias de dos poblaciones, etc.

Se presentan a continuación algunos ejemplos de los alcances que se pueden hacer con la herramienta análisis de datos.

## 2.1. ANÁLISIS DESCRIPTIVO DE DATOS

Una de las aplicaciones básicas para cualquier curso de estadística, ya sea de nivel escolar o universitario es el resumen descriptivo de datos. La herramienta “análisis de datos” proporciona dos opciones que permiten obtener resúmenes completos de datos. La primera de ellas es la opción

Histograma, esta opción permite obtener una tabla de frecuencias de los datos acompañada de una serie de gráficos disponibles en el menú de opciones. La tabla de frecuencias puede ser de frecuencias absolutas, porcentaje acumulado y también permite la opción de obtener la misma tabla pero ordenada de mayor a menor frecuencia que es utilizada para la confección del diagrama de Pareto. Adicional a esto, se puede generar un histograma de frecuencias, un histograma con una ojiva (gráficos de porcentaje acumulado) y un diagrama de Pareto.

La segunda opción denominada “Estadística Descriptiva” permite obtener un resumen descriptivo de los datos, la opción proporciona medidas de tendencia central (media, mediana y moda), medidas de posición (mínimo y máximo), medidas de dispersión (rango, varianza, desviación estándar y error típico) y medidas de forma (asimetría y curtosis).

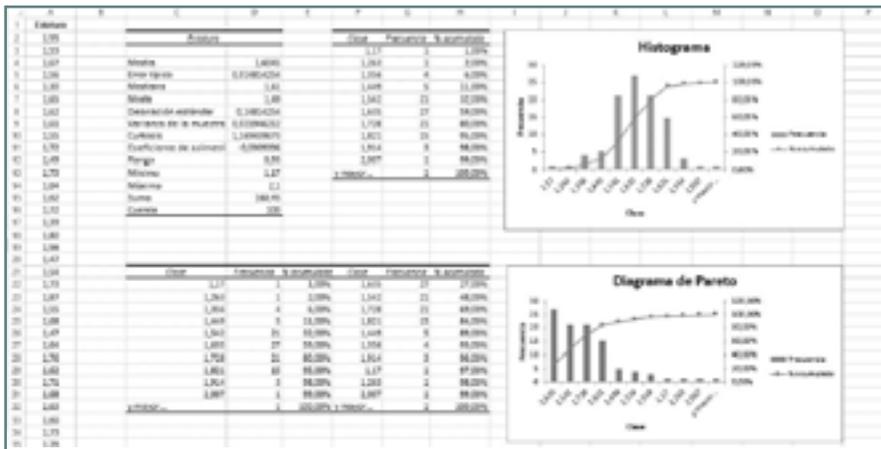


Figura 2. Salidas obtenidas de las opciones histograma y Estadística Descriptiva.

## 2.2. VISUALIZACIÓN DEL TEOREMA CENTRAL DEL LÍMITE

El Teorema Central del Límite es uno de los resultados más importantes obtenidos en estadística, suele ser visto como introducción a los cursos de Inferencia estadística o luego de estudiar las distribuciones de variable aleatoria. Una definición de este teorema que proporcionan los textos de nivel universitario es el siguiente: “Cuando se seleccionan muestras aleatorias simples de tamaño  $n$  de una población, la distribución muestral de la media muestral puede aproximarse mediante una distribución normal a medida que el tamaño de la muestra se hace grande” (Anderson, Sweeney y Williams, 2008, p. 274).

Esta definición si bien es simple, podría no comprenderse a cabalidad por los alumnos sin la visualización del teorema mediante gráficos de apoyo. Para ello, se puede utilizar la opción “Gene-

ración de números aleatorios” disponible en la herramienta análisis de datos de Excel. En la Figura 3 (parte superior) se muestra el gráfico de frecuencias de una variable Poisson (no es el gráfico teórico, es decir, el derivado del cálculo de las respectivas probabilidades), esta variable fue simulada con la opción Generación de número aleatorios (el gráfico se obtuvo con 1000 datos correspondientes a una distribución de Poisson con parámetro  $\lambda = 5$ ), luego, para mostrar el Teorema Central del Límite se realizó una simulación de 200 columnas de 1000 datos de la distribución Poisson, para cada columna de datos se determinó la media muestral y luego se graficaron estos 200 promedios para obtener el histograma con forma acampanada (gráfico de la parte inferior de la Figura 3) del cual, se observa que es aproximadamente una distribución Normal comprobando visualmente el Teorema Central del Límite.

De esto, también puede desprenderse que el gráfico está centrado en un valor cercano a 5, valor que es esperado, porque según la teoría de estimadores, el estimador del parámetro  $\lambda$  de la distribución de Poisson es la media muestral y la distribución de la media muestral es aproximadamente Normal centrada en tal valor  $\lambda$ .

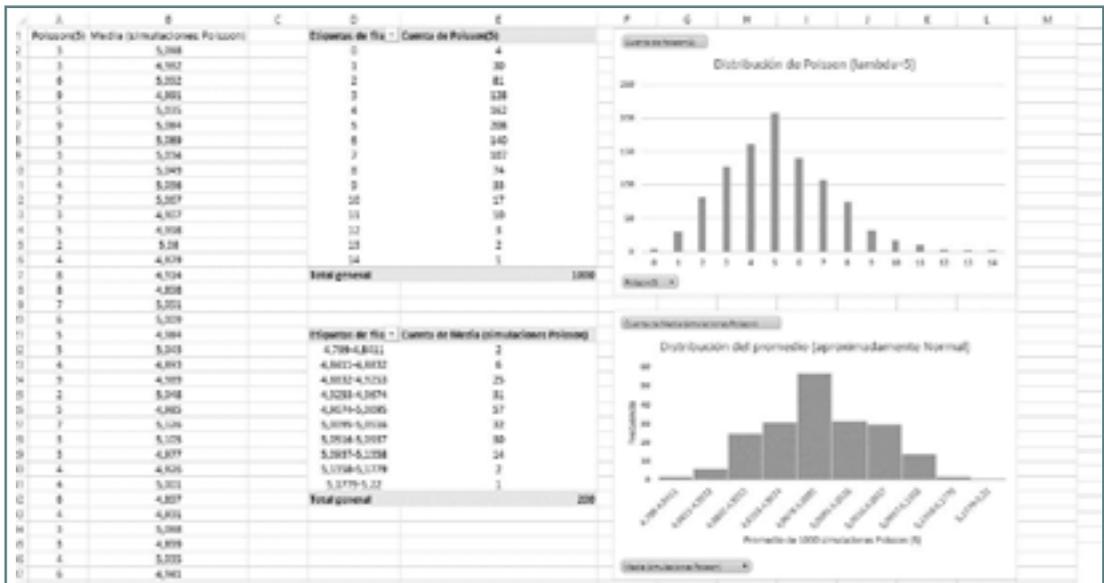


Figura 3. Uso de la función Generación de números aleatorios para mostrar el Teorema Central del Límite.

## 2.3. EL SIGNIFICADO DE LA CONFIANZA EN EL INTERVALO DE CONFIANZA

Una aplicación para el trabajo con intervalos de confianza es mostrar visualmente qué significa la confianza. La teoría indica que la confianza o también denominada grado de confianza o coeficiente de confianza corresponde a la “probabilidad de seleccionar una variable aleatoria que produzca un intervalo que contenga al parámetro” (Walpole, Myers, Myers y Ye, 2007, p. 273). Esta definición puede resultar confusa para los estudiantes e incluso para el docente que intenta explicarla. Una solución, es utilizar la opción “Generación de números aleatorios” de la herramienta Análisis de Datos. Como primer paso, se generan  $N$  muestras aleatorias de tamaño  $n$ . En el ejemplo de la Figura 4, se generaron  $N = 200$  muestras de tamaño  $n = 200$  de una distribución, luego, con cada muestra se determinó el respectivo intervalo de confianza de 95% para la media con varianza conocida obteniendo 200 intervalos de confianza, finalmente, para cada intervalo y mediante la creación de una función lógica se determinó si este intervalo contenía el valor del parámetro a estimar, es decir, si el intervalo de confianza de 95% para la media contenía el valor de la media poblacional. En este ejemplo se obtuvo que el 95,4% de los intervalos contenían a la media poblacional, es decir, aproximadamente el 95% de los intervalos contenían al parámetro, mostrando de este modo el significado en la construcción del intervalo de confianza el concepto de confianza.

|    | Intervalo inferior | Intervalo superior | 95% confianza al parámetro mu=1,07 | Simulación   |
|----|--------------------|--------------------|------------------------------------|--|
| 1  | 1,06179134         | 1,07929866         | SI                                 | 200 muestras aleatorias de tamaño n = 200, asociada a una variable con distribución Normal (mu=1,07, sigma=0,05) |
| 2  | 1,06222318         | 1,07277682         | SI                                 |  |
| 3  | 1,06102114         | 1,07977886         | SI                                 | Intervalo de confianza para la media con varianza conocida   |
| 4  | 1,06229218         | 1,07367886         | SI                                 | Para cada simulación de tamaño 200 se determinó un intervalo de confianza (IC) de 95%                            |
| 5  | 1,06302118         | 1,07397886         | SI                                 |  |
| 6  | 1,06197114         | 1,08177886         | SI                                 | Porcentaje de Intervalos de confianza que contienen el valor de la media poblacional (parámetro estimado)        |
| 7  | 1,06272118         | 1,07877886         | SI                                 | Porcentaje = 0,954   |
| 8  | 1,06279118         | 1,07827886         | SI                                 |  |
| 9  | 1,06402118         | 1,07827886         | SI                                 |  |
| 10 | 1,06272118         | 1,07522886         | SI                                 |  |
| 11 | 1,06272118         | 1,08227886         | SI                                 |  |
| 12 | 1,06272118         | 1,08227886         | SI                                 |  |
| 13 | 1,07872118         | 1,08472886         | NO                                 |  |
| 14 | 1,06872118         | 1,08072886         | SI                                 |  |

Figura 4. Uso de la función Generación de números aleatorios para la comprensión del concepto de confianza.

## 3. PROPUESTAS DE PROBLEMAS A DESARROLLAR EN EL AULA

Se presenta a continuación una propuesta de actividades que podrían ser desarrolladas durante una clase de Estadística en donde se utilice la herramienta Análisis de Datos de Excel para la resolución del problema en cuestión.

### 3.1. RESUMEN DE DATOS

Dar la indicación a los alumnos que busquen bases de datos en la web para una o más

variables de interés o dos grupos distintos para una misma variable. Estas bases de datos se pueden encontrar por ejemplo en el banco central, ministerios de educación, de salud, etc.

Ya contando con la base de datos de interés, utilizar la herramienta análisis de datos en la opción Estadística Descriptiva para obtener:

- Estadísticos descriptivos de centralización, (Media, Mediana y Moda)
- Estadísticos descriptivos de dispersión (rango y desviación estándar)
- Coeficiente de asimetría y curtosis

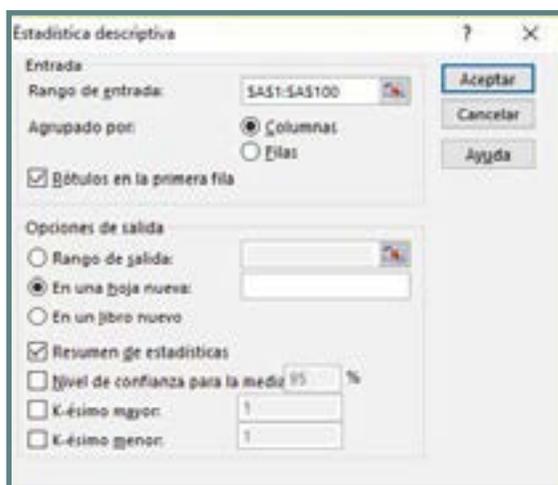


Figura 5.  
Menú Estadística Descriptiva

Será interesante para el alumno verificar, por ejemplo, asimetría en la distribución de frecuencias de los datos, haciendo la comparación entre las medidas de tendencia central y el coeficiente de asimetría, también verificar la relación de los datos que presenten una alta variabilidad con una curtosis negativa y una baja variabilidad con una curtosis positiva.

Utilizando la opción histograma, el alumno puede obtener un gráfico de frecuencias para los datos, construido en base a la cantidad de clases creada por Excel, se recomienda indagar sobre la construcción de las clases y cómo estas se determinan. El docente puede pedir al alumno que estudie la asimetría de la distribución de los datos y corrobore su apreciación del gráfico con lo que entregó el coeficiente de asimetría y la comparación de las medidas de centralización.

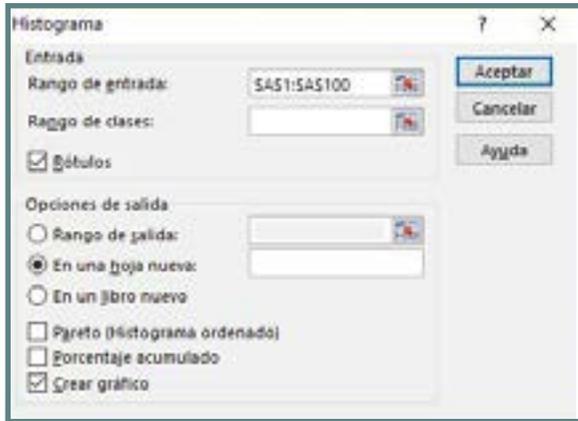


Figura 6. Menú Histograma

La herramienta del histograma, también permite construir un diagrama de Pareto con las frecuencias acumuladas, aquí el docente puede pedir al alumno la interpretación del diagrama en el contexto de la variable, de acuerdo al principio de Pareto.

### 3.2. ESTUDIO DE LAS DISTRIBUCIONES DE VARIABLES ALEATORIAS

Para esta actividad se utilizará la opción Generación de Números aleatorios de la Herramienta Análisis de datos.

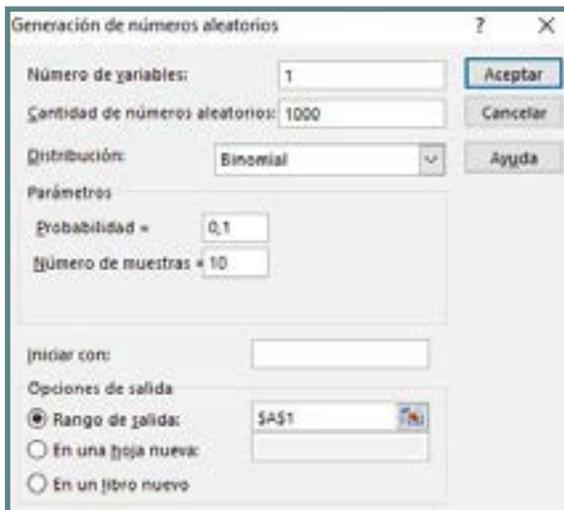


Figura 7..  
Menú Generación  
de números aleatorios.

La idea de la actividad es presentar a los alumnos las distintas distribuciones de probabilidad (discretas y continuas) disponibles en la opción Análisis de Datos. Esto es de utilidad, ya que, complementa la presentación teórica de las distribuciones de probabilidad (notación y función de densidad de probabilidad), permitiendo al estudiante explorar qué ocurre con la distribución si se modifican los parámetros que la definen.

El estudio podría estar centrado en responder las siguientes preguntas:

¿Qué tipo de parámetro está asociado a la distribución estudiada? ¿Un parámetro de localización (asociado a una media) o de escala (asociado a una varianza)?

En este problema se espera que el alumno manipule el o los parámetros asociados a la distribución seleccionada para una cierta cantidad de simulaciones de la variables aleatoria y observe los cambios en ésta. Además, se espera que se discuta respecto a la aproximación de la distribución empírica (simulada) a la distribución teórica (probabilidades determinadas mediante cálculo) a medida que se incrementa la cantidad de valores simulados.

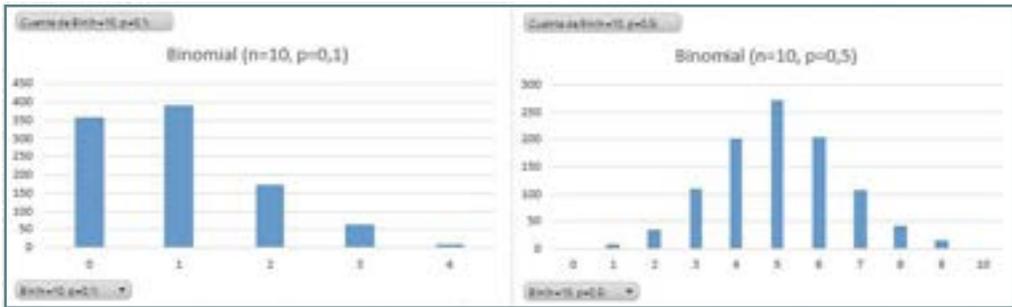


Figura 8. Comparación de distribuciones Binomiales.

### 3.3. DISTRIBUCIONES DE MUESTREO

Como problema, se propone estudiar la distribuciones de muestreo de la media y la proporción muestral.

Utilizando la opción generación de números aleatorios, simular N columnas de tamaño n de una distribución Normal con media y desviación estándar definidas. Para el estudio de

la media muestral, para cada una de las N columnas obtener un promedio para los n datos que la componen, graficar estos N promedios con la opción Histograma y observar la forma de la distribución. Sería interesante comprobar que esta distribución es Normal y que la media de esta Normal es prácticamente la media ingresada en la opción generación de números aleatorios (mientras más grande sea n, este valor será más cercano) y que la desviación estándar es cercana a la desviación estándar ingresada en la opción, pero dividida por la raíz de n.

| <i>Promedio de N= 200, Normal(mu=550, sigma=50)</i> |             |
|---|-------------|
| Media   | 550,1562329 |
| Error típico  | 0,103304435 |
| Mediana   | 550,0876076 |
| Moda  | #N/A        |
| Desviación estándar                                 | 1,460945328 |
| Varianza de la muestra                              | 2,13436125  |
| Curtosis  | 0,2062752   |
| Coefficiente de asimetría                           | 0,156313202 |
| Rango   | 9,10619616  |
| Mínimo  | 546,1671924 |
| Máximo  | 555,2733885 |
| Suma  | 110031,2466 |
| Cuenta  | 200         |

Figura 9.  
Estadísticos descriptivos  
para 200 promedios  
de una distribución Normal.

Para el caso de la proporción, se puede definir un criterio que asigne el valor 1 si se cumple la condición definida y 0 en otro caso. Aplicar este criterio a las N columna de tamaño n, se tendrán ahora N columnas con n datos dicotómicos (0 y 1). Obtener el promedio para cada una de las N columnas, estos promedios corresponderán a las proporciones muestrales de cada columna, es importante que el alumno comprenda que un promedio de variables dicotómicas genera una proporción muestral, además, que cada uno de los resultados (0 ó 1), pueden ser modelados por una distribución de Bernoulli de parámetro (probabilidad de éxito, es decir, probabilidad de que el dato cumpla con el criterio especificado) y que la distribución de la proporción es Normal (porque la proporción muestral es un promedio) con media igual a  $p$  y varianza igual a  $p(1-p)$ . También se puede abordar este problema simulando N columnas de n datos utilizando la distribución Bernoulli con un parámetro definido.

## 4. CONCLUSIONES

La herramienta análisis de datos de Excel ofrece una gran variedad de opciones para el tratamiento de distintos temas que habitualmente se presentan de forma teórica en los cursos de Estadística en distintos niveles educativos. El manejo de la planilla de cálculo de Excel facilita el aprendizaje de los conceptos estadísticos, ya que, al evitar el cálculo manual se puede dar énfasis en el entendimiento de los conceptos y de la interpretación de los resultados obtenidos.

Si bien, Excel presenta limitaciones en los alcances que puede entregar la herramienta Análisis de Datos, las herramientas disponibles son suficientes para poder abordar un curso de Estadística Matemática y de Estadística Inferencial, también, permiten al docente mostrar teoremas o conceptos, tales como, las distribuciones de variables aleatorias discretas y continuas, el teorema de la función inversa, el Teorema central del Límite, el significado de la confianza en un intervalo de confianza, entre otros.

## AGRADECIMIENTOS

Proyecto de Docencia 2105-2016, Vicerrectoría Académica, Universidad Bernardo O'Higgins

## REFERENCIAS BIBLIOGRÁFICAS

Anderson, D., Sweeney, D. y Williams, T. (2008). *Estadística para administración y Economía*. México: Cengage Learning.

Batanero, C. (2009). *Retos para la formación estadística de los profesores*. II Encontro de Probabilidade e Estatística na Escola. Universidade do Minho, 2009, Braga, Portugal. Recuperado el 30 de enero de 2016 de <http://www.ugr.es/~batanero/pages/ARTICULOS/Formprofesores.pdf>.

Batanero, C., y Díaz, C. (2015). *Aproximación informal al contraste de hipótesis*. II Jornadas de Didáctica de la Estadística la Probabilidad y la Combinatoria, Universidad de Granada, 2015, España. Recuperado el 16 de abril de 2016 de <http://www.ugr.es/~batanero/documentos/Aproximacion.pdf>

Brenes, G. S., y Vanegas, F. N. (2014). *Simulación en Excel: Buscando la probabilidad de un evento*. Revista Digital: Matemática, Educación e Internet, 12(2).

Cao, R. y Naya, S. (2010). *The use of statistical software to teach nonparametric curve estimation: from Excel to R*. 8th International Conference on Teaching Statistics. Ljubljana, Slovenia. Recuperado el 23 de diciembre de

2016 de [http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8\\_4B1\\_CAO.pdf](http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_4B1_CAO.pdf).

Coll Serrano, V., y Blasco Blasco, O. (2010). *El uso de gráficos interactivos en Excel para facilitar la comprensión de conceptos básicos de Estadística*. @tic. Revista d'Innovació Educativa, (5), 30-34.

Espinoza, C. y Fernández, J. (2014). *Importance of statistical software teaching and learning at the university of Carabobo (Venezuela)*. Aula de encuentro 16(1), 89-102.

Giles, O. (2002). *Using excel to teach statistics in New Zealand secondary schools*. ICOTS 2002. Recuperado el 20 de mayo de 2016 de [http://iase-web.org/documents/papers/icots6/7g1\\_gile.pdf](http://iase-web.org/documents/papers/icots6/7g1_gile.pdf)

Heiser, D. A., (2006). *Microsoft Excel 2000 and 2003 faults, problems, workarounds and fixes*. Computational Statistics & Data Analysis, 51 (2), pp. 1442-1443.

Hernández, S. y Cuevas, J. (2013). *Programas informáticos de uso libre y su aplicación en la enseñanza de la Estadística*. Revista Investigación Operacional 34(2), 166-174.

Levine, David M., Stephan, David F. y Szabat, Kathryn A. (2010). *Statistics for Managers using MS Excel*. New Jersey: Prentice Hall.

MacCullough, B.D. y Heiser, D. A. (2008). *On the accuracy of statistical procedures in Microsoft Excel 2007*. Computational Statistics & Data Analysis, 52, pp.4570-4578.

Martín Martín, Q., Cabero, M. T. y de Paz Santana, Y. (2008). *Tratamiento estadístico de datos con SPSS*. España: Thomson.

Simonoff, Jeffrey S. (2008). *Statistical analysis using Microsoft Excel*. Recuperado el 18 de abril de 2016 de <http://pages.stern.nyu.edu/~jsimonof/classes/1305/pdf/excelreg.pdf>.

Walpole, R. E., Myers, R., Myers, S. y Ye, K. (2007). *Probabilidad y Estadística para Ingeniería y Ciencias*. México: Prentice Hall.