

## ELEMENTOS DE INFERENCIA ESTADÍSTICA EN EXCEL

Álvaro Toledo

alvaro.toledo@ubo.cl

Universidad Bernardo O'Higgins, Chile

Tema: Pensamiento probabilístico-estadístico

Modalidad: Taller

Nivel educativo: Terciario - Universitario

Palabras clave: Inferencia Estadística– Excel – Análisis de datos

### Resumen

*En este taller se mostrarán elementos de inferencia estadística mediante el uso del complemento Análisis de Datos de Excel. Se explicará como mostrar conceptos base de inferencia, tales como, las distribuciones de muestreo asociadas a la media, proporción y varianza mediante la opción “generación de números aleatorios”, además de cómo construir intervalos confianza utilizando la teoría de distribución de muestreo. En este mismo punto se reflexionará sobre el significado de la confianza en un intervalo de confianza mostrando con elementos de la herramienta análisis de dato su significado. Finalmente, se explorarán las opciones de Excel dedicadas a intervalos de confianza y ANOVA de un factor.*

### Introducción

El desarrollo de las tecnologías de la información (TIC) han sido de gran relevancia durante los últimos años como complemento en la enseñanza en distintos ámbitos de la ciencias y las humanidades, “la tecnología computacional ha mostrado un enorme potencial para ayudar a los estudiantes a comprender conceptos difíciles” (NCTM, 2000; Ben-Zvi, 2000; Mills, 2002; Chance y Rossman 2006). El desarrollo de software en el área de las matemáticas ha proporcionado por ejemplo un gran avance respecto a la aplicación de conceptos teóricos que hasta hace algunos años atrás su explicación no escapaba del ámbito teórico. Lamentablemente como advierte Hernández (2013) el profesorado ha enfrentado problemas para incorporar TIC en su práctica docente, y agrega que “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” (Hernández, 2013, p.166)

En el ámbito de la Estadística distintos actores han destacado el uso de software como complemento a la enseñanza de la estadística Batanero (2009) destaca 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.

A través de la simulación el estudiante puede explorar y comprender conceptos y principios (por ejemplo las distribuciones de probabilidad, muestreo aleatorio y distribuciones de estadísticos) que de otro modo serían mucho más abstractos, contribuyendo con ello a mejorar la experiencia estocástica y la intuición probabilística. Por su parte, en estadística, la computadora puede ser de gran ayuda en la automatización de cálculos laboriosos (como el de medidas descriptivas como la desviación estándar y el coeficiente de correlación), en la exploración de datos y en la construcción de gráficas (Inzunsa, 2010, p.425)

Para el desarrollo de este taller se ha seleccionado el software Microsoft Excel, el porqué de su uso se fundamenta en el hecho de que Excel está disponible en la mayoría de los computadores personales o laboratorios de computación de las instituciones de educación, además, “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, 2013, p.2).

Excel cuenta con la herramienta Análisis de datos el cual puede ser activada desde las opciones de complementos del mismo. Esta herramienta posee un conjunto de opciones que abarcan desde elementos de estadística descriptiva (resumen de estadísticas,

histogramas y medidas de relación), simulación de números aleatorios, inferencia y modelos de regresión lineal entre otras. Este taller se concentrará en las opciones dedicadas a distribuciones de muestreo e inferencia Estadística.

### **Algunos concepto de distribuciones de muestreo e inferencia Estadística**

Se definen a continuación algunos conceptos de inferencia que serán tratados en el taller de Excel. Las definiciones han sido extraídas del texto Probabilidad y Estadística para Ingeniería y Ciencias de Walpole, Myers, Myers y Ye (2007).

- Distribución de Muestreo:

La distribución de probabilidad de un estadístico se llama distribución muestreo.

La distribución de probabilidad de  $\bar{X}$  se llama distribución muestral de la media .  
La distribución muestral de un estadístico depende del tamaño de la población, del tamaño de las muestras y del método de elección de éstas.

- Teorema Central del Límite:

Si es la media de una muestra aleatoria de tamaño  $n$  tomada de una población con media  $\mu$  y varianza finita  $\sigma^2$  , entonces la forma límite de la distribución de:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

conforme  $n \rightarrow \infty$  , es la distribución normal estándar  $N(z ; 0, 1)$ .

- Intervalo de Confianza (definición general, adaptada de Walpole et al.(2007)):

Sean  $\hat{\theta}_L$  y  $\hat{\theta}_U$  estimadores inferior y superior respectivamente del parámetro  $\theta$ , tales que:

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha$$

para  $0 < \alpha < 1$ , tenemos entonces una probabilidad de  $1 - \alpha$  de seleccionar una variable aleatoria que produzca un intervalo que contenga  $\theta$ . El intervalo  $\hat{\theta}_L < \theta < \hat{\theta}_U$ , que se calcula a partir de la muestra seleccionada, se llama entonces intervalo de confianza de  $(1 - \alpha)100\%$ , la fracción  $1 - \alpha$  se llama coeficiente de confianza o grado de confianza, y los extremos,  $\hat{\theta}_L$  y  $\hat{\theta}_U$ , se denominan límites de confianza inferior y superior.

### Desarrollo del taller

El taller abordará cuatro puntos que se consideran relevantes para la enseñanza de la estadística utilizando Excel. El primer punto consiste en la activación de la herramienta análisis de datos de Excel, una vez activada, se mostrará a los participantes las opciones disponibles y los alcances que pueden lograr con cada una de éstas. Como segundo punto se realizará el trabajo con la opción generación de números aleatorios de la herramienta análisis de datos. Se mostrará lo importante de esta opción en la comprobación del Teorema Central del Límite y la relación de éste con las distribuciones de muestreo. Se mostrará además la importancia de estos conceptos para el desarrollo de la Inferencia Estadística. El tercer punto corresponde a la construcción de intervalos de confianza, se mostrará mediante el uso de la herramienta análisis de datos el significado de la confianza en un intervalo de confianza, además, se mostrarán las diferencias entre el concepto de construcción y estimación del intervalo de confianza. Se indicarán errores habituales en la interpretación de estos y se intuirá el concepto de test de Hipótesis. Como punto final se mostrarán las opciones disponibles

en la herramienta análisis de datos dedicadas a inferencia estadística, específicamente a las opciones de comparación de medias (bajo el supuesto de varianzas desconocidas y desconocidas pero iguales) y razón varianzas poblacionales, también se explorará la opción de Análisis de la Varianza (ANOVA) de un factor.

### **Algunas actividades propuestas**

Se presentan a continuación un conjunto de problemas que guiarán en el desarrollo de los conceptos mencionados anteriormente. Los problemas 1 y 2 se obtuvieron de Toledo (2016).

#### 1. Visualización del Teorema Central del Límite.

Considere el experimento aleatorio de extraer de una urna 3 esferas con reposición. La urna está compuesta de 5 esferas azules y 3 esferas rojas. Nos interesa el número de esferas azules extraídas.

- i) Simule el experimento descrito en el enunciado.
- ii) Repita el experimento aleatorio 1000 veces
- iii) Realice una tabla de frecuencias para el punto anterior, luego grafique. Comente.
- iv) Simule 500 veces la repetición de 1000 veces del experimento aleatorio.
- v) Determine el promedio para cada una de las 500 columnas obtenidas anteriormente.
- vi) Grafique los promedios. ¿Qué observa?

2. El papel de la confianza en el intervalo de confianza.

Considere una variable aleatoria  $X$ : Estatura de un latinoamericano (metros).

- i) Simule 1000 datos de una distribución Normal con media 1,68 y desviación estándar 0,05 (¿qué implica esta desviación estándar?). Repita esta simulación 200 veces.
- ii) Para cada una de las repeticiones determine su respectivo intervalo de confianza para la media poblacional (puede considerar el caso de varianza poblacional conocida).
- iii) Finalmente, determine la proporción de intervalos de confianza que contienen a la media poblacional del enunciado.

3. Otras aplicaciones: ANOVA de un factor (Ejercicio extraído de Anderson, Sweeney y Williams, T (2008))

En la publicidad de cuatro pinturas se dice que tienen el mismo tiempo de secado. Para verificar esto, se prueban cinco muestras de cada una de las pinturas. Se registra el tiempo en minutos necesario para que el secado sea suficiente para la aplicación de una segunda mano. Los datos obtenidos son los siguientes.

Pintura 1	Pintura 2	Pintura 3	Pintura 4
128	144	133	150
137	133	143	142
135	142	137	135
124	146	136	140
141	130	131	153

- i) Plantee la hipótesis asociada al problema
- ii) Con  $\alpha = 0.05$  como nivel de significancia, realice una prueba para determinar si la media de los tiempos de secado es la misma en todas las pinturas.
- iii) Si la hipótesis nula es rechazada, indique cómo podría verificar cuáles y cuántas pinturas tienen un tiempo de secado distinto.

### 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.  
<http://www.ugr.es/~batanero/pages/ARTICULOS/Formprofesores.pdf>.  
Consultado el 30/10/2016.
- Ben-Zvi, D. (2000). "Toward understanding the Role of Technological Tools in Statistical Learning", *Mathematical Thinking and Learning*, 2(2), 127-155.
- 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. Disponible en [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). Consultado el 23/01/2016.
- Chance, B. y Rossman, A. (2006). "Using simulation to teach and learn statistics", en A. Rossman y B. Chance (eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. [cd-rom]. Voorburg: The Netherlands: International Statistical Institute.
- 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.
- Inzuna, S. (2010). Entornos Virtuales de Aprendizaje: Un enfoque alternativo para la enseñanza y aprendizaje de la inferencia estadística. *Revista Mexicana de Investigación Educativa* 15(45), 423-452.

<http://www.comie.org.mx/v1/revista/visualizador.php?articulo=ART45005&critério=http://www.comie.org.mx/documentos/rmie/v15/n045/pdf/ART45005.pdf> .  
Consultado: 24/02/17

Mills, J. (2002). “Using computer simulation methods to teach statistics: A review of the literature”, *Journal of Statistics Education* 10(1).

<http://www.amstat.org/publications/jse/v> . Consultado: 11/11/2016

NCTM (2000). Principles and Standards for School Mathematics. Reston, VA: NCTM.

Toledo, Á. (2016). El uso de la herramienta “Análisis de datos” de Excel como complemento para la enseñanza de la estadística en el aula, CUREM 6, <http://semur.edu.uy/curem6/actas/pdf/27.pdf> . Consultado: 29/11/2016.

Walpole, R., Myers, R., Myers, S. y Ye, K. (2007). *Probabilidad y Estadística para Ingeniería y Ciencias y Economía*. México: Pearson Education.