

SIMULACIÓN DE DISTRIBUCIONES MUESTRALES

Simulation of sampling distributions

EPISTEMUS
ISSN: 2007-8196 (electrónico)
ISSN: 2007-4530 (impresa)

Víctor A. Noriega Ortiz

Recibido: 12 de septiembre de 2016,
Aceptado: 30 de noviembre del 2016

Autor de Correspondencia:
Víctor A. Noriega Ortiz
Correo: victornoriega7@gmail.com

Resumen

Por lo general los cursos de estadística en una licenciatura presuponen que los estudiantes cuentan con conocimientos matemáticos que facilitarán la comprensión de diversas herramientas estadísticas. Al realizar inferencias acerca de un parámetro se utiliza la información de una muestra y generalmente se cuenta con un estimador de éste. Los estimadores son variables aleatorias y tienen asociadas una distribución de muestreo, a la cual nos acercamos por medio de herramientas de simulación, las cuales juegan un papel fundamental, ya que mediante ellas el estudiante comprende la distribución de muestral de varios estimadores y le es fácil conectar aspectos inferenciales a partir de los resultados computacionales obtenidos. El ahondar sobre las herramientas matemáticas necesarias para entender algunas de las distribuciones de muestreo en un curso de estadística, y el desarrollo de una aplicación que pueda facilitar esta enseñanza, son los motivos fundamentales que dieron lugar al presente material.

Palabras clave: distribución de muestreo, estimador, simulación, Shiny.

Abstract

Generally, basic statistic courses at undergraduate level suppose that students have certain mathematical background that will facilitate use and comprehension of diverse statistical tools. When making inferences for certain parameter, the information in a sample will allow, throughout certain estimator, to obtain inferences on the parameter of interest. These estimators are random variables and have associated a sampling distribution to which we can be close by means of simulations tools; these play an essential role because through them, students can learn about different sampling distributions making easy to link them with some inferential aspects from the computational results. The mathematical tools needed to understand some sampling distributions in a basic statistics course and the development of an app that could be helpful in these courses, are the main reasons of developing the material here presented.

Keywords: sampling distribution, estimator, simulation, Shiny.

INTRODUCCIÓN

En la actualidad es de lo más común ver estudios o bien reportes que utilizan técnicas estadísticas, ya sea desde un aspecto descriptivo o bien para estimar o predecir ciertos resultados. Esta tendencia, que al parecer seguirá creciendo, requiere que los cursos que se imparten a nivel universitario contemplen un aspecto práctico que permita a los estudiantes poseer los conocimientos básicos sobre la materia y cómo aplicar dichos conocimientos en la práctica. En la mayoría de las carreras universitarias se imparten cursos de estadística a distintos niveles. La actualización de temarios y la inclusión de nuevas tecnologías de la información, las grandes cantidades de datos que hoy en día se debe saber trabajar, obligan a descartar ciertos temas teóricos, cuya inclusión forzaría el extender los temarios o bien adicionar otros cursos de estadística, lo cual no siempre es posible.

Un tema que se cubre en todo curso de estadística, aunque a diferentes niveles, es el relacionado con los conceptos de población, muestra, muestreo y distribuciones de muestreo. Hacemos uso de muestras para poder inferir acerca de una población, pues usualmente resulta muy complicado y/o costoso obtener la información de cada elemento de la población. El parámetro o característica de interés de la población puede ser variado. Un caso común es tratar de inferir sobre la media de una población. Por ejemplo, un problema estadístico podría ser estimar la media del cociente intelectual del mexicano, la cual imaginemos está representada por la Figura 1. Claro está lo complejo de efectuar una prueba de IQ a todos los mexicanos para poder llegar a conocer esta distribución; es por ello que hacemos uso de técnicas de muestreo. Una característica de interés que en este caso nos permitiría estimar este parámetro es la media muestral, que es un estimador muy utilizado en la práctica.

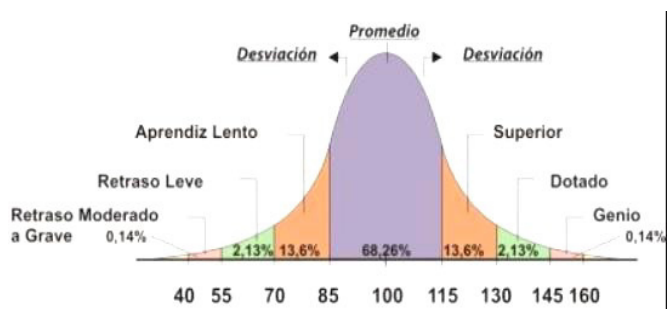


Figura 1. Distribución de Cocientes Intelectuales con media 100 y desviación estándar de 15, tomado y adaptado de S. B. Kathleen [1].

Una pregunta que resulta de inmediato es cómo se distribuye la media muestral, ese estimador que decidimos utilizar y que nos permitirá inferir sobre esta media poblacional. Con este ejemplo sencillo podemos ver que el concepto de distribución muestral de un estadístico es crucial en los cursos de estadística. Sobra decir que



inferencias acerca de una varianza son muy comunes en los cursos básicos de estadística y es por ello que en el presente manuscrito se considerará el muestreo de poblaciones normales y particularmente se trabajará con la distribución muestral de la media y varianza muestral. Tener conocimiento sobre estas distribuciones permitirá realizar inferencias sobre la media y varianzas poblacionales. Para el caso particular de la media, se trabajará suponiendo tanto conocida como desconocida.

Es importante mencionar que, en los cursos universitarios de estadística básica, no es de esperarse que los estudiantes tengan bases sólidas de programación, pero sí la facilidad de aprender ciertos comandos en un software amigable, que les permita crear sus propios programas o bien ejecutar y entender los programas creados por otros. Como se dijo anteriormente, tampoco es de esperarse que en sus cursos previos los estudiantes hayan cubierto todas las herramientas matemáticas necesarias para un curso de estadística. Existen temas como el relativo a funciones generadoras de momentos y teoremas asociados a éstas que facilitarían la exposición del tema de distribuciones muestrales. Sin embargo, si estos temas no se han cubierto, existe la alternativa de motivar a estudiantes con facilidades matemáticas o computacionales, para que a través de trabajo extra-clase presenten ante el grupo, los desarrollos matemáticos que justifican la distribución de ciertos estimadores, o bien algoritmos de simulación que permitan visualizar la distribución de muestreo de éstos. El trabajo que aquí se desarrolla inició de esta manera.

Como se mencionó anteriormente, actualmente es necesario incluir herramientas de programación y afortunadamente existe mucho software estadístico tanto propietario como libre, y por medio de éste es posible simular varios procedimientos que se cubren en curso básico de estadística. Por ello la importancia de familiarizarse con herramientas de programación. Sin embargo, ¿qué hacer cuando los estudiantes del curso de estadística no cuentan con las bases más elementales de programación? Aunque existe software estadístico que no requiere saber programación, éste no siempre es accesible, o bien, lo que se requiere no viene incluido en el mismo. Es

ahí donde se pueden crear o utilizar aplicaciones elaboradas especialmente para ciertas tareas determinadas. En este artículo se presentan las herramientas matemáticas que serían necesarias para ver rápidamente el tema de distribuciones muestrales en un curso de estadística y se desarrolla una aplicación alternativa que permitiría ver este tema por medio de simulaciones. Para facilitar la comprensión de este escrito, el tema relativo a momentos de una distribución se presenta en la Sección 2, dando después una breve descripción, en la Sección 3, de algunas distribuciones de probabilidad. La sección 4 está dedicada a las funciones generadoras de momentos, ya que en muchas situaciones facilitan el conocer la distribución muestral de un estadístico, aspecto que se cubre en la Sección 5. Finalmente, en la Sección 6 se concluye con la propuesta de una aplicación desarrollada en un software estadístico del cual se proveerá un enlace para poder utilizarla.

MOMENTOS

Dado que en el transcurso de este escrito utilizaremos el concepto de momento de una variable aleatoria, es necesario establecer que, si X es una variable aleatoria con función de distribución $F(X)$ y su esperanza existe, su r -ésimo momento, el cual denotaremos por μ'_r se define como:

$$\mu'_r = E[X^r] \quad (1)$$

Nótese que

$$\mu'_1 = E(X^1) = E(X) \quad (2)$$

De igual manera puede definirse el r -ésimo momento con respecto a la media (momento central), de la siguiente manera, considerando también que la esperanza existe:

$$\mu_r = E[(x - \mu)^r] \quad (3)$$

Podemos observar que el segundo momento central de X es la varianza,

$$\mu_2 = E[(x - \mu)^2] = \sigma^2 \quad (4)$$

Estos momentos se conocen como momentos poblacionales. Ahora, también se tiene el equivalente de momentos muestrales. Así, si $X_1, X_2, X_3, \dots, X_n$ es una muestra aleatoria con densidad $f(\cdot)$, su r -ésimo momento muestral centrado en el origen, que denotaremos como M'_r , se calcula como:

$$M'_r = \frac{1}{n} \sum_{i=1}^n X_i^r \quad (5)$$

Los momentos muestrales centrados con respecto a la media muestral \bar{X} se definen así:

$$M'_r = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^r \quad (6)$$



En particular, cuando $r = 1$ obtenemos la media muestral, la cual se suele denotar como \bar{X}_n y se calcula como:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (7)$$

FUNCIONES DE DENSIDAD NORMAL, JI-CUADRADA Y t-STUDENT

Algunas de las distribuciones de probabilidad ampliamente utilizadas en los cursos básicos de estadística son la distribución normal, la Ji-cuadrada y la t-Student, mismas que serán utilizadas en el transcurso de este escrito.

La función de densidad de una variable aleatoria X que se distribuye normalmente con parámetros $\mu \in \mathbb{R}$ y $\sigma > 0$, es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}, \quad (8)$$

donde $x \in \mathbb{R}$; μ , la media de la distribución es un parámetro de localización y σ , su desviación estándar, es un parámetro de escala.

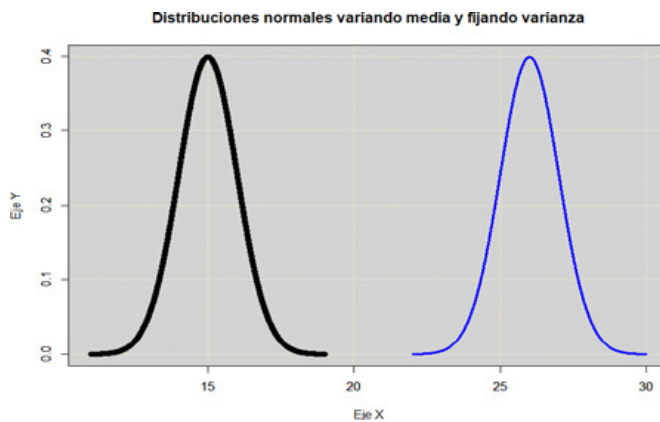


Figura 2. Distribuciones normales con diferente media pero misma varianza.

La forma que toma la distribución normal es acampanada, como puede verse en la Figura 2, donde se muestran dos distribuciones normales: ambas tienen $\sigma=1$, pero una tiene media $\mu = 15$, mientras que en la otra $\mu=26$. Por otra parte, en la figura 3, podemos observar dos distribuciones normales, ambas con $\mu = 15$, pero con diferente desviación estándar σ . Recordando lo expuesto

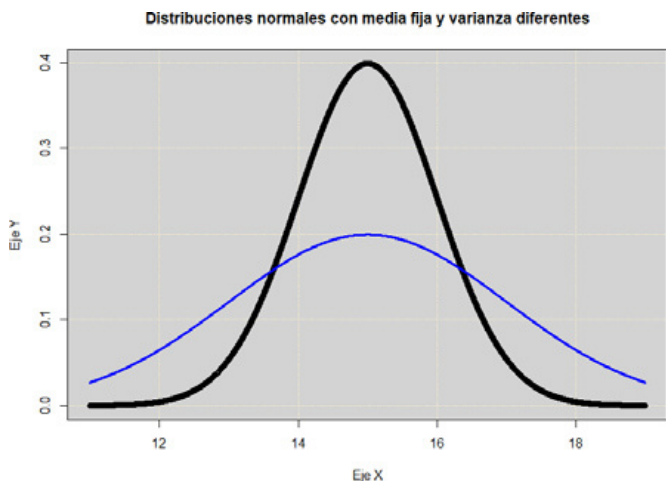


Figura 3. Distribuciones normales con misma media y diferente varianza.

en la Sección 2, podemos notar que el primer momento de la distribución normal es μ , mientras que el segundo momento de esta es σ^2 .

Por otra parte, la función de densidad de una variable aleatoria X que se distribuye como Ji-Cuadrada, con k grados de libertad, es:

$$g(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}, \quad (9)$$

donde $x \geq 0$, y la función $\Gamma(z)$ está dada por:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad (10)$$

Como puede verse en la Figura 4, la forma de la función de densidad de una variable aleatoria que se distribuye como una Ji-cuadrada, varía según sus grados de libertad.

Figura 4. Funciones de densidad de variables aleatorias con distribución Ji-Cuadrada con k grados de libertad.

Ahora, la función de densidad de una variable aleatoria X que se distribuye como una t-Student, con k grados de libertad está dada por

$$f(x) = \frac{\Gamma((k+1)/2)}{\sqrt{k\pi}\Gamma(k/2)} (1 + x^2/k)^{-\frac{(k+1)}{2}} \quad (11)$$

donde $x \in \mathbb{R}$ y la función $\Gamma(z)$ es la considerada anteriormente. En la Figura 5, puede observarse que la forma que toma esta densidad también depende de sus

grados de libertad. Es importante aclarar que las funciones de densidad aquí presentadas no son integrables y por ello la necesidad de utilizar tablas estadísticas, software estadístico o bien alguna aplicación que permita aproximar mediante algún método numérico, el área bajo estas curvas.

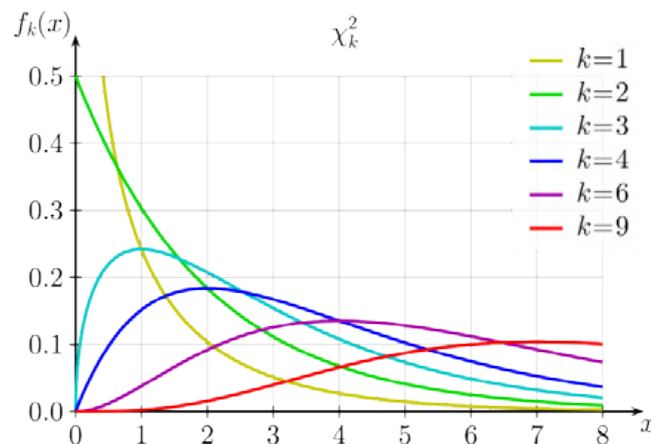


Figura 5. Funciones de densidad de variables aleatorias con distribución -Student, con dos grados distintos de libertad.

En la sección siguiente utilizaremos dos de las distribuciones antes presentadas para ejemplificar la utilidad de las funciones generadoras de momentos al analizar la distribución de ciertos estadísticos, aspecto muy importante en la inferencia estadística.

FUNCIONES GENERADORAS DE MOMENTOS

En esta sección veremos cómo las funciones generadoras de momentos nos permiten probar, de una manera relativamente sencilla, si dos distribuciones son iguales. Este material que en ocasiones no se cubre en los cursos de probabilidad es crucial para demostrar, de una manera inmediata, la distribución que siguen ciertos estadísticos que se utilizan en estadística inferencial.

La función generadora de momentos (fgm) de una variable aleatoria continua X es una función de valores reales, definida como:

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^\infty e^{tx} f(x) dx \quad (12)$$

donde $f(x)$ es la función de densidad de X . Un teorema que facilitará demostrar la distribución de ciertos estadísticos es el presentado por Mood [2] que se presenta a continuación.

Teorema 1. Sean X y Y variables aleatorias con densidades $f_X(\cdot)$ y $f_Y(\cdot)$, respectivamente. Supongamos que $M_X(t)$ y $M_Y(t)$ existen y son iguales para todo t en el intervalo $-h < t < h$ para algún $h > 0$. Entonces las dos funciones de distribución acumulada y $F_X(\cdot)$ y $F_Y(\cdot)$ son iguales.

Este teorema, como veremos en lo subsecuente, será de gran utilidad.

A continuación, se presentan las funciones generadoras de momentos para el caso de una variable aleatoria X que se distribuye normalmente, así como también para una variable aleatoria con distribución Ji-cuadrada.

Si X es una variable aleatoria que se distribuye normal, con media μ y varianza σ^2 , su función generadora de momentos puede expresarse como:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1(x-\mu)^2}{2\sigma^2}} dx \quad (13)$$

En el caso de que X se distribuya como una normal estándar sustuiremos los valores correspondientes de $\mu = 0$ y $\sigma^2 = 1$ y, de esta manera,

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2 + t^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} e^{-\frac{t^2}{2}} dx \\ &= e^{-\frac{t^2}{2}}. \end{aligned} \quad (14)$$

Ahora, supongamos que la variable aleatoria X es tal que $X \sim N(\mu, \sigma^2)$, entonces su fgm está dada por (13) y haciendo

$$z = \frac{x - \mu}{\sigma},$$

tenemos que $x = z\sigma + \mu$, y $dx = dz\sigma$, y con una simple sustitución obtenemos:

$$\begin{aligned} M_Z(t) &= \int_{-\infty}^{\infty} e^{t(z\sigma + \mu)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tz\sigma - \frac{z^2}{2}} dz \end{aligned} \quad (15)$$

Completando cuadrados en los exponentes tendremos que $tz\sigma - \frac{z^2}{2} = -\frac{1}{2}(z - t\sigma)^2 + \frac{1}{2}t^2\sigma^2$ y sustituyendo esto en (15) obtenemos,

$$\begin{aligned} M_Z(t) &= e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t\sigma)^2 + \frac{1}{2}t^2\sigma^2} dz \\ &= e^{t\mu + \frac{1}{2}t^2\sigma^2} \end{aligned} \quad (16)$$

A continuación, se muestra la función generadora de momentos del cuadrado de una variable aleatoria normal



estándar, que se denotará como $M_{Z^2}(t)$ y se calculará de forma muy análoga a la anterior.

$$\begin{aligned} M_{Z^2}(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz^2} e^{-\frac{z^2}{2}} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}(1-2t)} dz \end{aligned} \quad (17)$$

Sea $u = z\sqrt{1-2t}$, entonces,

$$\begin{aligned} M_U(t) &= (1-2t)^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du \\ &= (1-2t)^{-\frac{1}{2}} \end{aligned} \quad (18)$$

Esta última ecuación corresponde a la fgm del cuadrado de una normal estándar.

Para encontrar la función generadora de momentos de la suma de los cuadrados de normales estándar, recurriremos al siguiente teorema presentado en [3], y que establece:

Teorema 2. La función generadora de momentos de una suma de n variables aleatorias independientes, es igual al producto de las funciones generadoras de momentos de estas variables aleatorias, esto es,

$$M_{X_1 + \dots + X_n}(\theta) = M_{X_1}(\theta) M_{X_2}(\theta) \dots M_{X_n}(\theta) \quad (19)$$

Es claro que este teorema es bastante útil, pues una vez que conocemos la función generadora de momentos del cuadrado de una normal estándar, es posible encontrarla para la suma de los cuadrados de normales estándar.

Sea $W = \sum_{i=1}^n Z_i^2$ donde cada Z_i es una normal estándar, entonces

$$M_W(t) = (1-2t)^{-n/2} \quad (20)$$

Por otra parte, podemos calcular la función generadora de momentos de una variable aleatoria X , que sigue una distribución Ji-cuadrada con k grados de libertad, de la siguiente manera:

$$M_X(t) = \int_0^{\infty} e^{tx} \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2} dx$$

$$= \frac{1}{2^{k/2} \Gamma(k/2)} \int_0^{\infty} e^{-\frac{z}{2}(1-2t)} x^{\frac{k}{2}-1} dx \quad (21)$$

Sea $y = \frac{(1-2t)}{2}$, entonces $dx = \frac{2dy}{1-2t}$

$$M_Y(t) = \frac{1}{2^{k/2} \Gamma(k/2)} \int_0^{\infty} e^{-y} \left(\frac{2y}{1-2t}\right)^{\frac{k}{2}-1} \frac{2}{1-2t} dy$$

$$= \frac{(1-2t)^{-\frac{k}{2}}}{\Gamma(\frac{k}{2})} \int_0^{\infty} e^{-y} y^{\frac{k}{2}-1} dy \quad (22)$$

Y por la definición de la función Gamma presentada en (10), identificamos que lo que se encuentra dentro de esta integral es $\Gamma(\frac{k}{2})$ por lo que la función generadora de momentos de una variable aleatoria que se distribuye como una Ji-cuadrada con k grados de libertad es:

$$M_Y(t) = (1 - 2t)^{-\frac{k}{2}} \quad (23)$$

El material previamente presentado permitirá que en la siguiente sección obtengamos de manera muy sencilla la distribución muestral de \bar{x} y de $\frac{(n-1)s^2}{\sigma^2}$, los cuales permitirán realizar inferencias sobre la media y varianza de una población normal, tópicos que usualmente se enseñan en los cursos básicos de estadística.

DISTRIBUCIONES MUESTRALES DE \bar{X} Y DE $\frac{(N-1)s^2}{\sigma^2}$

Algunos de los estimadores más utilizados en los cursos básicos de estadística son la media y varianza muestrales. Conocer la distribución muestral de éstos nos permite realizar inferencias sobre los parámetros que están estimando, pudiendo así tomar decisiones. No se estudiarán aquí las propiedades de estos estimadores, pues éstas por lo general se cubren en los cursos de estadística.

Distribución muestral de \bar{X} , con σ^2 conocida

En un curso básico de estadística por lo general se enuncia el siguiente teorema y se dan algunas referencias donde puede verse su demostración, como puede ser [2]. Por medio de este teorema podemos conocer cómo se distribuye la media de muestras de tamaño n , esto es la distribución de \bar{X}_n .

Teorema 3. Si \bar{X}_n es la media de una muestra aleatoria de tamaño n tomada de una distribución normal con media μ y varianza σ^2 , entonces $\bar{X}_n \sim (\mu, \sigma^2/n)$

la demostración de este teorema resulta inmediata si utilizamos lo presentado sobre funciones generadoras de

momentos, pues

$$M_{\bar{X}_n}(t) = E(e^{t\bar{X}_n}) = E\left(e^{\frac{t\sum X_i}{n}}\right) = E\left(\prod_{i=1}^n e^{\frac{tX_i}{n}}\right)$$

$$= \prod_{i=1}^n E\left(e^{\frac{tX_i}{n}}\right) = \prod_{i=1}^n M_{X_i}(t/n) = \prod_{i=1}^n e^{\frac{\mu t}{n} + \frac{1}{2} \left(\frac{\sigma t}{n}\right)^2} = e^{\mu t + \frac{1}{2} \frac{(\sigma t)^2}{n}} \quad (24)$$

que resulta la función generadora de momentos de una distribución normal con media μ , pero con varianza $\frac{\sigma^2}{n}$.

Distribución muestral de $\frac{(n-1)S^2}{\sigma^2}$

Problemas que surgen de manera muy natural son los relacionados con estimar una varianza o bien realizar alguna prueba acerca de ésta. El apoyo para realizar esto es mostrar que si $s^2 = \frac{1}{(n-1)} \sum (X_i - \bar{X})^2$ es la varianza de una muestra aleatoria tomada de una distribución normal, con media μ y varianza σ^2 , entonces $u = \frac{(n-1)s^2}{\sigma^2}$ tiene una distribución Ji-cuadrada con $(n-1)$ grados de libertad.

Para demostrar lo anterior recurriremos a comparar sus fgm como lo hicimos en un caso anterior. Nótese que:

$$\sum Z_i^2 = \sum (Z_i - \bar{Z} + \bar{Z})^2$$

$$= \sum (Z_i - \bar{Z})^2 + n\bar{Z}^2$$

Es decir,

$$\sum Z_i^2 = \sum (Z_i - \bar{Z})^2 + n\bar{Z}^2 \quad (25)$$

y con esto,

$$\sum_{i=1}^n (Z_i - \bar{Z})^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \quad (26)$$

entonces $\frac{(n-1)s^2}{\sigma^2}$ se distribuye Ji-cuadrada con $n - 1$ g.l.

Con la finalidad de comprender el resultado anterior, recordemos que conocemos la fgm de $\sum Z_i^2$ y que, ahora buscamos conocer la de $\sum (Z_i - \bar{Z})^2$ para poder corroborar que efectivamente U sigue una distribución Ji-cuadrada.

En la ecuación (25) obtuvimos una suma de variables aleatorias, por lo que utilizando el Teorema 2 tenemos que:

$$M_{\sum Z_i^2}(t) = M_{\sum (Z_i - \bar{Z})^2}(t) M_{n\bar{Z}^2}(t)$$

y haciendo un sencillo despeje llegamos a

$$M_{\sum (Z_i - \bar{Z})^2}(t) = \frac{M_{\sum Z_i^2}(t)}{M_{n\bar{Z}^2}(t)} \quad (27)$$

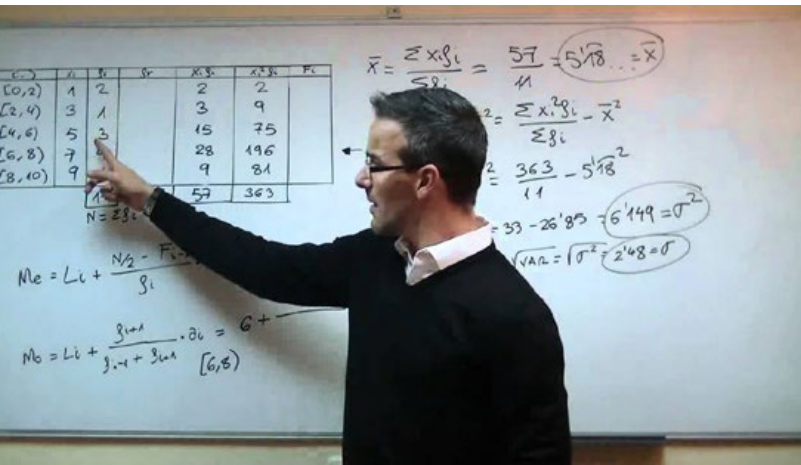
Conocemos la fgm del numerador en (28) y para el denominador sabemos que Z se distribuye como una normal, pero con varianza $\frac{\sigma^2}{n}$, donde $\sigma^2=1$, y al ser multiplicada por n obtendremos la misma fgm que para una variable aleatoria Z^2 . Luego entonces, tanto numerador como denominador se distribuyen como una Ji-cuadrada:

$$M_{\sum (Z_i - \bar{Z})^2}(t) = \frac{(1-2t)^{\frac{n}{2}}}{(1-2t)^{\frac{1}{2}}} = \left(\frac{1}{1-2t}\right)^{(n-1)/2} \quad t < 1/2, \quad (28)$$

SIMULACIÓN DE DISTRIBUCIONES MUESTRALES EN UN CURSO BÁSICO DE ESTADÍSTICA

Software estadístico: R

Las funciones generadoras de momentos son de gran importancia en probabilidad y estadística, ya que nos proporcionan un medio para probar la distribución de estadísticos y/o transformaciones de éstos. Aun así, volvemos a destacar que no en todos los cursos de estadística se puede esperar bases tan sólidas en matemáticas, ya que recordemos que la estadística se estudia en muchas áreas de servicio donde no se cuentan con estas bases. Por ello, para justificar lo que hemos visto acerca de las distribuciones de estadísticos y sus transformaciones se suele recurrir a simularlos computacionalmente. En la actualidad, las simulaciones a partir del software computacional son muy populares debido a lo práctico que esto resulta ser. En nuestro caso, podemos hacer uso de software estadístico que sea sencillo para los estudiantes, y a su vez, con las herramientas necesarias para nuestro objetivo. Tal es el caso de R [4], un software estadístico de entre los más populares en la actualidad, completamente libre, con una vasta documentación en línea y una gran comunidad. R es tan sencillo de usar que incluso en áreas como economía se ha implementado, donde no se espera experiencia en programación por parte de los estudiantes.



demonstrando con lo anterior que la fgm de $\sum(Z_i - \bar{Z})^2$ es igual a la fgm de una Ji-cuadrada con $n - 1$ grados de libertad.

Distribución muestral de \bar{x} , σ^2 desconocida

Es muy común que surja el problema de realizar inferencias acerca de la media de una población normal y que la varianza sea desconocida. En dicha situación no es posible recurrir a las funciones generadoras de momentos para estudiar la distribución muestral de la media, ya que la cantidad pivotal $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ resulta tener una distribución t -Student, la cual no posee función generadora de momentos. Sin embargo, en estos casos es posible auxiliarse de lo que vimos anteriormente y otros resultados como el siguiente:

Teorema 4: Si Z sigue una distribución normal estándar y U sigue una distribución Ji-cuadrada con $n - 1$ grados de libertad, donde Z y U son independientes, entonces $\frac{Z}{\sqrt{U/(n-1)}}$ sigue una distribución t -Student con $n - 1$ grados de libertad. Utilizando este teorema podemos ver que la cantidad pivotal $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ sigue una distribución t -Student con $n - 1$ grados de libertad pues,

$$\frac{\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}. \quad (29)$$

Esta cantidad permitirá construir intervalos de confianza para la media μ , así como realizar pruebas de hipótesis o bien pruebas de significancia para este parámetro.

En la siguiente sección abordaremos el tema de la simulación de distribuciones muestrales, ya que con ello podemos reconocer más fácilmente cómo se comportan ciertos estadísticos de interés en un curso básico de estadística. Sin embargo, hay que tener en consideración que para ello hay mucha teoría que lo sustenta, y que las funciones generadoras de momentos forman parte de las bases que hay que conocer si queremos hacer responsablemente simulación en estadística.

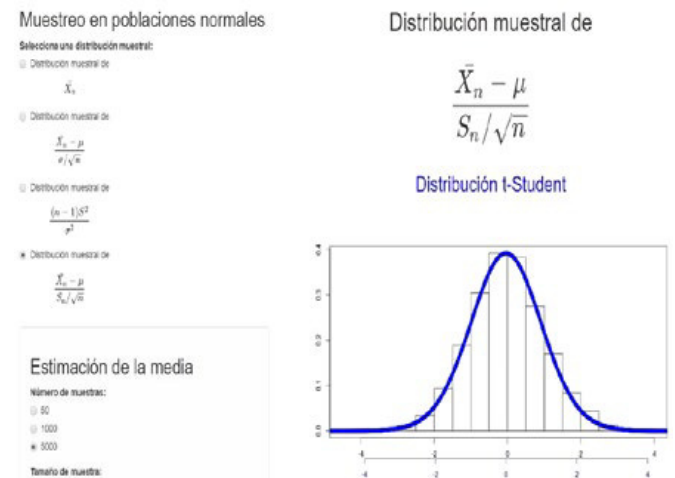


Figura 6. Interfaz gráfica de una aplicación web desarrollada con el paquete Shiny de R.

Shiny: Diseño de aplicaciones con R

Una útil herramienta cuando estamos trabajando en R es Shiny [5], una librería que permite el soporte de aplicaciones mucho más amigables para el usuario, como lo muestra la Figura 6. Así, si nuestro objetivo es mostrar distribuciones muestrales en un curso básico de estadística, con pocas bases de matemáticas y programación en software estadístico, el desarrollo de una aplicación con tal enfoque resulta ser sencillo a través de esta librería.



```

1 |
2 | # This is the user-interface definition of a shiny web application. You can
3 | # run the application by clicking 'Run App' above.
4 |
5 | # find out more about building applications with shiny here:
6 | #
7 | # http://shiny.rstudio.com/
8 | #
9 |
10 | library(shiny)
11 | s1 <- withMathjax("$$\bar{X}_{(n)}$$")
12 |
13 | shinyUI(fluidPage(
14 |   titlePanel("Muestreo en poblaciones normales"),
15 |   withMathjax(radioButtons("inRadio", "Selecciona una distribución muestral: ",
16 |     choiceNames = c("distribución muestral de $$\bar{X}_{(n)}$$",
17 |       "distribución muestral de $$\frac{\bar{X}_{(n)} - \frac{\mu}{\sqrt{n}}}{\sigma/\sqrt{n}}$$",
18 |       "distribución muestral de $$\frac{(n-1)S^2}{\sigma^2}$$",
19 |       "distribución muestral de $$\frac{\bar{X}_{(n)} - \frac{\mu}{\sqrt{n}}}{S_{(n)}/\sqrt{n}}$$"),
20 |     choicevalues = c("distrib3", "distrib4", "distrib2", "distrib1")
21 |   )),
22 |   sidebarLayout(
23 |     conditionalPanel(

```

Figura 7. Implementación de la vista de la aplicación mostrada en la figura 6.

Esto por supuesto implica que el profesor o desarrollador conozca R y Shiny, con el fin de desarrollar una aplicación para estudiantes, considerando que estos últimos no cuentan con el interés de programar, como se puede dar en muchas áreas. Shiny es una herramienta que nos sirve para el desarrollo de aplicaciones destinadas a estar en la red, por lo cual cuenta con funciones de HTML, algunos estilos de CSS, e incluso con la posibilidad de enlazar nuestros scripts con códigos de JavaScript, uno de los lenguajes más populares en la actualidad, como se ve en la Figura 7.

CONCLUSIONES

Es importante mencionar que los contenidos curriculares de los programas educativos universitarios se revisan regularmente y al adaptarlos a los cambios científicos y tecnológicos, en ocasiones se prescinde de temas que pueden ser fundamentos teóricos importantes en materias subsecuentes. Es tarea pues del maestro, encontrarlas herramientas que permitan a los estudiantes

comprender temas como el presentado en este escrito, incluso sin demostración matemática alguna. Esto generalmente implicará adaptarse a los tiempos actuales y aprovechar las nuevas tecnologías para ofrecer un contenido más didáctico y de fácil comprensión. En ocasiones esto no es tarea sencilla; sin embargo, el esfuerzo es gratamente recompensado. Herramientas como la aquí presentada permitirán que los estudiantes, sin importar sus bases de programación, computación o matemáticas, puedan comprender claramente el tema de distribuciones muestrales. Por tal razón, la aplicación aquí presentada, se encuentra disponible en línea y con libre de acceso, en el siguiente enlace: <https://victornoriega.shinyapps.io/distrib/>

Por otra parte, el trabajo colaborativo, extra-clase, entre maestros y alumnos es fundamental para lograr aplicaciones como la aquí presentada, la cual surgió del interés despertado en quien esto escribe, cuando cursé la materia de estadística con la Dra. Gudelia Figueroa, en el Departamento de Matemáticas. Su asesoría en este tema fue crucial ya que se logró no sólo este manuscrito, sino también una aplicación amigable y sencilla que permite a cualquier estudiante de un curso básico de estadística comprender el tema de distribuciones muestrales.

BIBLIOGRAFÍA

- [1] S. B. Kathleen, Psicología del desarrollo: Infancia y adolescencia, New York: Editorial Médica Panamericana, 2007.
- [2] A. M. Mood, F. A. Graybill and D. C. Boes, Introduction to the Theory of Statistics, New York: McGraw-Hill, 1974.
- [3] P. G. Hoel, Introduction to Mathematical Statistics, New York: John Wiley and Sons, 1964.
- [4] T. R. Foundation, "The R Project for Statistical Computing," Agosto 1993. [Online]. Disponible: <https://www.r-project.org/>. [Ingreso en 2019].
- [5] "Shiny," The R Foundation, [Online]. Disponible: <https://shiny.rstudio.com/>. [Ingreso en 2019].