

Distribución muestral

En la siguiente dirección de Internet se pueden realizar ejercicios de simulación de la distribución muestral de varios estadísticos: proporción, mediana, desviación típica o varianza. Veamos un ejemplo practicando con la distribución **muestral del estadístico de la media**. La aplicación para ejecutar la simulación de la distribución muestral (*Sampling Distribution Simulation*) se ofrece en la siguiente dirección de Internet: http://onlinestatbook.com/stat_sim/index.html

Una vez nos encontremos en la pantalla de la simulación, el usuario especifica:

1. la distribución de la población de la que se van a tomar las muestras aleatorias (distribución normal o gaussiana, distribución uniforme o rectangular, distribución asimétrica). También se puede cambiar la distribución normal que por defecto siempre aparece clicando con el ratón sobre la distribución original y de este modo se eliminan o añaden datos a la distribución (ver Figura 1). A la izquierda del primer histograma aparece la media, la mediana, la desviación típica (se trata del error típico o estandard ya que representa la desviación o variabilidad de los datos de la distribución muestral del estadístico), la asimetría, la curtosis de la distribución de la población. En el eje de las abscisas del histograma, una línea azul representa a la media de la población y una línea rosa a la mediana. Si la media y la mediana coinciden sólo aparece una línea ya que se superponen. La amplitud de la desviación típica de la media muestral se señala con un intervalo en rojo graficando una desviación típica a cada lado de la media. Los colores se corresponden con los utilizados a la izquierda de la distribución para los parámetros (media en azul, mediana en rosa y la desviación típica en rojo).

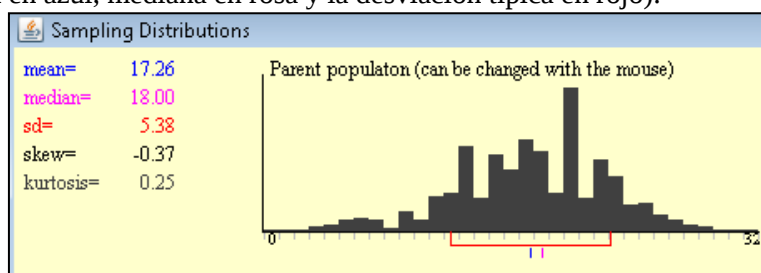


Figura 1. Distribución poblacional no normal

2. el tamaño de la muestra y el estadístico. Los estadísticos que pueden ser computados en la simulación son la media, la desviación típica de la media (N como denominador), la varianza de la muestra (N como denominador), la estimación insesgada de la varianza (N-1 como denominador), el valor absoluto medio de la desviación de la media y la amplitud. Y el tamaño de cada muestra puede ser de 2, 5, 10, 16 y 20 elementos. No hay que confundir tamaño de la muestra con número de muestras tomadas o réplicas.

Conviene recordar que la asimetría (sesgo de la distribución, skewness) y la curtosis (apuntamiento, kurtosis) es cero cuando la distribución es normal. Si la distribución muestral del estadístico se obtiene con tamaños de muestra mayor (ejemplo $N=25$) la asimetría y la curtosis será menor que si se obtienen con medias muestrales de menor tamaño (por ejemplo $N=5$).

El usuario debe especificar un tamaño concreto para las muestras que serán extraídas de forma aleatoria de la población, por ejemplo de 25 (Mean $N=25$) y clicar a continuación sobre "Sample: Animated". El *applet* de la simulación devuelve un segundo histograma con las frecuencias de los 25 datos y grafica en el histograma tercero la primera media muestral obtenida con esos 25 datos y progresivamente irá construyendo la distribución muestral del estadístico (ver Figura 2).

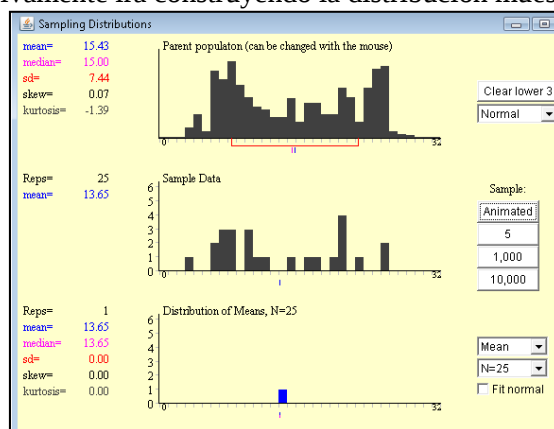


Figura 2. Representación en el segundo histograma de los 25 primeros datos de una media muestral ($N=25$) y representación de la primera media muestral obtenida con esos 25 datos en el tercer histograma que representa la distribución muestral de las medias

Este proceso de muestreo aleatorio se puede repetir una y otra vez para estimar la distribución muestral del estadístico. Se observará que a medida que aumenta el tamaño de medias muestrales, la distribución muestral de las medias se asemeja a la distribución normal (ver Figura 3 donde se han graficado 10003 medias muestrales con un tamaño de 25 elementos, $N=25$). Más adelante veremos por qué sucede este fenómeno gracias al Teorema del límite central.

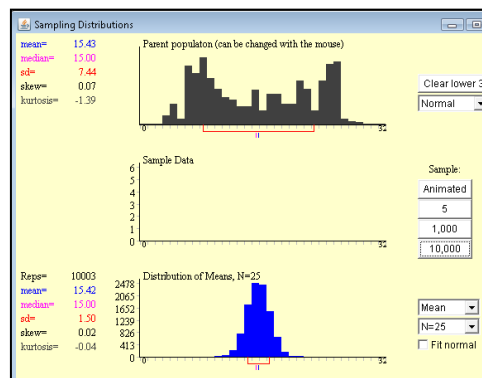


Figura 3. Representación en el tercer histograma de 10003 medias muestrales

En la figura 8 se representa la pantalla de la simulación y se detalla cada uno de sus elementos. Los datos de la figura 8 indican que se ha seleccionado la distribución normal como población para extraer las muestras. El estadístico que se va a estimar en la distribución es la media (Mean). Es decir, la distribución poblacional subyacente es la distribución normal con media y mediana de 16, desviación típica de 5 y la asimetría (skewness) y el apuntamiento (kurtosis) es igual a cero. En la distribución de medias que se ha graficado en el tercer histograma se han utilizado 1001 muestras (réplicas) siendo el tamaño de cada muestra de 10. La distribución muestral de medias obtenida con tamaños de muestra de 10 datos ($N = 10$) tiene una media de 15.97, una mediana de 16, una desviación típica (error típico) de 1.59, la asimetría es de 0.06 y la curtosis es igual a -0.08.

El programa permite ejecutar al mismo tiempo una segunda distribución muestral donde se puede cambiar el tamaño de las muestras (aparecería en un cuarto histograma) y realizar comparaciones entre las dos distribuciones muestrales.

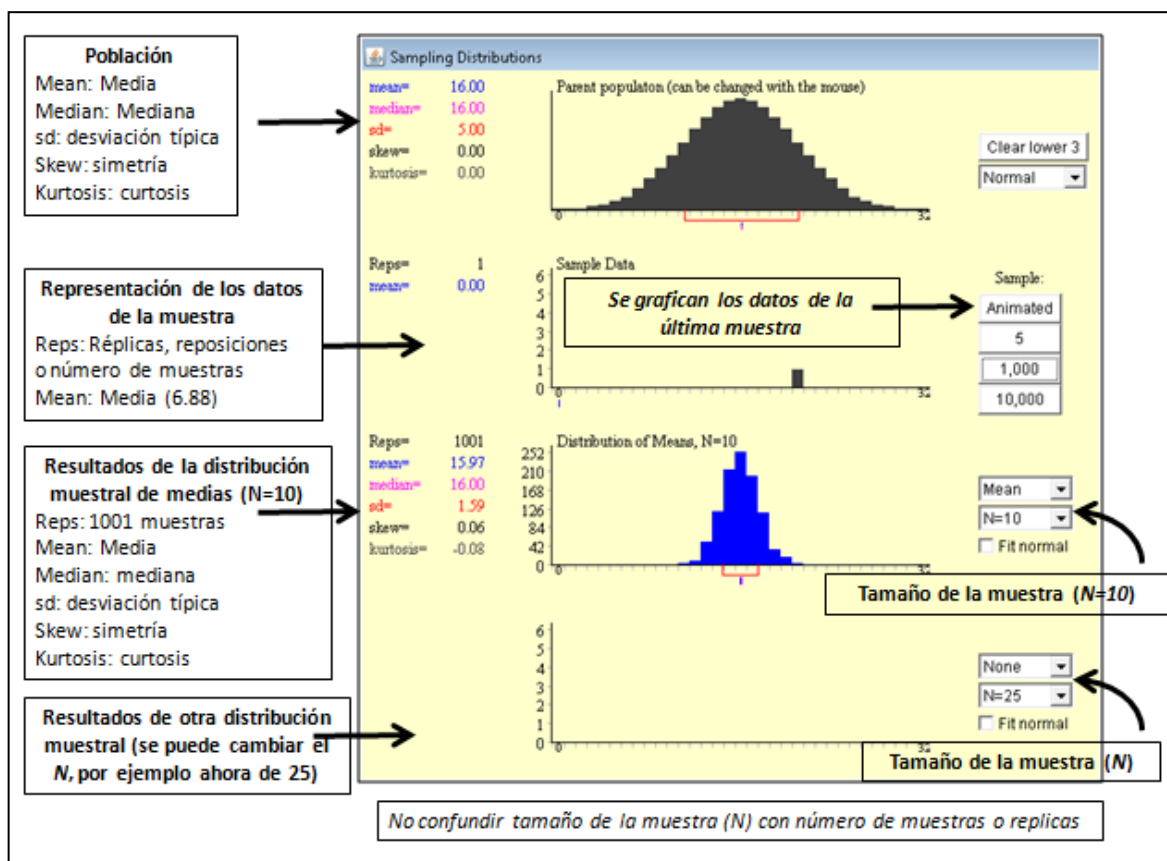


Figura 4. Simulación de la distribución muestral de la media