Universidad del Valle
Facultad de Ingeniería
Escuela de Estadística
Programa Académico de Estadística
Simulación Estadística
Cali, mayo de 2018

TEOREMA CENTRAL DEL LIMITE
PRUEBAS DE NORMALIDAD.

Kevin Steven García[a], Cesar Andres Saavedra[b]

1. TEOREMA CENTRAL DEL LIMITE.

El teorema del límite central es un teorema fundamental de probabilidad y estadística. El teorema describe la distribución de la media de una muestra aleatoria proveniente de una población con varianza finita. Cuando el tamaño de la muestra es lo suficientemente grande, la distribución de las medias sigue aproximadamente una distribución normal. El teorema se aplica independientemente de la forma de la distribución de la población. El teorema de límite central le permite aplicar estos procedimientos útiles a poblaciones que son considerablemente no normales. El tamaño que debe tener la muestra depende de la forma de la distribución original. Si la distribución de la población es simétrica, un tamaño de muestra de 5 podría producir una aproximación adecuada. Si la distribución de la población es considerablemente asimétrica, es necesario un tamaño de muestra más grande.

1.1. Teorema del límite central.

Si $\bar{X}$ es la media de una muestra aleatoria de tamaño n obtenida desde una población con cualquier distribución con media μ_X y varianza σ_X^2, entonces cuando n es grande, $\bar{X}$ sigue una distribución normal con $\mu_{\bar{X}} = \mu_X$ y $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$

Si $\bar{X}$ es la media de una muestra aleatoria de tamaño n que se toma de una población $N(\mu_X, \sigma_X^2)$ conocida, entonces:

$$Z = \frac{\bar{X} - \mu_X}{\frac{\sigma_x}{\sqrt{n}}}$$

Entonces Z se aproxima a una variable normal N(0,1), mejorándose la calidad de la aproximación a medida que el n aumenta.

- El teorema del límite central garantiza una distribución aproximadamente normal cuando n es suficientemente grande.
- Existen diferentes versiones del teorema, en función de las condiciones utilizadas para asegurar la convergencia. Una de las más simples establece que es suficiente que las variables que se suman sean independientes, idénticamente distribuidas, con valor esperado y varianza finitas.
- La aproximación entre las dos distribuciones es, en general, mayor en el centro de las mismas que en sus extremos o colas, motivo por el cual se prefiere el nombre "teorema del límite central".

[a]Código: 1533173. E-mail: kevin.chica@correounivalle.edu.co
[b]Código: 1628466. E-mail: cesar.saavedra@correounivalle.edu.co

2. VARIABLES ALEATORIAS.

2.1. Introducción.

En esta sección se evidencian los resultados de las simulaciones de las distribuciones de probabilidad para tamaños de muestra y distintos parámetros, así poder observar su comportamiento y dar respuesta a los siguientes interrogantes:

- Para que valor de n se puede afirmar el cumplimiento del teorema.
- La distribución de probabilidad de X afecta la convergencia.
- Los parámetros de la distribución de probabilidad de X afecta la convergencia.

El teorema referenciado en el numeral 1 del presente documento se llevara a cabo haciendo uso de dos distribuciones, donde: la distribución Poisson y la distribución Logística fueron simuladas por medio de las funciones ya incluidas en el programa estadístico r. Teniendo en cuenta estas simulaciones se aplico la prueba Cramér?Von Mises y se comparó los resultados para así saber si las simulaciones para los distintos tamaños de muestra y parámetros de verdad cumplen con el Teorema Central Del Limite y saber si estas convergen o no a la distribución normal. Posteriormente, se concluye acerca de los métodos de simulación usados teniendo en cuenta los resultados de la prueba.

2.2. Descripción general de las distribuciones y las pruebas de bondad de ajuste.

2.2.1. Distribución Poisson.

Esta distribución es una de las más importantes distribuciones de variable discreta. Sus principales aplicaciones hacen referencia a la modelización de situaciones en las que interesa determinar el número de hechos de cierto tipo que se pueden producir en un intervalo de tiempo o de espacio, bajo presupuestos de aleatoriedad. Su función de densidad está dada por:

$$f(x,\lambda) = \frac{e^{-\lambda}\cdot\lambda^{x}}{x!}; x \in \{0,1,2,3,...\}$$

2.2.2. Distribución Logística.

La función de distribución de la logística se utiliza como modelo de crecimiento. Por ejemplo, con un nuevo producto a menudo encontramos que el crecimiento es inicialmente lento, entonces gana impulso, y finalmente se ralentiza cuando el mercado está saturado o se alcanza alguna forma de equilibrio. Esta distribución tiene uso en los modelos logit, ya que estos modelos trabajan con su función de distribución acumulada denominada función logistica, por la simpleza de esta para encontrar probabilidades. Su función de distribución acumulada está dada por:

$$F(x,a,b) = \frac{1}{1+e^{-\left(\frac{x-a}{b}\right)}}$$

2.2.3. Prueba Cramer-Von Mises.

3. SIMULACIÓN DE VARIABLES ALEATORIAS Y PRUEBAS DE NORMALIDAD

3.1. Simulación de las observaciones:

3.1.1. Distribución Poisson:

Es importante recordar que los parámetros sobre los cuales se van a generar las observaciones son ($\lambda = 1$), ($\lambda = 5$), ($\lambda = 10$), con 20000 simulaciones para las cuales se sacaron muestras de tamaño ($n = 1$), ($n = 10$), ($n = 25$), ($n = 50$) datos. La generación de estos números se hizo por medio del comando rpois del software estadístico R.

3.1.2. Distribución Logística:

El uso de esta distribución en la problemática del articulo es una función derivada de ella, es decir, no se usa directamente la función de densidad logística para describir el fenómeno, se usa su función de distribución acumulada, denominada también función logística. Teniendo en cuenta lo anterior, se generaron datos con distribución Logística de parametros $(a = 0, b = 1)$, $(a = 9, b = 4)$, $(a = 15, b = 6)$, con 20000 simulaciones para las cuales se sacaron muestras de tamaño $(n = 1)$, $(n = 10)$, $(n = 25)$, $(n = 50)$ datos.

Prueba Cramer-Von Mises.

4. Resultados

4.0.1. Distribucción Poisson:

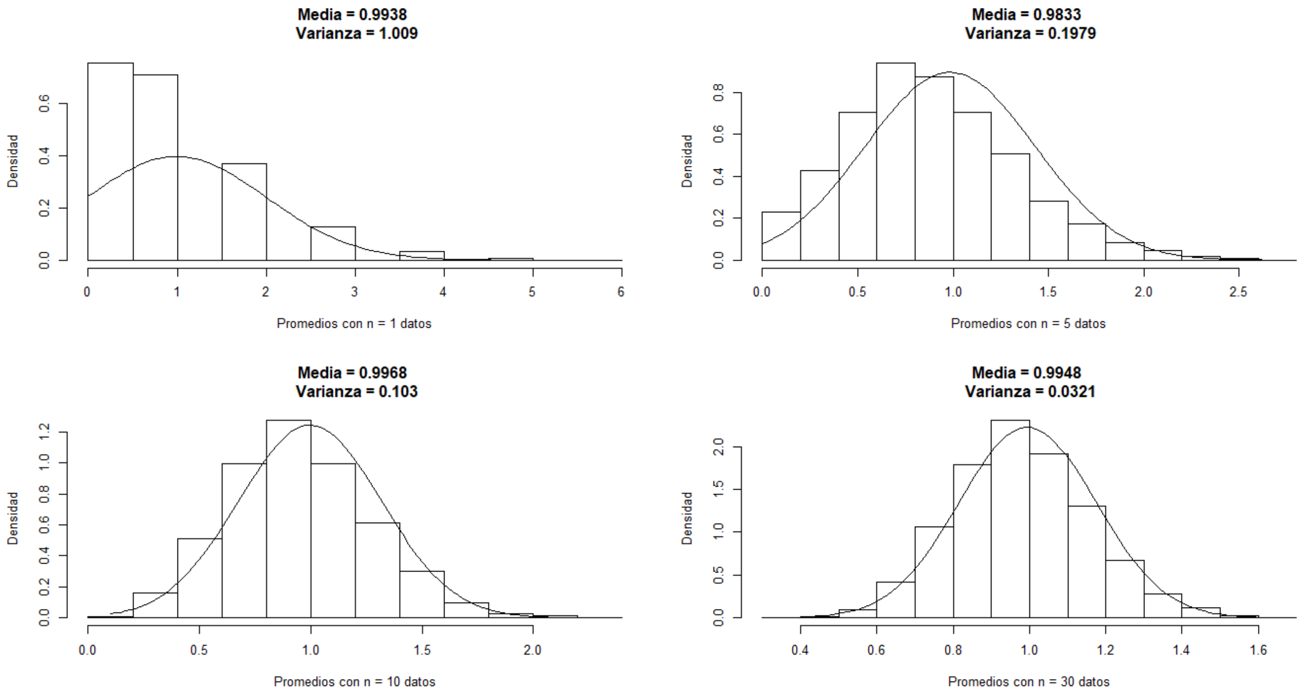

FIGURA 1: Simulaciones en R de la distribución Poisson($\lambda = 1$) con muestras de tamaño ($n = 1$), ($n = 10$), ($n = 25$), ($n = 50$)

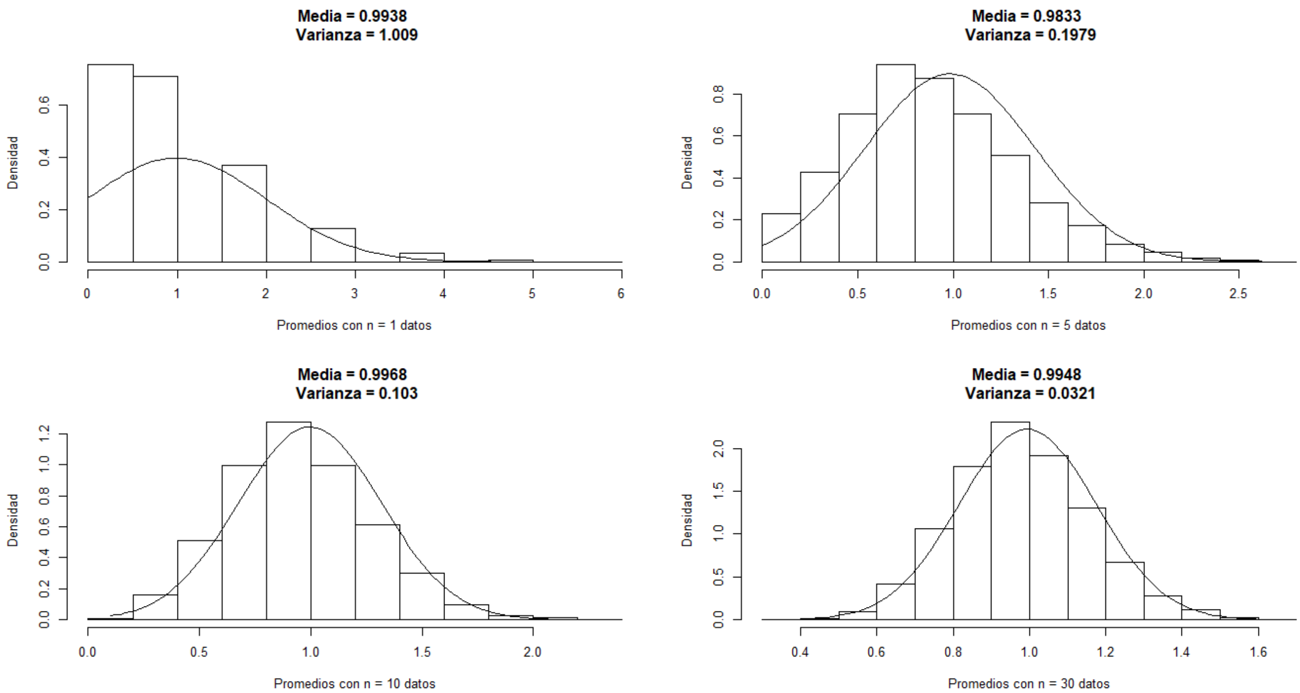

FIGURA 2: Simulaciones en R de la distribución Poisson($\lambda = 1$) con muestras de tamaño ($n = 1$), ($n = 10$), ($n = 25$), ($n = 50$)

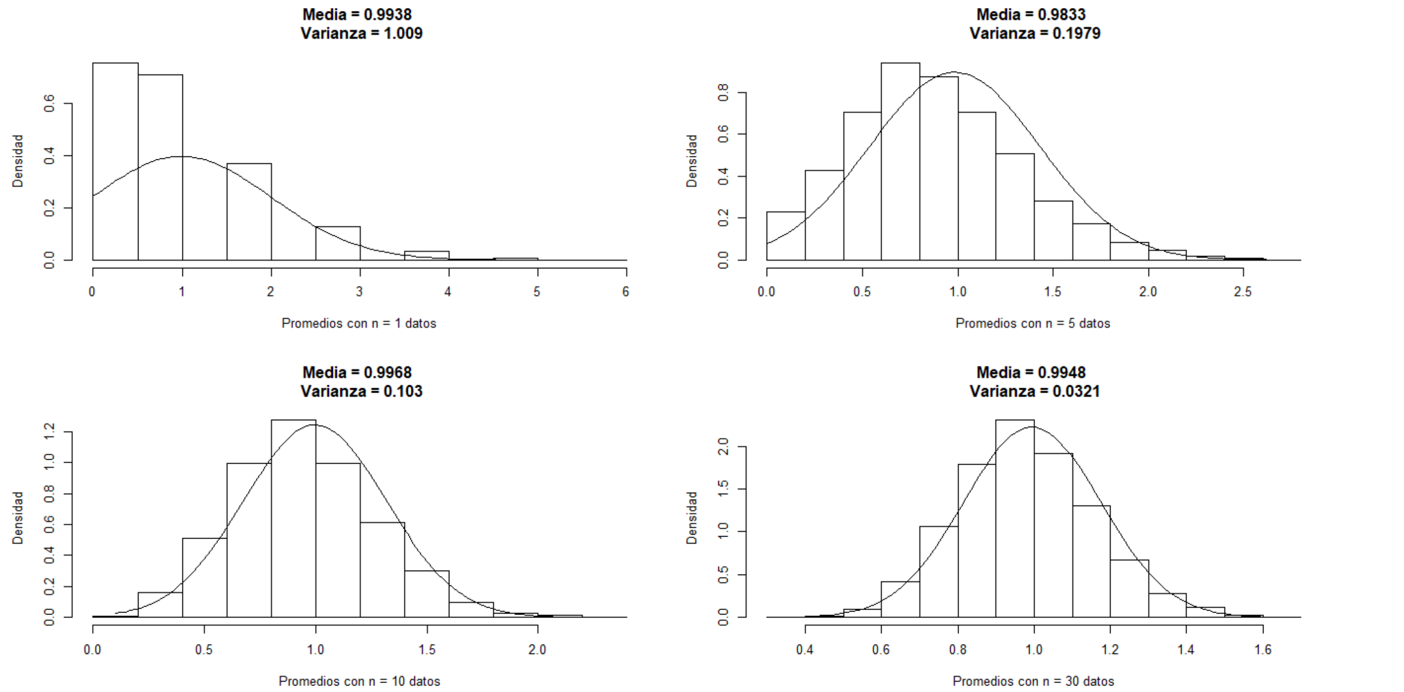

FIGURA 3: Simulaciones en R de la distribución Poisson($\lambda = 1$) con muestras de tamaño ($n = 1$), ($n = 10$), ($n = 25$), ($n = 50$)

En esta gráfica se puede observar como los valores obtenidos para la distribucion de poisson por el generador de R son casi identicos a los valores obtenidos por el generador propio de la distribución Poisson, por lo cuál se espera que la prueba de bondad de ajuste arroje que la distribución de los datos generados por el método propio de la Poisson realmente sigue una distribución Poisson($\lambda = 1$), ($\lambda = 5$), ($\lambda = 10$)

Prueba Cramer-Von Mises.

4.0.2. Distribucion Logistica:

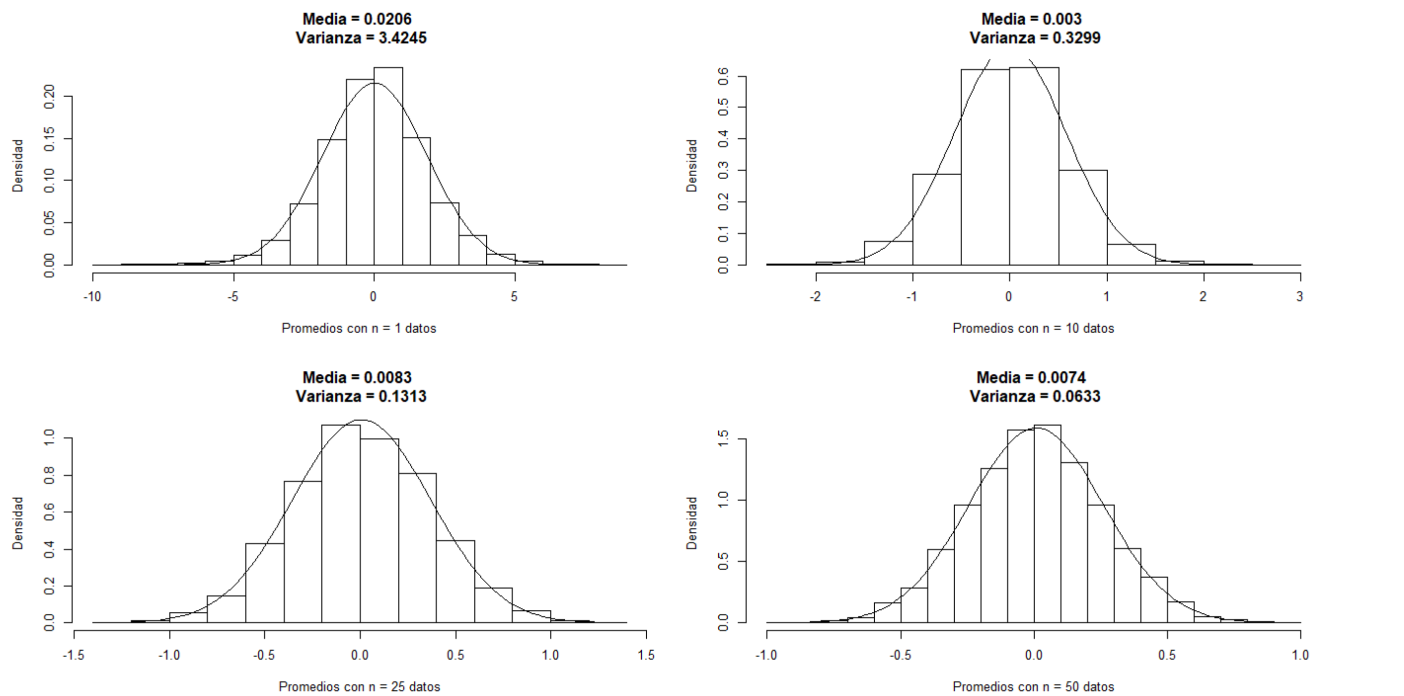

FIGURA 4: Simulaciones en R de la distribución Logística($a = 0, b = 1$) con muestras de tamaño ($n = 1$), ($n = 10$), ($n = 25$), ($n = 50$)

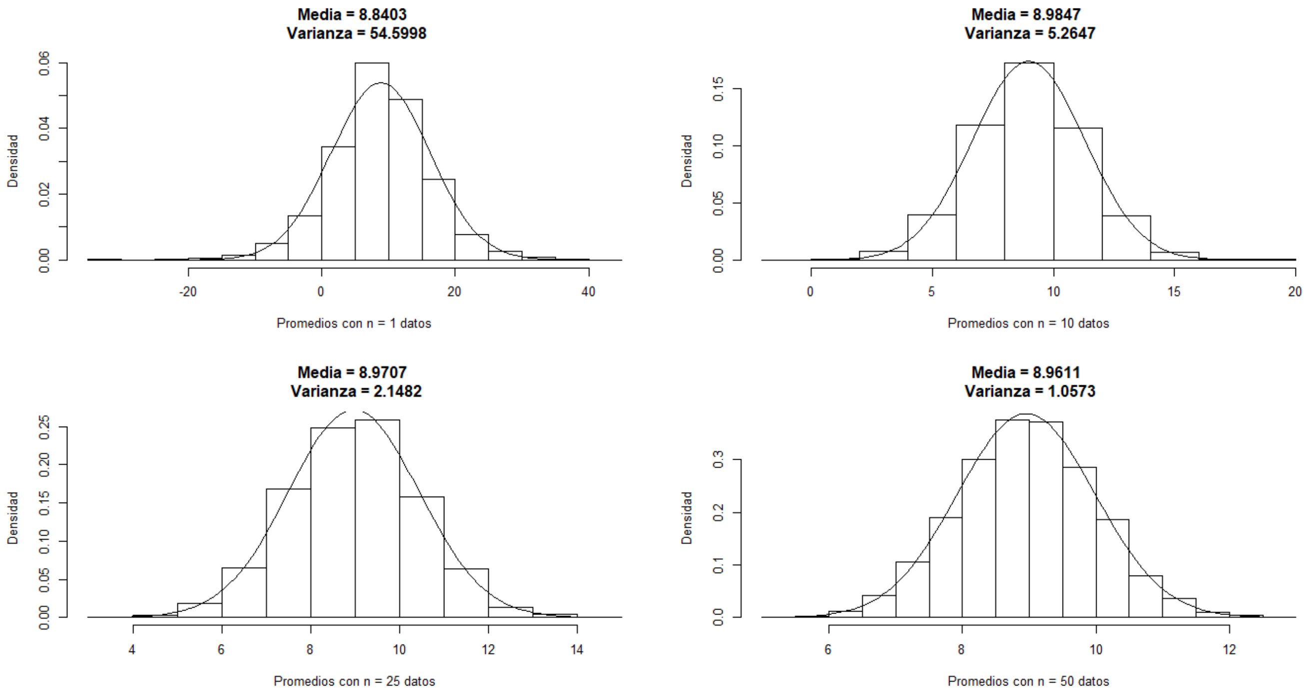

FIGURA 5: Simulaciones en R de la distribución Logística($a = 9, b = 4$) con muestras de tamaño ($n = 1$), ($n = 10$), ($n = 25$), ($n = 50$)

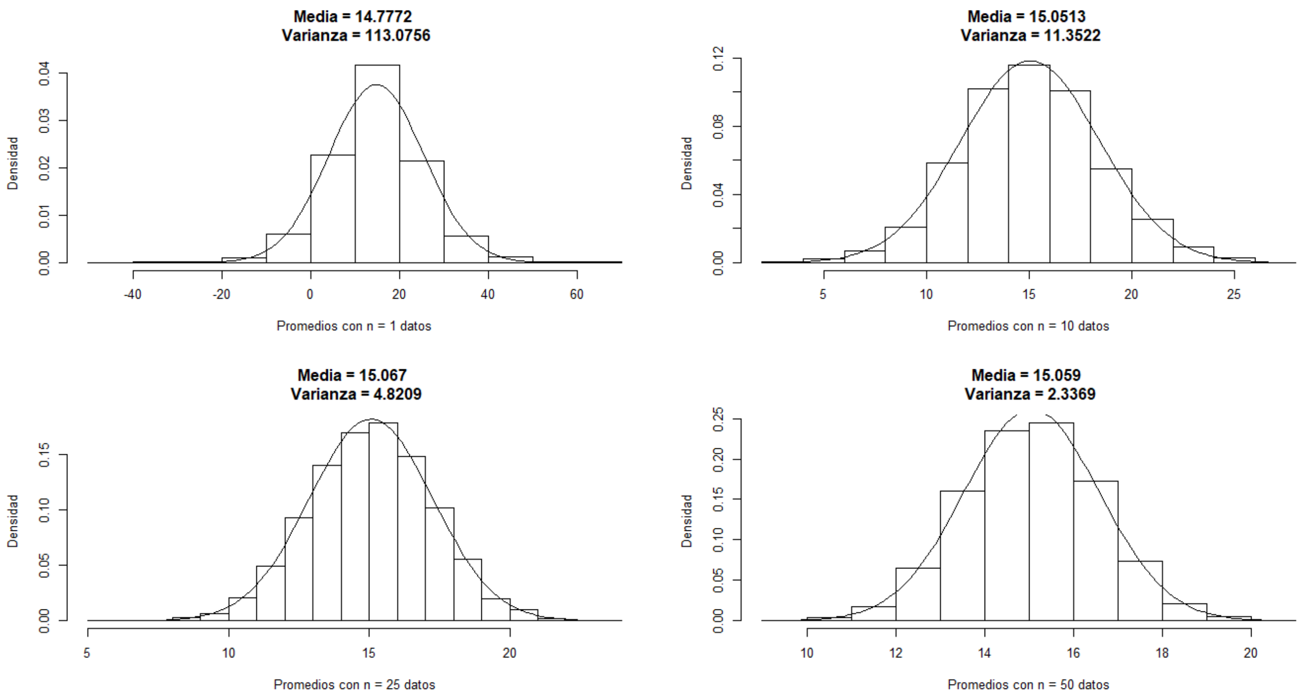

FIGURA 6: Simulaciones en R de la distribución Logística($a = 15, b = 6$) con muestras de tamaño ($n = 1$), ($n = 10$), ($n = 25$), ($n = 50$)

En esta gráfica se puede ver claramente como el histograma de los datos generados de la distribución Logistica muestra un comportamiento que sigue realmente la distribución esperada.

Prueba Cramer-Von Mises.

5. Conclusiones

-

5.1. Bibliografía

-
-
-

6. Scripts