

Actividad 2

Adrada Isabel, De la Peña Juan, Terán Federico, Troncoso Samuel
Pontificia Universidad Javeriana Cali

Resumen

Este estudio explora las propiedades estadísticas de una población simulada de 1000 números aleatorios y sus muestras, validando el Teorema del Límite Central (TLC). Mediante el análisis de distribuciones de frecuencias, histogramas y medias muestrales

Key words

Teorema del Límite Central, Distribución muestral, Media muestral, Inferencia estadística.

Introducción

En estadística, el Teorema del Límite Central (TLC) es un pilar fundamental que justifica la aproximación normal de las medias muestrales, independientemente de la distribución de la población original [1]. Este trabajo analiza una población simulada de 1000 datos generados aleatoriamente (0-1000) y sus muestras. En esta actividad aplicamos conceptos de estadística descriptiva para analizar una población generada aleatoriamente. Calculamos medidas como la media, mediana, moda y construimos tablas de frecuencias e histogramas para estudiar su comportamiento. Además, obtuvimos muestras aleatorias para comparar sus promedios con el de la población total y analizar la variabilidad entre ellas. Complementamos este trabajo con el análisis de dos artículos científicos relacionados con las distribuciones probabilísticas: una discreta, enfocada en la optimización de recursos en redes inalámbricas, y otra continua, aplicada al modelado de sistemas de colas. Esto nos permitió conectar la teoría con aplicaciones reales, reforzando nuestra comprensión del papel que juegan estas distribuciones en diferentes contextos de la ingeniería.

Métodos

En primer lugar, se generaron aleatoriamente 1000 números entre 1 y 100000, los cuales representan la población objetivo del presente estudio mediante la función `sample`.

```
“ r poblacion <- sample(0:1000, 1000, replace = TRUE) “
```

Esta población generada aleatoriamente se guardó en un archivo `datos.csv` a través del código presentado en el archivo `Poblacion.R`, por lo cual en este documento se trabajará con el data frame `datos`, cargado en el presente entorno a partir del archivo `csv` generado.

El promedio de la población objetivo de los 1000 números se obtuvo utilizando la función `mean` sobre el vector `datos` generado anteriormente.

```
“ r prom <- mean(datos) “
```

Para generar una tabla de frecuencia de la población, se realizaron k intervalos, donde k es igual a 10, con un ancho de $(\text{Max} - \text{Min})/k$ por intervalo. El conteo de la cantidad de datos dentro de un determinado conforma la frecuencia absoluta, se presenta además la frecuencia relativa, frecuencia absoluta acumulada y frecuencia relativa acumulada.

Por otro lado, se graficaron los datos en un histograma para determinar de manera visual el tipo de distribución probabilística de los datos analizados.

Para obtener 5 muestras sin reposición de tamaño 10 de la población de los 1000 números aleatorios se utilizó la función `sample` para obtener un vector con una muestra aleatoria.

```
“ r muestra1 <- sample(datos, size = 10, replace = FALSE) “
```

Para obtener el promedio muestral de cada una de las muestras obtenidas se utilizó la misma

metodología del promedio de la población con la función mean.

```
“ r prom1 <- mean(muestra1) “
```

Posteriormente, se graficó un histograma con las frecuencias de los promedios de las muestras de tamaño 10 con el objetivo de evaluar la forma de la distribución.

Resultados

Utilizando el método planteado anteriormente, el promedio de la población objetivo de los 1000 números es de 486.287.

En la Tabla 1 se presentan las frecuencias de los

datos distribuidas en 10 intervalos, donde se observa como todas las frecuencias absolutas para cada uno se encuentran en un rango de 82 a 113, por lo cual es posible que los datos presenten una distribución probabilista relativamente uniforme. Esto es reafirmado por la frecuencia relativa, la cual varía entre 0.08 y 0.11.

Las afirmaciones anteriores llevan a afirmar que la distribución no es normal, sin embargo para rectificar esta afirmación se calculó la mediana y moda, las cuales son 480 y 653, 696, 940 respectivamente. Esto evidencia que aunque la media y la mediana tienen valores relativamente similares, la moda difiere significativamente de ambos valores, por lo cual se rectifica que la distribución de los datos no es normal.

Table 1: Frecuencia de los datos

Intervalo	F.Absoluta	F.Relativa	F.Abs.Acum	F.Rel.Acum
[0, 100]	105	0.10	105	0.10
(100, 200]	109	0.11	214	0.21
(200, 300]	113	0.11	327	0.33
(300, 400]	97	0.10	424	0.42
(400, 500]	95	0.10	519	0.52
(500, 600]	108	0.11	627	0.63
(600, 700]	100	0.10	727	0.73
(700, 800]	82	0.08	809	0.81
(800, 900]	95	0.10	904	0.90
(900, 1000]	96	0.10	1000	1.00

A través de la Figura 1 se presenta la distribución de los datos en un histograma para verificar visualmente la distribución propabilística de los datos. En esta se observa como los datos en los diferentes intervalos tienen frecuencias similares, sin embargo no es una distribución exactamente uniforme debido a las diferencias en las frecuencias del histograma y se muestra una leve tendencia de disminución en la frecuencia a medida que aumenta el intervalo.

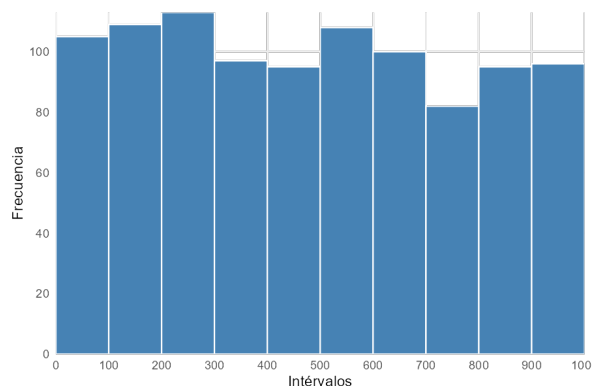


Figura 1. Distribución de los datos.

A continuación se presentan 5 muestras aleatorias sin reposición de tamaño $n = 10$:

Muestra 1: 517, 355, 383, 478, 57, 318, 995, 478, 572, 347

Muestra 2: 172, 319, 400, 589, 246, 259, 824, 681, 854, 608

Muestra 3: 650, 869, 76, 166, 324, 918, 113, 739, 891, 34

Muestra 4: 517, 276, 543, 318, 140, 102, 528, 189, 982, 956

Muestra 5: 125, 256, 184, 93, 347, 542, 958, 454, 680, 26

Estas muestras tienen los promedios muestrales 450, 495.2, 478, 455.1 y 366.5 respectivamente, mostrando una alta variabilidad de la media de las muestras obtenidas.

Los promedios de las muestras de tamaño 10 siguen una distribución aproximadamente normal (en forma de campana de Gauss), a pesar de que la población original no necesariamente tenga esta forma. Esto se debe al Teorema del Límite Central (TLC), que establece que, bajo condiciones generales, la distribución de las medias muestrales se aproxima a una normal a medida que aumenta el tamaño de la muestra, incluso si los datos originales no son normales.

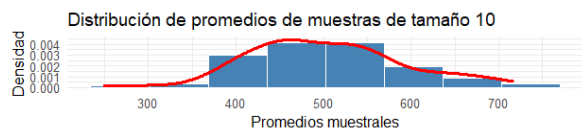


Figura2: Distribución de frecuencias de la población.

El promedio de las medias muestrales resulta casi idéntico al de la población, confirmando que la media muestral es un estimador insesgado. Por otro lado, la dispersión de las medias muestrales es notablemente menor que la de la población original, lo que concuerda con la teoría estadística: la varianza de las medias disminuye proporcionalmente al tamaño de la muestra ($1/n$). Esto explica por qué muestras más grandes producen estimaciones más precisas. Con muestras pequeñas (como $n \geq 10$), la aproximación a la normal puede no ser perfecta, pero ya se observa una tendencia hacia la simetría y concentración alrededor de la media poblacional. Si el tamaño de muestra fuera mayor (ej. $n \geq 30$), la distribución sería aún más claramente normal.

Análisis de resultados

A partir de los resultados obtenidos se puede determinar que la distribución de los datos generados aleatoriamente como población tiene una dis-

tribución probabilística relativamente similar a la distribución uniforme.

Al obtener muestras de 10 datos de la población se encuentra que las medias de las diferentes muestras presentan una significativa variación entre sí.

Este ejercicio práctico valida tres principios clave: (1) el Teorema del Límite Central, pues las medias muestrales siguen una distribución normal aunque la población no lo haga; (2) la media muestral es un estimador robusto de la media poblacional; y (3) el muestreo reduce la variabilidad, permitiendo inferencias confiables incluso con muestras pequeñas. Estos hallazgos justifican el uso de técnicas estadísticas basadas en muestras para analizar poblaciones grandes

Análisis artículo: distribución continua

En el artículo An Application to Continuous Probability Distribution, escrito por B. V. Dhandra y S. S. Kalyani, se presenta un enfoque práctico sobre el uso de distribuciones de probabilidad continuas para analizar el comportamiento de sistemas de colas. Este trabajo se enfoca en modelar la acumulación de personas en distintos escenarios usando funciones de densidad, destacando su aplicabilidad en contextos urbanos y rurales.

El artículo presenta una aplicación de las distribuciones de probabilidad continuas para modelar la cantidad de personas que se acumulan en un sistema de colas durante un intervalo de tiempo. A través del uso de funciones de densidad, se logra estimar con mayor precisión la probabilidad de que cierto número de personas esté presente en el sistema, lo cual es esencial para evaluar la eficiencia de distintos entornos de atención, como sucursales bancarias urbanas y rurales.

Se realizaron comparaciones entre dos sistemas de colas con diferentes tasas de llegada (λ), observando que un mayor valor de λ genera una mayor acumulación de personas, lo que indica menor eficiencia operativa. A través de esta técnica se pueden tomar decisiones basadas en datos para mejorar la asignación de recursos y la planificación de atención al cliente.

Este estudio resulta especialmente relevante para estudiantes y profesionales de Ingeniería de Sistemas, ya que proporciona herramientas matemáticas para el modelado de sistemas reales. Además,

promueve el uso de distribuciones probabilísticas como base para simulaciones, análisis de rendimiento y optimización de procesos, habilidades claves en áreas como redes, ingeniería del software y sistemas inteligentes.

Conclusiones

En este estudio, se analizó una población aleatoria de 1000 datos para evaluar sus características estadísticas y distribución probabilística. Los resultados indican que la distribución es relativamente uniforme, aunque con ligeras variaciones entre los intervalos. Las medidas estadísticas, como la media, mediana y moda,⁴ confirmaron que los datos no presentan una distribución normal. Además, las muestras aleatorias extraídas de la población mostraron una alta variabilidad en sus medias, destacando la importancia de la representatividad muestral. Por otro lado, se revisó el artículo científico “Method for Optimal Allocation of Network Resources Based on Discrete Probability Model” de Zhengqiang Song y Guo Hao. Este trabajo propone un modelo innovador para optimizar recursos en

redes inalámbricas, mejorando la eficiencia y minimizando el consumo energético mediante análisis probabilísticos y algoritmos adaptativos. Su relevancia radica en los avances que aporta al Internet de las Cosas (IoT) y las redes dinámicas.

En conclusión, tanto el análisis estadístico como la revisión del artículo destacan el valor de los modelos probabilísticos para abordar problemas complejos, desde la evaluación de datos hasta la optimización de sistemas tecnológicos.

Referencias

- [1] W. Navidi, Statistics for Engineers and Scientists w/ CD-ROM. McGraw-Hill Sci./Eng./Math, 2004.
- [2] Z. Song and G. Hao, “Method for optimal allocation of network resources based on discrete probability.
- [3] B. V. Dhandra and S. S. Kalyani, ”An Application to Continuous Probability Distribution,” Jul. 2023. model,” Wireless Networks, 2022. DOI: 10.1007/s11276-021-02727-7.