Actividad 2

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

Resumen

...

Key words

...

Introducción

... ... según Navidi [1] ...

Métodos

En primer lugar, se generaron aleatoriamente 1000 números entre 1 y 100000, los cuáles 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 cuál 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 intérvalos, donde k es igual a 10, con un ancho de (Max - Min)/k por intérvalo. 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 delos 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 intérvalos, donde se observa como todas las frecuencias absolutas para cada uno se encuentran en un rango de 82 a 113, por lo cuál es posible que los datos presenten una distribución probabilista relativamente uniforme. Esto es reafirmado por la frecuencia relativa, la cuál 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 cuáles 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 cuál 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 intérvalos 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 intérvalo.

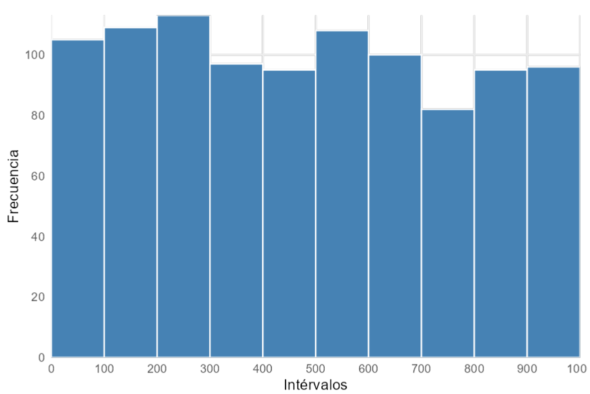

Figura 1. Distribución de los datos.

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

Muestra 1: 242, 130, 594, 449, 11, 853, 906, 355, 22, 474

Muestra 2: 536, 407, 715, 385, 954, 717, 185, 284, 443, 689

Muestra 3: 724, 726, 531, 534, 485, 49, 881, 631, 383, 967

Muestra 4: 844, 125, 465, 421, 464, 60, 511, 567, 872, 861

Muestra 5: 881, 803, 478, 117, 988, 988, 559, 412, 869, 597

Estas muestras tienen los promedios muestrales 403.6, 531.5, 591.1, 519 y 669.2 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.

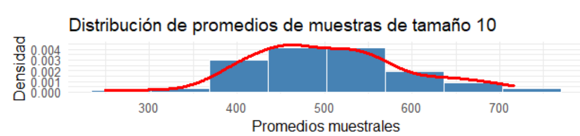

Figure 1: Mi gráfico

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 ($^2/n$). Esto explica por qué muestras más grandes producen estimaciones más precisas. Con muestras pequeñas (como n >=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 > =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 distribución prbabilística 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

Conclusiones

...

Referencias

[1] W. Navidi, Statistics for Engineers and Scientists w/ CD-ROM. McGraw-Hill Sci./Eng./Math, 2004.