

Actividad 2

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

Resumen

...

Key words

...

Introducción

...

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 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) ““
```

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 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 probabilística 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.

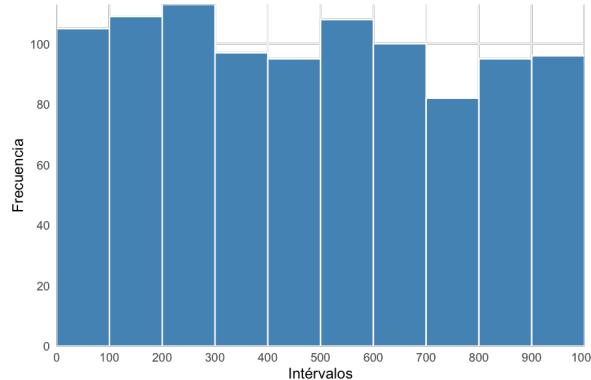


Figura 1. Distribución de los datos.

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

Muestra 1: 581, 270, 961, 584, 14, 238, 320, 301, 436, 812

Muestra 2: 36, 223, 615, 517, 894, 982, 734, 214, 681, 199

Muestra 3: 134, 855, 301, 14, 5, 503, 675, 160, 245,

195

Muestra 4: 950, 556, 857, 458, 138, 853, 876, 255, 167, 365

Muestra 5: 616, 148, 639, 698, 96, 884, 161, 580, 68, 578

Estas muestras tienen los promedios muestrales 451.7, 509.5, 308.7, 547.5 y 446.8 respectivamente, mostrando una alta variabilidad de la media de las muestras obtenidas.

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 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

Conclusiones

...

Referencias

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