

Variables aleatorias: Ejercicios

Bioinformática - Pasto 2017

Guillermo Torres

Ejercicios

Para estos ejercicios vamos a utilizar el siguiente dataset:

```
library(downloader)
url <- "https://raw.githubusercontent.com/guigotoe/Workshop-Pasto-2017/master/extdata/"
filename <- "femaleControlsPopulation.csv"
download(paste0(url,filename), destfile=filename)
x <- read.csv(filename)
## o
dir <- "extdata/"
filename <- "femaleControlsPopulation.csv"
x <- read.csv(paste0(dir,filename))
```

x representa los pesos de la población entera.

1. Cual es el promedio de los pesos?
2. Usar `set.seed(1)`. Tomar una muestra aleatoria de tamaño 5 (usar `sample()`). Cual es el valor absoluto (usar `abs()`) de la diferencia entre el promedio de la muestra y el promedio de todos los valores?
3. Usar `set.seed(5)`. tomar una muestra aleatoria de tamaño 5. Cual es el valor absoluto (usar `abs()`) de la diferencia entre el promedio de la muestra y el promedio de todos los valores?
4. Porqué las respuestas 2 y 3 son diferentes?
 - A) por que hicimos un error en el código
 - B) por que el promedio de x es aleatorio
 - C) por que el promedio de la muestra es un variable aleatoria
 - D) Todas las anteriores.
5. Usar `set.seed(1)`. Usar un for-loop y tomar una muestra aleatoria de 5 ratones 1.000 veces. Guardar los promedios. Que porcentaje de estos 1.000 promedios son más de 1kg mayor al promedio de x?
6. Ahora vamos a ir incrementando el número de veces que muestreamos, de 1.000 a 10.000. Usar `set.seed(1)`, ejecutar un for-loop y tomar una muestra aleatoria de 5 ratones 10.000 veces y guardar los promedios. Qué porcentajes de estos 10.000 promedios son mas de 1kg mayor al promedio de x?
7. Notece que las respuestas 5. y 6. cambian ligeramente. Esto es esperado dado que la forma de pensar la distribución de una variable aleatoria es como la distribución de una lista de valores obtenidos si repetimos el experimento un número infinito de veces. No obstante en el computador no podemos realizar una iteración infinita, por lo que consideramos 1.000 o 10.000 repetición lo suficientemente grandes para representar el infinito. Ahora, vamos a cambiar el tamaño de la muestra aleatoria, y así su distribución.

Usaremos `set.seed(1)` y ejecutaremos un for-loop en el cual tomaremos una muestra aleatoria de 50 ratones 1.000 veces y guardaremos los promedios de la muestra aleatoria. Qué porcentaje de los 1.000 promedios es mas de 1kg mayor al promedio de x?

8. Usar un histograma (`hist()`) para mirar la distribución de promedios que conseguimos con una muestra de tamaño 5 y la de tamaño 50. Cómo podría definir la diferencia?

- A) Son las mismas?
 - B) Las dos son similares a una distribución normal, pero con un tamaño de 50 la dispersión es más pequeña.
 - C) Las dos son similares a una distribución normal, pero con un tamaño de 50 la dispersión es más grande.
 - D) La distribución de tamaño 50 no parece una distribución normal.
9. Para el último set de promedios, aquellos obtenidos con el tamaño de muestra de 50, cuál es el porcentaje entre 23 y 25?
10. Ahora la misma pregunta pero a una distribución normal con media de 23.9 y desviación estandar de 0.43.
11. Por que las respuestas 9 y 10 son muy similares?