

Datos cuantitativos agrupados

Ramon Ceballos

31/1/2021

Introducción

Aunque no seamos completamente conscientes de ello, tendemos a agrupar datos cuantitativos constantemente.

Sin ir más lejos, calificamos de excelente a todas las notas que están sobre el 9. También decimos que una persona tiene 20 años cuando se encuentra en el intervalo $[20,21)$. Es decir, cuando ha cumplido los 20 pero aún no tiene los 21.

Primer motivo para agrupar datos

En estadística, existen innumerables motivos por los cuales nos interesa agrupar los datos cuando estos son cuantitativos. Uno de estos motivos puede ser perfectamente que los datos sean muy heterogéneos. En este caso, nos encontraríamos con que las frecuencias de los valores individuales serían todas muy similares, lo que daría lugar a un diagrama de barras muy difícil de interpretar, tal y como mostramos en el siguiente ejemplo.

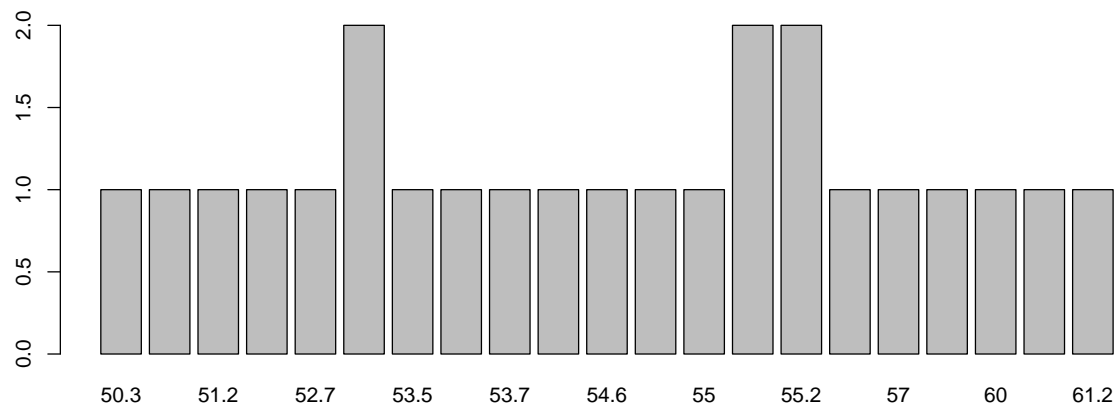
Ejemplo 1

Consideremos la siguiente muestra de 24 pesos de estudiantes:

```
pesos = c(55.2, 54.0, 55.2, 53.7, 60.2, 53.2, 54.6, 55.1, 51.2, 53.2, 54.8, 52.3, 56.9, 57.0, 55.0, 53.5, 50.9, 55.1, 53.0, 54.5, 56.0, 52.0, 55.5, 53.0)
```

El diagrama de barras de sus frecuencias absolutas, tomando como posibles niveles todos los pesos entre su mínimo y máximo se muestra en la siguiente diapositiva.

```
barplot(table(pesos))
```

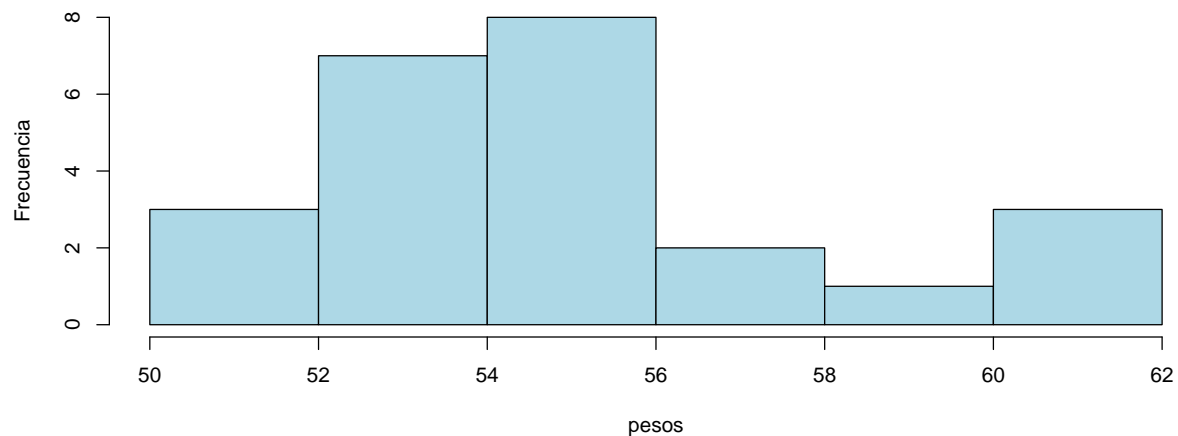


Como vemos, todas estas frecuencias se encuentran entre 0 y 2, cosa que no nos da mucha información.

En cambio, si dividiésemos todos estos posibles valores que puede tomar la variable cuantitativa en intervalos y tomásemos como sus frecuencias las de todos los valores que caen en dicho intervalo, la cosa cambia.

En este caso, sería mucho más fácil interpretar los resultados, ya que estos darán mucha más información. Más adelante veremos como crear estos intervalos.

```
hist(pesos, breaks = 6, right = FALSE, main = "", ylab = "Frecuencia", xlab = "pesos", col = "lightblue")
```



2. Segundo motivo para agrupar datos

Otro de los motivos por el que necesitamos muchas veces agrupar los datos cuantitativos es porque, como ya dijimos en temas anteriores, la precisión infinita no existe. Por tanto, esta imposibilidad de medir de manera exacta muchas de las magnitudes continuas (tiempo, peso, altura...) nos obliga a trabajar con

aproximaciones o redondeos de valores reales y que cada uno de estos represente todo un intervalo de posibles valores.

3. Resumen general de los motivos para agrupar datos

Por lo general, existen 3 situaciones en las cuales conviene sin lugar a dudas agrupar datos cuantitativos en intervalos, también llamados **clases**.

- Cuando los datos son continuos, su redondeo ya define un agrupamiento debido a la inexistencia de precisión infinita.
- Cuando los datos son discretos, pero con un número considerablemente grande de posibles valores.
- Cuando tenemos muchísimos datos y estamos interesados en estudiar las frecuencias de sus valores.