



Medidas de tendencia central

LECCIÓN 3

Es común que las distribuciones de probabilidad tengan algunos datos con más repeticiones que los demás. Estos comportamientos se describen mediante una función que nos ayuda a conocer cuales son las características de esa variable aleatoria. Para establecer dicha función es necesario conocer la posición y la dispersión de la variable. Conociendo la distribución de frecuencias podemos tener una idea del comportamiento para modelarlo con la medida de posición o de tendencia central y la medida de dispersión. De esta forma es posible comparar y operar variables aleatorias para realizar cálculos, propagación de las distribuciones, y validación de experimentos.

En el caso de las variables cuantitativas se tienen las siguientes medidas de tendencia central:

Moda:

Es el valor que tiene la máxima frecuencia (preguntar la moda de la edad en el titanic.). En ocasiones la moda no difiere de otras medidas, es decir, los valores de varias frecuencias son iguales o similares. En este caso se puede decir que la variable puede ser bimodal (dos modas) o multimodal (varias modas). Cuando todos los valores tienen la misma frecuencia no existe la moda (tal es el caso de una distribución uniforme). Cuando la escala de una variable se divide en clase, se puede definir la clase modal, como aquella que reúne la máxima frecuencia.

Cuantiles:

Un cuantil es un valor de una variable a la cual corresponde una determinada frecuencia relativa acumulada. El cuantil α de una distribución de frecuencias es un valor de la variable al cual corresponde la frecuencia relativa acumulada $f_{ra} = \alpha$. Por ejemplo, el cuantil 0,15 es un valor de la variable al cual corresponde la frecuencia relativa acumulada $f_{ra} = 0,15$. Es común referirse a percentiles que no son otra cosa que los cuantiles identificados por el valor de α expresado en porcentaje. Por ejemplo, en lugar de cuantil 0,15 podemos decir percentil 15. Los cuantiles 0,25, 0,50 y 0,75 se denominan respectivamente primer cuartil, segundo cuartil o mediana y tercer cuartil. La mediana es el valor que corresponde a la mitad de la distribución de frecuencias. Por eso decimos que la mediana es una medida de posición central.

Los gráficos de cajas o de bigotes (box plots) resumen la distribución de frecuencias a partir de unos cuantiles, como se muestra en la figura 1. En estos gráficos, los bordes de la caja indican el primer y tercer cuartil, la línea horizontal que corta la caja indica la mediana y los extremos muestran el valor mínimo y máximo de la variable. Los valores de las medidas de posición se aprecian en el eje vertical de la caja.

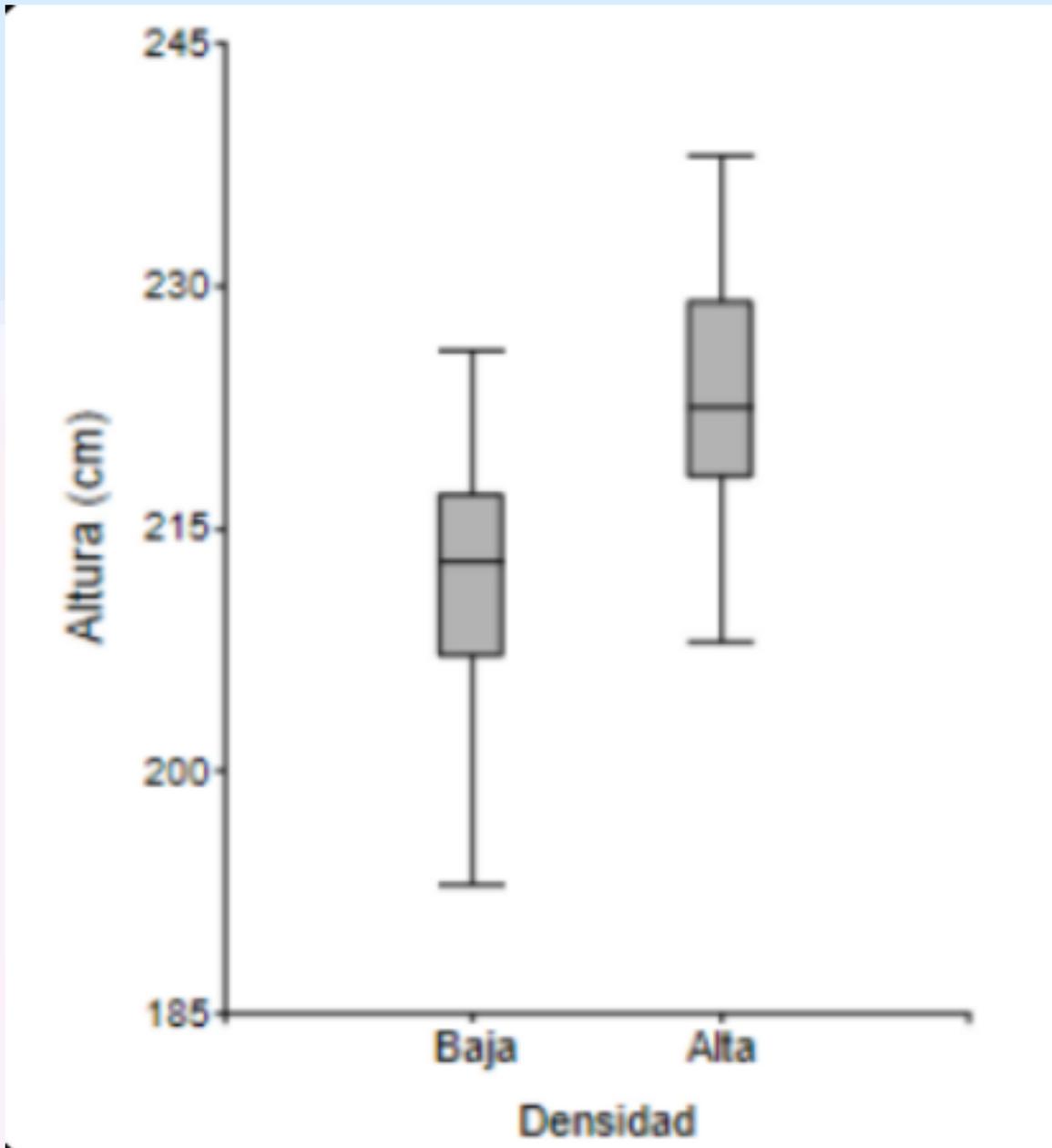


Figura 1. Ejemplo de diagrama de cajas o de bigotes.

Media aritmética

La media aritmética (promedio) es una de las medidas de tendencia central más conocidas de la estadística. Esta se calcula como el cociente entre la suma de los valores de la variable aleatoria y el número de mediciones. Generalmente se denota como una variable con una barra superior en la escritura como se muestra en la ecuación 1.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^N x_i$$

Ecuación 1: cálculo de la media aritmética.

La media aritmética brinda información sobre la posición central de una distribución de frecuencias. Sin embargo, la media no nos informa acerca de cómo se distribuyen los datos. Con lo que no es posible aun comparar muchas distribuciones de frecuencia solo con la media aritmética. Por eso se necesitan conocer algunas medidas de dispersión (se verán en la siguiente lección). Si se conoce una medida de tendencia central y una medida de dispersión, es posible comparar variables aleatorias y modelarlas dados estos dos comportamientos.