

Agrupación de datos por intervalos

Cuando la variable estadística es cuantitativa continua o bien cuantitativa discreta pero con muchos valores cercanos, es una buena simplificación dejar de considerar cada dato individualmente y agrupar los datos que estén entre dos valores.

Ejemplo 1a. La estatura de las personas es una variable estadística cuantitativa continua, pero realmente hay poca diferencia entre medir 183 y 184 centímetros, así que, si tenemos que manejar muchas estaturas, podemos agruparlas de cinco en cinco centímetros, con lo que las estaturas 183 y 184 quedarían englobadas en el intervalo de 180 a 185 centímetros.

Ejemplo 2a. El número de puntos anotados por un jugador o jugadora de baloncesto en una temporada es una variable estadística cuantitativa discreta, pero realmente hay poca diferencia entre anotar 2593 y 2596 puntos, así que si tenemos que manejar muchos jugadores o jugadoras, podemos agrupar sus anotaciones de cien en cien puntos, con lo que los puntajes 2593 y 2596 quedarían englobados en el intervalo de 2500 a 2599 puntos.

Definición del intervalo

Para definir cada intervalo se utiliza una notación que estudiaremos con detalle en el nivel 4. Consiste en escribir los valores extremos de menor a mayor, con un carácter separador, entre paréntesis o corchetes.

Ejemplo 1b. El intervalo de números entre 180 y 185, incluyendo el 180 y excluyendo el 185 se escribe $[180,185)$. Explicación:

- El corchete indica que el 180 sí está incluido.
- El paréntesis indica que el 185 no está incluido.
- La coma es el separador entre los dos números.

Podemos usar intervalos así formados para agrupar los datos del ejemplo (1a).

Ejemplo 2b. El intervalo de números entre 2500 y 2599, incluyendo ambos, se escribe $[2500,2599]$. Podemos usar intervalos así formados para agrupar los datos del ejemplo (2a).

Clase y marca de clase

- * Cada intervalo que usemos para agrupar datos se llama una **clase**.
- * La **marca de clase** es un número que representa a todos los valores de una clase. Siempre se calcula como la media de los dos extremos del intervalo que define la clase.

Ejemplo 1c. El intervalo $[180,185)$ representa una clase. Su marca de clase es:
 $(180+185) : 2 = 182,5$.

Ejemplo 2c. El intervalo $[2500,2599]$ representa una clase. Su marca de clase es:
 $(2500+2599) : 2 = 2549,5$.

Separación de los intervalos

Cuando se utilizan intervalos para agrupar los datos, los intervalos ni se solapan ni dejan huecos entre ellos, sino que distribuyen los datos de una manera continua.

Ejemplo 1d. Para agrupar estaturas podemos utilizar intervalos de esta forma:
... $[160,165)$, $[165,170)$, $[170,175)$, $[175,180)$, $[180,185)$, $[186,190)$, $[190,195)$...

Ejemplo 2d. Para agrupar puntos podemos utilizar intervalos de esta forma:
... $[2300,2399]$, $[2400,2499]$, $[2500,2599]$, $[2600,2699]$, $[2700,2799]$,...