1. MEDIDAS DE TENDENCIA CENTRAL

Lo importante en una tendencia central es calcular un valor central que actúe como resumen numérico para representar al conjunto de datos. Estos valores son las **medidas o índices de tendencia central.**

Las medidas de tendencia central más utilizadas son 3..

1) La media aritmética

- Es la más utilizada
- Es el valor central alrededor del cual están la mayoría de las observaciones
- Sólo puede calcularse para variables cuantitativas (números)

Se calcula sumando todos los valores de la variable divididos por el número total de observaciones.

$$\bar{X} = \Sigma X_i / n$$

$\bar{X}$ = Media aritmética
$\sum$ = Símbolo sumatorio (indica que se están sumando todas las X que hay)
X_i= Valor que toma la variable u observación del sujeto i
n = Número total de observaciones

Si el número de observaciones es muy grande, la media aritmética se puede calcular a partir de las **frecuencias absolutas** (n_i) o de las **frecuencias relativas** (p_i) (recordemos que las frecuencias relativas también las llamábamos proporciones)

Para frecuencias absolutas:

$$\bar{X} = \Sigma n_i X_i / \Sigma n_i$$

Para frecuencias relativas:

$$\bar{X} = \Sigma p_i X_i$$

En realidad da igual cual de los dos procedimientos escojamos, ya que ante una misma distribución de frecuencias el resultado va a ser el mismo aunque se haga por la fórmula de frecuencias absolutas o mediante la fórmula de frecuencias relativas. Recordemos que $\mathbf{p_i = n_i / n}$ (esto lo vimos en el primer tema)

Propiedades de la media aritmética

1. En una distribución, la suma de las desviaciones de cada valor con respecto a su media es igual a cero.

$$\Sigma (X_i - \bar{X}) = 0$$

2. Si a los valores de la variable X les aplicamos la siguiente transformación lineal:

$\mathbf{Y_i = bX_i + a}$ la media de los nuevos valores Y será $\mathbf{\overline{Y} = b\overline{X} + a}$

Con tanta letra esto suena un poco dificil pero básicamente explica que si cada uno de los valores antiguos (X) lo multiplicamos por un número cualquiera y le sumamos otro número, obtenemos nuevos valores (Y) pero con la misma proporción anterior. Es decir, si tenemos dos valores: X=4 y X=8 y los multiplicamos por 10, tendremos Y=40 y Y=80, pero la proporción se sigue manteniendo, el segundo valor sigue siendo el doble que el primero. El libro pone el ejemplo (pág 63) más extenso con el ejercicio pero la idea básica es esta.

Limitaciones de la media aritmética

a) Cuando los datos están agrupados en intervalos, la media no se puede calcular si el intervalo máximo no tiene límite superior y/o el intervalo mínimo no lo tiene inferior.

b) Puede que a la hora de tomar los datos nos encontremos con valores muy extremos (asimétricos), en ese caso debemos valorar si son errores, entonces deberemos eliminarlos para realizar la media. Pero si por el contrario estos valores extremos son importantes para nuestro estudio, deberemos escoger otros índices de medición, como por ejemplo la mediana que veremos a continuación.

2) La mediana

- La utilizaremos cuando la distribución sea muy asimétrica.
- Los valores extremos no le afectan (a diferencia de la media aritmética) ya que para su cálculo sólo se toman los valores que ocupan posiciones centrales.
- Se puede calcular para todo tipo de variables, excepto cualitativas.

La Mediana de una variable X, representada por **Md**, se define como el valor de la variable que divide la distribución de frecuencias en dos partes iguales, conteniendo cada una el 50% de las observaciones.

Cálculo de mediana con pocos casos (muy fácil, ☺ Pág. 65 y 66)

1º - Se ordenan las puntuaciones n de mayor a menor
2º - Se Observa si el número de observaciones n es par o impar

- Si es impar, la mediana es el valor de la posición central
- Si es par, la mediana es la media aritmética de los dos valores centrales.

CAPITULO 2: MEDIDAS DE TENDENCIA CENTRAL Y POSICIÓN

Cálculo de la mediana con muchos casos (esto ya no es tan fácil…☹)

Al ser muchos casos, los datos vienen presentados en intervalos. El intervalo en el que se encuentra la mediana se llama **Intervalo crítico** y se corresponde con aquel en el que la frecuencia absoluta acumulada n_a es igual o superior a n/2.

$$Md = L_i + \frac{(n/2 - n_d)}{n_c} I$$

L_i = Límite exacto inferior del intervalo crítico
n = Número de observaciones
n_d = Frecuencia absoluta acumulada por debajo del intervalo crítico
n_c = Frecuencia absoluta del intervalo crítico
I = Amplitud del intervalo crítico

En la página 68 hay un ejemplo numérico para hallar la mediana resuelto de dos maneras diferentes. La primera manera consiste simplemente en aplicar la fórmula, hacerlo paso por paso y comprendiendo qué es cada cosa es bastante sencillo. La segunda manera consiste en una regla de tres. Los dos procedimientos dan el mismo resultado.

Casos especiales (pág 69-71) Si hemos entendido el caso anterior, esto es más de lo mismo pero con un par de puntualizaciones)

A) Cuando se trata de una distribución de frecuencias pero los datos no están agrupados en intervalos. Como los datos son unitarios, la amplitud (I) será = 1
Es prácticamente lo mismo que el ejemplo anterior pero con I = 1

B) No se puede calcular la mediana cuando los datos están agrupados en intervalos y existe un intervalo abierto en el que se encuentra la mediana.

3) La moda (la más fácil de todas las medidas de tendencia central)

- Se puede calcular para variables cualitativas y cuantitativas.
- Se representa por **Mo** y se define como el valor o categoría de la variable con mayor frecuencia absoluta.
- Si los datos están agrupados en intervalos, la moda se puede calcular excepto si el intervalo modal coincide con el intervalo abierto.

- Si la variable es **cualitativa**, la moda es la categoría con la máxima frecuencia.
- Si la variable es **cuantitativa**, la moda es el valor con la máxima frecuencia absoluta.
- Si la variables es **cuantitativa** con datos en **intervalos**, se localiza el intervalo modal (intervalo con la frecuencia máxima) y la moda es el punto medio de dicho intervalo.

Podemos tener distinto número de modas en la distribución (1= unimodal) (2=bimodal) (3=trimodal); etc.

Distribución amodal = cuando no hay moda ya que todos los valores tienen la misma frecuencia absoluta.

La elección de una medida de tendencia central

- En primer lugar probaremos con la **media aritmética** (menos en los casos en que la distribución sea muy asimétrica, o el nivel de medida sea nominal u ordinal, o existan datos agrupados en intervalos con intervalos abiertos.

- Si vemos que por alguna de estas razones la media aritmética no funciona probaremos con la **mediana.** Sin embargo la mediana no podremos utilizarla cuando el nivel de la variable sea nominal o la mediana se encuentre en el intervalo abierto.

- Entonces, descartando las dos medidas anteriores, probaremos con la **moda.** Aunque no podremos calcularla si la distribución es amodal o el intervalo abierto coincide con el intervalo modal.

Hoy en día se recomienda calcular las 3 (siempre que se pueda).

- Variables cualitativas: Solo podremos calcular la moda
- Variables ordinales: moda y mediana
- Variables cuantitativas: Podemos calcular las 3.

2. MEDIDAS DE POSICIÓN

La medida de tendencia central buscaba un indicador para representar a un conjunto de datos. En cambio, las medidas de posición buscan un indicador para representar a un sujeto o a un dato en particular.

Medidas de tendencia central: Nota media de un examen en una clase de 30 niños
Medidas de posición: ¿Qué nota debe sacar un alumno para superar al 50% de compañeros?

Las medidas o índices de posición (también llamados cuantiles), informan acerca de la posición relativa de un sujeto con respecto a su grupo de referencia.

Dependiendo de cuántos valores de la variable utilicemos para dividir la distribución, hablaremos de percentiles, cuartiles o deciles.

Percentiles

También llamados “centiles”, son los 99 valores de la variable que dividen en 100 partes iguales la distribución de frecuencias.

Percentil k (P_k): Es un valor de la variable de interés que deja por debajo de sí un porcentaje k de sujetos, donde k = 1,2…99

Ejemplo P_{40} = 25 (25 es la puntuación en un examen)
Quiere decir que los sujetos con X= 25 están por encima del 40% de los sujetos.

En los percentiles, la mediana coincide con el percentil 50 (es justo la mitad). Por esta razón el percentil lo calcularemos a partir de las fórmulas para la mediana.

Cálculo de los percentiles

Utilizaremos la misma fórmula que la mediana pero sustituyendo **nk/100** en lugar de **n/2**

$$P_k = L_i + \frac{(nk/100 - n_d)}{n_c} I$$

L_i = Límite exacto inferior del intervalo crítico
n = Número de observaciones
n_d = Frecuencia absoluta acumulada por debajo del intervalo crítico
n_c = Frecuencia absoluta del intervalo crítico
I = Amplitud del intervalo

Ejemplo numérico en la página 78. Tenemos que tener en cuenta que lo primero que hay que hacer es hallar el intervalo crítico mediante la fórmula (nk/100) a diferencia de cómo lo hallábamos para la mediana que era mediante la fórmula (n/2)

Aparte de hallar el valor de los percentiles, también puede ocurrir que ya tengamos un valor y queramos saber qué posición ocupa ese valor en la distribución (es como lo de antes pero al revés, es decir, antes tenía una posición y quería hallar un valor. Ahora tengo un valor y quiero hallar una posición. Por lo tanto la fórmula cambia un poco.

$$k = \frac{(P_k - L_i)\, n_c / I + n_d}{n} \cdot 100$$

Si el resultado de k es decimal se redondea.

Cuartiles y deciles

Cuartiles: 3 valores de la distribución que dividen en 4 partes de igual frecuencia a la distribución.

Primer Cuartil (Q_1) = Deja por debajo de sí al 25% y por encima al 75% ($Q_1 = P_{25}$)
Segundo Cuartil (Q_2) = Deja por debajo de sí al 50% y por encima al 50% ($Q_2 = P_{50} = Md$)
Tercer Cuartil (Q_3) = Deja por debajo de sí al 75% y por encima al 25% ($Q_3 = P_{75}$)

Por lo tanto como cada cuartil se corresponde con un percentil, utilizaremos las fórmulas de los percentiles para hallar cada cuartil.

Deciles: Son 9 valores que dividen en 10 partes iguales a la distribución. Se representan por D_1, D_2 ... D_9 …y al igual que los cuartiles, también se corresponden con cada percentil: $D_1 = P_{10}$; $D_2 = P_{20}$ etc… Por lo tanto también utilizaremos las fórmulas de los percentiles para calcular los deciles.