

### 3. MEDIDAS DE TENDENCIA CENTRAL

Con estas medidas se persigue reducir en pocas cifras significativas el conjunto de observaciones de una variable y describir con ellas ciertas características de los conjuntos, logrando una comparación más precisa de los datos que la que se puede conseguir con tablas y gráficas.

#### 3.1. DISPERSIÓN Y FORMA

Los promedios son una medida de posición que dan una descripción compacta de como están centrados los datos y una visualización más clara del nivel que alcanza la variable, pueden servir de base para medir o evaluar valores extremos o raros y brinda mayor facilidad para efectuar comparaciones.

Es importante poner en relieve que la notación de promedio lleva implícita la idea de variación y que este número promedio debe cumplir con la condición de ser representativo de conjunto de datos.

El promedio como punto típico de los datos es el valor al rededor del cual se agrupan los demás valores de la variable.

#### 3.2. MEDIA ARITMÉTICA

Es una medida matemática, un número individual que representa razonablemente el comportamiento de todos los datos.

Para datos no agrupados  $X = \sum x_i / n$

Para datos agrupados  $X = \sum f_i X_i / \sum f_i$

donde  $X_i$  es la marca de clase para cada intervalo y  $f_i$  es la frecuencia de clase

Características de la Media:

1. En su cálculo están todos los valores del conjunto de datos por lo que cada uno afecta la media.
2. La suma algebraica de las desviaciones de los valores individuales respecto a la media es cero.
3. La suma del cuadrado de las desviaciones de una serie de datos a cualquier número  $A$  es mínimo si  $A = X$
4. Aunque es confiable porque refleja todos los valores del conjunto de datos puede ser afectada por los valores extremos, y de esa forma llegar a ser una medida menos representativa, por lo que si la distribución es asimétrica, la media aritmética no constituye un valor típico.

## **LA MODA**

Es el valor de un conjunto de datos que ocurre más frecuentemente, se considera como el valor más típico de una serie de datos.

Para datos agrupados se define como Clase Modal el intervalo que tiene más frecuencia.

La moda puede no existir o no ser única, las distribuciones que presentan dos o más máximos relativos se designan de modo general como bimodales o multimodales.

Características de la Moda.

1. Representa más elementos que cualquier otro valor
2. No está afectada por los valores extremos pero para datos continuos es dudoso su cálculo.
3. La moda para una distribución de frecuencias de datos agrupados no puede ser calculada exactamente, el valor de la moda puede ser afectado por el método de agrupación de los intervalos de clase.
4. La moda no permite conocer la mayor parte de los datos
5. Algunas veces el azar interviene de manera importante y hace que un valor no representativo se repita frecuentemente.
6. Puede usarse para datos cuantitativos como cualitativos
7. La moda como estadístico, varía mucho de una muestra a otra
8. Cuando se tienen dos o más modas es difícil su interpretación
9. Tiene la ventaja de que los datos desproporcionados con respecto al resto no la distorsionan, pero no se presta para un tratamiento matemático.

## **LA MEDIANA**

Es el valor de la observación que ocupa la posición central de un conjunto de datos ordenados según su magnitud. Es el valor medio o la media aritmética de los valores medios. La mediana es un valor de la variable que deja por debajo de él un número de casos igual al que deja por arriba.

Geoméricamente la mediana es el valor de la variable que corresponde a la vertical que divide al histograma en dos áreas iguales.

Cuando determinados valores de un conjunto de observaciones son muy grandes o pequeños con respecto a los demás, entonces la media aritmética se

puede distorsionar y perder su carácter representativo, en esos casos es conveniente utilizar la mediana como medida de tendencia central.

#### Características de la mediana

1. Es un promedio de posición no afectado por los valores extremos.
2. No está definida algebraicamente
3. Cuando la localización del elemento central puede ser determinada y los límites de clase mediana son conocidos, la mediana para la distribución de frecuencias puede ser calculada por interpolación, no importando que ésta contenga intervalos abiertos, cerrados, iguales o diferentes.
4. La suma de los valores absolutos, sin considerar el signo, de las desviaciones individuales respecto a la mediana es mínimo.
- 5 La mediana en caso de una distribución asimétrica, no resulta desplazado del punto de tendencia central.
6. Si el universo tiene curtosis excesiva la mediana como estadístico, varía menos que cualquier otra medida.
7. Si la mediana se calcula por interpolación y hay lagunas en los valores de la clase mediana o los datos son irregulares, esta medida no es buena ya que su ubicación puede resultar falsa.
8. Si se desea ubicar las condiciones de un elemento en una clase, la mediana resulta se indicada, ya que por comparación pone en evidencia si un elemento está en la mitad superior a ella o en la inferior.

#### **MEDIA ARITMETICA PONDERADA**

En ésta, para cada uno de los valores de  $x_i$  se asigna un factor  $w_i$  de peso, que depende de la importancia que el investigador desee darle.

$$X_p = \frac{\sum (x_i w_i)}{\sum w_i}$$

#### **MEDIA GEOMÉTRICA**

Útil cuando la variable cambia a lo largo del tiempo, esto es, en el cálculo del promedio de tasas, razones, proporciones geométricas y relaciones de variables. Se utiliza en Matemáticas Financieras y Finanzas para promediar números índices, tasas de cambio, etc.

La media Geométrica de una serie de números es la raíz  $n$ -ésima del producto de esos números

$$M = \sqrt[n]{x_1 * x_2 * x_3 * \dots * x_n}$$

Se ve afectada por todos los números y valores extremos pero en menor grado que la Media Aritmética, su valor siempre es menor que el de ésta.

### MEDIA ARMÓNICA

Se utiliza para el promedio de rendimientos y velocidades. La Media Armónica de una serie de números es el recíproco de la media aritmética del recíproco de esos números.

$$1 / MH = [ \sum 1 / x_i ] / n$$

### MEDIA CUADRÁTICA

Es la raíz cuadrada de la media aritmética de los cuadrados de los números, se usa eficientemente para promediar los errores o desviaciones porque es más susceptible a los mismos.

$$MC = \sqrt{ \sum x_i^2 / n }$$

### LOS CUANTILES

Son valores que dividen a la distribución en n partes iguales

Cuartiles, cuatro partes iguales: Q1, Q2, Q3

Deciles, diez partes iguales : D1, D2.....D9

Percentiles o centiles, cien partes iguales: P1, P2.....P99

Los cuantiles permiten hacer un análisis minucioso de la distribución, se utilizan generalmente cuando se quiere ubicar un dato dentro del conjunto. Por ejemplo. Pertenece el dato x al 50% superior ?, al 10% inferior? , al 50 % central?, etc.

### MEDIDAS DE DISPERSIÓN

Un rasgo principal de los datos es su dispersión o amplitud, que se refiere a su variabilidad, a la evaluación de cuán separados o extendidos están estos datos o bien cuanto difieren unos de otros.

Variación: es el grado en que los datos numéricos tienden a extenderse al rededor de un valor, generalmente el valor medio

¿Por qué es importante la variación?

1. Al menudo una medida de posición de un conjunto de datos se vincula con la indicación de cuán típico o representativo es para la población y para ello es necesario contar con la información que proporcionan las medidas de variación. Solo el conocimiento de un estadístico de tendencia central no aclara o define toda la distribución, además que no existe un valor de tendencia central ideal, por lo que es significativo tener una idea de la dispersión de los valores y determinar si es mucha o poca al rededor de la media, pues si la variación es muy grande entonces esta medida de tendencia central no es buena selección como valor típico.

2. La medida de tendencia central no indica la relación de un dato con los otros, es necesario para ello las medidas de variabilidad o dispersión.

3. Al tratar problemas con datos dispersos se requiere conocer que problemas puede esto traer, hasta que punto la dispersión tiene un riesgo aceptable o inaceptable en la toma de decisiones.

4. Al comparar dos distribuciones por lo general centramos la atención en la posición y en la dispersión.

### **3.3. RANGO**

Mide la dispersión de la totalidad de los datos. Es la más obvia de las medidas ya que es la distancia entre los valores máximo y mínimo.

El rango o recorrido da alguna idea del grado de variación que ocurre en la población, pero con frecuencia los resultados pueden ser engañosos, pues este depende de los valores extremos e ignora la variación de las demás observaciones. Está afectado por ocurrencias raras o extraordinarias.

### **INTERVALO INTERDECIL**

Mide la dispersión del 80% de los datos centrales y se obtiene de la diferencia entre el decil 9 y el decil 1, evitando así los puntos extremos.

### **INTERVALO INTERCUARTIL :**

Cuando aumenta la dispersión de una distribución de frecuencias, se amplía la distancia entre los cuartiles, por lo que esta distancia puede usarse como base de una medida de variabilidad

El intervalo intercuartil, es el recorrido entre el cuartil 3 y el cuartil 1. Es el intervalo en el cual está comprendido el 50% de los datos centrales.

### **DESVIACIÓN CUARTÍLICA**

Mide el intervalo promedio de un cuarto de los datos  $[(Q3-Q1)/2]$

Si la distribución es perfectamente simétrica, los dos cuartiles Q1 y Q3 equidistan de la mediana y la mitad de la distancia entre los cuartiles representa la distancia promedio entre ellos y la mediana.

Si en una distribución simétrica se mide una distancia igual a la desviación cuartílica a ambos lados de un punto ubicado en el centro de los cuartiles, el 50% de los valores estarán incluidos dentro de esos límites y el valor del punto medio coincide con la mediana.

La ventaja de la desviación cuartílica es que evita los valores extremos utilizando únicamente la mitad intermedia de los datos.

### 3.4. DESVIACIÓN MEDIA

La desviación Media o Desviación absoluta promedio, es la media aritmética de las desviaciones absolutas de cada una de las observaciones con respecto a su valor central, la media aritmética, o la mediana

Cuanto mayor es su valor, mayor es la dispersión de los datos

$$DM = \frac{\sum |x_i - X|}{n}$$

$$DM = \frac{\sum f_i |x_i - X|}{\sum f_i}$$

Las características de esta media de dispersión son:

1. Su valor depende del valor de cada observación.
2. Se puede calcular al rededor de la media o de la mediana.
3. La desviación promedio respecto a la mediana es un mínimo
4. Mide la desviación de una observación sin notar si está por encima o por debajo del promedio.

### 3.4. VARIANZA

Otro tratamiento para evadir la suma cero de las desviaciones de las observaciones respecto a su Media Aritmética, consiste en recurrir al proceso de elevar al cuadrado estas desviaciones y sumar los cuadrados, dividiendo la suma por el número de casos, a esta cantidad se le denomina varianza, y es la más importante de las medidas de variación porque tiene la ventaja de no prescindir de los signos de las desviaciones, pero al igual que la desviación media los valores extremos pueden distorsionarla

$$s^2 = \frac{\sum (x_i - X)^2}{n}$$

$$s^2 = \frac{\sum f_i (x_i - X)^2}{\sum f_i}$$

$$S^2 = \frac{\sum (x_i - X)^2}{(n)}$$

$$S^2 = \sum f_i (x_i - X)^2 / \sum f_i$$

$$S^{2*} = \sum (x_i - X)^2 / (n-1)$$

$$S^{2*} = \sum f_i (x_i - X)^2 / (\sum f_i - 1)$$

En inferencia, con una muestra tomada de una población grande se pretende descubrir cuanto varían los datos al rededor de la media poblacional, si embargo cuando no se conoce la media de la población se estima a partir de la media aritmética de la muestra y esto hace que parezca menos variable de o que es en realidad, al dividir por n-1 se está compensando por la variabilidad más pequeña que se observa en la muestra, por lo que  $S^{2*}$ , la suma de cuadrados dividida por n-1 es considerado un estimador más eficiente para la varianza poblacional.

## **DESVIACION ESTANDAR**

Cuando se utiliza la varianza como medida de dispersión, para salvar el problema de trabajar con distintas dimensiones en la media y en la medida de variabilidad es necesario definir la Desviación estándar como la raíz cuadrada de la varianza.

La Desviación Estándar es útil para describir cuanto se apartan de la media de la distribución los elementos individuales. Una medida de ello se denomina puntuación estándar número de desviaciones a las que determinada observación se encuentra con respecto a la media.

$$\text{Puntuación estándar de } x_i = (x_i - X) / s$$

Al comparar distribuciones también hacemos uso de la calificación estándar.

Característica de la Desviación Estándar:

1. Es afectada por el valor de cada observación
2. Como consecuencia de considerar desviaciones cuadráticas pone mayor énfasis en las desviaciones extremas que en las demás desviaciones.
3. Si en el eje X de la distribución de frecuencias normal, se mide a ambos lados de la media una distancia igual a :

Una desviación estándar se forma un intervalo en el cual se encuentra el 68.27% de los valores centrales de la variable

Dos desviaciones estándar, se forma un intervalo donde se encuentra el 95.43% de los valores centrales

Tres desviaciones estándar, se forma un intervalo que contiene el 99.73% de los valores centrales

4. Al construir la tabla de frecuencias de una variable discreta y calcular a partir de ella la desviación estándar no hay pérdida de información por lo que la desviación para los datos observados es igual que para los datos tabulados.

En la construcción de una tabla de una variable continua hay pérdida de información por el agrupamiento de los valores en intervalos y se traduce en la discrepancia entre el valor de la desviación observada y tabulada.

### MEDIADAS DE DISPERSIÓN RELATIVAS

Cuando se necesita comparar dos o más series de datos a veces no es posible hacerlo con las medidas absolutas, ya sea porque las unidades son diferentes o porque tienen diferente media, en éstos casos deben utilizarse cantidades relativas definida generalmente como:

Dispersión relativa = Dispersión absoluta / media

### 3.5. COEFICIENTE DE VARIACIÓN

Es la medida de dispersión relativa más usada y se define como el cociente de la desviación estándar entre el promedio aritmético, expresado en porcentaje y es adimensional

$$V = S / X$$

### MEDIDAS DE ASIMETRÍA Y CURTOSIS

#### MEDIDAS DE SESGO O ASIMETRIA

En las distribuciones que no toman la forma de una curva acampanada Normal, interesa muchas veces obtener dos medias adicionales, las de asimetría y curtosis. Las medidas de asimetría muestran si en la distribución hay concentración de datos en un extremo, superior o inferior, y se denomina Sesgo positivo o a la derecha si la concentración es en el extremo inferior y Sesgo Negativo o a la izquierda si la concentración es en el superior.

## BIBLIOGRAFÍA

MARTÍN-GUZMAN, P; TOLEDO, I.; LÓPEZ ORTEGA, F.J. y BELLIDO, N. “Manual de estadística: descriptiva” Ed. Thompson Civitas. 2006. Capítulos 1 y 2.

MARTIN-GUZMAN, P; et al. “Curso básico de estadística Económica”. Ed. AC. 3ª edición. 1989. Capítulos 1 y 2.

RUBIO, L.; MARCO, R. “Problemas resueltos de Estadística Descriptiva”. Ed. Nueva Visión. 2004. Capítulo 1.