**ESTADÍSTICA APUNTE N°1 - INTRODUCCIÓN**   
  
1.Introducción:   
La estadística se compone de dos grandes áreas, la descriptiva y la inferencial.   
Por medio de la estadística descriptiva se analizan propiedades de un conjunto de datos referidas al contexto en el que ellos se encuentran inmersos y no como entidades aisladas. Por ese motivo es muy importante antes del análisis descriptivo, conocer y clarificar el entorno del que provienen los datos.

Cuando hablamos del entorno estamos haciendo referencia a la población en estudio (o a una parte de ella) y a sus unidades de análisis y a los atributos que interesa estudiar, siempre conociendo la calidad del dato que se puede extraer.

En los estudios de campo se da con frecuencia y por distintas causas la imposibilidad de acceder a toda la población y es en esos casos donde los datos tienen que extraerse de un subconjunto de ella al que denominamos muestra.

Si el objetivo del análisis de la muestra es realizar inferencias respecto a la población, necesariamente debe ser una muestra representativa que comportándose como una imagen reducida de esa población de estudio refleje sus características. Estas muestras deben ser obtenidas por métodos de muestreo aleatorio. Entendiendo esto, podemos afirmar que la estadística inferencial tiene como objetivo realizar inferencias sobre la población en un ambiente de incertidumbre producto del azar y la aleatoriedad.

La parte de la estadística que trata solamente de describir y analizar un grupo dado sin sacar   
conclusiones o inferencias de un grupo mayor se llama estadística descriptiva. Por otro lado, si una muestra es representativa de una población, se pueden deducir importantes conclusiones acerca de esta, a partir del análisis estadístico de la misma. La parte de la estadística que trata esto es la estadística inferencial.

1. Algunos conceptos fundamentales:   
   Población vs Muestra:   
   El conjunto hipotético de todas las observaciones posibles del tipo que se está investigando, se conoce como universo o población de valores. Cualquier conjunto parcial, limitado, de estas observaciones puede considerarse como una muestra, extraída de ese universo.   
   En la práctica se analizan muestras extraídas de una población debido a que a menudo es   
   imposible, poco práctico y/o muy costoso observar la totalidad de los individuos.   
     
   Muestreo:   
   La tarea de extraer una muestra de una población se denomina muestreo y para que ella   
   represente a la población, es necesario que el muestreo se practique al azar. Muestra al azar o aleatoria es un conjunto de objetos o especímenes extraídos al azar de una población previamente definida, y en la que todos los objetos han tenido igual opción de ser considerados, es decir, igual probabilidad de ser extraídos.   
     
   Tamaño de la muestra (n):   
   El tamaño de la muestra es el número de especímenes (objetos) que la forman; a medida que este tamaño aumenta, sus propiedades se van pareciendo cada vez más a las del universo o población que representa.   
     
   Unidad de Observación o Unidad Experimental:   
   Es el objeto sobre el cual se efectúan las mediciones o se intenta clasificar en categorías. Sería, cada individuo de ese universo.   
     
   Variable (xi):   
   Es cualquier característica o atributo, que varía de una unidad de observación a otra en   
   la población o en la muestra.

3.Etapas en el análisis estadístico   
  
**I. Planteo de la investigación y recolección de los datos:**   
Planear la búsqueda y obtención de la información. Es la etapa del diseño de la   
investigación, en la que se define cómo se llevará a cabo, a fin de responder a las   
preguntas planteadas. Aquí se define cuál es la población objetivo, cuáles serán los   
métodos de selección de las muestras, cuántos individuos las integrarán, etcétera.   
  
**II. Organizar y sistematizar la información para su descripción y análisis.**   
Es la etapa del resumen y exploración de datos, cuando se pueden confeccionan   
tablas de frecuencia, gráficos y se calculan las medidas de resumen de los datos   
(estadistica descriptiva).   
  
**III. Efectuar las predicciones a través de la estimación y contrastación de hipótesis**A partir de la información organizada en la etapa II y con el uso de métodos   
estadísticos podemos cuantificar la posibilidad de cometer error en las estimaciones   
y predicciones (estadistica inferencial).

=======================================================================

**ESTADÍSTICA APUNTE N°2 - Los Datos y su Organización.**

Los datos son tomados de una cierta población (o universo) objeto de estudio que llamaremos la población objetivo.

Por ejemplo, todos los alumnos de una determinada universidad constituyen una población objetivo.

Y en ese caso, cada individuo de ese universo –cada alumno– es lo que se denomina una unidad de observación.

**2.Tipos de variables**

La variable es una característica o atributo de cada una de las unidades de observación que se quiere estudiar.

**Las variables se pueden clasificar en:**

**- Cualitativas o categóricas:**cuando los valores que asumen no son números. Por ejemplo, el estado civil, la ciudad de residencia o la nacionalidad.

**- Cuantitativas:** cuando los valores que toman son numéricos. Por ejemplo, la edad o la estatura.

**Las variables cualitativas, a su vez, pueden ser:**

**- Ordinales:**Admiten un orden según algún criterio, por ejemplo, el nivel de escolaridad alcanzado por las personas.

**- Nominales:**No admiten un orden, como la nacionalidad.

**Las variables cuantitativas, se clasifican en:**

**- Discretas:** cuando los valores que asumen son números enteros.   
Por ejemplo, cantidad de miembros en la familia.

**- Continuas:** cuando los valores que asumen son números reales, pueden tomar cualquier número entre dos valores enteros.   
Por ejemplo, el peso de un bebe al nacer.

**3.Tablas de frecuencia**

Para cada variable, se puede construir una tabla de frecuencias.

Ubicando en la primera columna, los valores posibles y en la segunda, la cantidad de personas que contestaron a esa variable con este valor. A ese número lo llamaremos frecuencia absoluta.

**La frecuencia absoluta de un valor de la variable es la cantidad de veces que ese valor está en el conjunto de datos.**

Cada valor de la variable tiene su propia frecuencia.

Ejemplo N°2: 25 estudiantes votaron por el destino de su viaje de egresados entre: Salta (S),

Bariloche (B), Cataratas (C), La Rioja (R) y Mar del Plata (M).

Los resultados fueron:

M M S R S M S R C B S B C S R M S C C S B S B C S

Los valores de la variable “Destino de preferencia” son S, B, C, R y M

Salta fue elegida 9 veces; por tanto 9 es la frecuencia absoluta del valor “Salta” que toma la variable.

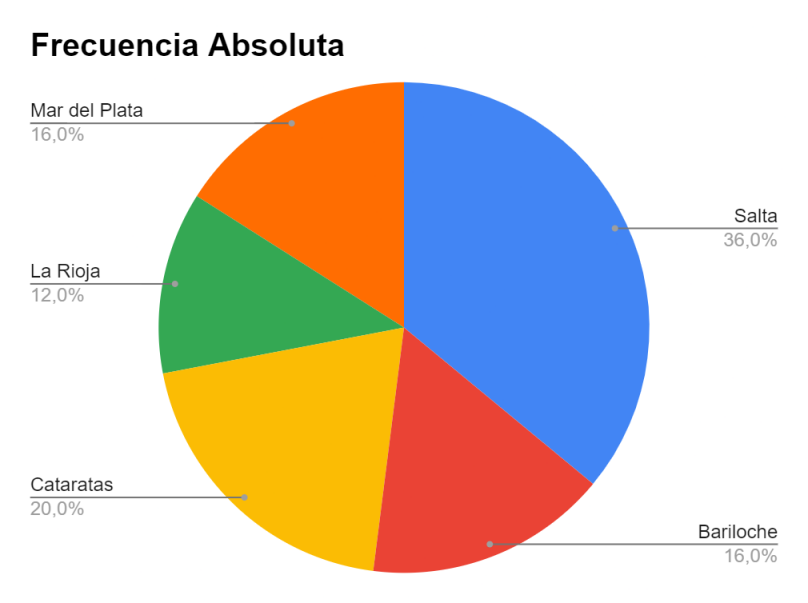
Así, la frecuencia absoluta de “Bariloche” es 4, de “Cataratas” es 5, de “La Rioja” es 3 y de “Mar del

Plata” es 4.

|  |  |
| --- | --- |
| Destino de Preferencia | Frecuencia Absoluta |
| Salta | 9 |
| Bariloche | 4 |
| Cataratas | 5 |
| La Rioja | 3 |
| Mar del Plata | 4 |
| Total | 25 |

**4.Representación gráfica**

Las representaciones gráficas de los datos pueden realizarse automáticamente con el programa Excel. Se seleccionan los datos que se desea representar y, mediante el menú “Insertar”, se elige el tipo de gráfico. De ese modo se realizó el siguiente “diagrama de torta”.



En la tercera columna de la tabla de frecuencias podemos calcular las frecuencias relativas.

La frecuencia relativa de un valor de la variable es la proporción de veces que ese valor está en el conjunto de datos.

La frecuencia relativa se calcula como el cociente entre la frecuencia absoluta y el total de individuos observados (tamaño del conjunto de datos) y expresa el “peso” que cada valor tiene en el total de observaciones.

En nuestro ejemplo1, la frecuencia relativa de “Salta” es 9/25 = 0,36; la de “Bariloche” es 4/25 = 0,16; la de “Cataratas” es 5/25 = 0,2; la de “La Rioja” es 3/25 = 0,12 y la de “Mar del Plata” es 4/25 = 0,16.

Salta fue elegida por 9 entre 25; eso es lo que significa el cociente 9/25 o 0,36. Pero es habitual referirse a este resultado como que “El 36% de los estudiantes eligieron como destino a Salta”; eso motiva la definición de Frecuencia Porcentual.

|  |  |  |  |
| --- | --- | --- | --- |
| Destino de Preferencia | Frecuencia Absoluta | Frecuencia Relativa | Frecuencia Relativa |
| Salta | 9 | 0,36 | 36% |
| Bariloche | 4 | 0,16 | 16% |
| Cataratas | 5 | 0,2 | 20% |
| La Rioja | 3 | 0,12 | 12% |
| Mar Del Plata | 4 | 0,16 | 16% |
| Total | 25 | 1 | 100% |

La frecuencia porcentual de un valor de la variable es el porcentaje de veces que ese valor está en el conjunto de datos.

Lo que hacemos en la cuarta columna es expresar la frecuencia relativa en unidades de 100; es decir, multiplicar por100 la frecuencia relativa. Igual que la frecuencia relativa, expresa el “peso” que cada valor tiene en el total de observaciones.

Volviendo al Ejemplo 1, el porcentaje de “Salta” es 0,36x100=36; el de “Bariloche” es

0,16x100=16; el de “Cataratas” es 0,2x100=20; el de “La Rioja” es 0,12x100=12 y el de “Mar del Plata” es 0,16x100=16.

La frecuencia porcentual no da una información diferente de la frecuencia relativa, sólo la expresa en otras unidades (de modo similar a cuando decimos 0,5 metros o 50 centímetros).

Para resumir todo lo visto, con el objetivo de analizar y comprender más directamente la información, conviene organizarla en una tabla, de la siguiente manera.

Obsérvese que las frecuencias absolutas deben sumar el total de datos (o de observaciones, en nuestro ejemplo n = 25), las frecuencias relativas deben sumar siempre 1 y las porcentuales100.

Cuando la variable es cuantitativa o cualitativa, pero está medida en un nivel ordinal, y, por tanto, tiene sentido considerar un orden de sus valores, puede ser útil tener en cuenta la frecuencia de todos los valores menores o iguales a un valor dado xi; es decir, la frecuencia acumulada (F) en x.

Se puede acumular cualquier tipo de frecuencias (absoluta, relativa o porcentual).

=======================================================================

**Medidas Descriptivas de Tendencia Central.**

**Media Aritmética, Mediana y Moda.**

**1.Funciones de Cálculo para Datos Individuales**

Se conocen varias medidas que cumplen con el requisito de posicionar a un conjunto de datos.

De allí que también se las denomina **Medidas de Posición**, siendo las más comúnmente utilizadas la **media aritmética**, la **mediana** y la **moda**.

El cálculo de estas medidas descriptivas difiere de acuerdo con el tipo de variable con que se trabaja y presenta modificaciones según se disponga de datos individuales o agrupados en categorías o intervalos de clase.

**Media Aritmética -**

Expresada de forma más intuitiva, podemos decir que la media (aritmética) es la cantidad total de la variable distribuida a partes iguales entre cada observación.

Por ejemplo, si en una habitación hay tres personas, la media de dinero que tienen en sus bolsillos sería el resultado de tomar todo el dinero de los tres y dividirlo a partes iguales entre cada uno de ellos.

También la media aritmética puede ser denominada como centro de gravedad de una distribución, el cual no es necesariamente la mitad.

Ejemplo. La media aritmética de 8, 5 y -1 es igual a (8 + 5 + (-1)) / 3 = 4.

Se simboliza:

Media aritmética poblacional: **µ**

Media aritmética muestral:

Se conoce comúnmente como **promedio**. Se calcula como la suma de los valores que toma la variable en estudio dividida por el número total de unidades experimentales observadas.

**Media aritmética para datos individuales.**

Su fórmula es la siguiente:

**Ejemplo 1**. Calcular la media de los siguientes datos: 11, 6, 7, 7, 4.

=7

**Mediana para datos individuales**

**Ejemplo 1** (serie impar de valores). Calcular la mediana de los siguientes datos: 11, 6, 7, 7, 4

Se ordenan los datos de menor a mayor: 4, 6, 7, 7, 11

La mediana es el dato que se encuentra al centro: 4, 6, **7**, 7, 11.

Me = 7

**Ejemplo 2** (serie par de valores). Calcular la mediana de los siguientes datos: 3, 6, 7, 9, 4, 4.

Se ordenan los datos de menor a mayor: 3, 4, **4, 6**, 7, 9.

La **mediana es la media entre los dos valores centrales.**

**Moda - Mo**

**La moda es el valor de la variable de máxima frecuencia (el que más se repite).** También podemos decir que la moda es el valor de la variable con mayor frecuencia absoluta.

**Ejemplo 1**. Calcular la moda de los siguientes datos: 11, 6, **7, 7**, 4.

El valor que más se repite es el 7, con frecuencia absoluta de 2.

**Mo = 7**

**Ejemplo 2**. En un examen calificado del 0 al 10, 3 personas obtuvieron 5 de nota, 5 personas

obtuvieron 4 de nota, y 2 personas obtuvieron 3 de nota. Calcular la moda.

Los datos son los siguientes: 5, 5, 5, **4, 4, 4, 4, 4**, 3, 3.

El valor que más se repite es el 4, con frecuencia absoluta 5.

**Mo = 4**

**Distribuciones multimodales**. Si en un grupo de datos, dos o más valores tienen la misma

frecuencia, y es la frecuencia máxima, entonces la distribución tiene dos o más modas y decimos que

es **bimodal (2 modas), o multimodal (varias modas).**

**Ejemplo.**Calcular la moda de los siguientes datos: 3, 4, 4, 6, 7, 7, 9, 11.

Hay 2 valores que se repiten 2 veces, el 4 y el 7.

**Mo = 4 y 7.**

=====================================================================

**2. Funciones de Cálculo para Datos Ordenados enTablas de Frecuencias sin**

**Intervalos (agrupamiento en categorías)**

**Media aritmética para datos agrupados en categorías (variables cuantitativas discretas)**

Cuando se busca calcular la media para datos agrupados sin intervalos o agrupados puntualmente, es necesario realizar algunos pasos adicionales y conocer algunas funciones.

La **media aritmética** es el valor que se obtiene al sumar todos los datos multiplicados por su

frecuencia y dividir el resultado entre la cantidad de datos.

**Ejemplo 1**. Calcular la media de la siguiente distribución de frecuencias de la variable Cantidad de Personas por Familia en una muestra de 40 socios de una cooperativa:

|  |  |  |
| --- | --- | --- |
| **Categoria** | **Frec. Abs.** | **Frec. Ab. Acum.** |
| **xi** | **ni=fi** | **Ni=Fi** |
| 1 | 9 | 9 |
| 2 | 6 | 15 |
| 3 | 10 | 25 |
| 4 | 10 | 35 |
| 5 | 3 | 38 |
| 6 | 1 | 39 |
| 7 | 1 | **40** |
|  | **40** |  |

Para calcular la media aritmética, vamos a agregar una columna adicional, en la que

**multiplicaremos el valor de la variable (x) por la frecuencia absoluta (f).**

|  |  |  |  |
| --- | --- | --- | --- |
| **Categoria** | **Frec. Abs.** | **Cat.\*Frec.** | **Frec. Ab. Acum.** |
| **xi** | **ni=fi** | **xi\*fi** | **Ni=Fi** |
| 1 | 9 | 9 | 9 |
| 2 | 6 | 12 | 15 |
| 3 | 10 | 30 | 25 |
| 4 | 10 | 40 | 35 |
| 5 | 3 | 15 | 38 |
| 6 | 1 | 6 | 39 |
| 7 | 1 | 7 | **40** |
|  | **40** | **119** |  |

**Mediana**

Es un valor de la variable que divide a un conjunto de observaciones (ordenadas en forma creciente o decreciente) en dos subconjuntos que contienen la misma cantidad de datos.

La **mediana** ocupa la posición central si **n es impar**, y es el promedio de los dos datos centrales si **n es par**, cuando todos los datos están ordenados.

Cuando buscamos encontrar la mediana para datos agrupados sin intervalos o agrupados

puntualmente, necesitamos realizar algunos pasos adicionales y conocer un par de fórmulas.

Una manera rápida de encontrar la mediana es encontrar el dato que ocupa la siguiente posición, y luego, ubicar dicho **valor de la variable (x)** a partir de la columna de frecuencias acumuladas (**Ni = Fi**).

**Procedimiento**. Calcular las frecuencias absolutas acumuladas (**Ni = Fi**) correspondientes a cada valor de la variable. Luego establecer el orden de localización de la mediana efectuando el cociente **(n + 1) / 2**, siendo **n** el número total de datos. En la columna de frecuencias acumuladas se busca la frecuencia que no supera a **(n + 1) / 2.** El valor de x correspondiente a dicho valor es el valor Mediano.

**Ejemplo 1** (n impar). Calcular la mediana de la siguiente distribución de frecuencias

|  |  |  |
| --- | --- | --- |
| **Categoria** | **Frec. Absoluta** | **Frec. Acumulada** |
| **xi** | **ni=fi** | **Ni=Fi** |
| 3 | 9 | 9 |
| 4 | 11 | 20 |
| 5 | 13 | **33** |
|  | **33** |  |

En este ejemplo la posiciónde la mediana es 17 (contenido en el valor acumulado 20 de la tabla), que se corresponde con el valor de la variable 4 que es el valor mediano.

**Me = 4.**

**Moda.**

La moda es el valor que más se repite. También podemos decir que la moda es el valor con mayor frecuencia absoluta o el valor que ocurre con más frecuencia, es decir, es el valor más común.

La moda puede no existir, incluso si existe puede no ser única (distribuciones bimodales, trimodales, etc.). El valor de la moda cuando los datos están agrupados en categorías (variables cualitativas o cuantitativas discretas), corresponde al o a los valores de la variable con máxima frecuencia absoluta.

La moda se representa con las letras:

**Mo.**

**Ejemplo 1**.

Calcular la moda de la siguiente distribución de frecuencias de la variable Nivel de Educación del Jefe de Familia en una muestra de 40 socios de una cooperativa:

|  |  |  |
| --- | --- | --- |
| **Categoria** | | **Frec. Abs** |
| **xi** | | **ni=fi** |
| 1) Primaria Completa o Incompleta | | 9 |
| 2) Secundaria Incompleta | | 14 |
| 3) Secundaria Completa | | 7 |
| 4) Universitaria Incompleta | | 4 |
| 5) Universitaria Completa | | 6 |
|  |  | **40** |

**Mo = Secundaria Incompleta**

**Ejemplo 2**. Calcular la Moda de la siguiente distribución de frecuencias de la variable Cantidad de Personas por Familia en una muestra de 40 socios de una cooperativa:

|  |  |
| --- | --- |
| **Categoria** | **Frec. Abs.** |
| **xi** | **ni=fi** |
| 1 | 9 |
| 2 | 6 |
| 3 | 10 |
| 4 | 10 |
| 5 | 3 |
| 6 | 1 |
| 7 | 1 |
|  | **40** |

**Mo = 3 y 4 Personas por familia. Distribución Bimodal.**

**3. Funciones de Cálculo para Datos Ordenados enTablas de Frecuencias por**

**Intervalos de Clase.**

**Media aritmética** para datos agrupados en clases de frecuencia (variables cuantitativas

continuas)

Cuando se busca calcular la media para datos agrupados en intervalos de clases, es necesario realizar algunos pasos adicionales y conocer algunas funciones.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Intervalos de Frecuencias** | | **Marca de Clase** | **Frec. Absolutas** | **Marca\*Frecuencia** |
| **Li** | **Ls** | **mi=xi** | **ni=fi** | **xi\*fi** |
| Li1 | Ls1 | x1 | f1 | x1\*f1 |
| Li2 | Ls2 | x2 | f2 | x2\*f2 |
| Li3 | Ls3 | x3 | f3 | x3\*f3 |
| Li4 | Ls4 | x4 | f4 | x4\*f4 |
| Li5 | Ls5 | x5 | f5 | x5\*f5 |
| Li6 | Ls6 | x6 | f6 | x6\*f6 |
| ... | ... | x. | f. | x.\*f. |
| Lis | Lss | xs | fs | xs\*fs |
|  |  |  | **Efi** | **E=1an(xi\*fi)** |

**Ejemplo.**

Calcular la media de la siguiente distribución de frecuencias de la variable calificaciones

de 50 estudiantes de un curso de Estadística.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Intervalos de Frecuencias** | | **Marca de Clase** | **Frec. Absolutas** | **Marca\*Frecuencia** |
| **Li** | **Ls** | **mi=xi** | **ni=fi** | **xi\*fi** |
| 29,5 | 39,5 | 34,5 | 4 | 138 |
| 39,5 | 49,5 | 44,5 | 6 | 267 |
| 49,5 | 59,5 | 54,5 | 8 | 436 |
| 59,5 | 69,5 | 64,5 | 12 | 774 |
| 69,5 | 79,5 | 74,5 | 9 | 670,5 |
| 79,5 | 89,5 | 84,5 | 7 | 591,5 |
| 89,5 | 99,5 | 94,5 | 4 | 378 |
|  |  |  | **50** | **3255** |

El valor de la media es Puntos

**Mediana**

La mediana en una distribución de frecuencias es un valor que divide a la distribución en dos partes iguales. En una tabla de Frecuencias por Intervalos de Clase, es el valor que corresponde al 50% en la curva de frecuencias acumuladas.

Para estimar la mediana, hay que seguir los siguientes pasos:

1.- Construir la Tabla de Frecuencias por clases con sus frecuencias absolutas y acumuladas:

2.- Encontrar el intervalo en el que se encuentra la mediana y calcular la misma usando las funciones:

Forma de Cálculo:

**Donde**:

**Li**: límite inferior del intervalo en el cual se encuentra la mediana.

**n**: número de datos del estudio. Es la sumatoria de las frecuencias absolutas.

**Fi-1**: frecuencia acumulada del intervalo anterior al que se encuentra la mediana.

**Ai**: amplitud del intervalo en el que se encuentra la mediana.

**fi**: frecuencia absoluta del intervalo en el que se encuentra la mediana.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Intervalos de Frecuencias** | | **Marca de Clase** | **Frec. Absolutas** | **Marca\*Frecuencia** | **Frec. Acumulada** |  |
| **Li** | **Ls** | **mi=xi** | **ni=fi** | **xi\*fi** | **Ni=Fi** |  |
| 29,5 | 39,5 | 34,5 | 4 | 138 | 4 |  |
| 39,5 | 49,5 | 44,5 | 6 | 267 | 10 |  |
| 49,5 | 59,5 | 54,5 | 8 | 436 | 18 | **Fi-1** |
| **59,5** | **69,5** | **64,5** | **12** | **774** | **30** | **Fi-1** |
| 69,5 | 79,5 | 74,5 | 9 | 670,5 | 39 |  |
| 79,5 | 89,5 | 84,5 | 7 | 591,5 | 46 |  |
| 89,5 | 99,5 | 94,5 | 4 | 378 | 50 |  |
|  |  |  | **50** | **3255** |  |  |

**Posicion: (50+1)/2=25,5**

Al ser una **serie par** de valores (50 alumnos), la posición de la Mediana de la variable calificaciones, corresponde a las unidades de observación (alumnos que rindieron el examen), ordenados en la posición 25 y 26 (25,5), correspondiendo al valor **30 (Fi**) de la columna de frecuencias acumuladas.

En tanto que el valor acumulativo anterior corresponde a **18 (Fi-1)**.

El intervalo que contiene al valor mediano es el comprendido entre las calificaciones de **59,5 y 69,5** puntos. Su límite inferior de clase es **59,5 (Li),** la amplitud de cada clase a **10** puntos **(Ai)** y su frecuencia absoluta de 12 alumnos **(fi).**

Reemplazando en la fórmula

Me = 59.5 + ((25-18) / 12) \* 10 = 65,83

**Moda - Mo**

Para estimar la moda cuando se trata de variables cuantitativas continuas, ordenadas en clases de

frecuencia, se siguen los siguientes pasos:

1.- Construir la Tabla de Frecuencias por clases con sus frecuencias absolutas:

|  |  |  |  |
| --- | --- | --- | --- |
| **Intervalos de Frecuencias** | | **Marca de Clase** | **Frec. Absolutas** |
| **Li** | **Ls** | **mi=xi** | **ni=fi** |
| Li1 | Ls1 | x1 | f1 |
| Li2 | Ls2 | x2 | f2 |
| Li3 | Ls3 | x3 | f3 |
| Li4 | Ls4 | x4 | f4 |
| Li5 | Ls5 | x5 | f5 |
| ... | ... | ... | ... |
| lis | lss | xs | fs |
|  |  |  |  |

2.- Encontrar el intervalo en el cual se encuentra la moda, que es el intervalo con mayor frecuencia

absoluta.

3.- Usar la siguiente fórmula para estimar el valor de la moda:

**Donde**:

**Li**: límite inferior del intervalo en el cual se encuentra la moda.

**fi-1**: frecuencia absoluta del intervalo anterior en el que se encuentra la moda.

**fi**: frecuencia absoluta del intervalo en el que se encuentra la moda.

**fi+1**: frecuencia absoluta del intervalo siguiente en el que se encuentra la moda.

**Ai** : amplitud del intervalo en el que se encuentra la moda.

**Ejemplo**. Encontrar la moda de la siguiente distribución de frecuencias de la variable calificaciones de 50 estudiantes de un curso de Estadística.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Intervalos de Frecuencias** | | **Marca de Clase** | **Frec. Absolutas** |  |
| **Li** | **Ls** | **mi=xi** | **ni=fi** |  |
| 29,5 | 39,5 | 34,5 | 4 |  |
| 39,5 | 49,5 | 44,5 | 6 |  |
| 49,5 | 59,5 | 54,5 | 8 | fi-1 |
| **59,5** | **69,5** | **64,5** | **12** | fi |
| 69,5 | 79,5 | 74,5 | 9 | fi+1 |
| 79,5 | 89,5 | 84,5 | 7 |  |
| 89,5 | 99,5 | 94,5 | 4 |  |
|  |  |  | **50** |  |

**Intervalo de Clase Modal**: 59.5 /69.5

**Mo = mi = Marca de Clase =** 64,5

**Li**: 59,5

**fi-1**: 8

**fi**: 12

**fi+1**: 9

**Ai**: 10

**Mo = Li + {** (fi **-** fi-1) / **[** (fi **-** fi-1) **+** (fi **-** fi+1) **]** **} \*** **Ai**

**Mo =** 59.5 **+** **{** (12 - 8) / **[**(12 - 8) **+** (12 **-** 9)**] } \*** 10 **= 65,71**

**Apunte 4**

**Medidas de Orden**

**Medidas Descriptivas de Orden**

**Funciones de Cálculo**

Se conocen varias medidas que cumplen con el requisito de posicionar a un conjunto de datos. De allí que se las denomina **Medidas de Posición**, siendo las más comúnmente utilizadas los **Cuartiles,** los**Deciles** y los**Percentiles**.

El cálculo de estas medidas descriptivas difiere de acuerdo al tipo de variable con que se trabaja y presenta modificaciones según se disponga de datos individuales o agrupados en categorías o intervalos de clase.

**Cuartiles**. Son valores que dividen a la distribución en cuatro partes iguales. Se representan por **Q1 , Q2 , Q3** ; se llaman primero, segundo y tercer cuartil (corresponden al 25%, 50%, 75% de la distribución). El segundo cuartil es igual a la mediana.**Q2=Me**

**Deciles**. Son valores que dividen a la distribución de frecuencias en diez partes iguales (D1, D2, etc.). El quinto decil es igual a la mediana. **D5=Mediana Percentiles**.

Son valores que dividen a los datos en cien partes iguales (P1, P2, etc.). El Percentil

50 es igual a la mediana. **P50=Mediana.**

**Cuartiles**

**Cuartil 1 (Q1).** Es el valor de la variable que divide a un conjunto de observaciones (ordenadas en forma creciente o decreciente) en dos subconjuntos que contienen el 25 % de las observaciones de la serie con valores menores o iguales al Primer Cuartil y el otro 75 % con valores mayores al mismo.

**Cuartil 2 (Q2).** Es un valor de la variable que divide a un conjunto de observaciones (ordenadas en forma creciente o decreciente) en dos subconjuntos que contienen la misma cantidad de datos. El 50 % de las observaciones de la serie son menores o iguales al valor Mediano y el otro 50 % mayores al mismo.

**Cuartil 3 (Q3).** Es el valor de la variable que divide a un conjunto de observaciones (ordenadas en forma creciente o decreciente) en dos subconjuntos que contienen el 75 % de las observaciones de la serie con valores menores o iguales al Primer Cuartil y el otro 25 % con valores mayores al mismo.

Para poder calcular los cuartiles es necesario conocer la posición del valor de la variable

correspondiente a cada Cuartil, de acuerdo a:

**Medidas de dispersión o variabilidad y medidas de forma.**

Aunque para algunos propósitos un promedio puede ser una descripción suficiente de una población o de una muestra, generalmente se necesita mayor información acerca del conjunto de observaciones en estudios.

Una característica muy importante de los datos estadísticos es su variabilidad, pues las mediciones efectuadas en cualquier estudio difieren de unidad experimental a unidad experimental es su variabilidad.

Otra característica importante es la forma de la representación gráfica de la distribución de frecuencias, esto es la forma simétrica o asimétrica respecto de un eje y la altura de dicha distribución de frecuencias.

Las medidas de **tendencia central (posición – centro)**, como la media, la mediana y la moda, solo describen el centro de los datos, pero no nos dicen nada acerca de la **dispersión (variabilidad)** de los datos, siendo que, en ocasiones, es muy importante conocer que tan dispersos o separados se encuentran los datos, y esto se consigue con las medidas de dispersión o variabilidad. En igual sentido respecto de la **forma** de la distribución.

**Dispersión o Variación.**

Es el grado en que los datos numéricos tienden a extenderse alrededor de un valor medio.

Intuitivamente podemos darnos cuenta de que una medida de posición es más representativa del total de observaciones, si la variabilidad del conjunto es pequeña.

**Ejemplo.**

Las estaturas correspondientes a tres (3) personas son 1,71; 1,72 y 1,73 metros, y su media aritmética es 1,72 metros.

Pero si las alturas de las tres (3) personas hubiesen sido 1,66; 1,72 y 1,78 metros, su altura media también hubiera sido 1,72 metros.

1,72 metros, como media aritmética, describe mejor el primer conjunto de datos, pues las tres (3) alturas son parecidas entre sí, y a su vez cercanas al promedio.

Para poder determinar si los datos del primer grupo son más parecidos entre sí que los del segundo, deberíamos tomar un punto de referencia y medir las diferencias entre cada valor observado y el punto de referencia establecido.

Generalmente el punto de referencio elegido corresponde a una medida de posición adecuada. Si por ejemplo tomamos la Media Aritmética (Punto de Respuesta o de Referencia):

Primera muestra: n1; x̄1 = 1,72

D1 = (x11 - x̄) = 1,71 – 1,72 = - 0,01

D1 = (x12 - x̄) = 1,72 – 1,72 = 0,00

D1 = (x13 - x̄) = 1,73 – 1,72 = 0,01

Segunda muestra: n2; x̄2 = 1,72

D2 = (x21 - x̄) = 1,71 – 1,72 = - 0,06

D2 = (x22 - x̄) = 1,72 – 1,72 = 0,00

D2 = (x23 - x̄) = 1,73 – 1,72 = 0,06

Para poder comparar las dispersiones de ambas muestras, deberíamos establecer una medida que las resuma, y esa medida podría ser el promedio de esas diferencias dentro de cada muestra.

Tendríamos:

D1 = (-0,01 + 0,00 + 0,01) / 3 = 0 D2 = (-0,06 + 0,00 + 0,06) / 3 = 0

Las diferencias negativas se compensan con las positivas y el promedio siempre es 0.

Lo expresado responde a una propiedad de la media aritmética que dice: ¨La suma de las desviaciones

de cada valor de la variable con respecto a la media aritmética es 0¨.

Esta propiedad nos indica que el simple promedio de las desviaciones no nos sirve para determinar la

variabilidad de un conjunto de datos, siendo necesario recurrir a otros indicadores que actúen como

medidas resúmenes.

Las más empleadas son el rango, la desviación media, la varianza y la desviación típica.

1. Rango, amplitud o recorrido.

2. La desviación media.

3. Varianza.

4. Desviación estándar.

5. Coeficiente de variación.

Rango:

El rango es la diferencia entre el valor máximo y mínimo de un conjunto de datos. Es la medida de dispersión o variabilidad más sencilla de calcular y tiene las mismas unidades que los datos. También es llamado amplitud o recorrido.

R = Valor Máximo – Valor Mínimo

El rango, amplitud total, recorrido o extensión de un conjunto de números es la diferencia entre el mayor y menor de todos ellos.

Ejemplo 1. Supongamos que hay dos grupos de estudiantes, A y B, y que ambos tienen una media de 70 puntos; con la información que nos proporciona la media parece no existir diferencia entre los dos grupos. Pero si se da la información adicional de los puntajes más alto y más bajo, en el grupo A son 99 y 25 y en el grupo B son 73 y 66, se ve inmediatamente que, aunque ambos grupos tienen la misma media, ellos son muy diferentes por la variabilidad de los puntajes.

µ1 = x̄1= 70 µ2 = x̄2 = 70

R1 = 99 – 25 = 74 R2 = 73 - 66 =7

Esta observación supone que los valores del grupo A están repartidos entre 99 y 25.

Puede ocurrir, sin embargo, que haya un sólo puntaje de 99 y otro de 25 y que el resto se halle entre 69 y 74 puntos.

El rango no puede darnos esta información porque se funda exclusivamente en valores extremos.

Características del rango:

1. Tiene la misma unidad de medida que las observaciones.

2. Se utiliza para tener una idea rápida del grado de dispersión de un conjunto de datos.

3. Es poco confiable.

4. El rango muestral es muy inestable.

5. El valor del rango no varía cuando se suma una constante K a cada observación de un conjunto de datos.

6. El valor del rango si varía cuando se multiplica por constante K a cada observación de un conjunto de datos.

**Desviación Media**

1. Datos Individuales (Variables cuantitativas Discretas y Continuas).

La desviación media de un conjunto de datos individuales es la media aritmética de los valores absolutos

de lo que se desvía cada valor respecto a la media aritmética.

D.M.: Desviación Media.

x̄ = µ: Media aritmética de los datos.

x1, x2, x3, …, xn: datos.

xi: cada uno de los datos.

n: número de datos.

Ejemplo 1. Las calificaciones de 10 estudiantes de un curso se dan en la tabla siguiente.

Calcular la Desviación Media.

|  |  |  |
| --- | --- | --- |
| **Calificaciones** | **Desvios** | **Desv. Abs** |
| **xi** | **(xi-x¨)** | **|xi-x¨|** |
| 60 | -1,9 | 1,9 |
| 30 | -31,9 | 31,9 |
| 85 | 23,1 | 23,1 |
| 52 | -9,9 | 9,9 |
| 65 | 3,1 | 3,1 |
| 47 | -14,9 | 14,9 |
| 84 | 22,1 | 22,1 |
| 65 | 3,1 | 3,1 |
| 57 | -4,9 | 4,9 |
| 74 | 12,1 | 12,1 |
| **619** | **0** | **0** |

n = 10

µ = x̄ = 619 / 10 = 61,90 puntos

1. M. = 127 / 10 = 12,70 puntos

2. **Datos agrupados en categorías o clases (Variables cuantitativas discretas).**

Se calcula a partir de una tabla de frecuencias con datos agrupados en categorías, por tratarse de una

variable discreta que toma valores puntuales, No intervalos de valores.

D.M.: desviación media.

fi: frecuencia absoluta de cada valor, es decir, número de veces que aparece el valor en el estudio.

x1, x2, x3, …, xn: datos.

xi: cada uno de los datos.

k: número de clases.

x̄: media aritmética de los datos.

n: número de datos.

Ejemplo 1. Cantidad de Personas por Familia en una cooperativa de 200 socios. Calcular la Desviación

Media.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Categoria** | **Frec. Abs.** | **Cat. \* Frec.** | **Desvio** | **Desv. Abs.** |
| **xi** | **ni = fi** | **xi \* fi** | **fi \* ( xi - x¨)** | **fi \* | xi - x¨ |** |
| 1 | 9 | 9 | -18 | 18 |
| 2 | 6 | 12 | -6 | 6 |
| 3 | 10 | 30 | 0 | 0 |
| 4 | 10 | 40 | 10 | 10 |
| 5 | 3 | 15 | 6 | 6 |
| 6 | 1 | 6 | 3 | 3 |
| 7 | 1 | 8 | 5 | 5 |
|  | **40** | **120** | **0** | **48** |

n = 40

µ = x̄ = 120 / 40 = 3 personas por familia

D.M. = 48 / 40 = 1,20 personas por familia = 1 persona por familia (dato discreto)