NOTAS SOBRE ESTADÍSTICA APLICADA A LA CALIDAD

1. CONCEPTO DE ESTADÍSTICA:

Es la ciencia que estudia la interpretación de datos numéricos.

a) Proceso estadístico:

Es aquél que a partir de unos datos numéricos, obtenemos unos resultados mediante unas reglas y unas operaciones.

Proceso estadístico:

- Elección de la unidad estadística.
- Recogida, análisis y presentación de los datos.
- Ordenación de los datos.
- Cálculo de las medidas de posición y dispersión.
- Representación gráfica.
- Análisis y predicción de resultados.
- Análisis de errores. Significación. Fiabilidad

Dentro de la estadística se distingue dos **ramas**:

- Estadística Inductiva
- Estadística Descriptiva.

Nota: En estas notas, sólo veremos la Estadística Descriptiva.

b) Estadística Descriptiva:

Es un modelo que permite acumular información, analizarla y sintetizarla, para describir un fenómeno.

2. POBLACIÓN, CARACTERÍSTICA Y MUESTRA:

El objetivo formal de la descripción estadística es la **masa estadística.**

Masa estadística: Es el conjunto de unidades que tienen características de identificación comparables en cada estudio. Se llama frecuentemente **población**, aunque estas poblaciones no se limitan a un conjunto de personas, sino que podrán se, por ejemplo, un conjunto de coches, un conjunto de productos agrícolas, etc.

El objetivo de un **proceso estadístico** es observar y comentar las distintas características de una población estadística, una vez que ésta se encuentra bien definida.

De esta forma la **característica** se divide o clasifica a la población original en masas parciales o **subpoblaciones estadísticas,** por ejemplo, de una población de coches los de 2.000 cc, etc.

Subpoblación: Existe una subpoblación, cuando sus elementos reúnen características especiales que no se presentan en los demás elementos de la población.

Es evidente que no siempre se puede hacer una investigación exhaustiva de la población, es decir, sobre todos y cada uno de sus elementos, y que, por tanto hay que recurrir a la encuesta por sondeo; se dice entonces que el objeto de la investigación es una **muestra** o subconjunto del total de casos de la población. Por ejemplo, de una población de estudiantes, queremos saber cuantos han asistido a un concierto que se han celebrado en la ciudad.

Para que una muestra sea representativa debe cumplirse que todos los elementos de la población tengan la misma posibilidad de formar parte de la muestra.

La **muestra** puede ser:

- **Aleatoria:** cuando la recogida de datos es al azar.
- **No aleatoria:** cuando los datos se obtiene según unos criterios:
    - ✓ Entrevistas.
    - ✓ Encuestas.

3. CONCEPTOS: INDIVIDUO Y MODALIDAD:

Individuos: son los elementos que componen la población estudiada.

Cada individuo puede describirse según uno o varios caracteres o características, que elegiremos atendiendo a los aspectos de la población sobre los que estemos interesados. Por ejemplo de una población de personas, una característica puede ser la estatura, la edad, etc.

Modalidades: son las diferentes situaciones posibles de un carácter.

Los caracteres de un individuo pueden presentar dos o más modalidades. Por ejemplo, las modalidades del carácter sexo en la población de personas son: masculino y femenino; las modalidades de la característica estatura, por ejemplo, seria 160 cm, 170 cm y 180 cm, etc.

4. CLASES DE CARACTERES: CUALITATIVOS Y CUANTITATIVOS:

En la investigación estadística podemos distinguir dos clases de caracteres o características: cualitativos y cuantitativos.

a) **Características cualitativas o atributos:** Son las que expresan una cualidad que generalmente no tiene representación numérica. De ellas lo único que puede determinarse es la frecuencia con que aparecen. Este tipo de características clasifica la población en categorías.

b) **Características cuantitativas:** Son las que permiten asignar a cada elemento de la población un número real. Ejemplos: la edad de un colectivo de individuos, el número de alumnos de cada clase, etc.

Variable: desde el punto de vista matemático, es aquel ente que puede tomar un valor cualquiera de un conjunto determinado de valores. La característica cuantitativa es una variable matemática, que nosotros cuando nos referimos a la característica cuantitativa la denominaremos **variable estadística.**

Las variables o características cuantitativas pueden ser de dos tipos:

1) **Discreta:** Cuando la variable sólo puede tomar valores numéricos aislados. Por ejemplos: el número de automóviles que pasan por una calle en una hora; el número de alumnos de una clase, ya que no podrá tomar el valor, por ejemplo, 25,37.

2) **Continua:** Es el caso contrario, cuando puede tomar cualquier valor real. Por ejemplo, la estatura de las personas. En general todas las magnitudes relacionadas con el tiempo (edad, duración de un fenómeno,...), la masa (volumen, peso,...) y el espacio (longitud, superficie,...) o una combinación de éstos (velocidad, densidad, capacidad,...) son varables continuas.

Finalmente, llamaremos **dominio** de una variable estadística al conjunto de valores que ésta puede tomar dentro del fenómeno estudiado (es equivalente al dominio de una variable matemática)

5. INTERVALO DE CLASE:

Para estudiar un hecho en el que la amplitud de la población es grande, donde la variable continua puede tomar un número elevado de valores, se define las **clases de valores o intervalo de clases.**

Clases de valores o intervalo de clases: son subconjuntos del conjunto de valores que puede tomar una variable continua. Estas clases pueden tener una amplitud constante o variable.

Por ejemplo, si se trata del estudio de la estatura de una población de individuos, es conveniente dividir en clases las posibles estaturas de los individuos investigados. Podría hacerse de la siguiente forma:

- Menos de 160 cm.
- De 160 a 170 cm.
- De 170 a 180 cm.
- Igual o más de 180 cm.

Se puede observar que la primera y última clase tiene una amplitud indeterminada, mientras las otras dos tienen una amplitud constante de 10 cm. Los extremos de los intervalos son 160 cm, 170 cm y 180cm; y las marcas de clase, 165 cm y 175 cm.

Límites del intervalo: Son los valores extremos de dicho intervalo. Por ejemplo, dado el intervalo del ejercicio anterior (160-170 cm), el 160 cm es el límite inferior y el 170 cm el límite superior.

Los criterios que se debe seguir para elegir las amplitudes de las clases y el número de estas, son los siguientes:

- **En cuanto a la amplitud de cada clase:** estará condicionada por la preocupación de obtener efectivos comparables de una clase a otra. La elección acertada suele ser: elegir clases mayor amplitud en las regiones donde el carácter es más raro (en nuestro ejemplo, las clases de los extremos tienen amplitud ilimitada), y de menor amplitud en el resto.
- **En cuanto al número de clases a adoptar:** dependerá de la precisión de las medidas y del efectivo de la población estudiada. Es conveniente que no sea inferior a 4 ó 6, pues si el número de clases es muy reducido nos conduce a una pérdida de información (el número de clases optimo se sitúa entre 10 y 15 clases).

6. SERIES ESTADÍSTICAS:

A. Concepto de serie estadística:

Una **serie estadística** es un conjunto de observaciones o medidas realizadas en una población, atendiendo a una o varias características determinadas.

Habitualmente las series estadísticas se disponen en tablas que llamaremos Tablas estadísticas, que sirven para contener los datos de la serie de una forma ordenada y fácil de consultar.

B. Frecuencias:

a) Frecuencia absoluta:

Dada una población estadística de **"n"** individuos, y considerada una modalidad concreta de una variable estadística, se denomina **frecuencia absoluta** (o simplemente **frecuencia** de la modalidad) al número de veces que esa modalidad aparece en el total de casos posibles que se presentan en la muestra.

Dada una modalidad $\mathbf{A_i}$, a la frecuencia con que se repite la notaremos por $\mathbf{n_i}$. Por ejemplo, si consideramos una población formada por 50 mujeres, y elegimos como variable “el color del pelo”, y como modalidad “el color negro”, la frecuencia de esta modalidad será el número de individuos que tienen pelo negro.

Es de destacar que <u>la suma de las frecuencias absolutas nos dará el número de casos posibles o total de individuos de la muestra</u>.

b) **Frecuencia relativa:**

Se denomina **frecuencia relativa** de una modalidad al valor de una fracción cuyo numerador es la frecuencia absoluta de esa modalidad y cuyo denominador es el número de individuos de la población. La frecuencia relativa $\int_i$ estará siempre comprendida entre 0 y 1.

Si $\mathbf{n}_i$ es la frecuencia de una cierta modalidad $\mathbf{A_i}$, y $\mathbf{n}$ es el número de individuos, la frecuencia $\int_i$ se calcula:

$$\int_i = \frac{n_i}{n}$$

c) **Frecuencia porcentual:**

Si la frecuencia relativa la expresamos mediante porcentajes, encontramos la frecuencia porcentual. Se calcula multiplicando por 100 el valor de la frecuencia relativa. La frecuencia porcentual estará comprendida lógicamente entre 0 y 100 y la representaremos por ρ_i

$$\rho_i = \int_i \cdot 100 = \frac{n_i}{n} \cdot 100$$

d) **Frecuencia acumulada:**

Considerada una modalidad determinada, la **frecuencia acumulada** nos da la suma de las frecuencias de las modalidades anteriores a ésta, es decir, la frecuencia que se ha acumulado después de considerar las modalidades anteriores.

Como es lógico, podemos hablar de frecuencia absoluta acumulada, frecuencia relativa acumulada y frecuencia porcentual acumulada; sin embargo, es más usual calcular sólo la frecuencia absoluta acumulada para cada modalidad; en el caso de las otras dos frecuencias suele calcularse la frecuencia total acumulada que puede servir como comprobante de no haber incurrido en error en los cálculos anteriores

C. **<u>Tipos de serie estadísticas</u>:**

No todas las series estadísticas se ocupan de características de la misma índole.

a) **Series estadísticas simples y agrupadas:**

Esta clasificación corresponde a la forma en que se puede presentar los datos.

(a). **Series estadísticas simples:**

Nos referimos a las series estadísticas <u>en las que cada dato del hecho estudiado se le asigna de forma unívoca el valor extraído de la observación</u>.

Ejemplos: de estas series son los números de habitantes de cada país de un continente, el número de piezas que cada empleado de una fabrica construye cada día, etc.

(b). **Series estadísticas agrupadas:**

Se refiere fundamentalmente a variables estadísticas continuas o discretas con un gran número de valores

En éstas series estadísticas los datos se agrupan por clases, y a cada clase se le asigna un valor llamado marca de clase que suele coincidir con el valor central.

Ejemplos de estas series son las estaturas de los individuos de una muestra o población, las dimensiones de las piezas de una máquina, etc.

b) **Series cronológicas o temporales:**

Se ocupa del comportamiento de los hechos a lo largo del tiempo.

c) **Cuadros estadísticos:**

Se denomina así a las tablas estadísticas resultante de agrupar varias series estadísticas.

Los cuadros estadísticos se diferencian del resto de las series estadísticas en que consideran más de un carácter. Los cuadros estadísticos tienen doble entrada y podrían considerarse incluso con más de dos.

Ejemplo de cuadros de doble entrada en los que los caracteres examinados corresponden: el primero de ellos, a las filas del cuadro y, el segundo, a las columnas. Distribución por colores y sexo de un conjunto de palomas de una familia.

	Blancas	**Negras**	**Otros colores**	**Total**
Machos	23	13	20	56
Hembras	17	16	25	58
Totales	40	29	45	114

7. MEDIDAS DE LA DISTRIBUCIÓN:

En general, se llaman **medidas de la distribución** a ciertos valores característicos que representan los aspectos más destacables de dicha distribución y facilitan su estudio.

Podemos distinguir fundamentalmente tres pipos de medidas de distribución:

- Medidas de posición.
- Medidas de dispersión.
- Medidas de deformación.

A. Medidas de posición:

También llamadas de centralización o de **"tendencia central".** Sirven para estudiar las características de los valores centrales de la distribución atendiendo a distintos criterios.

Supongamos que queremos describir de una forma breve y precisa los resultados obtenidos por un conjunto de alumnos en un cierto examen; diríamos:

(a). La nota media de la clase es de 6,5.
(b). La mitad de los alumnos han obtenido una nota inferior a 5.
(c). La nota que más veces se repite es el 4,5.

En la expresión (a) se utiliza como medida la **media aritmética** o simplemente **media.** En la (b) se emplea como medida la **mediana**, que es el valor promedio que deja por debajo de ella la mitad de las notas y por encima la otra mitad. Y en la (c) se usa el valor de la nota que más veces se ha repetido en ese examen; este valor es la **moda.**

Estos tres valores, media, mediana y moda, son sin duda los más usados en estadística. Veamos como se calcula cada uno de ellos.

a) **Media aritmética:**

Normalmente se suele distinguir entre **media aritmética simple y media aritmética ponderada.**

(a). **Media aritmética simple:**

Es la suma de todos los elementos de la serie dividida por el número de ellos.

Se calcula así:

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

Ejemplo: Hallar la media aritmética simple de los siguientes valores: 5, 7, 8, 10, 15.

Solución: $\sum x = 5 + 7 + 8 + 10 + 15 = 45 \,; \quad n = 5 \,; \; \bar{x} = \frac{\sum x}{n} = \frac{45}{5} = 9 \text{ de media.}$

(b). **Media aritmética ponderada:**

Por lo general, en estadística, los datos se nos presentan agrupados mediante una distribución de frecuencias que hace que no todos los elementos de la serie tengan el mismo peso específico, y eso influye a la hora de calcular la media; por eso se llama **media ponderada.**

La media ponderada se define como la suma de los productos de cada elemento de la serie por su frecuencia respectiva dividida por el número de elementos de la serie.

Se calcula así:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{n}$$

Ejemplo: Durante el mes de septiembre del 2002 los salarios recibidos por un obrero fueron:

Salarios en euros	Frecuencia en días
6	5
6,61	15
8,41	4

Hallar el salario medio durante ese mes.

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{n} = \frac{(6 \times 5) + (6{,}61 \times 15) + (8{,}41 \times 4)}{24} = 6{,}78$$

(c). **Cálculo de la media aritmética a partir de datos agrupados en clases:**

Hay dos métodos principalmente para calcular la media de una distribución con datos agrupados: método directo (o largo) y método abreviado (o corto), de este último existe una variante llamado método clave. Nosotros solo veremos en método directo.

i. Método directo:

Consiste en aplicar la fórmula ya vista para el cálculo de la media ponderada, con la única salvedad de que se toman como valores representativos de la variables los puntos medios de cada intervalo, que se denotan con $\mathbf{X_m}$

$$\bar{x} = \frac{\sum X_m \cdot n_i}{n}$$

b) **Mediana:**

Una vez dispuestos todos los valores que toma la variable en una serie creciente o decreciente, el valor central de esa serie, si existe, es la mediana. Así pues, la mediana deja el mismo número de valores a su izquierda que a su derecha. Cuando no existe un valor central se puede definir como la media aritmética de los dos valores medios.

Para su cálculo distinguiremos tres casos:

(a) Cálculo de la mediana con datos no agrupados:

Se ordenan los elementos en orden creciente o decreciente, y la mediana es el valor que ocupa el lugar $\frac{n+1}{2}$

Ejemplos:

Ejemplo 1º: Sea la serie: 5, 6, 9, 11, 15, 19, 23, 26, 27

$$\frac{n+1}{n} = \frac{9+1}{2} = 5; \qquad \textit{La Mediana es 15}$$

Ejemplo 2º: Sea la serie: 5, 7, 10, 15, 20, 21, 24, 27

En este caso el número de datos es par. Entonces el valor medio seria

$$\frac{n+1}{n} = \frac{8+1}{2} = 4{,}5;$$

Es decir, estaría comprendido entre el cuarto y el quinto; como el elemento cuarto es 15 y el quinto es 20, la mediana será el valor medio de ambos

$$\frac{15+20}{2} = 17{,}5; \qquad \textit{La Mediana es 17,5}$$

Ejemplo 3º:

En los ejemplos anteriores ocurría que la frecuencia de cada elemento era 1. Pero no siempre sucede así.

Sea ahora la serie: 3, 4, 4, 4, 6, 8, donde el elemento 4 tiene una frecuencia de 3.

Consideremos el intervalo que cada elemento desde 0,5 unidades a la izquierda hasta 0,5 unidades a la derecha. En nuestra serie, los tres elementos "4" se distribuyen entre el 3,5 y 4,5. Los representamos en el eje real de la siguiente forma:

3 4 4 4 6 8

3,5 3,83 4,16 4,5

Vemos que el valor 4,16 deja a su izquierda tres elementos 3, 4 y 4 y a su derecha otros tres 4, 6 y 8, luego la **Mediana es 4,16**

Lo que hemos hecho es dividir el intervalo unidad 3,5-4,5 en tres partes iguales de 0,33 unidades cada una.

Veamos **otro ejemplo** de este tercer caso.

Hallar la mediana de 5, 6, 8, 8, 8, 8, 10, 12, 13, done el elemento 8 tiene una frecuencia 4; y se distribuyen los 4 elementos en el intervalo unidad que va de 8,5 a 9,5 de la siguiente forma:.

5 6 8 8 8 8 10 12 13

7,5 7,75 8 8,25 8,5

En este caso es el tercer ocho el que deja 4 elementos a su izquierda 5, 6, 8 y 8, y cuatro a su derecha 8, 10, 12 y 13. Y los extremos del intervalo de ese tercer ocho son 8 y 8,25. Entonces:

$$\frac{8+8{,}25}{2}=8{,}125; \qquad \textit{La Mediana es } 8{,}125$$

(b) Cálculo de la mediana con datos agrupados (en intervalos y frecuencias):

Cuando los datos conviene agruparlos por intervalos debido al elevado número de ellos, la mediana se calcula de la siguiente forma:

1) Se calcula $\frac{n}{2}$.

2) A la vista de las frecuencias acumuladas, se halla el intervalo que contiene a la mediana. (El que deje el mismo número de elementos a ambos lados suyos).

3) Se calcula la frecuencia del intervalo que contiene la mediana.

4) Se halla uno cualquiera de los limites exactos (el superior o el inferior) del intervalo que contiene a la mediana. (Llamábamos límites exactos de un intervalo a-b, a los números a – 0,5 y b + 0,5.).

5) Se halla la frecuencia de los valores que quedan "por debajo", (a), del intervalo que contiene la mediana, o la frecuencia de los valores que quedan "por encima", (b), y según hayamos decidido (a) o (b), calculamos la mediana por alguna de estas dos fórmulas:

$$(a) \quad M = l + \frac{I}{\int_M} \cdot \left(\frac{n}{2} - \int_i\right)$$

$$(b) \quad M = L + \frac{I}{\int_M} \cdot \left(\frac{n}{2} - \int_s\right)$$

siendo:

- M = Mediana.
- l = Límite inferior del intervalo de la medina.
- L = Límite superior del intervalo de la mediana.
- I = Amplitud del intervalo de la mediana.
- $\int_M$ = Frecuencia del intervalo de la medina.
- $\int_i$ = Frecuencia acumulada de los valores inferiores al intervalo de la mediana.
- $\int_s$ = Frecuencia acumulada de los valores superiores al intervalo de la mediana.
- n = Número total de valores.

(c) Cálculo de la mediana con datos agrupados por frecuencias, pero no en intervalos:

Se puede decir que es un caso particular del método anterior. El procedimiento es el siguiente: una vez calculado el número alrededor del cual se encuentra la mediana, se considera este número como centro de un intervalo de amplitud "1"; a continuación se aplica la fórmula anterior para el cálculo con datos agrupados en intervalos.

Ejemplo:

x	n	$\int_a$
1	5	5
2	7	12
3	6	18
4	12	30
5	20	50
6	15	65
7	11	76
8	6	82
9	5	87
10	2	89

$$n=\frac{89}{2}=44,5$$

Por tanto, la mediana es un valor próximo a 5 .

$$M=4,5+\frac{1}{20}(44,5-30)=5,225$$

c) Moda:

La moda de una serie de números es el valor que se presenta con mayor frecuencia; es decir, el que se repite un mayor de número de veces. Es, por tanto, el valor más común.

En muchos caso la moda es única, pero puede ocurrir también que en una distribución haya dos más modas. Entonces se habla de **distribución bimodal, trimodal,** etc. Incluso no existir moda, como en la serie 2, 3, 4, 5, 7, 10.

(a) Cálculo de la moda con datos agrupados:

En el caso de una distribución de frecuencias con datos agrupados, si hiciéramos una gráfica o curva de frecuencias, la moda sería el valor (o valores) de la variable correspondiente al máximo (o máximos) de la curva.

La moda se puede calcular aplicando la siguiente fórmula:

$$MODA=Mo=l+\left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right)\cdot I$$

donde:

- l = Límite inferior de la clase que contiene a la moda. (Clase modal).
- Δ_1 = Diferencia entre la frecuencia de la clase modal y la frecuencia de la clase contigua inferior.
- Δ_2 = Diferencia entre la frecuencia de la clase modal y la frecuencia de la clase contigua superior.
- I = Amplitud del intervalo de la clase modal. (Y de las demás clases).

Ejemplo:
Hallar la moda de la siguiente distribución de frecuencia:

Clase	Frecuencia
10-20	11
20-30	14
30-40	21
40-50	30
50-60	18
60-70	15
70-80	7
80-90	3
	119

$$l \quad = 40 \,(\text{extremo inferior de 40-50})$$
$$\Delta_1 = 30 - 21 = 9$$
$$\Delta_2 = 30 - 18 = 12$$
$$I \quad = 10$$
$$\textit{luego}: \quad \mathrm{Mo} = 40 + \frac{9}{9+12} \times 10 = 44{,}28$$

B. Medidas de Dispersión:

A pesar de la gran importancia de las medidas de tendencia central y de la cantidad de información que aportan individualmente, no hay que dejar señalar que en muchas ocasiones esa información, no sólo no es completa, sino que puede introducir errores en su interpretación. Veamos algunos ejemplos.

Consideramos dos grupos de personas extraídas como muestra respectivas de dos poblaciones distintas: el primero está compuesto por 100 personas que asisten a la proyección de una película para niños, y el segundo de 100 personas elegidas entre los asistentes a una discoteca juvenil. Pudiera ocurrir que, aun siendo las distribuciones de las edades de ambos grupos distintas, la media y la mediana coincidieran para ambas.

Igualmente ocurre en este otro ejemplo. La caja de un restaurante registra las siguientes entradas en miles de euros, a lo largo de dos semanas correspondientes a épocas distintas del año.

1ª semana	2ª semana
10	30
20	40
30	50
50	50
60	60
80	60
100	.60
350	350

La Media:

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i}{n} = \frac{350}{7} = 50 \quad (\textit{en}\ \text{ambos casos})$$

La mediana:
En ambos casos es 50.

Como vemos la media y la mediana de ambas distribuciones coinciden (el valor de ambas es 50 en los dos casos) y, sin embargo, las consecuencias que se podrían derivar de una y otra tabla son distintas.

A la vista de estos ejemplos, se comprenden la necesidad de conocer otras medidas, a parte de los valores de centralización, que nos indique la mayor o menor desviación de cada observación respecto aquellos valores.

a) **Rango, amplitud total o recorrido:**

El rango se suele definir como la diferencia entre los dos valores extremos que toma la variable.

Comparemos, por ejemplo, estas dos series:

- Serie 1: 1 5 7 7 8 9 9 10 17.
- Serie 2: 2 4 6 8 10 12 14 16 18.

$$1)\ R=X_{max}-X_{min}=17-1=16$$
$$2)\ R=X_{max}-X_{min}=18-2=16$$

b) **Desviación media:**

En teoría, la desviación puede referirse a cada una de las medidas de tendencia central: media, mediana o moda; pero la más utilizada es la medida de desviación con respecto a la media, que llamaremos **desviación media.**

Puede definirse como la media aritmética de las desviaciones de cada uno de los valores con respecto a la media aritmética de la distribución.

Y se indica así, en el caso de datos sin agrupar:

$$DM=\frac{\sum|x-\overline{x}|}{n}$$

Se toma las desviaciones en **valor absoluto,** es decir, que la fórmula no distingue si la diferencia de cada valor de la variable con la media es en más o en menos.

Ejemplo:

Las calificaciones de 10 alumnos en el examen de Estadística han sido las siguientes: 2, 2, 4, 4, 4, 5, 6, 7, 8, 8. Averiguar la desviación media de estos valores. Su cálculo es el siguiente:

x	$x-\overline{x}$	$\lvert x-\overline{x}\rvert$
2	-3	3
2	3	3
4	-1	1
4	-1	1
4	-1	1
5	0	0
6	1	1
7	2	2
8	3	3
8	3	3
50		18

$$\overline{x}=\frac{\sum_{i=1}^{k}x_i}{n}=\frac{50}{10}=5 \qquad DM=\frac{\sum|x-\overline{x}|}{n}=\frac{18}{10}=1{,}8$$

Veamos ahora el caso de datos agrupados en intervalos:

$$DM = \frac{\sum |n_i(x - \bar{x})|}{n}$$

Además, las desviaciones son de cada centro, o marca de clase, a la media aritmética, Es decir:

$$DM = \frac{\sum |n_i(x_m - \bar{x})|}{n}$$

Ejemplo:
Hallar la desviación media de las edades de los 100 empleados de una cierta empresa:

Clase	n_i	X_m	$n_i \cdot X_m$	$\lvert X_m - \bar{X} \rvert$	$n_i \cdot \lvert X_m - \bar{X} \rvert$
16-20	2	18	36	16,72	33,44
20-24	8	22	176	12,72	101,76
24-28	8	26	208	8,72	69,76
28-32	18	30	540	4,72	84,96
32-36	20	34	680	0,72	14,40
36-40	18	38	684	3,28	59,84
40-44	15	42	630	7,28	109,20
44-48	8	46	368	11,28	90,24
48-52	3	50	150	15,28	45,84
	100		3.472		609,44

$$\bar{X} = \frac{\sum n_i \cdot X_m}{N} = \frac{3472}{100} = 34{,}72$$

$$DM = \frac{\sum n_i \cdot (X_m - \bar{X})}{N} = \frac{609{,}44}{100} = 6{,}09$$

La desviación media viene a indicar el grado de concentración o de dispersión de los valores de la variable. Si es muy alta, indica gran dispersión; si es muy baja refleja un buen agrupamiento y que los valores son parecidos entre sí.

c) **Desviación típica**:

Es sin duda la medida de dispersión más importante, ya que además sirve como medida previa al cálculo de otros valores estadśiticos.

La desviación típica se define como la raíz cuadrada de la media de los cuadrados de las desviaciones con respecto a la media de la distribución, es decir, **para datos sin agrupar:**

$$S=\sqrt{\frac{\sum (X_i-\bar{X})^2}{N}}$$

(a) Cálculo de la desviación típica para datos agrupados en clses y agrupados por frecuencias:

1) Método largo:

$$S=\sqrt{\frac{\sum (X_m-\bar{X})^2\cdot n}{N}}$$

2) Método abreviado o corto:

$$S=I\cdot\sqrt{\frac{\sum n\cdot d^2}{N}-(\frac{\sum n\cdot d}{N})^2}$$

donde:

- I = amplitud de las clases (constante).
- d = distancia en clases desde cada una en concreto a la clase que contiene a la media supuesta: A.

d) **Varianza**:

La varianza se defne como el cuadrado de la desviación típica, es decir, S^2, Se utiliza frecuentemente en lugar de la desviación típica.

Cuando la desviación típica se calcula para todos los elementos de la población, se utiliza el signo **σ** para designarla, reservando **S** para cuando medimos únicamente los elementos de la muestra. Análogamente llamaríamos $\mathbf{S^2}$ a la varianza muestral y $\boldsymbol{\sigma^2}$ a la varianza poblacional.