

MATEMÁTICAS APLICADAS A LAS CIENCIAS SOCIALES I

1º DE BACHILLERATO

BLOQUE TEMÁTICO I: ESTADÍSTICA Y PROBABILIDAD.

TEMA 1.- DISTRIBUCIONES UNIDIMENSIONALES .

1.- EL ORIGEN DE LA ESTADÍSTICA.

La palabra Estadística procede del vocablo Estado, pues era función principal de los Gobiernos de los Estados establecer registros de población, nacimientos, defunciones, impuestos, cosechas... La necesidad de poseer datos cifrados sobre la población y sus condiciones materiales de existencia han debido hacerse sentir desde que se establecieron sociedades humanas organizadas.

Es difícil conocer los orígenes de la Estadística. Desde los comienzos de la civilización han existido formas sencillas de estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o ciertas cosas. Hacia el año 3.000 a.C. los babilonios usaban ya pequeñas tablillas de arcilla para recopilar datos en tablas sobre la producción agrícola y los géneros vendidos o cambiados mediante trueque. Los egipcios ya analizaban los datos de la población y la renta del país mucho antes de construir la pirámides. Los libros bíblicos de números y crónicas incluyen, en algunas partes, trabajos de estadística. El primero contiene dos censos de la población de Israel y el segundo describe el bienestar material de las diversas tribus judías. En China existían los censos chinos ordenados por el emperador Tao hacia el año 2.200 a.C.

Posteriormente, hacia el año 500 a.C., se realizaron censos en Roma para conocer la población existente en aquel momento. Se erigió la figura del censor, cuya misión consistía en controlar el número de habitantes y su distribución por los distintos territorios.

En la Edad Media, en el año 762, Carlomagno ordenó la creación de un registro de todas sus propiedades, así como de los bienes de la iglesia. Después de la conquista normanda de Inglaterra en 1.066, el rey Guillermo I, el Conquistador, elaboró un catastro que puede considerarse el primero de Europa. Los Reyes Católicos ordenaron a Alonso de Quintanilla en 1.482 el recuento de fuegos (hogares) de las provincias de Castilla.

En 1.662 un mercader de lencería londinense, John Graunt, publicó un tratado con las observaciones políticas y naturales, donde Graunt pone de manifiesto las cifras brutas de nacimientos y defunciones ocurridas en Londres durante el periodo 1.604-1.661, así como las influencias que ejercían las causas naturales, sociales y políticas de dichos acontecimientos. Puede considerarse el primer trabajo estadístico serio sobre la población.

Curiosamente, Graunt no conocía los trabajos de B. Pascal (1.623-1.662) ni de C. Huygens (1.629-1.695) sobre estos mismos temas. Un poco más tarde, el astrónomo Edmund Halley (1.656- 1.742) presenta la primera tabla de mortalidad que se puede considerar como base de los estudios contemporáneos. En dicho trabajo se intenta establecer el precio de las anualidades a satisfacer a las compañías de seguros. Es decir, en Londres y en París se estaban construyendo, casi de manera simultánea, las dos disciplinas que actualmente llamamos estadística y probabilidad.

En el siglo XIX, la estadística entra en una nueva fase de su desarrollo con la generalización del método para estudiar fenómenos de las ciencias naturales y sociales. Galton (1.822-1.911) y Pearson (1.857-1936) se

pueden considerar como los padres de la estadística moderna, pues a ellos se debe el paso de la estadística deductiva a la estadística inductiva.

2.- ESTADÍSTICA. NOCIONES GENERALES.

Normalmente se entiende por Estadísticas los conjuntos de datos distribuidos en tablas, gráficos publicados en los diarios, etc. Sin embargo, la Estadística es también un método de toma de decisiones, por eso se utiliza actualmente en todas las ramas del saber. Por ejemplo:

- ¿Cómo decidir si un nuevo producto comercial tendrá éxito?
- ¿Qué podrá pronosticar un sociólogo a partir de una encuesta sobre voluntad de voto?
- ¿Cómo podrá un experto en geografía humana calcular la composición de la población del año 2.500?
- ¿Cuáles serán las necesidades de puestos escolares de Bachillerato para los próximos años?

¿Influye el índice de precios al consumo en la tasa de desempleo?

Podemos decir que la **Estadística es un método científico** muy utilizado en la actualidad, que tiene por objeto el desarrollo de técnicas para el conocimiento numérico de un conjunto de datos empíricos (recogidos mediante experimentos o encuestas) y sirve preferentemente para tomar decisiones.

En nuestros días la estadística se ha convertido en un método efectivo para describir valores de datos económicos, políticos, sociales, psicológicos, biológicos y físicos, y sirve como herramienta para relacionar y analizar dichos datos, así como establecer descripciones, predicciones, comparaciones y generalizaciones sobre una población a partir de los datos obtenidos de una muestra.

Medicina: uso de medicamentos en distintos países, análisis de la curva de crecimiento, cómo contrastar la eficacia comparada de dos medicamentos contra el SIDA.

Ciencias económicas y sociales: censos, tierras cultivadas de pastos, bosques, producción total de sustancias alimenticias, coeficiente de natalidad y mortalidad, paro obrero, nivel de vida, estudio de mercados, opiniones públicas.....

Control de calidad en la producción industrial: La fabricación en serie no permite el control de todas las piezas producidas; existen métodos estadísticos para mantener la calidad de producción dentro de unos límites de tolerancia admitidos.

Meteorología: cómo establecer las zonas de mayor índice de pluviometría.

Tecnología: cómo estimar la duración de vida de un componente hardware de un ordenador.

Literatura: la frecuencia de frases de cierta longitud, la frecuencia de verbos, etc., han permitido determinar la paternidad de ciertas obras literarias.

Gracias al uso de los ordenadores, la rápida computación de los datos recogidos en disciplinas como la Medicina, la Meteorología, la Sociología y otras ciencias permite la mejor aplicación de los métodos estadísticos; es posible realizar un sondeo con miles de cuestionarios.

3.- POBLACIÓN, INDIVIDUO Y MUESTRA

Se llama **población** al conjunto de todos los elementos cuyo conocimiento nos interesa y que serán sujetos de nuestro estudio. Todos los elementos, por tanto, cumplen una determinada característica. A cada elemento de la población se le llama **Individuo** (por su origen demográfico) o unidad estadística.

Se llama **Muestra** a cualquier subconjunto de la población, cuyo estudio sirve para inferir características de toda la población. Al número de elementos o individuos de una muestra se llama tamaño de la muestra.

Muestreo es seleccionar una muestra de la población.

Ejemplo: En un sondeo para conocer la intención de voto de los habitantes de una ciudad tendríamos:

–Población: todos los ciudadanos con derecho a voto.

- Individuo: Cada uno de los ciudadanos con derecho a voto.
- Muestra: Un conjunto de personas con derecho a voto, a las que se entrevistará.

Para realizar el estudio estadístico de una población, es muy importante que la **muestra** elegida sea **representativa**, es decir, que los individuos que la componen hayan sido bien elegidos.

4.– CLASIFICACIÓN DE LAS VARIABLES ESTADÍSTICAS:

Clasificación de las variables estadísticas.

- **Cualitativas:** Son aquellas que no toman valores numéricos. Ejemplos: La profesión, el estado civil, el color de los ojos...

En los caracteres cualitativos se pueden establecer diferencias, llamadas **modalidades**. Ej: las modalidades del carácter profesión serían: economista, ingeniero, sociólogo, empleado de banca,....

- **Cuantitativas:** Son aquellas que sí toman valores numéricos. Ejemplos: La talla de los individuos, peso, renta per cápita de cada comunidad autónoma.... Hay de dos tipos:

– **Cuantitativas Discretas:** Toman valores numéricos aislados. Ejemplos: número de empleados de una fábrica, nº de hijos de 20 familias, nº de cuentas corrientes abiertas al mes en una oficina de caja de ahorros,....

– **Cuantitativas Continuas:** Pueden tomar todos los valores posibles dentro de un intervalo. al. Ejemplos: Talla, peso, diámetros de ruedas de varios coches, medida del cráneo de 100 recién nacidos, etc.

5.– TIPOS DE ESTADÍSTICA (DOS RAMAS DE LA ESTADÍSTICA):

- **Estadística Descriptiva** (o deductiva):

Trata de describir y analizar algunos de los caracteres de un grupo dado, sin extraer conclusiones para un grupo mayor. Para este estudio se siguen los pasos siguientes:

- Selección de caracteres dignos de estudio.
- Recogida de datos.
- Clasificación y organización de los datos en tablas.
- Cálculo de ciertos valores numéricos (los parámetros estadísticos) a partir de los datos obtenidos.

B) Estadística Inferencial (o inductiva):

Establece previsiones y conclusiones generales sobre una población a partir de una muestra y se apoya fuertemente en el Cálculo de Probabilidades.

(NOTA: Inferir significa sacar consecuencia o deducir una cosa de otra).

Si un funcionario de la universidad organiza, tabula y representa gráficamente los datos correspondientes a todos los alumnos (por ejemplo, la distribución de sus edades) está haciendo estadística descriptiva.

Si un estadístico toma una muestra de 720 alumnos y estudia en ellos el tipo de lectura que prefieren con el fin de sacar consecuencias válidas para toda la universidad, estará realizando estadística inferencial.

• TABLAS DE FRECUENCIAS:

Las tablas de frecuencias sirven para ordenar y organizar los datos estadísticos. Con ellas, una masa amorfa de datos pasa a ser una colección ordenada. Por ejemplo:

Las calificaciones obtenidas por los 25 alumnos de una clase en una prueba escrita son las siguientes:

7 3 5 4 8 2 9 7 6 7 4 3 6 4 5 6 8 9 3 6 5 2 7 8 9

Estos datos los vamos a ordenar y clasificar para poder estudiarlos mejor.

Se llama **frecuencia absoluta** f_i de un valor x_i , al número de veces que se repite dicho valor.

Se llama **frecuencia absoluta acumulada** F_i del valor x_i , a la suma de las frecuencias absolutas de todos los valores anteriores a x_i más la frecuencia absoluta de x_i : $F = f_1 + f_2 +$

$+ f_3 + \dots + f_{n-1} + f_n$.

Se llama **frecuencia relativa** f_{ri} del valor x_i , al cociente entre la frecuencia absoluta de x_i y el número total de datos:

$$f_{ri} = \frac{f_i}{N}$$

$N = n^\circ$ total de datos.

Se llama **frecuencia relativa acumulada** F_{ri} del valor x_i , al cociente entre la frecuencia absoluta acumulada de x_i y el número total de datos.

Ejemplo 1: Completa la tabla de las calificaciones obtenidas por los 25 alumnos:

NOTA	F. ABSOLUTA	F. RELATIVA	F. ABS. ACUM.	F. REL. ACUM.	F. PORCEN- TUAL
2	2				
3	3				
4	3				
5	3				
6	4				
7	4				
8	3				
9	3				

$$N=25=\sum f_i$$

$$\sum \frac{f_i}{N}=1 \quad \sum \%=100$$

Nota: Los datos siempre se ordenan de menor a mayor.

Indica si son ciertas o falsas las siguientes afirmaciones:

- La frecuencia absoluta de 7 es 3.
- La nota que tiene una frecuencia absoluta mayor es 9.
- La frecuencia relativa de la nota 8 es 0,12.
- El 68% de los alumnos de clase han superado la prueba.

Σ es un símbolo matemático que se llama sumatorio y sirve para expresar abreviadamente una suma de muchos sumandos. (Es una letra griega que se llama sigma).

Observa que $\sum f_i = N$, es decir, la suma de todas las frecuencias absolutas de la tabla es igual al número de observaciones N .

También ocurre que $\sum f_i = 1$, es decir, la suma de todas las frecuencias relativas de la tabla es igual a la unidad. Esto no es casualidad, sino que ocurre en todas las tablas estadísticas, es por tanto una propiedad de las frecuencias.

Para calcular los % (también llamados frecuencias porcentuales) se multiplican las frecuencias relativas por 100.

Tablas con datos agrupados:

Cuando la variable estadística es continua, o bien discreta pero con un número muy grande de datos, es aconsejable agruparlos en intervalos llamados **clases**. A los puntos medios de cada intervalo se les llama **marcas de clase**.

Para agrupar los datos se realizan los siguientes pasos: (criterio de Norcliffe)

- Se localizan los valores extremos a y b y se halla su diferencia $r = b - a$ (r = recorrido).

2) Se halla $\sqrt{N}$

y este número va a ser el número de intervalos. ($N = n^\circ$ total de datos).

- 3) Se toma un intervalo r' , de longitud algo mayor que el recorrido r y que sea múltiplo del número de intervalos.
- 4) Se forman los intervalos de modo que el extremo inferior del primero sea algo menor que a y el extremo superior del último sea algo superior a b .

Hay que tener en cuenta que los intervalos deben ser cerrados por la izquierda y abiertos por la derecha y todos deben tener el mismo tamaño o amplitud.

Ejemplo 2: Tabla de frecuencias de una variable estadística continua:

Vamos a elaborar la tabla de frecuencias con la estatura de 40 chicos y chicas de Bachillerato

168 160 167 175 175 175 167 168 158 149 160 178 166 158 163 171 162 165 163 156 174 160 165 154 163
165 161 162 166 163 159 170 165 150 167 164 165 173 164 169 170

El número de valores distintos que hay es grande (mayor que 20). Por eso, lo adecuado es agruparlo en intervalos. Para ello procedemos así:

- Se localizan los valores extremos: el menor es 149 y el mayor 178 y se halla su diferencia $r = 178 - 149 = 29$ (r = recorrido).

2) Se halla $\sqrt{N}$

$$= \sqrt{40}$$

= 6'32 y este 6 va a ser el número de intervalos. (N = nº total de datos).

3) Se toma un intervalo $r' = 30$, de longitud algo mayor que el recorrido $r = 30$ y que sea múltiplo del número de intervalos ($6 \times 5 = 30$).

4) Se forman los intervalos de modo que el extremo inferior del primero sea algo menor que 149 y el extremo superior del último sea algo superior a 178.

Completa la tabla siguiente:

INTERVA-LOS	MARCAS DE CLASE	AMPLITUD	F. ABSOLUTA	F. RELATIVA
[148'5,153'5)			2	
[153'5,158'5)			4	
[158'5,163'5)			11	
[163'5,168'5)			14	
[168'5,173'5)			5	
[173'5,178'5)			4	

$$N = 40 = \sum F_i$$

I

$$h = 1$$

• GRÁFICOS ESTADÍSTICOS:

Pese a que las tablas estadísticas contienen toda la información, se suelen acompañar de gráficos para que esta información sea más clara y evidente. Muchas veces, incluso se prescinde de las tablas y se presenta directamente la información en uno o varios gráficos.

Los tipos de gráficos estadísticos que podemos encontrar son los siguientes:

1) DIAGRAMA DE BARRAS

Consiste en barras estrechas situadas sobre los puntos y separadas unas de otras. Se utiliza para variables cuantitativas **discretas**.

Si en el diagrama de barras de frecuencias absolutas se unen los extremos de las barras, se forma una línea quebrada que se llama Polígono de frecuencias absolutas. Si en lugar de representar las frecuencias absolutas,

representamos las frecuencias absolutas acumuladas, obtenemos el diagrama de barras de frecuencias absolutas acumuladas, y uniendo los extremos, el polígono de frecuencias absolutas acumuladas.

2) HISTOGRAMAS:

Se utiliza para distribuciones de variable estadística continua o para las discretas con gran número de datos que se han agrupado en intervalos.

Para construirlos, se representan sobre el eje de abscisas los límites de los intervalos, construyendo unos rectángulos que tienen por base la amplitud del intervalo, y por altura, la frecuencia absoluta de cada intervalo, siempre que todos los intervalos tengan la misma amplitud. Si las amplitudes no son iguales, las alturas de los rectángulos han de ser calculadas teniendo en cuenta que sus áreas deben ser proporcionales a las frecuencias de cada intervalo.

(es decir, lo que marca la frecuencia no es la altura de cada barra, sino el área de la misma).

El polígono de frecuencias correspondiente se halla uniendo los puntos medios de las bases superiores de cada dos rectángulos consecutivos.

3) DIAGRAMA DE SECTORES

Se utiliza para representar variables cualitativas. Consiste en dividir un círculo en sectores circulares de forma que el área de cada sector ha de ser proporcional a la frecuencia absoluta correspondiente.

4) PICTOGRAMAS

Son dibujos alusivos a la distribución que se pretende estudiar y que mediante su forma, tamaño, etc., ofrecen una descripción lo más expresiva posible de la distribución estadística. Se utilizan para representar caracteres cualitativos, pero es un método de poca precisión.

5) PIRÁMIDES DE POBLACIÓN

Se utilizan para estudiar conjuntamente la variable edad y el atributo sexo.

La gráfica se obtiene representando en la ordenada el grupo de edad y en la abscisa el sexo. Para la modalidad mujer se toma el semieje positivo, y para el hombre, el semieje negativo.

En las pirámides de población, es posible analizar la distribución de la población por edad y sexo y deducir las vicisitudes sufridas (guerras, catástrofes, etc), conocer el comportamiento demográfico y realizar previsiones sobre el futuro.

6) SERIES TEMPORALES O DIAGRAMAS LINEALES

Se utilizan para mostrar fluctuaciones de un determinado carácter estadístico con el paso del tiempo. En este diagrama sólo nos interesa la altura de la línea referida a la base del diagrama. Se pueden representar varios diagramas lineales a la vez.

• CARTOGRAMAS

Son útiles para mostrar los valores de la variable en las diferentes zonas de un mapa.

Denominamos cartogramas a las representaciones de una variable estadística en un mapa. Se consiguen

utilizando colores, distintas tonalidades de un mismo color, diferentes rayados, etc., para cada una de las zonas.

EJERCICIO: Realiza el diagrama de barras del Ejemplo 1 y el histograma del Ejemplo 2.

• **PARÁMETROS ESTADÍSTICOS**

Sirven para resumir en un número un aspecto importante de la distribución. Se suelen agrupar en dos categorías denominadas medidas de centralización y medidas de dispersión.

8a) MEDIDAS DE CENTRALIZACIÓN:

Son medidas o parámetros que tienden a situarse hacia el centro del conjunto de datos ordenados. Indican un valor central alrededor del cual oscilan y se concentran los valores de la variable. Son la media aritmética, la moda y la mediana.

• **la media aritmética :**

Se llama media aritmética de una variable estadística, a la suma de todos los valores de dicha variable dividido entre el número total de valores. La media aritmética, o simplemente media, se representa por $\bar{x}$.

Cálculo de $\bar{x}$
:

Si la variable estadística toma los valores x_1

, x_2

, x_3

, ... x_n

, con frecuencias absolutas f_1

, f_2

, f_3

, ... f_n

y el número total de observaciones es N , la media es:

$$\begin{aligned}\bar{x} &= \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + x_3 \cdot f_3 + \dots + x_n \cdot f_n}{N} \\ &= \frac{\sum_{i=1}^n x_i \cdot f_i}{N}\end{aligned}$$

Ejemplo 3:

Hallar la calificación media de Matemáticas de un alumno que tuvo las siguientes notas:

4, 7, 5, 8, 3, 9 y 6. (SOL: 6) *****

Ejemplo 4:

Hallar la calificación media de Matemáticas de los 40 alumnos de una clase, cuyos resultados fueron:

7 4 1 3 2 9 1 4 3 5 6 4 8 9 3 5 4 5 7 9 6 5 6 6 2 3 8 6 5 8 5 6 4 6 5 8 6 6 5 7

Empezaremos construyendo la tabla correspondiente:

$$\begin{aligned} \text{Luego } \bar{x} &= \frac{212}{40} \\ &= 5.3 \end{aligned}$$

Si los datos se encuentran agrupados en intervalos, se toman las marcas de clase como x_i

.

Ejemplo 5: Se ha aplicado un test de satisfacción en el trabajo a 88 empleados de una fábrica, obteniéndose los siguientes resultados:

Puntuaciones	[38, 44)	[44, 50)	[50, 56)	[56, 62)	[62, 68)	[68, 74)	[74, 80)
Nº Trabajadores	7	8	15	25	18	9	6

Calcular la puntuación media.

Para ello, empezamos construyendo la siguiente tabla:

Puntuaciones	MARCAS DE CLASE	F. ABSOLUTA	x_i f_i
[38, 44)	41	7	287
[44, 50)	47	8	376
[50, 56)	53	15	795
[56, 62)	59	25	1.475
[62, 68)	65	18	1.170
[68, 74)	71	9	639
[74, 80)	77	6	462

$$N=88=\sum f_i$$

I

$$\sum x_i f_i$$

$$= 5.204$$

$$= 5.204$$

$$\text{Luego } \bar{x} = \frac{5.204}{88}$$

$$= 59.14$$

$$= 59.14$$

- Ventajas:

El cálculo de $\bar{x}$

es sencillo y tiene en cuenta todos los datos de la distribución. Es la que más se utiliza.

- Inconvenientes:

Se distorsiona alterándose su significado cuando la distribución posee valores extremos (ej: si hallamos la talla media de los alumnos de una clase y hay un alumno que mide 2'10 m., este valor alterará notablemente la talla media de la clase).

No siempre se puede calcular, por ejemplo cuando el carácter es cualitativo o los datos están agrupados en intervalos y uno de ellos es abierto, por ejemplo, mayor de 70 años o menor de 6 años. En estos casos se utiliza como medida de centralización la moda o la mediana.

- **la moda :**

Se llama moda de una variable estadística al valor de dicha variable que presenta mayor valor absoluta. Se representa por M_o y no es única, es decir, una misma variable puede tener varias modas.

Ejemplo 6: En la tabla

x_i	1	2	3	4	5	6
f_i	6	7	14	10	14	9

Hay dos modas: $M_o = 3$ y $M_o' = 5$. Se dice que es bimodal.

Cálculo de la M_o si los datos se encuentran agrupados en intervalos:

En este caso, existe un intervalo modal que es el que tiene mayor frecuencia absoluta y para averiguar el valor de la M_o dentro de ese intervalo modal se utiliza la siguiente fórmula:

$$M_o = Li + C \cdot \frac{D_1}{D_1 + D_2}$$

Li es el límite inferior de la clase modal

C es la amplitud del intervalo modal

D_1

es la diferencia entre la frecuencia absoluta de la clase modal y la de la clase anterior

D_2

es la diferencia entre la frecuencia absoluta de la clase modal y la de la clase siguiente

Ejemplo 5: Calcular la moda de:

Puntuaciones	[38, 44)	[44, 50)	[50, 56)	[56, 62)	[62, 68)	[68, 74)	[74, 80)
Nº Trabajadores	7	8	15	25	18	9	6

- Intervalo modal= [56, 62), luego Li = 56
- C= 62-56 =6
- D1
= 25-15 =10
- D2
= 25-18= 7, entonces

$$Mo = 56 + 6 \cdot \frac{10}{10 + 7}$$

=59'5

Ejercicio: Calcula la moda del Ejemplo

INTERVALOS	F. ABSOLUTA
[148'5,153'5)	2
[153'5,158'5)	4
[158'5,163'5)	11
[163'5,168'5)	14
[168'5,173'5)	5
[173'5,178'5)	4

Cálculo gráfico de la Moda:

Cuando tenemos un histograma de frecuencias absolutas, en papel milimetrado para mayor exactitud, podemos calcular la moda gráficamente, uniendo con líneas de puntos los extremos de la clase modal con las contiguas y el punto de corte es la Mo.

Ejercicio: Dibuja el histograma del ejemplo anterior y calcúlala.

- Ventajas: La Mo representa el valor dominante de la distribución, así en unas elecciones, la Mo es el partido más votado. Es útil en caracteres cualitativos.
- Inconvenientes: Hay distribuciones que no tienen Mo, no intervienen todos los datos de la distribución y es menos representativa que la media. Además, no se sitúa siempre en la zona central de la distribución.

3) **La mediana** :

Se llama mediana de una variable estadística, a un valor de la variable, tal que el número de observaciones menores que él es igual al número de observaciones mayores que él. Es decir, si los datos se ordenan de menor a mayor, el valor del dato que ocupa el lugar intermedio se llama mediana. Así pues, la mediana deja tantos datos por debajo de ella como por encima. Se representa por Me.

Cálculo de la Mediana:

- Variable estadística discreta:

a1) Datos simples: Se ordenan éstos de menor a mayor y la Me será el término central.

- Si el número de datos es impar, la Me es única.

Ejemplo: Halla la mediana de los siguientes datos: 12, 3, 2, 9, 6, 11, 5.

Solución: Me =

- Si el número de datos es par, la Me es la semisuma de los dos valores centrales.

Ejemplo: Halla la mediana de los siguientes datos: 12, 3, 2, 9, 6, 11, 5, 13.

Solución: Me =

Cálculo de la Me si los datos se encuentran en tablas:

Se calculan las frecuencias acumuladas F_i y la Me viene dada por el primer valor de la variable cuya frecuencia absoluta acumulada exceda a la mitad del número de datos. Ejemplo:

Construimos la siguiente tabla:

NOTAS	F. ABSOLUTA	F.ABS. ACUM.
1	2	2
2	2	4
3	4	8
4	5	13 < 20
5	8	21 > 20, luego Me = 5
6	9	30
7	3	33
8	4	37
9	3	40

$$N=40=\sum f_i$$

$$\frac{N}{2}$$

$$= 20$$

En el caso de que la mitad del número de datos $\frac{N}{2}$

coincida con una frecuencia absoluta acumulada F_i

, la Me es la media entre ese valor de la variable y el siguiente de la tabla.

Ejemplo:

x_i	F. ABSOLUTA	F. ABS. ACUM.
3	15	15
6	20	35
7	15	50

8	40	90
9	10	100

$N = 100$

$$\frac{N}{2}$$

= 50, luego $Me = \frac{7+8}{2}$

= 7'5

Cálculo de la Me si los datos se encuentran agrupados en intervalos:

En este caso, buscamos el intervalo mediano de forma análoga a lo anterior, y para averiguar el valor de la Me dentro de ese intervalo mediano se utiliza la siguiente fórmula:

$$Me = Li + C \cdot \frac{\frac{N}{2} - F_{i-1}}{f_i}$$

Li es el límite inferior de la clase mediana

C es la amplitud del intervalo mediano

N es el número total de datos

F_{i-1}

es la frecuencia absoluta acumulada del intervalo anterior al intervalo mediano

f_i

es la frecuencia absoluta del intervalo mediano

Ejemplo: Calcular la mediana de:

Puntuaciones	[38, 44)	[44, 50)	[50, 56)	[56, 62)	[62, 68)	[68, 74)	[74, 80)
Nº Trabajadores	7	8	15	25	18	9	6

Para ello, empezamos construyendo la siguiente tabla:

Puntuaciones	MARCAS DE CLASE	F. ABSOLUTA	F. ABS. ACUM.
[38, 44)	41	7	7
[44, 50)	47	8	15
[50, 56)	53	15	30
[56, 62)	59	25	55
[62, 68)	65	18	73
[68, 74)	71	9	82

[74, 80)	77	6	88
----------	----	---	----

$$N=88=\sum f_i$$

$$\frac{\sum f_i}{2}$$

$$= 44$$

- Intervalo mediano = [56, 62), luego Li = 56
- C= 62-56 =6

$$- F_{i-1}$$

$$= 30$$

$$\bullet f_i$$

$$= 25, \text{ entonces } Me = 56 + 6 \cdot \frac{44 - 30}{25}$$

= 59'63, que, como es lógico, pertenece al intervalo mediano.

Ejercicio: Calcula la mediana del Ejemplo 3

INTERVALOS	F. ABSOLUTA
[148'5,153'5)	2
[153'5,158'5)	4
[158'5,163'5)	11
[163'5,168'5)	14
[168'5,173'5)	5
[173'5,178'5)	4

. Geométricamente y en un histograma de frecuencias, la Me es un valor de la variable tal que la vertical levantada sobre el mismo, divide al histograma en dos partes de igual área.

-Ventajas:

La Me es especialmente útil en los siguientes casos:

a) Cuando existe algún dato extremo que altere el significado de la $\bar{x}$

.

- Cuando los datos están agrupados en intervalos y alguno de ellos es abierto, pues en este caso no podríamos calcular la $\bar{x}$

.

- Inconvenientes:

- Al ser la Me un valor que depende del orden de los datos y no de su valor, es poco sensible a los valores que tenga la variable estadística.

4) Los Cuantiles:

Son parámetros que dividen a los datos de la distribución en función de determinadas cuantías. Los más importantes son:

4.1) Cuartiles:

Son tres valores que dividen a la serie de datos en cuatro partes iguales. Se representan por Q_1

= primer cuartil; Q_2

= segundo cuartil y Q_3

= tercer cuartil.

Q_1

deja por debajo de él al 25% de los datos y por encima al 75%

Q_2

deja por debajo de él al 50% de los datos y por encima al 50%, por tanto Q_2

= Me.

Q_3

deja por debajo de él al 75% de los datos y por encima al 25%

4.2) Deciles:

Son nueve valores que dividen a la serie de datos en 10 partes iguales. Se representan por D_1

, D_2

, D_3

, D_4

, D_5

, D_6

, D_7

, D_8

y D_9

.

4.3) Percentiles o Centiles:

Son 99 valores que dividen a la serie de datos en 100 partes iguales. Se representan por P_1

, P_2

, P_3

, P_4

..... P_{98}

y P_{99}

.

P_1

deja por debajo de él al 1% de los datos y por encima al 99%

P₂

deja por debajo de él al 2% de los datos y por encima al 98%

Etc.

Observa que según las definiciones se cumple que

$$\begin{aligned} \text{Me} &= Q_2 \\ &= D_5 \\ &= P_{50} \end{aligned}$$

Para variables estadísticas discretas se utiliza, para hallar P_s

la fórmula $\frac{N s}{100}$

.

Para tablas con datos agrupados en intervalos, se utiliza, para hallar P_s

la fórmula:

$$\begin{aligned} P_s &= Li + C \cdot \frac{\frac{N s}{100} - F_{i-1}}{f_i} \end{aligned}$$

Los cuantiles son muy utilizados en las Ciencias Sociales y nos proporcionan información sobre la estructura interna de los datos.

Cálculo de los Cuantiles:

Su cálculo es análogo al de la Me.

a) Variable estadística discreta:

Ejemplo: Las calificaciones de Matemáticas de 40 alumnos son:

Calcular

- Los cuantiles 1º y 3º.
- Los Percentiles 30 y 70.

Construimos la siguiente tabla:

NOTAS	F. ABSOLUTA	F.ABS. ACUM.
1	2	2
2	2	4
3	4	8

4	5	13
5	8	21
6	9	30
7	3	33
8	4	37
9	3	40

$$N=40=\sum f_i$$

$$\frac{N}{2}$$

$$= 20$$

– Q1

: Como Q1

$$= P_{25}$$

, calculamos $\frac{N \cdot s}{100}$

$$= \frac{40 \cdot 25}{100}$$

= 10, luego Q1

= 4, porque la primera frecuencia absoluta acumulada F_i

que excede a 10 es 13, y corresponde a x_i

= 4.

• Q3

Como Q3

$$= P_{75}$$

, calculamos $\frac{N \cdot s}{100}$

$$= \frac{40 \cdot 75}{100}$$

= 30, luego, como coincide con una frecuencia absoluta acumulada F_i

, hallamos la media entre el valor x_i

correspondiente a ella y el siguiente. Por tanto, Q3

$$= \frac{6 + 7}{2}$$

= 6'5.

• P30

$$= D_3$$

: calculamos $\frac{N \cdot s}{100}$

$$= \frac{40 \cdot 30}{100}$$

= 12, luego P_{30}

= 4, porque la primera frecuencia absoluta acumulada F_i que excede a 12 es 13, y corresponde a x_i

= 4.

• P_{70}

= D_7

: calculamos $\frac{N \cdot s}{100}$

$$= \frac{40 \cdot 70}{100}$$

= 28, luego P_{70}

= 6, porque la primera frecuencia absoluta acumulada F_i que excede a 28 es 30, y corresponde a x_i

= 6.

b) Variable estadística con los datos agrupados en intervalos:

Ejemplo: Puntuaciones de test de satisfacción de trabajo a 88 empleados:

Puntuaciones	[38, 44)	[44, 50)	[50, 56)	[56, 62)	[62, 68)	[68, 74)	[74, 80)
Nº Trabajadores	7	8	15	25	18	9	6

Calcular

- Los cuartiles 1º y 3º.

b) Los Percentiles 40 y 90. Hállalos tú, sabiendo que los resultados son: P_{40}

= 57'25 y P_{90}

= 72'13

Para ello, empezamos construyendo la siguiente tabla:

Puntuaciones	MARCAS DE CLASE	F. ABSOLUTA	F. ABS. ACUM.
[38, 44)	41	7	7
[44, 50)	47	8	15
[50, 56)	53	15	30
[56, 62)	59	25	55
[62, 68)	65	18	73
[68, 74)	71	9	82
[74, 80)	77	6	88

$N=88=\sum f_i$

$$\frac{I}{N}$$

$$= 44$$

Q1

: Como Q1

$$= P_{25}$$

, calculamos $\frac{N s}{100}$

$$= \frac{88 \cdot 25}{100}$$

$$= 22, \text{ luego } Q1$$

pertenece al intervalo [50, 56), porque la primera frecuencia absoluta acumulada F_i que excede a 22 es 30, y corresponde a ese intervalo.

P_s

$$= Li + C \cdot \frac{\frac{N s}{100} - F_{i-1}}{f_i}$$

; P₂₅

$$= 50 + 6 \cdot \frac{22 - 15}{15}$$

$$= 52.8 = Q1$$

Q3

: Como Q3

$$= P_{75}$$

, calculamos $\frac{N s}{100}$

$$= \frac{88 \cdot 75}{100}$$

$$= 66, \text{ luego } Q3$$

pertenece al intervalo [62, 68), porque la primera frecuencia absoluta acumulada F_i que excede a 66 es 73, y corresponde a ese intervalo.

P_s

$$= Li + C \cdot \frac{\frac{N s}{100} - F_{i-1}}{f_i}$$

; P₇₅

$$= 62 + 6 \cdot \frac{66 - 55}{18}$$

$$= 65'67 = Q_3$$

8b) MEDIDAS DE DISPERSIÓN

Son parámetros que indican la mayor o menor concentración de los valores o datos en torno a las medidas de centralización. Los más importantes son: Recorrido o rango, Varianza y Desviación Típica.

Ejemplo: En una oficina A:

Sueldo del empleado A₁
es 200.000 pesetas.

Sueldo del empleado A₂
es 80.000 pesetas.

Sueldo del empleado A₃
es 32.000 pesetas.

Luego , sueldo medio $\bar{x}_A$
= 104.000 pesetas

En otra oficina B:

Sueldo del empleado B₁
es 152.000 pesetas.

Sueldo del empleado B₂
es 100.000 pesetas.

Sueldo del empleado B₃
es 60.000 pesetas.

Luego , sueldo medio $\bar{x}_B$
= 104.000 pesetas

Los sueldos medios de las oficinas A y B son iguales, pero es evidente que en la oficina B el reparto es más igualitario que en la oficina A. Es decir, en B los sueldos están más próximos unos de otros, o lo que es lo mismo, la dispersión es menor. (Ya veremos que la desviación típica de la oficina A es 70.654'08 y la de la oficina B es 37.665'19172).

• Recorrido o Rango:

Es la diferencia entre el mayor y el menor valor de la variable estadística.

En el ejemplo anterior:

$$RA = 200.000 - 32.000 = 168.000 \text{ pesetas}$$

R_B

= 152.000 – 60.000 = 92.000 pesetas. Luego R_A

> R_B

Observaciones:

- Cuanto menor es el recorrido de una distribución, mayor es el grado de representatividad de las medidas de centralización.
- Tiene la ventaja de que su cálculo es sencillo.
- Tiene el inconveniente de que sólo depende de los valores extremos, con lo cual si uno de ellos se separa mucho, el recorrido se ve sensiblemente afectado. Para paliar este inconveniente, se utilizan:

- El Rango Intercuartílico: $Q = Q_3 - Q_1$
- El Rango entre Percentiles: $P = P_{90} - P_{10}$

• **Varianza y Desviación Típica:**

- Se llama **varianza** de una variable estadística a la media aritmética de los cuadrados de las desviaciones respecto a la media. Se representa por s^2 o S^2 .
- Se llama **desviación típica** de una variable estadística a la raíz cuadrada positiva de la varianza. Se representa por s o S .
- Cálculo de la varianza:

$$\begin{aligned}\text{Varianza} &= s^2 \\ &= S^2 \\ &= \frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_n(x_n - \bar{x})^2}{N} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{N}\end{aligned}$$

Siendo:

N = número total de datos

f_i

= frecuencias absolutas

x_i

= datos

$\bar{x}$

= media aritmética de la distribución.

Como este cálculo resulta muy laborioso, se suele utilizar otra expresión equivalente pero más fácil de calcular:

$$\begin{aligned} \text{Varianza} &= \frac{1}{N} \sum_{i=1}^n f_i x_i^2 - \bar{x}^2 \\ &= S^2 \end{aligned}$$

Con lo que:

$$\begin{aligned} \text{Desviación típica} &= S = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i} \\ &= \sqrt{\frac{1}{N} \sum_{i=1}^n f_i x_i^2 - \bar{x}^2} \end{aligned}$$

Cálculo de la Varianza y Desviación Típica:

a) Variable estadística discreta:

Ejemplo: Las calificaciones de Matemáticas de 40 alumnos son:

Calcular el rango, la varianza y la desviación típica.

Construimos la siguiente tabla:

NOTAS	F. ABSOLUTA	$f_i \cdot x_i$	$f_i \cdot x_i^2$
1	2	2	2
2	2	4	8
3	4	8	24
4	5	13	52
5	8	21	105
6	9	30	180
7	3	33	231
8	4	37	296
9	3	40	360

$$N=40=\sum f_i$$

I

$$188 \cdot 1.258$$

$$\begin{aligned} \bar{x} &= \frac{18}{40} \\ &= 4.7 \end{aligned}$$

$$\text{Rango} = 9 - 1 = 8$$

$$\begin{aligned} \text{Varianza} &= \frac{2}{40} \\ &= 1.258 \\ &- 4.7^2 \\ &= 31.45 - 22.09 = 9.36 \end{aligned}$$

$$\begin{aligned} \text{Desviación típica} &= \sqrt{9.36} \\ &= 3.059. \end{aligned}$$

b) Variable estadística con los datos agrupados en intervalos:

Ejemplo: Puntuaciones de test de satisfacción de trabajo a 88 empleados:

Puntuaciones	[38, 44)	[44, 50)	[50, 56)	[56, 62)	[62, 68)	[68, 74)	[74, 80)
Nº Trabajadores	7	8	15	25	18	9	6

Calcular el rango, la varianza y la desviación típica.

Construimos la siguiente tabla:

Puntuaciones	MARCAS DE CLASE	F. ABSOLUTA	$f_i \cdot x_i$	$\frac{f_i \cdot x_i^2}{2}$
[38, 44)	41	7	287	11.767
[44, 50)	47	8	376	17.672
[50, 56)	53	15	795	42.135
[56, 62)	59	25	1.475	87.025
[62, 68)	65	18	1.170	76.050
[68, 74)	71	9	639	45.369
[74, 80)	77	6	462	35.574

$$N=88=\sum f_i$$

I

$$5.204 \cdot 315.592$$

$$\begin{aligned}\bar{x} &= \frac{5.204}{88} \\ &= 59'14\end{aligned}$$

$$\text{Rango} = 80 - 38 = 42$$

$$\begin{aligned}\text{Varianza} &= \frac{2}{88} \\ &= \frac{315.592}{88} \\ &= 59'14^2 \\ &= 88'73\end{aligned}$$

$$\begin{aligned}\text{Desviación típica} &= \sqrt{88'73} \\ &= 9'4.\end{aligned}$$

Observaciones:

- 2

y dependen de todos los valores de la distribución y de la media.

- Cuando no se puede calcular $\bar{x}$ tampoco es posible calcular $\frac{2}{n_i}$.

- 2

tiene el inconveniente de que no viene expresada en las mismas unidades que los datos, pues se obtiene de valores elevados al cuadrado (si los datos vienen en cm., la varianza son cm²); pero sí se expresa en las mismas unidades, y por eso se utiliza más que 2.

-

- **Desviación Media:**

Se llama desviaciones respecto a la media a las diferencias entre cada valor de la variable y la media de la distribución: $x_i - \bar{x}$

$$- \bar{x}$$

$$, x_2 - \bar{x}$$

$$- \bar{x}$$

$$, \dots, x_n - \bar{x}$$

$$- \bar{x}$$

$$, \text{ donde } \bar{x}$$

es la media aritmética de la distribución, x_i

= datos o valores de la variable estadística.

Se llama **desviación media** a la media aritmética de los valores absolutos de las desviaciones de cada dato respecto a la media:

$$\text{D.M.} = \frac{\sum_{i=1}^n |x_i - \bar{x}| \cdot f_i}{N}$$

En el caso continuo se toman como x_i las marcas de clase.

Ejemplo:

El número de horas que dedica un alumno de 1º de Bachillerato al estudio, durante la semana, es el siguiente: 3'5, 5, 4, 6, 5'5, 3. Calcula la desviación media. ($\bar{x} = 4'5$)

x_i	$x_i - \bar{x}$	$x_i - \bar{x}$ $\emptyset$
3	$3 - 4'5 = -1'5$	1'5
3'5	$3'5 - 4'5 = -1$	1
4	$4 - 4'5 = -0'5$	0'5
5	$5 - 4'5 = 0'5$	0'5
5'5	$5'5 - 4'5 = 1$	1
6	$6 - 4'5 = 1'5$	1'5

" $x_i - \bar{x}$
 $\emptyset = 6$

$$\text{D.M.} = \frac{\sum_{i=1}^n |x_i - \bar{x}| \cdot f_i}{N}$$

$$= \frac{6}{6}$$

$$= 1$$

(Nota: En este ejemplo todas las f_i valen 1 y $N = 6$).

D.M. = 1 significa que como promedio, los valores se apartan una unidad de la $\bar{x}$.

• Coeficiente de variación de Pearson:

Se llama coeficiente de variación de Pearson al cociente entre la desviación típica y la media.

$$\text{CV} = \frac{s}{\bar{x}}$$

- En los casos en que la media sea muy próxima a cero, no debe usarse, ya que el denominador es muy pequeño y puede dar un grado erróneo de dispersión, dependiendo del valor de la desviación típica.

- Cuanto menor es el coeficiente de variación, menor dispersión tiene la distribución; por tanto la media será más representativa.

Ejemplo: Calcular el coeficiente de variación del ejemplo relativo a las puntuaciones del test de satisfacción de trabajo a 88 empleados.

Solución: CV =

9) COMPARACIÓN DE PUNTUACIONES OBTENIDAS EN DISTINTAS DISTRIBUCIONES. PUNTUACIONES TÍPICAS:

Si tenemos una variable estadística X que toma los valores x_1

, x_2

, ..., x_n

, llamados puntuaciones directas, y $\bar{x}$

y s son su media y su desviación típica. Se llaman puntuaciones típicas a los valores:

$$\frac{x_1 - \bar{x}}{s}, \frac{x_2 - \bar{x}}{s}, \dots, \frac{x_n - \bar{x}}{s}$$

Ejemplo:

Un alumno ha contestado a dos test, obteniendo las siguientes puntuaciones:

Test A: 50 puntos

Test B: 32 puntos

El profesor ha calculado para el test A: $\bar{x}_A$

$\bar{x}_A$

= 45 y s_A

= 6, y para el test B: $\bar{x}_B$

$\bar{x}_B$

= 26 y s_B

= 2. ¿En cuál de los dos test ha obtenido mejor resultado?.

Test A: $\frac{50 - 45}{6}$

= 0'83 = P.T.

Test B: $\frac{32 - 26}{2}$

= 3 = P.T. Luego como P.T. $\bar{x}_B$

> P.T. $\bar{x}_A$

, obtuvo mejor puntuación en el test B pues la puntuación típica es mayor y por tanto sobresale más de la media de la clase.

• **UTILIZACIÓN CONJUNTA DE $\bar{x}$**
y : TEOREMA DE Tchebyscheff:

La media aritmética de un conjunto de datos, se encuentra aproximadamente hacia el centro de la distribución, y la desviación típica nos informa sobre la dispersión que tienen los datos respecto a la media.

Utilizando ambos parámetros conjuntamente, podemos obtener resultados como éstos:

En distribuciones unimodales, simétricas o ligeramente asimétricas, se verifica que:

1º) En el intervalo $(\bar{x} - , \bar{x} +)$ se encuentra el 68% de los datos.

2º) En el intervalo $(\bar{x} - 2 , \bar{x} + 2)$ se encuentra el 95% de los datos.

3º) En el intervalo $(\bar{x} - 3 , \bar{x} + 3)$ se encuentra el 99% de los datos.

Estas afirmaciones se basan en el Teorema o desigualdad de un matemático ruso del siglo XIX llamado Tchebyscheff.

I.E.S. Guanarteme Curso 2.001– 2.002

Matemáticas aplicadas a las Ciencias Sociales I 1º de Bachillerato

Tema 1: Estadística descriptiva: distribuciones unidimensionales 19