

Estadística Descriptiva

La estadística es la parte de las matemáticas que se dedica a buscar regularidades entre las distintas características de personas, objetos o entes en general. Esto quiere decir, que su principal objetivo es el estudio de grandes conjuntos de personas, objetos o entes intentando encontrar similitudes o relaciones.

El estudio de la estadística se inicia por la estadística descriptiva, la misma se utiliza para describir las regularidades detectadas en los estudios realizados. Estos apuntes tienen la intención de introducirnos en el mundo de la estadística a través de la estadística descriptiva; o sea, la parte de la estadística que describe esas irregularidades.

CONCEPTOS BÁSICOS

Población: Llamamos población o universo al conjunto de personas, objetos o entes en general sobre el que se realiza el estudio estadístico.

Individuo: Llamamos individuo o unidad estadística a cada uno de los componentes de la población, y al número total de componentes de la misma se le llama tamaño de la población.

Datos: Los datos de un estudio estadístico lo constituyen el conjunto de informaciones que es preciso obtener para llevarlo a cabo. Es decir, cada una de las mediciones que se hace sobre los individuos.

Muestra: Llamamos muestra a la parte de la población de la que se obtienen los datos. La muestra tiene que ser representativa de la población, es decir, tiene que tener características similares a la población en estudio.

* Veamos con más profundidad el concepto de muestra representativa. A priori, cualquier subconjunto de la población es una muestra, pero lo que nos interesa es que sea representativa. Si lo que nos interesa es saber que edad tiene la mayoría de la población de Tenerife, no podremos coger para la muestra sólo personas del municipio de Arafo, tampoco podremos coger sólo a escolares o bien, sólo mujeres... sin embargo nos va dar igual que las personas escogidas sean rubias, o altas, u hombres,... ya que esos datos no interfieren en la edad.

Vamos a continuación a ver estos conceptos a través de un ejemplo.

Consideramos el siguiente supuesto: "Un Ayuntamiento quiere ver cuál es la estatura media de la población de entre 20 y 30 años de su municipio". El proceso seguido por el Ayuntamiento es el siguiente:

1º Comprueba en el censo del municipio que hay 4000 personas de esa edad. Tendremos que la **población** la forman 4000 **individuos**. Y por ello el **tamaño** de la población es de 4000.

2º Selecciona al azar una **muestra** de 100 individuos, ya que es excesivamente costoso medir a los 4000 individuos. Por ser seleccionada al azar se supone que es representativa. Existen otras formas de selección y es importante tenerlo en cuenta.

3º Efectúa las mediciones a los 100 individuos, con lo que se obtienen los **datos** necesarios para el estudio.

Variable estadística: A la cualidad o cantidad que es objeto de estudio en todos los individuos de la población se le llama carácter o variable. Las variables estadísticas pueden ser:

a) **Cualitativas**, son los valores de las observaciones quedan expresados por características o atributos: sexo, color de los ojos, color del pelo, estado civil,...

b) **Cuantitativas**, si los valores de las observaciones son cuantificables expresándose mediante números. Estas variables a su vez pueden aparecer de dos formas:

1. - **Discretas:** Cuando toman valores concretos, como por ejemplo número de hermanos, la edad,...

2. - **Continuas:** Cuando puede tomar cualquier valor de un cierto intervalo, como por ejemplo peso, estatura...

Estas variables se agrupan por **intervalos**, lo que se suele hacer es considerar la franja de posibles datos y dividirla en partes iguales según convenga. Cuando la variable es continua y medida en intervalos se suele considerar, como representante del intervalo, la **marca de clase** que coincide con el dato que se encuentra en medio del intervalo.

En el ejemplo anterior, la variable estadística en estudio es la estatura, por lo tanto será una variable estadística cuantitativa continua y la franja de posibles datos se sitúa entre los 150 cm y los 190 cm, por ello podremos dividirla en intervalos de amplitud 10 cm.

Frecuencia: la frecuencia es en general el número de veces que se repite un mismo dato. La frecuencia puede ser:

a) **Absoluta:** cuando se cuenta exactamente el número de veces que aparece un dato repetido.

b) **Relativa:** cuando se divide la frecuencia relativa entre el número de datos totales.

c) **En porcentajes:** cuando la frecuencia relativa se expresa en tantos por ciento, es decir, se multiplica la frecuencia relativa por cien.

Así, una tabla estadística es aquella tabla donde se representan todos los conceptos estudiados hasta el momento, o también, la tabla que se elabora con aquellos datos que necesitamos para hallar los parámetros que estudiaremos posteriormente.

Veamos algunos ejemplos:

* Se han anotado las tallas, en centímetros, de los 40 alumnos de una clase y se han obtenido los siguientes resultados: 160, 167, 163, 148, 151, 158, 166, 166, 157, 153, 151, 151, 150, 155, 164, 162, 166, 171, 167, 165, 152, 150, 147, 152, 162, 155, 158, 158, 158, 164, 157, 155, 160, 154, 153, 156, 160, 159, 159, 158, 163, 161.

Variable	Marca de clase	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia en Porcentaje	
[145,150)	147.5	2	0.05	5%	
[150,155)	152.5	9	0.225	22.50%	
[155,160)	157.5	12	0.3	30%	
[160,165)	162.5	10	0.25	25%	
[165,170)	167.5	6	0.15	15%	
[170,175]	172.5	1	0.025	2.50%	
		40	1	100%	TOTAL

Este es un ejemplo de tabla estadística de una variable continua.

* Se han lanzado 20 veces un dado con las caras numeradas del 1 al 6 y se han obtenido los siguientes resultados: 5, 4, 3, 6, 2, 1, 3, 4, 5, 6, 1, 2, 4, 3, 1, 2, 2, 5, 4, 6

Variable	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia en Porcentaje	
1	3	0.15	15%	
2	4	0.20	20%	
3	3	0.15	15%	
4	4	0.20	20%	
5	3	0.15	15%	
6	3	0.15	15%	
	20	1.00	100%	TOTAL

Este es un ejemplo de tabla estadística de una variable continua.

MEDIDAS DE CENTRALIZACIÓN

Las medidas de centralización son unos parámetros que representan la muestra. Son útiles porque hacen una estimación que indica alrededor de que valore se aglutinan los datos. Es decir, son los datos de la variable que se encuentran más cerca del resto de los datos.

Las principales medidas centrales son: la media aritmética (\bar{x}), la moda (M) y la mediana (m). Veamos en que consiste cada una de ellas:

Media: Se consigue sumando todos los valores y dividiéndolo entre el número de datos.

$$\bar{x} = \frac{\sum f_i \cdot x_i}{N} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + f_3 \cdot x_3 + \dots}{N}$$

Moda: La moda es el valor de la variable que más se repite.

Mediana: La mediana es el valor de la variable que deja por debajo a la mitad de los datos y por encima a la otra mitad. La forma de calcularla es ordenar todos los datos, contar el número de datos y localizar aquel dato que deja por debajo a la mitad de los datos.

Veamos algunos ejemplos según la variable sea continua o discreta. Para ello utilizaremos los ejemplos anteriores.

* Se han anotado las tallas, en centímetros, de los 40 alumnos de una clase y se han obtenido los siguientes resultados:

160, 167, 163, 148, 151, 158, 166, 166, 157, 153, 151, 151, 150, 155, 164, 162, 166, 171, 167, 165, 152, 150, 147, 152, 162, 155, 158, 158, 158, 164, 157, 155, 160, 154, 153, 156, 160, 159, 159, 158, 163, 161.

Variable	Marca de clase	Frecuencia Absoluta	$f_i \cdot x_i$	
[145,150)	147.5	2	295.00	
[150,155)	152.5	9	1372.50	
[155,160)	157.5	12	1890.00	
[160,165)	162.5	10	1625.00	
[165,170)	167.5	6	1005.00	
[170,175]	172.5	1	172.50	
		40	6360.00	TOTAL

Tendremos que:

$$\bar{x} = \frac{6330}{40} = 155.75$$

- a) $\bar{x} = 155.75$
b) M = 157.5
c) m = 157.5

Este es un ejemplo en una variable continua.

* Se han lanzado 20 veces un dado con las caras numeradas del 1 al 6 y se han obtenido los siguientes resultados:

5, 4, 3, 6, 2, 1, 3, 4, 5, 6, 1, 2, 4, 3, 1, 2, 2, 5, 4, 6

Variable	Frecuencia Absoluta	$f_i \cdot x_i$	
1	3	3.00	
2	4	8.00	
3	3	9.00	
4	4	16.00	
5	3	15.00	
6	3	18.00	
	20	69.00	TOTAL

Tendremos que:

$$\bar{x} = \frac{69}{20} = 3.405$$

- a) $\bar{x} = 3.405$
b) M = puede ser 2 ó 4 ya que ambos números se repiten el mismo número de veces. En estos casos se elige el que más convenga, o sea, el que más se acerque a las otras medidas de centralización. Hasta ahora nos da lo mismo que sea 2 que 4, veamos lo que sucede con la mediana.
c) m = también podría elegir entre 3 y 4, resulta más conveniente elegir 3 por el resultado que hemos obtenido en la media. Como vemos la mediana no arroja ninguna luz sobre la moda.

Este es un ejemplo de una variable discreta.

MEDIDAS DE DISPERSIÓN

Las medidas de dispersión son las que nos indican si las medidas de centralización representan o no a la muestra, es decir, si los valores de la variable se agrupan en torno a las medidas de centralización o no.

Las principales medidas o parámetros de dispersión son Varianza (V), Desviación Media (DM), Desviación Típica (σ) y Recorrido o Rango (Re). Veamos en que consiste cada uno de ellos:

$$V = \frac{\sum f_i \cdot (x_i - \bar{x})^2}{N} = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2$$

Varianza: La varianza nos indica si la separación entre los valores de la variable y la media es grande o pequeña. Es decir, si los datos se aglutinan o reúnen en torno a la media o no. Lo mismo nos indica tanto la Desviación Media como la Desviación Típica.

$$DM = \frac{\sum f_i \cdot |x_i - \bar{x}|}{N}$$

Desviación Media:

Desviación Típica: $\sigma = \sqrt{V}$

Recorrido: El recorrido se define como la diferencia entre el valor más grande de la variable y el más pequeño. Lo que nos indica el recorrido es lo distante que están los datos entre sí.

Veamos cómo se calculan los parámetros en los ejemplos que estamos utilizando:

* Se han anotado las tallas, en centímetros, de los 40 alumnos de una clase y se han obtenido los siguientes resultados:

160, 167, 163, 148, 151, 158, 166, 166, 157, 153, 151, 151, 150, 155, 164, 162, 166, 171, 167, 165, 152, 150, 147, 152, 162, 155, 158, 158, 158, 164, 157, 155, 160, 154, 153, 156, 160, 159, 159, 158, 163, 161.

Frecuencia Absoluta	$f_i \cdot x_i$	$f_i \cdot x_i^2$	$ x_i - \bar{x} $	$f_i \cdot x_i - \bar{x} $
2	295.00	43512.50	8.25	16.5
9	1372.50	209306.25	3.25	29.25
12	1890.00	297675.00	1.75	21
10	1625.00	264062.50	6.75	67.5
6	1005.00	168337.50	1.2	7.2
1	172.50	29756.25	16.75	16.75
40	6360.00	1012650.00	48.75	223

a) Calculamos la Varianza. Sabiendo que:

- $\bar{x} = 155.75 \Rightarrow \bar{x}^2 = 24258.0625$
- $\sum f_i \cdot x_i^2 = 1012650$
- $N = 40$

$$V = \frac{1012650}{40} - 24258.0625 = 1058.1875$$

Tendremos que:

b) Calculamos la Desviación Media. Sabiendo que:

$$\frac{\sum f_i \cdot |x_i - \bar{x}|}{N}$$

* Se han lanzado 20 veces un dado con las caras numeradas del 1 al 6 y se han obtenido los siguientes resultados:

5, 4, 3, 6, 2, 1, 3, 4, 5, 6, 1, 2, 4, 3, 1, 2, 2, 5, 4, 6

Variable	Frecuencia Absoluta	$f_i \cdot x_i$
1	3	3,00
2	4	8,00
3	3	9,00
4	4	16,00
5	3	15,00
6	3	18,00
	20	69,00

Coefficiente de variación : es el cociente entre la desviación típica y la media aritmética . Valores muy bajos indican

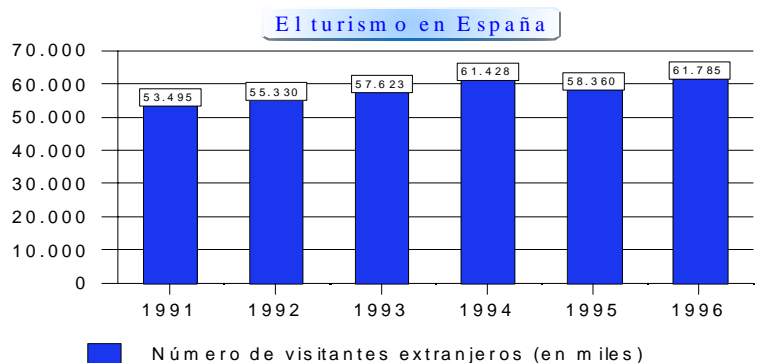
muestras muy concentradas . $C.V. = \frac{\sigma}{\bar{x}}$

PROBLEMAS:

ANEXO I

1.- Observa, calcula y contesta:

- Cuántos extranjeros visitaron España en 1992?) Y en 1993?
- En qué año visitaron España menos visitantes extranjeros?
- En qué año tuvo España más visitantes extranjeros?
- En qué año aumentó más el número de visitantes extranjeros con respecto al año anterior?
- Cuántos extranjeros visitaron España en los tres últimos años?
- Cuál fue la media de turistas por año en los años 1991, 1992 y 1993?) Y en los tres años siguientes?
- Cuál es la media de visitantes extranjeros en los seis años?) Es la media de las dos anteriores?



■ Número de visitantes extranjeros (en miles)

h) La media de turistas por mes durante el año 1995 fue menor que la media de turistas por mes durante el año 1996. ¿Puedes asegurar con este dato que cada mes del año 1995 tuvo menos turistas que el correspondiente mes del año 1996?

2.- Indica las variables que son cuantitativas y las que son cualitativas: talla, color del pelo, número de hermanos, estado civil, lugar de nacimiento, deporte preferido, marca de automóvil, calificación en Matemáticas.

3.- Indica las variables cuantitativas que son discretas y las que son continuas: número de hijos, ingresos diarios de una cafetería, edades de los vecinos de tu edificio, número de ascensores de un edificio, número de calzado de tus compañeros de clase.

4.- Clasifica las siguientes variables en cuantitativas y cualitativas, y da posibilidades de los valores que pueden tomar:

- Temperatura registrada cada hora del día.
- En un centro, número de alumnos de cada grupo de clase.
- En un centro, grupo musical preferido por los alumnos.
- En un centro, dinero del que dispone semanalmente cada alumno.
- En un centro, lugar preferido para ir en viaje de fin de curso.
- Opinión de los ciudadanos sobre una decisión política del alcalde.
- Duración de las llamadas telefónicas hechas desde una cabina.

5.- ¿Cuál es la población estudiada en cada uno de los casos del problema anterior?

6.- En las siguientes formas de escoger una muestra, se cometen errores que pueden motivar que los resultados obtenidos den una visión errónea de lo que se quiere observar. Di qué errores se cometen, cómo podrían distorsionar los resultados y qué cambios podríamos hacer en el mecanismo de recogida de datos para mejorarlo:

- Para controlar la calidad de las piezas producidas en una nave de fabricación de bombillas, se cogen 100 bombillas hechas por una de las 7 máquinas de la nave.
- Para recoger opiniones de tipo político, un periódico hace una encuesta entre sus suscriptores.
- Para estudiar la opinión política de los barceloneses, se entrevista durante un día a los usuarios del METRO en diferentes estaciones.
- Para hacer una encuesta sobre los hábitos de lectura de los habitantes de una población, un grupo de 30 alumnos sorteó previamente 30 calles y 5 viviendas en cada calle; y, una mañana, en horas de clase, pasan la encuesta a todos los mayores de 12 años que encuentran en las viviendas seleccionadas.

7.- Halla la media de los siguientes datos: 5, 6, 4, 5, 5, 7, 3.

8.- ¿Cuál es el valor de x para que la media del conjunto de datos 4, 5, x sea 5?

9.- Las calificaciones obtenidas por 7 amigos en Lengua han sido: Suficiente, Sobresaliente, Insuficiente, Notable, Bien, Insuficiente y Notable. ¿Qué calificación los representa?

10.- Las calificaciones obtenidas en inglés por 8 alumnos han sido:

10, 6, 7, 4, 5, 8, 2, 8. ¿Es representativa la media aritmética? ¿Y la mediana? Explícalo.

10.- Las edades de los componentes de un compañía de teatro juvenil son:

15 17 14 19 17 16 13 12 15 16 13 12 19 13 12 18 17 16 15 14 13 12 15 16 11 10

Realiza la tabla estadística de las frecuencias absoluta, relativa y en porcentajes. Calcula la media aritmética.

11.- Las temperaturas máximas en León durante el mes de mayo son las siguientes:

12 13 16 11 15 20 18 13 19 17 19 18 15 19 18 15 13 21 20 19 18 17 16 15 14 13 19 19 18 20 21

Realiza la tabla estadística de las frecuencias absoluta, relativa y en porcentajes. Calcula la media aritmética.

12.- Las alturas en cm de 20 plantas de una determinada especie son:

6.1 5.3 6.2 5.6 4.8 4.9 5.2 5.6 6.1 6.2 5.9 5.8 5.7 5.1 4.9 5.2 5.3 6.1 5.9 5.8

Agrupar las medidas en cuatro intervalos y realiza la tabla estadística de las frecuencias absoluta, relativa y en porcentajes. Calcula la media aritmética.

13.- De entre tres personas A, B y C, debe elegirse una para ocupar un puesto de trabajo. Para ello, se hacen 9 preguntas que puntúan de 1 a 10. Supón que obtienen:

Persona A 7 6 5 4 7 6 4 8 7

Persona B 3 7 5 4 8 5 9 5 9

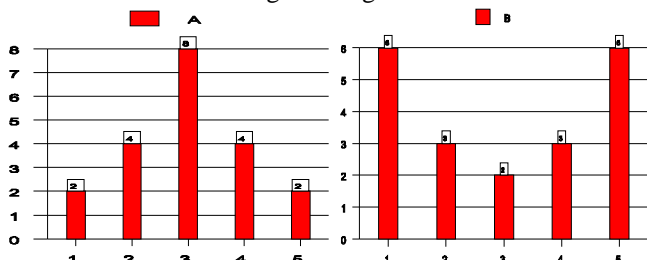
Persona C 2 3 6 6 9 4 9 9 6

Si resulta escogida la persona C y no ha habido favoritismos, ¿qué parámetro estadístico han valorado más los examinadores? ¿Y si la persona escogida es B?

14.- La nota media de un alumno en 6 pruebas, puntuadas de 0 a 10, ha sido 5.5. Si cinco de las notas eran 3, 4, 7, 9, 3, ¿cuál es la nota que falta?

15.- Calcula la moda y la mediana de los ejercicios 4, 5, 6, 7, 8, 10.

16.- Observa los siguientes gráficos:



- ¿Qué media corresponde al gráfico A? ¿Y al B?
- ¿Dónde hay mayor dispersión, en A o en B?
- Calcula la desviación al A y en B.