Menú principal

Estadística descriptiva. Introducción.

Esquema

Población e individuo

En Estadística, con el termino de **población** nos referimos a un conjunto de elementos cada uno de los cuales cumple una determinada característica. La definición que se establezca de una determinada población debe concretarse de tal manera que considerado un elemento arbitrario, podamos decidir si pertenece o no a la misma. Por ejemplo, al definir *la población de los alumnos del I.E.S "López Neyra"* tenemos que considerar si al elegir un individuo, en el ejemplo un alumno matriculado libre o un alumno que se dé de baja durante el curso, pertenece o no a dicha población. En particular, en la definición de la población deben recogerse todos aquellos casos en los que por estar próximos a la misma puedan presentarse dudas sobre su pertenencia. Otro ejemplo: si se trata de estudiar la población de los ciudadanos españoles, en la definición que demos hemos de saber si se incluye dentro a un ciudadano español residente en el extranjero, o a un ciudadano extranjero residente en España, o un ciudadano con doble nacionalidad; o, cualquier otra circunstancia que se pueda presentar en la práctica. A veces la definición no está totalmente cerrada, pues pueden presentarse casos no recogidos en la misma, cuando ello sucede no hay ningún inconveniente en ampliar la definición para poder decidir sobre su posible inclusión. Otras veces, la mayor parte, la definición se ha hecho con total precisión y no caben ambigüedades de interpretación aunque es conveniente hacerla explícita desde el principio para evitar en lo posible malentendidos. Téngase en cuenta que la definición de una determinada población puede cambiar en relación a las característica que se quieran estudiar y ello para un mismo autor o un mismo gabinete de estudios o centro. La definición depende en grado sumo del contexto en el cual se establezca y por tanto, para entendernos, conviene explicitarla todo cuanto nos sea posible.

Con el termino **individuo** designamos a cada unidad o elemento de la población. Un individuo no tiene por qué ser una persona o un objeto puede muy bien ser un conjunto de ellos. Por ejemplo, si pretendemos estudiar el conjunto de las familias españolas según sus rentas, la unidad , el individuo, sería cada una de las familias.

Muestra designa a un subconjunto de la población. Normalmente cuando las poblaciones son muy grandes se suelen elegir para estudiarlas muestras representativas de las mismas. Los datos obtenidos de la muestra, bajo ciertas restricciones pueden extrapolarse al conjunto de la población.

[Volver al principio]

Caracteres y modalidades.

Los individuos de una población pueden ser descritos según uno o varios caracteres, por ejemplo los alumnos matriculados en el I.E.S “López Neyra" pueden describirse según el número de asignaturas suspensas en la 1ª Evaluación. También podrían estudiarse según el curso de pertenencia y el número de asignaturas suspensas, etc,. Cada uno de estos caracteres pueden presentar una o varias **modalidades**. En nuestro caso, el número de asignaturas suspensas puede presentar los valores de 0, 1, 2, 3, 4, 5, 6, 7 ,8, 9. Las modalidades de un carácter son las diferentes situaciones posibles del carácter. Las modalidades de un carácter tienen que ser **exhaustivas** e **incompatibles.** Esto quiere decir:

Exhaustivas: Cada individuo de la población presenta al menos una de las modalidades del carácter.

Incompatible: Un mismo individuo no puede presentar dos modalidades diferentes.

En nuestro ejemplo, la **población** estaría formada por todos los alumnos del I.E.S “López Neyra” evaluados en la 1ª Evaluación. El **carácter** según el cual se describen a los alumnos de la población es “número de asignaturas suspensas”. Las **modalidades** del carácter son: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9..., los diferentes valores que puede tomar. Escogido un alumno de la población, necesariamente ha de presentar una de estas modalidades, es decir el número de asignaturas suspensas ha de estar comprendido entre 0 y 9. Un mismo alumno no puede presentar dos modalidades diferentes, es decir un alumno clasificado por ejemplo con 2 asignaturas suspensas no puede también entrar a formar parte de los alumnos clasificados o con 3 asignaturas suspensas. Esto se resume diciendo que las modalidades de un carácter son exhaustivas e incompatibles.

[Volver al principio]

Caracteres cualitativos y cuantitativos.

Un carácter diremos que es **cualitativo** si sus diversas modalidades no son medibles, es decir no se les puede asignar un número. Por ejemplo, si se trata de describir a la población extranjera residente en España según su nacionalidad, las diferentes modalidades del carácter serían : franceses, portugueses, ingleses, alemanes, italianos, marroquíes, holandeses, etc,. Si se trata de describir a la población española según su estado civil, las diferentes modalidades del carácter serían: Casado, soltero, viudo, divorciado, etc,. Ninguna de ellas es medible o por lo menos significativamente medible. Se tratan de caracteres cualitativos.

Un carácter diremos que es **cuantitativo** si sus diversas modalidades son medibles, es decir se les puede asignar un número, son cuantificables. Este número –variable con la modalidad pero específico de cada modalidad- se llama **variable estadística**. Por ejemplo, el conjunto de los alumnos evaluados en la 1ª evaluación del I.E.S “López Neyra” según el número de asignaturas suspensas. Otro ejemplo, alumnos matriculados en el I.E.S “López Neyra” según su estatura, o según su peso, etc. Dentro de los caracteres cuantitativos distinguiremos entre caracteres **discretos y continuos.** Son discretos cuando los valores que pueden tomar las modalidades, la variable estadística, son números aislados. Por ejemplo, el número de asignaturas suspensas, el número de hijos, etc, donde los valores posibles son 0, 1, 2, … sin que puedan tomar ningún valor decimal. Son continuos aquellos otros caracteres en los que la variable puede tomar, al menos *a priori*, cualquier valor dentro de un determinado intervalo, como por ejemplo la estatura de los alumnos de I.E.S “López Neyra”, o su peso o su velocidad en las pruebas de 100 m lisos, etc,. A veces un carácter discreto suele tratarse como si fuera continuo, ello es frecuente sobre todo cuando el conjunto de valores que puede tomar la variable estadística sea muy amplio, por ejemplo si vamos a estudiar las familias españolas según sus ingresos mensuales, puesto que la unidad monetaria mas pequeña es la peseta, se trataría de un carácter discreto, pero al ser tan amplio el conjunto de valores posibles es recomendable y casi obligatorio tratarlo como si fuera continuo.

La distinción entre una variable discreta y otra continua es un poco arbitraria. En realidad toda medida es siempre discreta, debido a la precisión siempre limitada de los instrumentos de medida.

Para estudiar las variables estadísticas continuas, se definen las clases (o grupos) de valores posibles que son las modalidades del carácter. Estas clase pueden tener una amplitud variable o constante. La elección de las amplitudes de clase está condicionada por la preocupación de tener efectivos comparables. Se consideran en general clases de amplitud variable: de pequeña amplitud donde le carácter el frecuente y de mayor amplitud donde el carácter es raro. En la estadística industrial, por simplicidad se consideran clases de igual amplitud.

[Volver al principio]

Distribuciones estadísticas de un carácter. Tablas. Representación gráfica.

Tablas. Frecuencias absolutas y relativas.

Consideremos una población de *n* individuos descrita según el carácter C, cuyas k modalidades son: C_1, C_2, ..., C_k.

Designemos con n_i el número de individuos que presentan la modalidad C_i del carácter C: n_i es la frecuencia absoluta de la modalidad C_i y la proporción:

$$\frac{n_i}{N} = fi$$

es la frecuencia relativa.

Obviamente, se tiene:

$$\sum_{i=1}^{k} n_i = N, \quad \sum_{i=1}^{k} f_i = 1$$

La tabla tiene la forma:

Modalidades del carácter	Frecuencias absolutas	Frecuencias relativas
C_1	n_1	f_1
C_2	n_2	f_2
…	…	…
C_k	n_k	f_k
Total	N	1

Respecto del carácter C, los n_i individuos que presentan la modalidad C_i son equivalentes.

[Volver al principio]

Representación de tablas estadísticas cualitativas.

La representación se hace mediante el **diagrama de barras**. Cada barra vertical corresponde a una cualidad de la variable y su longitud es proporcional a la correspondiente intensidad.

A veces, para hacer más llamativa la representación, se dibuja en vez de barras, pilas de pesetas, hombres, etc,. Objetos alusivos a lo que se trata de representar. Tales representaciones se conocen con el nombre de **pictogramas.**

Gráfico de sectores. Un circulo o semicírculo se ha dividido en sectores de áreas proporcionales a las intensidades con que se presentan las diversas modalidades del carácter. Asimismo se emplean frecuentemente los **cartogramas**, que son mapas divididos en zonas de distintos colores que corresponden a distintas intensidades del valor del carácter (población, número de fábricas, etc,.)

Hay que tener cuidado con las representaciones gráficas, pues hábilmente manejadas pueden dar lugar a una impresión exagerada respecto de la realidad.

[Volver al principio]

Representación gráfica de tablas estadísticas cuantitativas.

Muy frecuentemente se presentan en estadística variables cuantitativas, susceptibles de tomar valores numéricos distintos. Por ejemplo en la tabla figuran las calificaciones obtenidas por 100 operarios al aplicarles un cierto test, estando la variable, que es la nota, comprendida entre 0 y 10 puntos.

0	0	3	8	4	8	5	4	8	2
1	3	7	8	9	8	9	7	6	6
5	6	4	8	9	4	9	10	2	4
8	7	4	3	3	5	6	7	6	8
7	7	6	9	10	9	5	7	8	5
3	6	7	9	7	6	4	6	3	7
8	2	1	6	1	7	1	8	9	1
2	2	5	6	5	4	5	4	9	5
3	3	4	2	2	7	6	5	6	4
5	4	7	3	8	2	5	6	4	4

Nº de puntos	Repetición	Frecuencia relativa
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Se tiene una idea mejor de cómo se distribuyen los operarios en relación con su habilidad, formando la tabla de frecuencias. En ella aparecen en la primera columna, los valores de la variable estadística; en la segunda columna las frecuencias absolutas, o sea el número de operarios que hay en cada clase; y, en la tercera, las frecuencias relativas.

La representación gráfica se hace mediante el diagrama de barras, que representa en coordenadas cartesianas los valores de la variable estadística como abscisas, y construyendo en cada punto una ordenada, que mida la frecuencia absoluta correspondiente a dicho valor.

Análogamente se puede construir un diagrama de frecuencias relativas tomando estas como ordenadas.

Curva acumulativa o de distribución.

Sea F(x) la proporción de individuos de la población cuyo carácter es inferior a x, es decir:

$$F(x) = \sum_{j=1}^{i} f_j$$

para

$$x_i < x \leq x_{i+1}.$$

Esta función, llamada acumulativa o de distribución es una función escalonada, constante entre cada dos valores consecutivos de la variable, no decreciente y acotada entre 0 y 1.

[Volver al principio]

Representación gráfica de tablas de variables cuantitativas continuas.

Cuando la variable estadística es continua; es decir, puede tomar, teóricamente , todos los valores de un cierto intervalo. Por ejemplo, los pesos en gramos (aproximados hasta 0,01 gr) de 70 comprimidos fabricados automáticamente por una máquina son:

1,78	1,64	1,86	1,73	1,55	1,66	1,69	1,81	1,62	1,73
1,82	1,75	1,63	1,50	1,64	1,94	1,81	1,53	1,77	1,77
1,56	1,76	1,68	1,69	1,65	1,72	1,71	1,58	1,67	1,73
1,72	1,65	1,70	1,59	1,76	1,70	1,69	1,78	1,68	1,77
1,71	1,68	1,69	1,68	1,61	1,67	1,64	1,70	1,79	1,69
1,66	1,69	1,66	1,74	1,62	1,69	1,70	1,68	1,60	1,75
1,72	1,59	1,65	1,70	1,72	1,67	1,60	1,63	1,76	1,81

Una representación gráfica de estos datos puede hacerse mediante el diagrama de puntos tomando como abscisas los valores de los pesos y marcando un punto en la ordenada, por cada comprimido que tiene el peso que se considera.

Otra forma es el diagrama acumulativo, que se obtiene inmediatamente del anterior acumulando a cada ordenada los puntos situados en la precedente. Este gráfico nos da para la abscisa x el número de comprimidos cuyo peso es menor o igual que x.

Cuando hay un número de observaciones superior a 30 es conveniente construir la tabla de frecuencias agrupadas. Para esto se divide el intervalo total en que estan repartidas las observaciones en intervalos parciales, generalmente iguales, llamados intervalos de clase. Se suelen tomar estos intervalos de modo que las cifras de sus extremos tengan una cifra decimal más que las observaciones, para que se sepa sin dudar en qué intervalo se encuentra cada observación.

Intervalos	Marcas	n_i	f_i	N_i	F_i
1,475-1,525	1,50	1	0,014	1	0,014
1,525-1,575	1,55	3	0,043	4	0,057
1,575-1,625	1,60	8	0,114	12	0,171
1,625-1,675	1,65	14	0,199	26	0,370
1,675-1,725	1,70	23	0,322	49	0,693
1,725-1,775	1,75	12	0,170	61	0,863
1,775-1,825	1,80	7	0,099	68	0,962
1,825-1,875	1,85	1	0,014	69	0,976
1,875-1,925	1,90	0	0,000	69	0,976
1,925-1,975	1,95	1	0,014	70	1,000
	Total	70			

La representación de esta distribución de frecuencias se hace mediante **el histograma de frecuencias**.

Se obtiene construyendo sobre cada intervalo de clase de la variable estadística un rectángulo cuya área es proporcional a la frecuencia correspondiente al intervalo. La suma de las áreas de los rectángulos debe ser la unidad.

Cuando los intervalos no son todos de la misma amplitud hay que tener cuidado a la hora de representar el histograma: las alturas de los rectángulos son las frecuencias medias por unidad de amplitud.

Otra representación muy importante es el **polígono de acumulativo** que se obtiene del histograma anterior, partiendo del extremo izquierdo del primer intervalo con una ordenada igual a cero, trazamos un segmento que intercepte una ordenada en el otro extremo del intervalo que sea igual al área del correspondiente rectángulo. Por este punto se traza un segmento que determine una ordenada en el extremo del segundo intervalo igual a la suma de las áreas de los dos primeros rectángulos, etc.

Si se unen por una poligonal los puntos medios de las bases superiores de los rectángulos del histograma se tiene el llamado **polígono de frecuencias**.

Curva de distribución.

De manera análoga al caso de variables discretas, la función de distribución o acumulativa F(x) es la proporción de individuos de la población cuyo carácter es inferior a x. Esta función sólo se conoce para los valores de x que son extremos de clase:

$$x = e_1, e_2, \ldots, e_k.$$

$$F(e_i) = \sum_{j=1}^{i} f_j$$

La curva de distribución es la curva que pasa por los puntos (e_i , $F(e_i)$). Esta función es no decreciente y sus valores están comprendidos entre 0 y 1.

[Volver al principio]

Ejercicios

1. Lanzar 100 veces un dado al aire y formar la tabla de frecuencias correspondientes a los resultados de estas experiencias. Construir el diagrama de puntos y el histograma correspondiente.

2. Representar gráficamente las cotas x (medidas con aproximación de 0,05 milímetros) de 500 piezas fabricadas por una máquina automática cuya distribución de frecuencias es:

x	N.º de Observaciones	x	N.º de observaciones
16,85	0	17,30	42
90	1	35	39
95	6	40	38
17,00	25	45	32
05	35	50	10
10	49	55	20
15	54	60	13
20	70	65	0

3. Representar gráficamente las siguientes distribuciones de frecuencias relativas a tiempos de combustión de mechas para pólvora de 1 m de longitud.

Los tiempos están en segundos. La primera tabla se refiere a combustión al aire de 100 mechas, y la segunda a 1.200 mechas en una atmósfera de oxígeno.

EN EL AIRE		EN OXIGENO	
Tiempo	N.º de observaciones	Tiempo	N.º de observaciones
85	5	83	8
86	23	84	19
87	60	85	26
88	137	86	60
89	208	87	101
90	225	88	156
91	169	89	118
92	103	90	215
93	49	91	110
94	15	92	129
95	6	93	62
		94	10
		95	15

		96	1
	1.000		1.030

4. En un sondeo realizado para estudiar el mercado potencial de COCHES se obtuvo el siguiente resultado a las preguntas que se indican.

Precio en miles de pesetas	Porcentaje de individuos que estiman que a este precio tienen medios de comprar un coche	Porcentaje de individuos que estiman que a este precio EL COCHE no puede ser satisfactorio
500	97 %	90%
600	96	86
700	95	80
800	91	65
900	90	41
1.000	87	33
1.500	74	22
2.000	48	10
3.000	35	1
4.000	15	0
5.000	6	0

Obtener una tabla que represente para los diferentes precios señalados, porcentaje de personas que consideran que a tal precio tienen posibilidad de comprar un coche satisfactorio. Representaciones gráficas.

5. Las puntuaciones obtenidas en un test por 20 alumnos son las siguientes:

16,22,21,20,23,22,17,15,13,22,17,18,20,17, 22, 16, 23, 21, 22, 18.

a) Construir la tabla de frecuencias.

b) Representar el diagrama de barras de frecuencias absolutas y de frecuencias absolutas acumuladas.

6. Durante el mes de julio, en una determinada ciudad de la costa levantina se han registrado las siguientes temperaturas máximas:

32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29.

a) Construir la tabla de frecuencias.

b) Representar el diagrama de barras de frecuencias absolutas y de frecuencias absolutas acumuladas.

7. Se ha aplicado un test de capacidad espacial compuesto por 100 preguntas a un grupo de 100 alumnos, habiéndose obtenido los siguientes resultados:

Número de Número

preguntas correctas	de alumnos
[0-15)	10
[15-30)	15
[30-45)	25
[45-60)	20
[60-75)	20
[75-90)	10

a) Formar la tabla de frecuencias.

b) Representar el histograma de frecuencias absolutas y el hístograma de frecuencias absolutas acumuladas.

8. Se ha controlado el peso de 50 recién nacidos, en dos ciudades A y B, obteniéndose los siguientes resultados:

Peso (en kg)	**N.^{0}** de niños en A	N.^{0} de niños en B
[2,5-3)	6	11
[3-3,5)	23	26
[3,5-4)	12	9
[4-4,5)	9	4

Para cada ciudad:

a) Formar la tabla de frecuencias.

b) Representar el histograma de frecuencias absolutas de frecuencias absolutas acumuladas.

c) A partir de los histogramas, ¿qué se puede deducir?

9. La tabla siguiente indica la edad de los 40 socios de un club:

Edad	15	16	17	18	19
Número	5	8	2	20	5

Hacer el histograma o diagrama de barras correspondiente.

10. Se ha aplicado un test a los empleados de una fábrica obteniéndose la siguiente tabla:

x	[38,44)	[44,50)	[50,56)	[56,62)	[62,68)	[68,74)	[74,80)
Nº de trabajadores.	7	8	15	25	18	9	6

Se pide:

Histograma y polígono de frecuencias absolutas acumuladas.

11. Dada la distribución de frecuencias:

Intervalos	[0,3)	[3,6)	[6,9)	[9,12)	[12,15)	[15,18)
Frecuencias	2	7	12	13	4	3

Se pide:

Dibujar el histograma y el polígono de frecuencias.

12. De una muestra de 75 pilas eléctricas, se han obtenido los siguientes datos sobre duración en horas:

Duración	[25,30)	[30,35)	[35,40)	[40,45)	[45,55)	[55,70)
Nº de pilas	3	5	21	28	12	6

Representar gráficamente estos datos.

13. Se preguntó a 62 personas cuánto tiempo ha dedicado a ver la televisión durante un cierto fin de semana. Los datos obtenidos son los siguientes:

Tiempo en Horas	[0-0,5)	[0,5-1,5)	[1,5-2,5)	[2,5-4)	[4-8)
Nº de personas	10	10	18	12	12

Dibújese el histograma de frecuencias.

14. La siguiente tabla corresponde a la distribución tallas de 100 alumnos:

Tallas (cm)	**Frecuencias**
[140-150)	3
[150-160)	11
[160-170)	25
[170-180)	30
[180-190)	16

[190-200)	12
[200-210)	3

Se pide:

Histograma y polígono de frecuencias.

15. La dirección de tráfico ha recogido la siguiente información relativa al número de multas diarias que sus agentes han impuesto a los conductores que circulan por una autopista:

N.^{0} de multas	**Días**
[0-5)	6
[5-10)	14
[10-15)	20
[15-20)	10

Calcular y representar gráficamente las frecuencias acumuladas de la distribución anterior.

16. Las edades de las actrices y los actores ganadores de los premios Óscar de los últimos 30 años, han sido las siguientes:

Hombres: 32, 51, 33, 61, 35, 45, 55, 39, 76, 37, 42, 40, 32, 60, 38, 56, 48, 48, 40, 43, 62, 43, 42, 44, 41, 56, 39, 46, 31, 47.

Mujeres: 80, 26, 41, 21, 61, 38, 49, 33, 74, 30, 33, 41, 31, 35, 41, 42, 37, 26, 34, 34, 35, 26, 61, 60, 34, 24, 30, 37, 31, 27.

Construir un diagrama de tallos y hojas, de manera que los tallos sean comunes y las hojas de cada conjunto de datos queden a la izquierda y a la derecha de los tallos.

[Volver al principio]