

1. INTRODUCCIÓN

1.1. LA ESTADÍSTICA: CONCEPTO Y FUNCIÓN EN LA INVESTIGACIÓN PEGAGÓGICA EXPERIMENTAL.

La palabra "estadística" suele utilizarse bajo dos significados distintos, a saber:

Esto es el significado más vulgar de la palabra estadística. Se sobrentiende que dichos datos numéricos han de estar presentados de manera ordenada y sistemática. Una información numérica cualquiera puede no constituir una estadística, para merecer este apelativo, los datos han de constituir un conjunto coherente, establecido de forma sistemática y siguiendo un criterio de ordenación.

Tenemos muchos ejemplos de este tipo de estadísticas. El Anuario Estadístico publicado por el Instituto Nacional de Estadística, El Anuario de Estadísticas del Trabajo,...

Como ciencia.- En este significado, La Estadística estudia el comportamiento de los fenómenos de masas. Como todas las ciencias, busca las características generales de un colectivo y prescinde de las particulares de cada elemento. Así por ejemplo al investigar el sexo de los nacimientos, iniciaremos el trabajo tomando un grupo numeroso de nacimientos y obtener después la proporción de varones. Es muy frecuente enfrentarnos con fenómenos en los que es muy difícil predecir el resultado; así, no podemos dar una lista ,con las personas que van a morir con una cierta edad, o el sexo de un nuevo ser hasta que transcurra un determinado tiempo de embarazo,...

Por tanto, el objetivo de la estadística es hallar las regularidades que se encuentran en los fenómenos de masa.

Población, elementos y caracteres.

Es obvio que todo estudio estadístico ha de estar referido a un conjunto o colección de personas o cosas. Este conjunto de personas o cosas es lo que denominaremos **población**.

Las personas o cosas que forman parte de la población se denominan **elementos**. En sentido estadístico un elemento puede ser algo con existencia real, como un automóvil o una casa, o algo más abstracto como la temperatura, un voto, o un intervalo de tiempo.

A su vez, cada elemento de la población tiene una serie de características que pueden ser objeto del estudio estadístico. Así por ejemplo si consideramos como elemento a una persona, podemos distinguir en ella los siguientes **caracteres**:

Sexo, Edad, Nivel de estudios, Profesión, Peso, Altura, Color de pelo, Etc.

Luego por tanto de cada elemento de la población podremos estudiar uno o más aspectos cualidades o caracteres.

La población puede ser según su tamaño de dos tipos:

Población finita: cuando el número de elementos que la forman es finito, por ejemplo el número de alumnos de un centro de enseñanza, o grupo clase.

Población infinita: cuando el número de elementos que la forman es infinito, o tan grande que pudiesen considerarse infinitos.. Como por ejemplo si se realizase un estudio sobre los productos que hay en el mercado. Hay tantos y de tantas calidades que esta población podría considerarse infinita.

Ahora bien, normalmente en un estudio estadístico, no se puede trabajar con todos los elementos de la población sino que se realiza sobre un subconjunto de la misma. Este subconjunto puede ser **una muestra**, cuando se toman un determinado número de elementos de la población, sin que en principio tengan nada en común; o **una subpoblación**, que es el subconjunto de la población formado por los elementos de la población que comparten una determinada característica, por ejemplo de los alumnos del centro la subpoblación formada por los alumnos de 3º ESO, o la subpoblación de los varones.

Variables y atributos.

Como hemos visto, los caracteres de un elemento pueden ser de muy diversos tipos, por lo que los podemos clasificar en: dos grandes clases:

Variables **Cuantitativas**.

Variables **Cualitativas o Atributos**.

Las variables cuantitativas son las que se describen por medio de números, como por ejemplo el peso, Altura, Edad, Número de Suspensos...

A su vez este tipo de variables se puede dividir en dos subclases:

- **Cuantitativas discretas.** Aquellas a las que se les puede asociar un número entero, es decir, aquellas que por su naturaleza no admiten un fraccionamiento de la unidad, por ejemplo número de hermanos, páginas de un libro, etc.
- **Cuantitativas continuas:** Aquellas que no se pueden expresar mediante un número entero, es decir, aquellas que por su naturaleza admiten que entre dos valores cualesquiera la variable pueda tomar cualquier valor intermedio, por ejemplo peso, tiempo. etc.

No obstante en muchos casos el tratamiento estadístico hace que a variables discretas las trabajemos como si fuesen continuas y viceversa.

Los atributos son aquellos caracteres que para su definición precisan de palabras, es decir, no le podemos asignar un número. Por ejemplo Sexo Profesión, Estado Civil, etc.

A su vez las podemos clasificar en:

- **Ordenables:** Aquellas que sugieren una ordenación, por ejemplo la graduación militar, El nivel de estudios, etc.
- **No ordenables:** Aquellas que sólo admiten una mera ordenación alfabética, pero no establece orden por su naturaleza, por ejemplo el color de pelo, sexo, estado civil, etc.

1.2. ESTADÍSTICA DESCRIPTIVA

CONCEPTOS BÁSICOS

La investigación cuya finalidad es: el análisis o experimentación de situaciones para el descubrimiento de nuevos hechos, la revisión o establecimiento de teorías y las aplicaciones prácticas de las mismas, se basa en los principios de Observación y Razonamiento y necesita en su carácter científico el análisis técnico de Datos para obtener de ellos información confiable y oportuna. Este análisis de Datos requiere de la Estadística como una de sus principales herramientas, por lo que los investigadores de profesión y las personas que de una y otra forma la realizan requieren además de los conocimientos especializados en su campo de actividades, del manejo eficiente de los conceptos, técnicas y procedimientos estadísticos.

ESTADÍSTICA

Es el conjunto de procedimientos y técnicas empleadas para recolectar, organizar y analizar datos, los cuales sirven de base para tomar decisiones en las situaciones de incertidumbre que plantean las ciencias sociales o naturales.

ESTADÍSTICA INDUCTIVA Y DEDUCTIVA

Uno de los problemas fundamentales de la Estadística es el estudio de la relación existente entre una población y sus muestras. Según la dirección de tal relación la Estadística puede ser:

Deductiva, cuando a partir del conocimiento de la población se trata de caracterizar cada muestra posible.

Inductiva, cuando a partir del conocimiento derivado de una muestra se pretende caracterizar la población.

Ejemplo de la ordenación de datos en Estadística Descriptiva.

Una de las ramas de la Estadística más accesible a la mayoría de la población es la Descriptiva. Esta parte se dedica única y exclusivamente al ordenamiento y tratamiento mecánico de la información para su presentación por medio de tablas y de representaciones gráficas, así como de la obtención de algunos parámetros útiles para la explicación de la información.

La Estadística Descriptiva es la parte que conocemos desde los cursos de educación primaria, que se enseña en los siguientes niveles y que, por lo general, no pasa a ser un análisis más profundo de la información. Es un primer acercamiento a la información y, por esa misma razón, es la manera de presentar la información ante cualquier lector, ya sea especialista o no. Sin embargo, lo anterior no quiere decir que carezca de metodología o algo similar, sino que, al contrario, por ser un medio accesible a la mayoría de la población humana, resulta de suma importancia considerar para así evitar malentendidos, tergiversaciones o errores.

Representación de tronco y hoja

Un método para iniciar el análisis exploratorio de los datos, previo al uso de los métodos estadísticos tradicionales, y que además proporciona información rápida, visual y es relativamente nueva, es la **representación gráfica de tronco y hoja**. Esta representación se basa en la ordenación de los datos a manera de gráfico, pero sin llegar a ello, utilizando las decenas y las unidades.

Esta técnica se puede encontrar en el libro de Freund y Simon, pero comentaremos su uso a través del siguiente ejemplo que contiene las calificaciones obtenidas en una prueba de matemáticas:

78	93	61	100	70	83	88	74	97	72
66	73	76	81	83	64	91	70	77	86

Ahora pensaremos en cada uno de los datos separando las decenas de las unidades, es decir, el número 51 se verá como 5 | 1. De esta manera las decenas se pondrán en una columna, en forma vertical, y las unidades a su derecha:

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

Para entenderle un poco más, hemos de decir que el primer renglón que dice 6 | 1 6 4 quiere decir que entre la lista de datos se encuentran los valores 61, 66 y 64.

Esta es la representación gráfica **tronco y hoja**, donde cada renglón es una posición de **tronco** y cada dígito de la derecha es una **hoja**.

El procedimiento para realizarla es primero empezar con los troncos, es decir la columna de la izquierda, y después dato por dato ir llenando las hojas a la derecha de la línea vertical, en el tronco correspondiente.

Además, si se desean tener los datos ordenados, y hay gente que lo prefiere así, se pueden ordenar las hojas en cada renglón para que la representación quede como sigue:

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

En realidad una representación de tronco y hojas presenta la misma información que la lista original de datos, pero de una manera mucho más compacta (especialmente si la lista de datos es más grande) y manejable.

Sin embargo, información más compleja resulta un poco más difícil de manejar, por lo que en ocasiones conviene redondear los datos, ignorar sus partes decimales o utilizar las centenas u otras posiciones de los números para las troncos. En cada uno de esos casos conviene hacer alguna anotación, o poner una nota, a fin que los lectores puedan identificar las adecuaciones realizadas y así poder interpretar lo que se quiere transmitir.

Para mostrar la información de manera más clara, es posible modificar el número de posiciones del tronco, aumentándola o disminuyéndola de acuerdo a las necesidades particulares de cada problema. Por ejemplo, con los datos del examen anterior, se pueden dividir en dos cada posición del tronco, utilizando la primera posición para disponer las hojas 0, 1, 2, 3 y 4, y la segunda posición para las hojas restantes. De esta manera, se obtiene la **representación gráfica de doble tronco**:

6-	1	4
6+	6	
7-	0 0	2 3 4
7+	6	7 8
8-	1	3 3
8+	6	8
9-	1	3
9+	7	
10-	0	

Con esto se han duplicado el número de posiciones del tronco, con la intención de buscar una mayor claridad en la presentación.

Esta manera de representación inicial de los datos no la profundizaremos más, sino que la utilizaremos más adelante en algunos casos para, precisamente, presentar una representación inicial de la información obtenida.

1.2. ESTADÍSTICA DESCRIPTIVA E INFERENCIAL

Estadística Descriptiva se refiere a la recolección, presentación, descripción, análisis e interpretación de una colección de datos, esencialmente consiste en resumir éstos con uno o dos elementos de información (medidas descriptivas) que caracterizan la totalidad de los mismos. La estadística Descriptiva es el método de obtener de un conjunto de datos conclusiones sobre si mismos y no sobrepasan el conocimiento proporcionado por éstos. Puede utilizarse para resumir o describir cualquier conjunto ya sea que se trate de una población o de una muestra, cuando en la etapa preliminar de la Inferencia Estadística se conocen los elementos de una muestra.

Estadística Inferencial se refiere al proceso de lograr generalizaciones acerca de las propiedades del todo, población, partiendo de lo específico, muestra las cuales llevan implícitos una serie de riesgos. Para que éstas generalizaciones sean válidas la muestra deben ser representativa de la población y la calidad de la información debe ser controlada, además puesto que las conclusiones así extraídas están sujetas a errores, se tendrá que especificar el riesgo o probabilidad que con que se pueden cometer esos errores.

La estadística inferencial es el conjunto de técnicas que se utiliza para obtener conclusiones que sobrepasan los límites del conocimiento aportado por los datos, busca obtener información de un colectivo mediante un metódico procedimiento del manejo de datos de la muestra.

En sus particularidades la Inferencia distingue la Estimación y la Contrastación de Hipótesis. Es estimación cuando se usan las características de la muestra para hacer inferencias sobre las características de la población. Es contrastación de hipótesis cuando se usa la información de la muestra para responder a interrogantes sobre la población.

ANÁLISIS ESTADÍSTICO

El análisis estadístico es todo el proceso de organización, procesamiento, reducción e interpretación de datos para realizar inferencias.

DATOS Y VARIABLES

Cuando se consideran los métodos de organización, reducción y análisis de datos estadísticos, se hace necesario aclarar los siguientes conceptos.

Variabes: es toda característica que varía de un elemento a otro de la población.

Datos: son medidas o valores de las características susceptibles de observar y contar, se originan por la observación de una o más variables de un grupo de elementos o unidades

CLASIFICACIÓN DE VARIABLES

Las variables pueden clasificarse en: categóricas o cualitativas (atributos), no tienen ningún grado de comparación numérica, ejemplo: sexo, estado civil; y numéricas o cuantitativas, son características factibles de expresar por medio de números, estas pueden ser Discretas, que solo pueden tomar ciertos valores aislados en un intervalo, y Continuas, que pueden tomar cualquier valor en un intervalo.

1.3. ORDENACIÓN Y CATEGORIZACIÓN DE LOS DATOS.

Los datos son colecciones de un número cualquiera de observaciones relacionadas entre si, para que sean útiles se deben organizar de manera que faciliten su análisis, se puedan seleccionar tendencias, describir relaciones, determinar causas y efectos y permitan llegar a conclusiones lógicas y tomar decisiones bien fundamentadas; por esa razón es necesario conocer lo métodos de Organización y Representación, la finalidad de éstos métodos es permitir ver rápidamente todas las características posibles de los datos que se han recolectado.

Representación Tabular:

Presenta las variable y las frecuencias con que los valores de éstas se encuentran presentes en el estudio.

Representación Gráfica :

Se llaman gráficas a las diferentes formas de expresar los datos utilizando los medios de representación que proporciona la geometría.

REPRESENTACIÓN DE DATOS

Arreglo de Datos. Es una forma de presentar los datos en un arreglo ascendente o descendente. Ofrece las ventajas siguientes: describe los valores mínimos y máximos, en él se pueden dividir los datos fácilmente en secciones, permite darse cuenta de los valores que aparecen más de una vez, se puede observar la distancia entre valores consecutivos.

Diagrama de Puntos. Muestra la frecuencia con que aparece cada uno de los valores

Diagrama de Tallo y Hoja. Es útil para realizar una exploración preliminar del conjunto, genera una imagen adecuada de ellos sin perder información.

Distribución de Frecuencias. Es una forma de sintetizar los datos y consiste en valerse de una tabla para clasificar los datos según su magnitud, en ella se señala el número de veces que aparece cada uno de los valores. Cuando se dispone de un gran número de valores discretos o cuando las variables son continuas, tiene sentido formar una tabla que presente la distribución de frecuencias de los datos agrupados en intervalos o clases, de igual tamaño si es posible, sin embargo una tabla de este tipo supone una concentración de datos que produce pérdida de información.

Aquí se nos muestra un ejemplo de la ordenación de los datos según el tipo de categorización:

Ejercicio 1. Las edades de las personas que acuden a la consulta de un determinado psicólogo en un mes se recogen en la siguiente tabla:

3	6	5	34	23	6	12	14	4	35
8	22	6	9	16	14	4	7	6	9
8	42	10	5	6	3	7	9	7	6
21	32	17	15	11	14	17	3	5	4
13	15	12	6	14	8	9	12	25	11

Construir la tabla de frecuencias.

Datos ordenados:

3	4	6	6	8	9	12	14	17	25
3	5	6	7	8	10	12	14	17	32
3	5	6	7	9	11	13	15	21	34
4	5	6	7	9	11	14	15	22	35
4	6	6	8	9	12	14	16	23	42

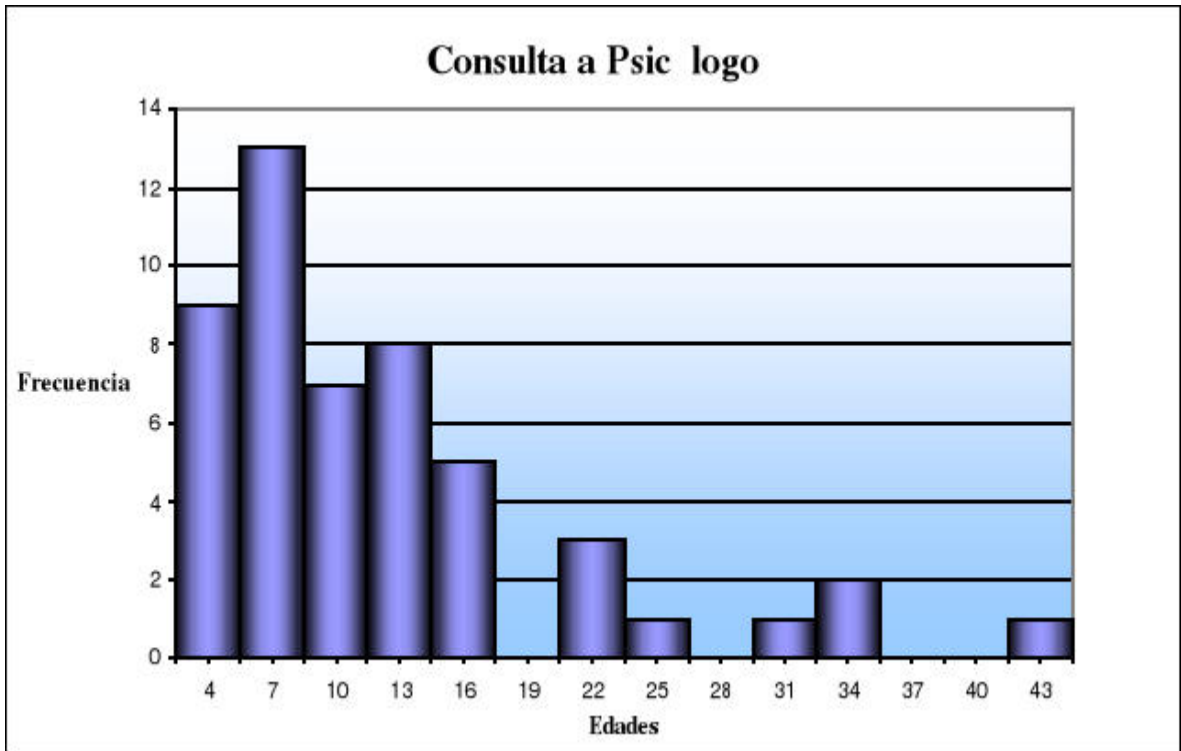
Tabla de frecuencias:

Intervalos	f.absoluta	f.relativa	f.a.acumulad	f.r.acumulada
3-5	9	0,18	9	0,18
6-8	13	0,26	22	0,44
9-11	7	0,14	29	0,58
12-14	8	0,16	37	0,74
15-17	5	0,1	42	0,84
18-20	0	0	42	0,84
21-23	3	0,06	45	0,9
24-26	1	0,02	46	0,92
27-29	0	0	46	0,92
30-32	1	0,02	47	0,94
33-35	2	0,04	49	0,98
36-38	0	0	49	0,98
39-41	0	0	49	0,98
42-44	1	0,02	50	1
	50	1		
intervalo crítico				

• **Gráfico de tallo y hojas.**

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

• **Realizar la representación gráfica mediante histograma.**



d. Hallar la mediana de la distribución.

$Mdn = (n+1) / 2$	$Mdn = (50 + 1) / 2$	$Mdn = 25,5$
-------------------	----------------------	--------------

La mediana estaría entre el dato 25 y el dato 26, es decir, entre 9 y 9. Por lo tanto

<u>MEDIANA = 9</u>

BIBLIOGRAFIA

- QUESADA V., ISIDORO, LÓPEZ: "Curso y Ejercicios de Estadística", Ed. Alhambra, 1989. (Problemas-Teoría)
Notas. Existen ejemplares en: biblioteca de Matemáticas, Biología, Inf./Est. Tiene muchos problemas resueltos de Estadística Descriptiva, Probabilidad e Inferencia Paramétrica (también problemas propuestos). Contiene un breve resumen teórico de cada tema.
- RUIZ CAMACHO M., MORCILLO AIXELÁ M.C., GARCÍA GALISTEO J., CASTILLO VÁZQUEZ C.: "Curso de Probabilidad y Estadística", Ed. Universidad de Málaga / Manuales, 2000. (Teoría-Problemas)
Notas. Existen ejemplares en: biblioteca de Matemáticas. Un libro de teoría muy "didáctico" de Estadística Descriptiva y Probabilidad. Contiene problemas resueltos y propuestos, al final de cada bloque teórico.
- SARABIA VIEJO A., MATE JIMÉNEZ C.: "Problemas de Probabilidad y Estadística. Elementos teóricos, cuestiones, aplicaciones con Statgraphics", Ed. CLAGSA, 1993.(Problemas)
Notas. Existen ejemplares en: biblioteca de Matemáticas. Tiene muchos problemas resueltos de Estadística Descriptiva, Probabilidad e Inferencia Paramétrica (también problemas propuestos). Resuelve algunos problemas con el programa Statgraphics en su versión de MSDOS.
- WALPOLE R.E., MYERS R.H., MYERS S.L.: "Probabilidad y Estadística para Ingenieros", Ed. Prentice Hall, 1998, 6ª edición.(Teoría)
Notas. Existen ejemplares en: ¿?. Un libro de teoría muy "didáctico" de Estadística Descriptiva, Probabilidad, Inferencia Paramétrica, Inferencia No Paramétrica y Diseño de Experimentos. Contiene problemas propuestos después de cada apartado tratado (solución corta de problemas impares) y con muchos ejemplos prácticos resueltos en el desarrollo del temario.