

CBS

Colegio Bautista Shalom



Estadística Descriptiva

Quinto BADG

Primer Bimestre

Contenidos

ESTADÍSTICA

- ✓ ¿QUÉ ES ESTADÍSTICA?
- ✓ OBJETIVOS.
- ✓ TIPOS DE ESTADÍSTICA.
- ✓ UTILIDAD E IMPORTANCIA.
- ✓ APLICACIÓN DE LA ESTADÍSTICA.
- ✓ PRESENTACIÓN DE LOS DATOS.
- ✓ NÚMERO DE ÍNDICES.
- ✓ ESTADÍSTICAS COMUNES.
- ✓ CONCEPTOS BÁSICOS DE ESTADÍSTICA.
 - POBLACIÓN.
 - MUESTRA.
 - INDIVIDUO.
 - VALOR.
 - DATO Y VARIABLE.
 - MUESTREO.
- ✓ TIPOS DE VARIABLES ESTADÍSTICAS.
- ✓ RECOLECCIÓN DE DATOS.
 - TÉCNICAS DE RECOLECCIÓN DE DATOS ESTADÍSTICOS.
- ✓ DISTRIBUCIÓN DE FRECUENCIAS.
- ✓ TABLAS ESTADÍSTICAS.
- ✓ CÁLCULO DE PROMEDIO DATOS NO AGRUPADOS.
- ✓ CÁLCULO DE MEDIANA DATOS NO AGRUPADOS.
- ✓ CÁLCULO DE MODA DATOS NO AGRUPADOS.
- ✓ CÁLCULO DE PROMEDIO DATOS AGRUPADOS.
- ✓ CÁLCULO DE MEDIANA DATOS AGRUPADOS.
- ✓ CÁLCULO DE MODA DATOS AGRUPADOS.
- ✓ EL GRÁFICO.
 - TIPOS DE GRÁFICAS.

NOTA: conforme tu catedrático/a vaya avanzando con el contenido del presente curso, te indicará la actividad o ejercicio a realizar. Sigue sus instrucciones.

ESTADÍSTICA

¿QUÉ ES ESTADÍSTICA?

La estadística es la ciencia que estudia cómo debe emplearse la información y cómo dar una guía de acción en situaciones prácticas que entrañan incertidumbre. (*Gutiérrez, p.23*)

La estadística es la ciencia de los datos, la cual implica su recolección, clasificación, síntesis, organización, análisis e interpretación, para la toma de decisiones frente a la incertidumbre. (*Ángel, p. 28*)

La estadística es la rama del conocimiento humano que tiene como objeto el estudio de ciertos métodos inductivos aplicables a fenómenos susceptibles de expresión cuantitativa. (*López, p.1*)

OBJETIVOS

El objetivo básico de la estadística es hacer inferencia acerca de una población basada en la información contenida en una muestra. Inferir significa predecir, suponer, asegurar. Es decir, se pretende establecer inferencia acerca de una población. Entendiendo a la población como un conjunto de individuos, organismos o entes inanimados de los cuales queremos conocer alguna o algunas características para que nos ayuden a tomar una decisión u obtener alguna conclusión de suma importancia.

El objetivo de la estadística es mejorar la comprensión de hechos a partir de datos. (*Moore, p.267*)

El principal objetivo de la estadística es hacer inferencias acerca de una población, con base en la información contenida en una muestra. (*Pérez, p.172*)

TIPOS DE ESTADÍSTICA

Básicamente se tienen dos tipos de estadística, a saber:

Estadística descriptiva. Se puede definir como un método para describir numéricamente conjuntos numerosos. Por tratarse de un método de descripción numérica, utiliza el número como medio para describir un conjunto, que debe ser numeroso, ya que las permanencias estadísticas no se dan en los casos raros. No es posible sacar conclusiones concretas y precisas de los datos estadísticos. (*Vargas, p.33*)

Objetivo de la estadística descriptiva. La finalidad última de la estadística descriptiva es resumir la información de conjuntos más o menos numerosos de datos. Para ello se asienta en un concepto inmediato a la tarea de recuento: la frecuencia, medida empírica de la ocurrencia de los distintos estados que puede presentar una variable. (*SGT, p.16*)

Estadística inferencial, analítica o deductiva. Estudia la probabilidad de éxito de las diferentes soluciones posibles a un problema en las diferentes ciencias en las que se aplica y para ello utiliza los datos observados en una o varias muestras de la población. Mediante la creación de un modelo matemático infiere el comportamiento de la población total partiendo de los resultados obtenidos en las observaciones de las muestras. (*Fernández et.al, p.17*)

Objetivo de la estadística inferencial. La inferencia estadística intenta tomar decisiones basadas en la aceptación o el rechazo de ciertas relaciones que se toman como hipótesis. Esta toma de decisiones va acompañada de un margen de error, cuya probabilidad está determinada. (*Vargas, p.33*)

La estadística inferencial tiene dos objetivos básicos; a) obtener conclusiones válidas acerca de una población sobre la base de una muestra, es decir, que las conclusiones que obtengamos de una muestra se puedan extrapolar a la población que dio origen a esa muestra y b) poder medir el grado de incertidumbre presente en dichas inferencias en términos de probabilidad. (*Díaz, p.287*)

UTILIDAD E IMPORTANCIA

Podemos decir que la función principal de la estadística es justamente la recolección y agrupamiento de datos de diverso tipo para construir con ellos informes estadísticos que nos den idea sobre diferentes y muy variados temas, siempre desde un punto de vista cuantitativo y no cualitativo. Esto es muy importante de remarcar ya que la estadística se convierte entonces en una ciencia que nos habla de cantidades (por ejemplo, cuántas personas viven en un país por metro cuadrado) pero no nos da información directa sobre la calidad de vida de esas personas. En

este sentido podemos decir que se presentan varias limitaciones ya que no permite conocer más que numéricamente aspectos que requieren un trabajo más complejo y profundo.

Los métodos estadísticos tradicionalmente se utilizan para propósitos descriptivos, para organizar y resumir datos numéricos. La estadística descriptiva, por ejemplo, trata de la tabulación de datos, su presentación en forma gráfica o ilustrativa y el cálculo de medidas descriptivas. Ahora bien, las técnicas estadísticas se aplican de manera amplia en mercadotecnia, contabilidad, control de calidad y en otras actividades; estudios de consumidores; análisis de resultados en deportes; administradores de instituciones; en la educación; organismos políticos; médicos; y por otras personas que intervienen en la toma de decisiones.

APLICACIÓN DE LA ESTADÍSTICA

Como parte de las ciencias de investigación y análisis de conjuntos de datos surge la incógnita ¿utilizaron la estadística Galileo, Newton y Einstein como herramienta en sus investigaciones?

En la Biología, Ciencias Humanas y en algunos campos de la Física podemos encontrar el concepto de “*experimento aleatorio*” el cual consiste en un experimento que siendo repetido en las mismas condiciones no produce el mismo resultado y asociado a este concepto encontramos también presente el de variable aleatoria. En este sentido una **variable no aleatoria** es aquella asociada al resultado de una experiencia que sí produce el mismo resultado y está caracterizada por un valor para cada condición.

Una **variable aleatoria** está caracterizada por la llamada función densidad de probabilidad, a partir de la cual se obtienen las probabilidades para sus posibles valores para cada condición. Para comprender la aplicación de la función de probabilidad es importante saber que la variable aleatoria permite definir la probabilidad como una función numérica (de variable real) y no como una función de conjunto.

Por ejemplo: Se tira 3 veces una moneda. Como solo se tienen 2 opciones de resultado (cara o escudo) se representa cara por “c” y escudo por “e”.

Entonces, se tiene la probabilidad elemental de cada suceso que es $1/8$.

$$\Omega = \{ccc, cce, cec, ecc, cee, ece, eec, eee\}$$

Ya que la probabilidad de que caiga cara en cada tirada es de $1/2$ y cada una de las tiradas de moneda son independientes.

NOTA: Para encontrar el número de posibles soluciones se eleva 2 a la variable “n” (2^n), siendo la variable $n = 3$, porque 3 veces será tirada la moneda. Entonces, $2^3 = 8$.

Ahora, definimos la variable aleatoria por “X”: número de caras, que puede tomar los valores de [0, 1, 2, 3]. Procedemos a buscar todos los puntos muestrales que den lugar a cada valor de la variable y a ese mismo le asignaremos la probabilidad de suceso correspondiente.

x	Sucesos	p_x
0	{eee}	$1/8$
1	{cee, ece, eec}	$3/8$
2	{cce, cec, ecc}	$3/8$
3	{ccc}	$1/8$

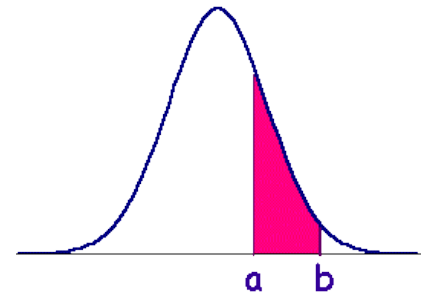
La función densidad de probabilidad no funciona de igual forma con variables discretas como con variables continuas. En el caso de las variables discretas, como en el ejemplo, es una función que para cada valor de la variable da su probabilidad.

Por ejemplo. Supongamos la variable *tipo histológico* de un tumor, con los valores 1, 2, 3, 4. Si la *función densidad de probabilidad* fuera:

x	f(x)
1	0,22
2	0,27
3	0,30
4	0,21

...daría a entender que la probabilidad del tipo 2 es 0,27; y así, continuamente.

Para variables continuas la probabilidad de que una variable tome *cualquier valor* concreto es 0, por lo tanto la *función densidad de probabilidad* sólo permite calcular la probabilidad para un intervalo del tipo $(a < X < b)$, mediante el área bajo la curva de la *función densidad de probabilidad*. En conclusión diremos que, la estadística es la ciencia que estudia los métodos que permiten realizar este proceso para variables aleatorias. Estos métodos permiten resumir datos y acotar el papel de la casualidad (azar).



La Estadística, como todas las ciencias, no surgió de improviso, sino mediante un proceso largo de desarrollo y evolución, desde hechos de simple recolección de datos hasta la diversidad y rigurosa interpretación de los datos que se dan hoy en día. Así pues, el origen de la Estadística se remonta a los comienzos de la historia y esto se sabe tanto a través de crónicas, datos escritos, como de restos arqueológicos, y esto es explicable por cuanto en ese tiempo se estaba formado recién la sociedad y es algo inherente la necesidad de saber cosas elementales como:

Cuántos habitantes tiene a tribu, con cuántos bienes cuenta, entre otros aspectos...

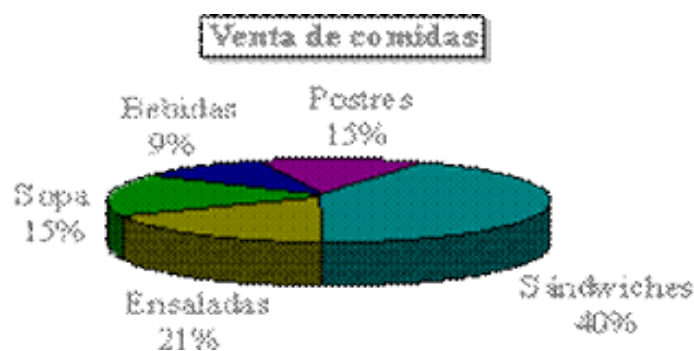
PRESENTACIÓN DE LOS DATOS

Los datos estadísticos se presentan generalmente expresando el valor de la **frecuencia absoluta** que toman las variables significativas de un estudio, ya correspondan a una población o a una muestra. La frecuencia absoluta de un valor o de una modalidad de una variable estadística es el número de datos observados que presentan ese valor o modalidad. El cociente entre la frecuencia absoluta de un valor o modalidad y el número total de datos es llamado **frecuencia relativa**. También suelen presentarse los datos en forma de **porcentaje** (es decir, en forma de **razón** de denominador 100).

Una razón se obtiene como el cociente entre dos cantidades numéricas comparables. Si el cociente se refiere a dos cantidades que se indican en unidades distintas, la razón recibe el nombre de **tasa**. Un ejemplo de tasa es la **densidad de población**, que se define como el número de habitantes por kilómetro cuadrado y que se aplica habitualmente en los estudios demográficos.

Dada una suma de varios sumandos, si el cociente hace referencia a la división numérica entre uno de los sumandos y la suma total, la cantidad expresada se denomina proporción.

Gráficos de sectores que reflejan el Índice de Consumo (IPC).



NÚMERO DE ÍNDICES

Otros métodos de presentación de datos utilizados en estadística se basan en el empleo de **números índices**. Tales números reflejan la evolución que experimenta con el paso del tiempo una variable estadística de interés. Así, se toma como referencia del índice el valor de la variable en un instante dado, de manera que sus valores posteriores se expresan como una razón de cambio con respecto a dicha referencia (a menudo, en forma de porcentaje).

Un ejemplo típico de empleo de números índices es el índice bursátil, cuya definición obedece a criterios diferentes en cada país (**índice Dow Jones**, en la Bolsa de Nueva York; **índice Nikkei**, en Tokio, etcétera).

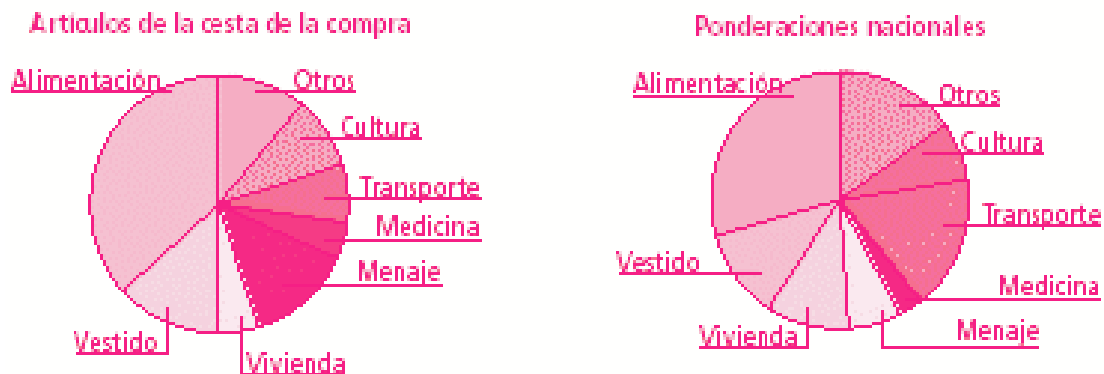
El gráfico de evolución de la Bolsa de Valores constituye un ejemplo de número índice.

ESTADÍSTICAS COMUNES

Varios estudios estadísticos comunes que aparecen con frecuencia en los medios de comunicación son los siguientes:

1. **Encuesta de Población Activa (EPA)**, elaborada por el Instituto Nacional de Estadística (INE) con periodicidad trimestral, según recomendaciones de la Organización Internacional del Trabajo (OIT), para obtener y clasificar datos sobre la actividad de la población. Esta encuesta se realiza por muestreo, y los resultados se ordenan por edad, sexo, nivel de estudios, profesión y otros parámetros.
2. **Índice de Precios al Consumo (IPC)**, que mide por medios estadísticos la evolución experimentada por los precios de los bienes y servicios consumidos por la población española. Se basa en la Encuesta de Presupuestos Familiares (EPF), y selecciona varios centenares de artículos, clasificados en ocho grupos, que se consideran representativos de la evolución de los precios. Los artículos seleccionados componen lo que se denomina **cesta de la compra**, considerada en la encuesta.
3. **Producto Interior Bruto (PIB)**, que registra la producción nacional de un país en bienes y servicios asociados a procesos considerados productivos.
4. **Poder adquisitivo**, que maneja combinadamente datos del Salario Mínimo Interprofesional (SMI) y el IPC.

Gráfico de sectores, o circular, que refleja estadísticamente la aplicación de las matemáticas según el resultado de una encuesta de opinión.



CONCEPTOS BÁSICOS DE ESTADÍSTICA

Vamos a imaginarnos que estamos en un colegio con 200 alumnos y se necesita realizar un estudio estadístico sobre los alumnos que aprueban matemáticas en un año.

Con respecto a los elementos del estudio los conceptos básicos de estadística son:

1. Población.
2. Muestra.
3. Individuo.

Por otro lado, en referencia a los resultados que se obtienen en el estudio tenemos:

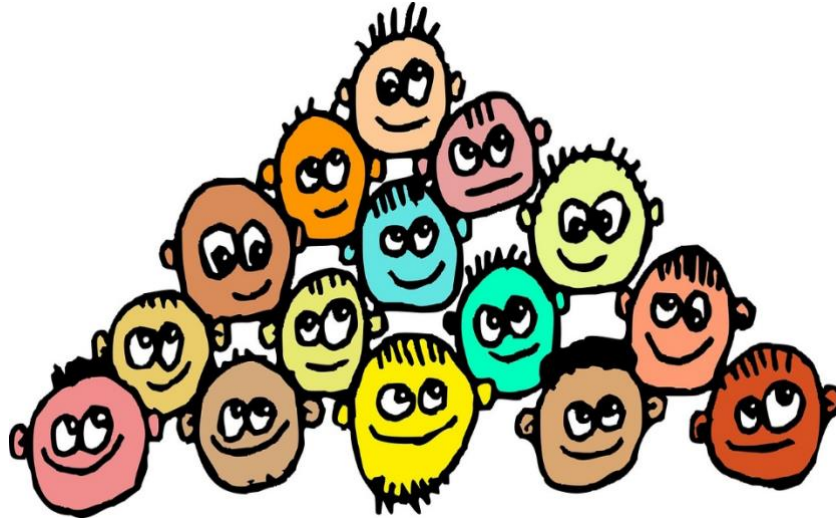
1. Valor.
2. Dato.
3. Muestreo.

Vamos a explicar cada uno de ellos:

POBLACIÓN

A todos los elementos a los que le realizaremos en el estudio se le llama **población**.

En nuestro ejemplo, la población es el conjunto de todos los alumnos, es decir, los 200 alumnos.

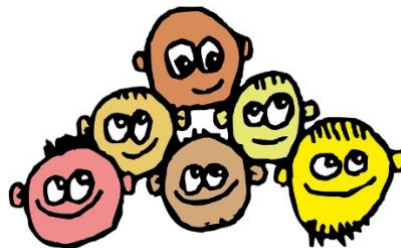


No debes confundir este concepto con la población de una ciudad, por ejemplo. Las poblaciones no tienen por qué ser siempre personas. **Población** son **todos los elementos** a los que le vamos a hacer un estudio, independientemente de lo que sea, ya sean piezas de una fábrica, animales, datos de cualquier tipo.

MUESTRA

Se decide que, de los 200 alumnos, van a escoger solamente a 50 para realizarles una serie de preguntas. Estos 50 alumnos serían una **muestra** de la población, que eran 200 alumnos.

Por tanto, se le llama **muestra** a una parte que es representativa de la población. La **muestra** siempre será más pequeña que la población.



INDIVIDUO

A cada uno de los alumnos del colegio, estadísticamente hablando se les llaman individuos.



Las muestras y las poblaciones están formadas por individuos.

La muestra está formada por 50 individuos, es decir, por 50 alumnos y la población por 200 individuos.

VALOR

El **valor** es el **resultado** que puede cada uno de los datos del estudio.

En nuestro ejemplo, estamos realizando un estudio sobre la cantidad de alumnos que aprueban las matemáticas.

Entonces, podemos tener dos valores diferentes por cada dato:

1. Sí aprueba.
2. No aprueba.

DATO Y VARIABLE

Se le llama **dato** a cada uno de los valores obtenidos después de realizar el estudio estadístico y **variable** al **tipo de dato**, que son una determinada característica de la población (número de hijos, estatura, peso, color, profesión, etc).

Por ejemplo, empezamos estudiando a 3 alumnos y obtenemos estos datos:

1. Alumno 1: Aprobado
2. Alumno 2: Suspenso
3. Alumno 3: Suspenso

En total, 1 «Aprobado» y 2 «Suspenso».

Tenemos 3 datos y cada uno de los datos puede tomar 2 valores («Aprobado» o «Suspenso»). La variable estudiada en este caso es la calificación de la asignatura.

Dentro de una población se pueden recoger datos de diferentes variables. Por ejemplo, a los vecinos de un edificio se les puede preguntar su peso, el color de ojos, si son hombre o mujer

MUESTREO

Se le llama muestreo al conjunto de datos obtenidos de la muestra.

En nuestro estudio, el conjunto de 50 datos, de los 50 alumnos de la muestra sería el muestreo.

Ahora ya conoces los **conceptos básicos de estadística**, o por lo menos ya te suenan más que antes de leer esto. Estos conceptos los verás repetidos en los ejercicios de estadística, por lo que ya estás un paso más cerca de saber resolver tu ejercicio.

TIPOS DE VARIABLES ESTADÍSTICAS

Ejemplos:

Vamos a ver ahora los tipos de variables estadísticas que existen según la forma de expresar sus valores. Tenemos:

1. Variables cuantitativas (numéricas).
 - Discreta (valores aislados).
 - Continua (todos los valores).
2. Variables cualitativas (no numéricas).

Vamos a ver cada una de ellas más despacio.

Variables cuantitativas: discretas y continuas

Las **variables cuantitativas** son las que se **expresan con números** o cantidades.

Una variable cuantitativa es una **variable discreta** cuando sólo admite valores aislados, es decir, no hay ninguna cantidad intermedia.

Por ejemplo, el número de hijos puede ser 1, 2, 3... pero no puedes tener un valor intermedio.

Una variable cuantitativa es una **variable continua** cuando puede tomar cualquier valor entre un intervalo de valores cualquiera. Por ejemplo, la estatura de una persona puede ser 1,85 m o el peso puede ser 76,8 kg. No tienen por qué ser valores concretos.

Variables cualitativas

Las **variables cualitativas** no se expresan con números, sino mediante una **cualidad**.

Podemos distinguir 3 tipos de variables cualitativas:

1. **Variable cualitativa nominal:** Son variables cuyas categorías que no siguen ningún criterio de orden. Por ejemplo: colores de ojos (verdes, marrones, azules...), profesiones (abogado, médico, carpintero...), etc.
2. **Variable cualitativa ordinal:** Sus categorías siguen un orden. Por ejemplo, calificaciones (suspense, aprobado, notable, sobresaliente), puesto en una carrera (primero, segundo, tercero...), etc.
3. **Variable cualitativa binaria:** Sólo tiene dos categorías. Por ejemplo, chico y chica, abierto y cerrado, correcto e incorrecto...

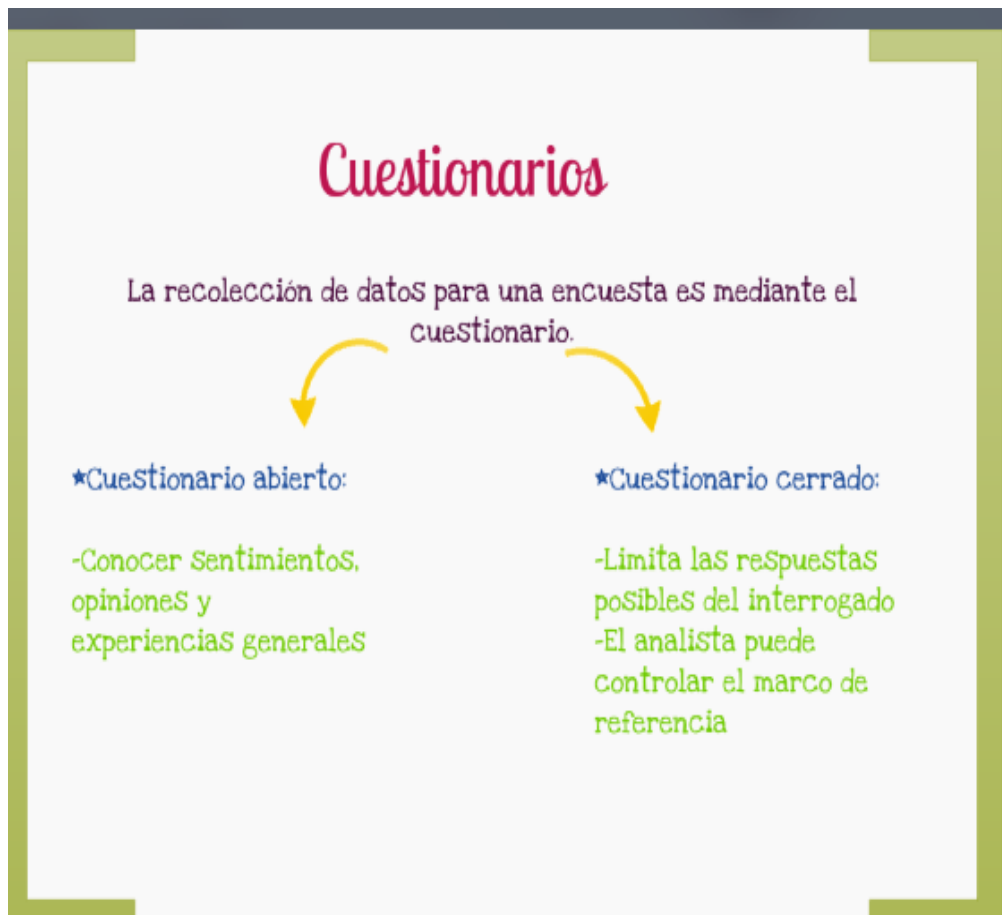


RECOLECCIÓN DE DATOS

La recolección de datos se refiere al uso de una gran diversidad de técnicas y herramientas que pueden ser utilizadas por el analista para desarrollar los sistemas de información, los cuales pueden ser la entrevista, la encuesta, el cuestionario, la observación, el diagrama de flujo y el diccionario de datos.

Todos estos instrumentos se aplicarán en un momento en particular, con la finalidad de buscar información que se útil a una investigación en común.

TÉCNICAS DE RECOLECCIÓN DE DATOS ESTADÍSTICOS



La Observación

Otra técnica útil para el analista en su progreso de investigación, consiste en observar a las personas cuando efectúan su trabajo.

Como técnica de investigación, la observación tiene amplia aceptación científica



La Entrevista

Las entrevistas se utilizan para recabar información en forma verbal. La estructura de la entrevista varía dependiendo de el tipo de información que se requiera.





Una **encuesta** es un procedimiento dentro de los diseños de una investigación descriptiva en el que el investigador recopila datos mediante un cuestionario previamente diseñado, sin modificar el entorno ni el fenómeno donde se recoge la información ya sea para entregarlo en forma de tríptico, gráfica o tabla. Los datos se obtienen realizando un conjunto de preguntas normalizadas dirigidas a una muestra representativa.

DISTRIBUCIÓN DE FRECUENCIAS

Llamamos distribución de frecuencias al conjunto de valores que puede presentar una variable junto con sus frecuencias. Según la naturaleza de la variable estudiada las distribuciones de frecuencias pueden ser:

NO AGRUPADAS: se presentan cuando el número de valores que puede presentar la variable no es muy elevado, y en ese caso podemos observar todos los valores de esa variable. Este caso se presenta cuando la variable es discreta y no presenta excesivos valores.

AGRUPADAS EN INTERVALOS: se presenta cuando la variable es continua o cuando es discreta, pero con elevado número de valores. en esta situación se agrupan dichos valores en intervalos o clases. Los intervalos se notan: $ei-1$ - ei es intervalo i -ésimo.

Se llama amplitud del intervalo a la distancia que existe entre los extremos, y se nota ai :

$$ai = ei - ei-1$$

Se llama marca de clase al punto medio de un intervalo. Este punto es importante porque es el representante del intervalo. Se nota xi :

$$xi = (ei + ei-1)/2$$

Se llama densidad de frecuencia de un intervalo a la frecuencia correspondiente a cada unidad de la variable en dicho intervalo, se nota di :

$$di = ni / ai$$

Los intervalos se suelen tomar abiertos por la izquierda y cerrados por la derecha, salvo el primero que se toma cerrado por los dos lados.

En este tipo de distribuciones se pierde parte de la información al agruparlas en intervalos, ya no se puede hablar de valores concretos sino de intervalos. Cuanto mayor sea la amplitud de los intervalos menos intervalos habrá, y por tanto menos precisión tendremos. En cambio, cuanto menor sea la amplitud de los intervalos menos intervalos habrá, y mayor será la precisión, sin embargo la distribución será más grande y más difícil de manejar.

Ejemplo:

Las dos primeras actividades dan lugar a distribuciones de frecuencia no agrupadas, ya que son variables discretas y presentan pocos valores, 11 en la primera actividad y 4 la segunda.

la tercera actividad da lugar a una distribución de frecuencia agrupada en intervalos ya que aunque es una variable discreta, presenta muchos valores, entre el 1,58 que es el valor más pequeño que presenta, y el 1,85 que es el más grande, hay 27 valores. La elección de los intervalos depende de nosotros, teniendo en cuenta que siempre es preferible que los intervalos sean todos de la misma amplitud.

TABLAS ESTADÍSTICAS

Ya hemos introducido la terminología adecuada, ahora vamos a utilizarla para ordenar y agrupar la información. Lo primero que vamos a hacer es construir tablas estadísticas, en las que va a aparecer toda la información de forma ordenada.

Llamamos tabla estadística a la disposición de forma ordenada y agrupada de los valores y frecuencias de una distribución. Distinguiremos entre tablas estadísticas de distribuciones no agrupadas y tablas de distribuciones agrupadas.

TABLAS DE DISTRIBUCIONES NO AGRUPADAS.

En las tablas de distribuciones no agrupadas aparecen las siguientes columnas: la primera contiene los valores de la distribución, ordenados de menor a mayor si son caracteres cuantitativos; la segunda contiene las frecuencias absolutas, la tercera las frecuencias relativas. Cuando las frecuencias acumuladas se pueden definir se añaden otras dos columnas, una para las frecuencias absolutas acumuladas y otra para las relativas acumuladas.

Ejemplo:

TABLA ESTADÍSTICA DE LA ACTIVIDAD 1ª, TOMANDO COMO POBLACIÓN TODA LA CLASE:

xi	f	fa	fr	Fra
0	2	2	0,04	0,04
1	3	5	0,06	0,10
2	6	11	0,12	0,22
3	6	17	0,12	0,34
4	3	20	0,06	0,40
5	5	25	0,10	0,50
6	5	30	0,10	0,60
7	8	38	0,16	0,76
8	6	44	0,12	0,88
9	4	48	0,08	0,96
10	2	50	0,04	1

Una vez construida la tabla es muy fácil responder a las tres primeras preguntas:

¿Cuántos alumnos han sacado un tres? La respuesta es n_3 que vale 6.

¿Cuántos alumnos han suspendido? La respuesta es N_4 que vale 20.

¿Cuántos alumnos han aprobado? La respuesta es $50 - N_4$ que vale $50 - 20 = 30$.

Para responder a la última pregunta: ¿Han aprobado más alumnos o alumnas? se podrían construir dos tablas una tomando como población los alumnos y otra tomando como población las alumnas y comparar las frecuencias acumuladas de la modalidad 5.

TABLA ESTADÍSTICA DE LA 2ª ACTIVIDAD TOMANDO COMO POBLACIÓN TODA LA CLASE

X_i	F	fr
Rubio	6	0,18
Pelirrojo	1	0,04
Moreno	12	0,36
Castaño	14	0,32
	33	

Con esta tabla es fácil responder a las dos primeras preguntas: El color de pelo que tiene menos gente es el pelirrojo que sólo hay 1 y el que tiene más gente es el castaño que lo tienen 14 alumnos.

Para responder a las otras preguntas vamos a construir la tabla correspondiente a considerar sólo los alumnos y la tabla correspondiente a considerar sólo las alumnas.

TABLA DE ALUMNOS

x_i	n_i	f_i
Rubio	2	0,13
Pelirrojo	1	0,07
Moreno	6	0,4
Castaño	6	0,4

TABLA DE ALUMNAS

x_i	n_i	f_i
Rubia	4	0,23
Pelirroja	0	0
Morena	6	0,33
Castaña	8	0,44

Ahora es muy fácil viendo estas dos tablas responder a las dos últimas preguntas de esta actividad: Hay más niñas morenas, 6, que rubias, 4. Y hay más niñas rubias, 4, que niños, 2.

TABLAS DE DISTRIBUCIÓN AGRUPADAS EN INTERVALOS

En las tablas estadísticas de distribuciones de frecuencia agrupadas por intervalos aparecen las siguientes columnas: la primera con los intervalos, la segunda con las amplitudes de los intervalos, la tercera con las marcas de clase, la cuarta con las frecuencias absolutas de cada intervalo, la quinta con las densidades de frecuencia y la sexta con las frecuencias relativas; además suelen aparecer también dos columnas más con las frecuencias acumuladas. Si se toman los intervalos con la misma amplitud no se ponen las columnas de amplitud ni de densidades, sólo se señala cual es la amplitud de todos los intervalos.

Responda las preguntas que se le hacen a continuación.

1. ¿Cuál es el intervalo de la tabla anterior? _____
2. ¿Cuál es la fórmula para sacar el límite real inferior? _____
3. ¿Cuál es la fórmula para sacar el Límite real superior? _____
4. ¿Cuál es la fórmula para trabajar Fr _____
5. ¿Cuál es la fórmula para trabajar Xi _____

EJERCICIO 1.

Con los datos que se le dan elabore proceso estadístico completo con datos agrupados, calcule rango, struger, arme sus intervalos, elabore tabla de frecuencias, graficas lineales, barras y circular. Luego elabore un informe en relación

a la nota promedio obtenida, cuántos alumnos están por debajo de la nota promedio y cuántos están por debajo de la nota promedio:

DATOS DESORDENADOS

67	37	74	63	58
76	66	79	58	60
68	73	78	56	82
38	73	61	65	39
47	68	67	25	58
77	76	55	42	39
51	79	73	81	66
78	54	78	46	48
70	77			

DATOS ORDENADOS

25	37	38	39	39
42	46	47	48	51
54	55	56	58	58
58	60	61	63	65
66	66	67	67	68
68	70	73	73	73
74	76	76	77	77
78	78	78	79	79
81	82			

Instrucciones: elabore tabla de frecuencias no agrupadas.

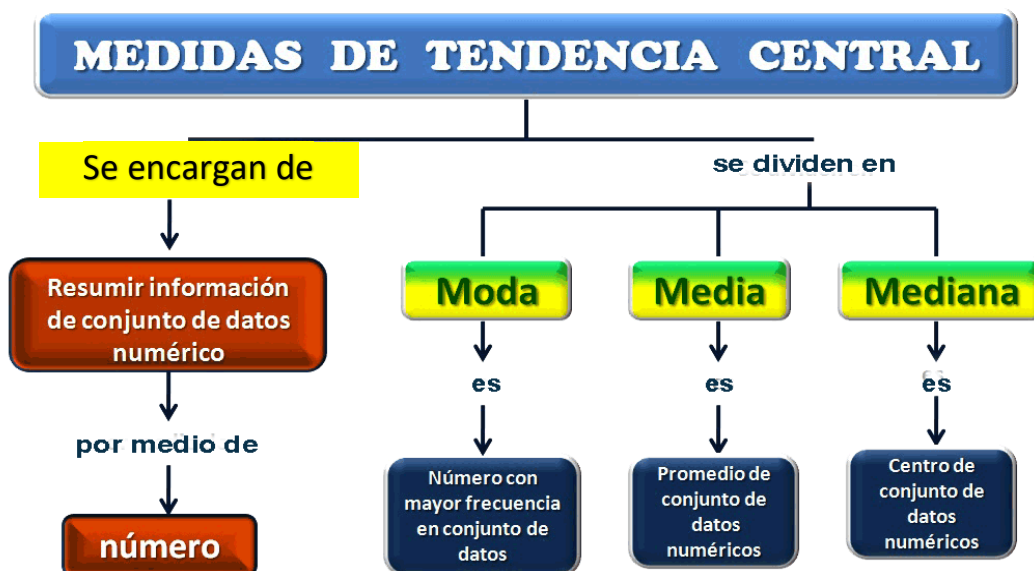
25, 32, 20, 21, 29, 26, 30, 25, 19, 22, 17, 28, 30, 21, 55

Instrucciones: responda las preguntas que le hacen a continuación.

- Se hace una encuesta a los estudiantes de una universidad para determinar qué clase de música es la más popular.
 - Población _____
 - Muestra _____
- En una ciudad se realiza un estudio sobre la cantidad de personas que usan diariamente el transporte público.
 - Población _____
 - Muestra _____

Instrucciones: ordene datos luego elabore tabla de frecuencias límites inferiores, superiores, marca de clase .

61	80	82	62	79	72	82	69
44	35	46	62	85	100	83	57
79	63	73	43	82	82	79	74
70	91	94	82	71	42	42	51



CÁLCULO DE PROMEDIO DATOS NO AGRUPADOS

Ejemplo ilustrativo:

1. Calcular la media aritmética de las siguientes calificaciones de Estadística tomadas de una muestra de 20, sin agrupar, agrupando en tablas de frecuencias y agrupando en intervalos.

4, 8, 10, 10, 5, 10, 9, 8, 6, 8, 10, 8, 5, 7, 4, 4, 8, 8, 6 y 6

Solución:

Sin agrupar

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{4 + 8 + 10 + 10 + 5 + 10 + 9 + 8 + 6 + 8 + 10 + 8 + 5 + 7 + 4 + 4 + 8 + 8 + 6 + 6}{20}$$

$$\bar{x} = \frac{144}{20} = 7,2$$

Ejemplo 2:

$$\bar{X} = \frac{2 + 2 + 4 + 4 + 8 + 10}{6} = \frac{30}{6}$$

$$\bar{X} = 6$$

CÁLCULO DE MEDIANA DATOS NO AGRUPADOS

- a) Si el número n de datos es impar, la mediana es el dato que se encuentra a la mitad de la lista. Para calcular su posición se aplica la siguiente ecuación:

Ejemplo ilustrativo:

Calcular la mediana de las siguientes calificaciones del curso de Estadística evaluadas sobre diez: 10, 8, 6, 4, 9, 7, 10, 9 y 6

Solución:

1) Se ordena los datos de menor a mayor:

4	6	6	7	8	9	9	10	10
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9

2) Se aplica la ecuación:

$$Md = x_{\frac{n+1}{2}}$$

$$Md = x_{\frac{9+1}{2}} = Md = x_5$$

Ejemplo ilustrativo:

Dados los siguientes 20 números:

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

- 1) Agrupar los datos en tabla de frecuencia.
- 2) Calcular la mediana.

x	f
1	1
2	3
3	2
4	4
5	8
6	2
Total	20

Solución:

Calculando la posición de la mediana se obtiene:

$$Md = \frac{n+1}{2} = \frac{20+1}{2} = 10,5$$

CALCULO DE MODA DATOS NO AGRUPADOS

Ejemplo ilustrativo N° 1

Determinar la moda del conjunto de datos 2, 4, 6, 8, 8 y 10

Solución:

Mo = 8, porque es el dato que ocurre con mayor frecuencia. A este conjunto de datos se le llama unimodal

Ejemplo ilustrativo N° 2

Determinar la moda del conjunto de datos: 2, 4, 6, 8 y 10

Solución:

Este conjunto de datos no tiene moda, porque todos los datos tienen la misma frecuencia.

Por lo tanto, se le denomina amodal

Ejemplo ilustrativo N° 3

Determinar la moda del conjunto de datos: 2, 4, 6, 6, 8, 8 y 10

Solución:

Este conjunto de datos tiene dos modas, 6 y 8, y se llama bimodal.

CALCULO DE PROMEDIO DATOS AGRUPADOS

Edad	Marca clase (X_i)	Frecuencia absoluta (f_i)	Frecuencia acumulada (F_i)
[0 - 10)	5	3	3
[10 - 20)	15	6	9
[20 - 30)	25	7	16
[30 - 40)	35	12	28
[40 - 50)	45	3	31

$$N = 31$$

$$\bar{X} = \frac{5 \times 3 + 15 \times 6 + 25 \times 7 + 35 \times 12 + 45 \times 3}{31} =$$

$$\bar{X} = \frac{15 + 90 + 175 + 420 + 135}{31} = \frac{835}{31} = 26,94$$

$$\bar{X} = 26,94$$

Si los datos vienen agrupados en una tabla de frecuencias, la expresión de la media es:

$$\bar{X} = \frac{X_1 f_1 + X_2 f_2 + X_3 f_3 + \dots + X_n f_n}{N}$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i f_i}{N}$$

R/. 43,33

Ejemplo:

	x_i	f_i	$x_i \cdot f_i$
[10, 20)	15	1	15
[20, 30)	25	8	200
[30, 40)	35	10	350
[40, 50)	45	9	405
[50, 60)	55	8	440
[60, 70)	65	4	260
[70, 80)	75	2	150
		42	1 820

CALCULO DE MEDIANA DATOS AGRUPADOS

Edad	Marca clase (X_i)	Frecuencia absoluta (f_i)	Frecuencia acumulada (F_i)
[0 - 10)	5	3	3
[10 - 20)	15	6	9
[20 - 30)	25	7	16
[30 - 40)	35	12	28
[40 - 50)	45	3	31

$$N = 31$$

$$N = 31$$

Ahora reemplazamos los datos en la fórmula:

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

$$Me = 20 + \frac{15,5 - 9}{7} \cdot 10$$

$$Me = 20 + 9,29$$

$$Me = 29,285$$

CALCULO DE MODA DATOS AGRUPADOS

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot t_i$$

$$Mo = 30 + \frac{12 - 7}{(12 - 7) + (12 - 3)} \cdot 10$$

$$Mo = 30 + 3,57$$

$$Mo = 33,6$$

EJERCICIO.

1. Calcular la media aritmética, la mediana y la moda de la siguiente serie de números: 5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5, 4.
2. Las puntuaciones obtenidas por un grupo en una prueba han sido: 15, 13, 16, 15, 19, 18, 15, 14, 18. Calcular la moda, la mediana y la media aritmética.
3. El número de estrellas de los hoteles de una ciudad viene dado por la siguiente serie: 3, 3, 4, 3, 4, 3, 1, 3, 4, 3, 3, 3, 2, 1, 3, 3, 3, 2, 3, 2, 2, 3, 3, 3. Hallar la moda, la mediana y la media aritmética.
4. Las calificaciones de 36 alumnos en Matemáticas han sido las siguientes: 5, 2, 4, 9, 7, 4, 5, 6, 5, 7, 7, 5, 5, 8, 2, 10, 5, 6, 10, 4, 7, 6, 7, 3, 5, 6, 9, 6, 1, 4, 6, 3, 5, 5, 6, 7. Calcular la moda, la mediana y la media aritmética.

5. En un estudio que se realizó en un asilo de ancianos, se tomó las edades de los envejecientes que pueden caminar sin dificultades. Buscar la media, la mediana y la moda de las siguientes edades. 69 73 65 70 71 74 65 69 60 62.

6. Se tiene las notas de 11 alumnos en un examen de matemática: **10 ; 12 ; 09 ; 12 ; 08 ; 14 ; 12 ; 10 ; 11 ; 12 ; 08.**

A) ¿Cuál es la moda?

a) 8 b) 10 c) 11 d) 12 e) 9

B. ¿Cuál es la mediana?

a) 9 b) 10,5 c) 10 d) 11 e) 12

C. ¿Cuál es la mediana de las notas restantes?

a) 10,5 b) 10 c) 11 d) 12 e) 11,5

7. Se tiene a continuación las edades de 20 alumnos de la I.E ROGERIO VELASQUEZ **16 18 20 21 19 19 20 18 17 18 21 16 21 19 16 16 17 18 16 18** se puede decir entonces que la moda es:

A) Unimodal B) Bimodal C) Amodal D) Trimodal E) Multimodal

8. calcule las medidas de tendencia central de los datos agrupados.

Intervalo (Horas)	Frecuencia Absoluta (ni)
11-15	9
16-20	20
21-25	13
26-30	10
31-35	13
36-40	18
41-45	9
46-50	8
51-55	0
56-60	0
TOTAL	100

EJERCICIO.

Instrucciones: con los datos que se le dan determine si las medidas de tendencia central coinciden con la respuesta. Si fueran falsas coloque la respuesta correcta.

- a. 2,2,2,2,2= Md 2
- b. 3,5,2,7,9= promedio 8
- c. 1,1,0,2,2,= promedio y Md son iguales
- d. 3,3,1,3,3,3= promedio y Md son iguales
- e. 2,3,4,5,6,7= promedio y Md son iguales

Instrucciones: calcule las medidas de tendencia central.

25, 32, 20, 21, 29, 26, 30, 25, 19, 22, 17, 28, 30, 21, 55

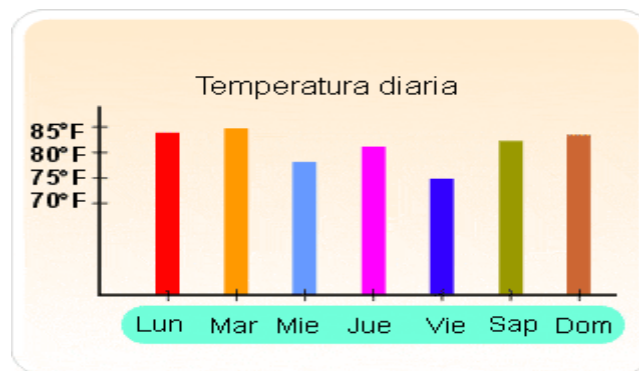
Instrucciones: ordene datos luego elabore tablas compuestas y determine las medidas de tendencia central.

61	80	82	62	79	72	82	69
44	35	46	62	85	100	82	57
79	63	73	43	82	82	79	74
70	91	94	82	42	42	42	51

Instrucciones: Determine las medidas de tendencia central con los datos que se dan en la gráfica.



Instrucciones: en base a la gráfica que se le da responda las preguntas.



1. ¿Cuál es el promedio? _____
2. ¿Cuál es la moda? _____
3. ¿La mediana es igual al promedio? _____
4. ¿Quién es Xi? _____
5. ¿Cuál es el valor de n? _____

EJERCICIO.

Instrucciones: ordena la serie de datos, elabore una tabla de frecuencias simples y calcule la media, la mediana, la moda y elabore una gráfica circular, lineal y de barras.

75, 69, 50, 85, 89, 60, 75, 69, 85, 85, 85, 89, 60, 75, 69, 69, 85, 85.

1. Ordene los datos y agrúpelos, elabore tabla de frecuencias, límites reales inferiores, límites reales superiores, marca clase, por último elabore histograma.

51, 28, 68, 72, 71, 80, 49, 53, 80, 31, 40
 24, 23, 33, 38, 21, 43, 47, 64, 49, 50, 48
 63, 57, 54, 44, 58, 65, 79, 71, 76, 80, 76
 80, 78, 57, 44, 80, 48, 65

2. Complete tabla y luego calcule medidas de tendencia central por fórmula

intervalos	f
21-30	4
31-40	4
41-50	8
51-60	6
61-70	5
71-80	12
suma	39

3. Ordene los datos y agrúpelos, elabore tabla de frecuencias, límites reales inferiores, límites reales superiores, marca clase, por último elabore histograma.

35, 18, 32, 40, 8, 12, 58, 35, 42, 15, 18, 36, 45
 60, 56, 45, 40, 36, 32, 35, 48, 55, 28, 30, 18, 15
 32, 40, 45, 45

4. Ordene los datos y agrúpelos, elabore tabla de frecuencias, límites reales inferiores, límites reales superiores, marca clase, por último elabore histograma.

146, 147, 148, 150, 151, 151, 151, 151, 151, 153, 153, 154, 156, 156, 156, 156, 157, 157, 157, 158, 158, 159,
 160, 161, 161, 161, 162, 163, 165, 165
 166, 166, 166, 169, 171, 172, 173, 173, 174, 175, 176, 176

EL GRÁFICO

A un nivel estadístico y matemático, denominamos "gráfica" a **aquella representación visual a partir de la cual pueden representarse e interpretarse** valores generalmente numéricos. De entre las múltiples informaciones extraíbles de la observación de la gráfica podemos encontrar la existencia de relación entre variables y el grado en que se da, las frecuencias o la proporción de aparición de determinados valores.

Esta representación visual sirve de apoyo a la hora de mostrar y comprender de manera sintetizada los datos recabados durante la investigación, de manera que puede tanto los investigadores que llevan a cabo el análisis como otros **puedan comprender los resultados y resulte sencillo utilizarlo como referencia**, como información a tener en cuenta o como punto de contraste ante la realización de nuevas investigaciones y meta análisis.

TIPOS DE GRÁFICAS

Existen muy diversos tipos de gráficas, generalmente aplicándose unas u otras en función de lo que se pretenda representar o simplemente de las preferencias del autor. A continuación, indicamos algunas de las más conocidas y comunes.

1. **Gráfico de barras.** El más conocido y utilizado de todos los tipos de gráficos es el gráfico o diagrama de barras. En éste, se presentan los datos en forma de barras contenidas en dos ejes cartesianos (coordenada

y abscisa) que indican los diferentes valores. **El aspecto visual que nos indica los datos es la longitud de dichas barras**, no siendo importante su grosor.

2. **Gráfico circular o por sectores.** El también muy habitual gráfico en forma de “quesito”, en este caso la representación de los datos se lleva a cabo mediante la división de un círculo en tantas partes como valores de la variable investigada y teniendo cada parte **un tamaño proporcional a su frecuencia dentro del total de los datos**. Cada sector va a representar un valor de la variable con la que se trabaja.

Este tipo de gráfico o diagrama es habitual cuando se está mostrando la proporción de casos dentro del total, utilizando para representarlo valores porcentuales (el porcentaje de cada valor).

3. **Histograma.** Aunque a simple vista muy semejante al gráfico de barras, el histograma es uno de los tipos de gráfica que a nivel estadístico resulta más importante y fiable. En esta ocasión, también se utilizan barras para indicar a través de ejes cartesianos la frecuencia de determinados valores, pero en vez de limitarse a establecer la frecuencia de un valor concreto de la variable evaluada refleja todo un intervalo. Se observa pues un rango de valores, que además **podrían llegar a reflejar intervalos de diferentes longitudes**.

Ello permite observar no solo la frecuencia sino también la dispersión de un continuo de valores, lo que a su vez puede ayudar a inferir la probabilidad. Generalmente se utiliza ante variables continuas, como el tiempo.

4. **Gráfico de líneas.** En este tipo de gráfico se emplean líneas para **delimitar el valor de una variable dependiente respecto a otra independiente**. También puede usarse para comparar los valores de una misma variable o de diferentes investigaciones utilizando el mismo gráfico (usando diferentes líneas). Es usual que se emplee para observar la evolución de una variable a través del tiempo.

5. **Gráfico de dispersión.** El gráfico de dispersión o gráfico xy es un tipo de gráfico en el cual mediante los ejes cartesianos se representa en forma de puntos todos los datos obtenidos mediante la observación. **Los ejes x e y muestran cada uno los valores de una variable dependiente y otra independiente** o dos variables de la que se esté observando si presentan algún tipo de relación.

Los puntos representados el valor reflejado en cada observación, lo que a nivel visual dejará ver una nube de puntos a través de los cuales podemos observar el nivel de dispersión de los datos.

6. **Gráfico de caja y bigotes.** Los gráficos de caja son uno de los tipos de gráficas que tienden a utilizarse de cara a observar la dispersión de los datos y cómo éstos agrupan sus valores. Se parte del cálculo de los cuartiles, los cuales son los valores que **permiten dividir los datos en cuatro partes iguales**. Así, podemos encontrar un total de tres cuartiles (el segundo de los cuales se corresponderían con la mediana de los datos) que van a configurar la “caja” en cuestión. Los llamados bigotes serían la representación gráfica de los valores extremos.

Este gráfico **es útil a la hora de evaluar intervalos**, así como de observar el nivel de dispersión de los datos a partir de los valores de los cuartiles y los valores extremos.

7. **Gráfico de áreas.** En este tipo de gráfico se observa, de manera semejante lo que ocurre con los gráficos de líneas, la relación entre variable dependiente e independiente. Inicialmente **se hace una línea que une los puntos que marcan los diferentes valores de la variable** medida, pero también se incluye todo lo situado por debajo: este tipo de gráfica nos permite ver la acumulación (un punto determinado incluye a los situados por debajo).

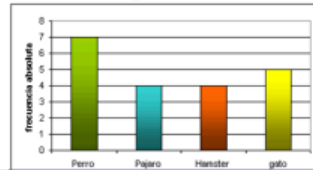
A través de él se pueden medir y comparar los valores de diferentes muestras (por ejemplo, comparar los resultados obtenidos por dos personas, compañías, países, por dos registros de un mismo valor). Los diferentes resultados pueden apilarse, observándose fácilmente las diferencias entre las diversas muestras.

8. **Pictograma.** Se entiende por pictograma a un gráfico en el que, en vez de representar los datos a partir de elementos abstractos como barras o círculos, **se emplean elementos propios del tema que se está investigando**. De este modo se hace más visual. Sin embargo, su funcionamiento es semejante al del gráfico de barras, representando frecuencias de la misma manera.

9. **Cartograma.** Este gráfico resulta de utilidad en el terreno de la epidemiología, indicando las zonas o áreas geográficas en las que aparece con mayor o menor frecuencia un determinado valor de una variable. Las

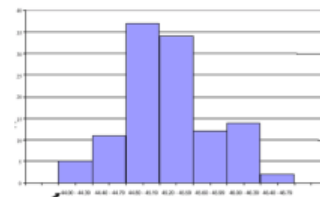
frecuencias o rangos de frecuencias se indican mediante el uso del color (requiriéndose una leyenda para comprenderse) o el tamaño.

Tipos de gráficas



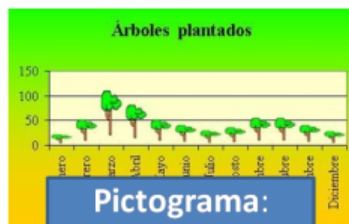
De barras:

Se usa para clases sin relación cuantitativa. Las barras se separan a la misma distancia.



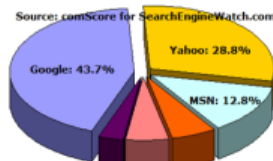
Histograma:

Barras verticales unidas por su relación numérica. En el eje X van las clases, en el Y va la frecuencia.



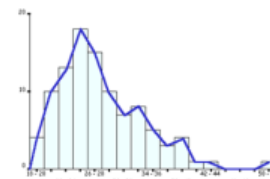
Pictograma:

Representan magnitudes con figuras.



Circular o de pastel:

Solo datos cuantitativos.
Porcentaje de la frecc.
Relativa.

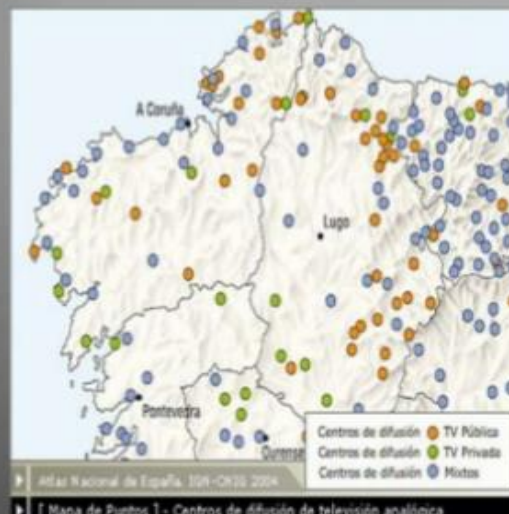


Polígono de frecuencias:

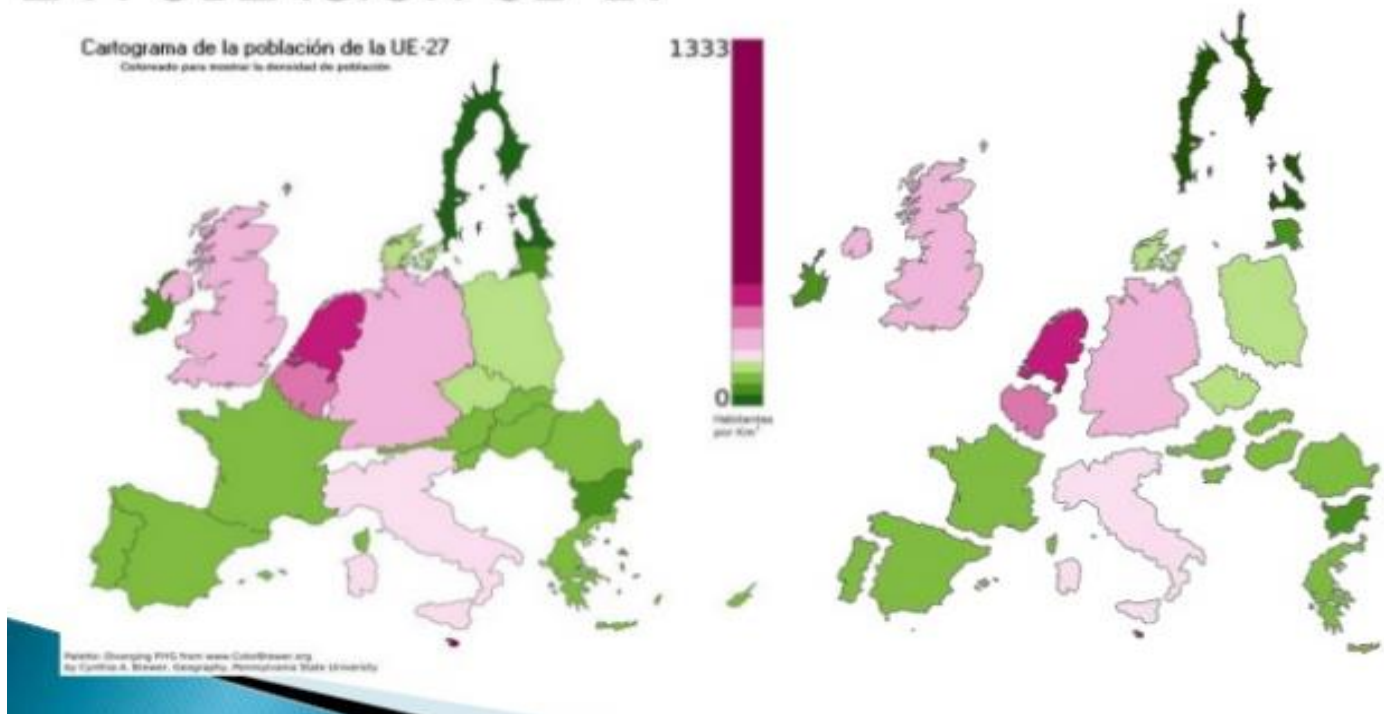
Distribución de frecc.
Unidas por puntos.

CARTOGRAMAS DE VARIAS VARIABLES

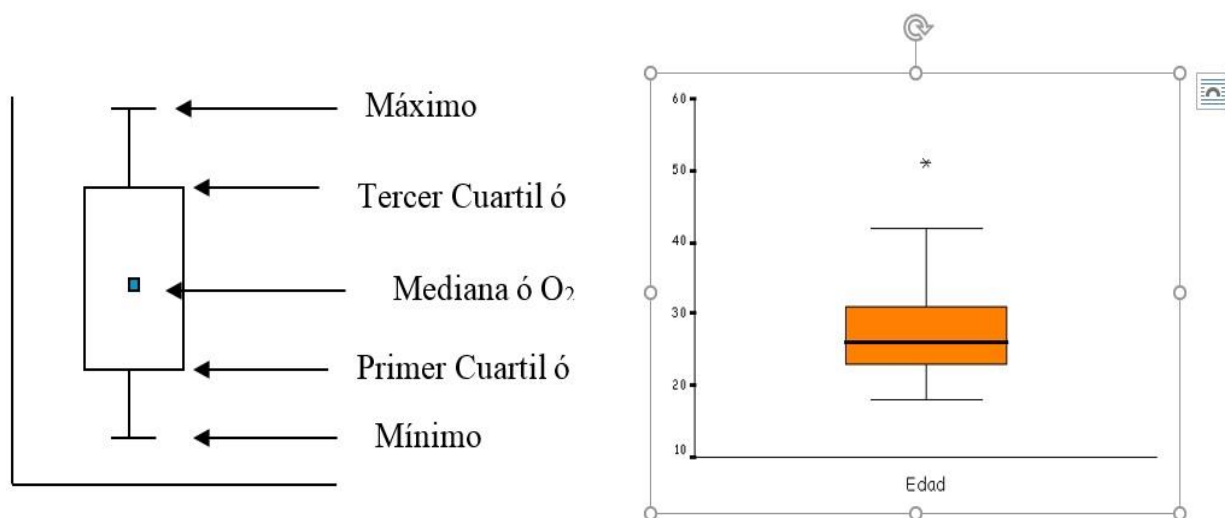
Si realizamos un cartograma en el que las unidades de enumeración tienen un tamaño proporcional a su población, podemos representar sobre el otras variables como en un mapa de coropletas.

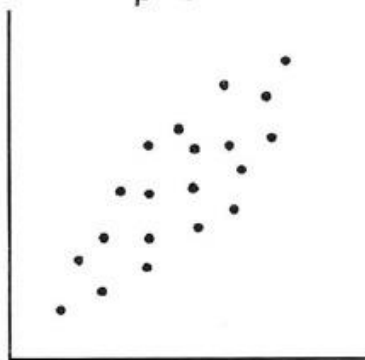
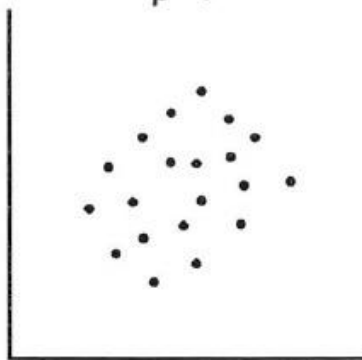
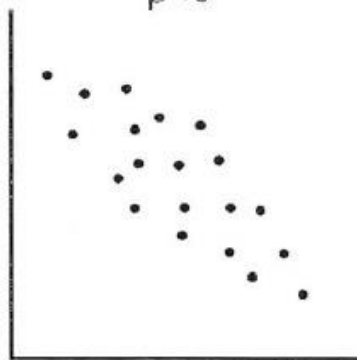


CARTOGRAMA CON Y SIN CONTIGUIDAD DE LA POBLACION UE-27



Gráfica de Cajas y Bigotes.



CORRELACION POSITIVA
 $\rho > 0$ CORRELACION NULA
 $\rho = 0$ CORRELACION NEGATIVA
 $\rho < 0$ CORRELACION IDEAL POSITIVA
 $\rho = 1$ CORRELACION IDEAL NEGATIVA
 $\rho < -1$ 