

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

---

### EXPECTATIVAS DE LOGRO:

- Informarse sobre el aporte y la importancia de la estadística en las disciplinas científicas que componen la carrera.
- Identificar los distintos términos estadísticos.
- Distinguir las diferentes clasificaciones de datos estadísticos.

### CONTENIDOS:

- 1.1-Proceso de investigación estadística. Estadística descriptiva. Noción de estadística inferencial.
- 1.2-VARIABLES estadísticas. Variables: Cualitativas, cuantitativas. Variables discretas y variables continuas. Unidades de estudio. Escala de Medición. Noción de Población y muestra- Métodos de muestreo: Probabilístico y No Probabilístico.
- 1.3-Recopilación y Ordenamiento de datos.
- 1.4-Tratamiento descriptivo de los valores observados de una variable cuantitativa.
- 1.5-Construcción de tablas de frecuencias. Frecuencias absolutas y relativas. Frecuencias acumuladas
- 1.6-Gráficos: de bastones, escalonados, histogramas, polígonos de frecuencias, ojivas

### RESEÑA HISTÓRICA

Históricamente el crecimiento y desarrollo de la estadística se remonta a tres fenómenos: la necesidad de los gobiernos de recopilar datos acerca de sus ciudadanos, el desarrollo matemático de las teorías de probabilidad, el surgimiento de la computadora y la evolución de la tecnología.

Desde que los pueblos se organizaron como estado sus gobernantes han recopilado datos sobre algunos aspectos relativos a la cantidad o distribución de la población, producción agrícola o ganadera, bienes muebles o inmuebles con el propósito primordial de cobrar impuestos y reclutar soldados. En la Edad media la iglesia llevaba registros de nacimientos, muertes, bautismos y casamientos. Se atribuye a Achenwall (1719-1772) la introducción del nombre Estadística para designar a esta "ciencia de las cosas que pertenecen al Estado". En nuestro país, se registró el primer censo durante la presidencia de D. F. Sarmiento, repitiéndose cada 10 años.

La tecnología computacional ha transformado radicalmente el campo de la estadística en los últimos 40 años. En nuestros días, la estadística se ha convertido en un método efectivo para describir con exactitud los valores de datos económicos, políticos, sociales, psicológicos, biológicos y físicos, y sirve como herramienta para relacionar y analizar dichos datos. El trabajo del estadístico no consiste ya sólo en reunir y tabular los datos, sino sobre todo en el proceso de **interpretación** de esa información. El desarrollo de la teoría de la probabilidad ha aumentado el alcance de las aplicaciones de la estadística.

### ¿POR QUÉ UN DISEÑADOR NECESITA CONOCER ESTADÍSTICA?

Hace un siglo H. G. Wells comentó: " Algún día el pensamiento estadístico será tan necesario como la habilidad para leer y escribir". Al comienzo del nuevo milenio los profesionales cuentan con una gran cantidad de información, dado el crecimiento exponencial en las comunicaciones; siendo uno de los mayores problemas cómo utilizar esa información disponible para tomar las decisiones más adecuadas.

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

---

Desde la perspectiva de la toma de decisiones informada, cabe preguntarse ¿por qué un Diseñador necesita saber estadística? Un profesional debe comprender conceptos básicos de estadística por las siguientes razones fundamentales:

- Para presentar y describir la información en forma adecuada.
- Para obtener conclusiones sobre poblaciones grandes basándose solamente en la información obtenida de las muestras para poder realizar sus diseños.
- Para comprender y mejorar los procesos.
- Para obtener pronósticos confiables.

Es común emplear métodos estadísticos en las distintas áreas que contribuyen a la producción y comercialización de un producto.

En Ergonomía, la búsqueda de datos y su procesamiento estadístico se encuentra presente en las fases de planificación y diseño de nuevos productos, como también en el rediseño de los ya existentes. El uso de tablas antropométricas donde el problema de la dispersión de las medidas del cuerpo humano exige que el Diseñador Industrial adopte criterios de diseño basados en conceptos estadísticos

En finanzas, se utilizan para elegir entre carteras alternativas de inversión y para detectar las tendencias en las medidas financieras a través del tiempo.

En administración, se recurre a los métodos estadísticos para mejorar la calidad de los productos manufacturados de los servicios que presta alguna organización. A partir de la década del 50 se desarrolló un enfoque que hace hincapié en la calidad y la mejora continua en los productos y servicios dando una gran importancia a la mejora de los procesos y a la optimización del sistema completo.

En mercadotecnia, se aplica la estadística para estimar la proporción de clientes que prefieren un producto en lugar de otro, además para saber el porqué de su preferencia. También se utiliza para obtener conclusiones y determinar la estrategia publicitaria que resultará más útil para aumentar las ventas de un producto.

### **ESTADÍSTICA**

**DEFINICIÓN:** *La Estadística como ciencia, es la rama de la Matemática que estudia los métodos para recopilar, organizar, resumir y analizar, e interpretar una gran cantidad de datos, así como sacar conclusiones válidas y tomar decisiones razonables basadas en tal análisis. Su objetivo es el **análisis de datos y la interpretación de resultados** y no el cálculo por sí mismo.*

**La Estadística es una forma de pensamiento bajo incertidumbre. Este surge cuando se analizan variables aleatorias (sus valores no se pueden predecir con exactitud)**

Los métodos estadísticos se aplican a cualquier área del comportamiento humano, donde el análisis de datos numéricos permite realizar un proceso de toma de decisión.

Dentro de la Estadística podríamos considerar dos grandes ramas:

- Estadística descriptiva
- Estadística inferencial.

## **ESTADÍSTICA DESCRIPTIVA**

La necesidad de recopilar datos, ordenarlos, presentarlos y analizarlos tuvo una relación estrecha con el desarrollo de la estadística descriptiva.

***La estadística descriptiva puede definirse como los métodos y procedimientos que involucran la recopilación, presentación y análisis de un conjunto de datos, con el fin de describir en forma apropiada varias de sus características.***

Su objetivo es examinar a todos los individuos de un conjunto y explorar el conjunto de datos para poder descifrar lo que expresan. Para poder interpretar esa información contenida en los datos de una forma más sencilla y presentable posible, se obtienen así los parámetros que distinguen las características de un conjunto de datos (lo que se conoce como estadísticos). Pertenecen al ámbito de la estadística descriptiva las tablas de frecuencias, a partir de las cuales se obtienen los estadísticos:

- Medidas de centralización: la media en todas sus variantes (aritmética, geométrica, ponderada), la moda y la mediana
- Medidas de dispersión: la varianza, la desviación típica (raíz cuadrada de la varianza) y el rango
- Medidas de tendencia central: los cuantiles y sus desgloses (percentiles, cuantiles, deciles, etc.)
- Medidas de forma: son aquellas que nos muestran si una distribución de frecuencia tiene características especiales como simetría, asimetría, nivel de concentración de datos y nivel de apuntamiento que la clasifiquen en un tipo particular de distribución.
- Entre otras.

## **ESTADÍSTICA INFERENCIAL**

Si bien, los métodos de la estadística descriptiva son importantes para caracterizar y presentar los datos, el desarrollo de los métodos de inferencia estadística es lo que llevó a su aplicación en todos los campos de la investigación. El desarrollo inicial de la teoría de la probabilidad surgido de la investigación de los juegos de azar durante el renacimiento, fue aprovechado a comienzo del siglo XX, cuando especialistas en estadística como Pearson, Fisher, Gosset, Neyman y otros, iniciaron el desarrollo de inferencia estadística, que en la época actual tiene una amplia aplicación en muchos campos.

***La inferencia estadística se define como aquellos métodos que hacen posible la estimación de una característica de una población, o la toma de una decisión con respecto a una población basada solo en resultados muestrales.***

**Por tanto, un aspecto importante de inferencia estadística es usar métodos estadísticos muestrales, para obtener conclusiones respecto a los parámetros de la población.** Provee conclusiones o inferencias, basándose en los datos simplificados y analizados; detectando las interrelaciones que pueden unirlos, las leyes que los rigen y eliminando las influencias del azar; llegando más allá de las verificaciones físicas posibles. **Sobre la base de la muestra estudiada saca conclusiones, o sea, hace inferencia o inducción, en cuanto al universo o población, de donde se obtuvo dicha muestra.**

### ¿QUÉ ES LO QUE DISTINGUE A LA ESTADÍSTICA DESCRIPTIVA DE LA INFERENCIAL?

En primer lugar, lo que las distingue es la **naturaleza de los datos**. Mientras que la estadística descriptiva sirve tanto para una población como para una muestra (un subconjunto de esa población cuyos elementos son elegidos al azar), en la estadística inferencial se trabaja con muestras a partir de las cuales intenta extraer conclusiones sobre la población.

Esta práctica se conoce como inferir, y es importante recalcar la diferencia en la naturaleza de los datos, ya que es un error muy común, el de extraer conclusiones de un conjunto, cuyas conclusiones, son los mismos datos en sí. Para explicar dicho error, conviene explicar la **principal diferencia** teórica entre estadística descriptiva e inferencial. La **descriptiva**, al ser únicamente una descripción de los datos, no asume que éstos tengan alguna propiedad más allá de las que **se pueden describir con los estadísticos** ya mencionados. En cambio, **la inferencial** asume que los datos se rigen bajo un fenómeno aleatorio subyacente que es el que hace que tomen un valor u otro. Es por esto por lo que los datos pasarían a denominarse **variables aleatorias**. Al existir incertidumbre, se puede igualmente describir la población de la que sale esa muestra, pero debemos entonces asumir un cierto error derivado de la naturaleza probabilística de los datos.

Un ejemplo práctico: si recogemos una muestra de alturas de 100 argentinos, y obtenemos una media de 1,85, podemos asumir que es una variable aleatoria, y que por lo tanto, si la media de la muestra está en 1,85, es muy probable que la media de altura de todos los argentinos esté en torno a esa cifra. Por otro lado, si hacemos un censo de las alturas de todos los habitantes de Argentina, no hace falta asumir un riesgo o error para concluir con que la media es equivalente a una cifra concreta: al 100%, la media poblacional es la que obtenemos de ese censo.



La necesidad del uso de los métodos de inferencia estadística se deriva en la utilización de **muestreo**. A medida que una población crece, por lo general resulta demasiado costoso, lento y tedioso obtener información de toda la población. Las decisiones respecto a las características de la población deben basarse en la información que está contenida en una muestra de la población.

### POBLACIÓN

Se llama **población** (o universo) a todo conjunto de elementos, finito o infinito, definido por una o más características, de las que gozan todos los elementos que lo componen, y sólo ellos. Cada uno de esos elementos es un individuo.

En estadística se usa la palabra población para referirse no sólo a personas sino a todos los elementos que han sido escogidos para su estudio.

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

Cada **individuo** puede ser descrito mediante uno o varios **caracteres**. Por ejemplo, si los individuos son personas, el sexo, el estado civil, el número de hermanos o su estatura son caracteres. Y si el individuo es una reacción química, el tiempo de reacción, la cantidad de producto obtenido o si éste es ácido o básico serán posibles caracteres que pueden analizarse.

La población puede ser finita o infinita.

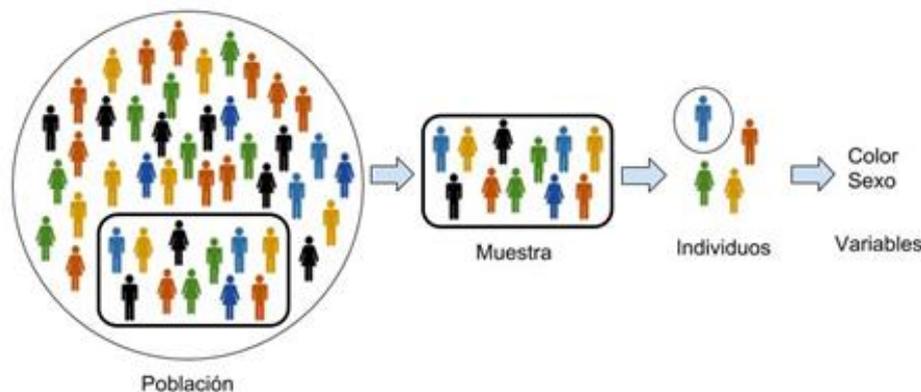
- **Una población finita es aquella que puede ser físicamente listada.**
- **Una población infinita es aquella que, en la práctica, no puede ser listada.** Ejemplo: En investigaciones biológicas, la mayoría de las poblaciones con que se trabaja son consideradas infinitas a causa de la imposibilidad física de obtener un listado de todos sus elementos. Los niños con problemas visuales en un determinado radio, la población de un determinado insecto que ataca a los cultivos, la cantidad de vacas de una cierta raza productora de leche, todos ellos son ejemplos de poblaciones infinitas y, por lo tanto hipotéticas.

Un **parámetro** es una medida de resumen que describe una característica de toda una población

### MUESTRA

Se dice que se está trabajando con una **muestra**, cuando se considera solamente un subconjunto de individuos u objetos de cierta población que componen una parte representativa de dicha población.

#### INTERPRETACIÓN GRÁFICA DE LOS CONCEPTOS DE INDIVIDUO – POBLACIÓN - MUESTRA



### MÉTODOS DE MUESTREO

El propósito de un estudio estadístico suele ser, extraer conclusiones acerca de la naturaleza de una población. Al ser la población grande y no poder estudiarla en su integridad en la mayoría de los casos, las conclusiones obtenidas deben basarse en el examen de una parte de ésta, lo que nos lleva, en primer lugar a la justificación, necesidad y definición de las diferentes técnicas de muestreo.

### MUESTREO

Es la actividad por la cual se toman ciertos subconjuntos de la población, éste será tomado siguiendo ciertos criterios de decisión, el muestreo es importante porque a través de él podemos hacer análisis de situaciones de una población haciendo uso de la inferencia estadística.

*En estadística*, el muestreo, es el proceso por el cual se seleccionan los individuos que formarán una muestra. Para que se puedan obtener conclusiones fiables para la población a partir de la muestra, es

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

importante **tanto su tamaño** como **el modo** en que han sido seleccionados los individuos que la componen.

### **TERMINOLOGÍA BÁSICA PARA EL MUESTREO**

Los nuevos términos, los cuales son frecuentemente usados en inferencia estadística son:

- **ESTADÍSTICO:** es una medida usada para describir alguna característica de una muestra, tal como una media aritmética, una mediana- o una desviación estándar de una muestra, los estadísticos emplean letras latinas minúsculas para denotar estadísticas y muestras.
- **PARÁMETRO:** es una medida usada para describir alguna característica de una población, tal como una media aritmética, una mediana o una desviación estándar de una población.

Cuando los dos nuevos términos de arriba son usados, por ejemplo, el proceso de estimación en inferencia estadística puede ser descrito como el proceso de estimar un parámetro a partir del estadístico correspondiente, tal como usar una media muestral (un estadístico) para estimar la media de la población (un parámetro).

- **TAMAÑO DE LA MUESTRA** depende de la precisión que se quiera conseguir en la estimación que se realice a partir de ella. Para su determinación se requieren técnicas estadísticas superiores, pero resulta sorprendente cómo, con muestras notablemente pequeñas, se pueden conseguir resultados suficientemente precisos.
- **UNIDAD DE MUESTREO:** Elemento o grupo de elementos que se toman de una población para un muestreo simple o multietápico (que puede combinar dos o más diseños muestrales).
- **MARCO DE MUESTREO.** Una lista física de todas las unidades de una población o un procedimiento para producir resultados comparables a los de una lista completa.

### **RAZONES PARA OBTENER UNA MUESTRA**

1. *Una muestra requiere menos tiempo que un censo.*
2. *El costo de una muestra es menor que la recolección de información de una población*
3. *En algunos casos el elemento elegido de la muestra debe ser sometidos a ensayos destructivos; por lo tanto es imprescindible el muestreo.*

### **CONVENIENCIA DEL MUESTREO**

El muestreo es conveniente cuando constituye la solución de mayor eficacia, es decir que se obtenga información empleando recursos mínimos o recursos prefijados.

Se extraerán muestras en los siguientes casos:

1. Cuando la población sea infinita o sin serlo se haga imposible el estudio de la población.
2. Cuando la población sea suficientemente uniforme, o bien su homogeneidad permita que cualquier muestra dé una buena información de la población.
3. Cuando el proceso de información sea destructivo de los elementos de estudio disminuyan su valor o utilidad.

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

4. Cuando los recursos son limitados (económicos o de tiempo); Cuando la población es muy grande suele ser muy costoso obtener la información de todos los individuos, como así también se requiere de tiempo para la planificación de recolectar la información y procesamiento de la misma.
5. Se puede mejorar la precisión de la observación en una muestra, ya que se pueden usar instrumentos más adecuados, personal más entrenado y mejor supervisado.

### LIMITACIONES DEL MUESTREO

Uno de los riesgos del muestreo consiste en inferir una característica de la población a partir de una muestra, ya que equivale a obtener conclusiones en la ignorancia parcial de dicha característica.

El muestreo implica menor trabajo en la obtención de la información, pero exige mayor refinamiento y preparación para el correcto empleo de la inferencia estadística.

### MÉTODOS DE SELECCIÓN DE MUESTRAS.

Una muestra debe ser representativa si va a ser usada para estimar las características de la población. Los métodos para seleccionar una muestra representativa son numerosos, dependiendo del tiempo, dinero y habilidad disponibles para tomar una muestra y la naturaleza de los elementos individuales de la población. Por lo tanto, se requiere un gran volumen para incluir todos los tipos de métodos de muestreo.

Estos métodos se pueden clasificar **según la forma de obtener la muestra, o según la cantidad de muestras obtenidas.**

### SEGÚN LA FORMA DE OBTENER LOS DATOS

Existen dos tipos básicos de muestreo: **No Probabilístico** y **Probabilístico**.

El siguiente cuadro describe esta clasificación.



### MÉTODOS DE MUESTREO NO PROBABILÍSTICOS.

A veces, para estudios exploratorios, el muestreo probabilístico resulta excesivamente costoso y se acude a métodos no probabilísticos, aun siendo conscientes de que no sirven para realizar generalizaciones a toda la población, pues no se tiene certeza de que la muestra extraída sea representativa, ya que no todos los sujetos de la población tienen la misma probabilidad de ser elegidos. En general se seleccionan a los sujetos siguiendo determinados criterios procurando que la muestra se distribuya uniformemente. Este tipo de muestras tiene algunas ventajas como son conveniencia, rapidez y bajo costo. Por otro lado, dos

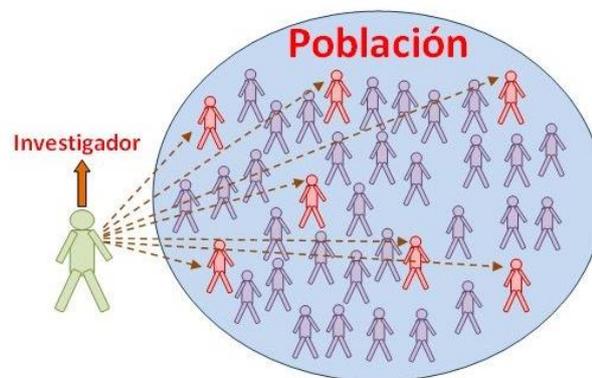
## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

desventajas importantes: falta de precisión debido al sesgo en la selección y falta de capacidad para generalizar los resultados.

Los muestreos No Probabilísticos se clasifican en:

- **MUESTREO SUBJETIVO O DE JUICIO:** cuando el investigador selecciona a los individuos a través de su criterio profesional. También, puede basarse en la experiencia de otros estudios anteriores o en su conocimiento sobre la población y el comportamiento de éste frente a las características que se estudian. Una muestra de juicio es llamada una muestra no probabilística, puesto que este método está basado en los puntos de vista subjetivos de una persona y la teoría de la probabilidad no puede ser empleada para medir el error de muestreo. Las principales ventajas de una muestra de juicio son la facilidad de obtenerla y que el costo usualmente es bajo.

Es aconsejable cuando la población es reducida y el investigador sabe con precisión que la muestra que utilizará será útil para el estudio porque es conocida por él mismo.



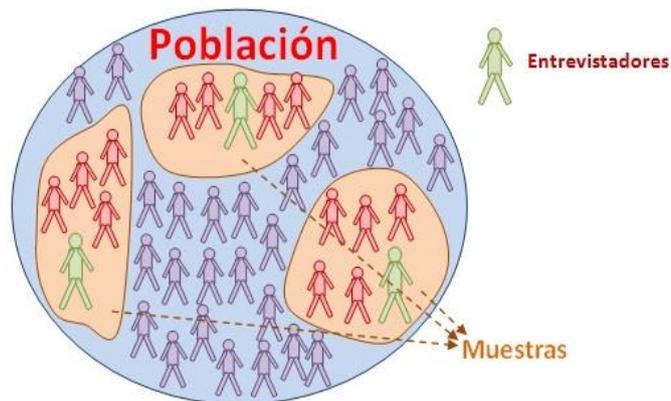
- **MUESTREO POR CUOTAS:** También denominado en ocasiones "accidental" y consiste en predeterminar la cantidad de elementos de cada categoría que la integran. Se asienta generalmente sobre la base de un buen conocimiento de los estratos de la población y/o de los individuos más "representativos" o "adecuados" para los fines de la investigación. Mantiene, por tanto, semejanzas con el muestreo aleatorio estratificado, pero no tiene el carácter de aleatoriedad de aquél.

En este tipo de muestreo se fijan unas "cuotas" que consisten en un número de individuos que reúnen unas determinadas condiciones, por ejemplo: 20 individuos de 25 a 40 años, de sexo femenino y residentes en Santa Lucía. Una vez terminada la cuota se eligen los primeros que se encuentren que cumplan esas características. Este método se utiliza mucho en las encuestas de opinión. Ej. Medición de rating televisivo.

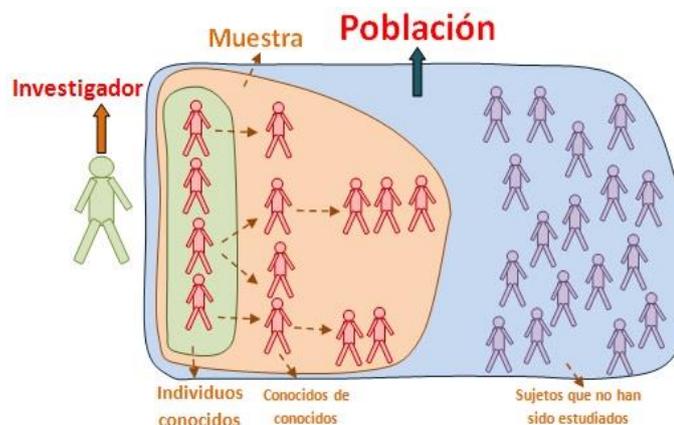


## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

- **MUESTREO CASUAL O INCIDENTAL:** En algunos casos también llamado por conveniencia. Se trata de un proceso en el que el investigador selecciona directa e intencionalmente los individuos de la población, dando por supuesto, que ésta será representativa de la población de referencia. El caso más frecuente de este procedimiento es el utilizar como muestra los individuos a los que se tiene fácil acceso (los profesores de universidad emplean con mucha frecuencia a sus propios alumnos para saber si entendieron los conceptos dados en la clase anterior, realiza una pregunta un caso particular es el de los voluntarios).



- **BOLA DE NIEVE:** Se localiza a algunos individuos, los cuales conducen a otros, y estos a otros, y así hasta conseguir una muestra suficiente. Este tipo se emplea muy frecuentemente cuando se hacen estudios con poblaciones "marginales", delincuentes, sectas, determinados tipos de enfermos, etc.



### MÉTODOS DE MUESTREO PROBABILÍSTICOS O ALEATORIO.

Una muestra se dice que es extraída al azar, cuando la manera de selección es tal, que cada elemento de la población **tiene igual oportunidad de ser seleccionado**. Una muestra aleatoria es también llamada una muestra probabilística.

Los elementos de la muestra se eligen de acuerdo a las probabilidades conocidas. La importancia de este muestreo se debe a dos razones la primera permite realizar inferencias estadísticas objetivas y además permite separar y evaluar objetivamente las fuentes de error.

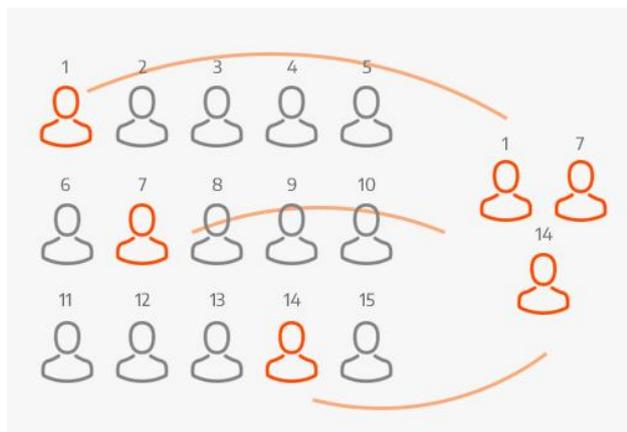
## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

Los muestreos Probabilísticos se clasifican en:

- **MUESTREO ALEATORIO SIMPLE (m.a.s.):** El muestreo aleatorio simple es el método de selección de  $n$  unidades de una población de tamaño  $N$  de tal modo que cada una de las muestras posibles tenga la misma oportunidad de ser elegida.

Para obtener una muestra aleatoria simple se enumeran las unidades de la población de 1 a  $N$  y posteriormente se extrae una serie de  $n$  números aleatorios entre 1 y  $N$  (tarea que se puede realizar usando una tabla de números aleatorios o mediante un programa de computación que produce una tabla semejante). Las unidades cuya numeración coincide con la serie de números seleccionados conformarán la muestra aleatoria. La probabilidad de seleccionar a cualquier miembro en particular de la población la primera vez es  $1/N$ .

Ejemplo: Se quiere conocer el comportamiento de una población conformada por 15 estudiantes de la Asignatura Estadística de la Carrera de Diseño Industrial de la UNSJ, para extraer una muestra el docente se vale de la lista a la cual a cada elemento de la población se le asigna un número. Eligiendo al azar el alumno 1, 7 y 14 de la lista.

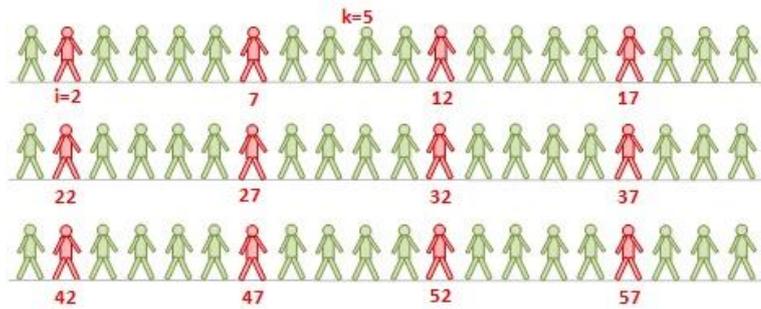


- **MUESTREO SISTEMÁTICO:** Una muestra sistemática es obtenida cuando los elementos son seleccionados en una manera ordenada. La manera de la selección depende del número de elementos incluidos en la población y el tamaño de la muestra. El número de elementos en la población es, primero, dividido por el número deseado en la muestra. El cociente indicará si cada décimo, cada onceavo, o cada centésimo elemento en la población va a ser seleccionado.

El primer elemento de la muestra es seleccionado al azar. Por lo tanto, una muestra sistemática puede dar la misma precisión de estimación acerca de la población, que una muestra aleatoria simple cuando los elementos en la población están ordenados al azar.

Ejemplo: De una población de 60 individuos deseo obtener una muestra de 12, al realizar el cociente obtenemos que la selección de la muestra se realizará cada 5 individuos. El primero es seleccionado es al azar, en este caso sería el individuo número 2 y a partir de este, selecciono los elementos de la muestra cada 5 individuos, obteniendo el número 7, 12, 17,... y así continuando hasta completar la muestra.

UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO.  
ANÁLISIS EXPLORATORIO.



- **MUESTREO ESTRATIFICADO:** Para obtener una muestra aleatoria estratificada, primero se divide la población en grupos, llamados estratos, que son más homogéneos que la población como un todo. Los elementos de la muestra son entonces seleccionados al azar o por un método sistemático de cada estrato. Las estimaciones de la población, basadas en la muestra estratificada, usualmente tienen mayor precisión (o menor error muestral) que si la población entera es muestreada mediante muestreo aleatorio simple. El número de elementos seleccionados de cada estrato puede ser proporcional o desproporcional al tamaño del estrato en relación con la población.

Los estratos (subconjuntos de unidades muestrales) deben tomarse de manera tal que los valores de la variable sean más homogéneos dentro de los mismos que entre ellos. Al existir un patrón de estratificación de las unidades respecto a la variable en estudio, se obtiene una caracterización más precisa de dicha variable, con muestreo estratificado que aplicando un **m.a.s.**

La principal ventaja de este procedimiento es que los resultados tienen mayor precisión, porque disminuye la variabilidad de los datos dentro de cada estrato. A mayor homogeneidad en los datos, mayor precisión en la estimación.



Ejemplo: Supongamos que hay en una fábrica 25 directivos, 130 empleados con algún cargo de mando y 913 operarios. El tamaño de la muestra a seleccionar es de 120 personas.

Esta selección puede ser proporcional a cada estrato. Es decir la suma de todos los trabajadores da un total de 1068 personas. Por simple regla de 3 se tiene:

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

$$E1(\text{directivos}) = \frac{25}{1068} \times 120 = 2,8 \cong 3 \text{ directivos}$$

$$E2(\text{Empleados}) = \frac{130}{1068} \times 120 = 14,6 \cong 15 \text{ empleados}$$

$$E3(\text{Operarios}) = \frac{913}{1068} \times 120 = 102,6 \cong 102 \text{ Completar}$$

Los 3 directivos, los 15 empleados t los 102 operarios se seleccionaran usando un M.A.S.

- **MUESTREO POR CONGLOMERADOS O ÁREAS:** Se aconseja su uso cuando por razones de costos, de practicidad u otras consideraciones no es conveniente enumerar y acceder a cada unidad de muestreo, y éstas se pueden agrupar en subconjuntos, denominados conglomerados, de forma tal que haya heterogeneidad entre las unidades de un mismo conglomerado y homogeneidad entre conglomerados. El concepto de homogeneidad entre conglomerados se refiere a que las medidas que se pueden calcular para cada conglomerado y difieren poco de uno a otro.

Al existir un patrón de conglomerados de las unidades muestrales, se obtiene una estimación más precisa si se muestrean aleatoriamente un número determinado de conglomerados y se censan todas las unidades muestrales que los constituyen que si se realizara un M.A.S.

Al igual que en el muestreo por estratos, la información de los distintos conglomerados incluidos en la muestra debe combinarse para obtener la información deseada; una forma es promediando la información de cada uno de ellos. Por ejemplo, si se desea estimar la producción de un cultivo en un departamento de la Provincia de Córdoba, es necesario visitar una gran cantidad de campos. En este caso, hacer un M.A.S. es muy costoso ya que se deben recorrer muchos kilómetros para recaudar la información. Dividiendo el departamento en áreas o conglomerados y seleccionando aleatoriamente algunos de ellos, y luego censándolos (es decir, se visita a todos los productores), se obtendrá la misma información con menor costo operativo. Esta práctica permite revisar más unidades muestrales con el mismo esfuerzo o costo que con un M.A.S., ya que al pertenecer todos los campos a un área determinada, las distancias a recorrer son cortas.

Debe quedar claro que este método presupone que la información dentro de cada conglomerado es más variable que entre conglomerados, de lo contrario no es aconsejado.



### SEGÚN LA CANTIDAD DE MUESTRAS EXTRAÍDAS

Se pueden clasificar en:

- **MUESTREO SIMPLE:** Este tipo de muestreo toma solamente una muestra de una población dada para el propósito de inferencia estadística. Puesto que solamente una muestra es tomada, el tamaño de muestra debe ser lo suficientemente grande para extraer una conclusión. Una muestra grande muchas veces cuesta demasiado dinero y tiempo.
- **MUESTREO DOBLE:** Bajo este tipo de muestreo, cuando el resultado del estudio de la primera muestra no es decisivo, una segunda muestra es extraída de la misma población. Las dos muestras son combinadas para analizar los resultados. Este método permite a una persona principiar con una muestra relativamente pequeña para ahorrar costos y tiempo. Si la primera muestra arroja un resultado definitivo, la segunda muestra puede no necesitarse. Por ejemplo, al probar la calidad de un lote de productos manufacturados, si la primera muestra arroja una calidad muy alta, el lote es aceptado; si arroja una calidad muy pobre, el lote es rechazado. Solamente si la primera muestra arroja una calidad intermedia, será requerirá la segunda muestra. Un plan típico de muestreo doble puede ser el siguiente: Al probar la calidad de un lote consistente de 3,000 unidades manufacturadas, cuando el número de defectos encontrados en la primera muestra de 80 unidades es de 5 o menos, el lote es considerado bueno y es aceptado; si el número de defectos es 9 o más, el lote es considerado pobre y es rechazado; si el número está entre 5 y 9, no puede llegarse a una decisión y una segunda muestra de 80 unidades es extraída del lote. Si el número de defectos en las dos muestras combinadas (incluyendo  $80 + 80 = 160$  unidades) es 12 o menos, el lote es aceptado si el número combinado es 13 o más, el lote es rechazado.
- **MUESTREO MÚLTIPLE:** El procedimiento bajo este método es similar al expuesto en el muestreo doble, excepto que el número de muestras sucesivas requerido para llegar a una decisión es más de dos muestras.

### MÉTODO ESTADÍSTICO

A continuación se analiza someramente las etapas del proceso mediante el cual se puede llegar al enunciado de conclusiones por el camino de la estadística.

El método de investigación estadística se encuentra íntimamente relacionado al **Método Científico** y podemos resumirlo en las siguientes fases:

#### 1. PLANEAMIENTO

Se limita a la redacción de las instrucciones para recabar los datos, definición precisa de los datos que se necesitan; diseño de formularios y planillas, planificación y organización del trabajo en el espacio y en el tiempo.

Esta etapa comprende diversos pasos como son:

- Planteo y formulación del problema.
- Formulación estadística. Descripción de la información, determinación de población y/o muestra, variables, métodos y procedimientos de recolección de información.

## **2. RECOPIACIÓN DE LOS DATOS**

En ella se recaban los datos necesarios para la investigación, mediante encuestas, muestreos, censos, o se toman de fuentes secundarias o registros y publicaciones.

La eficiencia con que se realice esta etapa generará la calidad de todo el trabajo de investigación estadística.

## **3. TABULACIÓN Y AGRUPAMIENTO DE DATOS. GRÁFICOS**

Los datos se suelen presentar de diversas formas: tabular, en cuadros de doble entrada, proporcionales, porcentuales, o en valores promedios. También suelen representarse gráficamente, mediante: Histogramas, polígonos, diagramas figurados, pictogramas, dibujos acotados, entre otros.

## **4. MEDICIÓN DE DATOS**

En esta etapa comienza la elaboración matemática y el procesamiento de los datos.

Se obtienen ciertos parámetros o medidas de posición, como así también se analiza la dispersión respecto de estas medidas de posición. Se definen entonces los parámetros o medidas de dispersión.

## **5. ANÁLISIS E INTERPRETACIÓN DE LA INFORMACIÓN.**

Esta etapa consiste en obtener conclusiones:

- a) Si se trabajó con la población, empleando los Métodos Básicos de la Estadística Descriptiva
- b) Inferencia estadística. Conclusiones referidas a la población. En esta etapa se deducen y aplican las leyes de inferencia que permiten predecir el comportamiento de una población investigada, estudiando solo una parte o muestra de ella.

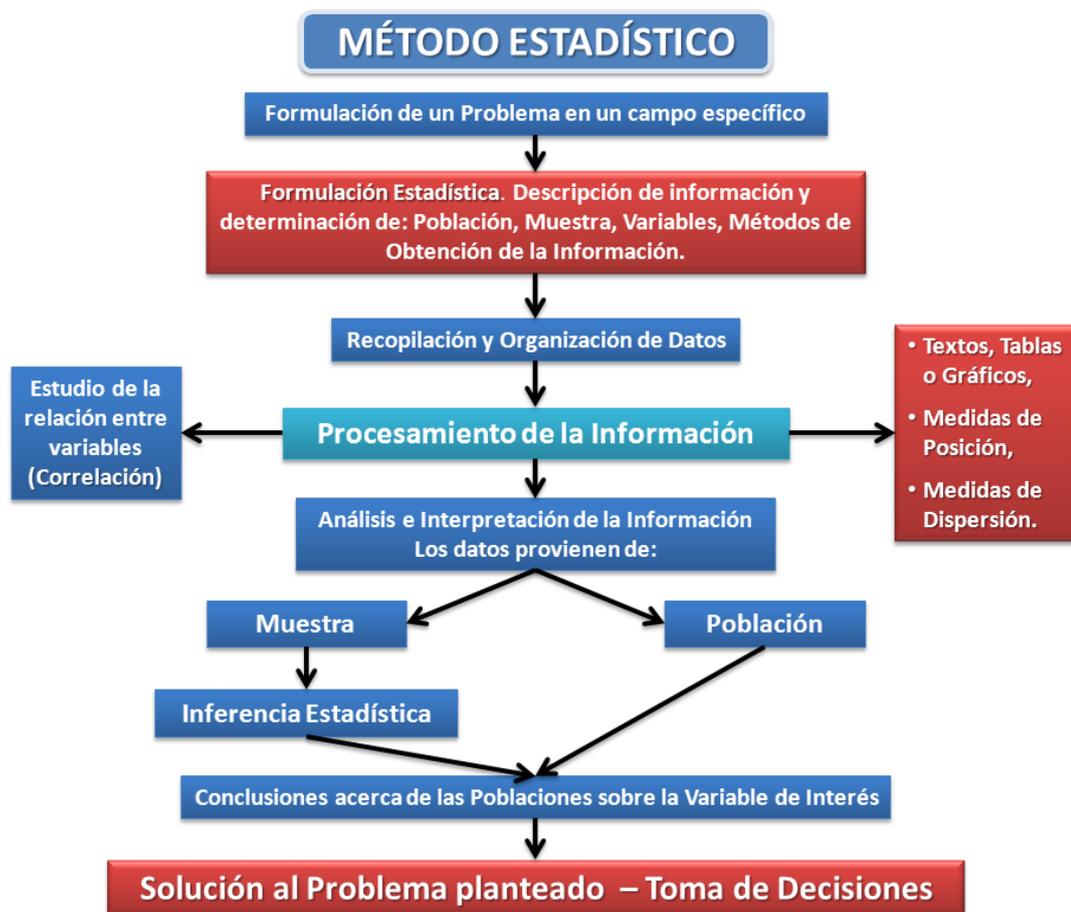
La estadística inferencial provee conclusiones o inferencias, basándose en los datos simplificados y analizados; detectando las interrelaciones que pueden unirlos, las leyes que los rigen y eliminando las influencias del azar; llegando más allá de las verificaciones físicas posibles. Sobre la base de la muestra estudiada saca conclusiones, o sea, hace inferencia o inducción, en cuanto al universo o población, de donde se obtuvo dicha muestra.

- c) El estudio de las relaciones entre las distintas variables. ( Correlación entre variables).

## **6. SOLUCIÓN AL PROBLEMA PLANTEADO Y TOMA DE DECISIONES.**

Tanto en el análisis descriptivo de los datos poblacionales como en la inferencia estadística esta etapa es una de las más importantes ya que la interpretación de los resultados del procedimiento estadístico utilizado, permitirá obtener conclusiones y tomar decisiones acertadas al problema planteado al comienzo de la investigación.

A continuación se presenta un esquema sobre el método estadístico.



### OBTENCIÓN DE DATOS

La obtención de información apropiada es esencial para manejar una investigación. Los datos son la información que permitirá tomar una mejor decisión en una situación específica. Existen muchas situaciones en que los datos son necesarios.

- Un investigador de mercado necesita evaluar las características de los productos para distinguir entre varios de ellos.
- Un fabricante de bombillas eléctricas debe determinar si una nueva bombilla tiene más horas útiles que las anteriores.
- Un administrador desea vigilar un proceso en forma periódica para saber si la calidad del servicio que proporciona (o la calidad de los productos que fabrica) cumple con los estándares de la compañía.
- Un empresario desea revisar la calidad de un producto a fin de averiguar si esta cumple o no los estándares de calidad de fabricación establecidos.
- Un inversionista potencial desea determinar que compañías y cuales industrias tienen posibilidades de crecimiento acelerado en un periodo de recuperación económica.
- Un estudiante quiere obtener datos sobre los grupos de música favoritos de sus compañeros para satisfacer su curiosidad.

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

Basándose en los ejemplos anteriores se dan a continuación algunas razones para recopilar datos.

### **RAZONES PARA OBTENER DATOS**

1. Los datos se necesitan como entrada de una encuesta.
2. Los datos se necesitan como entrada de un estudio.
3. Los datos se necesitan para medir el desempeño de un servicio o un proceso de producción en marcha.
4. Los datos se necesitan para evaluar la conformidad con los estándares
5. Los datos se necesitan en la formulación de estrategias alternativas en un proceso de toma de decisiones.
6. Los datos se necesitan para satisfacer la curiosidad.

Es muy importante empezar el análisis con la identificación de las fuentes de datos más adecuadas.

### **FUENTES DE INFORMACIÓN**

- **Fuentes Primarias:** son aquellas en las que la información se encuentra en su origen, y por lo tanto debe ser elaborada por primera vez y de forma específica
- **Fuentes Secundarias:** son aquellas que presentan información ya elaborada, o existente, que fue generada anteriormente con otra finalidad que no tiene por qué coincidir con la nuestra.

*Los recopiladores de datos son fuentes primarias; los compiladores de datos son fuentes secundarias.*

### **NATURALEZA DE LAS FUENTES DE INFORMACIÓN**

La información a utilizar en una Investigación Estadística debe cumplir con un conjunto de condiciones entre las que se debe destacar:

- **Fiabilidad:** por su fuente (origen), tratamiento y representatividad
- **Homogeneidad:** por la utilización de las mismas variables y criterios en su elaboración que permite la comparación
- **Actualidad:** por el menor desfase temporal con el momento de decisión o con el fenómeno que se desea analizar
- **Pertinente:** que realmente sea útil y adecuada para el análisis a llevar a cabo.

La información debe guardar relación clara con los objetivos del estudio, en caso contrario no es necesaria.

### **FUENTES CLAVES PARA LA OBTENCIÓN DE DATOS**

1. Se pueden obtener datos ya publicados por fuentes gubernamentales, industriales o individuales.
2. Se pueden diseñar experimentos para obtener datos necesarios.
3. Se puede realizar una encuesta.
4. Se puede realizar un estudio observacional.

En nuestro país el organismo gubernamental responsable de recopilar datos sobre empleo, costos, índices, etc. es el Instituto Nacional de Estadísticas y Censos: INDEC. También hay empresas como bancos,

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

---

compañías financieras e industrias que proporcionan datos a sus clientes que les permite comparar sus productos con la competencia.

La segunda fuente para recopilar datos es la **experimentación**. En un experimento se ejerce un control estricto sobre los tratamientos. En general los diseños experimentales adecuados son tema de estudio a nivel más avanzado, pues con frecuencia involucran procedimientos estadísticos elaborados.

La tercera fuente de datos se obtiene de realizar **una encuesta**. En ella se hacen preguntas sobre diferentes características y sus respuestas se editan, codifican y tabulan para su análisis. En la Unidad N° 5 profundizaremos sobre el tema de encuestas y entrevistas.

La cuarta fuente de datos que consiste en el **estudio observacional** ocurre cuando el investigador observa el comportamiento de manera directa, casi siempre en el entorno natural. En muchos campos del conocimiento científico como la astronomía o la geología, la mayor parte del conocimiento se ha desarrollado de esta manera, donde la experimentación o las encuestas son impracticables o imposibles.

**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, la observación y el diagrama de flujo.

Todos estos instrumentos se aplicarán en un momento en particular, con la finalidad de buscar información que será útil a una investigación en común. En la presente investigación trata con detalle los pasos que se debe seguir en el proceso de recolección de datos, con las técnicas ya antes nombradas.

### TÉCNICAS PARA OBTENER DATOS

Los analistas utilizan una variedad de métodos a fin de recopilar los datos sobre una situación existente, como entrevistas, cuestionarios o encuestas, inspección de registros (revisión en el sitio) y observación. Cada uno tiene ventajas y desventajas. Generalmente, se utilizan dos o tres para complementar el trabajo de cada una y ayudar a asegurar una investigación completa.

### ENTREVISTA

Es un intercambio verbal en el cual existen dos roles bien definidos, el del entrevistador y del entrevistado, donde el primero intenta obtener información sobre opiniones, creencias, ideas, actitudes, conocimientos, etc., del segundo y este responde en intenta adecuarse a las demandas del primero. Por lo tanto... la entrevista es un proceso de interacción con un fin determinado. Las entrevistas se utilizan para recabar información en forma verbal, a través de preguntas que propone el analista. Quienes responden pueden ser seleccionados de algunas de las formas ya vistas. El analista puede entrevistar al personal en forma individual o en grupos. Sin embargo, las entrevistas no siempre son la mejor fuente de datos de aplicación.

**Dentro de una investigación social, la entrevista es la técnica más significativa y productiva** de que dispone el analista para recabar datos. En otras palabras, la entrevista **es un intercambio de información** que se efectúa cara a cara.

**UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO.  
ANÁLISIS EXPLORATORIO.**

**VENTAJAS**

| Entrevista Estructurada   | Entrevista No Estructurada   |
|---|--|
| <ul style="list-style-type: none"> <li>• Asegura la elaboración uniforme de las preguntas para todos los que van a responder.</li> <li>• Fácil de administrar y evaluar.</li> <li>• Evaluación más objetiva tanto de quienes responden como de las respuestas a las preguntas.</li> <li>• Se necesita un limitado entrenamiento del entrevistador.</li> <li>• Resulta en entrevistas más pequeñas.</li> </ul> | <ul style="list-style-type: none"> <li>• El entrevistador tiene mayor flexibilidad al realizar las preguntas adecuadas a quien responde.</li> <li>• El entrevistador puede explotar áreas que surgen directamente durante la entrevista.</li> <li>• Puede producir información sobre área que se minimizaron o en las que no se pensó que fueran importantes.</li> </ul> |

**DESVENTAJAS**

| Entrevista Estructurada   | Entrevista No Estructurada   |
|---|--|
| <ul style="list-style-type: none"> <li>• Alto costo de preparación.</li> <li>• Los que responden pueden no aceptar un alto nivel en la estructura y el carácter mecánico de las preguntas.</li> <li>• Un alto nivel en la estructura puede no ser adecuado para todas las situaciones.</li> <li>• El alto nivel en las estructuras reduce responder en forma espontánea, así como la habilidad del entrevistador para continuar con comentarios hacia el entrevistado.</li> </ul> | <ul style="list-style-type: none"> <li>• Puede utilizarse negativamente el tiempo, tanto de quien responde como del entrevistador.</li> <li>• Los entrevistadores pueden introducir sus sesgos en las preguntas o al informar de los resultados.</li> <li>• Puede recopilarse información extraña</li> <li>• El análisis y la interpretación de los resultados pueden ser largos.</li> <li>• Toma tiempo extra recabar los hechos esenciales.</li> </ul> |

**ENCUESTAS**

Nuestra "sociedad", requiere un rápido y preciso flujo de información sobre las preferencias, necesidades y comportamiento de sus miembros. Es en respuesta a esta necesidad crítica de información por el gobierno, el comercio y las instituciones sociales que tanta confianza se pone en las encuestas. Las encuestas proveen una fuente importante de conocimiento científico básico.

Hoy en día la palabra "encuesta" se usa más frecuentemente para describir un método de obtener información de una muestra de individuos. Esta "muestra" es usualmente sólo una fracción de la población bajo estudio. Por ejemplo, un fabricante hace una encuesta al mercado potencial antes de introducir un nuevo producto o una entidad del gobierno comisiona una encuesta para obtener información para evaluar legislación existente o para preparar y proponer nueva legislación.

No tan sólo las encuestas tienen una gran variedad de propósitos, sino que también pueden conducirse de muchas maneras, incluyendo por teléfono, por correo o en persona.

A diferencia de un censo, donde todos los miembros de la población son estudiados, las encuestas recogen información de una porción de la población de interés, dependiendo el tamaño de la muestra en el

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

---

propósito del estudio. La muestra es seleccionada científicamente de manera que cada persona en la población tenga una oportunidad medible de ser seleccionada. De esta manera los resultados pueden ser proyectados con seguridad de la muestra a la población mayor. La información es recogida usando procedimientos estandarizados de manera que a cada individuo se le hacen las mismas preguntas en más o menos la misma manera. La intención de la encuesta no es describir los individuos particulares quienes, por azar, son parte de la muestra sino obtener un perfil compuesto de la población.

El estándar de la industria para todas las organizaciones respetables que hacen encuestas es que los participantes individuales nunca puedan ser identificados al reportar los hallazgos. Todos los resultados de la encuesta deben presentarse en resúmenes completamente anónimos, tal como tablas y gráficas estadísticas.

### Algunas consideraciones para obtener buenos resultados al realizar una encuesta:

- No fiarse de la memoria del entrevistado: las preguntas deben limitarse al pasado inmediato, ayudando al entrevistado a retroceder paso a paso en el tiempo, recordando la información que interesa.
- No hacer preguntas que obliguen al entrevistado a hacer cálculos. Es necesario concretar la pregunta para obtener contestaciones más exactas con mayor facilidad.
- Evitar preguntas embarazosas o formularlas de manera que se logre la respuesta en forma indirecta.
- No hacer preguntas que impliquen la respuesta.
- Insertar un conjunto de posibles respuestas entre las que pueda elegir la que más concuerde con su opinión, para los casos en que la pregunta no pueda ser contestada con exactitud.
- Incluir preguntas que puedan complementarse, de forma tal que el “sí” o “no” de primera pregunta excluya o requiera las siguientes.
- Incluir preguntas control, para controlar la veracidad de las respuestas.
- Evitar preguntas abiertas, porque dan lugar a una gran cantidad situaciones.
- Ordenar las preguntas del cuestionario siguiendo la secuencia lógica del pensamiento.
- No confeccionar cuestionarios extensos: es preferible tener poca información de buena calidad y no mucha de calidad poco satisfactoria.
- Tener cuidado con la redacción.

### ALGUNOS MÉTODOS COMUNES

Las encuestas pueden ser clasificadas en muchas maneras. Una dimensión es por tamaño y tipo de muestra. Las encuestas pueden ser usadas para estudiar poblaciones humanas o no humanas (por ejemplo, objetos animados o inanimados, animales, terrenos, viviendas). Mientras que muchos de los principios son los mismos para todas las encuestas, el foco aquí será en métodos para hacer encuestas a individuos.

Las encuestas pueden ser clasificadas por su método de recolección de datos. Las encuestas por correo, telefónicas y entrevistas en persona son las más comunes. Extraer datos de récords médicos y otros se hace también con frecuencia. En los métodos más nuevos de recoger datos, la información se entra

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

---

directamente a la computadora ya sea por un entrevistador adiestrado o aún por la misma persona entrevistada. Un ejemplo bien conocido es la medición de audiencias de televisión usando aparatos conectados a una muestra de televisores que graban automáticamente los canales que se observan.

**Las encuestas por correo pueden** ser de costo relativamente bajo. Como con cualquier otra encuesta, existen problemas en usar este método si no se presta suficiente atención a obtener niveles altos de cooperación. Estas encuestas pueden ser más efectivas cuando se dirigen a grupos particulares, tal como suscriptores a una revista especializada o a miembros de una organización profesional.

**Las encuestas por internet**, son la herramienta del momento, Actualmente, las encuestas por internet son muy populares entre los investigadores, por las siguientes razones, costo: enviar por correo una encuesta a cientos de personas, definitivamente será mucho más barato que imprimir los cuestionarios para repartir entre todos los participantes; ahorro de tiempo: tomará mucho más tiempo y esfuerzo procesar una encuesta en papel, en cambio, con las encuestas online permiten tener informes y estadísticas en tiempo real; mayor alcance: con las encuestas por internet podrás llegar a una gran cantidad de público, ya que no enfrentas las barreras geográficas como con las encuestas en papel.

**Las entrevistas telefónicas** son una forma eficiente de recoger ciertos tipos de datos y se están usando con cada vez mayor frecuencia. Se prestan particularmente bien a situaciones donde es necesario obtener resultados oportunos y cuando el largo de la encuesta es limitado.

**Las entrevistas** en persona en el hogar, oficina, etc. de un participante, son mucho más caras que las encuestas telefónicas o por correo. Estas pueden ser necesarias especialmente cuando se debe recoger información compleja.

Algunas encuestas combinan varios métodos. Por ejemplo, una encuestadora puede usar el teléfono para identificar participantes elegibles (tal como localizar individuos mayores elegibles para Medicare) y luego hacer cita para una entrevista en persona.

### OBSERVACIONES

Como técnica de investigación, la observación tiene amplia aceptación científica, generalmente esta consiste en observar a las personas cuando efectúan su trabajo. Se utiliza con el fin de estudiar a las personas en sus actividades de grupo y como miembros de la organización. El propósito de la organización es múltiple: permite determinar que se está haciendo, como se está haciendo, quien lo hace, cuando se lleva a cabo, cuanto tiempo toma, dónde se hace y por qué se hace.

### TIPOS DE OBSERVACIÓN

- Puede observar a una persona o actitud sin que el observado se dé cuenta.
- Segundo, el analista puede observar una operación sin intervenir para nada, pero estando la persona observada enteramente consciente de la observación.
- Por último, puede observar y a la vez estar en contacto con las personas observadas. La interacción puede consistir simplemente en preguntar respecto a una tarea específica, pedir una explicación, etc.

### NOCIÓN ACERCA DEL TAMAÑO DE UNA MUESTRA

El tamaño de muestra requerido en una encuesta depende en parte de la calidad estadística necesaria para los establecer los hallazgos; esto a su vez, está relacionado en cómo esos hallazgos serán usados.

Aun así, no hay una regla simple para el tamaño de muestra que pueda ser usada en todas las encuestas. Mucho de esto depende de los recursos profesionales y fiscales disponibles. Los analistas frecuentemente

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

---

encuentran que una muestra de tamaño moderado es suficiente estadística y operacionalmente. Por ejemplo, las muy conocidas encuestas nacionales frecuentemente usan cerca de 1,000 personas para obtener información razonable sobre actitudes y opiniones nacionales.

Cuando nos damos cuenta que una muestra apropiadamente seleccionada de sólo 1,000 individuos puede reflejar varias características de la población total, es fácil apreciar el valor de usar encuestas para tomar decisiones informadas en una sociedad compleja como la nuestra. Las encuestas proveen medios rápidos y económicos de determinar la realidad de nuestra economía y sobre los conocimientos, actitudes, creencias, expectativas y comportamientos de las personas.

### **TIPO DE DATOS**

Los especialistas en estadística desarrollan encuestas para manejar una amplia variedad de características llamadas **variables aleatorias**. Los datos que son los resultados observados de esas variables aleatorias, indudablemente serán distintos de una respuesta a otra.

**VARIABLE [1]:** es la cualidad o cantidad medible que se estudia de las unidades de análisis y que varían de una unidad a otra. Por ejemplo: edad, ingreso de un individuo, sexo, cantidad de lluvia caída, etc.

Un **carácter** puede ser **cuantitativo** si es **medible** numéricamente o **cualitativo** si no admite medición numérica. El número de hermanos y la estatura son caracteres **cuantitativos** mientras que el sexo y el estado civil son caracteres **cualitativos**.

Los distintos valores que puede tomar un carácter cuantitativo configuran una **variable estadística**. La variable estatura, en cierta población estadística, toma valores en el intervalo 147-205; y la variable número de hermanos toma los valores 0, 1, 2, 3, 4, 5, 6, 7 y 8. Una variable estadística como esta última es **discreta**, ya que sólo admite valores aislados. Una variable estadística es **continua** si admite todos los valores de un intervalo, como ocurre con la estatura.

Las variables se dividen en:

**VARIABLES CUANTITATIVAS (NUMÉRICAS O MEDIBLES):** son todas aquellas cuyos diferentes valores pueden expresarse en forma numérica ordenada, pueden ser de dos clases: variables continuas o variables discretas.

- ✓ **VARIABLES CONTINUAS:** pueden tomar un número infinito de valores entre dos puntos determinados, como por ejemplo, las longitudes, áreas, volúmenes, pesos, ángulos, temperaturas, períodos de tiempo, porcentajes y velocidades. **Las variables continuas producen respuestas numéricas que surgen de procesos de medición.** Al medir la estatura de una persona se obtiene como respuesta cualquier valor en un continuo o intervalo, según la precisión del instrumento de medición.
- ✓ **VARIABLES DISCRETAS O DISCONTINUAS:** solo tienen valores numéricos fijos (enteros) sin posibles valores intermedios. son ejemplos el número de crías, el número de plantas en un cuadrado determinado.

---

[1] Algunos autores hacen la siguiente clasificación de las observaciones: A toda observación de tipo cualitativo se le llama **carácter** y a las observaciones de tipo cuantitativo se llama **variable** o **variada**. En nuestro caso llamaremos variables cualitativas y variables cuantitativas.

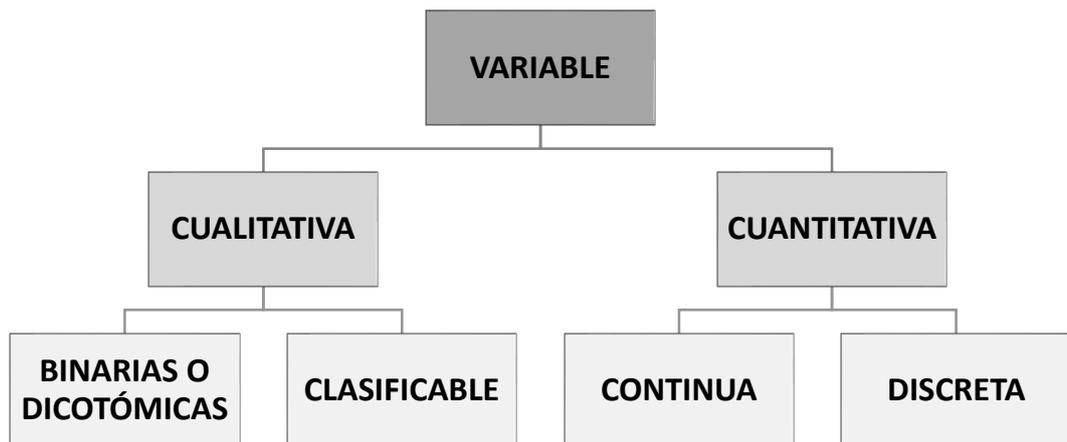
## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

- **Variables cualitativas (categóricas) o Atributos:** no pueden medirse sino que deben expresarse cualitativamente. todas ellas son propiedades tales como: sexo, color, nivel de ocupación, nivel de instrucción.

Si una variable categórica solo admite dos categorías, se llaman **dicotómicas o binarias**. Ej: Varón – Mujer; Estudia – No estudia.

Si la variable puede ordenarse o alinearse por su magnitud se llaman variables **clasificables**. Ej: Categoría de los hoteles, nivel de instrucción.

### CLASIFICACIÓN DE VARIABLES



### MEDICION DE VARIABLES

Antes que una variable sea tratada estadísticamente debe ser observada y medida para un conjunto de unidades observacionales, las unidades observacionales son aquellas entidades que se observan, cuando las observaciones se cuantifican (es decir se expresan numéricamente) se dice que las variables son medibles, una medición es una observación que se expresa físicamente o en forma numérica, es decir cuando se le otorga un valor determinado según su magnitud.

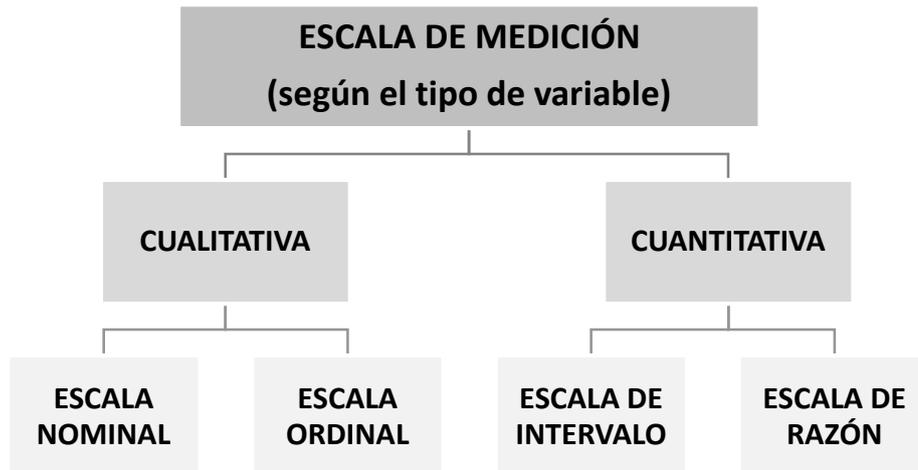
De lo anteriormente mencionado podemos deducir que existen diferentes formas de medir variables, algunas se las puede medir directamente por ejemplo la temperatura, la tensión arterial, la inteligencia, y otros en forma indirecta, como el grado de aprovechamiento de una determinada asignatura, la eficiencia, eficacia, el amor, ser romántico, ser buen o mal estudiante, etc.

### ESCALAS DE MEDICIÓN

Una medición consiste en el establecimiento de números o códigos a las observaciones mediante el empleo de una escala apropiada.

Una escala se define como un instrumento de medición. Las escalas de medición se diferencian por propiedades de orden y de distancia.

S. S. Stevens en 1946 desarrolló una clasificación para las escalas de medición en: **nominal, ordinal, de intervalo y de razón**. Estas escalas, las clasificaremos según al tipo de variable a la cual pertenezcan.



### **VARIABLE CUALITATIVA**

- **Escala nominal:** Es aquella en la que no se hace ningún supuesto respecto de las relaciones que existen entre los valores de las variables. Cada valor se asigna a una categoría diferente, la que se define mediante un nombre. Por ejemplo, el lugar de nacimiento, religión, establecimiento escolar, etc.
- **Escala Ordinal:** Es aquella que surge al ordenar todas las categorías de la variable de acuerdo a algún criterio. Por ejemplo, clasificar a las personas de acuerdo a su máximo nivel de instrucción alcanzado en: primario, secundario, universitario, etc. Otros casos pueden ser: nivel socioeconómico, categoría de empleo, grado de conformidad de un individuo respecto a alguna situación en particular.

### **VARIABLE CUANTITATIVA**

- **Escala de intervalo:** Es aquella que tiene una propiedad de asignar una medición de distancia entre los valores de la variable. Una de las características de la escala de intervalo es que el punto de origen o punto cero en la escala de medición es un punto arbitrario o punto convencional. Algunos de ejemplos de variables que se pueden medir con escala de intervalo son: temperatura, tiempo, inteligencia, altura geográfica de un lugar.
- **Escala de razón:** Es aquella que presenta todas las propiedades de orden y distancia de una escala de intervalo adicionando el punto cero o punto de origen. El cero es real, es absoluto, esto implica que hay un punto en la escala en que no existe la propiedad medida por la variable. Ej: peso, estatura y medidas de distancia, todas estas tienen un cero natural y son medibles con escalas racionales.

### **PRESENTACION SE DATOS**

Existen tres formas de presentaron de los datos ya organizados y procesados de cualquier estudio estadísticos.

- **Texto**
- **Cuadros**
- **Gráficos**

Cada una de estas formas presenta ventajas y desventajas, de acuerdo al tipo de variable en análisis y el objetivo perseguido con la presentación.

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

---

### TEXTO

Es una combinación de cifras y texto. Esta forma de presentación permite llamar la atención sobre las comparaciones de importancia y resaltar ciertas cifras, pero solo puede usarse cuando los datos a presentar son pocos.

### CUADROS O TABLAS

Este tipo de presentación permite valorar un gran número de datos en forma resumida, haciendo fácil y clara su lectura.

Generalmente un cuadro debe incluir las siguientes partes:

- **Título:** Debe enunciar que datos están incluidos en el cuerpo del cuadro, el lugar y el momento de referencia.
- **Encabezados:** En general se debe explicitar la variable seleccionada y las categorías correspondientes. En los cuadros de doble entrada generalmente en la columna se ubica la variable “independiente” y en las filas la variable “dependiente”, esta última es la variable que se trata de explicar.
- **Cuerpo del cuadro:** Contiene los datos clasificados de acuerdo a las variables del encabezado.
- **Notas de pie:** Son utilizados para realizar aclaraciones sobre partes incluidas en el cuadro. Por ejemplo: “cifras redondeadas”, “no se posee dato”, etc.
- **Fuente:** Da cuenta de la fuente de la cual fueron extraídos los datos, en el caso de ser secundarios. Cuando se trabaja con datos primarios, generalmente en fuente se coloca “elaboración propia”.

Las tablas de frecuencias sirven para ordenar los datos de una muestra y permitir que se pueda leer la información de forma más clara. En una tabla de frecuencia encontramos la siguiente información:

- **$n$** : es el número total de observaciones o tamaño de la muestra.
- **FRECUENCIA ABSOLUTA:** es el número de veces que se repite cada valor de la variable se denomina frecuencia absoluta. Simbolizaremos a las frecuencias absolutas con  $f_i$ .
- **FRECUENCIA RELATIVA:** es la proporción de veces que ocurre dicha categoría con respecto a la totalidad de las observaciones. Simbolizaremos a las frecuencias relativas con  $fr_i$
- La frecuencia relativa se calcula como la frecuencia absoluta correspondiente a una categoría de la variable dividida por el número total de observaciones ( $n$ ).

Se calcula mediante la fórmula:

$$fr_i = \frac{f_i}{n}$$

Este tipo de frecuencia relativa también se suele llamar **proporción**.

- **FRECUENCIA ACUMULADA:**

La frecuencia acumulada para un valor dado de la variable es la suma de las frecuencias (absolutas o relativas) de los valores menores o iguales al valor que se está considerando. En símbolos  $F_{ia}$  = (frecuencia absoluta acumulada) y  $F_{ra}$  = frecuencia relativa acumulada.

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

- **Frecuencia absoluta acumulada** indica la frecuencia absoluta que se acumula hasta esa fila de la tabla y se calcula, para una fila k cualquiera, mediante la siguiente fórmula:

$$f_{ia} = \sum_{i=1}^k f_i$$

- **Frecuencia relativa acumulada** indica la frecuencia relativa que se acumula hasta esa fila de la tabla y se calcula, para una fila k cualquiera, mediante la siguiente fórmula:

$$f_{ra} = \sum_{i=1}^k fr_i$$

La suma de las frecuencias relativas es siempre igual a 1. Las frecuencias relativas también se expresan como porcentajes al multiplicarlas por 100. En el ejemplo se tiene un cuadro que incluye las frecuencias relativas.

### ORGANIZACIÓN DE DATOS CUALITATIVOS.

Cuando se trabaja con variables cualitativas el análisis de la información es mucho más sencillo que cuando la variable es cuantitativa.

La primera operación a realizar con variables cualitativas es contabilizar el número de casos que pertenecen a una misma categoría de la variable.

Cuando se tiene una **gran cantidad** de datos, el primer paso consiste en ordenarlos en un cuadro o tabla que consta de dos columnas, la primera de ellas conteniendo la identificación de la unidad de observación y la segunda el dato correspondiente a la cantidad de conteos de dicha unidad.

#### Ejemplo:

Se está haciendo un estudio previo a las elecciones estudiantiles para saber por quiénes votarán a los alumnos que se encuentran en una mesa de la confitería. Se puede presentar la siguiente lista hipotética, donde la unidad de observación será cada uno de los integrantes de la reunión y la variable es la agrupación estudiantil a la que votarán.

**Tabla N° 1:** Agrupación estudiantil preferida por un grupo de estudiantes de la FAUD para las elecciones 2011.

| Alumno | Agrupación estudiantil |
|--------|------------------------|
| 1      | Participación Activa   |
| 2      | Renovación             |
| 3      | Participación Activa   |
| 4      | Cambio y Progreso      |
| 5      | Renovación             |
| 6      | Participación Activa   |
| 7      | Voto en blanco         |
| 8      | Cambio y Progreso      |
| 9      | Cambio y Progreso      |
| 10     | Participación Activa   |
| 11     | Renovación             |
| 12     | Voto en blanco         |
| 13     | Cambio y Progreso      |
| 14     | Participación Activa   |

Con estos datos se puede construir una tabla de frecuencias. Cuando se pasa de la lista original a la tabla de frecuencias se ignora toda la información excepto las categorías y cuantas veces se observa cada una de ellas, es decir, su frecuencia, por lo tanto, esta tabla resume la información original.

**Fuente:** Centro de estudiantes FACSO

**UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO.  
ANÁLISIS EXPLORATORIO.**

**Tabla N° 2:** Agrupación estudiantil preferida por un grupo de estudiantes de la FAUD para las elecciones 2011.

| <b>Categoría (Agrupación estudiantil preferida)</b> | <b>Cantidad de alumnos que votaran cada categoría</b> |
|---|---|
| Participación Activa                                | 5   |
| Renovación  | 3   |
| Cambio y Progreso                                   | 4   |
| Voto en blanco                                      | 2   |

**Fuente: Elaboración propia**

La tabla resumen también se denomina tabla de distribución de frecuencias pues muestra cómo se distribuyen los individuos en las categorías.

Observando la tabla de frecuencia, se puede determinar que la agrupación Participación activa será votada por 5 estudiantes, la agrupación Renovación por 3, etc. Los números 5, 3, 4, 2 son frecuencias absolutas. La suma de las frecuencias absolutas es igual al número total de observaciones.

Además de las frecuencias absolutas se pueden calcular frecuencias relativas.

**Tabla N° 3:** Agrupación estudiantil preferida por un grupo de estudiantes de la FAUD para las elecciones 2011.

| <b>Categoría (Agrupación estudiantil preferida)</b> | <b>Frecuencia absoluta (<math>f_i</math>) (Cantidad de alumnos que votaran cada categoría)</b> | <b>Frecuencia relativa (<math>f_{ri}</math>) (Proporción)</b> | <b>Frecuencia relativa en porcentajes</b> |
|---|--|---|---|
| Participación Activa                                | 5  | $5/14=0,36$   | 36%                                       |
| Renovación  | 3  | $3/14=0,21$   | 21%                                       |
| Cambio y Progreso                                   | 4  | $4/14=0,29$   | 29%                                       |
| Voto en blanco                                      | 2  | $2/14=0,14$   | 14%                                       |
|   | $\sum f_i = 14$  | $\sum f_{ri} = 1$   | $\sum f_{ri} = 100\%$                     |

Estas frecuencias pueden ser representadas utilizando para ello un gráfico de barras, modificando convenientemente la escala correspondiente al eje vertical.

**ORGANIZACIÓN DE DATOS NUMÉRICOS.**

La organización de datos correspondientes a variable cuantitativas difiere en algunos aspectos según se trate de variables discretas o continuas.

**VARIABLES DISCRETAS**

EJEMPLO: Supongamos que en el mismo grupo del ejemplo anterior alguien quiere saber cuántas materias tienen rendidas cada uno de los presentes en la mesa.

**UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO.  
ANÁLISIS EXPLORATORIO.**

**Tabla N°4**

| <b>Alumno</b> | <b>N° de Materias aprobadas.</b> |
|---------------|----------------------------------|
| 1             | 3                                |
| 2             | 6                                |
| 3             | 2                                |
| 4             | 7                                |
| 5             | 5                                |
| 6             | 6                                |
| 7             | 4                                |
| 8             | 3                                |
| 9             | 2                                |
| 10            | 8                                |
| 11            | 2                                |
| 12            | 4                                |
| 13            | 6                                |
| 14            | 3                                |

Con estos datos se puede construir una tabla de frecuencias.

Una tabla de frecuencias para variables numéricas es una tabla que asocia cada valor de la variable, con la cantidad de veces que se observa dicho valor. También pueden calcularse las frecuencias relativas siguiendo el procedimiento explicado anteriormente.

Otras frecuencias importantes para describir un conjunto de son las frecuencias absolutas y relativas acumuladas.

*Fuente: Elaboración Propia*

**Tabla N°5: Descripción de las materias rendidas por un grupo de alumnos de la FAUD en julio de 2010**

| <b>N° de materias aprobadas</b> | <b>Frecuencia Absoluta (<math>f_i</math>)</b> | <b>Frecuencia relativa (<math>f_{ir}</math>)</b> | <b>Frecuencia absoluta acumulada (<math>f_{ia}</math>)</b> | <b>Frecuencia relativa acumulada (<math>f_{ira}</math>)</b> |
|---------------------------------|---|--|--|---|
| 2                               | 3   | 0,214  | 3  | 0,214   |
| 3                               | 3   | 0,214  | 6  | 0,428   |
| <b>4</b>                        | <b>2</b>                                      | <b>0,143</b>                                     | <b>8</b>   | <b>0,571</b>  |
| 5                               | 1   | 0,071  | 9  | 0,642   |
| 6                               | 3   | 0,214  | 12   | 0,856   |
| 7                               | 1   | 0,071  | 13   | 0,927   |
| 8                               | 1   | 0,071  | 14   | 0,998 $\cong$ 1   |
|                                 | $\sum f_i = 14$                               | $\sum f_{ir} \cong 1$                            |  |   |

*Fuente: Elaboración Propia*

Si por ejemplo, nos posicionamos en la tercera fila de la tabla, la columna 2° y 3° establece que **dos** alumnos aprobaron 4 materias, o el 14,3% de los presentes, en la 4° columna se establece que 8 alumnos tiene aprobadas 4 o menos materias, y la 5° columna establece el mismo resultado pero se puede llevar a porcentajes, es decir que el 57,1% de los alumnos aprobó 4 o menos materias.

Estas frecuencias también pueden ser representadas por medio de un gráfico usando el eje horizontal para indicar las frecuencias.

**VARIABLES CONTINUAS.**

Veremos ahora como se resumen las variables continuas. Cuando se realiza un relevamiento de datos continuos, generalmente se tiene un registro desordenado de las mediciones efectuadas sobre cada unidad experimental.

**UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO.  
ANÁLISIS EXPLORATORIO.**

EJEMPLO: En un experimento industrial donde se desea estudiar la capacidad de una máquina para producir piezas dentro de especificaciones, se han medido 100 tornillos, registrándose su longitud en mm. Con los siguientes resultados:

|      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|
| 6,4  | 6,8  | 6,3  | 6,9  | 7,8  | 6,7  | 6,3  | 6,8  | 12,6 | 10,8 |
| 6,9  | 8,6  | 7,8  | 7,5  | 9,3  | 7,8  | 6,9  | 6,9  | 13,6 | 11,4 |
| 6,9  | 8,6  | 7,9  | 7,8  | 9,9  | 9,6  | 8,6  | 7    | 14   | 13,7 |
| 7,3  | 8,7  | 8,4  | 8,5  | 10   | 9,8  | 8,7  | 7,9  | 15   | 14,6 |
| 8,1  | 8,9  | 8,9  | 8,5  | 10,5 | 9,9  | 9,5  | 10,3 | 11,1 | 11,3 |
| 8,3  | 8,9  | 9,7  | 9,3  | 10,6 | 10,6 | 10,1 | 10,9 | 11,3 | 11,4 |
| 8,7  | 9,3  | 10,3 | 9,7  | 10,6 | 10,6 | 10,5 | 11,3 | 13,2 | 11,8 |
| 9,5  | 10,6 | 10,6 | 10,1 | 10,6 | 10,9 | 11   | 11,4 | 14,9 | 15   |
| 10,2 | 11,6 | 11,3 | 10,6 | 10,7 | 10,9 | 12,3 | 13,2 | 11,6 | 11,8 |
| 10,3 | 12,8 | 12,9 | 11,1 | 12,6 | 11,9 | 14,1 | 13,1 | 15,3 | 12,8 |

Al observar estos datos puede verse que el mayor valor que toma la variable longitud de tornillo es **15.3** y el menor valor es **6.3**; determinándose el **rango** que es **15.3-6.3=9**.

**RANGO**

*El rango se define como la diferencia que existe entre el mayor y el menor valor observado que toma la variable en estudio.*

Este intervalo que contiene a todos los valores registrados puede dividirse, por ejemplo en 10 intervalos de 1mm de longitud cada uno. Se pueden agrupar los datos construyendo una tabla de frecuencias donde en lugar de escribir los valores originales de la variable se los reemplaza por 10 intervalos de longitud 1. Para establecer la frecuencia de cada variable se cuentan los valores de la variable que caen en cada uno de ellos.

**Tabla de frecuencias de la longitud de tornillos producidos por una máquina.**

| <i>Intervalos de clase</i> | <i>Frecuencia Absoluta</i><br>$f_i$ | <i>Frecuencia Relativa</i><br>$f_{ir}$ | <i>Frecuencia Absoluta Acumulada</i> $f_{ia}$ | <i>Frecuencia Relativa Acumulada</i><br>$f_{ira}$ |
|----------------------------|-------------------------------------|--|---|---|
| [6-7)                      | 11                                  | 0,11                                   | 11  | 0,11  |
| [7-8)                      | 9                                   | 0,09                                   | 20  | 0,2   |
| [8-9)                      | 14                                  | 0,14                                   | 34  | 0,34  |
| [9-10)                     | 11                                  | 0,11                                   | 45  | 0,45  |
| [10-11)                    | 22                                  | 0,22                                   | 67  | 0,67  |
| [11-12)                    | 14                                  | 0,14                                   | 81  | 0,81  |
| [12-13)                    | 7                                   | 0,07                                   | 88  | 0,88  |
| [13-14)                    | 5                                   | 0,05                                   | 93  | 0,93  |
| [14-15)                    | 4                                   | 0,04                                   | 97  | 0,97  |
| [15-16)                    | 3                                   | 0,03                                   | 100   | 1   |

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

Al realizar el agrupamiento de datos de una variable continua, en una tabla de frecuencias, **se pierde información** correspondiente a los valores individuales de la variable y los mismos quedan representados por un intervalo denominado **intervalo de clase**.

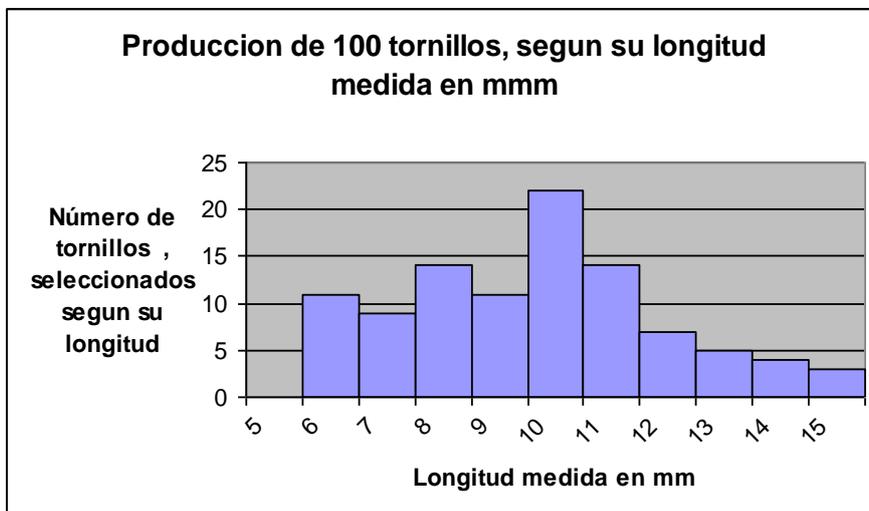
Algunas recomendaciones acerca de la elección de los intervalos de clase:

- La cantidad de intervalos y la amplitud de cada uno de ellos es arbitraria, pero a mayor cantidad de intervalos la amplitud será menor y se perderá menos información pero se incrementará la laboriosidad de los cálculos.
- Debemos tratar de que no nos queden intervalos de clase con frecuencia nula.
- Todas las variables deben quedar incluidas en alguno de los intervalos.
- Debe quedar definido cual extremo de cada intervalo es abierto o cerrado. Por ejemplo si hay un tornillo de 8mm, en que intervalo lo incluyo, en el **7-8** o en el **8-9**. Es por ello que conviene determinar cuál de los dos intervalos incluye al 8.

En este caso se pueden efectuar un gráfico que indique la distribución de la variable en estudio: longitud de tornillos. En el eje horizontal se ubicarán los distintos intervalos de clase y en el eje vertical las correspondientes frecuencias absolutas.

Además permite la identificación de valores **típicos y atípicos**<sup>2</sup> de una distribución. A partir de este histograma de frecuencias, en el caso de los tornillos se puede observar que una cantidad muy importantes de piezas presentan longitudes comprendidas entre 10 y 11mm, con una mayor concentración de tornillos más cortos y un rápido decrecimiento en las longitudes mayores de los mismos.

### GRÁFICO

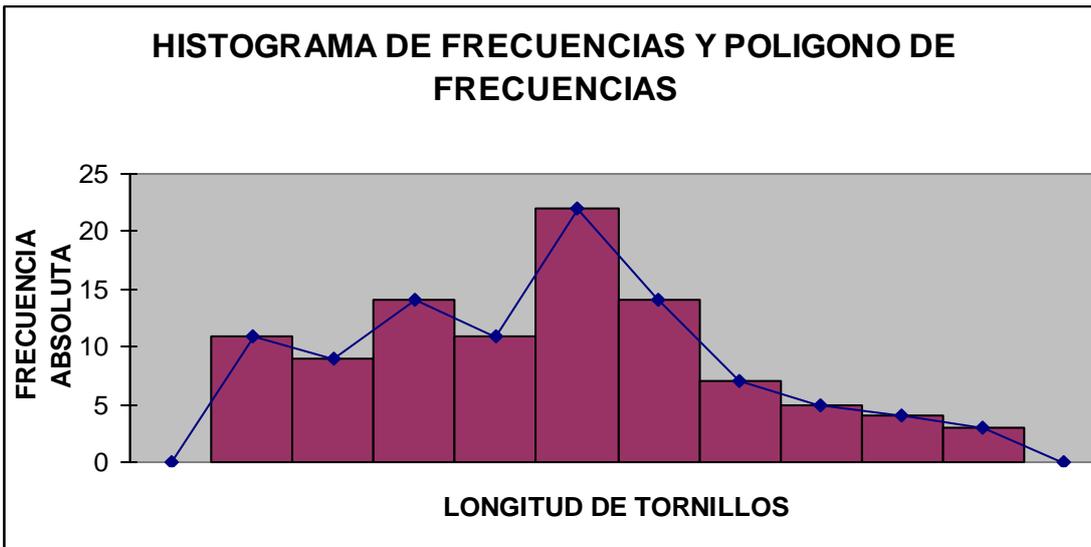


Cuando se estudian variables continuas, las frecuencias se representan mediante **áreas o superficies**, pues la escala en las que están medidas las mismas permiten tomar valores en cualquier punto del eje horizontal del gráfico.

El histograma de frecuencias es una representación simplificada de una gran masa de datos y debe ser parte del inicio de cualquier análisis más complejo.

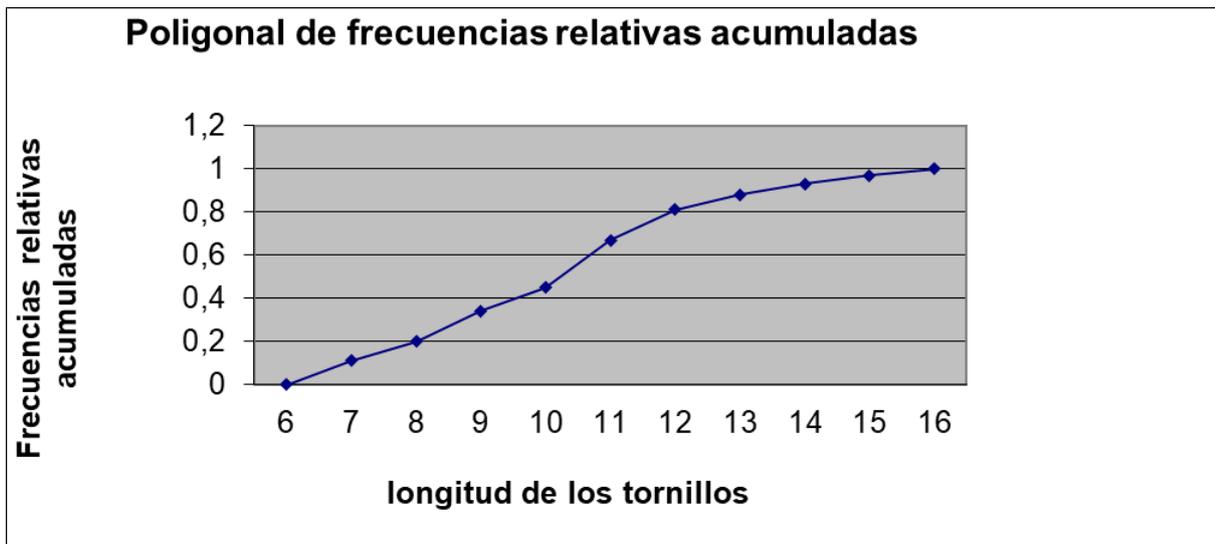
Si unimos los puntos medios (marca de clase) del histograma se obtiene el **polígono de frecuencias**

<sup>2</sup> En el análisis de datos se llamará **valor atípico** a aquel que se diferencia de los demás o que llama la atención del investigador. Este deberá averiguar su veracidad antes de proseguir con cualquier análisis.



**IMPORTANTE:** Si se quiere cerrar el polígono se agregan dos intervalos: uno anterior al primero y otro posterior al último se prolonga el polígono hasta el punto medio de estos intervalos.

Otro gráfico útil, conocido con el nombre de **polígono de frecuencias acumuladas u ojiva**, se utiliza para representar a las frecuencias acumuladas.



Observando el gráfico puede verse, por ejemplo, que el 20% de los tornillos mide aproximadamente 8mm o menos y que el 50% mide aproximadamente 10,3mm o menos.

**EJEMPLO DE TABLAS COMPARATIVAS:** Se ha contado el número de vainas por planta de soja para dos variedades A y B, obteniéndose las siguientes distribuciones de frecuencias. Para cada variedad.

**Tabla de frecuencias comparativa de la cantidad de vainas de soja por variedad.**

**UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO.  
ANÁLISIS EXPLORATORIO.**

| <i>Cantidad de vainas</i> | <i>Variedad A</i> | <i>Variedad B</i> | <i>Variedad A</i> | <i>Variedad B</i> |
|---------------------------|-------------------|-------------------|-------------------|-------------------|
| 0                         | 3                 | 7                 | 0,058             | 0,116             |
| 1                         | 5                 | 10                | 0,096             | 0,167             |
| 2                         | 8                 | 15                | 0,154             | 0,25              |
| 3                         | 9                 | 12                | 0,173             | 0,2               |
| 4                         | 12                | 7                 | 0,231             | 0,117             |
| 5                         | 8                 | 2                 | 0,154             | 0,033             |
| 6                         | 5                 | 4                 | 0,096             | 0,067             |
| 7                         | 2                 | 3                 | 0,038             | 0,05              |
| <b>Total</b>              | <b>52</b>         | <b>60</b>         | <b>1</b>          | <b>1</b>          |

Cuando se cuenta con distinto número de plantas de cada variedad es más conveniente utilizar la frecuencia relativa para comparar las distribuciones de frecuencias de la variable cantidad de vainas por planta de soja.

### GRÁFICOS ESTADÍSTICOS

La representación gráfica de los datos contenidos en un estudio estadístico tiene como finalidad ofrecer una visión del conjunto del fenómeno sometido a investigación, perceptible más rápidamente que la observación directa de los datos numéricos. Es por ello que las representaciones gráficas son un medio eficaz para el análisis de las estadísticas, ya que las magnitudes y regularidades se aprecian y recuerdan con más facilidad cuando se examinan gráficamente.

Las representaciones gráficas pueden hacerse usando un sistema geométrico de representación, en cuyo caso gozan de rigurosidad y precisión, o bien pueden utilizarse símbolos alusivos al tema en estudio (casas, árboles, figuras). Mediante este último sistema de representación no se persigue una rigurosa exactitud, sino impresionar a quien esté leyendo la información. Existen una gran cantidad de gráficos. Su elección depende de las variables en estudio, y de las características que se quieren resaltar.

Para la construcción de gráficos no hay reglas únicas, Siempre se debe tener presente que un gráfico da información más rápida pero menos precisa que el cuadro. El gráfico solo da información aproximada de la realidad.

Un gráfico completo debe reunir las siguientes partes:

- **Título:** Debe describir correctamente el contenido de la gráfica.
- **Diagrama:** Representa los datos mostrados en la gráfica.
- **Variable:** En todo gráfico debe colocarse la variable que se está representando. Si se usa un sistema de ejes, esta se coloca sobre el eje de las abscisas.
- **Escala:** Todo gráfico debe tener una escala de medida que de cuenta de las magnitudes que se están representando. Esta se coloca generalmente sobre el eje de ordenada.
- **Fuente:** debe ser colocada al pie de la gráfica e indicar cual fue el origen de los datos con los cuales la gráfica fue construida.

### RINCIPIOS DE EXCELENCIA GRÁFICA.

La propuesta más conocida para la adecuada presentación de datos en gráfica es la del profesor Edward R. Tufte, quien describe métodos adecuados para su diseño.

Las **características básicas** de una representación gráfica adecuada incluyen:

- Mostrar los datos.
- Indicar claramente las variables y los valores alcanzados por estas.
- Hacer que el observador se concentre en lo sustancial de la gráfica, y no en cómo se desarrolló.
- Evitar distorsiones.
- Facilitar la comparación de los datos.
- Cumplir con un objetivo claro.
- Que esté integrada las descripciones estadísticas y explicativas de la gráfica.

Los **principios de excelencia** grafica establecidos por Tufle son los siguientes:

- La excelencia gráfica es una presentación bien diseñada de los datos que proporcionan sustancia, estadística y diseño.
- La excelencia gráfica comunica ideas complejas con claridad, precisión y eficiencia.
- La excelencia gráfica proporciona al observador el mayor número de ideas en el menor tiempo y con el mínimo de tinta.
- La excelencia gráfica casi siempre involucra varias dimensiones.
- La excelencia gráfica requiere decir la verdad acerca de los datos.

A continuación se detallan los gráficos que se utilizan con mayor frecuencia:

- **Barras**
- **Gráficos de Líneas**
- **Diagrama Circular**
- **Histograma de Frecuencias**
- **Polígono de Frecuencias**
- **Ojiva.**

### TIPOS DE GRÁFICOS

- **GRAFICO A BASTONES.**

Este gráfico se conoce con el nombre de gráfico a bastones. La frecuencia se representa con barras o bastones aislados, pues entre dos valores sucesivos de la variable, **no existen** valores intermedios.

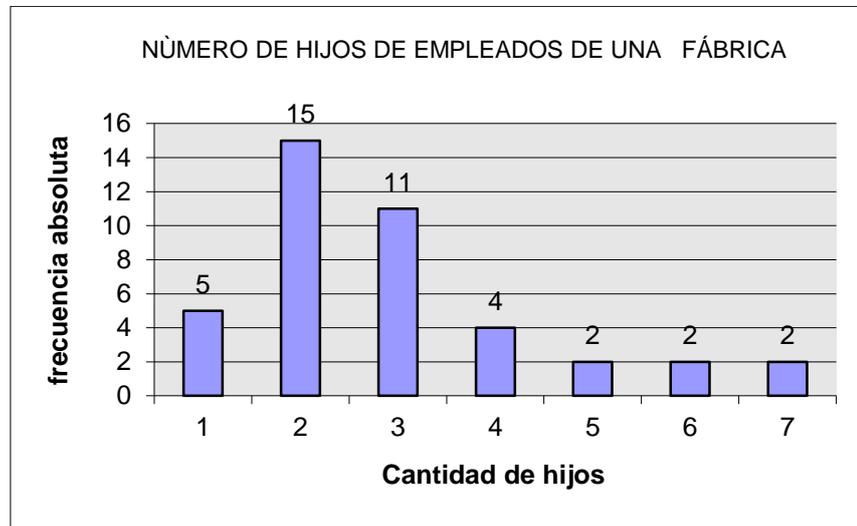
Se asocia a una tabla de frecuencias donde las barras son de igual ancho, pueden ser horizontales o verticales y conservan la misma distancia de separación. Se utiliza para variables cualitativas o continuas

**UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO.  
ANÁLISIS EXPLORATORIO.**

discretas. En el eje horizontal se colocan las categorías y en el eje de la ordenada los valores que asume la variable es decir la frecuencia absoluta.

Ejemplos:

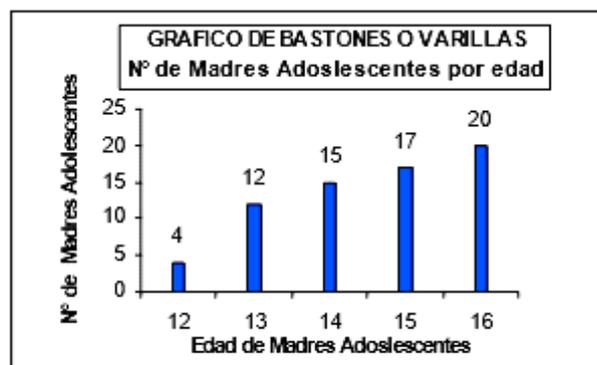
| Variable:<br>Número<br>de Hijos | Frecuencia<br>Absoluta<br>$f_i$ |
|---------------------------------|---------------------------------|
| 1                               | 5                               |
| 2                               | 15                              |
| 3                               | 11                              |
| 4                               | 4                               |
| 5                               | 2                               |
| 6                               | 2                               |
| 7                               | 2                               |
| TOTAL                           | 41                              |



Fuente: Elaboración Propia

| Edad | N° de Madres Adolescentes |
|------|---------------------------|
| 12   | 4                         |
| 13   | 12                        |
| 14   | 15                        |
| 15   | 17                        |
| 16   | 20                        |

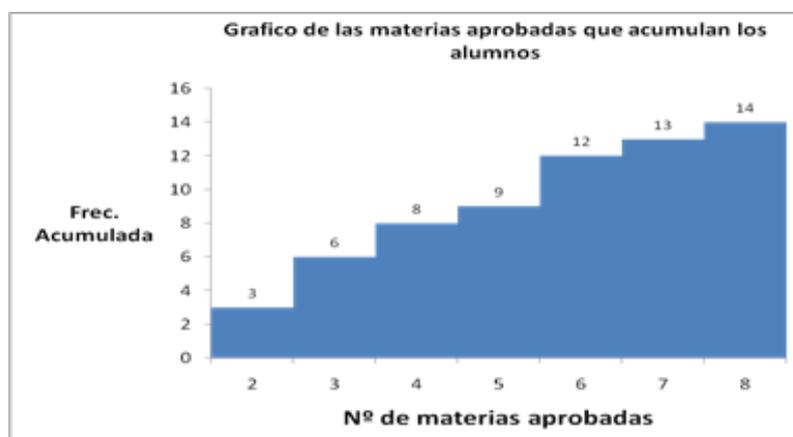
FUENTE : Maternidad de Lima  
ELABORACION : ESTADISTICA APLICADA /JMR



FUENTE : Maternidad de Lima  
ELABORACION : ESTADISTICA APLICADA /JMR

La frecuencia absoluta acumulada puede graficarse de la siguiente manera:

- **GRÁFICO A ESCALONES**



## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

Cada valor de la variable experimenta un salto de magnitud igual a la frecuencia relativa. De la misma manera se grafican las frecuencias absolutas acumuladas.

### • GRÁFICOS DE BARRAS

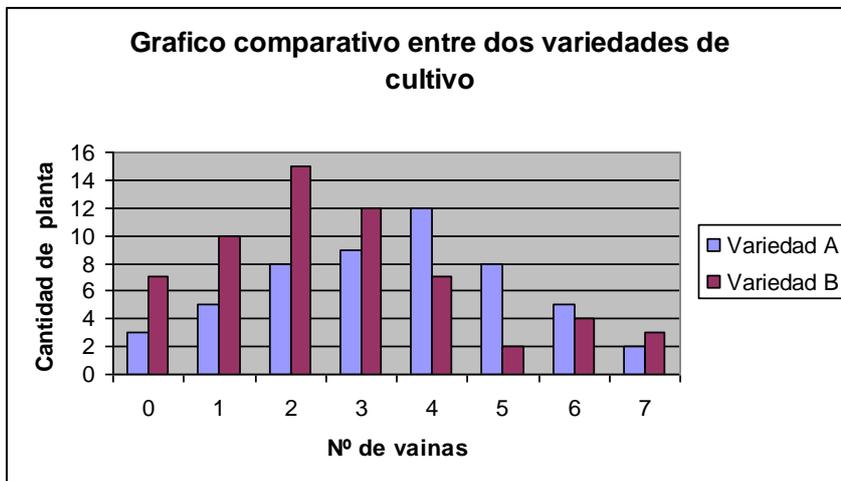
Representan valores usando trazos verticales u horizontales, aislados o no unos de otros, según la variable a graficar sea discreta o continua.

Pueden usarse para representar:

- una serie
- dos o más series (también llamado de barras comparativas)

La distribución de frecuencias también puede ser útil en la comparación de conjuntos de datos.

Efectuaremos esta comparación mediante la realización de un **gráfico de bastones por ser la variable de estudio de tipo numérica discreta.**

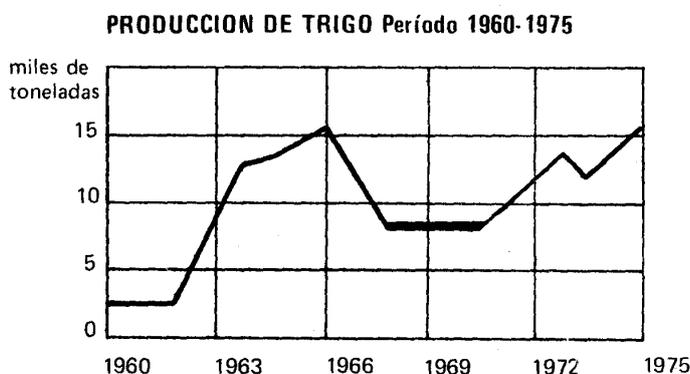


A simple vista puede observarse que la variedad A tiene mayor cantidad de vainas por plantas que la variedad B, ya que la variedad A tiene mayor frecuencia relativa para plantas con más vainas.

### • GRÁFICO DE LÍNEAS

Usado básicamente para mostrar el comportamiento de una variable cuantitativa a través del tiempo. El gráfico de líneas consiste en segmentos de líneas unidas entre sí que resaltan las variaciones de la variable en el tiempo

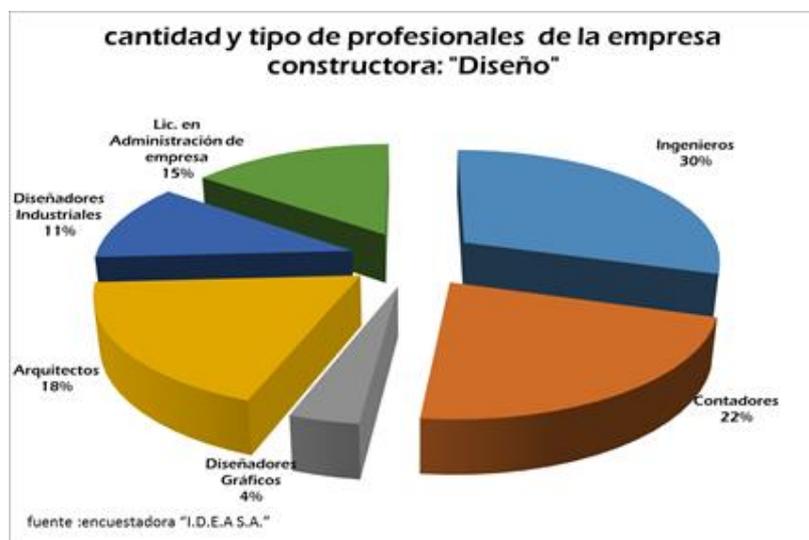
Ejemplo:



## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

### • DIAGRAMA CIRCULAR

Se utiliza para representar variables cualitativas en porcentajes. La circunferencia tiene en su interior 360° grados que comprende al 100% de los datos. Se trata de representar, mediante un círculo, la totalidad de los datos de la muestra y, en el mismo, mediante sectores circulares, proporcionalmente, se representarán los valores parciales.



| PROFESIÓN                         | PERSONAL (f) |
|-----------------------------------|--------------|
| Ingenieros                        | 8            |
| Contadores                        | 6            |
| Diseñadores Gráficos              | 1            |
| Arquitectos                       | 5            |
| Diseñadores Industriales          | 3            |
| Lic. en Administración de empresa | 4            |

### HISTOGRAMA DE FRECUENCIAS

Se asocia a una tabla de frecuencia de datos asociados en intervalos. La variable representada es cuantitativa continua.

Son diagramas de barras o rectángulos cuyas bases están sobre un eje horizontal (el eje X) con centros en las marcas de clase y longitud igual al tamaño de los intervalos de clase. Las alturas están dadas por las frecuencias absolutas o relativas. Se utiliza para graficar variables cuantitativas continuas.

#### Ejemplo:

En la Escuela primaria Sarmiento donde hay 4 salitas se decidió construir columpios en el patio de los jardines de infantes, por esto se hizo un relevamiento para detectar cuál era el peso en Kgs. de los alumnos que concurrían a la salita AZUL.

| Salita Azul |      |
|-------------|------|
| 14,1        | 16,9 |
| 14,2        | 17,7 |
| 14,3        | 18,4 |
| 14,4        | 18,6 |
| 14,9        | 19,2 |
| 15,4        | 19,8 |
| 15,5        | 20,5 |
| 15,8        | 21,7 |
| 16,3        | 22,9 |
| 16,6        | 23,9 |

| Sala azul |              |      |     |      |
|-----------|--------------|------|-----|------|
| Peso (kg) | Fi Sala Azul | Fir  | Fia | Fira |
| [14;15)   | 5            | 0,25 | 5   | 0,25 |
| [15;16)   | 3            | 0,15 | 8   | 0,4  |
| [16;17)   | 3            | 0,15 | 11  | 0,55 |
| [17;18)   | 1            | 0,05 | 12  | 0,6  |
| [18;19)   | 2            | 0,1  | 14  | 0,7  |
| [19;20)   | 2            | 0,1  | 16  | 0,8  |
| [20;21)   | 1            | 0,05 | 17  | 0,85 |
| [21;22)   | 1            | 0,05 | 18  | 0,9  |
| [22;23)   | 1            | 0,05 | 19  | 0,95 |
| [23;24)   | 1            | 0,05 | 20  | 1    |
| [24;25]   | 0            | 0    | 20  | 1    |
| Total     | 20           |      |     |      |

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

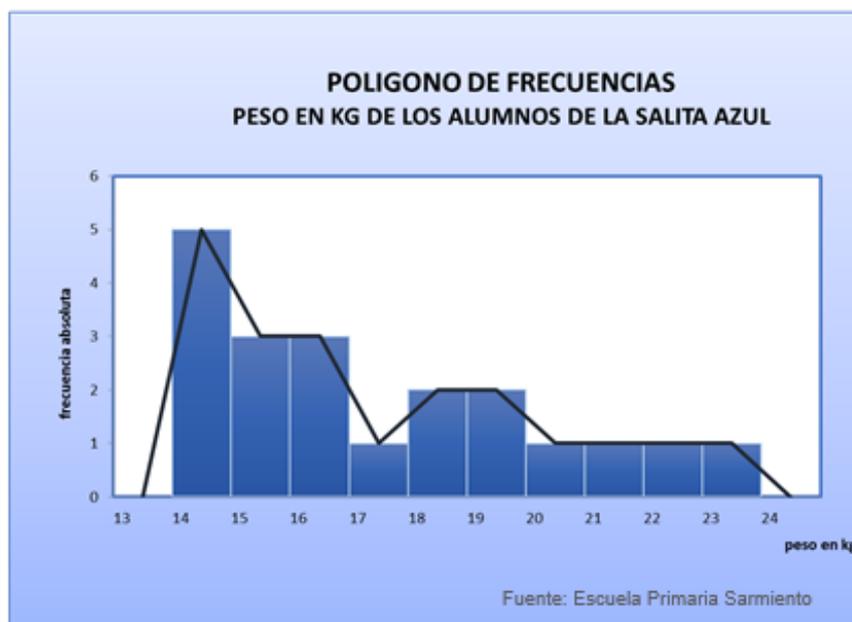


Fuente: Escuela Primaria Sarmiento

- **POLÍGONO DE FRECUENCIAS**

Un polígono de frecuencias es un gráfico de línea trazado sobre las marcas de clase. Puede obtenerse uniendo los puntos medios de los techos de los rectángulos en el histograma.

Tiene como base la marca de clase del intervalo ( $X_i$ ) y como altura la frecuencia absoluta respectiva ( $f_i$ ). Se cierra el polígono extendiendo la amplitud del intervalo en ambos lados. Al lado izquierdo restando a la marca de clase inicial el ancho de clase y al lado derecho agregando a la marca de clase final el ancho de clase.

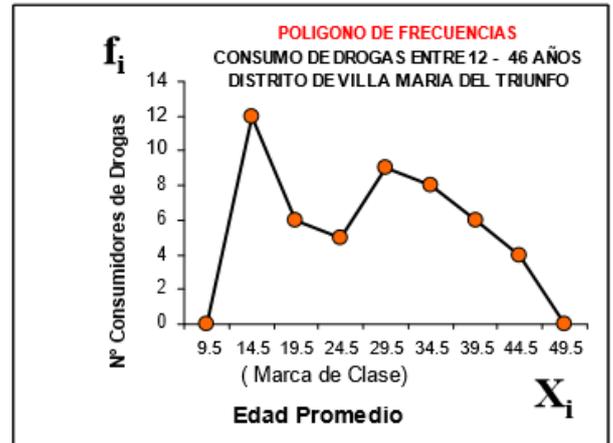


Fuente: Escuela Primaria Sarmiento

En un operativo en el distrito de Villa María se efectuó la detención de 50 consumidores de drogas observándose que sus edades fluctúan entre 12 años y 46 años siendo su tabla de distribución de frecuencia la siguiente:

**UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO.  
ANÁLISIS EXPLORATORIO.**

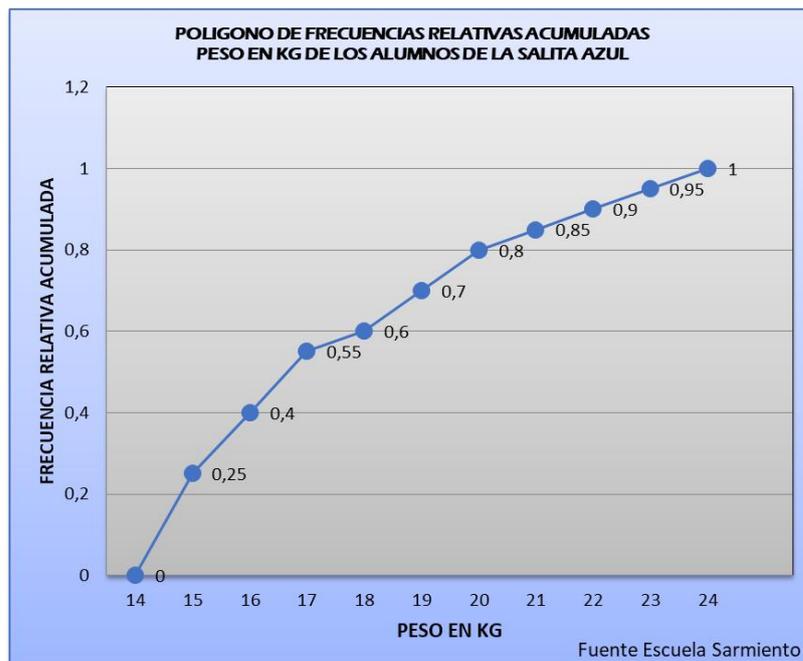
| <i>Intervalos Edades</i> | <i><math>X_i</math> ( Marca de Clase)</i> | <i><math>f_i</math><br/>N° de Consumidores</i> |
|--------------------------|---|--|
| [ 12- 17)                | 14.5                                      | 12   |
| [ 17- 22)                | 19.5                                      | 6  |
| [ 22- 27)                | 24.5                                      | 5  |
| [ 27- 32)                | 29.5                                      | 9  |
| [ 32- 37)                | 34.5                                      | 8  |
| [37- 42)                 | 39.5                                      | 6  |
| [ 42- 47)                | 44.5                                      | 4  |



FUENTE : Datos Hipotéticos  
ELABORACION : ESTADÍSTICA APLICADA / JMR

• **OJIVA:**

La base representa los intervalos y las aluras las frecuencias absolutas acumuladas ( $f_i$ ) o la frecuencia relativa acumulada ( $f_{ri}$ ). A cada límite superior le corresponde la frecuencia acumulada de cada intervalo iniciando con el límite inferior del primer intervalo asignándole como frecuencia = 0. Es utilizada para variables continuas.



UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO.  
ANÁLISIS EXPLORATORIO.

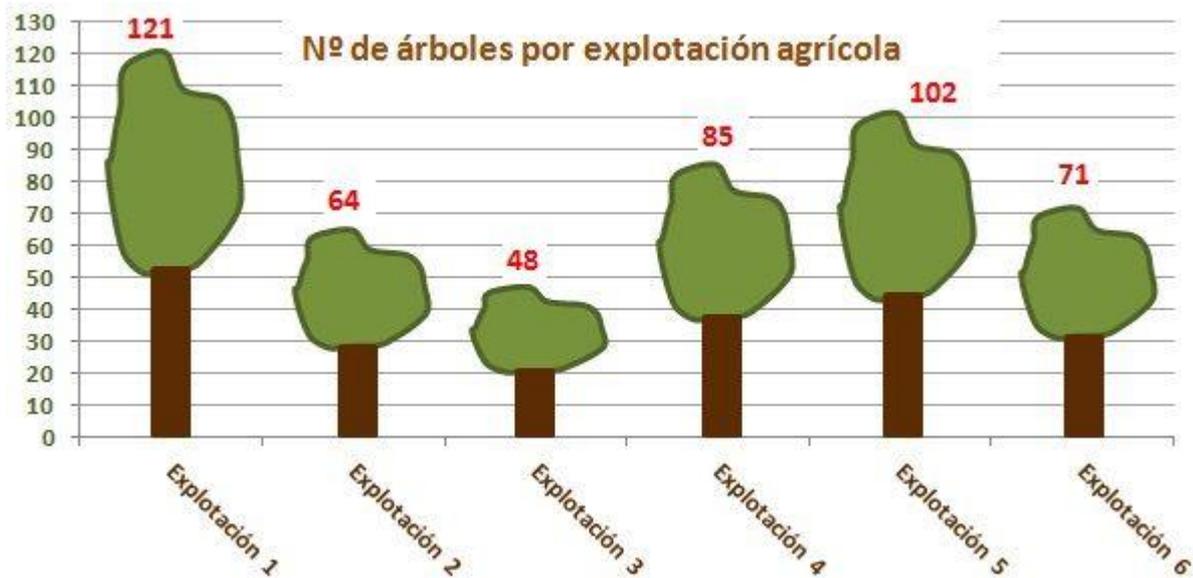
• PICTOGRAMAS

Los pictogramas son gráficos que emplean un dibujo en una determinada escala para expresar la unidad de medida de los datos. Los gráficos son dibujos alusivos al carácter que se está estudiando y cuyo tamaño es proporcional a las frecuencias que representan.

TIPOS DE PICTOGRAMAS:

DE BARRAS

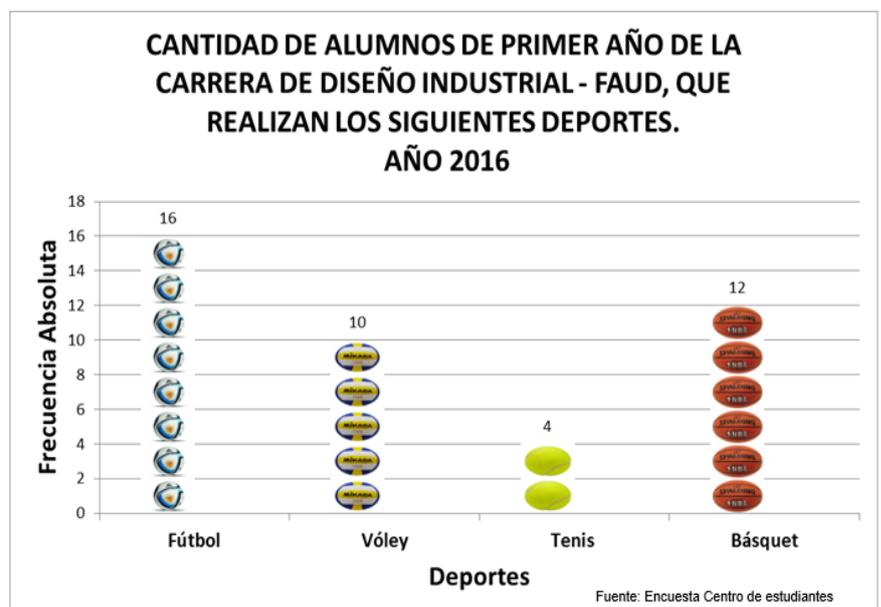
Las barras son dibujos o símbolos de **longitud proporcional** a la **frecuencia** de cada categoría o valor.



DE DIBUJOS

Cada uno de los dibujos o símbolos equivale a ciertas unidades de frecuencia **relativa** o **absoluta**. Cuantos más dibujos haya en cada categoría, mayor será la frecuencia.

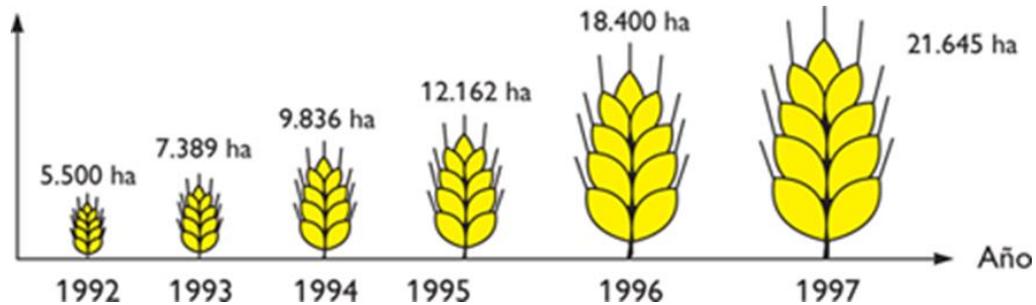
| Deportes | Cantidad de alumnos |
|----------|---------------------|
| Fútbol   | 16                  |
| Vóley    | 10                  |
| Tenis    | 4                   |
| Básquet  | 12                  |



## DE ÁREAS

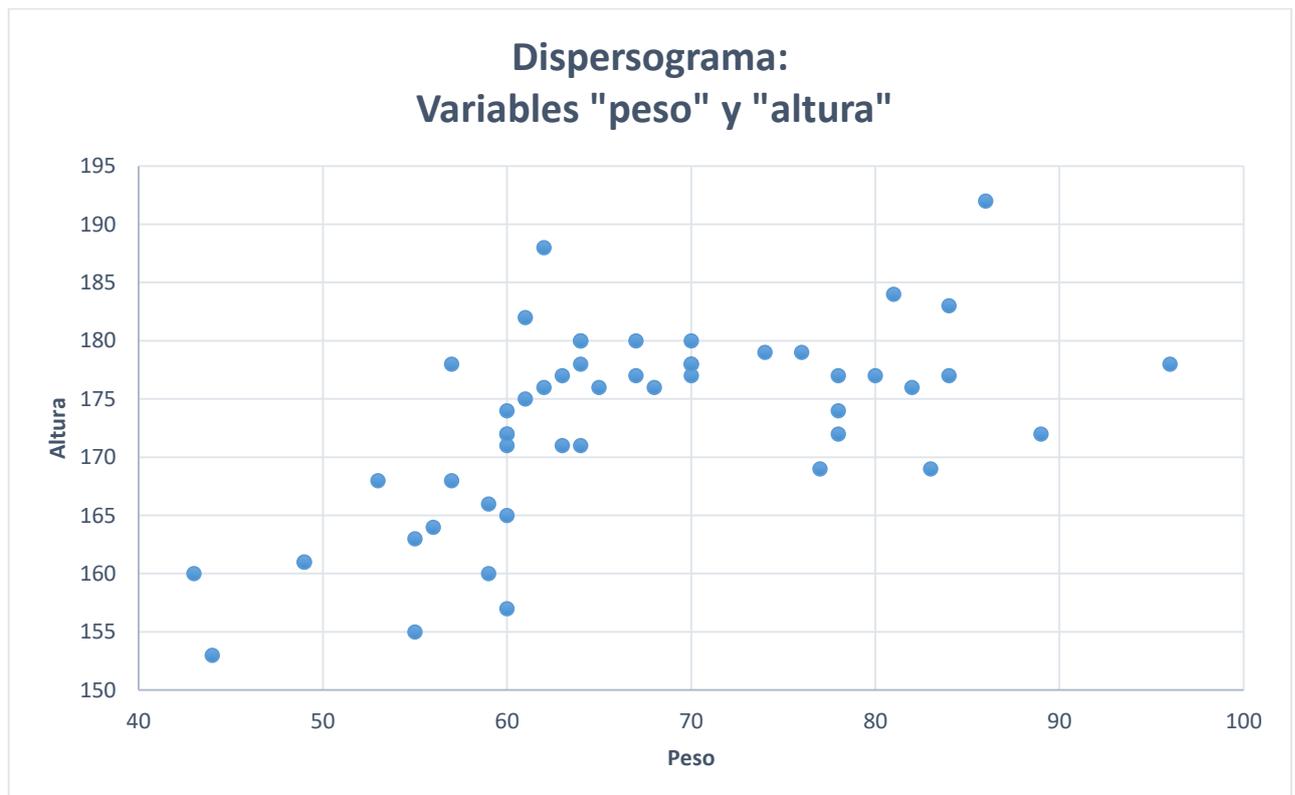
El dibujo o símbolo varía de tamaño proporcionalmente a la frecuencia. Cuanto mayor sea el **área** del dibujo, mayor será su frecuencia o cantidad.

El siguiente gráfico representa la evolución del número de hectáreas de sembradas de trigo en un país.



- **DISPERSOGRAMAS**

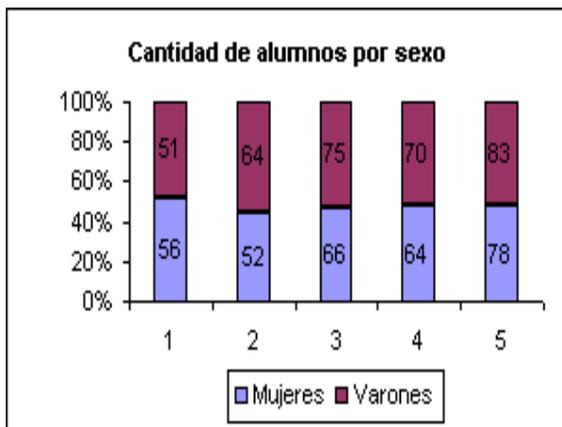
Son gráficos que se construyen sobre dos ejes ortogonales de coordenadas, llamados cartesianos, cada punto corresponde a un par de valores de datos  $x$  e  $y$  de un mismo elemento suceso. En el siguiente ejemplo se ha graficado para cada individuo dos variables. En el eje horizontal ( $x$ ) se graficó la variable peso de la muestra 3 y en el eje vertical ( $y$ ) la variable ancho de la muestra 3. Este gráfico es de gran utilidad cuando se desea determinar si hay correlación entre las dos variables.



• GRÁFICOS DE BARRAS HORIZONTALES PARA VARIABLES CATEGÓRICAS

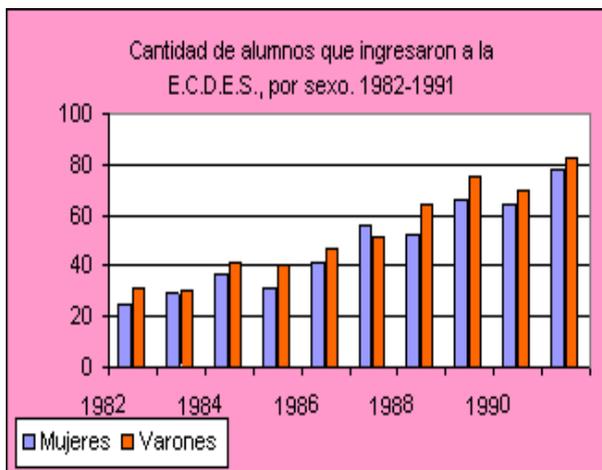


• GRÁFICOS DE BARRAS PROPORCIONALES



Se usan cuando lo que se busca es resaltar la representación de los porcentajes de los datos que componen un total. Las barras pueden ser: verticales u horizontales

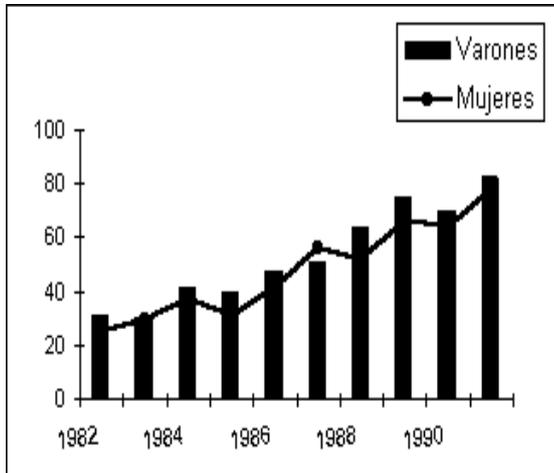
• GRÁFICOS DE BARRAS COMPARATIVAS PARA VARIABLES CATEGÓRICAS



Se utilizan para comparar dos o más series, para comparar valores entre **categorías**

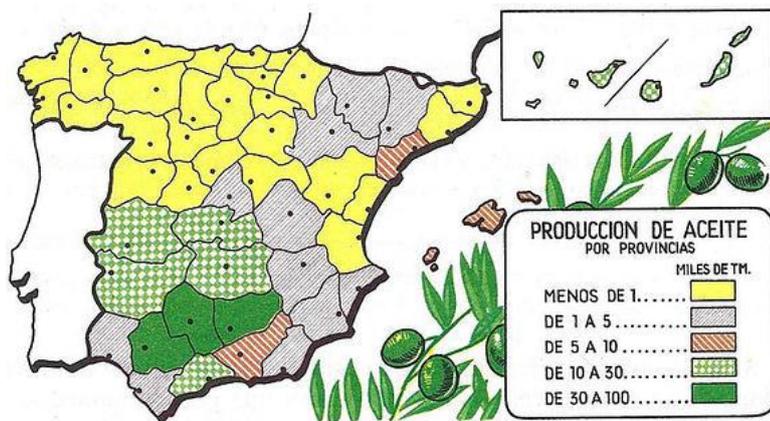
• **GRÁFICOS MIXTOS**

En estos tipos de gráficos se representan dos o más series de datos, cada una con un tipo diferente de gráfico. Son gráficos más vistosos y se usan para resaltar las diferencias entre las series.



• **OTROS GRÁFICOS<sup>3</sup>**

Existen infinidad de gráficos, que se usan para presentar la información en forma más atractiva y de fácil comprensión. Generalmente se usa en publicaciones de nivel popular y masivo.



Este sería un cartograma que además grafica la variable a mostrar.

Si bien en el desarrollo de la unidad se han explicado algunos gráficos para las diferentes tipos de variables, existen muchos más.

**ALGUNAS CONSIDERACIONES RESPECTO A LOS GRÁFICOS ESTADÍSTICOS.**

Es interesante tener en cuenta que a veces los gráficos pueden ser usados en forma conveniente dependiendo de los intereses de quien presente los datos como se verá en el siguiente artículo que tiene por título ¿Las estadísticas mienten? Se analizan algunas situaciones muy interesantes.

También se muestran algunos gráficos que no reflejan los datos que se pretenden mostrar. Estos son solo algunos de los tantos gráficos con los que se presenta la información, es por ello que el profesional debe estar atento a estos “errores”.

<sup>3</sup> Observe que en el caso de representar un dato obtenido de otra fuente, está debe ser mencionada.

## UNIDAD N° 1: CONCEPTO Y APLICACIÓN DE LA ESTADÍSTICA. CONCEPTO BÁSICO DE MUESTREO. ANÁLISIS EXPLORATORIO.

---

Esta tabla de frecuencias se puede representar por medio de un gráfico conocido como gráfico de barras o bastones.

Considerando un sistema de coordenadas cartesianas, sobre el eje horizontal se representan las distintas categorías de la variable en estudio (agrupación estudiantil preferida) sobre el eje vertical se construye una escala adecuada para representar la frecuencia correspondiente a cada una de estas categorías. Sobre cada categoría de la variable se levanta una barra de altura igual a la frecuencia de la categoría en cuestión. Estas barras deben presentar cierto espacio libre entre una categoría y otra.