**Probabilidad y Estadística**

**Estadística:** es una ciencia que nos permite obtener datos, organizarlos, procesarlos con el fin de sacar conclusiones para la toma de decisiones ante situaciones de incertidumbre.

**Dato Estadístico**: es el resultado de la medición. Se dice también que es el valor de la variable. Los números son útiles en conjunto y no como resultados aislados.  
  
**Existen 2 Ramas:**  
 **Estadística Descriptiva:** En ella se aprende una serie de técnicas para organizar, presentar y analizar un conjunto finito de observaciones, que, según el objetivo del estudio, constituyen una población o una muestra.

**Estadística Inferencial:** Nos ayuda a tomar decisiones para a la **población** a partir de la observación de una **muestra**.

*Conceptos a saber:*

* **Población**
* **Muestra**
* **Variable 🡪 2 tipos de variables: cuantitativa y cualitativa**
* **Parámetro**
* **Estadístico**

**Concepto de población y muestra:**

**Población:** conjunto de las unidades elementales a las cuales nos está interesando estudiarle algo. (Ej.: Todos los alumnos de un salón)

**Muestra:** subconjunto finito de una población y de carácter aleatorio. (Ej.: grupo de alumnos)

Contiene: **unidad elemental** (ej.: el alumno) y una característica de interés que llamaremos **variable** (ej.: edad del alumno).

**Técnicas de muestreo:**

Muestreo aleatorio o probabilístico. Este tipo de muestreo, al diseñar el grupo muestra de manera aleatoria, reconoce en cada individuo iguales probabilidades de ser seleccionados, de manera que se tenga un enfoque confiable y riguroso, desde el punto de vista metodológico, pues se considera exento de cualquier sesgo que pueda comprometer los resultados del estudio. Las técnicas de muestreo de este tipo son:

1. **Simple**

Todos los individuos de la población tienen la misma probabilidad de ser seleccionados. Este método es muy sencillo y es el más utilizado, pero es el menos eficaz en poblaciones muy extensas, porque el tamaño de la muestra es muy grande.

1. **Sistemático**

Se enumeran todos los individuos de la población, luego se elige el primero de la muestra de manera aleatoria y los siguientes de manera periódica. Pueden darse casos en los que en la población haya ciclos o periodicidades y que al utilizar este método haya una elevada homogeneidad en la muestra y que, por lo tanto, esta no sea representativa de la población.

1. **Estratificado**

Los individuos de la población se agrupan en estratos, en función de características esenciales de interés para el estudio a realizar (homogeneidad) y se seleccionan los individuos de la muestra de cada uno de los estratos de manera aleatoria. El número de individuos de la muestra pertenecientes a cada estrato es proporcional al número de individuos de cada uno de ellos con respecto a la población total. Debido a que se consigue aumentar la representatividad de la muestra los individuos son más confiables.

1. **Conglomerados**

Las poblaciones están formadas generalmente por determinados grupos o conglomerados, por ejemplo, un municipio está integrado por pueblos. En esta técnica, en vez de seleccionar a los individuos de la muestra directamente, los individuos que se escogen de manera aleatoria son algunos de estos grupos o conglomerados.

**Clasificación de variables:**

Las **variables** son las características observables, tanto cualitativas o cuantitativas que poseen las unidades experimentales o cada elemento individual de una población o muestra. Se utilizan para representar datos.

Una variable es **cualitativa** cuando expresa un atributo o cualidad de la unidad elemental que se observa. Nos devuelve un atributo o palabra. Su **escala de medida** puede ser **nominal** (no existe orden, color de ojos) u **ordinal** (existe orden, jerarquías de trabajo).

Una variable es **cuantitativa** cuando se expresa numéricamente ya que nos devuelven un número o una cantidad. Su **escala de medida** puede ser **discretas** (entre 2 valores consecutivos no existen valores intermedios, n° de manzanas) y **continuas** (entre 2 valores consecutivos existen valores intermedios, altura).

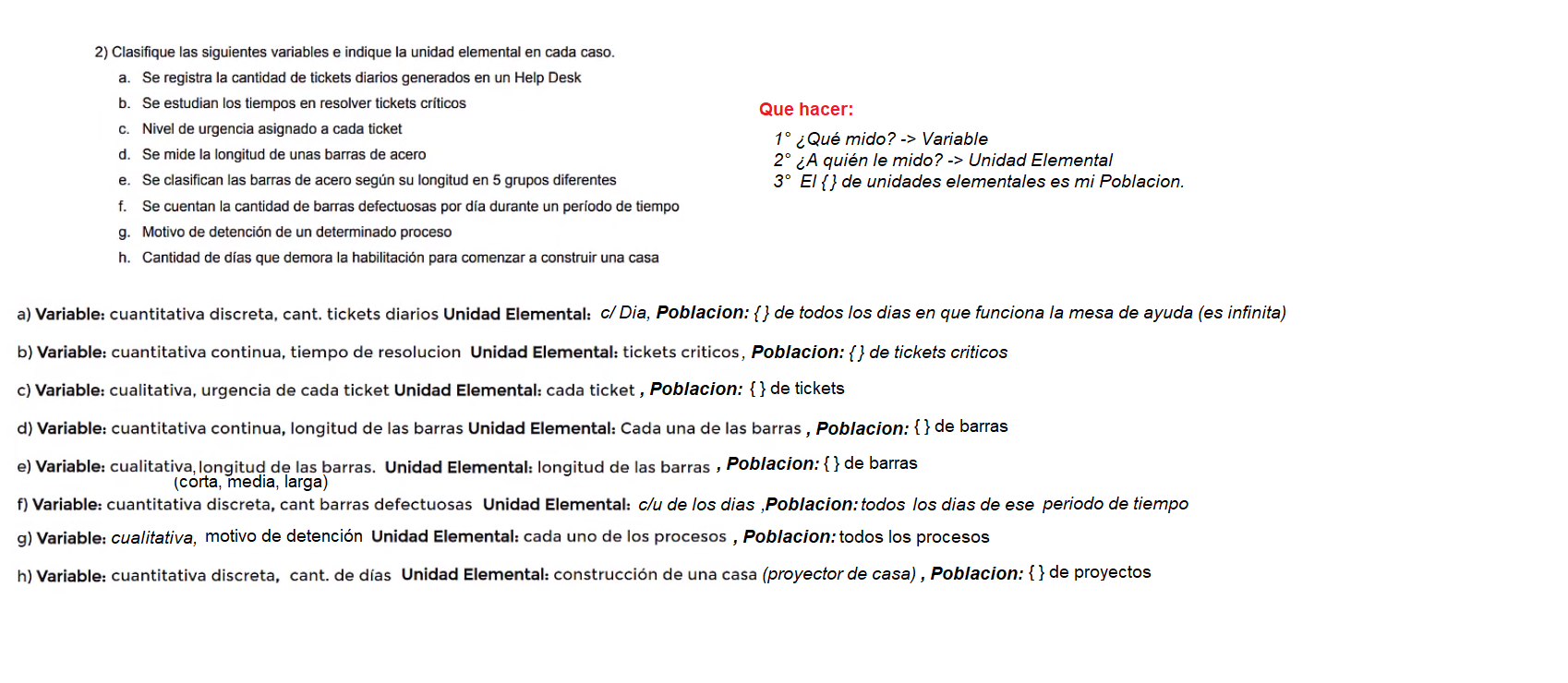
**Valores característicos:**

Los **parámetros** son las características numéricas de una **población.**

* Es una medida de resumen de una variable estudiando a la población entera.
* Es fijo.
* Usaremos:
* media/promedio poblacional (μ)
* desvío estándar poblacional (σ)
* varianza poblacional (σ2)
* proporción poblacional (π)

Los **estadísticos** son las características numéricas de una **muestra.**

* Es una medida de resumen para una variable y que se calcula estudiando los valores de la muestra.
* Es variable.
* Usaremos:
* media/promedio muestral (X̅)
* desvío estándar muestral (*S*)
* varianza muestral (*S* 2)
* proporción muestral (p̂)

**Ejercicio:**

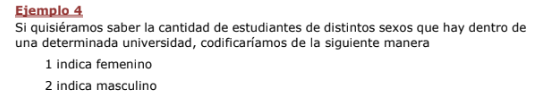
**Organizar y presentar datos**

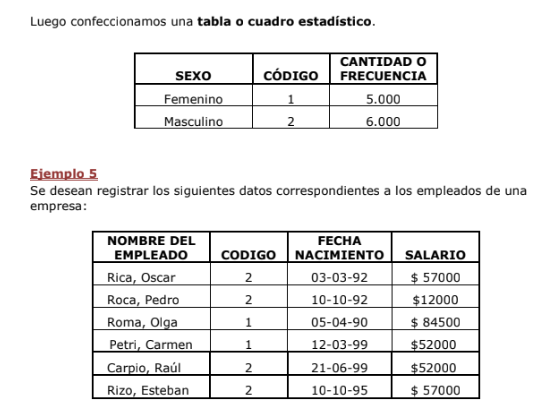
Para realizar el análisis de los datos que recolectamos, éstos deben ser organizados. Una vez hecho esto será posible su presentación a través de Tablas o Gráficos.

**Tablas:**

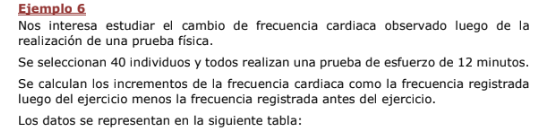
Una manera de organizarlos es ingresarlos a un ordenador y luego presentarlos mediante una tabla u hoja de cálculo. Algunos datos precisan ser codificados, es decir adjudicarles un símbolo para que puedan ser considerados como una variable cuantitativa y procesados por el ordenador. Veamos primero un ejemplo de una tabla que incluye codificaciones y luego otras que no la incluyen:

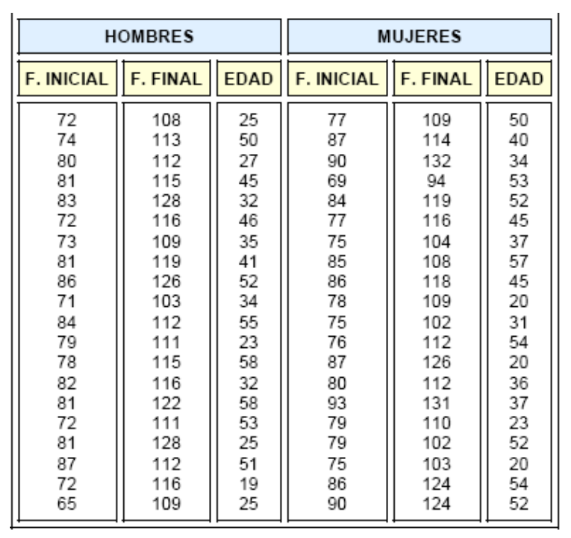
**Con codificación:**

****

****

**Sin codificación:**

****

****

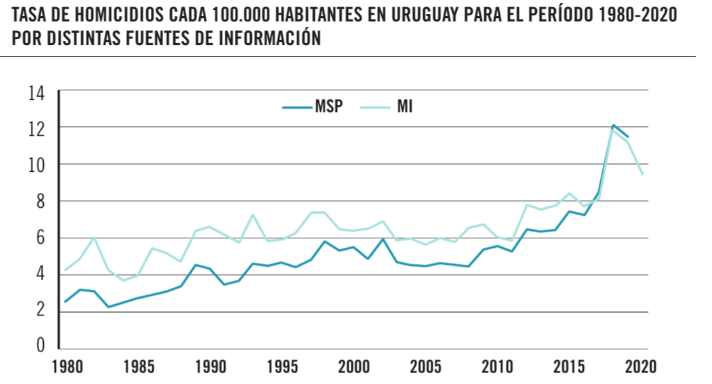
**Gráficos:**

A continuación, retomaremos los ejemplos propuestos para mostrarle cómo pueden ser organizados en otras formas de representación: los Gráficos.

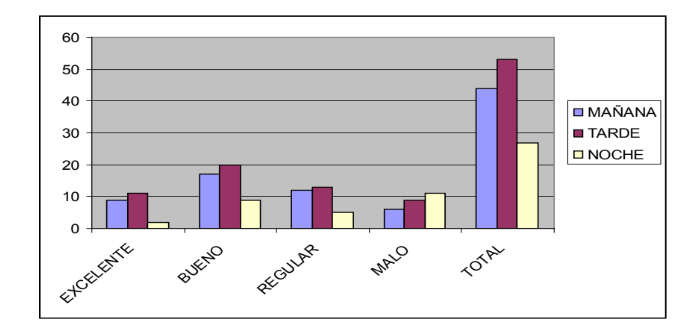
Existen distintos tipos de gráficos y la selección de uno o varios de ellos responde a que dichos gráficos permitan visualizar de la manera más clara aquello que queremos mostrar. A continuación, veremos tres tipos de ellos:

* Gráfico Lineal
* Gráfico Circular
* Gráfico de Barra

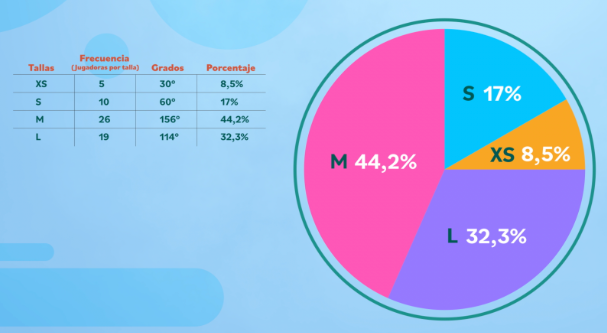
**Gráfico Lineal**: Se utiliza por lo general para **representar valores de dos variables cuantitativas** utilizando coordenadas cartesianas.



**Gráfico de Barras:** Se utiliza por lo general **cuando una de las variables es cualitativa**.Está formado por rectángulos cuya base se encuentra indistintamente en cualquiera de los dos ejes cartesianos. Estas barras pueden estar segmentadas o agrupadas.



**Gráfico Circular: Se utiliza por lo general para comparar partes referidas a un total y su evolución en el tiempo o el espacio.** **Las partes de este gráfico circular se denominan sectores circulares y el ángulo que le corresponde a cada sector es proporcional a la parte del total que se quiere representar.**

****

**Distribución de frecuencias**

Los datos (valores de la variable) se van recopilando a medida que los hechos ocurren o se observan por lo que son registrados en una forma desordenada. Decimos entonces que estos datos **no están agrupados**.

Muchos textos dividen los datos de esta forma, en agrupados y no agrupados. Lo cierto es que con los datos desordenados o no agrupados es muy difícil trabajar ya que no se ponen en evidencia fácilmente ciertas **regularidades** que precisa la estadística y, en general, toda la matemática, para obtener información. Es preciso entonces agrupar los datos a medida que estos aparecen.

A la relación que hace corresponder a los valores de una variable la cantidad de veces que estos aparecen o son observados se la denomina **Distribución de frecuencias** y se la representa en una tabla o en un gráfico estadístico.

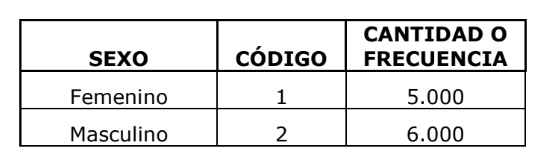
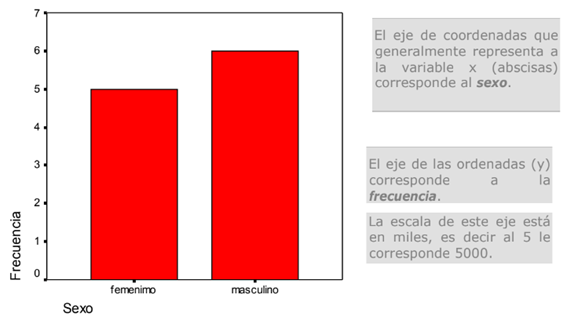


Tabla a 🡪 gráfico de barra



**Este tipo de gráfico se denomina Histograma de Frecuencia. Se construye así:**

1. En el eje horizontal se marcan los valores que toma la variable de estudio.
2. En el eje vertical se representa la frecuencia con que ocurren esos valores o la cantidad de veces que ocurren esos valores.
3. Para cada valor de la variable de estudio se dibujan rectángulos adyacentes de igual ancho y con una altura igual a la frecuencia con que ocurre ese valor.

PARA MI ESTÁ MAL PORQUE EL HISTOGRAMA TIENE LA VARIABLE REPRESENTADA EN INTERVALOS Y EL DE BARRA NO.