

# Probabilidad y Estadística

# 1

## Actividades de Aprendizaje

### Conceptos y definiciones de esta clase:



- Orígenes de la Estadística
- Ramas de la Estadística
- Experimento
- Unidad Experimental
- Medir
- Población
- Muestra

- Variables
- Dato Estadístico
- Tipos de datos
- Uso de tablas
- Gráficos
- Distribución de frecuencias

## 1. Un poco de historia

Los orígenes de la estadística son muy antiguos. Se han encontrado pruebas de recolección de datos en civilizaciones como: la egipcia, la china, y la sumeria.

Desde que los pueblos se organizaron como Estados, sus gobernantes necesitaron estar informados sobre su población: cantidad de nacimientos, muertes, bienes de cada miembro de la población, producción agrícola, ganadera, con el objeto de recaudar impuestos o de analizar las condiciones de vida de la gente, para poder tomar decisiones. De allí viene el nombre de **Estadística**, ya que esta ciencia fue una importante herramienta del *Estado*.



Tradicionalmente se ha clasificado a la estadística en dos ramas:

- La *estadística descriptiva*

- La *estadística inferencial*.

La **estadística descriptiva** se utiliza para organizar, resumir y presentar datos de forma que sean fáciles de interpretar.

La **estadística inferencial** trata de obtener conocimientos sobre una *población*, con el objetivo de tomar decisiones. Estos conocimientos los obtiene a partir de la información disponible de un subconjunto de dicha población, llamada *muestra*. La estadística inferencial utiliza, para poder elaborar conclusiones el cálculo de Probabilidades.

A partir de los años '60 la enseñanza de la matemática se caracterizó por una fuerte matematización, y la probabilidad y la estadística no escaparon a esto. Esto provocó, que no se pusiera gran énfasis en el estudio de casos prácticos.

Actualmente, con el gran desarrollo de la informática y la utilización de paquetes estadísticos, se pueden manejar una gran cantidad de datos y no es precisa la utilización de algunas distribuciones de probabilidad conocidas, lo que no implica que no deban comprenderse. Si podemos decir que estamos ante una nueva filosofía en el estudio de esta disciplina y que podríamos encararla desde *el análisis exploratorio de datos*.

## 2. Recolectar datos

### 2.1 Introducción de algunos conceptos

En esta compleja realidad que nos toca vivir, los usos de la estadística son ilimitados. Si pensamos en las ciencias al realizar un experimento precisamos recolectar datos. Si pensamos en auditar una empresa también es indispensable recolectar datos y lo mismo en una investigación de mercados.

Para poder comenzar el estudio de la estadística debemos conocer algunos conceptos básicos.

Introduciremos los siguientes conceptos básicos:

- **Experimento o encuesta:** es la observación planeada de un fenómeno cualquiera con el objetivo de conocer o aproximarnos a conocer acerca de su comportamiento para poder describirlo y luego tomar decisiones. Podemos decir que el experimento incluye las actividades para obtener datos.
- **Unidad Experimental:** es cada uno de los entes que son observados en el experimento
- **Medir:** es asignar un símbolo o un número a determinadas características de las unidades experimentales.

Sigamos introduciendo conceptos básicos:

***Población:*** Es el conjunto de individuos, objetos o eventos cuyas propiedades serán estudiadas. Las poblaciones pueden ser finitas o infinitas. Los valores de una variable que se estudia en un universo. Si el universo es finito la población es finita. Si es infinito entonces las poblaciones que se originan son infinitas.

**Muestra:** Es un subconjunto de una población. Generalmente se habla de muestra de tamaño "n".

### **Ejemplo 1**

- 1- Población: Todos los alumnos de una universidad.
- Muestra: Los alumnos de la carrera de medicina.

En este ejemplo, alguna de las variables de estudio podría ser:

- 1- La edad de los alumnos.
- 2- Nacionalidad de los alumnos.

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**.

**Dato Estadístico:** es el resultado de la medición. Se dice también que es el valor de la variable.

Ahora que ya hemos definido algunos conceptos básicos para acercarnos al lenguaje estadístico vayamos a los datos estadísticos.

---

## 2.2 Tipos de datos

---

Pueden ser de dos tipos:

- **Datos Estadísticos Cualitativos:** son las propiedades categóricas que pueden usarse para describir un ente. Un ejemplo de dato cualitativo es la nacionalidad de una persona.
- **Datos Estadísticos Cuantitativos:** son los que muestran las diferencias posibles entre los valores. Un ejemplo de un dato cuantitativo es la presión arterial de una persona.

Para comenzar a integrar los conceptos a situaciones posibles le proponemos la realización de la siguiente actividad.

### Actividad para la Facilitación de los Aprendizajes

---

#### **Situación Problemática 1**

Un productor agropecuario debe tomar decisiones acerca de que tipo de insecticidas aplicar para combatir la mosca del durazno considerando el nivel de enfermedad que presentan los árboles.

- Identifique la unidad experimental.

#### **Situación Problemática 2**

Imagine una situación problemática que le sea significativa respecto de su actividad laboral o cotidiana, e identifique la unidad experimental.

#### **Situación Problemática 3**

En las siguientes situaciones identificar la unidad experimental:

1. Quiero obtener información acerca de ciertas características de los recién nacidos cuyas madres tienen HIV.
2. Quiero obtener información acerca de la rentabilidad de un monte que tiene 2000 plantas de durazno

#### **Situación Problemática 4**

Una consultora económica quiere estudiar las condiciones laborales de empresas con menos de 40 empleados.

Selecciona 60 empresas a las que les consultará acerca del monto de remuneraciones que abona.

1. Identifique la población y la muestra del estudio.

#### **Situación Problemática 5**

Las autoridades académicas de la Universidad han decidido estudiar el perfil de los estudiantes que cursan en la Facultad de Medicina que cuenta con 600 alumnos.

Se les consultará para definir el perfil las siguientes características: barrio donde vive, estado civil, cantidad de hijos, peso, estatura, edad, grado de satisfacción con la carrera, cantidad de materias aprobadas del total de materias de la carrera, ocupación, ingreso.

1. Identifique en qué consiste el experimento
2. Identifique la unidad experimental
3. Identifique la población

Identifique cada una de las variables.

---

### 2.3 Tipos de variables

---

En esta sección, comenzaremos con un ejemplo para su comprensión.

#### **Ejemplo 2**

Supongamos que preguntamos en una escuela a 50 alumnos de diferentes años cuantas materias no tiene aprobadas, obtenemos el siguiente cuadro:

2	1	2	2	1	2	4	2	1	1
2	3	2	1	1	1	3	4	2	2
2	2	1	2	1	1	1	3	2	2
3	2	3	1	2	4	2	1	4	1
1	3	4	3	2	2	2	1	3	3

Podemos observar que la variable que en este caso será la cantidad de materias desaprobadas y toma valores comprendidos entre 1 y 4

Resumamos estos datos en la siguiente tabla:

Materias desaprobadas	Cantidad de Alumnos
1	16
2	20
3	9
4	5
<b>Total</b>	<b>50</b>

### **Ejemplo 3**

Supongamos que en una clase de 30 alumnos se ha preguntado el número de hermanos que tienen cada uno, el resultado ha sido el siguiente:

2	1	1	0	1	2	1	5	3	6
4	2	0	3	0	1	1	2	3	4
1	2	1	1	1	2	0	3	1	1

Podemos observar que la variable que en este caso es la cantidad de hermanos, toma valores comprendidos entre 0 y 6

Si presentamos esta información en una tabla:

Número de Hermanos	Cantidad de Alumnos
0	4
1	12
2	6
3	4
4	2

5	1
6	1
<b>Total</b>	<b>30</b>

Entonces en los ejemplos anteriores las variables toman valores que resultan de un experimento estadístico se denominan **variables estadísticas**. Estas variables estadísticas de estudio se clasifican en:

<b>Variables cualitativas</b>		<p>Producen respuestas categóricas es decir relacionadas con <b>categorías</b>. Toman valores que no constituyen un espacio métrico, es decir que no puedo establecer una distancia entre dos valores cualquiera de ellos. Se miden en escala nominal u ordinal.</p> <p>Por ejemplo la respuesta a una pregunta por si o por no.</p>
	•	<p><b>Variables cualitativas Nominales.</b> Cuando los valores que toma permiten clasificar a la población en <b>categorías</b> o <b>clases</b> mutuamente excluyentes, sin que éstas tengan un orden específico.</p> <p>Por ejemplo: las variables <i>Grupo Sanguíneo</i>, <i>Sexo</i>, <i>Estado Civil</i>.</p>
	•	<p><b>Variables cualitativas Ordinales.</b> Cuando los valores que toma pueden ser ordenados, pues existe una <b>relación de orden</b> entre las distintas clases o categorías.</p> <p>Por ejemplo: la variable <i>Estado de un paciente</i> que puede tomar los valores: <i>menos grave</i>, <i>grave</i>, <i>muy grave</i>.</p>
	•	<p><b>Variables cualitativas Ordinales de Escala.</b> Son variables cualitativas ordinales a las cuales se les asocia un valor numérico. Por ejemplo: la variable <i>Calificación de un alumno</i> que toma los valores <i>M</i>, <i>R</i>, <i>B</i>, <i>MB</i>, <i>E</i>. A cada uno de estos valores le asociamos un número:</p> <p><math>M \rightarrow 0</math> <math>R \rightarrow 1</math> <math>B \rightarrow 2</math> <math>MB \rightarrow 3</math> <math>E \rightarrow 4</math></p>
<b>Variables cuantitativas</b>		<p>Toman <b>valores</b> que constituyen un espacio métrico. Se pueden realizar operaciones algebraicas entre ellos. Se miden en escala de intervalo o de razón.</p>
	•	<p><b>Variables cuantitativas discretas:</b> pueden tomar solo algunos <b>valores de un determinado intervalo de números reales</b>. Se originan en conteos o cuando establecemos ciertas restricciones al medir magnitudes.</p> <p>Por ejemplo: <i>La cantidad de pacientes</i> que ingresaron a un hospital en un día determinado. <i>El número de bacterias</i> que existe en 1 litro de caldo de cultivo. <i>El sueldo de una persona</i> "sin centavos", esto es una restricción.</p>

**Variables cuantitativas continuas:** Cuando pueden tomar **todos los valores** de un determinado intervalo de números reales. Se originan cuando se miden magnitudes: volumen, longitud, superficie, tiempo, dinero, etc. El recorrido es infinito.

Por ejemplo: *La altura de un grupo de alumno, el tiempo de duración de una lámpara.*

Esperamos que la siguiente actividad le ayude a interpretar los conceptos explicados.

### Actividad para la Facilitación de los Aprendizajes

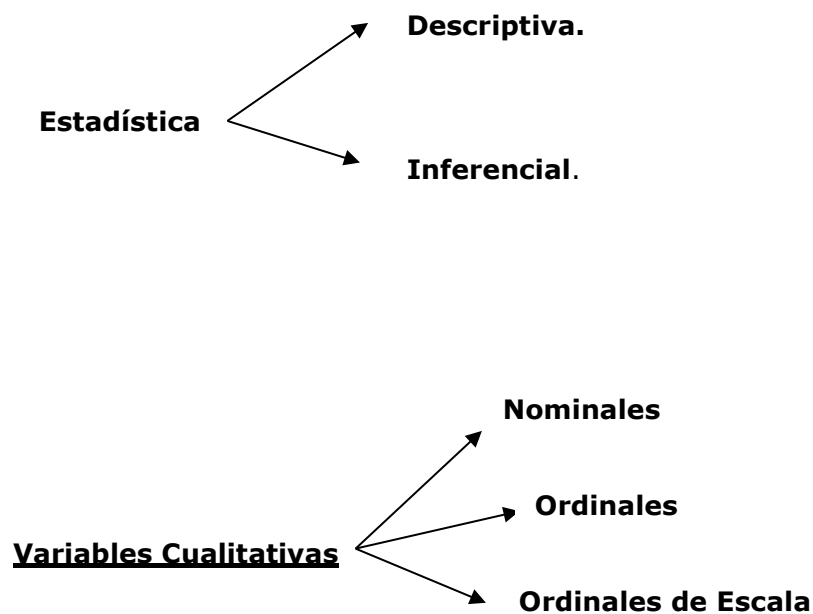
#### Situación Problemática 6

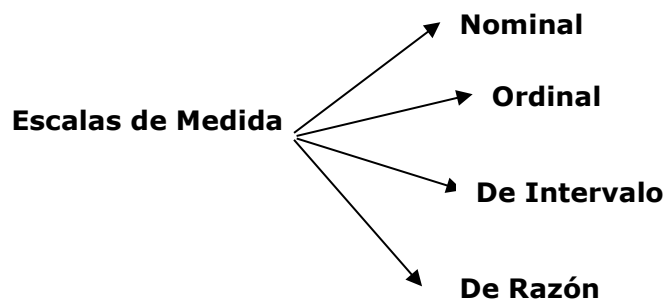
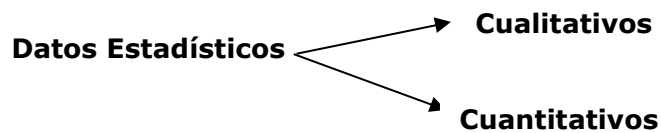
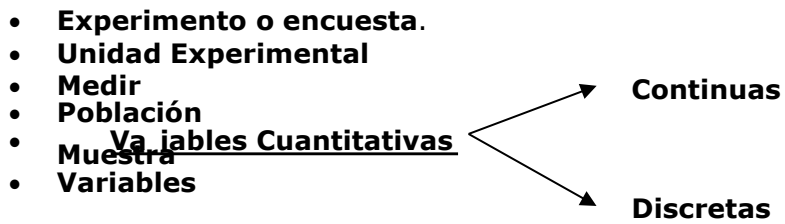
Las autoridades académicas de la Universidad han decidido estudiar el perfil de los estudiantes que cursan en la Facultad de Medicina que cuenta con 600 alumnos.

Se les consultará para definir el perfil las siguientes características: barrio donde vive, estado civil, cantidad de hijos, peso, estatura, edad, grado de satisfacción con la carrera, cantidad de materias aprobadas del total de materias de la carrera, ocupación, ingreso.

1. Identifique cada una de las variables. Clasifíquelas.

En el siguiente esquema con flechas a manera de síntesis de lo visto, hay algunos conceptos sin flechas. Intente relacionarlos.





### 3. 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.

#### 3.1 Tablas y Gráficos

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:

#### **Ejemplo 4**

Si quisiéramos saber la cantidad de estudiantes de distintos sexos que hay dentro de una determinada universidad, codificaríamos de la siguiente manera

- 1 indica femenino
- 2 indica masculino



Luego confeccionamos una **tabla o cuadro estadístico**.

<b>SEXO</b>	<b>CÓDIGO</b>	<b>CANTIDAD O FRECUENCIA</b>
Femenino	1	5.000
Masculino	2	6.000

### **Ejemplo 5**

Se desean registrar los siguientes datos correspondientes a los empleados de una empresa:

<b>NOMBRE DEL EMPLEADO</b>	<b>CODIGO</b>	<b>FECHA NACIMIENTO</b>	<b>SALARIO</b>
Rica, Oscar	2	03-03-92	\$ 57000
Roca, Pedro	2	10-10-92	\$12000
Roma, Olga	1	05-04-90	\$ 84500
Petri, Carmen	1	12-03-99	\$52000
Carpio, Raúl	2	21-06-99	\$52000
Rizo, Esteban	2	10-10-95	\$ 57000

### **Ejemplo 6**

Nos interesa estudiar el cambio de frecuencia cardiaca observado luego de la realización de una prueba física.

Se seleccionan 40 individuos y todos realizan una prueba de esfuerzo de 12 minutos.

Se calculan los incrementos de la frecuencia cardiaca como la frecuencia registrada luego del ejercicio menos la frecuencia registrada antes del ejercicio.

Los datos se representan en la siguiente tabla:

HOMBRES			MUJERES		
F. INICIAL	F. FINAL	EDAD	F. INICIAL	F. FINAL	EDAD
72	108	25	77	109	50
74	113	50	87	114	40
80	112	27	90	132	34
81	115	45	69	94	53
83	128	32	84	119	52
72	116	46	77	116	45
73	109	35	75	104	37
81	119	41	85	108	57
86	126	52	86	118	45
71	103	34	78	109	20
84	112	55	75	102	31
79	111	23	76	112	54
78	115	58	87	126	20
82	116	32	80	112	36
81	122	58	93	131	37
72	111	53	79	110	23
81	128	25	79	102	52
87	112	51	75	103	20
72	116	19	86	124	54
65	109	25	90	124	52

### Actividad para la Facilitación de los Aprendizajes

Responda las siguientes consignas:

1. Clasifique las características observadas en los ejemplos dados, según su **escala de medida** y decidir el **tipo de variable**.
2. Si agrupamos a los empleados de la empresa involucrada en el ejemplo en salario: alto, mediano y bajo ¿Qué tipo de escala de medida estamos utilizando?

### Ejemplo 7

El departamento de capacitación de una empresa desea evaluar el desempeño de los operarios a su cargo. Para ello se recogieron los siguientes datos:

- En el turno **mañana** trabajan 9 operarios de **excelente** rendimiento, 17 de **buen** rendimiento, 12 de rendimiento **regular** y 6 de **mal** rendimiento.
- En el turno **tarde** trabajan 11 operarios de **excelente** rendimiento, 20 de **buen** rendimiento, 13 de rendimiento **regular** y 9 de **mal** rendimiento.
  - En el turno **noche** trabajan 2 operarios de **excelente** rendimiento, 9 de **buen** rendimiento, 5 de rendimiento **regular** y 6 de **mal** rendimiento.

Representamos los datos en un cuadro estadístico.

RENDIMIENTO	MAÑANA	TARDE	NOCHE
EXCELENTE	9	11	2
BUENO	17	20	9
REGULAR	12	13	5
MALO	6	9	11
<b>TOTAL</b>	<b>44</b>	<b>53</b>	<b>27</b>

A continuación, retomaremos los ejemplos propuestos para mostrarle como 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

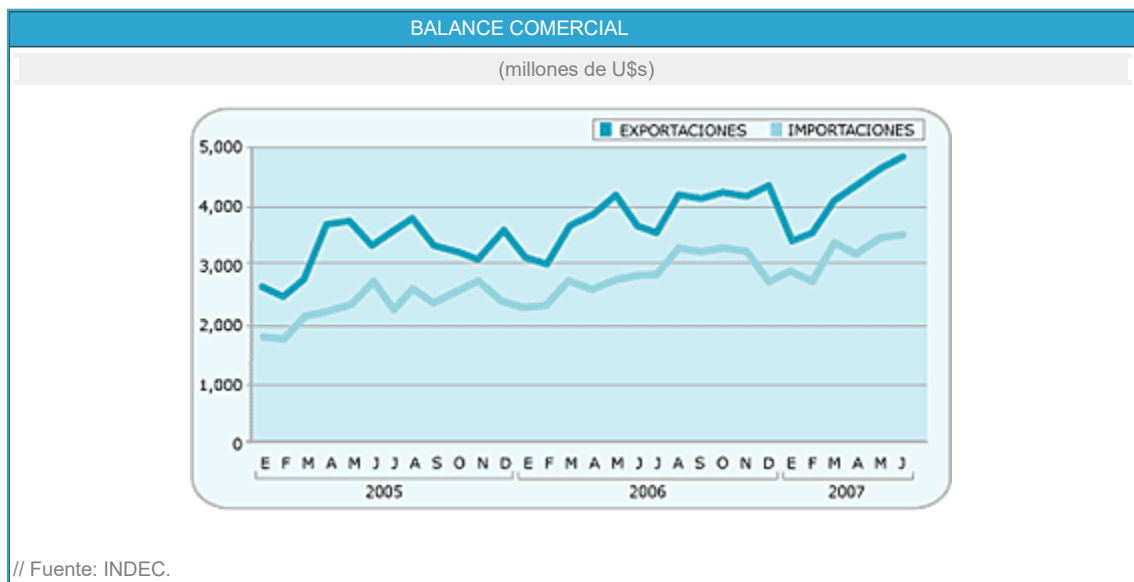
Veamos las características de cada uno.

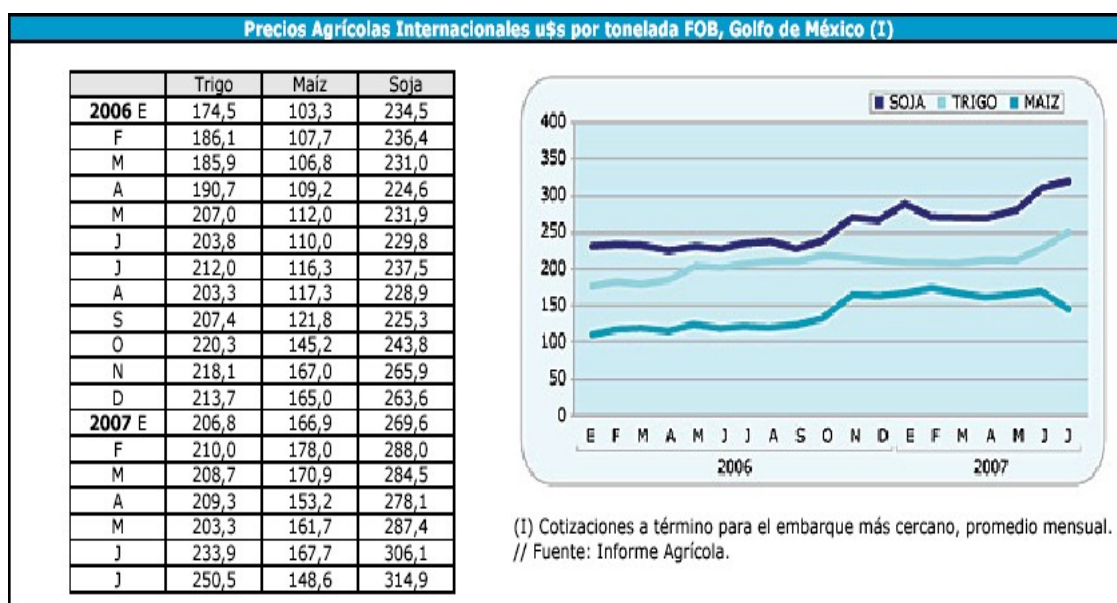
#### ➤ Gráfico Lineal

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

#### **Ejemplo 8**

De acuerdo a la información suministrada en [http://www.consejo.org.ar/frames/index\\_iec.htm](http://www.consejo.org.ar/frames/index_iec.htm) podemos realizar un gráfico lineal.





## Actividad para la Facilitación de los Aprendizajes

### Situación Problemática 7

- ¿Cuáles son las variables involucradas en el gráfico del Balance comercial?
- De junio del 2006 a junio del 2007 ¿Cuáles fueron los meses de mayor exportación? ¿Y de Importación?
- ¿En qué meses se exportaron entre 2.000 y 3.000 USD?
- ¿En qué meses se importaron entre 3.000 y 4.000 USD?

### Situación Problemática 8

- ¿Cuáles son las variables involucradas en el gráfico de Precios Agrícolas Internacionales?
- ¿En qué mes y año la soja alcanza su máximo valor?, ¿el trigo?, ¿Y el maíz?
- En septiembre del 2006 ¿Cuánto fue la diferencia de precio entre la Soja y el Trigo?
- ¿Cuál es el año y mes en donde el maíz comienza a bajar su precio?

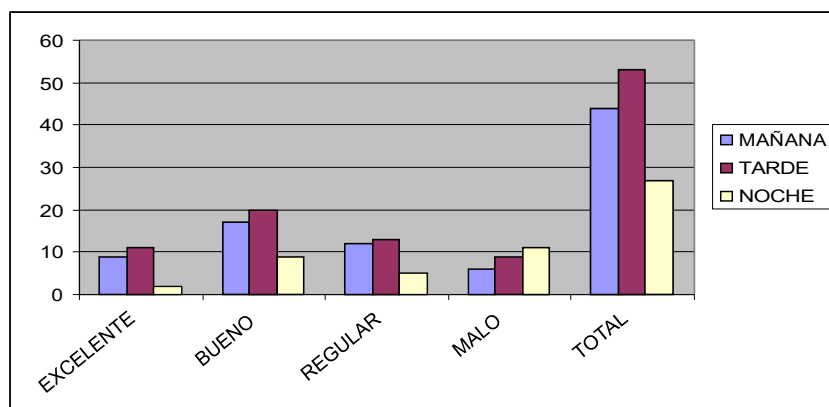
## ■ Gráfico de Barras

Se utiliza por lo general cuando una de las variables es cualitativa.

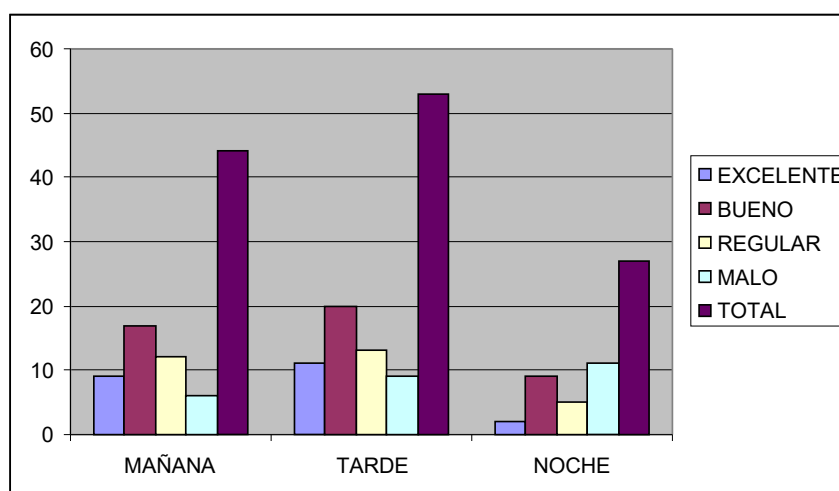
Está formado por rectángulos cuya base se encuentra indistintamente en cualquiera de los dos ejes cartesianos. Estas barras pueden estar segmentadas o agrupadas.

Del Ejemplo 7 obtenemos los siguientes gráficos de barras.

- Representación del rendimiento de los operarios.



b) Representación de los operarios según el turno



### Actividad para la Facilitación de los Aprendizajes

Observe los gráficos que le presentamos más arriba.

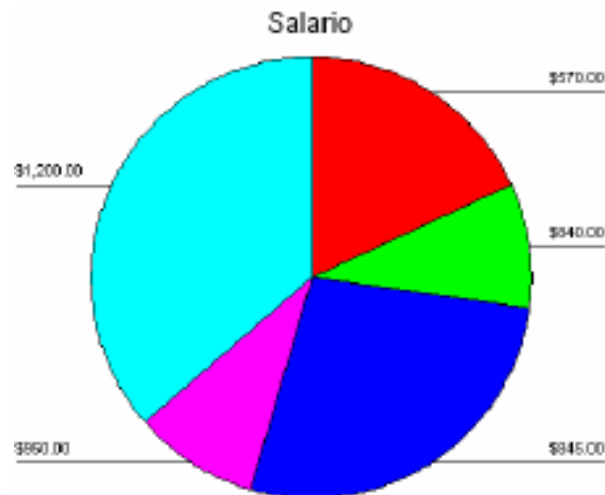
1. Enumere las diferencias del gráfico de barras del inciso a) y del inciso b)
2. ¿Qué se compara en el gráfico del inciso a)?
3. ¿Qué se compara en el gráfico del inciso b)?

### • 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.

Analice el siguiente grafico



¿Cuál es la variable que representa el gráfico? Clasifíquela.

¿Qué información obtiene mediante este gráfico? Enúnciela.

Realice un Grafico Circular que sea significativo para usted.

Existen otro tipo de gráficos para representar distintas situaciones. Lo más importante es poder determinar el tipo de gráfica que se elige de acuerdo a lo que tenemos que mostrar. De la elección dependerá si el lector comprende lo que quiere mostrarse.

### 3.2 Distribución de frecuencias

Los datos - o sea los 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**.

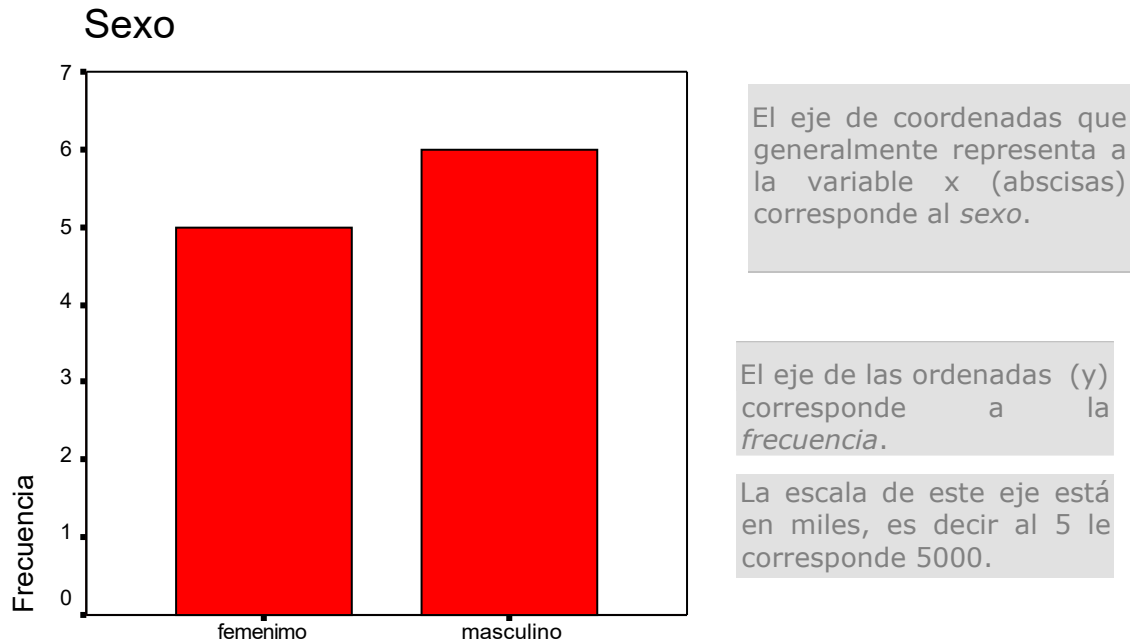
Muchos textos dividen los datos de esta forma, en agrupados y no agrupados. Lo cierto es que con los datos desordenados o no agrupados es muy difícil trabajar ya que no se ponen en evidencia fácilmente ciertas **regularidades** que precisa la estadística y, en general toda la matemática, para obtener 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 éstos aparecen o son observados se la denomina **Distribución de frecuencias** y se la representa en una tabla o en un gráfico estadístico.

Recordemos esta tabla ya presentada anteriormente:

SEXO	CÓDIGO	CANTIDAD O FRECUENCIA
Femenino	1	5.000
Masculino	2	6.000

Representemos esta misma situación en un gráfico de barras.



**Sexo**

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

- a. En el eje horizontal se marcan los valores que toma la variable de estudio.
- b. En el eje vertical se representa la frecuencia con que ocurren esos valores o la cantidad de veces que ocurren esos valores.
- c. 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.

En este histograma de frecuencia ¿qué tipo de variables intervinieron?

