

# Estadística

Podemos dividir el quehacer de la estadística en tres grandes áreas:

- ✓ Estadística Descriptiva.
- ✓ Estadística Inferencia (Prueba de hipótesis)
- ✓ Teoría de Muestreo.



La base matemática de la estadística es la teoría de la Probabilidad.

**Probabilidad:** El objetivo fundamental de la teoría de la probabilidad es desarrollar modelos y estudiar matemáticos para “experimentos= fenómenos (aleatorios)” cuyos resultados no pueden predecirse con exactitud.

Dichos modelos matemáticos son llamados “Espacios de probabilidad” ( espero dar una idea formal de este concepto el la 5 o 6 semana). Esto dependerá de ver su madures matemática.

## Fenómenos aleatorias vs Fenómeno aleatorio

**Fenómeno aleatoria:** fenómeno del que no se sabe que es lo que va a ocurrir ( es decir, no se sabe con certeza cual de los posibles resultados va a ocurrir). Por ejemplo lanzar una moneda, extraer de un lote de fabrica, alguna pieza para determinar si esta bien o esta defectuosa, lanzar un dado,

Fenómeno determinista: Se sabe de antemano cual será el resultado del experimento. Las leyes de la física clásica (Mecánica) son de esta naturaleza.

Ejemplos: Del día viene la noche, lanzar una piedra y esta cae al suelo.

Ganarme la lotería, ¿Es un fenómeno determinismo?.

Un experimento es una “acción” que se realiza (obteniendo resultados= Llamados **Población**) y tiene como ultimo fin el determinar la probabilidad de que ocurra o no un un determinado subconjunto de resultados (Este subconjunto se le llama **Evento**).

1. Se puede repetir indefinidamente, siempre en las mismas condiciones;
2. Antes de realizarlo, no se puede predecir el resultado que se va a obtener;
3. El resultado que se obtenga, s, pertenece a un conjunto conocido previamente de resultados posibles.

Ejemplo 0 : lanzar una moneda

1. Conjunto de resultados  $\{1,0\}$  1= Águila , 0 = Sol.

Notemos que se cumplen las tres condiciones anteriores.

Ejemplo 1: Lanzar un dado:

Conjunto de resultados = {1, 2, 3, 4, 5}, ¿Es un experimento aleatorio?.

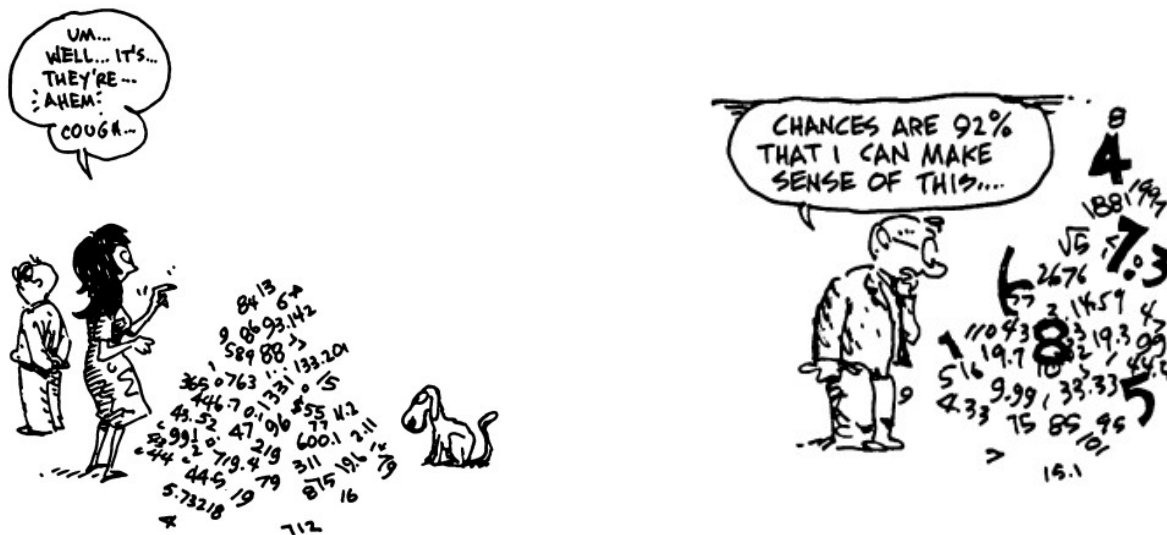
Más ejemplos:

- A. El tiempo de llegada de un avión.
- B. La calificación en un examen de un alumno
- C. La cantidad de gramos que contiene un frasco de café
- D. La cantidad de defectos de un artículo

## Estadística Descriptiva

¿Pero que es la estadística descriptiva?

### Descripción de Datos



En esta parte nos centraremos en la descripción de datos y en representarlos de una forma útil.

1. - ¿Que necesitamos para hacer esto?.

2. - ¿Cómo podemos ver el patrón subyacente en un montón de números desnudos?

3. ¿Cómo podemos resumir la forma básica de los datos?

Así, para describir datos , lo primero que necesitamos son .....!!!! **DATOS!!!!**.

**Ejemplo 0.** En una universidad de Estados Unidos se pide a 92 estudiantes que proporcionen su peso (en libras), con lo cual se obtiene la siguientes datos.

**Hombres:**

140 145 160 190 155 165 150 190 195 138 160 155 153 145 170 175 175 170 180 135  
170 157 130 185 190 155 170 155 215 150 145 155 155 150 155 150 180 160 135 160  
130 155 150 148 155 150 140 180 190 145 150 164 140 142 136 123 155

**Mujeres:**

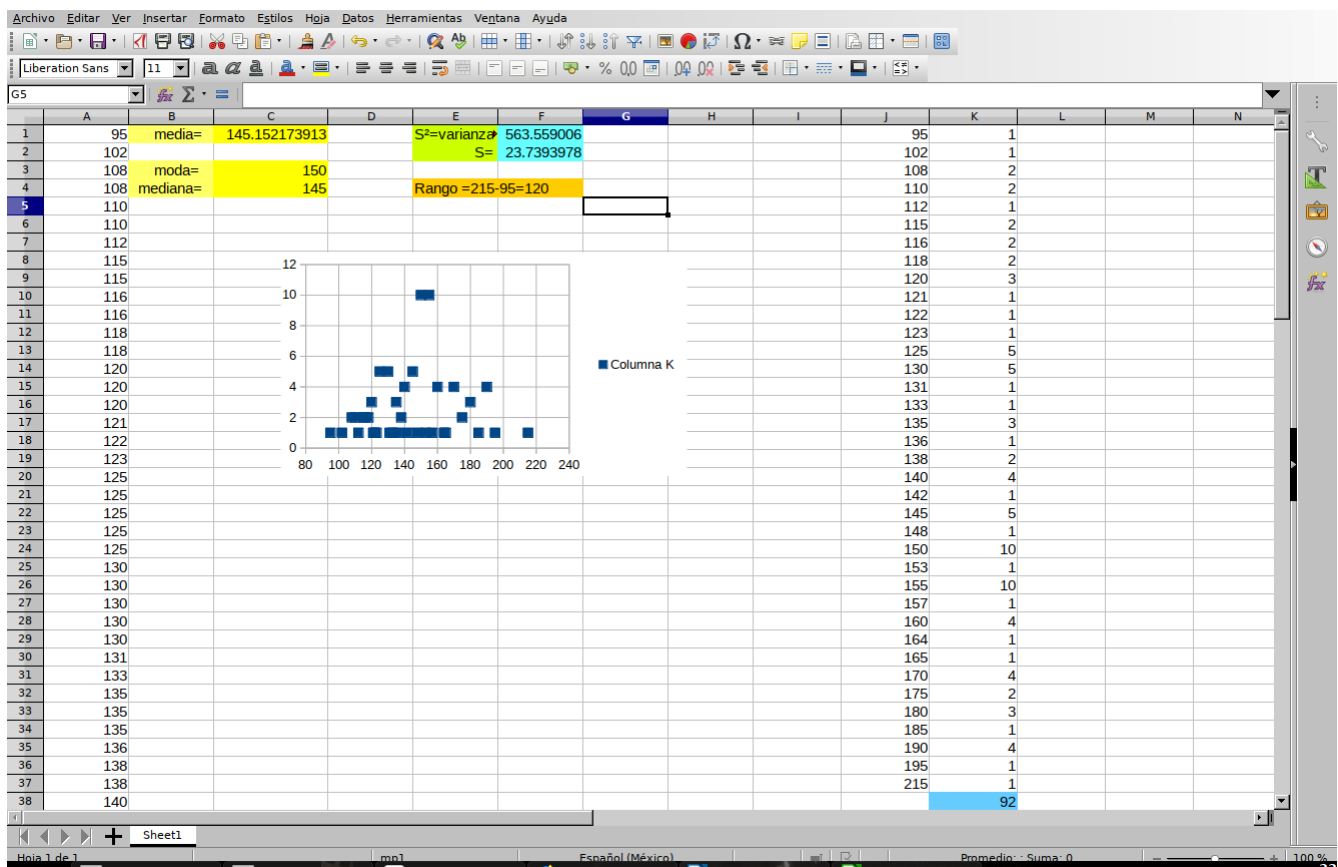
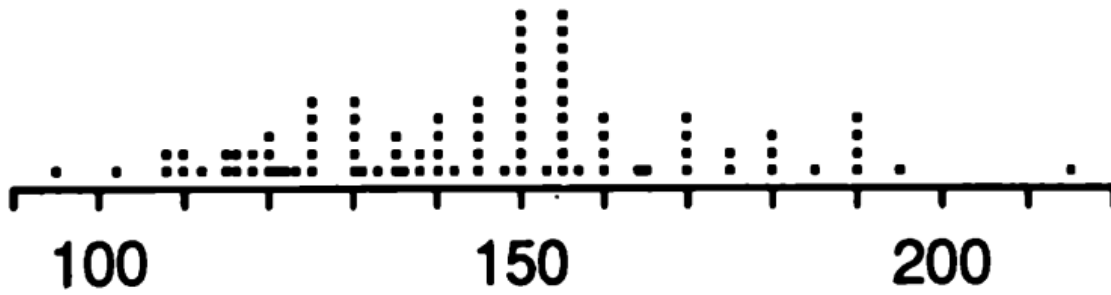
140 120 130 138 121 125 116 145 150 112 125 130 120 130 131 120 118 125 135 125  
118 122 115 102 115 150 110 116 108 95 125 133 110 150 108

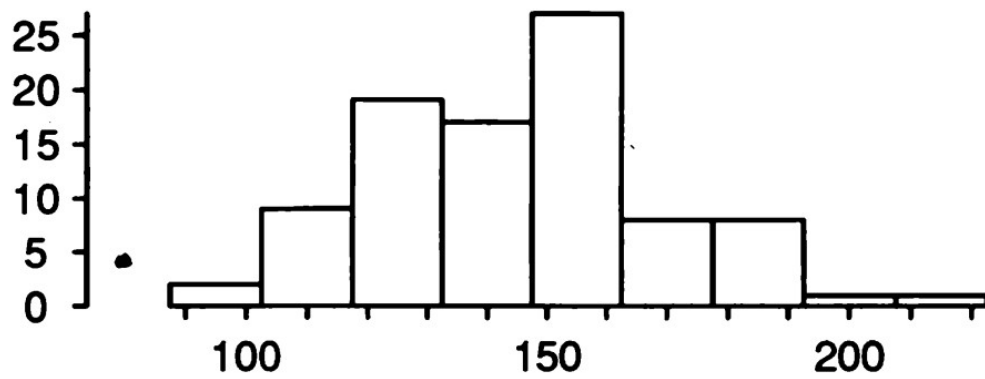
## **Tablas de Frecuencias.....**

**La Tabla de frecuencia de los datos de peso:** Dividamos la recta numérica en intervalos y contamos el número de pesos de los estudiantes, dentro de cada intervalo.

**Guia para formar los intervalos de clase:**

- I) Use intervalos de igual longitud con puntos medios en números redondos convenientes.
- II) Para un número pequeño de datos, usar un número pequeño de intervalos.
- III) Para un número grande de datos , usar mas intervalos...
- IV) se recomienda usar para **n** datos su **raíz cuadrada**. Para nuestro ejemplo sería de 9 a 10 intervalos.





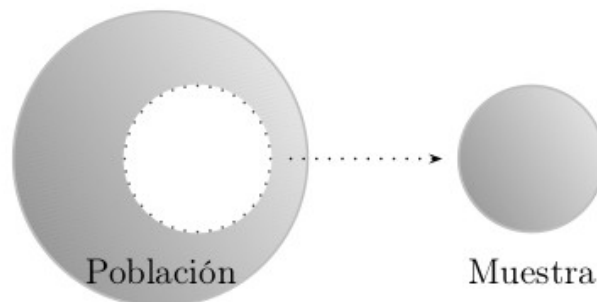
**Histograma**

## Conceptos elementales

### Población y muestra

**Def:** Una **población** es un conjunto de personas, objetos o eventos, de los cuales nos interesa estudiar algunas de sus características.

**Def:** Una muestra es cualquier subconjunto de una población. Al número de elementos de la muestra se le llama tamaño de la muestra.



**Nota:** Dicha muestra debe de ser “Representativa”, se suele decir que es “Muestra aleatoria”.

¿Pero que quiere decir que la muestra sea representativa o aleatoria?.

## Variables y datos

A lo que nos interesa medir y registrar en cada elemento de una población le llamaremos variable. En general, una variable es una característica que varía de un elemento a otro de la población.

**Def:** Una **variable** es una característica de interés que posee cada elemento de una población y que podemos medir.

**Ojo:** Una variable también puede considerarse como una pregunta que se le hace a cada elemento de la población.

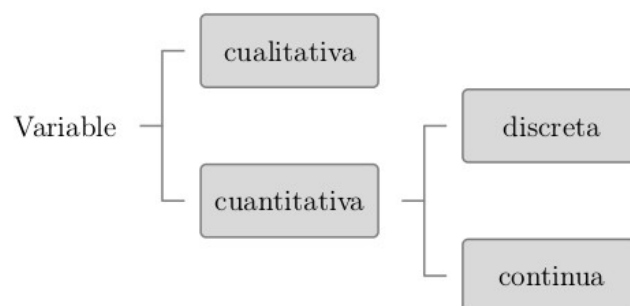
**Ejemplo:** ¿Usted fuma? y obtener como respuesta “sí” o “no”.

**Def:** Mediante el término **datos** se entiende al conjunto de observaciones de una o varias variables de interés para todos los elementos de una muestra.

## Clasificación de variables

**Def:** Una **variable es cuantitativa** si sus valores son números y representan una cantidad.

**Def:** Una **variable es cualitativa** si sus valores representan una cualidad, un atributo o una categoría. Se les llama también variables categóricas.



- Una **variable cuantitativa** es discreta si el conjunto de todos sus posibles valores tiene un número finito de elementos, o bien es infinito, pero se pueden numerar uno por uno de acuerdo al conjunto de números naturales.

Ejemplo la colección  $\{0, 1, 2, \dots, 120\}$  puede ser el conjunto de valores de una variable cuantitativa discreta pues este conjunto tiene un número finito de elementos. Puede corresponder al número de hijos de una persona o el número de años promedio que le quedan por vivir a una persona.

- Una **variable cuantitativa** es continua si puede tomar todos los valores dentro de un intervalo  $(a,b)$  de números reales y no toma valores aislados.

## Tipos de datos

Tipo	Escala	Descripción	Ejemplos
Discreto	Nominal	Categorico no ordenado Sin información de intervalo	Datos binarios (dicotómicos) Categorías Policotómicos
Discreto	Ordinal	Categorico ordenado Sin información de intervalo	Escala de Likert Escala del 1 al 5 Pequeño/mediano/grande
Continuo	Intervalo	Pueden tomar cualquier valor en un intervalo No hay cero absoluto	Temperatura (grados) Posición
Continuo	Proporción	Pueden tomar cualquier valor en un intervalo Sí hay cero absoluto	Dimensiones Presión Concentración

**Def: 1.** Una escala **nominal** es una escala de medición en la cual los números sirven como “etiquetas” solamente para identificar o clasificar un objeto.

Una escala de medición nominal normalmente trata sólo con variables no numéricas (no cuantitativas).

La escala ordinal es uno de los niveles de medición que nos otorga la clasificación y el orden de los datos sin que realmente se establezca el grado de variación entre ellos. La escala de medición ordinal es la segunda de las 4 escalas de medición.

Los datos ordinales son básicamente datos estadísticos que tienen la misma naturalidad pero existe una diferencia entre ellos que es desconocida. Estos datos pueden ser agrupados o clasificados.

Por ejemplo, una escala ordinal puede responder preguntas como:

¿Qué tan satisfecho estás con nuestros productos?

1. Totalmente satisfecho
2. Satisfecho
3. Neural
4. Insatisfecho
5. Totalmente insatisfecho

¿Qué tan feliz estás con el servicio al cliente?

1. Muy infeliz
2. Infeliz
3. Neutral
4. Feliz
5. Muy feliz

**Ahora si, podemos dar una definición aproximada del quehacer de la estadística descriptiva**

## **Medidas de Tendencia Central o medidas de localización**

**Medidas de  
localización**

- Media
- Moda
- Mediana

**Def:** Sean La media o media aritmética es simplemente el promedio de los números definida por



$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

## Moda

Ahora definiremos el concepto de moda. A diferencia de la media, la moda se puede calcular tanto para valores numéricos como no numéricos.

La **moda** es el valor que aparece con mayor frecuencia en el conjunto de datos, si lo hubiera.

Por ejemplo, consideremos el siguiente conjunto de edades de 6 personas.

Edades en años	
$x_1 = 15$	$x_4 = 20$
$x_2 = 20$	$x_5 = 25$
$x_3 = 15$	$x_6 = 15$

ejemplo de una variable cualitativa que mide las condiciones de un producto y que tiene como posibles valores: Malo, Regular, Bueno. Suponga que tenemos el siguiente conjunto de 6 observaciones:

Condiciones de un producto	
$x_1 = \text{Malo}$	$x_4 = \text{Regular}$
$x_2 = \text{Bueno}$	$x_5 = \text{Malo}$
$x_3 = \text{Bueno}$	$x_6 = \text{Bueno}$

## Mediana

Otra medida de tendencia central para datos numéricos es la mediana. Supongamos nuevamente que tenemos una colección de  $n$  números

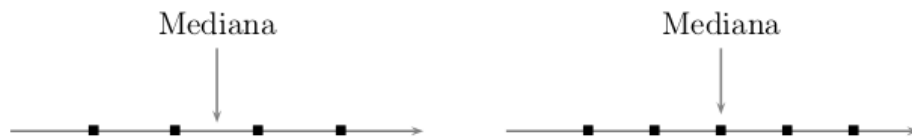
$$x_1, \dots, x_n.$$

Podemos ordenar estos números de menor a mayor, incluyendo repeticiones, y obtener la colección

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

La **mediana** es el dato ordenado de en medio, esto es,

- Si el número de datos  $n$  es par, entonces existen dos datos ordenados de en medio y la mediana es el promedio de estos dos números, esto es  $(x_{(n/2)} + x_{(n/2)+1})/2$ .
- Si el número de datos  $n$  es impar, entonces el dato ordenado de en medio es  $x_{(n+1)/2}$  y esta es la mediana.



**Ejemplos:** Estatura de 6 personas

Estaturas en centímetros					
165	172	170	165	174	182

Ordenando estos números de menor a mayor, incluyendo repeticiones, se obtiene el siguiente arreglo:

165	165	170	172	174	182
-----	-----	-----	-----	-----	-----

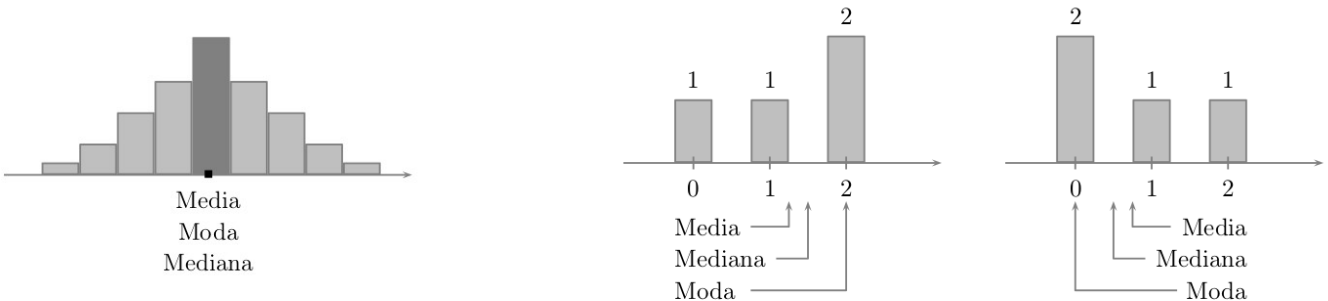
Como se trata de un número par de datos, la mediana es el promedio de los dos datos centrales, esto es,

$$\begin{aligned}\tilde{x} &= \frac{170 + 172}{2} \\ &= 171.\end{aligned}$$

## ¿Qué medida de localización es mejor?

No existe tal cosa. Cada una de las medidas de localización que hemos mencionado mide de manera diferente la centralidad de un conjunto de datos numéricos. Sin embargo, **la media** es la medida de localización que con mayor frecuencia se utiliza en los estudios estadísticos.

En gráficas de frecuencias se puede ver lo siguiente.



## Medidas de dispersión

**Objetivo:** estudiaremos algunas cantidades que permiten medir el grado de dispersión de un conjunto de datos numéricos

- En casi todas estas medidas de dispersión es necesario considerar un valor central de los datos como punto de referencia. Como tal valor central tomaremos a la media  $\bar{x}$ .

## Varianza (Poblacional, ¡ojo con la n!)

La varianza es un promedio de la distancia al cuadrado de cada uno de los datos  $x_i$  respecto de la media  $\bar{x}$  y es la medida de dispersión más comúnmente usada.

**Definición:** Se calcula de la forma siguiente

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

**Nota:** para muestras, se divide entre  $n-1$ .  
Ejemplo: Consideremos el peso de 6 personas:

Pesos en kilogramos	
$x_1 = 70$	$x_4 = 66$
$x_2 = 68$	$x_5 = 70$
$x_3 = 75$	$x_6 = 65$

Puede comprobarse que la media es  $\bar{x} = 69$  y por lo tanto la varianza se obtiene como sigue:

$$\begin{aligned}
 s^2 &= \frac{1}{6}(70 - 69)^2 + \frac{1}{6}(68 - 69)^2 + \frac{1}{6}(75 - 69)^2 \\
 &\quad + \frac{1}{6}(66 - 69)^2 + \frac{1}{6}(70 - 69)^2 + \frac{1}{6}(65 - 69)^2 \\
 &= 10.666.
 \end{aligned}$$

## Desviación estándar

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

## Rango

Ahora definiremos el rango de una colección de números  $x_1, \dots, x_n$ . Para calcular esta cantidad es necesario identificar el dato más pequeño  $x_{(1)}$  y el dato más grande  $x_{(n)}$ . El rango de la colección de números dada se denota por la letra  $r$  y es simplemente el dato mayor menos el dato menor.

$$r = x_{(n)} - x_{(1)}$$