

## ¿Qué es la Estadística?

- **La Estadística** es una ciencia que proporciona un conjunto de métodos que se utilizan para recolectar, resumir, clasificar, analizar e interpretar el comportamiento de los “**datos**” con respecto a una característica materia de estudio o investigación.
- **La Estadística** es la postulación de un modelo plausible que explica el mecanismo que genera los datos.

# ¡Todo comienza con DATOS!



**Obtener DATOS ...**

**Analizar DATOS**



**Presentación de los DATOS**



## Tipos de DATOS

- **Datos Nominales**  
Son números o letras que representan categorías donde no interesa el orden (ejm., 0=masculino, 1=fememino)
- **Datos Ordinales**  
Son números o letras que representan categorías donde el orden interesa (ejm., lesión fatal=1, lesión severa=2, lesión moderada=3, etc.)
- **Datos Discretos**  
Son aquellos que surgen por el procedimiento de conteo. Es decir, los datos discretos toman valores enteros (ejm., el número de hijos por familia; el número de automoviles que pasan por una avenida en una hora, etc.)
- **Datos Continuos**  
Son aquellos que surgen cuando se mide alguna característica. Es decir, toman al menos teóricamente cualquier valor dentro de un intervalo (ejm., el peso, la estatura, la tensión arterial de las personas, etc.)

## Población y Muestra

- **Población** es la colección de todos los individuos, objetos u observaciones que poseen al menos una característica común (ejm., las edades de los estudiantes de Chile.)
- **Parámetro** es una medida resumen que describe una característica de toda la población.
- **Muestra** es una parte o un subconjunto representativo de la población. Y al proceso de obtener una muestra se llama **muestreo**.

## Métodos para obtener muestras

- **Muestras Probabilísticas** son aquellas en que todos los elementos de la población tienen una posibilidad (una probabilidad conocida) de ser incluida en la muestra. Naturalmente no es necesario que todos tengan la misma posibilidad, basta que tengan alguna posibilidad.
- **Tabla de Números Aleatorios:** un procedimiento para extraer una muestra aleatoria, es empleando una tabla de números aleatorios, el cual puede construirse, empleando el método de la urna con 10 tarjetas numeradas: 0,1,2,...,9. Después que se mezclan las tarjetas, se extrae aleatoriamente una tarjeta y se registra su número. La tarjeta se reemplaza antes de la extracción siguiente, se mezclan bien las tarjetas y así, sucesivamente. En la práctica, se realiza mediante una computadora, por un procedimiento completamente al azar.

### métodos para obtener muestras (continuación)

- **Muestreo Aleatorio Simple** este método o esquema de muestreo, se caracteriza porque todos los elementos de la población tienen la misma probabilidad de ser incluidos en la muestra, o en otros términos, porque todas las posibles muestras de un tamaño fijo son igualmente probables.
- **Muestreo Sistemático** en este procedimiento, se selecciona una muestra, tomando cada  $k$ -ésima unidad de la población una vez que las unidades de muestreo están numeradas o arregladas en alguna forma.
- **Muestreo Estratificado** este se usa cuando la población no es homogénea, sino que pueden en ella identificarse clases definidas por algún atributo o característica relacionada con la variable que se estudia. Este procedimiento implica dividir la población en clases o grupos homogéneos relativos a las características que van a estudiarse, llamados **estratos**. Después se toma una submuestra de cada estrato.

## métodos para obtener muestras (continuación)

- **Muestreo por Conglomerados** se usa en poblaciones grandes y muy dispersos desde el punto de vista geográfico, y en las cuales el muestreo aleatorio simple sería poco económico debido a que daría lugar a muestras igualmente dispersas. En este tipo de muestreo, en lugar de seleccionar directamente los elementos de la población se hace una selección inicial de grupos o **conglomerados**, que son agrupaciones de elementos que deben ser lo más heterogéneo posible a diferencia de los estratos.

## Formas de resumir los DATOS

- **Tablas**

- Tablas de frecuencias (distribución de frecuencias)
  - Frecuencias relativas (porcentaje del total)

- **Gráficos**

- Histogramas
  - Polígonos de Frecuencias
  - Diagrama de Puntos
  - Diagrama de Cajas (Box plots)

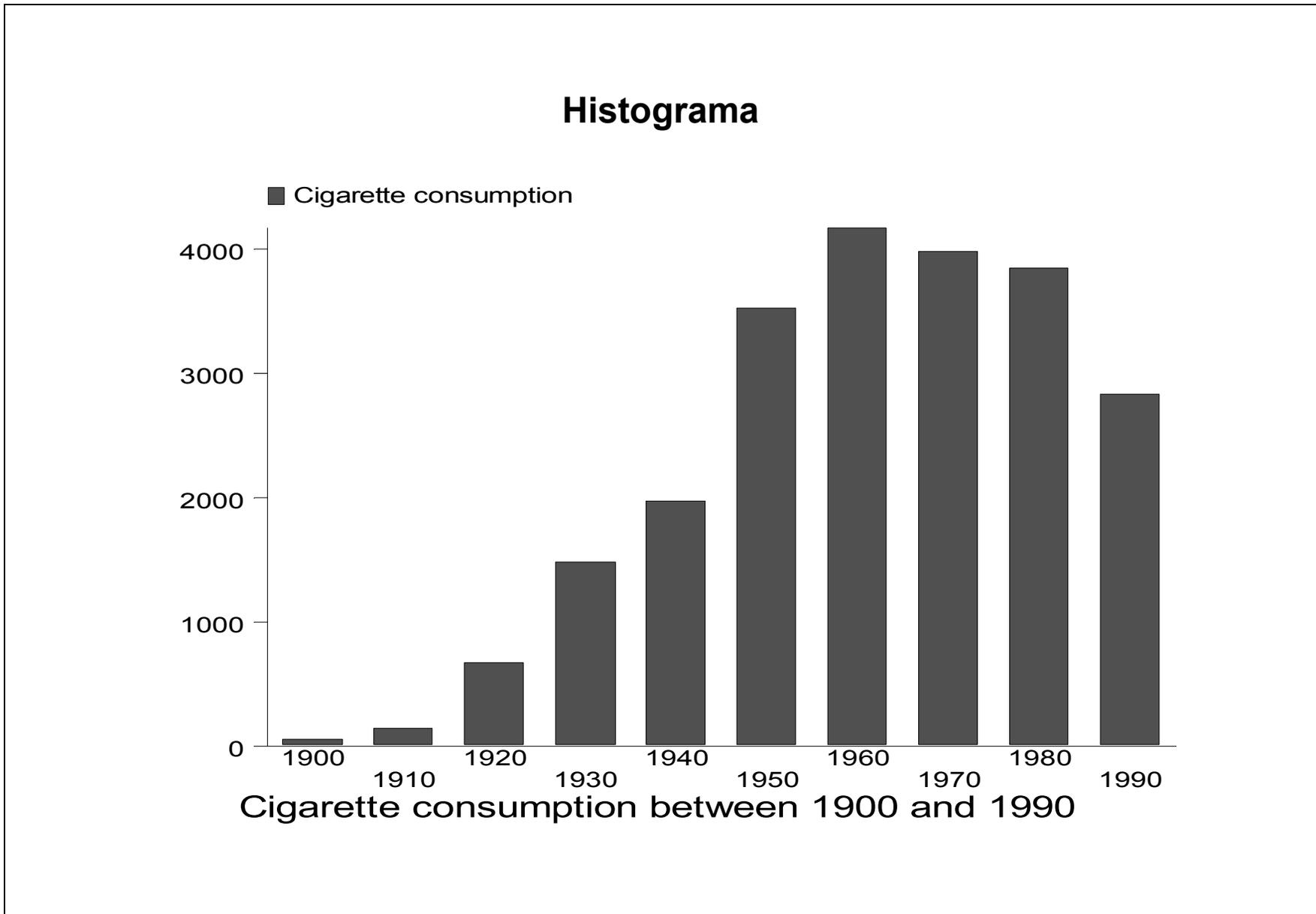
### Frecuencias de los niveles del colesterol

Nivel de Colesterol (mg/100 ml)	Frecuencia	Frecuencia Acumulada	Frecuencia Relativa (%)	Frecuencia Relativa Acumulada (%)
80-119	13	13	1.2	1.2
120-159	150	163	14.1	15.3
160-199	442	605	41.4	56.7
200-239	299	904	28.0	84.7
240-279	115	1019	10.8	95.5
280-319	34	1053	3.2	98.7
320-360	9	1062	0.8	99.5
360-399	5	1067	0.5	100.0
Total	1067		100.0	

**Nota. La elección del número de intervalos en una tabla de frecuencias es muy importante. Desafortunadamente, no hay reglas establecidas para determinarlos.**

Consumo de cigarrillos por persona (mayores de 18 años) en los EE.UU.,  
1900-1990

	Año	Número de Cigarillos
1.	1900	54
2.	1910	151
3.	1920	665
4.	1930	1485
5.	1940	1976
7.	1950	3522
8.	1960	4171
9.	1970	3985
10.	1980	3851
11.	1990	2828



## Una aplicación

Considere el siguiente conjunto de datos. Note que este es un ejemplo de datos nominales:

```
1 5 3 1 2 4 1 3 1 5
2 1 1 5 3 1 2 1 4 1
4 1 3 1 5 1 2 1 1 2
5 1 1 5 1 5 3 1 2 1
2 3 1 1 2 1 5 1 5 1
1 2 5 1 1 2 3 4 1 1
1 1 2 1 1 2 1 1 2 3
3 3 1 5 2 3 5 1 3 4
1 1 2 4 5 4 1 5 1 5
5 1 1 5 1 1 5 1 1 5
```

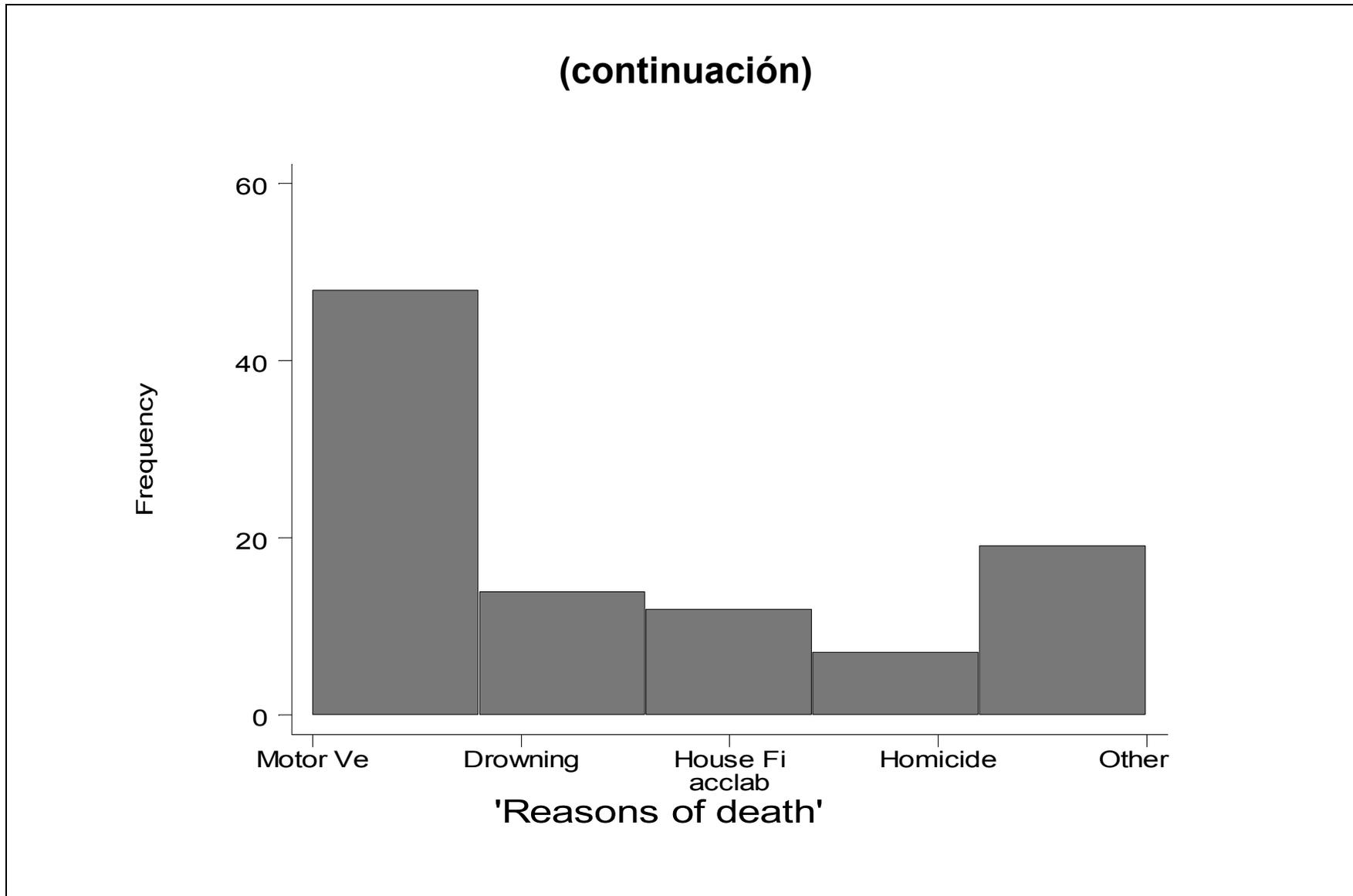
**donde:**

1. Vehicular	2. Inmersión
3. Incendio	4. Homicidio
5. Otros	

**Tabla resumen**

```
. tab accident
```

acc_lab	Freq.	Percent	Cum.
Motor Ve	48	48.00	48.00
Drowning	14	14.00	62.00
House Fi	12	12.00	74.00
Homicide	7	7.00	81.00
Other	19	19.00	100.00
Total	100	100.00	



## Otro Ejemplo

En cierto distrito de Guatemala, en el año 1969, se entrevistó a un cierto número de mujeres casadas nacidas entre los años 1935 y 1944 y se les preguntó a qué edad contrajeron matrimonio. A continuación se muestran los datos correspondientes a 50 de estas mujeres:

15	17	25	15	16	11	15	13	12	10
15	14	16	14	17	13	14	20	29	19
16	18	10	18	12	11	20	34	13	22
19	14	17	16	16	15	12	24	25	9
21	15	13	23	24	10	10	16	14	18

Describamos este conjunto de datos gráficamente haciendo uso del diagrama de puntos.

**Diagrama de Puntos**

