

# Universo, muestra, tamaño muestral y análisis estadístico



# En nuestros días ¿Qué es estadística?

Es la disciplina que se ocupa de:

- “La recolección, organización y procesamiento de datos, para obtener inferencias hacia un universo cuando se observa sólo una parte de este universo”
- “Identificar la variabilidad de un fenómeno y tratar de explicarla”
- “Tomar decisiones bajo incertidumbre”

# Método estadístico:

Proporciona las técnicas necesarias para:  
Recolectar y Analizar la información requerida.

Distingue dos etapas:

## I Planificación:

- Definición de objetivos: es la descripción formal del problema que da origen a la investigación. Se debe señalar detalladamente lo que se va a investigar: **qué, cómo, dónde, cuándo y por qué.**
- Universo del estudio: es la definición del conjunto desde el cual se extraerá la información y hacia el que se generalizarán las conclusiones obtenidas.
- Diseño de la muestra: la Teoría de Muestreo garantiza que la información que generaremos nos permita proyectar válidamente las conclusiones al Universo de interés.

# Método estadístico:

- Definición de las unidades de observación (que objetos observaremos), sus atributos y sus las escalas de medida.
- Preparación del Plan de Tabulación y Análisis de la información: especificación de las formas de presentar y analizar la información recolectada

## II Ejecución

- Recolección de la información
- Elaboración de la información
- Análisis de los resultados

# La unidad de análisis y sus atributos

- **Unidad de análisis:** Una vez definido el problema que se va a investigar, se definen naturalmente los objetos que serán observados o estudiados:
  - Seres humanos
  - Animales
  - Células
  - Órganos
  - Etcétera

# La unidad de análisis y sus atributos

- **Atributos:** Teniendo definidas las unidades de análisis, obviamente ellas presentan características que nos importan para nuestro estudio: Si nuestro estudio es antropométrico en chilenos adultos entre 15 y 90 años de edad, podemos consignar algunas características esenciales tales como:
  - Sexo
  - Estatura
  - Raza
  - Peso
  - Circunferencia de cintura

# Variables

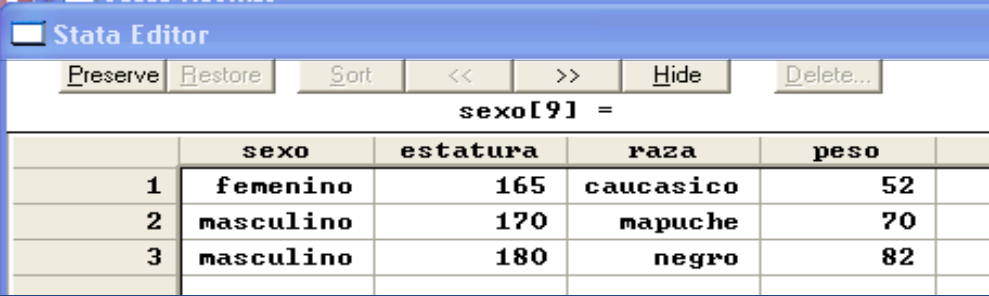
**Variables:** Cuando se han definido los atributos a estudiar, podemos ya observar unidades de análisis especificadas y los atributos quedan consignados como características únicas del objeto que estamos estudiando. Si observamos una persona en particular podemos consignar:

Sexo: Femenino

Estatura: 165 centímetros

Raza: Caucásico

Peso: 52 kilogramos



The screenshot shows the Stata Editor interface. At the top, there are buttons for 'Preserve', 'Restore', 'Sort', '<<', '>>', 'Hide', and 'Delete...'. Below these buttons, the text 'sexo[9] =' is displayed. The main part of the screenshot is a data table with the following structure:

	sexo	estatura	raza	peso
1	femenino	165	caucasico	52
2	masculino	170	mapuche	70
3	masculino	180	negro	82

Cuando los atributos ya han sido evaluados, reciben el nombre de Variables del estudio.

# Escalas de medida

Cuando procedemos a medir las variables del estudio, debemos tener presente que estamos consignando valores y por consiguiente introduciendo escalas de medición. Estas escalas pueden ser: Nominales, Ordinales o Intervalares (o de Razón). Estas escalas tienen diferente Poder de Clasificación



# Escalas de medida

Escala de medida	Capacidad
<b>Nominal</b>	Sólo es capaz de nombrar o etiquetar la unidad de análisis. Por ejemplo: Sexo, raza, nacionalidad
<b>Ordinal</b>	Es capaz de nombrar pero además introduce una jerarquía en las unidades observadas. Por ejemplo: Grado que se cursa en el sistema escolar, nivel económico, escala analógica para el dolor
<b>Intervalar y de razón</b>	Es capaz de nombrar, jerarquizar pero además permite hacer comparaciones matemáticas entre las unidades de análisis. Por ejemplo: Temperatura en grados Celcius (intervalar). Peso, estatura (de razón).  <b>En las escalas de razón el cero indica ausencia del atributo.</b>

# Escalas de medida

- **Las escalas de medida se pueden bajar pero nunca subir.** Es decir una variable en escala intervalar se puede dejar en escala ordinal y una en escala ordinal se puede dejar en escala nominal, pero una variable en escala nominal no se puede dejar en escala ordinal y una en escala ordinal no se puede dejar en escala intervalar

# Escalas de medida

Las variables medidas en escala intervalar pueden ser:

- Discretas: Asociadas a los números naturales, es decir sólo cuentan, por ejemplo: Número de hijos, células por campo
- Continuas: Asociadas a los números reales, es decir miden, por ejemplo: Peso, temperatura, edad

**Una variable continua se puede discretizar, pero una variable discreta no se puede continuizar**

Por ejemplo: la edad es una medida de tiempo y de naturaleza continua, sin embargo se registra en años cumplidos que es de naturaleza discreta. Resulta poco cómodo registrar la edad de alguien como: 30.2130 años (30 años con 2 meses, 16 días, 16 horas y 19 minutos ) es mejor contar la cantidad de velas que apagó en la torta en su último cumpleaños, 30 velas = 30 años

# Población y muestra

**Población:** Es el Conjunto Universo de las unidades de análisis, la población puede ser de finito o infinito:

- El conjunto de estrellas del cosmos es una población infinita
- El conjunto de eritrocitos en el torrente sanguíneo de un ser humano es una Población Finita, pero es prácticamente infinita (no factible).
- Las personas residentes en Chile con VIH+ son una población es finita (factible).

# Población y muestra

## MUESTRA:

Es un **SUBCONJUNTO FINITO** y **FACTIBLE** de la Población, que debe cumplir características ineludibles para lograr que las conclusiones de la inferencia estadística sean válidas.

- **Aleatoria:** garantiza que los elementos que componen la muestra fueron escogidos completamente al azar, es decir no hay predilección alguna por incluir o excluir determinada unidad de análisis (todos los sujetos de una población tienen la misma probabilidad de integrar la muestra)
- **El tamaño de la muestra,** que es el número de unidades de análisis que se deben escoger, debe ser lo suficientemente grande para garantizar la calidad de la estimación de la característica poblacional que se desea conocer.

# La Estadística descriptiva

- Se llama estadística descriptiva, al conjunto de técnicas que permiten ordenar, resumir y representar la información recolectada.
- Esta sólo pretende hacer una descripción o resumen, de los datos disponibles y no pretende generalizar o proyectar sus hallazgos .

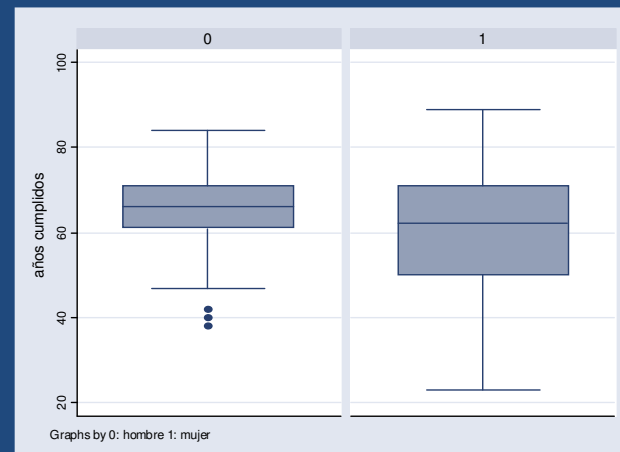
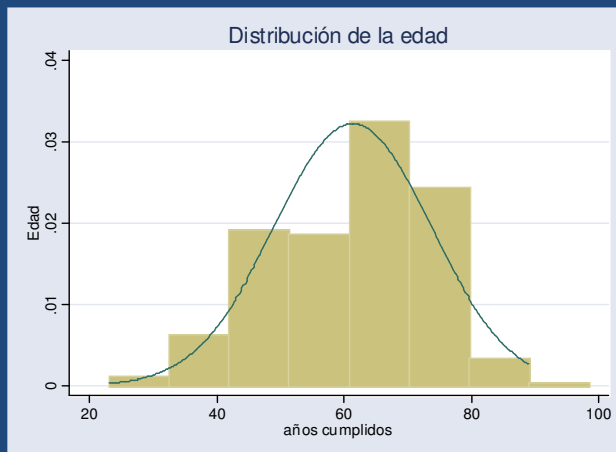
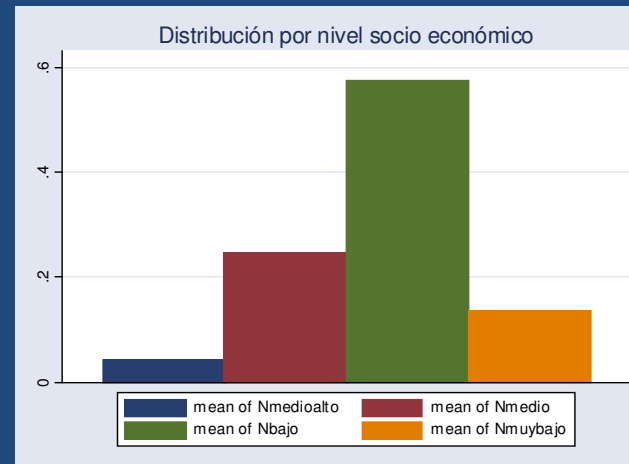
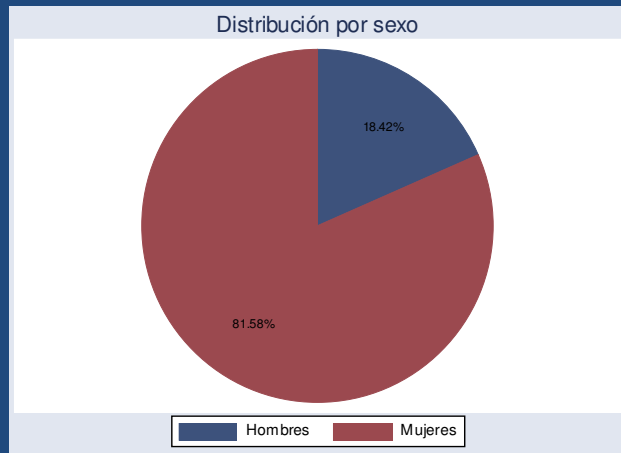
# La Estadística descriptiva

Para satisfacer su objetivo, ofrece:

- Tablas
- Estadígrafos o números resúmenes: frecuencias, proporciones , percentiles, mediana , medias (aritmética, geométrica y armónica)
- Gráficos

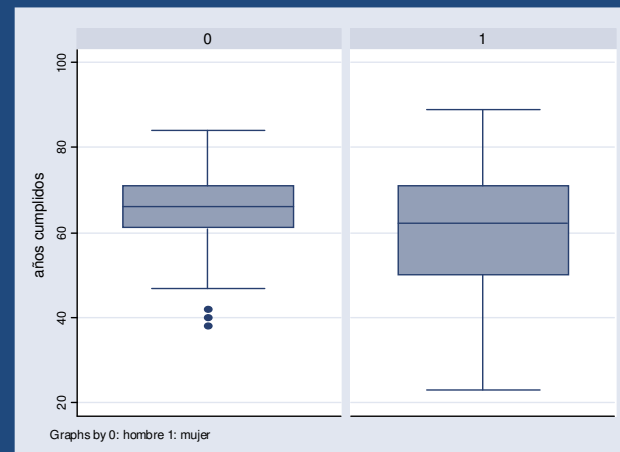
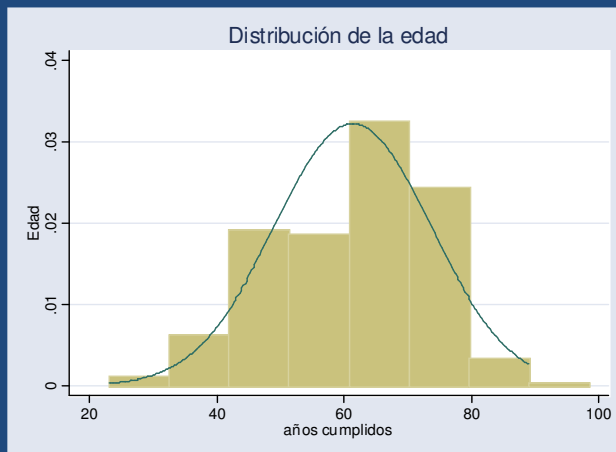
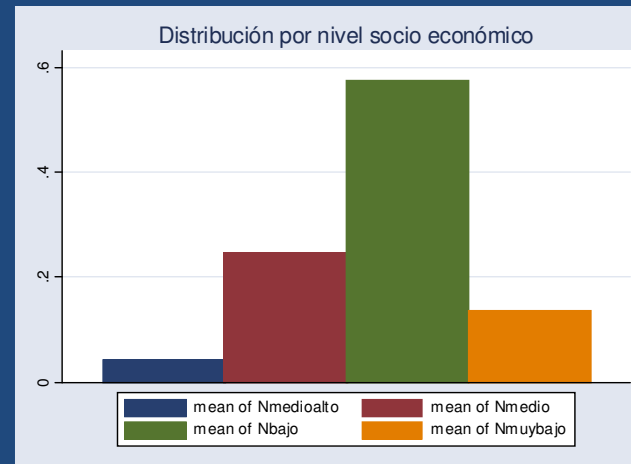
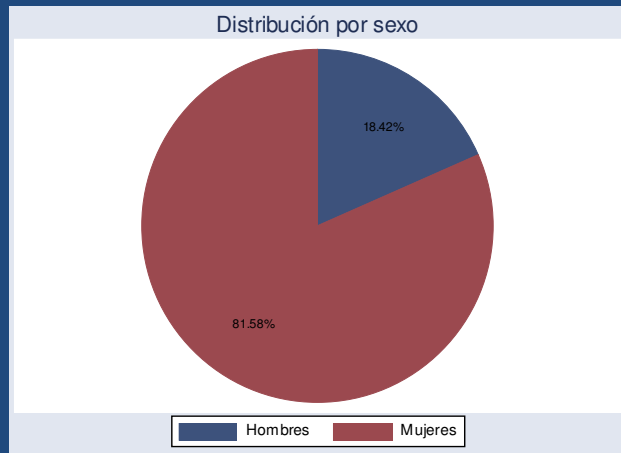
Constituye la herramienta básica de cualquier análisis estadístico y cuando se describe un fenómeno nuevo es la que permite sugerir hipótesis de investigación dándonos la primera aproximación al conocimiento de la población

# La Estadística descriptiva

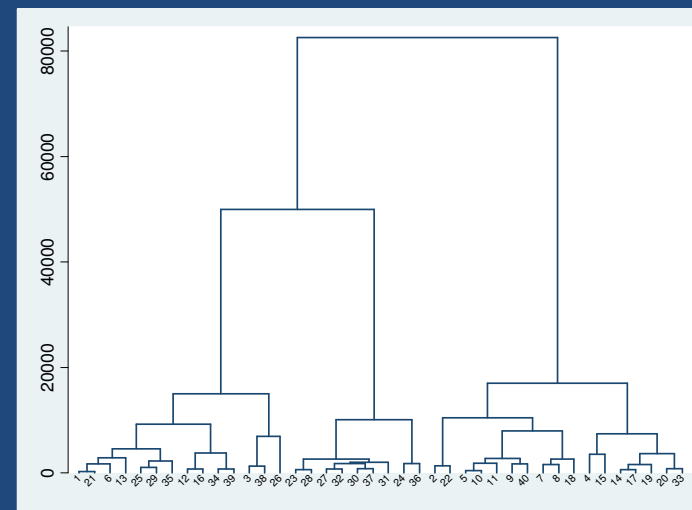
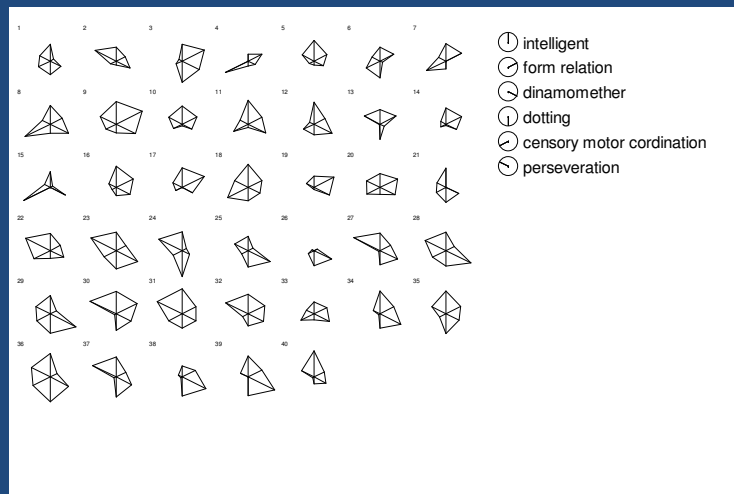
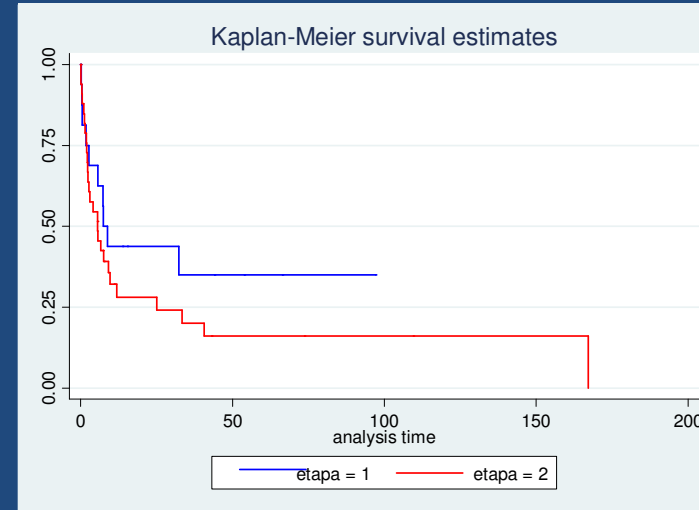
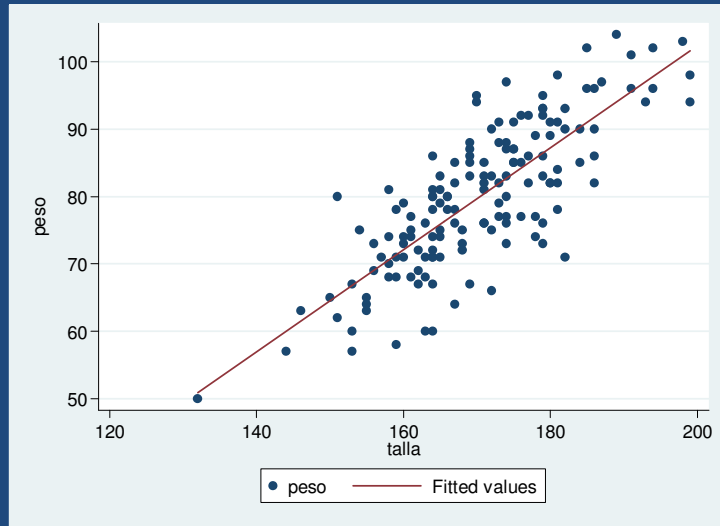




# La Estadística descriptiva



# La Estadística descriptiva



# La Estadística inferencial

El objetivo de la estadística inferencial es proyectar las características de una MUESTRA a la POBLACION que la contiene.

Este objetivo se cumple si la muestra es escogida en forma aleatoria de la población y su tamaño es lo suficientemente grande que permite una exactitud impuesta de antemano por el investigador.

Los resultados son los estimadores de los respectivos parámetros poblacionales y sus intervalos de confianza

# Dótimas de hipótesis

- **Introducción**

Hipótesis estadística es una afirmación respecto de una característica poblacional (forma de ella o valor de sus parámetros); esta sentencia puede ser “docimada” (probada) en base a una muestra aleatoria extraída de esa población.

# Dótimas de hipótesis

**Necesitamos desarrollar un procedimiento que nos permita tomar una decisión acerca de  $H_0$ , como esta decisión es tomada en base a información muestral está sujeta a errores probables, debido a que no se sabe como es realmente la naturaleza y sólo tenemos una percepción de ella. Cruzando este efecto con la decisión tenemos:**

# Dótimas de hipótesis

		Estado de la naturaleza	
		$H_0$ es Verdad	$H_0$ es Falsa
Percepción de la naturaleza	Se rechaza $H_0$	Error tipo I	Decisión correcta
	No se rechaza $H_0$	Decisión correcta	Error tipo II

	VERDADERO ESTADO DE LA NATURALEZA		
		SUJETO ENFERMO	SUJETO SANO
APRECIACIÓN DE LA NATURALEZA	EXAMEN POSITIVO	APRECIACIÓN CORRECTA	
	EXAMEN NEGATIVO		APRECIACIÓN CORRECTA

# Dótimas de hipótesis

**Llamamos:**

$\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ es Verdad})$  , tamaño del Error tipo I

$\beta = P(\text{No rechazar } H_0 \mid H_0 \text{ es Falsa})$  , tamaño del Error tipo II

**nos interesa que  $\alpha$  sea pequeño (generalmente 5% o menos).**

**$\alpha$  se llama significación de la dόcima y  $1-\beta$  se llama potencia de la dόcima, la potencia depende de la hipótesis alternativa que estemos proponiendo.**



# Diseño experimental

El diseño experimental tiene como objetivo asignar el valor de la verdad a una hipótesis de investigación, por ejemplo:

**“EL TRATAMIENTO ACTIVO OCASIONA UNA TASA DE MEJORÍA UN 30% MAYOR QUE EL TRATAMIENTO PLACEBO”**

**o**

**“LA DIFERENCIA ENTRE EL TRATAMIENTO ACTIVO Y EL PLACEBO ES DE UN 70% EN LA REDUCCIÓN DEL NIVEL DE TRIGLICERIDOS”**

# Diseño experimental

Las consideraciones del diseño experimental son las siguientes:

- Las unidades de análisis son asignadas a los distintos tratamientos en forma aleatoria
- La hipótesis nula siempre expresa el STATUS CUO de la naturaleza y la hipótesis alternativa contiene la hipótesis de investigación
- Si la hipótesis nula es rechazada se le da CREDIBILIDAD a la hipótesis alternativa (de investigación)
- Si la hipótesis nula Las unidades de análisis NO SON UNA MUESTRA ALEATORIA DE LA POBLACIÓN
- no es rechazada es que “no hay evidencia para darle CREDIBILIDAD a la hipótesis alternativa”

# Diseño experimental

- El p-value es la probabilidad de haber observado nuestro resultado si consideramos que la hipótesis nula es cierta.
- Y la potencia es el nivel de CREDIBILIDAD que nuestra experiencia le otorga a la hipótesis alternativa.

**MUCHAS GRACIAS**