

**UNIVERSIDAD
DE LOS ANDES**

Universidad de Los Andes
Instituto de Estadística Aplicada y Computación (IEAC)
Mérida, Venezuela



Métodos Estadísticos Básicos

Curso de Nivelación
Maestría en Estadística

Prof. Gudberto J. León R.
gudberto@ula.ve

Programa de la asignatura

1. Introducción
2. Análisis descriptivo de datos
3. Elementos básicos de probabilidad
4. Algunos modelos probabilísticos
5. Distribuciones en el muestreo
6. Inferencia estadística
7. Análisis de regresión y correlación lineal simple

Estadística: Conceptos básicos

Los conceptos y métodos que proporciona la estadística son de invaluable utilidad en la toma de decisiones ante **situaciones de incertidumbre**. En otras palabras, la estadística se encarga del estudio de los fenómenos aleatorios.

La estadística provee potentes herramientas analíticas que se emplean en una gran variedad de situaciones: en el gobierno, en la empresa privada, en los negocios, en la industria, en investigaciones: médicas, económicas, sociológicas, biológicas, agrícolas, genéticas, físicas, etc.

Definición de Estadística

La estadística es un conjunto de conocimientos y métodos que se utilizan para el diseño del proceso de la recolección, organización, presentación y análisis de los datos relativos a un fenómeno o hecho determinado con el fin de extraer conclusiones que le *permitan al investigador tomar decisiones en situaciones donde está presente la incertidumbre*.

Estadística: Conceptos básicos

Universo Estadístico

Generalmente, existe un conjunto de elementos claramente definido en el que el investigador está interesado. Este conjunto se llama *universo*.

Es un conjunto, finito o infinito de seres vivos, elementos o cosas, sobre las cuales están definidas características o variables que interesa analizar.

Los elementos individuales que conforman el universo se llaman **Unidades Elementales** (también se conocen como unidades individuales o unidades de observación) Las unidades elementales poseen las características de interés, las cuales pueden ser de naturaleza cuantitativa o cualitativa.

Estadística: Conceptos básicos

Ejemplos

- 1) El Instituto de Investigaciones Ambientales lleva a cabo un estudio para determinar el grado de contaminación de los ríos en la ciudad de Mérida.
Los elementos que poseen las características a estudiar son los ríos de Mérida y por tanto estos conforman el Universo Estadístico de esta investigación.
- 2) Un estudio sobre los ingresos mensuales de los hogares de la Región Andina es llevado a cabo por el Instituto de Investigaciones Económicas de la ULA.
El conjunto de elementos que poseen las variables a medir en el estudio, es decir, el universo, está conformado por todos los hogares de la Región Andina.
- 3) El Ministerio de Salud desea conocer si como consecuencia por el uso del teléfono celular, existen problemas de salud en los venezolanos.
En este caso, el Universo estadístico está compuesto por las personas venezolanas que usan teléfono celular.
- 4) Se lleva a cabo una investigación para determinar la eficiencia en el consumo de combustible de los automóviles con caja dual de marcas asiáticas. Universo:

- 5) La oficina de registros estudiantiles de la ULA quiere llevar a cabo una investigación sobre el rendimiento estudiantil en el primer semestre de las carreras de la universidad. Universo:_____.

Población Estadística

La *población*, es un conjunto *de valores* asociados con los elementos del universo. Es la colección de todas las posibles mediciones que pueden hacerse de la característica en estudio.

Al número de elementos en la población se le denomina *tamaño de la población* y, en el caso finito, este tamaño se denota con la letra N .

Estadística: Conceptos básicos

Ejemplos

En relación con los ejemplos anteriores se tiene que:

- 1) El Instituto de Investigaciones Ambientales lleva a cabo un estudio para determinar el grado de contaminación de los ríos en la ciudad de Mérida.

La población es el conjunto de valores que miden el grado de contaminación de los ríos de la ciudad de Mérida.

- 2) Un estudio sobre los ingresos mensuales de los hogares de la Región Andina es llevado a cabo por el Instituto de Investigaciones Económicas de la ULA.

Los ingresos mensuales (en bolívares) de los hogares de la Región Andina conforman una población en esta situación.

- 3) El Ministerio de Salud desea conocer si como consecuencia por el uso del teléfono celular, existen problemas de salud en los venezolanos.

Población: _____

- 4) Se lleva a cabo una investigación para determinar la eficiencia en el consumo de combustible de los automóviles con caja dual de marcas asiáticas

Población: _____

- 5) La oficina de registros estudiantiles de la ULA quiere llevar a cabo una investigación sobre el rendimiento estudiantil en el primer semestre de las carreras de la universidad.

Población: _____

Estadística: Conceptos básicos

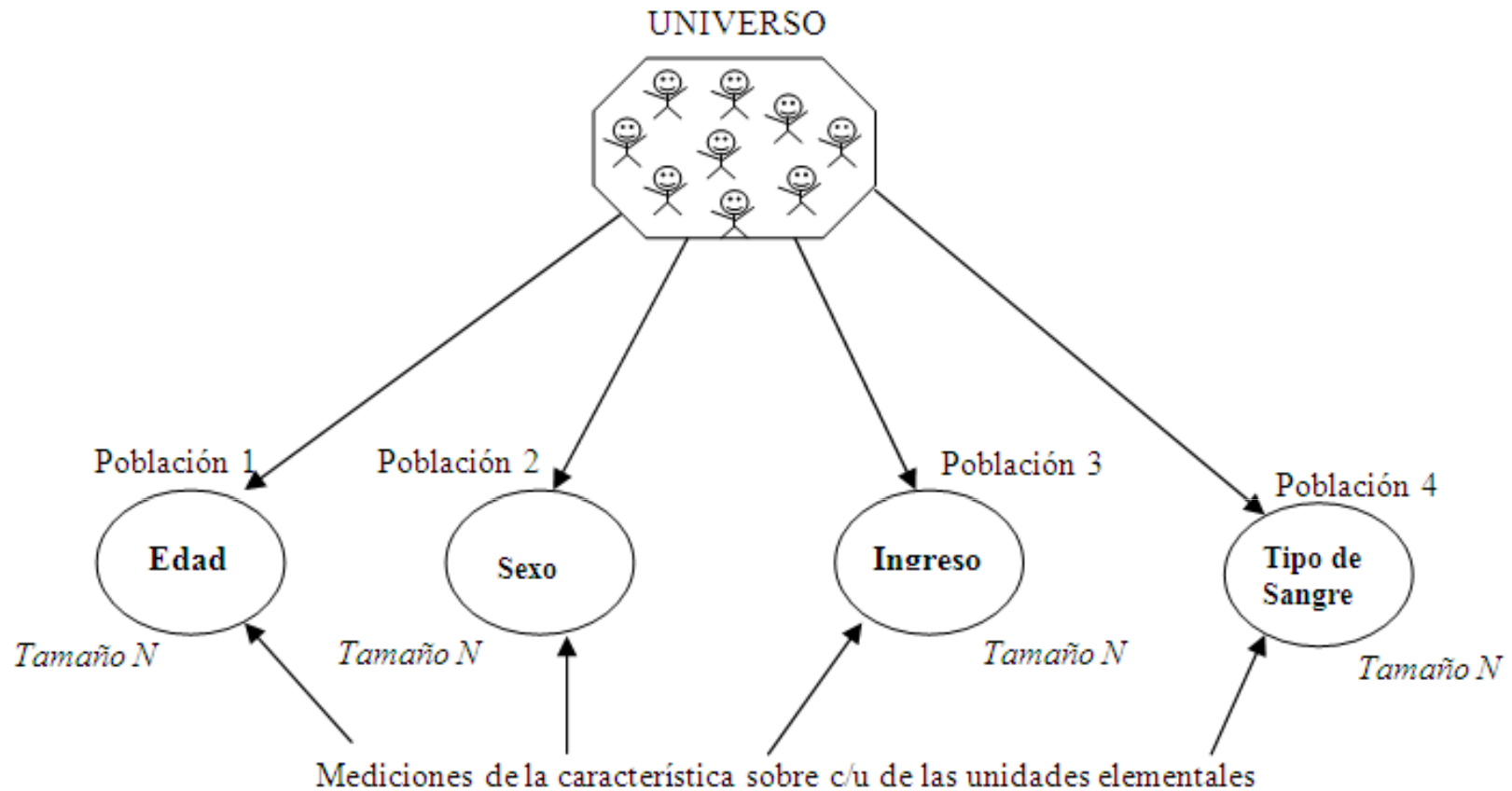


Figura 1. Ilustración didáctica de un universo con varias poblaciones

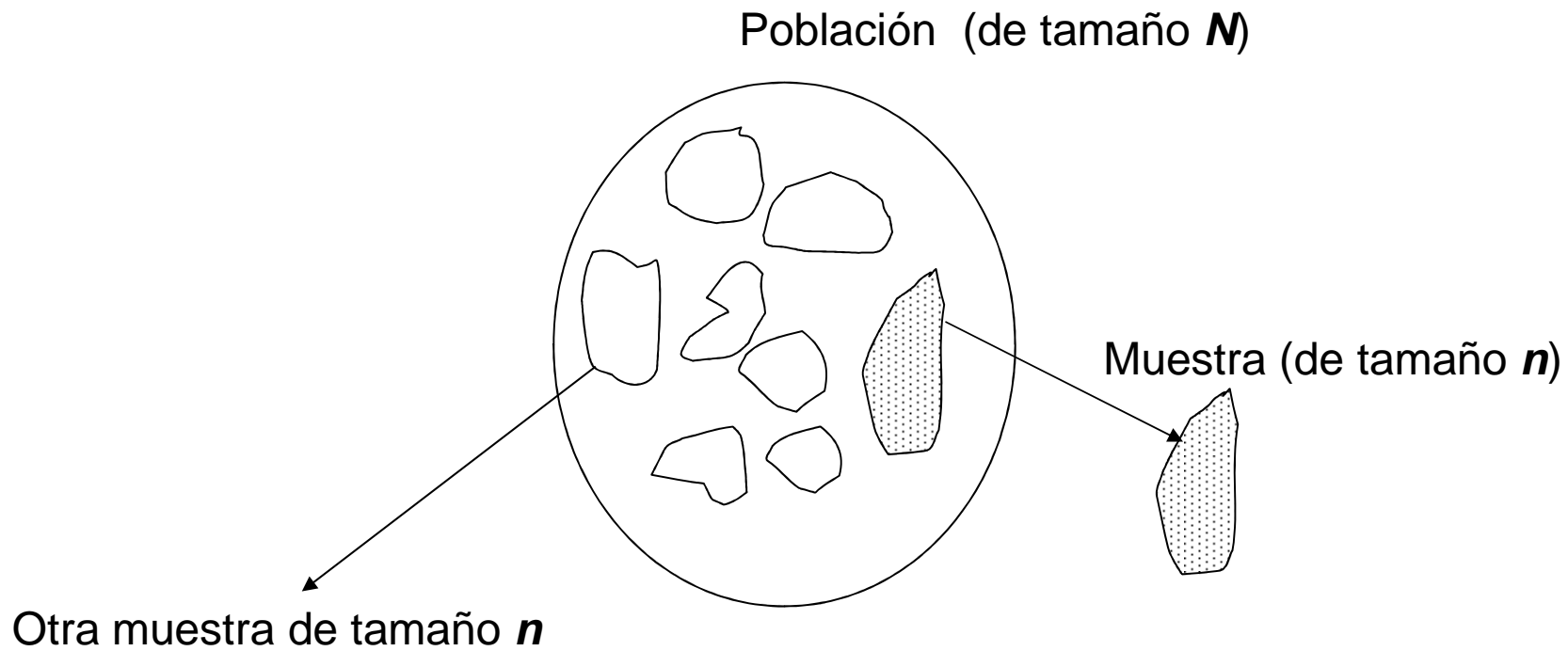
Muestra

Frecuentemente es imposible obtener o medir todos los valores en una población. Un subconjunto de valores de la población se conoce como una *muestra*. Es decir, una muestra es una parte de una población, por tanto, ésta va a estar constituida también por mediciones de la característica de interés.

Así, una muestra está compuesta por n mediciones sobre las unidades elementales. En otras palabras, n representa el tamaño de la muestra y por lo tanto $n \leq N$.

Estadística: Conceptos básicos

Fácilmente se puede deducir que de una misma población pueden seleccionarse diferentes muestras:



Estadística: Conceptos básicos

Ejemplo

En el siguiente ejemplo determinar: unidades elementales, universo, población, tipo de población (finita o infinita) y muestra.

Mediante un estudio se quiere conocer la opinión de los estudiantes de la ULA sobre el servicio de comedor que presta esta universidad. Con este fin se piensa entrevistar 500 estudiantes seleccionados aleatoriamente para conocer su opinión al respecto:

Característica en estudio:

Opinión sobre el servicio del comedor.

Unidad elemental:

Estudiante de la ULA

Universo:

Todos los estudiantes de la ULA que asisten regularmente al comedor.

Población:

La opinión sobre el servicio de comedor de cada uno de los estudiantes de la universidad que asiste regularmente al comedor.

Tipo de población:

Finita

Muestra:

Las opiniones de los 500 estudiantes seleccionados al azar.

Censo

Se dice que se ha realizado un *censo* y se habla de enumeración completa, cuando una investigación es exhaustiva en el sentido de analizar toda la población estadística.

Muestreo

Cuando el estudio se hace sobre la base de una muestra de la población estadística, se habla de una *investigación por muestreo* o enumeración parcial.

Estadística: Conceptos básicos

Razones del uso del muestreo

Las razones que determinan la conveniencia de tomar muestras son entre otras las siguientes:

- Menor costo que un censo
- Mayor control en la recolección de la información y en consecuencia mejor calidad de la misma.

En una muestra se puede dedicar más atención a la calidad de los datos, al entrenar al personal y realizar un seguimiento de quienes no contestan la encuesta. Es mucho mejor tener buenas mediciones en una muestra representativa que mediciones poco confiables sobre toda la población.

- Mayor rapidez en los resultados.

Una estimación de la tasa de desempleo del año 2017 no es muy útil si para entrevistar a cada familia, en esta etapa del estudio se demora hasta el 2019.

- El que la población sea excesivamente grande o infinita lo cual imposibilita cubrirla totalmente. Por ejemplo:
 - Una evaluación de los recursos camaroneros del litoral venezolano
 - O la evaluación de los recursos forestales de la región sur del estado Bolívar.

Estadística: Conceptos básicos

- El que la población sea suficientemente homogénea.
Este hecho permite que una muestra muy pequeña sea suficiente para inferir en la población con un margen de seguridad muy alto.
- Que el proceso de medición sea auto destructivo en el sentido de ocasionar daño o pérdida de la unidad sobre la cual se mide.

Por ejemplo:

- Cuando una galleta debe pulverizarse para determinar el contenido de grasa.
- Al probar los cinturones de seguridad para conocer su punto de ruptura, evidentemente se destruye el producto. Si todos se probaran de esa manera, no quedaría ninguno para vender.

Razones del uso del censo

- La población es muy pequeña
Por ejemplo, si se quiere conocer el historial de empleo de los graduados en Estadística de la Universidad de los Andes en el año 2010, se podría establecer contacto con ellos.
- Si el tamaño de la muestra es relativamente grande con respecto al tamaño de la población, el esfuerzo adicional requerido para hacer un censo puede ser pequeño
- Si se requiere una exactitud completa, un censo es la única forma de alcanzarla.
Por ejemplo, un gerente bancario no tomaría una muestra al azar del dinero en las cajas para saber de cuánto efectivo dispone el banco, sino que contaría todo el dinero depositado en ellas.

Clasificación de la estadística

Estadística Descriptiva

Esta rama de la estadística se encarga de organizar, resumir, simplificar, presentar los datos en cuadros y gráficos y del cálculo de medidas numéricas que permitan destacar los aspectos más importantes de los datos.

Promedios de bateo, índices de accidentes, tasas de mortalidad, promedio de goles como visitante (en fútbol), el promedio industrial Dow-Jones, el índice de desempleo, el costo de la vida, la precipitación pluvial, el rendimiento medio de un auto en kilómetros por litro y los promedios de calificación, quedan todos en esta categoría.

Nota:

Un análisis descriptivo puede realizarse en una muestra o en toda una población.

Inferencia Estadística

Consiste en el análisis e interpretación de una muestra de datos.

Más formalmente la *inferencia estadística* se encarga de estudiar las características y las leyes propias de la población mediante una muestra seleccionada de ella.

La idea básica en el muestreo es medir una porción pequeña pero *típica (representativa)* de alguna población y posteriormente utilizar dicha información para inferir (conjeturar inteligentemente) qué características tiene la población total.

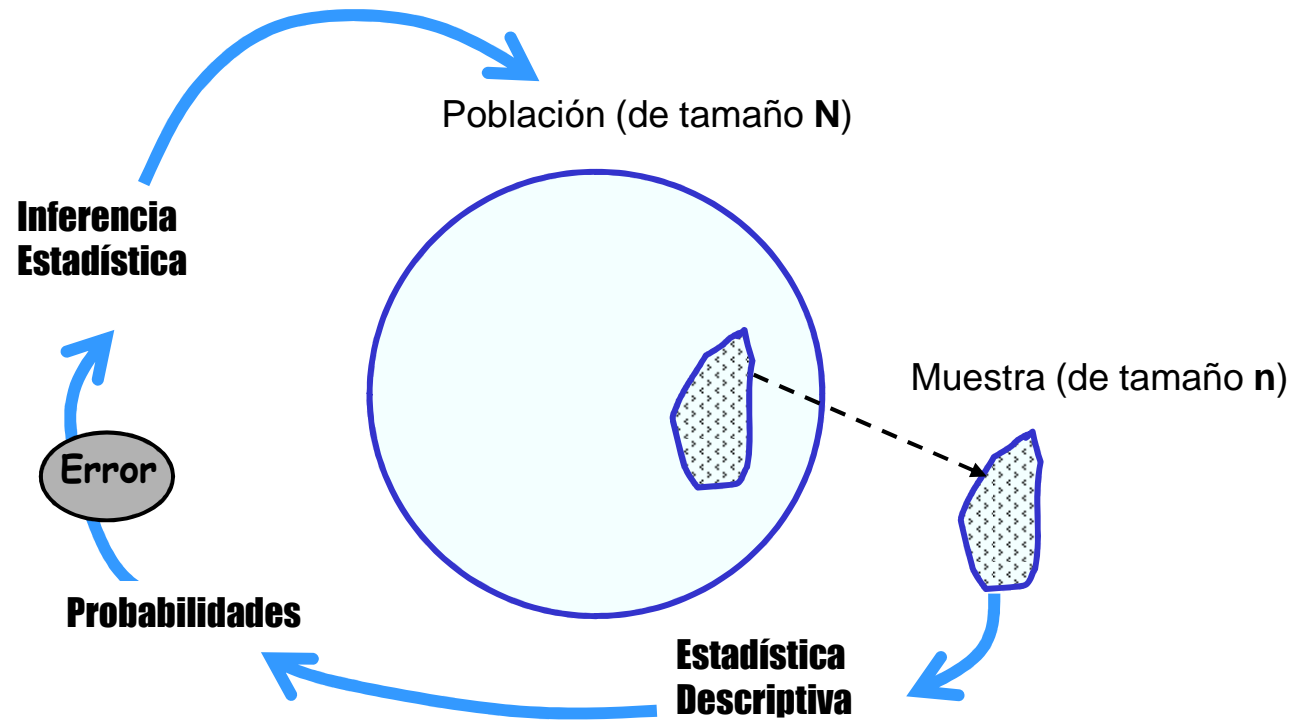
Clasificación de la estadística

Ejemplos

- Un estudio cinematográfico somete a diversas pruebas a algunos actores y actrices antes de decidir quién interpretará cada papel.
- Las fábricas suelen producir un pequeño número de piezas (producción piloto) antes de pasar a la producción en gran escala.
- Muchas compañías almacenan cientos de artículos en inventario y, mediante técnicas de muestreo, pueden estimar su valor en unidades monetarias sin tener que contar por completo todos los artículos.
- Algunas veces se llevan a cabo estudios de mercado en ciudades claves, para establecer el grado de aceptación por el consumidor.

Clasificación de la estadística

Veamos de una manera ilustrativa la definición de inferencia Estadística:



Datos estadísticos

Los datos estadísticos se obtienen mediante un proceso que comprende la observación o medición de conceptos o características como:

- Ingresos anuales en una comunidad
- Calificaciones de exámenes
- Cantidad de café por taza despachada por una máquina vendedora
- Sexo
- Resistencia a la rotura de fibras de plástico
- Porcentaje de azúcar en cereales, etc.

Tales conceptos también reciben el nombre de **variables**, ya que producen valores que tienden a mostrar cierto grado de variabilidad, al efectuarse mediciones sucesivas.

Datos estadísticos

Notación:

Se acostumbra denotar a las variables por letras latinas mayúsculas, en general las últimas del alfabeto: X, W, Y, Z, etc.

A los valores que toma la variable se habitúa denotar con la misma letra en minúscula enumerada con un subíndice.

Por ejemplo, si Y representa a la variable Edad, entonces y_3 indica la edad que toma el tercer individuo.

Escalas de medida

Se define como *escala de medida* de una variable al conjunto de valores que puede tomar esa variable o el conjunto de modalidades que puede presentar una característica.

Tipos de escalas de medida

Las escalas se clasifican de acuerdo a su proximidad o lejanía entre los valores y de acuerdo a las operaciones matemáticas que pueden realizarse entre esos valores.

Nota:

Dependiendo del tipo de escala que se utilice, se aplican unos u otros de los procedimientos estadísticos.

Escalas de medida

Escala Nominal

Se caracteriza porque la única relación que está definida entre los valores que puede tomar la variable es la igualdad o diferencia. Las operaciones aritméticas entre valores de la variable no están definidas.

Los valores se designan por palabras, letras, números o símbolos cualesquiera.

Al usar letras el orden según el alfabeto no tiene ningún significado especial.

En caso de usar números, estos constituyen un código de distinción cualitativa. De esta manera, el orden y la distancia entre esos números no tiene sentido ni interpretación alguna.

Ejemplos

- La escala correspondiente al sexo, cuyos valores se pueden denotar por M y F.
- Las posibles modalidades que puede tomar el estado civil de una persona son: soltero, casado, viudo, divorciado, concubinato (también se le pueden asignar códigos numéricos).

Escalas de medida

Escala Ordinal

Se caracteriza porque entre dos valores de la variable, además de la relación de igualdad o diferencia se pueden dar las relaciones “mayor que” o “menor que” ($=, \neq, >, <$).

Dados dos valores, se puede decir si son iguales o diferentes y además jerarquizar a los valores.

Los valores de la variable se denotan por letras, números o palabras, pero en este caso existe un orden de los valores que en el caso de las letras viene dado por el alfabeto, en el de los números por los conjuntos numéricos.

En el caso de usar números, la distancia entre ellos y las operaciones aritméticas no tienen sentido.

Ejemplo

Los rangos o jerarquías de mando de los oficiales del ejército: General en jefe, general de división, general de brigada, coronel, teniente-coronel, mayor, capitán, y teniente.

Escalas de medida

Escala de intervalos

En esta escala los valores de la variable son números y entre ellos tienen sentido las relaciones de igualdad, de orden y las distancias, es decir que la única operación aritmética que tiene sentido entre los valores es la diferencia o resta. Además, estas diferencias pueden sumarse entre sí, restarse o multiplicarse por un número real.

Propiedades

- 1) Existe una unidad de medida cuyo significado se mantiene constante para todos los valores que puede tomar la variable
- 2) Esta escala poseen un cero u origen relativo, esto quiere decir que se escoge de forma arbitraria y en consecuencia el cero en la escala no significa ausencia de la característica.

Escalas de medida

Ejemplo

Suponga que se tienen cuatro cuerpos: A, B, C y D a las temperaturas (en grados centígrados) 0° , 4° , 12° , 28° respectivamente.

Entonces, tiene sentido decir que:

- El cuerpo B tiene 4° de temperatura más que el A
- La diferencia de temperaturas entre C y D es el cuádruple de la diferencia entre A y B.

No se puede decir que:

- La temperatura del C el triple que la de B (no tienen sentido los múltiplos)
- La temperatura de B más la temperatura de D resulta en una temperatura de 32° .

Si se pasa de grados centígrados a grados fahrenheit, se mantienen las mismas conclusiones anteriores.

Nota:

Por ser el origen arbitrario, 0°C no significa que no hay temperatura (equivale a 32°F)

Escalas de medida

Escala proporcional o de razón

Es la más completa de las escalas de medida y se caracteriza porque los valores de la variable son números entre los cuales tienen sentido las relaciones de: igualdad, orden, diferencias, las sumas y las proporciones o múltiplos.

Las escalas de razón tienen un cero absoluto.

Ejemplos

El peso, la estatura, el nivel de colesterol, el tiempo de traslado entre dos puntos, la longitud, etc.

Escalas de medida

Escalas absolutas

Se caracterizan porque los valores que puede tomar la variable son el resultado de un conteo y en consecuencia las escalas absolutas están constituidas por todos los números enteros positivos y el cero.

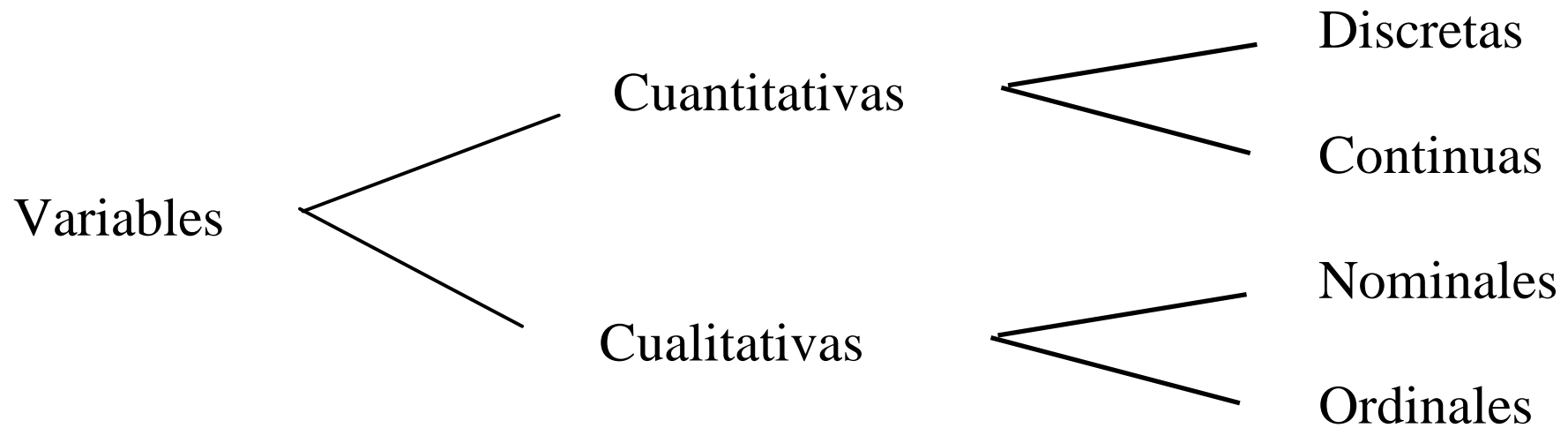
Las escalas absolutas tienen todas las propiedades de las escalas de razón, salvo que no se puede hacer una transformación de escala multiplicando por un número.

Ejemplo

El número de desempleados de un país

Escalas de medida

Tipos de variables



Serie de tiempo

Series Cronológicas o Series de Tiempo

Una *serie cronológica* o *serie de tiempo* es una sucesión de observaciones tomadas secuencialmente en el tiempo. Así, una *serie de tiempo* refleja las variaciones de una variable en el tiempo.

Ejemplos

- a. Producción anual de petróleo (en número de barriles) en Venezuela
- b. La cotización diaria del dólar
- c. El índice mensual de precios al consumidor
- d. Las pruebas de electrocardiograma en un hospital

Variables Univariantes y Multivariantes

Cuando las variables se presentan y analizan individualmente, se habla de *variable univariante*.

Alternativamente, cuando se analizan simultáneamente dos, tres o más variables se habla de *variable bivariante, trivariante o multivariante*.

Ejemplo

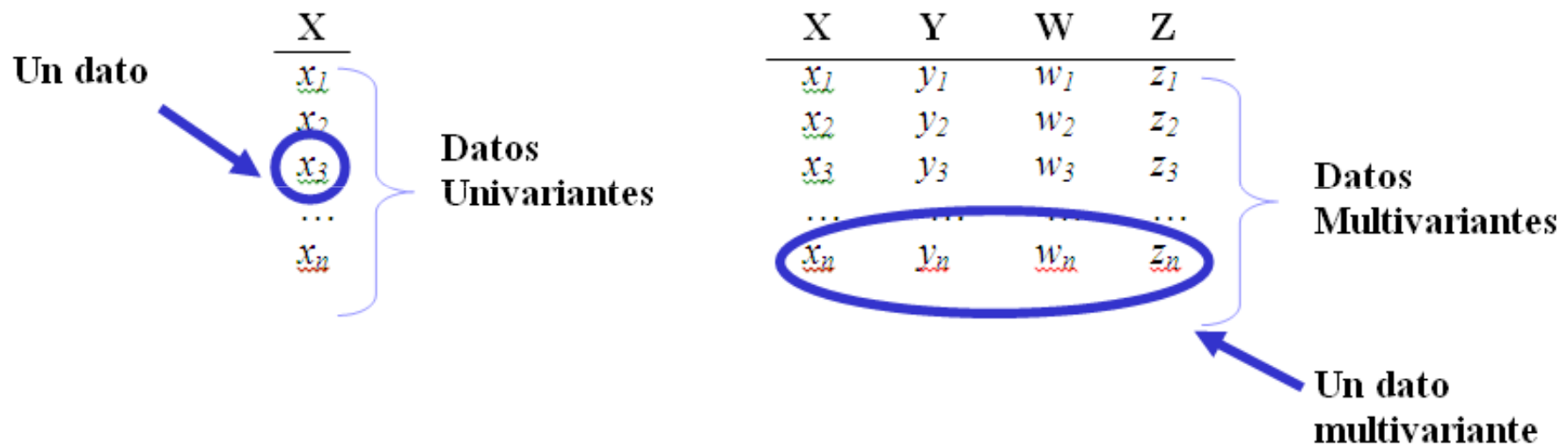
De una encuesta se obtienen los datos sobre tipo de sangre, peso, ingreso y sexo de los estudiantes de Métodos Estadísticos Básicos; y se analiza cada una de estas variables separadamente. En este caso se tienen cuatro variables univariantes.

Por otro lado, si es de interés analizar conjuntamente las variables tipo de sangre y peso se está ante la presencia de una variable bivariante. Pero, si se analizan simultáneamente las cuatro variables entonces se habla de una variable multivariante.

Variables Univariantes y Multivariantes

Sea,

X: Tipo de sangre, Y: Peso, W: Ingreso, Z: Sexo.



Estudio descriptivo de una colección de datos

Organización de datos cualitativos

VARIABLE			
Modalidad 1	...	Modalidad k	Total
f_1	...	f_k	n

Organización de datos cualitativos

Tabla de doble entrada o tabla de contingencia

		VARIABLE A				TOTALES
		a_1	a_2	...	a_i	
VARIABLE B	b_1					
	b_2					
	⋮					
	b_j					
TOTALES						

Organización de datos cuantitativos

Distribución de frecuencias

Intervalos de clase	frecuencia
$LI_1 - LS_1$	f_1
$LI_2 - LS_2$	f_2
...	...
$LI_i - LS_i$	f_i
...	...
$LI_k - LS_k$	f_k
Total de observaciones	

Nota:

Cuando los datos se presentan en distribuciones de frecuencias, se habla de *datos agrupados*, mientras que cuando se presentan individualmente, se habla de *datos no agrupados*.

Distribución de frecuencias

Pasos para la construcción de una distribución de frecuencias

1. Determinar el valor máximo y el valor mínimo de los datos.
2. Calcular el rango (o recorrido) de la variable. Este viene dado por la diferencia entre el valor máximo y el valor mínimo. El rango se denota por ***R***.
3. Determinar el *número de clases* (***K***) y las *amplitudes de clase* (***C_i***):

La *amplitud de clase* de un intervalo viene dada por la diferencia entre el límite superior y el límite inferior de dicho intervalo.

Podemos determinar la amplitud o el número de clases tomando en cuenta lo siguiente:

a. Si se conoce el número de clases:
$$C_i = \frac{R}{K}$$

b. Si se conoce la amplitud de las clases:
$$K = \frac{R}{C_i}$$

Distribución de frecuencias

c. Regla de Sturges:

$$K = 1 + 3,3 * \text{Log } n$$

Nota:

- i. La fórmula de Sturges sólo proporciona una orientación sobre cuál debe ser el número de clases. También se puede usar la regla de la raíz cuadrada: $K = \sqrt{n}$
- ii. Pueden existir clases abiertas, es decir, clases que sólo tienen un límite superior o solamente un límite inferior. Si ese es el caso, a esta clase abierta no se le podrá determinar la amplitud.
- iii. En la práctica no se conoce de antemano el número de clases y la amplitud de estas. Sin embargo existen dos recomendaciones importantes al construir una distribución de frecuencias:
 - a. Que el número de clases no sea inferior a 5 ni mayor que 15.
 - b. De ser posible es deseable que todas las clases tengan la misma amplitud.

Distribución de frecuencias

4. Proceder a construir los intervalos de clase.
 - a. Establecer el límite inferior del primer intervalo de clase.
 - b. Fijado el primer límite inferior se le suma a este la amplitud de la primera clase, C_1 , y se obtiene el límite superior de esta primera clase, el cual se constituye a la vez como el límite inferior de la segunda clase, a este se le suma la amplitud C_2 y se obtiene el límite superior de la segunda clase. Y de la misma manera se construyen los K intervalos. El último intervalo de clase debe incluir el valor máximo de los datos.
 - c. Para calcular la frecuencia de cada intervalo, se debe asumir lo siguiente: En términos matemáticos los intervalos de clase van a ser intervalos cerrados por su límite inferior y abiertos por su límite superior. Es decir, el intervalo de la i -ésima clase será de la forma:

$$[LI_i - LS_i)$$

con $i = 1, \dots, K$

Distribución de frecuencias

5. Determinar el número de datos contenidos en cada clase. Es decir, determinar las *frecuencias absolutas* de clase (f_i).

Se debe cumplir que $\sum_{i=1}^K f_i = n$, siendo n el número total de datos.

6. Determinar el resto de las frecuencias:

$$fr_i = \frac{f_i}{n}$$

$$F_i$$

$$Fr_i = \frac{F_i}{n}$$

$$m_i = \frac{LI_i + LS_i}{2}$$

Distribución de frecuencias

Ejercicio

Completar la siguiente distribución de frecuencias*:

Clases	m_i	f_i	fr_i	F_i	Fr_i
---	15	--	---	--	0,16
[20 -30)	--	--	0,08	--	---
[30 - 40)	--	6	---	12	---
[40 - 50)	45	8	---	20	---
---	65	--	---	--	---
Totales		--	---		

* Tomado de Armas, J.M. (1988). *Estadística Sencilla. Descriptiva*. Mérida-Venezuela: FACES-Universidad de Los Andes.

Distribución de frecuencias

Distribución de frecuencias cuyas clases son valores individuales de la variable en estudio

En este tipo de distribución de frecuencias no existen límites de clase, amplitudes y las marcas de clase m_i coinciden con las clases.

En las distribuciones de frecuencias cuyas clases son valores individuales, se puede reconstruir fácilmente la colección de datos originales. Esto no es posible cuando las clases son intervalos.