

# Unidad 1

Probabilidad y estadística | Alejandro Jarillo Silva



## DEFINICIONES

**Datos** son el conjunto de información recolectada (como mediciones, géneros, respuestas de encuestas).

**Estadística** es la ciencia que se encarga de planear estudios y experimentos, obtener datos y luego organizar, resumir, presentar, analizar e interpretar la información para extraer conclusiones basadas en los datos.

**Población** es el conjunto completo de todos los elementos (puntuaciones, personas, mediciones, etcétera) que se someten a estudio. El conjunto es completo porque incluye a *todos* los sujetos que se estudiarán.

**Censo** es el conjunto de datos de *cada uno* de los miembros de la población.

**Muestra** es un *subconjunto* de miembros seleccionados de una población.

## Ojo mucho ojo

- Los datos muestrales deben reunirse de una forma adecuada, como a través de un proceso de selección *aleatoria*.
- Si los datos muestrales no se reúnen de forma adecuada, resultarán tan inútiles que ningún grado de tortura estadística podrá salvarlos.



**Concepto clave** En esta sección se introducen los principios básicos del pensamiento estadístico que se utilizarán a lo largo del libro. Ya sea que realicemos un análisis estadístico con datos que hemos reunido, o que examinemos uno realizado por alguien más, no debemos aceptar a ciegas los cálculos matemáticos; es necesario tomar en cuenta los siguientes factores:

- Contexto de los datos
- Fuente de los datos
- Método de muestreo
- Conclusiones
- Implicaciones prácticas



**Tabla 1-1 Datos utilizados para análisis**

$x$	56	67	57	60	64
$y$	53	66	58	61	68

Las siguientes son preguntas fundamentales que se deben plantear para obtener esa información: ¿Cuál es el contexto de los datos? ¿De qué fuente se obtuvieron? ¿Cómo se recabaron? ¿Qué se puede concluir a partir de la información? Con base en conclusiones estadísticas, ¿qué implicaciones prácticas resultan del análisis?

### EJEMPLO 1

**Contexto para la tabla 1-1** Los datos de la tabla 1-1 se tomaron del conjunto de datos 3 del apéndice B, y representan los pesos (en kilogramos) de estudiantes de la Universidad de Rutgers, en Nueva Jersey. Los valores  $x$  son los pesos registrados en el mes de septiembre de su primer año de estudios, y los valores  $y$  son los pesos correspondientes registrados en abril del siguiente semestre. Por ejemplo, el primer estudiante pesó 56 kg en septiembre y 53 kg en abril. Estos pesos están incluidos en un estudio descrito en el artículo “Changes in Body Weight and Fat Mass of Men and Women in the First Year of College: A Study of the ‘Freshman 15’”, de Hoffman, Policastro, Quick y Lee, *Journal of American College Health*, vol. 55, núm. 1. El título del artículo nos indica el objetivo del estudio: determinar si los estudiantes universitarios realmente aumentan 15 libras de peso durante el primer año de estudios, de acuerdo con la leyenda llamada “Freshman 15”.



### EJEMPLO 2

**Fuente de los datos de la tabla 1-1** Las mediciones de la tabla 1-1 fueron realizadas por investigadores respetables del Departamento de Ciencias de la Nutrición de la Universidad de Rutgers. Los investigadores no tienen razones para distorsionar o modificar los resultados con la finalidad de apoyar alguna postura de beneficio personal; no ganan ni pierden si alteran los resultados. No recibieron un pago de una compañía que pudiera beneficiarse de resultados favorables. Podemos tener la confianza de que estos investigadores son imparciales y de que no alteraron los resultados.





### EJEMPLO 3

**Muestreo utilizado para los datos de la tabla 1-1** Los pesos de la tabla 1-1 provienen de la muestra más grande de pesos incluida en el conjunto de datos 3 del apéndice B. Los investigadores obtuvieron los datos de sujetos que participaron como voluntarios en una evaluación de salud realizada en septiembre, durante su primer año de estudios. Los 217 estudiantes que participaron en la evaluación de septiembre fueron invitados para un estudio de seguimiento en primavera; de ellos, 67 aceptaron la invitación para ser pesados de nuevo durante las últimas dos semanas de abril. Se trata de una muestra de respuesta voluntaria. Los investigadores reportaron que “la muestra obtenida no fue aleatoria y podría haber un sesgo de autoselección”. Los autores profundizaron en el potencial de sesgo al hacer una lista específica de fuentes potenciales de sesgo; por ejemplo, es posible que “solo aquellos estudiantes que se sintieron suficientemente cómodos con su peso [aceptaran] someterse a medición en ambas ocasiones”.



### EJEMPLO 4

**Conclusiones a partir de los datos de la tabla 1-1** La tabla 1-1 incluye los pesos, antes y después, de cinco sujetos tomados del conjunto de datos 3 del apéndice B. El análisis de tales pesos llevó a las conclusiones reportadas en el artículo “Changes in Body Weight and Fat Mass of Men and Women in the First Year of College: A Study of the ‘Freshman 15’”, de Hoffman, Policastro, Quick y Lee, *Journal of American College Health*, vol. 55, núm. 1. Al analizar los datos de la tabla 1-1, los investigadores concluyeron que los estudiantes aumentan de peso durante el primer año de estudios universitarios. Sin embargo, también comentaron que en el pequeño grupo no aleatorio estudiado, el incremento de peso fue menor que 15 libras, y que esta cantidad no era universal. Concluyeron que la leyenda “Freshman 15” sobre el aumento de peso es un mito.



### EJEMPLO 5

**Implicaciones prácticas de los datos de la tabla 1-1** En su análisis de los datos recolectados para el estudio de “Freshman 15”, los investigadores señalan algunas implicaciones prácticas de sus resultados. Ellos afirmaron: “Quizá sea más importante que los estudiantes reconozcan que los cambios aparentemente mínimos e incluso inofensivos en los hábitos alimenticios y de ejercicio podrían producir grandes cambios en el peso y en el contenido de grasa corporal durante un periodo largo”. Los estudiantes que inician el primer año de la universidad deben reconocer que las rutinas de alimentación y de ejercicio radicalmente diferentes pueden tener consecuencias graves en la salud.





## Ejercicio



- Una empresa de software evaluó el rendimiento académico y los hábitos de estudio de 120 estudiantes de Tecnologías de la Información en una universidad. Los resultados mostraron que los estudiantes dedican en promedio 12 horas semanales al estudio independiente. El 65 % cuenta con equipo de cómputo propio, mientras que el resto usa laboratorios de la escuela. Los alumnos con computadora propia alcanzaron un promedio de 8.7, frente a 7.9 en quienes solo usan laboratorios. Además, el 75 % declaró estar satisfecho o muy satisfecho con la carrera.

## Uno más

- En la Facultad de Ingeniería se encuestó a 200 estudiantes sobre sus hábitos de sueño. Los resultados mostraron que duermen en promedio 6.2 horas diarias. El 55 % reportó sentirse cansado durante clases y el 25 % dijo que ha faltado a alguna clase por desvelo relacionado con tareas. Se observó que los alumnos que duermen menos de 6 horas tienen en promedio un promedio escolar de 7.5, mientras que los que duermen 7 horas o más alcanzan un promedio de 8.6.

La **significancia estadística** de un estudio difiere de su **significancia práctica**. Es posible que, con base en los datos muestrales disponibles, se utilicen métodos estadísticos para llegar a la conclusión de que algún tratamiento o hallazgo es eficaz, aunque el sentido común sugiera que no hay una diferencia suficiente debida al tratamiento para justificar que su uso sea práctico.

#### EJEMPLO 6

**Significancia estadística y significancia práctica** En una prueba del programa Atkins para perder peso, 40 sujetos registraron una pérdida de peso promedio de 2.1 libras después de someterse al programa durante un año (según datos de “Comparison of the Atkins, Ornish, Weight Watchers and Zone Diets for Weight Loss and Heart Disease Risk Reduction”, de Dansinger *et al.*, *Journal of the American Medical Association*, vol. 293, núm. 1). Utilizando métodos formales de análisis estadísticos, podemos concluir que la pérdida media de peso de 2.1 es estadísticamente significativa; es decir, con base en criterios estadísticos, parece que la dieta es eficaz. Sin embargo, el sentido común indica que no vale la pena seguir un programa de pérdida de peso que produzca resultados tan insignificantes. Lo más probable es que una persona que inicie un programa de pérdida de peso quiera perder mucho más de 2.1 libras. Aunque la pérdida media de peso de 2.1 es estadísticamente significativa, no tiene una significancia práctica. El análisis estadístico sugiere que el programa es eficaz, pero las consideraciones prácticas sugieren que el programa es básicamente ineficaz.

En los ejercicios 23 a 26, dé una conclusión sobre la significancia estadística. No realice ningún cálculo formal. Utilice los resultados presentados o haga juicios subjetivos acerca de estos.

**24. Experimentos genéticos de Mendel** Uno de los famosos experimentos de hibridación con guisantes (o chícharos) de Gregor Mendel produjo 580 vástagos, de los cuales 152 (o el 26%) tuvieron vainas amarillas. Según la teoría de Mendel, el 25% de los guisantes vástagos deberían tener vainas amarillas. ¿Los resultados del experimento difieren del porcentaje establecido por Mendel del 25% en una cantidad estadísticamente significativa?

**25. Encuesta sobre el tabaquismo pasivo** En una encuesta que aplicó Gallup a 1038 adultos seleccionados al azar, el 85% dijo que el tabaquismo pasivo es un poco dañino o muy dañino, pero un representante de la industria tabacalera afirma que solo el 50% de los adultos creen que el tabaquismo pasivo es un poco dañino o muy dañino. ¿Existe evidencia estadísticamente significativa en contra de la aseveración del representante? ¿Por qué?

**Concepto clave** Un objetivo de la estadística es realizar inferencias o generalizaciones acerca de una población. Además de los términos *población* y *muestra*, que se definieron al principio de este capítulo, necesitamos conocer el significado de los conceptos *parámetro* y *estadístico*. Estos nuevos términos se utilizan para distinguir entre los casos en que contamos con los datos de una población completa y los casos en los que solo contamos con los datos de una muestra.



### DEFINICIONES

**Parámetro** es una medición numérica que describe algunas características de una *población*.

**Estadístico** es una medición numérica que describe algunas características de una *muestra*.





## Según la relación de las variables

- Se puede definir como el factor de riesgo o la variable que va a condicionar o detonar a la variable dependiente; es decir, aquella que antecede a los efectos.

Variable independiente



- Es la que está condicionada por la variable o variables independientes; es decir, la del efecto o resultado.

Variable dependiente



Fumar  
(Variable independiente)



Cáncer pulmonar  
(Variable dependiente)

# Clasificación de acuerdo a la naturaleza de la variable

## Variable cualitativa

Se refiere a la variable en la cual su serie de valores se mide en cualidades; por ejemplo, alto, bajo, bonito, feo, etcétera.

Algunas de las propiedades con las que deben cumplir las variables cualitativas son que deben ser **mutuamente excluyentes**; es decir, que dentro de las opciones de respuesta que puede tomar la variable, un sujeto no puede permanecer a la vez en varias categorías de dicha variable.

## Variable cuantitativa

Es aquella en la que la escala de medición se puede expresar en números.

Además, debe ser **colectivamente exhaustiva**; es decir, dentro de las categorías se encuentran todas las posibilidades en donde se puede clasificar determinada variable. En otras palabras, se cuenta con todas las posibles respuestas para la variable.

# De acuerdo a su escala de medición

## Nominal

Es aquella variable categórica en la que no se sigue un orden o jerarquía.

Ejemplo: El estado civil.

## Ordinal

Es aquella variable categórica en la que se sigue un orden o jerarquía.

Ejemplo: El grado de escolaridad.

## Discretas

Son aquellas variables numéricas que tienen un número finito de valores entre dos valores cualquiera; por lo tanto, pueden tomar valores enteros únicamente.

Ejemplo: El número de hijos.

## Continua

Son aquellas variables numéricas que tienen un número infinito de valores entre dos valores cualquiera; por lo tanto, toman valores con decimales.

Ejemplo: El peso.

# De acuerdo a su contenido

## Variables unidimensionales

- Son aquellas variables que sólo tienen una posible dimensión, es decir, una posible respuesta, generalmente su única característica medible. Al momento que se junta con otras variables de este tipo, pueden formar una variable multidimensional.

## Variables multidimensionales

- Son aquellas variables que son compuestas por varias unidimensionales; en medicina y economía, los ejemplos claros que se tienen son los distintos índices que se utilizan.





# Escalas de medición

**Tabla 1-2 Niveles de medición**

<b>De razón:</b>	Hay un punto de partida cero natural y las razones tienen significado.	<i>Ejemplo:</i> distancias
<b>De intervalo:</b>	Las diferencias tienen un significado, pero no hay punto de partida cero natural, y las razones no tienen significado.	<i>Ejemplo:</i> temperaturas corporales en grados Fahrenheit o Celsius
<b>Ordinal:</b>	Las categorías están ordenadas, pero no hay diferencias o estas carecen de significado.	<i>Ejemplo:</i> las clasificaciones de las universidades en el <i>U.S. News and World Report</i>
<b>Nominal:</b>	Solo categorías. Los datos no pueden acomodarse en un esquema de orden.	<i>Ejemplo:</i> el color de los ojos

# En los ejercicios 5 a 12, determine si el valor dado es un estadístico o un parámetro

**5. Ingreso y educación** En una muestra grande de hogares, el ingreso anual medio por hogar para los individuos graduados de bachillerato es de \$19,856 (según datos de la Oficina de Censos de Estados Unidos).

**6. Política** De los senadores que conforman el Congreso estadounidense actual, el 44% son demócratas.

**7. Titanic** En un estudio de los 2223 pasajeros del *Titanic*, se encontró que 706 sobrevivieron cuando el transatlántico se hundió.

**8. Dispositivos para cruzar las calles** En la ciudad de Nueva York hay 3250 dispositivos para cruzar las calles; los peatones tienen que presionar un botón de dichos dispositivos ubicados en las intersecciones de tránsito para controlar las luces del semáforo. Se descubrió que el 77% de los aparatos no funcionan (según datos del artículo “For Exercise in New York Futility, Push Button”, de Michael Luo, *New York Times*).

**9. Áreas de estados** Si se suman las áreas de los 50 estados de EUA, y la suma se divide entre 50, el resultado es de 196,533 kilómetros cuadrados.

**10. Tabla periódica** El peso atómico promedio (o la media del peso atómico) de todos los elementos de la tabla periódica es de 134.355 unidades de masa atómica unificada.

**11. Voltaje** El autor midió el voltaje suministrado a su hogar 40 días diferentes, y el valor promedio (o medio) es de 123.7 volts.

**12. Ganancias de películas** El autor eligió al azar 35 películas y calculó la cantidad de dinero que obtuvieron por las ventas de boletos. El promedio (o la media) es de \$123.7 millones.

En los ejercicios 22 a 28, determine cuál de los cuatro niveles de medición (nominal, ordinal, de intervalo, de razón) es el más adecuado.

- 22.** Tipos de películas (drama, comedia, aventura, documental, etcétera).
- 23.** Críticas de películas en una escala de 0 a 4 estrellas.
- 24.** Temperaturas reales (en grados Fahrenheit), como aparecen en el conjunto de datos 11 del apéndice B.
- 25.** Las compañías (Disney, MGM, Warner Brothers, Universal, 20th Century Fox) que produjeron las películas incluidas en el conjunto de datos 7 del apéndice B.
- 26.** Cantidades medidas de gases de invernadero (en toneladas por año) emitidas por los automóviles del conjunto de datos 16 del apéndice B.
- 27.** Años en que se proyectaron las películas, según se indica en el conjunto de datos 9 del apéndice B.
- 28.** Las calificaciones de los automóviles evaluados por Consumer's Union.

# ESTADÍSTICA DESCRIPTIVA

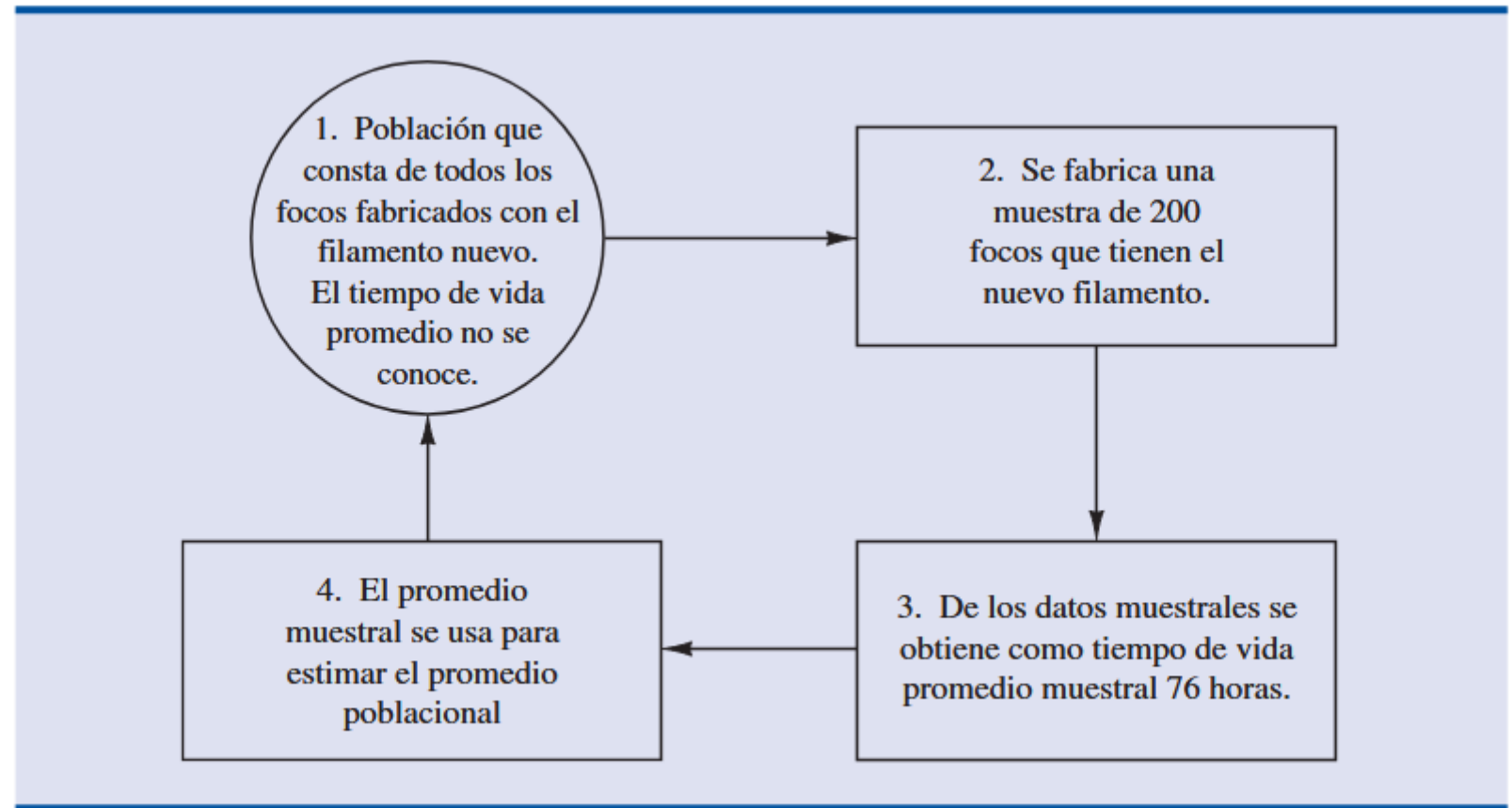
Se ocupa de la recopilación, organización y análisis de datos sobre alguna característica de ciertos individuos pertenecientes a la población o universo.



# ESTADISTICA INFERENCIAL

## Inferencia estadística

Es un método que permite hacer estimaciones y predicciones sobre una población completa basándose en el análisis de una muestra representativa.





# Muestreo



## DEFINICIÓN

Una **muestra aleatoria simple** de  $n$  sujetos se selecciona de manera que cada posible *muestra del mismo tamaño  $n$*  tenga la misma posibilidad de ser elegida.

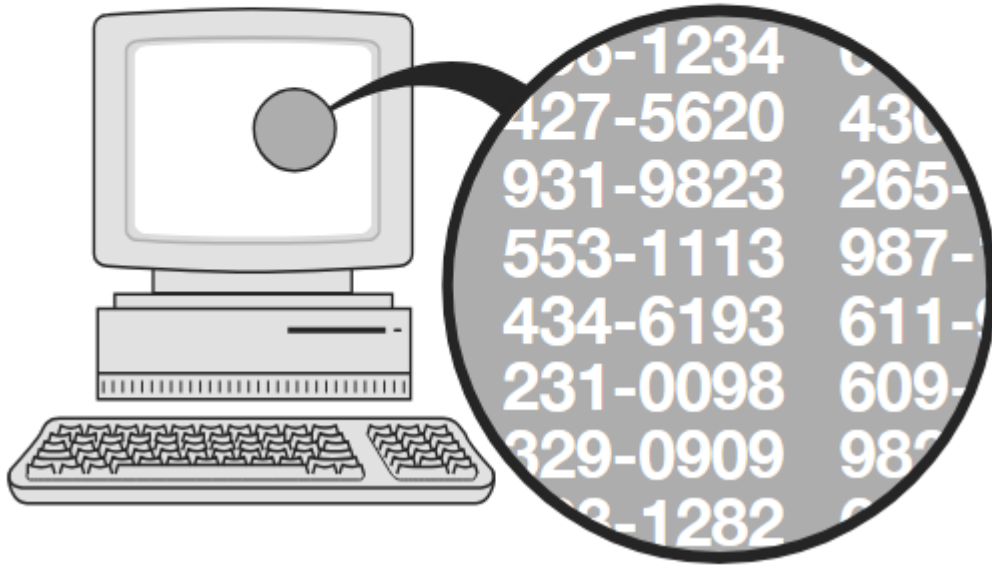


## DEFINICIONES

En una **muestra aleatoria** los miembros de la población se seleccionan de forma que cada *miembro individual* tenga la misma posibilidad de ser elegido.

Una **muestra probabilística** implica seleccionar a miembros de una población de forma que cada miembro tenga una posibilidad conocida (aunque no necesariamente la misma) de ser elegido.

# Tipos de muestreo



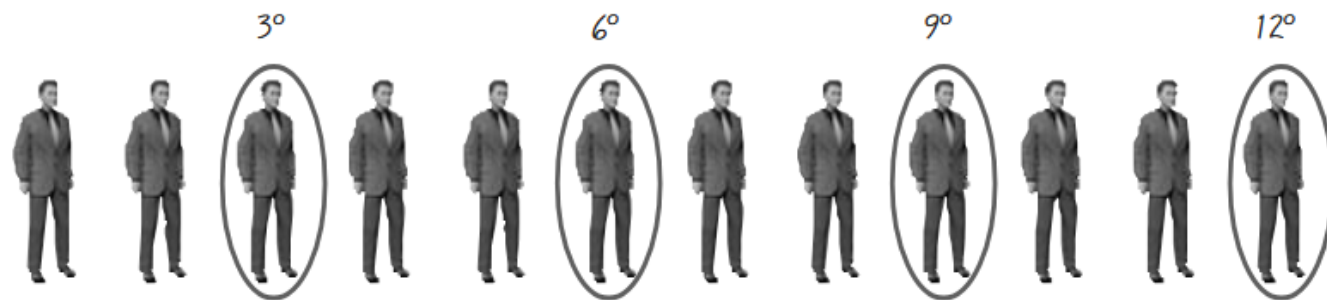
## *Muestreo aleatorio:*

*Cada miembro de la población tiene la misma probabilidad de resultar seleccionado.*

*A menudo se usan computadoras para generar números telefónicos aleatorios.*

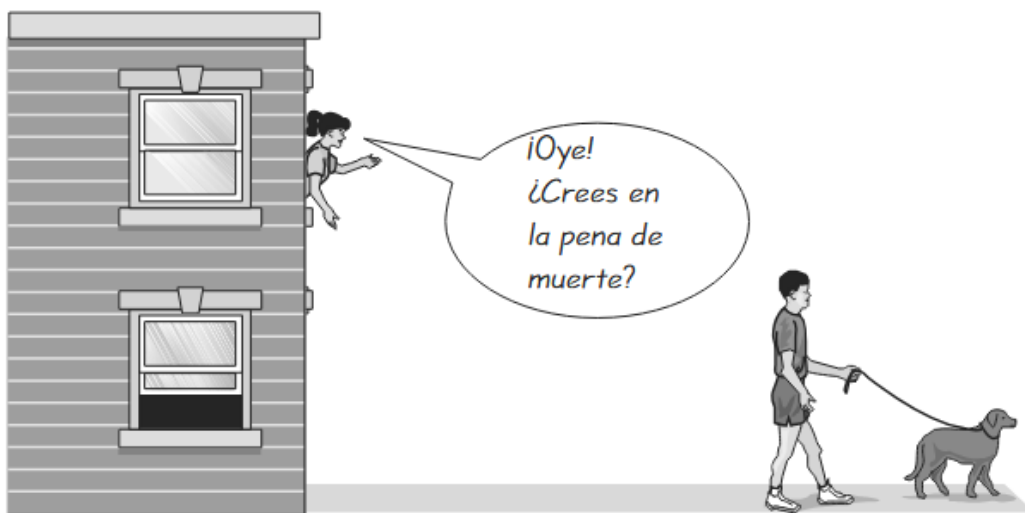
## *Muestreo aleatorio simple:*

*Se selecciona una muestra de  $n$  sujetos, de manera que cada posible muestra del mismo tamaño  $n$  tenga la misma posibilidad de ser elegida.*



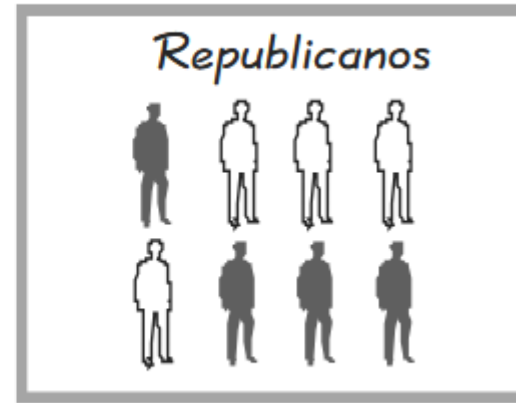
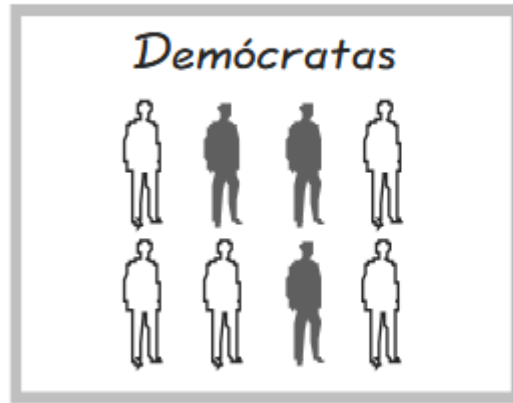
### *Muestreo sistemático:*

*Se selecciona un punto de partida, después se elige cada  $k$ -ésimo (por ejemplo, cada quincuagésimo) elemento de la población.*



### *Muestreo de conveniencia:*

*Se utilizan resultados que son fáciles de obtener.*



### *Muestreo estratificado:*

*Se subdivide a la población en al menos dos subgrupos (o estratos) diferentes, de manera que los sujetos del mismo subgrupo compartan las mismas características (como el género o la categoría de edad), y después se obtiene una muestra de cada subgrupo.*

### *Muestreo por conglomerados:*

*Se divide a la población en secciones (o conglomerados), luego se eligen al azar algunos de estos conglomerados, y después se elige a todos los miembros de los conglomerados seleccionados.*

#### **Todas las clases en una universidad:**

Arquitectura Sección 1

Historia del arte Sección 1

Historia del arte Sección 2

Biología Sección 1

Biología Sección 2

Biología Sección 3



Zoología Sección 1

Se encuesta a **todos** los estudiantes en clases seleccionadas al azar.

# Tipos de estudio



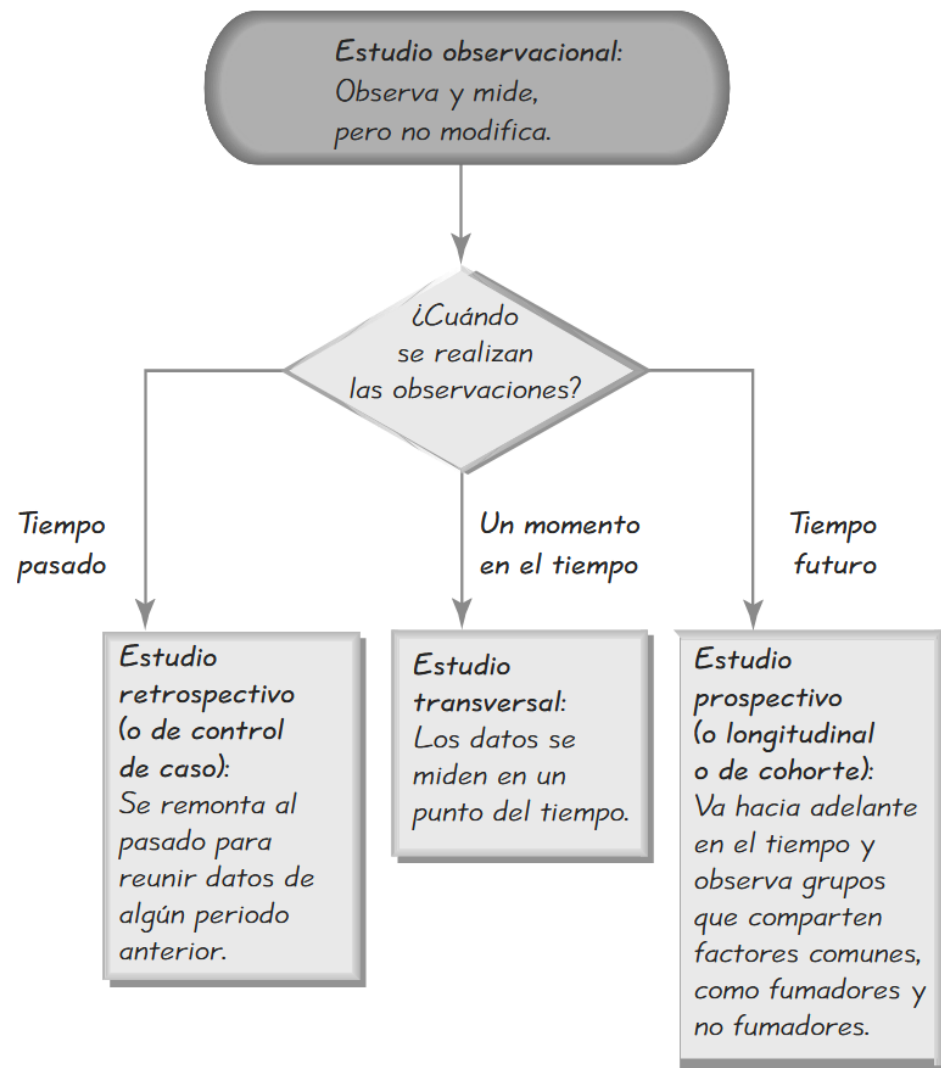
## DEFINICIÓN

En un **estudio transversal**, los datos se observan, miden y reúnen en un solo momento.

En un **estudio retrospectivo** (o **de control de caso**), los datos se toman del pasado (mediante el examen de registros, entrevistas y otros recursos).

En un **estudio prospectivo** (o **longitudinal** o **de cohorte**), los datos se reunirán en el futuro y se toman de grupos (llamados *cohortes*) que comparten factores comunes.





## DEFINICIÓN

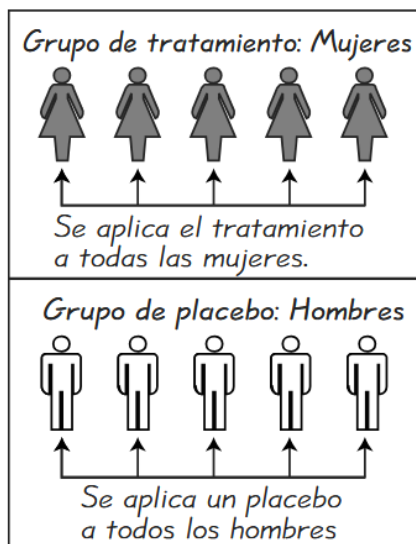
En un experimento ocurre **confusión** cuando uno no es capaz de distinguir entre los efectos de diferentes factores.

**Trate de planear el experimento de manera que no haya lugar para la confusión.**

**Diseño experimental inadecuado:**

Tratar a todas las mujeres y no tratar a los hombres.

(Problema: No se sabe si los efectos se deben al género o al tratamiento).



a)

**Diseño experimental completamente aleatorizado:**

De manera aleatoria se determina quién recibirá el tratamiento.

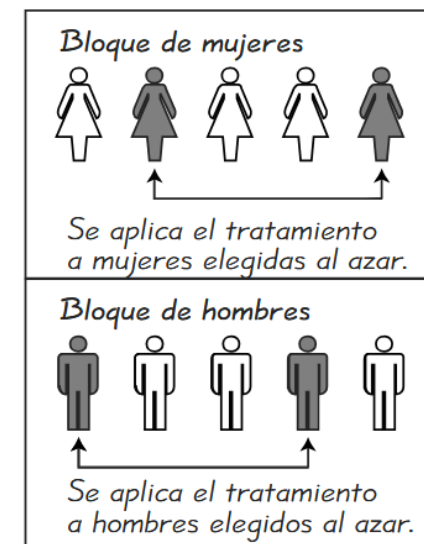


b)

**Diseño de bloques aleatorizados:**

1. Se forma un bloque de mujeres y un bloque de hombres.

2. Dentro de cada bloque se seleccionan al azar los sujetos que recibirán el tratamiento.



c)

# Errores de muestreo



## DEFINICIONES

Un **error de muestreo** es la diferencia entre el resultado de una muestra y el verdadero resultado de la población; este error es consecuencia de las fluctuaciones por el azar.

Un **error que no es de muestreo** sucede cuando los datos muestrales se obtienen, registran o analizan de forma incorrecta (como cuando se selecciona una muestra sesgada, cuando se emplea un instrumento de medición defectuoso o cuando se registran los datos de forma incorrecta).

# Características de los datos

1. **Centro:** Valor promedio o representativo que indica la localización de la mitad del conjunto de los datos.
2. **Variación:** Medida de la cantidad en que los valores de los datos varían entre sí.
3. **Distribución:** La naturaleza o forma de la distribución de los datos sobre el rango de valores (como en forma de campana, distribución uniforme o sesgada).
4. **Valores atípicos:** Valores muestrales que están muy alejados de la vasta mayoría de los demás valores de la muestra.
5. **Tiempo:** Características cambiantes de los datos a través del tiempo.

# Tabulación de los datos

En los experimentos estadísticos los datos recolectados pueden corresponder a una **población** o **muestra**. En ambos casos los procedimientos de resumen de datos son análogos y designaremos por:

$N$  = Tamaño de la población estudiada

$n$  = Tamaño de la muestra (parte de la población)

Con el objeto de realizar un mejor estudio de los datos es necesario organizar éstos, mediante el uso de distribuciones de frecuencia.

Una **distribución de frecuencia** es una tabla resumen en la que se disponen los datos divididos en grupos ordenados numéricamente y que se denominan **clases** o **categorías**.

# Tabulación de datos cualitativos

La construcción de una distribución de frecuencia de atributos o distribución de frecuencia de variable cualitativa es simple, basta enumerar los diversos atributos con su respectiva frecuencia de ocurrencia.

**Frecuencia absoluta :** (  $f_i$  ) indica el número de veces que se repite un atributo.

Ejemplo:

Considérese una muestra 400 trabajadores de una cierta empresa de la región los cuales han sido encuestados sobre su actual estado civil. La información es tabulada de la siguiente manera:

Estado Civil	$f_i$
Soltero	75
Casado	200
Viudo	50
Separado	75
Total	400

$n = 400$  (tamaño de la muestra)

$m = 4$  (número de clases)

## Tabulación de variable discreta (que toma un conjunto pequeño de datos distintos)

Las tablas de frecuencia de variable discreta llevan cinco columnas donde los elementos que participan son los siguientes:

a) **Frecuencia absoluta** : (  $f_i$  ) indica el número de veces que se repite una variable.

b) **Tamaño de la muestra** : (  $n$  ) indica la cantidad de elementos que conforman la muestra, se obtiene sumando todas las frecuencias absolutas.

$$n = \sum_{i=1}^m f_i \quad m = \text{número de clases distintas}$$

c) **Frecuencia relativa** : (  $h_i$  ) es la proporción de datos que se encuentra en una clase, se obtiene dividiendo la frecuencia absoluta de la clase por el tamaño de la muestra.

$$h_i = \frac{f_i}{n} \quad \text{Obs:} \quad \text{a) } \sum h_i = 1$$

b)  $0 \leq h_i \leq 1$

d) **Frecuencia absoluta acumulada** : (  $F_i$  ) indica la cantidad de datos que se encuentran hasta cierta clase.

$$F_i = \sum_{j=1}^i f_j$$

e) **Frecuencia relativa acumulada** : (  $H_i$  ) es la proporción de datos acumulados que se encuentran hasta cierta clase.

$$H_i = \sum_{j=1}^i h_j \quad \text{Obs:} \quad \text{a) } H_m = 1$$

b)  $0 \leq H_i \leq 1$



# Ejercicio

Una empresa que tiene 50 trabajadores se propone reestructurar las remuneraciones, se estudia los años de servicio de los trabajadores determinándose los siguientes resultados:

4	5	4	6	7	9	7	7	5	8
8	7	6	7	7	4	6	8	8	9
6	8	9	5	6	5	4	7	9	6
7	6	5	4	4	4	6	8	8	7
8	9	5	5	4	6	7	9	5	4

$N = 50$  (tamaño de la población)

Se pide:

1. — Tabular la información.
2. — ¿Qué cantidad de trabajadores tiene 8 años de servicio ?.
3. — ¿Qué porcentaje de trabajadores tiene 6 años de servicio ?.
4. — Si aquellos trabajadores que tengan a lo menos siete años de servicio reciben un aumento del 8% .¿ Qué porcentaje de los trabajadores recibió dicho aumento?.
5. — Si todos los trabajadores que tengan a lo más cinco años de servicio reciben una bonificación de \$20.000 .¿ Qué cantidad de trabajadores recibió dicha bonificación?.
6. — Si la empresa decide otorgar una bonificación especial de \$13.200 por cada año de servicio.¿ Cuánto será el dinero necesario para cumplir dicha bonificación?.

# Tabulación de variable continua o discreta

Para tabular una variable continua o discreta (que tome un gran número de datos distintos) se necesitan los siguientes elementos:

**a) Rango o recorrido :** Es la diferencia entre el valor máximo y valor mínimo que toma la variable.

$$R = x_{\max} - x_{\min}$$

**b) Número de intervalos o clases ( $m$ ) :** Es el número de grupos en que es posible dividir los valores de la variable.

El número de clases no debe ser ni muy grande ni muy pequeño, un número pequeño de clases puede ocultar la naturaleza general de los datos y un número muy grande puede ser demasiado detallado como para revelar alguna información útil. Como regla general se recomienda que el número de clases esté entre cinco y veinte. Hay una regla llamada **Regla de Sturges** que puede dar una aproximación razonable para el número de clases, ella es:

$$m = 1 + 3,3 \log(n) \text{ donde } n \text{ es el número de datos de la muestra.}$$

**c) Amplitud del intervalo o amplitud de la clase ( $a$ ) :**

$$a = \frac{\text{Recorrido}}{\text{Nº de clases}} = \frac{R}{m}$$

**d) Límites de un intervalo :** Son los valores extremos de una clase. El menor valor es considerado como el **límite inferior** y el valor que se obtiene sumando al límite inferior la amplitud del intervalo es el **límite superior** de la segunda clase.

**e) Límites reales de un intervalo :** Se obtienen calculando el promedio entre el límite superior de una clase y el límite inferior de la clase siguiente.

**f) Marca de clase : ( $x_i$ )** Es el punto medio de un intervalo.

**g) Frecuencia absoluta : ( $f_i$ )** indica el número de observaciones que pertenece a un intervalo dado.

**Observación:** 
$$n = \sum_{i=1}^m f_i \quad n = \text{tamaño de la muestra}$$

**h) Frecuencia relativa : ( $h_i$ )** es la proporción de datos que se encuentra en un intervalo, se determina dividiendo la frecuencia absoluta del intervalo por el tamaño de la muestra.

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

**i) Frecuencia absoluta acumulada : ( $F_i$ )** indica el número de datos de la muestra menores o iguales al límite real superior del intervalo  $i$ .

$$F_i = \sum_{j=1}^i f_j \quad \text{Obs: } F_m = n$$

**j) Frecuencia relativa acumulada : ( $H_i$ )** indica la proporción de datos de la muestra menores o iguales al límite real superior del intervalo  $i$ .

$$H_i = \sum_{j=1}^i h_j$$

Los siguientes datos corresponden a las notas obtenidas por 100 alumnos en un curso de Estadística :

100	87	54	82	93	47	40	53	88	58
84	65	57	66	25	70	85	36	61	34
33	33	100	69	77	88	63	17	42	55
98	70	68	70	65	70	84	52	60	54
57	47	57	86	25	66	40	100	32	39
90	83	64	95	85	100	67	60	42	65
82	85	62	72	65	76	23	96	30	45
77	55	100	80	55	52	85	68	53	82
55	51	47	47	64	75	65	60	45	75
62	93	98	58	95	83	33	70	51	60

1. — Construya la correspondiente distribución de frecuencia.

2. — ¿En qué clase se concentra el mayor número de notas?

3. — ¿Cuál es la frecuencia absoluta del cuarto intervalo?. Interprete el resultado .

4. — ¿Qué porcentaje de los alumnos tienen una nota inferior a 57?

5. — ¿Cuántos alumnos tienen una nota superior a 46?

6. — Interprete la frecuencia acumulada del sexto intervalo.

7. — Interprete la frecuencia relativa acumulada del quinto intervalo.