

# ESTADÍSTICA Y PROBABILIDAD

## GUÍA DE CLASE

PROFESOR: Alonso Arroyo A.

CORREO: [alroyo@hotmail.com](mailto:alroyo@hotmail.com)

Cali, Julio 2018

## Guía 0: Para tener en cuenta (I): ¿Cómo aprendemos?

Para responder a esta pregunta, presento las ideas de William Glasser (1925-2013), psiquiatra estadounidense quien planteo la “Teoría de la Elección” aplicada a la educación. Según Glasser, el aprendizaje se da en forma piramidal como se ilustra a continuación:



## (II) La palabra ESTADISTICA...

¿Qué significado tiene para usted la palabra **ESTADISTICA**?

¿Por qué es importante?

¿Cuál es el sentido que usted le da al concepto de **ESTADISTICA**?

¿Qué problemas resuelve la **ESTADISTICA**?

¿Cuál es la relación de la **ESTADISTICA** con otras áreas del conocimiento (Administración, Biología, Economía, Ingeniería, etc.)?

## (III) ¿Para qué sirve un curso de ESTADISTICA?

- Adquirir las herramientas necesarias para hacer razonamiento Estadístico en el desempeño profesional y/o personal. Las principales herramientas son de:



- **Descripcion:** indicadores de tendencia, variacion y forma
- **Asociacion:** tablas de de frecuencias, indicadores de covarianza y correlacion
- **Graficas:** histograma, barras, circulares, tallo y hojas, caja, etc.
- **Inferencia:** estimacion de intervalos, prueba de hipotesis, diseño de experimentos
- **Modelacion estadistica:** modelos de regresion, series de tiempo
- **Simulacion.**

- Permitir a los estudiantes realizar toma de decisiones con base en un analisis estadistico.



La toma de decisiones puede ser de dos tipos:

- **No científica:** basarse en vaticinios o pronósticos sustentados en creencias o suposiciones personales.
- **Una visión integradora de conceptos de múltiples disciplinas:** La Economía, La Administración, La Política, La Ética, La Estadística, etc.

- Entender conceptos fundamentales como la Aleatorización, el efecto de variables no controladas (confusión) y las réplicas independientes

#### (IV) ¿Para qué sirve un curso de PROBABILIDAD?

- Medir el nivel de incertidumbre en una situación determinada
- Dar soporte a los conceptos estadísticos

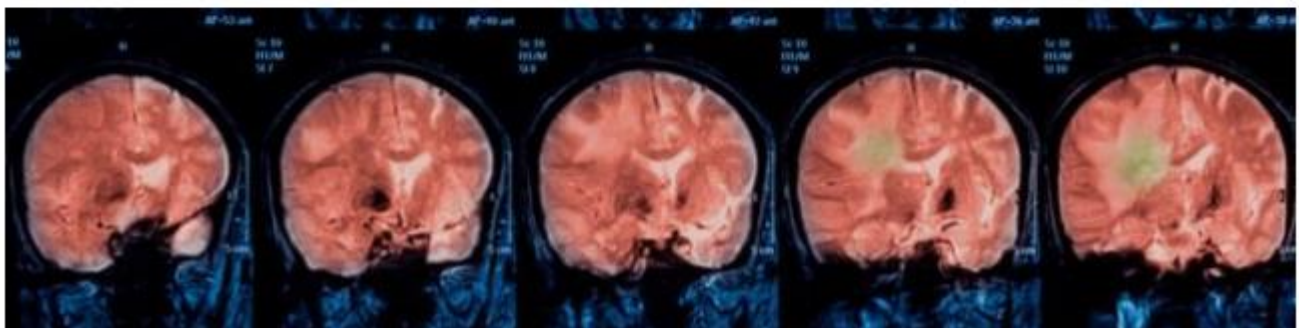
#### (IV) Un ejemplo

# La falla en los programas que leen los escáneres cerebrales que puede invalidar 15 años de investigación científica

Redacción  
BBC Mundo

🕒 7 julio 2016

🔗 Compartir



Desde que hace más de 20 años se empezaron a usar imágenes por resonancia magnética funcional (IRMf) para entender el comportamiento de nuestro cerebro, se han publicado unos 40.000 trabajos en el área de la neurociencia.

Ahora, un estudio sobre los programas que se utilizan para interpretar esas imágenes de los escáneres cerebrales **pone en duda la veracidad de los resultados** de esas investigaciones.

El equipo del investigador Anders Eklund, de la universidad de Linköping, en Suecia, decidió poner a prueba los tres programas informáticos más usados -SPM, FSL y AFNI- y encontró errores en su *software*.

"Quedé sorprendido con el resultado", le dijo a BBC Mundo. "Yo me esperaba encontrar **márgenes de error de 5%, pero no del 20, 40 y hasta 70%**".

"No esperé que fuera a ser tan malo", agregó.

*Esto es serio. Esto significa que algunos estudios están errados".*

*Anders Eklund, investigador de la universidad de Linköping, en Suecia*

Descubrió que los software llegaron a arrojar tasas de falsos positivos de hasta el 70%, donde se mostraba actividad en partes del cerebro donde en realidad no había.

"Esto es serio", sentencia Eklund. **"Esto significa que algunos estudios están errados"**.

En el trabajo publicado en el *Proceedings of the National Academy of Sciences*, los investigadores escriben que sus hallazgos **"cuestionan la validez de unos 40.000 estudios** con IRMf y pueden tener un gran impacto en la interpretación de los resultados de imágenes neuronales".

Uno de los problemas que descubrieron fue una falla informática en los programas que parece haber estado durante 15 años.

La buena noticia es que para el momento en que se escribió el trabajo, en 2015, esta falla se había arreglado.

La no tan buena noticia es que **"encontramos muchos métodos estadísticos que estaban basados en suposiciones que no eran correctas"**.

## **(V) ¿Cómo explicar estos resultados?**

La Estadística actual es el resultado de la unión de dos disciplinas que evolucionaron de manera independiente hasta confluir en el siglo 19, que son:

- Cálculo de probabilidades (nace en el siglo 17)
- Estadística (o ciencia del Estado, del latín Status) que estudia la descripción de los datos.

En el diccionario de la Real Lengua Española aparece la palabra **Estadística** bajo el significado de “censo o recuento de la población, de la producción, del tráfico o de cualquier otra entidad colectiva”.

Hay dos tipos de fenómenos o experimentos que se pueden observar y estudiar, pero la **Estadística** no se ocupa del estudio de ambos, sólo de uno de ellos, que son los fenómenos aleatorios o de azar. Los dos tipos de fenómenos son:

- **Fenómenos causales o determinísticos.**- Son los que están sometidos a leyes, repetidos en las mismas condiciones nos dan los mismos resultados. (Así, el tirar un objeto desde una cierta altura muchas veces, con la misma masa, tarda siempre el mismo tiempo).
- **Fenómenos aleatorios o de azar.** No están sometidos a leyes. Se caracterizan por la imposibilidad de conocer los resultados cuando se repiten en iguales condiciones. Sólo podemos predecir lo que sucederá después de repetir el fenómeno o experimento varias veces). Ejemplo: el valor promedio del peso de un producto en un proceso productivo.

El mundo en la actualidad nos ofrece una gran cantidad de datos que podemos evaluarlos para tomar buenas decisiones de negocios, resolver crímenes, analizar tendencias, identificar patrones de comportamiento etc. etc. etc. ¿Y Cómo hacerlo? La respuesta está en aprender a utilizar los métodos y las herramientas que nos ofrece la Estadística y esta será nuestra principal motivación para dedicarnos al estudio de esta disciplina.





## UN COMIC CON MUCHA VERDAD

**¡YA NO  
QUIERO  
ESTUDIAR MÁS!**

EL GOBIERNO TE QUIERE IGNORANTE, LA  
IGLESIA SUMISO, LOS EMPRESARIOS  
AGACHADO, LOS BANCOS FUNDIDO, LA  
JUSTICIA PRESO... POR ESO CUÁNDO TE PIDO  
QUE ESTUDIES ES PORQUE TE QUIERO VER  
LIBRE! EL CONOCIMIENTO ES TU ARMA.



[www.fb.com/sabiasquep](http://www.fb.com/sabiasquep)

## Guía de estudio 1: CONCEPTUALIZACION DE ESTADISTICA

### Unidad 1: Introducción a estadística descriptiva.

- ☐ Clasificación de los datos, introducción al muestreo.
- ☐ Agrupación en tablas de frecuencia, construcción de gráficos.
- ☐ Indicadores de centro, dispersión y forma, indicadores de posición, diagrama de caja

#### 1.1 Conceptos Básicos

El primer concepto a formalizar es el de la definición de Estadística: ¿Qué es Estadística? ¿Se podrán diferenciar varias clases de Estadísticas? Para responder a estas preguntas, se tomara como punto de partida una definición encontrada en la literatura actual:

***“Estadística es la ciencia de la interpretación de los datos y de la toma de decisiones en entornos de variabilidad e incertidumbre”<sup>1</sup>.***

Esta definición hace referencia a un concepto fundamental en todo tratamiento de datos que es el de la Variabilidad, ¿Cómo explicar la variabilidad en un resultado de laboratorio en dos especímenes sometidos a las mismas condiciones? Situación valida en cualquier análisis de datos estadístico, pero ¿Cómo se tomaron los datos? ¿Cuáles herramientas de análisis se utilizaron?. Para responder a estos interrogantes se deben analizar varias alternativas sobre el concepto de Estadística.

Cuando se trata de definir el concepto de Estadística se tiene que diferenciar entre tres campos de acción en el análisis de datos, estos son: la Estadística Descriptiva, la Estadística Inferencial y la Estadística Multivariada. Las diferencias entre la Estadística Descriptiva y la Estadística Inferencial están en la metodología de recolección de información y en las herramientas de análisis. En la Estadística Inferencial se recurre a una muestra tomada bajo condiciones de aleatoriedad que garanticen representatividad de la población objeto del problema. En la Estadística Descriptiva se recurre no necesariamente a una muestra aleatoria o se estudia a la población completa. A continuación se mencionan las herramientas objeto de análisis de cada una de estas estadísticas.

El interés en la Estadística Descriptiva se centra en tres elementos:

- El análisis de patrones o tendencias,
- La medición de la variación,
- La identificación de la forma o distribución de los datos.

Para el estudio de estos elementos se recurre a una variedad de opciones presentes en tablas, gráficos e indicadores estadísticos.

Las herramientas principales de la Estadística Inferencial son:

- El muestreo
- Las distribuciones muestreo
- La estimación de parámetros
- La prueba de hipótesis
- Análisis de correlación
- Los modelos estadísticos de regresión
- El diseño de experimentos

---

<sup>1</sup> Myra L., Jeffrey A., Andrew A. Fundamentos de Estadística para las Ciencias de la Vida, cuarta edición, 2012.

Cada una de estas herramientas tiene sus procedimientos de análisis que van desde el cálculo del tamaño de muestra hasta la proyección de los resultados sobre la población objeto de análisis.

La Estadística Multivariada maneja simultáneamente muchas variables, lo que no hacen las otras estadística mencionadas, y posee sus propias herramientas de análisis tal como análisis de correspondencias, análisis factorial, regresión lineal múltiple, etc.

**Definición 1.1: Población o universo.** La población es el conjunto de todos los individuos o elementos sobre los cuales se desea extender los resultados de una investigación.

**Definición 1.2: Parámetro.** Término con el cual se identifica a un indicador que hace referencia a la población.

**Ejemplo:** La población total en Colombia reportada en el censo del año 2005 por el Dane<sup>2</sup> fue de 42'888,592 personas. Las regiones con mayor población fueron Bogotá (6'840,116) Antioquia (5'682,276) y Valle del Cuaca (4'161,425). Todos estos indicadores son ejemplos de parámetros estadísticos.

**Definición 1.3: Muestra.** Muestra es una parte de una población que se espera provea los datos estadísticos necesarios para estimar las características de la población con unos errores controlados y conocidos.

**Definición 1.4: Estadística.** Es una medida o indicador asociada a una muestra.

**Ejemplo:** lea la siguiente noticia publicada en el diario EL TIEMPO el 18 de mayo del 2018:

 **EL TIEMPO**

ELECCIONES COLOMBIA 2018 | PRESIDENCIALES CONGRESO SERVICIOS

# Duque y Petro puntan en nueva encuesta de intención de voto

Así lo revela la encuesta de Guarumo. Fajardo y Vargas Lleras los siguen en esta medición.

---

<sup>2</sup> Fuente: página Web del Dane: <http://www.dane.gov.co>

Por: **ELTIEMPO.COM** | 18 de mayo 2018 , 12:49 p.m.

A una semana de la primera vuelta presidencial, los candidatos Iván Duque y Gustavo Petro puntúan en los dos primeros lugares en una nueva medición de la intención de voto, realizada por la firma encuestadora Guarumo para EL TIEMPO y la W Radio.

A la pregunta: "**¿Si las elecciones a la Presidencia fueran mañana usted por quién votaría?**", el 37,6 % por ciento de los encuestados se inclinó por Iván Duque, el 24,2% por Gustavo Petro, un 16 % por Sergio Fajardo, el 11 % por Germán Vargas Lleras, y el 2,3 % por Humberto de la Calle. El voto en blanco ajusta un 5,8 por ciento de la intención de voto, según esta medición.

La encuesta fue realizada por la empresa Guarumo S.A.S. en los días 12 al 16 de mayo del 2018 con un tamaño de muestra de 7,182 personas adultas ponderadas por nivel socio económico, género y rangos de edad, con base en una población habilitada para votar de 36'011,052 según datos de la Registradora Nacional del Estado civil a diciembre 11 del 2017.

Los valores de 37.6% y 24.2% de intención de voto para los candidatos Duque y Petro respectivamente son Estadísticas ya que se obtuvieron de un muestreo.

**Definición 1.5: Variable.** Es la característica que interesa medir.

## 1.2 Clasificación de variables

De acuerdo al tipo de dato existen varias clasificaciones de las variables. Las variables pueden ser Cualitativas o Cuantitativas, si la característica de interés es un atributo se dice que la variable es cualitativa, por ejemplo, el Color del empaque de un producto. Si la característica es una cantidad, se tiene una variable cuantitativa, por ejemplo, Peso de un producto.

Las variables cuantitativas pueden ser Discretas o Continuas, se clasifica una variable como discreta si sus valores surgen de un proceso de conteo, si los valores provienen de un proceso de medición se tiene una variable continua.

## Definición 1.6 Escalas de medición

De acuerdo al tipo de dato se definen las siguientes escalas de medición:

- **Escala Nominal:** variables cualitativas a las que solo se pueden establecer categorías o grupos.

### Ejemplos:

- a. Grupo sanguíneo de una persona: A, B, AB, O.
- b. En una encuesta realizada a cada uno de los empleados administrativos de una organización grande, si es propietario o no de auto. Dos categorías: sí es propietario, no es propietario.
- c. Tipo de vehículo que ingresa al parqueadero de una empresa. Varias categorías: automóvil, camión, bus, moto, etc.

- d. Clasificación de las unidades de producción en una jornada laboral: Defectuosas y No Defectuosas.

En esta escala la única operación posible es el conteo, las otras operaciones no se pueden efectuar o no tienen sentido.

**Ejemplo:** En una empresa distribuidoras de alimentos se clasifican los empleados en: Empacadores, Conductores, Almacenistas, Vigilantes y Administradores. Aquí la única operación permitida para la variable Clasificación de empleado es el conteo, en esta empresa actualmente se tienen 15 empacadores, 4 conductores, 2 almacenistas, 3 vigilantes y 5 administradores.

- **Escala ordinal:** Variables cualitativas en las que se pueden hacer comparaciones de la forma “mayor que”, “igual que” o “menor que”.

**Ejemplos:**

- a. Calidad de los productos disponibles para el público en un estante en un hipermercado. Las categorías: muy bueno, bueno, regular, malo, permiten establecer comparaciones de mayor, igual o menor calidad.
- b. Respuesta de un paciente a una terapia: ninguna, parcial, completa
- c. Calificaciones de los 40 estudiantes de un curso de inglés. En algunas instituciones educativas las notas son letras: E: excelente, B: bueno, A: aceptable e I: insuficiente.

- **Escala de intervalo:** Variables cuantitativas en que todos los números están referidos a un punto cero arbitrario. Tal es el caso de la temperatura medida en grados centígrados (el punto cero grados centígrados no significa ausencia de temperatura).

**Ejemplos:**

- a. El tiempo calendario. Depende de la cultura, de la civilización que se quiere mirar: los chinos tienen un sistema muy distinto al sistema occidental, luego el año 2014 para la cultura china corresponde a otro año.
- b. El cociente intelectual, CI. El CI es un test para medir la inteligencia de una persona. Los usados actualmente tienen un promedio de 100 y una desviación estándar de 15.

En la escala de intervalo, la suma y la resta son posibles pero no la multiplicación ni la división.

- **Escala de Razón:** Es la escala más completa, es posible hacer todas las operaciones y se diferencia de la escala de intervalo en el hecho de que el cero es significativo, por ejemplo, para una persona que no tiene ninguna entrada de dinero, el valor 0 indica ausencia de dinero en cualquier sistema monetario.

**Ejemplos:**

- a. El salario de un empleado. En Colombia el salario mínimo diario para el año 2018 es de \$ 26,041.4, en el año 2003 era de \$ 12,317.
- b. Total de exportación de café por año en Colombia. En el año 2016 Colombia exportó 12'844,833 sacos de 60 kilogramos, en el año 2012 se exportaron 7'168,998 sacos.
- c. Otros ejemplos en esta escala son: Peso de una persona, en kilogramos; Tiempo de atención, en segundos, por cliente en un banco; ingresos por servicio prestados en una institución, etc.

Es importante identificar el nivel de medición de las variables porque dependiendo de dicho nivel se selecciona el tipo de análisis estadístico.

**Ejemplo:** existe un coeficiente de correlación para variables ordinales (coeficiente de correlación polícórica) y otro para variables en escala de razón (coeficiente de correlación de Pearson).

**Nota:** es bueno advertir que en las variables de intervalo y de razón se pueden aplicar análisis estadísticos utilizados para variables cualitativas pero no al contrario.

### 1.3 Taller

1. Formalice una definición del concepto de Estadística Descriptiva.
2. Formalice una definición del concepto de Estadística Inferencial. Construya un diagrama de flujo que represente este concepto.
3. (tarea) El jefe de un taller de mantenimiento mecánico de una compañía grande realizó un estudio estadístico para determinar la capacidad operativa de sus empleados y el estado de funcionamiento de las máquinas disponibles en el taller. Se tuvieron en cuenta las siguientes variables:
  - **Categoría de empleado:** Jefe, subjefe, operario
  - **Experiencia en labores de mantenimiento mecánico** (en años)
  - **Certificaciones Académicas:** Bachiller, Tecnólogo, Profesional
  - **Número de máquinas en mantenimiento** (en unidades)
  - **Costo anual de mantenimiento de cada máquina** (en dólares)
  - **Tipo de máquina:** Rotor Rf"345", Torno, Motor, Fresa, Cortadora
  - **Estado de la máquina:** De baja, en mantenimiento, funcionando óptimamente

En el estudio se hizo un inventario de todas las máquinas y herramientas utilizadas, encontrándose que el 90% de todas las máquinas estaban funcionando en condiciones óptimas.

Resuelva:

- a. Clasifique cada una de las variables de acuerdo al tipo (cualitativa, cuantitativa) y a la escala de medición (nominal, ordinal, intervalo, razón)
- b. A) Indique el tipo de estudio realizado: Exploratorio, Descriptivo o Inferencial. B) Identifique la población o la muestra utilizada C) ¿El valor de 90% hace referencia a una Estadística o un Parámetro? Justifique cada una de sus respuestas.
4. Clasifique las siguientes variables según sean: Cualitativas o Cuantitativas, Discretas o Continuas y según la escala de medición: nominal, ordinal, de intervalo o de razón.
  - a. Consumo de energía (en kilovatios) de una planta procesadora de alimentos
  - b. Temperatura (en grados centígrados) de un refrigerador de carnes en un hipermercado local
  - c. Tipos de carreras en la facultad de administración en una universidad local (Administración, Economía, Contaduría)
  - d. Autonomía de vuelo<sup>3</sup> de los aviones de la Fuerza Aérea Colombiana
  - e. Tipos de las lesiones tratadas en el departamento médico de un club deportivo en la ciudad de Cali en el último mes
  - f. Tiempo (en minutos) que permanece un automóvil en un negocio de parqueo en un día laboral
  - g. Número de actividades realizadas por un asistente administrativo en una jornada laboral
  - h. Nivel educativo alcanzado por cada uno de los integrantes de una familia (secundaria, técnica, profesional y postgrado)

---

<sup>3</sup> **Autonomía:** tiempo en que una aeronave se mantiene en vuelo

- i. Lectura del velocímetro de un automóvil en una autopista local
  - j. Monto de la deuda externa en Colombia durante los últimos cinco años (en millones de pesos)
5. Las siguientes son algunas estadísticas reportadas por la entidad Cali Cómo Vamos en una encuesta realizada en diciembre del 2017. Resuelva las preguntas que siguen a continuación.

¿Qué medio de transporte usa principalmente para desplazarse?		
	2016	2017
Mío	34%	35%
Moto	22%	24%
Vehículo particular	15%	14%
Bicicleta	6%	5%
Taxi	5%	5%
Bus / buseta tradicional	7%	5%
A pie	4%	5%
Transporte informal	7%	7%

Se siente inseguro, según zona		
	Barrio	Ciudad
Nororiental	37%	51%
Noroccidente	22%	40%
Distrito Aguablanca	42%	37%
Sur	28%	32%
Oriente	22%	29%

## Estos son los resultados de la encuesta Cali Cómo Vamos 2017

19 dic 2017

La encuesta de Percepción Ciudadana de Cali Cómo Vamos en 2017, realizada por Ipsos Napoleón Franco se aplicó a 1.503 personas, mayores de edad residentes habituales en todas las comunas de Cali y Jamundí.

¿Se siente inseguro en...?	
Cali	

2017	
Barrio	Ciudad
22%	26%

¿Ud fue víctima de algún delito el último año?		
	2016	2017
Sí	18%	22%

¿De qué delito fue víctima?		
	2016	2017
Atraco- robo - raponazo	78%	85%

Preguntas:

- a. Identifique el tipo de estudio: Descriptivo o Inferencial realizado por en la investigación.
- b. Identifique la población objeto de estudio en esta investigación.
- c. Identifique las variables presentadas y haga una clasificación de cada una de ellas de acuerdo al tipo y la escala de medición.
- d. Identifique parámetros o estadísticos reportados en esta investigación.

### 1.4 Introducción al muestreo

La Inferencia Estadística se apoya ciento por ciento en el muestreo, una buena muestra es garantía de la calidad de las conclusiones a que se lleguen al final del estudio, esta y otras muchas razones hacen necesario un análisis de las condiciones para tomar la muestra de la población en cuestión. Pero antes de tomar la decisión de hacer muestreo cabe la pregunta: ¿es realmente necesaria la muestra? o equivalentemente ¿la información obtenida por la muestra habrá de ayudar a resolver el problema objeto de estudio? Si estas respuestas son afirmativas se justifica plenamente la realización de un plan de muestreo. En este capítulo se formalizarán algunos conceptos básicos del muestreo y deducirán las condiciones para lograr una muestra representativa de la población objeto de estudio.

#### 1.4.1 Muestreo no probabilístico

Los muestreos no probabilísticos o determinísticos se definen como aquellos muestreos en los cuales no se usa el azar para seleccionar los elementos. Los criterios de selección son circunstanciales y se dejan a quien está a cargo de la investigación.

Los principales tipos de muestreos no probabilísticos son: i. **Muestreo por Juicio, Selección Experta o Selección Intencional**, ii. **Muestreo casual o fortuito**, iii. **Muestreo de cuota**, iv. **Muestreo de poblaciones móviles**, v. **Muestreo bola de nieve**, entre otros.

**Ejercicio 1:** Investigue cada una de las características de los tipos de muestreos no probabilísticos listados en el párrafo anterior.

**Ejercicio 2:** Identifique la clase de muestreo no probabilístico en cada uno de los casos siguientes:

- a. El jefe de mercadeo de un hipermercado encuesta a varios clientes que salen del negocio para conocer la opinión sobre un determinado producto.
- b. En el diario El País de Cali se pide la opinión de los lectores sobre si considera justa la condena de 17 años contra Andrés Felipe Arias por el escándalo de Agro Ingreso Seguro.
- c. Un investigador toma muestras del carbón extraído de una mina, tomando intencionalmente trozos de carbón de la parte superior de cada carro.
- d. La cooperativa Coomeva desea dar a conocer un nuevo plan vacacional para lo cual rifa un BMW entre las personas que lleven cinco personas más a una charla sobre el nuevo plan vacacional.

**Ejercicio 3:** Identifique casos de la vida real donde se recurra a muestreos no probabilísticos. (hasta aquí taller)

#### 1.4.2 Los métodos de muestreo probabilísticos

Los métodos de muestreo probabilísticos garantizan la representatividad de la muestra para la población objeto de estudio. A continuación se listan los principales métodos de muestreo, luego se describen cada uno de ellos indicando las condiciones de su aplicación.

- Muestreo Aleatorio Simple (MAS)
- Muestreo Aleatorio Estratificado
- Muestreo Aleatorio de Conglomerados
- Muestreo Sistemático
- Muestreo aleatorio Multi-etápico
- Muestreo de aceptación

##### a. Muestreo aleatorio simple (MAS)



Las características y condiciones para utilizar este método de muestreo son:

- La población debe ser homogénea en la característica de interés
- Debe definirse un marco de muestreo
- Cada elemento de la población tiene la misma probabilidad de selección
- Cada elemento es independiente de los demás
- La muestra se toma mediante un procedimiento aleatorio

Un procedimiento para seleccionar una muestra aleatoria simple es el siguiente:

- Seleccionar un marco de muestreo adecuado
- Asignar a cada elemento de la población un número de 1 hasta  $N$
- Generar  $n$  números aleatorios diferentes comprendidos en el intervalo 1 hasta  $N$
- Los números así generados indican los elementos o unidades de la población que se deben incluir en la muestra.

**Ejercicio 4:** Considere lista de clase del curso actual de Estadística, proponga varios procedimientos aleatorios para seleccionar una muestra de tamaño  $n = 5$ .

#### **b. Muestreo aleatorio sistemático**

Se caracteriza por lo siguiente:

- La población se ubica en una base de datos, lista, directorio telefónico, mapa, etc. de tal manera que se pueda identificar una secuencia ordenada de acuerdo a una característica de interés y que se puedan seleccionar elementos a intervalos constantes
- Si la población es de tamaño  $N$  y la muestra  $n$ , se define un número  $k = \frac{N}{n}$
- Se genera un solo número aleatorio entre 1 y  $k$ , al que se denominará  $L$ .  $L$  será el primer elemento de la muestra.
- Los  $(n - 1)$  elementos restantes de la muestra seguirán la secuencia:

$$L + k; L + 2k; L + 3k; \dots L + (n-1)k$$

**Nota 1:** si el ordenamiento identificado en la población es aleatorio (es decir los elementos de la población están ordenados al azar) el muestreo sistemático se puede considerar equivalente al muestreo aleatorio simple, y por lo tanto todas las formulas del MAS son equivalentes para el sistemático.

**Nota 2:** Si el esquema de ordenamiento en la población sigue un orden de magnitud de acuerdo a la característica analizada, el muestreo sistemático es más conveniente que el muestreo aleatorio simple.

**Ejercicio 5:** Las directivas de una institución universitaria compuesta de 510 empleados quiere consultar la opinión de sus empleados con respecto a las jornadas de salud implementadas en el último año, mas concretamente se quiere determinar si estas jornadas cumplen con las expectativas de la mayoría de los empleados. Para conocer la opinión de los empleados se recurre a una encuesta la cual será aplicada a una muestra aleatoria de empleados. Indique el procedimiento para seleccionar una muestra aleatoria sistemática de 30 empleados. Realice una simulación del proceso de selección de la muestra con el programa Excel.

#### **c. Muestreo aleatorio estratificado**

Cuando la población se divide en grupos relativamente homogéneos con relación a la característica principal del estudio. Las características de este método de muestreo son las siguientes:

- En la población se identifican grupos denominados Estratos, cada grupo debe ser homogéneo internamente y claramente diferenciable de los otros grupos.
- Se debe definir una variable de estratificación para identificar los estratos.
- Los estratos pueden o no estar compuestos del mismo número de unidades, por tal razón la fracción de muestreo ( $f = n/N$ ) puede variar de un estrato a otro.

Existen tres formas alternativas de realizar la estratificación:

- **Asignación igual:** Muestras de tamaño igual en todos los estratos
- **Asignación proporcional:** Los tamaños de muestra en cada estrato son proporcionales a los tamaños de los estratos poblacionales
- **Asignación óptima:** Cada tamaño por estrato es determinado de acuerdo a los costos y al grado de variabilidad, de tal forma que el error de estimación sea mínimo para un costo total dado.

**Ejercicio 6:** Una entidad de investigación socioeconómica está interesada en establecer los ingresos promedio en un determinado municipio, y decide realizar una encuesta por muestreo para estimar los ingresos promedio por familia en los hogares del mismo. El municipio se compone de una cabecera municipal (estrato 1), una zona rural (estrato 2) y una zona industrial (estrato 3). La cabecera municipal obtiene ingresos principalmente del comercio, la zona rural es fundamentalmente cultivadora de café y la zona industrial corresponde a una gran fábrica productora de azúcar y alcohol carburante. El estrato 1 está compuesto de 155 hogares, el 2 de 62 y el 3 de 93. Analice los méritos de usar muestreo estratificado en esta situación.

#### d. Muestreo aleatorio por conglomerados

Los conglomerados son agrupaciones con características similares entre ellos pero con una variación importante dentro de cada grupo, se puede pensar que cada conglomerado representa las características de la población. Las agrupaciones (o los conglomerados) se forman frecuentemente por zonas geográficas. A modo de ejemplo, suponga que se quiere investigar la acogida de una nueva marca de bebida gaseosa en la ciudad de Cali, una manera de hacerlo sería identificar agrupaciones naturalmente formadas como son los grandes centros comerciales, cada centro sería un conglomerado, se seleccionarían algunos de ellos y se haría la encuesta a los clientes que se encuentren allí.

Las principales características son:

- En la población se identifican grupos que, a diferencia de los del muestreo estratificado, se espera que internamente sean heterogéneos en la variable de agrupación
- Generalmente los conglomerados los forman unidades cercanas geográficamente
- Se usa un muestreo por conglomerado principalmente para reducir costos y comodidad en la toma de datos más que por lograr márgenes de error menores que otros métodos de muestreo.

**Ejercicio 7:** Investigue un caso de muestreo por conglomerado

**Ejercicio 8:** Un Profesor de Estadística quiere investigar sobre el tiempo diario de estudio de 30 estudiantes de una clase:

Estudiantes de Ingeniería	Número de horas	Estudiantes de Administración	Número de horas	Estudiantes de Contaduría	Número de horas
Miguel	3.0	Juan	2.0	Amparo	1.0
Victoria	2.5	Alicia	2.0	Astrid	1.5

María	2.5	Pedro	1.5	Beatriz	1.0
Fernanda	3.0	Marcos	2.0	Cindy	0.5
Julio	3.5	Alberto	1.5	Rafael	0.5
Rosa	2.5	Jorge	2.0	Eduardo	0.0
Fabián	2.0	José	2.0		
Ana	2.5	Carlos	2.0		
Laura	2.0	Eval	1.5		
Enrique	2.5				
Carmen	3.0				
Marcelo	3.0				
Patricia	4.0				
Dalila	2.5				
Juliana	3.0				

Resuelva indicando el procedimiento seguido para la selección de la muestra.

- Elija e indique una Muestra no aleatoria de tamaño 5 de esta población y estime el número de horas promedio de estudio. Comente sobre las desventajas de esta muestra.
- Suponga que se puede hacer muestreo aleatorio simple en esta población, elija e indique una Muestra Aleatoria Simple de tamaño 5 de esta población y estime el número de horas promedio de estudio. Comente sobre las desventajas de esta muestra.
- Elija e indique una Muestra Aleatoria sistemática de tamaño 5 de esta población y estime el número de horas promedio de estudio. Comente sobre las ventajas de esta muestra.
- Elija e indique una Muestra Aleatoria Estratificada de tamaño 10 de esta población usando asignación proporcional y estime el número de horas promedio de estudio.
- Concluya sobre el procedimiento de muestreo más adecuado para estimar el número de horas promedio.

## 1.5 Tabla de frecuencias

### Definición 1.7: Tabla de frecuencias

La tabla de frecuencias resume grandes volúmenes de información disponiéndola en categorías o agrupamientos (llamadas clases para datos cuantitativos).

**Ejemplo:** El director de un gran centro comercial en una ciudad local reporta el tiempo de permanencia de los clientes que ingresan con carro al centro comercial en un día corriente de la semana. La información esta resumida en la siguiente tabla de frecuencias:

Tiempo (en minutos)	No. vehículos	% de vehículos
[ 0 ; 30]	24	8.0
( 30 ; 60]	80	26.7
( 60 ; 90]	130	43.3
( 90 ; 120]	46	15.3
(120 ; 180]	20	6.7
Total	300	

Pasos en la construcción de una tabla de frecuencias:

**a. número de agrupamientos o clases**

El número apropiado de agrupamientos o clases depende del número de observaciones y de si la variable es discreta o continua. La experiencia en el manejo de datos indica que este número oscila entre 5 y 20, el número de agrupamientos se representa con la letra  $k$ . Existen varias fórmulas empíricas que sirven para estimar el número de clases; algunas de ellas son:

- $k \approx 1 + 3.3\text{Log}(N)$
- $2^k \geq N$  se despeja  $k$ , obteniéndose  $k = \frac{\text{Ln}(N)}{\text{Ln}(2)}$
- $k \approx \sqrt{N}$

donde  $N$  es el tamaño de la población.

**b. Ancho de clase**

Se deben definir los anchos de cada clase. Esto se puede hacer arbitrariamente, pero en la práctica se acostumbra a manejar las clases con igual amplitud lo cual facilita la interpretación de la tabla de frecuencias.

Si se toman clases de igual amplitud, la fórmula es:

$$\text{Amplitud} \approx \frac{\text{Valor mayor de los datos} - \text{Valor menor de los datos}}{\text{Numero de clases}}$$

En forma simplificada seria:

$$C \approx \frac{X_{\text{mayor}} - X_{\text{menor}}}{k}$$

donde  $C$ : amplitud o ancho de clase,  $X_{\text{mayor}}$ : valor mayor de los datos y  $X_{\text{menor}}$ : valor menor de los datos.

**c. Los límites de cada clase**

Cada clase tiene un límite inferior y un límite superior. En la primera clase el límite inferior es el valor del dato menor:  $\text{limite inferior} = X_{\text{menor}}$ , a este valor se le suma el ancho de clase ( $C$ ) resultando el límite superior para la primera clase:  $\text{limite superior} = X_{\text{menor}} + C$ . Para la segunda clase, el límite inferior será igual al límite superior de la primera clase más una cierta cantidad y el límite superior de esta segunda clase se forma tomando el límite inferior acabado de formar más el ancho de clase ( $C$ ). Las demás clases se forman igual que la segunda clase.

**d. Marca de clase**

La marca de clase es el valor que se encuentra justamente en la mitad entre los límites de cada clase. Será un valor que representará a los datos de la clase en el cálculo de indicadores o en la representación gráfica de la variable. Se obtiene mediante la fórmula:

$$\text{Marca de la clase } i \approx \frac{\text{limite inferior de la clase } i + \text{limite superior de la clase } i}{2}$$

**e. Frecuencia absoluta**

La frecuencia absoluta es el número de datos contados en cada clase.

**f. Frecuencia relativa**

La frecuencia absoluta permite determinar el comportamiento de los datos a través de todas las clases o agrupamientos, se visualiza la tendencia y la variación del conjunto de datos en cuestión, pero tiene el inconveniente de que no permite comparaciones con datos de poblaciones similares. Para poder hacer esto se recurre a la frecuencia relativa, la cual da un número entre 0 y 100% permitiendo hacer dichas comparaciones.

$$\text{Frecuencia relativa de la clase } i = \frac{\text{Frecuencia absoluta de la clase } i}{\text{Numero total de observaciones}} * 100\%$$

#### **g. Frecuencia acumulativa**

En algunas situaciones interesa observar cómo se acumula la frecuencia absoluta por lo cual se construyen la frecuencia acumulativa absoluta y la frecuencia acumulativa relativa. La frecuencia acumulativa se calcula sumando los valores de la frecuencia absoluta.

**Ejercicio 9:** (tarea) (ejercicio 1.22 del texto de Walpole) Los siguientes datos son las mediciones del diámetro de 36 cabezas de remache en centésimos de una pulgada.

6.72, 6.77, 6.82, 6.70, 6.78, 6.70, 6.62, 6.75, 6.66, 6.66, 6.64, 6.76, 6.73, 6.80, 6.72, 6.76, 6.76, 6.68, 6.66, 6.62, 6.72, 6.76, 6.70, 6.78, 6.76, 6.67, 6.70, 6.72, 6.74, 6.71, 6.79, 6.78, 6.66, 6.76, 6.76, 6.72.

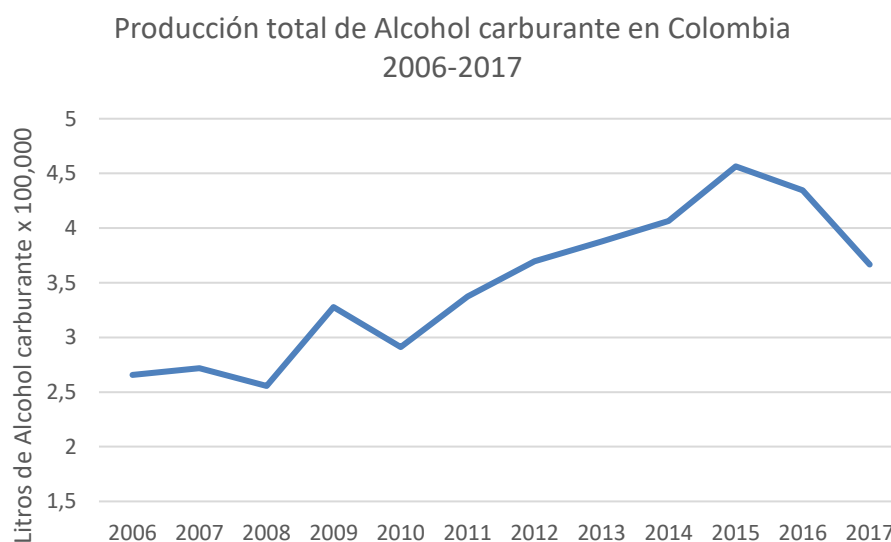
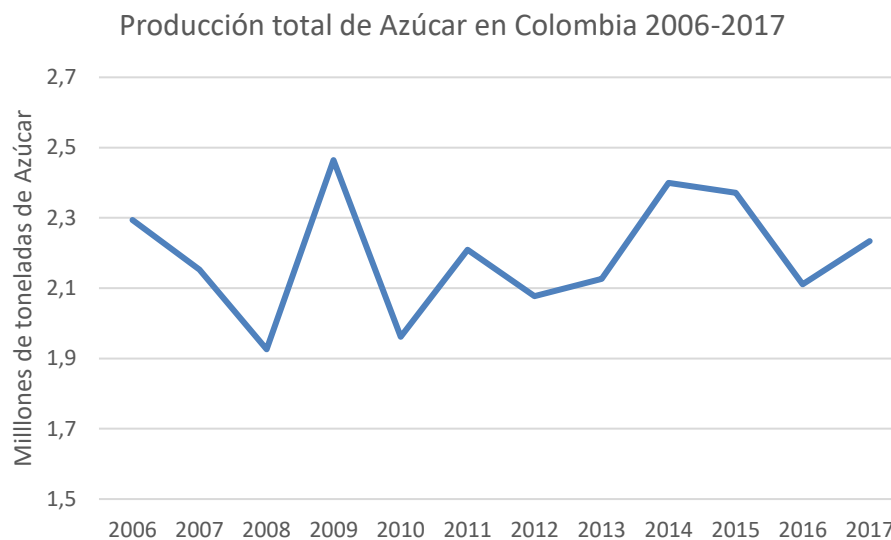
- construya la tabla de frecuencias para los datos
- Construya un histograma de frecuencias relativas para los datos
- Comente si existe o no una indicación clara de que la muestra proviene de una población que tiene una distribución en forma de campana.

### **1.6 Indicadores de tendencia**

La tendencia es una de las características de análisis de la Estadística Descriptiva. Cada situación o problema objeto del análisis estadístico requiere de la elección adecuada de los indicadores y/o herramientas gráficas. Por ejemplo, si se quiere determinar si el tiempo promedio de terminación de una tarea en un proceso de fabricación ha cambiado al implementar un cambio en una de las herramientas de trabajo, se debe recurrir a indicadores de tendencia central y a la construcción de un histograma; por otro lado, si se quiere evaluar el impacto que tiene el precio del dólar en las exportaciones de petróleo en Colombia en los últimos cinco años, se debe analizar la tendencia de las exportaciones a través del tiempo y una gráfica de dispersión (o una gráfica de serie de tiempo) sería la más adecuada.

#### **Definición 1.8: Serie de tiempo**

La gráfica de serie de tiempo se construye cuando se quiere mostrar la variación de los datos a través del tiempo. Variables típicas de este comportamiento son: ventas, número de tareas realizadas, accidentes laborales ocurridos en un determinado tiempo, etc. En las gráficas adjuntas se muestra el comportamiento de la producción de caña de azúcar y la producción de alcohol carburante en Colombia en un periodo de tiempo.



Fuente: Asocaña. Documento: Balance azucarero colombiano Asocaña 2006 - 2017

**Ejercicio 10:** Analice en términos de la Estadística Descriptiva las gráficas anteriores. ¿Qué implicaciones tiene para la región el comportamiento de las producciones de caña de azúcar y de alcohol carburante?

La estadística descriptiva tiene disponibles indicadores para determinar la tendencia central de un conjunto de datos que no tengan secuencia en el tiempo, por ejemplo, en un turno de trabajo se hace un control de calidad al peso de un producto empacado de 500 gramos, se espera que el peso de todos los productos tengan una tendencia hacia el valor de 500 gramos, el promedio o media aritmética del peso sería el mejor indicador para representar dicha tendencia. Por otro lado, si el interés es el precio diario de exportación de este producto, ya no interesa la tendencia central sino las posibles fluctuaciones a través del tiempo del precio en el mercado internacional, luego el promedio del precio podría no ser un buen indicador de la tendencia del precio.

**Definición 1.9: Media aritmética:** para un conjunto de datos numéricos, es la suma de cada uno de los datos dividida por el total de datos. Simbólicamente es:  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$  (cuando es una muestra)  $\mu = \frac{\sum_{i=1}^N x_i}{N}$  (cuando es una población).

**Definición 1.10: Mediana:** es el valor central de un conjunto ordenado de datos.

**Definición 1.11: Moda:** es el dato de mayor frecuencia en un conjunto de datos. En algunos casos cuando hay dos datos con marcada frecuencia se identifican dos modas (se dice que el conjunto de datos es bimodal).

La mejor manera de identificar la tendencia central de un conjunto de datos es con una gráfica, en muchas situaciones se tienen pocos datos y se debe recurrir a técnicas como diagrama de Tallo y Hojas, diagramas de Cajas y Bigotes y/o diagrama de Puntos. A continuación se describe la construcción de un diagrama de tallo y hojas.

**Definición 1.12: Diagrama de tallo y hojas:** Es una técnica gráfico-numérica que permite presentar y organizar un conjunto pequeño de datos (por lo general menos de 50).

**Ejercicio 11:** Se desea describir gráficamente los datos que se muestran a continuación; estos corresponden a las notas de un parcial de un curso de Estadística.

3.0, 0.8, 1.2, 1.6, 2.4, 2.8, 3.5, 4.0, 4.5, 5.0, 1.9, 2.0, 2.6, 3.0, 3.8, 4.1, 4.8, 2.1, 2.7, 3.2, 3.9, 4.1, 2.8, 3.3, 3.4.

Investigue como se construye una gráfica de tallo y hojas y elabore la respectiva gráfica para estos datos.

**Ejercicio 12: Comportamiento de compra de los clientes en un hipermercado de la ciudad de Cali.**

El gerente de una cadena de almacenes en la ciudad de Cali quiere determinar cuál es el comportamiento de compra de los clientes en cada una de las tres sucursales ubicadas en tres sitios distintos de la ciudad. Se tienen las ventas netas (en millones de pesos) durante 19 días de prueba:

**Sucursal 1:** 12.9, 13.4, 13.6, 14.0, 14.5, 15.0, 15.5, 13.4, 13.8, 14.0, 14.7, 15.3, 13.9, 14.2, 14.8, 14.2, 13.7, 14.2, 14.5

**Sucursal 2:** 12.8, 13.0, 13.5, 14.0, 14.5, 15.4, 20.9, 13.3, 13.5, 14.0, 14.9, 13.3, 13.6, 14.1, 13.4, 13.7, 13.4, 13.8, 13.9.

**Sucursal 3:** 12.4, 12.8, 12.9, 13.4, 13.4, 13.4, 13.4, 13.5, 13.6, 14.8, 15.4, 15.4, 15.8, 15.9, 15.9, 16.0, 16.0, 16.5.

El propósito es hacer un análisis estadístico de los datos y generar un reporte al gerente. Para hacer esto debe realizar las actividades propuestas a continuación.

1. Realice diagramas de tallo y hojas para cada muestra e identifique la tendencia en cada caso.
2. Trate de asociar la tendencia identificada en 1) con un valor numérico (o dos si es el caso).
3. Calcule la media aritmética, la mediana y la moda para cada sucursal. Compare estos resultados con los obtenidos en 2).
4. Establezca reglas para la utilización de cada indicador de tendencia central.
5. ¿Cuál de las tres sucursales brinda mayor confianza en cuanto al volumen de venta? Justifique su respuesta.

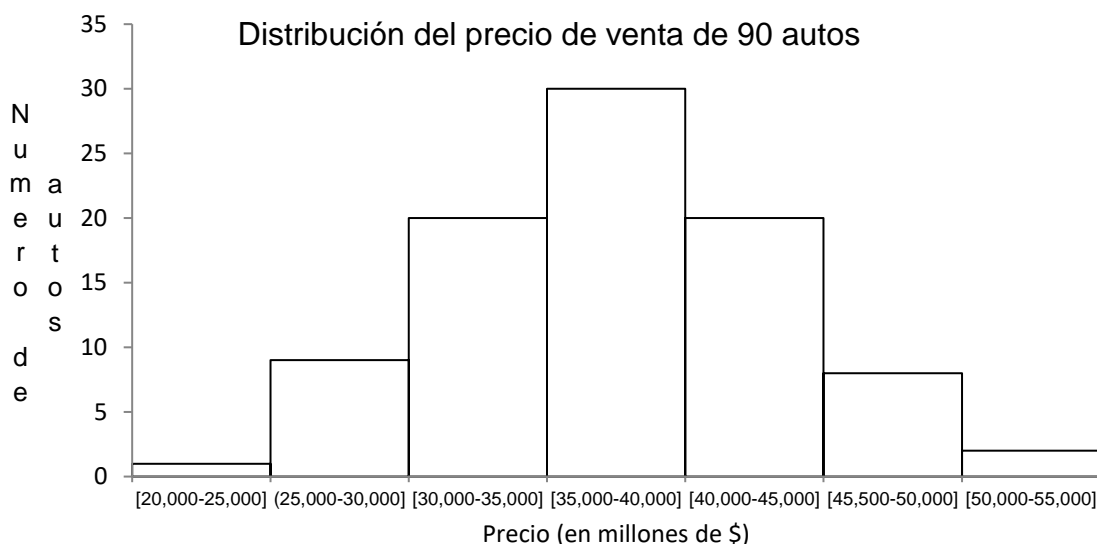
## 1.7 Taller

1. Los consumos de 10 estudiantes en una cafetería de la universidad fueron los siguientes, en miles de pesos:

\$3.5, 4.0, 5.0, 3.6, 4.2, 4.5, 5.0, 4.3, 3.3, 3.0

- a. Determine el valor promedio del consumo de los 10 estudiantes.
- b. Compruebe que  $\sum_{i=1}^{10} (x_i - \bar{x}) = 0$ . Generalice una regla para  $n$  datos. De un nombre a esta regla.
- c. (tarea) Por política del director de la cafetería, en cada compra superior a \$ 2000 deben cobrar \$500 adicionales por persona para donar a los damnificados por el fenómeno del niño. Calcule el valor promedio del gasto de los 10 estudiantes teniendo en cuenta este impuesto. Generalice y de un nombre a esta propiedad.

- d. (tarea) 16% es el impuesto del IVA para comestibles, ¿Cuánto es el valor real promedio (después de descontar impuestos) que ingresan a la cafetería? Generalice y de un nombre a esta propiedad.
2. (tarea) Un concesionario local de autos acaba de publicar la distribución de los precios de venta de 90 carros en el último mes. Ver gráfica adjunta. Determine:
- La tabla de frecuencias completa
  - El precio promedio de venta de los 90 carros.
  - Sugiera un procedimiento (analítico o gráfico) para determinar la mediana de los precios de venta.
  - ¿Cuál es la tendencia en los precios de venta en este concesionario?



3. Un Administrador de un centro de computación, interesado en optimizar el desempeño del sistema que usa recolectó datos sobre el tiempo, en segundos, entre solicitudes de un servicio de procesamiento en particular y obtuvo los siguientes valores:

2.8, 4.2, 3.5, 9.1, 2.1, 5.9, 1.6, 6.7, 21.7, 3.1, 2.9, 11.8, 4.7, 14.2, 1.6, 9.9, 20.8,  
 6.6, 1.8, 7.0, 1.9, 4.2, 6.6, 15.5, 4.7, 8.5, 10.7, 20.1, 16.7, 2.6, 3.6, 2.5, 4.2, 4.0,  
 2.4, 1.0, 2.3, 14.0, 15.2, 1.6, 5.7, 5.4, 2.4, 2.4, 0.9, 3.1, 19.5, 11.4, 8.8, 3.4.

- Construya la tabla de frecuencias
- Construya el histograma
- Identifique la forma y la tendencia de los datos
- ¿Cuál, entre la media, la mediana y la moda representa mejor la tendencia del tiempo de desempeño del sistema? ¿Qué valor tiene esta medida?

## 1.8 Indicadores de variación

Ya se sabe medir la tendencia de un conjunto de datos, pero ¿será suficiente con estos indicadores para resumir el comportamiento global del proceso? No, hay que determinar otros indicadores que tengan en cuenta la variabilidad de los datos. Se tienen dos tipos de indicadores de variación, unos cuantifican la variación según la distancia entre los datos y otros miden la variación alrededor de la media.

**Ejercicio 13:** Investigue las definiciones de Cuartiles y las respectivas fórmulas para obtener sus valores.



**Ayuda:** Use la fórmula:  $L_p = (n + 1) \frac{p}{100}$ , donde  $p$  representa el porcentaje a localizar, es decir,  $p = 25\%$  si es el primer cuartil y  $p = 75\%$  si es el tercer cuartil.

**Ejercicio 14:** Veinticinco empleados públicos fueron evaluados en su desempeño en una escala de 20 a 90 con los siguientes resultados:

24, 34, 40, 50, 63, 39, 88, 49, 34, 40, 53, 66, 74, 24, 36, 44, 55, 38, 45, 58, 39, 48, 49, 70, 25.

Obtenga los cuartiles.

Rtas.  $Q_1 = X_7 = 38$ ,  $Q_2 = \text{Mediana} = X_{13} = 45$  y  $Q_3 = X_{20} = 58$

**Ejercicio 15:** Investigue el diagrama de Caja y extensión, sus usos y su construcción.

**Ejercicio 16:** Investigue las definiciones de varianza y sus fórmulas.

**Ejercicio 17:** Investigue las definiciones de desviación estándar y sus fórmulas.

### Ejercicio 18: El proceso de empaclado

En un proceso de empaclado de 500 gramos, se toma una muestra de 10 productos para determinar si el proceso está empaclando correctamente. Un registro de los pesos de 10 productos mostró los siguientes resultados: 501, 500, 499, 500, 499, 501, 501, 500, 498, 500. Es evidente que no todos los productos pesan igual, surgen entonces las preguntas: ¿el proceso de empaclado está fallando? ¿Se debería parar el proceso? Realice las actividades planteadas a continuación para tomar una decisión.

1. ¿Qué tan conveniente es utilizar los indicadores de dispersión definidos en la sección 5.2 para este caso?
2. Complete la siguiente tabla:

$x_i$	$x_i - \bar{x}$	$ x_i - \bar{x} $	$(x_i - \bar{x})^2$
501			
500			
499			
500			
499			
501			
501			
500			
498			
500			
Total:			

3. Calcule la varianza y la desviación estándar de los pesos del producto empaclado.
4. Tome una decisión sobre si el proceso de empaclado está funcionando bien o no.
5. Comente sobre las ventajas y desventajas del uso de los indicadores de variación, es decir, determine cuando se pueden usar y cuando no.

En el análisis de la forma de poblaciones es muy recurrente la forma simétrica, una regla universalmente aceptada para este tipo de poblaciones es la Regla empírica que describe en forma porcentual la ubicación de los datos alrededor de la media.

**Ejercicio 19:** Investigue la Regla empírica y aplíquela al siguiente ejercicio.

**Ejercicio 20:** El consumo diario (en pesos) de un refrigerador industrial de 400 Watios en una empresa comercial sigue una forma de campana con una media de \$ 2,500 y una desviación estándar de \$ 250. Establezca los intervalos donde estarán el 68.2%, el 95.5% y el 99.7% de los consumos de electricidad.

## 1.9 Taller

- (Ejercicio 1.14) Un fabricante de neumáticos quiere determinar el diámetro interior de un neumático de cierto grado de calidad. Idealmente el diámetro sería de 570 mm. Los datos son los siguientes:

572    572    573    568    569    575    565    570

- Calcule la media y la mediana de la muestra
  - Obtenga la varianza, la desviación estándar y el rango de la muestra.
  - Con base en los resultados en a) y b) ¿Qué comentaría acerca de la calidad de los neumáticos?
- Dos entidades financieras han ofrecido tasas de rendimiento para inversión en los últimos doce meses con los siguientes valores:

Entidad 1: 19, 20, 20, 21, 21, 21, 21, 22, 22, 23, 24, 21%

Entidad 2: 11, 16, 16, 19, 20, 21, 21, 24, 25, 29, 25, 28%

¿Cuál de las dos entidades financieras debiera elegirse para una próxima inversión si se quiere asumir un riesgo mayor en la inversión?

- La cotización en la bolsa de valores de New York del petróleo crudo (dólares por barril) y del Oro (Pesos colombianos por gramo fino) mostro los siguientes valores mensuales en promedio para el año 2015<sup>4</sup>:

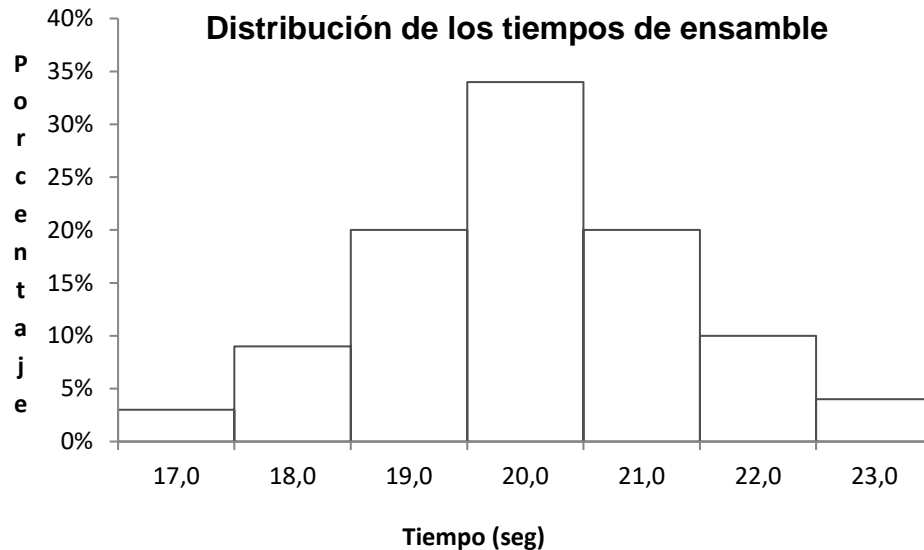
Mes	Nov.	Octubre	Sept.	Agosto	Julio	Junio	Mayo	Abril	Marzo	Febrero	Enero
Petróleo	42,91	46,37	45,50	43,13	51,84	59,76	59,45	53,75	47,83	50,81	48,14
Oro	96503	100306	102375	99480	91383	89343	86426	88710	90191	88097	88155

- Comente sobre la posibilidad de usar la desviación estándar para comparar la variación en los precios de los dos productos.
  - Proponga otro indicador de variación usando la desviación estándar y la media aritmética. De un nombre a este indicador.
- Demuestre que la varianza de una constante es cero (**Ayuda:** use la fórmula de varianza remplace para  $n$  valores constantes:  $k, k, k, \dots, k$  con media  $\bar{x} = k$ )
  - Demuestre que la varianza es positiva o igual a cero
  - Si a cada dato de una muestra se le suma o resta un valor constante, la varianza de la muestra no cambia. **Ayuda:** use para  $x_1, x_2, \dots, x_n$  y un valor constante  $k$ ,  $y_1 = x_1 + k$ ,  $y_2 = x_2 + k$ ,  $\dots$ ,  $y_n = x_n + k$ , y reemplace en la formula de varianza.
  - Si a cada uno de los valores de una muestra se le multiplica por una constante  $k$ , la varianza de la muestra queda multiplicada por  $k^2$ .
  - (Datos del ejercicio 1.22) Los siguientes datos son las mediciones del diámetro de 36 cabezas de remache en centésimos de una pulgada.

6.72    6.77    6.82    6.70    6.78    6.70    6.62    6.75    6.66    6.66    6.64    6.76    6.73    6.80  
6.72    6.76    6.76    6.68    6.66    6.62    6.72    6.76    6.70    6.78    6.76    6.67    6.70    6.72  
6.74    6.81    6.79    6.78    6.66    6.76    6.76    6.72

<sup>4</sup> Datos tomados de la base de datos del Banco de la Republica de Colombia.

- Construya graficas de tallo y hoja, histograma y diagrama de caja. ¿Cuál de las tres opciones refleja el comportamiento global de la muestra?
  - Calcule los indicadores de tendencia central. ¿Qué indicador marca la tendencia central de los datos?
  - Calcule los indicadores de variación, comente sobre la variación de los datos.
9. Considere los siguientes datos sobre el tiempo (en segundos) de ejecución de una tarea rutinaria en una línea de ensamble de motos realizada en una muestra de 200 chasises (media aritmética o promedio: 19.8833, desviación estándar: 1.5243):



- Sobre la gráfica ubique los siguientes rangos  $\bar{x} \pm 1s$ ,  $\bar{x} \pm 2s$  y  $\bar{x} \pm 3s$ . Estime los porcentajes de datos para cada rango.

Respuestas:  $19.8833 \pm 1.5243$ : [18.359; 21.4076]: 68.3%

$19.8833 \pm 2 * 1.5243$ : [ 16.8347; 22.9319]: 95.5%

- Calcule  $Z = \frac{X - \bar{x}}{s}$ , para cada  $X$ , donde  $X$  son los extremos de cada rango, ¿Cuáles son los valores máximo y mínimos de  $Z$ ?

Respuesta:  $Z = \frac{23.5 - 19.8833}{1.5243} = 2.37$      $Z = \frac{16.5 - 19.8833}{1.5243} = -2.21$

- Si  $x = 25$ , cuánto vale  $Z$ ? ¿es viable este valor
- Establezca una regla para determinar datos atípicos usando los conceptos de simetría y el valor  $Z$ .