



# **PROBABILIDAD Y ESTADÍSTICA**

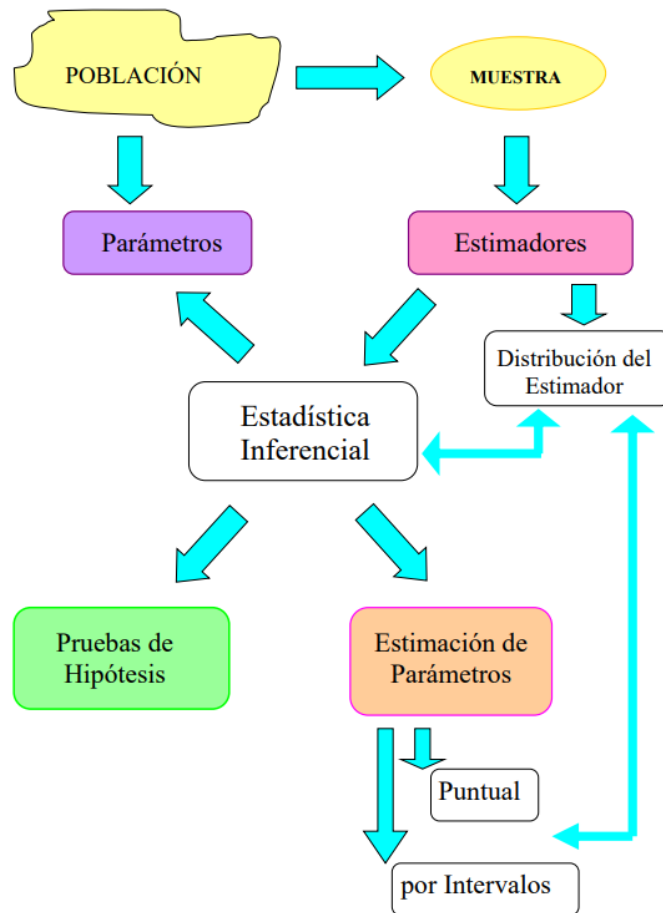
## **GUÍA DE ACTIVIDADES PRÁCTICAS**

Parte 2  
Unidades 5, 6 y 7

2023

## Unidad 5: Distribucion en el muestre y Estimacion de parametros

Esquema Conceptual.

**Objetivos.**

- Distinguir los conceptos de parámetro y estimador.
- Interpretar las diferencias entre parámetros y estimadores. Definir las propiedades de un estimador.
- Distinguir los diferentes tipos de estimadores.
- Estimacion por intervalos

**Contenidos.**

- Conceptos de parámetro y estimador.
- Distribución de la media muestral
- Distribucion de la varianza muestral
- Estimación puntual de parámetros. Propiedades de un buen estimador.
- Estimación por intervalos de confianza

## CUESTIONARIO TEÓRICO

1. ¿Qué es un parámetro y qué es un estimador? ¿Cuál es la diferencia entre ambos conceptos? Explicar cuál es la razón por la que se dice que los estimadores son variables aleatorias.
2. Suponga que estudiamos la variable continua: “Tiempo de reparación de una máquina, ante un cierto tipo de fallas (en minutos)”. Suponga, además, que la verdadera media poblacional es  $\mu = 135\text{min}$ . Se toma una muestra de veinte observaciones de esa variable. Analice si es posible que el promedio de la muestra sea exactamente igual a 135min. Justifique conceptualmente la respuesta.
3. ¿Cómo es la distribución de la media aritmética de muestras extraídas de una población Normal de parámetros conocidos? ¿Qué pasa si la variable no tiene distribución Normal, o su distribución es desconocida?
4. Representar gráficamente una variable con distribución Normal con media 200 y con desvío 20. Superpuesta con la figura anterior, graficar la distribución de la media aritmética de muestras de tamaño  $n=4, 25$  y 100, de esa población.
5. El promedio de una muestra se considera un buen estimador de la media de la población. Explicar los motivos de la afirmación anterior.
6. Analizando el comportamiento de los estimadores, se acostumbra calificarlos como: Insesgados; Consistentes; Eficientes y Suficientes. Explicar bajo qué condiciones se otorga cada una de esas calificaciones. Indicar en particular cuáles son las medidas muestrales que pueden utilizarse como estimadores puntuales insesgados de los parámetros  $\mu$  y  $\sigma^2$ .
7. Construya una lista de estimadores convenientes para los parámetros de los siguientes modelos: Binomial; Poisson; Normal; Exponencial.
8. Explique las principales características de la distribución Chi-cuadrado.
9. Explique las principales características de la distribución t de Student.
10. Los parámetros de una población pueden estimarse en forma puntual y por intervalo. Explicar en qué consiste cada una y cuáles son las ventajas de la estimación por intervalos de confianza.
11. Realice la deducción de los siguientes intervalos: o Intervalos de confianza para la media de una Normal con varianza conocida. o Intervalos de confianza para la media de una Normal con varianza no conocida. o Intervalos de confianza para la varianza de una Normal. Aplique para ello el método del pivote.
12. Cuando se hace una estimación por intervalo, la aproximación es más precisa cuanto menor es la longitud. ¿De qué depende la longitud de un intervalo de confianza? ¿Cómo se puede hacer para disminuir dicha longitud, conservando el nivel de confianza?
13. El Nivel de Confianza es una decisión importante a la hora de realizar una estimación por intervalos. Explique qué es lo que representa el Nivel de Confianza.
14. Se considera que el tiempo de duración de cierta batería se distribuye con media  $\mu$  y varianza 1. Se toma una muestra aleatoria de tamaño 7 con cuya información se desea realizar una estimación puntual de la duración promedio poblacional. Con ese fin se sugieren los siguientes dos estimadores:

$$\mu_1 = \frac{X_1 + X_2 + \dots + X_7}{7} \quad \mu_2 = \frac{2X_1 - X_6 + X_4}{2}$$

¿Cuál estimador es el “mejor”? ¿En qué sentido es mejor?

### EJERCICIOS SOBRE DISTRIBUCION DE LOS ESTADÍSTICOS MUESTRALES

1. Una compañía de electrónica fabrica resistores que tienen una resistencia promedio de  $100\ \Omega$  y una desviación estándar de  $10\ \Omega$ . Si se sabe que la distribución de la resistencia es Normal:
  - a. ¿Cuál es la probabilidad de que al tomar una muestra de 49 resistores, la resistencia promedio sea menor de  $97\ \Omega$ ?
  - b. Si en lugar de tomar una muestra de 49 se considera una de 10 resistores ¿cuál será el valor de la probabilidad antes mencionada?
  - c. ¿Es importante la hipótesis de normalidad de la variable en ambos casos? ¿por qué?
2. Calcule la probabilidad de que una muestra aleatoria de 25 observaciones, de una población normal con varianza  $\sigma^2 = 6$ , tenga una varianza muestral  $S^2$ 
  - a. mayor que 9.1
  - b. entre 3.462 y 10.745.
3. La concentración de un ingrediente activo en el producto de una reacción química es fuertemente influida por el catalizador que se usa en la reacción. Se considera que cuando se utiliza el catalizador A la concentración media de la población excede es de 60ppm. Se sabe que la desviación estándar es  $\sigma = 0,5\text{ppm}$ . Determine la probabilidad que la media muestral calculada para 30 reacciones independientes sea:
  - a)  $\bar{x} > 58\text{ ppm}$
  - b)  $\bar{x} < 55\text{ ppm}$
  - c)  $59 < \bar{x} < 59,9\text{ ppm}$
4. La resistencia a la rotura X de cierto remache que se utiliza en el motor de una máquina tiene una media de 5000 psi y varianza poblacional desconocida. Se toma una muestra aleatoria de 36 remaches donde se calculó un desvío muestral  $s = 400\text{psi}$ .
  - a) ¿Cuál es la probabilidad de que la media de la muestra caiga entre 4800 psi y 5200 psi?
  - b) ¿Cuál es la probabilidad que la media muestral este entre 4900 y 5100 psi?
  - c) Cuáles son los valores de media muestral simétricos respecto de  $\mu$  que encierran entre si una probabilidad del 95%
5. Las densidades de la celulosa tienen una distribución normal de  $\mu = 1,5\text{ g/cm}^3$  y varianza desconocida. Si se toma una muestra de tamaño 16 para este polímero, cuya desviación estándar muestral  $S = 0,1$ :
  - a) Con que probabilidad la media muestral es de  $0,9\text{ g/cm}^3$
  - b) Cuáles son los valores de media muestral simétricos con respecto a  $\mu$  que encierran entre si una probabilidad del 70%
6. La densidad de la quitina (Polímero natural) tienen distribución normal de media  $\mu = 1,44\text{ g/ml}$  con una desvío poblacional  $\sigma = 0,098\text{ g/ml}$ . Si se toma una muestra de tamaño 16, calcular la probabilidad de que la varianza muestral sea:
  - a) Menor a 0,08
  - b) Un valor comprendido entre 0,07 y 0,01
  - c) Mayor a 0,09

**EJERCICIOS DE ESTIMACIÓN POR INTERVALOS DE PARAMETROS****Estimación de  $\mu$  con  $\sigma^2$  conocido**

1. Una máquina llena bolsas de cemento. Se estudia la variable X: contenido de una bolsa. Si todo funciona normalmente, esta variable debe tener distribución Normal, con un desvío de 0.5 Kg. se extrae una muestra de 10 bolsas, que tienen los siguientes pesos (en Kg.):

49,67 49,79 49,79 51,25 49,18 50,36 48,75 50,12 50,24 50,40

- Realizar una estimación puntual para la varianza poblacional
  - Realizar estimaciones por intervalos para los siguientes niveles de confianza: 90%, 95% y 99%
  - Interprete los resultados
  - Interprete el significado del nivel de confianza
2. Especialistas en Geología han determinado, gracias a muchos estudios, que la densidad del oro puede ser modelada con una distribución Normal, con media 19,3 gramos por centímetro cúbico y con desvío 0,3 gramos por centímetro cúbico. Por otra parte, para evaluar la densidad de cualquier material, se realizan varias mediciones con un equipo denominado espectrómetro y se determina el promedio de las mismas. Suponiendo que se coloca en el espectrómetro un trozo de mineral y después de realizar 20 mediciones se determina que el promedio de las mismas es 19,05 gr./cm<sup>3</sup>. Asumiendo que la dispersión no varía, estimar un intervalo del 90% de confianza para la verdadera densidad media poblacional del metal analizado. En base al resultado obtenido, ¿es posible suponer que la muestra es de oro puro?
  3. Con relación al ejercicio anterior, determinar el tamaño que debería tener la muestra para duplicar la precisión de intervalo obtenido (es decir que la longitud del intervalo de confianza se reduzca a la mitad)
  4. Una empresa de material eléctrico fabrica bombillas que tienen una duración distribuida de forma normal, con una desviación estándar de 40 horas. Si una muestra de 30 bombillas tiene una duración promedio de 780 horas, calcule un intervalo de confianza del 95% para la media de la población de todas las bombillas producidas por esta empresa.
  5. Las resistencias a la compresión de un nuevo material de aleación tienen distribución normal de media  $\mu$  y  $\sigma=6.9$  KG/cm<sup>2</sup>. Se toma una muestra aleatoria de 50 barras de esta aleación, la misma tienen una media muestral de 174.5 KG/cm<sup>2</sup>
    - a) Construya un intervalo de confianza del 90% para la media poblacional de la resistencia
    - b) ¿Cómo se interpreta esta confianza del 90%?
    - c) ¿Cuál debería ser el tamaño de la muestra si queremos duplicar la precisión en la estimación? (es decir que la longitud del intervalo de confianza se reduzca a la mitad)

**Estimación de  $\mu$  con  $\sigma^2$  desconocido**

1. Una máquina produce piezas metálicas de forma cilíndrica cuyo diámetro tienen distribución normal de parámetros  $\mu$  y  $\sigma^2$ , ambos desconocidos. Se toma una muestra de las piezas y los diámetros en centímetros son:

1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03.

- a) Calcule un intervalo de confianza del 90% y 95% para la media del diámetro de las piezas que se manufacturan con esta máquina.
  - b) Realice una interpretación de estos resultados
  - c) ¿Qué tamaño de muestra será necesario para reducir a la mitad la longitud de los intervalos calculados?
  - d) ¿Cómo se puede mejorar la precisión de las estimaciones sin modificar los niveles de confianza utilizados?
2. La distribución de la variable aleatoria “Calorías liberadas” para un reactivo tienen una distribución normal de parámetros  $\mu$  y  $\sigma^2$  desconocidos. Una muestra aleatoria de tamaño 10 de este reactivo libero, en promedio, 230 calorías con una desviación estándar de 15 calorías cuando se redujo a ceniza.
- a) Construya un intervalo de confianza del 90% y del 99% para el contenido medio poblacional de calorías.
  - b) Analice la precisión para cada uno de los intervalos calculados
3. Al Gerente del Departamento de Servicios al Cliente de una compañía de gas le gustaría estimar el tiempo promedio que transcurre entre la solicitud de servicio y su conexión. Se seleccionó una muestra aleatoria de 6 casas de los registros disponibles del año anterior. Los resultados obtenidos (horas) se presentan de la manera siguiente:

349,05; 365,39; 357,55; 359,37; 355,03; 358,30

- a) Identificar con claridad la variable aleatoria estudiada.
  - b) Obtener estimaciones puntuales del promedio dicha variable
  - c) Realice una estimación para  $\mu$  con una confianza del 95%
4. La resistencia a la rotura de los cables producidos por una fabricante tiene una distribución normal de parámetros media de 180 Kg. y un desvío estándar de 10 Kg. Mediante una nueva técnica en el proceso de fabricación se aspira a que esta resistencia pueda ser incrementada. Se ensaya una muestra de 20 unidades producidas con el nuevo método y se encuentra que su resistencia media es de 185 Kg. y el desvío 12 Kg. ¿Puede sostenerse que, en efecto, existe un aumento significativo de la resistencia de los cables? Para dar respuesta a esta pregunta realice una estimación por intervalos para la media poblacional del nuevo metodo y analice los resultados. Utilice una confianza del 95%

#### Estimacion por intervalos para la varianza poblacional $\sigma^2$

1. Retomando el ejercicio 1 del apartado anterior. Una máquina produce piezas metálicas de forma cilíndrica cuyo diámetro tienen distribución normal de parámetros  $\mu$  y  $\sigma^2$ , ambos desconocidos. Se toma una muestra de las piezas y los diámetros en centímetros son:

1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03

- a) Realizar una estimación por intervalo utilizando un nivel de confianza del 90, 95 y 99% para la varianza poblacional.
  - b) Analice cuál de estos intervalos tienen una mayor precisión en la estimación.
2. Retomando el ejercicio 3 de la sección anterior: Al Gerente del Departamento de Servicios al Cliente de una compañía de gas le gustaría estimar el tiempo promedio que transcurre entre la solicitud de servicio y su conexión. Se seleccionó una muestra aleatoria de 6 casas de los registros disponibles del año anterior. Los resultados obtenidos (horas) se presentan de la manera siguiente:

349,05; 365,39; 357,55; 359,37; 355,03; 358,30

Realizar una estimación por intervalo utilizando un nivel de confianza del 90, 95 y 99% para la varianza poblacional.

3. Retomando el ejercicio 4 de la sección anterior: La resistencia a la rotura de los cables producidos por una fabricante tiene una distribución normal de parámetros media de 180 Kg. y un desvío estándar de 10 Kg. Mediante una nueva técnica en el proceso de fabricación se aspira a que esta resistencia pueda ser incrementada. Se ensaya una muestra de 20 unidades producidas con el nuevo método y se encuentra que su resistencia media es de 185 Kg. y el desvío 12 Kg. ¿Puede sostenerse que, la varianza del nuevo método se ha modificado? Para dar respuesta a esta pregunta realice una estimación por intervalos para la varianza poblacional del nuevo método y analice los resultados. Utilice una confianza del 95%

### Estimación por intervalos para una diferencia de medias poblacionales

1. El encargado del mantenimiento de una máquina compara el tiempo de reparación (en hrs.) de la misma mediante el procedimiento convencional de trabajo, con un nuevo método que resulta más costoso pero que tiende a reducir los tiempos de reparación. Para ello saca una muestra de 10 registros con el procedimiento actual, y otra de 15 valores con el nuevo método:

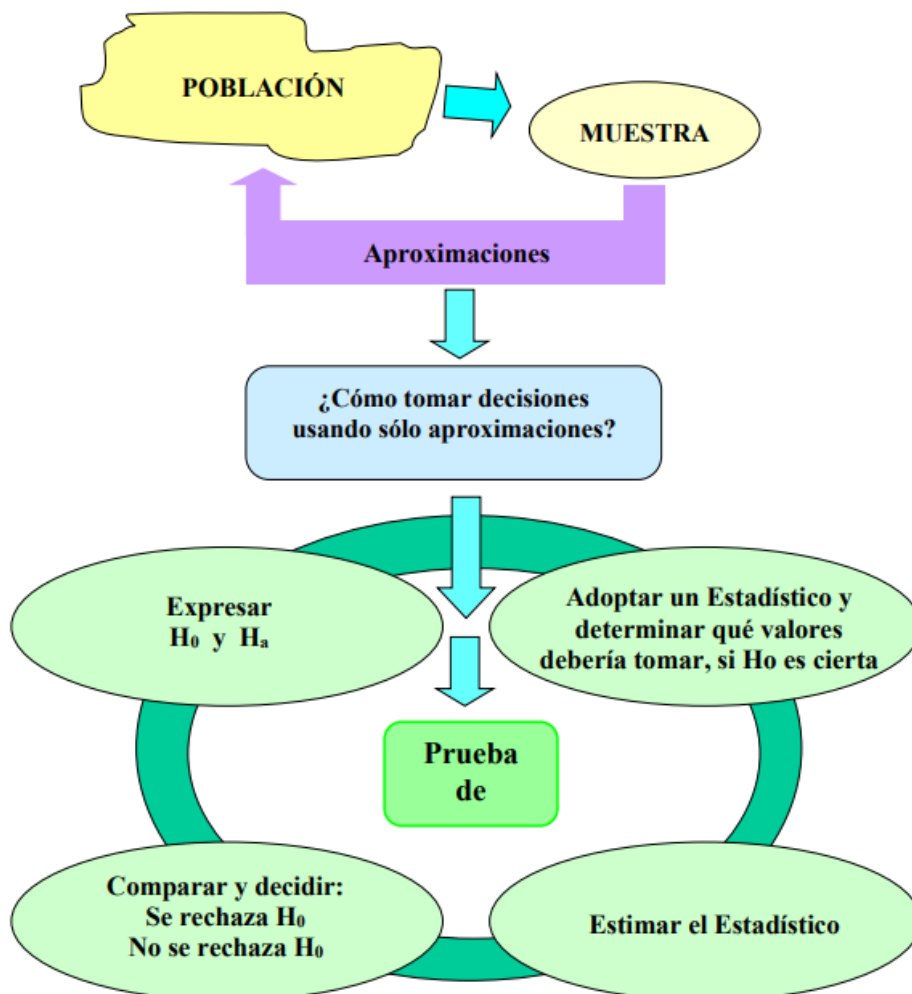
Viejo 28 25 25 29 24 29 31 28 31 27

Nuevo 26 25 24 24 25 22 27 26 27 27 21 23 30 26 25

- a) Determinar si puede suponerse que el nuevo procedimiento reduce significativamente el tiempo medio de reparación de la máquina.

## Unidad 5: Prueba de hipótesis

## Esquema Conceptual.



## Objetivos

- Describir el procedimiento de una prueba de hipótesis estadística
- Interpretar el fundamento metodológico de las pruebas de hipótesis
- Reconocer los errores que se pueden cometer en una prueba
- Manejar los diferentes tipos de pruebas
- Seleccionar la prueba adecuada para cada problema particular
- Inferir acerca del comportamiento de una y dos poblaciones.

## Contenidos

- Esquema general de una prueba de hipótesis
- Errores en una prueba
- Pruebas sobre los parámetros de una y dos poblaciones



## CUESTIONARIO TEÓRICO

1. Imagine a las pruebas de hipótesis como un proceso donde se desarrolla una secuencia de actividades. Elabore un diagrama de flujo que represente estas actividades y prevea las posibles respuestas
2. Todas las pruebas de hipótesis son muy similares en su estructura. Sin embargo, se diferencian entre sí en una cuestión fundamental. Especifique cuál es esa cuestión. Elabore ejemplos donde esa diferencia se ponga de manifiesto
3. Elabore un diagrama o representación gráfica que represente todas las pruebas de hipótesis que se analizan en la esta unidad temática
4. Las pruebas de hipótesis no proveen un resultado certero, sino que por el contrario pueden presentar errores. Especifique cuáles son los posibles errores y determine las probabilidades de cada uno. Además, analice cuáles son las condiciones que determinan dichas probabilidades y el modo en que se pueden calcular.

## EJERCICIOS PRACTICOS

### Prueba de hipotesis para $\mu$ con $\sigma^2$ conocido

1. Una máquina tiene una cierta producción media horaria, digamos  $\mu_0$ . Se está pensando en remplazar la máquina por una nueva. Explicar cómo formularía la Hipótesis Alternativa en cada uno de los siguientes casos:
  - a) Se hará el reemplazo siempre y cuando la nueva máquina no tenga una producción significativamente diferente de la actual
  - b) Se incorporará la nueva máquina solo si su producción horaria es significativamente mayor que la actual
  - c) Se hará el cambio siempre y cuando la nueva máquina no tenga una producción significativamente menor que la actual.
2. La variable aleatoria concentración de impurezas en la aleación de cierto metales realizada con una metodología estándar tienen distribución normal de media 50 ng y  $\sigma = 0,5$  ng (Metodo A). Se está estudiando un nuevo metodo (B) para estas aleaciones y se presentan las siguientes situaciones:
  - a) Se saca una muestra de 30 planchas de esta aleación realizadas con la metodología B, que arroja una media de 50,21 ng . Comprobar si la media del proceso es significativamente diferente a los 50 ng. para un nivel de significación del 1 %. Se verifico que el desvío se mantuvo sin variar.
3. Volviendo al ejemplo planteado en la Unidad anterior sobre la densidad del oro, que según estudios realizados puede ser modelada con una distribución Normal, con media 19,3 gramos por centímetro cúbico y con desvío 0,3 gramos por centímetro cúbico. El análisis de un trozo de mineral con espectrómetro había arrojado después de realizar 20 mediciones un promedio de 19,05 g/cm<sup>3</sup> .
  - a) Suponiendo que la dispersión no cambia, comprobar considerando un nivel de significación del 5%, si el material analizado puede considerarse de oro puro.
  - b) ¿Coincide la conclusión de esta prueba con la obtenida mediante intervalo de confianza calculado en la Unidad anterior?

### Prueba de hipotesis para $\mu$ (con $\sigma^2$ desconocido) y prueba de hipotesis para $\sigma^2$

1. Una empresa automotriz se dedica a producir unidades de gran confort y prestaciones. Uno de los aspectos que diferencian sus productos respecto a la competencia es la seguridad. Para verificar su sistema de frenos, se realiza el siguiente ensayo: con el vehículo a 60 Km./h., se activan los frenos y se mide la

distancia hasta que el vehículo se detiene. Con el sistema actual la variable distancia de frenado tiene distribución Normal con media 16,5 m. y desvío 0,5 m. Para mejorar esta prestación, se prueba un nuevo sistema con veinticinco autos, obteniéndose un promedio de 15,8 m. con un desvío de 0,55 m. Dado que el cambio de sistema requiere una fuerte inversión para el desarrollo de proveedores, se desea cambiar el sistema sólo si se encuentran evidencias de que la mejora es significativa. Para resolver el problema:

- a. Analizar si la dispersión realmente se incrementó, utilizando una prueba de hipótesis con un nivel de significancia de 0,10.
  - b. Decidir si puede suponerse que la distancia media de frenado se ha reducido, utilizando nuevamente un nivel de significancia de 0,10.
  - c. Realice un análisis de los errores de tipo 1 y 2 en este contexto.
2. Las autoridades de una región han establecido que el vuelco de agua con temperaturas promedio no superiores a los 56,5°C no afecta al ecosistema fluvial. Una planta industrial que está al lado del río posee un sistema de enfriamiento por agua y la descarga se realiza mediante una cañería de corto recorrido que desemboca en el río. Para verificar si el volcamiento de agua cumple con lo requerido, se mide la temperatura de 25 muestras de agua a la salida de la descarga en el río, obteniéndose los siguientes valores muestrales: Promedio = 57,5°C y Desvío = 2,5°C. verificar:
- a) si se cumple o no con lo requerido por las reglamentaciones. Analizar en este contexto los errores de tipo I y de tipo II.
  - b) Si el desvío estándar para la temperatura del agua de enfriamiento para esta planta industrial se puede considerar significativamente inferior a 3 °C
3. Recordemos que en la Unidad anterior la Gerencia de una compañía de gas quería estimar el tiempo promedio que transcurre entre la solicitud de servicio y su conexión (en hs), para lo cual se había extraído la siguiente muestra:

349,05; 365,39; 357,55; 359,37; 355,03; 358,30

- a. La dirección de la compañía había fijado como meta no superar significativamente las 360 hs para realizar la conexión con un desvío estándar inferior a 23,98 hs. Compruebe mediante una prueba de hipótesis apropiada si se cumplen esos objetivos.
  - b. En todos los casos, realice una interpretación de los errores de tipo 1 y 2 en el contexto del problema
4. La empresa “Recorridos S.A” realiza trayectos entre la ciudad de Córdoba y la localidad de Luque. Hace unos meses se han recibido una proporción considerable de reclamos por parte de los pasajeros respecto de incumplimientos en los horarios pautados por la misma. La empresa afirma que el tiempo promedio en que realiza el recorrido es de 100 minutos con un desvío de 5 minutos. En tanto los pasajeros expresan: ¡Nunca se sabe a la hora que pasan, a veces tardan más de dos horas! Atendiendo a esta situación la empresa decide realizar una investigación y obtiene la siguiente información referida a los tiempos de viaje entre Córdoba y Luque durante once días considerados al azar:

119	131	106	120	116	126	122	105	115	120	112
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Mediante las pruebas de hipótesis correspondientes y utilizando un nivel de significancia del 1% verifique si la empresa está en lo correcto al hacer sus afirmaciones.

### Prueba de hipótesis para la igualdad de medias y varianzas poblacionales

1. En una operación de mecanizado se realiza un agujero. El diámetro de dicho agujero es una característica de calidad clave, porque si falla el producto no se puede utilizar. Ahora bien, se pueden utilizar dos mechas denominadas A y B respectivamente. El problema es decidir cuál es la mecha más conveniente. Para ello se toman las siguientes muestras y se obtienen los siguientes resultados:

Mecha A	Mecha B
4,84	3,08
6,97	3,25
6,37	3,17
6,30	1,65
4,15	4,39
7,86	1,10
5,29	6,17
4,09	4,39
8,25	3,38
5,64	4,34

- a) Determinar si las varianzas y las medias son similares en ambos casos. Realice la prueba en infostat con un nivel de significancia del 1%
  - b) En base a los resultados anteriores, teniendo en cuenta además que el valor especificado para el diámetro en cuestión es de 6 mm. y que la mecha B es más barata, decidir cuál es la más conveniente.
2. Retomando el ejemplo considerado en la unidad anterior, sobre el mantenimiento de una máquina donde se comparaba el tiempo de reparación de la misma mediante dos procedimientos y recordando que los valores muestreados fueron:

Viejo 28 25 25 29 24 29 31 28 31 27

Nuevo 26 25 24 24 25 22 27 26 27 27 21 23 30 26 25

- a) Empleando la prueba apropiada, determinar si puede considerarse que los dos procedimientos de reparación tienen una dispersión similar. Realice los cálculos en infostat con un nivel de significancia del 1%
- b) Determinar si el tiempo de reparación con el nuevo procedimiento es significativamente menor que con el viejo. Realice los cálculos en infostat con un nivel de significancia del 1%
- c) En base al análisis realizado ¿conviene introducir el nuevo método de reparación?

### Datos emparejados

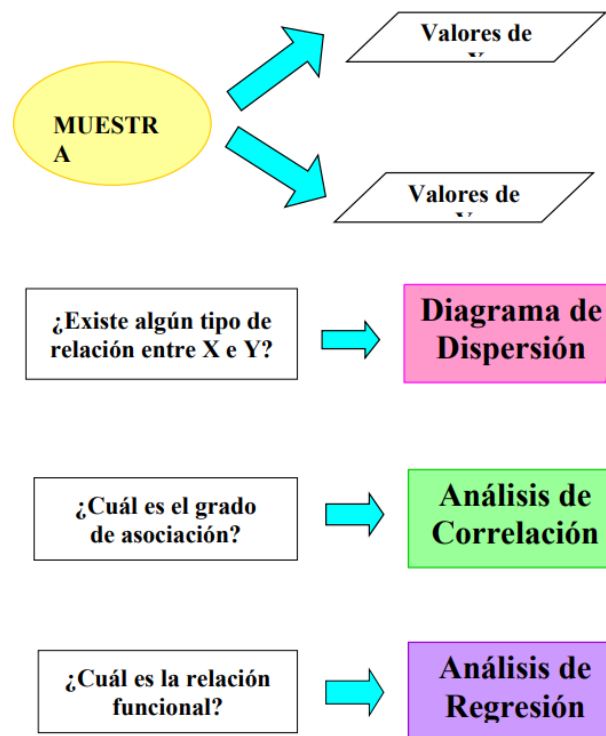
Se están evaluando dos metodologías (A y B) de enseñanza para la instrucción de los operarios en el manejo de máquinas compleja. Durante 3 meses una muestra de 10 operarios fue instruido con la metodología A y se midieron los tiempos (en horas) que estos tardaron en hacer un uso adecuado de una máquina compleja. Luego, estos mismos operarios fueron instruidos con el método B durante 3 meses y se midieron los tiempos que estos tardaron en hacer un uso adecuado de otra máquina compleja. Los resultados se muestran a continuación:

Metodo A	Metodo B
4,07	3,74
3,09	4,72
4,32	2,35
5,93	4,50
3,01	3,89
2,67	3,66
3,98	5,02
3,93	4,43
1,69	4,09
1,70	2,08

Determinar si existen diferencias significativas entre estos métodos de instrucción. Realizar el análisis en infostat con un nivel de significancia del 5%

## Unidad 7: Análisis de regresion y correlacion

Esquema Conceptual.



## Objetivos.

- Reconocer posibles relaciones entre variables.
- Medir el grado de asociación entre dos variables
- Identificar el tipo de vinculación entre dos variables
- Modelar la relación funcional que existe entre dos o más variables
- Evaluar la calidad del ajuste realizado.

## Contenidos

- Diagrama de dispersión
- Análisis de correlación
- Coeficiente de correlación
- Análisis de regresión. Ajuste por mínimos cuadrados
- Pruebas de hipótesis de la regresión

**Cuestionario teórico**

1. Mencione la principal diferencia entre un análisis de regresión y un análisis de correlación
2. ¿Cuáles son los estimadores puntuales de los coeficientes de regresión? ¿Cómo se obtienen estas estimaciones?
3. ¿Cuáles son las pruebas de hipótesis que se deben realizar en un análisis de correlación y en un análisis de regresión? ¿Cuál es el motivo por el cual se deben hacer dichas pruebas?
4. ¿Cómo se interpretan los coeficientes del modelo de regresión?
5. ¿Cómo se interpreta el coeficiente de determinación?
6. ¿Cuáles son los supuestos que se deben verificar en un análisis de regresión y de qué manera se hace?

**Ejercicios prácticos**

1. En una operación donde se aplica pintura esmalte con cargas electrostáticas, a puertas de automóviles, se miden las variables siguientes: X: Tiempo de aplicación de pintura con pistola pulverizadora, medido en segundos. Y: Espesor de la capa de pintura resultante (expresada en micrones). Los resultados obtenidos en este experimento, se presentan en la siguiente Tabla:

Tiempo	Espesor
14	55
8	38
16	50
8	46
14	47
10	49
16	58
12	52
10	41
12	44

- a) Construya un diagrama de dispersión con estos datos. ¿Qué nota en el gráfico?
  - b) Estime la pendiente, la ordenada al origen y la varianza residual, para un modelo de regresión lineal con ordenada al origen diferente de cero
  - c) Estime el espesor que se debe obtener, si se pulveriza pintura durante 11 segundos
  - d) Desarrolle una descomposición de la suma de cuadrados para este problema. Determine la suma de cuadrados total, la suma de cuadrados explicada y la suma de cuadrados de los errores. Determine los grados de libertad, de cada suma de cuadrados
  - e) Determine los coeficientes de correlación y de determinación. Interprete los resultados.
2. Se estudia la relación entre la superficie ( $m^2$ ) de una vivienda y el consumo mensual de kerosén empleado (Lts) para calefaccionar la misma, en un determinado barrio de la ciudad. Una muestra arrojó los siguientes resultados:

Sup	Consumo
69	20
135	40
121	35
103	35
97	26
78	24

- Identificar cuál es la variable dependiente en este estudio.
  - Dibujar el diagrama de dispersión y especificar si se observa algún tipo de relación entre las variables.
  - Calcular la covarianza y el coeficiente de correlación. Interpretar los valores obtenidos.
  - Ajustar un modelo lineal simple por mínimos cuadrados y dibujar el modelo obtenido en el diagrama de dispersión.
  - Calcular el coeficiente de determinación e interpretar el resultado obtenido.
  - En base a los resultados del análisis explicar si es razonable tratar de explicar el consumo de kerosén en ese barrio a partir de la superficie de las viviendas.
3. Se realiza un estudio similar al del caso anterior en otro barrio de la ciudad, de características socio-económicas diferentes. Una muestra de 10 viviendas produjo los siguientes resultados:

Superficie (m <sup>2</sup> .)	Consumo (lts.)
177	49
187	52
198	50
170	49
166	50
190	49
170	50
183	48
192	51
198	49

- Realizar el mismo análisis estadístico del caso anterior.
  - Comprobar si el coeficiente de correlación es significativo.
  - Comprobar si la pendiente del modelo es significativa.
  - ¿Resulta útil en este caso, el análisis conjunto de las variables?
4. Se almacenó maní tostado a temperatura ambiente durante 120 días. Cada 28 días se tomaron muestras y se determinaron dos variables químicas: Índice de Peróxido e Índice de p-Anisidina, para determinar su deterioro por oxidación. Los datos obtenidos fueron:

Tiempo (días)	Índice de Peróxido	Índice de p-Anisidina
0	1.30	0.10
28	11.27	1.01
56	24.08	2.54
84	36.74	3.26
112	50.30	4.15

- Identifique las variables dependientes e independientes.
  - Grafique los diagramas de dispersión para cada variable dependiente.
  - Determine si las variables dependientes se correlacionan significativamente. Interprete el resultado.
  - Ajuste un modelo lineal simple por mínimos cuadrados para cada variable dependiente en función de la variable independiente, calcule los respectivos coeficientes de determinación, e interprete los resultados obtenidos.
  - Considerando que el Código Alimentario Argentino fija para el Índice de Peróxido un valor máximo de 10, ¿Cuál sería la vida útil estimada (en días) de estas muestras de maní tostado, en estas condiciones de almacenaje?
5. En los últimos meses la empresa de transporte interurbano “Camino Cordobeses” ha recibido un número considerable de reclamos por parte de los pasajeros al respecto de incumplimientos en los horarios pautados de arribo a las distintas paradas uno de los tramos que realiza. Los pasajeros expresan que a veces se espera poco y otras veces mucho. Se decide entonces realizar una investigación que permita determinar si la percepción de los pasajeros es acertada y se enfoca la misma en el estudio de la posible relación de la variable Distancia (Km) entre paradas con el Tiempo de retraso (Min). Se obtiene la siguiente muestra aleatoria:

Distancia entre paradas (Km.)	10	15	20	25	30	35	40	45	50	55
Tiempo (min.)	2	3	6	8	9	14	17	19	20	22

A partir de la muestra indicada se obtiene la siguiente información del software Infostat. Observe que hay celdas incompletas.

Variable	N	R <sup>2</sup>
Tiempo (min)	10	¿?

Coeficientes de regresión y estadísticos asociados						
Coef	Est.	E.E.	LI(95%)	LS(95%)	T	p-valor
Const	-3,6	0,82	-5,49	-1,71	¿?	0,0023
Distancia (Km)	0,48	0,02	0,43	0,53	¿?	

- Complete los datos faltantes en la salida de Infostat expuesta. Justifique.
- Interprete los valores de los coeficientes de correlación y de determinación, obtenidos en este ajuste. Explique el concepto de los mismos.
- Realice el planteo de una prueba de hipótesis sobre la pendiente del modelo, Justifique el planteo y el resultado de la misma.
- Explique la evaluación que debe realizarse sobre los residuos para que el modelo sea aceptado y efectúe dicha evaluación.
- Elabore una conclusión respecto del análisis realizado.