ESTADISTICA GENERAL

- INFERENCIA ESTADISTICA
- Profesor: Celso Gonzales

Objetivos

- Entender los conceptos de estimación puntual y estimación por intervalos.
- Calcular e interpretar intervalos de confianza para una media poblacional con varianza conocida.
- Calcular e interpretar intervalos de confianza para una media poblacional con varianza desconocida.
- Calcular e interpretar intervalos de confianza para una proporción poblacional.
- Calcular e interpretar intervalos de confianza para una varianza poblacional .

INFERENCIA ESTADÍSTICA

Análisis, interpretación de resultados y conclusiones a partir de una muestra aleatoria

Estimación de Parámetros

Aproximación de los valores de los parámetros.

Estimador

Función de las observaciones muestrales

COMPRENDE:

- o Estimación de Parámetros
  - **Estimación Puntual**
  - **Estimación por intervalo**
- o **Prueba de hipótesis**

TIPOS DE ESTIMACION

- Estimación Puntual

Único valor

Propiedades

- **Insesgado**
- **Consistente**
- **Eficiente**
- **Suficiente**

ESTIMACIÓN DE PARÁMETROS.

- **Puntual.** $\hat{\theta}$, estimación puntual de θ, corresponde al valor del estimador Θ en $x_1, x_2, \ldots, x_n$.

$\bar{x}$, estimación puntual de μ, que corresponde al valor del estimador $\bar{X}$ en $x_1, x_2, \ldots, x_n$:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \sum_{i=1}^{n} \frac{x_i}{n}$$

PROPIEDADES DE LOS ESTIMADORES

Insesgado → $$E(\hat{\theta}) = \theta$$

Consistencia → $$\lim_{n\to\infty}(\hat{\theta}) = \theta$$

Eficiencia → $$Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$$

Suficiencia → $$\hat{\theta}_1 = \frac{X_1 + X_2 + .. + X_m}{m}$$

$$\hat{\theta}_2 = \frac{X_1 + X_2 + .. + X_n}{n}$$

$$n \succ m$$

TIPOS DE ESTIMACION

- Estimación por intervalo

Conjunto de valores contenidos en un intervalo

Tipos

- **Media**
- **Proporción**
- **Varianza, etc**

Estimación por intervalos.

- Consiste en la determinación de un intervalo, que contendrá el parámetro con una confianza **1- α** , número entre 0 y 1, fijado.

Se requiere:

- Una muestra aleatoria $\mathbf{X_1, X_2, \ldots, X_n}$ de tamaño n
- Un estimador **Θ** del parámetro poblacional **θ** , con distribución o función de probabilidad conocida.
- El nivel de confianza **1- α**

INTERVALO DE CONFIANZA DE LA MEDIA POBLACIONAL

Varianza conocida

$$\left\langle \overline{x} - Z_{(1-\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}} ; \overline{x} + Z_{(1-\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}} \right\rangle$$

Ejercicio

Suponga que la producción de clips metálicos por minuto de un determinado modelo de maquinaria industrial sigue una distribución normal con desviación estándar 18. En una muestra de 36 máquinas instaladas se ha obtenido una media de 145 clips por minuto. Construya un intervalo de confianza al 95% para la media poblacional

Ejercicio

Un comprador está interesado en la resistencia a la tensión de una fibra que se usa en la manufactura de telas. La experiencia indica que la desviación estándar de la resistencia es de 2 psi. Se selecciona una muestra aleatoria de ocho piezas de fibras y la resistencia media a la tensión resulta ser de 127 psi. Calcule e interprete con 95% de confianza para la verdadera resistencia media a la tensión

INTERVALO DE CONFIANZA DE LA MEDIA POBLACIONAL

Varianza Desconocida

$$\left\langle \overline{x} - t_{\left(1-\frac{\alpha}{2},n-1\right)} \frac{s}{\sqrt{n}} ; \overline{x} + t_{\left(1-\frac{\alpha}{2},n-1\right)} \frac{s}{\sqrt{n}} \right\rangle$$

Ejercicio

Se usa una máquina para llenar envases con cierto producto líquido. Es posible suponer que el volumen de llenado tiene distribución normal. Se selecciona una muestra aleatoria de 5 envases y se miden los contenidos netos, con los resultados que se muestran.

25.5	26.8	24.2	25	27.3

Estimar e interpretar un intervalo de confianza del 95 % para el volumen medio de llenado.

INFERENCIA PARA PROPORCIONES

Interés: Estimar la proporción p (o el porcentaje) de ocurrencia de un evento

Ejemplo:

El porcentaje de votantes que favorecen a un cierto candidato, etc.

- Cuando el tamaño de muestra es muy grande, entonces el estadístico es:

$$Z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}}$$

se distribuye aproximadamente como una normal estándar. Cuando es cercano a 0 ó a 1 se debe tomar un tamaño de muestra más grande para que la aproximación sea buena.

INTERVALO DE CONFIANZA DE UNA PROPORCIÓN

Un Intervalo de confianza aproximado del 100(1- α)% para la proporción poblacional π será:

$$IC(\pi) = \left\langle p - Z_{(1-\frac{\alpha}{2})}\sqrt{\frac{p(1-p)}{n}} ; p + Z_{(1-\frac{\alpha}{2})}\sqrt{\frac{p(1-p)}{n}} \right\rangle$$

Ejercicio

Una empresa quiere introducir un nuevo producto al mercado local, por tanto quiere estimar la proporción de clientes potenciales (dispuestos a adquirir el producto al precio que se ofrece), para tal efecto se entrevistó a 200 personas de las cuales 68 mostraron ser potenciales clientes.

Encuentre el porcentaje de personas dispuestas a adquirir el producto mínimo y máximo al 95% de confianza.

INTERVALO DE CONFIANZA DE UNA VARIANZA POBLACIONAL

$$IC(\sigma^2) = \left\langle \frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}} ; \frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}} \right\rangle$$

Ejercicio:

Una de las preocupaciones de los usuarios de sistemas interactivos es la magnitud de la varianza del tiempo de respuesta. Necesitamos comprar uno de estos sistemas y, en una versión de evaluación hemos obtenido las siguientes medidas de dicho tiempo, en ms:

20.1 22.9 18.8 20.9 22.7 21.4 20 25.8 32.1 33

Suponiendo que los tiempos de respuesta tienen distribución normal, obtener un intervalo de confianza para la varianza, con un nivel de confianza del 95%

PRUEBA DE HIPÓTESIS

Objetivos

- Diferenciar entre hipótesis nula y alternativa
- Definir los errores de tipo I y de tipo II
- Describir el procedimiento para realizar una prueba de hipótesis
- Realizar una prueba de hipótesis para la media poblacional
- Realizar una prueba de hipótesis para la proporción poblacional
- Realizar una prueba de hipótesis para la varianza poblacional
- Realizar una prueba de hipótesis para la razón de varianzas poblacionales
- Realizar una prueba de hipótesis para la diferencia de medias poblacionales.
- Realizar una prueba de hipótesis para la diferencia de proporciones poblacionales.

¿Qué es una hipótesis?

- Una creencia sobre la Población, principalmente sus parámetros:
  - Media
  - Varianza
  - Proporción
- NOTA: debe establecerse antes del análisis.

HIPÓTESIS ESTADÍSTICA

Es una afirmación que se hace acerca de un parámetro poblacional.

- ***Hipótesis nula*** es una afirmación que está establecida y que se espera sea rechazada después de aplicar una prueba estadística. Se representa por *Ho*.
- ***Hipótesis alternante,*** *es la* afirmación que se espera sea aceptada después de aplicar una prueba estadística y se representa por *Ha*.

PRUEBA DE HIPÓTESIS → **Procedimiento estadístico basado en la evidencia muestral y la teoría de probabilidad.**

TIPOS DE ERRORES

- ***Error tipo I***, que se comete cuando se rechaza una hipótesis nula que realmente es cierta.
- ***Error tipo II,*** que se comete cuando se acepta una hipótesis nula que realmente es falsa.

TIPOS DE ERROR AL PROBAR HIPÓTESIS

	Realidad	
Decisión H_0	H_0 cierta	H_0 Falsa
No Rechazo H_0	**Correcto**	**Error de tipo II** **P(Error de tipo II) =β**
Rechazo H_0	**Error de tipo I** **P(Error de tipo I)= α**	**Correcto**

- Para un tamaño de muestra fijo, no se pueden reducir a la vez ambos tipos de error.
- Para reducir β, hay que aumentar el tamaño de la muestra.
- El **nivel de significación,** representada por α, es la probabilidad de cometer *error tipo I*, y por lo general se asume que tiene un valor de 0.05 ó 0.01.
- La probabilidad de cometer *error tipo II*, representado por β y al valor *1-* β se le llama ***la potencia de la prueba****.* Una buena prueba estadística es aquella que tiene una potencia de prueba alta.

Formulación Ho, H1

↓

Elegir α

↓

Supuestos

Seleccionar la prueba estadística

↓

Criterios de Decisión

Cálculo de la prueba estadística

↓

Conclusión

IDENTIFICACIÓN DE HIPÓTESIS

- **Hipótesis nula H_o**
  - La que probamos
  - Los datos pueden refutarla
  - No debería ser rechazada sin una buena razón.

- **Hipótesis Alternante H_1**
  - Niega a H_0
  - Los datos pueden mostrar evidencia a favor
  - No debería ser aceptada sin una gran evidencia a favor.

$$\begin{cases} H_0 : \quad \mu = \mu_0 \\ H_1 : \quad \mu \neq \mu_0 \end{cases}$$

$= \, , \leq \, , \geq$

$\neq \, , > \, , <$

PRUEBA DE HIPÓTESIS PARA UNA MEDIA POBLACIONAL

σ conocido

$$Z_c = \frac{\bar{x} - \mu_o}{\sigma / \sqrt{n}}$$

Ejercicio

En una fábrica de conservas de frutas desea verificar de que si las latas tiene un peso promedio inferior a 1 kg. Se sabe que el tamaño de la fruta puede introducir una variación en los pesos de las latas de manera que estos se distribuyan normalmente con una desviación estándar de 0,08. Se toma una muestra de 100 latas en la que se determina los pesos, resultando un promedio de 980 gr. Deseamos saber si la muestra comprueba tal afirmación. Utilizar un nivel de significación igual al 2,5%.

PRUEBA DE HIPÓTESIS PARA UNA MEDIA POBLACIONAL

σ desconocido

$$t_c = \frac{\bar{x} - \mu_o}{S/\sqrt{n}}$$

Un informe publicado en el New England Journal of Medicine volvió a sembrar dudas al señalar que la peor de las grasas era la margarina. Esta revista norteamericana la acusaba de disminuir el llamado colesterol “bueno” o HDL propiciando la aparición de enfermedades cardiacas. El departamento médico de la UNALM decide tomar una muestra de estudiantes (hombres y mujeres) consumidores habituales de margarina para medir su nivel de colesterol en la sangre. Los valores (en miligramos) se muestran a continuación:

Descriptive Statistics: Hombres, Mujeres

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Hombres	27	200	200	200	1.15	0.22
Mujeres	31	199.33	199.15	199.31	1.08	0.19

El departamento médico de la UNALM afirma que en promedio una persona con un nivel de colesterol menor de 200 miligramos es considerada como una con bajo riesgo de tener complicaciones cardiacas. ¿Se puede afirmar que las mujeres poseen un bajo riesgo de poseer este tipo de complicaciones?. Use $\alpha = 0.01$.

PRUEBA DE HIPÓTESIS PARA UNA PROPORCION POBLACIONAL

$$z_c = \frac{p - \pi_o}{\sqrt{\dfrac{\pi_o(1-\pi_o)}{n}}}$$

Un periódico local de la ciudad de Lima, ha publicado recientemente una noticia con el siguiente titular: “Crece el porcentaje de ciudadanos que no tienen confianza en el sistema político del país.” Más adelante en la noticia, se explicaba que la información procedía de una encuesta de opinión hecha por una prestigiosa empresa investigadora, y que los resultados mostraban un aumento con respecto a la realizada el año pasado en el cual el 35% de ciudadanos declararon “no tener confianza” con el sistema político del país. Suponiendo que la reciente encuesta fue aplicada a 3000 personas de las cuales 1100 manifestaron no tener confianza con el sistema político del país.

¿Es posible refutar el titular publicado por el periódico? Use $\propto = 0.05$

Un estudio realizado sobre la duración de los circuitos, se recogió una muestra de 225 circuitos electrónicos para estudiar la proporción de circuitos que salían del mercado. Se sabe que de esos 225 circuitos 38 no superaron el control de calidad del cliente. Calcular:

a. Un intervalo de confianza del 95% de la proporción de circuitos que superan el control de calidad.

b. ¿Se puede afirmar que la proporción de circuitos que no supera el control de calidad es superior al 20%? (α=0,05).

PRUEBA DE HIPÓTESIS PARA UNA VARIANZA POBLACIONAL

$$\chi^2_c = \frac{(n-1)S^2}{\sigma_0^2}$$

EJERCICIO

La empresa *CONTAMINA S.A.*, dedicada a la fabricación de insumos químicos, tiene su planta industrial en el distrito de Comas. El Ministerio de Salud ha recibido una queja de los pobladores ya que esta empresa despide una gran cantidad de gases tóxicos y han notado la presencia de Mercurio en sus viviendas. La empresa decide detener su producción y tomar una muestra aleatoria de 35 pobladores (15 hombres y 20 mujeres) y someterlos a una serie de exámenes. El contenido de Mercurio (en miligramos) presente en la sangre de estos pobladores se muestra a continuación:

Variable	N	Mean	Median	TrMean	StDev	SE Mean	Minimum	Maximum	Q1	Q3
Mujeres	20	3.9905	3.965	3.9889	0.1867	0.0417	3.6	4.41	3.875	4.10
Hombres	15	4.307	3.9	4.175	0.969	0.25	3.31	7.02	3.590	4.73

¿Se puede afirmar que la variancia del contenido neto de Mercurio presente en la sangre de los hombres es mayor a 0.95?. Use $\alpha = 0.05$.

El peso de 12 latas de cerezas, en onzas, es:

11,9 12,3 12,6 11,8 12,1 11,5

12,7 11,3 11,9 12,0 11,8 12,1

La variación estándar especificada es de 1/2 onza. ¿Se cumple esta especificación? Use el nivel de significación del 1% y una prueba bilateral

PRUEBA DE HIPÓTESIS PARA UNA RAZON DE VARIANZAS POBLACIONALES

$$F_c = \frac{S_1^2}{S_2^2}$$

Del ejemplo de la empresa CONTAMINA:

Probar si existe homogeneidad de varianzas?. Usar un nivel de significación del 10 %

Mediante dos procesos se fabrican alambres galvanizados lisos para alambrados rurales. Los técnicos de la fábrica desean determinar si los dos procesos poseen diferentes efectos en la resistencia de la media de ruptura del alambre. Se someten varias muestras a los dos procesos dando los siguientes resultados:

Proceso 1 = 9 4 10 7 9 10

Proceso 2 = 14 9 13 12 13 8 10

Probar si existe homogeneidad de varianzas en los procesos con un α = 0,10.

PRUEBA DE HIPÓTESIS PARA UNA DIFERENCIA DE MEDIAS POBLACIONALES

$$Z_c = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

EJERCICIO

Un empresario desea comparar la productividad de dos tipos de obreros industriales de una región, supone que la productividad de ambos tipos de trabajadores es similar pero con mayor variabilidad en uno de ellos; desviación estándar 0,9 por hora en la industria A, con solo 0,3 en la industria B. Para comprobar esta suposición controla durante un cierto tiempo la producción de 200 obreros de A y 350 obreros de B obteniendo una productividad media por hora de 1 y 0,89 respectivamente. ¿Puede concluirse en base a estos resultados que la suposición del empresario era correcta? (α = 0,05)

PRUEBA DE HIPÓTESIS PARA UNA DIFERENCIA DE MEDIAS POBLACIONALES CON VARIANZAS POBLACIONALES DESCONOCIDAS

$$t_c = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Del ejercicio de Colesterol:

i. El informe publicado en el New England Journal of Medicine menciona además que en las universidades Americanas el nivel promedio de colesterol en los hombres es un miligramo mayor que el nivel promedio de colesterol en las mujeres. ¿Se puede afirmar lo mismo en la UNALM?. Use $\alpha = 0.02$.

ii. Suponga que al siguiente ciclo al momento de la matricula se midió el nivel de colesterol de todos los estudiantes de la UNALM encontrándose como valores promedio 201 y 198 miligramos para los hombres y las mujeres respectivamente. Asumiendo que no hubo cambios en el nivel de colesterol de los estudiantes. ¿Se cometió algún error?.

HIPÓTESIS PARA COMPARAR DOS MEDIAS POBLACIONALES (Varianzas heterogéneas)

$$\mathbf{H_0:} \quad \mu_1 - \mu_2 = D_0$$

$$\mathrm{H_a:} \quad \begin{cases} \mu_1 - \mu_2 > D_0 \\ \mu_1 - \mu_2 < D_0 \\ \mu_1 - \mu_2 \neq D_0 \end{cases}$$

$$t'_c = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad H = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}} - 2$$

$$\textbf{R.R.:} \quad \begin{cases} t'_c > t_H \\ t_c < t_H \\ \left| t'_c \right| > t_H \end{cases}$$

Del ejemplo de la empresa CONTAMINA

El Ministerio de Salud ha recomendado a la empresa iniciar un tratamiento para purificar la sangre de estos pobladores. Para esto es necesario aplicar una dosis de PURIFICOL durante un mes a cada una de las personas afectadas. La dosis diaria es 50 mg más 0.25 mg por cada miligramo de Mercurio presente en la sangre. El costo de PURIFICOL es de 1 nuevo sol por miligramo. ¿Se puede afirmar que en promedio el gasto diario en PURIFICOL por persona es el mismo para los hombres y las mujeres?. Use $\alpha = 0.10$.

PRUEBA DE HIPÓTESIS DE UNA DIFERENCIA DE PROPORCIONES POBLACIONALES

$$H_0: \pi_1 = \pi_2$$

$$H_a: \begin{cases} \pi_1 > \pi_2 \\ \pi_1 < \pi_2 \\ \pi_1 \neq \pi_2 \end{cases}$$

$$\mathbf{Z_c} = \frac{(p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\textbf{R.R.} \begin{cases} Z_c > Z_{(1-\alpha)} \\ Z_c < Z_{\alpha} \\ |Z_c| > Z_{(1-\frac{\alpha}{2})} \end{cases}$$

De una muestra de 450 votantes hombres, 105 se declararon simpatizantes del candidato A. De una muestra de 550 votantes mujeres 120 se declararon simpatizantes del mismo candidato. ¿ Proporcionan estos datos evidencia suficiente como para considera que las proporciones de los simpatizantes hombres y mujeres son iguales? $\propto = 0,05$

PRUEBA DE HIPÓTESIS DE UNA DIFERENCIA PROPORCIONES POBLACIONALES

$$H_0: \quad \pi_1 - \pi_2 = k$$

$$H_a: \begin{cases} \pi_1 - \pi_2 > k \\ \pi_1 - \pi_2 < k \\ \pi_1 - \pi_2 \neq k \end{cases}$$

$$\mathbf{Z}_c = \frac{(p_1 - p_2) - k}{\sqrt{\left(\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}\right)}}$$

$$\textbf{R.R.} \begin{cases} Z_c > Z_{(1-\alpha)} \\ Z_c < Z_{\alpha} \\ |Z_c| > Z_{(1-\frac{\alpha}{2})} \end{cases}$$