INFERENCIA ESTADISTICA

Es una rama de la Estadística que se ocupa de los procedimientos que nos permiten analizar y extraer conclusiones de una población a partir de los datos de una muestra aleatoria, mediante la teoría de las probabilidades y de las distribuciones muéstrales. Comprende:

a) Estimación de Parámetros -

- Puntual

- Intervalos

b) Prueba de hipótesis.

ESTIMACION DE PARAMETROS.

1) ESTIMACION PUNTUAL.- Si θ es un parámetro poblacional (valor desconocido); la estimación puntual o estimador es est(θ) es decir:

$$Est(\theta) = h(X1, X2, \ldots, Xn)$$

Ejemplo

Si $\theta=\mu$ (media poblacional), un estimador puede ser el promedio de una muestra. Otro estimador puede ser el punto medio entre el valor mas pequeño y el mas grande; sin embargo existe un mejor estimador, aquel que satisface ciertas propiedades de un buen estimador.

Propiedades de un buen estimador.-

1. Debe ser insesgado: $E [Est(\theta)] = \theta$

2. Debe ser eficiente: Si $\theta1$ y $\theta2$ son dos estimadores insesgados de θ, entonces $\theta1$ es más eficiente que $\theta2$ si y sólo si $Var(\theta1) < Var(\theta2)$.

3. Debe ser consistente: θ es un estimador consistente de θ, si y sólo si cuando el tamaño de muestra se incrementa, $Est(\theta)$ se aproxima a θ.

4. Debe ser suficiente: θ es un estimador suficiente de θ, si y sólo si $Est(\theta)$ contiene la información suficiente para estimar el parámetro θ.

El promedio y la variancia muestral son buenos estimadores de la media y variancia de la población.

2) ESTIMACION POR INTERVALOS.- Consiste en determinar un conjunto de valores, el cual contiene el valor verdadero del parámetro θ, con un nivel de confianza dado: $(1-\alpha)x100$

$$LIC < \theta < LSC$$

LIC: límite inferior de confianza
LSC: límite superior de confianza

Intervalo de confianza (θ) = <LIC, LSC>

Basado en los supuestos de distribución normal o aproximadamente normal de la población en estudio, se puede deducir los siguientes intervalos de confianza:

Intervalo de confianza para la media poblacional (μ)

Caso 1: Cuando σ^2 es conocida

$$\overline{X} \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

Caso 2: Cuando σ^2 es desconocida

$$\overline{X} \pm t_{\alpha}(n-1) \frac{S}{\sqrt{n}};$$

S es la desviación estándar de la muestra.

EJEMPLO.- Si de una población normal con media μ y σ^2 desconocida caso de la producción de madera contrachapada (miles de m3) en los años 1990 -1990

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
23.73	25.89	29.38	35.84	64.48	63.94	69.44	52.58	133.99	34.32

El promedio = 53,36
Desviación estándar = 33.12
n = 10
t0,05(9) =2.26

Limites: 29,67 y 77,05

El intervalo <29,67 , 77.0> brinda un 95% de confianza de contener el verdadero valor de μ de la producción madera contrachapada.

Intervalo de confianza para la variancia (σ^2)

Es útil también conocer los limites de la variancia, a una confianza determinada. Para esto se utiliza la distribución Chi Cuadrada.

Se requiere estimar la variancia de la población al 95% de confianza.

Limite inferior: $$\frac{(n-1)\,S^2}{\chi^2_{0.975}(n-1)}$$;

Limite superior: $$\frac{(n-1)\,S^2}{\chi^2_{0.025}(n-1)}$$

Variancia de la muestra (S^2) y el tamaño de muestra (n)

Por ejemplo, para la variancia de la madera contrachapada.

Limite inferior: $$\frac{(10-1)1097}{19.02} = 519.08$$;

Limite superior: $$\frac{(10-1)1097}{2.7} = 3656.67$$

Si se requiere los limites de confianza para la desviación estándar, a estos limites hallar su raíz cuadrada.

Limite inferior: 22.78; Limite superior 60.47

Prueba de Hipótesis

Hipótesis Estadística.- Es un supuesto sobre la distribución de una variable aleatoria, que necesita ser comprobada para su aceptación o rechazo.

Hipótesis Planteada o Nula (Hp o H0).- es la suposición que se hace acerca de que el parámetro pueda tomar determinado valor.

Hipotesis Alternante (Ha o H1).- es la negación de la hipótesis planteada, esta hipótesis se aceptada siempre y cuando la hipótesis planteada fue rechazada.

Prueba de Hipótesis.- Es un procedimiento estadístico de comprobación de una hipótesis y se realiza utilizando los valores observados que constituyen una muestra.

Tipos de Errores.- En el procedimiento de prueba de hipótesis se puede incurrir en dos tipos de errores:

- Error tipo I: cuando se rechaza una hipótesis planteada, siendo ésta realmente cierta
- Error tipo II: cuando se acepta una hipótesis planteada, siendo ésta realmente falsa.

Nivel de significación (α).- Es la probabilidad de cometer error tipo I; es decir, es la probabilidad de rechazar una hipótesis planteada verdadera.

La probabilidad de cometer error tipo II está representado por la letra griega ß.

Pasos necesarios para realizar una prueba de hipótesis

1. formulación de las hipótesis: Hp , Ha
2. Establecer el nivel de significación (α)
3. Determinar la prueba estadística (t, Z, χ^2, F) y las asunciones respectivas. En todos los casos las asunciones de la prueba son:

   - La población de donde se extrae la muestra, es normal
   - La muestra es extraída al azar
4. Determinar las regiones de aceptación y rechazo de Hp.
5. Realizar el calculo de la prueba estadística.
6. Establecer las conclusiones de la prueba.

Prueba de Hipótesis para la Media de una Población

1.- Hp: μ = k Hp: μ ≥ k Hp: μ ≤ k
Ha: μ ≠ k Ha: μ < k Ha: μ > k

2.- Establecer el nivel de significación, α

3.- Prueba estadística.- puede ser: "Z" si la variancia poblacional es conocida ó "t" si es desconocida. Para un cado u otro, los estadísticos son:

$$Zcalculado = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \;;$$

$$t\_calculado = \frac{\overline{X} - \mu}{\sqrt{\frac{S^2}{n}}}$$

Conclusiones.-

Si el valor calculado del estadístico cae en la zona de aceptación; entonces se acepta la hipótesis planteada.

La zona de aceptación de la hipótesis planteada (nula) por ejemplo para 0.05 de significancia:

a) Hp: $\mu = k$; Se acepta esta hipótesis si t calculado esta en el intervalo [t 0.025, t 0.975]

se desea probar si la media poblacional es diferente de "k" por eso que en la nulidad se pone que $\mu = k$ y se espera rechazar esta hipótesis para afirmar con un nivel de probabilidad de 0.05 de error que $\mu \neq k$.

b) Hp: $\mu \geq k$; Se acepta esta hipótesis si t calculado es superior a t 0.05

se desea probar si la media poblacional es inferior a "k", por eso que en la hipótesis de nulidad se pone $\mu \geq k$ y se espera rechazar esta hipótesis para afirmar con una probabilidad de 0.05 de error que $\mu < k$.

c) Hp: $\mu \leq k$; Se acepta esta hipótesis si t calculado es inferior a t 0.95

se desea probar si la media poblacional es superior a "k", por eso que en la hipótesis de nulidad se pone $\mu \leq k$ y se espera rechazar esta hipótesis para afirmar con una probabilidad de 0.05 de error que $\mu > k$.

En otros casos se rechaza la hipótesis y se acepta la alterna.

En los casos b) y c) se dice prueba de una cola y en el caso a) de dos colas.

Para ilustrar estas hipótesis se plantea los siguientes casos en el estudio de madera contrachapada.

1. Probar la hipótesis que la producción media contrachapada en el Perú a cambiado, según la muestra de estos últimos 10 años. (considere que la producción media estimada general es 45 mil m3 por año.

Hp: $\mu = 45$
Ha: $\mu \neq 45$
$\alpha = 0.05$

Según las estadísticas de la muestra (n=10)

El promedio = 53,36
Variancia muestral = 1097

Los valores críticos para la prueba, según la tabla de t-Student son:

t0,975(9) = 2.26 ; t0,025(9) =-2.26

Mediante Excel puede ser hallado con la función; TINV(0.05, 9) = 2.26 (la función esta programada para la prueba de dos colas)

$$t.calculado = \frac{53.36 - 45}{\sqrt{\frac{1097}{10}}} = 0.798$$

2. Probar la hipótesis que la producción media contrachapada en el Perú se a incrementado, según los reportes de estos últimos 10 años.

Hp: $\mu \leq 45$
Ha : $\mu > 45$
$\alpha = 0.05$

Las estadísticas de la muestra son las mismas; el valor critico es diferente: $t(0,95) = 1.83$.

Mediante Excel =Tinv(0.10,9) = 1.83

El valor t-calculado se determina de igual forma que para la hipótesis anterior.

t-calculado = 0.798

Este valor es menor que el critico (0.798 < 1.83) por lo tanto se acepta la hipótesis planteada, es decir que no hay evidencia estadística para afirmar que la producción media de madera contrachapada se ha incrementado.