11MIAR–Herramientas de Estadística (Complemento Formativo-3ECTS)

Dr. Walter A. Ortiz

Máster Universitario en Inteligencia Artificial
Universidad Internacional de Valencia

July 17, 2025

Inferencia estadística Son las afirmaciones válidas acerca de la población o proceso basadas en la información contenida en la muestra.

Estadístico Cualquier función de los datos muestrales que no contiene parámetros desconocidos.

Distribución de probabilidad de X Relaciona el conjunto de valores de X con la probabilidad asociada con cada uno de estos valores.

Estimador puntual Estadístico que estima un valor específico de un parámetro.

Ejemplo: La media poblacional es un estimador de la media muestral $\hat{\mu} = \bar{X}$

Se pueden proponer tantos estimadores para los parámetros como se quiera.

En lo que sigue se denotará cómo parámetro θ y estimador del parámetro $\hat{\theta}$.

Propiedades de los estimadores

- Un estimador es insesgado, si verifica $E[\hat{\theta}] = \theta$.
- En caso de no ser insesgado, es sesgado y este está dado por $E[\hat{\theta}] - \theta$
- Eficiencia :Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de θ. Diremos que $\hat{\theta}_1$ es mas eficiente que $\hat{\theta}_2$ si verifica que $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.
- Error cuadrático medio: $ECM(\hat{\theta}) = V(\hat{\theta}) + sesgo(\hat{\theta})^2$

Ejemplo: Verificar si los siguientes estimadores son insesgados. Si tiene por media poblacional, μ y desviación típica σ. Cuál es más eficiente?

- $\hat{\theta} = 0.5x_1 + 0.15x_2 + 0.25x_3 + 0.1x_4$
- $\hat{\theta} = \dfrac{x_1 + 2x_2 + 3x_3}{6}$
- $\hat{\theta} = \frac{x_3 - 4x_2}{-3}$

Estimación Intervalo de confianza para la varianza Para dicha estimación debe ser conocido el valor de la varianza o desviación estándar o típica (el estadístico); es decir s^2 donde.

$$\sigma^2 \in \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \right)$$

$\chi^2_{\alpha/2,n-1}$ (Chi cuadrado)

Estimación Intervalo de confianza para la varianza de dos poblaciones Para dicha estimación debe ser conocido el valor de las varianza o desviaciones estándar o típica (el estadístico); es decir s_1^2, s_2^2 donde.

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{s_1^2}{s_2^2} F_{1-\alpha/2, n_1-1, n_2-1}, \frac{s_1^2}{s_2^2} F_{\alpha/2, n_1-1, n_2-1} \right)$$

$F_{1-\alpha/2, n_1-1, n_2-1}$ (F-Fisher)

Tamaño de la muestra: Se considera el error, como

$$E = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \rightarrow n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2$$

Hipótesis estadística Es una afirmación sobre los valores de los parámetros de una población o proceso, que puede probarse a partir de la información contenida en una muestra.

Hipótesis nula: H_0 Es considerada como la que hace referencia al valor del parámetro que se quiere probar como verdadero.

Hipótesis alternativa: H_1 Corresponde a la falsedad o estableciendo que el parámetro puede ser mayor, menor o igual, de acuerdo con la propuesta hecha en la hipótesis nula.

Estadístico de prueba o contraste Número calculado a partir de los datos y de H_0, cuya magnitud permite discernir si se rechaza o no la hipótesis nula.

Región de rechazo Es el conjunto de posibles valores del estadístico de prueba que llevan a rechazar la hipótesis nula.

Nivel de significación α es el recicproco de la confianza, el cual debe ser fijado antes de escoger la muestra.

Observación: Los test de hipótesis pueden ser unilaterales o bilaterales

Test de hipótesis- Comparación de medias I

- $H_0 : \mu = \mu_0$
  $H_1 : \mu \neq \mu_0 (\mu \leq \mu_0 \;\; \mu \geq \mu_0)$
- $\alpha = 0.05, 0.01, 0.1$
- Estadistico de prueba o contraste:
  varianza poblacional conocida

$$Z_0 = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

- Región de rechazo $|Z_0| > Z_{\alpha/2}$, $Z_0 > Z_\alpha$, $Z_0 < -Z_\alpha$

Universidad Internacional de Valencia

Test de hipótesis- Comparación de medias I

- $H_0 : \mu_1 = \mu_2$
  $H_1 : \mu_1 \neq \mu_2 (\mu_1 \leq \mu_2 \;\; \mu_1 \geq \mu_2)$
- $\alpha = 0.05, 0.01, 0.1$
- Estadistico de prueba o contraste:
  varianzas poblacionales conocidas

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

- Región de rechazo $|Z_0| > Z_{\alpha/2}$, $Z_0 > Z_\alpha$, $Z_0 < -Z_\alpha$

Universidad Internacional de Valencia

Test de hipótesis- Comparación de medias II

- $H_0 : \mu = \mu_0$
  $H_1 : \mu \neq \mu_0 (\mu \leq \mu_0 \;\; \mu \geq \mu_0)$
- $\alpha = 0.05, 0.01, 0.1$
- Estadistico de prueba o contraste:
  varianza poblacional desconocida

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

- Región de rechazo $|t_0| > t_{\alpha/2,n-1} \;\; t_0 > t_{\alpha,n-1}, \; t_0 < -t_{\alpha,n-1}$

Test de hipótesis- Comparación de medias II

- $H_0 : \mu_1 = \mu_2$
  $H_1 : \mu_1 \neq \mu_2 (\mu_1 \leq \mu_2 \;\; \mu_1 \geq \mu_2)$
- $\alpha = 0.05, 0.01, 0.1$
- Estadistico de prueba o contraste:
  varianzas poblacionales conocidas

$$t_0 = \frac{\bar{x_1} - \bar{x_2}}{s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

donde $s_p = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

- Región de rechazo $|t_0| > t_{\alpha/2, n_1+n_2-2}$, $t_0 > t_{\alpha, n_1+n_2-2}$
  $t_0 < -t_{\alpha, n_1+n_2-2}$