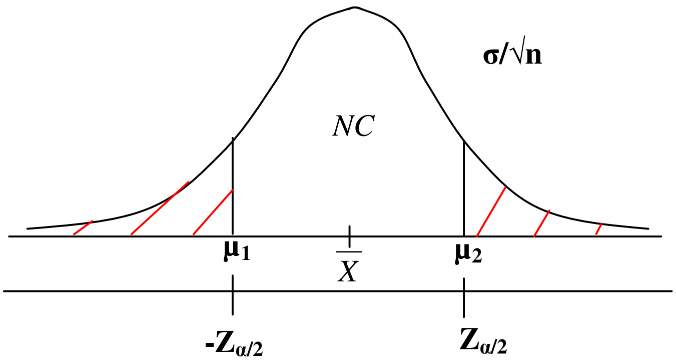
MODULO IV. ESTIMACIÓN POR INTERVALOS DE CONFIANZA – ANÁLISIS DE CASOS

DOCENTE: JUAN CARLOS VERGARA SCHMALBACH

ESTADÍSTICA INFERENCIAL

CASO 1: ESTIMACIÓN DE LA MEDIA CON DESVIACIÓN POBLACIONAL CONOCIDA

La primera estimación es para la μ a partir de un $\overline{X}$ (para una muestra de tamaño *n*) y σ conocida. Una vez definido un nivel de confianza (*NC*) o nivel de significancia (*α*), podremos obtener los estimadores críticos representados por $Z_{\alpha/2}$.



Los estimadores críticos representan un intervalo, donde, con un nivel de confianza (NC) encontraremos la media poblacional. Matemáticamente tenemos:

$$P(\mu_1 < \mu < \mu_2) = NC \text{ equivalente a } P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = NC \quad (1)$$

Donde Z representa:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \quad (2)$$

Remplazamos la ecuación completa de Z en (1):

$$P(-Z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}) = NC$$

Los valores conocidos son $Z_{\alpha/2}$, $\overline{X}$, n y σ , para lo cual despejaremos el valor de μ .

$$P\left(\frac{-Z_{\alpha/2}\sigma}{\sqrt{n}} < \overline{X} - \mu < \frac{Z_{\alpha/2}\sigma}{\sqrt{n}}\right) = NC$$

$$P\left(\frac{-Z_{\alpha/2}\sigma}{\sqrt{n}} - \overline{X} < -\mu < \frac{Z_{\alpha/2}\sigma}{\sqrt{n}} - \overline{X}\right) = NC$$

Multiplicamos por -1 las desigualdades:

$$P\left(\overline{X} + \frac{Z_{\alpha/2}\sigma}{\sqrt{n}} > \mu > \overline{X} - \frac{Z_{\alpha/2}\sigma}{\sqrt{n}}\right) = NC$$

Se determinan un intervalo de confianza para μ :

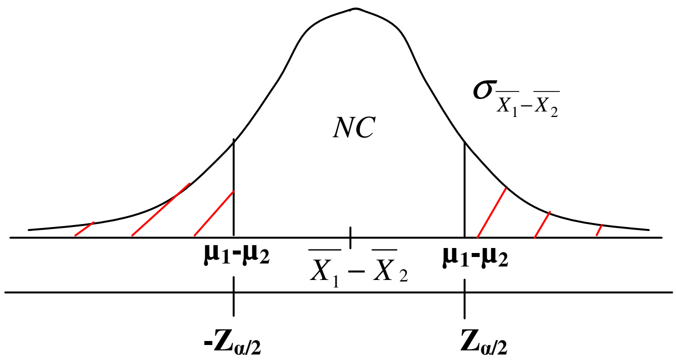
$$\mu = \overline{X} \pm \frac{Z_{\alpha/2}\sigma}{\sqrt{n}}$$

Identificamos el error de la estimación *e*:

$$e = \frac{Z_{\alpha/2}\sigma}{\sqrt{n}}$$

CASO 2: ESTIMACIÓN DE LA DIFERENCIA DE MEDIAS CON DESVIACIONES POBLACIONALES CONOCIDAS

La estimación de la diferencia de medias muestrales consiste en un ejercicio similar al primer caso, donde conocemos de antemano $Z_{\alpha/2}$, $\overline{X}_1$, $\overline{X}_2$, n_1, n_2, σ_1 y σ_2 .



$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = NC \quad (1)$$

Donde Z representa:

$$Z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (2)$$

Remplazamos la ecuación completa de Z en (1):

$$P\left(-Z_{\alpha/2} < \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < Z_{\alpha/2} \right) = NC$$

Los valores conocidos son $Z_{\alpha/2}$, $\overline{X}$, n y σ , para lo cual despejaremos el valor de μ .

$$P\left(-Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < (\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2) < Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) = NC$$

$$P\left(-Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} - (\overline{X_1} - \overline{X_2}) < -(\mu_1 - \mu_2) < Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} - (\overline{X_1} - \overline{X_2}) \right) = NC$$

Multiplicamos por -1 las desigualdades:

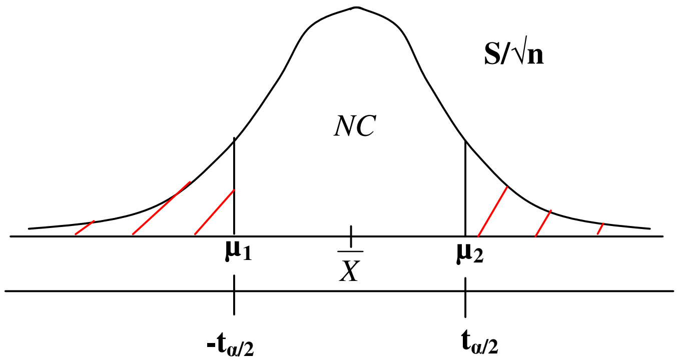
$$P\left((\overline{X_1} - \overline{X_2}) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} > \mu_1 - \mu_2 > (\overline{X_1} - \overline{X_2}) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) = NC$$

Se determinan un intervalo de confianza para μ :

$$\mu_1 - \mu_2 = (\overline{X}_1 - \overline{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

CASO 3: ESTIMACIÓN DE LA MEDIA CON DESVIACIÓN POBLACIONAL DESCONOCIDA

Como se desconoce σ, es necesario emplear la desviación muestral (S), tomando lcomo eje para resolver los ejercicios la distribución t - students. Una vez definido un nivel de confianza (*NC*), podremos obtener los estimadores críticos representados por $t_{\alpha/2}$.



Replicando los pasos de los casos anteriores tenemos:

$$P(\mu_1 < \mu < \mu_2) = NC \text{ equivalente a } P(-t_{\alpha/2} < t < t_{\alpha/2}) = NC \quad (1)$$

Donde t representa:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \quad (2)$$

Remplazamos la ecuación completa de t en (1):

$$P(-t_{\alpha/2} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}) = NC$$

Los valores conocidos son $t_{\alpha/2}$, $\overline{X}$, *n* y S, para lo cual despejaremos el valor de μ.

$$P(\frac{-t_{\alpha/2}S}{\sqrt{n}} < \overline{X} - \mu < \frac{t_{\alpha/2}S}{\sqrt{n}}) = NC$$

$$P(\frac{-t_{\alpha/2}S}{\sqrt{n}} - \overline{X} < -\mu < \frac{t_{\alpha/2}S}{\sqrt{n}} - \overline{X}) = NC$$

Multiplicamos por -1 las desigualdades:

$$P(\overline{X} + \frac{t_{\alpha/2}S}{\sqrt{n}} > \mu > \overline{X} - \frac{t_{\alpha/2}S}{\sqrt{n}}) = NC$$

Se determinan un intervalo de confianza para μ:

$$\mu = \overline{X} \pm \frac{t_{\alpha/2}S}{\sqrt{n}}$$

Identificamos el error de la estimación *e*:

$$e = \frac{t_{\alpha/2}S}{\sqrt{n}}$$

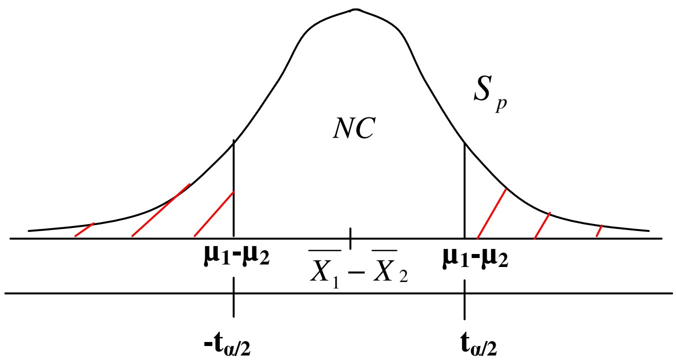
CASO 4: ESTIMACIÓN DE LA DIFERENCIA DE MEDIAS CON DESVIACIONES POBLACIONALES DESCONOCIDAS

CASO A: DESVIACIONES POBLACIONALES IGUALES

Se da por hecho que las desviaciones poblaciones son iguales (a pesar del desconocimiento de su valor). Para tal efecto se procede a calcula una desviación muestral a partir del ponderada (S_p).

$$S_p^2 = \frac{(n_1 - 1)S_1^2 - (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$ con grados de libertad $v = n_1 + n_2 - 2$

El gráfico quedaría:



$$P(-t_{\alpha/2} < t < t_{\alpha/2}) = NC \quad (1)$$

Donde t representa:

$$t = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2)$$

Remplazamos la ecuación completa de t en (1):

$$P\left(-t_{\alpha/2} < \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t_{\alpha/2}\right) = NC$$

$$P\left(-t_{\alpha/2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < (\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2) < t_{\alpha/2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = NC$$

$$P\left(-t_{\alpha/2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} - (\overline{X_1} - \overline{X_2}) < -(\mu_1 - \mu_2) < t_{\alpha/2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} - (\overline{X_1} - \overline{X_2})\right) = NC$$

Multiplicamos por -1 las desigualdades:

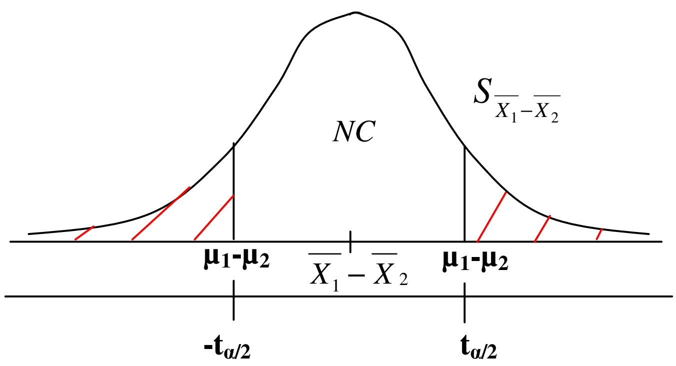
$$P\left((\overline{X_1} - \overline{X_2}) + t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} > \mu_1 - \mu_2 > (\overline{X_1} - \overline{X_2}) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) = NC$$

Se determinan un intervalo de confianza para $\mu_1 - \mu_2$:

$$\mu_1 - \mu_2 = (\overline{X_1} - \overline{X_2}) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

CASO B: DESVIACIONES POBLACIONALES DIFERENTES

Procedemos a trabajar con las desviaciones muestrales de forma independiente, realizando el cálculo del ponderaje solo para determinar los grados de libertad (*v*)



$$P(-t_{\alpha/2} < t < t_{\alpha/2}) = NC \quad (1)$$

Donde t representa:

$$t = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (2)$$

Remplazamos la ecuación completa de t en (1):

$$P\left(-t_{\alpha/2} < \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} < t_{\alpha/2} \right) = NC$$

Despejaremos el valor de $\mu_1 - \mu_2$.

$$P\left(-t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < (\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2) < t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right) = NC$$

$$P\left(-t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} - (\overline{X_1} - \overline{X_2}) < -(\mu_1 - \mu_2) < t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} - (\overline{X_1} - \overline{X_2}) \right) = NC$$

Multiplicamos por -1 las desigualdades:

$$P\left((\overline{X_1} - \overline{X_2}) + t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} > \mu_1 - \mu_2 > (\overline{X_1} - \overline{X_2}) - t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right) = NC$$

Se determinan un intervalo de confianza para $\mu_1 - \mu_2$:

$$\mu_1 - \mu_2 = (\overline{X_1} - \overline{X_2}) \pm t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

CASO 5: ESTIMACIÓN DE UNA PROPORCIÓN

El objetivo es determinar la Proporción Poblacional (*P*) a partir del análisis de una muestra.

Con los datos obtenidos de una muestra de tamaño *n*, podremos obtener los valores de *p* y *q*, y un $Z_{\alpha/2}$ a partir de un nivel de confianza (*NC*).

Recordemos que la desviación muestral de la proporción muestral tiende a parecerse a su desviación poblacional cuando trabajamos con poblaciones grandes (mayores a 30), por lo tanto, estableceremos la siguiente relación:

$$\sigma_p = \sqrt{\frac{PQ}{n}} \, equivalente \, a \, S_p = \sqrt{\frac{pq}{n}} \; para \, n > 30$$

Tenemos:

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = NC \quad (1)$$

Donde:

$$Z = \frac{p - P}{\sqrt{\frac{pq}{n}}} \quad (2)$$

Remplazando Z en la ecuación (1) y despejando P concluimos que:

$$P(-Z_{\alpha/2} < \frac{p-P}{\sqrt{\frac{pq}{n}}} < Z_{\alpha/2}) = P(p + Z_{\alpha/2}\sqrt{\frac{pq}{n}} > P > p - Z_{\alpha/2}\sqrt{\frac{pq}{n}}) = NC$$

$$P = p \pm Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

CASO 6: ESTIMACIÓN DE LA DIFERENCIA DE PROPORCIONES

Partimos del valor de **Z** para la diferencia de medias muestrales.

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \quad (1)$$

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = NC \quad (2)$$

Reemplazando (1) en (2) y despejando $P_1 - P_2$ se estima la diferencia de proporciones:

$$P(-Z_{\alpha/2} < \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} < Z_{\alpha/2}) = NC$$

$$P_1 - P_2 = (p_1 - p_2) \pm Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

CASO 7: ESTIMACIÓN DE LA VARIANZA PROPORCIONAL

Partimos del análisis del la distribución X^2 donde se conoce el tamaño de la muestra (*n*) y la varianza de la muestra (S^2).

Esto valores extremos serán denotados por el nivel de significancia ($\alpha = 1 - NC$).

Fórmula: $X^2 = \frac{(n-1)S^2}{\sigma^2}$

Gráfico:

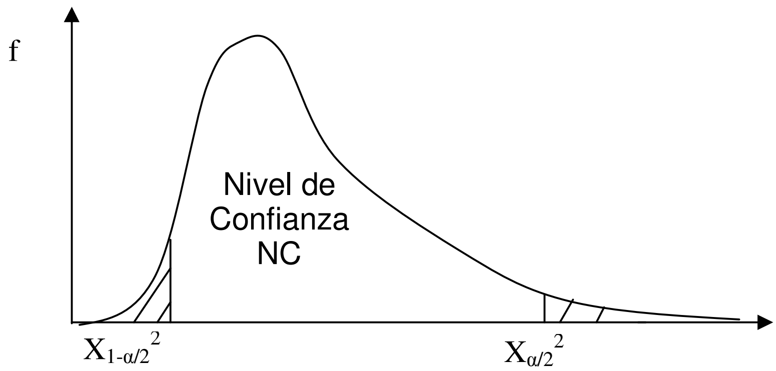


Figura 1. Distribución chi – cuadrado (X^2) (Grados de libertad v = n – 1)

Tenemos que la relación entre la varianza poblacional y la varianza muestral representan un X^2 comprendido entre $X^2{}_{1-\alpha/2}$ y $X^2{}_{\alpha/2}$ con una seguridad o probabilidad equivalente a *NC*, es decir:

$$P(X^2_{1-\alpha/2} < X^2 < X^2_{\alpha/2}) = NC$$

Remplazando X^2 tenemos que:

$$P(X^2_{1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < X^2_{\alpha/2}) = NC$$

Despejando la varianza poblacional:

$$P(\frac{(n-1)S^2}{X^2_{1-\alpha/2}} > \sigma^2 > \frac{(n-1)S^2}{X^2_{\alpha/2}}) = NC$$

Logrando un intervalo de confianza con una certeza de *NC*, de que la varianza poblacional se encontrará en dicho intervalo.

$$\sigma_1^2 = \frac{(n-1)S^2}{X^2_{1-\alpha/2}}$$

$$\sigma_2^2 = \frac{(n-1)S^2}{X^2_{\alpha/2}}$$