PyE – INTERVALOS DE CONFIANZA – Diferencias paramétricas– UTN

En esta segunda parte de Intervalos de Confianza, nos va a interesar estimar la diferencia entre los parámetros de dos variables aleatorias, con el objetivo de estudiar cuál será mayor y en qué grado o si bien podría llegar a ser iguales. Solo buscaremos estos intervalos para la diferencia entre dos esperanzas, o bien entre dos proporciones.

• Intervalos de Confianza para la diferencia de medias.

1er Caso **Dadas dos muestras aleatorias independientes,** $X_1, \ldots, X_{n_1}$**, con** $X_i \sim N(\mu_1, \sigma_1^2)$**, e** $Y_1, \ldots, Y_{n_2}$**, con** $Y_i \sim N(\mu_2, \sigma_2^2)$**, donde los valores** σ_1 **y** σ_2 **son conocidos:**

Se quiere construir un intervalo de confianza de nivel $1-\alpha$ para la diferencia

$$\mu_1 - \mu_2$$

Partimos de un estimador de esta diferencia:

$$\overline{X} - \overline{Y}$$

Como X_i e Y_i son v.a. normales e independientes para todo i, y, además tenemos que

$$E\left(\overline{X} - \overline{Y}\right) = \mu_1 - \mu_2 \quad \text{y} \quad V\left(\overline{X} - \overline{Y}\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

podemos plantear el **pivote**, a partir del cual obtendremos el intervalo deseado:

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0;1),$$

Hacemos el planteo a partir del cual despejaremos la diferencia $\mu_1 - \mu_2$

$$\begin{aligned}
1-\alpha &= P\left(-z_{1-\alpha/2} \leq \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_{1-\alpha/2}\right) \\
&= P\left(-\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}z_{1-\alpha/2} \leq \overline{X} - \overline{Y} - (\mu_1 - \mu_2) \leq \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}z_{1-\alpha/2}\right) \\
&= P\left(-(\overline{X} - \overline{Y}) - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}z_{1-\alpha/2} \leq -(\mu_1 - \mu_2) \leq -(\overline{X} - \overline{Y}) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}z_{1-\alpha/2}\right) \\
&= P\left(\overline{X} - \overline{Y} + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}z_{1-\alpha/2} \geq \mu_1 - \mu_2 \geq \overline{X} - \overline{Y} - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}z_{1-\alpha/2}\right)
\end{aligned}$$

Por lo tanto decimos que encontramos un

$$\boxed{\text{Intervalo de Confianza de nivel } 1-\alpha \text{ para } \mu_1-\mu_2\text{: } \left[\overline{X}-\overline{Y}-\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}z_{1-\alpha/2}\ ;\ \overline{X}-\overline{Y}+\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}z_{1-\alpha/2}\right]}$$

2do Caso **Dadas dos muestras aleatorias independientes, $X_1,\ldots,X_{n_1}$, con $X_i \sim N(\mu_1,\sigma^2)$, e $Y_1,\ldots,Y_{n_2}$, con $Y_i \sim N(\mu_2,\sigma^2)$, (es decir, ambas muestras tienen la misma varianza), pero el valor de σ es desconocido:**

Repitiendo el planteo que hicimos en el primer caso:

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}} \sim N(0;1), \quad \text{saco factor común } \sigma^2 \tag{1}$$

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0;1) \quad \text{distribuyo la raíz} \tag{2}$$

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0;1) \tag{3}$$

Pero σ es desconocido, por lo tanto debemos estimarlo, y lo haremos en base a ambas muestras a partir de un "promedio" entre las varianzas muestrales de cada una:

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

Por lo tanto el pivote resulta ser

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Planteamos la probabilidad:

$$1-\alpha = P\left(-t_{n_1+n_2-2;1-\alpha/2} \leq \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \leq t_{n_1+n_2-2;1-\alpha/2}\right)$$

De esta manera, despejamos $\mu_1-\mu_2$ como en el caso anterior y obtenemos:

Intervalo de Confianza de nivel $1-\alpha$ para $\mu_1-\mu_2$:

$$\boxed{\left[\overline{X}-\overline{Y}-S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{n_1+n_2-2;1-\alpha/2}\ ;\ \overline{X}-\overline{Y}+S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{n_1+n_2-2;1-\alpha/2}\right]}$$

- **Intervalo de Confianza para la diferencia de proporciones.**

Dadas dos muestras aleatorias independientes, $X_1, \ldots, X_{n_1}$**, con** $X_i \sim Be(p_1)$**, e** $Y_1, \ldots, Y_{n_2}$**, con** $Y_i \sim Be(p_2)$**, tales que** $n_1 > 30$ **y** $n_2 > 30$**:**

Se quiere construir un intervalo de confianza de nivel $1-\alpha$ para la diferencia

$$p_1 - p_2$$

Partimos de un estimador de esta diferencia:

$$\overline{X} - \overline{Y}$$

Como

$$E\left(\overline{X} - \overline{Y}\right) = p_1 - p_2 \quad \text{y} \quad V\left(\overline{X} - \overline{Y}\right) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

Tenemos que, por el Teorema Central del Límite:

$$\frac{\overline{X} - \overline{Y} - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \overset{(a)}{\sim} N(0;1),$$

Pero, al tener valores desconocidos en el denominador, no podemos despejar la diferencia entre las proporciones, por lo tanto las podemos estimar conservando la distribución aproximada:

$$\frac{\overline{X} - \overline{Y} - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \overset{(a)}{\sim} N(0;1), \ \text{donde } \hat{p}_1 = \overline{X} \text{ y } \hat{p}_2 = \overline{Y}$$

obteniendo el pivote a partir del cual podremos plantear la siguiente probabilidad:

$$1-\alpha = P\left(-z_{1-\alpha/2} \leq \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \leq z_{1-\alpha/2}\right)$$

Despejano $p_1 - p_2$, obtenemos

Intervalo de Confianza de nivel $1-\alpha$ para $p_1 - p_2$:

$$\boxed{\left[\hat{p}_1 - \hat{p}_2 - \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\, z_{1-\alpha/2} \ ; \ \hat{p}_1 - \hat{p}_2 + \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\, z_{1-\alpha/2}\right]}$$

donde $\hat{p}_1 = \overline{X}$ y $\hat{p}_2 = \overline{Y}$