INSTITUTO TECNOLÓGICO DE COSTA RICA
ESCUELA DE MATEMÁTICA
MA 3405 ESTADÍSTICA
II SEMESTRE 2025
Profesora: Mag. Nathalia Saborío Cordero

TEC | Tecnológico de Costa Rica

ESTIMACIÓN CON DOS POBLACIONES

2. Diferencia entre dos promedios

La estimación puntual de la diferencia entre $\mu_1 - \mu_2$ está dada por la diferencia entre las dos medias muestrales $\overline{X_1} - \overline{X_2}$. Recordemos que de una población pueden tomarse muchas muestras diferentes y esto genera toda una distribución de diferencias de estas medias muestrales.

TEOREMA 1:

(**Intervalo de confianza para la diferencia de promedios si estos son normales y se conocen las varianzas poblacionales**).

Considere la población X_1 con media poblacional μ_1 y varianza poblacional σ_1^2. Considere la población X_2 con media poblacional μ_2 y varianza poblacional σ_2^2. Si $\bar{X}_1$ y $\bar{X}_2$ siguen una distribución normal para muestras de tamaño n_1 y n_2 respectivamente, y se conocen σ_1 y σ_2 entonces

a) Un intervalo de confianza del $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ tiene extremos

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\propto/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Donde $\bar{x}_1$ es el valor de $\bar{X}_1$ para una muestra de tamaño n_1, $\bar{x}_2$ es el valor de $\bar{X}_2$ para una muestra de tamaño n_2.

b) Para encontrar un intervalo de confianza del $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ con un radio menor o igual a r, el tamaño de las muestras de cada población debe ser

$$n \geq \left(\frac{z_{\propto/2} \cdot \sqrt{\sigma_1^2 + \sigma_2^2}}{r} \right)^2$$

Observación:

Recuerde que para que $\bar{X}_1$ y $\bar{X}_2$ se distribuyan normalmente se debe cumplir con una de las siguientes opciones:

- Su población sigue una distribución normal (Teorema: el promedio de normales es normal).
- El tamaño de la muestra es mayor a 30 (Teorema del límite central). En este caso se puede utilizar la desviación estándar de la muestra como una buena aproximación de la desviación estándar poblacional.

Ejercicio 1

Una universidad analizó las capacidades de rendimiento de 2 tipos de variedades de cierto producto agrícola, así obtuvo los siguientes datos utilizando parcelas de igual tamaño:

Variedad	Tamaño de la muestra	Rendimiento promedio (kg por parcela)	Desviación muestral
A	32	43	15
B	30	47	22

a) Determine un IC del 90% para la diferencia entre el rendimiento promedio de las variedades A y B.

R/$I =]-11.917\ 3, 3.917\ 32[$

b) El ministro de agricultura afirma que la variedad 2 tiene mejor rendimiento. ¿Aceptaría esta afirmación?

R) Note que I tiene valores positivos y negativos, pues 0 $\in$ I. Por lo tanto, no se considera aceptable la afirmación del ministro.

Ejercicio 2:

Un intervalo de confianza (IC) de 90% para la diferencia de promedios $(\mu_1 - \mu_2)$ es $]165.5\ , 192.9[$. Si las muestras utilizadas en el cálculo del IC son ambas de tamaño 50, determine el valor de $\bar{x}_1 - \bar{x}_2$ y de $\sigma_1^2 + \sigma_2^2$.

R/ $\bar{x}_1 - \bar{x}_2 = 179.2$
$\sigma_1^2 + \sigma_2^2 = 3468.0019$

TEOREMA 2:

(Intervalo de confianza para la diferencia de promedios si las poblaciones son normales, se desconocen las varianzas poblacionales y se suponen iguales).

Considere la población 1 dada por la variable aleatoria X_1 con media poblacional μ_1 y la población 2 dada por la variable aleatoria X_2 con media poblacional μ_2. Si las poblaciones X_1 y X_2 siguen una distribución normal y se desconocen las desviaciones poblacionales σ_1 y σ_2 pero se suponen iguales, entonces un intervalo de confianza del $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ tiene extremos

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\upsilon} \cdot \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Donde $\bar{x}_1, s_1$ son valores de $\bar{X}_1$ y S_1 para una muestra de tamaño n_1. Además $\bar{x}_2, s_2$ son valores de $\bar{X}_2$ y S_2 para una muestra de tamaño n_2.

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

$t_{\alpha/2,\upsilon}$ es el valor de la distribución t con υ grados de libertad ($\upsilon = n_1 + n_2 - 2$).

TEOREMA 3:

(Intervalo de confianza para la diferencia de promedios si las poblaciones son normales, se desconocen las varianzas poblacionales y NO se suponen iguales).

Considere la población 1 dada por la variable aleatoria X_1 con media poblacional μ_1 y la población 2 dada por la variable aleatoria X_2 con media poblacional μ_2. Si las poblaciones X_1 y X_2 siguen una distribución normal y se desconocen las desviaciones poblacionales σ_1 y σ_2 pero no se suponen iguales, entonces un intervalo de confianza del $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ tiene extremos

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\upsilon} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Donde $\bar{x}_1, s_1$ son valores de $\bar{X}_1$ y S_1 para una muestra de tamaño n_1 ; $\bar{x}_2, s_2$ son valores de $\bar{X}_2$ y S_2 para una muestra de tamaño n_2.

$$\upsilon = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

$t_{\alpha/2,\upsilon}$ es el valor de la distribución $t\ de\ Student$ con υ grados de libertad.

Ejercicio 3:

Una ferretería tiene dos marcas de pintura color verde para portones: A y B. El vendedor asegura que la pintura B es más cara por que seca más rápidamente que la pintura A. Se obtienen mediciones de ambos tipos de pintura. Los tiempos de secado (en minutos) son los siguientes:

Pintura A:	*120*	*132*	*123*	*122*	*140*	*110*	*120*	*107*		
Pintura B:	*126*	*124*	*116*	*125*	*109*	*130*	*125*	*117*	*129*	*120*

1. Encuentre un intervalo de confianza del 95% para la diferencia entre los tiempos de secado promedio, suponiendo que las desviaciones estándar de éstos son iguales y que el tiempo de secado de ambas pinturas está distribuido de manera normal.

Sean

X_1: tiempo de secado de la pintura A
X_2: tiempo de secado de la pintura B

R/]−9. 008 89, 8. 308 89[

2. ¿Existe alguna evidencia que respalde la afirmación del vendedor?

No, el IC tiene valores positivos y negativos.

Note que hay un 95% de probabilidad de que $\mu_1 - \mu_2 \in]-9.\ 008\ 89,\ 8.\ 308\ 89[$. Si el IC fuera positivo entonces es muy probable que $(\mu_1 - \mu_2) > 0$ y entonces $\mu_1 > \mu_2$, así se tendría que la pintura A secaría más rápido que la B, en promedio. Similarmente si el IC fuera negativo, la pintura B secaría más rápido que la A, en promedio, lo cual respaldaría la afirmación del vendedor. Sin embargo, el IC determinado contiene al cero, por lo que no se puede inferir ningún resultado.

Observación:

Es importante diferenciar entre un IC para el **promedio de una diferencia** y un IC para la **diferencia de promedios**. El primer caso se refiere a un IC para un promedio visto anteriormente (cuando se estudió para una población), en este caso las variables involucradas se ***funden en una sola variable***, la variable diferencia, de la cuál interesa estimar su promedio, y para la muestra se debe realizar observaciones pareadas (es decir para cada individuo elegido en la muestra se determina el valor de las variables involucradas con el fin de obtener un valor de la variable diferencia).

El segundo caso, se trata del IC estudia en esta sección, en el cuál las ***poblaciones son independientes*** por lo que no es necesario que las observaciones sean pareadas.

Ejercicio 4:

Las calificaciones obtenidas por 10 estudiantes en los cursos de Probabilidad y Estadística son las siguientes:

Probabilidad	70	30	30	73	60	75	65	60	40	55
Estadística	100	85	90	95	90	95	85	85	80	90

Suponiendo que la distribución de las notas es normal:

a) Encuentre un intervalo de confianza del 95% para el promedio de diferencias entre la nota de probabilidad y de estadística.

R/Sea $d = Nota_{est} - Nota_{prob}, el\ IC\ para\ d\ es\ ]23.577\ ,43.8217[$

b) Encuentre un intervalo de confianza del 95% para la diferencia entre la nota de estadística y la de probabilidad.

R/El IC para $\mu_{est} - \mu_{prob}$: $]21.228\ ,46.172[$

c) ¿Aceptaría el hecho de que el rendimiento promedio en probabilidad es menor que el rendimiento promedio en estadística?

R/ Sí

Ejercicio 5:

Transfer Trucking trasporta remesas entre Chicago y Kansas City por dos rutas. Una muestra de 100 camiones enviados por la ruta del norte reveló un tiempo promedio de tránsito de $\bar{x}_N$=17.2 horas con una desviación estándar de $\bar{s}_N$ =5.3 horas, mientras que 75 camiones que utilizan la ruta del sur necesitaron un promedio de $\bar{x}_S$=19.4 horas con una desviación estándar de $\bar{s}_s$ = 4.5 horas. El despachador de Transfer Trucking, desea desarrollar un intervalo de confianza del 95% para la diferencia en el tiempo promedio entre estas dos rutas alternas. ¿Cuál sería dicho intervalo?

R/$]-3{,}654, -0.745[$

PRUEBA DE HIPÓTESIS CON DOS PROMEDIOS

Tomar en consideración la siguiente notación:

Población	PARÁMETROS		Tamaño de la muestra	ESTADÍSTICA	
	Media poblacional	Varianza poblacional		Media muestral	Varianza muestral
$\boldsymbol{X_1}$	μ_1	σ_1^2	n_1	$\bar{X}_1$	s_1^2
$\boldsymbol{X_2}$	μ_2	σ_2^2	n_2	$\bar{X}_2$	s_2^2

$\bar{X}_1 - \bar{X}_2$ es un estimador insesgado de $\mu_1 - \mu_2$, además $Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$.
La hipótesis nula para este tipo de pruebas es de la forma $\boldsymbol{H_0: \mu_1 - \mu_2 = d_0}$, donde d_0 se conoce como ***diferencia nula.*** Además, se sabe que:

1) Si $\bar{X}_1$ y $\bar{X}_2$ siguen una distribución normal para muestras de tamaño n_1 y n_2 respectivamente, y se conocen σ_1 y σ_2 entonces

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

HAY QUE RECORDAR QUE → Esta opción se utiliza en alguno de los siguientes casos:

a) n_1 y n_2 son mayores o iguales a 30. No es necesario conocer σ_1 y σ_2 pues se cumple que $\sigma_1 \approx s_1$ y $\sigma_2 \approx s_2$.
b) Las poblaciones $\bar{X}_1$ y $\bar{X}_2$ siguen una distribución normal y se conoce σ_1 y σ_2.
c) La población $\bar{X}_1$ sigue una distribución normal, $n_2 \geq 30$ y se conoce σ_1. No es necesario conocer σ_2 pues $\sigma_2 \approx s_2$.
d) La población $\bar{X}_2$ sigue una distribución normal, $n_1 \geq 30$ y se conoce σ_2. No es necesario conocer σ_1 pues $\sigma_1 \approx s_1$.

2) Si las poblaciones X_1 y X_2 siguen una distribución normal y se desconocen las desviaciones poblacionales σ_1 y σ_2 pero **se suponen iguales,** entonces

$$T = \frac{\bar{X}_1 - \bar{X}_2 - d_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

sigue una distribución $t\ de\ Student$ con $v = n_1 + n_2 - 2$ grados de libertad, donde s_p^2 es una estimación dada por el estadístico:

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

3) Si las poblaciones X_1 y X_2 siguen una distribución normal y se desconocen las desviaciones poblacionales σ_1 y σ_2 pero **no se suponen iguales**, entonces

$$T = \frac{\bar{X}_1 - \bar{X}_2 - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

sigue una distribución $t\ de\ Student$ con v grados de libertad, donde

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Ejemplo:

Se obtuvieron las estaturas de 20 mujeres y 30 hombres seleccionados aleatoriamente de la población de alumnos de cierta escuela. En la siguiente tabla se resumen los datos encontrados

	Número	Media	Desviación
Masculino (m)	30	69.8	1.92
Femenino (f)	20	63.8	2.18

Suponga que la estatura de las mujeres y la estatura de los hombres se distribuyen normalmente. ¿Puede concluirse, con un nivel de significancia de 4%, que el promedio de estaturas de los hombres de la escuela supera en más de 3cm el promedio de las mujeres? Suponga que $\sigma_1 = \sigma_2$.

SOLUCIÓN:

En este caso la afirmación es $u_m - u_f > 3$ *y las hipótesis son*

$$H_0 : u_m - u_f = 3 \ (\leq), \quad H_1 : u_m - u_f > 3$$

Datos:

$$n_1 = 30, \quad n_2 = 20, \quad \overline{x}_1 = 69.8, \quad \overline{x}_2 = 63.8, \quad s_1^2 = 1.92, \quad s_2^2 = 2.18.$$

Como las poblaciones son normales, se desconocen las varianzas poblacionales, $n_2 < 30$ *y se supone que* $\sigma_1 = \sigma_2$ *entonces*

$$T = \frac{\overline{X}_1 - \overline{X}_2 - d_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \sim t\,(v)$$

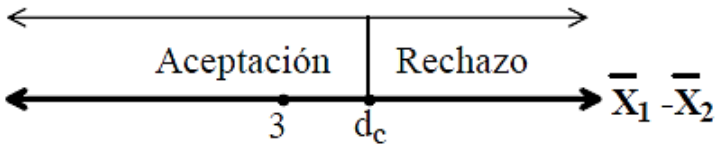
donde el valor de S_p^2 *observado es*

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2} = \frac{19 \cdot 2.18^2 + 29 \cdot 1.92^2}{48} = 4.1084$$

Varias formas de resolver el problema:

1

Utilizando regiones con estadístico $\overline{X}_1 - \overline{X}_2$. *Las regiones para* $\overline{X}_1 - \overline{X}_2$ *se definen a continuación*



Recuerde que para hallar el valor crítico se hace uso del nivel de significancia $\alpha = 0.04$:

$$0.04 = P(H_1|H_0) = P\left(\overline{X}_1 - \overline{X}_2 > d_c | u_m - u_f = 3\right)$$

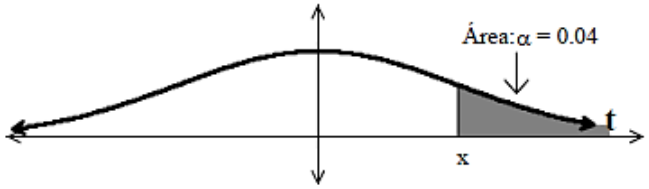
$$\Longrightarrow 0.04 = P\left(t > \frac{d_c - 3}{\sqrt{s_p^2/n_1 + s_p^2/n_2}}\right) = P\left(t > \frac{d_c - 3}{0.585\,121}\right)$$

Note que $x = \dfrac{d_c - 3}{0.585\,121}$ *es positivo pues* $d_c > 3$:

$$\Longrightarrow \frac{d_c - 3}{0.585\,121} = t_{0.96,48}$$

Como $t_{0.96,48} = -t_{0.04,48} = 1.78854$

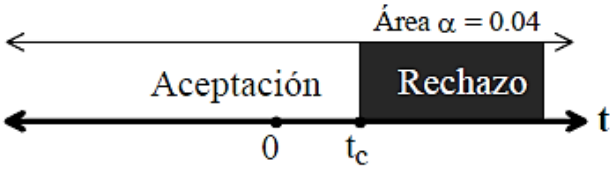
Entonces $d_c = 4.04648$



Como el valor observado $\overline{x}_1 - \overline{x}_2 = 69.8 - 63.8 = 6$ *está en la región de rechazo se rechaza* H_0.

2

Utilizando el enfoque clásico con estadístico T. Las regiones para t se definen a continuación



El valor t crítico es

$$t_c = t_{0.96,48} = -t_{0.04,48} = 1.78854$$

por lo tanto la región de aceptación para t es $]-\infty, 1.78854[$ *y de rechazo* $]1.78854, +\infty[$. *Por otro lado el valor t observado es*

$$t_{obs} = \frac{69.8 - 63.8 - 3}{\sqrt{\dfrac{4.108\,4}{20} + \dfrac{4.108\,4}{30}}} = 5.1271$$

Como $t_{obs} \in]1.78854, +\infty[$ *entonces se rechaza* H_0.

3 *Utilizando el valor P y estadístico* $\overline{X}_1 - \overline{X}_2$. *Dado que* $H_1 : u_m - u_f > 3$ *y el valor observado del estadístico* $\overline{X}_1 - \overline{X}_2$ *es*

$$d_{obs} = \overline{x}_1 - \overline{x}_2 = 6$$

entonces se tiene que

$$\begin{aligned} Valor\ P &= P\left(\overline{X}_1 - \overline{X}_2 \geq 6 | u_m - u_f = 3\right) \\ &= P\left(t \geq \frac{6-3}{0.5851}\right) = P\left(t \geq 5.12733\right) \end{aligned}$$

Utilizando la tabla de t *con* $v = 48$ *se tiene que*

$$Valor\ P = P\left(t \geq 5.12733\right) < 0.005 < \alpha = 0.04.$$

Por lo tanto se rechaza H_0.

4 *Utilizando el valor P y estadístico T. Bajo la hipótesis nula* $H_0 : u_m - u_f = 3$, *se tiene que el valor observado de T es*

$$t_{obs} = \frac{\overline{x}_1 - \overline{x}_2 - d_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{69.8 - 63.8 - 3}{\sqrt{\frac{4.1084}{20} + \frac{4.1084}{30}}} = 5.1271$$

y entonces

$$Valor\ P = P\left(T \geq t_{obs}\right) = P\left(t \geq 5.12733\right) < 0.005 < \alpha = 0.04$$

por lo tanto se rechaza H_0.

Así, en los 4 casos se rechaza H_0 *y se acepta la afirmación. Puede concluirse que no hay evidencia en contra de que el promedio de estaturas de los hombres de la escuela supera en más de 3cm el promedio de las mujeres*

Teorema:

Tomando muestras del mismo tamaño $n_1 = n_2 = n$ de modo que $\bar{X}_1 - \bar{X}_2$ se distribuya aproximadamente normal, para probar $\boldsymbol{H_0}: \boldsymbol{\mu_1 - \mu_2 = d_0}$ con un nivel de significancia de α y una potencia mínima de $1 - \beta$ para la hipótesis alternativa específica $H'_1: \boldsymbol{\mu_1 - \mu_2 = d_1}$ puede tomarse una muestra de tamaño

$n \geq \frac{(|z_\alpha| + |z_\beta|)^2 (\sigma_1^2 + \sigma_2^2)}{(d_1 - d_0)^2}$ si la prueba es de una cola

$n \geq \frac{(|z_{\alpha/2}| + |z_\beta|)^2 (\sigma_1^2 + \sigma_2^2)}{(d_1 - d_0)^2}$ si la prueba es de dos colas

Ejercicio 6:

Una muestra de países latinoamericanos y europeos revelo los siguientes resultados, donde n es el número de países, $\bar{x}$ el promedio de las esperanzas de vida en años y s la desviación estándar de las esperanzas de vida en años:

Países	n	$\bar{x}$	s
A	35	76.16	1.85
B	31	60.7	1.08

Pruebe la hipótesis de que la esperanza de vida en los países Europa es superior a la de los países Latinoamericanos en por lo menos 16 años.

R/ Valor P ≈ 0.07358, se acepta la afirmación.

Ejercicio 7:

Sea desea investigar la duración de dos tipos A y B de baterías AA no recargables, las cuales tiene precios similares. Un estudiante, que utiliza mucho batería AA, señala que la duración promedio de las baterías A supera en más de 10 minutos a la duración promedio de las baterías B. Se tomaron muestras de duraciones de ambos tipos de baterías, la información se resume en la siguiente tabla

Batería AA	Tamaño de la muestra	$\bar{x}$	s
Tipo A	21	5.3 horas	0.8 horas
Tipo B	19	5.1 horas	0.6 horas

Suponga que las varianzas son iguales. Con un nivel de significancia de 10%, ¿Es aceptable la afirmación del estudiante?

R/ $t_{obs} = 0.1478,\ t_c = 1.30423,\ \ valor\ P \approx 0{,}4416$, hay evidencia en contra de la afirmación.

Ejercicio 8:

Los siguientes datos se refieren a las unidades vendidas en una semana de un nuevo dispositivo para cargas inalámbricas en teléfonos móviles. Se consideraron locales de San José y de Cartago

San José	59	68	44	71	63	46	69	54	48
Cartago	50	36	62	52	70	41	-	-	-

Un ejecutivo de la empresa proveedora de este dispositivo afirma que, si bien es cierto que las ventas promedio semanales son mejores en San José, la diferencia entre estas está por debajo de 8 unidades. Suponiendo que las varianzas son iguales, ¿Respaldan las evidencias esta afirmación, con significancia del 3%?

R/La hipótesis nula se tolera, por tanto, los datos no respaldan la afirmación del ejecutivo de la empresa.

Texto base utilizado: Sanabria Brenes, Giovanni, 2011. **Comprendiendo la Estadística Inferencial**.