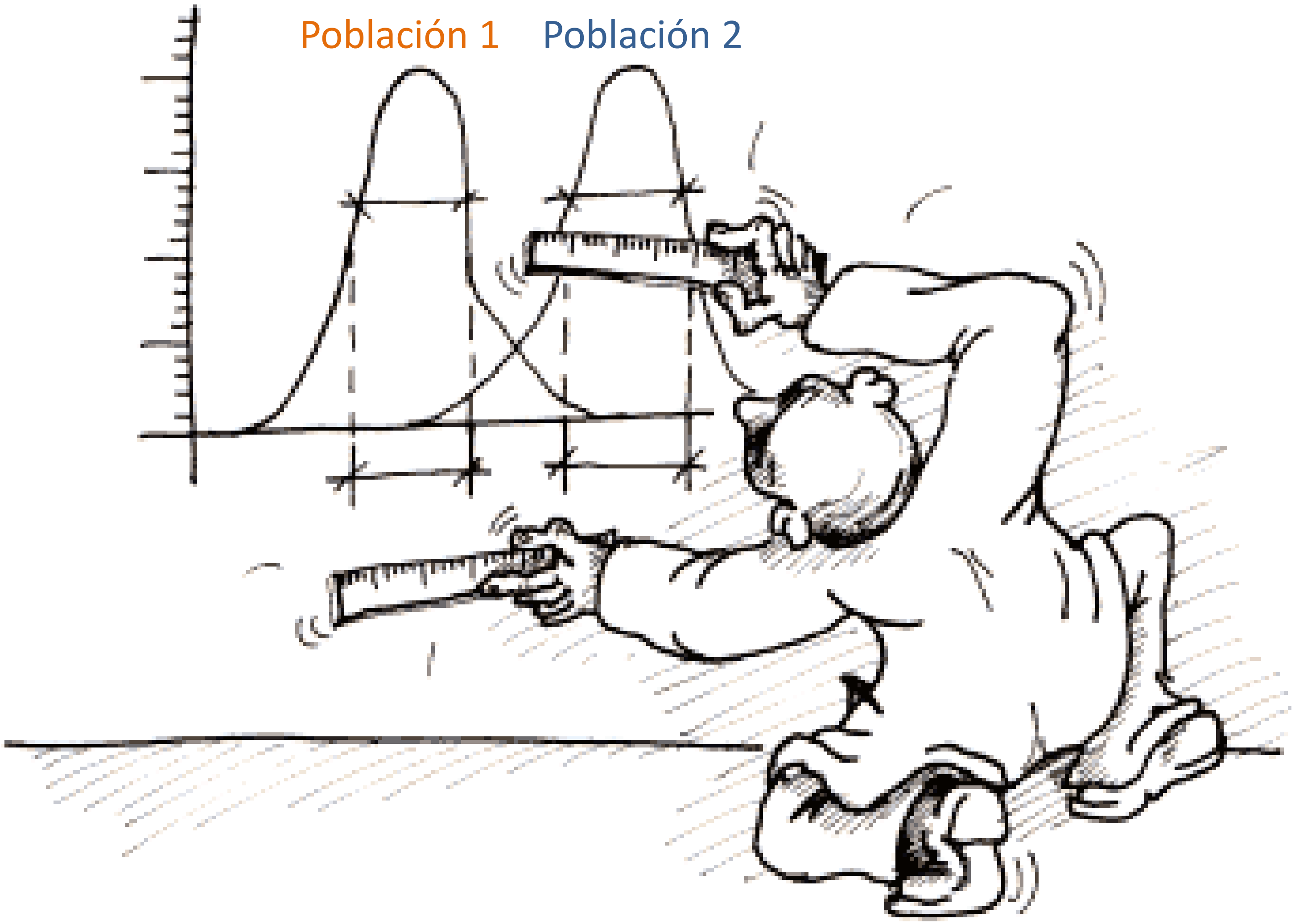
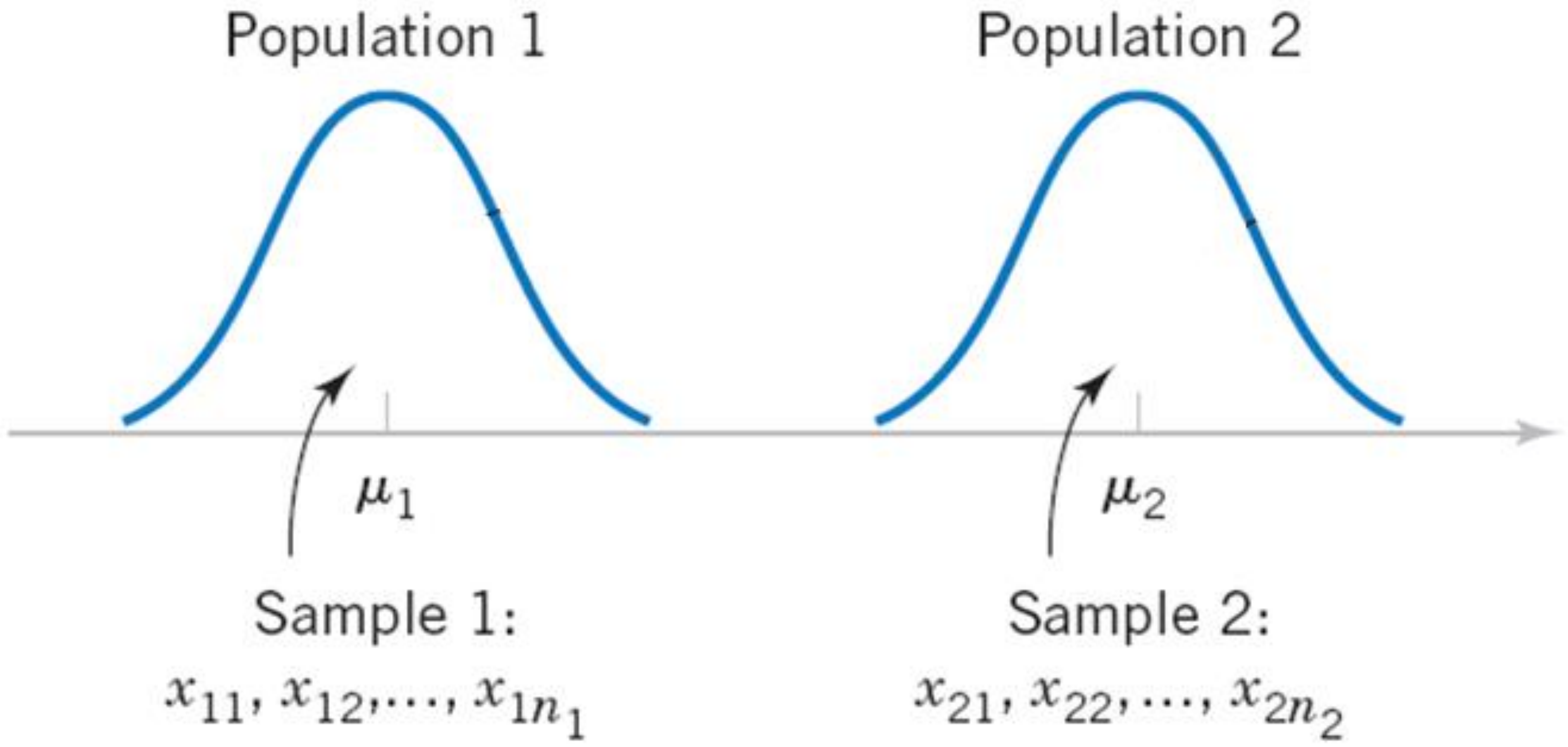
Clase 18: PH para el cociente de varianzas de poblaciones normales. PH para diferencia de medias de poblaciones normales.

Universidad Nacional de Colombia – sede Medellín

Comparación de varianzas de poblacionales normales



Prueba de hipótesis para diferencia de medias $\mu_1 - \mu_2$



Prueba de hipótesis para cociente de varianzas

Queremos probar:

$$H_0\colon \sigma_1^2 = \sigma_2^2 \qquad H_0\colon \sigma_1^2 = \sigma_2^2 \qquad H_0\colon \sigma_1^2 = \sigma_2^2$$
$$H_a\colon \sigma_1^2 < \sigma_2^2 \qquad H_a\colon \sigma_1^2 \neq \sigma_2^2 \qquad H_a\colon \sigma_1^2 > \sigma_2^2$$

El estadístico está dado por:

$$F_0 = \frac{S_1^2}{S_2^2} \sim f_{n_1-1,n_2-1}$$

$valor\text{–}P$

- Si $H_a\colon \sigma_1^2 < \sigma_2^2$

  $$\text{Valor} - \text{P} = \text{Prob}(F_{n_1-1,n_2-1} \leq f_0)$$

- Si $H_a\colon \sigma_1^2 \neq \sigma_2^2$

  $$\text{Valor} - \text{P} = 2\ \times\ min\{\text{P}\left(F_{n_1-1,n_2-1} \leq f_0\right), \text{P}(F_{n_1-1,n_2-1} \geq f_0)\}$$

- Si $H_a\colon \sigma_1^2 > \sigma_2^2$

  $$\text{Valor} - \text{P} = \text{Prob}(F_{n_1-1,n_2-1} \geq f_0)$$

Regiones de rechazo

- Si $H_a: \sigma_1^2 < \sigma_2^2$

  Se rechaza H_0 si $f_0 < f_{1-\alpha, n_1-1, n_2-1}$

- Si $H_a: \sigma_1^2 \neq \sigma_2^2$

  Se rechaza H_0 si $f_0 > f_{\alpha/2, n_1-1, n_2-1}$ o si $f_0 < f_{1-\alpha/2, n_1-1, n_2-1}$

- Si $H_a: \sigma_1^2 > \sigma_2^2$

  Se rechaza H_0 si $f_0 > f_{\alpha, n_1-1, n_2-1}$

Equivalencia

$$H_0 : \sigma_1^2 / \sigma_2^2 = 1 \qquad \Longleftrightarrow \qquad H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_a : \sigma_1^2 / \sigma_2^2 \neq 1 \qquad \qquad H_a : \sigma_1^2 \neq \sigma_2^2$$

Ejemplo 1

Se realiza un estudio para comparar dos tratamientos que se aplicarán a frijoles crudos con el objetivo de reducir el tiempo de cocción. El tratamiento T1 es a base de bicarbonato de sodio, el T2 es a base de cloruro de sodio o sal común. La variable respuesta es el tiempo de cocción en minutos. Los datos se muestran en la tabla. ¿Son las varianzas de los tiempos iguales o diferentes? Usar $\alpha = 3\%$.

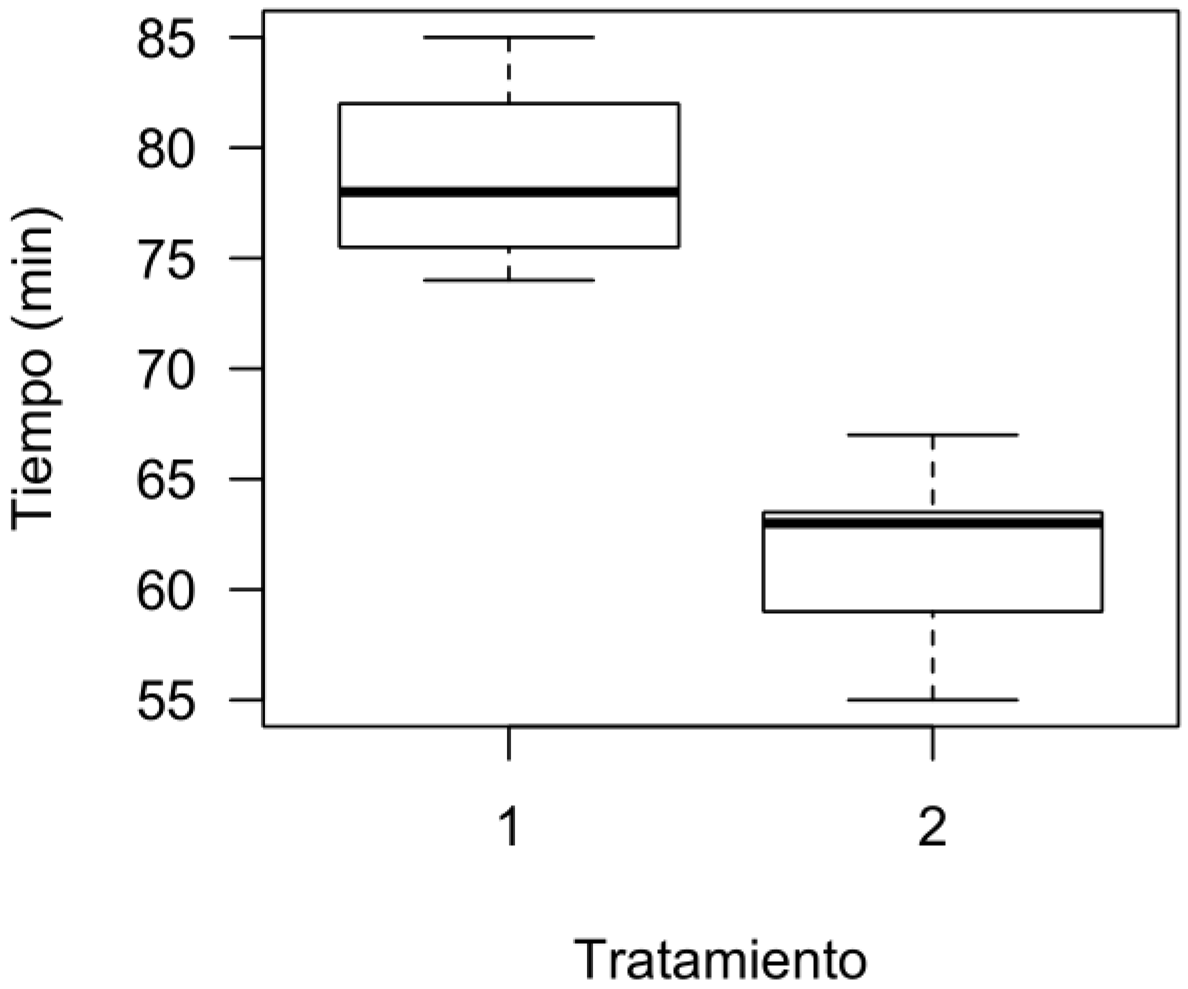
Ejemplo 1

Tratamiento	Tiempo (min)						
T1	76	85	74	78	82	75	82
T2	57	67	55	64	61	63	63

T1: bicarbonato

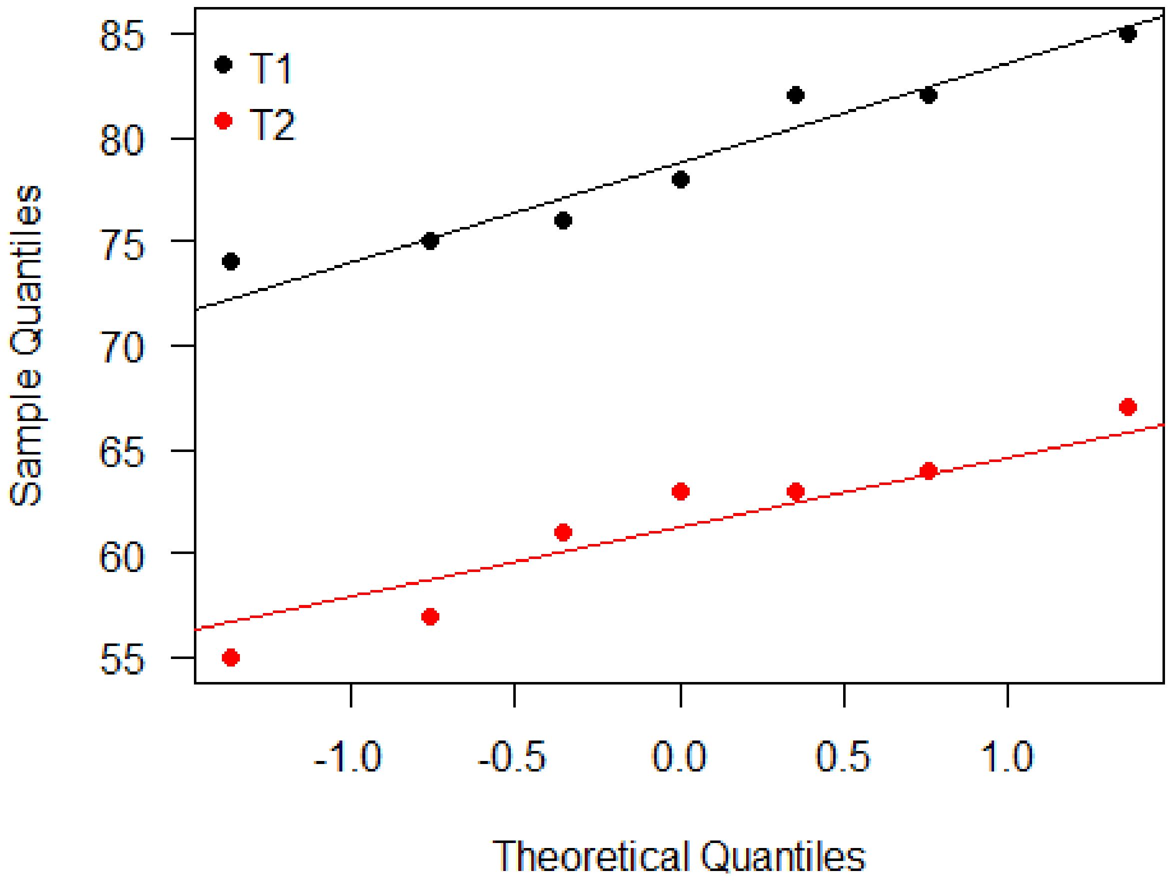
T2: sal

Ejemplo 1



Ejemplo 1

Paso 0. ¿Son las poblaciones normales?



Al aplicar la prueba de normalidad Kolmogorov-Smirnov se encontró que:

$valor\text{−}P$ para T1 fue 0.5205

$valor\text{−}P$ para T2 fue 0.3953

Ejemplo 1

Paso 1. Definir las hipótesis.

$$H_0:\ \sigma_{T1}^2 = \sigma_{T2}^2$$
$$H_a: \sigma_{T1}^2 \neq \sigma_{T2}^2$$

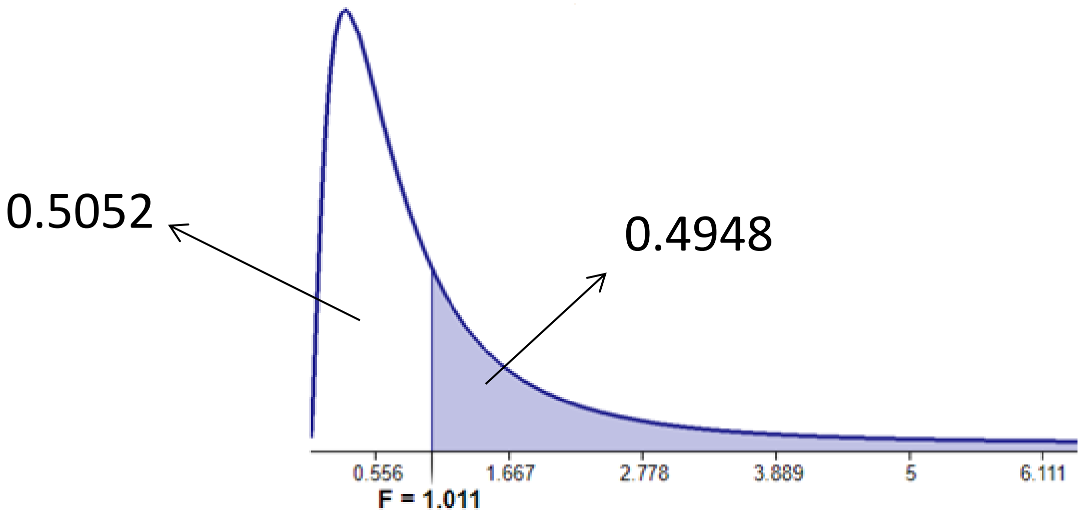
Paso 2. Calcular el estadístico

$$f_0 = \frac{s_{T1}^2}{s_{T2}^2} = 1.0111019$$

Ejemplo 1

Paso 3. Calcular el $valor{-}P$

En R `pf(q=1.011, df1=6, df2=6, lower.tail=FALSE)`



Distribución F con df1=6 y df2=6

$$\text{Valor} - \text{P} = 2\ \times\ min\{\text{P}(f_{n_1-1,n_2-1} \leq f_0)\ , \text{P}(f_{n_1-1,n_2-1} \geq f_0)\ \}$$

$$\text{Valor} - \text{P} = 2\ \times\ min\{0.5052, 0.4948\} = 2\ \times 0.4948 = 0.9896$$

Ejemplo 1

Paso 4. Tomar decisión

Como el $valor{-}P$ es 0.9896 no se puede rechazar la hipótesis nula y por lo tanto se concluye que las varianzas son iguales, es decir, no hay evidencias muestrales suficientes para pensar que no es cierto que $\sigma_{T1}^2 = \sigma_{T2}^2$.

Ejemplo 2

El arsénico en agua potable es un posible riesgo para la salud. Un artículo reciente reportó concentraciones de arsénico en agua potable en partes por billón (ppb) para 10 comunidades urbanas y 10 comunidades rurales. Los datos son los siguientes:

Urbana: 3, 7, 25, 10, 15, 6, 12, 25, 15, 7

Rural: 48, 44, 40, 38, 33, 21, 20, 12, 1, 18

Ejemplo 2

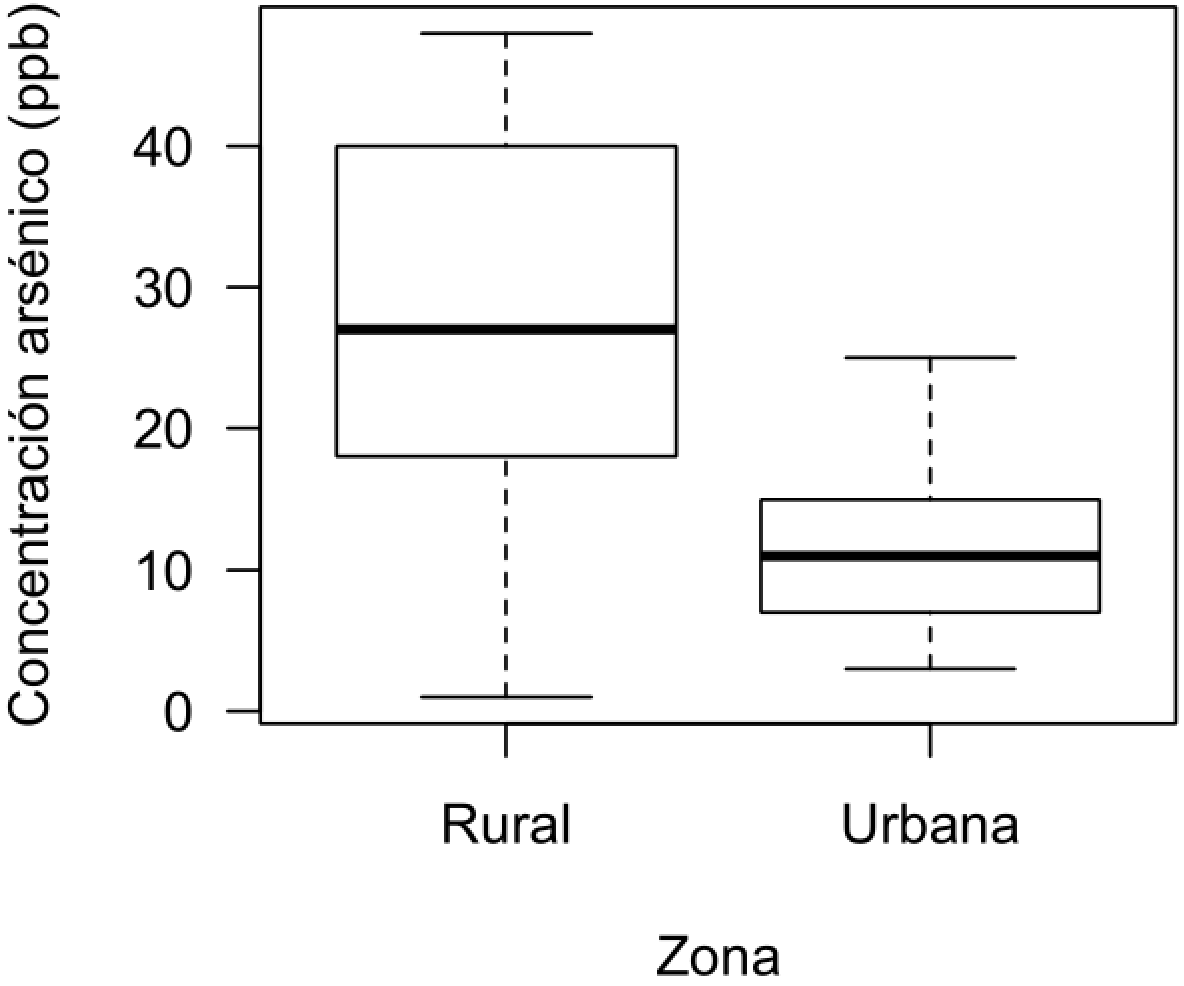
Considerando los datos, ¿se podría concluir que las varianzas del contenido de arsénico son diferentes para las dos ubicaciones?

Aplicar una prueba de hipótesis con $\alpha = 7\%$.

Urbana

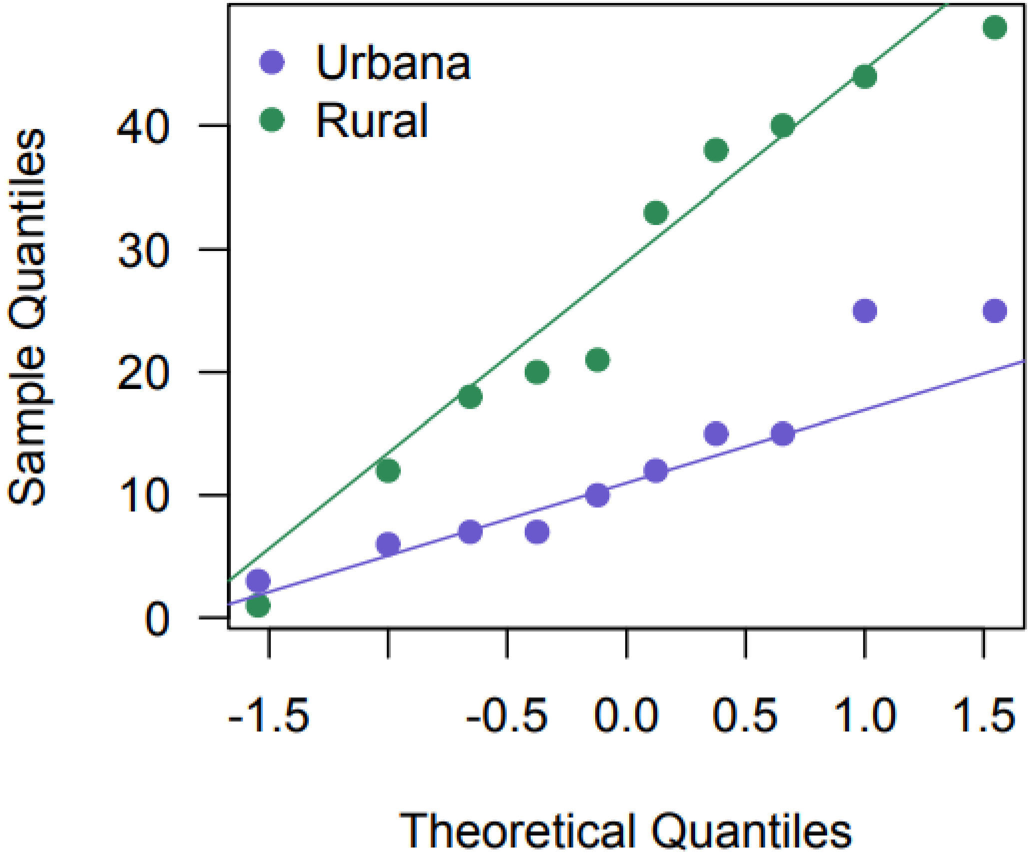
Rural

Ejemplo 2



Ejemplo 2

Paso 0. ¿Son las poblaciones normales?



Al aplicar la prueba de normalidad Kolmogorov-Smirnov
se encontró que:

$valor\text{−}P$ para urbana 0.5522

$valor\text{−}P$ para rural 0.6250

Ejemplo 2

Paso 1. Definir las hipótesis

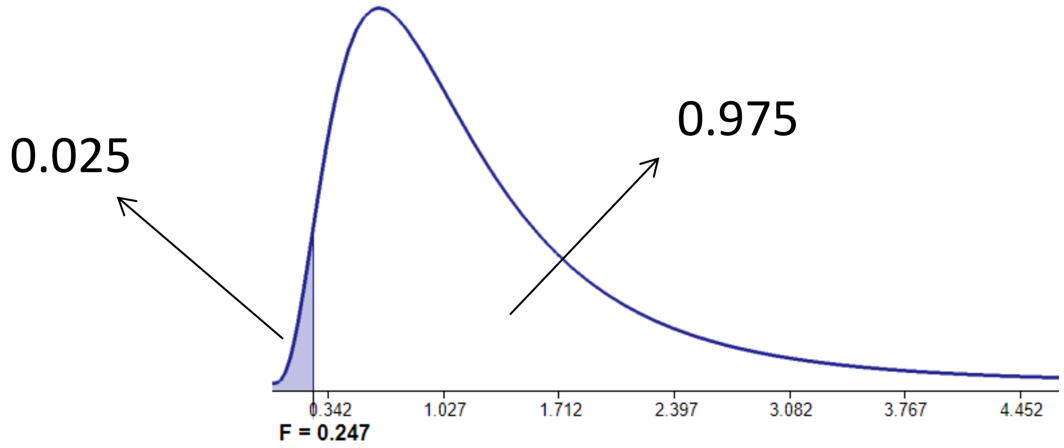
$$H_0\text{: } \sigma^2_{Urb} = \sigma^2_{Rur}$$
$$H_a\text{: } \sigma^2_{Urb} \neq \sigma^2_{Rur}$$

Paso 2. Calcular el estadístico

$$f_0 = \frac{s^2_{Urb}}{s^2_{Rur}} = 0.2473473$$

Ejemplo 2

Paso 3. Calcular el $valor\text{–}P$



Distribución F con df1=9 y df2=9

$$\text{Valor} - \text{P} = 2 \times min\{\text{P}(f_{n_1-1,n_2-1} \leq f_0)\,, \text{P}(f_{n_1-1,n_2-1} \geq f_0)\,\}$$

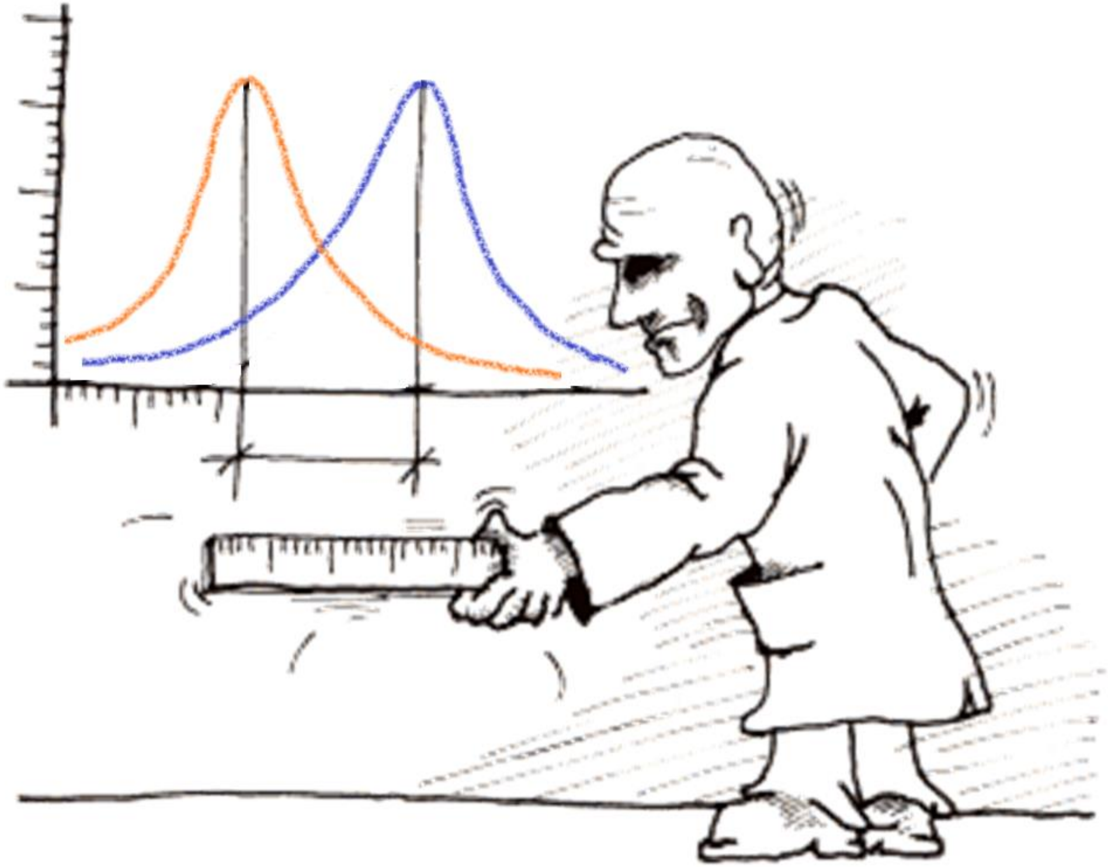
$$\text{Valor} - \text{P} = 2 \times min\{0.025, 0.975\} = 2 \times 0.025 = 0.050$$

Ejemplo 2

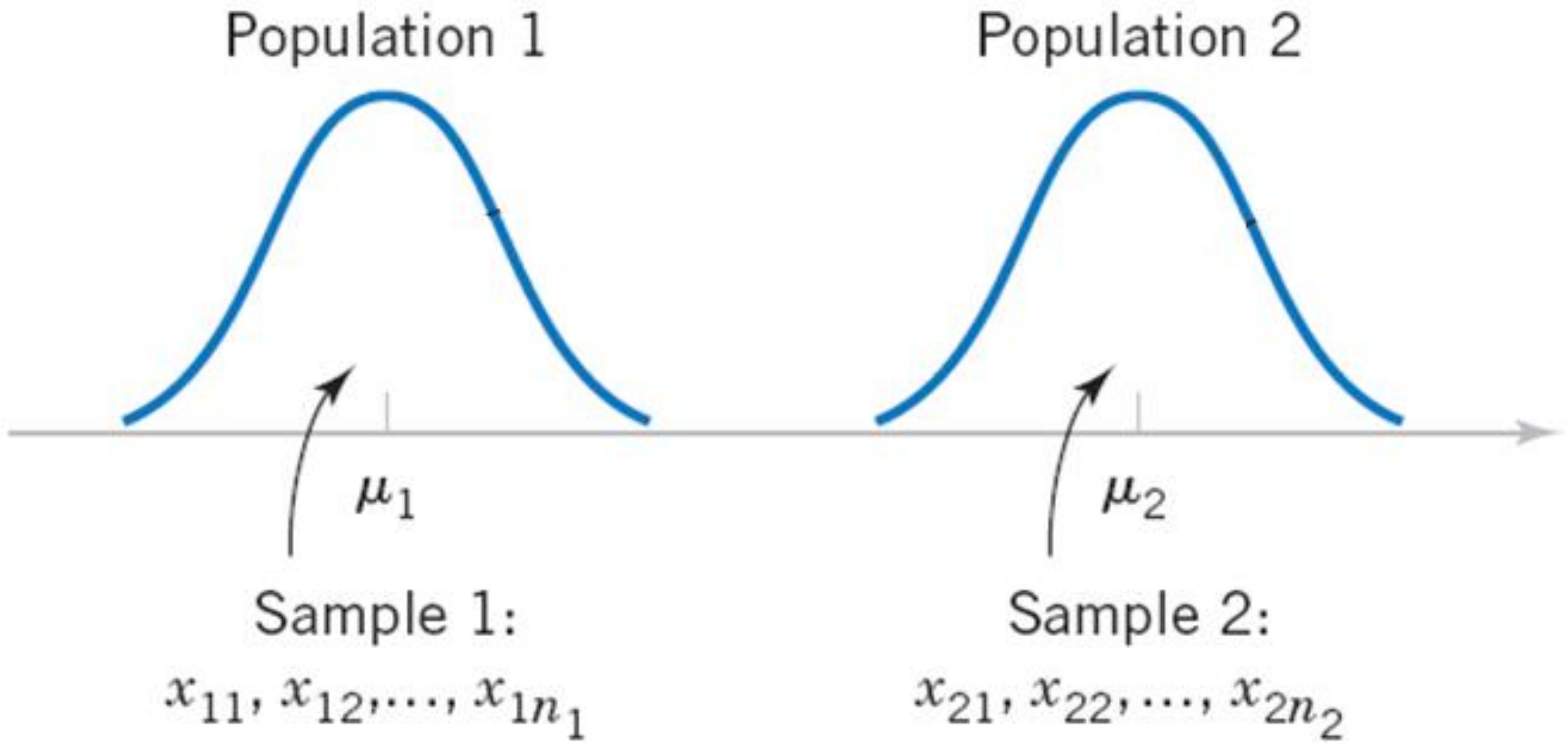
Paso 4. Conclusión: como el $valor{-}P$ es 0.05 y el nivel de significancia era 7%, se rechaza la hipótesis nula y por lo tanto se concluye que las varianzas son diferentes, es decir, hay evidencias muestrales suficientes para pensar que no es cierto que $\sigma_{Urb}^2 = \sigma_{Rur}^2$.

Comparación de medias de poblacionales normales

Población 1 Población 2



Prueba de hipótesis para diferencia de medias $\mu_1 - \mu_2$



Caso 1: Prueba de hipótesis para diferencia de medias $\mu_1 - \mu_2$ con poblaciones normales y varianzas iguales

Se quiere estudiar:

$$H_0\colon \mu_1 - \mu_2 = \delta_0 \qquad H_0\colon \mu_1 - \mu_2 = \delta_0 \qquad H_0\colon \mu_1 - \mu_2 = \delta_0$$
$$H_a\colon \mu_1 - \mu_2 < \delta_0 \qquad H_a\colon \mu_1 - \mu_2 \neq \delta_0 \qquad H_a\colon \mu_1 - \mu_2 > \delta_0$$

El estadístico está dado por:

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

donde $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ y $t_0 \sim t_{n_1+n_2-2}$

Two simple t-Test

$valor\text{–}P$

- Si $H_a: \mu_1 - \mu_2 < \delta_0$

$$\mathrm{Valor-P = Prob}(t_{n_1+n_2-2} \leq t_0)$$

- Si H_a: $\mu_1 - \mu_2 \neq \delta_0$

$$\mathrm{Valor-P} = 2 \times \mathrm{Prob}(t_{n_1+n_2-2} \geq |t_0|)$$

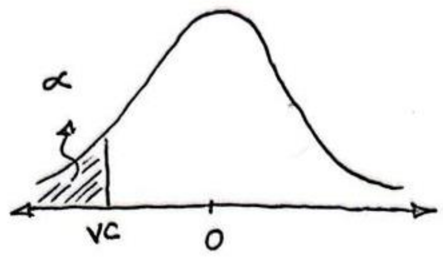
- Si H_a: $\mu_1 - \mu_2 > \delta_0$

$$\mathrm{Valor-P = Prob}(t_{n_1+n_2-2} \geq t_0)$$

Regiones de rechazo

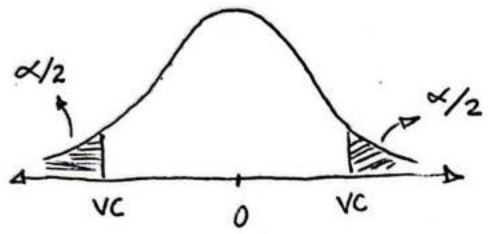
- Si H_a: $\mu_1 - \mu_2 < \delta_0$

  Se rechaza H_0 si $t_0 < -t_{\alpha,n_1+n_2-2}$



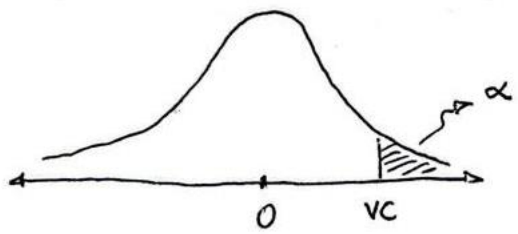
- Si H_a: $\mu_1 - \mu_2 \neq \delta_0$

  Se rechaza H_0 si $|t_0| > t_{\alpha/2,n_1+n_2-2}$



- Si H_a: $\mu_1 - \mu_2 > \delta_0$

  Se rechaza H_0 si $t_0 > t_{\alpha,n_1+n_2-2}$



Ejemplo

Se realiza un estudio para comparar dos tratamientos que se aplicarán a frijoles crudos con el objetivo de reducir el tiempo de cocción. El tratamiento T1 es a base de bicarbonato de sodio, el T2 es a base de cloruro de sodio o sal común. La variable respuesta es el tiempo de cocción en minutos. Los datos se muestran en la tabla. ¿Son las varianzas de los tiempos iguales o diferentes? Usar $\alpha = 3\%$.

Nota: de un ejemplo anterior se sabe que los datos provienen de poblaciones normales con varianzas iguales.

Ejemplo

Tratamiento	Tiempo (min)						
T1	76	85	74	78	82	75	82
T2	57	67	55	64	61	63	63

T1: bicarbonato

T2: sal

Ejemplo

¿Existen diferencias significativas entre los tiempos de cocción de los fríjoles con T1 y T2?

T1: bicarbonato

T2: sal

Tiempos promedios

78.86 min

61.43 min

Ejemplo

Paso 1. Definir las hipótesis de interés

$$H_0\colon \mu_{T1} - \mu_{T2} = 0$$
$$H_a\colon \mu_{T1} - \mu_{T2} \neq 0$$

T1: bicarbonato

T2: sal

Ejemplo

Paso 2. Definir el nivel de confianza, $\alpha = 3\%$

Paso 3. Calcular los estadísticos muestrales

Para tratamiento T1	Para tratamiento T2
$\bar{x}_1 = 78.86$	$\bar{x}_2 = 61.43$
$s_1^2 = 17.48$	$s_2^2 = 17.29$
$n_1 = 7$	$n_2 = 7$

Ejemplo

Paso 4. Calcular el valor de S_p^2

$$S_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{(7-1)17.48+(7-1)17.29}{7+7-2} = 17.39$$

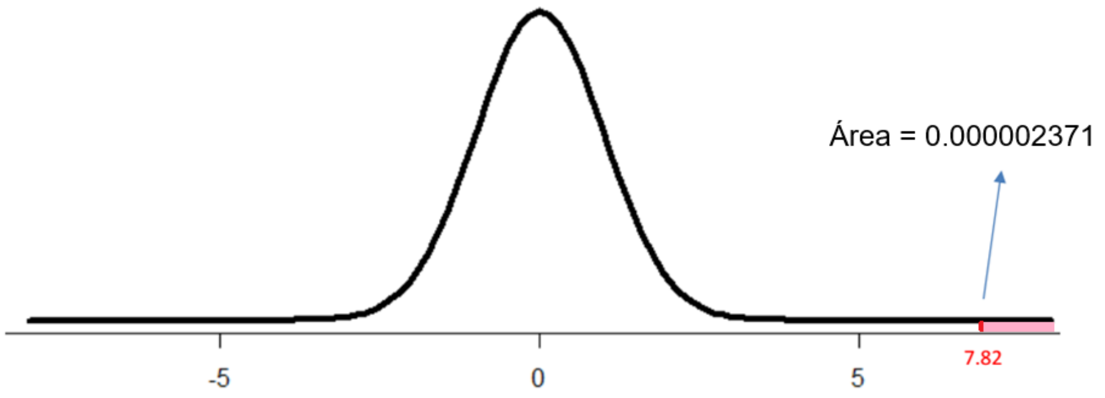
Así el estadístico es

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{78.86-61.43-0}{4.17\sqrt{\frac{1}{7}+\frac{1}{7}}} = 7.82$$

Ejemplo

Paso 5. Calcular el Valor-P

Como $t_0 = 7.82$ se busca el área a la derecha de una t-student con $n_1 + n_2 - 2$ grados de libertar, es decir, con 12 grados de libertad. En R `pt(q=7.82, df=12, lower.tail=FALSE)`



Como la prueba es bilateral entonces $ValorP = 2 \times Área < 0.0010$

Ejemplo

Paso 5. Conclusión:

- Como el $ValorP < 0.0010$ y $\alpha = 3\%$, entonces se concluye que se puede rechazar la hipótesis nula.

- En otras palabras, hay evidencia de que los tiempo promedios de cocción no son iguales.

- Como $\bar{x}_1 = 78.86$ y $\bar{x}_2 = 61.43$ entonces se puede afirmar que los fríjoles remojados en el tratamiento 2 tienen tiempos de cocción menores.

Caso 2: Prueba de hipótesis para diferencia de medias $\mu_1 - \mu_2$ con poblaciónes normales y varianzas diferentes

Se quiere estudiar

$$H_0: \mu_1 - \mu_2 = \delta_0 \qquad H_0: \mu_1 - \mu_2 = \delta_0 \qquad H_0: \mu_1 - \mu_2 = \delta_0$$
$$H_1: \mu_1 - \mu_2 < \delta_0 \qquad H_1: \mu_1 - \mu_2 \neq \delta_0 \qquad H_1: \mu_1 - \mu_2 > \delta_0$$

El estadístico es

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Welch's two simple t-Test

y se tiene que $t_0 \sim t_v$ donde v se calcula así:

$$v = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$$

$valor{-}P$

- Si $H_1 \colon \mu_1 - \mu_2 < \delta_0$

  $$\text{Valor} - \text{P} = \text{Prob}(t_v \leq t_0\ )$$

- Si $H_1 \colon\ \mu_1 - \mu_2 \neq \delta_0$

  $$\text{Valor} - \text{P} = 2\ \times\ \text{Prob}(t_v \geq |t_0|\ )$$

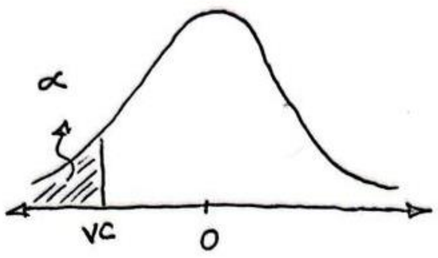
- Si $H_1 \colon\ \mu_1 - \mu_2 > \delta_0$

  $$\text{Valor} - \text{P} = \text{Prob}(t_v \geq t_0\ )$$

Regiones de rechazo

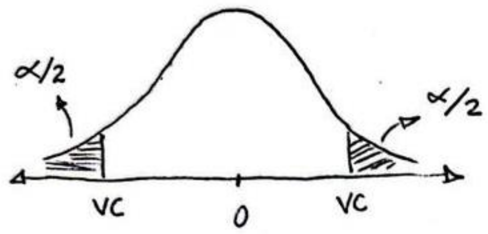
- Si H_a: $\mu_1 - \mu_2 < \delta_0$

  Se rechaza H_0 si $t_0 < -t_{\alpha,v}$



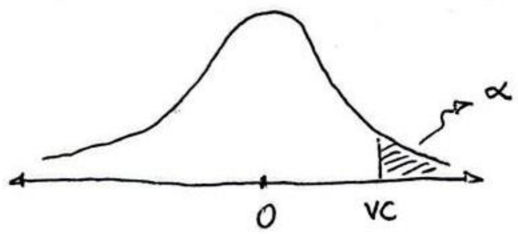
- Si H_a: $\mu_1 - \mu_2 \neq \delta_0$

  Se rechaza H_0 si $|t_0| > t_{\alpha/2,v}$



- Si H_a: $\mu_1 - \mu_2 > \delta_0$

  Se rechaza H_0 si $t_0 > t_{\alpha,v}$



Ejemplo

El arsénico en agua potable es un posible riesgo para la salud. Un artículo reciente reportó concentraciones de arsénico en agua potable en partes por billón (ppb) para 10 comunidades urbanas y 10 comunidades rurales. Los datos son los siguientes:

Urbana: 3, 7, 25, 10, 15, 6, 12, 25, 15, 7

Rural: 48, 44, 40, 38, 33, 21, 20, 12, 1, 18

Nota: de un ejemplo anterior se sabe que los datos provienen de poblaciones normales con varianzas diferentes.

Ejemplo

¿Existen diferencias entre las concentraciones medias de arsénico de la zona rural y urbana? Aplicar una prueba de hipótesis con $\alpha = 7\%$.

Concentraciones promedio

Urbana: 12.5 ppb

Rural: 27.5 ppb

Ejemplo

Paso 1. Definir las hipótesis

$$H_0\colon\ \mu_{Urb} - \mu_{Rur} = 0$$
$$H_1\colon\ \mu_{Urb} - \mu_{Rur} \neq 0$$

Urbana

Rural

Ejemplo

Paso 2. Calcular el estadístico

De las muestras se tiene

Para urbana

$$\bar{x}_1 = 12.50$$
$$s_1^2 = 58.28$$
$$n_1 = 10$$

Para rural

$$\bar{x}_2 = 27.50$$
$$s_2^2 = 235.61$$
$$n_2 = 10$$

Ejemplo

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{12.5 - 27.5 - 0}{\sqrt{\frac{58.28}{10} + \frac{235.61}{10}}} = -2.77$$

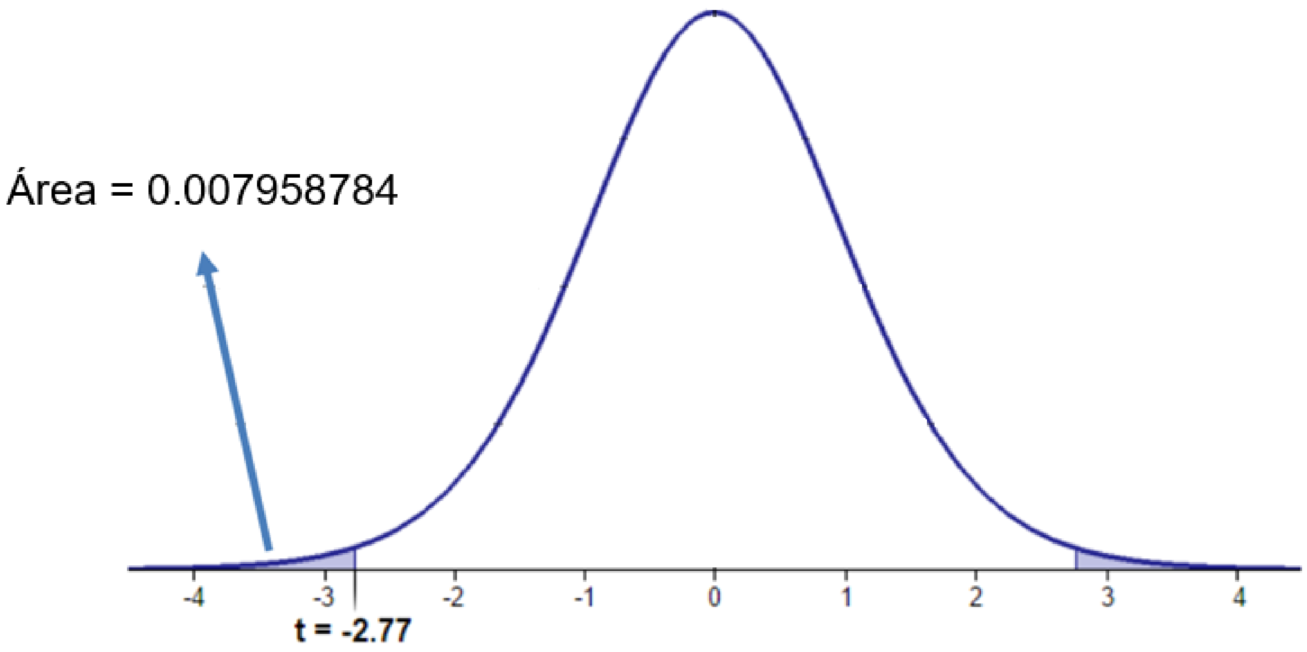
Los grados de libertad son:

$$v = \frac{\left({s_1^2}/{n_1} + {s_2^2}/{n_2} \right)^2}{\frac{\left({s_1^2}/{n_1} \right)^2}{n_1 - 1} + \frac{\left({s_2^2}/{n_2} \right)^2}{n_2 - 1}} \approx 13.2 = 13$$

Ejemplo

Paso 3. Calcular el $valor\text{–}P$.

En R `pt(q=-2.77, df=13, lower.tail=TRUE)`



Como la prueba es bilateral entonces $ValorP = 2 \times Área$ = 0.01591757

Ejemplo

Paso 4. Tomar la decisión.

Como el $valor\text{–}P$ es pequeño entonces se rechaza la hipótesis nula. En otras palabras, hay evidencias de que las concentraciones medias de arsénico son diferentes.

Caso 3: Prueba de hipótesis para diferencia de medias $\mu_1 - \mu_2$ con poblaciónes normales, varianzas conocidas

Se quiere estudiar:

$$H_0\colon \mu_1 - \mu_2 = \delta_0 \qquad H_0\colon \mu_1 - \mu_2 = \delta_0 \qquad H_0\colon \mu_1 - \mu_2 = \delta_0$$
$$H_a\colon \mu_1 - \mu_2 < \delta_0 \qquad H_a\colon \mu_1 - \mu_2 \neq \delta_0 \qquad H_a\colon \mu_1 - \mu_2 > \delta_0$$

El estadístico está dado por:

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

donde $Z_0 \sim N(0, 1)$.