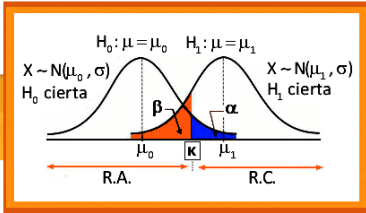




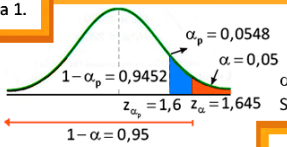
GUÍA p-valor

Cuanto más lejos se encuentra H_0 de H_1 menor es la probabilidad de cometer un Error Tipo II (β) y, por tanto, la Potencia se aproxima a 1.

$\alpha = P[\text{Error tipo I}]$
 $\beta = P[\text{Error tipo II}]$
Potencia = $1 - \beta$




$\alpha = P[\text{Rechazar } H_0 \mid H_0 \text{ es cierta}]$
 $\beta = P[\text{Aceptar } H_0 \mid H_0 \text{ es falsa}]$



$\alpha_p = 0,0548 > \alpha = 0,05$
Se acepta H_0

Potencia = $1 - \beta = P[\text{Aceptar } H_1 \mid H_1 \text{ es cierta}]$



La elección del nivel de significación α (riesgo) es en cierta forma arbitrario.
El p-valor es una cantidad que resume de manera objetiva el resultado del experimento.
El p-valor $\equiv \alpha_p$ es el riesgo más pequeño con el que se acepta la hipótesis alternativa H_1 en las observaciones muestrales.
La decisión a tomar con $\alpha_p \equiv p$ -valor que se obtiene de la muestra:

- { Se acepta la hipótesis nula H_0 : p-valor $\equiv \alpha_p > \alpha$
- { Se rechaza la hipótesis nula H_0 : p-valor $\equiv \alpha_p < \alpha$

$\alpha_p = p\text{-valor} = P[\text{Rechazar estadístico muestra} \mid H_0 \text{ es cierta}]$

Una empresa de neumáticos afirma que una nueva gama en promedio dura al menos 28.000 km. Las pruebas con 64 neumáticos dan como resultado una duración media de 27.800 km, con una desviación estándar de 1.000 km.

Bajo la normalidad de los neumáticos, se pide:

- Comprobar si hay evidencia suficiente para rechazar la afirmación de la empresa, un nivel de significación del 5%
- ¿Cuál es el p-valor?

Solución:

a) Sobre la población de los neumáticos se define la variable aleatoria $X = \text{"duración en kilómetros"}$ donde $X \sim N(28000, \sigma)$

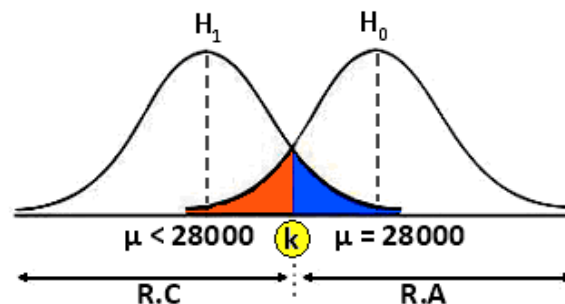
Las hipótesis sobre la media poblacional μ con σ^2 desconocida:

$$H_0: \mu_0 \geq 28000 \quad H_1: \mu_1 < 28000$$

Se trata de un contraste compuesto, unilateral por la izquierda.

Regla de decisión:

$$\begin{cases} \text{Si } \bar{x} \geq k & \text{Se acepta } H_0 \text{ (R.A)} \\ \text{Si } \bar{x} < k & \text{Se rechaza } H_0 \text{ (R.C)} \end{cases}$$



En el muestreo de una población normal con varianza desconocida, con muestras grandes $n > 30$ la media muestral $\bar{x} \sim N\left(\mu, \frac{s_x}{\sqrt{n}}\right)$

Bajo la hipótesis nula $\bar{x} \sim N\left(28000, \frac{1000}{\sqrt{64}}\right) \equiv N(28000, 125)$

El valor crítico k , bajo la hipótesis nula, se determina con el nivel de significación $\alpha = 0,05$:

$$\alpha = P[\text{Rechazar } H_0 \mid H_0 \text{ cierta}] = P[\bar{x} < k \mid H_0 \text{ cierta}] = P[\bar{x} < k \mid \mu_0 = 28000] =$$

$$= P\left(\frac{\bar{x} - 28000}{125} < \frac{k - 28000}{125}\right) = P\left[z < \frac{k - 28000}{125}\right] = 0,05$$

$$P\left[z < \frac{k - 28000}{125}\right] = P\left[z > -\frac{k - 28000}{125}\right] = 0,05$$

$$-\frac{k - 28000}{125} = 1,645 \quad \mapsto \quad k = 28000 - 125 \times 1,645 = 27794,375$$

Siendo $\bar{x} = 27800 > 27794,375$ se acepta la hipótesis nula, por tanto, se acepta la afirmación de la empresa con un nivel de confianza del 95%

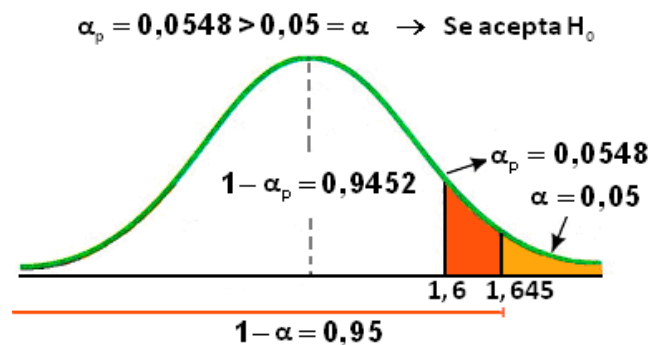
b) El p-valor (α_p) es el menor nivel de significación para el que se rechaza la hipótesis nula, es decir:

$$\alpha_p = \text{p-valor} = P[\text{Rechazar el estadístico muestral} / H_0 \text{ es cierta}]$$

Si $\alpha_p > \alpha$ se acepta la hipótesis nula H_0

$$\alpha_p = P[\bar{x} < 27800 | H_0 \text{ cierta}] = P[\bar{x} < 27800 | N(28000, 125)] =$$

$$= P\left[\frac{\bar{x} - 28000}{125} < \frac{27800 - 28000}{125}\right] = P[z < -1,6] = P[z > 1,6] = 0,0548$$



Para un nivel de significación $\alpha = 0,05$ el p-valor $\alpha_p = 0,0548 > 0,05 = \alpha \mapsto$
 \mapsto Se acepta H_0

Con una fiabilidad del 95% se acepta que la duración media de los neumáticos es de 28.000 km.

El propietario de un automóvil sospecha que su vehículo tiene un consumo medio de combustible en carretera superior a los 5,6 litros /100 km, que es lo que el fabricante indica en su publicidad.

Para apoyar empíricamente su sospecha observa el consumo medio en 11 viajes seleccionados aleatoriamente entre todos los que realiza en el año, obteniendo los siguientes resultados:

6,1 6,5 5,1 6 5,9 5,2 5,8 5,3 6,2 5,9 6,3

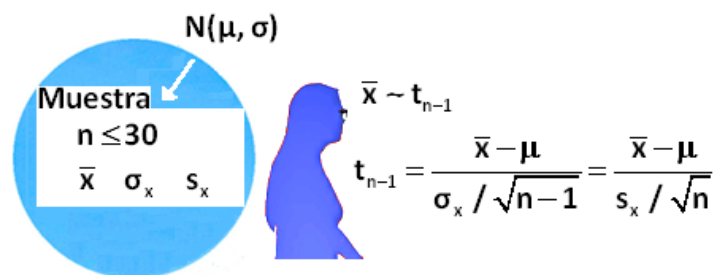
Se pide:

- ¿Están fundadas las sospechas del propietario a un nivel de significación del 1%?
- Calcula el p-valor.

Solución:

a) Se supone que el consumo medio del automóvil sigue una distribución normal $N(\mu, \sigma)$, siendo ambos parámetros desconocidos.

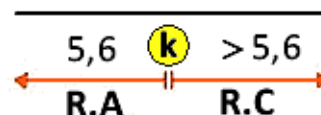
En el muestreo de una población normal con varianza desconocida, con muestras pequeñas $n \leq 30$, la media muestral $\bar{x} \sim t_{n-1}$



El fabricante afirma que $H_0 : \mu = 5,6$ y el propietario del vehículo cree que $H_1 : \mu > 5,6$

Se trata, pues, de un contraste unilateral, donde H_1 es compuesta

Regla decisión: $\begin{cases} \text{Si } \bar{x} \leq k \mapsto \text{R.A: Aceptar } H_0 \\ \text{Si } \bar{x} > k \mapsto \text{R.C: Rechazar } H_0 \end{cases}$



Bajo la hipótesis nula, con los datos muestrales ($\bar{x} = 5,8454$, $s_x = 0,4612$),

el muestreo sigue una distribución $t_{10} \left[5,6; \frac{0,4612}{\sqrt{11}} \right] \equiv t_{10} (5,6, 0,139)$

El valor crítico k , bajo la hipótesis nula, se calcula a partir del nivel de significación α :

$$\alpha = P[\text{Rechazar } H_0 \mid H_0 \text{ cierta}] = P[\bar{x} > k \mid H_0 \text{ cierta}] = P[\bar{x} > k \mid \mu_0 = 5,6] =$$

$$= P\left[\frac{\bar{x} - 5,6}{0,139} > \frac{k - 5,6}{0,139}\right] = P\left[t_{10} > \frac{k - 5,6}{0,139}\right] = 0,01 \Rightarrow \frac{k - 5,6}{0,139} = 2,764$$

$$\mapsto \boxed{k = 5,9842}$$

Siendo $\bar{x} = 5,8454 < 5,9842$ no se puede rechazar la hipótesis nula H_0 , con lo que se acepta las afirmaciones del fabricante sobre el consumo medio del automóvil.

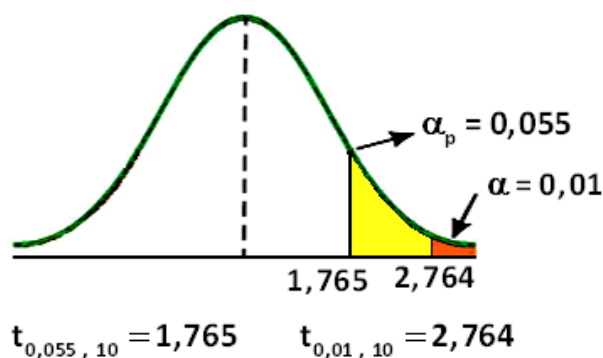
b) $\alpha_p = p - \text{valor} = P[\text{Rechazar el estadístico muestral} / H_0 \text{ es cierta}]$

Si $\alpha_p > \alpha$ se acepta la hipótesis nula H_0

$$\alpha_p = P[\bar{x} > 5,8454 \mid H_0 \text{ cierta}] = P[\bar{x} > 27800 \mid t_{10}(5,6, 0,139)] =$$

$$= P\left[\frac{\bar{x} - 5,6}{0,139} > \frac{5,8454 - 5,6}{0,139}\right] = P[t_{10} > 1,765] = 0,055$$

Tabla de la t-Student: $\frac{1,372 - 1,812}{0,10 - 0,05} = \frac{1,765 - 1,812}{x - 0,05} \mapsto x = 0,055$



$\alpha_p = 0,055 > 0,01 = \alpha \mapsto$ Se acepta H_0

Con un nivel de confianza del 99% se acepta que el consumo medio de combustible en carretera superior es de 5,6 litros /100 km.

Ventajas del p – valor frente al nivel de significación (riesgo α):

Obtenida una muestra, el p – valor es una medida verosímil que permite *decidir* que hipótesis se acepta según un nivel de riesgo α elegido, indicando cuánto contradice la muestra a la hipótesis alternativa H_1 . Esto no se produce cuando se utiliza el nivel de significación (un riesgo fijo α) para aceptar o rechazar la hipótesis nula H_0 .

$$\alpha_p = p - \text{valor} = P[\text{Rechazar estadístico muestral} / H_0 \text{ es cierta}]$$



Un banco quiere analizar si las comisiones que cobra a sus clientes por operaciones en el mercado bursátil difieren significativamente de las que cobra la competencia, cuya media es de 12 euros mensuales con una desviación estándar de 4,3 euros.

Este banco toma una muestra de 64 operaciones bursátiles y observa que la comisión promedio es de 13,6 euros. Contrastar, al nivel de significación del 5%, que este banco no difiere significativamente en el cobro de comisiones por operaciones en la Bolsa con respecto a la competencia.

Solución:

Sea la v. a. X ="Comisiones que se cobran por operaciones en el mercado bursátil", donde $X \sim N(\mu, 4,3)$



Con el nivel de significación $\alpha = 0,05$

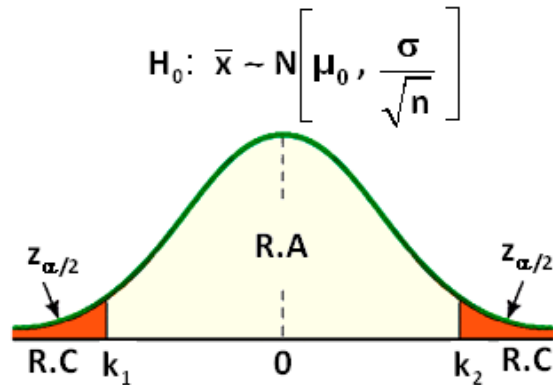
Se establecen las hipótesis:

$$H_0: \mu = 12 \quad H_1: \mu \neq 12$$

Como la hipótesis alternativa es $H_1: \mu \neq 12$ en la decisión deben ser válidos valores de μ tanto mayores o menores que 12, por lo que el contraste debe ser bilateral o de dos colas.

Regla de decisión:

$$\begin{cases} \text{Si } |\bar{x}| \leq k & \text{se acepta } H_0 \text{ (R.A)} \\ \text{Si } |\bar{x}| > k & \text{se rechaza } H_0 \text{ (R.C)} \end{cases}$$



Bajo la hipótesis nula: $\bar{x} \sim N\left[12, \frac{4,3}{\sqrt{64}}\right] \equiv N(12, 0,5375)$

Con el nivel de significación $\alpha = 0,05$ se calculan los valores críticos k_1, k_2 :

$$\begin{aligned} \alpha &= P[\text{Rechazar } H_0 | H_0 \text{ cierta}] = P[|\bar{x}| > k | N(12, 0,5375)] = P[(\bar{x} < k_1) \cup (\bar{x} > k_2)] = \\ &= P(\bar{x} < k_1) + P(\bar{x} > k_2) = 0,025 + 0,025 = 0,05 \end{aligned}$$

- $P(\bar{x} < k_1) = P\left[\frac{\bar{x} - 12}{0,5375} < \frac{k_1 - 12}{0,5375}\right] = P\left[z < \frac{k_1 - 12}{0,5375}\right] = P\left[z > -\frac{k_1 - 12}{0,5375}\right] = 0,025$

$$-\frac{k_1 - 12}{0,5375} = 1,96 \quad \mapsto \quad k_1 = 10,9465$$

- $P(\bar{x} > k_2) = P\left[\frac{\bar{x} - 12}{0,5375} > \frac{k_2 - 12}{0,5375}\right] = P\left[z > \frac{k_2 - 12}{0,5375}\right] = 0,025$

$$\frac{k_2 - 12}{0,5375} = 1,96 \quad \mapsto \quad k_2 = 13,0535$$

En consecuencia, la región de aceptación: $10,9465 < \bar{x} < 13,0535$

Como el banco cobra una comisión promedio de 13,6 euros, no se encuentra dentro de la región de aceptación, $13,6 \notin (10,9465, 13,0535)$, en consecuencia difiere significativamente de la competencia.

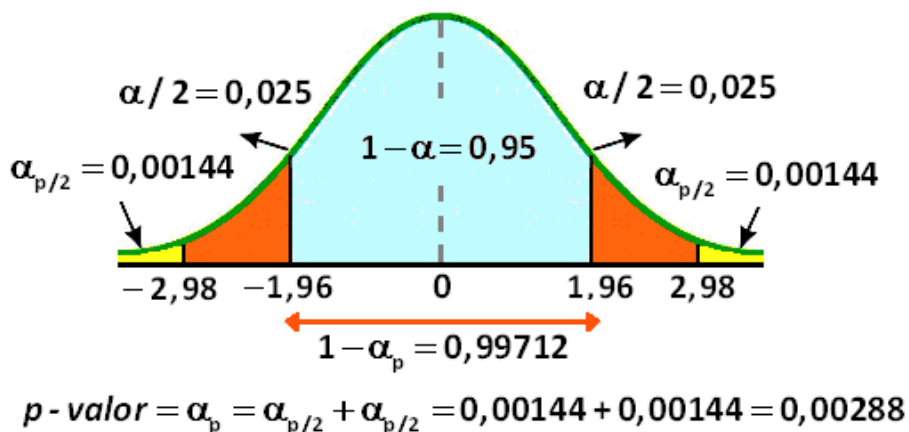


El p -valor con los datos de la muestra decide la decisión a tomar con la hipótesis nula $H_0: \mu = 12$
 muestra: $n = 64$ $\bar{x} = 13,6$

$$\alpha_p = \text{p-valor} = P[\text{Rechazar media muestral} / H_0 \text{ es cierta}]$$

Si $\alpha_p > \alpha$ se acepta la hipótesis nula H_0

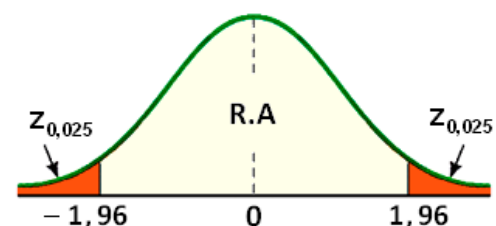
$$\begin{aligned} \alpha_p &= P[|\bar{x}| > 13,6 \mid N(12, 0,5375)] = P[(\bar{x} < -13,6) \cup (\bar{x} > 13,6)] = \\ &= P(\bar{x} < -13,6) + P(\bar{x} > 13,6) = 2 P(\bar{x} > 13,6) = \\ &= 2 P\left[\frac{\bar{x} - 12}{0,5375} > \frac{13,6 - 12}{0,5375}\right] = 2 P(z > 2,98) = 2 \times 0,00144 = 0,0029 \end{aligned}$$



Siendo $\alpha_p = 0,0029 < 0,05 = \alpha \rightarrow$ rechaza la hipótesis nula a un nivel de significación del 5%. Por tanto, existe evidencia estadística de que la comisión promedio que cobra este banco difiere significativamente de la competencia.



Se puede considerar la región de aceptación $(-z_{0,025}, z_{0,025})$, es decir, $(-1,96, 1,96)$



$$H_0: \bar{x} \sim N\left[\mu_0, \frac{\sigma}{\sqrt{n}}\right] \equiv N\left[12, \frac{4,3}{\sqrt{64}}\right]$$

La comisión promedio que cobra el banco: $\bar{x} \sim N \left[12, \frac{4,3}{\sqrt{64}} \right] \equiv N(12, 0,5375)$

Tipificando se obtiene el estadístico de contraste:

$$z = \frac{\bar{x} - 12}{0,5375} = \frac{13,6 - 12}{0,5375} = 2,98 \quad 2,98 \notin (-1,96, 1,96)$$

El valor del estadístico de contraste está en la zona de rechazo, por lo que se rechaza la hipótesis nula, afirmando que la comisión promedio que cobra este banco difiere significativamente de la competencia.



Estadístico de contraste: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Se acepta H_0 si $z = \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}} \leq z_{\alpha/2}$

Región rechazo $R = \left\{ |\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$

Intervalo confianza $I_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$



Un supermercado tiene dos formas diferentes de venta: tradicional o con caja rápida. En la tradicional, un operario registra cada artículo, lo pone en bolsas y efectúa el cobro. En la caja rápida, el cliente registra cada artículo, lo introduce en bolsas y paga en una máquina con tarjeta de crédito. La dirección del supermercado desea saber si el tiempo medio que un cliente se encuentra en la fila del proceso con el método tradicional es mayor que con la caja rápida, reuniendo la información adjunta:

Tipo cliente	Media muestral	Tamaño muestra	Varianza poblacional
Tradicional	6,1 minutos	49	0,49 minutos ²
Caja rápida	5,6 minutos	72	0,36 minutos ²

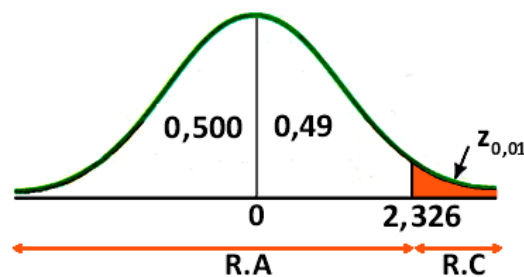
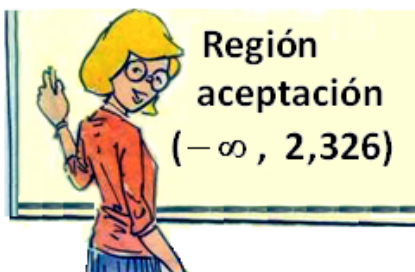
Suponiendo que el tiempo en espera en el proceso sigue una ley normal, con un nivel de confianza del 99%, ¿qué decisión debería adoptar el supermercado para agilizar las ventas?

Solución:

Sean las variables de respuesta: $X = \text{"tiempo método tradicional"}$ e $Y = \text{"tiempo caja rápida"}$, que siguen respectivamente $N(\mu_t, 0,7)$ y $N(\mu_{cr}, 0,6)$

El p -valor con los datos de la muestra facilita la decisión que debe tomar la dirección del supermercado, estableciendo las hipótesis:

$$H_0: \mu_t \leq \mu_{cr} \quad H_1: \mu_t > \mu_{cr}$$



$$H_0: \bar{x} - \bar{y} \sim N\left(0, \sqrt{\frac{\sigma_t^2}{n_t} + \frac{\sigma_{cr}^2}{n_{cr}}}\right)$$

Tipificando se obtiene el estadístico de contraste:

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_t^2}{n_t} + \frac{\sigma_{cr}^2}{n_{cr}}}} = \frac{6,1 - 5,6}{\sqrt{\frac{0,49}{49} + \frac{0,36}{72}}} = 4,082$$

estadístico teórico (valor crítico): $z_{0,01} = 2,326$

El valor del estadístico de contraste $z = 4,082 \notin (-\infty, 2,326)$ por lo que se rechaza la hipótesis nula, concluyendo que el supermercado debe adoptar la caja rápida.



$\alpha_p = p\text{-valor} = P[\text{Rechazar estadístico muestral} / H_0 \text{ es cierta}]$

$\alpha_p = p\text{-valor} < 0,500 - 0,490 < 0,01 = \alpha$

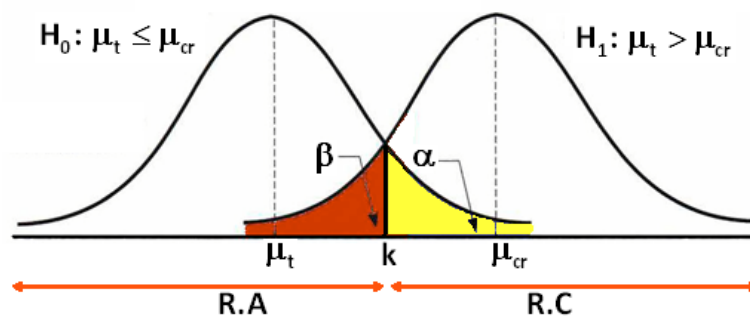
se rechaza la hipótesis nula H_0

Contraste unilateral de igualdad de medias de dos poblaciones normales de varianzas conocidas:

$H_0: \mu_1 \leq \mu_2$

$H_1: \mu_1 > \mu_2$

Se acepta H_0 si $z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_\alpha$



$$H_0: (\bar{x} - \bar{y}) \sim N\left(0, \sqrt{\frac{\sigma_t^2}{n_t} + \frac{\sigma_{cr}^2}{n_{cr}}}\right)$$

Regla Decisión $\begin{cases} \bar{x} - \bar{y} \leq k \text{ se acepta } H_0 \text{ (RA)} \\ \bar{x} - \bar{y} > k \text{ se rechaza } H_0 \text{ (RC)} \end{cases}$

Se calcula el valor crítico k con el nivel de significación $\alpha = 0,01$:

$$\alpha = P(\text{Error Tipo I}) = P[\text{Rechazar } H_0 / H_0 \text{ es cierta}] = 0,01$$

$$\begin{aligned}\alpha &= P[\bar{x} - \bar{y} > k \mid H_0 \text{ cierta}] = P\left[\bar{x} - \bar{y} > k \mid N\left(0, \sqrt{\frac{0,49}{49} + \frac{0,36}{72}}\right)\right] \\ &= P[\bar{x} - \bar{y} > k \mid N(0, 0,122)] = P\left[\frac{\bar{x} - \bar{y} - 0}{0,122} > \frac{k - 0}{0,122}\right] = P\left[z > \frac{k}{0,122}\right] = 0,01\end{aligned}$$

$$\frac{k}{0,122} = 2,326 \quad \mapsto \quad k = 0,284$$

$$\text{Regla Decisión} \quad \begin{cases} \bar{x} - \bar{y} \leq 0,284 & \text{se acepta } H_0 \text{ (RA)} \\ \bar{x} - \bar{y} > 0,284 & \text{se rechaza } H_0 \text{ (RC)} \end{cases}$$

Siendo, $\bar{x} - \bar{y} = 6,1 - 5,6 = 0,5 > 0,284 \rightarrow$ Se rechaza H_0

El supermercado debe adoptar la caja rápida.



William Sealy Gosset (1876 -1937) fue un estadístico empleado en el control de calidad de la compañía cervecera Guinness.

En 1908 descubrió la distribución t de Student, el contrato con la compañía estipulaba que no podía utilizar su nombre en sus publicaciones, con lo que recurrió al sobrenombre de Student.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \mapsto \quad \begin{array}{l} \text{diferencia a probar} \\ \text{desviación estándar de la diferencia} \\ \text{o error estándar} \end{array}$$



La directora del departamento de personal de una corporación está buscando empleados para un puesto en el extranjero. Durante el proceso de selección, la administración le pregunta cómo va la incorporación de empleados, y ella contesta que la puntuación promedio en la prueba de aptitudes será de 90 puntos. Cuando la administración revisa 19 de los resultados de la prueba, encuentra que la puntuación media es de 83,25 puntos con una desviación estándar de 11. Bajo el supuesto de normalidad, ¿lleva razón la directora, con un nivel de confianza del 90%?

Solución:

Se supone que la población de resultados de todos los candidatos siguen una distribución $X \sim N(\mu, \sigma)$, siendo ambos parámetros desconocidos. En el muestreo de la población normal con varianza desconocida, con

muestras pequeñas $n = 11 < 30$, la media muestral $\bar{x} \sim t_{10}$ con $t_{n-1} = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$

Se establecen las hipótesis: $H_0: \mu = 90$ $H_1: \mu \neq 90$

Como la hipótesis alternativa es $\mu \neq 90$ en la decisión deberán ser válidos valores de μ tanto mayores o menores que 90, por lo cual el contraste debe ser bilateral o de dos colas.

Regla decisión: $\begin{cases} \text{Si } |\bar{x}| > k \text{ se rechaza } H_0 \text{ (R.C.)} \\ \text{Si } |\bar{x}| \leq k \text{ se acepta } H_0 \text{ (R.A.)} \end{cases}$

$\alpha_p = p\text{-valor} = P[\text{Rechazar estadístico muestral} / H_0 \text{ es cierta}]$

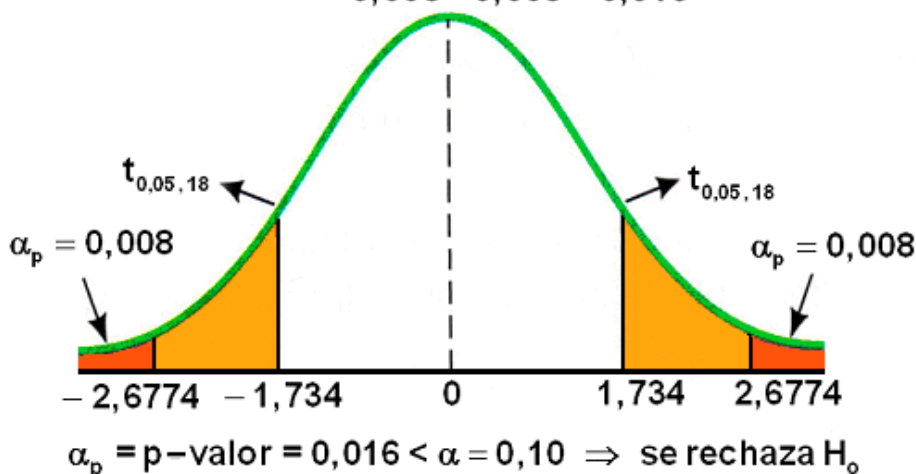
$\alpha_p = P[|\bar{x}| > 83,25 \mid t_{18}(90; 2,524)] =$

$= P[(\bar{x} > 83,25) \cup (\bar{x} < -83,25) \mid t_{18}(90; 2,524)] =$

$$\begin{aligned}
&= P\left[\bar{x} > 83,25 \mid t_{18}(90; 2,524)\right] + P\left[\bar{x} < -83,25 \mid t_{18}(90; 2,524)\right] = \\
&= P\left[\bar{x} > 83,25 \mid t_{18}(90; 2,524)\right] + P\left[\bar{x} > 83,25 \mid t_{18}(90; 2,524)\right] = \\
&= 2P\left[\bar{x} > 83,25 \mid t_{18}(90; 2,524)\right] = 2P\left[\frac{\bar{x} - 90}{2,524} > \frac{83,25 - 90}{2,524}\right] = \\
&= 2P\left[t_{18} > 2,6774\right] = 2(0,008) = 0,016
\end{aligned}$$

Tabla t-Student: $\frac{2,552 - 2,878}{0,01 - 0,005} = \frac{2,674 - 2,878}{x - 0,005} \mapsto x = 0,008$

p-valor = suma de ambas áreas =
 $= 0,008 + 0,008 = 0,016$



Siendo $\alpha_p = 0,016 < 0,10 = \alpha \mapsto$ Se rechaza la hipótesis nula.

En consecuencia, se rechaza la manifestación de la directora de la corporación.



Los paquetes de software estadístico denotan al p-valor por sig (bilateral). Conocido el p-valor, el usuario toma la decisión de aceptar o rechazar la hipótesis nula.

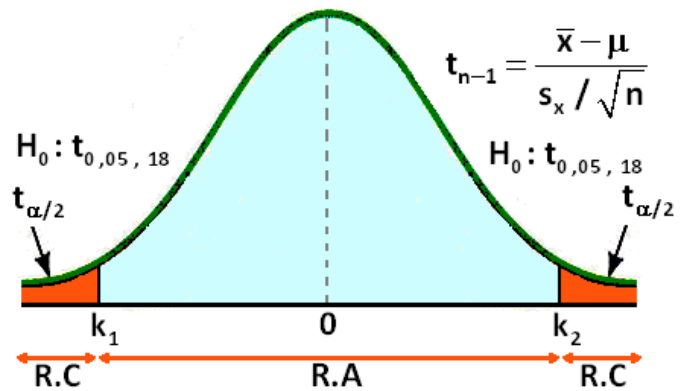
SPSS siempre calcula el p-valor para el contraste bilateral, en caso de realizar un contraste unilateral hay que dividir entre 2 el valor calculado por SPSS.

p-valor = sig (bilateral) = 0,025
 Contraste unilateral: $H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$
 p-valor = $0,025 / 2 = 0,0125$
 Si $\begin{cases} \alpha < 0,0125 \mapsto \text{Se acepta } H_0 \\ \alpha > 0,0125 \mapsto \text{Se rechaza } H_0 \end{cases}$



Regla decisión:

- { Si $|\bar{x}| \leq k$ se acepta H_0 (R.A.)
- { Si $|\bar{x}| > k$ se rechaza H_0 (R.C.)



El valor crítico k se calcula con el nivel de significación $\alpha = 0,10$:

$$\alpha = P[\text{Rechazar } H_0 | H_0 \text{ cierta}] = P[|\bar{x}| > k | t_{18}(90; 2,524)] = P[(\bar{x} < k_1) \cup (\bar{x} > k_2)] =$$

$$= P(\bar{x} < k_1) + P(\bar{x} > k_2) = 0,05 + 0,05 = 0,10$$

- $P(\bar{x} < k_1) = P\left[\frac{\bar{x} - 90}{2,524} < \frac{k_1 - 90}{2,524}\right] = P\left[t_{18} < \frac{k_1 - 90}{2,524}\right] = P\left[t_{18} > -\frac{k_1 - 90}{2,524}\right] = 0,05$

$$-\frac{k_1 - 90}{2,524} = 1,734 \quad \mapsto \quad k_1 = 85,62$$

- $P(\bar{x} > k_2) = P\left[\frac{\bar{x} - 90}{2,524} > \frac{k_2 - 90}{2,524}\right] = P\left[t_{18} > \frac{k_2 - 90}{2,524}\right] = 0,05$

$$\frac{k_2 - 90}{2,524} = 1,734 \quad \mapsto \quad k_2 = 94,37$$

Para aceptar la hipótesis nula la media muestral se tiene que encontrar en el intervalo $85,62 < \bar{x} < 94,37$

No encontrándose la media muestral observada $\bar{x} = 83,25$ en la región de aceptación, con un nivel de significación de $\alpha = 0,10$, se rechaza la manifestación de la directora de la corporación.



Los contrastes de hipótesis están muy relacionados con la teoría de los intervalos de confianza. En muchos casos se puede resolver la misma cuestión por las dos vías. Mientras que el contraste de hipótesis acepta o rechaza la hipótesis nula, el intervalo de confianza ofrece más información.

$$I_{1-\alpha}(\mu) = \left[\bar{x} \pm t_{\alpha/2, n-1} \frac{s_x}{\sqrt{n}} \right] \mapsto I_{0,90}(\mu) = \left[83,25 \pm t_{0,05, 18} \frac{11}{\sqrt{19}} \right]$$

$$I_{0,90}(\mu) = \left[83,25 \pm 1,734 \frac{11}{\sqrt{19}} \right] = [83,25 \pm 4,376] = [78,874, 87,626]$$

En consecuencia, $78,874 \leq \mu \leq 87,626$

La puntuación promedio $\mu_0 = 90$ que manifiesta la directora del departamento de personal no se encuentra dentro del intervalo de la media de la población con una fiabilidad del 90%, teniendo que rechazar su afirmación.

Un portal e-business sabe que el 60% de todos sus visitantes a la web están interesados en adquirir sus productos pero son reacios al comercio electrónico y no realizan finalmente la compra vía internet. Sin embargo, en la dirección del portal se piensa que en el último año el porcentaje de gente que está dispuesta a comprar por internet ha aumentado y esto se debe reflejar en los resultados empresariales. En esta línea, se tomó una muestra de 500 visitantes para conocer su opinión y se observó que el 55% no estaba dispuesta a realizar compras vía on-line.

Contrastar con el 2% de significación si el último año se ha reducido el porcentaje de personas que no está dispuesta a comprar vía internet.

Solución:

Sea p = "Proporción personas compran on-line"

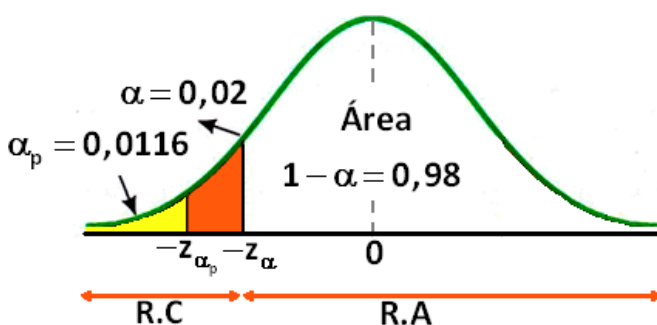
Al realizar el contraste sobre la proporción, se parte de una muestra aleatoria $(x_1, x_2, \dots, x_{500})$, donde $X \sim B(1, p)$

La distribución del estadístico muestral $\hat{p} = \sum_{i=1}^{500} \frac{x_i}{n}$ al ser el tamaño suficientemente grande $n = 500$ y estar \hat{p} definido como suma de variables independientes entre sí, según una distribución de Bernouilli $X \sim B(1, p)$, por el teorema central del límite

(TCL) se puede aproximar $\hat{p} = \sum_{i=1}^{500} \frac{x_i}{n} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$

Se establecen las hipótesis: $H_0: p \geq 0,6$ $H_1: p < 0,6$

Bajo la hipótesis nula, el estadístico muestral $\hat{p} \sim N\left(0,6, \sqrt{\frac{0,6 \times 0,4}{500}}\right)$



Contraste unilateral a la izquierda
(cola a la izquierda)

Región aceptación $(-z_\alpha, \infty)$

- Siendo que el estadístico $\hat{p} \sim N\left(0,6, \sqrt{\frac{0,6 \times 0,4}{500}}\right) \equiv N(0,6, 0,022)$, se obtiene el estadístico de contraste tipificando:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}} = \frac{0,55 - 0,60}{0,022} = -2,27$$

El estadístico de contraste $z = -2,27$ no se encuentra en la región de aceptación $(-z_{0,02}, \infty) \equiv (-2,05, \infty)$, rechazando la hipótesis nula.

En conclusión, existe evidencia empírica que la proporción de visitantes al portal que están dispuestos a comprar on-line ha disminuido, es decir, el porcentaje de visitantes que son reacios a comprar por internet ha aumentado.



O lo que es lo mismo, decidir a partir del p – valor:

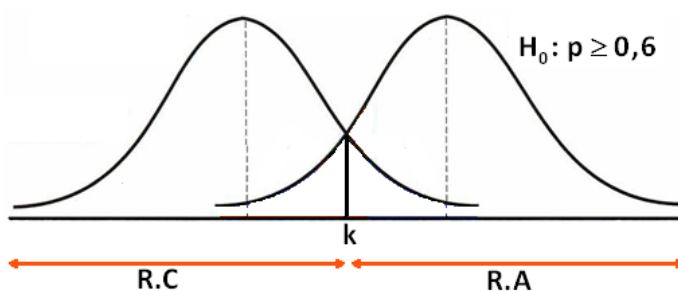
$$\alpha_p = \text{p – valor} = P[\text{Rechazar estadístico} / H_0 \text{ es cierta}]$$

$$\begin{aligned} \alpha_p &= P\left[\hat{p} < 0,55 \mid N(0,6, 0,022)\right] = P\left[\frac{\hat{p} - 0,6}{0,022} < \frac{0,55 - 0,6}{0,022}\right] = P(z < -2,27) = \\ &= P(z > 2,27) = 0,0116 \end{aligned}$$

Como $\alpha_p = 0,0116 < 0,02 = \alpha$ se rechaza la hipótesis nula a un nivel de significación del 2%.

$\alpha_p = p\text{-valor} = P[\text{Rechazar estadístico} / H_0 \text{ es cierta}]$
 $\alpha_p = p\text{-value} = P[\text{estadístico tipificado} \mid \text{distribución que sigue}]$
 Se acepta H_0 si $\alpha_p > \alpha$
 $\alpha_p = p\text{-valor} = \text{Sig (bilateral)}$
 contraste unilateral $\alpha_p = p\text{-valor} = \frac{\text{Sig (bilateral)}}{2}$

- Se determina el valor crítico k con el nivel de significación $\alpha = 0,02$



Contraste unilateral por la izquierda

regla de decisión: $\begin{cases} \hat{p} \geq k & \text{se acepta } H_0 \\ \hat{p} < k & \text{se rechaza } H_0 \end{cases}$

$$\alpha = P[\text{Rechazar } H_0 \mid H_0 \text{ cierta}] = P[\hat{p} < k \mid N(0,6, 0,022)] = P\left[\frac{\hat{p} - 0,6}{0,022} < \frac{k - 0,6}{0,022}\right] =$$

$$= P\left[z < \frac{k - 0,6}{0,022}\right] = 0,02 \quad \mapsto \quad \frac{k - 0,6}{0,022} = 2,05 \quad \Rightarrow \quad k = 0,645$$

El valor del estadístico muestral (evidencia empírica) $\hat{p} = 0,55 < 0,642$, por lo que se rechaza la hipótesis nula. En definitiva, ha aumentado el porcentaje de personas que son reacios a comprar por internet.

- Con el intervalo de confianza unilateral a la izquierda para la proporción, verificando si el estadístico $\hat{p} = 0,55$ pertenece al intervalo:

$$H_0: p \geq p_0 \quad H_1: p < p_0 \quad \mapsto \quad \left(p_0 - z_\alpha \sqrt{\frac{p_0 \cdot q_0}{n}}, \infty \right)$$

$$\left(0,6 - 2,05 \cdot \sqrt{\frac{0,6 \cdot 0,4}{500}}, \infty \right) \equiv (0,56, \infty)$$

Siendo $\hat{p} = 0,55 \notin (0,56, \infty)$ se rechaza la hipótesis nula.



Los contrastes bilaterales corresponden a intervalos centrados, mientras que los contrastes unilaterales derechos corresponden a estimaciones unilaterales por exceso y los unilaterales izquierdos a estimaciones por defecto.

- $H_0: p = p_0 \quad H_1: p \neq p_0 \quad \mapsto \left(p_0 - z_{\alpha/2} \sqrt{\frac{p_0 \cdot q_0}{n}}, p_0 + z_{\alpha/2} \sqrt{\frac{p_0 \cdot q_0}{n}} \right)$ bilateral

Se acepta H_0

o también: $z = \frac{|\hat{p} - p_0|}{\sqrt{\frac{p_0 \cdot q_0}{n}}} \leq z_{\alpha/2}$ $R.C = \left\{ |\hat{p} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0 \cdot q_0}{n}} \right\}$

estadístico contraste región rechazo

- $H_0: p \leq p_0 \quad H_1: p > p_0 \quad \mapsto \left(-\infty, p_0 + z_{\alpha} \sqrt{\frac{p_0 \cdot q_0}{n}} \right)$ unilateral derecho

Se acepta H_0

o también: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}} \leq z_{\alpha}$ $R.C = \left\{ \hat{p} - p_0 > z_{\alpha} \sqrt{\frac{p_0 \cdot q_0}{n}} \right\}$

estadístico contraste región rechazo

- $H_0: p \geq p_0 \quad H_1: p < p_0 \quad \mapsto \left(p_0 - z_{\alpha} \sqrt{\frac{p_0 \cdot q_0}{n}}, \infty \right)$ unilateral izquierdo

Se acepta H_0

o también: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}} \geq z_{1-\alpha}$ $R.C = \left\{ \hat{p} - p_0 < z_{1-\alpha} \sqrt{\frac{p_0 \cdot q_0}{n}} \right\}$ $z_{1-\alpha} = -z_{\alpha}$

estadístico contraste región rechazo



Dos muestras son independientes o dependientes entre sí en función si las observaciones se han obtenido de los mismos individuos u objetos o no.

Para estudiar si el consumo de tabaco tiende a provocar problemas de trombosis debido a un aumento en la capacidad de coagulación, se extrajeron muestras de sangre en 10 individuos antes y después de que fumasen un cigarrillo, midiendo la capacidad de agregación de las plaquetas:

Antes	25	27	44	30	67	53	53	52	60	28
Después	29	37	56	46	82	57	80	61	59	43

Bajo el supuesto de normalidad, ¿hay suficiente evidencia estadística (con un nivel de significación 0,01) para afirmar que los fumadores presentan la misma tendencia a la formación de coágulos?

Calcular el p-valor

Solución:

Sean las variables aleatorias:

X = "coágulos obtenidos antes de fumar", donde $X \sim N(\mu_1, \sigma_1)$, donde μ_1 es el porcentaje medio de agregación de plaquetas antes de fumar.

Y = "coágulos obtenidos después de fumar", con $Y \sim N(\mu_2, \sigma_2)$, donde μ_2 es el porcentaje medio de agregación de plaquetas después de fumar.

Destacar que los datos obtenidos antes de fumar, y los obtenidos después de fumar, no son independientes, dado que se obtienen de la misma persona.

No se puede llevar a cabo un contraste de igualdad de medias como se realizó cuando las muestras eran independientes, se puede considerar la variable aleatoria $d = X - Y$, y admitiendo que su distribución sea $N(\mu, \sigma)$, efectuar un contraste adecuado sobre $\mu = \mu_1 - \mu_2$

Se desea probar estadísticamente que $\mu_1 = \mu_2$, o lo que es igual, que $\mu = 0$ por lo que se trata de un contraste bilateral (o de dos colas).

Se establecen las hipótesis: $H_0: \mu = 0$ $H_1: \mu \neq 0$

- En otros casos, se podría estar ante contrastes unilaterales:

$H_0: \mu \geq 0$ frente a $H_1: \mu < 0$ o bien $H_0: \mu \leq 0$ frente a $H_1: \mu > 0$

- Si se tuviera la creencia de que fumar supone un aumento de formación de coágulos de 5 puntos, las hipótesis serían:

$H_0: \mu = \mu_1 - \mu_2 = -5$ frente a $H_1: \begin{cases} \mu \neq -5 \\ \mu > -5 \end{cases}$

Para continuar es necesario conocer la tabla de diferencias:

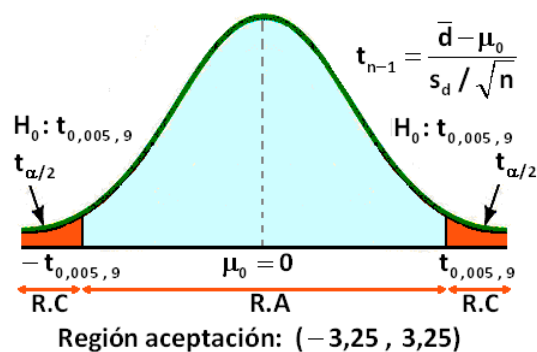
Antes	25	27	44	30	67	53	53	52	60	28
Después	29	37	56	46	82	57	80	61	59	43
d_i	-4	-10	-12	-16	-15	-4	-27	-9	1	-15

$$\bar{d} = \frac{1}{10} \sum_{i=1}^{10} d_i = -11,1 \quad s_d^2 = \frac{1}{9} \sum_{i=1}^{10} (d_i - \bar{d})^2 = 62,32 \quad s_d = 7,89$$

- En el muestreo de la población normal con varianza desconocida, con

muestras pequeñas $n = 10 < 30$, la media muestral $\bar{d} \sim t_9$ con $t_{n-1} = \frac{\bar{d} - \mu_0}{s_d / \sqrt{n}}$

como estadístico de contraste.



$$|t_{n-1}| = \frac{|\bar{d} - \mu_0|}{s_d / \sqrt{n}}$$

Región de aceptación:

$$(-t_{\alpha/2, n-1}, t_{\alpha/2, n-1})$$

Contraste bilateral, estadístico de contraste: $|t_{n-1}| = \frac{|\bar{d} - \mu_0|}{s_d / \sqrt{n}}$

Región de aceptación $R.A = \{t_{n-1}: |t_{n-1}| \leq t_{\alpha/2, (n-1)}\}$

Región crítica $R.C = \{t_{n-1}: |t_{n-1}| > t_{\alpha/2, (n-1)}\}$

Lo que equivale a decir que H_0 se acepta si $|t_{n-1}| = \frac{|\bar{d} - \mu_0|}{s_d / \sqrt{n}} \leq t_{\alpha/2, (n-1)}$

Bajo la hipótesis nula, el estadístico contraste: $|t_9| = \frac{|-11,1 - 0|}{7,89 / \sqrt{10}} = 4,449$

$t_{\alpha/2, n-1} = t_{0,005, 9} = 3,25 \mapsto$ Región aceptación: $(-3,25, 3,25)$

Siendo $4,449 \notin (-3,25, 3,25)$ se rechaza la hipótesis nula, pudiendo decir que los fumadores no presentan la misma tendencia a la formación de coágulos.

• $\alpha_p = p$ – valor = $P[\text{Rechazar estadístico} \mid H_0 \text{ es cierta}] = P[t_9 \geq 4,449] = 0,00084$

Tabla t-Student: $\frac{4,297 - 4,781}{0,001 - 0,0005} = \frac{4,449 - 4,781}{x - 0,0005} \mapsto x = 0,00084$

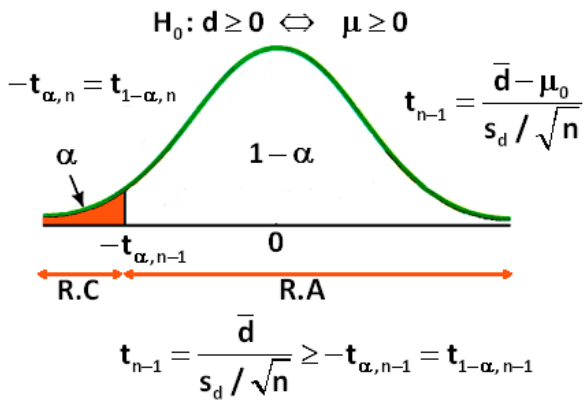
• El intervalo de confianza es: $I_{1-\alpha}(\mu) = \left[\bar{d} \pm t_{\alpha/2, n-1} \frac{s_d}{\sqrt{n}} \right]$

$$I_{0,99}(\mu) = \left[-11,1 \pm 3,25 \cdot \frac{7,89}{\sqrt{10}} \right] \equiv [-19,21, -2,99]$$

El intervalo no abarca el cero, $\mu \neq 0$, concluyendo que los fumadores no presentan la misma tendencia a la formación de coágulos.



¿Hay suficiente evidencia estadística (con un nivel de confianza del 99%) a favor de la hipótesis de que los fumadores presentan mayor tendencia a la formación de coágulos?



Se desea probar estadísticamente que $\mu_1 < \mu_2$ o bien que $\mu = \mu_1 - \mu_2 < 0$

Se contrasta: $H_0: \mu \geq 0$ frente a $H_1: \mu < 0$

Región de aceptación: $(-t_{\alpha, n-1}, \infty)$

$(-t_{0,01,9}, \infty) \equiv (-2,821, \infty)$

Contraste unilateral a la izquierda, estadístico de contraste: $t_{n-1} = \frac{\bar{d} - \mu_0}{s_d / \sqrt{n}}$

Región de aceptación $R.A = \{t_{n-1}: t_{n-1} \geq -t_{\alpha, n-1} = t_{1-\alpha, n-1}\}$

Región crítica $R.C = \{t_{n-1}: t_{n-1} < t_{1-\alpha, n-1}\}$

Bajo la hipótesis nula, el estadístico contraste: $t_9 = \frac{-11,1 - 0}{7,89 / \sqrt{10}} = -4,449$

Como $-4,449 \notin (-2,821, \infty)$ se rechaza la hipótesis nula, aceptando en consecuencia la hipótesis alternativa, concluyendo que los fumadores presentan mayor tendencia a la formación de coágulos.



- **Intervalo de confianza para el parámetro λ de una distribución de Poisson :**

Aproximación asintótica: Para valores del parámetro λ grandes, la distribución de Poisson puede aproximarse a una distribución normal según $P(\lambda) \mapsto N(\lambda, \sqrt{\lambda})$

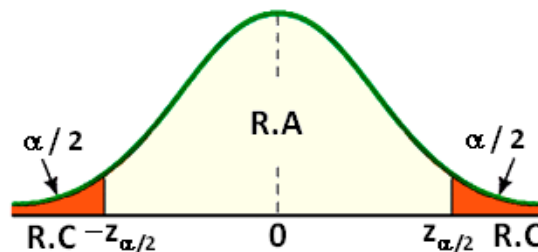
Dada una muestra de n observaciones independientes X_1, X_2, \dots, X_n , que siguen una distribución de Poisson de parámetro λ , $X_i \sim P(\lambda)$, al ser la distribución de

Poisson aditiva en λ se verifica $\sum_{i=1}^n X_i \sim N(n\lambda, \sqrt{n\lambda})$

En la muestra: $\hat{\lambda} = \bar{X} \xrightarrow{\text{aproximadamente}} N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)$

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}} \sim N(0, 1)$$

$$P\left(-z_{\alpha/2} \leq \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\hat{\lambda}}{n}}} \leq z_{\alpha/2}\right) = 1 - \alpha$$



$$I_{1-\alpha}(\lambda) = \left[\hat{\lambda} \pm z_{\alpha/2} \sqrt{\frac{\hat{\lambda}}{n}} \right] \quad \text{R.A} \equiv (-z_{\alpha/2}, z_{\alpha/2})$$

- En caso de no ser aplicables las condiciones para utilizar la aproximación asintótica, el intervalo exacto, que es válido siempre, viene dado:

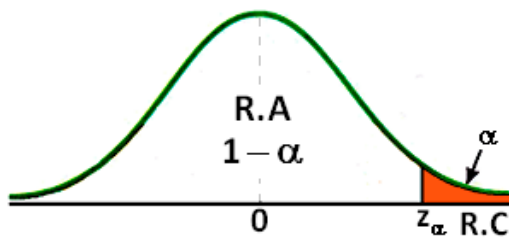
$$I_{1-\alpha}(\lambda) = \left[\frac{1}{2n} \left(2 \sum_{i=1}^n X_i \right) \chi_{1-\alpha/2}^2 ; \frac{1}{2n} \left(2 \sum_{i=1}^n X_i + 2 \right) \chi_{\alpha/2}^2 \right]$$

- **Contraste bilateral para el parámetro λ de una distribución de Poisson :**

$$H_0: \lambda = \lambda_0 \quad H_1: \lambda \neq \lambda_0 \quad R.A = \left\{ \frac{|\hat{\lambda} - \lambda_0|}{\sqrt{\lambda_0/n}} \leq z_{\alpha/2} \right\} \quad R.C = \left\{ \frac{|\hat{\lambda} - \lambda_0|}{\sqrt{\lambda_0/n}} > z_{\alpha/2} \right\}$$

- **Contraste unilateral a la derecha para el parámetro λ de una distribución de Poisson :**

$$H_0: \lambda \leq \lambda_0 \quad H_1: \lambda > \lambda_0$$



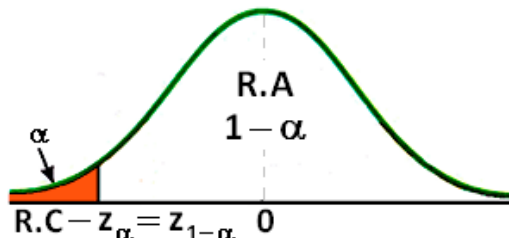
Región aceptación: $(-\infty, z_\alpha)$

$$R.A = \left\{ \frac{\hat{\lambda} - \lambda_0}{\sqrt{\lambda_0/n}} \leq z_\alpha \right\}$$

$$R.C = \left\{ \frac{\hat{\lambda} - \lambda_0}{\sqrt{\lambda_0/n}} > z_\alpha \right\}$$

- **Contraste unilateral a la izquierda para el parámetro λ de una distribución de Poisson :**

$$H_0: \lambda \geq \lambda_0 \quad H_1: \lambda < \lambda_0$$



Región aceptación: $(-z_\alpha, \infty)$

$$R.A = \left\{ \frac{\hat{\lambda} - \lambda_0}{\sqrt{\lambda_0/n}} \geq z_{1-\alpha} \right\}$$

$$R.C = \left\{ \frac{\hat{\lambda} - \lambda_0}{\sqrt{\lambda_0/n}} < z_{1-\alpha} \right\}$$

Se quiere comprobar la efectividad de una vacuna contra una alergia. Para ello se suministró la vacuna a cien personas y se les comparó con un grupo testigo de cien pacientes afectados también por la alergia en épocas pasadas. Entre los vacunados, ocho sufrieron alergia y entre los no vacunados veinticinco volvieron a sufrir alergia.

¿Se puede concluir que la vacuna es eficaz en disminuir la alergia?.

Utilizar un nivel de significación de 0,05.

Solución:

Sean las variables aleatorias:

X = "número pacientes vacunados afectados por la alergia"

Y = "número pacientes no vacunados afectados por la alergia"

donde, $X \sim B(p_1, n_1)$ e $Y \sim B(p_2, n_2)$

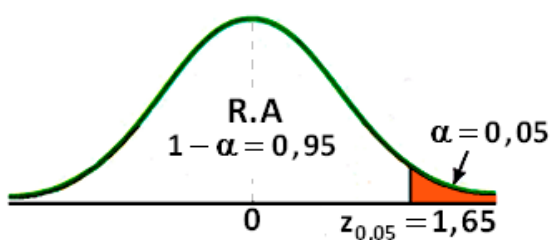
$$\text{En el muestreo: } \begin{cases} \hat{p}_1 = \frac{8}{100} = 0,08 & \hat{q}_1 = 0,92 & n_1 = 100 \\ \hat{p}_2 = \frac{25}{100} = 0,25 & \hat{q}_2 = 0,75 & n_2 = 100 \end{cases}$$

Las muestras $n_1 = n_2 = 100$ son grandes, en consecuencia, ambas distribuciones se pueden aproximar a normales:

$$X \sim N(n_1 p_1, \sqrt{n_1 p_1 q_1}) \quad , \quad Y \sim N(n_2 p_2, \sqrt{n_2 p_2 q_2})$$

$$\text{En el muestreo: } \hat{p}_2 - \hat{p}_1 \sim N\left(p_2 - p_1, \sqrt{\frac{p_2 q_2}{n_2} + \frac{p_1 q_1}{n_1}}\right)$$

Para ver si la vacuna es eficaz hay que contrastar si $p_2 \leq p_1$



Región aceptación: $(-\infty, 1,65)$

$$H_0: p_2 \leq p_1 \quad H_1: p_2 > p_1$$

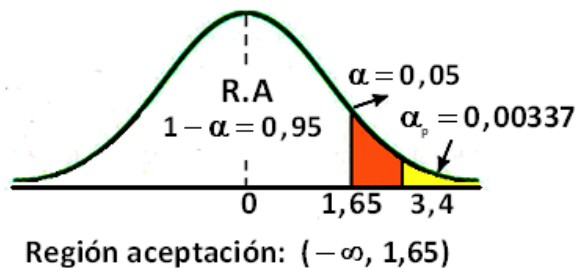
Región aceptación: $(-\infty, 1,65)$

- Bajo la hipótesis nula $\hat{p}_2 - \hat{p}_1 \sim N\left(0, \sqrt{\frac{p_2 q_2}{n_2} + \frac{p_1 q_1}{n_1}}\right)$, tipificando se obtiene el estadístico de contraste:

$$z = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\frac{p_2 q_2}{n_2} + \frac{p_1 q_1}{n_1}}} = \frac{0,25 - 0,08}{\sqrt{\frac{0,25 \cdot 0,75}{100} + \frac{0,08 \cdot 0,92}{100}}} = 3,4$$

El estadístico de contraste $z = 3,4$ no se encuentra en la región de aceptación $(-\infty, 1,65)$, rechazando la hipótesis nula y concluyendo que la vacuna es eficaz ($p_2 > p_1$)

- $\alpha_p = p - \text{valor} = P[\text{Rechazar estadístico} \mid H_0 \text{ es cierta}]$



$$\alpha_p = P(z \geq 3,4) = 0,00337$$

$$\alpha_p = 0,00337 < 0,05 = \alpha \rightarrow$$

\mapsto se rechaza la hipótesis nula a un nivel de significación del 5%.

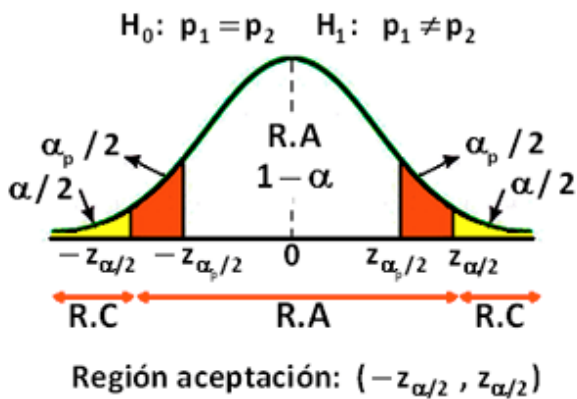
- Intervalo de confianza para la diferencia de parámetros ($p_1 - p_2$) de dos distribuciones binomiales.

$$I_{1-\alpha}(p_1 - p_2) = \left[(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right] \quad z_{0,025} = 1,96$$

$$I_{1-\alpha}(p_1 - p_2) = \left[(0,08 - 0,25) \pm 1,96 \sqrt{\frac{0,08 \cdot 0,92}{100} + \frac{0,75 \cdot 0,25}{100}} \right]$$

$$I_{1-\alpha}(p_1 - p_2) = [-0,269, -0,07] \quad \mapsto \quad -0,269 \leq p_1 - p_2 \leq -0,07$$

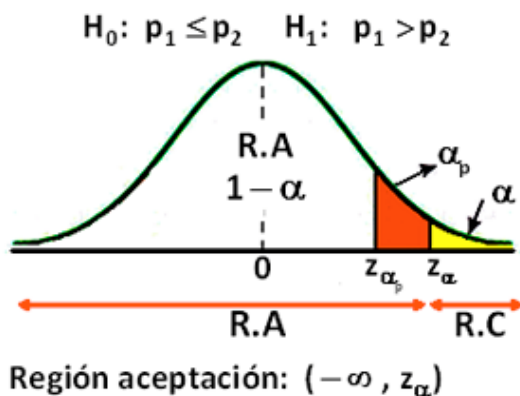
Como el intervalo no cubre el cero ($p_1 - p_2 \neq 0$) se rechaza que la vacuna no tenga eficacia. Por otra parte, ambos límites del intervalo de confianza son negativos, lo que significa que $p_1 < p_2$



estadístico contraste: $z^* = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$

p-valor = $P[|z| > z^*] = P\left[|z| > \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}\right]$

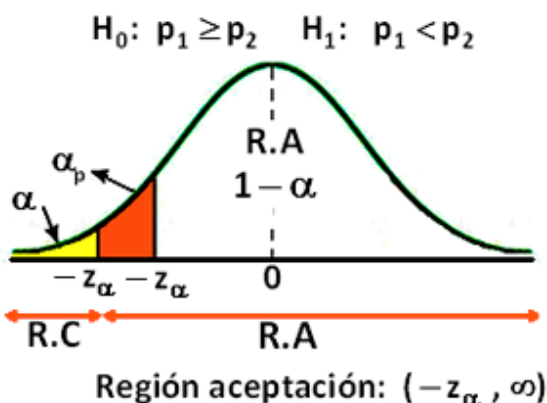
R.C = $\left\{ |\hat{p}_1 - \hat{p}_2| > z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \right\}$



estadístico contraste: $z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$

p-valor = $P[z > z^*] = P\left[z > \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}\right]$

R.C = $\left\{ \hat{p}_1 - \hat{p}_2 > z_{\alpha} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \right\}$



estadístico contraste: $z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$

p-valor = $P[z < z^*] = P\left[z < \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}\right]$

R.C = $\left\{ \hat{p}_1 - \hat{p}_2 < z_{1-\alpha} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \right\}$

Dos variables aleatorias X e Y siguen distribuciones normales con varianzas desconocidas



$$X \sim N(\mu_1, \sigma_1) \text{ con una muestra } n_1 \leq 30 \quad \mapsto \quad t_{n_1-1} = \frac{\bar{x} - \mu_1}{\sqrt{\frac{s_1^2}{n_1}}}$$

$$Y \sim N(\mu_2, \sigma_2) \text{ con una muestra } n_2 \leq 30 \quad \mapsto \quad t_{n_2-1} = \frac{\bar{y} - \mu_2}{\sqrt{\frac{s_2^2}{n_2}}}$$

$X - Y \sim N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ con una muestra $n_1 + n_2 \leq 30$:

$$t_{n_1+n_2-2} = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Con hipótesis nula $H_0: \mu_1 = \mu_2 \Leftrightarrow \mu_1 - \mu_2 = 0 \quad \mapsto \quad t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

• Si las cuasivarianzas muestrales $s_1^2 = s_2^2$: $t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$


s_p se utiliza una cuasidesviación estándar muestral ponderada:

$$\text{varianza muestral ponderada: } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Contraste de igualdad de medias de dos poblaciones normales de varianzas desconocidas pero iguales $\sigma_1^2 = \sigma_2^2$, con muestras pequeñas $n_1 + n_2 \leq 30$


- Hipótesis nula $H_0: \mu_1 = \mu_2$ Hipótesis alternativa $H_1: \mu_1 \neq \mu_2$

estadístico contraste $t_{n_1+n_2-2} = \frac{|\bar{x} - \bar{y}|}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$




- Hipótesis nula $H_0: \mu_1 \leq \mu_2$ Hipótesis alternativa $H_1: \mu_1 > \mu_2$

estadístico contraste $t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$



- Hipótesis nula $H_0: \mu_1 \geq \mu_2$ Hipótesis alternativa $H_1: \mu_1 < \mu_2$

estadístico contraste $t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$




- Contraste de igualdad de medias de dos poblaciones normales de varianzas desconocidas y distintas $\sigma_1^2 \neq \sigma_2^2$, con muestras pequeñas $n_1 + n_2 \leq 30$

Igual que el caso anterior, sustituyendo los grados de libertad del estadístico teórico $(n_1 + n_2 - 2)$ por la aproximación de Welch:

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


- Hipótesis nula $H_0: \mu_1 = \mu_2$ Hipótesis alternativa $H_1: \mu_1 \neq \mu_2$

estadístico contraste $t_{n_1+n_2-2} = \frac{|\bar{x} - \bar{y}|}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$




- Hipótesis nula $H_0: \mu_1 \leq \mu_2$ Hipótesis alternativa $H_1: \mu_1 > \mu_2$

estadístico contraste $t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$



- Hipótesis nula $H_0: \mu_1 \geq \mu_2$ Hipótesis alternativa $H_1: \mu_1 < \mu_2$

estadístico contraste $t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$



Se ha comprobado un micrómetro con una serie de piezas patrón. Bajo el supuesto de normalidad, con un nivel de significación de 0,01, se desea comprobar si el micrómetro tiene error en las lecturas realizadas:

Patrón	0,003	0,100	0,200	0,300	0,400	0,500	0,600	0,700
Lectura	0,002	0,09	0,198	0,297	0,403	0,498	0,602	0,698

Solución:

Se plantea el modelo: Sea que la variable de respuesta $X = \text{"valor en el patrón"}$ sigue una normal $N(\mu_1, \sigma_1)$ y la variable $Y = \text{"valor en la lectura"}$ sigue una normal $N(\mu_2, \sigma_2)$

$$X \sim N(\mu_1, \sigma_1) \text{ en la muestra: } t_{n_1-1} = \frac{\bar{x} - \mu_1}{\sqrt{s_1^2 / n_1}}$$

$$Y \sim N(\mu_2, \sigma_2) \text{ en la muestra: } t_{n_2-1} = \frac{\bar{y} - \mu_2}{\sqrt{s_2^2 / n_2}}$$

$X - Y \sim N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ con una muestra $n_1 + n_2 = 16 < 30$:

$$t_{n_1+n_2-2} = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Se trata de un contraste de hipótesis bilateral de igualdad de medias de poblaciones normales. Hipótesis del contraste:

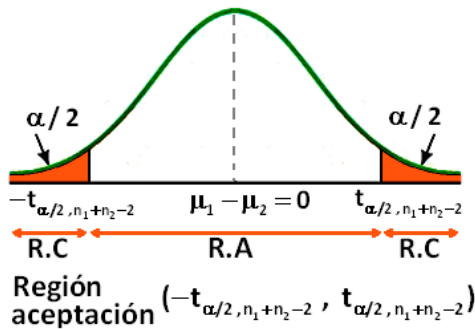
$$H_0: \mu_1 = \mu_2 \Leftrightarrow \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 \neq \mu_2 \Leftrightarrow \mu_1 - \mu_2 \neq 0$$

Bajo la hipótesis nula: $t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ estadístico contraste

$$\text{Muestras} \begin{cases} \bar{x} = 0,350 & s_1^2 = 0,06 & s_1 = 0,244 & n_1 = 8 \\ \bar{y} = 0,349 & s_2^2 = 0,061 & s_2 = 0,246 & n_2 = 8 \end{cases}$$

$$t_{n_1+n_2-2} = t_{14} = \frac{0,350 - 0,349}{\sqrt{\frac{0,06}{8} + \frac{0,061}{8}}} = 0,081$$

Como $\alpha_p = p\text{-valor} = P[t_{14} \geq 0,081] > 0,01 \mapsto$ se admite la hipótesis nula, pudiendo decir que el micrómetro no tiene error



$$R.A \equiv (-t_{\alpha/2, n_1+n_2-2}, t_{\alpha/2, n_1+n_2-2})$$

$$t_{\alpha/2, n_1+n_2-2} = t_{0,005, 14} = 2,977$$

$$(-2,977, 2,977)$$

Se observa que $0,081 \in (-2,977, 2,977)$ con lo que se admite la hipótesis nula, pudiendo concluir que el micrómetro no tiene error y las diferencia que se observan son debidas al azar o al observador.

La duración de una enfermedad sigue una ley normal. Para su curación se emplean dos tipos de antibióticos y se desea comprobar estadísticamente cuál de los dos es más eficaz respecto a la duración de la enfermedad. A tal fin se observan seis enfermos a los que se aplica el primer antibiótico y cinco enfermos a los que se aplica el segundo antibiótico.

Los datos obtenidos han sido:

$$\bar{x} = 12 \text{ días} \quad s_1^2 = 16 \quad n_1 = 6$$

$$\bar{y} = 15 \text{ días} \quad s_2^2 = 16 \quad n_2 = 5$$

¿Cuál de los dos antibióticos es más eficaz a un nivel de significación 0,05?

Solución:

La variable de respuesta $X =$ "duración de la enfermedad con el primer antibiótico" sigue una normal $N(\mu_1, \sigma_1)$ y la variable $Y =$ "duración de la enfermedad con el segundo antibiótico" sigue una normal $N(\mu_2, \sigma_2)$

$$X \sim N(\mu_1, \sigma_1) \text{ en la muestra: } t_{n_1-1} = \frac{\bar{x} - \mu_1}{\sqrt{s_1^2 / n_1}}$$

$$Y \sim N(\mu_2, \sigma_2) \text{ en la muestra: } t_{n_2-1} = \frac{\bar{y} - \mu_2}{\sqrt{s_2^2 / n_2}}$$

$X - Y \sim N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ con una muestra $n_1 + n_2 = 11 < 30$:

$$t_{n_1+n_2-2} = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

• Se trata de un contraste de hipótesis unilateral a la derecha de igualdad de medias de poblaciones normales. Hipótesis del contraste:

$$H_0: \mu_1 \leq \mu_2 \Leftrightarrow \mu_1 - \mu_2 \leq 0 \quad H_1: \mu_1 > \mu_2 \Leftrightarrow \mu_1 - \mu_2 > 0$$

Bajo la hipótesis nula: $t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ *estadístico contraste*

$$t_{n_1+n_2-2} = t_9 = \frac{12 - 15}{\sqrt{\frac{16}{6} + \frac{16}{5}}} = -1,24$$

$$\alpha_p = p\text{-valor} = P[t_9 \geq -1,24] = P[t_9 \leq 1,24] = 1 - P[t_9 \geq 1,24] = 1 - 0,1286 = 0,8714$$

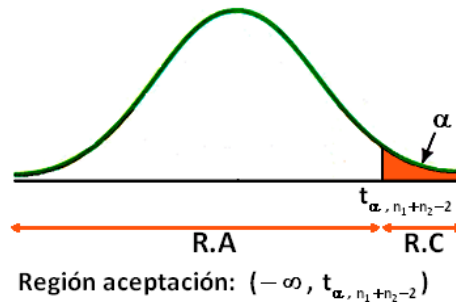
Tabla t-Student: $\frac{0,883 - 1,383}{0,20 - 0,10} = \frac{1,24 - 1,383}{x - 0,10} \mapsto x = 0,1286$

Como $\alpha_p = p\text{-valor} = 0,8714 > 0,05 \mapsto$ Se admite la hipótesis nula, concluyendo que el primer antibiótico tiene una duración media inferior en la curación de la enfermedad.

- $R.A \equiv (-\infty, t_{\alpha, n_1+n_2-2})$

$$t_{\alpha, n_1+n_2-2} = t_{0,05,9} = 1,833$$

$$R.A \equiv (-\infty, 1,833)$$



$$H_0: \mu_1 \leq \mu_2 \Leftrightarrow \mu_1 - \mu_2 \leq 0 \quad H_1: \mu_1 > \mu_2 \Leftrightarrow \mu_1 - \mu_2 > 0$$

Bajo la hipótesis nula: $t_{n_1+n_2-2} = t_9 = \frac{12-15}{\sqrt{\frac{16}{6} + \frac{16}{5}}} = -1,24$

$$-1,24 \in (-\infty, 1,833) \mapsto \text{Se admite } H_0: \mu_1 \leq \mu_2$$

El primer antibiótico tiene una duración media inferior en la curación de la enfermedad.

- Se trata de un contraste unilateral para las medias de dos poblaciones normales con muestras pequeñas, varianzas desconocidas pero iguales (debido a la información muestral $s_1^2 = s_2^2$). Hipótesis del contraste:

$$H_0: \mu_1 \leq \mu_2 \quad H_1: \mu_1 > \mu_2$$

Estadístico contraste: $t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ Estadístico teórico: t_{α, n_1+n_2-2}

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

$$s_p^2 = \frac{6 \cdot 16 + 5 \cdot 16}{6 + 5 - 2} = 16 \mapsto s_p = 4$$

$$t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{12 - 15}{4 \sqrt{\frac{1}{6} + \frac{1}{5}}} = \frac{-3}{2,42} = -1,24$$

Se acepta la hipótesis H_0 ya que $t = -1,24 < 1,833 = t_{0,05,9}$

