CPE (SEGUNDO CURSO)

PRÁCTICA 17 **SOLUCIONES** (**Curso 2023–2024**)

1.– El sistema de molido de áridos en una fábrica de cemento dispone de dos molinos con capacidad teórica aproximada de 260 t/h cada uno. Como hay ciertas dudas acerca de que los molinos tengan producciones medias iguales, se realiza un estudios de rendimento para poder comprobar dicha hipótesis (igualdad de rendimientos medios de los dos molinos), con un nivel de significación del 5 %. Para ello se medirá la producción horaria de n horas en cada molino, obteniéndose las medias muestrales $\bar{x}$ y $\bar{y}$. Se sabe que $\sigma_x^2 = 23\ (t/h)^2$ y que $\sigma_y^2 = 26\ (t/h)^2$. Para aceptar la igualdad de las medias a ese nivel de significación, ¿cuál ha de ser la relación entre la diferencia $|\bar{x} - \bar{y}|$ y el número n de horas de producción analizadas? Si decidimos analizar 100 horas de producción, ¿cuál ha de ser la diferencia máxima entre las medias muestrales, en valor absoluto, para poder aceptar la hipótesis primaria? Si el resultado en este caso fuera $|\bar{x}_a - \bar{x}_b| = 0.65$ t/h, ¿cuál sería el nivel p correspondiente?

—————————SOLUCIÓN—————————

El contraste de hipótesis a realizar es

$H_o : m_x = m_y$

$H_1 : m_x \neq m_y$

Datos: $\alpha = 0.1$; $\sigma_x^2 = 23$; $\sigma_y^2 = 26$; $n_x = n_y = n$

Sabemos que si las dos producciones tienen distribuciones normales (o aunque no las tuvieran, si el tamaño de las muestras es suficientemente grande) el estadístico

$$U = \frac{(\overline{x} - \overline{y}) - (m_x - m_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

tiene una distribución normal $N(0, 1)$. Entonces, aceptaremos H_o cuando $-a \le U \le a$, con $P[-a \le U \le a] = 1 - \alpha = 0.95$. De las tablas obtenemos $a = 1.96$. Es decir, aceptaremos H_o si

$$-a \le \frac{(\overline{x} - \overline{y}) - (m_x - m_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_x^2}{n_y}}} \le a$$

Como en H_o se verifica $m_x - m_y = 0$, y como además sabemos que $n_x = n_y = n$ podemos escribir que aceptaremos la hipótesis primaria si

$$-a\sqrt{\sigma_x^2 + \sigma_y^2} \le (\overline{x} - \overline{x})\sqrt{n} \le a\sqrt{\sigma_x^2 + \sigma_y^2}$$

Pero $\sqrt{\sigma_x^2 + \sigma_y^2} = 7$, luego la condición pedida es

$$|\overline{x} - \overline{y}|\sqrt{n} \le 13.72$$

Si $n = 100$, ha de verificarse que $|\overline{x} - \overline{y}| \le 1.372$ t/h.

Y si, tras realizar el ensayo, obtenemos además un valor de la diferencia de medias $|\overline{x}-\overline{y}| = 0.65$ t/h., el nivel p del test será

$$1 - \frac{p}{2} = F_U(0.65) = 0.7421 \qquad \Longrightarrow \qquad p = 51.6\,\%$$

2.– La British Coal Utilization Research Association realizó un extenso estudio sobre el contenido de partículas sólidas (polvo) en emisiones de gas. El sistema de tubos estudiado presentaba dos sectores A y B claramente diferenciados. Durante el estudio se intuyó que la cantidad X de polvo (en mg) variaba más en A que en B. De un muestreo se obtuvieron los siguientes resultados:

	Sector A	*Sector B*
Tamaño de la muestra	12	6
Media muestral	65 *mg*	44 *mg*
Desv. tipica muestral	26 *mg*	16 *mg*

¿Corrobora la muestra la suposición de mayor variabilidad en A? ¿Qué puede decirse de las medias poblacionales? (Suponer que X sigue una distribución normal)

SOLUCIÓN

Vamos a suponer que las desviaciones típicas dadas en el enunciado son las insesgadas.

a) Comparemos las varianzas

$H_o : \sigma_B \leq \sigma_A$

$H_1 : \sigma_B > \sigma_A$

Definamos el estadístico en la hipótesis primaria H_o

$$F = \frac{S_B^{*\,2}}{S_A^{*\,2}} \equiv F_{5,11}$$

La región crítica está situada en la cola superior de la distribución, y los correspondientes valores críticos valen

$\alpha = 0.1 \qquad c = 2.45$

$\alpha = 0.05 \qquad c = 3.20$

$\alpha = 0.01 \qquad c = 5.32$

El estadístico F vale en este caso $F = 0.3787$. Por lo que aceptaremos la hipótesis primaria a cualquier nivel habitual de significación

b) El contraste a realizar en este caso es

$H_o : m_A = m_B$

$H_1 : m_A \neq m_B$

Como no conocemos las varianzas y los tamaños de muestra no son muy grandes, aplicaremos el test de Welch. Sabemos que

$$t = \frac{\overline{x}_A - \overline{x}_B}{\sqrt{\frac{S_A^{*\,2}}{n_A} + \frac{S_B^{*\,2}}{n_B}}} \equiv t_\eta$$

donde el número de grados de libertad de la distribución t de Student es

$$\eta = \frac{\left(\frac{S_A^{*\,2}}{n_A} + \frac{S_B^{*\,2}}{n_B}\right)^2}{\frac{S_A^{*\,4}}{n_A^2(n_A-1)} + \frac{S_B^{*\,4}}{n_B^2(n_B-1)}} \approx 15$$

En nuestro caso además $t = 2.11058$. Podemos pues calcular el nivel p del contraste con esta muestra determinada obteniendo

$$1 - \frac{p}{2} = F_{t_{15}}(2.11058) = 0.974 \quad \Rightarrow \quad p \approx 5.2\,\%$$

Luego sólo podemos asegurar que las medias poblacionales son iguales a un nivel máximo de significación del 5 %, que es bajo.

3.– En un importante acceso a una gran ciudad se ha procedido a la regulación del tráfico mediante un sistema por control remoto centralizado. Con objeto de comprobar si el tráfico es más fluido con el nuevo sistema, se ha tomado como parámetro de control la velocidad V que rebasan el 80 por ciento de los vehículos en la hora de máxima intensidad de tráfico, a su paso por un determinado punto.

Antes de la implantación del sistema de regulación dicha velocidad, V, fué a lo largo de una semana (de lunes a domingo) de: 27, 35.5, 39, 34, 24, 42.5, 50 en km/h.

Después de la implantación del sistema, la velocidad ha sido a lo largo de una semana de registro (de lunes a domingo): 31.1, 37.2, 38, 33.5, 29.5, 45, 47.2 en km/h.

¿Ha mejorado el tráfico con la implantación del sistema de control?. ¿Con qué nivel de significación?

Nota: especificar todas las hipótesis realizadas

SOLUCIÓN

Sea V la velocidad antes de la implantación del sistema de regulación, y W la velocidad después de la implantación del sistema. Calculemos los parámetros muestrales.

	V	V²	W	W²
	27	729	31,1	967,21
	35,5	1260,25	37,2	1383,84
	39	1521	38	1444
	34	1156	33,5	1122,25
	24	576	29,5	870,25
	42,5	1806,25	45	2025
	50	2500	47,2	2227,84
Σ	252	9548,5	261,5	10040,39
Media	36		37,35714286	
Varianza(S^2)		68,07142857		38,78530612
Desv. Típica (S)		8,250541108		6,227785009
Varianza(S^{*2})		79,41666667		45,24952381
S*		8,911602924		6,726776628

Con los datos que tenemos, a la pregunta de si ha mejorado el tráfico con la implantación del sistema de control, sólo podemos contestar si ha mejorado, en media, la velocidad del tráfico semanal. Para ello plantearemos el siguiente contraste de hipótesis

$H_0 : m_w \geq m_v$

$H_1 : m_w < m_v$

que como sabemos es equivalente, desde el punto de vista operativo, a

$H_0 : m_w = m_v$

$H_1 : m_w < m_v$

o bien

$H_0 : m_w - m_v = 0$

$H_1 : m_w - m_v < 0$

Dado que no sabemos nada sobre las varianzas poblacionales, y dado que las muestras son pequeñas, realizaremos un test de Welch, utilizando como estadístico

$$t = \frac{(\overline{w} - \overline{v}) - (m_w - m_v)}{\sqrt{S_w^{*2}/n_w + S_v^{*2}/n_v}} = \frac{\overline{w} - \overline{v}}{\sqrt{S_w^{*2}/n_w + S_v^{*2}/n_v}} \text{ (en } H_0)$$

que, como sabemos, y si tanto W como V son normales, es una t de Student con, aproximadamente,

$$\eta = \frac{(S_w^{*2}/n_w + S_v^{*2}/n_v)^2}{\frac{S_w^{*4}}{n_w^2(n_w-1)} + \frac{S_v^{*4}}{n_v^2(n_v-1)}}$$

grados de libertad. Dada la hipótesis alternativa, la región crítica está a la izquierda de la distribución de t. Es decir, la región crítica será $R_C = [-\infty, c]$ determinándose c de forma que $F_T(c) = \alpha$ y calculándose el nivel p de forma que $p = F_T(t)$, siendo t el estadístico de contraste. En nuestro caso, $\eta = 11.16$ y $t = 0.3216$. Utilizando las tablas de la t de Student, e interpolando adecuadamente obtenemos $p = 62.3\,\%$, lo que significa que podemos asegurar que el tráfico ha mejorado (en el sentido indicado) con un nivel de significación alto.

4.– La tabla adjunta representa el rendimiento de un cereal en 10 pares de parcelas. En una de las parcelas de cada par se ha utilizado un nuevo abono fosfatado y, en la otra, el abono tradicional no fosfatado. Siendo las demás condiciones las mismas,

a) ¿Puede concluirse de estos datos que el abono fosfatado mejora el rendimiento? ¿Con qué nivel máximo de significación?

b) ¿Puede concluirse de estos datos que el abono fosfatado empeora el rendimiento? ¿Con qué nivel máximo de significación?

DATOS

Parcela	1	2	3	4	5	6	7	8	9	10
Con fosfato	6.6	6.5	5.6	6.6	6.1	5.8	6.0	6.1	6.3	6.1
Sin fosfato	6.0	5.4	5.8	5.4	5.8	5.7	5.4	5.7	6.0	5.3

A la vista de los resultados anteriores, ¿se debería recomendar la compra del nuevo abono? Especificar todas las hipótesis que sea necesario realizar. Supóngase que las varianzas de ambas poblaciones son iguales.

————————SOLUCIÓN————————

Llamemos X a la variable "rendimiento con fosfato" e Y a la variable "rendimiento sin fosfato". Los datos suministrados por las muestras son:

$n_x = 10,\ \overline{x} = 6.17,\ S_x^* = 0.33,\ S_x = 0.316$

$n_y = 10,\ \overline{y} = 5.65,\ S_y^* = 0.26,\ S_y = 0.246$

Como nos dicen que las varianzas son iguales para ambas poblaciones utilizaremos como estadístico una t de Student con $\eta = n_x + n_y - 2 = 18$ grados de libertad.

a)

El contraste de hipótesis que tenemos que realizar es

$H_o : m_x \geq m_y$

$H_1 : m_x < m_y$

El estadístico a utilizar para la hipótesis primaria señalada es

$$t = \frac{\overline{x} - \overline{y}}{R\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \equiv t_\eta, \quad \text{con } R^2 = \frac{(n_x - 1){S_x^*}^2 + (n_y - 1){S_y^*}^2}{n_x + n_y - 2}$$

En nuestro caso $t = 3.914$. La región crítica para un nivel p determinado es, obviamente la que indica la figura situada al final del ejercicio.

Nuestro estadístico vale $t = 3.914$. Entonces el nivel p es $p = F_{t_{18}}(3.914) > 0.999$, es decir muy alto, y aceptaremos la hipótesis primaria a cualquier nivel de significación.

b)

En este caso nuestro contraste es

$H_o : m_y \geq m_x$

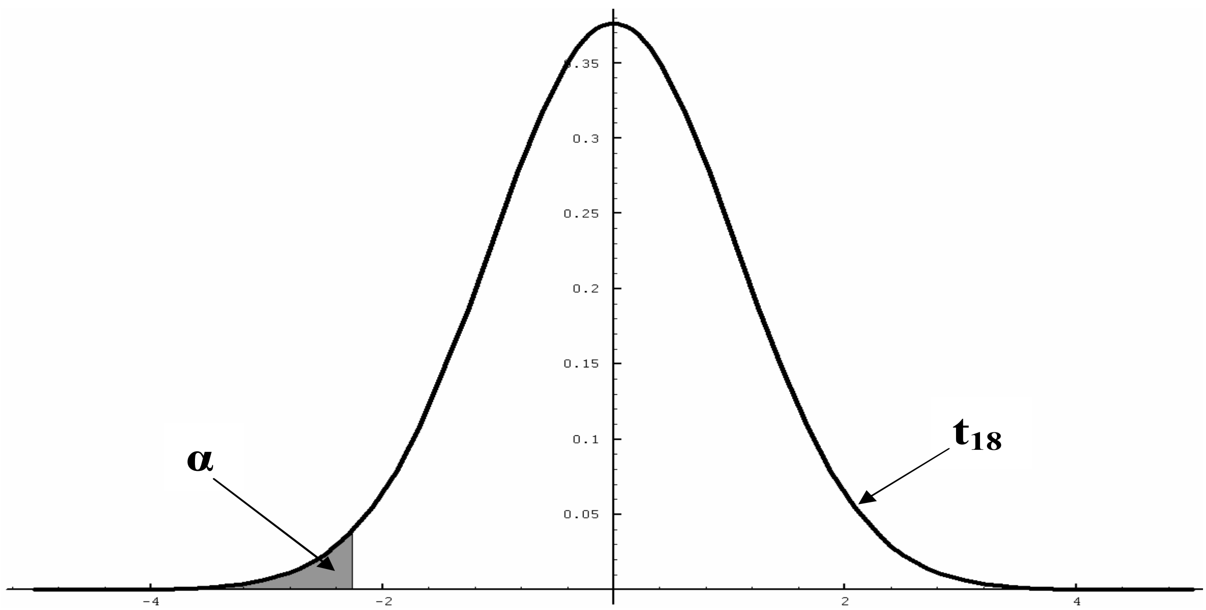
$H_1 : m_y < m_x$

La región critica es la misma, así como el número de grados de libertad, siendo nuestro estadístico

$$t = \frac{\overline{y} - \overline{x}}{R\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = -3.914$$

Y el correspondiente valor p es $p = F_{t_{18}}(-3.914) = 1 - F_{t_{18}}(3.914) \approx 0.0005 \approx 0$, por lo que no podemos aceptar en ningún caso la hipótesis primaria.

Como conclusión diremos que el nuevo abono mejora el rendimiento, y debe recomendarse su compra.



5.– Un sistema productivo puede emplear dos tipos de maquinaria en la fabricación de una cierta pieza cilíndrica de alta precisión. Interesa, entre otras cosas, que la producción sea lo más regular posible, es decir que la varianza de los diámetros de las piezas sea lo más pequeña posible. A la hora de elegir el tipo de maquinaria para su instalación definitiva, se han tomado muestras obteniéndose los siguientes resultados para cada tipo de maquinaria:

$Maquinaria\ Tipo\ 1: \quad n_1 = 41, \quad \bar{x}_1 = 10^{-4}\ m, \quad S_1^2 = 10^{-12}\ m^2$

$Maquinaria\ Tipo\ 2: \quad n_2 = 31, \quad \bar{x}_2 = 0.8 \times 10^{-4}\ m, \quad S_2^2 = 1.5 \times 10^{-12}\ m^2$

a) ¿Pueden considerarse iguales ambos dispositivos de fabricación? En caso contrario, ¿cuál debería adquirirse?

b) Un inventor afirma que con un procedimiento de su invención, la homogeneidad del producto fabricado por la Maquinaria Tipo 1 puede mejorarse hasta dividir por dos la desviación típica del diámetro de las piezas. Para comprobar la veracidad de esta afirmación, se ha tomado la siguiente muestra utilizando el nuevo procedimiento:

$$n_3 = 41, \quad \bar{x}_3 = 0.92 \times 10^{-4}\ m, \quad S_3^2 = 0.38 \times 10^{-12}\ m^2$$

Con estos datos, ¿puede concluirse que el inventor tiene razón?

———————SOLUCIÓN———————

a)

Realizaremos los siguientes dos contrastes de hipótesis:

$H_o : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

El estadístico correspondiente es, por ejemplo (podría usarse también el inverso dada la simetría del problema)

$$F_1 = \frac{S_2^{*2}}{S_1^{*2}} \equiv F_{30,40}$$

que en nuestro caso vale $F_1 = 1.512$.

La región crítica está obviamente localizada en ambas colas de la distribución F. Las regiones de aceptación correspondientes a distintos valores de α son, para el estadístico F_1:

$\alpha = 0.1 \qquad R_A = [0.5581, 1.7444]$

$\alpha = 0.05 \qquad R_A = [0.4978, 1.9429]$

$\alpha = 0.01 \qquad R_A = [0.3962, 2.4015]$

De hecho, el nivel p es del 22 %, por lo que aceptaríamos las dos hipótesis primarias a cualquier nivel de significación habitual.

b)

Nuestro test de hipótesis sería, en este caso

$H_o : \sigma_3 \leq \frac{1}{2}\sigma_1$

$H_1 : \sigma_3 > \frac{1}{2}\sigma_1$

Sabemos que

$$F = \frac{S_3^{*2}/\sigma_3^2}{S_1^{*2}/\sigma_1^2} \equiv F_{40,40}$$

y por lo tanto, en nuestra hipótesis primaria

$$F = \frac{S_3^{*2}}{S_1^{*2}} \frac{\sigma_1^2}{\sigma_3^2} = 4\frac{S_3^{*2}}{S_1^{*2}} = 4 \times 0.38 = 1.52$$

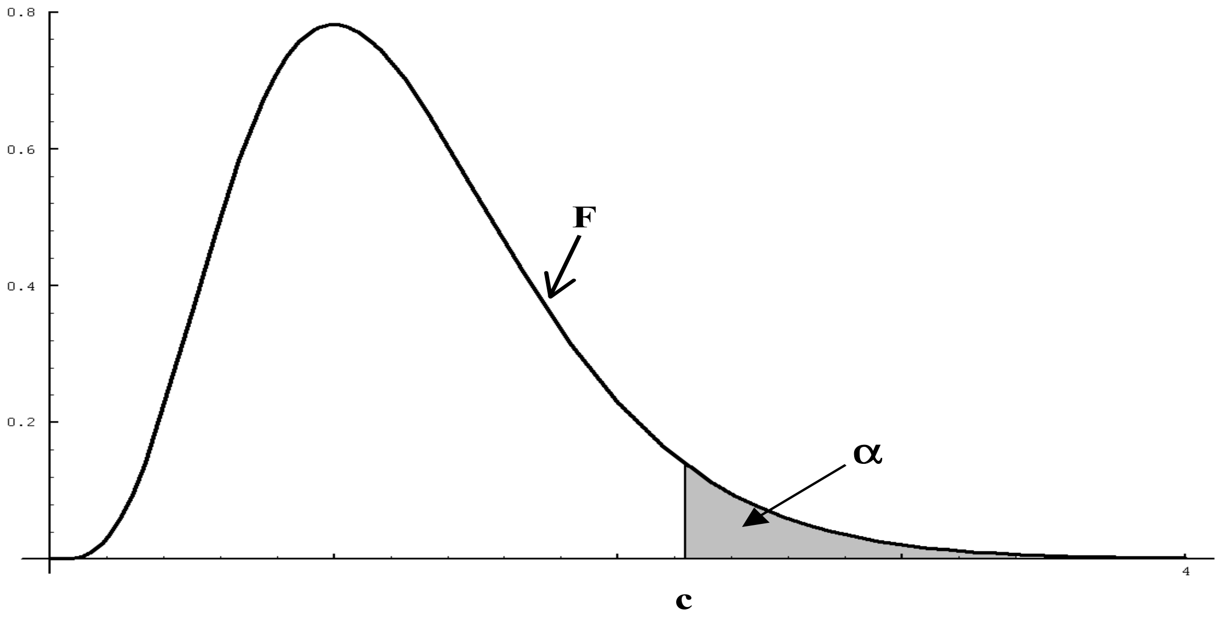
Los valores críticos correspondientes a una distribución $F_{40,40}$ son

$\alpha = 0.1 \qquad c = 1.51$

$\alpha = 0.05 \qquad c = 1.69$

$\alpha = 0.01 \qquad c = 2.11$

por lo que aceptaremos la hipótesis al 5 % y al 1 % de significación, y la rechazaremos al 10 %. De hecho, el nivel p vale 0.0949. Por lo tanto, no podemos confiar excesivamente en lo que dice el inventor.



6.– La duración media de una muestra de 10 conciertos recientes de los Rolling Stones es $\bar{x} = 2.2\ h$, con una desviación típica muestral de $S_x = 25\ min$. La duración media de una muestra de 12 conciertos antiguos de los Beatles es $\bar{y} = 2.15\ h$ con una desviación típica muestral de $S_y = 22\ min$.

1.- ¿Puede aceptarse con un nivel de significación 0.05 que las varianzas de la duración de los dos tipos de concierto son iguales?

2.- ¿Eran más largos los conciertos de los Beatles?

——————————SOLUCIÓN——————————

a) Vamos a realizar el siguiente contraste de hipótesis

$H_o : \sigma_x = \sigma_y$

$H_1 : \sigma_x \neq \sigma_y$

En H_o^1 sabemos que $Z = S_x^{*2}/S_y^{*2} \equiv F(n_x - 1, n_y - 1) = F(9, 11)$, donde $F()$ representa la distribución F de Fisher. El contraste que planteamos es por ambos lados, luego la región crítica estará a ambos lados de la distribución de Z con un área en cada parte de la RC de 0.025 Es decir, debemos calcular $F_Z(c_1) = 0,975$ y $F_Z(c_2) = 0,025$. De las tablas o de la calculadora obtenemos

$$c_1 = 3.5879, \qquad c_2 = 0.2556$$

Luego la región de aceptación de la hipótesis primaria es $RA = [0.2556, 3.5879]$. De los datos obtenemos las varianzas muestrales insesgadas de la siguiente forma:

$$S_x^{*2} = \frac{n_x}{n_x - 1} S_x^2 = 0.1929\ h^2 \text{ y equivalentemente } S_y^{*2} = \ 0,1467 h^2$$

con lo que nuestros estadísticos son

$$S_x^{*2}/S_y^{*2} = 1.3152 \in [0.2556, 3.5879]$$

Luego admitimos la hipótesis primaria a ese nivel de significación (de hecho, el nivel p es de 65.76 %.)

b) En este caso, nuestro contraste de hipótesis es

$H_o : m_y \geq m_x$

$H_1 : m_y < m_x$

No conocemos las varianzas, pero por lo obtenido en el apartado anterior podemos suponer que son iguales. Por tanto, en H_o, nuestro estadístico puede ser

$$t = \frac{(\bar{y} - \bar{x}) - (m_y - m_x)}{R\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{(\bar{y} - \bar{x})}{R\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \equiv t_{n_x+n_y-2}$$

donde

$$R^2 = \frac{(n_x - 1)S_x^{*2} + (n_y - 1)S_y^{*2}}{n_x + n_y - 2}$$

Obsérvese que el punto de H_o correspondiente a $m_y - m_x = 0$ es el límite inferior del intervalo que define a la hipótesis primaria. Si podemos aceptar que $m_y - m_x = 0$ aceptamos de hecho que $m_y - m_x \geq 0$.

La región critica, a la vista de las hipótesis y del estadístico que estamos utilizando, está en la cola inferior de la distribución $t_{n_x+n_y-2} = t_{20}$. Calculando los parámetros necesarios

$$R^2 = 0.1521, \quad R = 0.3900, \quad t = -0.3477$$

y entrando en las tabla de la t_{20} obtenemos que el nivel p vale $p = 36.6\,\%$. Es decir, en este caso, y a pesar de que los datos parece que sugieren lo contrario, podemos afirmar, con un 36.6 % de significación, que los conciertos de Los Beatles son estadísticamente más largos que los de los Rolling Stones

Obsérvese que si hubiéramos realizado el contraste complementario (los conciertos de los Rolling duran más que los de los Beatles), obtendríamos el siguiente resultado

$H_o : m_x \geq m_y$

$H_1 : m_x < m_y$

$$t = \frac{(\bar{x} - \bar{y}) - (m_y - m_x)}{R\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = 0.3477$$

La región crítica sigue estando en la cola inferior de la distribucióm t_{20}. Y por lo tanto, el nivel p es, en este caso, $p = 63.3\,\%$, como era obvio. Este resultado ya se adapta más a lo que nos dicen los datos por lo que se vuelve a comprobar que, en estadística, todo análisis factible debe ser realizado.