

## Tarea 2

**Ejercicio 1.** Sea  $(X_1, X_2, X_3)$  una muestra aleatoria simple de una variable  $X \sim B(1, p)$  con  $p \in (0, 1)$ . Probar que el estadístico  $X_1 + 2X_2 + 3X_3$  no es suficiente.

*Solución.* Sea  $T = X_1 + 2X_2 + 3X_3$ . Como  $X_i \sim \text{Bin}(1, p)$ , los valores que toma  $X_i$  son 1, con probabilidad  $p$ , y 0, con probabilidad  $1 - p$ ,  $i \in \{1, 2, 3\}$ . Se tiene que

$$P(T = 3) = P(X_1 = 1, X_2 = 1, X_3 = 0) + P(X_1 = 0, X_2 = 0, X_3 = 1) = p^2(1 - p) + p(1 - p)^2$$

Por tanto,

$$\begin{aligned} P(X_1 = 1, X_2 = 1, X_3 = 0 | T = 3) &= \frac{P(X_1 = 1, X_2 = 1, X_3 = 0, T = 3)}{P(T = 3)} \\ &= \frac{P(X_1 = 1, X_2 = 1, X_3 = 0)}{P(T = 3)} \\ &= \frac{P(X_1 = 1)P(X_2 = 1)P(X_3 = 1)}{P(T = 3)} \\ &= \frac{p^2(1 - p)}{p^2(1 - p) + p(1 - p)^2} \end{aligned}$$

Como este cociente depende de  $p$ , entonces  $T$  no es un estadístico suficiente. □

**Ejercicio 2.** Aplicando el teorema de factorización, encontrar un estadístico suficiente para cada una de las siguientes familias de distribuciones (en las familias biparamétricas, suponer el caso de solo un parámetro desconocido).

(a)  $U(-\theta/2, \theta/2)$  con  $\theta > 0$ .

(b)  $Ga(p, a)$  con  $p > 0$  y  $a > 0$ .

*Solución.*

(a) Sea  $X \sim U(-\theta/2, \theta/2)$  y sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de  $X$ . Entonces

$$L(X_1, \dots, X_n; \theta) = \prod_{i=1}^n f_\theta(X_i) \mathcal{I}_{(-\theta/2, \theta/2)}(X_i) = \prod_{i=1}^n \frac{1}{\frac{\theta}{2} - (-\frac{\theta}{2})} \mathcal{I}_{(0, \theta/2)}(|X_i|) = \frac{1}{\theta^n} \mathcal{I}_{(0, \theta/2)}\left(\max_{i=1, \dots, n} |X_i|\right)$$

Sean

$$T = \max_{i=1, \dots, n} |X_i|, \quad h(X_1, \dots, X_n) = 1, \quad g(T; \theta) = \frac{1}{\theta^n} \mathcal{I}_{(0, \theta/2)}(T)$$

Como  $h$  y  $g$  son no negativas y  $L(X_1, \dots, X_n; \theta) = g(T; \theta)h(X_1, \dots, X_n)$ , entonces, por el teorema de factorización,  $T$  es un estadístico suficiente para  $\theta$ .

(b) Supongamos que  $a$  es desconocido y  $p$  conocido. Entonces

$$L(X_1, \dots, X_n; a) = \prod_{i=1}^n \frac{a^p}{\Gamma(p)} X_i^{p-1} e^{-aX_i} \mathcal{I}_{(0, \infty)}(X_i) = \frac{a^{np}}{\Gamma(p)^n} e^{-a \sum_{i=1}^n X_i} \prod_{i=1}^n X_i^{p-1} \mathcal{I}_{(0, \infty)}(X_i)$$

Sean

$$T = \sum_{i=1}^n X_i, \quad h(X_1, \dots, X_n) = \prod_{i=1}^n X_i^{p-1} \mathcal{I}_{(0, \infty)}(X_i), \quad g(T; a) = \frac{a^{np}}{\Gamma(p)^n} e^{-aT}$$

Como  $h$  y  $g$  son no negativas y  $L(X_1, \dots, X_n; a) = g(T; a)h(X_1, \dots, X_n)$ , entonces, por el teorema de factorización,  $T$  es un estadístico suficiente para  $a$ . □

**Ejercicio 3.** Probar que las siguientes familias de distribuciones son exponenciales uniparamétricas y, considerando una muestra aleatoria simple de una variable con distribución en dicha familia, obtener, si existe, un estadístico suficiente.

(a)  $B(k_0, p)$  con  $p \in (0, 1)$ .

(b)  $P(\lambda)$  con  $\lambda > 0$ .

(c)  $Exp(\lambda)$  con  $\lambda > 0$ .

*Solución.*

(a) La distribución de una variable aleatoria  $X \sim B(k_0, p)$  viene dada por

$$\begin{aligned} P_p(X = x) &= \binom{k_0}{x} p^x (1-p)^{k_0-x} \mathcal{I}_{\{0,1,\dots,k_0\}}(x) = \binom{k_0}{x} e^{\log(p^x (1-p)^{k_0-x})} \mathcal{I}_{\{0,1,\dots,k_0\}}(x) \\ &= \binom{k_0}{x} e^{x \log(p) + (k_0-x) \log(1-p)} \mathcal{I}_{\{0,1,\dots,k_0\}}(x) = \binom{k_0}{x} e^{x \log(p) + k_0 \log(1-p) - x \log(1-p)} \mathcal{I}_{\{0,1,\dots,k_0\}}(x) \\ &= \binom{k_0}{x} e^{\log((1-p)^{k_0})} e^{x \log \frac{p}{1-p}} \mathcal{I}_{\{0,1,\dots,k_0\}}(x) = \binom{k_0}{x} (1-p)^{k_0} e^{x \log \frac{p}{1-p}} \mathcal{I}_{\{0,1,\dots,k_0\}}(x) \end{aligned}$$

Sean

$$a(p) = (1-p)^{k_0}, \quad b(x) = \binom{k_0}{x} \mathcal{I}_{\{0,1,\dots,k_0\}}(x), \quad c(p) = \log \frac{p}{1-p}, \quad d(x) = x$$

Tenemos que  $a$  y  $b$  son funciones no negativas, y además,

$$P_p(X = x) = a(p)b(x)e^{c(p)d(x)}$$

Esto demuestra que la distribución de  $X$  es exponencial uniparamétrica, y además, dada una muestra aleatoria simple  $(X_1, \dots, X_n)$  de  $X$ , un estadístico suficiente para  $p$  sería

$$T = \sum_{i=1}^n d(X_i) = \sum_{i=1}^n X_i$$

(b) La distribución de una variable aleatoria  $X \sim P(\lambda)$  viene dada por

$$P_\lambda(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \mathcal{I}_{\mathbb{N} \cup \{0\}}(x) = e^{-\lambda} e^{x \log(\lambda)} \frac{1}{x!} \mathcal{I}_{\mathbb{N} \cup \{0\}}(x)$$

Sean

$$a(\lambda) = e^{-\lambda}, \quad b(x) = \frac{1}{x!} \mathcal{I}_{\mathbb{N} \cup \{0\}}(x), \quad c(\lambda) = \log(\lambda), \quad d(x) = x$$

Tenemos que  $a$  y  $b$  son funciones no negativas, y además,

$$P_\lambda(X = x) = a(\lambda)b(x)e^{c(\lambda)d(x)}$$

Esto demuestra que la distribución de  $X$  es exponencial uniparamétrica, y además, dada una muestra aleatoria simple  $(X_1, \dots, X_n)$  de  $X$ , un estadístico suficiente para  $\lambda$  sería

$$T = \sum_{i=1}^n d(X_i) = \sum_{i=1}^n X_i$$

(c) La distribución de una variable aleatoria  $X \sim Exp(\lambda)$  viene dada por

$$P_\lambda(X = x) = \lambda e^{-\lambda x} \mathcal{I}_{(0,\infty)}(x)$$

Sean

$$a(\lambda) = \lambda, \quad b(x) = \mathcal{I}_{(0,\infty)}(x), \quad c(\lambda) = -\lambda, \quad d(x) = x$$

Tenemos que  $a$  y  $b$  son funciones no negativas, y además,

$$P_\lambda(X=x) = a(\lambda)b(x)e^{c(\lambda)d(x)}$$

Esto demuestra que la distribución de  $X$  es exponencial uniparamétrica, y además, dada una muestra aleatoria simple  $(X_1, \dots, X_n)$  de  $X$ , un estadístico suficiente para  $\lambda$  sería

$$T = \sum_{i=1}^n d(X_i) = \sum_{i=1}^n X_i \quad \square$$

**Ejercicio 4.** Sea  $X$  una variable aleatoria con función de densidad de la forma

$$f_\theta(x) = \theta x^{\theta-1}, \quad 0 < x < 1$$

Calcular un estimador de máxima verosimilitud para  $\theta$ .

*Solución.* Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de  $X$  con  $X_i \in (0, 1)$  para todo  $i \in \{1, \dots, n\}$ . Entonces

$$L(X_1, \dots, X_n; \theta) = \prod_{i=1}^n \theta X_i^{\theta-1} = \theta^n \prod_{i=1}^n X_i^{\theta-1}$$

Por tanto,

$$\log(L(X_1, \dots, X_n; \theta)) = n \log(\theta) + (\theta - 1) \sum_{i=1}^n \log(X_i)$$

Derivando e igualando a cero,

$$\frac{\partial \log(L(X_1, \dots, X_n; \theta))}{\partial \theta} = 0 \iff \frac{n}{\theta} + \sum_{i=1}^n \log(X_i) = 0 \iff \theta = -\frac{n}{\sum_{i=1}^n \log(X_i)}$$

Sea  $\hat{\theta} = -\frac{n}{\sum_{i=1}^n \log(X_i)}$ . Entonces

$$\frac{\partial^2 \log(L(X_1, \dots, X_n; \theta))}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} = -\frac{n}{\hat{\theta}^2} < 0$$

Concluimos que  $\hat{\theta}$  es un estimador de máxima verosimilitud para  $\theta$ .  $\square$

**Ejercicio 5.** Sean  $X_1, \dots, X_n$  observaciones independientes de una variable  $X \sim \text{Ga}(p, a)$  con  $p, a > 0$ . Estimar ambos parámetros mediante el método de los momentos.

*Aplicación.* Ciertos neumáticos radiales han tenido vidas útiles de 35200, 41000, 44700, 38600 y 41500 kilómetros. Suponiendo que estos datos son observaciones independientes de una variable aleatoria con distribución exponencial de parámetro  $\theta$ , dar una estimación de dicho parámetro por el método de los momentos.

*Solución.* Se tiene que

$$\alpha_1(p, a) = E(X) = \frac{p}{a}, \quad \alpha_2(p, a) = E(X^2) = \frac{\Gamma(p+2)}{\Gamma(p)a^2} = \frac{p(p+1)}{a^2} = \frac{p^2 + p}{a^2}$$

Por tanto,

$$\begin{aligned} \begin{cases} \alpha_1(p, a) = m_1 \\ \alpha_2(p, a) = m_2 \end{cases} &\iff \begin{cases} p = am_1 \\ \frac{a^2 m_1^2 + am_1}{a^2} = m_2 \end{cases} \iff \begin{cases} p = am_1 \\ a(am_1^2 + m_1 - am_2) = 0 \end{cases} \\ &\iff \begin{cases} p = am_1 = \frac{m_1^2}{m_2 - m_1^2} \\ a = \frac{m_1}{m_2 - m_1^2} \end{cases} \end{aligned}$$

Concluimos que la estimación por el método de los momentos es

$$\hat{p} = \frac{m_1^2}{m_2 - m_1^2}, \quad \hat{a} = \frac{m_1}{m_2 - m_1^2}$$

Según los datos del enunciado, si  $X \sim \text{Exp}(\theta) = \text{Ga}(1, \theta)$  y consideramos la muestra aleatoria simple

$$(X_1, X_2, X_3, X_4, X_5) = (35200, 41000, 44700, 38600, 41500),$$

entonces

$$m_1 = \bar{X} = \frac{1}{5}(35200 + 41000 + 44700 + 38600 + 41500) = 40200$$

Además,

$$m_2 = \overline{X^2} = \frac{1}{5}(35200^2 + 41000^2 + 44700^2 + 38600^2 + 41500^2) \approx 1.63 \cdot 10^9$$

Por tanto, la estimación mediante el método de los momentos de  $\theta$  sería

$$\hat{\theta} = \frac{m_1}{m_2 - m_1^2} \approx 2.88 \cdot 10^{-3}$$

□

**Ejercicio 6.** Sea  $\bar{X}$  la media de una muestra aleatoria simple de tamaño  $n$  de una población  $N(\mu, 16)$ . Encontrar el menor valor de  $n$  para que  $(\bar{X} - 1, \bar{X} + 1)$  sea un intervalo de confianza para  $\mu$  al nivel de confianza 0.9.

*Solución.* Se tiene que

$$\begin{aligned} IC(\mu, 0.9) = (\bar{X} - 1, \bar{X} + 1) &\iff P(\bar{X} - 1 \leq \mu \leq \bar{X} + 1) = 0.9 \\ &\iff P(-1 \leq \bar{X} - \mu \leq 1) = 0.9 \\ &\iff P\left(-\frac{\sqrt{n}}{4} \leq \frac{(\bar{X} - \mu)\sqrt{n}}{4} \leq \frac{\sqrt{n}}{4}\right) = 0.9 \\ &\stackrel{(*)}{\iff} 2P\left(Z \leq \frac{\sqrt{n}}{4}\right) - 1 = 0.9 \\ &\iff P\left(Z \leq \frac{\sqrt{n}}{4}\right) = 0.95, \end{aligned}$$

donde en (\*) se ha denotado  $Z = \frac{(\bar{X} - \mu)\sqrt{n}}{4}$ , y se tiene  $Z \sim N(0, 1)$  por ser  $X \sim N(\mu, 16)$ . Observando la tabla de la distribución normal estándar, debe cumplirse

$$\frac{\sqrt{n}}{4} \approx 1.65,$$

de donde se deduce que

$$n \approx 16 \cdot 1.65^2 = 43.56$$

Se concluye que el menor valor de  $n$  para el que  $(\bar{X} - 1, \bar{X} + 1)$  es un intervalo de confianza para  $\mu$  al nivel de confianza 0.9 es  $n = 44$ . □

**Ejercicio 7.** Dos muestras independientes, cada una de tamaño 7, de poblaciones normales con igual varianza, producen medias 4.8 y 5.4 y cuasivarianzas muestrales 8.38 y 7.62, respectivamente. Encontrar un intervalo de confianza para la diferencia de medias al nivel de confianza 0.95.

*Solución.* Sean  $X$  e  $Y$  variables aleatorias independientes con  $X \sim N(\mu_X, \sigma^2)$ ,  $Y \sim N(\mu_Y, \sigma^2)$ , y sean  $(X_1, \dots, X_7)$ ,  $(Y_1, \dots, Y_7)$  muestras aleatorias simples de  $X$  e  $Y$ , respectivamente, tales que

$$\bar{X} = 4.8, \quad \bar{Y} = 5.4, \quad S_{c,X}^2 = 8.38, \quad S_{c,Y}^2 = 7.62$$

Un intervalo de confianza para la diferencia de medias al nivel de confianza 0.95 sería

$$IC(\mu_X - \mu_Y, 0.95) = \left[ \bar{X} - \bar{Y} - S \cdot T_{1-\alpha/2}, \bar{X} - \bar{Y} + S \cdot T_{1-\alpha/2} \right],$$

donde  $T \sim t_{12}$  (pues  $n + m - 2 = 12$ ), luego  $T_{1-\alpha/2} = T_{0.975} = 2.179$ , y, por otra parte,

$$S = \sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{(n-1)S_{c,X}^2 + (m-1)S_{c,Y}^2}{n+m-2}} = \sqrt{\frac{1}{7} + \frac{1}{7}} \sqrt{\frac{6 \cdot 8.38 + 6 \cdot 7.62}{12}} \approx 1.51186$$

Concluimos que

$$IC(\mu_X - \mu_Y, 0.95) = [4.8 - 5.4 - 1.51186 \cdot 2.179, 4.8 - 5.4 + 1.51186 \cdot 2.179] \approx [-3.89434, 2.69434]$$

es un intervalo de confianza para la diferencia de medias al nivel de confianza 0.95.  $\square$

**Ejercicio 8.** Con objeto de estudiar la efectividad de un agente diurético, se eligen al azar 11 pacientes, aplicando dicho fármaco a seis de ellos y un placebo a los cinco restantes. La variable observada en esta experiencia fue la concentración de sodio en la orina a las 24 horas, que se supone tiene una distribución normal en ambos casos. Los resultados observados fueron:

**DIURÉTICO:** 20.4, 62.5, 61.3, 44.2, 11.1, 23.7.

**PLACEBO:** 1.2, 6.9, 38.7, 20.4, 17.2.

(a) Calcular un intervalo de confianza para el cociente de las varianzas al nivel de confianza 0.95.

(b) Suponiendo que las varianzas son iguales, calcular un intervalo de confianza para la diferencia de las medias al nivel de confianza 0.9.

**Solución.** Consideremos las variables aleatorias

$X \equiv$  concentración de sodio en la orina de los pacientes a los que se les aplica el fármaco;

$Y \equiv$  concentración de sodio en la orina de los pacientes a los que se les aplica el placebo

Tenemos que  $X \sim N(\mu_X, \sigma_X^2)$ ,  $Y \sim N(\mu_Y, \sigma_Y^2)$  y disponemos de las muestras aleatorias simples

$$(X_1, \dots, X_6) = (20.4, 62.5, 61.3, 44.2, 11.1, 23.7), \quad (Y_1, \dots, Y_5) = (1.2, 6.9, 38.7, 20.4, 17.2)$$

de  $X$  e  $Y$ , respectivamente.

(a) El intervalo de confianza pedido es

$$IC\left(\frac{\sigma_X^2}{\sigma_Y^2}, 1-\alpha\right) = \left[ \frac{S_{c,X}^2}{S_{c,Y}^2} \frac{1}{F_{1-\alpha/2}}, \frac{S_{c,X}^2}{S_{c,Y}^2} \frac{1}{F_{\alpha/2}} \right],$$

donde

$$F = \frac{S_{c,X}^2 \sigma_Y^2}{S_{c,Y}^2 \sigma_X^2} \sim F_{n-1, m-1}$$

Como  $n = 6$ ,  $m = 5$  y  $\alpha = 0.05$ , entonces, observando la tabla de la distribución  $F_{5,4}$ ,

$$F_{1-\alpha/2} = F_{0.975} \approx 9.365, \quad F_{\alpha/2} = F_{0.025} \approx 0.135$$

Por otra parte,

$$\bar{X} = 37.2, \quad \bar{Y} = 16.88,$$

luego

$$S_{c,X}^2 = \frac{1}{5} (16.8^2 + 25.3^2 + 24.1^2 + 7^2 + 26.1^2 + 13.5^2) = 483.12,$$

$$S_{c,Y}^2 = \frac{1}{4} (15.68^2 + 9.98^2 + 21.82^2 + 3.52^2 + 0.32^2) = 208.517$$

Por tanto,

$$IC\left(\frac{\sigma_X^2}{\sigma_Y^2}, 0.95\right) = \left[ \frac{483.12}{208.517 \cdot 9.365}, \frac{483.12}{208.517 \cdot 0.135} \right] \approx [0.2474, 17.1625]$$

(b) Supongamos que  $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ . El intervalo de confianza pedido es

$$IC(\mu_X - \mu_Y, 1 - \alpha) = \left[ \bar{X} - \bar{Y} - S \cdot T_{1-\alpha/2}, \bar{X} - \bar{Y} + S \cdot T_{1-\alpha/2} \right],$$

donde  $T \sim t_{n+m-2}$  y

$$S = \sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{(n-1)S_{c,X}^2 + (m-1)S_{c,Y}^2}{n+m-2}}$$

Sustituimos los datos del problema: como  $n = 6$ ,  $m = 5$ , observando la tabla de la distribución  $t_9$ ,

$$T_{1-\alpha/2} = T_{0.95} = 1.833$$

Por otra parte,

$$S = \sqrt{\frac{1}{6} + \frac{1}{5}} \sqrt{\frac{5 \cdot 483.12 + 4 \cdot 208.517}{9}} \approx 11.5063, \quad \bar{X} - \bar{Y} = 37.2 - 16.88 = 20.32$$

Concluimos que

$$IC(\mu_X - \mu_Y, 0.9) = [20.32 - 11.5063 \cdot 1.833, 20.32 + 11.5063 \cdot 1.833] \approx [-0.771, 41.411]$$

es un intervalo de confianza para la diferencia de medias al nivel de confianza 0.9.  $\square$

**Ejercicio 9.** Un profesor asegura que tiene un nuevo método de enseñanza mejor que el tradicional. Para comprobar si tiene razón se seleccionan de forma aleatoria e independiente dos grupos de alumnos, A y B, utilizándose el nuevo método con el grupo A y el tradicional con el B. Al final de curso se hace un examen a los alumnos, obteniéndose las siguientes puntuaciones:

GRUPO A: 6, 5, 4, 7, 3, 5.5, 6, 7, 6.

GRUPO B: 5, 4, 5, 6, 4, 6, 5, 3, 7.

Supuesto que las puntuaciones de cada grupo tienen distribución normal, ¿proporcionan estos datos evidencias para rechazar el nuevo método a nivel de significación 0.05?

*Solución.* Consideremos las variables aleatorias

$X \equiv$  puntuaciones de los alumnos del grupo A;

$Y \equiv$  puntuaciones de los alumnos del grupo B

Tenemos que  $X \sim N(\mu_X, \sigma_X^2)$ ,  $Y \sim N(\mu_Y, \sigma_Y^2)$  y disponemos de las muestras aleatorias simples

$$(X_1, \dots, X_9) = (6, 5, 4, 7, 3, 5.5, 6, 7, 6), \quad (Y_1, \dots, Y_9) = (5, 4, 5, 6, 4, 6, 5, 3, 7)$$

de  $X$  e  $Y$ , respectivamente. Hay que contrastar las hipótesis

$$H_0: \mu_X \geq \mu_Y, \quad H_1: \mu_X < \mu_Y$$

Consideremos el estadístico de contraste

$$T = \frac{(\bar{X} - \bar{Y})\sqrt{n}}{\sqrt{S_c^2}} \sim t_{n-1},$$

donde  $S_c^2$  es la cuasivarianza muestral de la muestra aleatoria simple

$$(X_1 - Y_1, \dots, X_9 - Y_9) = (1, 1, -1, 1, -1, -0.5, 1, 4, -1)$$

de la variable aleatoria  $X - Y$ . Se tiene que

$$\bar{X} - \bar{Y} = \overline{X - Y} = 0.5, \quad S_c^2 = \frac{1}{8} (0.5^2 + 0.5^2 + 1.5^2 + 0.5^2 + 1.5^2 + 1^2 + 0.5^2 + 3.5^2 + 1.5^2) = 2.625$$

Por tanto,

$$T = \frac{0.5 \cdot 3}{\sqrt{2.625}} \approx 0.92582, \quad T_{1-\alpha} = T_{0.95} \approx 1.860$$

Como  $T > -T_{1-\alpha}$ , entonces se debería aceptar la hipótesis  $H_0$ , es decir, la media de las puntuaciones del nuevo método es mejor que la media del método tradicional, así que se debería aceptar el nuevo método.  $\square$

**Ejercicio 10.** A partir de las siguientes observaciones de muestras independientes de dos poblaciones normales, contrastar, al nivel de significación 0.01, si la media de la primera población supera en al menos una unidad la media de la segunda.

MUESTRA 1: 132, 139, 126, 114, 122, 132, 141, 126.

MUESTRA 2: 124, 141, 118, 116, 114, 132, 145, 123, 121.

*Solución.* Sean  $X$  e  $Y$  variables aleatorias independientes con  $X \sim N(\mu_X, \sigma_X^2)$ ,  $Y \sim N(\mu_Y, \sigma_Y^2)$ , y sean

$$(X_1, \dots, X_8) = (132, 139, 126, 114, 122, 132, 141, 126),$$

$$(Y_1, \dots, Y_9) = (124, 141, 118, 116, 114, 132, 145, 123, 121)$$

muestras aleatorias simples de  $X$  e  $Y$ , respectivamente. Hay que contrastar las hipótesis

$$H_0: \mu_X - \mu_Y \geq 1, \quad H_1: \mu_X - \mu_Y < 1$$

Consideramos el estadístico de contraste

$$T = \frac{\bar{X} - \bar{Y} - \mu_X + \mu_Y}{\sqrt{\frac{(n-1)S_{c,X}^2 + (m-1)S_{c,Y}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\bar{X} - \bar{Y} - \mu_X + \mu_Y}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

Se tiene que

- $\bar{X} = \frac{1032}{8} = 129.$
- $\bar{Y} = \frac{1134}{9} = 126.$
- $S_{c,X}^2 = \frac{1}{7}(3^2 + 10^2 + 3^2 + 15^2 + 7^2 + 3^2 + 12^2 + 3^2) \approx 79.1429.$
- $S_{c,Y}^2 = \frac{1}{8}(2^2 + 15^2 + 8^2 + 10^2 + 12^2 + 6^2 + 19^2 + 3^2 + 5^2) = 121.$

Si suponemos  $\mu_X - \mu_Y = 1$ , entonces

$$T = \frac{\bar{X} - \bar{Y} - 1}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \approx \frac{129 - 126 - 1}{\sqrt{\frac{7 \cdot 79.1429 + 8 \cdot 121}{15}} \sqrt{\frac{1}{8} + \frac{1}{9}}} \approx 0.4086$$

Por otra parte, como  $T \sim t_{15}$  y  $\alpha = 0.01$ , en la tabla de la distribución de Student se observa que

$$T_{1-\alpha} = T_{0.99} = 2.602$$

Como  $T > -T_{1-\alpha}$ , entonces aceptamos la hipótesis  $H_0$ , es decir, se puede afirmar que la media de la primera población supera en al menos una unidad la media de la segunda.  $\square$