

EJERCICIO 1. Control 2, año 2003. Nancy Lacourly y Lorena Cerda

En el año 2000, en las salas cunas de la Región Metropolitana se obtuvieron datos sobre 622 niñas y 694 niños recién nacidos. Supongamos que  $X$  e  $Y$  son las variables aleatorias que representan los perímetros del cráneo de las niñas y niños que siguen las distribuciones normales  $N(\mu_1, \sigma_1^2)$  y  $N(\mu_2, \sigma_2^2)$  respectivamente. Se observó que en las muestras, las medias de cada grupo son  $\bar{x} = 34.643\text{cm}$ ,  $\bar{y} = 35.062\text{cm}$  y las varianzas  $S_1^2 = 2.18\text{cm}^2$ ,  $S_2^2 = 2.16\text{cm}^2$ .

a) Se sabe que en la totalidad del país, el promedio del cráneo de los recién nacidos de sexo masculino es 35 cm con una desviación estándar de 1.35 cm. ¿Los recién nacidos de sexo masculino de la RM poseen un perímetro de cráneo significativamente diferentes a la de la totalidad del país?. ¿Cuál es la probabilidad de que se equivoque en su respuesta.?

b) ¿Podemos admitir que  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ?

c) Compare los perímetros del cráneo de los recién nacidos niños y niñas de la RM.

a) Como no se tiene información de la media de los cráneos de niños de la RM la hipótesis nula no puede comparar las dos medias con desigualdades de tipo  $<$ ,  $>$ , sino que sólo estableciendo una diferencia entre las dos medias. En el caso del país, sí se tiene información del valor de su media, luego el planteamiento de las hipótesis queda como sigue:

$$H_0 : \mu_{rm} = \mu_{pais} = 35$$

$$H_1 : \mu_{rm} \neq 35$$

El ejercicio será resuelto calculando la región de rechazo, calcularlo mediante el p-valor queda propuesto... un ejemplo similar es el de la clase del lunes.

La región de rechazo está dada por:

$$W = \{\bar{y} < c_1\} \cup \{\bar{y} > c_2\}$$

sabemos que  $Y \longrightarrow N(\mu_2, \sigma_2^2)$ , como no conocemos  $\sigma_2^2$ , sino que sólo su estimador, debemos construir una t-student.

Se tiene:

$$P\{\bar{y} < c_1/\mu_2 = 35\} = P\{\bar{y} > c_2/\mu_2 = 35\} = 0.025$$

resolviendo el segundo término se tiene:

$$P\left\{\frac{(\bar{y} - \mu_2) \sqrt{n-1}}{S_2} > \frac{(c_2 - \mu_2) \sqrt{n-1}}{S_2} / \mu_2 = 35\right\} = 0.025$$

$$P\left\{t_{n-1} > \frac{(c_2 - 35) \sqrt{n-1}}{S_2}\right\} = 0.025$$

Para una  $P\{t_{n-1} > t\} = 0.025$ ,  $t = 1,97$

$$\Rightarrow \frac{(c_2 - 35) \sqrt{693}}{\sqrt{2.16}} = 1.97$$

despejando la ecuación se obtiene  $c_2 = 35.10$ , análogamente para  $c_1$ .

$$P\left\{t_{n-1} < \frac{(c_1 - 35) \sqrt{n-1}}{S_2}\right\} = 0.025 \Rightarrow P\{t_{n-1} < t\} = 0.025$$

estos se cumple para  $t = -1,97$ , luego se despeja  $c_1 = 34.89$ .

Finalmente la región de rechazo está dada por:

$$\Rightarrow W = \{\bar{y} < 34.89\} \cup \{\bar{y} > 35.109\}$$

Como la media obtenida es  $\bar{y} = 35.062 \notin W \Rightarrow$  aceptamos  $H_0$ , con una probabilidad de un 5% de equivocarnos.

b) Se quiere comparar dos desviaciones estandar, definir la hipótesis nula es bastante trivial, quedaría de la forma  $H_0 : \sigma_1^2 = \sigma_2^2$ , definir la hipótesis alternativa puede traer un poco más de confusión, si bien no se tiene información de las respectivas varianzas se conocen sus respectivos estimadores  $S_1^2$ ,  $S_2^2$  por lo que se puede establecer una comparación con desigualdad.

Finalmente la hipótesis alternativa queda de la forma:  $H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1$

Cuando se quiere comparar dos varianzas o desviaciones estandar en dos poblaciones, se debe formar la distribución de Fisher para calcular la región crítica o el p-valor.

#### 1) REGION DE RECHAZO

Está dada por la  $P(RECHAZAR H_0 | H_0 CIERTO)$

$$\begin{aligned} P\left(\frac{S_1^2}{S_2^2} > C / \frac{\sigma_1^2}{\sigma_2^2} = 1\right) &= \alpha \\ = P\left(\frac{n_1 S_1^2 (n_2 - 1) \sigma_2^2}{\sigma_1^2 (n_1 - 1) n_2 S_2^2} > \frac{C \cdot n_1 (n_2 - 1) \sigma_2^2}{\sigma_1^2 (n_1 - 1) n_2} / \frac{\sigma_1^2}{\sigma_2^2} = 1\right) &= 0.05 \\ = P\left(\frac{n_1 S_1^2 (n_2 - 1)}{(n_1 - 1) n_2 S_2^2} > \frac{C \cdot n_1 (n_2 - 1)}{(n_1 - 1) n_2}\right) &= 0.05 \\ = P(F_{621, 693} > F) &= 0.05 \end{aligned}$$

De la tabla  $F = 1.08 = \left(\frac{C \cdot n_1 (n_2 - 1)}{(n_1 - 1) n_2}\right) = \frac{C \cdot 622 \cdot 693}{694 \cdot 621} = 1.00017 \cdot C$ , despejando C, se obtiene  $C = 1.079$  Finalmente la Región de rechazo queda dada por  $W = \left(\frac{S_1^2}{S_2^2} > 1.079\right)$ , como  $\frac{\sigma_1^2}{\sigma_2^2} = 1.0920$  entonces no pertenece a la región de rechazo, luego se acepta  $H_0$

---

2) P-VALOR

$$\begin{aligned} P\left(\frac{S_1^2}{S_2^2} > C/\frac{\sigma_1^2}{\sigma_2^2} = 1\right) &= \alpha \\ &= P\left(\frac{n_1 S_1^2(n_2 - 1)\sigma_2^2}{\sigma_1^2(n_1 - 1)n_2 S_2^2} > \frac{C \cdot n_1(n_2 - 1)\sigma_2^2}{\sigma_1^2(n_1 - 1)n_2} / \frac{\sigma_1^2}{\sigma_2^2} = 1\right) \\ &= P\left(\frac{n_1 S_1^2(n_2 - 1)}{(n_1 - 1)n_2 S_2^2} > \frac{C \cdot n_1(n_2 - 1)}{(n_1 - 1)n_2}\right) = P(F_{n_1-1, n_2-1} > F) \end{aligned}$$

Al calcular el p-valor, lo que se está haciendo es calcular el error de tipo I, es decir, la  $P(\text{rechazar } H_0/H_0 \text{ cierto})$ , para poder calcular el p-valor, se calcula una aproximación del estadístico que se construyó con la información del enunciado, luego se tendrá:

$$\hat{F} = \frac{n_1 S_1^2(n_2 - 1)}{(n_1 - 1)n_2 S_2^2} = \frac{622 \cdot 693 \cdot 2.18}{621 \cdot 694 \cdot 2.16} = 1.0094$$

Notar que  $\hat{F}$  es un estimador de  $F_{n_1-1, n_2-1}$ , del lado izquierdo de la ecuación y que ya no es necesario calcular  $C$ . El paso siguiente es calcular  $P(F_{621, 693} > 1.009)$ , de las tablas sólo tenemos información de errores de 5% y 1%, la idea es sacar una relación del comportamiento de  $F$  para los grados de libertad que se tienen.

La  $P(F_{621, 693} > 1.08) = 0.05$  y  $P(F_{621, 693} > 1.12) = 0.01$ , entonces de  $P(F_{621, 693} > f) = x\%$  se puede decir que si aumenta el error  $x\%$ , entonces el valor de  $f$  disminuye.

En nuestro caso el valor  $f$  es de 1.009, que es muchísimo menor al  $f=1.08$  que existía con un error del 5% por lo que podemos concluir que el p-valor será más grande que 5% y por tanto no se puede rechazar  $H_0$ , se acepta.

---

## TEST UNIFORMEMENTE MÁS POTENTE (U.M.P)

Breve resumen para planteamiento de región crítica del test U.M.P.

Sea  $x_1, \dots, x_n$  una m.a.s tal que  $X \rightarrow N(\mu, \sigma^2)$

- Al realizar un test con Hipótesis alternativa  $H_1 : \mu < \mu_0$  el test U.M.P. está dado por  $W = (x_1, \dots, x_n)t.q.\bar{x} < c$  independiente de la hipótesis nula.
- Al realizar un test con Hipótesis alternativa  $H_1 : \mu > \mu_0$  el test U.M.P. está dado por  $W = (x_1, \dots, x_n)t.q.\bar{x} > c$  independiente de la hipótesis nula.
- Al realizar un test con Hipótesis alternativa  $H_1 : \sigma^2 < \sigma_0^2$ , con  $\sigma_0^2$  el test U.M.P. está dado por  $W = (x_1, \dots, x_n)t.q.S_n^2 < c$  independiente de la hipótesis nula.
- Al realizar un test con Hipótesis alternativa  $H_1 : \sigma^2 > \sigma_0^2$ , con  $\sigma_0^2$  el test U.M.P. está dado por  $W = (x_1, \dots, x_n)t.q.S_n^2 > c$  independiente de la hipótesis nula.

Sea  $x_1, \dots, x_n$  e  $y_1, \dots, y_m$  dos m.a.s tal que  $X \rightarrow N(\mu_1, \sigma_1^2)$  y  $Y \rightarrow N(\mu_2, \sigma_2^2)$

- Al realizar un test con Hipótesis alternativa  $H_1 : \mu_1 < \mu_2$  el test U.M.P. está dado por  $W = (x_1, \dots, x_n), (y_1, \dots, y_m)t.q.\bar{x} - \bar{y} < c$  independiente de la hipótesis nula.
- Al realizar un test con Hipótesis alternativa  $H_1 : \mu_1 > \mu_2$  el test U.M.P. está dado por  $W = (x_1, \dots, x_n), (y_1, \dots, y_m)t.q.\bar{x} - \bar{y} > c$  independiente de la hipótesis nula.
- Al realizar un test con Hipótesis alternativa  $H_1 : \sigma_1^2 < \sigma_2^2$  el test U.M.P. está dado por  $W = (x_1, \dots, x_n), (y_1, \dots, y_m)t.q.\frac{S_n^2}{S_m^2} < c$  independiente de la hipótesis nula.
- Al realizar un test con Hipótesis alternativa  $H_1 : \sigma_1^2 > \sigma_2^2$  el test U.M.P. está dado por  $W = (x_1, \dots, x_n), (y_1, \dots, y_m)t.q.\frac{S_n^2}{S_m^2} > c$  independiente de la hipótesis nula.