PH para igualdad de vectores de medias de dos Poblaciones $\underline{\mu}_1$, $\underline{\mu}_2$

Sea $\underline{\boldsymbol{x}_1}, \underline{\boldsymbol{x}_2}, \cdots, \underline{\boldsymbol{x}_n}$ una m.a de una población normal p-variada con vector de medias $\underline{\mu}_1$-desconocida y matriz de varianzas-covarianzas $\boldsymbol{\Sigma}_1$, ie. $\underline{\boldsymbol{x}_i} \sim N_p(\underline{\mu}_1, \boldsymbol{\Sigma}_1)$ y sean $\underline{\boldsymbol{y}_1}, \underline{\boldsymbol{y}_2}, \cdots, \underline{\boldsymbol{y}_m}$ una m.a de una población normal p-variada con vector de medias $\underline{\mu}_2$-desconocida y matriz de varianzaz covarianzas $\boldsymbol{\Sigma}_2$, ie. $\underline{\boldsymbol{y}_i} \sim N_p(\underline{\mu}_2, \boldsymbol{\Sigma}_2)$. Ambas m.a son independientes entre si.

1. $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$-Desconocida

Se desea contrastar las hipótesis: $\begin{cases} H_0 & : \ \underline{\mu}_1 - \underline{\mu}_2 = \underline{\boldsymbol{\delta}}_0 \\ & \\ H_a & : \ \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\boldsymbol{\delta}}_0 \end{cases}$

En este caso, se usa como estimador de $\boldsymbol{\Sigma}$ a la varianza ponderada dada por:

$$S_p = \hat{\boldsymbol{\Sigma}} = \frac{(n-1)S_1 + (m-1)S_2}{n+m-2},$$

Bajo H_0-cierto, el estadístico de prueba es:

$$T^2 = \frac{nm}{n+m}\left(\underline{\overline{\boldsymbol{x}}} - \underline{\overline{\boldsymbol{y}}} - \underline{\boldsymbol{\delta}}_0\right)^t \boldsymbol{S}_p^{-1}\left(\underline{\overline{\boldsymbol{x}}} - \underline{\overline{\boldsymbol{y}}} - \underline{\boldsymbol{\delta}}_0\right) \sim \frac{(n+m-2)p}{n+m-p-1} F_{p;\ n+m-p-1} = k$$

$$\text{con:} \quad k = \frac{(n+m-2)p}{n+m-p-1}$$

Rechazamos H_0 si:

$$T_0^2 > kF = \frac{(n+m-2)p}{n+m-p-1} F_{\alpha\ :\ p\ ,\ n+m-p-1}$$

O equivalentemente, rechazamos H_0 si:

$$F_0 = \frac{n+m-p-1}{(n+m-2)p} T_0^2 = \frac{1}{k} T_0^2 > F_{tabla} = F_{\alpha\ :\ p\ ,\ n+m-p-1}$$

Ejemplo-2 Cuatro pruebas psicológicas fueron aplicadas a 32 hombres y 32 mujeres. Las variables consideradas en dicha prueba fueron: X_1: Inconsistencias pictóricas, X_2: Reconocimiento de herramientas, X_3: Forma de emplear el papel y X_4: Vocabulario. Se pretende establecer si la respuesta media para las cuatro variables es similar en hombres y mujeres. El experimento fue llevado a cabo de tal manera que las observaciones para ambos grupos fueron independientes. Se asume además que cada grupo de personas es una m.a de una distribución normal 4-variada, con vectores de media $\underline{\mu}_H$ y $\underline{\mu}_M$ respectivamente, y matriz de varianzas-covarianzas $\mathbf{\Sigma}$-común pero desconocida.

Se desea probar las hipótesis:

$$\begin{cases} H_0 \; : \; \underline{\mu}_H = \underline{\mu}_M \\ \\ H_a \; : \; \underline{\mu}_H \neq \underline{\mu}_M \end{cases} \iff \begin{cases} H_0 \; : \; \underline{\mu}_H - \underline{\mu}_M = \underline{\mathbf{0}} \\ \\ H_a \; : \; \underline{\mu}_H - \underline{\mu}_M \neq \underline{\mathbf{0}} \end{cases} , \quad \underline{\boldsymbol{\delta}}_0 = \underline{\mathbf{0}}$$

La información recopilada permite obtener los siguientes cálculos resúmenes:

$$n = 32\,, \quad \underline{\bar{\boldsymbol{x}}}_H = \begin{bmatrix} 15.97 \\ 15.91 \\ 27.19 \\ 22.75 \end{bmatrix}, \quad \boldsymbol{S}_H = \begin{bmatrix} 5.192 & 4.545 & 6.522 & 5.250 \\ & 13.18 & 6.760 & 6.266 \\ & & 28.67 & 14.47 \\ & & & 16.65 \end{bmatrix}$$

$$m = 32\,, \quad \underline{\bar{\boldsymbol{x}}}_M = \begin{bmatrix} 12.34 \\ 13.91 \\ 16.59 \\ 21.94 \end{bmatrix}, \quad \boldsymbol{S}_M = \begin{bmatrix} 9.136 & 7.549 & 5.531 & 4.151 \\ & 18.60 & 5.446 & 5.466 \\ & & 13.55 & 13.55 \\ & & & 28.00 \end{bmatrix}$$

Usando estos resultados se tiene que la matriz de var-cov ponderada es:

$$S_p = \hat{\boldsymbol{\Sigma}} = \frac{(n-1)S_1 + (m-1)S_2}{n+m-2} = \begin{bmatrix} 7.164 & 6.047 & 6.027 & 4.701 \\ & 15.89 & 6.103 & 5.866 \\ & & 21.11 & 14.01 \\ & & & 22.325 \end{bmatrix}$$

Bajo H_0-cierto, el estadístico de prueba es:

$$T_0^2 = \left[\begin{pmatrix}15.97\\15.91\\27.19\\22.75\end{pmatrix} - \begin{pmatrix}12.34\\13.91\\16.59\\21.94\end{pmatrix}\right]^t$$

$$\left[\left(\frac{1}{32}+\frac{1}{32}\right)\begin{pmatrix}7.164 & 6.047 & 6.027 & 4.701\\ & 15.89 & 6.103 & 5.866\\ & & 21.11 & 14.01\\ & & & 22.325\end{pmatrix}\right]^{-1}\left[\begin{pmatrix}15.97\\15.91\\27.19\\22.75\end{pmatrix} - \begin{pmatrix}12.34\\13.91\\16.59\\21.94\end{pmatrix}\right]$$

$$= 136.67$$

Usando $\alpha = 0.05$, se tiene que:

$$F_{tabla} = F_{\alpha\ ;\ p,n+m-p-1} = F_{0.05\ ;\ 4,32+32-4-1} = F_{0.05\ ;\ 4,59} = 2.53,$$

luego:

$$kF = \frac{(n+m-2)p}{n+m-p-1} F_{\alpha\ ;\ p, n+m-p-1} = \frac{(62)4}{59} F_{0.05\ ;\ 4,59} = 10.635,$$

de donde, como:

$$T_0^2 = 136.67 > 10.635 = kF = \frac{(n+m-2)p}{n+m-p-1} F_{\alpha\ ;\ p, n+m-p-1},$$

entonces se rechaza H_0 y se concluye que los resultados medios son diferentes para hombres y mujeres a un nivel de significancia del 5%.

equivalentemente:

$$F_0 = \frac{1}{k} T_0^2 = \frac{59}{(62)4}(136.67) > 2.53 = F_{tabla}$$

Otro Ejemplo Usando R: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$-Desconocida, Pob. Normal

Resultados usando la función de usuario de Usuario
`HT2_sigmas_iguales()`

```
x<-grupo1[,1:3]
y<-grupo2[,1:3]
mu_0<-c(0,0,0)
res2p<-HT2_sigmas_iguales(x,y,mu_0,0.05)
kable(res2p)
```

T2	k	F0	df1	df2	F_Tabla	Valor_p
0.387667	3.11765	0.124346	3	51	2.78623	0.945298

Resultados Utilizando la función `HotellingsT2` del paquete `ICSNP` del R.

Existen varias funciones en distintos paquetes del R que se utilizan para esta prueba de hipótesis cuando $\Sigma_1 = \Sigma_2 = \Sigma$-Desconocida, Pob. Normal.

```
res_1<-HotellingsT2(x,y)
res_1
```

```
##
##  Hotelling's two sample T2-test
##
## data:  x and y
## T.2 = 0.12435, df1 = 3, df2 = 51, p-value = 0.9453
## alternative hypothesis: true location difference is not equal to c(0,0,0)
```

Estadístico= $F_0 \sim F$	gl_{num}	gl_{den}	p-Valor
0.12435	3	51	0.9453

Resultados Utilizando la función `T2.test` del paquete `rrcov` del R.

```
resT2<-T2.test(x,y)
resT2
```

```
##
##  Two-sample Hotelling test
##
## data:  x and y
## T2 = 0.38767, F = 0.12435, df1 = 3, df2 = 51, p-value = 0.9453
## alternative hypothesis: true difference in mean vectors is not equal to (0,0
## sample estimates:
##                      X1       X2       X3
## mean x-vector 4.504767 4.987719 5.178689
## mean y-vector 4.928179 5.102939 4.947521
```

$T2$	F	df_1	df_2	$Valor - p$
0.38767	0.12435	3	51	0.9453

Resultados Utilizando la función `hotelling.test` del paquete `Hotelling` del R.

```
resh<-hotelling.test(x, y)
resh
```

```
## Test stat:  0.38767
## Numerator df:  3
## Denominator df:  51
## P-value:  0.9453
```

$T2$	$1/K$	df_1	df_2	n_x	n_y	p
0.387666735927368	0.320754716981132	3	51	27	28	3

2. $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$-Desconocida (**Aproximación de: Nel and Van Der Merwe-1986**)

Se desea contrastar las hipótesis:

$$\begin{cases} H_0 & : \ \underline{\mu}_1 - \underline{\mu}_2 = \underline{\boldsymbol{\delta}}_0 \\ H_a & : \ \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\boldsymbol{\delta}}_0 \end{cases}$$

Bajo H_0-cierto, el estadístico de prueba es:

$$T^2 = (\underline{\overline{\boldsymbol{x}}} - \underline{\overline{\boldsymbol{y}}} - \underline{\boldsymbol{\delta}}_0)^t \left[\frac{S_1}{n} + \frac{S_2}{m}\right]^{-1} (\underline{\overline{\boldsymbol{x}}} - \underline{\overline{\boldsymbol{y}}} - \underline{\boldsymbol{\delta}}_0) \sim \frac{vp}{v-p+1} F_{p\ ;\ v-p+1} = kF$$

$$\text{con:} \quad k = \frac{vp}{v-p+1} \quad \text{y} \quad v = \frac{tr(S_e) + \left[tr(S_e)\right]^2}{\sum_{i=1}^{2} \frac{1}{n_i - 1} \left\{ tr(V_i) + [tr(V_i)]^2 \right\}}$$

$$V_i = \frac{S_i}{n_i} \quad \text{y} \quad S_e = V_1 + V_2 = \frac{S_1}{n} + \frac{S_2}{m}.$$

Rechazamos H_0 si: $T_0^2 > kF$ ó $F_0 = \frac{1}{k} T_0^2 > F_{tabla}$

Ejemplo-3 Se compararon dos tipos de suelos, uno de los cuales tiene un tipo de bacteria y el otro no. Las variables aleatorias de interés que fueron medidas son: X_1: PH del suleo, X_2: Cantidad de fosfato y X_3: Contenido de nitrógeno.

Los resultados de las mediciones se muestran en la siguiente tabla. Se desea verificar si hay similitud en ambos suelos, en relación con los vectores de medias asociados a estas tres variables medidas. ¿Cuál es la conclusión usando un $\alpha = 0.05$? $n = 13$ y $m = 10$.

X_1	X_2	X_3	X_1	X_2	X_3	X_1	X_2	X_3
8.0	60	58	8.3	57	60	5.7	42	14
8.0	156	68	7.0	94	43	6.2	40	23
8.0	90	37	8.5	86	40	6.4	49	18
6.1	44	27	8.4	52	48	5.8	31	17
7.4	207	31	7.9	146	52	6.4	31	19
7.4	120	32	6.2	49	30	5.4	62	26
8.4	65	43	5.6	31	23	5.4	42	16
8.1	237	45	5.8	42	22			

Se desea probar las hipótesis:

$$\begin{cases} H_0 \ : \ \underline{\mu}_1 = \underline{\mu}_2 \\ \\ H_a \ : \ \underline{\mu}_1 \neq \underline{\mu}_2 \end{cases} \quad \Longleftrightarrow \quad \begin{cases} H_0 \ : \ \underline{\mu}_1 - \underline{\mu}_2 = \underline{\mathbf{0}} \\ \\ H_a \ : \ \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\mathbf{0}} \end{cases}$$

es decir que: $\underline{\boldsymbol{\delta}}_0 = \underline{\mathbf{0}}$.

Realizando los cálculos respectivos, se obtiene que bajo H_0-cierto, el estadístico de prueba es:

$$T^2 = \left(\underline{\overline{\boldsymbol{x}}} - \underline{\overline{\boldsymbol{y}}} - \underline{\boldsymbol{\delta}}_0\right)^t \left[\frac{S_1}{n} + \frac{S_2}{m}\right]^{-1} \left(\underline{\overline{\boldsymbol{x}}} - \underline{\overline{\boldsymbol{y}}} - \underline{\boldsymbol{\delta}}_0\right) = 96.8178$$

Después de ciertos cálculos, se obtiene que el valor de v es:

$$v = \frac{tr(S_e) + \left[tr(S_e)\right]^2}{\sum_{i=1}^{2} \frac{1}{n_i - 1}\left\{tr(V_i) + [tr(V_i)]^2\right\}} = 12.874 \approx 13$$

Para $\alpha = 0.05$ se tiene que:

$$F_{tabla} = F_{\alpha\ :\ p\ ,\ v-p+1} = F_{0.05;3,13-3+1} = F_{0.05;3,11} = 3.59$$

y

$$kF = \frac{vp}{v-p+1} F_{\alpha\ :\ p\ ,\ v-p+1} = \frac{13(3)}{11} \times 3.59 = 3.54 \times 3.59 = 12.7083,$$

como $T_0^2 = 96.8178 > 12.73 = kF_{tabla}$,

luego rechazamos H_0 y se cuncluye que las características medias son diferentes para ambos tipos de suelo, aun nivel de significancia del 5%.

equivalentemente:

$$F_0 = \frac{1}{k} T_0^2 = \frac{11}{13(3)}(96.18) = 27.1276 > 3.59 = F_{tabla}$$

Usando R. $\Sigma_1 \neq \Sigma_2$-Desconocidas, Pob. Normal

```
##    x1  x2 x3
## 1   8  60 58
## 2   8 156 68
## 3   8  90 37
```

```
##     x1 x2 x3
## 1 6.2 49 30
## 2 5.6 31 23
## 3 5.8 42 22
```

Resultados usando la función de usuario `HT2_sigmas_diferentes()` (**Aproximación de: Nel and Van Der Merwe-1986**).

T_2	v	k	F0	df1	df2	F_Tabla	Valor_p
96.8178	13	3.54545	27.3076	3	11	3.58743	2.14103e-05

Resultados usando la función de usuario `HT2_sigmas_diferentes_texto_guia()` (**Aproximación de: Krishnamoorty and Yu-2004**)

T_2	v	k	F0	df1	df2	F_Tabla	Valor_p
96.8178	70	3.08824	31.3505	3	68	2.7395	7.69607e-13

Otro Ejemplo Usando R. $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$-Desconocidas, Pob. Normal

Resultados usando la función de usuario `HT2_sigmas_diferentes()` (**Aproximación de: Nel and Van Der Merwe-1986**).

```
mu_0<-c(0,0,0)
res_sigmasdif<-HT2_sigmas_diferentes(x,y,mu_0,0.05)
kable(res_sigmasdif)
```

T_2	v	k	F0	df1	df2	F_Tabla	Valor_p
0.386428	53	3.11765	0.123949	3	51	2.78623	0.94554

Otro Ejemplo Usando R. $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$-Desconocidas, Pob. Normal

Resultados usando la función de usuario `HT2_sigmas_diferentes_texto_guia()` (**Aproximación de: Krishnamoorty and Yu-2004**)

Para este caso el v se calcula como sigue:

$$v = \frac{p + p^2}{\frac{1}{n_1}\left\{tr\left[(V_1 S_e^{-1})^2\right] + [tr(V_1 S_e^{-1})]^2\right\} + \frac{1}{n_2}\left\{tr\left[(V_2 S_e^{-1})^2\right] + [tr(V_2 S_e^{-1})]^2\right\}}$$

```
mu_0<-c(0,0,0)
res_sigmasdif1<-HT2_sigmas_diferentes_texto_guia(x,y,mu_0,0.05)
kable(res_sigmasdif1)
```

T_2	v	k	F0	df1	df2	F_Tabla	Valor_p
0.386428	211	3.02871	0.127589	3	209	2.6478	0.943665