Sobre Pruebas de Hipótesis

Evaluación, Generación, Gráficos, PH etc.

Para: Estudiantes de IAM

2021-04-27

LECTURA DE DATOS

A continuación se da una vista previa del conjunto de datos:

Tabla 1: Emcabezado de Datos

	X1	X2	X3	Grupos
1	9.6838	8.7045	4.0509	1
2	5.9703	4.5724	8.3056	1
3	1.5094	6.3666	7.5676	1
4	.	.	.	NA
5	.	.	.	NA
6	.	.	.	NA
98	7.8864	6.0692	6.445	3
99	4.1332	4.5396	6.7288	3
100	4.4509	7.6449	8.2376	3

Resumen de los Datos

Tabla 2: Resumen de Datos

X1	X2	X3	Grupos
Min. :-1.969	Min. :-0.3808	Min. :-0.1823	1:27
1st Qu.: 3.776	1st Qu.: 3.7231	1st Qu.: 3.2236	2:28
Median : 5.186	Median : 5.1071	Median : 4.7204	3:45
Mean : 5.157	Mean : 5.1503	Mean : 4.9258	NA
3rd Qu.: 6.756	3rd Qu.: 6.6321	3rd Qu.: 6.5206	NA
Max. :11.137	Max. : 9.2089	Max. :11.4573	NA

DATOS POR GRUPOS

Tabla 3: Medias por Grupos

Grupos	X1	X2	X3
1	4.504767	4.987718	5.178689
2	4.928179	5.102939	4.947521
3	5.690567	5.277362	4.760589

Tabla 4: Desviaciones Estándar por Grupos

Grupos	X1	X2	X3
1	2.740672	1.809656	2.518015
2	2.634428	2.322268	2.916262
3	1.804084	1.706208	2.403054

Tabla 5: Varianzas por Grupos

Grupos	X1	X2	X3
1	7.511285	3.274856	6.340397
2	6.940210	5.392928	8.504582
3	3.254719	2.911147	5.774668

Tabla 6: Medianas por Grupos

Grupos	X1	X2	X3
1	4.3004	4.57240	4.83410
2	4.6078	5.03415	4.94875
3	5.5045	5.22570	4.36050

Pruebas de Normalidad Multivariadas

Tabla 7: Resultados de NM

Test	Statistic	p value	Result
Mardia Skewness	8.69541680950064	0.561232907640276	YES
Mardia Kurtosis	-0.485009811062382	0.627669386036795	YES
MVN	NA	NA	YES

Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	X1	0.9938	0.9308	YES
Shapiro-Wilk	X2	0.9882	0.5217	YES
Shapiro-Wilk	X3	0.9878	0.4962	YES

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
X1	100	5.156932	2.360555	5.18565	-1.9689	11.1373	3.776300	6.755725	-0.1529182	0.0052095
X2	100	5.150320	1.907849	5.10705	-0.3808	9.2089	3.723100	6.632100	-0.2036691	-0.3835989
X3	100	4.925817	2.565363	4.72040	-0.1823	11.4573	3.223625	6.520600	0.2568454	-0.2877569

Tabla 8: Resultados de NM

Test	Statistic	p value	Result
Mardia Skewness	24.3325702815579	0.00676492265803443	NO
Mardia Kurtosis	1.76902217293395	0.0768901744081527	YES
MVN	NA	NA	NO

Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	X1	0.9840	0.9391	YES
Shapiro-Wilk	X2	0.9521	0.2413	YES
Shapiro-Wilk	X3	0.9786	0.8304	YES

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
X1	27	4.504767	2.740672	4.3004	-1.9689	9.6838	2.76530	6.1330	-0.1252070	-0.4751280
X2	27	4.987718	1.809656	4.5724	-0.3808	8.7045	3.93525	6.1042	-0.4885750	1.0733485
X3	27	5.178689	2.518015	4.8341	0.0305	10.6845	3.84740	6.9798	0.0947103	-0.3876734

Pruebas de Shapito Wilk Multivariada

Método	Paquete	Función	Estadística	p-Valor
Shapiro-Wilk normality test	mvnormtest	mshapiro.test	0.9802882	0.1399768
Multivariate Shapiro-Wilk normality test	RVAideMemoire	mshapiro.test	0.9802882	0.1399768

Prueba M-Box para Matrices de Var-Cov ($\Sigma_1 = \Sigma_2$), $n \geq 20$)

M	U	C	$gl = df$	χ-Tabla	Valor-p
4.16033337146195	0.0613498718844631	3.9050974521261	6	12.591587243744	0.689517640870942

Resultados de esta PH Utilizando la función boxM del paquete biotools del R:

```
Box's M-test for Homogeneity of Covariance Matrices
```

data: datos[1:55, 1:3] Chi-Sq (approx.) = 3.9051, df = 6, p-value = 0.6895

Prueba T2-Hotelling para: $\underline{\mu}_1 - \underline{\mu}_2 = \underline{\delta}_0$.

$\Sigma_1 = \Sigma_2 = \Sigma$-Pob. Normal

Se desea contrastar las hipótesis:

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

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

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

Bajo H_0-cierto, el estadístico de prueba es:

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

con:

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

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:

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

Se crea una función de usuario llamada: *HT2_sigmas_iguales* la cual se utiliza a continuación.

```
mu_0 <- c(0, 0, 0)
# Uso de la función de usuario HT2_sigmas_iguales()
res2p <- HT2_sigmas_iguales(x, y, mu_0, 0.05)
```

$T2$	K	F_0	df_1	df_2	F_{Tabla}	Valor-p
0.387666735927368	3.11764705882353	0.124345934165382	3	51	2.78622881314677	0.945297863947044

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, las cuales se ilustran a continuación.

Resultados de esta prueba Utilizando la función HotellingsT2 del paquete ICNP del R:

```
library(ICSNP)

HotellingsT2(x, y)
```

```
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)

Resultados de esta prueba Utilizando la función T2.test del paquete rrcov del R:

```
library(rrcov)

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 (0,0,0) 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	$Valor - p$
0.3876667	0.1243459	3	51	0.9452979

Tabla 9: Medias de Resultados con T2.test

	X1	X2	X3
mean x-vector	4.504767	4.987718	5.178689
mean y-vector	4.928179	5.102939	4.947521

Resultados de esta prueba Utilizando la función hotelling.test del paquete Hotelling del R:

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

Test stat: 0.12435 Numerator df: 3 Denominator df: 51 P-value: 0.9453

Estadísticas de la función hotelling.test del paquete Hotelling del R

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

Ahora, Para Muestras Grandes, (igualmente para la PH con: $\Sigma_1 = \Sigma_2 = \Sigma$ -Desconocida)

Se desea contrastar las hipótesis:

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

Donde el tamaño de muestra utilizado n**-es grande**.

Bajo H_0-cierto, el estadístico de prueba es:

$$\chi^2 = \frac{nm}{n+m}\left(\underline{\overline{\mathbf{x}}} - \underline{\overline{\mathbf{y}}} - \underline{\delta}_0\right)^t \mathbf{S}_p^{-1}\left(\underline{\overline{\mathbf{x}}} - \underline{\overline{\mathbf{y}}} - \underline{\delta}_0\right) \sim \chi_p^2$$

Rechazamos H_0 si:

$$\chi_0^2 > \chi_{\alpha\ ;\ p}$$

Se crea una función de usuario llamada: *HT2_sigmas_iguales_ngrande* la cual se utiliza a continuación.

```
mu_0 <- c(0, 0, 0)
# Uso de la función de usuario HT2_sigmas_iguales_ngrande()
res2p_ngrande <- HT2_sigmas_iguales_ngrande(x, y, mu_0, 0.05)
```

χ_0^2	df	χ_{Tabla}	Valor-p
0.387666735927368	3	7.81472790325118	0.94277830815779

Resultados de esta PH Utilizando la función HotellingsT2 del paquete ICNP del R:

```
HotellingsT2(x, y, test = "chi")
```

```
Hotelling's two sample T2-test
```

data: x and y T.2 = 0.38767, df = 3, p-value = 0.9428 alternative hypothesis: true location difference is not equal to c(0,0,0)

$\Sigma_1 \neq \Sigma_2$-Pob. Normal

Se desea contrastar las hipótesis:

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

Bajo H_0-cierto, el estadístico de prueba es:

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

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

es decir:

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

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

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

Se crea una función de usuario llamada: *HT2_sigmas_diferentes* la cual se utiliza a continuación.

$T2$	v	$K = vp/(v-p+1)$	F_0	df_1	df_2	F_{Tabla}	Valor-p
0.386428446579277	38	3.16666666666667	0.122030035761877	3	36	2.86626555094018	0.946521235587968

Para este caso de $\Sigma_1 \neq \Sigma_2$-Desconocida, Pob. Normal, **No conozco si existen funciones en R** para realizar dicha prueba, si conocen alguna me la hacen saber por favor.

Prueba de Hipótesis para $\mu = \mu_0$.

Pob. Normal

Se desean contrastar las hipótesis:

$$\begin{cases} H_0 \ : \ \underline{\mu} = \underline{\mu_0} \\ \\ H_a \ : \ \underline{\mu} \neq \underline{\mu_0} \end{cases}$$

El estadístico de prueba es:

$$T^2 = \left(\underline{\overline{\mathbf{x}}} - \underline{\mu}_0\right)^t \left(\frac{1}{n}\mathbf{S}\right)^{-1} \left(\underline{\overline{\mathbf{x}}} - \underline{\mu}_0\right) = n(\underline{\overline{\mathbf{x}}} - \underline{\mu}_0)^t \mathbf{S}^{-1} (\underline{\overline{\mathbf{x}}} - \underline{\mu}_0),$$

Se utiliza el siguiente resultado:

$$T^2 \sim \frac{(n-1)p}{(n-p)} F_{p,n-p} = kF, \quad \text{con:} \ k = \frac{(n-1)p}{(n-p)}$$

o equivalentemente:

$$F = \frac{1}{k} T^2 = \frac{(n-p)}{(n-1)p} T^2 \sim F_{p,n-p},$$

donde, $F_p, n-p$-denota una v.a con distribución F con p y $n-p$ grados de libertad respectivamente.

Al nivel de significancia del $\alpha\%$, rechazamos $H_0 : \underline{\mu} = \underline{\mu}_0$, en favor de: $H_a : \underline{\mu} \neq \underline{\mu}_0$, si el valor de:

$$T_0^2 = n(\underline{\overline{\mathbf{x}}} - \underline{\mu}_0)^t \mathbf{S}^{-1} (\underline{\overline{\mathbf{x}}} - \underline{\mu}_0) > kF = \frac{(n-1)p}{(n-p)} F_{\alpha;p,n-p},$$

o equivalentemente, rechazamos H_0 si:

$$F_0 = \frac{(n-p)}{(n-1)p} T^2 = \frac{1}{k} T_0^2 > F_{\alpha;p,n-p},$$

Se crea una función de usuario llamada: *HT2_mu0* con la cual se obtinen los siguinetes resultados:

```r
mu_0 <- c(0, 0, 0)
# Uso de la función de usuario HT2_mu0()
res_mu0 <- HT2_mu0(grupo1[, 1:3], mu_0, 0.05)
```

$T2$	K	F_0	df_1	df_2	F_{Tabla}	Valor-p
307.063821111781	3.25	94.481175726702	3	24	3.00878657044736	1.98729921407903e-13

En R existen varias funciones en distintos paquetes o librerias, las cuales se utilizan para realizar este tipo de pruebas de hipóteis. A continuación se ilustran varias de ellas.

Se recomienda leer muy bien las ayudas que existen sobre estas funciones para utilizarlas de manera adecuada y definir de de forma apropiada sus respectivos argumentos.

Resultados de esta PH utilizando la función HottellingsT2 del paquete ICSNP del R.

```r
x <- subset(datos1, datos1$Grupos == 1)
x <- x[, 1:3]
HotellingsT2(x)
```

```
Hotelling's one sample T2-test
```

data: x T.2 = 94.481, df1 = 3, df2 = 24, p-value = 1.987e-13 alternative hypothesis: true location is not equal to c(0,0,0)

Resultados de esta PH Utilizando la función T2.test del paquete rrcov del R.

```
T2.test(x)
```

```
One-sample Hotelling test
```

data: x T2 = 307.064, F = 94.481, df1 = 3, df2 = 24, p-value = 1.987e-13 alternative hypothesis: true mean vector is not equal to (0, 0, 0)'

sample estimates: X1 X2 X3 mean x-vector 4.504767 4.987719 5.178689

Prueba de Hipótesis para $\mu = \mu_0$.

n-Grande.

En esta caso, el estadístico de prueba utilizado es:

$$\chi_0^2 = n(\underline{\overline{\mathbf{x}}} - \underline{\mu}_0)^t \mathbf{S}^{-1}(\underline{\overline{\mathbf{x}}} - \underline{\mu}_0) \underset{\text{Bajo } H_0}{\sim} \chi^2_{(p)}$$

La regla de decisión es: Rechazar H_0 si $\chi_0^2 > \chi^2_{\alpha;p}$.

Para este se creo una función de usuario llamada: *HT2_mu0_ngrande* la cual se utiliza a continuación.

```
mu_0 <- c(0, 0, 0)
# Uso de la función de usuario HT2_mu0_ngrande()
res_mu0ngrande <- HT2_mu0_ngrande(grupo1[, 1:3], mu_0, 0.05)
```

χ_0^2	*df*	χ_{Tabla}	Valor-p
307.063821111781	3	7.81472790325118	0

Igualmente, también se puede suar la función *HotellingsT2* del R de la siguiente forma.

Resultados utilizando la Función HotellingsT2 del R

```
HotellingsT2(grupo1[, 1:3], mu = mu_0, test = "chi")
```

```
Hotelling's one sample T2-test
```

data: grupo1[, 1:3] T.2 = 307.06, df = 3, p-value < 2.2e-16 alternative hypothesis: true location is not equal to c(0,0,0)

Prueba T2-Hotelling para contrastes de medias: $C\mu = \delta_0$.

Pob. Normal.

Se desea contrastar las hipótesis:

$$\begin{cases} H_0 \; : \; C\underline{\mu} = \underline{\gamma} \\ \\ H_0 \; : \; C\underline{\mu} \neq \underline{\gamma} \end{cases} \quad , \quad \text{con} \quad \underline{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{pmatrix} - \text{Vector de Constantes.}$$

Un estimador insesgado para $C\underline{\mu}$ es: $C\underline{\overline{\mathbf{x}}}$, el cual tiene la siguiente distribución:

$$C\underline{\overline{\mathbf{x}}} \sim N_k\left(C\underline{\mu}\ ,\ C\boldsymbol{\Sigma}_{\underline{\overline{\mathbf{x}}}}C^T\right) \quad , \quad \text{es decir :}$$

$$C\underline{\overline{\mathbf{x}}} \sim N_k\left(C\underline{\mu}\ ,\ \frac{1}{n}C\boldsymbol{\Sigma}C^t\right), \quad \text{pues :} \quad \boldsymbol{\Sigma}_{\underline{\overline{\mathbf{x}}}} = \frac{\boldsymbol{\Sigma}}{n}.$$

Como $\boldsymbol{\Sigma}$-es desconocida se usa la estadística de prueba:

$$T_0^2 = n\big(C\underline{\overline{\mathbf{x}}} - \underline{\gamma}\big)^t\big[C\mathbf{S}C^t\big]^{-1}\big(C\underline{\overline{\mathbf{x}}} - \underline{\gamma}\big) \sim \frac{(n-1)k}{n-k}F_{k,n-k} = cF$$

Se rechaza H_0 si: $T_0^2 > cF$ ó $F_0 = \frac{1}{c}T_0^2 > F_{tabla}$, $c = \frac{(n-1)k}{n-k}$.

Se crea una función de usuario llamada: *HT2_CU* con la cual se obtinen los siguinetes resultados:

$T2$	c	F_0	df_1	df_2	F_{Tabla}	Valor-p
212.867850345757	2.08	102.340312666229	2	25	3.38518996144917	9.12381281636954e-13

Prueba T2-Hotelling para contrastes de medias: $C\mu = \delta_0$.

n-Grande.

En este caso el estadístico de prueba es:

$$\chi_0^2 = n\big(C\underline{\overline{\mathbf{x}}} - \underline{\gamma}\big)^t\big[C\mathbf{S}C^t\big]^{-1}\big(C\underline{\overline{\mathbf{x}}} - \underline{\gamma}\big) \sim \chi_k^2$$

Se rechaza H_0 si: $\chi_0^2 > \chi_{Tabla} = \chi_{\alpha;k}$

Se crea una función de usuario llamada: *HT2_CU_ngrande* con la cual se obtinen los siguinetes resultados:

χ_0^2	df	χ_{Tabla}	Valor-p
212.867850345757	2	5.99146454710798	0

Prueba de Razón de Ver. de una Matriz de Var-Cov: $\Sigma = \Sigma_0$.

Pob. Normal.

Se tiene interés en la siguiente PH:

$$\begin{cases} H_0 \ : \ \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 \\ \\ H_a \ : \ \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0 \end{cases}$$

La Estadística de Razon de Verosimilitud para esta PH es:

$$\lambda = \frac{|\mathbf{S}|^{\frac{v}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{v}{2}}}\text{Exp}\left\{-\frac{1}{2}\left[v\text{tr}(\mathbf{S}\boldsymbol{\Sigma}_0^{-1}) - vp\right]\right\}$$

y haciendo $\lambda^{\star} = -2\log\lambda$, se tiene que:

$$\lambda^{\star} = v\left[\text{Log}|\mathbf{\Sigma}_0| - \text{Log}|\mathbf{S}| + \text{tr}(\mathbf{S}\mathbf{\Sigma}_0^{-1}) - p\right]$$

Bajo H_0-cierta, se tiene que:

$$\lambda^{\star} \sim \chi^2_k \ , \quad \textit{para} \quad n-1 \ \ \text{grande}$$

con, $k = \frac{p(p+1)}{2}$.

Rechazamos H_0 si.

$$\lambda^{\star} > \chi^2_{\alpha \ ; \ k}$$

Para este se creo una función de usuario llamada: *sigma_sigma0* la cual se utiliza a continuación.

$\lambda_0^{\star}$	*df*	χ_{Tabla}	Valor-p
21.7798732836933	6	12.591587243744	0.00132724441806598