

# Análisis Multivariado I - Práctica 3 - Parte 1

## Test de Hotelling para dos muestras

1. En la primera fase de un estudio sobre el costo de transporte de la leche desde las granjas hasta las lecherías, se tomó una muestra de empresas de transporte vinculadas al transporte de lácteos. En la Tabla 1 y en los archivos `P3-1-ej1-n-2019.txt` y `P3-1-ej1-d-2019.txt` se presentan los datos de costos por milla de

$$x_1 = \text{combustible}$$

$$x_2 = \text{reparaciones}$$

$$x_3 = \text{capital}$$

para  $n_1 = 36$  camiones nafteros y  $n_2 = 23$  camiones a diesel.

- (a) Testear si hay diferencias entre los vectores de costos esperados. Tomar  $\alpha = 0.01$ .
  - (b) Si la hipótesis de igual vector de costos medios es rechazada en la parte (a), hallar la combinación lineal de las componentes de las medias que es más responsable del rechazo.
  - (c) Construir intervalos de confianza de nivel simultáneo 0.99 para los pares de costos medios. Si los hay ¿Qué costos aparecen como muy distintos?
  - (d) Comentar la validez de los supuestos realizados.
2. Grogan y Smith (1981) describen dos especies de mosquitos (midges): *Amerohelea fasciata* (*Af*) y *A. pseudofasciata* (*Apf*) recientemente descubiertas. En la Tabla 2 y en el archivo `P3-1-ej2.txt` aparecen los datos correspondientes a las mediciones de la longitud de las antenas y de las alas de nueve insectos *Af* y seis *Apf*. Definamos dos nuevas variables

$$y_1 = \text{longitud de las antenas} + \text{longitud de las alas}$$

$$y_2 = \text{longitud de las alas.}$$

- (a) Calcular el estadístico  $T^2$  de Hotelling para testear la igualdad entre los vectores de medias de ambos grupos, basado en  $\mathbf{y} = (y_1, y_2)^T$ . Verificar que se obtiene el mismo valor que el del estadístico de Hotelling si uno usaba las variables originales. (Comparar con el ejercicio 3 de la práctica 2, parte 1). Concluir si se acepta o rechaza la hipótesis.
- (b) Mostrar que en los tests de  $t$  univariados de nivel de significación 0.05 realizados sobre cada variable  $y_j$ ,  $j = 1, 2$ , por separado se aceptaría la hipótesis de igualdad de medias.
- (c) Realizar un *scatterplot* de  $y_1$  versus  $y_2$  para los datos, marcando los puntos correspondientes a cada grupo con símbolos distintos, y explicar cómo puede suceder que en los tests univariados la hipótesis de la igualdad de medias es aceptada y sin embargo, en el caso multivariado, la hipótesis de igualdad de medias es claramente rechazada.

- (d) Dibujar la elipse de confianza de nivel 98% para el vector de diferencia de medias y mostrar que no cubre al vector  $\mathbf{0}$ . En el mismo gráfico, dibujar un rectángulo correspondiente a los intervalos de confianza univariados de nivel 99% ( $0.99^2 \cong 0.98$ ) para las diferencias entre las medias de  $y_1$  e  $y_2$ . ¿Cómo se puede ver en este gráfico que es mejor realizar un test de Hotelling multivariado que dos tests de  $t$  univariados?

3. Como parte de un estudio sobre el amor y el matrimonio a una muestra de maridos y esposas se les pidió que respondieran a estas preguntas:

- i. ¿Cuál es el nivel de amor apasionado que siente por su pareja?
- ii. ¿Cuál es el nivel de amor apasionado que su pareja siente por ud.?
- iii. ¿Cuál es el nivel de sentimiento de compañerismo que siente por su pareja?
- iv. ¿Cuál es el nivel de sentimiento de compañerismo que su pareja siente por ud.?

Las respuestas se registraron en una escala de 5 puntos:

1. nada
2. poco
3. algo
4. bastante
5. mucho

Las respuestas de 30 parejas figuran en la Tabla 3 y en el archivo `P3-1-ej3.txt`, donde  $x_i$  = respuesta (en la escala de 1-5) para la pregunta  $i$ -ésima.

- (a) Graficar los valores medios de hombres y mujeres como perfiles muestrales.
- (b) ¿Es el perfil para hombres paralelo al perfil para mujeres? Testear si los perfiles son paralelos con  $\alpha = 0.05$ . Si no se rechaza la hipótesis de paralelismo, testear si los perfiles son coincidentes, para el mismo nivel de significancia. Finalmente, si los perfiles son coincidentes, testear si el perfil común está nivelado, es decir si todas las variables tienen la misma media (siempre con  $\alpha = 0.05$ .)

4. Consideremos los datos que aparecen en las Tablas 4 y 5 y en los archivos `P3-1-ej4-C.txt` y `P3-1-ej4-TNT.txt` y que corresponden a los resultados de tomar un test de habilidad psicolingüística a dos grupos de 27 chicos de edades 8-9 años. Los primeros corresponden a chicos con una enfermedad neonatal (TNT) y los segundos a chicos normales que forman el grupo control. Las variables medidas son

- $x_1$  = recepción auditiva
- $x_2$  = recepción visual
- $x_3$  = memoria visual
- $x_4$  = asociación auditiva
- $x_5$  = memoria auditiva
- $x_6$  = asociación visual
- $x_7$  = oclusión visual
- $x_8$  = expresión oral
- $x_9$  = oclusión gramatical
- $x_{10}$  = destreza manual

Interesa estudiar las siguientes situaciones:

$H_{01}$ : los dos perfiles son similares

$H_{02}$ : los dos perfiles están al mismo nivel

$H_{03}$ : no hay diferencias entre las medias de los tests

- (a) Expresar las hipótesis anteriores matemáticamente.
- (b) Graficar (en un mismo gráfico) las medias de cada grupo en función de las variables, es decir los perfiles.
- (c) Testear  $H_{01}$ .
- (d) En caso de no rechazarse  $H_{01}$ , testear  $H_{03}$  e interpretar el significado de  $H_{01} \cap H_{03}$ . En el caso de rechazarse  $H_{01}$ , testear  $H_{02}$  e interpretar el significado de testear esto.

5. La Tabla 6 contiene las medidas de cinco variables biométricas sobre gorriones hembra, recogidos casi moribundos después de una tormenta. 21 de ellos sobrevivieron mientras que los 28 restantes no lo consiguieron.

Ver los archivos `P3-1-ej5-Superv-2019.txt` y `P3-1-ej5-N0-Superv-2019.txt`. Las variables medidas corresponden a

- $x_1$  = longitud total
- $x_2$  = extensión del ala
- $x_3$  = longitud del pico y de la cabeza
- $x_4$  = longitud del húmero
- $x_5$  = longitud del esternón

Se desea comparar el grupo de los supervivientes con los no supervivientes.

- (a) Testear si hay diferencias entre los vectores de medidas esperadas. Tomar  $\alpha = 0.01$ . Escriba claramente el modelo y los supuestos que está realizando.
- (b) Si la hipótesis de igual vector de medidas esperadas es rechazada en la parte (a), hallar la combinación lineal que es más responsable del rechazo.
- (c) Comentar la validez de los supuestos realizados.

Camiones nafteros			Camiones a diesel		
$x_1$	$x_2$	$x_3$	$x_1$	$x_2$	$x_3$
16.44	12.43	11.23	8.50	12.26	9.11
7.19	2.70	3.92	7.42	5.13	17.15
9.92	1.35	9.75	10.28	3.32	11.23
4.24	5.78	7.78	10.16	14.72	5.99
11.20	5.05	10.67	12.79	4.17	29.28
14.25	5.78	9.88	9.60	12.72	11.00
13.50	10.98	10.60	6.47	8.89	19.00
13.32	14.27	9.45	11.35	9.95	14.53
29.11	15.09	3.28	9.15	2.94	13.68
12.68	7.61	10.23	9.70	5.06	20.84
7.51	5.80	8.13	9.77	17.86	35.18
9.90	3.63	9.13	11.61	11.75	17.00
10.25	5.07	10.17	9.09	13.25	20.66
11.11	6.15	7.61	8.53	10.14	17.45
12.17	14.26	14.39	8.29	6.22	16.38
10.24	2.59	6.09	15.90	12.90	19.09
10.18	6.05	12.14	11.94	5.69	14.77
8.88	2.70	12.23	9.54	16.77	22.66
12.34	7.73	11.68	10.43	17.65	10.66
8.51	14.02	12.01	10.87	21.52	28.47
26.16	17.44	16.89	7.13	13.22	19.44
12.95	8.24	7.18	11.88	12.18	21.20
16.93	13.37	17.59	12.03	9.22	23.09
14.70	10.78	14.58			
10.32	5.16	17.00			
8.98	4.49	4.26			
9.70	11.59	6.83			
12.72	8.63	5.59			
9.49	2.16	6.23			
8.22	7.95	6.72			
13.70	11.22	4.91			
8.21	9.85	8.17			
15.86	11.42	13.06			
9.18	9.18	9.49			
12.49	4.67	11.94			
17.32	6.86	4.44			

Table 1: Costos de Transporte de la Leche. Corresponden a la Tabla 6.6 de Johnson y Wichern (1982)

Especies	Longitud de la Antena (en mm)	Ancho de la Antena (en mm)
Af	1.38	1.64
Af	1.40	1.70
Af	1.24	1.72
Af	1.36	1.74
Af	1.38	1.82
Af	1.48	1.82
Af	1.54	1.82
Af	1.38	1.90
Af	1.56	2.08
Apf	1.14	1.78
Apf	1.20	1.86
Apf	1.18	1.96
Apf	1.30	1.96
Apf	1.26	2.00
Apf	1.28	2.00

Table 2: Datos de mosquitos. Corresponden a la Tabla 1.1 de Flury (1997).

Hombre calificando a su mujer				Mujer calificando a su hombre			
$x_1$	$x_2$	$x_3$	$x_4$	$x_1$	$x_2$	$x_3$	$x_4$
2	3	5	5	4	4	5	5
5	5	4	4	4	5	5	5
4	5	5	5	4	4	5	5
4	3	4	4	4	5	5	5
3	3	5	5	4	4	5	5
3	3	4	5	3	3	4	4
3	4	4	4	4	3	5	4
4	4	5	5	3	4	5	5
4	5	5	5	4	4	5	4
4	4	3	3	3	4	4	4
4	4	5	5	4	5	5	5
5	5	4	4	5	5	5	5
4	4	4	4	4	4	5	5
4	3	5	5	4	4	4	4
4	4	5	5	4	4	5	5
3	3	4	5	3	4	4	4
4	5	4	4	5	5	5	5
5	5	5	5	4	5	4	4
5	5	4	4	3	4	4	4
4	4	4	4	5	3	4	4
4	4	4	4	5	3	4	4
4	4	4	4	4	5	4	4
3	4	5	5	2	5	5	5
5	3	5	5	3	4	5	5
5	5	3	3	4	3	5	5
3	3	4	4	4	4	4	4
4	4	4	4	4	4	5	5
3	3	5	5	3	4	4	4
4	4	3	3	4	4	5	4
4	4	5	5	4	4	5	5

Table 3: Datos Maritales. Corresponden a la Tabla 6.9 de Johnson y Wichern (1983)

Grupo Control

Individuo	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
1	40	32	16	20	38	37	28	43	32	30
2	35	30	41	44	39	38	32	36	41	27
3	30	42	48	26	42	34	30	36	43	36
4	22	27	34	19	40	24	29	14	30	31
5	21	38	46	28	48	33	28	22	34	26
6	39	40	47	37	43	40	34	31	49	42
7	39	39	27	42	36	36	34	47	43	38
8	22	23	34	16	33	30	33	20	21	30
9	44	33	43	31	40	40	32	41	40	27
10	34	34	41	41	51	32	41	46	38	38
11	30	43	34	46	50	34	32	45	38	33
12	26	34	32	20	38	20	26	28	28	33
13	44	42	54	48	54	44	34	52	43	44
14	36	39	49	24	49	36	35	50	36	36
15	30	35	32	28	43	32	34	37	39	24
16	18	25	38	24	32	22	30	8	36	23
17	27	28	35	25	42	25	36	16	30	24
18	30	36	42	28	15	26	32	39	28	23
19	33	16	38	35	51	40	33	40	35	40
20	26	37	54	32	36	41	38	27	37	32
21	31	33	33	32	47	37	32	22	36	28
22	29	31	29	26	38	28	21	27	27	22
23	34	29	40	26	33	31	30	39	34	26
24	36	27	34	21	31	27	23	35	36	37
25	42	40	36	31	41	38	35	41	36	31
26	32	38	37	42	44	49	32	43	32	40
27	38	40	40	32	36	41	36	43	41	28
Media	32.15	33.74	38.30	30.52	40.37	33.89	31.85	34.37	35.67	31.44
SD	7.00	6.58	8.33	8.82	8.09	7.05	4.28	11.57	6.03	6.35

Table 4: Resultados (en 10 categorías) de un test de habilidad psicolingüística (Illinois Test of Psycholinguistic Abilities) para el Grupo Control. Corresponden a la Tabla 3.7 de Seber (1984)

Grupo TNT

Individuo	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
1	26	38	44	26	37	38	22	28	30	33
2	31	30	26	23	43	35	28	24	27	33
3	28	36	36	16	46	28	33	28	28	39
4	19	43	39	16	46	43	33	28	40	28
5	31	33	26	29	41	33	23	30	23	24
6	35	36	43	32	46	22	37	48	34	43
7	37	35	38	31	44	29	37	41	46	31
8	41	37	36	29	32	30	32	39	48	27
9	35	40	34	49	37	34	34	36	34	40
10	29	19	34	42	32	32	30	40	22	37
11	18	26	27	26	21	22	26	12	27	26
12	38	40	27	30	44	40	29	36	42	38
13	31	23	40	28	37	37	30	45	39	24
14	26	21	31	29	41	23	24	22	28	26
15	33	20	44	34	23	37	28	17	39	35
16	27	36	39	19	28	28	23	38	29	36
17	22	32	35	26	41	38	25	24	38	26
18	37	38	36	47	46	38	36	34	32	40
19	24	40	28	16	36	34	27	24	34	38
20	11	27	36	19	38	22	23	31	41	35
21	24	30	30	26	30	36	30	38	35	32
22	28	14	31	36	48	36	27	38	31	41
23	17	25	31	35	46	25	30	42	36	27
24	31	29	34	28	32	31	24	26	29	31
25	34	38	42	31	43	26	27	27	30	30
26	41	46	40	39	44	46	36	42	41	40
27	23	26	37	21	32	34	22	20	31	18
Media	28.78	31.78	34.96	28.63	38.30	32.48	28.74	31.78	33.85	32.52
SD	7.53	8.09	5.51	8.69	7.42	6.50	4.76	9.07	6.64	6.42

Table 5: Resultados (en 10 categorías) de un test de habilidad psicolingüística (Illinois Test of Psycholinguistic Abilities) para el Grupo TNT. Corresponden a la Tabla 3.7 de Seber (1984)



gorriones supervivientes					gorriones no supervivientes				
$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
156	245	31.6	18.5	20.5	155	240	31.4	18.0	20.7
154	240	30.4	17.9	19.6	156	240	31.5	18.2	20.6
153	240	31.0	18.4	20.6	160	242	32.6	18.8	21.7
153	236	30.9	17.7	20.2	152	232	30.3	17.2	19.8
155	243	31.5	18.6	20.3	160	250	31.7	18.8	22.5
163	247	32.0	19.0	20.9	155	237	31.0	18.5	20.0
157	238	30.9	18.4	20.2	157	245	32.2	19.5	21.4
155	239	32.8	18.6	21.2	165	245	33.1	19.8	22.7
164	248	32.7	19.1	21.1	153	231	30.1	17.3	19.8
158	238	31.0	18.8	22.0	162	239	30.3	18.0	23.1
158	240	31.3	18.6	22.0	162	243	31.6	18.8	21.3
160	244	31.1	18.6	20.5	159	245	31.8	18.5	21.7
161	246	32.3	19.3	21.8	159	247	30.9	18.1	19.0
157	245	32.0	19.1	20.0	155	243	30.9	18.5	21.3
157	235	31.5	18.1	19.8	162	252	31.9	19.1	22.2
156	237	30.9	18.0	20.3	152	230	30.4	17.3	18.6
158	244	31.4	18.5	21.6	159	242	30.8	18.2	20.5
153	238	30.5	18.2	20.9	155	238	31.2	17.9	19.3
155	236	30.3	18.5	20.1	163	249	33.4	19.5	22.8
163	246	32.5	18.6	21.9	163	242	31.0	18.1	20.7
159	236	31.5	18.0	21.5	156	237	31.7	18.2	20.3
					159	238	31.5	18.4	20.3
					161	245	32.1	19.1	20.8
					155	235	30.7	17.7	19.6
					162	247	31.9	19.1	20.4
					153	237	30.6	18.6	20.4
					162	245	32.5	18.5	21.1
					164	248	32.3	18.8	20.9

Table 6: Medidas biométricas sobre gorriones