

# Estadística Multivariada

## Tarea 1

Carlos Cuauhtemoc Gutierrez Salazar

Fecha de entrega: Martes 31 de Enero de 2023

**Ejercicio 1.** (30 pts) Demuestre que la matriz de centrado  $\mathbf{P} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$  cumple las siguientes propiedades:

- Tiene rango  $(n - 1)$ , es decir, tiene  $n - 1$  columnas o renglones linealmente independientes
- Sus valores propios son 1 o 0.

**Ejercicio 2.** (10 pts) Suponga que  $\mathbf{X}$  tiene media cero y matriz de covarianza  $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ . Sea  $\mathbf{Y} = \mathbf{X}_1 + \mathbf{X}_2$ . Escriba  $\mathbf{Y}$  como una transformación lineal, es decir, encuentre la matriz de transformación  $\mathbf{A}$  y calcule  $\text{Var}(\mathbf{Y})$

**Ejercicio 3.** (30 pts) Considere las siguientes funciones:

$$\begin{aligned} f_1(x_1, x_2) &= 4x_1x_2 \exp(-x_1^2) & x_1, x_2 > 0 \\ f_2(x_1, x_2) &= 2 & 0 < x_1, x_2 < 1, \text{ y } x_1 + x_2 < 1 \\ f_3(x_1, x_2) &= \frac{1}{2} \exp(-x_1) & x_1 > |x_2| \end{aligned}$$

y diga si son pdf's, si es así calcule  $E(X)$ ,  $\text{Var}(X)$ ,  $E(X_1 | X_2)$  y  $E(X_2 | X_1)$

**Ejercicio 4.** (20 pts) Descargue de moodle los datos denotados como datos\_tarea1.dat, los cuales consisten de 42 mediciones de variables relacionadas con la calidad de aire de la ciudad de Los Angeles. Estos datos fueron registrados a las 12:00 de mediodía en diferentes días. (a) Grafique los diagramas de dispersión marginal para todas las variables. (b) Construya los arreglos  $\hat{\mathbf{x}}$ ,  $\mathbf{S}_n$  y  $\mathbf{R}$  (interprete las entradas de  $\mathbf{R}$ ).

**Ejercicio 5.** (10 pts) Agregue y discuta las actividades 1 y 2 vistas en la sesión de taller.

### Ejercicio 1.

- Tiene rango  $(n - 1)$ , es decir, tiene  $n - 1$  columnas o renglones linealmente independientes

Para una matrix idepotente, su rango es igual a la suma de sus elementos en su diagonal principal, esto es

$$\begin{aligned} &\rightarrow \\ \text{rank}(P) &= \left(1 - \frac{1}{n}\right) \cdot n \\ \text{rank}(P) &= n - 1 \end{aligned}$$

- Sus valores propios son 1 o 0.

$$\begin{aligned} \det(P - \lambda I) &= 0, \\ \det\left(I - \frac{1}{n}J_n - \lambda I\right) &= 0 \\ \text{si } \lambda &= 1 \\ &\rightarrow \det\left(-\frac{1}{n}J_n\right) = 0, \end{aligned}$$

Por propiedad de determinantes, podemos sacar el valor escalar del calculo de la determinante de la siguiente manera

$$B = kA, \det(B) = k^n \det(A) \quad (1)$$

por lo que, sabemos que

$$\begin{aligned} \det\left(-\frac{1}{n}J_n\right) &= 0, \\ \frac{-1}{n} \det(J_n) &= 0 \end{aligned}$$

donde podemos observar que se cumple que 0 es un valor propio de la matriz de centrado

Para  $\lambda = 0$

$$P - \lambda I = \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & & \vdots \\ \vdots & & \ddots & \vdots \\ -\frac{1}{n} & \cdots & \cdots & 1 - \frac{1}{n} \end{pmatrix}$$

Donde podemos observar que si a una fila especifica sumamos todas las demas filas ( $n - 1$  filas), obtenemos una fila de ceros. Por ello un valor propio de la matriz de centrado es el 0.

**Ejercicio 2.**

Buscamos una matriz  $A$  que cumpla  $AX = Y$ . Basta con tomar los valores de cada parametro o  $X_i$ , de manera que si

$$\begin{aligned} Y &= X_1 + X_2, \\ \rightarrow AX &= \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = X_1 + X_2 = Y \end{aligned}$$

Ahora, si sabemos que

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2ab \cdot Cov(X, Y) \quad (2)$$

Entonces

$$\begin{aligned} Var(Y) &= Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2) \\ \rightarrow Var(Y) &= 1 + 2 + 0 = 3 \end{aligned}$$

**Ejercicio 3.**

$$f_1(x_1, x_2) = 4x_1x_2 \exp(-x_1^2) \quad x_1, x_2 > 0$$

- Verificar que sea funcion de probabilidad

$$\begin{aligned} \int_{x_1} \int_{x_2} f(x_1, x_2) d_{x_1} d_{x_2} &= 1 \\ \rightarrow 1 &= \int_{x_1} \int_{x_2} 4x_1x_2 e^{-x_1^2} d_{x_2} d_{x_1} = 2 \int_0^\infty [x_1 e^{-x_1^2}] \frac{2x_2^2}{2} \Big|_0^\infty d_{x_1} \end{aligned}$$

Se puede observar que la integral no converge, por lo que no es funcion de probabilidad conjunta

$$f_2(x_1, x_2) = 2 \quad 0 < x_1, x_2 < 1, \text{ y } x_1 + x_2 < 1$$

Vale notar que el comportamiento para ambas  $x_i$  es el mismo, por lo que tienen el mismo valor esperado, varianza y funciones marginales.

- Verificar que sea funcion de probabilidad

$$\begin{aligned} \int_{x_1} \int_{x_2} f(x_1, x_2) d_{x_1} d_{x_2} &= 1 \\ \rightarrow 1 &= \int_0^1 \int_0^{1-x_1} 2 d_{x_2} d_{x_1} = 2 \int_0^1 (1-x_1) d_{x_1} \\ &= 2 \int_0^1 d_{x_1} - 2 \int_0^1 x_1 d_{x_1} = 2 - 1 = 1 \end{aligned}$$

- Obtener los siguientes valores:  $E[X], Var(X), E(X_1|X_2), E(X_2|X_1)$

$$\begin{aligned} E(X) &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} E[X_1] \\ E[X_2] \end{pmatrix} \\ E[X_1] &= \int_0^1 \int_0^{1-x_1} 2x_1 d_{x_2} d_{x_1} = 2 \int_0^1 x_1 d_{x_1} - 2 \int_0^1 x_1^2 d_{x_1} = 1 - \frac{2}{3} = \frac{1}{3} \\ \rightarrow E(X) &= \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} Var(X) &= \begin{pmatrix} E[(X_1 - \mu_1)^2] & E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] & E[(X_2 - \mu_2)^2] \end{pmatrix} \\ \rightarrow E[(X_1 - \mu_1)^2] &= \int_0^1 \int_0^{1-x_1} 2(x_1 - 1/3)^2 d_{x_2} d_{x_1} \\ (x_1 - 1/3)^2 \cdot (1 - x_1) &= -x_1^3 + \frac{5}{3}x_1^2 - \frac{7}{9}x_1 + \frac{1}{9} \\ E[(X_1 - \mu_1)^2] &= 2 \int_0^1 [-x_1^3 + \frac{5}{3}x_1^2 - \frac{7}{9}x_1 + \frac{1}{9}] d_{x_1} = 2 [-\frac{1}{4} + \frac{5}{9} - \frac{7}{18} + \frac{1}{9}] = \frac{1}{18} \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] &= 2 \int_0^1 \int_0^{1-x_1} (x_1 - \frac{1}{3})(x_2 - \frac{1}{3}) d_{x_2} d_{x_1} \\ u &= x_2 - 1/3 \\ d_u &= d_{x_2} \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] &= \int_0^1 (x_1 - \frac{1}{3}) (x_2 - \frac{1}{3}) \Big|_0^{1-x_1} d_{x_1} = \int_0^1 [(x - \frac{1}{3})(x - \frac{2}{3})^2] \\ &= \int_0^1 [x_1^3 - \frac{5}{3}x_1^2 + \frac{7}{9}x_1 - \frac{1}{9}] d_{x_1} = \frac{1}{4} - \frac{5}{9} + \frac{7}{18} - \frac{1}{9} = \frac{1}{36} \\ \rightarrow Var(X) &= \begin{pmatrix} 1/18 & -1/36 \\ -1/36 & 1/18 \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
E(X_1|X_2) &= \int_{x_1} f(x_1|x_2)dx_1 \\
f(x_1|x_2) &= \frac{f(x_1, x_2)}{f(x_2)} \\
f(x_2) &= \int_{x_1} f(x_1, x_2)dx_1 \\
\rightarrow f(x_2) &= \int_0^{1-x_2} 2dx_1 = 2(1-x_2) \\
x_1 + x_2 &< 1, 0 < x_1, x_2 < 1 \\
\rightarrow f(x_1|x_2) &= \frac{2}{2(1-x_2)} = \frac{1}{1-x_2} \\
x_1 + x_2 &< 1, 0 < x_1, x_2 < 1 \\
E(X_1|X_2) &= \int_0^1 \frac{x_1}{1-x_2} dx_1 \\
\rightarrow E(X_1|X_2) &= \frac{1}{2(1-x_2)} \\
x_1 + x_2 &< 1, 0 < x_1, x_2 < 1 \\
\rightarrow E(X_2|X_1) &= \frac{1}{2(1-x_1)} \\
x_1 + x_2 &< 1, 0 < x_1, x_2 < 1
\end{aligned}$$

Reitero que el comportamiento para ambas  $x_i$  es el mismo, por lo que tienen el mismo valor esperado, varianza y funciones marginales.

$$f_3(x_1, x_2) = \frac{1}{2} \exp(-x_1) \quad x_1 > |x_2|$$

- Verificar que sea función de probabilidad

$$\begin{aligned}
\int_{x_1} \int_{x_2} f(x_1, x_2) dx_1 dx_2 &= 1 \\
\rightarrow 1 &= \int_0^\infty \int_{-x_1}^{x_1} \frac{1}{2} \exp(-x_1) dx_2 dx_1 \\
&= \int_0^\infty x_1 \exp(-x_1) dx_1 (\sim \Gamma(\alpha = 2, \beta = 1)) \\
&= 1
\end{aligned}$$

- Obtener los siguientes valores:  $E[X], Var(X), E(X_1|X_2), E(X_2|X_1)$

$$\begin{aligned}
E(X) &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix} \\
E[X_1] &=
\end{aligned}$$

## Ejercicio 4.

Con las funciones de la libreria de *numpy*, se pueden manejar las matrices sin dificultad, como al usar la funcion para invertir entre otras.

```
1 import pandas as pd
2 import seaborn as sns
3 import numpy as np
4 from itertools import combinations
5
6 df = pd.read_csv(
7     'C:\\MCE 2nd\\Estadistica Multivariada\\Tareas\\Datos_tarea1.txt', sep=' ',
8     , header = None)
9 df.columns = ["a", "b", "c", "d", "e", "f", "g"]
10
11
12 sns.pairplot(df[:])
13
14
15 x = np.array(df)
16 x = x.astype(np.float32)
17
18 n = x.shape[0]
19 p = x.shape[1]
20
21 # 42 observaciones de un vector aleatorio X = (X_1, X_2, X_3, ...)
22
23 x_miu = sum(x)/n
24
25 x1 = x.copy()
26
27 print("X")
28 print(x)
29 for i in range(n):
30     x1[i] -= x_miu
31
32 print("centralized X")
33 print(x1)
34
35 # Sx = E[(x2-mu_x2)(x3-mu_x3)]
36 print("to get covariance matrix S")
37
38 S = np.zeros((p, p))
39
40 for i in range(p):
41     S[i, i] = sum(x1[:, i]**2)/n
42
43 for i in combinations(range(p), 2):
44     # centralized x1
45     a = sum(list(map(lambda x: x[0]*x[1], x1[:, [i[0], i[1]]])))/n
46     # print(i)
47     # print(a)
48     S[i[0], i[1]] = a
49     S[i[1], i[0]] = a
50
51 print(S)
52
53 print("to get matriz diagonal D")
54
55 D = np.zeros((p, p))
56 for i in range(p):
57     D[i, i] = S[i, i]
58 print(D)
59
60 print("to get correlation matrix R")
61
62 D1 = np.linalg.inv(D**0.5)
63
```

```

64 R = np.matmul(np.matmul(D1, S), D1)
65 print(R)
66
67 with open("Tillog.txt", "w") as f:
68     f.write("X\n")
69     f.write(str(x))
70     f.write("\ncentralized X\n")
71     f.write(str(x1))
72     f.write("\nto get covariance matrix S\n")
73     f.write(str(S))
74     f.write("\nto get matriz diagonal D\n")
75     f.write(str(D))
76     f.write("\nto get correlation matrix R\n")
77     f.write(str(R))

```

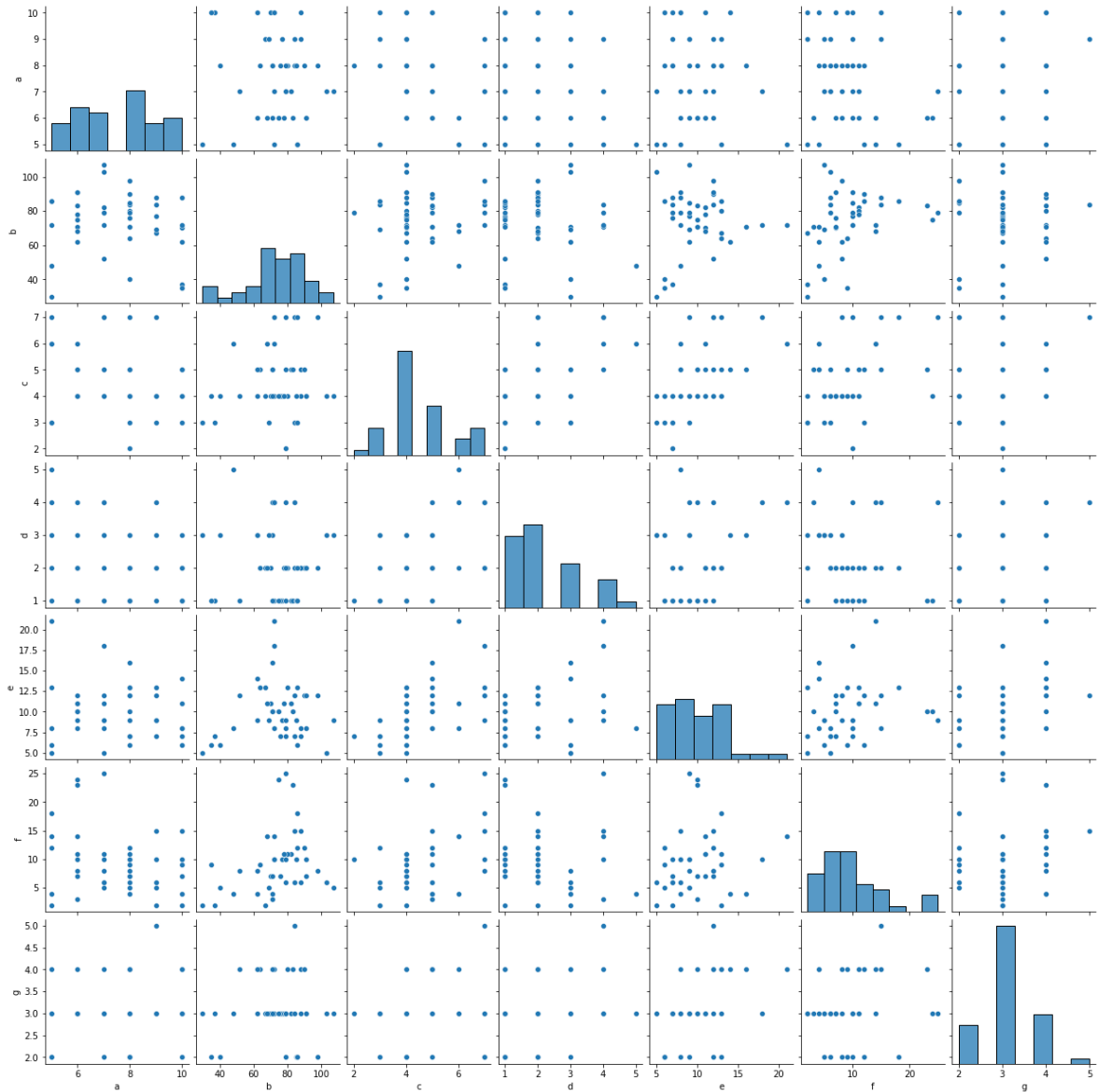


Figura 1: Correlacion muestral en pares de parametros.

Los graficos de correlacion no parecen tener una gran inclinacion de ningun tipo, esto sugiere que puede que no halla una relacion en los parametros de los datos, quiza a excepcion para la segunda variable con la 4ta y la 5ta. Esto se puede visualizar mas facilmente con la matriz de correlaciones.

Vale la pena notar que el maximo valor en la matrix  $R$  es de  $0,55 \sim$  entre la variable  $c$  y  $e$  que asumo corresponden a las variables de  $CO$  y  $NO_3$ .

<https://rpubs.com/bridgetmcgowan/471921>

```
1 import pandas as pd
2 import seaborn as sns
3 df = pd.read_csv(
4     'C:\\MCE 2nd\\Estadistica Multivariada\\Tareas\\Datos_tarea1.txt', sep=' ',
5     , header = None)
6 df.columns = ["a", "b", "c", "d", "e", "f", "g"]
7 sns.heatmap(df.corr(), cmap='PuOr')
```

Resultado en consola del codigo: <https://github.com/gutierrez310/MCE-EstadisticaMultivariada/blob/main/T1-log.txt>



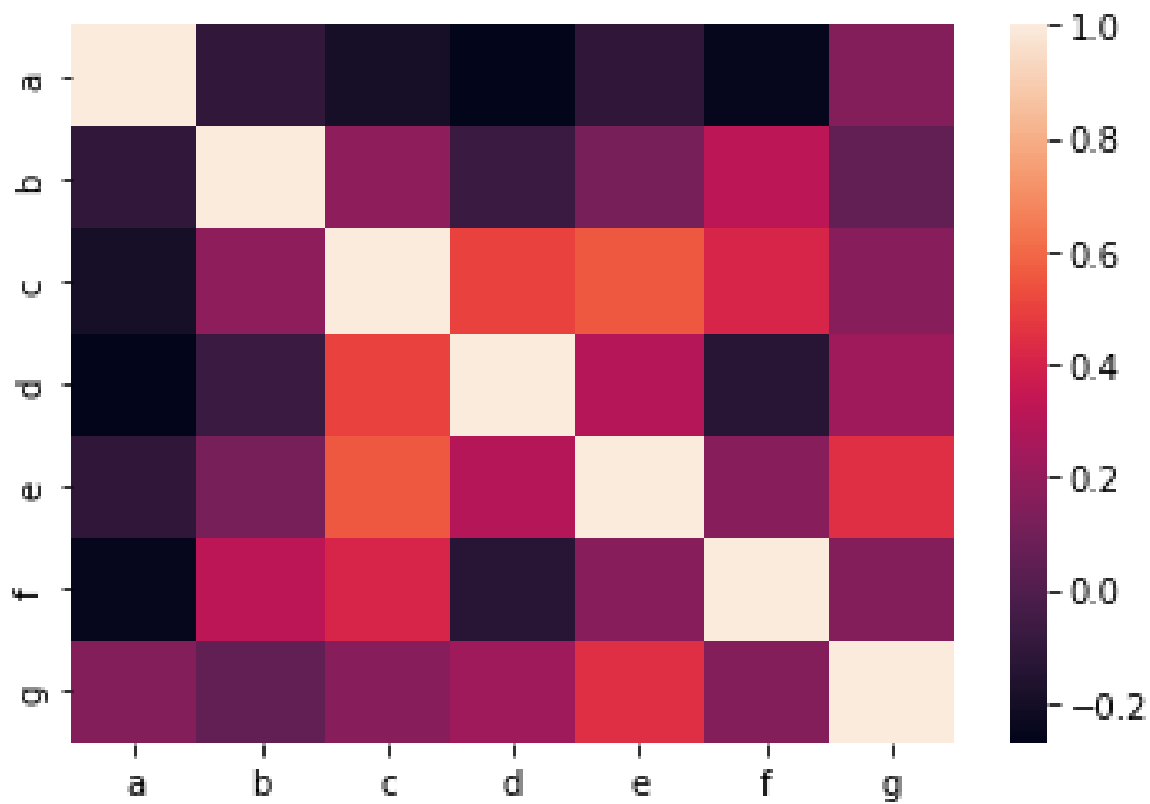


Figura 2: Mapa de calor para la correlacion de los datos

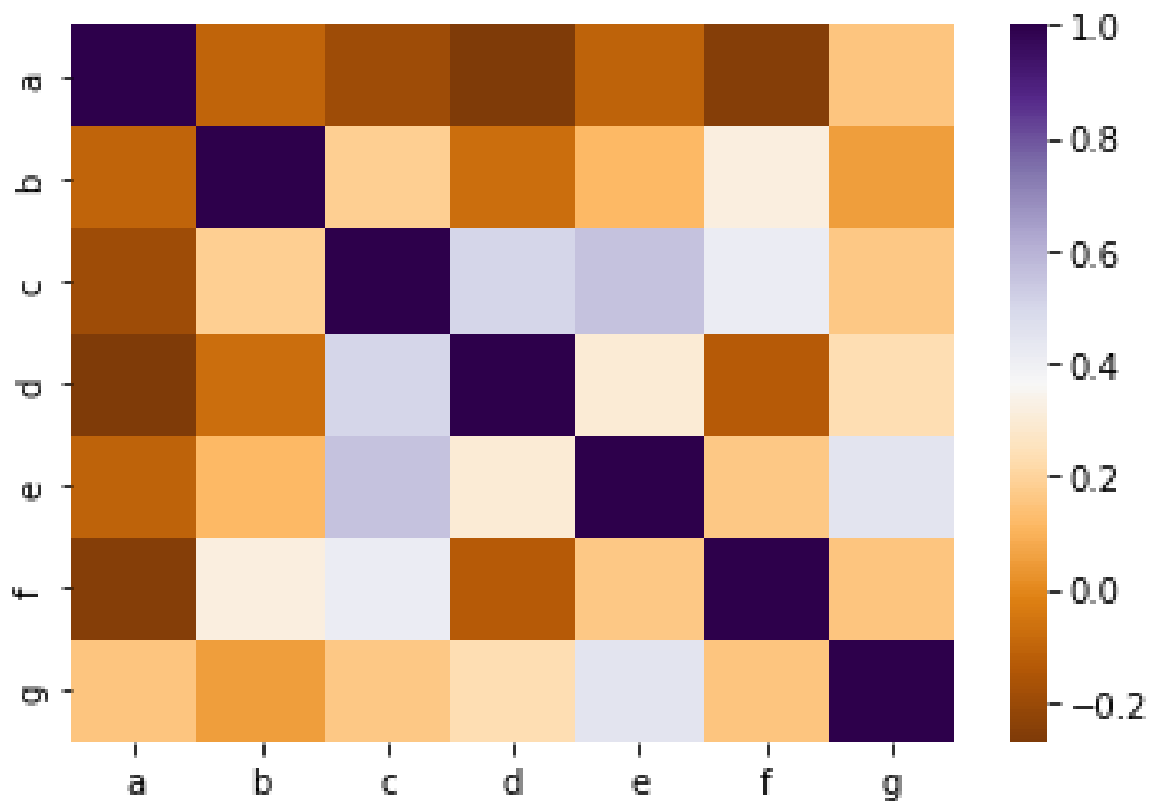


Figura 3: El mismo mapa de calor pero con diferentes colores para facilitar la visualizacion

## Ejercicio 5.

### ■ Ejercicio 1

- (a) Obtenga el vector de medias  $\bar{x}$  y la matriz de covarianzas muestrales  $S_x$
- (b) Obtenga la matriz de covarianzas muestrales de los datos estandarizados a media cero y varianza unidad.
- (c) Sea el vector aleatorio  $\mathbf{Y} = (Y_1, Y_2)$ , donde  $Y_1 = X_1 + 2X_2X_3$  y  $Y_2 = X_1 + X_2$ . Calcule la matriz de observaciones de  $\mathbf{Y}$  mediante una operacion matricial en la que aparezca la matriz de datos  $\mathbf{X}$ .
- (d) Calcule la matriz de covarianzas del vector aleatorio  $\mathbf{Z} = (Z_1, Z_2)$ , donde  $Z_1 = Y_1/\sqrt{6}$  y  $Z_2 = Y_2/\sqrt{2}$
- (e) Calcule las matrices de correlacion de  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $\mathbf{Z}$  y de la matriz de datos obtenida en el apartado (b)

$$X = \begin{pmatrix} -2 & 1 & 4 \\ 3 & 0 & -1 \\ 5 & 1 & 2 \\ -1 & 3 & 6 \\ 2 & -7 & 4 \\ -1 & 0 & -1 \end{pmatrix}$$

A decir verdad, el codigo del *Ejercicio 4* primero fue hecho para este ejercicio, por lo que se reusan ciertas lineas de codigo (el maestro menciona que estaba bien usar codigo para estos ejercicios)

```
1 import numpy as np
2 from itertools import combinations
3
4 x = np.array([[-2., 1., 4.],
5               [3., 0., -1.],
6               [5., 1., 2.],
7               [-1., 3., 6.],
8               [2., -7., 4.],
9               [-1., 0., -1.]])
10
11 n = x.shape[0]
12 p = x.shape[1]
13
14 # 6 observaciones de un vector aleatorio X = (X_1, X_2, X_3)
15
16 x_miu = sum(x)/n
17
18 x1 = x.copy()
19
20 print("X")
21 print(x)
22 for i in range(n):
23     x1[i] -= x_miu
24
25 print("centralized X")
26 print(x1)
27
28 print("standarized X")
29 vars_x = np.var(x, axis=0)*n/(n-1)
30 for i in range(x.shape[0]):
31     x1[i] /= vars_x
32 print(x1)
33
34 # Sx = E[(x2-mu_x2)(x3-mu_x3)]
35 print("to get covariance matrix S")
36
37 x1 = x.copy()
38 for i in range(n):
```

```

39     x1[i] -= x_miu
40
41 S = np.zeros((p, p))
42
43 for i in range(p):
44     S[i, i] = sum(x1[:, i]**2)/n
45
46 for i in combinations(range(p), 2):
47     # centralized x1
48     a = sum(list(map(lambda x: x[0]*x[1], x1[:, [i[0], i[1]]])))/n
49     # print(i)
50     # print(a)
51     S[i[0], i[1]] = a
52     S[i[1], i[0]] = a
53
54 print(S)
55
56 print("to get matriz diagonal D")
57
58 D = np.zeros((p, p))
59 for i in range(p):
60     D[i, i] = S[i, i]
61 print(D)
62
63 print("to get correlation matrix R")
64
65 D1 = np.linalg.inv(D**0.5)
66
67 R = np.matmul(np.matmul(D1, S), D1)
68 print(R)
69
70 # =====
71 # R = np.ones((p, p))
72 #
73 # for i in combinations(range(p), 2):
74 #     R[i[0], i[1]] = S[i[0], i[1]] / D[i[0], i[0]]**0.5 / D[i[1], i[1]]**0.5
75 #     R[i[1], i[0]] = S[i[0], i[1]] / D[i[0], i[0]]**0.5 / D[i[1], i[1]]**0.5
76 # =====
77 print("Y")
78 A = np.array([[ -1, 2, -1], # relaciones con X
79               [1, 1, 0]])
80
81 Y = np.matmul(x, np.transpose(A))
82
83 print(Y)
84
85 times = np.matmul
86
87 print("covariance matrix of Y")
88 S_y = times(times(A, S), np.transpose(A))
89 print(S_y)
90
91 print("to get matriz diagonal D_y")
92 p = len(A)
93 D_y = np.zeros((p, p))
94 for i in range(p):
95     D_y[i, i] = S_y[i, i]
96 print(D_y)
97
98 print("to get correlation matrix R_y")
99
100 D1 = np.linalg.inv(D_y**0.5)
101
102 R_y = np.matmul(np.matmul(D1, S_y), D1)
103 print(R_y)
104
105 print("Z")
106
107 B = np.array([[1/(6**0.5), 0], # relaciones con Y
108               [0, 1/(2**0.5)]])
109

```

```

110 Z = np.matmul(Y, np.transpose(B))
111
112 print(Z)
113
114 print("covariance matrix of Z")
115 S_z = times(times(B, S_y), np.transpose(B))
116 print(S_z)
117
118 print("to get matriz diagonal D_y")
119 p = len(B)
120 D_z = np.zeros((p, p))
121 for i in range(p):
122     D_z[i, i] = S_z[i, i]
123 print(D_z)
124
125 print("to get correlation matrix R_z")
126
127 D1 = np.linalg.inv(D_z**0.5)
128
129 R_z = np.matmul(np.matmul(D1, S_z), D1)
130 print(R_z)
131
132
133 print("std X")
134
135 x1 = x.copy()
136 p = 3
137 for i in range(n):
138     x1[i] -= x_miu
139
140 for i in range(x.shape[0]):
141     x1[i] /= vars_x**0.5
142 print(x1)
143
144 S_x1 = np.zeros((p, p))
145
146 for i in range(p):
147     S_x1[i, i] = sum(x1[:, i]**2)/n
148
149 for i in combinations(range(p), 2):
150     a = sum(list(map(lambda x: x[0]*x[1], x1[:, [i[0], i[1]]]))))/n
151     S_x1[i[0], i[1]] = a
152     S_x1[i[1], i[0]] = a
153
154 print("covariance matrix of X")
155 print(S_x1)
156
157 D_x1 = np.zeros((p, p))
158 for i in range(p):
159     D_x1[i, i] = S_x1[i, i]
160
161 print("to get matriz diagonal D_x")
162 print(D_x1)
163
164 print("to get correlation matrix R_x")
165
166 D1 = np.linalg.inv(D_x1**0.5)
167
168 R_x1 = np.matmul(np.matmul(D1, S_x1), D1)
169 print(R_x1)
170
171
172 with open("T1.1-log.txt", "w") as f:
173     f.write("X\n")
174     f.write(str(x))
175     f.write("\ncentralized X\n")
176     f.write(str(x1))
177     f.write("\nto get covariance matrix S\n")
178     f.write(str(S))
179     f.write("\nto get matriz diagonal D\n")
180     f.write(str(D))

```

```

181 f.write("\nto get correlation matrix R\n")
182 f.write(str(R))
183 f.write("\nY")
184 f.write(str(Y))
185 f.write("\ncovariance matrix of Y\n")
186 f.write(str(S_y))
187 f.write("\nto get matriz diagonal D_y\n")
188 f.write(str(D_y))
189 f.write("\nto get correlation matrix R_y\n")
190 f.write(str(R_y))
191 f.write("\nZ")
192 f.write(str(Z))
193 f.write("\ncovariance matrix of Z\n")
194 f.write(str(S_z))
195 f.write("\nto get matriz diagonal D_y\n")
196 f.write(str(D_z))
197 f.write("\nto get correlation matrix R_z\n")
198 f.write(str(R_z))
199 f.write("std X")
200 f.write(str(x1))
201 f.write("covariance matrix of X")
202 f.write(str(S_x1))
203 f.write("to get matriz diagonal D_x")
204 f.write(str(D_x1))
205 f.write("to get correlation matrix R_x")
206 f.write(str(R_x1))

```

Vale la pena notar que para el calculo de los valores de  $\mathbf{Y}$  en (c), se hace uso de los siguiente:

$$X \cdot A^T = Y^T$$

Resultado en consola del codigo: <https://github.com/gutierrez310/MCE-EstadisticaMultivariada/blob/main/T1.1-log.txt>

## ■ Ejercicio 2

- El diagrama de dispersion de las variables  $(x_1, x_2), (x_1, x_3), (x_2, x_3)$
- $\bar{x}, s_{11}, s_{22}, s_{12}, r_{12}$
- Varianza total, varianza generalizada, desviacion tipica generalizada

```

1 import pandas as pd
2 import seaborn as sns
3 import numpy as np
4 from itertools import combinations
5
6 df = pd.read_csv('C:\\MCE 2nd\\datos_ejercicio_semana_1.csv', sep=',',
7               , header = None)
8 df.columns = ["a", "b", "c"]
9
10 sns.pairplot(df[:])
11
12 x = np.array(df)
13 x = x.astype(np.float32)
14
15 n = x.shape[0]
16 p = x.shape[1]
17
18 print("X")
19 print(x)
20
21 print("mean of X")
22 x_miu = sum(x)/n
23 print(x_miu)
24
25 print("to get covariance matrix S")

```

```

26
27 x1 = x.copy()
28 for i in range(n):
29     x1[i] -= x_miu
30
31 S = np.zeros((p, p))
32
33 for i in range(p):
34     S[i, i] = sum(x1[:, i]**2)/n
35
36 for i in combinations(range(p), 2):
37     # centralized x1
38     a = sum(list(map(lambda x: x[0]*x[1], x1[:, [i[0], i[1]]])))/n
39     # print(i)
40     # print(a)
41     S[i[0], i[1]] = a
42     S[i[1], i[0]] = a
43
44 print(S)
45
46 print("to get matriz diagonal D")
47
48 D = np.zeros((p, p))
49 for i in range(p):
50     D[i, i] = S[i, i]
51 print(D)
52
53 print("to get correlation matrix R")
54
55 D1 = np.linalg.inv(D**0.5)
56
57 R = np.matmul(np.matmul(D1, S), D1)
58 print(R)
59
60 valores_propios, vectores_propios = np.linalg.eig(S)
61
62 print("varianza total")
63 # =====
64 # res=0
65 # for i in range(p):
66 #     res+=S[i,i]
67 # print(res)
68 # =====
69 print(total_var := sum(valores_propios))
70
71 print("varianza generalizada")
72 # =====
73 # res=1
74 # for i in valores_propios:
75 #     res *= i
76 # print(res)
77 # =====
78 print(gen_var := np.linalg.det(S))
79
80 with open("T1.2-log.txt", "w") as f:
81     f.write("X\n")
82     f.write(str(x))
83     f.write("\nmean of X\n")
84     f.write(str(x_miu))
85     f.write("\nto get covariance matrix S\n")
86     f.write(str(S))
87     f.write("\nto get matriz diagonal D\n")
88     f.write(str(D))
89     f.write("\nto get correlation matrix R\n")
90     f.write(str(R))
91     f.write("\nvarianza total\n")
92     f.write(str(total_var))
93     f.write("\nvarianza generalizada\n")
94     f.write(str(gen_var))

```

Ya que el determinante de la matriz de covarianzas, o su varianza generalizada no es 0, sabemos que

ninguna de sus variables tiene dependencia o colinealidad.

La varianza que mayor peso tiene en la varianza total es la de la primera variable o componente, por lo que es el "*primer componente*"

Del grafico de dispersiones junto con la matriz de correlacion podemos observar como entre la variable 1 y variable 3, apodadas ( $a$  y  $c$ ) en mi codigo, existe una correlacion muy cercana a  $-1$ , esto se puede observar en las dispersiones ya que hay una inclinacion en los datos que va del lado superior-izquierdo, al inferior-derecho.

Resultado en consola del codigo: <https://github.com/gutierrez310/MCE-EstadisticaMultivariada/blob/main/T1.2-log.txt>

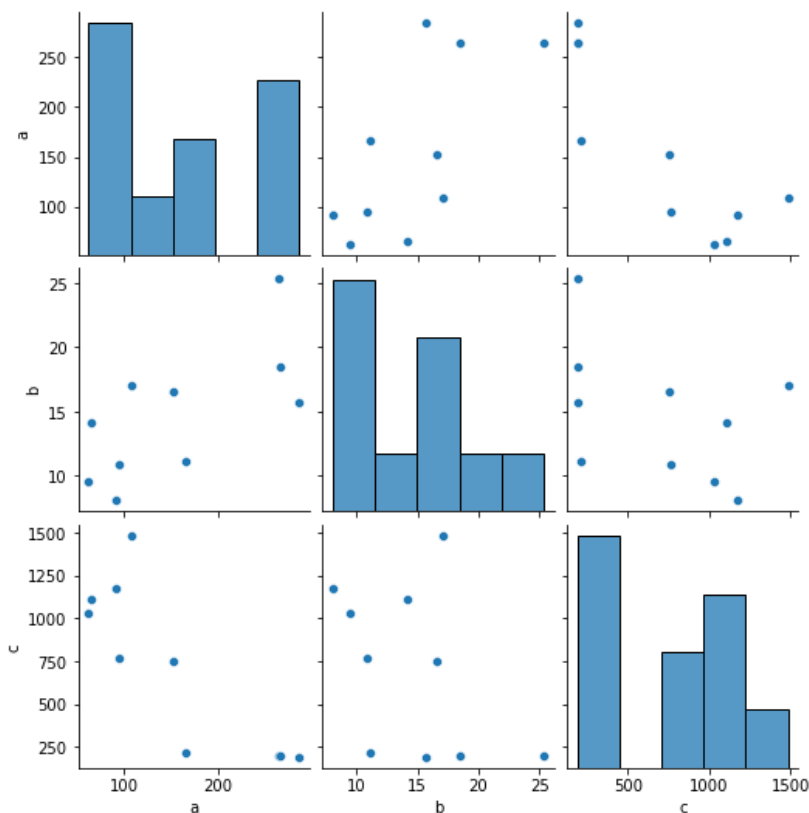


Figura 4: Correlacion muestral en pares de parametros.