Universidade Estadual Do Oeste do Paraná
PGEAGRI
Programa de Pós-Graduação em Engenharia Agrícola

ANÁLISE MULTIVARIADA
Prof.: Dra. Luciana Pagliosa

Resolução da 1ª Lista de Exercícios - 2018

Vanessa Mendes Pientosa
Willyan Goergen de Souza

Cascavel - PR
Setembro de 2018

RESOLUÇÃO

Questão 1) Considerando as matrizes $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix}$ e $\boldsymbol{B} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 1 & 2 & 4 \end{bmatrix}$, calcule:

a) **A** - 3**B**

R= $\begin{bmatrix} -2 & -4 & -6 \\ 1 & -5 & -7 \\ -3 & -5 & -10 \end{bmatrix}$

b) **|AB|**

R= -2

c) **A ⊗ B**

R= $\begin{bmatrix} 1 & 2 & 3 & 2 & 4 & 6 & 3 & 6 & 9 \\ 0 & 2 & 3 & 0 & 4 & 6 & 0 & 6 & 9 \\ 1 & 2 & 4 & 2 & 4 & 8 & 3 & 6 & 12 \\ 1 & 2 & 3 & 1 & 2 & 3 & 2 & 4 & 6 \\ 0 & 2 & 3 & 0 & 2 & 3 & 0 & 4 & 6 \\ 1 & 2 & 4 & 1 & 2 & 4 & 2 & 4 & 8 \\ 0 & 0 & 0 & 1 & 2 & 3 & 2 & 4 & 6 \\ 0 & 0 & 0 & 0 & 2 & 3 & 0 & 4 & 6 \\ 0 & 0 & 0 & 1 & 2 & 4 & 2 & 4 & 8 \end{bmatrix}$

d) **A ⊕ B**

R= $\begin{bmatrix} 1 & 2 & 3 & 0 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 & 4 \end{bmatrix}$

e) **B^{-1} e mostre que $BB^{-1} = I$**

R:

B^{-1}= $\begin{bmatrix} 1.0 & -1.0 & 0.0 \\ 1.5 & 0.5 & -1.5 \\ -1.0 & 0.0 & 1.0 \end{bmatrix}$ **e** $\mathbf{B}\boldsymbol{B}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ = ***I***

f) **Pesquise sobre o produto de Khatri-Rao e calcule A |⊗| B**

R: O produto Khatri-Rao é um produto de Kronecker, denotado por ⊗, é uma operação em duas matrizes de tamanho arbitrário, resultando em uma matriz de blocos, considerando as colunas, sendo introduzido por Khatri e Rao (1968).

Segundo Gomes (2014) Esta solução explora a propriedade de simetria dual do tensor de dados e pode ser aplicada em técnicas de processamento de sinais em arranjos de sensores baseadas em covariância**.** O produto de Khatri-Rao entre as matrizes A ∈ C I×R e B ∈ C J×R, representado por ⬨, é definido como: A⨀B= [a1⊗b1a2⊗b2…ak⊗bk].

$$\mathbf{A} \,|\otimes|\, \mathbf{B} = \begin{bmatrix} 1 & 4 & 9 \\ 0 & 4 & 9 \\ 1 & 4 & 12 \\ 1 & 2 & 6 \\ 0 & 2 & 6 \\ 1 & 2 & 8 \\ 0 & 2 & 6 \\ 0 & 2 & 6 \\ 0 & 2 & 8 \end{bmatrix}$$

<u>Questão 2)</u> Seja a matriz $\boldsymbol{A} = \begin{bmatrix} 9 & -5 \\ -5 & 6 \end{bmatrix}$, responda:

a) Verifique se *A* é positiva definida e escreva sua forma quadrática
R:

A Matriz A é simétrica, considerando que sua Matriz Transposta é análoga à original.

$$\boldsymbol{A^T} = \begin{bmatrix} 9 & -5 \\ -5 & 6 \end{bmatrix}$$

Sua forma quadrática é dada por:

$$x^T * A * x = [x_{(1)} \quad x_{(2)}] * \begin{bmatrix} 9 & -5 \\ -5 & 6 \end{bmatrix} * \begin{bmatrix} x_{(1)} \\ x_{(2)} \end{bmatrix}$$
$$(x^T * A) * x = [9x_{(1)} - 5x_{(2)} \quad -5x_{(1)} + 6x_{(2)}] * \begin{bmatrix} x_{(1)} \\ x_{(2)} \end{bmatrix}$$
$$\boldsymbol{A = [9x_{(1)}^2 - 10x_{(1)x_{(2)}} + 6x_{(2)}^2]}$$

Tendo em vista que $x^{T*}A*x \geq 0$, para todos os vetores não nulos da função acima (domínio: números reais; imagem: números reais não negativos, ou seja, sempre maior ou igual a zero), pode-se afirmar que a mesma é positiva definida.

b) Determine seus autovalores e autovetores (normalizados)
R:

<u>Autovalores:</u>
$\lambda_1 =$ 12,7202
$\lambda_2 =$ 2,2798

<u>Autovetores (normalizados):</u>
$e_1 = \begin{bmatrix} -0{,}8023 \\ 0{,}5969 \end{bmatrix}$ $e_2 = \begin{bmatrix} -0{,}5969 \\ -0{,}8023 \end{bmatrix}$

c) Escreva a decomposição espectral dessa matriz
R:

Sendo V: a matriz ortogonal dos autovetores padronizados de A; e Λ a matriz diagonal dos autovalores de A, a decomposição espectral se dá por:

$A = V * \Lambda * V^T = \begin{bmatrix} -0{,}8023 & -0{,}5969 \\ 0{,}5969 & -0{,}8023 \end{bmatrix} * \begin{bmatrix} 12{,}7202 & 0 \\ 0 & 2{,}2798 \end{bmatrix} * \begin{bmatrix} -0{,}8023 & 0{,}5969 \\ -0{,}5969 & -0{,}8023 \end{bmatrix} =$

$A = V * \Lambda * V^T = \begin{bmatrix} 9 & -5 \\ -5 & 6 \end{bmatrix}$

d) Determine os autovalores e autovetores (normalizados) da inversa dessa matriz
R:

A inversa da Matriz A:

$$\boldsymbol{A^{-1}} = \begin{bmatrix} 0,2069 & 0,1724 \\ 0,1724 & 0,3103 \end{bmatrix}$$

Autovalores:

$\lambda_1 = 0,4386$

$\lambda_2 = 0,0786$

Autovetores (normalizados):

$$e_1 = \begin{bmatrix} 0,5969 \\ 0,8023 \end{bmatrix} \; e_2 = \begin{bmatrix} -0,8023 \\ 0,5969 \end{bmatrix}$$

Questão 3) Numa exploração agrícola da Bélgica registraram-se os valores de 5 variáveis meteorológicas ao longo 11 anos agrícolas na década de 1920. As cinco variáveis são: x1 = precipitação total em Novembro e Dezembro (mm); x2 = temperatura média em Julho (°C); x3 = precipitação total em Julho (mm); x4 = radiação em Julho (mm de álcool); x5 = rendimento médio de colheita (quintais/hectare). Os valores observados foram:

Campanha	x_1	x_2	x_3	x_4	x_5
1920-21	87.9	19.6	101.0	1661	28.37
1921-22	89.9	15.2	90.1	968	23.77
1922-23	153.0	19.7	56.6	1353	26.04
1923-24	132.1	17.0	91.0	1293	25.74
1924-25	88.8	18.3	93.7	1153	26.68
1925-26	200.0	17.8	106.9	1286	24.29
1926-27	117.7	17.8	65.5	1204	28.00
1927-28	109.0	18.3	41.8	1400	28.37
1928-29	156.1	17.8	57.4	1222	24.96
1929-30	181.5	16.8	140.6	902	21.66
1930-31	181.4	17.0	74.3	1150	24.37

a) Construa o vetor de médias amostrais, a matriz de variâncias e covariâncias amostrais, a matriz média de correlação amostral e o gráfico das correlações amostrais. Realize o teste de hipótese para identificar quais correlações são significativas (com 5% de significância). Interprete os resultados obtidos.

R: Os valores para o vetor das médias amostrais são:

$$\mu = \begin{bmatrix} 136.1273 \\ 17{,}75455 \\ 83.5364 \\ 1235{,}6364 \\ 25{,}6591 \end{bmatrix}$$

Usando o vetor de médias amostrais, é possível verificar uma média de 136,1273 de precipitação total nos meses de novembro e dezembro (x1), já a temperatura média em julho foi de 17,75455°C (x2), a precipitação total em julho foi de 83,5364 mm (x3). O valor médio da radiação em julho corresponde a 1235,6364 mm de álcool (x4). O rendimento médio da colheita representa 25,6591 quintais/hectare (x5)

Tabela 1: Tabela da matriz de variância e covariância:

	x1	x2	x3	x4	x5
x1	**1666.6002**	-5.0156	249.4609	-2242.8091	-54.3525
x2	-5.0156	**1.6367**	-10.8242	208.1618	1.7347
x3	249.4609	-10.8242	**787.0406**	-2090.5155	-34.9714
x4	-2242.8091	208.1618	-2090.5155	**42584.2545**	338.0426
x5	-54.3525	1.7347	-34.9714	338.0426	**4.4958**

A matriz de variância e covariância, também conhecida como matriz S (tabela 1), mostrou que a variância para x1 foi de 1666,6002 mm para a precipitação total dos meses de novembro e dezembro.

Para x2 a variância é de 1,6367 °C com relação a temperatura média em julho. A variância para x3 que representa a precipitação total em julho foi de 787,0406 mm, enquanto que a variância de x4 que é a radiação em julho corresponde á 42584,2545 mm de álcool. No caso de x5 que é o rendimento médio de colheita, a variância representa 4,4958 de quintais/hectare.

As covariâncias estão apresentadas nos demais elementos fora da diagonal principal, representando a covariância das variáveis x1-x2, x1-x3, x1-x4, x1-x5, x2-x3, x2-x4, x2-x5, x3-x5 e x4-x5. Destaca-se ainda que existem doze, covariâncias negativas (x1-x2, x1-x4, x1-x5, x2-x1, x2-x3, x3-x2, x3-x4, x3-x5, x4-x1, x4-x3, x5-x1, x5-x3), ou seja, quando os valores que estão acima (abaixo) da média de uma variável tendem a estar relacionados com valores que estão abaixo (acima) da média da outra variável.

As covariâncias positivas, conforme pode ser visto na matriz S, implicam em valores que estão acima (abaixo) da média de uma variável tendem a estar relacionados com valores que também estão acima (abaixo)da média da outra variável.

Tabela 2: Matriz de correlação amostral

	x1	x2	x3	x4	x5
x1	**1.0000**	-0.0960	0.2178	-0.2662	-0.6279
x2	-0.0960	**1.0000**	-0.3016	0.7885	0.6395
x3	0.2178	-0.3016	**1.0000**	-0.3611	-0.5879
x4	-0.2662	0.7885	-0.3611	**1.0000**	0.7726
x5	-0.6279	0.6395	-0.5879	0.7726	**1.0000**

A correlação é uma medida que padroniza duas variáveis. Essa correlação não pode ser maior que 1 e menor que -1. Uma correlação próxima a zero, significa que não existe associação linear entre duas variáveis sendo que uma correlação positiva indica que as duas variáveis se movem juntas e a correlação negativa indica que as duas variáveis se movem em direções opostas.

De acordo com o material da disciplina disponível, os graus de consideração apresentados foram:

r = 0 significa que não há relacionamento entre as variáveis.
r = 0,20 significa que ocorre uma baixa relacionamento entre as variáveis.
r = 0,40 ocorre um moderado relacionamento entre as variáveis.
r = 0,70 significa um alto relacionamento entre as variáveis.
r = 1,00 Perfeito correspondência entre as variáveis.

Através da tabela 2 e do gráfico 1, verifica-se a matriz de correlação amostral. Observa - se que x1-x2 não apresentam correlação, x1 – x3, x1 – x4, x2 – x3 e x3 – x4, apresentaram baixa correlação linear, x1 – x5, x2 – x5, x3 – x5 apresentaram moderada correlação linear, enquanto que x2 – x4 e x4 – x5 apresentaram alta correlação linear.

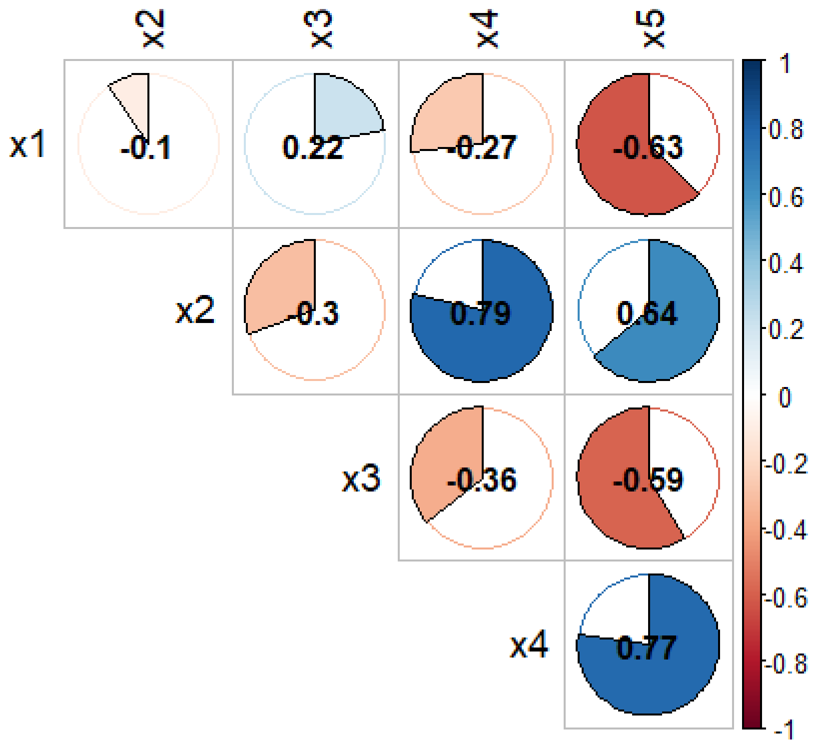


Figura 1: Gráfico de correlação amostral

Teste de hipótese de correlações (5% significância):

- Correlação amostral:

Foram realizados testes de hipótese de correlação linear para todos os pares de variáveis, a 5% de significância, com as seguintes hipóteses:

Ho: as variáveis são independentes (correlação não significativa);

Versus a hipótese alternativa

H1: as variáveis são dependentes (correlação significativa).

Nas tabelas a seguir, são apresentados os p-valor obtidos para as correlações amostrais ao nível de 5% de significância.

Tabela 3: Correlação linear de Pearson a 5% de significância entre X1-X2

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
X1 – x2	9	0.7788	-0.6580	0.5346	-0.0960

De acordo com o p-valor a 5% de significancia rejeita-se H1 e aceita-se H0, ou seja, não ocorre correlação entre essas duas variaveis.

Tabela 4: Correlação linear de Pearson a 5% de significância entre X1-X3

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
X1 – x3	9	0.5200	-0.4394	0.7232	0.2178

De acordo com o p-valor a 5% de significancia rejeita-se H1 e aceita-se H0, ou seja, não ocorre correlação entre essas duas variaveis.

Tabela 5: Correlação linear de Pearson a 5% de significância entre X1-X4

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
X1 – x4	9	0.4288	-0.7468	0.3970	-0.2667

De acordo com o p-valor a 5% de significancia rejeita-se H1 e aceita-se H0, ou seja, não ocorre correlação entre essas duas variaveis.

Tabela 6: Correlação linear de Pearson a 5% de significância entre X1-X5

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
X1 – x5	9	0.0385	-0.8918	0.0449	-0.6279

De acordo com o p-valor, aceita-se a Hipotese alternativa, ou seja, que as variáveis são dependentes, com correlação significativa.

Tabela 7: Correlação linear de Pearson a 5% de significância entre X2-X3

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
X2 – x3	9	0.3674	-0.7633	0.3641	-0.3015

De acordo com o p-valor a 5% de significancia rejeita-se H1 e aceita-se H0, ou seja, não ocorre correlação entre essas duas variaveis.

Tabela 8: Correlação linear de Pearson a 5% de significância entre x2 – x4

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
x2 – x4	9	0.0039	-0.3578	0.9425	0.7884

De acordo com o p-valor, aceita-se a Hipotese alternativa, ou seja, que as variáveis são dependentes, com correlação significativa.

Tabela 9: Correlação linear de Pearson a 5% de significância entre x2 – x5

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
X2 – x5	9	0.0341	0.0642	0.8957	0.6394

De acordo com o p-valor, aceita-se a Hipotese alternativa, ou seja, que as variáveis são dependentes, com correlação significativa.

Tabela 10: Correlação linear de Pearson a 5% de significância entre x3 – x4

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
X3 – x4	9	0.2752	-0.7898	0.53047	-0.3611

De acordo com o p-valor a 5% de significancia rejeita-se H1 e aceita-se H0, ou seja, não ocorre correlação entre essas duas variaveis.

Tabela 11: Correlação linear de Pearson a 5% de significância entre x3 – x5

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
X3 – x5	9	0.0571	-0.8781	0.0184	-0.5779

De acordo com o p-valor a 5% de significancia rejeita-se H1 e aceita-se H0, ou seja, não ocorre correlação entre essas duas variaveis.

Tabela 12: Correlação linear de Pearson a 5% de significância entre x4 – x5

Correlação Linear de Pearson (5%)	**Grau de liberdade**	**P-valor**	**Intervalo de confiabilidade (95%)**		**Coeficiente de Correlação**
X4 – x5	9	0.0053	0.3218	0.9378	0.7725

De acordo com o p-valor, aceita-se a Hipotese alternativa, ou seja, que as variáveis são dependentes, com correlação significativa.

b) **Determine para cada variável a sua respectiva variável padronizada. Verifique se a matriz de variâncias e covariâncias amostral das variáveis padronizadas é igual a matriz de correlação amostral das variáveis originais.**

R: As variáveis padronizadas constam na Tabela 9, a seguir, sendo que para cálculo usa-se cada valor amostral decrescido da sua respectiva média e dividida pelo desvio padrão:

Tabela 9: Variáveis padronizadas:

x1	x2	x3	x4	x5
-1.1813	1.4425	0.6225	2.0613	1.2785
-1.1324	-1.9968	0.2340	-1.2969	-0.8909
0.4133	1.5207	-0.9602	0.5687	0.1796
-0.0986	-0.5898	0.2660	0.2780	0.0382
-1.1593	0.4264	0.3623	-0.4004	0.4815
1.5646	0.0355	0.8328	0.2441	-0.6457
-0.4514	0.0355	-0.6429	-0.1533	1.1040
-0.6645	0.4264	-1.4877	0.7965	1.2785
0.4892	0.0355	-0.9316	-0.0661	-0.3297
1.1114	-0.7461	2.0340	-1.6168	-1.8861
1.1090	-0.5898	-0.3292	-0.4150	-0.6080

Tabela 10: Matriz de variâncias e covariâncias amostral das variáveis padronizadas:

	x1	x2	x3	x4	x5
x1	1	-0.0960	0.2178	-0.2662	-0.6279
x2	-0.0960	1	-0.3016	0.7885	0.6395
x3	0.2178	-0.3016	1	-0.3611	-0.5879
x4	-0.2662	0.7885	-0.3611	1	0.7726
x5	-0.6279	0.6395	-0.5879	0.7726	1

Tabela 11: Matriz de correlação amostral das variáveis originais:

	x1	x2	x3	x4	x5
x1	1	-0.0960	0.2178	-0.2662	-0.6279
x2	-0.0960	1	-0.3016	0.7885	0.6395
x3	0.2178	-0.3016	1	-0.3611	-0.5879
x4	-0.2662	0.7885	-0.3611	1	0.7726
x5	-0.6279	0.6395	-0.5879	0.7726	1

Observando-se as Tabelas 10 e 11, percebe-se então que a Matriz de variância e covariância, agora das variáveis padronizadas, corresponde com a Matriz de correlação amostral das variáveis originais.

c) **Construa diagramas de dispersão comparando as variáveis duas a duas. Compare os resultados obtidos nesses diagramas com a matriz de variâncias e covariâncias e a matriz de correlação**

R: Os diagramas de dispersão para cada par de variável (x1-x2, x1-x3, x1-x4, x1-x5, x2 – x3, x2- x4, x2- x5, x3 – x4, x3- x5, x4- x5) estão representados nas figuras a seguir:

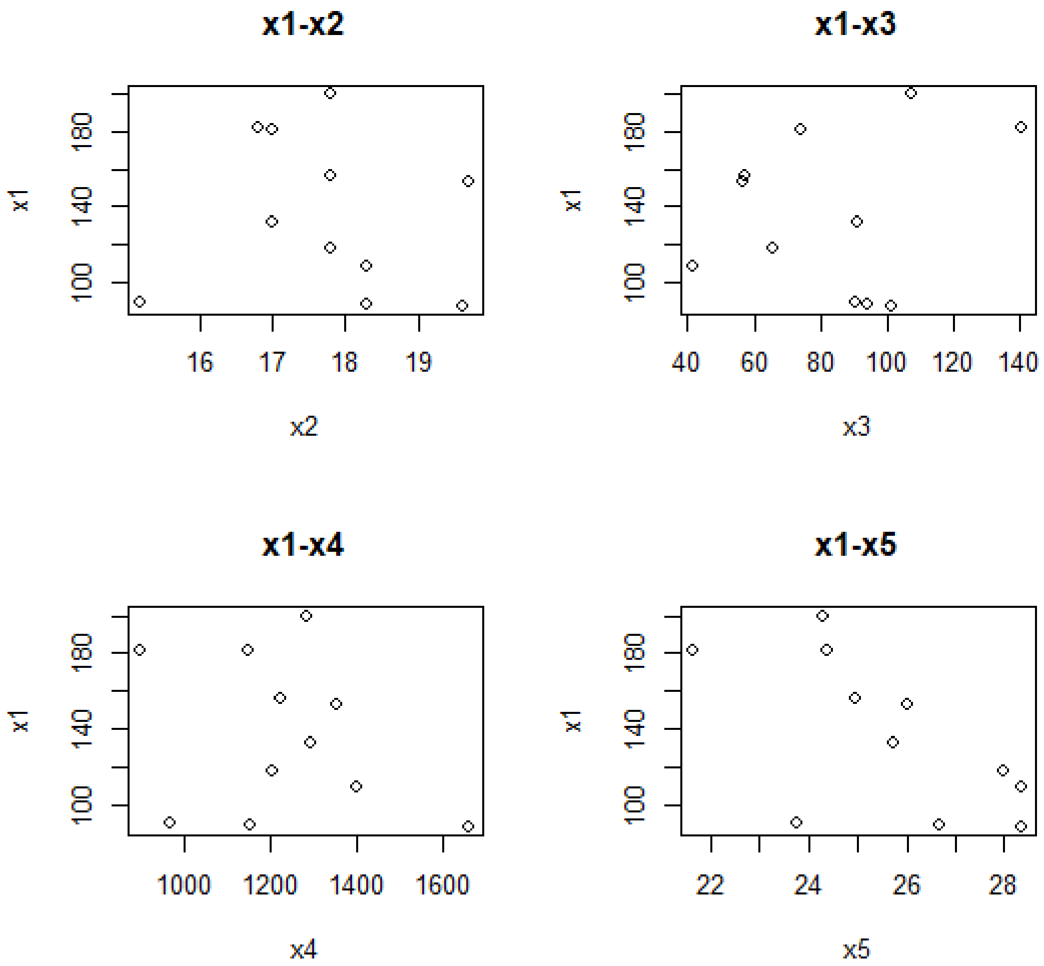


Figura 2: Diagrama de dispersão das variáveis x1 – x2, x1 – x3, x1 – x4, x1- x5.

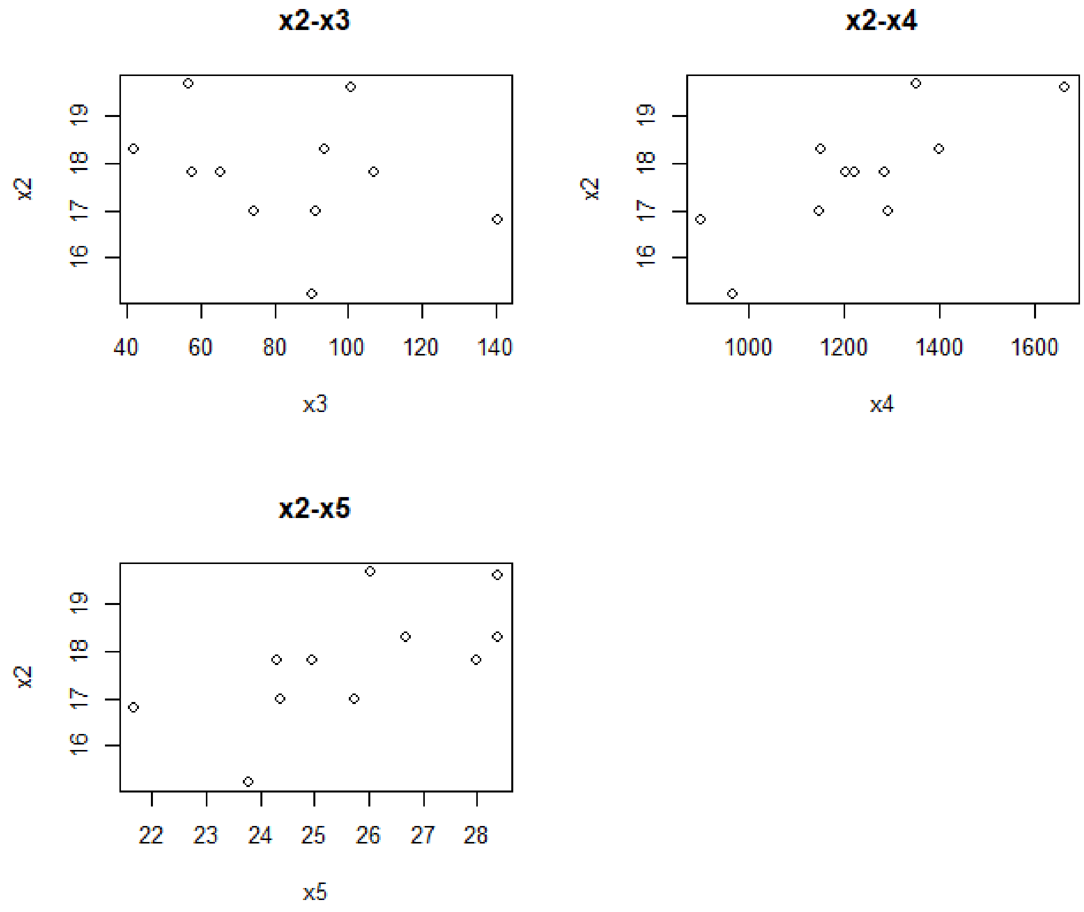


Figura 3: Diagrama de dispersão das variáveis x2- x3, x2- x4, x2- x5

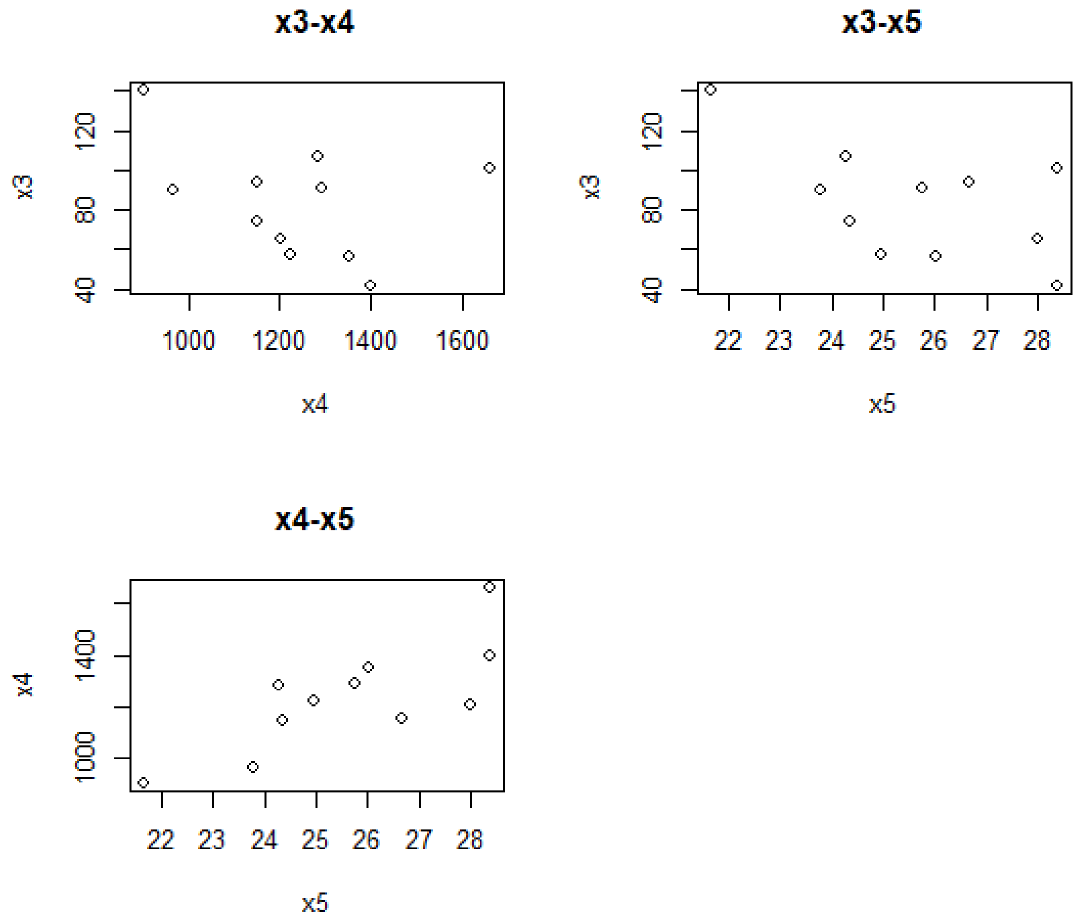


Figura 4: Diagrama de dispersão das variáveis x3- x4, x3- x5, x4- x5

Os gráficos x1- x2, x1- x3, x1- x4 apresentaram os menores valores de correlação linear e visualmente apresentaram-se mais dispersos dentre os diagramas de dispersão.

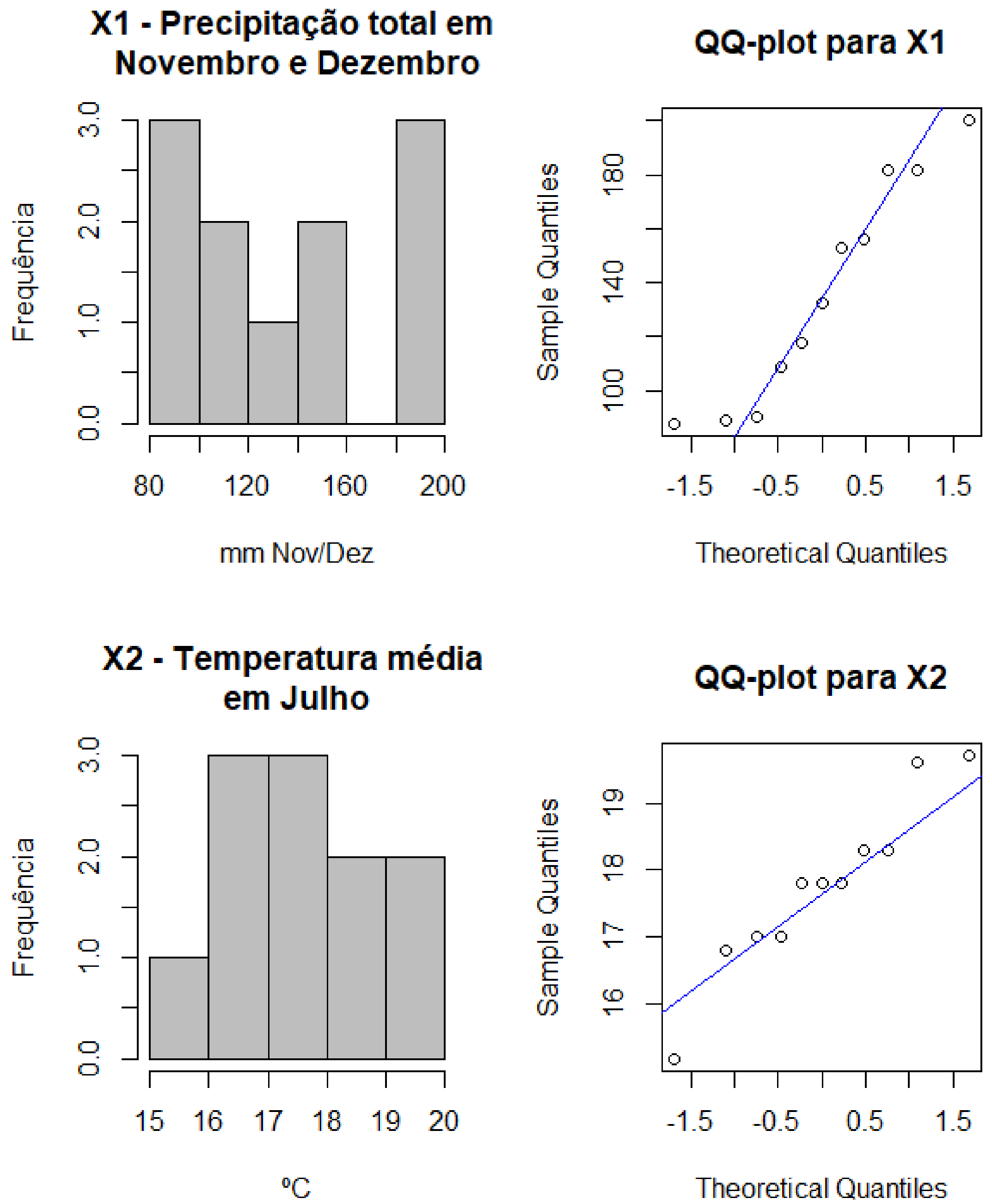
As covariâncias e correlações negativas (x1-x2, x1-x4, x1-x5, x2-x3, x3-x4, x3-x5), demonstraram padrões nos diagramas de dispersão que vão do canto superior esquerdo, evidenciando assim, essa relação inversa entre os valores das variáveis. Nesse sentido, as covariâncias e correlações positivasm apresentam diagramas de dispersão com pontos que se iniciam no canto inferior até o canto superior esquerdo.

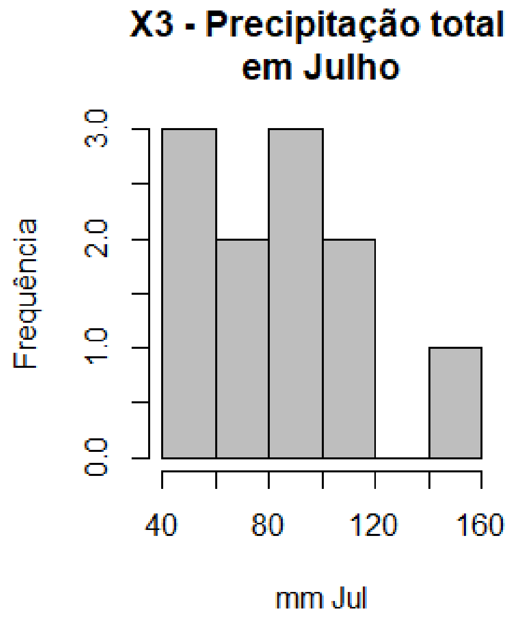
As correlações acima de 0,50 (consideradas como altas) possuem diagramas de dispersão visualmente melhor, caracterizando assim, sua distribuição espacial.

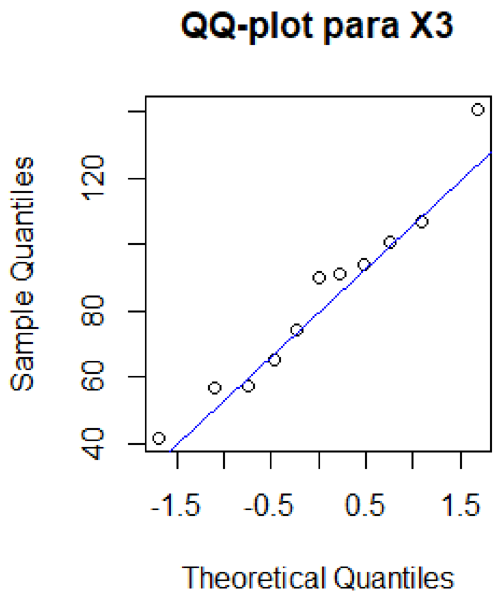
d) Avalie se cada variável tem distribuição normal de probabilidade (considere a 5% de significância)

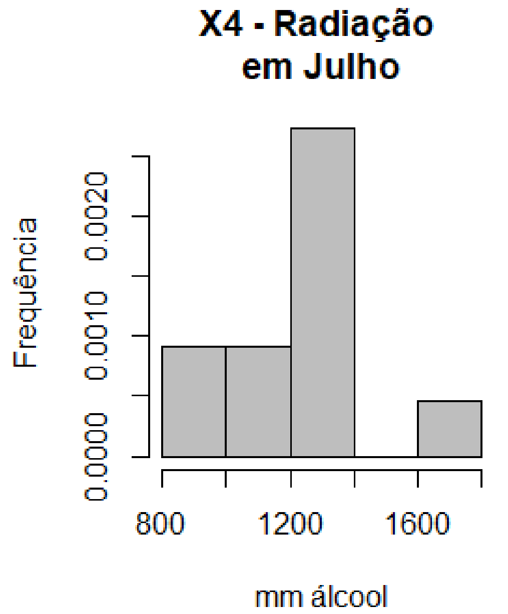
R: A fim de verificar a normalidade por métodos visuais pode-se observar os gráficos bloxplot e histogramas das variáveis na figura 5

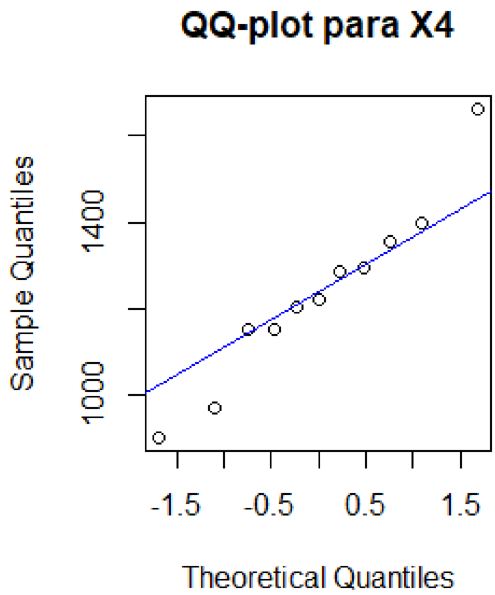
Figura 5: Histogramas e QQ-plot para cada variável

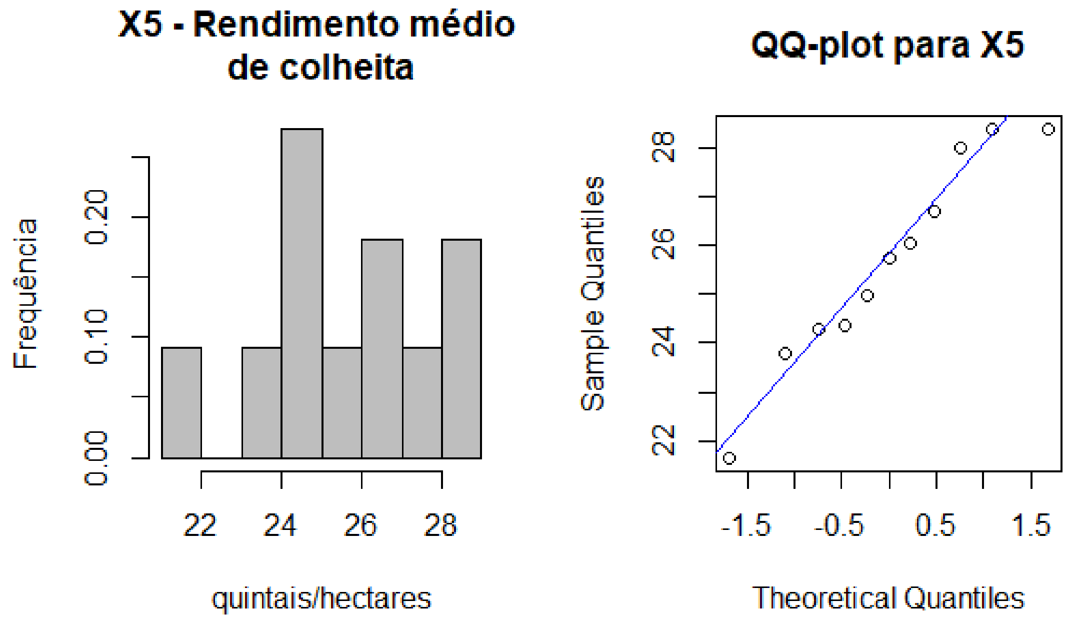



X3 - Precipitação total
em Julho
Frequência
0.0
1.0
2.0
3.0
40
80
120
160
mm Jul


QQ-plot para X3
Sample Quantiles
40
60
80
120
-1.5
-0.5
0.5
1.5
Theoretical Quantiles


X4 - Radiação
em Julho
Frequência
0.0000
0.0010
0.0020
800
1200
1600
mm álcool


QQ-plot para X4
Sample Quantiles
1000
1400
-1.5
-0.5
0.5
1.5
Theoretical Quantiles



Observa-se no histograma para a variável x1 (mm precipitação total em Novembro e Dezembro) (Figura 5) algumas faixas de valores da característica observada que ficaram isoladas da maioria dos dados, gerando barras ou pequenos agrupamentos separados. Pelo Boxplot (figura 6), é possível observar que a mediana é inferior a média, o que é característico de assimetria positiva dos dados.

Para o histograma x2 (Temperatura média em Julho °C) (figura 5), as maiores frequências de dados ficaram entre as temperaturas de 16 e 18 C, contendo o valor médio nesse intervalo. Pelo Boxplot (Figura 6), é possível visualizar que a mediana é superior a média, o que é característico de assimetria negativa de dados. Nota-se pontos bastante discrepantes com relação a linha mediana.

A variável de x3 (Precipitação total em Julho mm), apresenta histograma com maiores frequências para valores próximos da media. Algumas faixas de valores ficaram isoladas da maioria dos dados, gerando barras ou pequenos agrupamentos separados. Pelo Boxplot (figura 6), é possível observar que a mediana é superior a média, o que é característico de assimetria negativa dos dados.

Em relação a variável x4 (radiação solar em Julho mm de álcool) (Figura 5) as frequências foram observadas no intervalo de 1200 q 1400 kW/m^2, ocorrendo baixas frequências nos valores superiores a 1400 kW/m$^{2.}$ .Pelo Boxplot (figura 6) é possível observar que a mediana é superior a média, o que é característico de assimetria negativa. A amplitude superior e inferior se mostrou relativamente iguais.

Com relação a variável x5, ou seja, do rendimento médio de colheita (quintais/hectare), apresenta o histograma (figura 5) com maiores frequências para valores próximos a média. Algumas faixas de valores ficaram isoladas da grande maioria dos dados, gerando barras ou pequenos agrupamentos. No

Boxplot (Figura 6), ocorre a presença de leve simetria negativa, pois a mediana é pouco superior que a média.

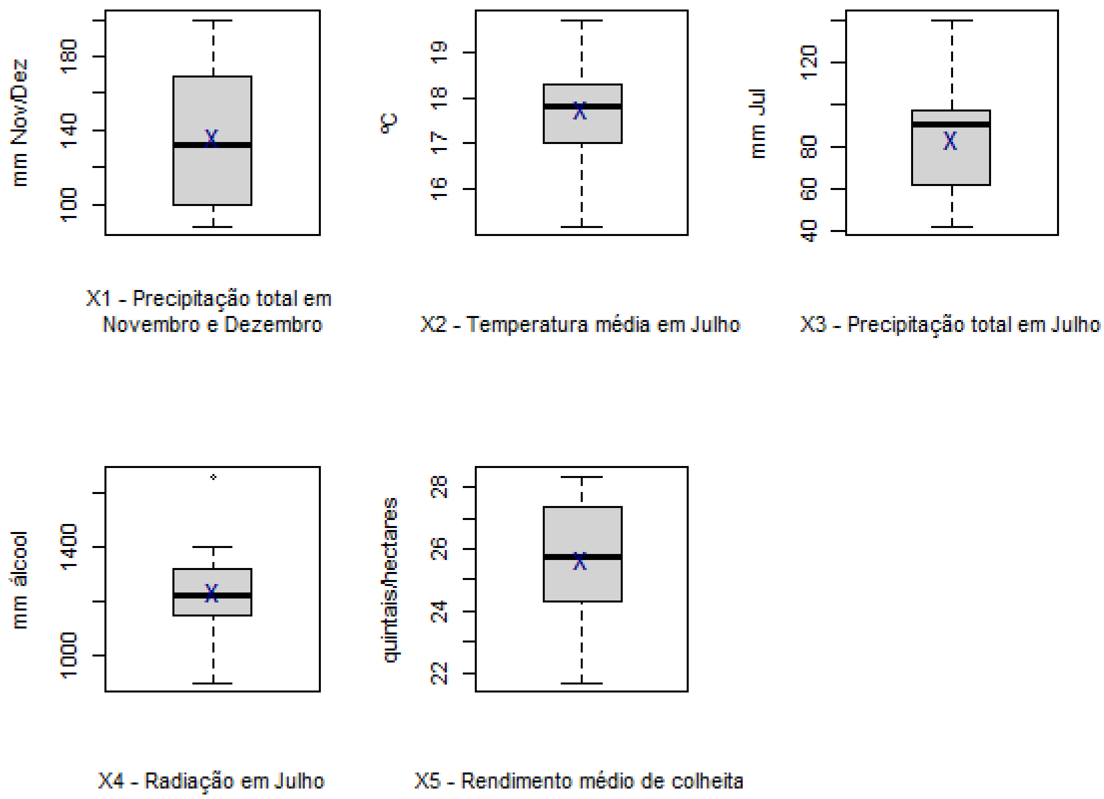


Figura 6: Boxplot das variáveis.

Teste de Shapiro – Wilks

Para verificar a suposição de normalidade dos dados, realizaram-se os testes de normalidade de Shapiro – Wilks, a 5% de significância para cada uma das variáveis de maneira independente.

Hipóteses testadas:

Ho: Os dados de cada variável possuem distribuição normal

Versus

H1: Os dados de cada variável não possuem distribuição normal

Como regra, se p-valor for maior que o nível de significância estipulado, 5% para o caso em questão, não se rejeita a hipótese nula e pode-se considerar que os dados da variável possuem distribuição normal.

Os valores de p-valores obtidos para o teste de normalidade univariado por Shapiro-Wilk, a 5% de significância, encontram-se na tabela a seguir.

Tabela 13: Teste de Shapiro-Wilks (normalidade) para cada variável:

Variável	p- valor	w
X1	0.2696	0.9137
X2	0.5549	0.9428
X3	0.8101	0.9631
X4	0.7920	0.9616
X5	0.6152	0.9477

Observa-se que todos os p-valores foram maiores que 0,05, por isso, não se rejeita Ho, podendo assim, considerer que os dados de cada uma das variáveis possuem distribuição normal.

e) Avalie se o conjunto de variáveis tem distribuição normal de probabilidade multivariada (considere 5% de significância)

R: Para verificar a suposição de normalidade dos dados de maneira multivariada, realizou-se o teste de normalidade multivariado de Shapiro – Wilks e Shapiro – Francia ambos a 5% de significância.

Hipóteses testadas:

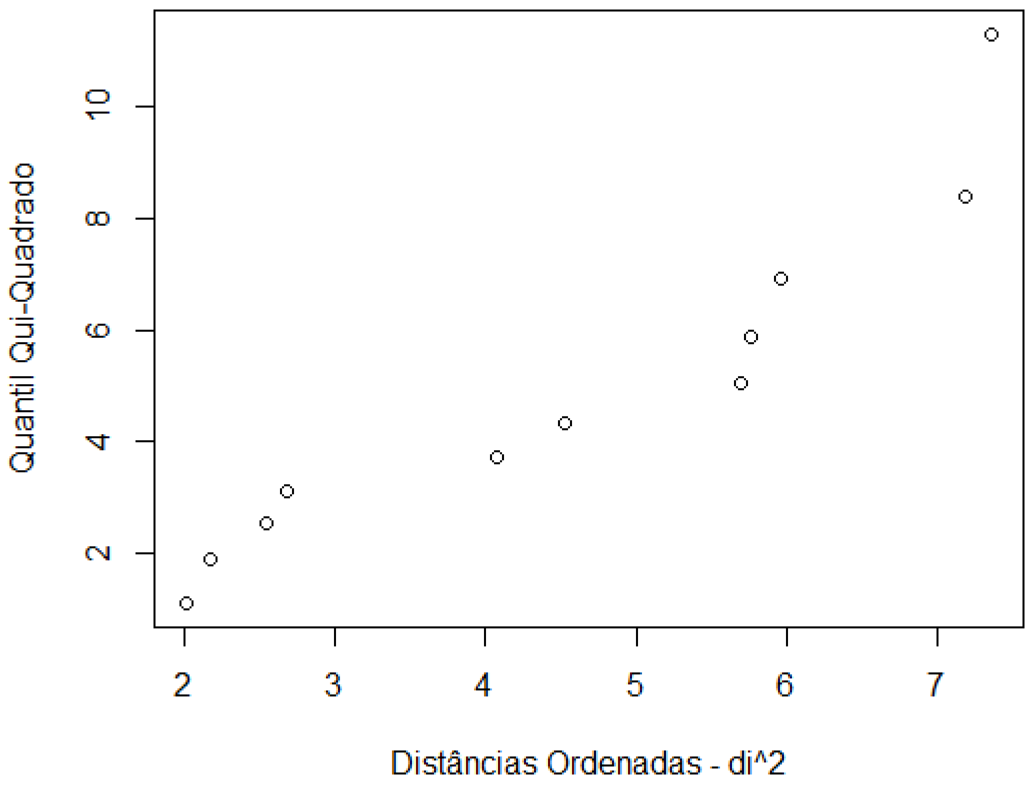
Ho: O conjunto de variáveis possui distribuição normal multivariada
Versus
H1: O conjunto de variáveis não possui distribuição normal multivariada

Tabela 14: Distribuição normal de probabilidade multivariada:

	Shapiro - Wilks	**Shapiro - Francia**
p- Valor	0.003025	0.00373

Os valores dos dois testes, tanto de Shapiro – Wilks, quanto de Shapiro – Francia, apresentaram valores menos que 0,05% rejeita-se Ho e pode-se dizer que o conjunto de variáveis não possui distribuição normal multivariada

Figura 7: Gráfico QQ-plot multivariado:



Na Figura 7 tem-se o grafico QQ-plot multivariada, onde observa-se que a nuvem de pontos não segue uma linha reta, com pontos bastante dispersos.

Questão 4) Para cada uma das funções de densidade de probabilidade conjunta a seguir apresente o vetor de médias e a matriz de variâncias e covariância e o valor da correlação entre as duas variáveis do vetor aleatório bi-dimensional. Além disso, responda se em cada caso as duas variáveis X e Y podem ser consideradas independentes.

Para as duas alternativas dessa questão, os respectivos cálculos se encontram no Anexo E.

a)

$$f(x,y) = \frac{1}{2\pi} * \exp\left\{\left[-\frac{1}{2}\{(x-1)^2 + (y-2)^2\}\right]\right\}$$

R:

Vetor de médias amostrais encontrado:

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \end{bmatrix}$$

A matriz de variância e covariância:

$$\boldsymbol{S} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

A correlação entre as duas variáveis:

$$\boldsymbol{\rho_{12}} = \mathbf{0}$$

A média para a variável X foi de 1u.m. e da variável Y 2u.m., e observando a matriz S nota-se que as variâncias (diagonal principal) foram 1u.m. para ambas. Enquanto que a covariância foi 0u.m. para o par X-Y, o que representa uma total falta de correlação entre as duas, o que se confirma pelo coeficiente calculado. Sendo assim, pode-se afirmar que X e Y, nesse caso, são variáveis independentes entre si.

b)

$$f(x,y) = \frac{1}{2{,}4} * \exp\left\{\left[-\frac{1}{0{,}72}\left\{\frac{x^2}{4} - 8{,}8xy + y^2\right\}\right]\right\}$$

R:

Conforme cálculo em anexo, não foi possível chegar nos valores de médias e variâncias, apenas na correlação das variáveis X e Y de $\boldsymbol{\rho_{12}} = \mathbf{0{,}8}$, indicando que provavelmente as duas sejam dependentes.

Questão 5) Para os dados abaixo provenientes de um delineamento inteiramente ao acaso, e assumindo que o conjunto de variáveis respostas tem distribuição normal de probabilidade multivariada e homocedasticidade da matriz de variâncias e covariâncias, construa a MANOVA e interprete os resultados obtidos usando alfa igual a 0,05. Sendo que:

Fator: Sexo, com níveis: macho e fêmea

Variável 1 (X1): Comprimento

Variável 2 (X2): Largura

Variável 3 (X3): Peso

Fêmea			Macho		
Comprimento	Largura	Peso	Comprimento	Largura	Peso
98	81	38	93	74	37
103	84	38	94	78	35
103	86	42	95	80	35
105	86	42	100	84	39
109	88	44	102	85	38
123	92	50	106	80	37
123	90	46	104	83	39
133	99	51	106	83	36
133	102	51	107	82	38
133	102	51	112	89	40
134	100	48	113	88	40
136	102	49	110	86	40
138	98	54	116	90	40
138	95	51	117	90	41
141	105	53	117	91	40
147	108	57	119	93	41
149	107	55	120	89	40
153	107	56	120	93	44
155	115	63	121	95	42
155	117	60	125	93	45
158	115	62	127	96	45
159	118	63	128	95	45
162	124	61	131	95	46
177	132	67	130	106	47

R: Considerando os dados amostrais serem provenientes de um delineamento ao acaso, bem como que o conjunto de variáveis respostas possuírem distribuição normal de probabilidade multivariada e homocedasticidade da matriz de variâncias e covariâncias, avaliou-se a influência do fator sexo (fêmea e macho) sobre três variáveis respostas, sendo: comprimento, largura e peso. A análise foi feita a partir da MANOVA com um nível de significância de 5%.

As hipóteses testadas foram:

Ho: as variáveis (comprimento, largura e peso) não sofrem influência do fator sexo em nenhum dos seus níveis;

Versus

H1: as variáveis (comprimento, largura e peso) sofrem influência do fator sexo em nenhum dos seus níveis.

Assim, observamos que na Tabela 15, encontra-se a análise de variância multivariada (MANOVA), considerando todas as três variáveis simultaneamente, pelo critério Lambda Wilks.

Tabela 15: Analise de variância multivariada (MANOVA) pelo critério de Lambda Wilks

	Df	**Wilks**	**approx F**	**num Df**	**den Df**	**Pr(>F)**
SEXO	1	0.3614	25.9220	3.0000	44.0000	8.189e-10 ***
Resíduo	46					

Observando os resultados da ánalise de variância multivariada (MANOVA), com 5 % de significância, nota-se que p-valor, foi inferior a 0,05%, rejeitando-se assim Ho, podendo afirmar que pelo menos uma das variáveis está sendo influenciado pelo fator sexo.

ANEXO A – Script software R (Exercício 1)

```
##########################################################
### 1a LISTA DE EXERCÍCIOS - ANÁLISE MULTIVARIADA 2018 ###
##########################################################



        ##### Questao 1 #####

A <- matrix(c(1,1,0,2,1,1,3,2,2), ncol = 3, nrow = 3)
B <- matrix(c(1,0,1,2,2,2,3,3,4), ncol = 3, nrow = 3)
A
B

### Alternativa a) ###
A-3*B

### Alternativa b) ###
A%*%B
det(A%*%B)

### Alternativa c) ###
kronecker(A,B)

### Alternativa d) ###
C <- matrix (c(1,1,0,0,0,0,2,1,1,0,0,0,3,2,2,0,0,0,0,0,0,1,0,1,0,0,0,2,2,2,0,0,0,3,3,4), ncol = 6,
nrow= 6)
C

### Alternativa e) ###
D <- solve(B)
D
B%*%D

### Alternativa f) ###
KhatriRao(A,B)
```

ANEXO B – Script software R (Exercício 2)

```
############################################################
### 1a LISTA DE EXERCÍCIOS - ANÁLISE MULTIVARIADA 2018 ###
############################################################


          ##### Questao 2 #####

A <-  matrix(c(9,-5,-5,6), ncol = 2, nrow = 2)
A

## Alternativa (a) ##
a <- t(A)
a

## xt * A * x



## Alternativa (b) ##

eigen(A)


## Alternativa (c) ##

Avet <- matrix(c(-0.8023, 0.5969, -0.5969, -0.8023), ncol=2, nrow=2)
Avet

Aval <- matrix(c(12.7202, 0, 0, 2.2798), ncol=2, nrow=2)
Aval

Adecomp <- Avet %*% Aval %*% t(Avet)
Adecomp

## Alternativa (d) ##

solve(A)
Ainv <- solve(A)
Ainv
eigen(Ainv)
```

ANEXO C – Script software R (Exercício 3)

```
#########################################################
### 1a LISTA DE EXERCÍCIOS - ANÁLISE MULTIVARIADA 2018 ###
#########################################################



        ##### Questao 3 #####


dados <- read.table("questao3_dados.txt", header = T)
dados

names(dados)
is.data.frame(dados)
c <- as.matrix(dados)
c

attach(dados)
dim(dados)



        ### Alternativa (a) ###


summary(dados)

u <-colMeans(dados) ## vetor de médias ##
u
S <- cov(dados,dados) ## matriz de variancia e covariancia ##
S
R <- cor(dados,dados) ## matriz de correlação ##
R

## gráfico das correlações amostrais ##

require(corrplot)
corrplot(R, method="pie", type = c("upper"), diag = F, tl.cex = 1.2,
tl.col = "black", addCoef.col = "black")

## teste de hipóteses das correlações cor.test ##

library(readr)
library(psych)
corr.test(dados)

cor.test(x1,x2)
cor.test(x1,x3)
cor.test(x1,x4)
cor.test(x1,x5)
cor.test(x2,x3)
cor.test(x2,x4)
cor.test(x2,x5)
cor.test(x3,x4)
cor.test(x3,x5)
cor.test(x4,x5)
```

```
### Alternativa (b) ###

## variável padronizada ##
dadosPad <- scale(dados)
dadosPad
cov(dadosPad)
cor(dados)
## são iguais ##


### Alternativa (c) ###

## Diagrama de dispersão - variáveis duas a duas ##

pairs(dados, c("X1", "X2", "X3", "X4", "X5"))

par(mfrow = c(2,2))
diagrama1 <- plot(x1~x2, main = "x1-x2")
diagrama2 <-plot(x1~x3, main = "x1-x3")
diagrama3 <-plot(x1~x4, main = "x1-x4")
diagrama4 <-plot(x1~x5, main = "x1-x5")

par(mfrow = c(2,2))
diagrama5 <-plot(x2~x3, main = "x2-x3")
diagrama6 <-plot(x2~x4, main = "x2-x4")
diagrama7 <-plot(x2~x5, main = "x2-x5")

par(mfrow = c(2,2))
diagrama8 <-plot(x3~x4, main = "x3-x4")
diagrama9 <-plot(x3~x5, main = "x3-x5")
diagrama10 <-plot(x4~x5, main = "x4-x5")


## gráficos de dispersão simultâneos com diagonal = histograma das variáveis ##
##  função que cria um histograma ##

panel.hist <- function(x, ...)
{
   usr <- par("usr"); on.exit(par(usr))
   par(usr = c(usr[1:2], 0, 1.5) )
   h <- hist(x, plot = FALSE)
   breaks <- h$breaks; nB <- length(breaks)
   y <- h$counts; y <- y/max(y)
   rect(breaks[-nB], 0, breaks[-1], y, col="cyan", ...)
}

pairs(dados, c("X1", "X2", "X3", "X4", "X5"), diag.panel=panel.hist)



### Alternativa (d) ###

## X1 ##
par(mfrow = c(1,2))
hist(x1, col="gray", main= "X1 - Precipitação total em
Novembro e Dezembro",
xlab = "mm Nov/Dez", ylab = "Frequência", prob=F)      ## Histograma
qqnorm(x1, main = "QQ-plot para X1")
qqline(x1, col = "blue")          ## qq-plot
```

```
## X2 ##
par(mfrow = c(1,2))
hist(x2, main = "X2 - Temperatura média
em Julho", col="gray",xlab = "ºC",
ylab = "Frequência", prob=F)                ## Histograma
qqnorm(x2, main = "QQ-plot para X2")
qqline(x2, col = "blue")          ## qq-plot

## X3 ##
par(mfrow = c(1,2))
hist(x3, main = "X3 - Precipitação total
em Julho", col="gray",
xlab = "mm Jul", ylab = "Frequência", prob=F)    ## Histograma
qqnorm(x3, main = "QQ-plot para X3")
qqline(x3, col = "blue")          ## qq-plot

## X4 ##
par(mfrow = c(1,2))
hist(x4, main = "X4 - Radiação
em Julho", col="gray",xlab = "mm álcool",
ylab = "Frequência", prob=T)                ## Histograma
qqnorm(x4, main = "QQ-plot para X4")
qqline(x4, col = "blue")          ## qq-plot

## X5 ##
par(mfrow = c(1,2))
hist(x5, main = "X5 - Rendimento médio
de colheita", col="gray",
xlab = "quintais/hectares", ylab = "Frequência", prob=T) ## Histograma
qqnorm(x5, main = "QQ-plot para X5")
qqline(x5, col = "blue")          ## qq-plot

## Boxplot - Variáveis ##
par(mfrow = c(2,3))
boxplot(x1, xlab = "X1 - Precipitação total em
Novembro e Dezembro",
ylab = "mm Nov/Dez", col=c("lightgray"))
points(mean(x1), pch='x', cex=1.5, col='darkblue')

boxplot(x2, xlab = "X2 - Temperatura média em Julho",
ylab = "ºC", col=c("lightgray"))
points(mean(x2), pch='x', cex=1.5, col='darkblue')

boxplot(x3, xlab = "X3 - Precipitação total em Julho",
ylab = "mm Jul", col=c("lightgray"))
points(mean(x3), pch='x', cex=1.5, col='darkblue')

boxplot(x4, xlab = "X4 - Radiação em Julho",
ylab = "mm álcool", col=c("lightgray"))
points(mean(x4), pch='x', cex=1.5, col='darkblue')

boxplot(x5, xlab = "X5 - Rendimento médio de colheita",
ylab = "quintais/hectares", col=c("lightgray"))
points(mean(x5), pch='x', cex=1.5, col='darkblue')

## TESTE DE SHAPIRO-WILKS UNIVARIADO ##
shapiro.test(x1)
shapiro.test(x2)
shapiro.test(x3)
shapiro.test(x4)
```

```
shapiro.test(x5)



        ### Alternativa (e) ###

## GRÁFICO QQ-PLOT MULTIVARIADO ##

S=var(c)
S
m=apply(c,2,mean)   ## indica que é p/ fazer os cálculos para cada coluna ##
m
invS=solve(S)
invS
d1_2 <- t(c[5,]-m) %*% invS %*% (c[5,]-m)
d1_2
(dim(c)[1]-1+0.5)/dim(c)[1]
1-((1-0.5)/dim(c)[1])
d = q = NULL
n= nrow(c)    ## nº de indivíduos ##
p= ncol(c)    ## nº de variáveis ##
for (i in 1:n){
  d = c(d, t(c[i,]-m) %*% invS %*% (c[i,]-m))
  prob = (i-0.5)/n
  q = c(q, qchisq(prob, df=p))
}
d=sort(d)
d
q

plot(d,q, xlab = "Distâncias Ordenadas - di^2", ylab = "Quantil Qui-Quadrado")


## TESTE DE SHAPIRO WILKS MULTIVARIDO ##

require(mvnormtest)
mshapiro.test(t(c))

## TESTE DE NORMALIDADE MULTIVARIADO DE SHAPIRO-FRANCIA ##
require(mvsf)
mvsf(t(c))
```

ANEXO D – Script software R (Exercício 5)

```
##########################################################
### 1a LISTA DE EXERCÍCIOS - ANÁLISE MULTIVARIADA 2018 ###
##########################################################


        ##### Questao 5 #####

x <- read.table("questao5_dados.txt", header = T)
x

attach(x)

## MANOVA ##
# test = c("Pillai" (defualt), "Wilks", "Hotelling-Lawley", "Roy") #

var <- as.matrix(x[,2:4])
var
fit <- manova(var ~ SEXO)
fit

# Resumo Manova #
summary.manova(fit)
summary.manova(fit, test = "Wilks")
```

ANEXO E – Roteiro de Cálculo (Exercício 4)

A função densidade conjunta de distribuição normal bivariada é:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2{}_{12})}} * \exp\left[-\frac{1}{2(1-\rho^2{}_{12})}\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right\}\right]$$

Alternativa A)

Comparando com a função dada pelo enunciado,

$$f(x_1, x_2) = \frac{1}{2\pi} * \exp\left\{\left[-\frac{1}{2}\{(x_1-1)^2 + (x_2-2)^2\}\right]\right\}$$

Fez-se algumas ponderações:

i) Se:

$-\frac{1}{2(1-\rho^2{}_{12})} = -\frac{1}{2} \quad então \ (1-\rho^2{}_{12}) = 1$, de onde deduz-se: $\boldsymbol{\rho_{12} = 0}$ (correlação)

ii) Se:

$\frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2{}_{12})}} = \frac{1}{2\pi} \quad então \ \sqrt{\sigma_{11}\sigma_{22}(1-\rho^2{}_{12})} = 1$, o que em conjunto com **(i)** leva à: $\begin{cases}\sigma_{11} = \pm 1\\ \sigma_{22} = \pm 1\end{cases}$

iii) Mas como

$\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 = (x_1-1)^2$, então $\begin{cases}\boldsymbol{\sigma_{11} = 1}\\ \boldsymbol{\sigma_{22} = 1}\end{cases}$ Que são os respectivos valores de variância de x1 e x2

iv) Substituindo as variáveis encontradas na equação inicial:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1*1(1-0^2)}} * \exp\left[-\frac{1}{2(1-0^2)}\left\{\left(\frac{x_1-\mu_1}{\sqrt{1}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{1}}\right)^2 - 2*0\left(\frac{x_1-\mu_1}{\sqrt{1}}\right)\left(\frac{x_2-\mu_2}{\sqrt{1}}\right)\right\}\right] =$$

$$f(x_1, x_2) = \frac{1}{2\pi} * \exp\left\{\left[-\frac{1}{2}\{(x_1-\mu_1)^2 + (x_2-\mu_2)^2\}\right]\right\}$$

Então tem-se que: $\begin{cases}\boldsymbol{\mu_1 = 1}\\ \boldsymbol{\mu_2 = 2}\end{cases}$ Cujos representam os valores das médias de x1 e x2

v) Por fim, sabendo que:

$\rho_{12} = \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$ então, $\boldsymbol{\sigma_{12}} = \boldsymbol{\sigma_{21}} = \rho_{12}\sigma_{11}\sigma_{22} = 0*1*1 = \boldsymbol{0}$ (covariância entre x1 e x2)

vi) Dessa forma o resultado do vetor médias é:

$$\mu = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} = \begin{bmatrix}1\\ 2\end{bmatrix}$$

vii) E a matriz de variância e covariância é:

$$S = \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

Alternativa B)

A função densidade conjunta de distribuição normal bivariada é:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2{}_{12})}} * \exp\left[-\frac{1}{2(1-\rho^2{}_{12})}\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right\}\right]$$

Comparando com a função dada pelo enunciado,

$$f(x_1, x_2) = \frac{1}{2{,}4} * \exp\left\{\left[-\frac{1}{0{,}72}\left\{\frac{x_1{}^2}{4} - 8{,}8x_1x_2 + x_2{}^2\right\}\right]\right\}$$

Fez-se algumas ponderações:

i) Se:

$\left(1-\rho^2{}_{12}\right) = z$ e $\sqrt{\sigma_{11}\sigma_{22}(z)} = w$

Então

$\frac{1}{2\pi * w} = \frac{1}{2{,}4}$, e assim $w = 0{,}382$ $z = \frac{0{,}146}{\sigma_{11}\sigma_{22}}$

ii) Como:

$$\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right\}$$
$$= \left\{\left(\frac{x_1^2 - 2x_1\mu_1 + \mu_1{}^2}{\sigma_{11}}\right) + \left(\frac{x_2^2 - 2x_2\mu_2 + \mu_2{}^2}{\sigma_{22}}\right) - 2\rho_{12}\left(\frac{x_1x_2 - x_1\mu_2 - x_2\mu_1 + \mu_1\mu_2}{\sqrt{\sigma_{11}\sigma_{22}}}\right)\right\}$$

iii) Então, comparando as duas equações:

$$-\frac{1}{0{,}72} * \frac{x_1{}^2}{4} = -\frac{1}{2\left(1-\rho^2{}_{12}\right)} * \left(\frac{x_1^2 - 2x_1\mu_1 + \mu_1{}^2}{\sigma_{11}}\right) \Rightarrow \frac{x_1{}^2}{2{,}88} = \frac{x_1^2 - 2x_1\mu_1 + \mu_1{}^2}{\sigma_{11} * z}$$

$$-\frac{1}{0{,}72} * x_2{}^2 = -\frac{1}{2\left(1-\rho^2{}_{12}\right)} * \left(\frac{x_2^2 - 2x_2\mu_2 + \mu_2{}^2}{\sigma_{22}}\right) \Rightarrow \frac{x_2{}^2}{0{,}72} = \frac{x_1^2 - 2x_1\mu_1 + \mu_1{}^2}{\sigma_{11} * z}$$

$$-\frac{1}{0{,}72} * -8{,}8x_1x_2 = -\frac{1}{2\left(1-\rho^2{}_{12}\right)} * -2\rho_{12}\left(\frac{x_1x_2 - x_1\mu_2 - x_2\mu_1 + \mu_1\mu_2}{\sqrt{\sigma_{11}\sigma_{22}}}\right)$$

iv) Se:

$-\frac{1}{0{,}72} = \frac{1}{2\left(1-\rho^2{}_{12}\right)}$, então $\boldsymbol{\rho_{12} = 0{,}8}$

Referências Bibliográficas

GOMES. P. R. B. **Métodos tensoriais para estimação cega de assinaturas espaciais**. 2014. 84 f. Dissertação (Mestrado em Engenharia de Teleinformática) - Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2014.

Khatri, C. G.; Rao, C. R. **Solutions to Some Functional Equations and Their Applications to Characterization of Probability Distributions**. Sankhya: Indian J. Statistics, Series A 30, 167–180,1968.