Lista 08 de FECD

Exercício 8

• A quantidade D^2 tem um valor tıpico (ou valor esperado): E(D^2) =??.

• D^2 possui um afastamento tıpico de seu valor esperado, seu DP. O desvio-padrao de D^2 é:raiz(V(D^2)) =??.

• Mais que isto, nao somente estes dois resumos da distribuiçao de D^2 sao conhecidos mas a propria distribuiçao de D^2 é conhecida. D^2 ∼??.

• Fixando uma constante c qualquer, o conjunto de pontos x ∈ R^p que satisfazem D^2 = c formam um elipsoide em p dimensoes. Isto e, os pontos x que estao a uma distancia D^2 igual a c do seu perfil esperado formam um elipsoide. Quais sao os eixos deste elipsoide e os seus tamanhos relativos?

• E possıvel mostrar que, com probabilidade 1 − α, o vetor aleatorio X deve cair dentro da elipse D^2 = c onde c = χ2p(α) e o quantil (1 − α)100% de uma distribuicao qui-quadrado com p graus de liberdade onde p e a dimensao do vetor X. No caso particular de um vetor bidimensional, o valor de c associado com a probabilidade 1 − α = 0.95 e igual a c = 9.21. Assim, se X = (X1, X2) estiver fora dessa elipse (isto e, se D^2 > 9.21), o ponto pode ser consirado um tanto anomalo ou extremo. O arquivo stiffness.txt contem quatro tipos de medicoes da rigidez de pranchas de madeira. A primeira e obtida aplicando-se uma onda de choque atraves da prancha, a segunda aplicando-se uma vibracao a prancha e as outas duas sao obtidas por meio de testes estaticos. Assuma que cada as 4 mediçoes em uma prancha sao instancias de um vetor N4(μ, Σ). Estime o vetor μ e a matriz4 × 4 Σ usando os dados do amostra. A seguir, usando estes valores estimados como se fossem os verdadeiros valores de μ e Σ, calcule o valor de D^2 para cada ponto da amostra. Quais pontos parecem extremos? Olhando as variaveis INDIVIDUALMENTE ou em pares atraves de scatterplots seria possıvel detectar estes pontos extremos? Faca scatterplot dos dados para entender sua resposta

Resposta de todas as perguntas acimas:

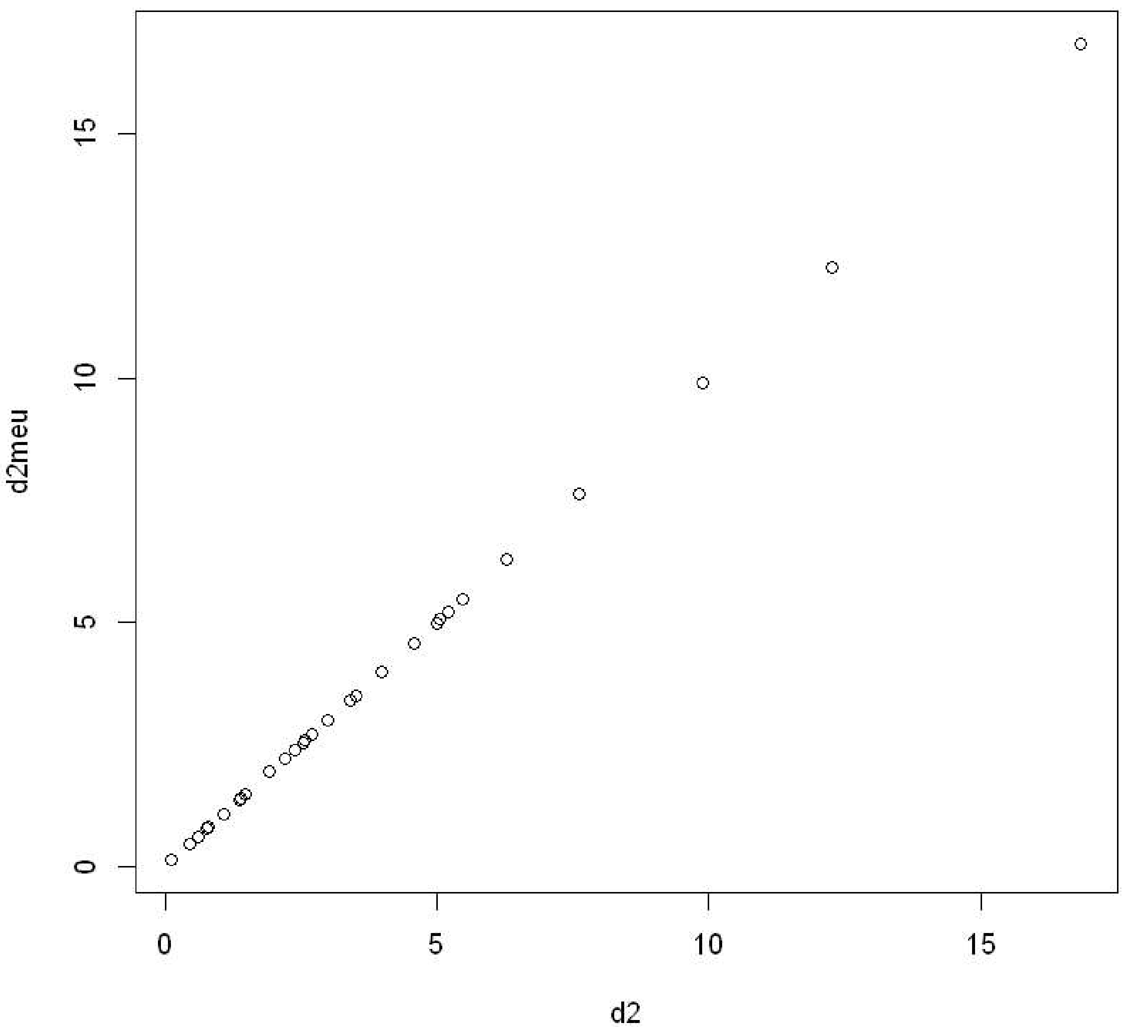
É possıvel deduzir que, se X ∼ Np(μ, Σ), entao D^2 segue uma distribuicao qui-quadrado com p graus de liberdade. Isto permite obter E(D^2) = p e tambem V(D^2) = 2p. Os eixos do elipsoide estao na direcao dos autovetores da matriz Σ e com tamanhos proporcionais a raiz quadrada de seus autovalores.

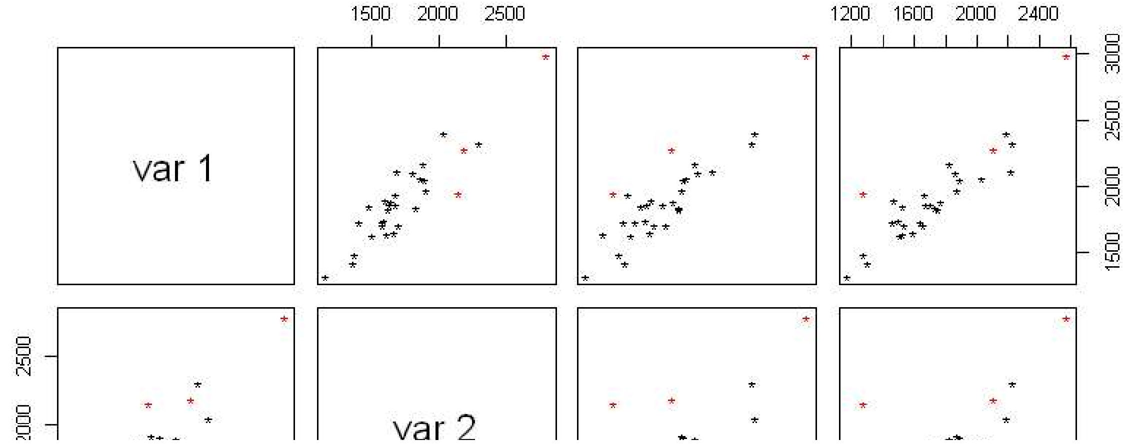
In [5]:
```
stiffness = matrix(scan("stiffness.txt"), ncol=5, byrow=T)
x = stiffness[,1:4]
mu = apply(x, 2, mean)
sigma = cov(x)

n = nrow(x)
desvio = x - matrix(mu, nrow=nrow(x), ncol=ncol(x), byrow=T)
d2meu = diag(desvio %*% (solve(sigma) %*% t(desvio)))
# comando acima calcula D2
# R possui um comando proprio (e mais eficiente) para isto: mahalanobis
d2 = mahalanobis(x, mu, sigma)

# verificando que meu comando ineficiente calculou a mesma coisa
plot(d2, d2meu)
# identificando as anomalias
anomalias = d2 > qchisq(0.95,4)
x[anomalias,]
nanom = sum(anomalias)
# plotando e marcando em vermelho as anomalias
pairs(rbind(x, x[anomalias,]), pch="*", col=rep(c("black", "red"), c(n, nanom)))
```

2983	2794	2412	2581
1954	2149	1180	1281
2276	2189	1547	2111


d2meu
0
5
10
15
0
5
10
15
d2



Exercicío 9

• Forneca uma estimativa para o vetor μ e para a matriz Σ e ρ.

• A partir da matriz de correlacao entre os pares de v.a.'s (e do plot de dispersao dos pontos), quais as variaveis que sao mais correlacionadas? E quais sao menos correlacionadas?

• Obtenha a distribuicao MARGINAL do sub-vetor X∗ = (X1, X2), o comprimento e largura da sepala.

• Obtenha a distribuicao CONDICIONAL do sub-vetor X∗ = (X1, X2) quando sao conhecidos os valores x3 e x4 das v.a.'s (X3, X4). Obtenha esta distribuicao para dois valores genericos x3 e x4. A seguir use dois valores especıficos: x3 = 1.8 e x4 = 0.6, dois valores relativamente altos para estas variaveis. Compare DP1 = raiz(V(X1)) com raiz(V(X1|X3 = 1.8, X4 = 0.6)), o desvio padrao da variavel X1 condicionada nos valores de X3 e X4.

• Obtenha agora a distribuicao CONDICIONAL do sub-vetor X∗ = (X1, X2) quando e conhecido apenas o valor de X3.

• Obtenha tambem distribuicao CONDICIONAL do sub-vetor X∗ = (X1, X2) quando e conhecido apenas o valor de X4.

• Comparando as tres ultimas respostas que voce forneceu, qual das duas variaveis isoladamente, X3 ou X4, diminui a incerteza acerca de X2 mais fortemente? Isto e, se voce tivesse de escolher apenas uma delas, X3 ou X4, qual voce iria preferir se seu objetivo fosse predizer o valor de X2?

• Considere a melhor preditora para X2 que voce escolheu, dentre X3 ou X4, na questao anterior. Digamos que tenha sido X4. Avalie quanto conhecer a outra vari´avel (neste caso, X3) reduz ADICIONALMENTE a incerteza acerca de X3. Isto e, compare V(X2|X4) com V(X2|X3, X4).

Resposta de todas as perguntas acima: (em ordem)

• Estimativas de μ e Σ:

In [6]:
```
setosa <- iris[iris$Species == "setosa", 1:4]
pairs(setosa) # plots de pares das 4 variaveis
apply(setosa, 2, mean) # media aritmetica de cada variavel
cov(setosa) # estimativa da matriz de covariancia
cor(setosa) # estimativa da matriz de correlacao
round(cor(setosa),2) # valores arrendondados em duas casas decimais
apply(setosa, 2, mean) # estimativa de mu
round(cov(setosa),3)
```

Sepal.Length	5.006
Sepal.Width	3.428
Petal.Length	1.462
Petal.Width	0.246

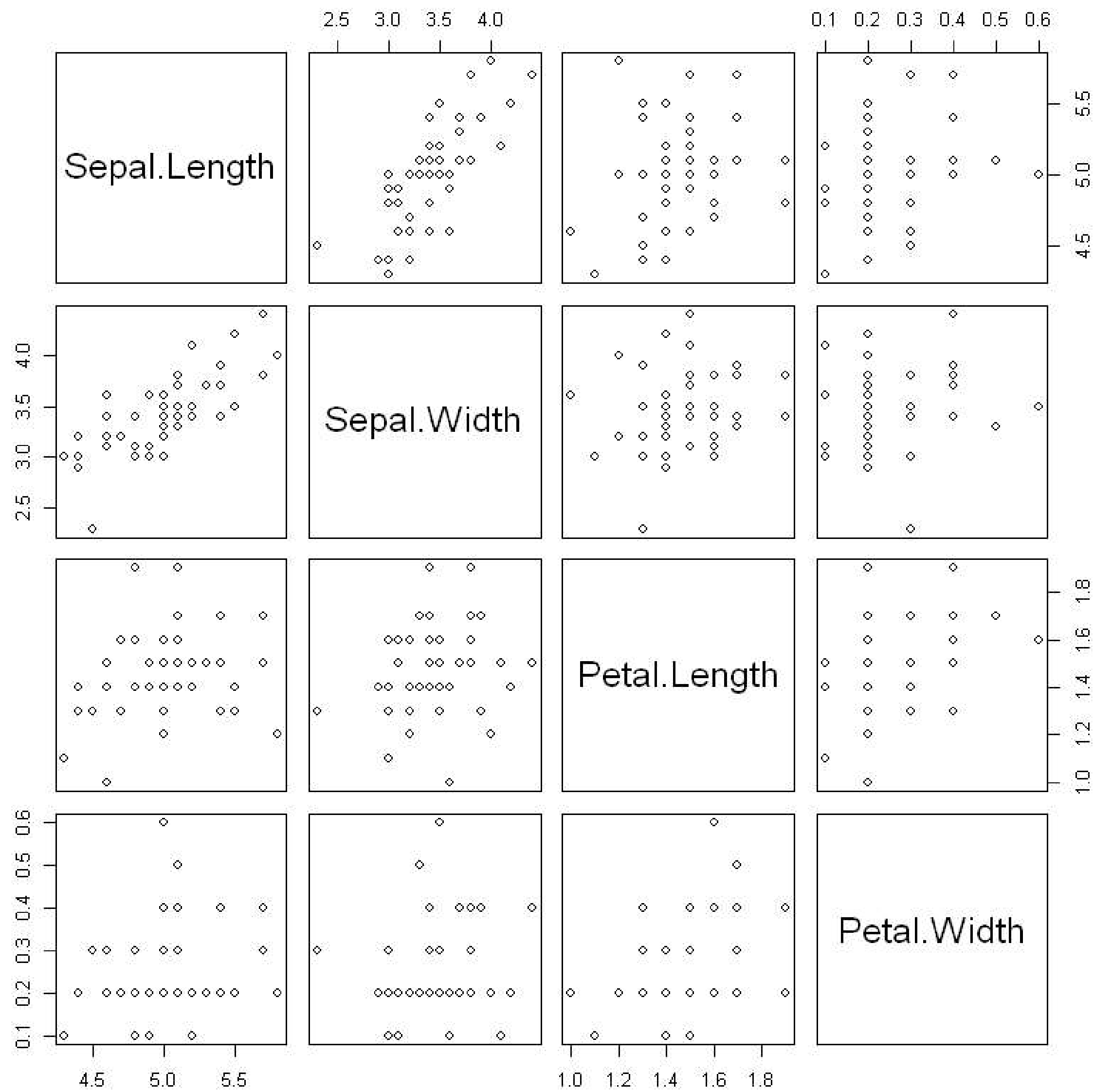
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.12424898	0.099216327	0.016355102	0.010330612
Sepal.Width	0.09921633	0.143689796	0.011697959	0.009297959
Petal.Length	0.01635510	0.011697959	0.030159184	0.006069388
Petal.Width	0.01033061	0.009297959	0.006069388	0.011106122

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	1.0000000	0.7425467	0.2671758	0.2780984
Sepal.Width	0.7425467	1.0000000	0.1777000	0.2327520
Petal.Length	0.2671758	0.1777000	1.0000000	0.3316300
Petal.Width	0.2780984	0.2327520	0.3316300	1.0000000

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	1.00	0.74	0.27	0.28
Sepal.Width	0.74	1.00	0.18	0.23
Petal.Length	0.27	0.18	1.00	0.33
Petal.Width	0.28	0.23	0.33	1.00

Sepal.Length	5.006
Sepal.Width	3.428
Petal.Length	1.462
Petal.Width	0.246

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.124	0.099	0.016	0.010
Sepal.Width	0.099	0.144	0.012	0.009
Petal.Length	0.016	0.012	0.030	0.006
Petal.Width	0.010	0.009	0.006	0.011



• A partir da matriz de correlacao cor(setosa), o comprimento X1 e a largura X2 das sepalas sao as variaveis mais correlacionadas: ρ12 = 0.74. A largura da sepala X2 e o comprimento da petala X3 sao as menos correlacionadas, com ρ23 = 0.18.

• A distribuicao do sub-vetor X∗ = (X1, X2), o comprimento e largura da sepala, vem diretamente dos elementos 1 e 2 de μ e do bloco da matriz Σ:

.

$$\mathbf{X}^* = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\boldsymbol{\mu}[1:2], \boldsymbol{\Sigma}[1:2,1:2]\right) = N_2\left(\begin{pmatrix} 5.006 \\ 3.428 \end{pmatrix}, \begin{bmatrix} 0.124 & 0.099 \\ 0.099 & 0.144 \end{bmatrix}\right)$$

•A distribuicao condicional de X∗ = (X1, X2) quando sao conhecidos os valores (x3, x4) das v.a.'s (X3, X4).

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \Big| \begin{pmatrix} X_3 = x_3 \\ X_4 = x_4 \end{pmatrix} \sim N_2\left(\mathbf{m}, \mathbf{V}\right)$$

onde, usando a notacao das notas de aula

$$\begin{aligned} \mathbf{m} &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\begin{pmatrix} x_3 - \mu_3 \\ x_4 - \mu_4 \end{pmatrix} \\ &= \begin{pmatrix} 5.006 \\ 3.428 \end{pmatrix} + \begin{bmatrix} 0.016 & 0.010 \\ 0.012 & 0.009 \end{bmatrix}\begin{bmatrix} 0.030 & 0.006 \\ 0.006 & 0.011 \end{bmatrix}^{-1}\begin{pmatrix} x_3 - 1.462 \\ x_4 - 0.246 \end{pmatrix} \\ &= \begin{pmatrix} 5.006 \\ 3.428 \end{pmatrix} + \begin{bmatrix} 0.399 & 0.712 \\ 0.247 & 0.702 \end{bmatrix}\begin{pmatrix} x_3 - 1.462 \\ x_4 - 0.246 \end{pmatrix} \end{aligned}$$

Para x3 = 1.8 e x4 = 0.6 temos

$$\mathbf{m} = \begin{pmatrix} 5.006 \\ 3.428 \end{pmatrix} + \begin{pmatrix} 0.387 \\ 0.332 \end{pmatrix} = \begin{pmatrix} 5.393 \\ 3.760 \end{pmatrix}$$

Quanto a matriz de covariancia V para a distribuicao condicional, temos ##

$$\begin{aligned} \mathbf{V} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \\ &= \begin{bmatrix} 0.124 & 0.099 \\ 0.099 & 0.144 \end{bmatrix} - \begin{bmatrix} 0.016 & 0.010 \\ 0.012 & 0.009 \end{bmatrix}\begin{bmatrix} 0.030 & 0.006 \\ 0.006 & 0.011 \end{bmatrix}^{-1}\begin{bmatrix} 0.016 & 0.012 \\ 0.010 & 0.009 \end{bmatrix} \\ &= \begin{bmatrix} 0.110 & 0.088 \\ 0.088 & 0.134 \end{bmatrix}. \end{aligned}$$

Temos DP1 =√V(X1) = √0.124 = 0.352 e √V(X1|X3 = 1.8, X4 = 0.6) = √0.110 = 0.332.

• Queremos a distribuicao condicional do sub-vetor X∗ = (X1, X2) quando e conhecido apenas o valor de X3 = 1.8. Neste caso, e como se a variavel X4 nao existisse: ela nao esta envolvida. Vamos obter a distribuicao conjunta do vetor (X1, X2, X3) e entao usar a nossa formula de condicional da normal mutivariada. Para obter a distribuicao marginal de (X1, X2, X3), basta olhar μ e Σ e ignorar as entradas associadas com X4. Temos

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}\right) = N_3\left(\begin{bmatrix} 5.006 \\ 3.428 \\ 1.462 \end{bmatrix}, \begin{bmatrix} 0.124 & 0.099 & 0.016 \\ 0.099 & 0.144 & 0.012 \\ 0.016 & 0.012 & 0.030 \end{bmatrix}\right)$$

Dividindo este vetor em dois blocos, representados por letras em negrito e com indexacao ligada aos blocos (e nao as variaveis), podemos usar as formulas derivadas em sala de aula:

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3 \left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} , \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \right) =\sim N_3 \left(\begin{bmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \end{bmatrix} , \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

Temos

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \Big| X_3 = 1.8 \sim N_2 (\mathbf{m}, \mathbf{V})$$

onde, usando a nota¸c˜ao das notas de aula,

$$\begin{aligned} \mathbf{m} &= \boldsymbol{m}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\boldsymbol{x}_2 - \boldsymbol{m}_2) \\ &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \end{bmatrix} [\sigma_{33}]^{-1} (1.8 - \mu_3) \\ &= \begin{pmatrix} 5.006 \\ 3.428 \end{pmatrix} + \begin{bmatrix} 0.016 \\ 0.012 \end{bmatrix} [0.030]^{-1} (1.8 - 1.462) \\ &= \begin{pmatrix} 5.006 \\ 3.428 \end{pmatrix} + \begin{bmatrix} 0.533 \\ 0.400 \end{bmatrix} (1.8 - 1.462) \\ &= \begin{pmatrix} 5.006 \\ 3.428 \end{pmatrix} + \begin{pmatrix} 0.180 \\ 0.135 \end{pmatrix} = \begin{pmatrix} 5.186 \\ 3.563 \end{pmatrix} \end{aligned}$$

e

$$\begin{aligned} \mathbf{V} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} - \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \end{bmatrix} [\sigma_{33}]^{-1} \begin{bmatrix} \sigma_{31} & \sigma_{32} \end{bmatrix} \\ &= \begin{bmatrix} 0.124 & 0.099 \\ 0.099 & 0.144 \end{bmatrix} - \begin{bmatrix} 0.016 \\ 0.012 \end{bmatrix} [0.030]^{-1} \begin{bmatrix} 0.016 & 0.012 \end{bmatrix} \\ &= \begin{bmatrix} 0.115 & 0.093 \\ 0.093 & 0.139 \end{bmatrix} \end{aligned}$$

• Temos V(X2|X4 = 0.6) = 0.136 < 0.139 = V(X2|X3 = 1.8). Assim, saber que X4 = 0.6 leva a uma menor incerteza acerca do valor de X2 que aquela que resta quando X3 = 1.8. Para predizer X2, saber o valor de X4 e melhor que saber o valor de X3. Observe que 0.139 = V(X2|X3 = 1.8) = V(X2|X3 = x) para todo x, bem como 0.136 = V(X2|X4 = 0.6) = V(X2|X4 = x) para todo x. Portanto, a conclusao sobre a maior reducao da incerteza de X2 alcancada pelo conhecimento do valor de X3, nao depende dos valores especıficos x3 = 1.8 e x4 = 0.6 usados no exercıcio. Terıamos a mesma conclusao com quaisquer dois valores para x3 e x4 pois as variancias condicionais nao variam com x3 e x4.

• Entre X3 e X4, a melhor preditora de X2 e X4. Acrescentar o conhecimento sobre o valor de X3 ao conhecimento de que X4 = 0.6 reduz pouco a variabilidade (ou incerteza) acerca de X2:

$$0.134 = \mathbb{V}(X_2|X_3 = 1.8, X_4 = 0.6) < \mathbb{V}(X_2|X_4 = 0.6) = 0.136 < \mathbb{V}(X_2) = 0.144$$

Exercício 10

10. Seja $\mathbf{X} = (X_1, X_2, X_3)'$ um vetor aleatório com distribuição normal multivariada com $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)' = [-1, 0, 2]'$ e

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Seja $\mathbf{b}$ um vetor k-dimensional e $\mathbf{C}$ uma matriz $k \times 3$ formada por constantes. Uma das propriedades da normal multivariada é que a distribuição do vetor $\mathbf{b} + \mathbf{CX}$ de dimensão k é normal com vetor de médias $\mathbf{b} + \mathbf{C}\mu$ e matriz de $k \times k$ covariância $\mathbf{C\Sigma C}^t$. Use esta propriedade para obter a distribuição das seguintes variáveis:

• Distribuicao marginal de X1, de X2 e de X3.

• Distribuicao de um indicador composto pelas 3 variaveis: T = 0.4X1 + 0.3X2 + 0.3X3.

• Distribuicao de um indicador composto pelas 3 variaveis normalizadas: T = 0.4(X1 − 10)/2 +0.3(X2 − 20)/√30 + 0.3(X3 + 50)/√94.

• Distribuicao conjunta de (X1 − X2, 4X1 + 2X2 − X3).

• Distribuicao conjunta de (X1, aX1 + bX2 + cX3). onde a, b, c s˜ao constantes reais. Em particular, encontre a covariancia entre X1 e o indicador Y = aX1 + bX2 + cX3 formado pela combinacao linear de X1, X2 e X3.

Exercício 3 - Capitulo 6

3. Consideramos um modelo probabilıstico para um problema de diagnostico de falhas. Uma variavel binaria C representa a integridade de uma unidade de disco: C = 0 significa que esta operando normalmente e C = 1 significa esta em estado de falha. Quando a unidade est´a em funcionamento, monitora-se continuamente usando um temperatura e sensor de choque, e registra duas caracterısticas binarias, X e Y . Temos X = 1 se o drive foi sujeito a choque (por exemplo, caiu), e X = 0, caso contrario. Temos Y = 1 se a unidade a temperatura j´a foi acima

de 70o e Y = 0, caso contrario. A tabela abaixo define a funcao de massa de probabilidade conjunta dessas tres variaveis aleatorias:

x	y	c	$p_{XYC}(x, y, c)$
0	0	0	0.10
0	1	0	0.20
1	0	0	0.20
1	1	0	0.10
0	0	1	0.00
0	1	1	0.10
1	0	1	0.05
1	1	1	0.25

Forneca o valor num´erico das probabilidades abaixo:

• Qual e a probabilidade P(C = 1)?

Solução: P(C = 1) = 0.0 + 0.1 + 0.5 + 0.25 = 0.85

• Qual e a probabilidade P(C = 0|X = 1, Y = 0)?

Solução: P(C = 0|X = 1, Y = 0) = 0.2

• Qual e a probabilidade P(X = 0, Y = 0)?

Solução: P(X = 0, Y = 0) = 0.1

• Qual e a probabilidade P(C = 0|X = 0)?

Solução: P(C = 0|X = 0) = 0.3

• Sao X e Y independentes? Justifique sua resposta.

Solução: 0.4*0.35 != 0.1 para o primeiro valor da tabela, e como isso difere concluimos que X e Y não sao independentes.

In [ ]:

In [ ]:

In [ ]:

In [ ]: