Lista de Exercício e Alguns Conceitos

20 de Agosto de 2007

As notas de aula apenas apresentam os tópicos de Interesse. O aluno pode consultar os livros abaixo para obter conhecimentos adicionais. O aluno deve tentar solucionar todos os exercícios destas notas antes das provas e procurar o professor para esclarecer suas dúvidas.

Livros Textos:

1. Anderson, T. A. (1984), An Introduction To Multivariate Statistical
2. Mardia, Kent and Bibby (1979) Multivariate Analysis
3. Johnson and Wichern (1982) Applied Multivariate Statistical Analysis

CONTEÚDO PROGRAMÁTICO

1. Revisão de vetores aleatórios: Função de distribuição vetorial, densidade conjunta. Momentos vetoriais: vetor esperança, matrizes de covariância e de correlação.
2. Distribuição normal multivariada: Distribuições marginais e condicionais. Correlação parcial e múltipla. Independência. Distribuição de funções lineares de um vetor aleatório normal multivariado, distribuição da média amostral. Estimadores do vetor esperança e da matriz de covariância. Inferência sobre a matriz de correlação.
3. Testes para médias de populações normais multivariadas: Distribuição de Wishart, estatística T2 de Hotelling e sua distribuição. Teste para média de uma população com matriz de covariância conhecida e desconhecida. Intervalos de confiança simultâneos para uma combinação linear qualquer das

componentes. Teste de comparação de médias de duas populações. Intervalos de confiança simultâneos para uma combinação linear qualquer das componentes da diferença de dois vetores normais multivariados. Teste de igualdade de matrizes de covariância entre K populações.

4. Coeficientes de correlação simples, parciais e múltiplos: Estimadores de máxima verossimilhança para os coeficientes de correlação simples. Distribuição do coeficiente de correlação simples no caso de população normal bivariada. Testes de hipótese sobre o coeficiente de correlação simples. Coeficientes de correlação parciais, propriedades, interpretação, fórmula de recorrência e estimadores de máxima verossimilhança. Coeficiente de correlação múltiplos, propriedades, interpretação e estimadores.

5. Análise de variância multivariada: Modelo linear geral multivariado. Análise de variância para critério único de classificação, blocos aleatoriezados. Análise de variância com dois critérios. Análise de regressão multivariada. Análise de covariância multivariada. Ajuste de curvas para medidas repetidas.

Avaliação: 2 provas e algumas questões para casa, onde o alunos serão sorteados em aula.

1 Introdução

Muitos elementos estudados na natureza como objetos, seres vivos e e outros possuem várias variáveis de interesse.

A análise multivariada é um conjunto de técnicas que são adequadas para situações onde várias variáveis estão envolvidas.

Suponha que estamos interessados nas medidas de um indivíduo que são relevantes para o estudo da pressão arterial. Algumas medidas de interesse são :idade, peso, superfície do corpo, medida do stress e outras variáveis que um pesquisador julgue relevante.

Na problemas do interesse da análise multivariada têm se a matrix de dados definida por

$$X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix},$$

Nome	Peso	Altura
João	73.9	170
Lisa	70.1	165
Pedro	84.8	180
Maria	59.9	155
Júnior	82.0	175
Paula	65.6	160
Sonia	68.2	165

Tabela 1: Dados sobre Peso e Altura

a i-ésima linha é corresponde a um indivíduo e a j-ésima coluna corresponde a uma variável. Assim, o elemento x_{ij} corresponde ao valor observado da j-ésima variável do i-ésimo indivíduo.

Seja x_i a i-ésima linha escrita como coluna, que é dada por

$$x_i = (x_{i1}, \ldots, x_{ip})',$$

e seja $x_{(j)}$ a j-ésima coluna de X

$$x_{(j)} = (x_{1j}, \ldots, x_{nj})'.$$

Na Tabela 1 são apresentados dados sobre peso e altura relativos a 7 pessoas. Dessa forma, as duas últimas colunas desta tabela representam uma matriz de dados de dimensão 7×2.

Similarmente ao caso univariado, a etapa inicial da análise estatística é a estatística descritiva, onde as duas principais medidas são extensões da média e da variância para o caso multivariado.

Vetor de médias é dado

$$\bar{x}' = (\bar{x}_1, \ldots, \bar{x}_p),$$

onde $\bar{x}_i = \frac{1}{n}\sum_{r=1}^{n} x_{ri}$.

E a matriz de covariância é dada por

$$S = (s_{ij}),$$

onde $s_{ij} = \frac{1}{n-1}\sum_{r=1}^{n}(x_{ri} - \bar{x}_i)(x_{rj} - \bar{x}_j)$.

Exemplo Para os dados da Tabela 1, têm-se

$$\bar{x} = \begin{pmatrix} 72.07 \\ 167.14 \end{pmatrix}$$

$$S = \begin{pmatrix} 78.81 & 75.57 \\ 75.57 & 73.80 \end{pmatrix}$$

código em R

```
altura<-c(170,165,180,155,175,160,165)
peso<- c(73.9 70.1 84.8 59.9 82.0 65.6 68.2)
mean(altura)
mean(peso)
var(altura)
var(peso)
cov(altura,peso)
```

Estas quantidades podem ser convientemente escritas em notação matricial conforme o texto apresentado abaixo.

$$\bar{x} = \frac{1}{n} X'1,$$

onde $1' = (1, \ldots, 1)$ é um vetor de dimensão n.

Para a matriz de covariância

$$S = \frac{1}{n}(X'X - \frac{1}{n}X'11'X),$$

ou ainda, se $H = I - \frac{1}{n}11'$,

$$S = \frac{1}{n}X'HX$$

Matriz de Correlação

$$R = (r_{ij}),$$

onde $r_{ij} = \frac{s_{ij}}{s_i s_j}$.

Problema 1. *O programa R dispõe de uma boa quantidade de banco de dados disponivéis em sua blibioteca. Para acessar estes bancos basta usar o comando data( ). Use o comando data(ToothGrowth) para ter acesso a este banco de dados. Digite a palavra ToothGrowth e a a matriz de dados irá surgir na tela do computador. Faça os seguintes cálculos:*

1. As médias de len mean(ToothGrowth[,1])e dose e as matrizes de covariância e correlação destas duas variáveis.

2. Faça um diagrama de dispersão. Que análises podem ser feita deste diagrama a partir da covariância e da correlação?

2 Algebra Matricial e Vetores Aletórios

Os resultados teóricos de análise multivariada precisam de alguns teoremas e axiomas de disciplinas como: probabilidade, geometria analítica e algebra linear. Sugere-se que o aluno consulte livros probabilidade, geometria analítica e algebra linear para obter informações mais detalhadas. Mardia et al (1979) apresenta um apêndice bastante extenso sobre estes tópicos.

2.1 Traço de Uma Matriz

O traço de uma matriz é utilizado na simplificação de várias provas feitas durante o curso de análise multivariada.

Sejam A, B e C matrizes. O traço da matriz A pode ser denominado de $tr(A)$ e é calculado por

$$tr(A) = \sum a_{ii},$$

isto é, a soma dos elementos da diagonal principal de A. Por exemplo, se $\mathbf{X}$ é igual a

$$\mathbf{X} = \begin{pmatrix} 7 & 3 & 2 \\ 4 & 6 & 4 \\ 2 & 3 & 5 \end{pmatrix},$$

o traço de $\mathbf{X}$ é igual a 18.

Existem várias propriedades interessantes do traço. Alguma delas são:

1. $tr\alpha A \pm B = \alpha trA \pm trB$

2. $$\begin{aligned} tr\Sigma x_i' A x_i &= \Sigma tr x_i' A x_i \\ &= \Sigma tr A x_i x_i' \\ &= tr A \Sigma x_i x_i' \end{aligned}$$

2.2 Autovalores e Autovetores

Autovalores e autovetores são conceitos fundamentais para análise multivariada. Por exemplo, a região de confiança para o vetor de médias, que será vista posteriormente, é baseada nestes dois conceitos.

Seja $\mathbf{A}$ uma matriz $p \times p$ e seja $\mathbf{b}$ um vetor $p \times 1$. $\mathbf{b}$ é um autovetor de $\mathbf{A}$ se $\mathbf{A}\mathbf{b} = \lambda\mathbf{b}$. A constante λ corresponde autovalor associado ao autovetor $\mathbf{b}$.

Para encontrar o autovetores de uma matriz $\mathbf{A}$, é necessário encontrar as raízes da equação

$$det(\mathbf{A} \; - \; \lambda I).$$

Se a matriz $\mathbf{A}$ tem dimensão $p \times p$ e é simétrica, esta matriz possue p pares de autovetores e autovalores que podem ser denotados por

$$e_1, \lambda_1 \; e_2, \lambda_2, \ldots, e_k, \lambda_k.$$

A decomposição espectral de uma matriz $\mathbf{A}$ corresponde a sua representação a partir de seus autovalores e autovetores. Se a matriz $\mathbf{A}$ é simétrica, esta pode ser escrita como

$$A = \lambda_1 \; e_1 e_1' \; + \; \ldots + \lambda_k e_k e_k'.$$

Problema 2. *Considere a matriz*

$$\begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}.$$

Obtenha os seguintes itens:

1. *Ache os autovalores de A e seus respectivos autovetores*
2. *Escreva a decomposição espectral de A.*
3. *Encontre a decomposição espectral de A^{-1}.*
4. *Compare as duas decomposições.*

Problema 3. *Se isto for possível, encontre a decomposição espectral da matriz*

$$\begin{pmatrix} 3 & 0 & -4 \\ 0 & 3 & 5 \\ 0 & 0 & -1 \end{pmatrix}.$$

Problema 4. *Se isto for possível, encontre a decomposição espectral da matriz*

$$\begin{pmatrix} 3 & -3 & -4 \\ 0 & 3 & 5 \\ 0 & 0 & -1 \end{pmatrix}.$$

Problema 5. *Encontre a decomposição espectral da matriz associado ao elipsóide* $2x^2 + 2y^2 + 2z^2 + 2xy + 2xz + 2yz$

sugestão: *escrever o elipsóide em forma matricial, isto é,*

$$\mathbf{u}' \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \mathbf{u},$$

.

onde $\mathbf{u}' = (x, y, z)$. *Basta indentificar na matriz acima quais são os coeficientes* a_{ij} *que correspondem aos termos de* $2x^2 + 2y^2 + 2z^2 + 2xy + 2xz + 2yz$, *o que é feito a partir do cálculo com produto acima.*

2.3 Formas Quadráticas

Uma forma quadrática para as variáveis $x_1, \ldots, x_n$ é uma expressão do tipo

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j,$$

ou em forma matricial

$$x' A x.$$

Uma matriz simétrica A e sua correspondente forma quadrática $x'Ax$ são denominadas positiva definida se $x'Ax > 0$. Se $x'Ax \geqslant 0$, a forma quadrática e a matriz são denominadas positivas definidas.

Problema 6. *Escreva a matrix da forma quadrática*

$$2x^2 - 2xy + y^2 + 4xz - 3z^2$$

e determine se esta é positiva definida.

Problema 7. *Para a matriz*

$$\mathbf{A} = \begin{pmatrix} 9 & -2 \\ -2 & 6 \end{pmatrix}$$

forneça:
a)Encontre a decomposição espectral de $\mathbf{A}$.
b)Mostre que $\mathbf{A}$ *é positiva definida.*

2.4 Matriz Raiz Quadrada

A partir da decomposição espectral de uma matriz, é possível definir o conceito de matriz raiz quadrada.

Seja A uma matriz $p \times p$ com decomposição espectral

$$A = \sum_{i=1}^{p} \lambda_i e_i e_i',$$

onde e_i é um autovetor de A com norma igual a 1.

Considere a matriz de autovetores $P = [e_1, \ldots, e_p]$.

A matriz A pode escrita como

$$A = \sum_{i=1}^{p} \lambda_i e_i e_i' = P \Lambda P',$$

onde Λ é uma matriz diagonal e os elementos da sua diagonal são os autovalores de A.

A matriz raiz quadrada de A é definida por

$$A^{1/2} = \sum_{i=1}^{p} \sqrt{\lambda_i} e_i e_i' = P \Lambda^{1/2} P',$$

$\Lambda^{1/2}$ é uma matriz diagonal cujo os elementos da diagonal principal são as raízes quadradas dos autovalores de A.

A matriz raiz quadrada tem a seguintes propriedades:

1. $(A^{1/2})' = A^{1/2}$
2. $A^{1/2} A^{1/2} = A$

3. $(A^{1/2})^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} e_i e_i' = P\Lambda^{-1/2}P'$, onde $\Lambda^{-1/2}$ é uma matriz diagonal e $1/\sqrt{\lambda_i}$ é o i-ésimo elemento diagonal de $\Lambda^{-1/2}$.

4. $A^{1/2}A^{-1/2} = A^{-1/2}A^{1/2}$ e $A^{-1/2}A^{-1/2} = A^{-1}$ onde $A^{-1/2} = (A^{1/2})^{-1}$.

Problema 8. *Considere a matriz $A^{1/2} = \sum_{i=1}^{p} \sqrt{\lambda_i} e_i e_i'$. Verifique as 4 propriedades da matriz raiz quadrada supracitada.*

2.5 Vetores e Matrizes Aleatórias

Um vetor aleatório é aquele cujos os elementos são variáveis aleatórias. Da mesma forma, uma matriz aleatória é aquela cujos os elementos são variáveis aleatórias.

Seja $X = \{x_{ij}\}$ uma matriz aleatória $(n \times p)$, o valor esperado de X é dado por

$$E(X) = \begin{pmatrix} E(X_{11}) & \dots & E(X_{1p}) \\ \vdots & \ddots & \vdots \\ E(X_{p1}) & \dots & E(X_{pp}) \end{pmatrix}$$

Problema 9. *Sejam A e B duas matrizes compatíveis, prove que*

$$E(X+Y) = E(X) + E(Y)$$

$$E(AXB) = AE(X)B$$

2.6 Vetor de Médias e Matriz de Covariância

Se X é um vetor aletório, cada elemento de X é uma variável aleatória que possue uma distribuição de probabilidade marginal.

As médias, μ_j, e as variâncias, σ_j^2, de X_j são definidas por $\mu_j = E(X_j)$ e $\sigma_j^2 = E(X_j - \mu_j)^2$.

Se $X = [X_1, \dots, X_p]$ é um vetor aleatório, a função de densidade conjunta de X é dada por

$$f(x_1, \dots, x_p) = f_1(x_1) \dots f_p(x_p).$$

Se as variáveis aleatórias $X = [X_1, \dots, X_p]$ são independentes

$$f(x_1,\ldots,x_p) = f_1(x_1)\ldots f_p(x_p).$$

Por outro lado, para garantir a independência de X_i e X_k temos que verificar se $Cov(X_i, X_k) = 0$.

A matriz de covariâncias é definida por $\Sigma = \{\sigma_{ij}\}$, onde $\sigma_{ij} = E(x_i - \mu_i)(x_j - \mu_j)$.

A matriz de correlação é dada por $\rho = \{\rho_{i,j}\}$ onde $\rho_{i,j} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$.

Problema 10. *Considere a densidade de probabilidade bivariada*

$f(x_1, x_2) = 2, \quad 0 < x_1 < x_2 < 1.$

Calcule as densidades marginais $f_1(x_1)$ *e* $f_2(x_2)$*. Neste caso,* x_1 *e* x_2 *são independentes?*

3 Normal Multivariada

É possível se obter a expressão da normal multivariada como uma extensão da normal univariada. Considere a normal univariada

$$f(x;\mu,\sigma) = k \exp -\frac{1}{2}(x-\mu)\sigma^{-1}(x-\mu)$$

As quantidades univariadas podem ser redefinidas para o caso multivariado

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_{1,1} & \ldots & \sigma_{1,m} \\ \vdots & \ddots & \vdots \\ \sigma_{k,1} & \ldots & \sigma_{k,m} \end{pmatrix}$$

Substituindo-se x, μ e σ por suas versões multivariadas, temos

$$f(x;\mu,\sigma) = k\exp^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}.$$

A única incógnita para determinar a distribuição de x é k.

3.1 Cálculo da Constante k

$$k = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\exp^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}\,dx_p\dots dx_1.$$

Usando-se o corolário A.1.6 (Vide Anderson, 1984, p. 586), se Σ é positiva definida, então existe uma matriz não singular C tal que

$$C'\Sigma^{-1}C = I,$$

I é a matriz identidade e C' é a tranposta de C.

Considere

$$x - \mu = Cy,$$

onde $y' = (y_1, \dots, y_p)$.

Temos que

$$(x-\mu)'\Sigma^{-1}(x-\mu) = y'C'\Sigma^{-1}Cy = y'y.$$

Como

$$J = |C|,$$

a constante de interesse é dada por

$$k^* = Mod|C|\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\exp^{\{-\frac{1}{2}y'y\}}\,dy_p\dots dy_1.$$

Simplificando-se o integrando, temos

$$\exp^{\{-\frac{1}{2}y'y\}} = \prod_{i=1}^{p}\exp^{-\frac{1}{2}y_i^2}.$$

O valor da constante de interesse é dado por

$$
\begin{aligned}
k &= Mod|C| \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp^{-\frac{1}{2}y_1^2} \dots \\
&\quad \exp^{-\frac{1}{2}y_p^2}\, dy_p \dots dy_1 \\
&= Mod|C| \int_{-\infty}^{\infty} \exp^{-\frac{1}{2}y_p^2}\, dy_p \dots \\
&\quad \int_{-\infty}^{\infty} \exp^{-\frac{1}{2}y_1^2}\, dy_1 \\
&= Mod|C|(\sqrt{2\pi})^p.
\end{aligned}
$$

Calculando-se o determinante de C, tem-se

$$|C'||\Sigma^{-1}||C| = I,$$

o que resulta em

$$Mod|C| = \frac{1}{\sqrt{|\Sigma^{-1}|}}.$$

Assim, a constante de interesse é

$$\frac{1}{k} = \sqrt{\Sigma^{-1}}(2\pi)^{-\frac{1}{2}p}.$$

Portanto, a função de densidade da normal multivariada é dada por

$$f(x) = \frac{\sqrt{|A|}(2\pi)^{\frac{1}{2}p}}{\exp}^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

3.2 Propriedades da Normal Multivariada

A normal multivariada apresenta várias propriedades de interesse.

Os contornos são as curvas de superfícies obtidas apartir de

$$(x-\mu)'\Sigma^{-1}(x-\mu) = c^2, \tag{1}$$

o que corresponde a um elipsóide centrado em μ.

Se λ_i e e_i, com $i = 1, \dots, p$, são os autovalores e autovetores de Σ, os eixos do elipsóide (1) são $\pm c\sqrt{\lambda_i}e_i$.

Se X tem distribuição normal multivariada, é possível observar as propriedades:

1. Combinações lineares das componentes de X são normalmente distribuídas.

2. Todos os subconjuntos das componentes de X tem distribuição normal.

3. Covariância nula entre componentes implica que estas são independemente distribuídas

4. A distribuição condicional das componentes são normais multivariadas.

Algumas destas propriedades podem ser melhor formalizadas com os seguintes teorema

Teorema 1. *Se a distribuição de X é normal multivariada com vetor de me´dias μ e matriz de covariâncias Σ, a função característica de X é dada por*

$$\phi_X(t) = exp(it'\mu - \frac{1}{2}t'\Sigma t).$$

A função característica é muito utilizada para encontrar a distribuição da soma de variáveis aleatórias. Isto deve-se a propriedade: se X e Y são variáveis aleatórias independentes, temos $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

Teorema 2. *Se $X \sim N_p(\mu, \Sigma)$ e $Y = AX + c$ então $Y \sim N_q(A\mu + c, A\Sigma A')$.*

Teorema 3. *A distribuição condicional de X_2 para um dado vetor X_1 é dada por*

$$X_2/X_1 \sim N_s(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \Sigma_{22,1})$$

Problema 11. *Considere X um vetor aleatório com média $\mu = [\mu_1, \mu_2, \mu_3]$ e matriz de covariância*

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 \\ & 1 & 0 \\ & & 1 \end{pmatrix}$$

Mostre que a distribuição condicional de (X_1, X_2) dado X_3 tem vetor de média $[\mu_1 + \rho^2(x_3 - \mu_3), \mu_2]$ e matriz de covariância

$$\begin{pmatrix} 1 - \rho^4 & \rho \\ \rho & 1 \end{pmatrix}$$

Problema 12.

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

Avalie se variáveis abaixo são independentes ou não:

1. X_1 *e* X_2
2. X_2 *e* X_3
3. (X_1, X_2) *e* X_3
4. $\frac{X_1+X_2}{2}$ *e* X_3.

Problema 13. *Seja* $X \sim N_3(\mu, \Sigma)$ *com* $\mu' = [2, 3, -1]$ *e*

$$\Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}$$

1. *Ache a distribuição de* $3X_1 - 2X_2 + X_3$.
2. *Ache a distribuição de* $X_1 + X_2$.

Problema 14. *Se* $X \sim N_p(\mu, \Sigma)$ *e* a *é um vetor fixado, mostre que*

$$f = \frac{a'(X - \mu)}{\sqrt{a'\Sigma a}} \sim N(0, 1)$$

Problema 15. *Se* X_1, X_2, X_3 *são independentes e identicamente distribuídas (i.i.d)* $N_p(\mu, \Sigma)$ *e se* $Y_1 = X_1 + X_2$, $Y_2 = X_2 + X_3$ *e* $Y_3 = X_1 + X_3$, *obtenha as distribuições destas variáveis e a distribuição condicional de* Y_1 *dado* Y_2 *e* Y_3.

4 Verificação da Normalidade Multivariada

Se X, um vetor aleatório de dimensão p, possui a distribuição marginal de seus componentes normal univariada, isto não implica que a distribuição de X é normal multivariada. Assim, é preciso procurar um método que seja apropriado para avaliar a distribuição multivariada de X.

Antes de considerar o caso multivariado, o caso bivariado pode ser avaliado, inicialmente, com os seguintes critérios:

1. Obter q-q plots e aplicar testes de normalidade (Kolmogorov ou outro), para cada variável individualmente.

2. Fazer diagramas de dispersão (XY) e verifique se o conjunto dos pontos possuem aproximadamente a forma de uma elipse.

3. Verificar se existem pontos aberrantes que precisam ser cuidadosamente analisados.

Para caso multivariado, um resultado muito importante é obtido a partir do cálculo de

$$d_j^2 = (x_j - \bar{x})' S^{-1} (x_j - \bar{x}) \qquad j = 1, \ldots, n,$$

onde $x_1, \ldots, x_n$ são as observações amostrais.

Para avaliar a normalidade multivariada, verifica-se por um Q-Q plot se os $d_j's$ seguem uma distribuição χ_p^2. Se os $d_j's$ seguem esta distribuição, a suposição de normalidade multivariada é adequada.

Problema 16. *Avaliando a normalidade multivariada no R.*

Tente entender e depois utilizar os comandos abaixo. Explique o cada comando faz. Explique o resultado esperado.

```
mu<-array(0,c(2,1))
Sigma<-diag(2,2,2)
dados<-mvrnorm(n = 30, mu, Sigma, tol = 1e-6, empirical = FALSE)
S<-cov(dados)
xbar<-array(0,c(2,1))
xbar[1,]<-mean(dados[,1])
xbar[2,]<-mean(dados[,2])
```

Problema 17. *Para calcular o termo:*

$$d_j^2 = (x_j - \bar{x})' S^{-1} (x_j - \bar{x}) \qquad j = 1, \ldots, n,$$

onde $x_1, \ldots, x_n$ são as observações amostrais, é necessário fazer um loop e para calcular os valores de d_j^2 e depois fazer um Q-Q Plot. Quais são os comandos para fazer isto no R? Implemente isto no R e analise o resultado.

5 Geometria da amostra e Amostragem Aleatória

Um dos objetivos desta seção é estudar os aspectos geométricos das quantidades amostrais.

Inicialmente, consideramos a interpretação das observações. Cada observação representa um ponto em $\mathbb{R}^{'}$ e a média é um ponto situado no centro das observações.

Problema 18. *Represente os pontos amostrais e a média da matriz de dados abaixo em* $\mathbb{R}^2$.

$$\Sigma = \begin{pmatrix} 3 & 1 \\ -1 & 2 \\ 3 & 5 \end{pmatrix}$$

Cada variável pode ser considerada como um vetor. Assim, este vetor é graficamente representado no espaço p-dimensional.

Problema 19. *Represente as variáveis da matriz de dados do exercício anterior como vetores em* $\mathbb{R}^3$.