**UFC/Universidade Federal do Ceará**
**Depto. de Estatística e Matemática Aplicada/CC - 1ª Avaliação Parcial – 2021/I – 16/07/2021**
**Disciplina: Análise Multivariada - Curso: Estatística/DEMA - Prafª: Sílvia Freitas**
**1ª Parte: Sem consulta**

**01)** Seja o vetor aleatório $\underset{\sim}{X}$ com vetor de médias $\underset{\sim}{\boldsymbol{\mu}}$ e matriz de variâncias-covariâncias $\underset{\sim}{\Sigma}$. Considere sua partição da seguinte forma: $\underset{\sim}{X}^T = (\underset{\sim}{X}_{qx1}^{(1)} \mid \underset{\sim}{X}_{(p-q)x1}^{(2)})$. Sejam os vetores $a_{qx1}$ e $b_{(p-q)x1}$, sendo $U = a^T \underset{\sim}{X}^{(1)}$ e $V = b^T \underset{\sim}{X}^{(2)}$. Encontre:

**(a)** E(U) e Var(U) **(b)** E(V) e Var(V) **(c)** Cov(U,V) **(d)** Mostre que: $Corr(U,V) = \dfrac{a^T \Sigma_{12} b}{\sqrt{a^T \Sigma_{11} a}\sqrt{b^T \Sigma_{22} b}}$

**02)** Suponha um vetor aleatório **X**=$(X_1,X_2,X_3)^T$, com $\Sigma_X = \begin{bmatrix} 6 & 0 & \\ & 4 & \\ 0 & 1 & 2 \end{bmatrix}$ e $\mu_X = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$. Considere as combinações lineares: $Z_1=X_1-3X_3$, $Z_2=X_1+2X_2$ e $Z_3=X_1+X_2+X_3$. Complete $\boldsymbol{\Sigma_X}$ e responda os itens abaixo:

**(a)** Organize estas combinações lineares na forma matricial em função do vetor aleatório **X**=$(X_1,X_2,X_3)^T$.
**(b)** Encontre a esperança e a variância da quantidade que você definiu em (a).
**(c)** Defina uma nova transformação linear de **X**, como sendo $\mathbf{Y}_{2x1}$, de modo que suas componentes sejam independentes. Justifique a sua resposta.

**03)** Seja o vetor aleatório $\boldsymbol{X}_{px1}$, com $Var(X) = \Sigma_X$ e $E(X) = \mu_X$. Considere a transformação ***Y=CX+d***, sendo $\boldsymbol{C}_{pxp}$ uma matriz não-singular de constantes e $\boldsymbol{d}_{px1}$ um vetor de constantes. Mostre que:

**(a)** $\mu_Y = \boldsymbol{C}\mu_X + \boldsymbol{d}$ **(b)** $\Sigma_Y = C\Sigma_X C^T$ **(c)** A igualdade $(Y-\mu_Y)^T \Sigma_Y^{-1}(Y-\mu_Y) = (X-\mu_X)^T \Sigma_X^{-1}(X-\mu_X)$ é satisfeita.

**04)** Os dados abaixo referem-se ao estudo de três índices de custo de vida relacionados aos anos de 1900-1910 nos Estados Unidos. Pesquisadores estavam em dúvida sobre qual dos índices adotar em um estudo econômico. Decidiram, então, formar uma combinação linear desses índices e usar esse resultado para representá-los como um todo.

**(a)** Se eles decidiram utilizar a técnica de componentes principais para encontrar esta combinação linear, e trabalharam com as variáveis padronizadas, que combinação linear eles obtiveram?
**(b)** Que percentual da variabilidade total é explicada por essa combinação linear obtida?
**(c)** Se você fosse um dos pesquisadores, você defenderia a posição de trabalhar apenas com uma combinação linear dos índices ou com mais de uma? Justifique sua resposta.

$$R = \begin{bmatrix} 1 & 0{,}9543 & 0{,}8927 \\ 0{,}9543 & 1 & 0{,}9580 \\ 0{,}8927 & 0{,}9580 & 1 \end{bmatrix}$$

Autovalores: $\lambda_1 = 2{,}8703$ ; $\lambda_2 = 0{,}1073$; $\lambda_3 = 0{,}0224$
Autovetores: $e_1^T$= [0,5726 0,5858 0,5858]
$e_2^T$ = [0,7175 -0,0198 -0,6962]
$e_3^T$ = [0,3965 -0,8101 0,4317]