Instituto de Matemática e Estatística
Universidade de São Paulo

Lista 8

MAE0330 - Análise Multivariada de Dados

Profª Lucia Pereira Barroso

Bruno Groper Morbin - nºUSP 11809875

Luigi Pavarini de Lima - nºUSP 11844642

São Paulo
19 de dezembro, 2022

Exercício 2

Considera-se o arquivo `vendedores.xlsx` para análise dos dados. Ajusta-se então um modelo de regressão multivariada, considerando o `Índice de crescimento de vendas`(X1) e `Índice de lucratividade`(X2) como variáveis respostas e `Criatividade`(X4) e `Habilidade matemática`(X7) como variáveis explicativas. Para o conjunto de dados, tem-se o seguinte dicionário:

- Análise do desempenho:
  - X1: índice de crescimento de vendas;
  - X2: índice de lucratividade e;
  - X3: índice de captação de novas vendas.
- Análise de habilidades:
  - X4: criatividade;
  - X5: raciocínio mecânico;
  - X6: raciocínio abstrato e;
  - X7: habilidade matemática.

Tabela 1: Conjunto de dados ‘vendedores.xlsx‘

Rotulo	X1	X2	X3	X4	X5	X6	X7
1	93.0	96.0	97.8	9	12	9	20
2	88.8	91.8	96.8	7	10	10	15
3	95.0	100.3	99.0	8	12	9	26
4	101.3	103.8	106.8	13	14	12	29
5	102.0	107.8	103.0	10	15	12	32
6	95.8	97.5	99.3	10	14	11	21
7	95.5	99.5	99.0	9	12	9	25
8	110.8	122.0	115.3	18	20	15	51
9	102.8	108.3	103.8	10	17	13	31
10	106.8	120.5	102.0	14	18	11	39
11	103.3	109.8	104.0	12	17	12	32
12	99.5	111.8	100.3	10	18	8	31
13	103.5	112.5	107.0	16	17	11	34
14	99.5	105.5	102.3	8	10	11	34
15	100.0	107.0	102.8	13	10	8	34
16	81.5	93.5	95.0	7	9	5	16
17	101.3	105.3	102.8	11	12	11	32
18	103.3	110.8	103.5	11	14	11	35
19	95.3	104.3	103.0	5	14	13	30
20	99.5	105.3	106.3	17	17	11	27
21	88.5	95.3	95.8	10	12	7	15
22	99.3	115.0	104.3	5	11	11	42
23	87.5	92.5	95.8	9	9	7	16
24	105.3	114.0	105.3	12	15	12	37
25	107.0	121.0	109.0	16	19	12	39
26	93.3	102.0	97.8	10	15	7	23
27	106.8	118.0	107.3	14	16	12	39
28	106.8	120.0	104.8	10	16	11	49
29	92.3	90.8	99.8	8	10	13	17
30	106.3	121.0	104.5	9	17	11	44
31	106.0	119.5	110.5	18	15	10	43
32	88.3	92.8	96.8	13	11	8	10
33	96.0	103.3	100.5	7	15	11	27
34	94.3	94.5	99.0	10	12	11	19
35	106.5	121.5	110.5	18	17	10	42
36	106.5	115.5	107.0	8	13	14	47
37	92.0	99.5	103.5	18	16	8	18
38	102.0	99.8	103.3	13	12	14	28
39	108.3	122.3	108.5	15	19	12	41
40	106.8	119.0	106.8	14	20	12	37
41	102.5	109.3	103.8	9	17	13	32
42	92.5	102.5	99.3	13	15	6	23
43	102.8	113.8	106.8	17	20	10	32
44	83.3	87.3	96.3	1	5	9	15
45	94.8	101.8	99.8	7	16	11	24
46	103.5	112.0	110.8	18	13	12	37
47	89.5	96.0	97.3	7	15	11	14
48	84.3	89.8	94.3	8	8	8	9
49	104.3	109.5	106.5	14	12	12	36
50	106.0	118.5	105.0	12	16	11	39

Obtenção dos coeficientes estimados

Tem-se o seguinte modelo de regressão multivariada:

$$Y_{(n\times p)} = X_{(n\times q)}B_{(q\times p)} + U_{(n\times p)}$$

onde $X \equiv$ matriz das variáveis explicativas, $Y \equiv$ matriz das variáveis respostas, $B \equiv$ coeficientes da regressão, e $U \equiv$ erro aleatório do modelo.

Tem-se também as premissas de $\mathbb{E}(U) = 0$, $Var(U_k) = \Sigma$, onde $U_k \equiv$ k-ésima linha de U, correspondente a k-ésima unidade amostral.

No caso em questão, tem-se $n = 50$ e $p = q = 2$, portanto obtem-se os coeficientes estimados por Mínimos Quadrados da regressão multivariada através do cálculo:

$$\hat{B} = (X^TX)^{-1}X^TY$$

```
> # Obtenção atráves do método multivariado
> Y = as.matrix(vendedores[c("X1","X2")])
> X = as.matrix(vendedores[c("X4","X7")])
>
> B_hat = solve(t(X)%*%X)%*%t(X)%*%Y
> colnames(B_hat) <- NULL; rownames(B_hat) <- NULL
>
> print(B_hat)
```

```
         [,1]     [,2]
[1,] 3.605998 3.638801
[2,] 1.756552 2.006603
```

```
> # Obtenção atráves do método univariado
> y = c(Y)
> x = kronecker(diag(1,nrow=2), X)
>
> b_hat = solve(t(x)%*%x)%*%t(x)%*%y
> print(b_hat)
```

```
         [,1]
[1,] 3.605998
[2,] 1.756552
[3,] 3.638801
[4,] 2.006603
```

Nota-se que `b_hat` (obtido pelo método univariado) é equivalente a `vec(B_hat)` obtido pelo método multivariado.

Verificação de normalidade multivariada

Para verificar uma das premissas do modelo que trata da normalidade multivariada do erro aleatório, calcula-se a distância de Mahalanobis obtida por:

$$(D_M)_i = \sqrt{(\epsilon_i - \mu_\epsilon)^T S^{-1} (\epsilon_i - \mu_\epsilon)}$$

sendo $\epsilon_i \equiv$ resíduos do ajuste da i-ésima unidade amostral, $\mu_\epsilon \equiv$ média amostral dos resíduos ajustados sendo representada por um vetor coluna, e $S \equiv$ matriz de covariância dos resíduos amostral, ou seja $S = \frac{1}{n-1}\sum_{j=1}^{n}(\epsilon_j - \mu_\epsilon)(\epsilon_j - \mu_\epsilon)^T$.

Portanto, visto que $\epsilon = Y - X\hat{B}$, obtém-se as seguintes distâncias de Mahalanobis por unidade amostral:

```
> e <- Y - X%*%B_hat
> mu_e <- t(c(mean(e[1]), mean(e[2]))); mu_e
```

```
         [,1]     [,2]
[1,] 25.41497 37.20973
```

```
> S <- cov(e); S
```

```
         X1       X2
X1 445.4927 439.7480
X2 439.7480 445.1363
```

```
> D_M <- apply(e, FUN=function(x) sqrt((x-mu_e)%*%solve(S)%*%t(x-mu_e)), MARGIN=1)
> # ou equivalentemente: > sqrt(mahalanobis(e,center = mu_e, cov = S))
>
> D_M
```

```
 [1] 4.237262 3.836750 4.012866 5.291878 4.431027 4.724733 4.347081 5.532128 4.417386 3.187828
[11] 4.323259 2.467348 4.149314 4.486935 4.513259 1.206779 5.026148 4.266663 3.149000 4.448741
[21] 2.726044 2.267494 3.329947 4.164182 3.351026 2.821335 3.834315 4.013288 5.337374 2.969026
[31] 4.223690 3.080665 3.460328 5.016111 3.765846 4.791342 3.263050 6.582452 3.416966 3.346177
[41] 4.078836 2.654091 3.525029 3.623647 3.309112 4.822890 2.743045 2.679526 5.230392 3.308187
```

Realiza-se o teste de Henze-Sirklers para verificar hipótese nula de normalidade multivariada.

```
> # Teste de Henze-Zirklers para normalidade multivariada
> par(cex=0.95)
> test_MN <- MVN::mvn(e,mvnTest = "hz",desc = F, multivariatePlot = "qq")
```

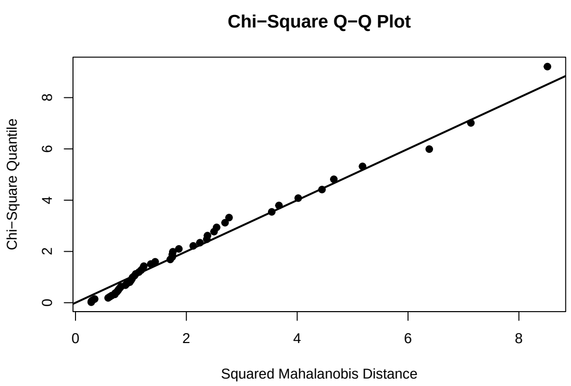

Figura 1: QQ-Plot para verificação da hipótese de normalidade multivariada

```
> # p-valor do teste
> library(ascii)
> print(ascii(test_MN$multivariateNormality), type = 'pandoc')
```

	Test	**HZ**	**p value**	**MVN**
1	Henze-Zirkler	0.53	0.38	YES

Com p-valor $> 0,05$ não rejeita-se hipótese nula e infere-se, juntamente com o gráfico QQ-Plot, que ϵ segue Normalidade Multivariada.

Teste de contrastes

Objetiva-se testar simultaneamente se os coeficientes associados a `Criatividade` nas duas funções de regressão são iguais a zero, ou seja:

$$\text{Tem-se que} \quad B = \begin{array}{c c} & \begin{array}{cc} X1 & X2 \end{array} \\ \begin{array}{c} X4 \\ X7 \end{array} & \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \end{array}$$

Logo, os coeficientes associados a `Criatividade`(X4) são β_{11} e β_{12} referentes às equações da variável resposta X1 e X2, respectivamente. Portanto, o objetivo do teste é analisar a hipótese $\{H_0 : \beta_{11} = \beta_{12} = 0.$

Para tal, considera-se $C_1BM_1 = D$ para o modelo $Y = XB + U$, sob a hipótese de normalidade multivariada.

Tem-se que dimensão$(C1) = (g \times q)$, dimensão$(M1) = (p \times r)$, dimensão$(D) = (g \times r)$, portanto, com $\{H_0 : \beta_{11} = \beta_{12} = 0\} \implies r = 2\text{restrições}, g = 1 \implies D = \begin{bmatrix} 0 & 0 \end{bmatrix}_{(1\times 2)} \implies (C_1)_{(1\times 2)}$ e $(M_1)_{(2\times 2)}$, resultando em:

$$C_1BM_1 = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

Ou seja, $M_1 = I$ e usando o material de apoio, tem-se que a estatística do TRV é dada por:

$$\lambda^{2/n} = \frac{|Y^TPY|}{|Y^TPY + Y_+^TP_2Y_+|}$$

que sob a hipótese nula, tem distribuição de Wilks $\Lambda(p, n-1, g)$. Além disso, $Y_+ = Y - XB_0$, onde B_0 tal que $C_1B_0 = D$ e $P = I - X(X^TX)^{-1}X^T$, com $P_2 = X(X^TX)^{-1}C_1^T[C_1(X^TX)^{-1}C_1^T]^{-1}C_1(X^TX)^{-1}X^T$. Portanto, toma-se então:

$$C_1B_0 = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 4 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix} \implies B_0 = \begin{pmatrix} 0 & 0 \\ 4 & 3 \end{pmatrix}$$

Calcula-se então a estatística do teste:

```
> n <- nrow(Y) # n = número de unidades amostrais
> Yt <- t(Y) # Y transposto
> Xt <- t(X) # X transposto
> C1 <- matrix(c(1,0),ncol=2, nrow=1) # C_1
> C1t <- t(C1)
> B0 <- matrix(c(0,0,4,3), byrow = T, ncol=2, nrow=2)
>
> P <- diag(1,nrow = n) - X%*%solve(Xt%*%X)%*%Xt # I - X(XtX)^{-1}Xt = P = projeção
> P2 <- X%*%solve(Xt%*%X)%*%C1t%*%solve(C1%*%solve(Xt%*%X)%*%C1t)%*%C1%*%solve(Xt%*%X)%*%Xt
> Yp <- Y - X%*%B0 # Y_{+}
> Ypt <- t(Yp)
>
> lambda_obs <- det(Yt%*%P%*%Y)/det(Yt%*%P%*%Y+Ypt%*%P2%*%Yp)# Estatística lambda^{2/n} do TRV
> lambda_obs
```

```
[1] 0.6715128
```

$$\text{Observa-se também que poderia ser considerado } B_0 = \begin{pmatrix} 0 & 0 \\ 32 & 55 \end{pmatrix},$$

e chegaria-se no mesmo resultado da estatística do teste observada $\lambda^{2/n}$, ou seja, não importa qual solução para B_0 tal que $C_1 B_0 = D$ escolha, pois a estatística do teste será a mesma:

```
> # alterando apenas as partes que dependem de B_0 e chega-se no mesmo resultado que anteriormente
> B0 <- matrix(c(0,0,32,55), byrow = T, ncol=2, nrow=2)
> Yp <- Y - X%*%B0 # Y_{+}
> Ypt <- t(Yp)
> lambda_obs <- det(Yt%*%P%*%Y)/det(Yt%*%P%*%Y+Ypt%*%P2%*%Yp)# Estatística lambda^{2/n} do TRV
> lambda_obs
```

```
[1] 0.6715128
```

Tem-se então que $\lambda^{2/n} \overset{H_0}{\sim} \Lambda(p, n-1, g)$, ou seja, com $p = 2, n = 50, g = 1 \implies \lambda^{2/n} \overset{H_0}{\sim} \Lambda(2, 49, 1)$. Pode-se usar a relação com a distribuição F, em que:

$$\frac{1 - \Lambda(p, m, 1)}{\Lambda(p, m, 1)} \sim \frac{p}{m - p + 1} F_{p, m-p+1}$$

Portanto,

$$\frac{1 - \Lambda(2, 49, 1)}{\Lambda(2, 49, 1)} \sim \frac{2}{49 - 2 + 1} F_{2,49-2+1} = \frac{2}{48} F_{2,48}$$

$$\implies \frac{48}{2} \frac{1 - \lambda^{2/n}}{\lambda^{2/n}} \overset{H_0}{\sim} F_{2,48}$$

Usando a estatística do teste observada, tem-se $\lambda_{obs}^{2/n} = 0.6715128$, então:

```
> transf_lambda_obs <- 24*(1-lambda_obs)/lambda_obs
> pvalor <- 1-pf(transf_lambda_obs, df1=2, df2=48) # quantil da estatística transformada observada
> pvalor
```

```
[1] 7.068114e-05
```

Portanto, encontrou-se um p-valor $< 0,05$ então rejeita-se hipótese nula dos coeficientes serem iguais a zero. Ou seja, rejeita-se $\{H_0 : \beta_{11} = \beta_{12} = 0$ a nível de 5% de significância.

Código

```
library(tidyverse)
library(dplyr)
library(readxl)
vendedores <- read_excel("vendedores.xlsx")

# Obtenção atráves do método multivariado
Y = as.matrix(vendedores[c("X1","X2")])
X = as.matrix(vendedores[c("X4","X7")])

B_hat = solve(t(X)%*%X)%*%t(X)%*%Y
colnames(B_hat) <- NULL; rownames(B_hat) <- NULL
print(B_hat)

# Obtenção atráves do método univariado
y = c(Y)
x = kronecker(diag(1,nrow=2), X)

b_hat = solve(t(x)%*%x)%*%t(x)%*%y
print(b_hat)

e <- Y - X%*%B_hat
mu_e <- t(c(mean(e[1]), mean(e[2]))); mu_e
S <- cov(e); S

D_M <- apply(e, FUN=function(x) sqrt((x-mu_e)%*%solve(S)%*%t(x-mu_e)), MARGIN=1)
# ou equivalentemente: > sqrt(mahalanobis(e,center = mu_e, cov = S))
D_M

# Teste de Henze-Zirklers para normalidade multivariada
par(cex=0.95)
test_MN <- MVN::mvn(e,mvnTest = "hz",desc = F, multivariatePlot = "qq")

# p-valor do teste
test_MN$multivariateNormality

n <- nrow(Y) # n = número de unidades amostrais
Yt <- t(Y) # Y transposto
Xt <- t(X) # X transposto
C1 <- matrix(c(1,0),ncol=2, nrow=1) # C_1
C1t <- t(C1)
B0 <- matrix(c(0,0,4,3), byrow = T, ncol=2, nrow=2)

P <- diag(1,nrow = n) - X%*%solve(Xt%*%X)%*%Xt # I - X(XtX)^{-1}Xt = P = projeção
P2 <- X%*%solve(Xt%*%X)%*%C1t%*%solve(C1%*%solve(Xt%*%X)%*%C1t)%*%C1%*%solve(Xt%*%X)%*%Xt
Yp <- Y - X%*%B0 # Y_{+}
Ypt <- t(Yp)

lambda_obs <- det(Yt%*%P%*%Y)/det(Yt%*%P%*%Y+Ypt%*%P2%*%Yp)# Estatística lambda^{2/n} do TRV
lambda_obs

# alterando apenas as partes que dependem de B_0 e chega-se no mesmo resultado que anteriormente
B0 <- matrix(c(0,0,32,55), byrow = T, ncol=2, nrow=2)
Yp <- Y - X%*%B0 # Y_{+}
Ypt <- t(Yp)
lambda_obs <- det(Yt%*%P%*%Y)/det(Yt%*%P%*%Y+Ypt%*%P2%*%Yp)# Estatística lambda^{2/n} do TRV
lambda_obs

transf_lambda_obs <- 24*(1-lambda_obs)/lambda_obs
pvalor <- 1-pf(transf_lambda_obs, df1=2, df2=48) # quantil da estatística transformada observada
pvalor
```

Bruno Groper Morbin - 11809875 | 19/12/22

MAE330 - Análise Multivariada de Dados

Lista 8

1.) $\underset{(3\times2)}{Y} = \underset{(3\times2)(2\times2)}{XB} + \overset{(3\times2)}{E}$, com $Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \end{pmatrix}$, $X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix}$, $B = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}$

e $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$, onde $\sigma_{11} \equiv$ variabilidade na coluna 1 de Y

$\sigma_{21} \equiv$ covariância entre colunas 1 e 2 de Y

a) $vec(Y) = y = \begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \\ y_{32} \end{pmatrix}_{(6\times1)}$, $vec(B) = \beta = \begin{pmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{12} \\ \beta_{22} \end{pmatrix}_{(4\times1)}$

$vec(E) = \underset{6\times1}{\varepsilon}$ e $X^* = I_2 \otimes X = \begin{bmatrix} X_{3\times2} & 0_{3\times2} \\ 0_{3\times2} & X_{3\times2} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & 0 & 0 \\ x_{21} & x_{22} & 0 & 0 \\ x_{31} & x_{32} & 0 & 0 \\ 0 & 0 & x_{11} & x_{12} \\ 0 & 0 & x_{21} & x_{22} \\ 0 & 0 & x_{31} & x_{32} \end{bmatrix}_{(6\times4)}$

$$\Omega = Var(\varepsilon) = \Sigma \otimes I_3 = \begin{array}{c} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \\ y_{32} \end{array} \begin{bmatrix} \sigma_{11} & 0 & 0 & \sigma_{12} & 0 & 0 \\ 0 & \sigma_{11} & 0 & 0 & \sigma_{12} & 0 \\ 0 & 0 & \sigma_{11} & 0 & 0 & \sigma_{12} \\ \sigma_{21} & 0 & 0 & \sigma_{22} & 0 & 0 \\ 0 & \sigma_{21} & 0 & 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{21} & 0 & 0 & \sigma_{22} \end{bmatrix}_{6\times6}$$

, nota-se que indivíduos da amostra (linhas de Y) diferentes são não-correlacionados.

b) $\mathbb{E}(\varepsilon) = \mathbb{E}(vec(E)) = vec(\mathbb{E}(E)) = vec(\underline{0})_{6\times1}$, pois $\mathbb{E}(E) = \underline{0}$

$Var(\varepsilon) = Var(vec(\varepsilon)) = \Omega = \Sigma \otimes I_3$

Nota-se: $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \otimes B = \begin{bmatrix} aB & bB \\ cB & dB \end{bmatrix}$

$$\underset{4\times1}{\beta} = \underbrace{(\underset{4\times6}{X^{*T}} \underset{6\times6}{\Omega^{-1}} \underset{6\times4}{X^*})^{-1}}_{4\times4} \underset{4\times6}{X^{*T}} \underset{6\times6}{\Omega^{-1}} \underset{6\times1}{y} \quad \text{I}$$

II

Usa-se então a propriedade do produto de Kronecker em que $(A\otimes B)(C\otimes D) = AC \otimes BD$

e além disso, $(A\otimes B)^T = A^T \otimes B^T$ bem como $(A\otimes B)^{-1} = A^{-1} \otimes B^{-1}$

$\therefore$ de I:

$$\beta = \left[(I_2 \otimes X)^T (\Sigma \otimes I_3)^{-1} (I_2 \otimes X) \right]^{-1} (I_2 \otimes X)^T (\Sigma \otimes I_3)^{-1} vec(Y)$$

$$= \left[(I_2 \otimes X^T)(\Sigma^{-1} \otimes I_3)(I_2 \otimes X) \right]^{-1} (I_2 \otimes X^T) \cdot (\Sigma^{-1} \otimes I_3)\, vec(Y)$$

$\begin{cases} I^T = I \\ e\ I^{-1} = I \end{cases}$

Usando (II):

$$\Rightarrow \beta = [\underbrace{(I_2\Sigma^{-1}\otimes X^T I_3)}_{=(\Sigma^{-1}\otimes X^T)}(I_2\otimes X)]^{-1}(I_2\Sigma^{-1}\otimes X^T I_3)\,vec(Y)$$

$$= (\Sigma^{-1}I_2\otimes X^TX)^{-1}(\Sigma^{-1}\otimes X^T)\,vec(Y)$$

$$= (\Sigma\otimes(X^TX)^{-1})(\Sigma^{-1}\otimes X^T)\,vec(Y)$$

$$= (\Sigma\cdot\Sigma^{-1}\otimes(X^TX)^{-1}\cdot X^T)\,vec(Y) = (I_2\otimes(X^TX)^{-1}X^T)\,vec(Y)$$

Logo, $\boxed{\beta = [I_2\otimes((X^TX)^{-1}X^T)]\,vec(Y)}$ (III), ou seja, não depende de Ω e $\therefore$ Σ.

c) Primeiramente prova-se o seguinte Teorema: $vec(AXB) = (B^T\otimes A)\,vec(X)$

Prova: Toma-se $B = [b_1\ b_2 \dots b_m]_{(m\times m)}$, $X = [x_1\ x_2 \dots x_m]_{(p\times m)}$, $A = [a_1 \dots a_p]$

Então a k-ésima coluna de AXB é:

$$(AXB)_{:,k} = AXb_k = A\sum_{i=1}^{m} x_i\, b_{(i,k)} = [b_{(1,k)}A\ \ b_{(2,k)}A \dots b_{(m,k)}A]\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_m\end{bmatrix}$$

$$= [(b_{1,k}\ b_{2,k}\dots b_{m,k})\otimes A]\,vec(X)$$

e

$$vec(AXB) = \begin{bmatrix}(AXB)_{:,1}\\ (AXB)_{:,2}\\ \vdots\\ (AXB)_{:,m}\end{bmatrix} = \begin{bmatrix} b_1^T\otimes A\\ b_2^T\otimes A\\ \vdots\\ b_m^T\otimes A\end{bmatrix}\cdot vec(X) = (B^T\otimes A)\cdot vec(X) \quad \blacksquare$$

Usando tal teorema em (III):

$$\beta = vec(B) = [I_2\otimes((X^TX)^{-1}X^T)]\cdot vec(Y) \overset{I_2^T = I_2}{=} vec(\underbrace{[(X^TX)^{-1}X^T]}_{\equiv A \text{ do Teorema}}\cdot \overset{\equiv X \text{ do Teorema}}{Y}\cdot \underbrace{I_2}_{\equiv B \text{ do Teorema}})$$

$\therefore vec(B) = vec((X^TX)^{-1}X^TY)$, pois $\underset{(3\times2)(2\times2)}{Y\cdot I_2} = \underset{3\times2}{Y}$

e então

$$\boxed{\underset{2\times2}{B} = \underbrace{(\underset{2\times3}{X^T}\underset{3\times2}{X})^{-1}}_{2\times2}\underset{2\times3}{X^T}\underset{3\times2}{Y}}$$