

Regressão Multivariada no R

Seja um modelo de regressão multivariada do tipo

$$Y = Z\beta + \epsilon$$

em que

- Y é uma matriz de variáveis respostas de dimensão $n \times m$,
- Z é uma matriz de variáveis explicativas de dimensão $n \times (p + 1)$, com uma coluna de 1's relativa ao intercepto do modelo,
- β é uma matriz de parâmetros de dimensão $(p + 1) \times m$,
- ϵ é um matriz de erros aleatórios de dimensão $n \times m$.

Exercício 7.9 de Johnson e Wichern (2007), p. 422.

```
Z <- matrix(c(1,1,1,1,1,-2,-1,0,1,2), ncol=2, byrow=FALSE)
Y <- matrix(c(5,3,4,2,1,-3,-1,-1,2,3), ncol=2, byrow=FALSE)
beta.chapeu <- solve(t(Z) %*% Z) %*% t(Z) %*% Y
print(round(beta.chapeu,2))
```

```
##      [,1] [,2]
## [1,]  3.0  0.0
## [2,] -0.9  1.5
```

```
Y.chapeu <- Z %*% beta.chapeu
print(Y.chapeu)
```

```
##      [,1]      [,2]
## [1,]  4.8 -3.00e+00
## [2,]  3.9 -1.50e+00
## [3,]  3.0  1.11e-16
## [4,]  2.1  1.50e+00
## [5,]  1.2  3.00e+00
```

```
residuo <- Y - Y.chapeu
print(residuo)
```

```
##      [,1]      [,2]
## [1,]  0.2  4.441e-16
## [2,] -0.9  5.000e-01
## [3,]  1.0 -1.000e+00
## [4,] -0.1  5.000e-01
## [5,] -0.2 -4.441e-16
```

Forma alternativa utilizando a função `lm` do R. Note que a matriz Z já possui a coluna referente ao intercepto, por isso o termo `-1` dentro da função `lm`. O R por default já inclui intercepto no modelo linear.

```
Z <- matrix(c(1,1,1,1,1,-2,-1,0,1,2), ncol=2, byrow=FALSE)
Y <- matrix(c(5,3,4,2,1,-3,-1,-1,2,3), ncol=2, byrow=FALSE)
modelo<-lm(Y ~ -1 + Z)
summary(modelo)
```

```
## Response Y1 :
##
## Call:
## lm(formula = Y1 ~ -1 + Z)
##
## Residuals:
##      1      2      3      4      5
##  0.2 -0.9  1.0 -0.1 -0.2
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## Z1      3.000      0.356   8.43  0.0035 **
## Z2     -0.900      0.252  -3.58  0.0374 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.796 on 3 degrees of freedom
## Multiple R-squared:  0.965, Adjusted R-squared:  0.942
## F-statistic: 41.9 on 2 and 3 DF,  p-value: 0.00642
##
##
## Response Y2 :
##
## Call:
## lm(formula = Y2 ~ -1 + Z)
##
## Residuals:
##      1      2      3      4      5
## -5.10e-16  5.00e-01 -1.00e+00  5.00e-01  1.76e-16
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## Z1 1.99e-16  3.16e-01   0.00  1.0000
## Z2 1.50e+00  2.24e-01   6.71  0.0068 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.707 on 3 degrees of freedom
## Multiple R-squared:  0.938, Adjusted R-squared:  0.896
## F-statistic: 22.5 on 2 and 3 DF,  p-value: 0.0156
```

Exercício

Obtenha os códigos para a MANOVA, análise de variância multivariada.

Referência

Johnson, R. A. and Wichern, D. W. (2007) Applied Multivariate Statistical Analysis. 5th edition. Prentice-Hall.