Lista 3 - MAE0328

Guilherme NºUSP: 8943160 e Leonardo NºUSP: 9793436

Exercício 2

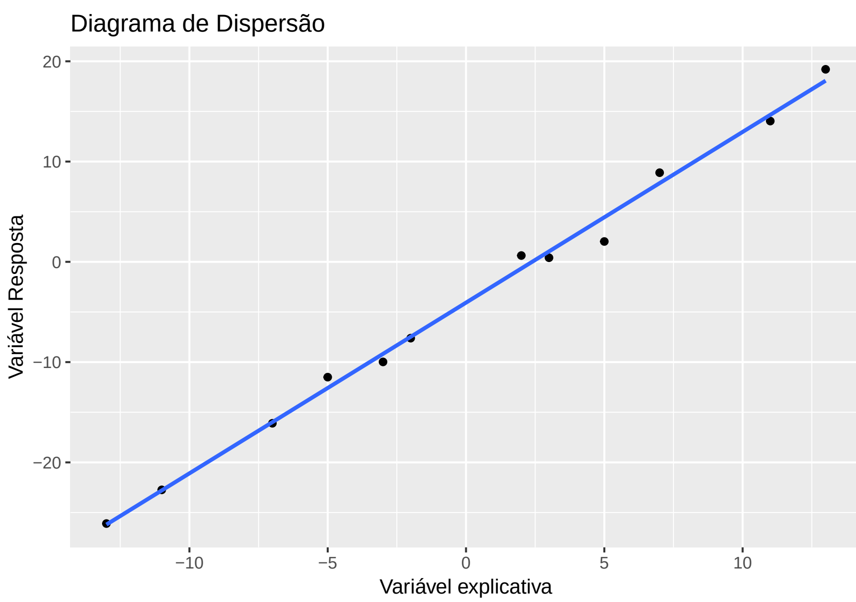
Considere o conjunto de dados fictícios da Tabela 1.

Tabela 1: Dados fictícios para análise de regressão

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	-13	-11	-7	-5	-3	-2	2	3	5	7	11	13
Y_i	-26.10	-22.74	-16.10	-11.50	-9.98	-7.61	0.63	0.41	2.03	8.89	14.04	19.20

(a) Assuma o modelo $Y = \beta_0 + \beta_1 + e_i$, em que $e_i \overset{iid}{\sim} N(0, \sigma^2),\ i \in \{1, ..., n\}$ Apresente as estimativas de máxima verossimilhança para β e σ^2. Faça o gráfico de dispersão com a curva de regressão estimada.

Resolução



Como foi visto em aula, quando $e_i \overset{iid}{\sim} N(0, \sigma^2),\ i \in \{1, ..., n\}$ os estimadores de mínomos quadrados são iguais aos de máxima verossimilhnaça, ou seja, $\widehat{\beta} = \begin{pmatrix} \widehat{\beta_0}^{MQ} \\ \widehat{\beta_1}^{MQ} \end{pmatrix} = \begin{pmatrix} \widehat{\beta_0}^{MV} \\ \widehat{\beta_1}^{MV} \end{pmatrix}$ assim, temos que:

$$\widehat{\beta}_{MV} = (X^T X)^{-1} X^T Y \Rightarrow \widehat{\beta}_{MV} = \begin{pmatrix} -4.069 \\ 1.7025 \end{pmatrix}$$

e agora o estimador de máxima verossimilhança para σ^2 que é

$$\widehat{\sigma}^2_{MV} = \frac{1}{n}(Y - X\widehat{\beta})^T (Y - X\widehat{\beta}) = \frac{12.5134}{12} = 1.0442$$

Porém sabemos que o estimador de máxima verossimilhança para σ^2 é viciado então uma solução para isso seria a estimativa $\widehat{\sigma}^2 = \frac{n}{n-2}\widehat{\sigma}^2_{MV} \Rightarrow \widehat{\sigma}^2 = \frac{12}{10}1.0442 = 1.2513$

(b) Apresente as estatísticas para testar as seguintes hipóteses:

$$(I) = \begin{cases} H_{00} : \beta_0 = -5 \\ H_{10} : \beta_0 \neq -5 \end{cases}$$

e

$$(II) = \begin{cases} H_{01} : \beta_1 = 0 \\ H_{11} : \beta_1 \neq 0 \end{cases}$$

Calcule os respectivos valores-p e teste cada uma das hipóteses considerando um nível de significância de 5%.

Resolução

Para realizar o teste de hipótese *(I)* a estatística a ser utilizada é a mesma do exercício 1 item d, que é:

$$F_{\widehat{\beta},C,c} = \frac{(C\widehat{\beta} - c)^T [C(X^T X)^{-1} C^T]^{-1} (C\widehat{\beta} - c)}{k\widehat{\sigma}^2}$$

onde $C = \begin{pmatrix} 1 & 0 \end{pmatrix}$, $c = -5$, $k = 1$, $\widehat{\beta} = \begin{pmatrix} -4.069 \\ 1.7025 \end{pmatrix}$ e $\widehat{\sigma}^2 = 1.2513$ (estimativa não viciada para $\widehat{\sigma}^2$) Assim, temos que:

$$F^{obs}_{\widehat{\beta},C,c} = 8.309$$

e seu valor-p= $\mathbb{P}(F_{k,n-2} > F^{obs}_{\widehat{\beta},C,c}) = \mathbb{P}(F_{1,10} > F^{obs}_{\widehat{\beta},C,c}) = 0.016$. Logo temos evidencias estatísticas para rejeitar a hipótese nula a um nível de significância de 5%.

Para realizar o teste de hipótese *(II)* a estatística a ser utilizada é a mesma do exercício anterior, onde: $C = \begin{pmatrix} 0 & 1 \end{pmatrix}$, $c = 0$, $k = 1$, $\widehat{\beta} = \begin{pmatrix} -4.069 \\ 1.7025 \end{pmatrix}$ e $\widehat{\sigma}^2 = 1.2513$ (estimativa não viciada para $\widehat{\sigma}^2$) Assim, temos que:

$$F^{obs}_{\widehat{\beta},C,c} = 1746.56$$

e seu valor-p= $\mathbb{P}(F_{k,n-2} > F^{obs}_{\widehat{\beta},C,c}) = \mathbb{P}(F_{1,10} > F^{obs}_{\widehat{\beta},C,c}) = 1.47 * 10^{-12} < 0.0001$. Logo temos evidencias estatísticas para rejeitar a hipótese nula a um nível de significância de 5%.

(c) Encontre os intervalos de confiança para β_0, β_1 e σ^2 utilizando um coeficiente de confiança de 90%.

Resolução

Como visto em aula, temos que os intervalos de confiança para β_0, β_1 e σ^2 com $\alpha = 90\%$ são:

IC(β_0, 0.9)=$\widehat{\beta}_0 \pm t_{\frac{\alpha}{2},n-2}\sqrt{\widehat{\sigma}^2_{\widehat{\beta}_0}}$ com $\widehat{\beta}_0 = -4.069$, $t_{\frac{\alpha}{2},n-2} = t_{0.95,10} = 1.812$ e $\sqrt{\widehat{\sigma}^2_{\widehat{\beta}_0}} = 0.323$. Assim:

IC(β_0, 0.9)=$[-4.069 - 1.812 * 0.323, -4.069 + 1.812 * 0.323] = [-4.654, -3.484]$

Agora,

IC(β_1, 0.9)=$\widehat{\beta}_1 \pm t_{\frac{\alpha}{2},n-2}\sqrt{\widehat{\sigma}^2_{\widehat{\beta}_1}}$ com $\widehat{\beta}_0 = 1.7025$, $t_{\frac{1+\alpha}{2},n-2} = t_{0.95,10} = 1.812$ e $\sqrt{\widehat{\sigma}^2_{\widehat{\beta}_0}} = 0.0407$. Assim:

IC(β_0, 0.9)=$[1.7025 - 1.812 * 0.0407, 1.7025 - 1.812 * 0.0407] = [1.628, 1.776]$

Agora,

IC(σ^2, 0.9)=$(\frac{\widehat{\sigma}^2(n-2)}{\chi^2_{\frac{1+\alpha}{2},n-2}}, \frac{\widehat{\sigma}^2(n-2)}{\chi^2_{\frac{1-\alpha}{2},n-2}}) = (\frac{1.251*10}{18.307}, \frac{1.251*10}{3.9402}) = (0.6835, 3.1758)$

(d) Encontre C e c da hipótese linear geral $H_0 : C\beta = c$ contra $H_1 : C\beta \neq c$ e faça o teste considerando 5% de significância estatística nos seguintes casos:

$$(I) = \begin{cases} H_0 : \frac{1}{2}\beta_0 = -\frac{5}{4}\beta_1 \\ H_1 : \frac{1}{2}\beta_0 \neq -\frac{5}{4}\beta_1 \end{cases}$$

$$(II) = \begin{cases} H_0 : \beta_0 = \beta_1 = 0 \\ H_1 : \beta_0 = \beta_1 \neq 0 \end{cases}$$

$$(III) = \begin{cases} H_0 : \frac{\beta_0}{\beta_1} = 3 \\ H_1 : \frac{\beta_0}{\beta_1} \neq 3 \end{cases}$$

$$(IV) = \begin{cases} H_0 : \begin{pmatrix} \beta_0 + \frac{5}{2}\beta_1 \\ \beta_1 - \beta_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix} \\ H_1 : \begin{pmatrix} \beta_0 + \frac{5}{2}\beta_1 \\ \beta_1 - \beta_0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 6 \end{pmatrix} \end{cases}$$

Resolução

Para realizar o teste de hipótese *(I)* a estatística a ser utilizada é a mesma do exercício 1 item d, que é:

$$F_{\widehat{\beta},C,c} = \frac{(C\widehat{\beta} - c)^T [C(X^T X)^{-1} C^T]^{-1} (C\widehat{\beta} - c)}{k\widehat{\sigma}^2}$$

onde $C = \begin{pmatrix} \frac{1}{2} & \frac{5}{4} \end{pmatrix}$, $c = 0$, $k = 1$, $\widehat{\beta} = \begin{pmatrix} -4.069 \\ 1.7025 \end{pmatrix}$ e $\widehat{\sigma}^2 = 1.2513$ (estimativa não viciada para $\widehat{\sigma}^2$) Assim, temos que:

$$F^{obs}_{\widehat{\beta},C,c} = 0.3055$$

e seu valor-p= $\mathbb{P}(F_{k,n-2} > F^{obs}_{\widehat{\beta},C,c}) = \mathbb{P}(F_{1,10} > F^{obs}_{\widehat{\beta},C,c}) = 0.592$. Logo temos evidencias estatísticas para NÃO rejeitar a hipótese nula a um nível de significância de 5%.

Para realizar o teste de hipótese *(II)* a estatística a ser utilizada é a mesma do exercício anterior onde $C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $c = \begin{pmatrix} 0 & 0 \end{pmatrix}$, $k = 2$, $\widehat{\beta} = \begin{pmatrix} -4.069 \\ 1.7025 \end{pmatrix}$ e $\widehat{\sigma}^2 = 1.2513$ (estimativa não viciada para $\widehat{\sigma}^2$) Assim, temos que:

$$F^{obs}_{\widehat{\beta},C,c} = 952.677$$

e seu valor-p= $\mathbb{P}(F_{k,n-2} > F^{obs}_{\widehat{\beta},C,c}) = \mathbb{P}(F_{1,10} > F^{obs}_{\widehat{\beta},C,c}) = 8.879 * 10^{-12} < 0.0001$. Logo temos evidencias estatísticas para rejeitar a hipótese nula a um nível de significância de 5%.

Para realizar o teste de hipótese *(III)* a estatística a ser utilizada é a mesma do exercício anterior onde $C = \begin{pmatrix} 1 & -3 \end{pmatrix}$, $c = 0$, $k = 1$, $\widehat{\beta} = \begin{pmatrix} -4.069 \\ 1.7025 \end{pmatrix}$ e $\widehat{\sigma}^2 = 1.2513$ (estimativa não viciada para $\widehat{\sigma}^2$) Assim, temos que:

$$F^{obs}_{\widehat{\beta},C,c} = 706.395$$

e seu valor-p= $\mathbb{P}(F_{k,n-2} > F^{obs}_{\widehat{\beta},C,c}) = \mathbb{P}(F_{1,10} > F^{obs}_{\widehat{\beta},C,c}) = 1.311 * 10^{-10} < 0.0001$. Logo temos evidencias estatísticas para rejeitar a hipótese nula a um nível de significância de 5%.

Para realizar o teste de hipótese *(IV)* a estatística a ser utilizada é a mesma do exercício anterior onde $C = \begin{pmatrix} 1 & \frac{5}{2} \\ -1 & 1 \end{pmatrix}$, $c = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$, $k = 2$, $\widehat{\beta} = \begin{pmatrix} -4.069 \\ 1.7025 \end{pmatrix}$ e $\widehat{\sigma}^2 = 1.2513$ (estimativa não viciada para $\widehat{\sigma}^2$) Assim, temos que:

$$F^{obs}_{\widehat{\beta},C,c} = 0.266$$

e seu valor-p= $\mathbb{P}(F_{k,n-2} > F^{obs}_{\widehat{\beta},C,c}) = \mathbb{P}(F_{1,10} > F^{obs}_{\widehat{\beta},C,c}) = 0.771$. Logo temos evidencias estatísticas para NÃO rejeitar a hipótese nula a um nível de significância de 5%.

Exercício 3

Considere $x_1, ..., x_{12}$ da Tabela 1 fixos e simule $N = 1000$ conjuntos de dados de tamanho $n = 12$ segundo a equação $Y_i = -4 + \frac{8}{5}x_i + e_i$, com $e_i \overset{iid}{\sim} N(0, 2)$ Para cada conjunto de dados simulado, estime β_0, β_1 e σ^2, calcule seus respectivos intervalos de confiança considerando um coeficiente de confiança de 97,5% e verifique se estes contém os verdadeiros valores dos parâmetros. Apresente a proporção dos intervalos que você obteve que contêm os verdadeiros valores e o algoritmo utilizado.

Resolução

```
N <- 1000 # Número de simulações
n <- length(Y)  # tamanho da amostra
V <- solve(t(X)%*%X) # (X'X)-¹
D <- V%*%t(X) # (X'X)-¹*X'
gamma <- 0.975 # Confiança do intervalo

c <- qt((1+gamma)/2,10) # Calculo do quartil da t-student com gl=10

g1 <- qchisq((1-gamma)/2,10) # Calculo do quartil da X² com gl=10
g2 <- qchisq(((1+gamma)/2),10) # Calculo do quartil da X² com gl=10

# Matrizes para armazenar os IC's
b0 <- matrix(NA,N,2)
b1 <- matrix(NA,N,2)
s2 <- matrix(NA,N,2)

for (i in 1:N){
  erro <- rnorm(n,0,sqrt(2)) # erro simulado com distribuição N(0,2) e n=12
  yy <- cbind(-4 + (8/5)*x + erro) # vetor coluna de Y simulados dado que os o vetor x está fixo
  beta.S <- D%*%yy # vetor coluna de simuações de beta0 e beta1
  sigma2.hatS <- as.numeric(t(yy)%*%(diag(n) - X%*%D)%*%yy/(n-2)) # vetor de simuações de sigma²
  Matrix.cov <- sigma2.hatS*V # Matriz de covariâncias
  b0[i,] <- c(beta.S[1] -c*sqrt(Matrix.cov[1,1]), beta.S[1]+c*sqrt(Matrix.cov[1,1])) # IC p/ B0
  b1[i,] <- c(beta.S[2] -c*sqrt(Matrix.cov[2,2]), beta.S[2]+c*sqrt(Matrix.cov[2,2])) # IC p/ B1
  s2[i,] <- c(sigma2.hatS*(n-2)/g2,sigma2.hatS*(n-2)/g1) # IC p/ sigma²
}

# Verifica se o verdadeiro valor está nos intervalos
aa <- ifelse(b0[,1]<= -4 & b0[,2] >= -4 ,1,0)
bb <- ifelse(b1[,1]<= 8/5 & b1[,2] >= 8/5 ,1,0)
ss <- ifelse(s2[,1]<= 2 & s2[,2] >= 2 ,1,0)

# Junta a variaável binária com a matriz de IC's
b0 <- cbind(b0,aa)
b1 <- cbind(b1,bb)
ss <- cbind(s2,ss)

# Proporção dos IC's que tem o verdadeiro valor de B0=-4
mean(b0[,3])
```

```
## [1] 0.968
```

```r
# Proporção dos IC's que tem o verdadeiro valor de B1=8/5
mean(b1[,3])
```

```
## [1] 0.971
```

```r
# Proporção dos IC's que tem o verdadeiro valor de sigma^2=2
mean(ss[,3])
```

```
## [1] 0.983
```