

Exercício 4

Um experimento foi conduzido para avaliar o número de colônias de 2 espécies de bactérias, A e B, presentes após 2 dias em meios de cultura com 4 níveis distintos de um antibiótico, a saber 0 μL , 2.5 μL , 6 μL e 10 μL . Os dados encontram-se na Tabela 1, com Y , a variável resposta, sendo o logaritmo do número de colônias.

Tabela 1: Dados do experimento de cultivo de bactérias da espécie A (esquerda) e B (direita).

ln (Nº de colônias)	Volume de Antibiótico	ln (Nº de colônias)	Volume de Antibiótico
17.93	0	15.69	0
18.82	0	17.04	0
15.39	0	16.21	0
14.21	2	16.96	2
14.16	2	14.83	2
14.87	2	15.07	2
9.67	6	15.33	6
13.41	6	14.46	6
11.06	6	12.70	6
8.55	10	13.11	10
5.12	10	11.23	10
7.91	10	11.32	10

- (a) Formalize um modelo de regressão linear múltiplo no qual ambas as espécies apresentam o mesmo intercepto, porém coeficientes angulares distintos. Apresente as suposições desse modelo e suas estimativas.
Dica: Use um modelo similar ao do exercício 3.

Resolução

O modelo a ser utilizado é o $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$, $i \in \{1, \dots, 24\}$ com $e_i \stackrel{iid}{\sim} N(0, \sigma^2)$, $\sum_{i=1}^{2n} x_{1i} = \sum_{i=1}^{2n} x_{2i}$, $\sum_{i=1}^n x_{2i} = \sum_{i=n+1}^{2n} x_{1i} = 0$ $i \in \{1, \dots, 2n\}$ com $n = 12$. Pelo Método dos mínimos quadrados temos que $\hat{\beta}_{MQ} = \begin{pmatrix} 16.726 \\ -0.943 \\ -0.471 \end{pmatrix}$

Assim sua reta estimada é

$$\hat{\mu}_i = 16.726 - 0.943x_{1i} - 0.471x_{2i}$$

- (b) Apresente a matriz de variâncias e covariâncias estimadas dos parâmetros da regressão. Usando-a, obtenha intervalos de confiança marginais de coeficiente $\gamma = 95\%$ para os parâmetros em questão.

Resolução

A matriz de Covariâncias estimada é:

$$\hat{\sigma}^2(X^T X)^{-1} = 1.567 \begin{pmatrix} 0.0988 & -0.0127 & -0.0127 \\ -0.0127 & 0.004 & 0.0016 \\ -0.0127 & 0.0016 & 0.004 \end{pmatrix} = \begin{pmatrix} 0.1549 & -0.0199 & -0.0199 \\ -0.0199 & 0.0063 & 0.0025 \\ -0.0199 & 0.0025 & 0.0063 \end{pmatrix}$$

Sabemos que o intervalo de confiança marginal para β_i com $i \in \{0, \dots, 2\}$ com coeficiente $\gamma = 95\%$ é:

$$IC(\beta_i, \gamma) = \hat{\beta}_i \pm t_{(\frac{1+\gamma}{2}, n-p-1)} \hat{\sigma}_{\hat{\beta}_i}$$

em que $T_{\hat{\beta}_i} = \frac{\beta_i MV - \beta_i}{\hat{\beta}_i} \sim t_{n-p-1}$ para $i = 0, \dots, p$ e $\hat{\sigma}_{\beta_i}^2 = \hat{\sigma}^2 C_i (X^T X)^{-1} C_i^T$ e $\hat{\sigma}^2 = \frac{1}{n-p-1} (Y - X\hat{\beta})^T (Y - X\hat{\beta})$

Para β_0 temos: $C_0 = (1, 0, 0)$ $\hat{\sigma}_{\beta_0}^2 = 0.1549 \Rightarrow \hat{\sigma}_{\beta_0} = 0.393$

$IC(\beta_0, 0.95) = \hat{\beta}_0 \pm t_{(0.975, 21)} \hat{\sigma}_{\beta_0} = 16.726 \pm 2.08 * 0.393 = 16.726 \pm 0.817$

Para β_1 temos: $C_0 = (0, 1, 0)$ $\hat{\sigma}_{\beta_1}^2 = 0.0063 \Rightarrow \hat{\sigma}_{\beta_1} = 0.0793$

$IC(\beta_1, 0.95) = \hat{\beta}_1 \pm t_{(0.975, 21)} \hat{\sigma}_{\beta_1} = -0.943 \pm 2.08 * 0.0793 = -0.943 \pm 0.1649$

Para β_2 temos: $C_0 = (0, 0, 1)$ $\hat{\sigma}_{\beta_2}^2 = 0.0063 \Rightarrow \hat{\sigma}_{\beta_2} = 0.0793$

$IC(\beta_2, 0.95) = \hat{\beta}_2 \pm t_{(0.975, 21)} \hat{\sigma}_{\beta_2} = -0.471 \pm 2.08 * 0.0793 = -0.471 \pm 0.1649$

- (c) Os pesquisadores postulam a hipótese de que “as espécies A e B *respondem de forma semelhante aos incrementos adotados do antibiótico*”. Traduza esta hipótese na linguagem da estatística, apresente o valor- p obtido e teste-a ao nível de 5% de significância.

Resolução

A hipótese a ser testada em linguagem estatística é:

$$\begin{cases} H_0 : \beta_1 = \beta_2 \\ H_1 : \beta_1 \neq \beta_2 \end{cases}$$

Escrevendo o teste acima na forma linear geral, temos:

$$\begin{cases} H_0 : C\beta = d \\ H_1 : C\beta_1 \neq d \end{cases}$$

Onde $C = (0, 1, -1)$ e $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$ e $d = 0$

Logo calculando a estatística de teste, temos:

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

Assim Sob H_0 , $F_{\hat{\beta}, C, d} \sim F_{k, n-p-1}$:

$$F_{\hat{\beta}, C, d} = \frac{(C\hat{\beta})^T [C(X^T X)^{-1} C^T]^{-1} (C\hat{\beta})}{k \hat{\sigma}^2} = 28.375$$

Assim seu P-valor = $\mathbb{P}(F_{1, 21} > F_{\hat{\beta}, C, d}^{Obs}) = \mathbb{P}(F_{1, 21} > 28.375) = 0.00002$

Logo temos evidências estatísticas para rejeitar H_0 , logo os valores de $\beta_1 \neq \beta_2$

- (d) Para uma nova observação x_0 da espécie A, com 10 μL antibiótico, mostre como obter e apresente os intervalos de confiança (e de predição, respectivamente) de coeficiente $\gamma = 95\%$ para

- i. $\mu(x_0) = x_0^T \beta$
- ii. $Y_0 = x_0^T \beta + e_0$

Resolução

Para i. $\mu(x_0) = x_0^T \beta$, Sabemos que

$$\hat{\mu}_0 \sim N(\mu_0, \sigma^2 x_0^T (X^T X)^{-1} x_0)$$

Assim temos que

$$\frac{\hat{\mu}_0 - \mu}{\sqrt{\sigma^2 x_0^T (X^T X)^{-1} x_0}} \sim t_{(n-p-1)}$$

e $\hat{\sigma}_{\mu_0}^2 = \hat{\sigma}^2 x_0^T (X^T X)^{-1} x_0$ Logo

$$IC(\mu_0, \gamma) = [\hat{\mu}_0 - t_{(\frac{1+\gamma}{2}, n-p-1)} \hat{\sigma}_{\mu_0}, \hat{\mu}_0 + t_{(\frac{1+\gamma}{2}, n-p-1)} \hat{\sigma}_{\mu_0}]$$

Utilizando nossos dados, logo $x_0^T = (1, 10, 0)$ e $\hat{\beta} = \begin{pmatrix} 16.726 \\ -0.943 \\ -0.471 \end{pmatrix} \Rightarrow \hat{\mu}_0 = x_0^T \hat{\beta} = 7.29$,

$$\hat{\sigma}_{\mu_0} = \sqrt{1.647 * 0.246} = 0.201$$

e $t_{(\frac{1+\gamma}{2}, n-p-1)} = t_{(0.975, 21)} = 2.08$ Assim o intervalo de confiança é:

$$IC(\mu_0, 0.95) = 7.29 \pm 2.08 * 0.201 = 7.29 \pm 0.418$$

Para ii. $Y_0 = x_0^T \beta + e_0$ Sabemos que um preditor não viciado para Y_0 é μ_0 , pois o preditor para o erro do modelo e_0 é zero. O intervalo de predição para Y_0 com coeficiente de confiança de γ é dado por:

$$IC(Y_0, \gamma) = [\hat{Y}_0 - t_{(\frac{1+\gamma}{2}, n-p-1)} \hat{\sigma}_{\hat{Y}_0}, \hat{Y}_0 + t_{(\frac{1+\gamma}{2}, n-p-1)} \hat{\sigma}_{\hat{Y}_0}]$$

em que $\hat{Y}_0 = \hat{\mu}_0$ e $\hat{\sigma}_{\hat{Y}_0}^2 = \hat{\sigma}^2(1 + x_0^T (X^T X)^{-1} x_0)$ Utilizando os dados: $x_0^T = (1, 10, 0)$ e $\hat{\beta} = \begin{pmatrix} 16.726 \\ -0.943 \\ -0.471 \end{pmatrix} \Rightarrow$

$$\hat{\mu}_0 = x_0^T \hat{\beta} = 7.29,$$

$$\hat{\sigma}_{\hat{Y}_0} = \sqrt{1.647 * 1.246} = 1.433$$

e $t_{(\frac{1+\gamma}{2}, n-p-1)} = t_{(0.975, 21)} = 2.08$ Assim o intervalo de confiança é:

$$IC(Y_0, 0.95) = 7.29 \pm 2.08 * 1.433 = 7.29 \pm 2.979$$

- (e) Apresente, em dois gráficos separados, as curvas estimadas de regressão para as espécies A e B com respectivos intervalos de confiança $\gamma = 95\%$ para as médias estimadas dentro da região experimental.

Resolução

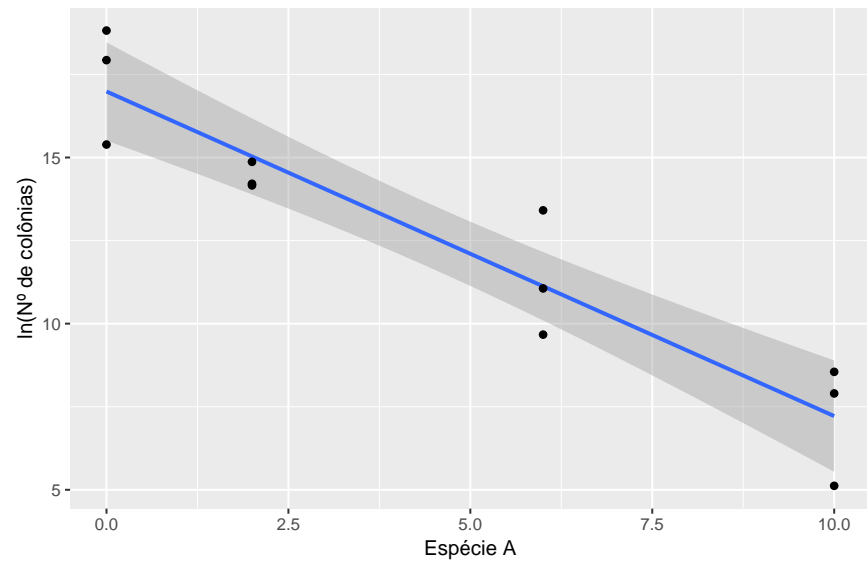
O modelo de regressão, pode ser escrito da seguinte forma

$$\begin{cases} Y_i = \beta_0 + \beta_1 x_{1i} + e_i, & \text{se } i \in \{1, \dots, 12\} \text{ (Especie A)} \\ Y_i = \beta_0 + \beta_2 x_{2i} + e_i, & \text{se } i \in \{13, \dots, 24\} \text{ (Especie B)} \end{cases}$$

Assim as retas estimadas são:

$$\begin{cases} Y_i = 16.726 - 0.943 x_{1i} & \text{se } i \in \{1, \dots, 12\} \text{ (Especie A)} \\ Y_i = 16.726 - 0.471 x_{2i} & \text{se } i \in \{13, \dots, 24\} \text{ (Especie B)} \end{cases}$$

curva estimada com IC de 95%



curva estimada com IC de 95%

