



HOMEWORK III

NOME COMPLETO: LORENA DE CARVALHO GONÇALVES
MARIA LISSA RODRIGUES COSTA
NUMERO DE MATRICULA: 567096
565974

[LINK REPOSITÓRIO GITHUB](#)

QUESTÃO 1

Assume-se que o tempo de vida X (medido em anos) de um computador segue uma distribuição exponencial com parâmetro desconhecido $\lambda > 0$. Uma amostra aleatória dos tempos de vida dos computadores é apresentada na Tabela 1. Os dados são fictícios e são utilizados apenas para fins ilustrativos.

0.99	2.31	10.85	6.15	10.81	3.72	5.75	4.15	9.27	7.84
2.31	10.85	6.15	1.81	3.72	5.75	10.40	10.04	4.15	9.27

Tabela 1: Dados usados na questão 1: Tempo de vida (em anos) dos computadores.

1. Escreva a função densidade de probabilidade da distribuição exponencial com parâmetro λ .
2. Dada uma amostra aleatória X_1, X_2, \dots, X_n :
 - (a) Escreva a função de verossimilhança $L(\lambda)$.
 - (b) Derive a correspondente função log-verossimilhança $\ell(\lambda)$.
 - (c) Determine o estimador de máxima verossimilhança (MLE, do inglês) $\hat{\lambda}$ de λ .
3. Utilizando os dados fornecidos na Tabela 1, calcule o valor numérico do MLE $\hat{\lambda}$.
4. Construa o gráfico da função log-verossimilhança $\ell(\lambda)$ com base nos dados observados, considerando um intervalo adequado de valores para λ . Indique claramente no gráfico o valor do estimador de máxima verossimilhança $\hat{\lambda}$.
5. Utilizando o parâmetro estimado $\hat{\lambda}$:
 - (a) Calcule o tempo médio de vida estimado de um computador.
 - (b) Calcule a probabilidade de que um computador funcione por mais de 5 anos.

6. A distribuição exponencial possui a *propriedade da falta de memória*, o que significa que a probabilidade de falha no futuro não depende do tempo que o computador já esteve em funcionamento.
- (a) Explique essa propriedade com suas próprias palavras.
 - (b) Discuta brevemente se essa suposição parece razoável para modelar o tempo de vida de computadores.

SOLUÇÃO DA QUESTÃO 1

1. Para a distribuição exponencial, a PDF (Probability Distribution Function) se expressa do seguinte modo, com parâmetro $\lambda > 0$, para uma variável aleatória X :

$$PDF(X = k) = \lambda e^{-\lambda k}$$

2. (a) A função de verossimilhança, ou Likelihood Function, $L(\theta)$, é dada por

$$\prod_{i=1}^n f(x_i|\theta)$$

, em que θ representa o parâmetro que descreve ou não os dados observados. Em suma, a função de verossimilhança mede o quão bem um certo parâmetro θ representa o conjunto de dados observados. Para a distribuição exponencial, para o qual $L(\lambda)$, temos:

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i}$$

A partir dessa equação, sabemos que a representação do produtório significa que:

$$(\lambda e^{-\lambda x_1}) * (\lambda e^{-\lambda x_2}) * \dots * (\lambda e^{-\lambda x_n})$$

- (b) A função log-verossimilhança, ou Log-Likelihood, $l(\theta)$, é o logaritmo da função Likelihood $L(\theta)$, ou seja:

$$l(\theta) = \ln(L(\theta))$$

Para a distribuição exponencial, temos: $l(\lambda) = \ln(\prod_{i=1}^n \lambda e^{-\lambda x_i})$

Pela propriedade do logaritmo $\ln(xy) = \ln x + \ln y$, podemos observar a expressão acima da seguinte maneira. Sabemos que $L(\lambda) = (\lambda e^{-\lambda x_1}) * (\lambda e^{-\lambda x_2}) * \dots * (\lambda e^{-\lambda x_n})$, logo:

$$\ln((\lambda e^{-\lambda x_1}) * (\lambda e^{-\lambda x_2}) * \dots * (\lambda e^{-\lambda x_n})) = \sum_{i=1}^n \ln(\lambda^n e^{-\lambda x_i}) = \ln(\lambda^n e^{-\lambda \sum_{i=1}^n x_i})$$

Utilizando da mesma propriedade, podemos manipulá-lo de maneira que:

$$\ln(\lambda^n e^{-\lambda \sum_{i=1}^n x_i}) = \ln(\lambda^n) + \ln(e^{-\lambda \sum_{i=1}^n x_i})$$

Em seguida, podemos referenciar a propriedade $\ln(a)^b = b\ln(a)$ dos logaritmos, e observar que:

$$\ln(\lambda^n) + \ln(e^{-\lambda \sum_{i=1}^n x_i}) = n\ln(\lambda) - \lambda \sum_{i=1}^n x_i \ln(e) = n\ln(\lambda) - \lambda \sum_{i=1}^n x_i$$

Ou seja, temos que:

$$l(\lambda) = n\ln(\lambda) - \lambda \sum_{i=1}^n x_i$$

(c) O estimador de máxima verossimilhança, o MLE (Maximum Likelihood Estimators) $\hat{\lambda}$, do parâmetro λ que estamos utilizando para a distribuição exponencial, é o valor de $\lambda = \hat{\lambda}$ tal que o valor da função verossimilhança seja o máximo possível.

Para encontrar esses valores, podemos fazer uso do teorema dos extremos locais, e da observação de pontos críticos, para a descoberta do parâmetro $\hat{\lambda}$ que maximiza a função verossimilhança. Para maior facilidade dos cálculos, podemos utilizar a log verossimilhança em vez da verossimilhança, já que essa manipulação ainda mantém o ponto máximo da verossimilhança.

Precisaremos derivar $n\ln(\lambda) - \lambda \sum_{i=1}^n x_i$ em λ , e igualar a zero. Sabemos que $\frac{d(\ln x)}{dx} = \frac{1}{x}$, logo:

$$\frac{d(n\ln(\lambda) - \lambda \sum_{i=1}^n x_i)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i$$

Em seguida, igualando a 0, temos que:

$$\frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \therefore \frac{n}{\lambda} = \sum_{i=1}^n x_i \therefore \lambda = \frac{n}{\sum_{i=1}^n x_i}$$

Logo, temos que o MLE $\hat{\lambda}$, o λ para o qual o valor da função verossimilhança é o maior, ou seja, o parâmetro que melhor representa o conjunto de dados, deve ser igual a:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i}$$

3. Descobrimos no ítem anterior que o estimador MLE $\hat{\lambda}$ se expressa como

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i}$$

para a distribuição exponencial. Na tabela 1, temos 20 dados, ou seja, $n = 20$. Para $\sum_{i=1}^n x_i$, temos que:

$\sum_{i=1}^n x_i = 0.99 + 2.31 + \dots + 9.27 = 126.29$
Logo, o Maximum Likelihood Estimator é:

$$\hat{\lambda} = \frac{20}{126.29} = 0.1583657$$

4. Sabemos que a expressão da log verossimilhança é

$$l(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i$$

Precisamos plotar um gráfico com eixo x em valores de λ , e eixo Y $l(\lambda)$. Como temos $\lambda > 0$, observaremos um range de λ partindo do 0 até 1.

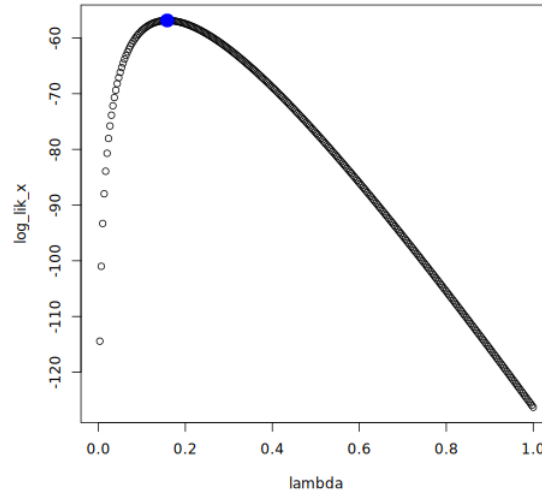


Figura 1: Gráfico da função log verossimilhança $l(\lambda)$ com base nos dados da Tabela 1.

O estimador de máxima verossimilhança para o conjunto de dados da tabela, como calculado na questão anterior, é igual a 0.1583657. Para esse $\hat{\lambda}$, o valor de $l(0.1583657)$ é igual a

$$\begin{aligned} l(0.1583657) &= 20 \ln(0.1583657) - 0.1583657 \sum_{i=1}^{20} x_i \\ &= 20 \ln(0.1583657) - 0.1583657 * (126.29) = -56.85697 \end{aligned}$$

Logo, indicamos o valor do estimador de máxima verossimilhança pelo ponto (0.1583657, -56.85697), de azul. Podemos observar, também, que esse é o ponto mais elevado do gráfico, reforçando a utilidade do MLE como maximizador das funções.

5. (a) Com o parâmetro $\hat{\lambda} = 0.1583657$, para calcularmos o tempo médio de vida esperado de um computador, podemos observar o valor esperado $E[X]$, já que ele representa

o valor médio que esperamos obter após os resultados de experimentos aleatórios.

Para a distribuição exponencial, temos que o valor esperado se expressa por:

$$E[X] = \frac{1}{\lambda}$$

Para o nosso valor de $\lambda = \hat{\lambda}$, temos que:

$$E[X] = \frac{1}{\hat{\lambda}} = \frac{1}{0.1583657} = 6.31449865722$$

Logo, o tempo médio de vida estimado de um computador é de aproximadamente 6.3 anos.

(b) Sabemos que, para a distribuição exponencial, a função densidade de probabilidade (PDF) é dada por

$$PDF(X = k) = \lambda e^{-\lambda k}$$

Como queremos a probabilidade para valores de $k > 5$, e estamos trabalhando com uma variável aleatória contínua, podemos obter essa estimativa por meio de uma integral nos valores de $X = k$ dados. Para $k > 5$, podemos observar que:

$$\begin{aligned} P(X > 5) &= \int_5^{\infty} \lambda e^{-\lambda k} dk = \lambda \int_5^{\infty} e^{-\lambda k} dk \\ &= \lambda * \left(-\frac{1}{\lambda}\right) [e^{-\lambda k}]_5^{\infty} = -[e^{-\lambda k}]_5^{\infty} \end{aligned}$$

Como $e^{-\lambda \infty}$ tende a 0, temos que

$$P(X > 5) = 0 - (-e^{-\lambda 5}) = e^{-\lambda 5}$$

Substituindo λ por $\hat{\lambda} = 0.1583657$ na expressão, temos que:

$$P(X > 5) = e^{-0.1583657*5} = 0.4530157$$

Logo, vemos que a probabilidade de um computador funcionar por mais de 5 anos é de 0.4530157.

6. (a) A propriedade de falta de memória se refere ao fato de que, para a distribuição exponencial, para cada tempo da variável aleatória, a probabilidade de um certo sistema ou item observado funcionar por mais tempo não depende do tempo passado. A taxa de acontecimento dos eventos é constante, não se altera nem com o aumento do tempo. Ou seja, a probabilidade de X durar mais tempo t , já que já se passou tempo $t_0 > 0$, é a mesma probabilidade de um X que tem tempo passado $t_0 = 0$.

(b) Para a modelagem do tempo de vida dos computadores, a utilização da distribuição exponencial pode ser adequada e certa para máquinas cuja idade ainda não seja tão avançada e próxima de zero, já que ainda não há desgaste significativo natural dos componentes.

No entanto, com o aumento dos anos de vida dos computadores e o envelhecimento dos materiais, a taxa de falha decorrente da propriedade da falta de memória aumenta gradativamente, já que a probabilidade real e lógica de uma maior duração da máquina diminui em decorrência do desgaste, e ela se afasta cada vez mais da probabilidade para tempo passado t_0 próximo a 0.

O código é reportado, por exemplo, no Listado 1 ou no Listado da apêndice ??.

Listado 1: Código desenvolvido na solução da questão 1.

```
#QUESTAO 01
#item 3
x <- c(0.99, 2.31, 10.85, 6.15, 10.81, 3.72, 5.75, 4.15, 9.27,
       7.84, 2.31, 10.85, 6.15, 1.81, 3.72, 5.75, 10.40, 10.04, 4.15,
       9.27)
sum_x <- sum(x)
n_x <- length(x)
mle_x <- n_x/sum_x

cat("O somatorio dos dados eh", sum_x, "\n")
cat("O numero de amostras eh", n_x, "\n")
cat("O estimador MLE eh", mle_x, "\n")

#item 4

lambda <- seq(0, 1, length.out = 300)
log_lik_x <- n_x*log(lambda) - lambda*sum_x
plot(lambda, log_lik_x)

log_lik_max <- n_x*log(mle_x) - mle_x*sum_x

points(mle_x, log_lik_max, col = "blue", pch = 19, cex = 2)

#item 5
#(a)
ev_x <- 1/mle_x
cat("O tempo esperado de vida de um computador eh de", ev_x,
    "\n")

#(b)
p_gt_5 <- exp(-0.1583657*5)
cat("A probabilidade de um computador viver mais que 5 anos eh",
    p_gt_5, "\n")
```

QUESTÃO 2

O conjunto de dados de `penguins`, na biblioteca `palmerpenguins`¹ do R, contém medidas para as três espécies de pinguins (figura 2): ilha no arquipélago Palmer na Antártica,

¹ <https://cran.r-project.org/web/packages/palmerpenguins/index.html>

tamanho (comprimento da nadadeira, massa corporal, dimensões do bico) e sexo. Importe o conjunto de dados² e familiarize com ele.



Figura 2: Espécies e características dos pinguins na questão 2.

1. Considere a massa corporal (`body_mass`) em gramas como variável independente, x , e o comprimento do bico (`bill_length`) em milímetros como variável dependente y . Construa um gráfico de dispersão entre x and y . Com base no gráfico, comente se uma relação linear entre as variáveis parece plausível.
2. Defina os parâmetros da reta de regressão com o método dos mínimos quadrados e verifique os resultados obtidos com o comando `lm()` no R. Adicione a reta de regressão no gráfico de dispersão.
3. Calcule os resíduos da regressão e apresente uma representação gráfica dos mesmos. Em seguida, calcule a raiz do erro quadrático médio (RMSE, do inglês) e o coeficiente de determinação R^2 . Comente sobre os resultados obtidos.
4. O conjunto de dados não apresenta outliers evidentes. Modifique esse conjunto introduzindo artificialmente uma observação extrema, seja por meio de um aumento ou de uma redução substancial no valor da massa corporal ou do comprimento do bico de um dos pinguins. Em seguida, ajuste um modelo de regressão linear utilizando o conjunto de dados modificado. Compare os coeficientes estimados da regressão, as retas ajustadas e os valores do RMSE e do R^2 com aqueles obtidos no item 2. Por fim, discuta a influência da observação artificialmente introduzida sobre os resultados da regressão.

SOLUÇÃO DA QUESTÃO 2

1: Ao observar o gráfico, notamos uma correlação positiva, pinguins mais pesados tendem a ter bicos mais longos. Embora os pontos não formem uma linha perfeita (há variabilidade entre indivíduos e espécies), o comportamento geral é linear: o bico cresce

² `install.packages("palmerpenguins"); library(palmerpenguins);`
`penguins_data <- na.omit(penguins) # desconsiderando os dados faltantes`

proporcionalmente ao peso. Portanto, um modelo de regressão linear é uma ótima escolha para a análise.

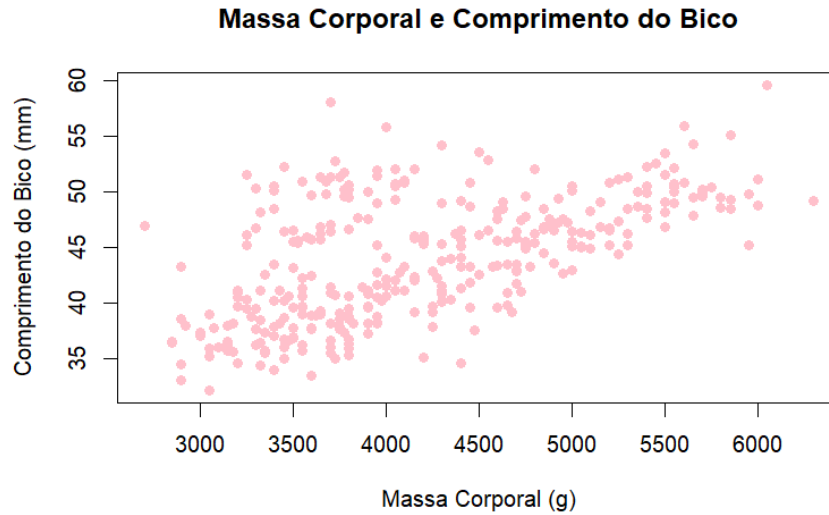


Figura 3: Gráfico de dispersão entre massa corporal e comprimento do bico.

2: O método dos mínimos quadrados busca minimizar a soma dos quadrados dos resíduos ($SQRes$). A função objetivo é:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

Para encontrar os parâmetros no R (que o comando `lm()` automatiza), as fórmulas matemáticas são:

Inclinação (β_1):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\beta_1 = 0.00400329$$

Este valor representa a taxa de variação. Matematicamente, para cada 1 grama adicional na massa corporal do pinguim, o comprimento do bico aumenta, em média, cerca de 0,004 mm.

Intercepto (β_0):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\beta_0 = 27.15072$$

Este é o valor esperado de y quando $x = 0$. Ele indica que a reta "começa" na altura de 26,9 mm.

Reta de regressão:

$$\hat{y} = \beta_0 + \beta_1 x$$

$$\hat{y} = 27.15072 + 0.00400329 \cdot x$$

A reta de regressão resume a relação entre as variáveis em uma única sentença matemática. O sinal positivo do coeficiente angular (+0.00400329) prova que existe uma relação diretamente proporcional: pinguins mais pesados têm, estatisticamente, bicos mais longos. Essa reta é o "melhor ajuste" possível porque ela minimiza a soma dos erros quadráticos entre as medidas reais feitas pelos cientistas e as previsões feitas pelo modelo.

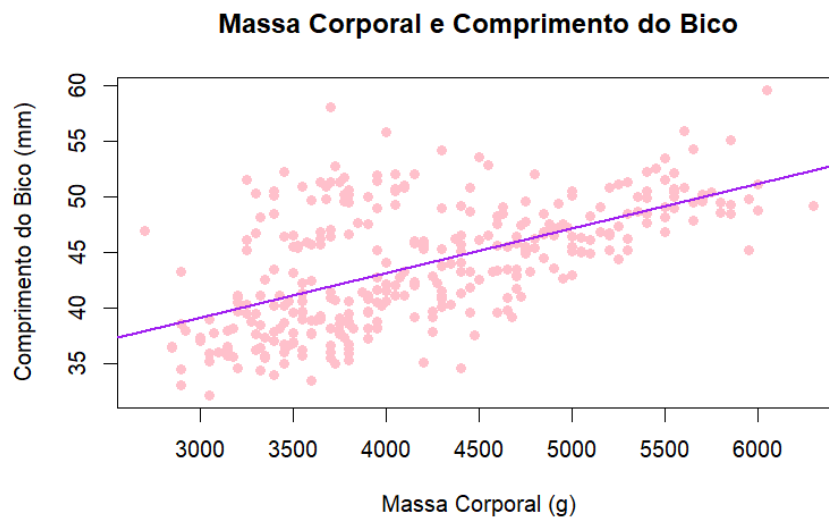


Figura 4: Gráfico de dispersão entre massa corporal e comprimento do bico com reta de regressão.

3: O RMSE é a medida de erro "padrão" do modelo, dada por:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

$$RMSE = 4.410974mm$$

Ou seja, em média, as previsões da nossa reta erram o comprimento do bico em cerca de 4,4 milímetros para mais ou para menos.

O R^2 mede a porcentagem da variação do bico que é explicada pelo peso. É dado

por:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$
$$R^2 = 0.3474526$$

O que significa que a massa corporal explica cerca de 35% da variação no comprimento do bico. Os outros 65% da variação se devem a outros fatores que não estão no modelo simples, como a espécie ou o sexo do animal.

O resíduo (e_i) é a diferença entre o valor real (y) e o valor que a reta previu (\hat{y}):

$$e_i = y_i - \hat{y}_i$$

Ao calcular isso para todos os pinguins no R, geramos o Gráfico de Resíduos. No caso dos pinguins, os resíduos estão bem distribuídos em torno da linha horizontal zero, o que valida o uso de uma reta. Não há um formato de "U" ou "funil" evidente, indicando que o erro é consistente ao longo de todos os pesos.

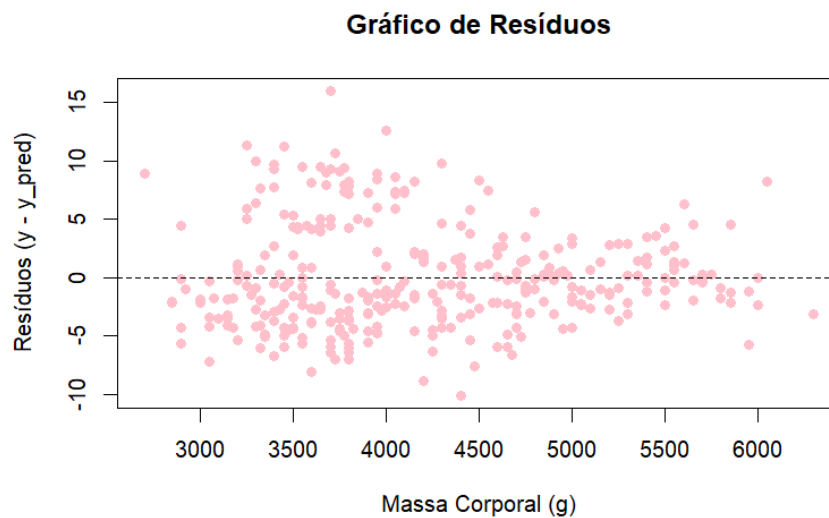


Figura 5: Representação gráfica dos resíduos da regressão.

4: Para a criação de um outlier, no código, criamos um pinguim artificial com valores impossíveis para a espécie. Massa Corporal = 10.000g (10kg), muito acima da máxima massa. Comprimento do Bico = 20mm, um bico extremamente curto para esse peso. Este ponto é uma observação extrema, pois ele não segue a tendência de "quanto mais pesado, maior o bico".

Tabela 2: Comparação dos Modelos de Regressão: Original vs. com Outlier

Métrica	Modelo Original	Modelo com Outlier
Intercepto (β_0)	27.15072	31.67621
Inclinação (β_1)	0.00400329	0.002901024
Coef. Determinação (R^2)	0.3474526	0.1995598
RMSE (mm)	4.410974	5.018952

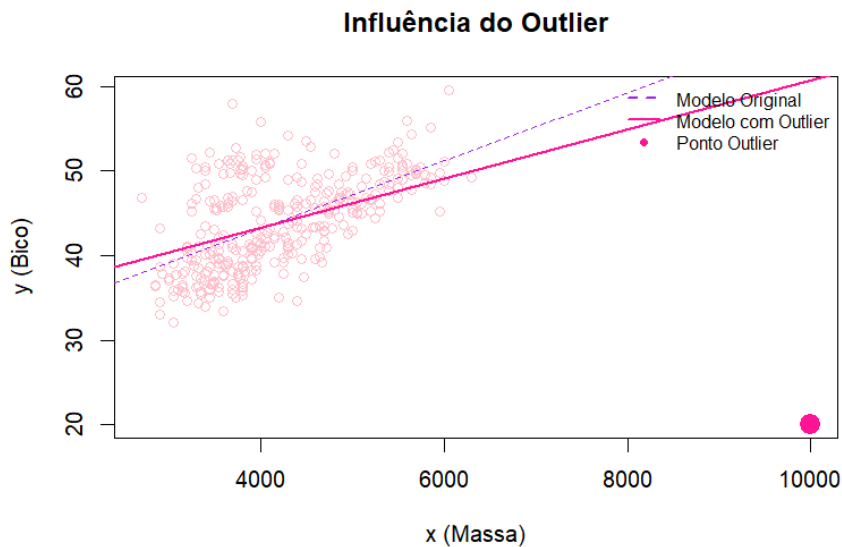


Figura 6: Representação gráfica dos resíduos da regressão.

Nota-se que ocorreu um achatamento da Inclinação (β_1), o valor original era 0,00400, e com o outlier ele caiu para 0,00290. Isso significa que o outlier "puxou" a reta para baixo na extremidade direita. O modelo passou a subestimar a importância da massa corporal no crescimento do bico.

Em relação ao Intercepto (β_0), temos uma elevação, para compensar a perda de inclinação, o ponto onde a reta "começa" subiu de 27,15 para 31,67. A reta rotacionou, perdendo sua fidelidade aos dados reais.

Analisando o coeficiente de determinação (R^2), percebe-se que ele caiu de 0,347 (34,7%) para 0,199 (19,9%). Antes, a massa corporal explicava quase 35% da variação dos bicos. Com a introdução de apenas uma medida errada ou extrema, a capacidade do modelo de explicar os dados caiu para menos de 20%. Isso mostra que o erro introduzido pelo outlier aumentou o ruído do sistema de forma drástica.

Ademais, em relação ao Erro Médio, nota-se que ele subiu de 4,41 mm para 5,01 mm. Isso indica que, ao tentar acomodar o outlier, o modelo tornou-se menos preciso para todos os pinguins da amostra. Um aumento de 0,6 mm no erro médio parece pouco, mas

em um estudo biológico de precisão, isso representa uma perda de confiança considerável nas predições do modelo.

Listado 2: Código desenvolvido na solução da questão 2

```
#item 1

x <- penguins_data$body_mass_g
y <- penguins_data$bill_length_mm

plot(x, y,
      main = "Massa Corporal e Comprimento do Bico",
      xlab = "Massa Corporal (g)",
      ylab = "Comprimento do Bico (mm)",
      pch = 19, col = "pink")

#item 2

beta1 <- sum((x - mean(x)) * (y - mean(y))) / sum((x - mean(x))^2)
beta0 <- mean(y) - beta1 * mean(x)

modelo <- lm(y ~ x)
coef_lm <- coef(modelo)

cat("Manual: Beta0=", beta0, "Beta1=", beta1, "\n")
cat("lm(): Beta0=", coef_lm[1], "Beta1=", coef_lm[2], "\n")

abline(modelo, col = "purple", lwd = 2)

#item 3

residuos <- y - (beta0 + beta1 * x)
plot(x, residuos,
      main = "Gráfico de Resíduos",
      xlab = "Massa Corporal (g)", ylab = "Resíduos (y - y_pred)",
      pch = 19, col = "pink")
abline(h = 0, lty = 2)

n <- length(y)
rmse <- sqrt(sum(residuos^2) / n)
sst <- sum((y - mean(y))^2)
ssr <- sum(residuos^2)
r_quadrado <- 1 - (ssr / sst)

cat("RMSE:", rmse, "\n")
cat("R2:", r_quadrado, "\n")

#item 4

x_out <- x
y_out <- y
```

```

x_out[1] <- 10000 #massa de 10kg (muito pesado)
y_out[1] <- 20    #bico de 20mm (muito curto)

modelo_outlier <- lm(y_out ~ x_out)
coef_lm_outlier <- coef(modelo_outlier)

rmse_out <- sqrt(mean(resid(modelo_outlier)^2))
r2_out <- summary(modelo_outlier)$r.squared

cat("Beta0_original:", coef_lm[1], "\u00a0Beta0_com_Outlier:", coef_lm_outlier
    [1], "\n")
cat("Beta1_original:", coef_lm[2], "\u00a0Beta1_com_Outlier:", coef_lm_outlier
    [2], "\n")
cat("R2_original:", r_quadrado, "\u00a0R2_com_Outlier:", r2_out, "\n")
cat("RMSE_original:", rmse, "\u00a0RMSE_com_Outlier:", rmse_out, "\n")

plot(x_out, y_out, main = "Influ\u00eancia do Outlier",
     xlab = "x (Massa)", ylab = "y (Bico)", col = "pink")
points(x_out[1], y_out[1], col = "deeppink", pch = 19, cex = 2)
abline(modelo_linear, col = "purple", lty = 2)
abline(modelo_outlier, col = "deeppink", lwd = 2)

legend("topright",
     legend = c("Modelo Original", "Modelo com Outlier", "Ponto Outlier"),
     col = c("purple", "deeppink", "deeppink"),
     lty = c(2, 1, NA),
     pch = c(NA, NA, 19),
     lwd = c(2, 2, NA),
     bty = "n",
     cex = 0.8)

```