

Exercício 26

Natanael Magalhães Cardoso

22/07/2021

Bibliotecas

```
library(ggplot2)
```

Item (a)

O código abaixo mostra a definição do conjunto de dados e o ajuste da curva

```
data <- data.frame(
  x=seq(1, 8, 1),
  y=c(74, 54, 52, 51, 52, 53, 58, 71)
)

formula <- y ~ x + I(x^2)

quadModel <- lm(formula, data=data)

summary(quadModel)

##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##      1      2      3      4      5      6      7      8
##  3.6250 -5.8036 -0.7679  1.7321  2.6964  0.1250 -1.9821  0.3750
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   84.4821     4.9036   17.229 1.21e-05 ***
## x             -15.8750     2.5001   -6.350  0.00143 **
## I(x^2)         1.7679     0.2712    6.519  0.00127 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.515 on 5 degrees of freedom
## Multiple R-squared:  0.8948, Adjusted R-squared:  0.8527
## F-statistic: 21.25 on 2 and 5 DF,  p-value: 0.003594
```

Pelo sumário da curva ajustada, a equação obtida para o ajuste foi

$$y = 84.4821 - 15.8750x + 1.7679x^2$$

```
ggplot(data, aes(x=x)) +
  geom_point(aes(y=y), shape=16, size=2) +
  stat_smooth(aes(y=y), method='lm', formula=formula, size=1, level=0.95, se=TRUE) +
  labs(
    title='Gráfico de dispersão',
    subtitle='Concentração de glucose em função do tempo',
    x='Tempo (dias)',
    y='Concentração de glucose (g/L)'
  )
```

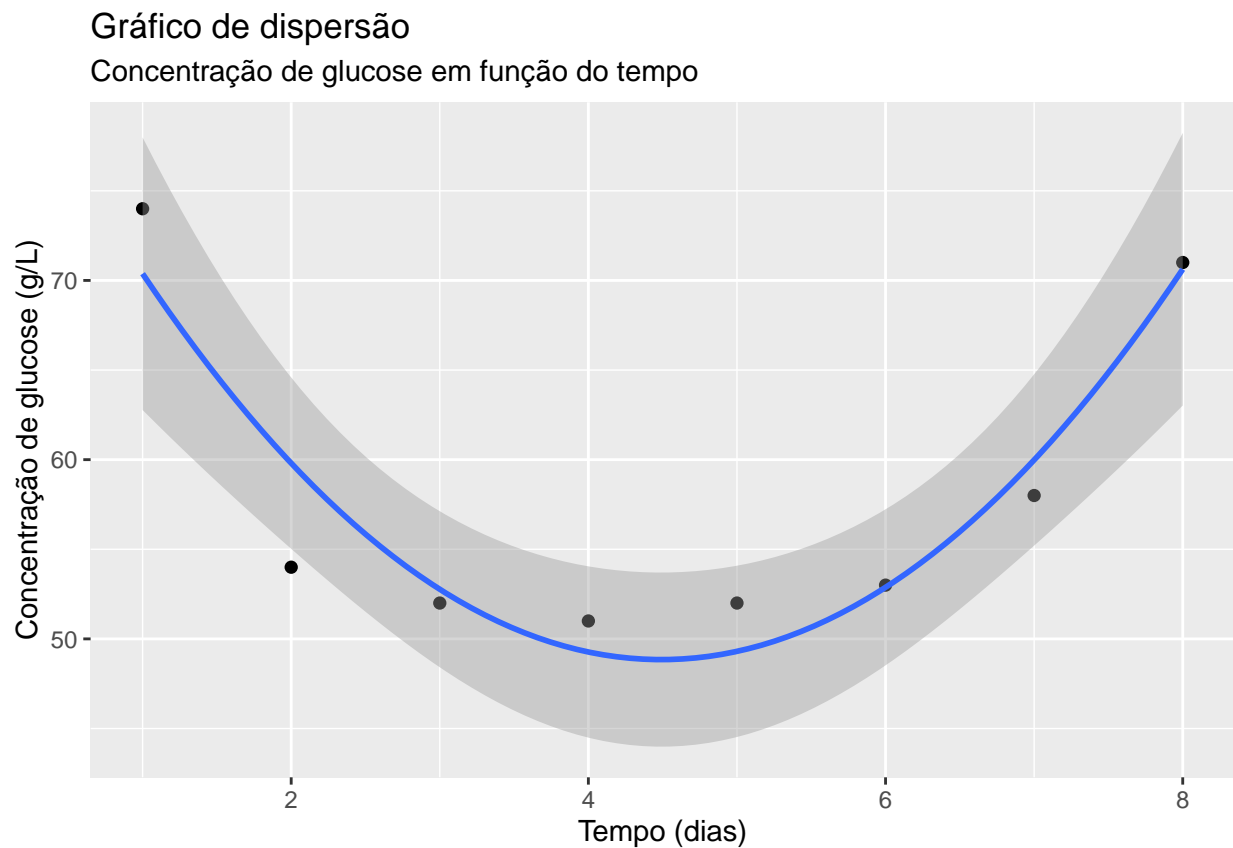


Figure 1: Gráfico de dispersão e curva ajustada da concentração de glucose em função do tempo.

Pela análise do gráfico da Figura 1, é notado que o modelo quadrático é um bom ajuste para os dados observados. A maioria das observações estão dentro do intervalo de confiança, ao nível de 95%.

Item (b)

O código abaixo mostra os valores de \hat{y} e ϵ para os dados observados.

```
predict(quadModel)

##          1          2          3          4          5          6          7          8
## 70.37500 59.80357 52.76786 49.26786 49.30357 52.87500 59.98214 70.62500

residuals(quadModel)

##          1          2          3          4          5          6          7
## 3.6250000 -5.8035714 -0.7678571 1.7321429 2.6964286 0.1250000 -1.9821429
##          8
## 0.3750000
```

Logo, $\hat{y}_6 = y(6) = 52.875$ g/L e $\epsilon_6 = 0.125$ g/L.

Item (c)

O valor do coeficiente de determinação R^2 será usado para determinar a proporção da variância na variável dependente que pode ser explicada pela variável independente. O código abaixo implementa o cálculo de R^2 , onde

$$R^2 = 1 - \frac{SQE}{SQT} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

```
sqe <- sum(residuals(quadModel)^2)
sqt <- sum((data$y - mean(data$y))^2)
r2 <- 1-(sqe/sqt)
r2
```

```
## [1] 0.8947513
```

Como $R^2 \approx 0.895$, aproximadamente 89.5% dos dados se ajustam ao modelo quadrático.