Exercicio9

higor lucas de Araujo Freitas

2024-10-29

R Markdown

Neste exercício, você irá realizar uma análise de regressão simples em dois conjuntos de dados que simulam exemplos reais. O primeiro dataset reflete uma relação linear simples entre as variáveis, enquanto o segundo apresenta um cenário onde a inclinação e o resíduo aumentam à medida que o valor da variável independente cresce. Vamos fornecer um anexo de referência para auxiliar no exercício.

Situação: Dataset 1: Este dataset simula a relação entre a quantidade de fertilizante (kg/ha) aplicado em uma plantação e o rendimento da colheita (toneladas) obtido. A relação entre essas duas variáveis é aproximadamente linear, com pouco ruído nos dados.

Dataset 2: Este dataset simula a relação entre o tamanho do motor de um carro (litros) e o consumo de combustível (litros/100 km). Neste caso, à medida que o tamanho do motor aumenta, o consumo de combustível também aumenta de forma não linear, com resíduos que aumentam com o tamanho do motor.

Tarefas: Geração e visualização dos dados:

Use os dois datasets fornecidos abaixo:

Dataset 1: Fertilizante vs. Rendimento da Colheita

10 25.3 20 31.5 30 36.7 40 40.2 50 46.9 60 50.7 70 55.4 80 59.3 90 63.1 100 68.7

```r
fertilizante <- c(10,20,30,40,50,60,70,80,90,100)
redimento_da_colheita <- c(25.3,31.5,36.7,40.2,46.9,50.7,55.4,59.3,63.1,68.7)

n <- length(fertilizante)


media_x <- sum(fertilizante) / n
media_y <- sum(redimento_da_colheita) / n


numerador <- sum((fertilizante - media_x) * (redimento_da_colheita - media_y))
denominador <- sum((fertilizante - media_x)^2)
beta_1 <- numerador/denominador

beta_0 <- media_y - beta_1 * media_x

cat("Calcule manualmente os coeficientes de regressão linear BETA 1", beta_1, "\n")
```

```
## Calcule manualmente os coeficientes de regressão linear BETA 1 0.4692121
```

```r
cat("Calcule manualmente os coeficientes de regressão linear BETA 0",beta_0,"\n")
```

```
## Calcule manualmente os coeficientes de regressão linear BETA 0 21.97333
```

```r
# Calcule o somatório dos quadrados dos erros (SSE)

# Valores previstos (Y_hat) usando a reta de regressão

Y_hat <- beta_0 + beta_1 * fertilizante

# Cálculo do SSE
SSE <- sum((redimento_da_colheita - Y_hat)^2)

cat(" Cálculo do SSE",SSE,"\n")
```

```
##  Cálculo do SSE 6.715879
```

```r
# Calcule o erro padrão dos coeficientes e os intervalos de confiança para beta_0 e beta_1.

# Erro Padrao dos Coeficientes beta_0 e beta_1
# Agora calculamos o erro padrao para os coeficientes.
SE_beta1 <- sqrt(SSE / (n - 2) / sum((fertilizante - media_x)^2))

SE_beta0 <- sqrt(SSE / (n - 2)) * sqrt(1/n + (media_x^2 / sum((fertilizante - media_x)^2)))

cat("Erro padrao de beta1:", SE_beta1, "\nErro padrao de beta0:", SE_beta0)
```

```
## Erro padrao de beta1: 0.0100874
## Erro padrao de beta0: 0.625907
```

```r
# Calcule o coeficiente de determinação (R²) e a estatística F, juntamente com o p-valor para avaliar a
alpha <- 0.05
t_crit <- qt(1 - alpha/2, df = n-2)
IC_beta1 <- c(beta_1 - t_crit * SE_beta1, beta_1 + t_crit * SE_beta1)
IC_beta0 <- c(beta_0 - t_crit * SE_beta0, beta_0 + t_crit * SE_beta0)
cat("Intervalo de confiança para beta1:", IC_beta1, "\nIntervalo de confiança para beta0:", IC_beta0,"\n
```

```
## Intervalo de confiança para beta1: 0.4459505 0.4924737
## Intervalo de confiança para beta0: 20.52999 23.41668
```

```r
# Calcule o coeficiente de determinação (R²) e a estatística F, juntamente com o p-valor para avaliar a
SS_tot <- sum((redimento_da_colheita - media_y)^2)
R2 <- 1 - SSE / SS_tot
cat("R^2:", R2,"\n")
```

```
## R^2: 0.9963161
```

```r
# Calculamos a estatística F e o p-valor para avaliar a significância do modelo.
F_stat <- ((SS_tot - SSE) / 1) / (SSE / (n - 2))
p_value <- 1 - pf(F_stat, 1, n - 2)
cat("Estatística F:", F_stat, "\nP-valor:", p_value)
```

```
## Estatística F: 2163.613
## P-valor: 5.043477e-11
```

```r
# Criar um dataset transformado - aplicando log em fertilizante, por exemplo
fertilizante_log <- log(fertilizante)
modelo_transformado <- lm(redimento_da_colheita ~ fertilizante_log)

cat("Resultados com lm() - Dataset Transformado (log(fertilizante)):\n")
```

```
## Resultados com lm() - Dataset Transformado (log(fertilizante)):
```

```r
print(summary(modelo_transformado))
```

```
##
## Call:
## lm(formula = redimento_da_colheita ~ fertilizante_log)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.2576 -2.6346 -0.7586  2.0732  6.1014
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -23.550      7.124  -3.305   0.0108 *
## fertilizante_log   18.707      1.838  10.177 7.44e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.042 on 8 degrees of freedom
## Multiple R-squared:  0.9283, Adjusted R-squared:  0.9193
## F-statistic: 103.6 on 1 and 8 DF,  p-value: 7.444e-06
```

```r
# Cálculo manual dos coeficientes no dataset transformado
media_x_log <- mean(fertilizante_log)
numerador_log <- sum((fertilizante_log - media_x_log) * (redimento_da_colheita - media_y))
denominador_log <- sum((fertilizante_log - media_x_log)^2)
beta_1_log <- numerador_log / denominador_log
beta_0_log <- media_y - beta_1_log * media_x_log
SSE_log <- sum((redimento_da_colheita - (beta_0_log + beta_1_log * fertilizante_log))^2)

cat("\nResultados Manuais - Dataset Transformado (log(fertilizante)):\n")
```

```
##
## Resultados Manuais - Dataset Transformado (log(fertilizante)):
```

```r
cat("beta_0:", beta_0_log, "\n")
```

```
## beta_0: -23.54993
```

```r
cat("beta_1:", beta_1_log, "\n")
```

```
## beta_1: 18.70691
```

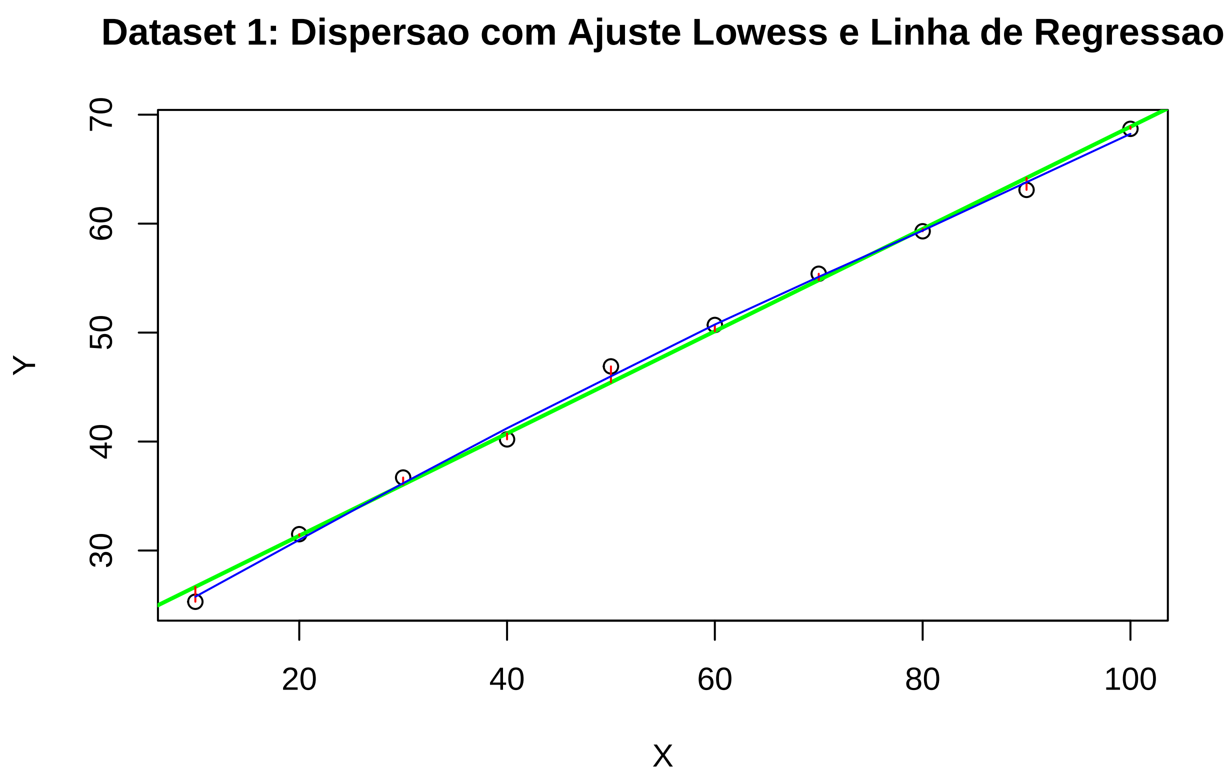
```r
cat("SSE:", SSE_log, "\n")
```

```
## SSE: 130.7153
```

```r
model1 <- lm(redimento_da_colheita ~ fertilizante)
summary(model1)
```

```
##
## Call:
## lm(formula = redimento_da_colheita ~ fertilizante)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36545 -0.45894 -0.02606  0.57985  1.46606
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21.97333    0.62591   35.11 4.74e-10 ***
## fertilizante  0.46921    0.01009   46.52 5.04e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9162 on 8 degrees of freedom
## Multiple R-squared:  0.9963, Adjusted R-squared:  0.9959
## F-statistic:  2164 on 1 and 8 DF,  p-value: 5.043e-11
```

```r
# Gráfico para o Dataset 1 com linha de regressao e resíduos
plot(fertilizante, redimento_da_colheita, main="Dataset 1: Dispersao com Ajuste Lowess e Linha de Regres
abline(model1, col="green", lwd=2) # Linha de regressao
segments(fertilizante, redimento_da_colheita, fertilizante, Y_hat, col="red") # Resíduos
lines(lowess(fertilizante, redimento_da_colheita), col="blue") # Ajuste Lowess
```



Dataset 2: Tamanho do Motor vs. Consumo de Combustível

1.2 7.2 1.4 8.1 1.6 9.5 1.8 11.4 2.0 12.8 2.2 13.3 2.4 15.0 2.6 17.5 2.8 19.0 3.0 20.4

```
tamanho_motor <- c(1.2,1.4,1.6,1.8,2,2.2,2.4,2.6,2.8,3)
cosumo_de_combustivel <- c(7.2,8.1,9.5,11.4,12.8,13.3,15,17.5,19,20.4)

n <-  length(tamanho_motor)
media_x <- mean(tamanho_motor)
media_y <- mean(cosumo_de_combustivel)

nums <- sum((tamanho_motor - media_x)*(cosumo_de_combustivel - media_y))
dem <- sum((tamanho_motor - media_x)^2)

beta_1 <- nums/dem

beta_0 <- media_y - beta_1 * media_x


cat("Calcule manualmente os coeficientes de regressão linear BETA 1", beta_1, "\n")
```

```
## Calcule manualmente os coeficientes de regressão linear BETA 1 7.466667
```

```
cat("Calcule manualmente os coeficientes de regressão linear BETA 0",beta_0)
```

```
## Calcule manualmente os coeficientes de regressão linear BETA 0 -2.26
```

```r
# Calcule o somatório dos quadrados dos erros (SSE)


Y_hat <- beta_0 + beta_1 * tamanho_motor

# Cálculo do SSE
SSE <- sum((cosumo_de_combustivel - Y_hat)^2)

cat("Cálculo do SSE",SSE,"\n")
```

```
## Cálculo do SSE 1.857333
```

```r
SE_beta1 <- sqrt(SSE / (n - 2) / sum((tamanho_motor - media_x)^2))

SE_beta0 <- sqrt(SSE / (n - 2)) * sqrt(1/n + (media_x^2 / sum((tamanho_motor - media_x)^2)))

cat("Erro padrao de beta1:", SE_beta1, "\nErro padrao de beta0:", SE_beta0)
```

```
## Erro padrao de beta1: 0.2652424
## Erro padrao de beta0: 0.5774736
```

```r
# Calcule o coeficiente de determinação (R²) e a estatística F, juntamente com o p-valor para avaliar a
alpha <- 0.05
t_crit <- qt(1 - alpha/2, df = n-2)
IC_beta1 <- c(beta_1 - t_crit * SE_beta1, beta_1 + t_crit * SE_beta1)
IC_beta0 <- c(beta_0 - t_crit * SE_beta0, beta_0 + t_crit * SE_beta0)
cat("Intervalo de confiança para beta1:", IC_beta1, "\nIntervalo de confiança para beta0:", IC_beta0)
```

```
## Intervalo de confiança para beta1: 6.855017 8.078317
## Intervalo de confiança para beta0: -3.591657 -0.9283435
```

```r
# Calcule o coeficiente de determinação (R²) e a estatística F, juntamente com o p-valor para avaliar a
SS_tot <- sum((cosumo_de_combustivel - media_y)^2)
R2 <- 1 - SSE / SS_tot
cat("R^2:", R2,"\n")
```

```
## R^2: 0.9900055
```

```r
# Calculamos a estatística F e o p-valor para avaliar a significância do modelo.
F_stat <- ((SS_tot - SSE) / 1) / (SSE / (n - 2))
p_value <- 1 - pf(F_stat, 1, n - 2)
cat("Estatística F:", F_stat, "\nP-valor:", p_value,"\n")
```

```
## Estatística F: 792.4422
## P-valor: 2.739313e-09
```

```r
# Criar um dataset transformado - aplicando log em tamanho_motor
tamanho_motor_log <- log(tamanho_motor)
modelo_transformado <- lm(cosumo_de_combustivel ~ tamanho_motor_log)

cat("Resultados com lm() - Dataset Transformado (log(tamanho_motor)):\n")
```

```
## Resultados com lm() - Dataset Transformado (log(tamanho_motor)):
```

```r
print(summary(modelo_transformado))
```

```
##
## Call:
## lm(formula = cosumo_de_combustivel ~ tamanho_motor_log)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3856 -0.5271 -0.1729  0.6901  1.3608
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         3.1783     0.8049   3.949  0.00424 **
## tamanho_motor_log  14.5948     1.0604  13.764 7.49e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9702 on 8 degrees of freedom
## Multiple R-squared:  0.9595, Adjusted R-squared:  0.9544
## F-statistic: 189.4 on 1 and 8 DF,  p-value: 7.493e-07
```

```r
# Cálculo manual dos coeficientes no dataset transformado
media_x_log <- mean(tamanho_motor_log)
numerador_log <- sum((tamanho_motor_log - media_x_log) * (cosumo_de_combustivel - media_y))
denominador_log <- sum((tamanho_motor_log - media_x_log)^2)
beta_1_log <- numerador_log / denominador_log
beta_0_log <- media_y - beta_1_log * media_x_log
SSE_log <- sum((cosumo_de_combustivel - (beta_0_log + beta_1_log * tamanho_motor_log))^2)

cat("\nResultados Manuais - Dataset Transformado (log(tamanho_motor)):\n")
```

```
##
## Resultados Manuais - Dataset Transformado (log(tamanho_motor)):
```

```r
cat("beta_0:", beta_0_log, "\n")
```

```
## beta_0: 3.178279
```

```r
cat("beta_1:", beta_1_log, "\n")
```

```
## beta_1: 14.59475
```

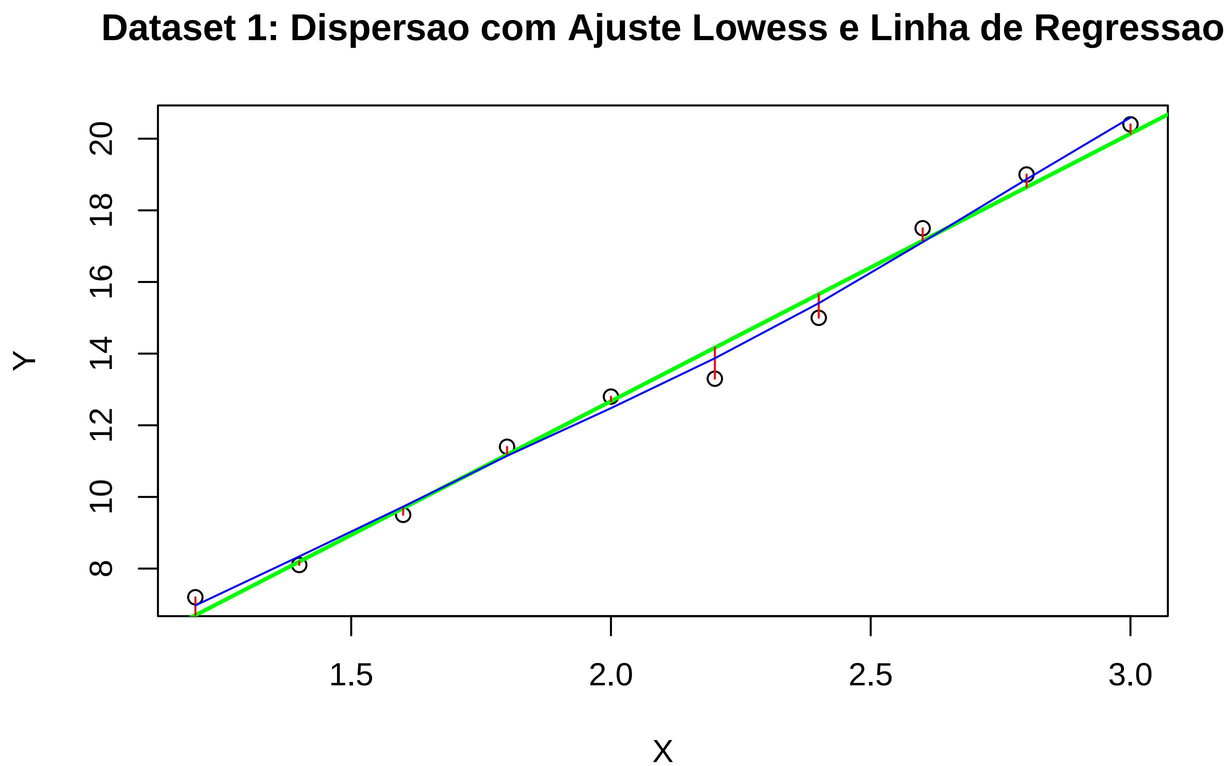
```r
cat("SSE:", SSE_log, "\n")
```

```
## SSE: 7.529945
```

```r
model1 <- lm(cosumo_de_combustivel ~ tamanho_motor)
summary(model1)
```

```
##
## Call:
## lm(formula = cosumo_de_combustivel ~ tamanho_motor)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8667 -0.1633  0.1733  0.3250  0.5000
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.2600     0.5775  -3.914  0.00446 **
## tamanho_motor   7.4667     0.2652  28.150 2.74e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4818 on 8 degrees of freedom
## Multiple R-squared:   0.99,  Adjusted R-squared:  0.9888
## F-statistic: 792.4 on 1 and 8 DF,  p-value: 2.739e-09
```

```r
# Gráfico para o Dataset 1 com linha de regressao e resíduos
plot(tamanho_motor, cosumo_de_combustivel, main="Dataset 1: Dispersao com Ajuste Lowess e Linha de Regre
abline(model1, col="green", lwd=2) # Linha de regressao
segments(tamanho_motor, cosumo_de_combustivel, tamanho_motor, Y_hat, col="red") # Resíduos
lines(lowess(tamanho_motor, cosumo_de_combustivel), col="blue") # Ajuste Lowess
```



Cálculos manuais para o Dataset 1 (Fertilizante vs. Rendimento da Colheita):

Calcule manualmente os coeficientes de regressão linear (beta_0 e beta_1). Calcule o somatório dos quadrados dos erros (SSE). Calcule o erro padrão dos coeficientes e os intervalos de confiança para beta_0 e beta_1. Calcule o coeficiente de determinação (R^2) e a estatística F, juntamente com o p-valor para avaliar a significância dos parâmetros.

Transformação e Regressão para o Dataset 2 (Tamanho do Motor vs. Consumo de Combustível):

Dado que a inclinação e os resíduos aumentam com o tamanho do motor, aplique uma transformação adequada. Realize os cálculos manuais de regressão linear para os dados transformados. Calcule os coeficientes, intervalos de confiança e a estatística F, assim como no Dataset 1. Comparação com a Função lm():

Compare os resultados manuais com os gerados pela função lm() para ambos os datasets (original e transformado). Verifique se os resultados batem. Visualização Gráfica:

Para cada dataset (original e transformado), faça um gráfico de dispersão dos dados com a linha de regressão ajustada. Mostre também os resíduos graficamente, desenhando linhas conectando cada ponto de dados ao valor previsto pela regressão. Utilize o ajuste Lowess para suavizar os dados e sobreponha esta curva no gráfico. Interpretação dos Resultados:

Interprete os coeficientes da regressão para os dois datasets. O que os coeficientes beta_0 e beta_1 representam no contexto de fertilizante e rendimento da colheita, e de tamanho do motor e consumo de combustível? Discuta a importância de realizar transformações nos dados quando a relação entre as variáveis é não linear.