

Atividade 4 - Modelos de Regressão Linear Simples aplicado ao banco de dados do consumo de combustível

Disciplina: Regressão I

Schênia Taynna Medeiros Silva - 20190156798

2023-08-14

Dados sobre o consumo de combustível

Os dados da *Tabela B.3* do livro do Montgomery podem ser encontrados no pacote MPV do R como o nome `table.b3`. Estes dados consistem de 32 observações sobre o consumo de combustível de diferentes automóveis, com variáveis sobre o desempenho da quilometragem da gasolina e características físicas e/ou de performance.

y = Consumo de combustível em Milhas/galão (do inglês Miles/gallon);

x1 = Cilindrada ou volume de deslocamento do motor (cubic in.) (do inglês Engine displacement);

x2 = Cavalos de potência (do inglês Horsepower);

x3 = Torque (ft-lb);

x4 = Taxa de compressão (do inglês Compression ratio);

x5 = Razão do eixo traseiro (do inglês Rear axle ratio);

x6 = Número de carburadores (do inglês Carburetor (barrels));

x7 = Velocidades de transmissão (do inglês No. of transmission speeds);

x8 = Comprimento total (in.) (do inglês Overall length);

x9 = Largura (in.) (do inglês Width);

x10 = Peso (lbs)(do inglês Weight);

x11 = Tipo de transmissão (1=automática, 0>manual) (do inglês Transmission).

Considerando estes dados e fixando o nível de significância em 5%, responda as questões abaixo:

- a) Estime um modelo de regressão linear que relaciona o consumo de combustível em milhas/galão, y, com o volume de deslocamento do motor (cilindrada), x1 (in³). Expresse o modelo estimado e interprete os parâmetros destes modelos.

```
#library('MPV') # Biblioteca do pacote
dados <- MPV::table.b3 # Dados
modelo <- lm(y~x1, data = dados) # Ajustar o modelo de regressão linear simples
summary(modelo) # Resultado
```

```
##
## Call:
## lm(formula = y ~ x1, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.7923 -1.9752  0.0044  1.7677  6.8171
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.722677   1.443903   23.36 < 2e-16 ***
## x1          -0.047360   0.004695  -10.09 3.74e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.065 on 30 degrees of freedom
## Multiple R-squared:  0.7723, Adjusted R-squared:  0.7647
## F-statistic: 101.7 on 1 and 30 DF,  p-value: 3.743e-11
```

A inclinação da reta é -0.0473, com a redução esperada no consumo de combustível em milhas/galão ao acréscimo de uma unidade no volume de deslocamento do motor.

Dessa forma, conforme o volume de deslocamento do moto aumenta, há uma redução de 0.047 no consumo médio de combustível. Já O intercepto obtido é 33.72, é representado como valor esperado do consumo de combustível quando o volume de deslocamento do motor estiver igual a 0.

- b) Construa a tabela ANOVA para o modelo estimado e explique cada um dos resultados da tabela e conclua sobre a significância da regressão.

```
anova(modelo)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1          1  955.72   955.72   101.74 3.743e-11 ***
## Residuals  30  281.82     9.39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ao realizar a análise de variância (ANOVA) utilizamos o teste F para verificar se existe uma relação significativa entre a variável resposta (consumo de combustível) e a variável preditora (volume de deslocamento do motor).

Dessa forma, obtendo o valor F de 101.74 e o valor p associado ao F (3.743e-11) confirma que há uma relação estatisticamente significativa entre o consumo de combustível e o volume de deslocamento do motor (x1). Também podemos avaliar o p-valor (3.743e-11) como um O valor extremamente baixo, assim indicando que há uma forte evidência estatística de que o volume de deslocamento do motor afeta significativamente o consumo de combustível.

- c) Qual o percentual da variabilidade total do consumo de combustível é explicado por esse modelo?

```
summary(modelo)
```

```
##
## Call:
## lm(formula = y ~ x1, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.7923 -1.9752  0.0044  1.7677  6.8171
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.722677   1.443903   23.36 < 2e-16 ***
## x1          -0.047360   0.004695  -10.09 3.74e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.065 on 30 degrees of freedom
## Multiple R-squared:  0.7723, Adjusted R-squared:  0.7647
## F-statistic: 101.7 on 1 and 30 DF,  p-value: 3.743e-11
```

Obtemos que o percentual da variabilidade total do consumo de combustível explicado por esse modelo ajustado é de 77,23%

d) Construa um intervalo de confiança para o consumo médio de combustível se a cilindrada é de 275 in³

```
predict(modelo,newdata=data.frame(x1=275), level = 0.90, interval="confidence")
```

```
##           fit          lwr          upr
## 1 20.69879 19.77571 21.62188
```

Dessa forma, obtemos que o valor esperado do consumo médio de combustível é 20.699 quando a cilindrada (volume de deslocamento do motor) atinge 275, com um intervalo de confiança variando de 19.77 a 21.62.

e) Suponha que desejamos prever o consumo de combustível obtidos de carros com 275 in³ cilindradas. Qual a estimativa pontual para este consumo? Calcule o intervalo de predição para o consumo de combustível, com 95% de confiança.

```
predict(modelo,newdata=data.frame(x1=275), level = 0.95, interval="prediction")
```

```
##           fit          lwr          upr
## 1 20.69879 14.34147 27.05611
```

O consumo de combustível obtidos de carros com 275 in³ cilindradas, está entre aproximadamente 14.34 e 27.06, com 95% de confiança.

f) Compare os dois intervalos obtidos em (d) e (e). Explique a diferença entre eles. Qual deles é o maior, e por quê?

A diferença entre é :

o intervalo de confiança se concentra na média da população, enquanto o intervalo de predição considera a variabilidade geral dos dados e a incerteza na previsão de valores individuais. Por isso que o intervalo de predição é maior!

- g) Ajuste agora um modelo de regressão linear que relaciona a variável consumo de combustível em milhas/galão, y , com o peso do carro, x_{10} .

```
modelo_2 <- lm(y~x10, data = dados) # Ajustar o modelo de regressão linear simples
summary(modelo_2) # Resultado
```

```
##
## Call:
## lm(formula = y ~ x10, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.4093 -2.0352 -0.6714  3.0807  7.2026
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.8524315  2.2789916  17.926  < 2e-16 ***
## x10         -0.0057516  0.0006149  -9.353 2.12e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.246 on 30 degrees of freedom
## Multiple R-squared:  0.7446, Adjusted R-squared:  0.7361
## F-statistic: 87.48 on 1 and 30 DF,  p-value: 2.121e-10
```

- h) Considere agora o modelo de regressão linear que relaciona o consumo de combustível, y , com o comprimento total do veículo, x_8 , e o seu peso, x_{10} . Baseado na comparação dos dois modelos, qual das variáveis regressoras podemos concluir que é a melhor? Justifique, apresentando os resultados.

```
modelo_3 <- lm(y~ x8+ x10, data = dados) # Ajustar o modelo de regressão linear simples
summary(modelo_3) # Resultado
```

```
##
## Call:
## lm(formula = y ~ x8 + x10, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.1776 -1.8607 -0.8257  2.6454  7.0030
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.379687  10.061703   1.827   0.0781 .
## x8           0.189992   0.083130   2.285   0.0298 *
## x10          -0.009654   0.001802  -5.357 9.39e-06 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.039 on 29 degrees of freedom
## Multiple R-squared:  0.7836, Adjusted R-squared:  0.7687
## F-statistic: 52.51 on 2 and 29 DF,  p-value: 2.295e-10
```