

# Atividade 4

## Dados sobre o consumo de combustível

Paulo Ricardo Seganfredo Campana

11 de agosto de 2023

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. 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),  $x_1$  ( $\text{in}^3$ ). Expresse o modelo estimado e interprete os parâmetros destes modelos.
- b) Construa a tabela ANOVA para o modelo estimado e explique cada um dos resultados da tabela e conclua sobre a significância da regressão.
- c) Qual o percentual da variabilidade total do consumo de combustível é explicado por esse modelo?
- d) Construa um intervalo de confiança para o consumo médio de combustível se a cilindrada é de 275  $\text{in}^3$ .
- e) Suponha que desejamos prever o consumo de combustível obtidos de carros com 275  $\text{in}^3$  cilindradas. Qual a estimativa pontual para este consumo? Calcule o intervalo de predição para o consumo de combustível, 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?
- g) Ajuste agora um modelo de regressão linear que relaciona a variável consumo de combustível em milhas/galão,  $y$ , com o peso do carro,  $x_{10}$ .
- 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.

```
data <- MPV::table.b3
fit1 <- lm(y ~ x1, data)
summary(fit1)
```

#### 1: Termos do modelo de regressão linear

| Termo       | Estimativa | Erro padrão | Estatística | p-valor    |
|-------------|------------|-------------|-------------|------------|
| (Intercept) | 33.723     | 1.444       | 23.355      | 8.4559e-21 |
| x1          | -0.047     | 0.005       | -10.086     | 3.7430e-11 |

Temos um modelo de regressão linear expresso pela reta  $\hat{y} = 33.723 - 0.047 \times x1$ , isso significa que, para cada polegada cúbica extra na cilindrada do motor, o consumo de combustível cai em 0.047 milhas por galão (veículos de maior cilindrada são menos eficientes). O intercepto não tem interpretação pois não temos observações com cilindrada 0.

```
anova(fit1)
```

#### 2: Análise de variância

| Termo     | gl | SS     | MS      | Estatística | p-valor |
|-----------|----|--------|---------|-------------|---------|
| x1        | 1  | 955.72 | 955.720 | 101.736     | 4e-11   |
| Residuals | 30 | 281.82 | 9.394   |             |         |

A tabela ANOVA nos mostra a contribuição das variáveis da regressão na variação do consumo de combustível, a soma dos quadrados explicada pela cilindrada é de 955.72, enquanto pelos resíduos é 281.82, o teste  $F$  mostra que a regressão é significativa, com p-valor muito baixo.

```
summary(fit1)$r.squared
## [1] 0.77227
```

O coeficiente de determinação  $R^2$  traz a informação sobre o quanto da variância do consumo de combustível é explicada pelo modelo, nesse caso 77%, enquanto o resto é explicada pelos resíduos.

```
predict(fit1, newdata = data.frame(x1 = 275), interval = "confidence")
predict(fit1, newdata = data.frame(x1 = 275), interval = "prediction")
```

3: Consumo médio de combustível

| Estimativa | IC inferior | IC superior |
|------------|-------------|-------------|
| 20.699     | 19.588      | 21.81       |

4: Consumo de combustível previsto

| Estimativa | IC inferior | IC superior |
|------------|-------------|-------------|
| 20.699     | 14.341      | 27.056      |

O primeiro intervalo de confiança é sobre o valor esperado para o consumo de combustível de um carro com 275 in<sup>3</sup> cilindradas segundo o modelo, enquanto o segundo intervalo refere aos possíveis valores previstos do consumo de combustível para uma observação com 95% de confiança.

O intervalo de predição é maior, pois é um intervalo para a variável aleatória  $Y$  enquanto o primeiro intervalo é sobre a média da mesma,  $\bar{Y}$ .

```
fit2 <- lm(y ~ x10, data)
summary(fit2)
```

5: Termos do modelo de regressão linear

| Termo       | Estimativa | Erro padrão | Estatística | p-valor    |
|-------------|------------|-------------|-------------|------------|
| (Intercept) | 40.852     | 2.279       | 17.926      | 1.4092e-17 |
| x10         | -0.006     | 0.001       | -9.353      | 2.1207e-10 |

```
fit3 <- lm(y ~ x8 + x10, data)
summary(fit3)
```

6: Termos do modelo de regressão linear

| Termo       | Estimativa | Erro padrão | Estatística | p-valor    |
|-------------|------------|-------------|-------------|------------|
| (Intercept) | 18.38      | 10.062      | 1.827       | 7.8056e-02 |
| x8          | 0.19       | 0.083       | 2.285       | 2.9782e-02 |
| x10         | -0.01      | 0.002       | -5.357      | 9.3946e-06 |

A variável  $x10$  referente ao peso do veículo é mais significativa, pois o teste de hipótese mostra menos evidências em relação ao seu coeficiente ser igual a 0.