

Atividade 4

Teresa Nascimento

- 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 (in3). Expresse o modelo estimado e interprete os parâmetros destes modelos.

```
library('MPV')
dados <- MPV::table.b3
modelo <- lm(y~x1, data = dados)
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
```

A inclinação da reta -0.047 é a redução esperada no consumo de combustível em milhas/galão ao acréscimo de uma unidade no volume de deslocamento do motor (cilindrada). O intercepto 33.72 é o valor esperado do consumo de combustível quando o volume de deslocamento do motor (cilindrada) for 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
```

A soma dos quadrados associada ao volume de deslocamento do motor (cilindrada) é 955.72, ou seja, a variação total explicada por essa variável; e a variação não explicada pelo modelo é de 281.82. O P-valor associado ao teste F é muito baixo, portanto há evidências para rejeitar a hipótese nula de que a variável independente tem coeficiente igual a zero, ou seja, o modelo de regressão é estatisticamente significativo. A variável volume de deslocamento do motor (cilindrada) tem efeito significativo no 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
```

O percentual da variabilidade total do consumo de combustível explicado por esse modelo é de 77,23%

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

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

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

O valor esperado do consumo de combustível quando a cilindrada é de 275 é de 20.699 com um intervalo de confiança variando entre 19.775 e 21.622

e) Suponha que desejamos prever o consumo de combustível obtidos de carros com 275 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
```

- 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