


Atividade: concessionária

Você trabalha em uma concessionária de carros e deseja

- entender como características técnicas dos veículos influenciam o consumo de combustível (mpg - miles per gallon)
- Você tem à disposição a base de dados mtcars, que contém informações de 32 modelos de carro, com variáveis como:
 - mpg: consumo (milhas por galão)
 - hp: potência do motor (horsepower)
 - wt: peso do carro
 - cyl: número de cilindros
 - am: tipo de transmissão (0 = automática, 1 = manual)


```
1 install.packages("repr")
```



Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

```
1 library(repr)
2 options(repr.plot.width = 16, repr.plot.height = 8)
```

```
1 # Parte 0 - Estatística descritiva
2 # Use summary(mtcars) e str(mtcars) para explorar a base.
3 # Gere gráficos de dispersão com plot() ou pairs().
4
5 data(mtcars)
6 summary(mtcars)
7 str(mtcars)
8
9 plot(mtcars)
```

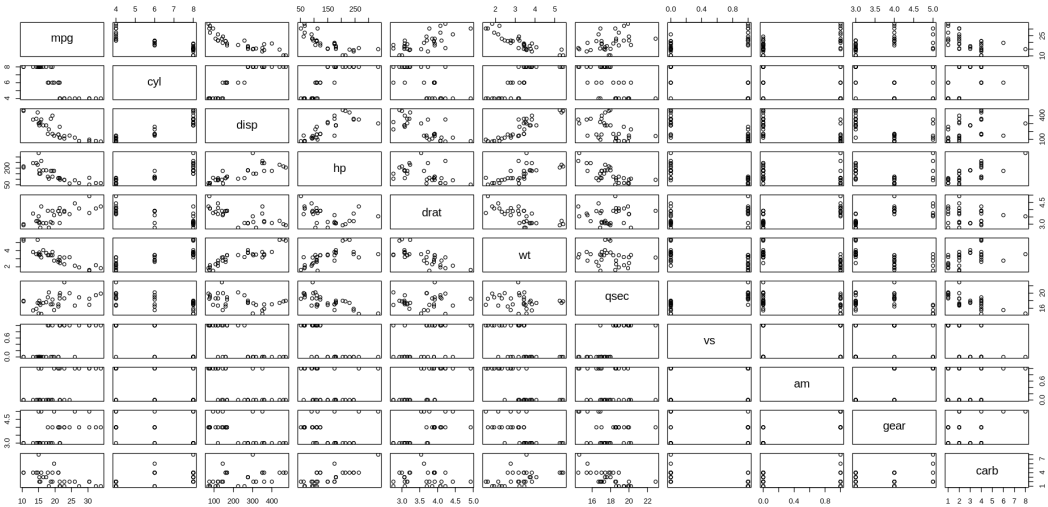


| mpg | cyl | disp | hp |
|---------------|---------------|---------------|---------------|
| Min. :10.40 | Min. :4.000 | Min. : 71.1 | Min. : 52.0 |
| 1st Qu.:15.43 | 1st Qu.:4.000 | 1st Qu.:120.8 | 1st Qu.: 96.5 |
| Median :19.20 | Median :6.000 | Median :196.3 | Median :123.0 |
| Mean :20.09 | Mean :6.188 | Mean :230.7 | Mean :146.7 |
| 3rd Qu.:22.80 | 3rd Qu.:8.000 | 3rd Qu.:326.0 | 3rd Qu.:180.0 |
| Max. :33.90 | Max. :8.000 | Max. :472.0 | Max. :335.0 |

| drat | wt | qsec | vs |
|---------------|---------------|---------------|----------------|
| Min. :2.760 | Min. :1.513 | Min. :14.50 | Min. :0.0000 |
| 1st Qu.:3.080 | 1st Qu.:2.581 | 1st Qu.:16.89 | 1st Qu.:0.0000 |
| Median :3.695 | Median :3.325 | Median :17.71 | Median :0.0000 |
| Mean :3.597 | Mean :3.217 | Mean :17.85 | Mean :0.4375 |
| 3rd Qu.:3.920 | 3rd Qu.:3.610 | 3rd Qu.:18.90 | 3rd Qu.:1.0000 |
| Max. :4.930 | Max. :5.424 | Max. :22.90 | Max. :1.0000 |

| am | gear | carb |
|----------------|---------------|---------------|
| Min. :0.0000 | Min. :3.000 | Min. :1.000 |
| 1st Qu.:0.0000 | 1st Qu.:3.000 | 1st Qu.:2.000 |
| Median :0.0000 | Median :4.000 | Median :2.000 |
| Mean :0.4062 | Mean :3.688 | Mean :2.812 |
| 3rd Qu.:1.0000 | 3rd Qu.:4.000 | 3rd Qu.:4.000 |
| Max. :1.0000 | Max. :5.000 | Max. :8.000 |

'data.frame': 32 obs. of 11 variables:
\$ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
\$ cyl : num 6 6 4 6 8 6 8 4 4 6 ...
\$ disp: num 160 160 108 258 360 ...
\$ hp : num 110 110 93 110 175 105 245 62 95 123 ...
\$ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
\$ wt : num 2.62 2.88 2.32 3.21 3.44 ...
\$ qsec: num 16.5 17 18.6 19.4 17 ...
\$ vs : num 0 0 1 1 0 1 0 1 1 1 ...
\$ am : num 1 1 1 0 0 0 0 0 0 ...
\$ gear: num 4 4 4 3 3 3 3 4 4 ...
\$ carb: num 4 4 1 1 2 1 4 2 2 4 ...



```

1 # Parte 1 - Regressão Linear Simples
2 # Ajuste um modelo de regressão linear simples prevendo mpg a partir de hp
3
4 model_simple <- lm(mpg ~ hp, data = mtcars)
5 summary(model_simple)

```

```

Call:
lm(formula = mpg ~ hp, data = mtcars)

Residuals:
    Min       1Q   Median       3Q      Max
-5.7121 -2.1122 -0.8854  1.5819  8.2360

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.09886    1.63392   18.421 < 2e-16 ***
hp          -0.06823    0.01012   -6.742 1.79e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.863 on 30 degrees of freedom
Multiple R-squared:  0.6024,    Adjusted R-squared:  0.5892
F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07

```

```

1 # Parte 2 - Regressão Linear Múltipla
2 # Adicione o peso do carro como segunda variável
3
4 model_multi <- lm(mpg ~ hp + wt, data = mtcars)
5 summary(model_multi)

```

```

Call:
lm(formula = mpg ~ hp + wt, data = mtcars)

Residuals:
    Min       1Q   Median       3Q      Max
-3.941 -1.600 -0.182  1.050  5.854

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.22727    1.59879   23.285 < 2e-16 ***
hp          -0.03177    0.00903   -3.519 0.00145 **
wt          -3.87783    0.63273   -6.129 1.12e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.593 on 29 degrees of freedom
Multiple R-squared:  0.8268,    Adjusted R-squared:  0.8148
F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12

```

```

1 install.packages("scatterplot3d")

```

```

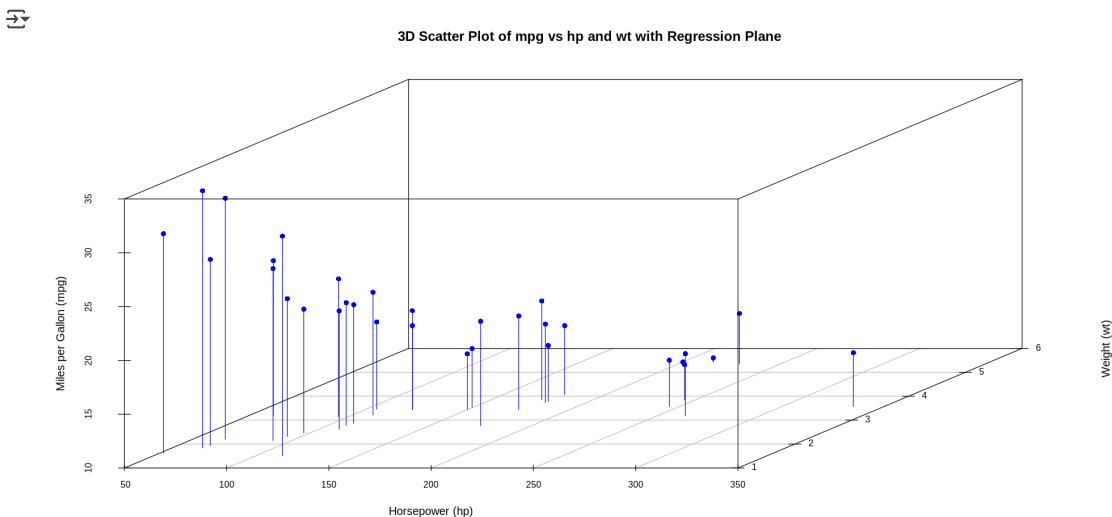
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

```

```

1 # Parte 3 - gráfico 3D
2 # Existe algum ponto extremo (outlier) que parece não se ajustar bem à reta/plano de regressão?
3
4 library(scatterplot3d)
5
6 scatterplot3d(mtcars$hp, mtcars$wt, mtcars$mpg,
7               xlab = "Horsepower (hp)",
8               ylab = "Weight (wt)",
9               zlab = "Miles per Gallon (mpg)",
10              main = "3D Scatter Plot of mpg vs hp and wt with Regression Plane",
11              color = "blue",
12              pch = 16,
13              type="h") # add vertical lines to the plane

```



```

1 # Perform ANOVA test to compare the simple and multiple regression models
2 anova_result <- anova(model_simple, model_multi)
3
4 # Display the ANOVA test results
5 print(anova_result)

```

```
↺ Analysis of Variance Table

Model 1: mpg ~ hp
Model 2: mpg ~ hp + wt
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1      30 447.67
2      29 195.05   1    252.63 37.561 1.12e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Parte 4 – Interpretação

1. O que significa o sinal dos coeficientes?

- No modelo de regressão linear simples (`model`), o coeficiente para `hp` tem um sinal negativo. Isso indica que, à medida que a potência do motor (`hp`) aumenta, o consumo de combustível (`mpg`) tende a diminuir.
- No modelo de regressão linear múltipla (`model_multi`), o coeficiente para `hp` também é negativo, indicando a mesma relação inversa com `mpg`. O coeficiente para `wt` (peso do carro) também é negativo, sugerindo que carros mais pesados tendem a ter menor `mpg`. O coeficiente constante (intercepto) representa o `mpg` esperado quando `hp` e `wt` são zero (embora isso não tenha um significado prático real neste contexto, pois `hp` e `wt` não podem ser zero).

2. O modelo com duas variáveis é melhor que o modelo simples? Justifique usando o anova()

- Após executar a função `anova()` comparando `model_simple` e `model_multi`, observamos o valor-p (`Pr(>F)`) na saída para a linha correspondente a `model_multi`.
 - Se o valor-p for menor que o nosso nível de significância escolhido (geralmente 0.05), rejeitamos a hipótese nula. A hipótese nula, neste caso, é que a variável adicional (`wt`) em `model_multi` não melhora significativamente o ajuste do modelo em comparação com `model_simple`. Rejeitar a hipótese nula significa que a inclusão de `wt` melhorou significativamente o modelo.
 - Se o valor-p for maior que 0.05, falhamos em rejeitar a hipótese nula, sugerindo que adicionar `wt` não melhorou significativamente o ajuste do modelo.
- Com base na saída do ANOVA, se o valor-p para `model_multi` for menor que 0.05, podemos concluir que o modelo com duas variáveis (`hp` e `wt`) é estatisticamente significativo melhor do que o modelo simples (apenas com `hp`) na explicação da variância em `mpg`. Isso ocorre porque a variável adicional `wt` explica uma quantidade significativa da variância restante não explicada após considerar `hp`.

3. Que variáveis poderiam ser incluídas para melhorar o modelo?

- Observando a descrição da base de dados e os gráficos de dispersão, outras variáveis que poderiam ser incluídas para potencialmente melhorar o modelo incluem:
 - `cyl` (número de cilindros): É razoável supor que o número de cilindros pode influenciar o consumo de combustível.
 - `am` (tipo de transmissão): O tipo de transmissão (automática vs. manual) também é conhecido por afetar a eficiência do combustível.
 - Outras variáveis como `disp` (cilindrada), `drat` (relação do eixo traseiro), `qsec` (tempo no quarto de milha), `vs` (tipo de motor - V-shape vs. straight), e `gear` (número de marchas) também poderiam ser consideradas, dependendo de sua correlação com `mpg` e sua multicolinearidade com as variáveis já incluídas. A seleção de variáveis adicionais deve ser feita com base em análise exploratória, conhecimento do domínio (carros e consumo de combustível) e técnicas de seleção de modelos (como `stepwise regression`, análise de VIF para multicolinearidade, etc.).