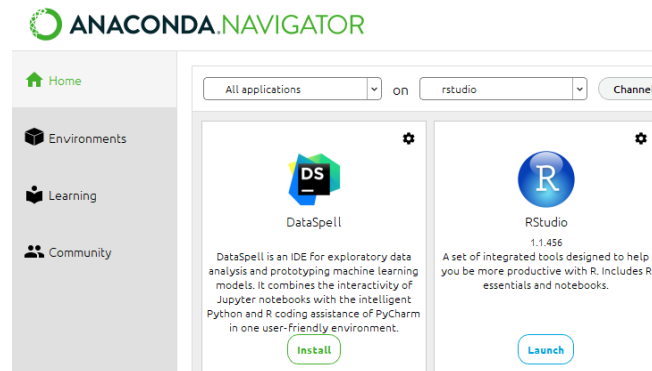


Aluno: Fábio Manuel Martins Troncão

versão do Rstudio:



Você trabalha para a Motor Trend, uma revista sobre a indústria automobilística. Olhando para um conjunto de dados de uma coleção de carros, eles estão interessados em explorar a relação entre um conjunto de variáveis e milhas por galão (MPG) (resultado). Eles estão particularmente interessados nas duas questões a seguir:

"Uma transmissão automática ou manual é melhor para MPG (Milhas por Galão), a mesma ideia de km/l (quilômetros por litro)

"Quantificar a diferença MPG entre transmissões automáticas e manuais"

code:

```
library(datasets)
```

```
data(mtcars)
```

```
names(mtcars)
```

```
> library(datasets)
> data(mtcars)
> names(mtcars)
[1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear" "carb"
```

```
>
```

```
summary(mtcars)
```

```
> summary(mtcars)
      mpg      cyl      disp      hp      drat      wt      qsec      vs
Min.   :10.40  Min.   :4.000  Min.   : 71.1  Min.   : 52.0  Min.   :2.760  Min.   :1.513  Min.   :14.50  Min.   :0.0000
1st Qu.:15.43  1st Qu.:4.000  1st Qu.:120.8  1st Qu.: 96.5  1st Qu.:3.080  1st Qu.:2.581  1st Qu.:16.89  1st Qu.:0.0000
Median :19.20  Median :6.000  Median :196.3  Median :123.0  Median :3.695  Median :3.325  Median :17.71  Median :0.0000
Mean   :20.09  Mean   :6.188  Mean   :230.7  Mean   :146.7  Mean   :3.597  Mean   :3.217  Mean   :17.85  Mean   :0.4375
3rd Qu.:22.80  3rd Qu.:8.000  3rd Qu.:326.0  3rd Qu.:180.0  3rd Qu.:3.920  3rd Qu.:3.610  3rd Qu.:18.90  3rd Qu.:1.0000
Max.   :33.90  Max.   :8.000  Max.   :472.0  Max.   :335.0  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
```

Análise exploratória de dados:

Começamos a análise de dados exploratórios observando o gráfico de dispersão par a par entre todas as variáveis. (consulte o apêndice para gráficos)

Com a distribuição da variável dependente: mpg, e veja se atende aos pressupostos da regressão. Tanto no histograma quanto na densidade do kernel, é aproximadamente normal (consulte o apêndice para gráficos).

Uma transmissão automática ou manual é melhor para MPG?

Traçamos um boxplot de MPG por tipos de transmissão (consulte o apêndice para gráficos).

A partir desse enredo simples, parece que a transmissão manual é melhor em MPG do que a transmissão automática. Em seguida, realizamos um teste para confirmar essa hipótese.

```
t.test(mtcars$mpg~mtcars$am,conf.level=0.95)
```

```
> t.test(mtcars$mpg~mtcars$am,conf.level=0.95)

      welch Two Sample t-test

data:  mtcars$mpg by mtcars$am
t = -3.7671, df = 18.332, p-value = 0.001374
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -11.280194  -3.209684
sample estimates:
mean in group 0 mean in group 1
    17.14737      24.39231
```

Com valor $p = 0,001374$, rejeitamos a hipótese nula de que não há diferença em MPG, e a transmissão manual parece melhor em MPG do que a transmissão automática, desde que todas as outras condições sejam as mesmas.

Quantifique a diferença de MPG entre transmissões automáticas e manuais.

Nesta seção, pretendemos quantificar a diferença de MPG entre os tipos de transmissão e descobrir se existem outras variáveis que explicam as diferenças de MPG.

Primeiro, tentamos fazer uma regressão linear multivariada com todas as variáveis:

```
mlr = lm(data = mtcars, mpg ~ .)
summary(mlr)
```

```

> mlr = lm(data = mtcars, mpg ~ .)
> summary(mlr)

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

Residuals:
    Min       1Q   Median       3Q      Max
-3.4506 -1.6044 -0.1196  1.2193  4.6271

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.30337   18.71788   0.657  0.5181
cyl         -0.11144    1.04502  -0.107  0.9161
disp          0.01334    0.01786   0.747  0.4635
hp          -0.02148    0.02177  -0.987  0.3350
drat          0.78711    1.63537   0.481  0.6353
wt         -3.71530    1.89441  -1.961  0.0633 .
qsec          0.82104    0.73084   1.123  0.2739
vs            0.31776    2.10451   0.151  0.8814
am            2.52023    2.05665   1.225  0.2340
gear          0.65541    1.49326   0.439  0.6652
carb         -0.19942    0.82875  -0.241  0.8122
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.65 on 21 degrees of freedom
Multiple R-squared:  0.869,    Adjusted R-squared:  0.8066
F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07

```

A partir dos coeficientes, parece que wt é o único fator que muda significativamente com mpg. No entanto, incluir todas as variáveis possivelmente resultará em overfitting, e precisamos testar diferentes modelos com diferentes variáveis exploratórias.

R tem uma função de seleção automática de variável (etapa). Usamos essa função de escolha automática de modelo para escolher os melhores modelos de regressão linear com base no AIC.

```

bestmodel = step(lm(data = mtcars, mpg ~ .), trace=0)
summary(bestmodel)

```

```

>
> bestmodel = step(lm(data = mtcars, mpg ~ .), trace=0)
> summary(bestmodel)

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

Residuals:
    Min       1Q   Median       3Q      Max
-3.4811 -1.5555 -0.7257  1.4110  4.6610

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.6178     6.9596   1.382 0.177915
wt          -3.9165     0.7112  -5.507 6.95e-06 ***
qsec          1.2259     0.2887   4.247 0.000216 ***
am            2.9358     1.4109   2.081 0.046716 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.459 on 28 degrees of freedom
Multiple R-squared:  0.8497,    Adjusted R-squared:  0.8336
F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11

```

Parece que o melhor modelo é aquele que inclui wt, qsec e am, o que significa que além dos tipos de transmissão, peso e aceleração também precisam ser considerados. O peso muda negativamente com mpg, e qsec e am mudam positivamente. Cada aumento de peso de lb/1000 causará uma diminuição de aproximadamente 4 mpg, cada aumento de 1/4 milha no tempo causará um aumento de 1,2 mpg e, em média, a transmissão manual é 2,9 mpg melhor do que a transmissão automática. O modelo é capaz de explicar 85% da variância. As parcelas residuais também parecem estar espalhadas aleatoriamente (consulte o apêndice).

Conclusão:

Em média, a transmissão manual é melhor que a transmissão automática em 2,9 mpg. No entanto, o tipo de transmissão não é o único fator responsável por MPG, peso e aceleração (tempo de 1/4 de milha) também precisam ser considerados.

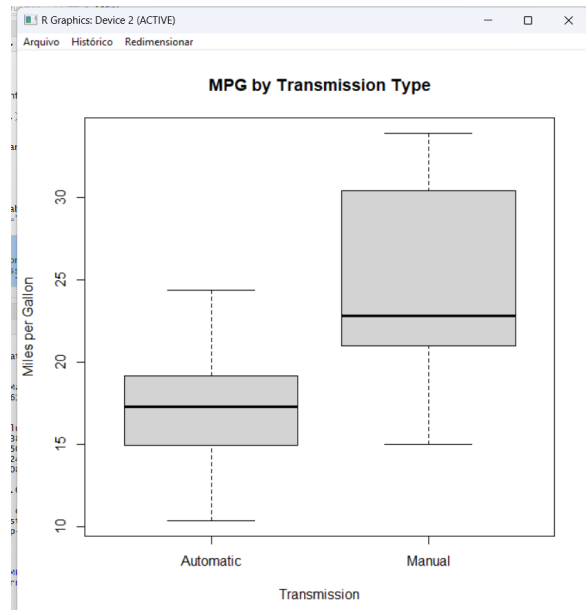
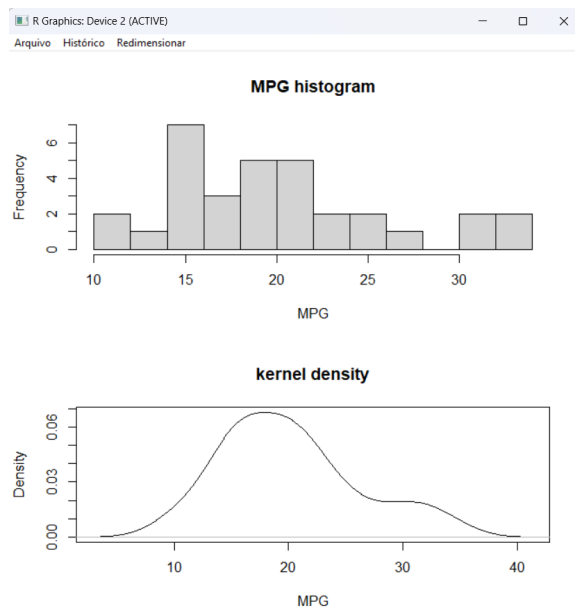
Apêndice

gráficos de dispersão pareados:

`pairs(mtcars)`



```
par(mfrow=c(2,1))
hist(mtcars$mpg, breaks=10, xlab="MPG", main="MPG histogram")
plot(density(mtcars$mpg), main="kernel density", xlab="MPG")
boxplot(mpg~am, data = mtcars,
        xlab = "Transmission",
        ylab = "Miles per Gallon",
        main = "MPG by Transmission Type",
        names = c("Automatic", "Manual"))
```



```
par(mfrow = c(2,2))
plot(bestmodel)
```

