

## Exercício de Regressão Múltipla

Caros e caras estudantes,

Hoje vamos tentar uma regressão múltipla usando um dos conjuntos de dados do pacote MASS do R.

Lembrem-se que a regressão é uma das técnicas mais clássicas de ajuste de modelos, usando o método de mínimos quadrados para ajustar uma equação linear entre as variáveis dependentes e independentes. Vamos fazer o ajuste mais simples, usando variáveis do tipo contínuo independentes e dependentes.

Usando o comando `library` carreguem o pacote MASS

Usando o comando `View` visualizem o banco de dados `Cars93`

Este banco de dados representa as características de diversos modelos de carros vendidos nos EUA no ano de 1993. Tem 93 linhas e 27 colunas. Cada linha é um modelo de carro. Cada coluna é uma variável observada sobre aquele modelo de carro. Há variáveis de todos os tipos (lembrem dos tipos de variáveis que discutimos logo no início do curso?).

No caso atual estamos interessados em saber quais atributos do carro influenciam o preço de venda do carro. Portanto nossa variável dependente (a variável que queremos explicar) é o preço de venda - neste banco de dados - dataframe - é a variável `"Price"` do tipo numérico, portanto contínua.

Quais seriam as variáveis independentes? Isto depende das nossas hipóteses de trabalho. Vamos propor que o preço do carro seja dependente de sua economia de combustível (variável `"MPG.highway"`) e do espaço interno (variável `"Rear.seat.room"`). Traduzindo: os carros com maior economia de combustível seriam mais baratos; e os carros mais espaçosos seriam mais caros. Estamos a princípio sugerindo uma relação negativa entre economia de combustível e preço, e uma relação positiva entre espaço interno e preço. E achamos que estas duas variáveis explicam a maior parte da variação de preço entre os diversos modelos.

Vamos fazer uma inspeção visual antes de fazer a regressão para ver se nossa idéia tem mérito.

Use a função `plot` para fazer 2 gráficos: das variáveis `Price` vs `MPG.highway` e `Price` vs `Rear.seat.room`.

Verifiquem se as premissas da regressão linear são atendidas verificando que não há correlação alta entre as variáveis independentes usando a função `cor.test`. Qual é o resultado?

Montem agora o ajuste de modelo usando a função `lm` e as variável dependente e independentes do parágrafo acima. Criem um objeto com o resultado da regressão.

Use a função `summary` sobre o objeto para ver o resultado da regressão. Respondam aos seguintes quesitos:

- a) Qual é a percentagem da variância do preço de venda explicado pela regressão múltipla ( $r$  quadrado).
- b) Qual é a contribuição relativa de cada variável independente para o ajuste final (usem a função `model`, a função `anova` e a função `summary` para verem)

Discutam o ajuste e os resultados.

Script feito hoje dia 28 abril 2020 com Marcela Ponce e Ana Terra

```
library(MASS)
View(Cars93)
plot(Cars93$MPG.highway, Cars93$Price)
plot(Cars93$Rear.seat.room, Cars93$Price)
cor.test(Cars93$MPG.highway, Cars93$Rear.seat.room)
regpreco<-lm(Cars93$Price~ Cars93$MPG.highway+Cars93$Rear.seat.room)
summary(regpreco)
anova(regpreco)
plot(regpreco)
coef(regpreco)
```