

Lista 3

Aluno: Gustavo Luiz Bispo dos Santos - 117210400

Aluno: Diego Amancio Pereira - 116210716

Aluno: Gilmar Gonzaga da Silva - 119211123

Aluno: Anderson Kleber Dantas - 117110537

2022-08-24

Desenvolvimento de uma Pesquisa com Análise de Correlação e regressão Linear Simples

- 1) Apresente/Descreva um conjunto de dados que contenha duas (2) variáveis do tipo quantitativa (numérica) cujo interesse é investigar sobre a existência de uma relação linear entre elas. Descreva o contexto ao qual a base de dados está inserida.

Os dados foram extraídos da revista Motor Trend US de 1974 e abrangem o consumo de combustível e 10 aspectos do design e desempenho do automóvel para 32 automóveis (modelos de 1973 a 74).

As variáveis que serão usadas são mpg (Milhas/galão) e hp (potência bruta).

hp: é o valor medido no eixo motor, com os acessórios necessários para ligá-lo e funcionar autonomamente.

galão: 1 galão = 3,78544 litros

```
dados <- mtcars
dados
```

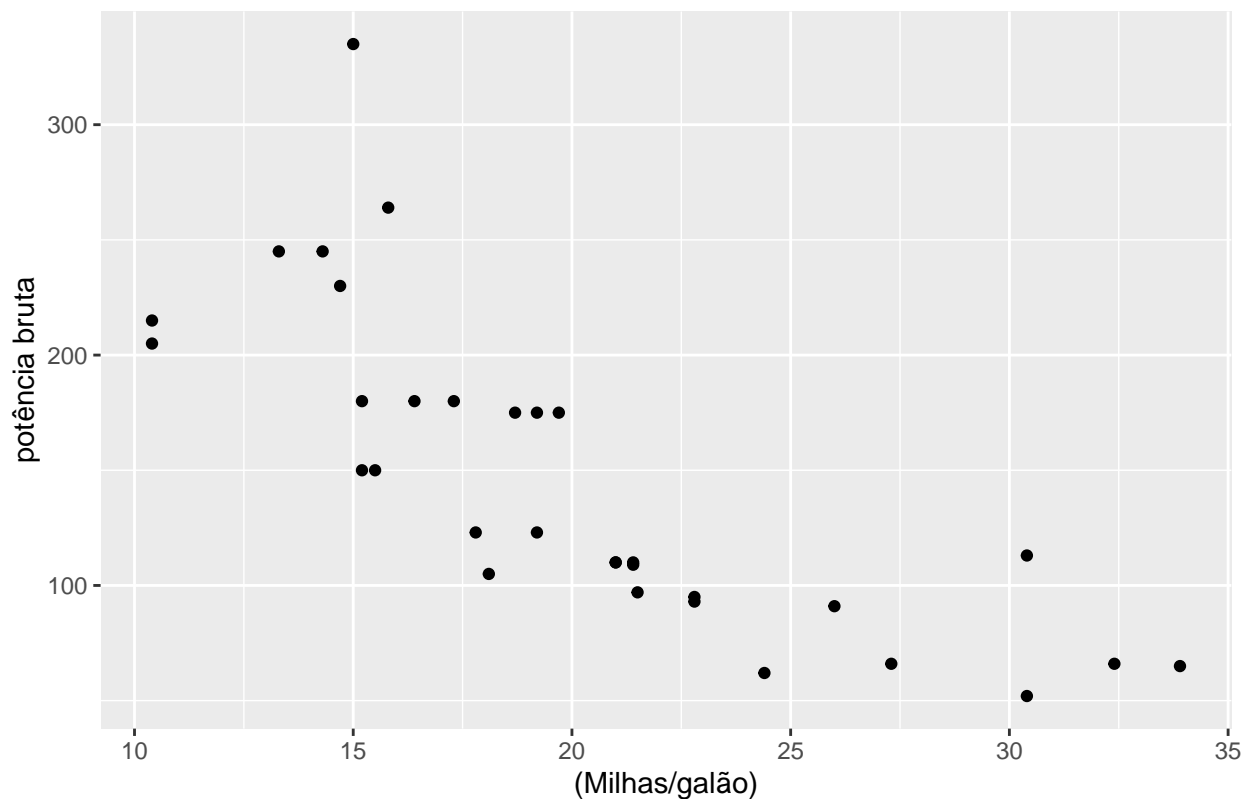
##	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
## Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
## Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
## Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
## Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
## Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
## Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
## Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
## Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
## Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
## Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
## Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
## Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
## Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
## Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
## Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
## Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
## Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1

## Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
## Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
## Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
## Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
## AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
## Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
## Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
## Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
## Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
## Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
## Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
## Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
## Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
## Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

2) A partir das duas (2) variáveis adotadas para análise:

- Desenvolva uma breve análise exploratória/descriptiva das mesmas;
- Desenvolva e interprete de forma prática uma análise de correlação.

Relação entre (Milhas/galão) e potência bruta



Como visto no gráfico de dispersão quanto menor a potencia do automovel menos potencia temos, mas para podermos validar essa correlação, faremos a correlação de pearson.

```
# Teste de hipotese sobre correlacao nula
cor.test(dados$mpg, dados$hp)
```

```
##
## Pearson's product-moment correlation
##
## data: dados$mpg and dados$hp
## t = -6.7424, df = 30, p-value = 1.788e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.8852686 -0.5860994
## sample estimates:
## cor
## -0.7761684
```

Dado o intervalo de confiança -0.8852686 -0.5860994, é demonstrado que não contém a correlação nula.

- c) Desenvolva e interprete de forma prática uma análise de regressão linear simples, incluindo a análise de resíduos e previsões para alguns valores estabelecidos para a variável independente, $X = x$.

Estimacao dos Parametros do Modelo de Regressao Linear Simples (MRLS)

Os coeficientes estimados

```
mod <- lm(hp ~ mpg, data=dados)
mod

##
## Call:
## lm(formula = hp ~ mpg, data = dados)
##
## Coefficients:
## (Intercept)      mpg
##      324.08      -8.83
```

inferencias

```
summary(mod)

##
## Call:
## lm(formula = hp ~ mpg, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -59.26 -28.93 -13.45  25.65 143.36
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   324.08     27.43   11.813 8.25e-13 ***
## mpg           -8.83      1.31   -6.742 1.79e-07 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.95 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
```

Desvio padrão relacionado ao valor 324.08 é -8.83

Normalidade dos resíduos:

```
shapiro.test(mod$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mod$residuals
## W = 0.89154, p-value = 0.003799
```

Outliers nos resíduos

```
summary(rstandard(mod))
```

```
##      Min.    1st Qu.    Median      Mean   3rd Qu.      Max.
## -1.372665 -0.669729 -0.317109  0.006667  0.605565  3.354610
```

Independencia dos resíduos (Durbin-Watson)

```
durbinWatsonTest(mod)
```

```
## lag Autocorrelation D-W Statistic p-value
## 1      0.1186843      1.736674    0.392
## Alternative hypothesis: rho != 0
```

Homocedasticidade (Breusch-Pagan)

```
bptest(mod)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  mod
## BP = 0.86843, df = 1, p-value = 0.3514
```

Análise do modelo

```
summary(mod)
```

```
##
## Call:
## lm(formula = hp ~ mpg, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -59.26 -28.93 -13.45  25.65 143.36
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   324.08      27.43   11.813 8.25e-13 ***
## mpg           -8.83       1.31   -6.742 1.79e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.95 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
```

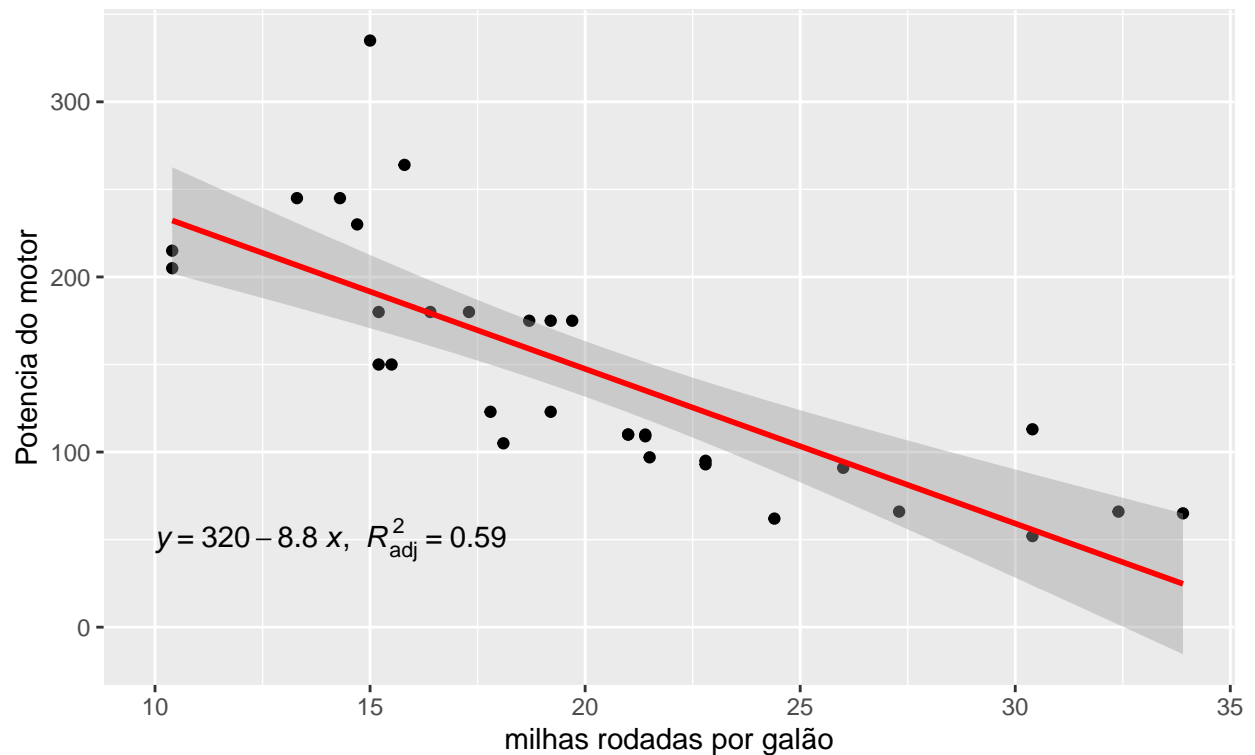
Apresentacao Grafica

```
ggplot(data = dados, mapping = aes(x=mpg, y=hp)) +
  geom_point() +
  geom_smooth(method = "lm", col = "red") +
  stat_regline_equation(aes(label = paste(..eq.label.., ..adj.rr.label..,
                                          sep = "*plain(\"\",\"\")~~\"")),
                        label.x = 10, label.y = 50) +
  labs(x='milhas rodadas por galão',y='Potencia do motor',
       title='Ajuste de um Modelo de Regressão Linear Simples',
       subtitle = 'Potencia do motor x Milhas rodadas por galão')
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

Ajuste de um Modelo de Regressão Linear Simples

Potencia do motor x Milhas rodadas por galão



Predição

```
df.teste <- data.frame(mpg = c(21))
df.teste
```

```
##   mpg
## 1   21
```

```
predict(mod, df.teste)
```

```
##      1
## 138.658
```

Prevendo varios valores:

```
df.teste <- data.frame(mpg = c(29,45,38))
df.teste
```

```
##   mpg
## 1   29
## 2   45
## 3   38
```

```
predict(mod, df.teste)
```

```
##      1      2      3
## 68.02012 -73.25558 -11.44746
```