R para Ciência de Dados

Regressão Linear Simples e Múltipla

Profa Carolina Paraíba e Prof Gilberto Sassi

Departamento de Estatística
Instituto de Matemática e Estatística
Universidade Federal da Bahia

Abril de 2024

Curso R para Ciência de Dados: Regressão Linear Simples e Múltipla

- Introdução
- Modelo de Regressão Linear Múltipla
- Estimação
- Testes de Hipóteses
- Multicolinearidade

Introdução

Nesta aula, passaremos do modelo de regressão linear simples, com um único preditor, para o modelo de regressão linear múltipla com dois ou mais preditores.

O adjetivo *simples* denotava que nosso modelo possuia apenas um preditor.

O adjetivo *múltipla* irá indicar que nosso modelo possui pelo menos dois preditores.

Em regressão múltipla, devido ao número potencialmente grande de preditores, é mais eficiente usar matrizes para definir o modelo de regressão e as análises subsequentes. Por esta razão, vamos considerar algumas das fórmulas de regressão múltipla mais importantes em forma matricial.

Introdução

A boa notícia é que tudo que aprendemos sobre o modelo de regressão linear simples se estende (com as modificações necessárias) ao modelo de regressão linear múltipla. Em particular, fazemos as mesmas suposições de linearidade e normalidade, homoscedasticidade e independência dos erros. A diferença é que enquanto na regressão linear simples pensamos na distribuição de erros em um valor fixo do preditor único, em regressão linear múltipla temos que pensar na distribuição dos erros em um conjunto fixo de valores para todos os preditores.

Todos os procedimentos de verificação de modelo (análise de resíduos e diagnóstico de influência) que aprendemos anteriormente são úteis na estrutura de regressão linear múltipla, embora o processo se torne um pouco mais complicado, pois agora teremos vários preditores.

Modelo de Regressão Linear Múltipla

O modelo de regressão linear múltipla é definido pela função linear

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon,$$

onde

- Y é a variável resposta;
- $x_1, \ldots, x_p$ são as covariáveis (variáveis explicativas, variáveis independentes ou variáveis regressoras);
- $\beta_0, \ldots, \beta_p$ são os parâmetros do modelo;
- ϵ é o erro aleatório do modelo.

Interpretação de β_j: é o efeito de x_j em Y quando as demais covariáveis são mantidas constantes.

Quando $p = 1$, temos o modelo de regressão linear simples.

Modelo de Regressão Linear Múltipla

Interesse: estimar os parâmetros $\beta_0, \ldots, \beta_p$.

Considere n observações $y_1, \ldots, y_n$ de Y.

Cada observação y_i está associada às observações de p variáveis explicativas, $x_{i1}, \ldots, x_{ip}$.

Dados observados: $(y_i, x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$.

Então, o modelo de regressão linear amostral é dado por

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, 2, \ldots, n \quad \text{e} \quad n > p.$$

Observação: Um modelo de regressão linear especifica uma relação funcional entre a variável resposta e as covariáveis que é linear nos parâmetros.

Modelo de Regressão Linear Múltipla

Suposições do modelo de regressão linear múltipla:

No modelo de regressão linear múltiplo usual, os ϵ_i's são variáveis aleatórias sujeitas às seguintes condições:

- $E(\epsilon_i) = 0$.
- $Var(\epsilon_i) = \sigma^2$.
- $Cov(\epsilon_i, \epsilon_j) = 0, \; i \neq j, \; j = 1, 2, \ldots, n$.

Modelo de Regressão Linear Múltipla

Os $x_{i1}, \ldots, x_{ip}$ são apenas números, ou seja, valores das respectivas covariáveis $x_1, \ldots, x_p$ que não são aleatórias.

Cada y_i depende da quantidade aleatória ϵ_i. Portanto, cada y_i é uma variável aleatória}}.

Então, o modelo de regressão linear múltipla descreve uma situação onde a média da distribuição de y_i, dado $x_{i1}, \ldots, x_{ip}$ conhecidos, é dada por

$$E(Y_i|x_{i1}, \ldots, x_{ip}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}.$$

Interpretação de β_j: é a mudança na resposta média de Y associada com a mudança de uma unidade em x_j quando as demais covariáveis são mantidas constantes.

A variância de y_i é

$$Var(Y_i|x_{i1}, \ldots, x_{ip}) = \sigma^2.$$

Modelo de Regressão Linear Múltipla

Modelo linear geral:

Em notação matricial, temos o modelo linear geral que é expresso da forma

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}}_{\boldsymbol{X}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}_{\boldsymbol{\epsilon}},$$

onde

- $\boldsymbol{y}$: vetor $n \times 1$ das variáveis $y_1, \dots, y_n$;
- $\boldsymbol{X}$: matriz $n \times (p+1)$;
- $\boldsymbol{\beta}$: vetor $(p+1) \times 1$ dos coeficientes de regressão;
- $\boldsymbol{\epsilon}$: vetor $n \times 1$ de erros aleatórios.

Modelo de Regressão Linear Múltipla

Reescrevendo as suposições do modelo em notação matricial, temos:

$$E(\boldsymbol{\epsilon}) = \mathbf{0} \quad \text{e } Cov(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n.$$

Consequentemente,

$$E(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{X}\beta,$$

e

$$Cov(\boldsymbol{y}|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n.$$

Observação: $\boldsymbol{I}_n$ é a matriz identidade de dimensão $n \times n$.

Estimação dos parâmetros via Método de Mínimos Quadrados:

Método de mínimos quadrados (MMQ): obtém os estimadores de $\boldsymbol{\beta}$ minimizando a soma dos quadrados entre as medições e o modelo. Isto é, o método consiste em minimizar a soma de quadrados dos erros com respeito a $\boldsymbol{\beta}$.

A soma de quadrados dos erros (SQE) é definida por

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \epsilon_i^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}.$$

Usando o MMQ, a estimativa de $\boldsymbol{\beta}$ é

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

Estimação

Modelo ajustado: o modelo de regressão múltipla ajustado (vetor de valores ajustados ou vetor de valores preditos) correspondente ao vetor de valores observados $\boldsymbol{y}$ é dado por

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\beta} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{H}\boldsymbol{y}.$$

A matriz $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ de dimensão $n \times n$ é geralmente chamada de matriz Chapéu.

Resíduos: diferença entre o valor observado $\boldsymbol{y}$ o seu valor ajustado $\hat{\boldsymbol{y}}$,

$$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}.$$

Estimação

Estimação de σ^2:

Soma de quadrado de resíduos (SQR):

$$SQR = \boldsymbol{e}'\boldsymbol{e} = \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}.$$

Quadrado médio dos resíduos (QMR):

$$QMR = \frac{SQR}{n - p - 1}.$$

Um estimador não viciado de σ^2 é definido por $S^2 = QMR$.

Testes de Hipóteses

Em regressão linear, é importante avaliar se existe uma boa "relação" entre a resposta e as variáveis explicativas.

Para responder esta questão, podemos utilizar o procedimento de testes de hipóteses para os parâmetros.

Ao realizar testes de hipóteses em regressão linear, fazemos a suposição de que os erros ϵ_i's são independentes e identicamente distribuídos de acordo com a distribuição $N(0, \sigma^2)$.

Isto é,

$$\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n).$$

$$\boldsymbol{y} \sim N_n(\boldsymbol{X}\beta, \sigma^2 \boldsymbol{I}_n).$$

Testes de Hipóteses

Teste de significância do modelo:

Teste para determinar se existe uma relação linear entre a variável resposta Y e quaisquer das covariáveis $x_1, \ldots, x_p$.

É visto como um teste geral ou global de adequação do modelo.

A rejeição da hipótese nula implica que pelo menos uma das covariáveis contribui significativamente para o modelo.

Suponha que desejamos testar

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0 \\ H_1 : \beta_j \neq 0, \text{ para pelo menos um } j, \end{cases}$$

Testes de Hipóteses

Teste de significância do modelo:

A estatística de teste é

$$F = \frac{\frac{SQM}{p}}{\frac{SQR}{n-p-1}} = \frac{QMM}{QMR}.$$

Considerando um nível de significância α, devemos rejeitar H_0 se $F > F^{\alpha}_{p,n-p-1}$.

Analogamente, usando o p-valor do teste de hipóteses com nível de significância α pré-fixado, rejeitamos H_0 se p-valor do teste for menor do que α.

Testes de Hipóteses

Análise de variância (ANOVA):

Uma vez obtido o modelo estimado, estamos interessados em saber qual é a proporção da variabilidade de $\boldsymbol{y}$ que está sendo explicada pelas covariáveis. Esta variabilidade por ser particionada como

$$SQT = SQM + SQR,$$

onde

- $SQT = \boldsymbol{y}^{'}\boldsymbol{y} - \hat{\beta}^{'}\boldsymbol{X}^{'}\boldsymbol{y}$: soma de quadrados total (a variabilidade total em y);
- $SQM = \hat{\beta}^{'}\boldsymbol{X}^{'}\boldsymbol{y} - n\overline{Y}^{2}$: soma de quadrados da regressão (a variabilidade em $\boldsymbol{y}$ induzida pelas covariáveis);
- $SQT = \boldsymbol{y}^{'}\boldsymbol{y} - n\overline{Y}^{2}$: soma de quadrados residual (a variabilidade em $\boldsymbol{y}$ não induzida pelas covariáveis).

Testes de Hipóteses

Tabela ANOVA:

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	F_0
Modelo	SQM	p	QMM	$\dfrac{QMM}{QMR}$
Resíduo	SQR	$n-p-1$	QMR	
Total	SQT	$n-1$		

Testes de Hipóteses

Uso da ANOVA para comparar dois modelos aninhados:

Em termos práticos, quando utilizamos desejamos testar um modelo restrito contra um modelo completo, podemos utilizar a análise de variância para testar:

$$\begin{cases} H_0: & \text{modelo reduzido,} \\ H_1: & \text{modelo completo.} \end{cases}$$

Esse teste corresponde ao teste F derivado anteriormente. Assim, valores grande da estatística F fornecem evidências contra H_0.

Testes de Hipóteses

Teste individual para cada parâmetro:

Se o teste F para $H_0 : \boldsymbol{\beta} = \mathbf{0}$ é significativo, então faz sentido verificar porque H_0 foi rejeitada.

Para isto, podemos testar cada restrição $\beta_j = 0$, $j = 1, \ldots, p$, separadamente.

Considere, então, a hipótese nula

$$H_0 : \beta_j = 0 \quad vs. \quad H_1 : \beta_j \neq 0.$$

Para este caso, podemos definir a estatística de teste T, que é dada por

$$T = \frac{\hat{\beta}_j}{\sqrt{S^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{jj}}} \sim t_{n-p-1}.$$

Rejeitamos H_0 a um nível de significância $\alpha \in (0, 1)$ se $|T| \geq t^{\alpha/2}_{n-p-1}$.

Multicolinearidade

Multicolinearidade:

A multicolinearidade ocorre quando dois ou mais preditores em um modelo de regressão são moderada ou altamente correlacionados.

A multicolinearidade pode prejudicar nossa análise e, assim, limitar as conclusões de pesquisa que podemos tirar.

Multicolinearidade

Consequências de multicolinearidade entre as covariáveis são:

- a estimativa do coeficiente de qualquer covariável depende de quais outras covariáveis estão incluídas no modelo;
- a precisão dos coeficientes estimados diminui à medida que mais preditores são adicionados ao modelo;
- a contribuição marginal de qualquer covariável na redução da soma dos quadrados dos resíduos depende de quais outros preditoras já estão no modelo;
- testes de hipóteses podem produzir conclusões diferentes dependendo de quais preditoras estão no modelo.

Multicolinearidade

Tipos de multicolinearidade:

- Multicolinearidade estrutural:
  - causada pelo criação de novas covariáveis a partir de outras covariáveis. Por exemplo, quando criamos a preditora x^2 usando a preditora x.
- Multicolinearidade baseada em dados:
  - resulta de um experimento mal projetado, da dependência de dados puramente observacionais ou da incapacidade de manipular o sistema no qual os dados são coletados.

Multicolinearidade

Detecção de multicolinearidade:

- As estimativas dos coeficientes variam de modelo para modelo.
- O teste-t para cada uma dos parâmetros não são significativos mas o teste-F de significância do modelo é significativo.
- As correlações entre pares de variáveis preditoras são moderadas/altas.

Multicolinearidade

Fator de inflação da variância (VIF, em inglês): métrica que avalia o quanto a variância de um coeficiente de regressão estimado aumenta se as preditoras estiverem correlacionadas. Definida por:

$$VIF_j = \frac{1}{1 - R_j^2}, \quad j = 1, \ldots, p.$$

onde R_j^2 é o valor do coeficiente de determinação obtido da regressão da j-ésima covariável nas demais preditoras.

- Se $VIF_j = 1$, não há correlação entre a j-ésima preditora e as demais covariáveis.
- Regra geral: VIF's superiores a 4 justificam uma investigação mais aprofundada, enquanto VIF's superiores a 10 são sinais de multicolinearidade grave que requer correção.

Modelo de Regressão Linear Múltipla no R

```
df_mtcarros <- read_xlsx("../dados/mtcarros.xlsx")

df_rlm_mtcarros <- df_mtcarros[,c(1,3,4,5,6,7)]
glimpse(df_rlm_mtcarros)
```

```
## Rows: 32
## Columns: 6
## $ mpg  <dbl> 21.0, 21.0, 22.8, 21.4, 18.7, 18.1,
## $ disp <dbl> 160.0, 160.0, 108.0, 258.0, 360.0,
## $ hp   <dbl> 110, 110, 93, 110, 175, 105, 245, 6
## $ drat <dbl> 3.90, 3.90, 3.85, 3.08, 3.15, 2.76,
## $ wt   <dbl> 2.620, 2.875, 2.320, 3.215, 3.440,
## $ qsec <dbl> 16.46, 17.02, 18.61, 19.44, 17.02,
```

Modelo de Regressão Linear Múltipla no R

```
fit <- lm(mpg ~ ., data = df_rlm_mtcarros)

fit0 <- lm(mpg ~ 1, data = df_rlm_mtcarros)

anova(fit0, fit)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ 1
## Model 2: mpg ~ disp + hp + drat + wt + qsec
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     31 1126.05
## 2     26  170.13  5    955.92 29.218 6.892e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
```

Modelo de Regressão Linear Múltipla no R

```r
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq  F value   Pr(>F)
## disp       1 808.89  808.89 123.6185 2.23e-11 ***
## hp         1  33.67   33.67   5.1449 0.031854 *
## drat       1  30.15   30.15   4.6073 0.041340 *
## wt         1  70.51   70.51  10.7754 0.002933 **
## qsec       1  12.71   12.71   1.9422 0.175233
## Residuals 26 170.13    6.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
```

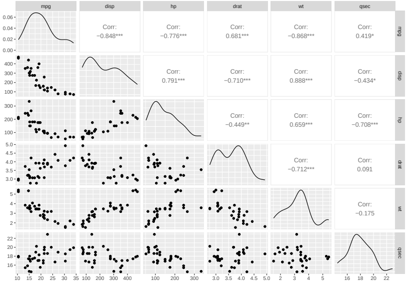
Modelo de Regressão Linear Múltipla no R

```r
tidy(fit)
```

```
## # A tibble: 6 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept) 16.5       11.0         1.51  0.144
## 2 disp         0.00872    0.0112      0.779 0.443
## 3 hp          -0.0206     0.0153     -1.35  0.189
## 4 drat         2.02       1.31        1.54  0.136
## 5 wt          -4.39       1.24       -3.53  0.00158
## 6 qsec         0.640      0.459       1.39  0.175
```

Modelo de Regressão Linear Múltipla no R

```r
ggpairs(df_rlm_mtcarros)
```



Modelo de Regressão Linear Múltipla no R

```r
library(olsrr)

ols_vif_tol(fit)
```

```
##   Variables Tolerance      VIF
## 1      disp 0.1097590 9.110869
## 2        hp 0.1922399 5.201833
## 3      drat 0.4305997 2.322343
## 4        wt 0.1425987 7.012686
## 5      qsec 0.3132892 3.191939
```

Modelo de Regressão Linear Múltipla no R

```r
fit1 <- lm(mpg ~ hp + drat + wt + qsec,
           data = df_rlm_mtcarros)

anova(fit1, fit)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ hp + drat + wt + qsec
## Model 2: mpg ~ disp + hp + drat + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     27 174.10
## 2     26 170.13  1    3.9744 0.6074 0.4428
```

Modelo de Regressão Linear Múltipla no R

```r
ols_vif_tol(fit1)
```

```
##   Variables Tolerance      VIF
## 1        hp 0.2031712 4.921958
## 2      drat 0.4912864 2.035473
## 3        wt 0.2791205 3.582683
## 4      qsec 0.3476912 2.876115
```

Modelo de Regressão Linear Múltipla no R

```r
fit2 <- lm(mpg ~ drat + wt + qsec,
           data = df_rlm_mtcarros)

anova(fit2, fit1)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ drat + wt + qsec
## Model 2: mpg ~ hp + drat + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 183.52
## 2     27 174.10  1    9.4181 1.4606 0.2373
```

Modelo de Regressão Linear Múltipla no R

```r
ols_vif_tol(fit2)
```

```
##   Variables Tolerance      VIF
## 1      drat 0.4912866 2.035472
## 2        wt 0.4802849 2.082097
## 3      qsec 0.9672265 1.033884
```

```r
fit3 <- lm(mpg ~ wt + qsec, data = df_rlm_mtcarros)

anova(fit3, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + qsec
## Model 2: mpg ~ drat + wt + qsec
##   Res.Df    RSS Df Sum of Sq     F Pr(>F)
## 1     29 195.46
## 2     28 183.52  1    11.942 1.822 0.1879
```

Modelo de Regressão Linear Múltipla no R

```r
fit4 <- lm(mpg ~ qsec, data = df_rlm_mtcarros)

anova(fit4, fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ qsec
## Model 2: mpg ~ wt + qsec
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 928.66
## 2     29 195.46  1    733.19 108.78 2.519e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
```

Modelo de Regressão Linear Múltipla no R

```
fit_step <- stepAIC(fit)
```

```
## Start:  AIC=65.47
## mpg ~ disp + hp + drat + wt + qsec
##
##        Df Sum of Sq    RSS    AIC
## - disp  1     3.974 174.10 64.205
## <none>              170.13 65.466
## - hp    1    11.886 182.01 65.627
## - qsec  1    12.708 182.84 65.772
## - drat  1    15.506 185.63 66.258
## - wt    1    81.394 251.52 75.978
##
## Step:  AIC=64.21
## mpg ~ hp + drat + wt + qsec
##
##        Df Sum of Sq    RSS    AIC
## - hp    1     9.418 183.52 63.891
## - qsec  1     9.578 183.68 63.919
## <none>              174.10 64.205
## - drat  1    11.956 186.06 64.331
## - wt    1   113.882 287.99 78.310
##
## Step:  AIC=63.89
## mpg ~ drat + wt + qsec
##
##        Df Sum of Sq    RSS    AIC
## <none>              183.52 63.891
## - drat  1    11.942 195.46 63.908
## - qsec  1    85.720 269.24 74.156
## - wt    1   275.686 459.21 91.241
```

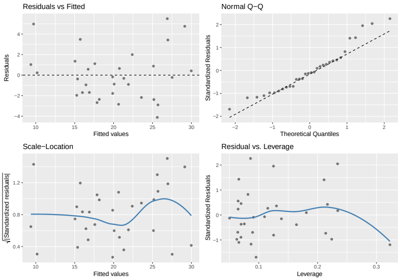
Modelo de Regressão Linear Múltipla no R

```r
summary(fit_step)
```

```
##
## Call:
## lm(formula = mpg ~ drat + wt + qsec, data = df_rlm_mtcarros)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1152 -1.8273 -0.2696  1.0502  5.5010
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.3945     8.0689   1.412  0.16892
## drat          1.6561     1.2269   1.350  0.18789
## wt           -4.3978     0.6781  -6.485 5.01e-07 ***
## qsec          0.9462     0.2616   3.616  0.00116 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.56 on 28 degrees of freedom
## Multiple R-squared:  0.837,  Adjusted R-squared:  0.8196
## F-statistic: 47.93 on 3 and 28 DF,  p-value: 3.723e-11
```

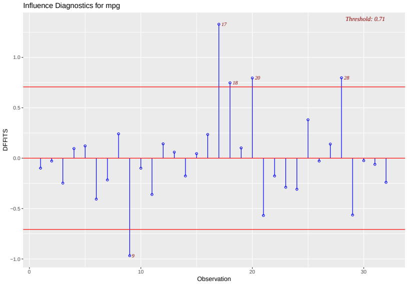
Modelo de Regressão Linear Múltipla no R

```r
fit <- lm(mpg ~ drat + wt + qsec,
          data = df_rlm_mtcarros)

df_fit_resid <- augment(fit)

gglm(fit)
```



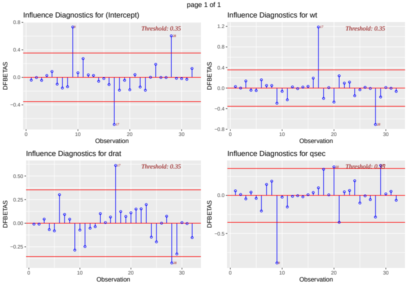
Modelo de Regressão Linear Múltipla no R

```
ols_plot_dffits(fit)
```



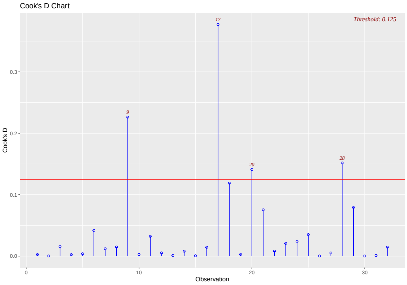
Modelo de Regressão Linear Múltipla no R

```
ols_plot_dfbetas(fit)
```



Modelo de Regressão Linear Múltipla no R

```
ols_plot_cooksd_chart(fit)
```



Modelo de Regressão Linear Múltipla no R

```r
shapiro.test(fit$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit$residuals
## W = 0.94389, p-value = 0.09659
```

Modelo de Regressão Linear Múltipla no R

```r
ks.test(df_fit_resid$.std.resid, "pnorm")
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  df_fit_resid$.std.resid
## D = 0.10604, p-value = 0.8276
## alternative hypothesis: two-sided
```

Modelo de Regressão Linear Múltipla no R

```r
ad.test(fit$residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  fit$residuals
## A = 0.59522, p-value = 0.1125
```

Modelo de Regressão Linear Múltipla no R

```r
bptest(fit)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  fit
## BP = 3.6078, df = 3, p-value = 0.307
```

Modelo de Regressão Linear Múltipla no R

```r
dwtest(fit)
```

```
##
##  Durbin-Watson test
##
## data:  fit
## DW = 1.8587, p-value = 0.2564
## alternative hypothesis: true autocorrelation is great
```