

# Lista 7 [AD-UFPE-2019]

*Antonio Fernandes*

*28 de maio de 2019*

## Conteúdo

<b>Apresentação</b>	<b>1</b>
Questão 1 . . . . .	2
a) Análise descritiva de todas as variáveis da base de dados . . . . .	2
b) Modelo de regressão . . . . .	9
Resultado da regressão . . . . .	10
Capacidade explicativa do modelo . . . . .	10
Ajuste do modelo . . . . .	10
c) Modelo de regressão com mais de uma VI . . . . .	13
Resultados do modelo . . . . .	14
Capacidade explicativa do modelo . . . . .	14
RMSE . . . . .	14
Comparando o resultados dos modelos . . . . .	14
Ajuste do modelo . . . . .	15
d) Modelo de regressão com a VI War . . . . .	17
Resultados do modelo . . . . .	18
Capacidade explicativa do modelo . . . . .	18
RMSE . . . . .	18
Comparando o resultados dos modelos . . . . .	19
Ajuste do modelo . . . . .	20
Questão 2 . . . . .	22
Citação do artigo . . . . .	22
Modelo . . . . .	22

## Apresentação

Este documento apresenta as respostas da lista de exercícios 7 da disciplina de Análise de dados.

O link está disponível no GitHub: [https://github.com/alvesat/AD\\_7](https://github.com/alvesat/AD_7)

A lista envolve a execução de diversos modelos de regressão linear envolvendo um banco de dados específicos.

## Questão 1

O primeiro passo para responder a questão é abrir o banco de dados contendo as informações necessárias. É importante verificar que os dados estão no formato *.dta*, que é o formato relacionado ao *software* Stata. Devido a isso, teremos que abrir o pacote **haven** e executar o comando para abrir o banco.

```
library(haven)
```

```
fair <- read_dta("~/Dados/Listas/AD_7/DATA_L7/fair.dta")
```

### a) Análise descritiva de todas as variáveis da base de dados

Vamos verificar o nome das variáveis presentes no banco

```
names(fair)
```

```
## [1] "YEAR"      "VOTE"       "PARTY"      "PERSON"     "DURATION"   "WAR"
## [7] "GROWTH"    "INFLATION" "GOODNEWS"
```

Podemos observar que o banco apresenta 9 variáveis: *Year*, *Vote*, *Party*, *Person*, *Duration*, *War*, *Growth*, *Inflation* e *Good news*.

Um outro passo envolvendo o processo de descrição das variáveis é identificar a estrutura de cada variável:

```
str(fair)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   32 obs. of  9 variables:
## $ YEAR      : num  1880 1884 1888 1892 1896 ...
## ..- attr(*, "format.stata")= chr "%8.0g"
## $ VOTE       : num   50.2 49.8 50.4 48.3 47.8 ...
## ..- attr(*, "format.stata")= chr "%9.0g"
## $ PARTY      : num   -1 -1 1 -1 1 -1 -1 -1 -1 1 ...
## ..- attr(*, "format.stata")= chr "%8.0g"
## $ PERSON     : num    0 0 1 1 0 1 0 0 1 1 ...
## ..- attr(*, "format.stata")= chr "%8.0g"
## $ DURATION   : num    1.75 2 0 0 0 0 1 1.25 1.5 0 ...
## ..- attr(*, "format.stata")= chr "%9.0g"
## $ WAR        : num    0 0 0 0 0 0 0 0 0 0 ...
## ..- attr(*, "format.stata")= chr "%8.0g"
## $ GROWTH     : num    3.88 1.59 -5.55 2.76 -10.02 ...
## ..- attr(*, "format.stata")= chr "%9.0g"
## $ INFLATION  : num    1.974 1.055 0.604 2.274 3.41 ...
## ..- attr(*, "format.stata")= chr "%9.0g"
## $ GOODNEWS   : num    9 2 3 7 6 7 5 8 8 3 ...
## ..- attr(*, "format.stata")= chr "%8.0g"
```

É possível identificar que todas as variáveis presentes no banco são numéricas e que o banco possui 32 observações. Agora vamos fazer uma análise descritiva de cada variável:

## Year

A variável ano apresenta os anos de cada observação. Por meio do comando `fivenum`, podemos obter um sumário da variável (valor mínimo, 1º quartil, mediana, 3º quartil e valor máximo). No caso dessa variável, o importante é verificarmos o valor inicial e final.

```
fivenum(fair$YEAR)
```

```
## [1] 1880 1910 1942 1974 2004
```

Em relação a variável Year, percebemos que a primeira observação é em 1880 e a última é em 2004.

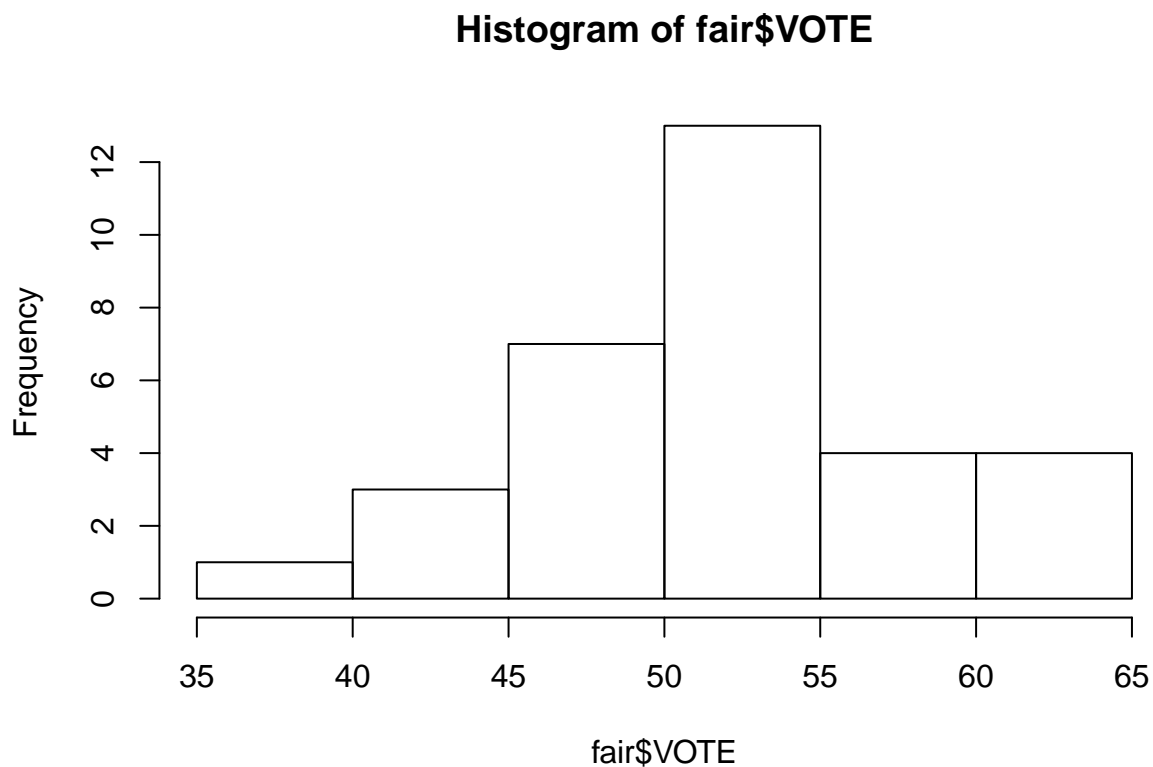
## Vote

```
fivenum(fair$VOTE)
```

```
## [1] 36.1190 49.2720 52.0260 56.3815 62.4580
```

Em relação a variável vote, ela representa a porcentagem de votos recebidas pelo partido do incumbente. Observando os valores obtidos do comando `fivenum`, o valor mínimo da variável foi de 36.12, o 1º quartil é de 49.27, a mediana é de 52.03, o 3º quartil é de 56.38 e o valor máximo é de 62.46. Podemos analisar a variável vote por meio de um histograma:

```
hist(fair$VOTE)
```



Com o histograma, verificamos que os valores estão mais concentrados entre 45 e 55, ou seja, no centro da distribuição.

## Party

A variável party é dicotômica, apresentando como valores -1 e 1.

```
fable(fair$PARTY)
```

```
##  -1  1  
##  
##  18 14
```

Com o comando `fable`, podemos fazer uma tabela de frequência da variável party e identificar que das 32 observações do banco, 18 são do valor -1 e 14 são do valor 1.

## Person

A variável person também é dicotômica, apresentando como valores 1 e 0.

```
fable(fair$PERSON)
```

```
##    0  1  
##  
##   13 19
```

O resultado mostra que 13 observações apresentam o valor 0 e 19 observações apresentam o valor 1.

## Duration

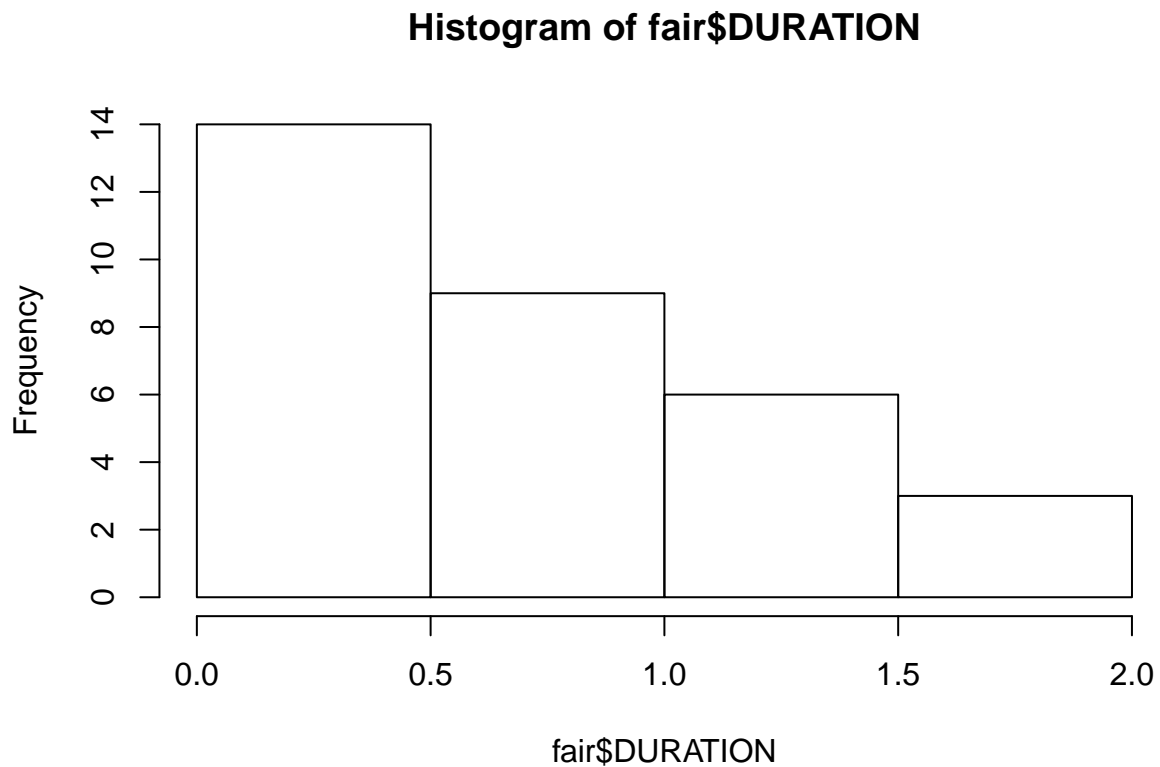
A variável duration apresenta os valores de 2 a 0. Com o comando `fivenum` é possível obter uma análise descritiva da variável:

```
fivenum(fair$DURATION)
```

```
## [1] 0.00 0.00 1.00 1.25 2.00
```

O valor mínimo e do primeiro quartil é 0, a mediana é 1, o 3º quartil é 1.25 e o valor máximo da distribuição é 2.

```
hist(fair$DURATION)
```



Pelo histograma, vemos que a maior parte das observções são 0 e 1 (25)

#### War

A variável war é uma variável dicotômica que é melhor observada por meio do comando `fTable`:

```
fTable(fair$WAR)
```

```
##    0    1
##
## 29    3
```

é possível verificar que 29 casos possuem valor 0 e 3 casos apresentam valor 1

#### Growth

A variável growth contém informações acerca da variação de crescimento do PIB por ano:

```
fivenum(fair$GROWTH)
```

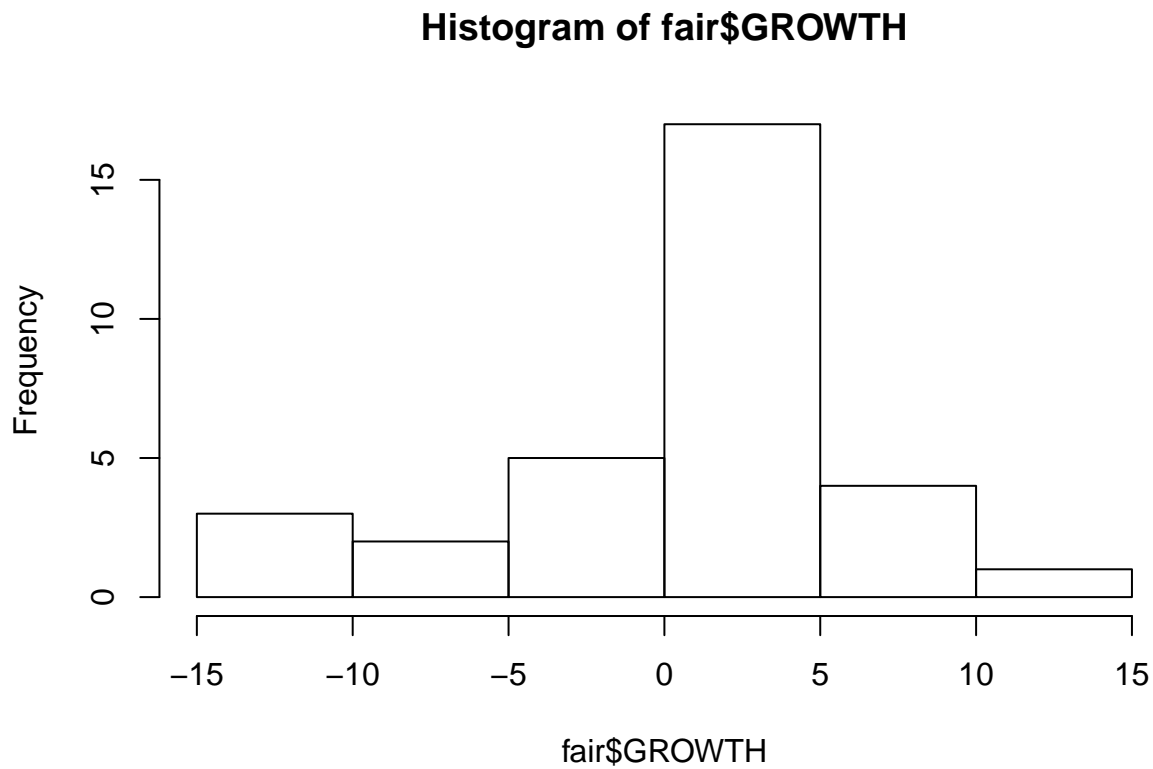
```
## [1] -14.557 -1.923  2.245  4.095 11.677
```

```
mean(fair$GROWTH)
```

```
## [1] 0.628
```

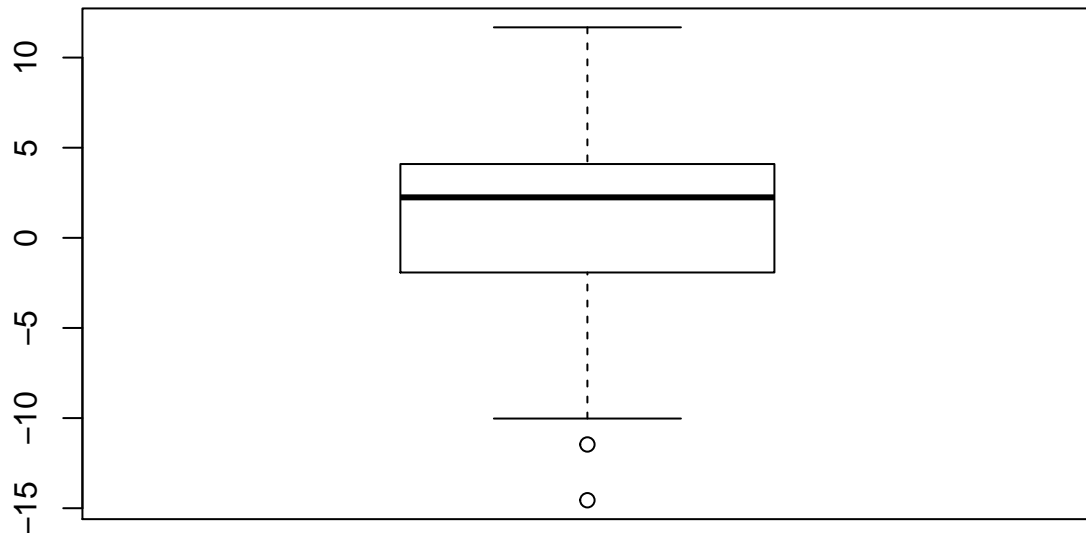
A descrição por meio do comando `fivenum` mostra que o valor mínimo da distribuição foi uma contração de 14.56 no PIB, enquanto que o 1º quartil apresenta um valor de -1.92, a mediana é de 2.24, o 3º quartil é 4.1 e o valor máximo da distribuição é de 11.68. Por meio de um histograma é possível uma melhor visualização da distribuição. A média de variação do crescimento no período foi de 0.628

```
hist(fair$GROWTH)
```



Com o histograma, percebemos que boa parte da distribuição está localizada no centro, com casos entre -5 e 5. Com um box plot, podemos identificar a presença ou não de outliers na distribuição.

```
boxplot(fair$GROWTH)
```



Por meio do Boxplot, podemos verificar que existem dois valores que são outliers: -14.55 e -11.46. Ambos resultados de crises econômicas (1920 e 1929)

## Inflation

Vamos verificar a descrição da variável inflation:

```
fivenum(fair$INFLATION)
```

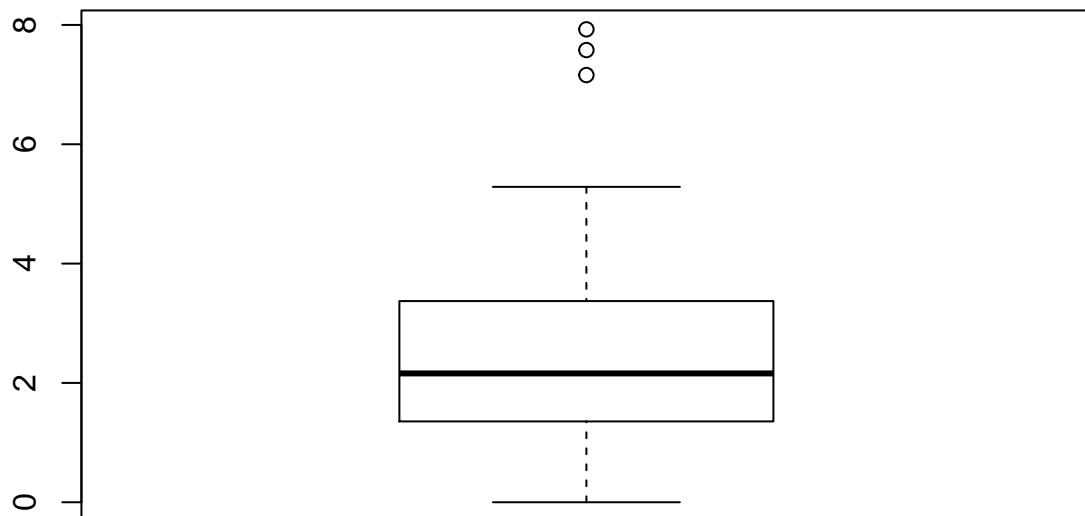
```
## [1] 0.0000 1.3545 2.1590 3.3715 7.9260
```

```
mean(fair$INFLATION)
```

```
## [1] 2.656812
```

O valor mínimo da distribuição é 0, o 1º quartil é 1.35, a mediana é de 2.16, o 3º quartil é de 3.38 e o valor máximo é de 7.93. A média no período foi uma inflação de 2.65.

```
boxplot(fair$INFLATION)
```



O boxplot da variável mostram 3 outliers: 7.92 (em 1980), 7.58 (em 1976) e 7.16 (em 1932)

### Goodnews

Abaixo vemos os cinco valores da variável goodnews, que representa a quantidade de boas notícias no campo da economia (no ano observado)

```
fivenum(fair$GOODNEWS)
```

```
## [1] 0.0 3.5 5.0 7.5 10.0
```

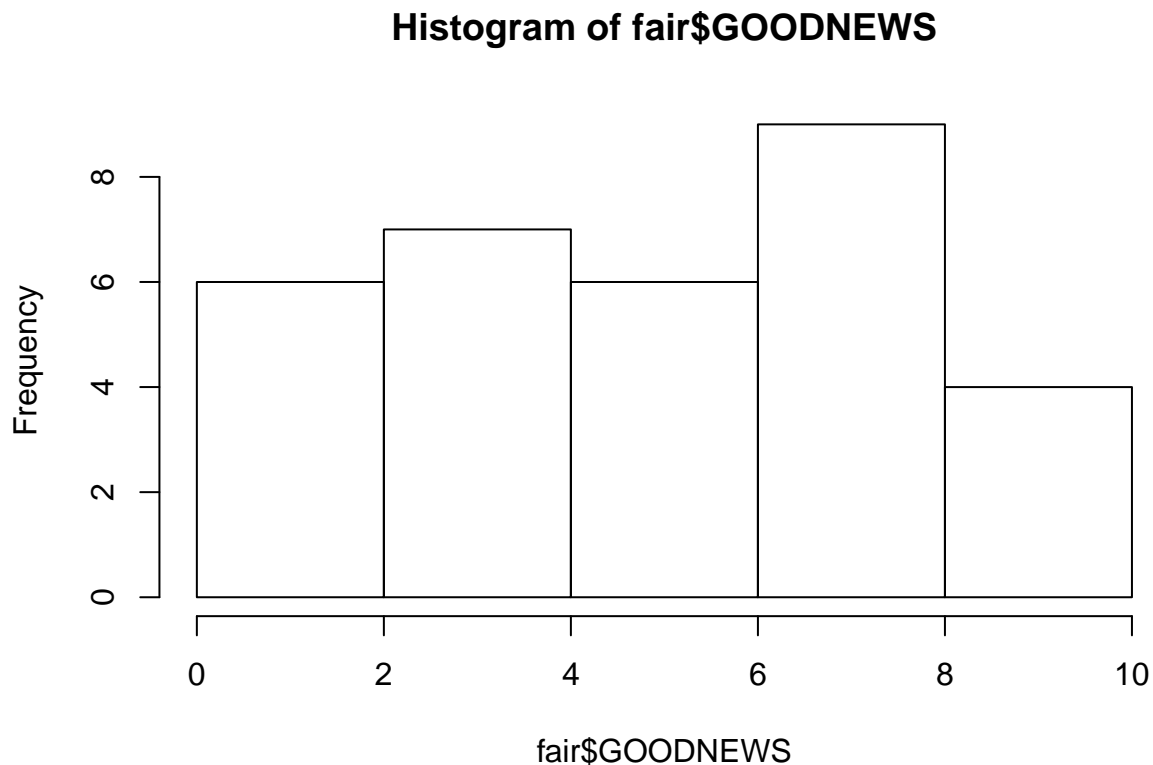
```
mean(fair$GOODNEWS)
```

```
## [1] 5.28125
```

O valor mínimo da variável é 0, o 1º quartil é 3.5, a mediana é 5, o 3º quartil é 7.5 e o valor máximo é 10. A média é de 5.28.

```
hist(fair$GOODNEWS)
```





O histograma mostra que a distribuição dos valores é bem balanceada.

#### b) Modelo de regressão

Para executar o modelo, com a variável Vote como VD e Growth como VI, vamos utilizar o seguinte comando:

```
Linear <- lm(VOTE ~ GROWTH, data = fair)
summary(Linear)
```

```
##
## Call:
## lm(formula = VOTE ~ GROWTH, data = fair)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.2487 -3.3330 -0.4282  3.1425  9.7286
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   51.8598     0.8817  58.821  < 2e-16 ***
## GROWTH         0.6536     0.1607   4.068 0.000316 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.955 on 30 degrees of freedom
## Multiple R-squared:  0.3555, Adjusted R-squared:  0.3341
```

```
## F-statistic: 16.55 on 1 and 30 DF,  p-value: 0.0003165
```

## Resultado da regressão

Como é possível observar no resultado, a variação de uma unidade na VI (Growth) leva a um aumento de 0.65 na VD (Votes). Nesse caso, o modelo de regressão utilizado é bivariado, dado a presença de apenas duas variáveis no modelo.

Em relação aos resíduos do modelo, o valor mínimo é de -8.25, o 1º quartil é de -3.33, a mediana é de -0.423, o 3º quartil é de 3.14 e o valor máximo é de 9.72.

Quando a VI assume o valor de 0, espera-se que o valor da VD seja de 51.86 (Valor do intercepto). Já em relação a VI, além do valor do coeficiente, o erro padrão foi de 0.16 e o teste-f que testa a hipótese nula de que não há relação entre as variáveis apresentou um p-valor menor que 0.001. Ou seja, rejeita-se a hipótese nula de que não há relação entre as variáveis.

## Capacidade explicativa do modelo

No que se refere a capacidade explicativa do modelo, o  $R^2$  ajustado foi de 0.33, significando que o modelo consegue explicar 33% da variância da variável dependente. O erro padrão do resíduo foi de 4.955.

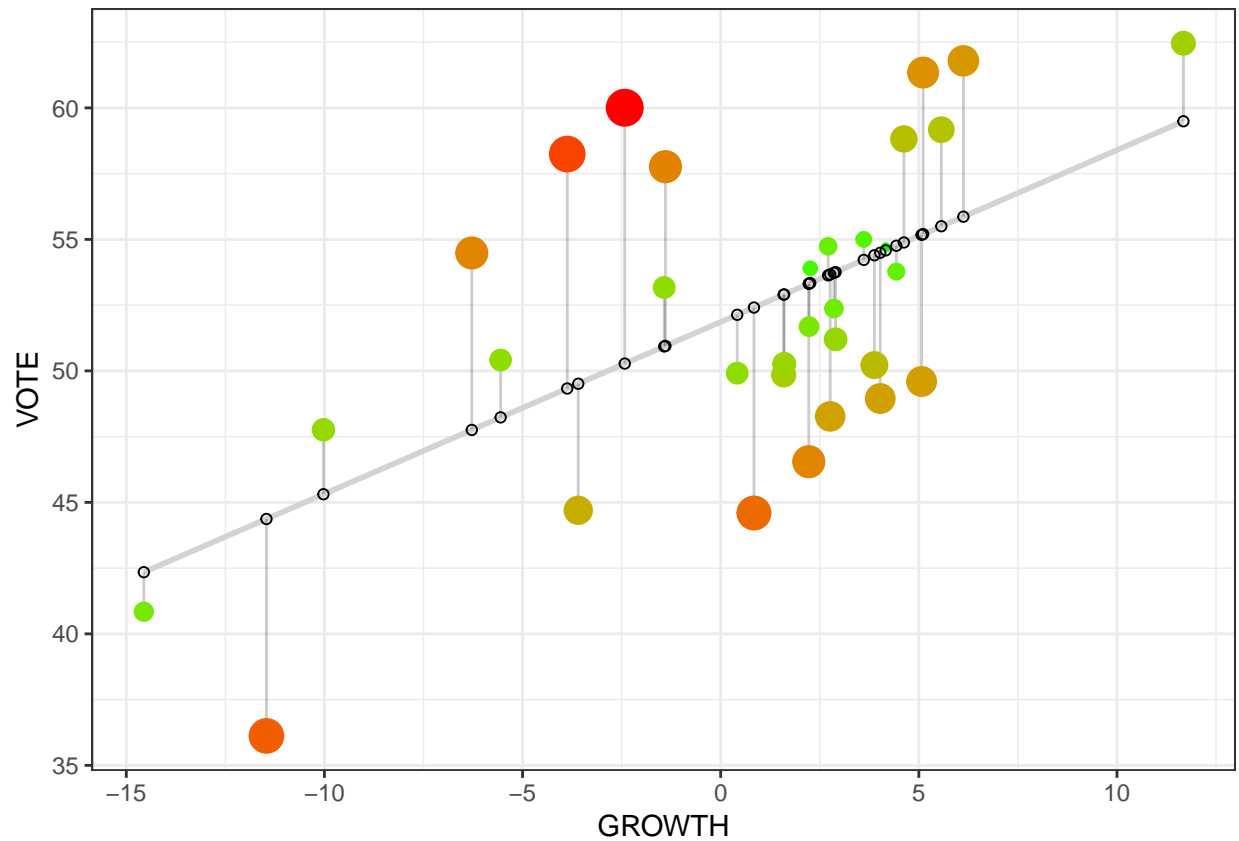
## Ajuste do modelo

Já no tocante ao ajuste, podemos analisar alguns gráficos e verificar a adequabilidade do modelo:

```
library(ggplot2)
```

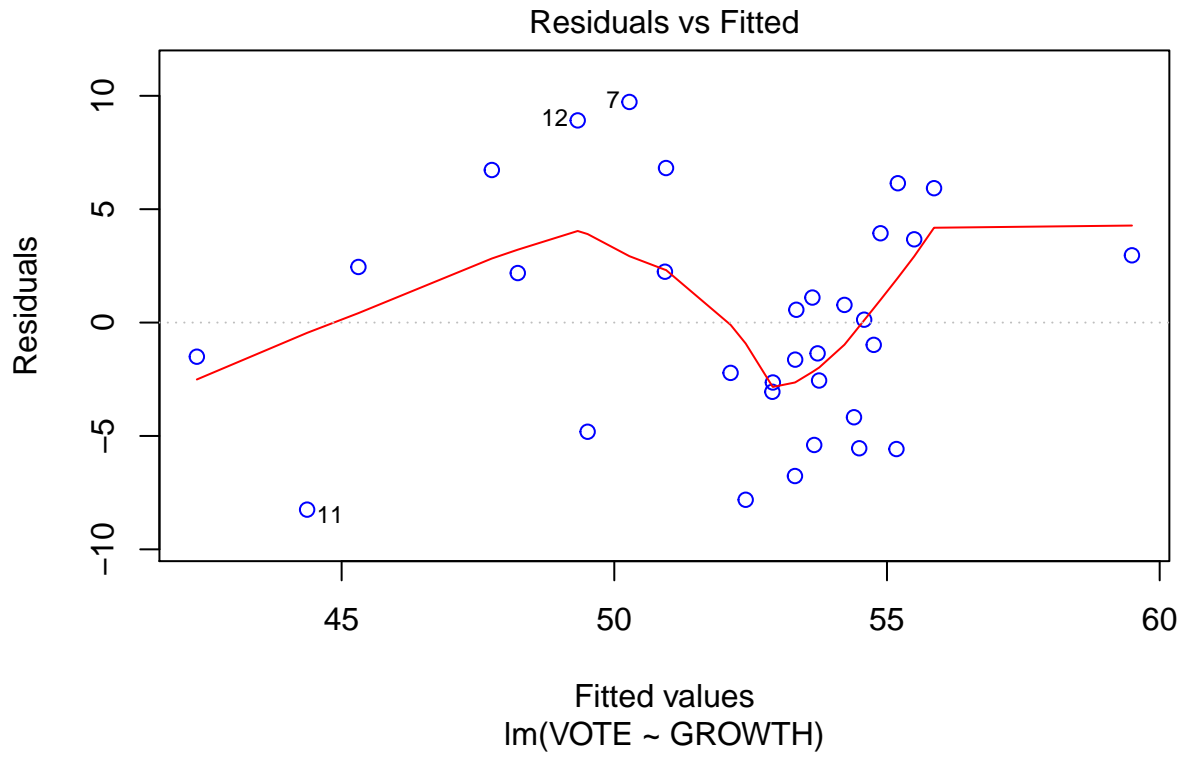
O gráfico abaixo apresenta o tamanho do resíduo por meio do tamanho do ponto e cor (quanto mais vermelho maior e quanto mais verde menor o resíduo). O tamanho do resíduo é a distancia entre o ponto e a linha de regressão.

```
fair$predicted <- predict(Linear)
fair$residuals <- residuals(Linear)
ggplot(fair, aes(x = GROWTH, y = VOTE)) +
  geom_smooth(method = "lm", se = FALSE, color = "lightgrey") +
  geom_segment(aes(xend = GROWTH, yend = predicted), alpha = .2) +
  geom_point(aes(color = abs(residuals), size = abs(residuals))) +
  scale_color_continuous(low = "green", high = "red") +
  guides(color = FALSE, size = FALSE) +
  geom_point(aes(y = predicted), shape = 1) +
  theme_bw()
```



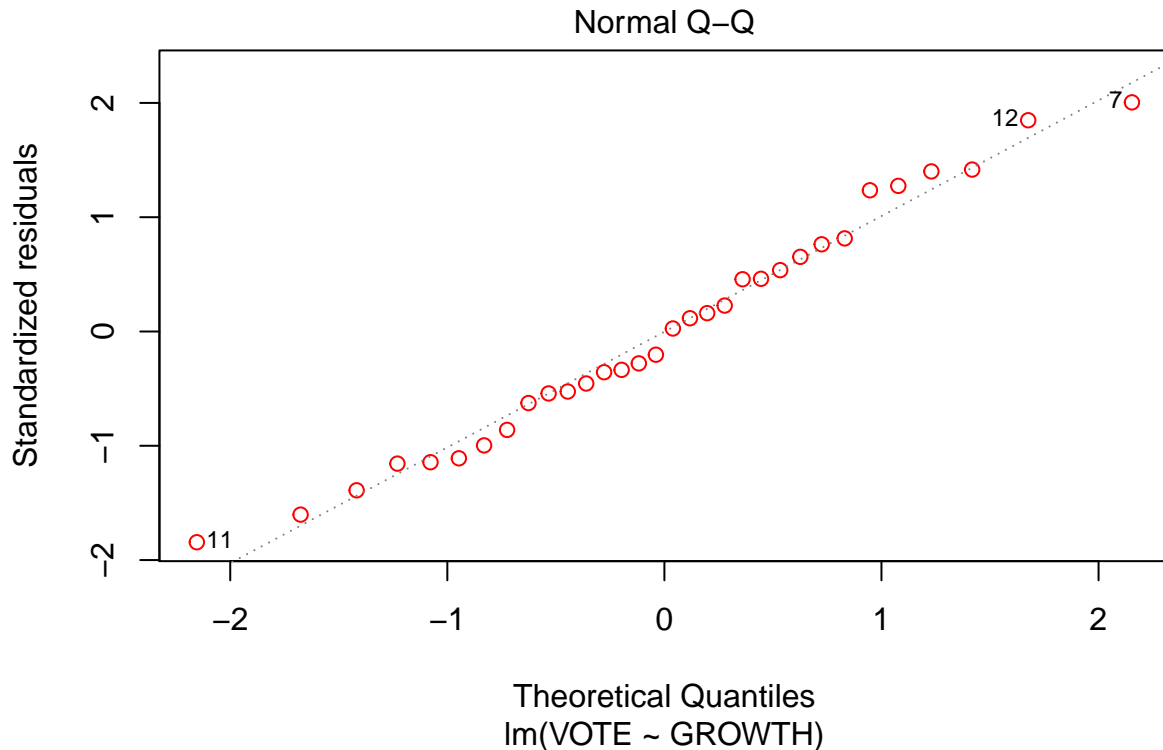
Podemos por meio do gráfico observar a distribuição dos resíduos do modelo. Por meio do gráfico de resíduos, é possível verificar a adequabilidade dos dados. Dado que o resíduo é aquela parte do modelo que não é explicada, espera-se que este não possua nem um padrão. Em suma, a distribuição do resíduo deve ser aleatória.

```
plot(Linear, which=1, col=c("blue"))
```



Quando observamos o gráfico, é possível verificar uma certa concentração de observações na parte um pouco a direita do centro do gráfico, o que pode apresentar uma não normalidade dos dados. Por isso será executado um gráfico de quantis (Q-Q). Caso os resíduos sigam a linha de regressão, é uma boa indicação de uma distribuição normal.

```
plot(Linear, which=2, col=c("red"))
```



Observa-se que os resíduos seguem a linha da regressão de maneira bastante alinhada.

Com base nessas análises, é possível concluir que o modelo é adequado.

### c) Modelo de regressão com mais de uma VI

Para executar o modelo de regressão com VOTES como VD e Growth e Goodnews como VI, executamos o seguinte comando:

```
Linear_2 <- lm(VOTE ~ GROWTH + GOODNEWS, data = fair)
summary(Linear_2)
```

```
##
## Call:
## lm(formula = VOTE ~ GROWTH + GOODNEWS, data = fair)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-8.3125	-3.9191	0.4876	3.0489	9.6846

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	48.1202	1.7476	27.535	< 2e-16 ***
GROWTH	0.5730	0.1527	3.752	0.000781 ***
GOODNEWS	0.7177	0.2964	2.421	0.021947 *

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.596 on 29 degrees of freedom
## Multiple R-squared:  0.4639, Adjusted R-squared:  0.4269
## F-statistic: 12.55 on 2 and 29 DF,  p-value: 0.0001185
```

## Resultados do modelo

A variável adicionada ao modelo é a de GoodNews, relacionada a quantidade de notícias boas na área econômica.

Ao analisar o modelo, percebemos que ambas as variáveis independentes apresentam um efeito na variável dependente. Entretanto, a variável Growth apresenta um P-valor menor que 0.001 enquanto que a variável Goodnews só apresenta um efeito significativo na VD quando consideramos um P-valor menor que 0.05. Mesmo assim, ambas as variáveis estão abaixo do patamar de 0.05, permitindo rejeitar a hipótese nula de que não há relação entre as variáveis. Observando o coeficiente da variável Growth, podemos interpretar que o aumento de 1% na variação do PIB leva a um aumento de 0.57 na VD (mantendo todo o resto constante). Já em relação a GOODNEWS, o aumento em 25% de notícias econômicas positivas, leva ao aumento de 0.71 na VD (mantendo todo o resto constante).

Os valores do resíduo mostram um resíduo mínimo de -8.3125, com o 1º quartil igual a -3.9191, Mediana de 0.4876, 3º quartil de 3.05 e valor máximo de 9.68.

## Capacidade explicativa do modelo

No que se refere a capacidade explicativa, o modelo apresenta um  $R^2$  ajustado de 0.43, significando uma explicação de 43% de variação na VD. O erro padrão do resíduo foi de 4.596.

## RMSE

Para analisar o RMSE (Raiz quadrada do erro médio) será necessário primeiro abrir o pacote `sjstats`

```
library("sjstats")
```

Assim é possível verificar o valor da RMSE:

```
rmse(Linear_2)
```

```
## [1] 4.375343
```

O RMSE é calculado para mensurar a diferença entre os valores preditos pelo modelo e os valores observados. Também é conhecido como erro padrão do modelo e é calculado elevando ao quadrado cada erro do modelo, somando-os, dividindo pelo número de casos e obtendo a raiz quadrada. Nesse caso, o valor obtido foi de 4.37. O RMSE tem a mesma unidade que a VD e quanto menor, melhor.

## Comparando o resultados dos modelos

Para poder comparar os resultados desse modelo com o anterior, é necessário padronizar os coeficientes de ambos os modelos. Para isso será utilizado o pacote `lm.beta`:

```
library('lm.beta')
```

```
lm.beta(Linear)
```

```
##  
## Call:  
## lm(formula = VOTE ~ GROWTH, data = fair)  
##  
## Standardized Coefficients::  
## (Intercept)      GROWTH  
##  0.0000000    0.5962706
```

```
lm.beta(Linear_2)
```

```
##  
## Call:  
## lm(formula = VOTE ~ GROWTH + GOODNEWS, data = fair)  
##  
## Standardized Coefficients::  
## (Intercept)      GROWTH    GOODNEWS  
##  0.0000000    0.5227285    0.3373268
```

Por meio dos coeficientes padronizados, é possível comparar os modelos. No primeiro modelo executado, verifica-se que para o aumento de 1 desvio padrão da VI Growth, espera-se o aumento de .60 desvio padrão na VD.

Já no segundo modelo, o aumento de 1 desvio padrão da VI Growth leva a um aumento de 0.52 desvio padrão na VD. Ou seja, houve uma redução no efeito da VI Growth do modelo 1 para o modelo 2.

```
confint(Linear_2)
```

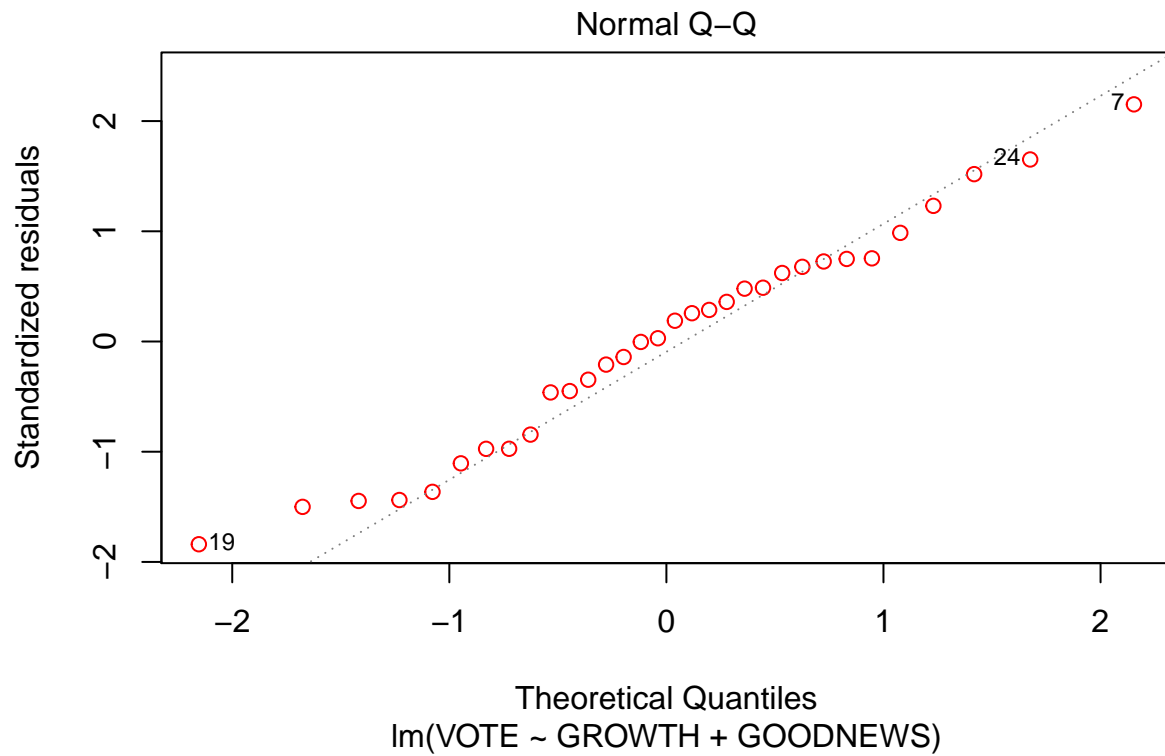
```
##              2.5 %      97.5 %  
## (Intercept) 44.5460566 51.6944312  
## GROWTH      0.2606609 0.8852903  
## GOODNEWS    0.1114819 1.3238397
```

Já o intervalo de confiança apresenta o intervalo para cada coeficiente do modelo. Na VI Growth, o valor mínimo é 0.26 e máx é 0.88, na VÍ Goodnews o valor vai de 0.11 a 1.32.

## Ajuste do modelo

Para verificar o ajuste do modelo, os resíduos serão analisados graficamente.

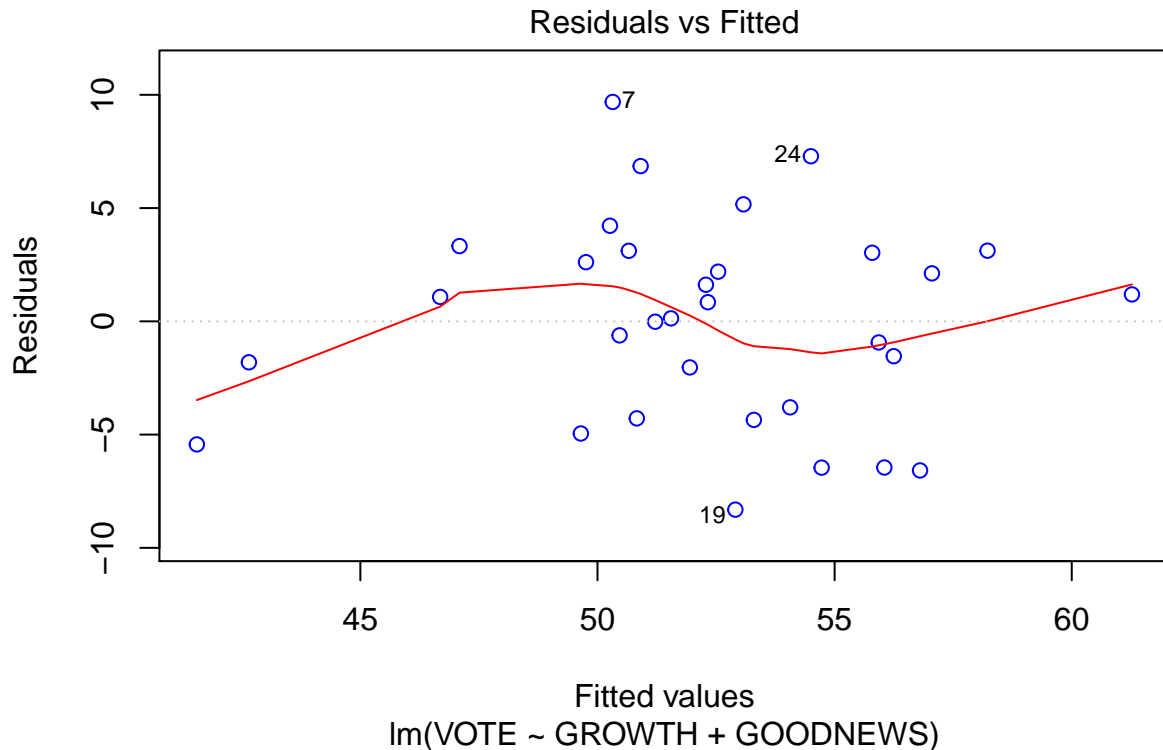
```
plot(Linear_2, which=2, col=c("red"))
```



O gráfico Q-Q mostra um ajuste adequado do modelo dado a tendência dos pontos em relação a linha de regressão.

```
plot(Linear_2, which=1, col=c("blue"))
```





Já o gráfico dos resíduos apresenta valores tanto acima quanto abaixo da linha, representando também um bom ajuste do modelo.

Por fim, o teste de Shapiro permite verificar a normalidade da distribuição:

```
res <- residuals(Linear_2)

shapiro.test(res)

##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.97866, p-value = 0.7598

mean(res)

## [1] 4.26742e-16
```

Os resultados do teste apontam um p-valor de 0.76, levando a não rejeitar a hipótese nula de que os dados testados não estão normalmente distribuídos. A média dos resíduos obtida foi de 4.27. Dado os gráficos analisados e teste realizado, é possível concluir que há homocedasticidade no modelo.

#### d) Modelo de regressão com a VI War

Para executar o modelo de regressão com VOTES como VD e com Growth, Goodnews e War como VI, executamos o seguinte comando:

```
Linear_3 <- lm(VOTE ~ GROWTH + GOODNEWS + WAR, data = fair)
summary(Linear_3)
```

```
##
## Call:
## lm(formula = VOTE ~ GROWTH + GOODNEWS + WAR, data = fair)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.3043 -3.9080  0.5065  3.0190  9.7032
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  48.0479     2.3386  20.545 < 2e-16 ***
## GROWTH        0.5729     0.1554   3.686 0.000969 ***
## GOODNEWS      0.7284     0.3761   1.937 0.062925 .
## WAR           0.1698     3.5622   0.048 0.962322
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.677 on 28 degrees of freedom
## Multiple R-squared:  0.464, Adjusted R-squared:  0.4065
## F-statistic: 8.078 on 3 and 28 DF, p-value: 0.0004951
```

## Resultados do modelo

A variável adicionada ao modelo é a de War, uma variável dicotômica que indica que se naquele ano o país estava em guerra ou não.

Ao analisar o modelo, percebemos que apenas a variável Growth tem efeito sobre a VD. Ao observar o P-valor dos coeficientes, o P-valor de Growth é menor que 0.001 enquanto que os coeficientes de Goodnews e War passam do limiar de 0.05, não sendo assim possível rejeitar a hipótese nula de que não existe relação entre as variáveis. Observando o coeficiente da variável Growth, podemos interpretar que, o aumento de 1% na variação do PIB leva a um aumento de 0.57 na VD (mantendo todo o resto constante). Os coeficientes da variável Goodnews e War são 0.73 e 0.17 respectivamente.

Os valores do resíduo mostram um resíduo mínimo de -8.3043, com o 1º quartil igual a -3.9080, Mediana de 0.5065, 3º quartil de 3.0190 e valor máximo de 9.70.

## Capacidade explicativa do modelo

No que se refere a capacidade explicativa, o modelo apresenta um  $R^2$  ajustado de 0.40, significando uma explicação de 40% de variação na VD. O erro padrão do resíduo foi de 4.677.

## RMSE

```
rmse(Linear_3)
```

```
## [1] 4.375166
```

O RMSE é calculado para mensurar a diferença entre os valores preditos pelo modelo e os valores observados. Também é conhecido como erro padrão do modelo e é calculado elevando ao quadrado cada erro do modelo, somando-os, dividindo pelo número de casos e obtendo a raiz quadrada. Nesse caso, o valor obtido foi de 4.37. O RMSE tem a mesma unidade que a VD e quanto menor, melhor.

### Comparando o resultados dos modelos

```
library('lm.beta')
```

```
lm.beta(Linear)
```

```
##
## Call:
## lm(formula = VOTE ~ GROWTH, data = fair)
##
## Standardized Coefficients::
## (Intercept)      GROWTH
##  0.0000000    0.5962706
```

```
lm.beta(Linear_2)
```

```
##
## Call:
## lm(formula = VOTE ~ GROWTH + GOODNEWS, data = fair)
##
## Standardized Coefficients::
## (Intercept)      GROWTH    GOODNEWS
##  0.0000000    0.5227285    0.3373268
```

```
lm.beta(Linear_3)
```

```
##
## Call:
## lm(formula = VOTE ~ GROWTH + GOODNEWS + WAR, data = fair)
##
## Standardized Coefficients::
## (Intercept)      GROWTH    GOODNEWS      WAR
## 0.000000000 0.522617627 0.342359350 0.008281798
```

Por meio dos coeficientes padronizados, é possível comparar os modelos. No primeiro modelo executado, verifica-se que para o aumento de 1 desvio padrão da VI Growth, espera-se o aumento de .60 desvio padrão na VD.

Já no segundo modelo, o aumento de 1 desvio padrão da VI Growth leva a um aumento de 0.52 desvio padrão na VD. Ou seja, houve uma redução no efeito da VI Growth do modelo 1 para o modelo 2.

No terceiro modelo, o valor do coeficiente padronizado da VI Growth permanece o mesmo, enquanto que em relação a VI Goodnews, no terceiro modelo há um aumento de 0.01 no coeficiente. Entretanto, no terceiro modelo a VI Goodnews não possui um efeito significativo. Já a variável War apresenta um coeficiente padronizado menor que 0.

Comparando os três modelos, percebemos que a variável que mais explica a variação na VD é Growth.

```
confint(Linear_3)
```

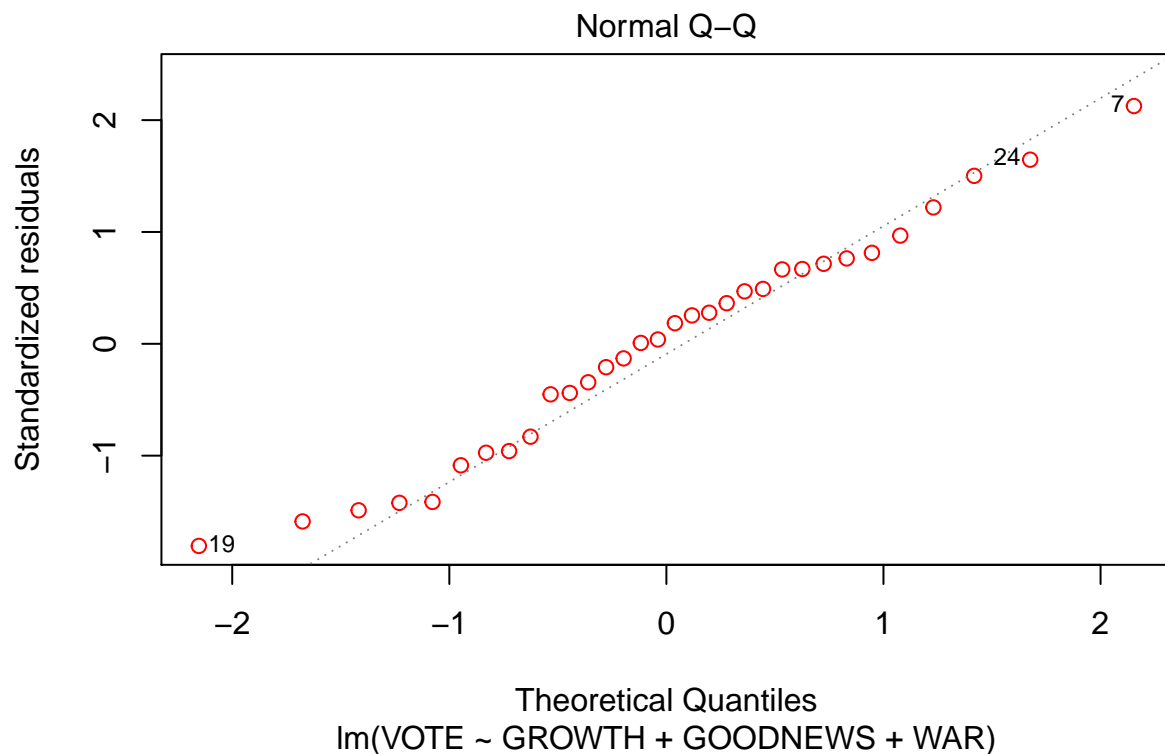
```
##                2.5 %      97.5 %  
## (Intercept) 43.25739623 52.8383197  
## GROWTH      0.25448754 0.8912206  
## GOODNEWS   -0.04198698 1.4987219  
## WAR        -7.12709113 7.4666712
```

Já o intervalo de confiança apresenta o intervalo para cada coeficiente do modelo. Na VI Growth, o valor mínimo é 0.25 e máx é 0.89, na VI Goodnews o valor vai de -0.04 a 1.50 e na VI War os valores são -7.13 de mínimo e 7.47 de máximo.

### Ajuste do modelo

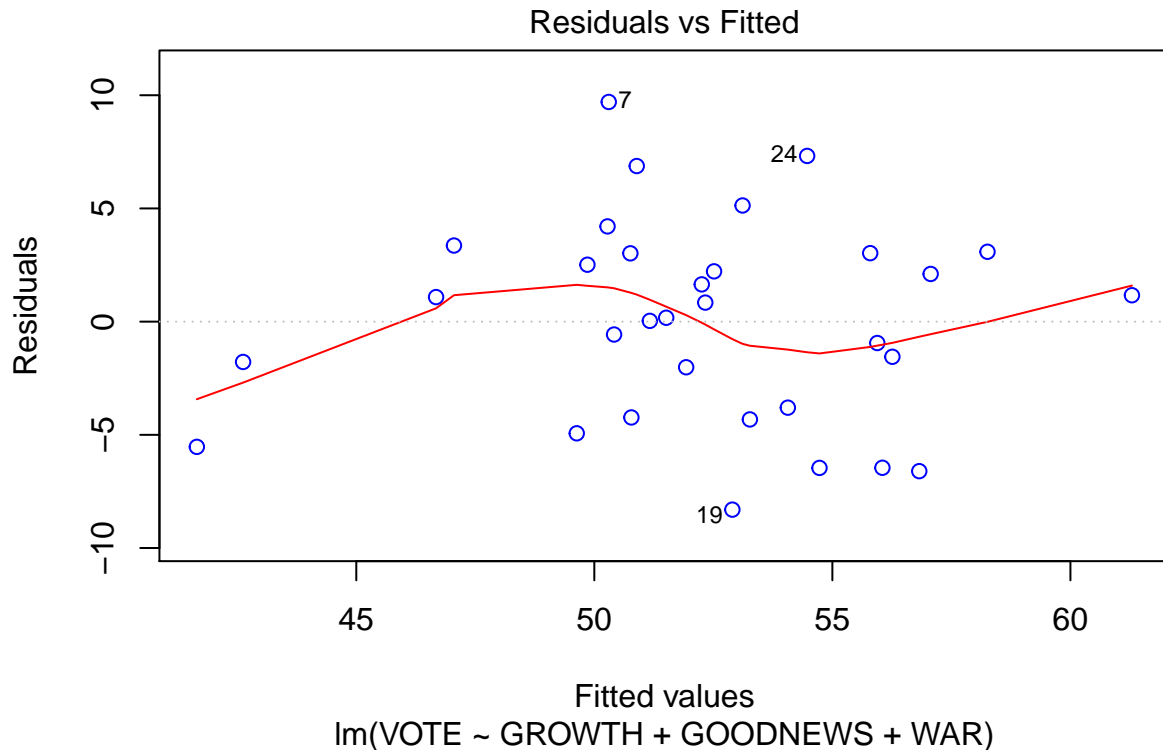
Para verificar o ajuste do modelo, os resíduos serão analisados graficamente.

```
plot(Linear_3, which=2, col=c("red"))
```



O gráfico Q-Q mostra um ajuste adequado do modelo dado a tendência dos pontos em relação a linha de regressão.

```
plot(Linear_3, which=1, col=c("blue"))
```



Já o gráfico dos resíduos apresenta valores tanto acima quanto abaixo da linha, representando também um bom ajuste do modelo.

Por fim, o teste de Shapiro permite verificar a normalidade da distribuição:

```
res <- residuals(Linear_3)

shapiro.test(res)

##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.97851, p-value = 0.7554
```

```
mean(res)

## [1] 3.330669e-16
```

Os resultados do teste apontam um p-valor de 0.75, levando a não rejeitar a hipótese nula de que os dados testados não estão normalmente distribuídos. A média dos resíduos obtida foi de 3.30. Dado os gráficos analisados e o teste realizado, é possível concluir que há homocedasticidade no modelo.

## Questão 2

### Citação do artigo

BATISTA, Mariana. Who Gets What and How Does It Matter? Importance-Weighted Portfolio Allocation and Coalition Support in Brazil. *Journal of Politics in Latin America*, v. 10, n. 3, p. 99-134, 2018.

### Modelo

A variável dependente do modelo é disciplina partidária. O que a autora busca analisar é como a alocação de ministérios importa para o apoio legislativo dos parceiros de coalização no sistema presidencialista brasileiro. Na tabela 4 do artigo, são apresentados 4 modelos de regressão linear tendo como VD a variável de disciplina partidária. O objetivo é verificar o efeito da alocação do ministério na disciplina partidária. Por meio do modelo, é possível verificar os fatores que levam a maior disciplina partidária e qual o papel que a alocação dos ministérios tem sobre a VD.

No modelo I, as VIs do modelo são: Distância Ideológica, Dimensão política do ministério, Segundo mandato, Tempo até a eleição e Popularidade. Ao verificar os resultados dos coeficientes, apenas os efeitos da VI de distância ideológica e Tempo até a eleição apresentam um P-valor menor que 0.01 enquanto que Segundo mandato e popularidade são significativos apenas com um P-valor menor que 0.1 e Dimensão Política do Min. não apresenta um P-valor significativo.

Em relação aos coeficientes, o da VI Distância Ideológica é de -10.131 (aumento de 1 unidade leva a redução de -10.131 na VD), Dimensão política do ministério é 8.341 (aumento de 1 unidade leva a aumento de 8.341 na VD), Segundo mandato é 3.439 (aumento de 1 unidade leva a aumento de 3.439 na VD), Tempo até a eleição 4.456 (aumento de 1 unidade leva a aumento de 4.456 na VD) e Popularidade 0.139 (aumento de 1 unidade leva a aumento de 0.139). A constante tem um coeficiente de 65.685.

O  $R^2$  do modelo é de 0.335. Ou seja, 33% da variação da VD é explicada pela VI e o erro padrão do resíduo é de 9.344.