

# Análise de Dados - UFPE/2019 - Lista 7

*Maria Eduarda R. N. Lessa*

*28 de maio de 2019*

## Questão 1:

```
# Definir diretório:
setwd("C:/Users/Duda/Desktop/PPGCP/Análise de Dados/lista_07")

options(repos=c(CRAN="<something sensible near you>"))

# Instalar pacote "foreign" para ler dta:
install.packages("foreign")

## Installing package into 'C:/Users/Duda/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)

## Warning: unable to access index for repository <something sensible near you>/src/contrib:
##   scheme not supported in URL '<something sensible near you>/src/contrib/PACKAGES'

## Warning: package 'foreign' is not available (for R version 3.5.2)

## Warning: unable to access index for repository <something sensible near you>/bin/windows/contrib/3.5
##   scheme not supported in URL '<something sensible near you>/bin/windows/contrib/3.5/PACKAGES'

library("foreign")

# Carregar base "fair.dta", atribuir a objeto "vote_growth":
vote_growth <- read.dta("fair.dta")
```

## letra a)

```
# Checar variáveis (colunas) e visualizar a base:

colnames(vote_growth)

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

View(vote_growth)

# Apresentar a análise descritiva de todas as variáveis da base "vote_growth":
summary(vote_growth)
```

```
##      YEAR      VOTE      PARTY      PERSON
## Min.   :1880   Min.   :36.12   Min.   :-1.000   Min.   :0.0000
## 1st Qu.:1911   1st Qu.:49.43   1st Qu.: -1.000   1st Qu.:0.0000
## Median :1942   Median :52.03   Median :-1.000   Median :1.0000
## Mean   :1942   Mean   :52.27   Mean   :-0.125   Mean   :0.5938
## 3rd Qu.:1973   3rd Qu.:55.69   3rd Qu.: 1.000   3rd Qu.:1.0000
## Max.   :2004   Max.   :62.46   Max.   : 1.000   Max.   :1.0000
##      DURATION      WAR      GROWTH      INFLATION
## Min.   :0.0000   Min.   :0.00000   Min.   :-14.557   Min.   :0.000
## 1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.: -1.674   1st Qu.:1.398
## Median :1.0000   Median :0.00000   Median : 2.245   Median :2.159
## Mean   :0.7031   Mean   :0.09375   Mean   : 0.628   Mean   :2.657
## 3rd Qu.:1.2500   3rd Qu.:0.00000   3rd Qu.: 4.061   3rd Qu.:3.352
## Max.   :2.0000   Max.   :1.00000   Max.   : 11.677   Max.   :7.926
##      GOODNEWS
## Min.   : 0.000
## 1st Qu.: 3.750
## Median : 5.000
## Mean   : 5.281
## 3rd Qu.: 7.250
## Max.   :10.000
```

---

letra b)

```
# Análise de regressão de VOTE (VD) e GROWTH (VI):
reg <- lm(VOTE ~ GROWTH, data = vote_growth)
summary(reg)
```

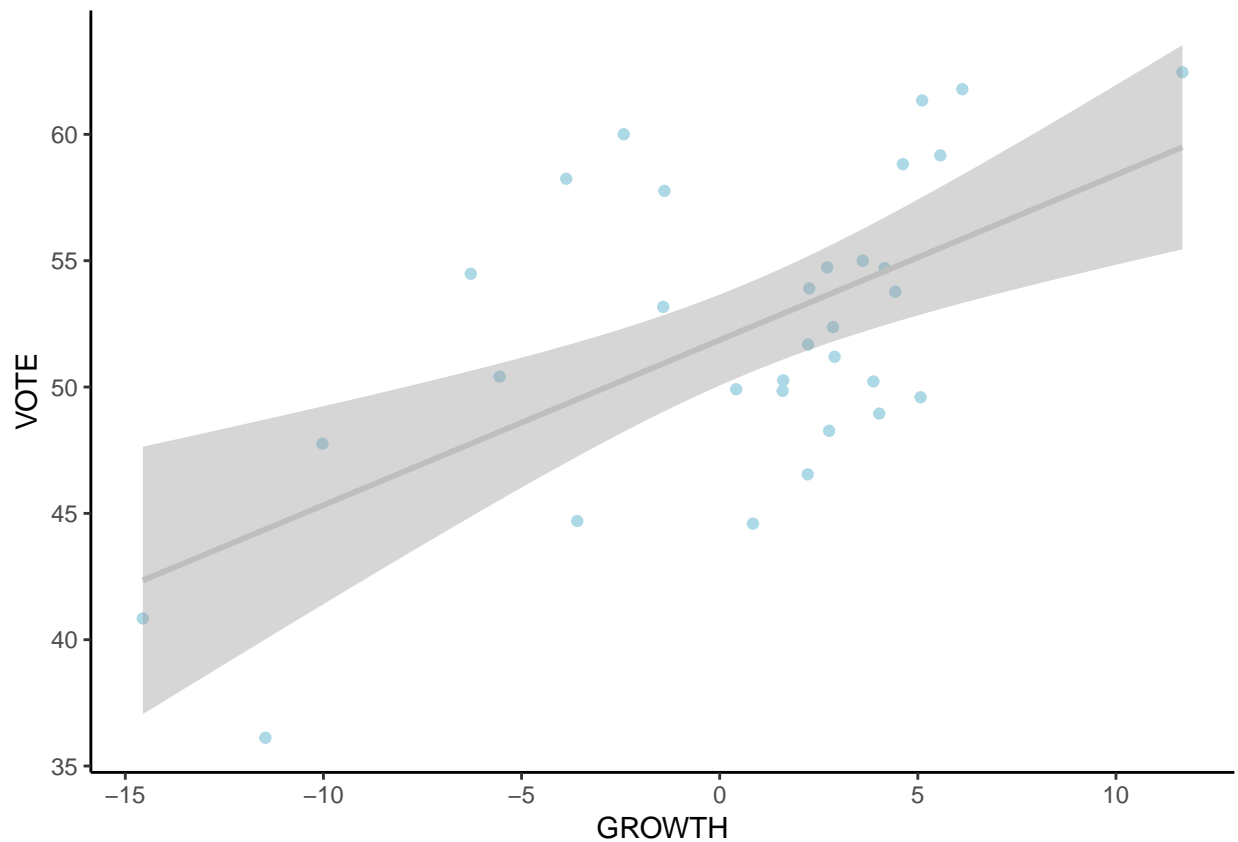
```
##
## Call:
## lm(formula = VOTE ~ GROWTH, data = vote_growth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.2487 -3.3330 -0.4282  3.1425  9.7286
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  51.8598      0.8817  58.821 < 2e-16 ***
## 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
```

```
# Plotar dados e reta de regressão com IC:
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
ggplot(data = vote_growth, aes(y = VOTE, x = GROWTH)) +  
  geom_point(color = "lightblue") +  
  theme_classic() +  
  geom_smooth(method="lm", color = "gray", se = TRUE)
```



i)

As variáveis utilizadas foram “VOTE” e “GROWTH”. O modelo analisa o efeito do crescimento econômico (a variável independente), sobre o percentual de votos recebidos pelo candidato do partido incumbente (a variável dependente).

ii)

```
# Imprimir resultados do modelo (tabela de regressão):  
summary(reg)
```

```
##
```

```
## Call:
```

```
## lm(formula = VOTE ~ GROWTH, data = vote_growth)
##
## 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
```

```
# Calcular RMSE:
require(magrittr)
```

```
## Loading required package: magrittr
```

```
## Warning: package 'magrittr' was built under R version 3.5.3
```

```
mean(residuals(reg)^2) %>% sqrt()
```

```
## [1] 4.797286
```

- Valor de  $\hat{a} = 51.860$ , desvio padrão = 0.882, p-valor < 0.001

- Valor de  $\hat{b} = 0.654$ , desvio padrão = 0.161, p-valor < 0.001

- Valor do R quadrado = 0.356

- Valor do erro padrão dos resíduos (RSE) = 4.95, gl = 30

- Valor do RMSE: 4.797

iii)

Os valores, tanto para o intercepto (a), quanto para o coeficiente de variação (b), são estatisticamente significativos, já que apresentaram p-valor < 0.001. O valor de  $\hat{b}$ , considerando um índice de confiança de 95%, não inclui o zero e assume um valor positivo (0.654), o que mostra que o efeito do crescimento econômico sobre o voto no candidato do partido incumbente é significativo. Quanto aos valores do  $R^2$  e do RMSE, 0.356 e 4.797, respectivamente, apontam que o crescimento econômico explica aproximadamente 0.35 da variação dos votos no candidato do partido incumbente e o erro padrão do resíduo de 4.95 pontos em uma escala de 0 a 100 (já que é medido na mesma unidade da variável dependente, que neste caso é a porcentagem de voto), indica que o ajuste do modelo é satisfatório.

letra c)

```
# Análise de regressão de VOTE (VD), GROWTH (VI) e GOODNEWS (VI):
reg_multi <- lm(VOTE ~ GROWTH + GOODNEWS, data = vote_growth)
summary(reg_multi)
```

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

i)

```
# Verificar o coeficiente de correlação das VIs:
cor(x = vote_growth$GOODNEWS, y = vote_growth$GROWTH, method = c("pearson"))
```

```
## [1] 0.2180145
```

A variável adicionada ao modelo é a “GOODNEWS”, que avalia o crescimento forte (ou acima do esperado) em cada um dos trimestres da administração do candidato incumbente. Estas boas performances da economia são consideradas “boas notícias”, já que tendem a influenciar o comportamento do eleitor, segundo estudos da psicologia política. A inclusão desta variável tem como objetivo analisar o seu efeito que sobre os votos no candidato do partido incumbente (a VD), para que seja possível, então, analisar o efeito da primeira VI (“GROWTH”) sobre a VD, controlando para a VI adicionada (“GOODNEWS”).

ii)

```
# Imprimir resultados do modelo (tabela de regressão):
summary(reg_multi)
```

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

```
# Calcular RMSE:
mean(residuals(reg_multi)^2) %>% sqrt()
```

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

- Valor de  $\hat{a} = 48.120$ , desvio padrão = 1.748, p-valor < 0.001
- Valor de  $\hat{b}_X$  (GROWTH) = 0.573, desvio padrão = 0.153, p-valor < 0.001
- Valor de  $\hat{b}_Z$  (GOODNEWS) = 0.718, desvio padrão = 0.296, p-valor = 0.01
- Valor do  $R^2 = 0.464$
- Valor do erro padrão dos resíduos = 4.6, gl = 29
- Valor do RMSE: 4.375

iii)

Com a inclusão da variável GOODNEWS, o  $R^2$  mostra que 0.464 da variação dos votos no candidato do partido incumbente é explicada pelas duas VIs analisadas (GROWTH e GOODNEWS). O valor do coeficiente de variação ( $\hat{b}_X$ ) para GROWTH neste modelo é de 0.573, que é menor do que aquele encontrado no modelo bivariado, visto que este modelo apresenta os resultados de GROWTH sobre os votos, controlando para o efeito de GOODNEWS, com p-valor < 0.001. Para GOODNEWS, controlando para o efeito de GROWTH, o coeficiente de variação ( $\hat{b}_Z$ ) encontrado foi de 0.718 com p-valor = 0.01. Os resultados do p-valor e o fato dos ICs de 95% não incluírem o 0, mostra que os resultados são estatisticamente significantes. Finalmente, o RMSE de 4.375 mostra que este modelo é mais ajustado do que o bivariado apresentado na questão anterior.

iv)

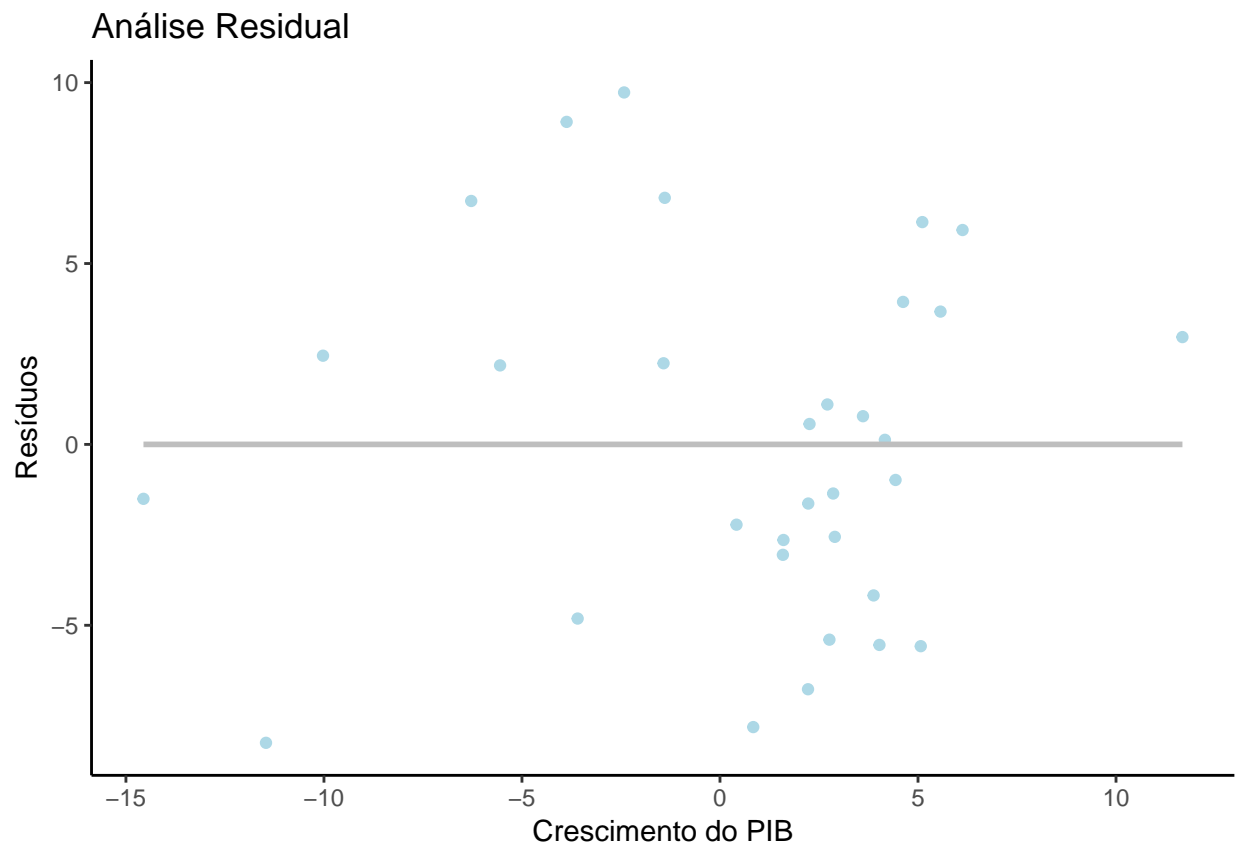
O modelo multivariado apresentou RMSE menor do que o modelo bivariado, o que significa que o erro padrão dos resíduos foi menor neste último modelo, ou seja, a diferença entre os valores observados e os valores estimados foi menor. Também, ao analisar os valores do  $R^2$ , nota-se que o modelo multivariado explicou uma parte maior da variação dos votos no candidato do partido incumbente.

v)

```
# Analisar média dos resíduos do modelo 1:  
resid_reg <- resid(reg)  
mean(resid_reg)
```

```
## [1] 2.63678e-16
```

```
# Plotar análise residual:  
ggplot(data = vote_growth, aes(y = resid_reg, x = GROWTH)) +  
  geom_point(color = "lightblue") +  
  theme_classic() +  
  geom_smooth(method="lm", color = "gray", se = F) +  
  labs(title = "Análise Residual", y = "Resíduos", x = "Crescimento do PIB")
```



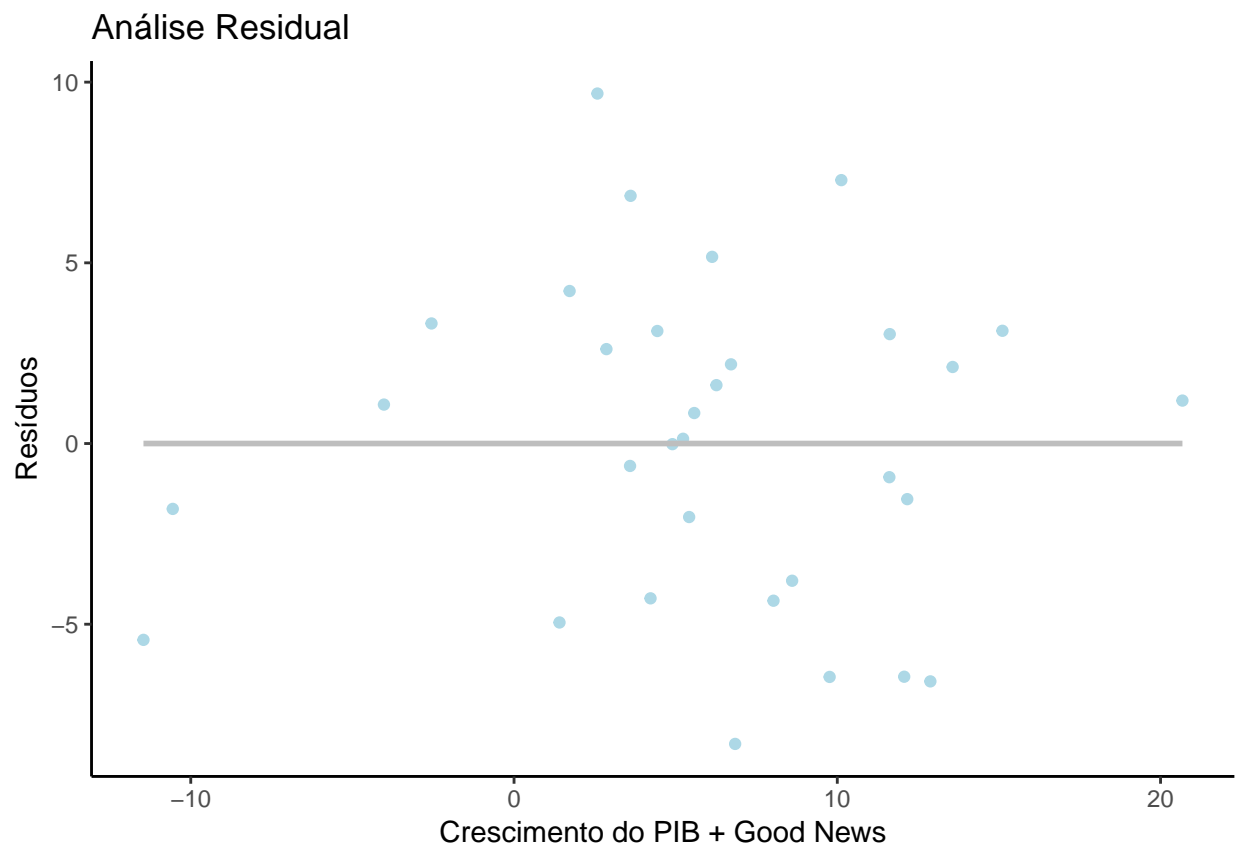
O modelo 1 parece ser homocedástico, já que a dispersão dos resíduos não parece diminuir ou aumentar com o aumento de X (GROWTH).

## Analisar resíduos modelo 2:

```
resid_reg_multi <- resid(reg_multi)
mean(resid_reg_multi)
```

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

```
# Plotar análise residual:
ggplot(data = vote_growth, aes(y = resid_reg_multi, x = GROWTH + GOODNEWS)) +
  geom_point(color = "lightblue") +
  theme_classic() +
  geom_smooth(method="lm", color = "gray", se = F) +
  labs(title = "Análise Residual", y = "Resíduos",
       x = "Crescimento do PIB + Good News")
```



O modelo 2 também parece ser homocedástico, já que a dispersão dos resíduos não parece diminuir ou aumentar com o aumento de X e Z (GROWTH e GOODNEWS).



letra d)

```
# Análise de regressão de VOTE (VD), GROWTH (VI) e WAR (VI):
reg_multi_2 <- lm(VOTE ~ GROWTH + WAR, data = vote_growth)
summary(reg_multi_2)

##
## Call:
## lm(formula = VOTE ~ GROWTH + WAR, data = vote_growth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.1789 -3.6648  0.3595  3.0116  9.2815
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  52.2460     0.9186  56.876 < 2e-16 ***
## GROWTH        0.6284     0.1598   3.932 0.000481 ***
## WAR         -3.9510     2.9894  -1.322 0.196609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.894 on 29 degrees of freedom
## Multiple R-squared:  0.3922, Adjusted R-squared:  0.3502
## F-statistic: 9.355 on 2 and 29 DF,  p-value: 0.0007329
```

i)

```
# Verificar o coeficiente de correlação das VIs:
cor(x = vote_growth$WAR, y = vote_growth$GROWTH, method = c("pearson"))

## [1] -0.1190938
```

A variável “WAR” apresenta fraca correlação com “GROWTH” (-0.12). A inclusão da variável servirá para analisar e controlar os efeitos dos períodos de guerra sobre “GROWTH” e sobre a variação do percentual de voto recebido pelo candidato do partido incumbente.

ii)

```
# Imprimir resultados do modelo (tabela de regressão):
summary(reg_multi_2)

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

```
##      Min      1Q  Median      3Q      Max
## -8.1789 -3.6648  0.3595  3.0116  9.2815
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  52.2460      0.9186  56.876 < 2e-16 ***
## GROWTH        0.6284      0.1598   3.932 0.000481 ***
## WAR          -3.9510      2.9894  -1.322 0.196609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.894 on 29 degrees of freedom
## Multiple R-squared:  0.3922, Adjusted R-squared:  0.3502
## F-statistic: 9.355 on 2 and 29 DF,  p-value: 0.0007329
```

```
# Calcular RMSE:
mean(residuals(reg_multi_2)^2) %>% sqrt()
```

```
## [1] 4.65902
```

- Valor de  $\hat{a} = 52.246$ , desvio padrão = 0.919, p-valor < 0.001
- Valor de  $\hat{b}_X$  (GROWTH) = 0.628, desvio padrão = 0.160, p-valor < 0.001
- Valor de  $\hat{b}_W$  (WAR) = -3.951, desvio padrão = 2.989, p-valor = 1
- Valor do  $R^2 = 0.392$
- Valor do erro padrão dos resíduos = 4.98, gl = 29
- Valor do RMSE: 4.659

iii)

Com a inclusão da variável WAR, o  $R^2$  mostra que 0.392 da variação dos votos no candidato do partido incumbente é explicada pelas duas VIs analisadas (GROWTH e WAR). O valor do coeficiente de variação ( $\hat{b}_W$ ) encontrado, controlando para “GROWTH”, foi de -3.951 com p-valor = 1. O resultado encontrado não é estatisticamente significativo, já que o valor do coeficiente de variação ( $\hat{b}_W$ ) apresenta p-valor = 1 e o IC de 95% inclui o zero. Finalmente, o RSE de 4.89 e o RMSE de 4.659, mostram que este modelo é menos ajustado do que aquele apresentado na letra c.

iv)

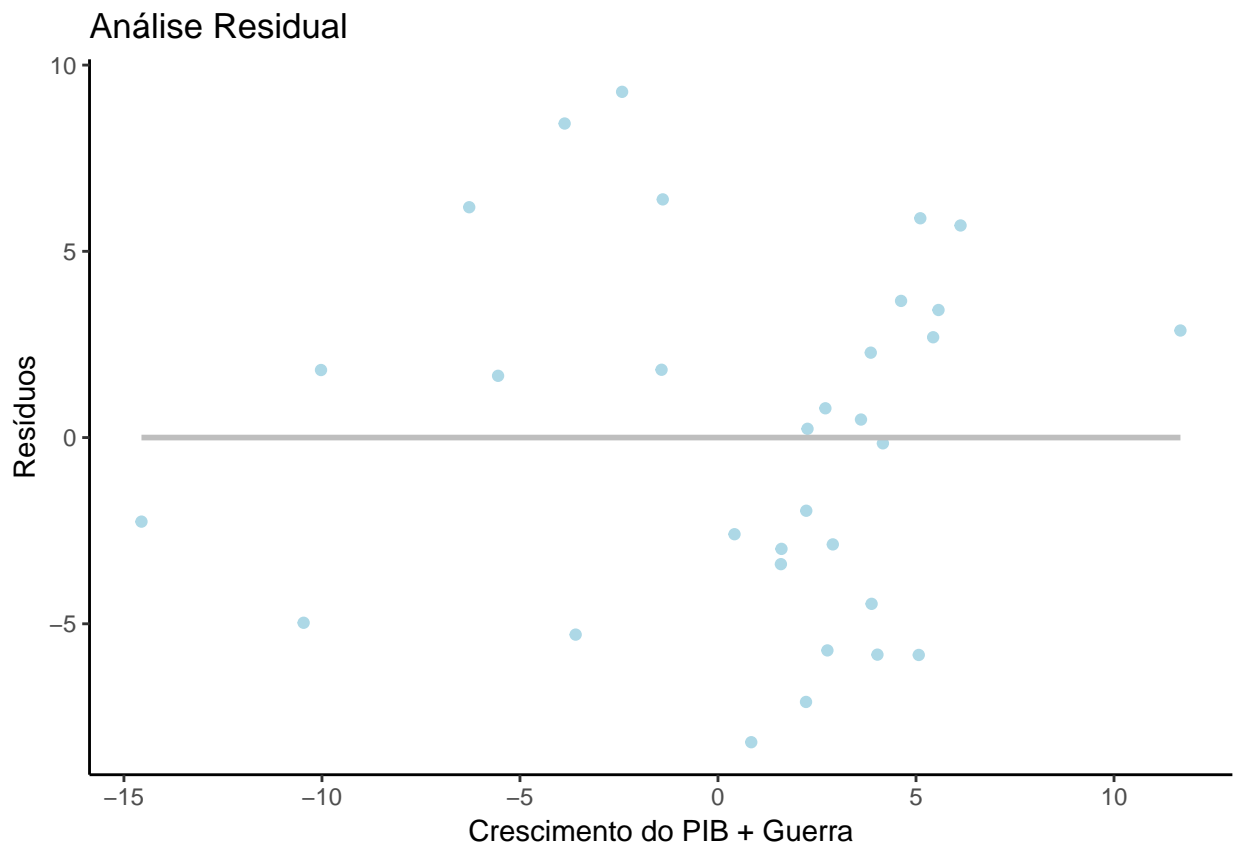
Neste terceiro modelo o valor do RMSE aumentou, o que significa que ele é menos ajustado do que o modelo anterior. A inclusão de WAR pouco afetou o efeito de GROWTH sobre a variação dos votos, mas afetou o efeito.

v)

```
# Analisar resíduos modelo 3:  
resid_reg_multi_2 <- resid(reg_multi_2)  
mean(resid_reg_multi_2)
```

```
## [1] 6.036838e-16
```

```
# Plotar análise residual:  
ggplot(data = vote_growth, aes(y = resid_reg_multi_2, x = GROWTH + WAR)) +  
  geom_point(color = "lightblue") +  
  theme_classic() +  
  geom_smooth(method="lm", color = "gray", se = F) +  
  labs(title = "Análise Residual", y = "Resíduos",  
       x = "Crescimento do PIB + Guerra")
```



O modelo 3 também parece ser homocedástico, já que a dispersão dos resíduos não parece apresentar um padrão (de diminuição ou aumento da variância) com o aumento de X e W (GROWTH e WAR, respectivamente).

vi)

Padronizar os  $\hat{b}$ :

```
# Modelo 2 (GROWTH e GOODNEWS):  
0.573 * (sd(vote_growth$GROWTH) / sd(vote_growth$VOTE))
```

```
## [1] 0.5227507
```

```
0.718 * (sd(vote_growth$GOODNEWS) / sd(vote_growth$VOTE))
```

```
## [1] 0.3374863
```

```
# Modelo 3 (GROWTH e WAR):  
0.628 * (sd(vote_growth$GROWTH) / sd(vote_growth$VOTE))
```

```
## [1] 0.5729275
```

```
-3.951 * sd(vote_growth$WAR) / (sd(vote_growth$VOTE))
```

```
## [1] -0.1927168
```

A variável que tem maior efeito sobre o percentual de votos no candidato do partido incumbente é “GROWTH”, visto que, no segundo modelo, o incremento de uma unidade nesta VI é responsável por uma variação de cerca de 0.523 na VD (versus 0.337 de “GOODNEWS”) e 0.573 no terceiro modelo (versus -0.193 de “WAR”).

## Questão 2:

letra a)

Pérez, E. O., & Tavits, M. (2017). Language Shapes People’s Time Perspective and Support for Future-Oriented Policies. *American Journal of Political Science*, 61(3), 715–727. Disponível em: doi:10.1111/ajps.12290.

letra b)

A variável dependente é a “percepção temporal” dos cidadãos. Os autores buscam compreender se a língua falada influencia o apoio dos cidadãos às políticas “*future-oriented*”. Descubrem que nas línguas “*futureless*” os eleitores têm uma percepção temporal diferente do que aqueles de línguas “*futured*”; os primeiros têm maior probabilidade de endossar as políticas de longo prazo, já que percebem o futuro como um “tempo” mais próximo do presente. No primeiro estudo, os autores avaliam o impacto de uma língua “*futureless*” - o estoniano - e desenvolvem cinco modelos de análise, cada um com uma variável dependente diferente, mas todas com o mesmo objetivo de avaliar a “percepção temporal” dos eleitores. Neste estudo, os autores esperam encontrar maior coeficiente de variação para VDs que representem percepções de longo prazo, já que o estoniano, por ser uma língua “*futureless*”, faz com que os seus falantes tenham percepções menos imediatistas e mais voltadas para o futuro (e esperam encontrar o efeito contrário para o russo, que é a língua “*futured*” avaliada neste primeiro estudo). No

segundo estudo os autores expandem a análise para 60 países e incluem uma série de variáveis de controle, como gênero, idade, educação, renda, situação empregatícia, estado civil, apoio à democracia e capital social. As línguas são classificadas em “*futured*” ou “*futureless*” de acordo com a teoria de “future time reference” (FTR; Chen 2013) de uma língua, recebe valor 1 se fortemente “*futured*” e 0 se fortemente “*futureless*”. Os autores desenvolvem seis modelos, os três primeiros têm como VD a probabilidade de os cidadãos apoiarem políticas que poupem o meio ambiente e os três últimos, a VD é a probabilidade de o indivíduo poupar dinheiro. Neste segundo estudo, os autores esperam um coeficiente de variação ( $\hat{b}$ ) negativo para países com línguas consideradas “*futured*”, já que postulam que os cidadãos falantes destas línguas estão mais propensos a apoiar políticas de curto prazo.

---

letra c)

No primeiro estudo os autores encontram, assim como esperado, um  $\hat{b}$  positivo e significativo para o efeito do estoniano (uma língua “*futureless*”) sobre as percepções de longo prazo; e um  $\hat{b}$  negativo e significativo para o efeito do estoniano sobre as percepções de curto prazo. No segundo estudo, os modelos 3 e 6 agregam todas as variáveis políticas e demográficas mencionadas na letra b) e o resultado que é apresentado, mais uma vez, corresponde aos postulados dos autores: o resultado do  $\hat{b}$  para o efeito de línguas “*futured*” sobre o apoio a compromissos de longo prazo, mostrou-se negativo e significativo.