

Lista_7_Caio_Rios

Caio Rios

27 de maio de 2019

1. A base de dados fair.dta (disponível na pasta do OSF) possui variáveis adicionais ao banco de dados sobre voto econômico nos EUA. Com ela desenvolva as atividades abaixo e comente seu código:

a) Apresente uma análise descritiva de todas as variáveis da base de dados.

```
Variaveis <- c("Year", "Vote", "Party", "Person", "Duration", "War", "Growth",  
              "Inflation", "Good News")  
Minimo <- c(1880, 36.12, -1, 0, 0, 0, -14.56, 0, 0)  
PQ <- c(1911, 49.43, -1, 0, 0, 0, -1.67, 1.40, 3.75)  
Mediana <- c(1942, 52.03, -1, 1, 1, 0, 2.24, 2.16, 5)  
Media <- c(1942, 52.27, -0.12, 0.59, 0.70, 0.09, 0.63, 2.66, 5.29)  
TQ <- c(1973, 55.69, 1, 1, 1.25, 0, 4.06, 3.35, 7.25)  
Maximo <- c(2004, 62.46, 1, 1, 2, 1, 11.68, 7.93, 10)  
  
desc <- data.frame(Variaveis, Minimo, PQ, Mediana, Media,  
                  TQ, Maximo)  
kable(desc)
```

Variaveis	Minimo	PQ	Mediana	Media	TQ	Maximo
Year	1880.00	1911.00	1942.00	1942.00	1973.00	2004.00
Vote	36.12	49.43	52.03	52.27	55.69	62.46
Party	-1.00	-1.00	-1.00	-0.12	1.00	1.00
Person	0.00	0.00	1.00	0.59	1.00	1.00
Duration	0.00	0.00	1.00	0.70	1.25	2.00
War	0.00	0.00	0.00	0.09	0.00	1.00
Growth	-14.56	-1.67	2.24	0.63	4.06	11.68
Inflation	0.00	1.40	2.16	2.66	3.35	7.93
Good News	0.00	3.75	5.00	5.29	7.25	10.00

b) Apresente um modelo de regressão linear bivariado no qual VOTE seja variável dependente e GROWTH variável independente.

```
regbi <- lm(VOTE ~ GROWTH, data = ecovote)
```

I) Descreva as variáveis utilizadas no modelo e sua relação;

A variável dependente mede a porcentagem do voto do incumbente e a variável independente mede o crescimento econômico. Logo, a relação esperada, de acordo com a teoria do voto econômico, é que o crescimento econômico tenha um impacto positivo nos votos do incumbente. Isto é, eleitores tendem a recompensar o incumbente quando a economia vai bem.

II) Apresente os resultados do modelo;

```
summary(regbi)
```

```
##
## Call:
## lm(formula = VOTE ~ GROWTH, data = ecovote)
##
## 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
```

A relação encontrada corrobora a hipótese do voto econômico. Quando há crescimento econômico, os eleitores tendem a votar no candidato incumbente. A relação foi estatisticamente significativa a um p-valor menor que 0,000 e quanto a magnitude do efeito, ao aumentar uma unidade da variável crescimento, o incumbente receberia 0,65% a mais dos votos.

III) Avalie o modelo quanto a seu ajuste e sua capacidade explicativa;

O R^2 ajustado do modelo foi de 0,3341. Isto quer dizer que o modelo explica cerca de 33,5% dos casos. Por ser um modelo bivariado, sua capacidade explicativa é limitada.

c) Apresente um modelo de regressão multivariada adicionando ao modelo do item a) a variável **GOODNEWS** como variável independente.

```
regmu <- lm(VOTE ~ GROWTH + GOODNEWS, data = ecovote)
```

I) Descreva a variável adicionada ao modelo e sua relação com as demais;

A nova variável acrescentada no modelo de regressão foi GOODNEWS. Esta mede a quantidade de notícias positivas em relação ao governo vigente. Ela deve interagir positivamente com a variável dependente (% de votos do incumbente) pois boas notícias refletem um bom governo e o eleitor tende a recompensar bons governos.

II) Apresente os resultados do modelo;

```
summary(regmu)
```

```
##
## Call:
## lm(formula = VOTE ~ GROWTH + GOODNEWS, data = ecovote)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.3125 -3.9191  0.4876  3.0489  9.6846
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##              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
```

A relação entre crescimento econômico e votos do incumbente permaneceu positiva e estatisticamente significativa. A variável GOODNEWS também tem uma relação positiva com voto, contudo com um p-valor mais modesto de 0,022. Ainda sim, podemos dizer que existe significância estatística se considerarmos, como ponto de corte, o p-valor de 0,05. A cada boa notícia a mais, o voto do incumbente cresce em 0,71%.

III) Avalie o modelo quanto a seu ajuste e sua capacidade explicativa. Comente e interprete os coeficientes estimados, os intervalos de confiança do modelo e o RMSE e o R².

Ao acrescentarmos mais uma variável ao modelo, ele ficou mais ajustado do que o primeiro. O R² ajustado foi de 0,4269, isto é, o modelo explica, aproximadamente, 42,7% dos casos. Em consonância com o R², o erro padrão dos resíduos foi de 4,596. Este valor também nos informa o quão ajustado é o nosso modelo com os dados reais. Ele calcula a raiz quadrada da média dos resíduos ao quadrado dividido pelo grau de liberdade, sendo seu valor “perfeito” igual a 0.

IV) Compare os resultados do item anterior com os adquiridos com o modelo do item b).

No modelo 1, o efeito de crescimento econômico sobre o voto do incumbente foi de 0,65 enquanto no modelo 2, ao adicionarmos a variável GOODNEWS, o efeito caiu para 0,57. O p-valor em ambos os modelos foram estatisticamente significantes, mas o modelo 2 apresentou um ajuste melhor aos dados.

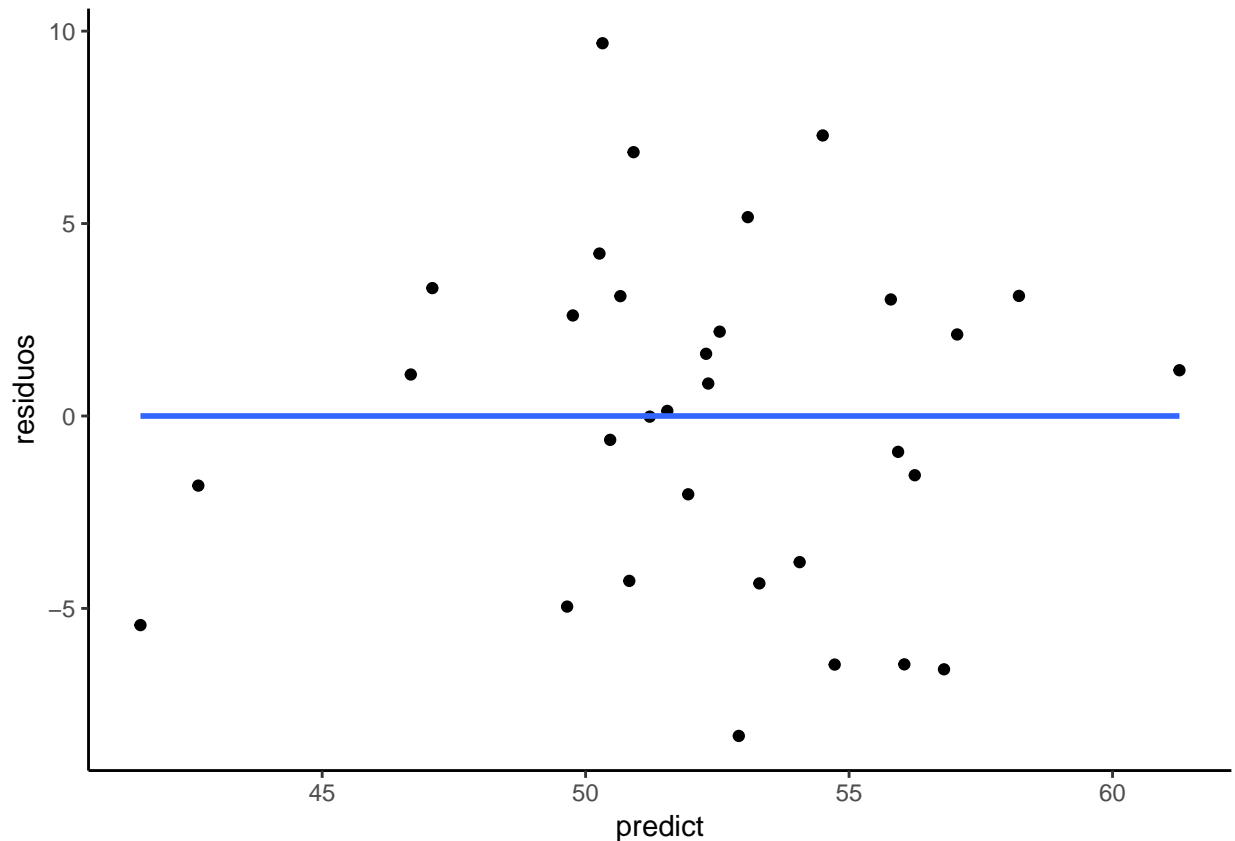
V) Faça uma análise dos resíduos (ui), apresente sua média e indícios para concluir se há ou não há homocedasticidade no modelo.

```
options(scipen=999)
ecovote$residuos <- residuals.lm(regmu)
ecovote$predict <- predict(regmu)
print(mean(ecovote$residuos))
```

```
## [1] 0.0000000000000000426742
```

A média dos resíduos foi aproximadamente 0. Para observar se existe heterocedasticidade no modelo, podemos ver se existe alguma relação entre os resíduos e os valores preditos pelo modelo. Caso os dados estejam aleatoriamente distribuídos ao redor do 0, podemos dizer que o modelo não tem heterocedasticidade. De fato, como mostra o gráfico abaixo, não há relação entre essas variáveis.

```
theme_set(theme_classic())
ggplot(ecovote, aes(x = predict, y = residuos)) +
  geom_point() +
  geom_smooth(method = "lm", se = F)
```



d) Apresente um modelo de regressão multivariada adicionando ao modelo do item b) a variável WAR como variável independente.

```
regmu2 <- lm(VOTE ~ GROWTH + GOODNEWS + WAR, data = ecovote)
```

I) Descreva a variável adicionada ao modelo e sua relação com as demais;

A variável WAR é uma variável binária que representa a presença (ou não) de guerras por ano. É provável que em tempos de guerra, o crescimento econômico seja afetado negativamente. Porém, a relação com votos no incumbente é incerta, pois os eleitores podem apoiar o incumbente almejando estabilidade que ajudaria na resolução da guerra, ou, punir o candidato mandatário por causa da insatisfação de ter o país em guerra. Essa variável pode estar relacionada a boas notícias, pois em tempos de guerra, é provável que a quantidade de boas notícias sejam bem menores.

II) Apresente os resultados do modelo;

```
summary(regmu2)
```

```
##
## Call:
## lm(formula = VOTE ~ GROWTH + GOODNEWS + WAR, data = ecovote)
##
## 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 < 0.0000000000000002 ***
## 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
```

A relação de crescimento econômico e votos do incumbente continua positiva (0,57) e estatisticamente significativa a um p-valor menor que 0,000. As demais variáveis (GOODNEWS e WAR) estão positivamente correlacionadas a variável dependente contudo o nível de significância foram, de ambas, menor que 0,05.

III) Avalie o modelo quanto a seu ajuste e sua capacidade explicativa. Comente e interprete os coeficientes estimados, os intervalos de confiança do modelo e o RMSE e o R2.

O modelo apresenta um R^2 ajustado de 0,4065. Isto quer dizer que o modelo explica aproximadamente 40,5% dos casos. O erro padrão dos resíduos foi de 4,677 que representa o quão próximo os dados reais estão do modelo. Quanto menor o seu valor, melhor.

IV) Compare os resultados do item anterior com os adquiridos com o modelo do item c).

Ambos os modelos mostram que o crescimento econômico e a quantidade de votos obtidos pelo incumbente são positivos e estatisticamente significantes. Ao acrescentarmos uma variável ao modelo (GOODNEWS), o R^2 ajustado foi de aproximadamente 0,43, contudo, esse valor cai para 0,41 ao adicionarmos a variável WAR. Isto é, o modelo ficou levemente menos ajustado. Outro ponto interessante é que a variável GOODNEWS foi estatisticamente significativa a um p-valor menor que 0,05, mas ao adicionarmos WAR no modelo, seu nível de significância caiu para 0,06, tornando-se não significativa a um teste de corte de 0,05.

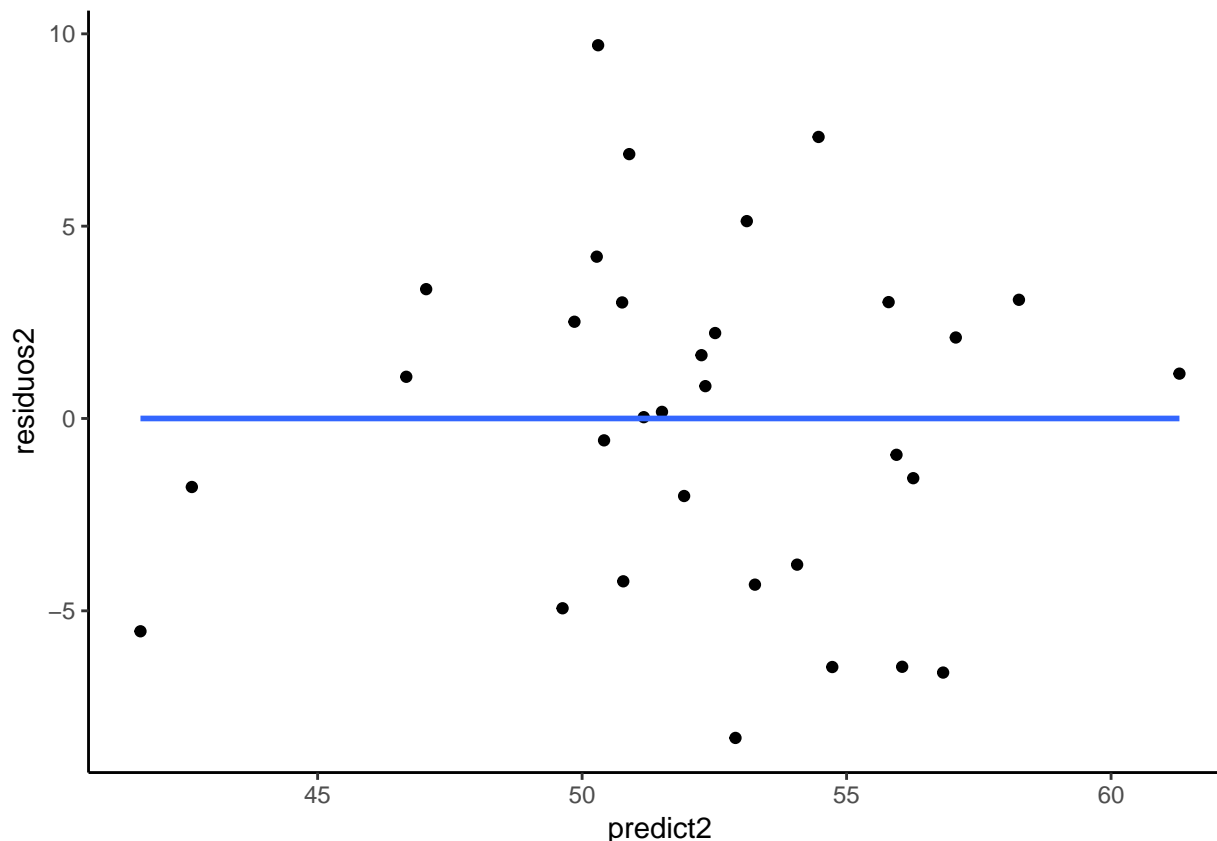
V) Faça uma análise dos resíduos (ui), apresente sua média e indícios para concluir se há ou não há homocedasticidade no modelo.

```
ecovote$residuos2 <- residuals.lm(regmu2)
ecovote$predict2 <- predict(regmu2)
print(mean(ecovote$residuos2))
```

```
## [1] 0.0000000000000003330669
```

A média do resíduo foi aproximadamente 0. Para observar a presença de heterocedasticidade vamos observar a relação entre resíduos e valores preditos. Como o gráfico abaixo mostra, os resíduos se dispõem de forma aleatória ao redor do 0. Isso indica presença de homocedasticidade no modelo.

```
ggplot(ecovote, aes(x = predict2, y = residuos2)) +
  geom_point() +
  geom_smooth(method = "lm", se = F)
```



VI) Qual variável tem maior efeito descritivo sobre a variável dependente? Apresente sua análise.

A variável que apresenta maior efeito sobre os votos no incumbente foi GOODNEWS. Esta apresentou um efeito de 0,72% para cada boa notícia que a mídia apresenta. Porém essa relação obteve um p-valor maior que 0,05. A variável crescimento econômico é mais robusta vista que o p-valor foi menor que 0,000 ainda que a magnitude do seu efeito (0,5729) seja menor do que a de GOODNEWS.

2. Numa das revistas listadas abaixo, identifique um artigo que apresente resultados de um modelo de regressão multivariada e realize as seguintes tarefas:

a) Faça a citação do artigo.

KOOS, Carlos. 2018. “Decay or Resilience? The long-Term social consequences of conflict-related sexual Violence in sierra leone”.World Politics. v. 70. Issue 2.

b) Qual a variável dependente do modelo? Apresente o modelo utilizado no artigo e a justificativa para tal verificando qual a relevância do uso do modelo para a pergunta de pesquisa do artigo.

A variável dependente do modelo é “association” que mede se algum membro da família pertence a associações comunitárias. Por ser uma variável binária, o autor apresentou uma regressão logística.

A pergunta de pesquisa do artigo é: Os conflitos de violência sexual destrói o comportamento social das comunidades, ou, as vítimas e famílias conseguem superar esses eventos traumáticos e demonstrar resiliência? A tabela acima mostra os principais resultados do artigo. Ele mostra, exatamente, o efeito de conflitos

TABLE 2
EFFECT OF CRSV ON MEMBERSHIP IN ASSOCIATIONS^a

	(1) <i>Association</i>	(2) <i>Association</i>	(3) <i>Association</i>	(4) <i>Association</i>
CRSV in HH	0.347** (0.122)	0.343** (0.121)	0.321* (0.124)	0.390** (0.139)
Humanitarian aid			1.189*** (0.157)	1.314*** (0.206)
CRSV in HH*Humanitarian aid				-0.314 (0.262)
Homicide in HH		-0.224 (0.146)	-0.186 (0.148)	-0.189 (0.148)
Mutilation in HH		0.755** (0.242)	0.528* (0.228)	0.559* (0.230)
HH ever displaced		-0.151 (0.150)	-0.165 (0.152)	-0.154 (0.152)
RUF camps		-0.028 (0.035)	-0.016 (0.035)	-0.016 (0.034)
Household controls	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes
District dummies	yes	yes	yes	yes
F-stat	1.221	1.437	0.292	0.339
p-value	0.278	0.168	0.977	0.962
Model	logit	logit	logit	logit
Observations	5475	5475	5475	5475

Robust standard errors clustered by primary sampling unit in parentheses; *p<0.10, **p<0.05, ***p<0.01,
***p<0.001

^aOutcome variable *association* measures whether someone in the household is member of a community-based association.

Figure 1: Tabela de Regressão

relativos à violência sexual no engajamento da família em associações comunitárias. O autor ainda verifica a relação entre os conflitos e ajudas humanitárias através do termo interativo entre essas variáveis.

c) Interprete os resultados do modelo.

O autor mostra que conflitos relativos a violência sexual não inibe a família de participar ativamente da comunidade. Ele mostra que existe um efeito de resiliência por parte dos violentados. Isso é comprovado, no modelo, pois o efeito desses conflitos (CRSV in HH) na participação em associações comunitárias é positivo e estatisticamente significativo em todos os modelos da tabela. O autor testa também se esse efeito só ocorre caso haja ajudas humanitárias na comunidade. De fato, quando existem esses tipos de ajuda, há mais membros nas associações comunitárias, porém o seu efeito interativo com os conflitos não foi significativo, mostrando, assim, uma independência nessa relação. Ou seja, o efeito de resiliência não depende das ajudas humanitárias.