

# Lista 3

## Econometria

### Diogo Wolff Surdi

June 9, 2020

#### Questão 6

(a)

```
library(foreign)
library(tidyr)
library(dplyr)
library(lmtest)
library(sandwich)

dados <- read.dta('(...) / fertil2.dta')

NAs <- dados %>%
  filter_at(vars(children, age, educ, electric, urban), any_vars(is.na(.)))

dados <- anti_join(dados, NAs)
dados$agesquared <- dados$age^2

reg <- lm(children ~ age+agesquared+educ+electric+urban, dados)
erropadrao <- coef(summary(reg))[ ,2]

errorobusto <- coeftest(reg, vcov = vcovHC(reg, type="HC0"))[ ,2]

erros <- cbind(erropadrao, errorobusto)

> erros
erropadrao  errorobusto
(Intercept) 0.2401887611 0.2436830685
age          0.0165082295 0.0191614558
agesquared   0.0002717715 0.0003502695
```

```
educ          0.0062966083  0.0063033633
electric      0.0690044704  0.0639041125
urban         0.0465062118  0.0454396178
```

O erro usual não necessariamente é maior (menor) do que o robusto.

**(b)**

```
library(car)
```

```
reg2 <- lm(children ~ age+agesquared+educ+electric+urban+
            spirit+protest+catholic, dados)
```

```
nullhp<-c('spirit', 'protest', 'catholic')
```

```
> linearHypothesis(reg2, nullhp)
Linear hypothesis test
```

```
Hypothesis:
spirit = 0
protest = 0
catholic = 0
```

Model 1: restricted model

Model 2: children ~ age + agesquared + educ + electric + urban + spirit + protest + catholic

```
Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     4352 9176.4
2     4349 9162.5  3      13.88 2.1961 0.08641 .
```

O p-valor do teste usual é 0.086. Para o teste robusto, temos, utilizando o método do livro:

```
u <- resid(reg)
```

```
spi <- resid(lm(spirit ~ age+agesquared+educ+electric+urban, dados))
pro <- resid(lm(protest ~ age+agesquared+educ+electric+urban, dados))
cat <- resid(lm(catholic ~ age+agesquared+educ+electric+urban, dados))
```

```
p1 <- u*spi
p2 <- u*pro
p3 <- u*cat
```

```
magia <- rep(1, 4358)
```

```
reg3 <- lm(magia ~ 0+p1+p2+p3)
```

```
ssr <- sum(resid(reg3)^2)
```

```
> dim(dados)[1] - ssr
[1] 6.465107
```

```
> pchisq(dim(dados)[1] - ssr, df=3, lower.tail=FALSE)
[1] 0.0910488
```

Logo o p-valor é 0.091. Assim, para ambos os testes, rejeitamos a hipótese ao nível 0.1, mas não a rejeitamos a níveis usuais menores.

**(c)**

```
fit <- fitted(reg)
fit2 <- fitted(reg2)
fit2sq <- fit2^2
```

```
itemc <- lm(resid^2 ~ fit2 + fit2sq)
> itemc
```

Call:

```
lm(formula = resid^2 ~ fit2 + fit2sq)
```

Coefficients:

```
(Intercept)          fit2          fit2sq
0.3126      -0.1489      0.2668
```

```
nula <- c('fit2', 'fit2sq')
```

```
> linearHypothesis(itemc, nula)
Linear hypothesis test
```

Hypothesis:

```
fit2 = 0
```

```
fit2sq = 0
```

Model 1: restricted model

Model 2: `resid^2 ~ fit2 + fit2sq`

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	4357	76589			
2	4355	57436	2	19153	726.11 < 2.2e-16 ***

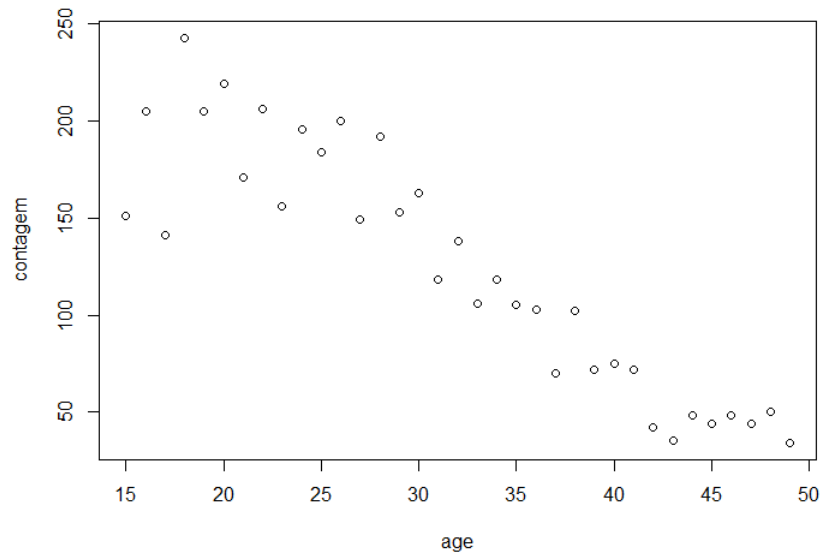
Como pode-se ver, rejeitamos a hipótese de que o quadrado do erro não é correlacionado com essas duas variáveis, logo a variância do erro apresenta correlação com elas, e há heteroscedasticidade.

**(d)**

```

cont <- dados %>% group_by(age) %>% summarise(contagem = sum(age))
cont$contagem <- cont$contagem/cont$age
plot(cont)

```



Como pode-se ver, há muito menos dados ao longo do tempo, então afirmo que a variância do erro ser correlacionada com a idade não é algo muito relevante, pois há muito menos dados conforme a idade aumenta, logo é natural que o erro da regressão seja maior.