Cap.8: HETEROCEDASTICIDADE

EXEMPLO 8.

No exemplo em questão, é feito o teste Breush-Pagan.

1º: estima-se a regressão e salva-se os resíduos:

```
reg1 <- lm(V1~V4+V5+V3)
```

Obtendo os resíduos da regressão:

```
reg1$resid
```

**nome da regressão$resid.

Como vamos fazer uma regressão com o quadrado dos resíduos, podemos criar um nome para os resíduos:

```
resid <- reg1$resid
```

**mas nada impede de ser usado reg1$resid, para se referir aos resíduos.

2º: Estima-se uma nova regressão, desta vez com o quadrado dos resíduos e as variáveis exógenas da primeira regressão:

```
resid1 <- lm(resid^2~V4+V5+V3)
```

```
summary(resid1)
```

O que nos interessa é o R^2 desta regressão (Multiple R-squared), para calcular a estatística LM. O procedimento é o seguinte:

```
nrow(hprice1)*summary(resid1)$r.squared,
```

ou, de uma forma mais rápida:

88*0.1601 (número de linhas * R^2)

O valor obtido é utilizado para computar o p-valor da estatística LM, por meio de uma distribuição de probabilidade Qui-quadrado:

```
pchisq(c(14.09239), df=3, lower.tail=FALSE)
```

Lembrando que a hipótese nula é:

Os erros são homocedásticos,

Um p-valor abaixo de 0.05 implica que rejeitamos a hipótese nula, implicando a presença de heterocedasticidade no modelo estimado.

Para tentar corrigir a heterocedasticidade, basta estimar um modelo onde todas as variáveis serão transformadas em logarítmos. A regressão a ser estimada ficará da seguinte forma:

Logreg <- lm(log(V1)~log(V4)+log(V5)+V3)

Os procedimentos são os mesmos descritos para a outra regressão.

**Usamos log quando nos referimos ao logarítmo natural. Portanto, o R não reconhece o comando ln.

**Calculando o teste Breush-Pagan de forma direta:

library(lmtest, pos=4)

bptest(log(V1) ~ V3 + log(V4) + log(V5), studentize=TRUE)

EXEMPLO 8.5

Um outro teste para verificar heterocedasticidade é o teste de White. No entanto este teste não vem em nenhum pacote do R diferentemente do teste de Breush-Pagan. Então estimamos o teste em 3 estapas:

1º Estimamos a regressão e salvamos os resíduos e os valores ajustados.

Call:

lm(formula = log(V1) ~ log(V4) + log(V5) + V3)

Residuals:

Min	1Q	Median	3Q	Max
-0.68422	-0.09178	-0.01584	0.11213	0.66899

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.29704	0.65128	-1.992	0.0497 *
log(V4)	0.16797	0.03828	4.388	3.31e-05 ***
log(V5)	0.70023	0.09287	7.540	5.01e-11 ***

V3	0.03696	0.02753	1.342	0.1831

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1846 on 84 degrees of freedom

Multiple R-squared: 0.643,   Adjusted R-squared: 0.6302

F-statistic: 50.42 on 3 and 84 DF, p-value: < 2.2e-16

*obtendo os resíduos:

```
resid1 <- Logreg$resid
```

***como no exemplo anterior já tinha sido usado o nome resid, temos que diferenciar o nome dos resíduos desta regressão, por isso foi usado o nome resid1

*para encontrarmos os valores ajustados

```
fitted <- Logreg$fitted
```

2º Regredimos o $\hat{u}^2$ em nos valores

```
residlog <- lm(resid1^2~fitted+I(fitted^2))
```

**É necessário colocarmos a letra I na frente do Segundo termo, senão a regressão não será estimada corretamente.

```
summary(residlog)
```

Call:

lm(formula = resid^2 ~ fitted + I(fitted^2))

Residuals:

Min	1Q	Median	3Q	Max
-0.060042	-0.031773	-0.013643	0.005276	0.418079

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)

(Intercept)	5.0468	3.3450	1.509	0.135
fitted	-1.7092	1.1633	-1.469	0.145
I(fitted^2)	0.1451	0.1010	1.437	0.154

Residual standard error: 0.07299 on 85 degrees of freedom

Multiple R-squared: 0.03917,   Adjusted R-squared: 0.01657

F-statistic: 1.733 on 2 and 85 DF, p-value: 0.1830

3º Multplica-se o número de linhas pelo R^2 da regressão estimada, para encontrar a estatística LM,

```
> 88*0.0392
[1] 3.4496
```

4º Depois calcula-se a distribuição Qui-quadrado:

```
pchisq(c(3.45), df=2, lower.tail=FALSE)
[1] 0.1781731
```

EXEMPLO 8.7

Estimando a regressão via Minímos Quadrados Generalizados Factíveis.

O processo consiste em:

1- estimar a regressão convencional, via MQO;

2- Salvar os resíduos, e estimar uma nova regressão com o log dos resíduos ao quadrado, e as mesmas variáveis do modelo de MQO;

3- Salvar os valores ajustados $\hat{g}$ desta nova regressão, e usá-los como ponderação da seguinte forma: $\hat{h} = \exp(\hat{g})$.

4- Reestimar novamente a equação via MQO, porém usando como ponderação 1/ $\hat{h}$.

Este processo no R pode ser feito da seguinte forma:

1º: Importe os dados SMOKE.RAW, já transformado em arquivo texto.

2º: Estime a seguinte regressão:

```
> reg87 <- lm(V6~V8+V10+V1+V4+V9+V7)
> summary(reg87)
```

Call:

lm(formula = V6 ~ V8 + V10 + V1 + V4 + V9 + V7)

Residuals:

Min	1Q	Median	3Q	Max
-15.819	-9.381	-5.975	7.922	70.221

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3.639868	24.078661	-0.151	0.87988
V8	0.880269	0.727784	1.210	0.22682
V10	-0.750854	5.773343	-0.130	0.89656
V1	-0.501498	0.167077	-3.002	0.00277 **
V4	0.770694	0.160122	4.813	1.78e-06 ***
V9	-0.009023	0.001743	-5.176	2.86e-07 ***
V7	-2.825085	1.111794	-2.541	0.01124 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.4 on 800 degrees of freedom

Multiple R-squared: 0.05274,   Adjusted R-squared: 0.04563

F-statistic: 7.423 on 6 and 800 DF, p-value: 9.499e-08

3º: Faça o teste para detectar a presença de heterocedasticidade,

```
> library(lmtest, pos=4)

> bptest(V6 ~ V1 + V4 + V7 + V8 + V9 + V10, studentize=TRUE, data=smoke)
```

4º: Salve os residuos da regressão estimada:

```
> resid87 <- reg87$resid
```

5º: Estime novamente a regressão usando como variável explicativa o logarítmo natural dos resíduos ao quadrado.

```
> regresid <- lm(log(resid87^2)~V8+V10+V1+V4+V9+V7)

> summary(regresid)
```

6º: salve os valores ajustados desta regressão:

```
> fitted <- regresid$fitted
```

7º: Reestime a equação usando ponderação: weights=1/exp(fitted)

```
> reg <- lm(V6~V8+V10+V1+V4+V9+V7,weights=1/exp(fitted))

> summary(reg)


Call:

lm(formula = V6 ~ V8 + V10 + V1 + V4 + V9 + V7, weights = 1/exp(fitted))


Residuals:

    Min          1Q           Median          3Q          Max

-1.9036     -0.9532        -0.8099        0.8415      9.8556


Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)  5.6353434 17.8031310   0.317 0.751678

V8           1.2952413  0.4370119   2.964 0.003128 **

V10         -2.9402848  4.4601431  -0.659 0.509934

V1          -0.4634462  0.1201586  -3.857 0.000124 ***

V4           0.4819474  0.0968082   4.978 7.86e-07 ***

V9          -0.0056272  0.0009395  -5.990 3.17e-09 ***
```

```
V7          -3.4610662  0.7955046  -4.351 1.53e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.579 on 800 degrees of freedom
Multiple R-squared: 0.1134,     Adjusted R-squared: 0.1068
F-statistic: 17.06 on 6 and 800 DF,  p-value: < 2.2e-16
```

Exemplo 8.1

O exemplo consiste em usar a metodologia de correção de White para o desvio padrão de cada um dos coeficientes da regressão.

O exemplo utiliza variáveis dummy no modelo de regressão:

1º Precisamos criar as variáveis marrmale (homens casados), marrfem(mulheres casadas), singfem(mulheres solteiras).

Sabemos que a dummy female, vale 1 se mulher e 0 se homem, e que a dummy married, vale 1 se casado e 0 se solteiro

marr <- V7==1 (casado)

female <- V6==1 (mulher)

male <- V6==0 (homem)

sing <- V7==0 (solteiro)

educ <- V2

exper <- V3

expersq <- V23

ten <- V4

tensq <- V24

marrmale <- marr*male

marrfem <- marr*female

```
singfem <- sing*female

lwage <- V22

reg3 <- lm( lwage~marrmale+marrfem+singfem+educ+exper+expersq+ten+tensq)

summary(reg3)
```

Para calcular erros robustos da variável educ:

calculamos uma nova regressão tendo educ como variável endógena, da seguinte forma:

```
reg4 <- lm(educ~marrmale+marrfem+singfem+exper+expersq+ten+tensq)

> summary(reg4)
```

Precisamos dos residuos ao quadrado da regressão 3 e 4 deste exemplo,

```
resid3 <- reg3$resid^2

resid4 <- reg4$resid^2

> dp.educ <- sqrt(sum(resid4*resid3)/(sum(resid4))^2)

> dp.educ

[1] 0.007350961
```