Exercício semana 12

Vinícius Araújo Kluska - GRR20183601

25/07/2021

# Getting data related to the Exercise\_Week\_12

Open data base “MEAP93”

T\_MEAP93<-read.table("C:/Users/klusk/Projects/R-studies/Examples/MEAP93.txt", header= TRUE, sep="")

# Exercise\_1

Utilizando a tabela MEAP93.txt, estime as regressões a seguir:

Reg1: math10 = B1 + B2enroll + B3sttaff + B4totcomp + u

Reg2: math10 = B1 + B2log(enroll) + B3log(sttaff) + B4log(totcomp) + u

Reg3: math10 = B1 + B2log(enroll) + B3log(sttaff) + B4log(totcomp) + B5lnchprg + u

Analise os testes t e R2 ajustados obtidos. Qual das regressões apresenta melhores resultados? Por quê?

# Exercise\_2

Baseado nos testes t de Reg3, exclua a(s) variável(is) que não seja(m) significativa(s). Estime a nova regressão, que chamaremos de Reg4.

# Exercise\_3

Faça os testes de autocorrelação dos resíduos de Breuch-Godfrey e Durbin Watson em Reg1 e Reg4. Houve melhora na regressão? Por quê?

# Exercise\_4

Aplique o teste de White com e sem termos cruzados na Reg1 e Reg4. O que você conclui a respeito da hipótese de homocedasticidade?

# Exercise\_5

Utilize as variâncias robustas de White na Reg4. Quais seus efeitos sobre os testes t?

## Resolução

1)

reg1\_MEAP93=lm(math10~enroll+staff+totcomp, data = T\_MEAP93)  
reg2\_MEAP93=lm(math10~log(enroll)+log(staff)+log(totcomp), data = T\_MEAP93)  
reg3\_MEAP93=lm(math10~log(enroll)+log(staff)+log(totcomp)+lnchprg, data = T\_MEAP93)  
print(summary(reg1\_MEAP93))

##   
## Call:  
## lm(formula = math10 ~ enroll + staff + totcomp, data = T\_MEAP93)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22.235 -7.008 -0.807 6.097 40.689   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.2740214 6.1137938 0.372 0.710   
## enroll -0.0001976 0.0002152 -0.918 0.359   
## staff 0.0479199 0.0398140 1.204 0.229   
## totcomp 0.0004586 0.0001004 4.570 6.49e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.24 on 404 degrees of freedom  
## Multiple R-squared: 0.05406, Adjusted R-squared: 0.04704   
## F-statistic: 7.697 on 3 and 404 DF, p-value: 5.179e-05

print(summary(reg2\_MEAP93))

##   
## Call:  
## lm(formula = math10 ~ log(enroll) + log(staff) + log(totcomp),   
## data = T\_MEAP93)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22.735 -6.838 -0.835 6.139 39.718   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -207.6649 48.7031 -4.264 2.50e-05 \*\*\*  
## log(enroll) -1.2680 0.6932 -1.829 0.0681 .   
## log(staff) 3.9800 4.1897 0.950 0.3427   
## log(totcomp) 21.1550 4.0555 5.216 2.92e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.18 on 404 degrees of freedom  
## Multiple R-squared: 0.06538, Adjusted R-squared: 0.05844   
## F-statistic: 9.42 on 3 and 404 DF, p-value: 4.974e-06

print(summary(reg3\_MEAP93))

##   
## Call:  
## lm(formula = math10 ~ log(enroll) + log(staff) + log(totcomp) +   
## lnchprg, data = T\_MEAP93)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -26.069 -6.379 -0.674 4.993 40.560   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -112.63268 46.61662 -2.416 0.01613 \*   
## log(enroll) -1.82805 0.64625 -2.829 0.00491 \*\*   
## log(staff) 4.85533 3.88546 1.250 0.21216   
## log(totcomp) 12.88530 3.89286 3.310 0.00102 \*\*   
## lnchprg -0.30529 0.03727 -8.191 3.46e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.439 on 403 degrees of freedom  
## Multiple R-squared: 0.1988, Adjusted R-squared: 0.1908   
## F-statistic: 24.99 on 4 and 403 DF, p-value: < 2.2e-16

Análise:

A primeira (reg1\_MEAP93), apresenta somente “totcomp” como significativa a 5% de significância e R Quadrado ajustado igual a 0,04704.

A segunda (reg2\_MEAP93), apresenta o intercepto, “log(enroll)” e “log(totcomp)” significativos a 5% de significância e R Quadrado ajustado igual a 0,05844.  
A terceira (reg3\_MEAP93), apresenta o intercepto, “log(enroll)”, “log(totcomp)” e “lnchprg” significativos a 5% de significância e R Quadrado ajustado igual a 0.1908.

Portanto, concluísse pela preferência por “reg3\_MEAP93”, dado a maior quantidade de variáveis explicativas significativas a 5% e ao aumento de R Quadrado ajustado pela inclusão da última variável.

2)

reg4\_MEAP93=lm(math10~log(enroll)+log(totcomp)+lnchprg, data = T\_MEAP93)  
print(summary(reg4\_MEAP93))

##   
## Call:  
## lm(formula = math10 ~ log(enroll) + log(totcomp) + lnchprg, data = T\_MEAP93)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25.930 -5.849 -0.504 4.971 42.709   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -79.56073 38.40213 -2.072 0.03892 \*   
## log(enroll) -1.89236 0.64465 -2.935 0.00352 \*\*   
## log(totcomp) 11.91087 3.81661 3.121 0.00193 \*\*   
## lnchprg -0.30401 0.03728 -8.154 4.48e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.446 on 404 degrees of freedom  
## Multiple R-squared: 0.1957, Adjusted R-squared: 0.1897   
## F-statistic: 32.76 on 3 and 404 DF, p-value: < 2.2e-16

3)

library(lmtest)  
BGtest\_1 <- bgtest(reg1\_MEAP93)  
BGtest\_4 <- bgtest(reg4\_MEAP93)  
print(BGtest\_1)

##   
## Breusch-Godfrey test for serial correlation of order up to 1  
##   
## data: reg1\_MEAP93  
## LM test = 9.4643, df = 1, p-value = 0.002095

print(BGtest\_4)

##   
## Breusch-Godfrey test for serial correlation of order up to 1  
##   
## data: reg4\_MEAP93  
## LM test = 0.58733, df = 1, p-value = 0.4435

Análise:

Sobre os testes de Breuch-Godfrey:

Recusa-se H0 em “reg1\_MEAP93”, pois o P-valor é inferior a 5% (0.002095), denotando assim, possível presença de autocorrelação dos resíduos pois não é possível aceitar que os ros sejam iguais a 0.

Recusa-se H1 em “reg4\_MEAP93”, pois o P-valor é superior a 5% (0.4435), denotando assim, a ausência de autocorrelação dos resíduos pois é possível aceitar que os ros sejam iguais a 0.

Sobre os testes de Durbin Watson:

Recusa-se H0 em “reg1\_MEAP93”, pois Dw calculado (1.6689) é inferior a Dl (1.821), denotando assim, possível presença de autocorrelação positiva.

Recusa-se H1 em “reg4\_MEAP93”, pois Dw calculado (1.9091) é superior a Du (1.821) e inferior a 4-Du (2.179), denotando assim, a falta de evidências para existência de autocorrelação.

Resultado: Dada a análise, é correto admitir melhora na regressão. Passou de um modelo com problemas relativos à autocorrelação, para um modelo melhor ajustado sem evidências de autocorrelação.

4)

library(skedastic)  
White\_test\_1=white\_lm(reg1\_MEAP93, interactions=FALSE, statonly=FALSE)  
White\_test\_crossed\_1=white\_lm(reg1\_MEAP93, interactions=TRUE, statonly=FALSE)  
White\_test\_4=white\_lm(reg4\_MEAP93, interactions=FALSE, statonly=FALSE)  
White\_test\_crossed\_4=white\_lm(reg4\_MEAP93, interactions=TRUE, statonly=FALSE)  
print(White\_test\_1)

## # A tibble: 1 x 5  
## statistic p.value parameter method alternative  
## <dbl> <dbl> <dbl> <chr> <chr>   
## 1 42.7 0.000000133 6 White's Test greater

print(White\_test\_4)

## # A tibble: 1 x 5  
## statistic p.value parameter method alternative  
## <dbl> <dbl> <dbl> <chr> <chr>   
## 1 39.6 0.000000536 6 White's Test greater

print(White\_test\_crossed\_1)

4)

## # A tibble: 1 x 5  
## statistic p.value parameter method alternative  
## <dbl> <dbl> <dbl> <chr> <chr>   
## 1 43.5 0.00000178 9 White's Test greater

print(White\_test\_crossed\_4)

## # A tibble: 1 x 5  
## statistic p.value parameter method alternative  
## <dbl> <dbl> <dbl> <chr> <chr>   
## 1 42.5 0.00000265 9 White's Test greater

Análise:

Para todos os testes realizados, tanto para “reg1\_MEAP93” e”reg4\_MEAP93”, há indícios de heterocedasticidade. Isto da-se pelo fato de que todos os p-valores encontrados são inferiores há 5%, caracterizando a recusa de H0: homoscedasticidade.

5)

library(lmtest)  
library(sandwich)  
T\_robust=coeftest(reg4\_MEAP93, vcov=vcovHC(reg4\_MEAP93, type = 'HC0'))  
print(T\_robust)

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -79.560726 39.542679 -2.0120 0.04488 \*   
## log(enroll) -1.892356 0.791969 -2.3894 0.01733 \*   
## log(totcomp) 11.910866 4.034712 2.9521 0.00334 \*\*   
## lnchprg -0.304006 0.033897 -8.9686 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Análise:

O teste mantém o estabelecido pelo teste t: todos os parâmetros são significantes a 5%¨. A variância de White é uma forma de conferir maior peso para as informações que apresentam menor variância (informações mais próximas a média populacional).