1. a) Teste RESET

b) Equação:

$$\log(\text{salário}_i) = \beta_0 + \beta_1 \cdot educ_i + \beta_2 \cdot anosemp_i + \delta_1 \cdot \widehat{\log(\text{salário}_i)}^2 + \delta_2 \cdot \widehat{\log(\text{salário}_i)}^3 + \varepsilon_i$$

$H_0: \delta_1 = \delta_2 = 0$

$H_A: \delta_1 \neq 0$ e/ou $\delta_2 \neq 0$

$$F = \frac{(R^2_{ir} - R^2_R)/2}{(1 - R^2_{ir})/(n-k-3)} \sim F_{[2;\, n-k-3]}$$

c)

Ramsey RESET Test
Equation: UNTITLED
Specification: LOG(SALARIO) C EDUC ANOSEMP
Omitted Variables: Powers of fitted values from 2 to 3

	Value	df	Probability
F-statistic	1.839905	(2, 41)	0.1717
Likelihood ratio	3.953684	2	0.1385

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	0.062142	2	0.031071
Restricted SSR	0.754523	43	0.017547
Unrestricted SSR	0.692381	41	0.016887
Unrestricted SSR	0.692381	41	0.016887

LR test summary:

	Value	df
Restricted LogL	29.26597	43
Unrestricted LogL	31.24281	41

Unrestricted Test Equation:
Dependent Variable: LOG(SALARIO)
Method: Least Squares
Date: 05/27/14 Time: 21:48
Sample: 1 46
Included observations: 46

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4404.914	2712.429	1.623974	0.1120
EDUC	30.26849	18.60505	1.626897	0.1114
ANOSEMP	10.65851	6.551183	1.626959	0.1114
FITTED^2	-62.22936	38.49690	-1.616477	0.1137
FITTED^3	1.936675	1.204158	1.608323	0.1154

R-squared	0.770726	Mean dependent var	10.55832
Adjusted R-squared	0.748358	S.D. dependent var	0.259053
S.E. of regression	0.129951	Akaike info criterion	-1.140992
Sum squared resid	0.692381	Schwarz criterion	-0.942226
Log likelihood	31.24281	Hannan-Quinn criter.	-1.066533
F-statistic	34.45641	Durbin-Watson stat	1.457412
Prob(F-statistic)	0.000000		

Como p-valor > α(5%), não há evidências para rejeitar H_0. Assim, com 95% de confiança, encontramos evidências a favor da especificação do modelo.

2. a) Modelo estimado

Dependent Variable: LO_SALAR
Method: Least Squares
Date: 05/07/14 Time: 22:09
Sample: 1 177
Included observations: 177

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.571977	0.253466	18.03781	0.0000
LO_SALES	0.187787	0.040003	4.694340	0.0000
LO_MKT	0.099872	0.049214	2.029345	0.0440
PROFMARG	-0.002211	0.002105	-1.050132	0.2951
COMTEN	-0.009238	0.003337	-2.767983	0.0063
CEOTEN	0.017104	0.005540	3.087309	0.0024

R-squared	0.352537	Mean dependent var	6.582848
Adjusted R-squared	0.333606	S.D. dependent var	0.606059
S.E. of regression	0.494744	Akaike info criterion	1.463759
Sum squared resid	41.85601	Schwarz criterion	1.571425
Log likelihood	-123.5427	Hannan-Quinn criter.	1.507424
F-statistic	18.62159	Durbin-Watson stat	2.010839
Prob(F-statistic)	0.000000		

Teste RESET

Ramsey RESET Test
Equation: UNTITLED
Specification: LO_SALAR C LO_SALES LO_MKT PROFMARG COMTEN CEOTEN
Omitted Variables: Powers of fitted values from 2 to 3

	Value	df	Probability
F-statistic	4.553727	(2, 169)	0.0119
Likelihood ratio	9.290434	2	0.0096

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	2.140290	2	1.070145
Restricted SSR	41.85601	171	0.244772
Unrestricted SSR	39.71572	169	0.235004
Unrestricted SSR	39.71572	169	0.235004

LR test summary:

	Value	df
Restricted LogL	-123.5427	171
Unrestricted LogL	-118.8975	169

Unrestricted Test Equation:
Dependent Variable: LO_SALAR
Method: Least Squares
Date: 05/07/14 Time: 22:10
Sample: 1 177
Included observations: 177

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	448.0912	185.4682	2.416001	0.0168
LO_SALES	36.00905	14.75667	2.440188	0.0157
LO_MKT	19.21709	7.858795	2.445297	0.0155
PROFMARG	-0.422816	0.173468	-2.437432	0.0158
COMTEN	-1.773576	0.726437	-2.441472	0.0157
CEOTEN	3.282331	1.344547	2.441216	0.0157
FITTED^2	-28.22893	11.81594	-2.389054	0.0180
FITTED^3	1.387449	0.590909	2.347991	0.0200

R-squared	0.385645	Mean dependent var	6.582848
Adjusted R-squared	0.360199	S.D. dependent var	0.606059
S.E. of regression	0.484772	Akaike info criterion	1.433870
Sum squared resid	39.71572	Schwarz criterion	1.577425
Log likelihood	-118.8975	Hannan-Quinn criter.	1.492090
F-statistic	15.15505	Durbin-Watson stat	2.062631
Prob(F-statistic)	0.000000		

P-valor < α → rejeitar H_0
Há evidências contrárias à especificação do modelo proposto.

b)

Dependent Variable: LO_SALAR
Method: Least Squares
Date: 05/07/14 Time: 22:14
Sample: 1 177
Included observations: 177

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.423713	0.265604	16.65532	0.0000
LO_SALES	0.185673	0.039802	4.664887	0.0000
LO_MKT	0.101761	0.048719	2.088757	0.0382
PROFMARG	-0.002575	0.002088	-1.233406	0.2191
COMTEN	-0.006063	0.011892	-0.509860	0.6108
CEOTEN	0.047716	0.014156	3.370634	0.0009
COMTEN^2	-5.39E-05	0.000253	-0.213141	0.8315
CEOTEN^2	-0.001119	0.000481	-2.323751	0.0213

R-squared	0.374475	Mean dependent var	6.582848
Adjusted R-squared	0.348565	S.D. dependent var	0.606059
S.E. of regression	0.489160	Akaike info criterion	1.451889
Sum squared resid	40.43785	Schwarz criterion	1.595444
Log likelihood	-120.4922	Hannan-Quinn criter.	1.510109
F-statistic	14.45327	Durbin-Watson stat	2.019652
Prob(F-statistic)	0.000000		

$R^2_{a_1} = 0,333$ $\quad$ $R^2_{a_2} = 0,3486$

$R^2_1 = 0,3525$ $\quad$ $R^2_2 = 0,3746$

Podemos observar que, individualmente, ceoten² é significante a 5%, porém comten² não é utilizando nenhum α usual. Ainda, observamos que tanto o R^2 quanto R^2_a aumentaram, evidenciando uma melhora do poder explicativo do modelo.

c)

Ramsey RESET Test
Equation: UNTITLED
Specification: LO_SALAR C LO_SALES LO_MKT PROFMARG COMTEN CEOTEN COMTEN^2 CEOTEN^2
Omitted Variables: Powers of fitted values from 2 to 3

	Value	df	Probability
F-statistic	3.168073	(2, 167)	0.0446
Likelihood ratio	6.591292	2	0.0370

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	1.478169	2	0.739084
Restricted SSR	40.43785	169	0.239277
Unrestricted SSR	38.95968	167	0.233291
Unrestricted SSR	38.95968	167	0.233291

LR test summary:

	Value	df
Restricted LogL	-120.4922	169
Unrestricted LogL	-117.1965	167

Unrestricted Test Equation:
Dependent Variable: LO_SALAR
Method: Least Squares
Date: 05/07/14 Time: 22:23
Sample: 1 177
Included observations: 177

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	333.6449	155.8752	2.140462	0.0338
LO_SALES	28.22537	13.09094	2.156099	0.0325
LO_MKT	15.50846	7.181237	2.159580	0.0322
PROFMARG	-0.390423	0.181244	-2.154136	0.0327
COMTEN	-0.926698	0.428891	-2.160685	0.0321
CEOTEN	7.256487	3.364837	2.156564	0.0325
COMTEN^2	-0.008116	0.003791	-2.140856	0.0337
CEOTEN^2	-0.170153	0.078903	-2.156484	0.0325
FITTED^2	-22.37616	10.59405	-2.112143	0.0362
FITTED^3	1.101398	0.529464	2.080215	0.0390

R-squared	0.397340	Mean dependent var	6.582848
Adjusted R-squared	0.364861	S.D. dependent var	0.606059
S.E. of regression	0.483003	Akaike info criterion	1.437249
Sum squared resid	38.95968	Schwarz criterion	1.616692
Log likelihood	-117.1965	Hannan-Quinn criter.	1.510024
F-statistic	12.23388	Durbin-Watson stat	2.071773
Prob(F-statistic)	0.000000		

P-valor $< \alpha \rightarrow$ rejeitar H_0
Há evidências contrárias a especificação do modelo proposto.

b) $H_0: \beta_2 = \beta_3 = 0$

H_A: pelo menos 1 diferente de zero

Modelo restrito: $TestsCR_i = \beta_0 + \beta_1 \cdot STR_i + \varepsilon_i$

Em seguida estimar:

$\hat{\varepsilon}_{R_i} = \beta_0 + \beta_1 STR_i + \beta_2 El\_PCT_i + \beta_3 Meal\_PCT_i + \eta_i$

$R^2_{\hat{\varepsilon}_R} = 0,7623$

$n = 420$

$LM = 420 \cdot 0,7623 = 320,166$

$\chi^2_{(2)}$ crítico $= 4,6$ $\quad RC = [4,6; +\infty)$

$LM_{obs} \in RC \rightarrow$ há evidências p/ rejeitar H_0.

③ a) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

modelo Restrito: $TestScr_i = \beta_0 + e_i$

↳ Teste LM (fazendo no eviews)

i) Estime o modelo restrito e salve os resíduos $\hat{e}_i$

ii) Estime o modelo Abaixo e guarde seu $R^2_{\hat{e}_i}$

$$\hat{e}_i = \delta_0 + \delta_1 STR_i + \delta_2 EL\_PCT_i + \delta_3 MEAL\_PCT_i + \epsilon_i$$

iii) Compute a estatística LM

$$LM = n R^2_{\hat{e}_i} \sim \chi^2_{(q)} \quad \text{em que } q = 3$$

Fazendo isso temos $LM_{obs} = 420 \cdot 0,7745 = 325,29$

$$LM_{crit} = 6,2513$$

como $LM_{obs} > LM_{crit}$ → rej. H_0

b) $H_0: \beta_2 = \beta_3 = 0$

Repetimos os passos para o teste LM e chegamos em

$LM_{obs} = 420 \cdot 0,7623 = 320,166$

$LM_{crit} = 5,9914$

como $LM_{obs} > LM_{crit}$ → rej. H_0

(4) a) (Estimação no Eviews)

modelo auxiliar do teste:

$$y_i = \beta_0 + \beta_1 K_i + \beta_2 L_i + \gamma_1 \hat{y}_i^2 + \gamma_2 \hat{y}_i^3 + \epsilon_i$$

$H_0: \gamma_1 = \gamma_2 = 0$ (Teste F parcial)

$$\left.\begin{array}{l} F_{obs} = 2{,}5404 \\ F^{10\%}_{crítico\,(2,295)} = 2{,}32 \end{array}\right\} \text{rej } H_0 \rightarrow \text{modelo mal especificado}$$

b) passando o $\ln(\cdot)$ dos dois lados do modelo

$$\ln(y_i) = \beta_0 + \beta_1 \ln(k_i) + \beta_2 \ln(l_i) + \varepsilon_i$$

modelo auxiliar do teste:

$$\ln(y_i) = \beta_0 + \beta_1 \ln(k_i) + \beta_2 \ln(l_i) + \gamma_1 \widehat{\ln(y_i)}^2 + \gamma_2 \widehat{\ln(y_i)}^3 + \epsilon_i$$

$H_0: \gamma_1 = \gamma_2 = 0$ (Teste F parcial)

$$\left.\begin{array}{l} F_{obs} = 0{,}644 \\ F^{10\%}_{crítico\,[2,295]} = 2{,}32 \end{array}\right\} \text{não rej. } H_0 \rightarrow \text{modelo bem especificado.}$$

c) queremos testar $H_0: \beta_1 + \beta_2 = 1$

Podemos escrever a Hipótese na forma matricial:

$$\begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = 1 \iff R\beta = r$$

Podemos usar a estatística do teste LM:

$$LM = \frac{1}{\hat{\sigma}^2_R} \left(R\hat{\beta} - r\right)' \left[R (X'X)^{-1} R' \right]^{-1} \left(R\hat{\beta} - r\right) \sim \chi^2_q$$

que pode ser facilmente computada com a ajuda do Eviews:

$$\left.\begin{aligned} LM &= 3{,}44 \\ \chi^{2\ 10\%}_{(1)} &= 2{,}7055 \end{aligned}\right\}$$

rejeita H_0 $\Rightarrow$ Retornos não são ~~decrescentes~~ de escala. constantes

(5) ~~[illegible]~~ (todos os estimativas foram feitos no Eviews)

a) Estimando ~~o~~ modelo:

$$Buchanan_i = \beta_0 + \beta_1\, gore + \varepsilon_i$$

e fazendo histograma dos ~~[illegible]~~ resíduos, notamos a presenso de outliers. Também podemos observar isso também pelo gráfico de dispersão gore x Buchanan

b) O número de votos errados pode ser ~~[illegible]~~ estimado colocando uma dummy para o município de Palm Beach. A interpretasão do coeficiente estimado associado à ela é "número de votos que Buchanan recebeu a mais em Palm Beach, além do que seria predito pelos outros 66 municípios". Temos então o modelo

$$Buchanan_i = \beta_0 + \beta_1\, gore_i + \beta_2\, DumPalm_i + \varepsilon_i$$

e queremos testar

$$H_0: \beta_2 \leq 975$$
$$H_a: \beta_2 > 975$$

A rejeisão da ipótese nula nos diria que o número de votos errados seria o suficiente para mudar o resultado da eleisão no estado da Flórida.

$$t_{calc} = \frac{\hat{\beta}_2 - 975}{Se(\hat{\beta}_2)} = \frac{2614,51 - 975}{149,32} = 10,98 \Rightarrow p\text{-valor} \approx 0,0000$$

rejeitamos H_0

c) Teste BP no modelo do item (a).

i) Estime o modelo $Buchanan_i = \beta_0 + \beta_1 Gore_i + \varepsilon_i$ e compute os quadrados dos resíduos $\hat{\varepsilon}_i^2$.

ii) Estime o modelo $\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 Total_i + \epsilon_i$ e guarde o $R^2_{\hat{\varepsilon}_i^2}$

iii) compute a estatística LM do teste BP

$$n . R^2_{\hat{\varepsilon}_i^2} = 67 . 0{,}2217 = 14{,}85 \iff p\text{-valor} = 0{,}0001$$

ou seja, rejeitamos a hipótese nula de que os termos de erro do modelo do item (a) são homocedásticos.

d) Agora temos em mente o modelo:

$$\frac{Buchanan_i}{Total_i} = \beta_0 + \beta_1 \left(\frac{Gore_i}{Total_i} \right) + \epsilon_i$$

Para testar se o número de votos a mais em palm beach seria o suficiente para mudar o resultado da eleição, temos que colocar a variável $DUMPALM_i$ no modelo. Só que agora, nossa variável resposta é uma proporção, então nosso teste sobre o parâmetro da Dummy deve ser modificado.

$$total_{(50)} = 432.286 \qquad \frac{975}{432\,286} = 0{,}002255$$

$H_0: \beta_2 \leqslant 0{,}002255$ $\qquad t_{obs} = \dfrac{0{,}005694 - 0{,}002255}{0{,}003187} = 1{,}078 \rightarrow p\text{-valor} = 0{,}1423$

$H_1: \beta_2 > 0{,}002255$

não rejeitamos H_0

Seguindo os mesmos passos que o caso anterior, estimamos o modelo auxiliar do teste BP:

$$\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 \frac{1}{Total_i} + \epsilon_i$$

$H_0: \delta_1 = 0$ (Erros Homocedásticos)

$H_1: \delta_1 \neq 0$ (Erros Heterocedásticos)

$LM = n R^2_{\hat{\varepsilon}_i^2} = 67 \cdot 0{,}1522 = 10{,}2 \qquad p\text{-valor} = 0{,}0014$

Rejeitamos H_0 de que os erros são Homocedásticos.

e) Sabemos que um município possui uma população finita. Mas, por motivos didáticos, vamos assumir que a população de cada município é grande o suficiente para gerarmos boas aproximações da distribuição hipergeométrica pela distribuição binomial.

O modelo a seguir ilustra como o tamanho da população de um município induz Heterocedasticidade.

Seja p a proporção de pessoas que vota em Buchanan e seja n_i a população votante do i-ésimo município.

Sabemos que o número de votos esperados de Buchanan no i-ésimo município é $n_i p$ com variância $n_i p(1-p)$ (distribuição binomial). Se chamarmos $p(1-p) = \sigma^2$ temos que a variância do número de votos no i-ésimo município é $n_i \sigma^2$, ou seja a variância aumenta com o tamanho da população. Então podemos propor o modelo

$$E(\varepsilon_i^2) = Var(\varepsilon_i) = n_i \sigma^2$$

Isso explica porque o modelo auxiliar do teste BP do item (b) tem aquela forma ao dividirmos o total de votos pelo tamanho da população do município temos

$$Var\left(\frac{\varepsilon_i}{n_i}\right) = \frac{1}{n_i^2}\, Var(\varepsilon_i) = \frac{n_i\,\sigma^2}{n_i^2} = \frac{\sigma^2}{n_i}$$

ou seja, agora a variância do erro é inversamente relacionada com a população do município. Isso explica a forma funcional do teste BP do item (d).

Sabendo que $E(\varepsilon_i^2) = n_i\,\sigma^2$, se quisermos fazer alguma transformação nos dados que torne os erros homocedásticos, então queremos encontrar a tal que

$$Var(a\,\varepsilon_i) = \sigma^2$$

$$a^2\, Var(\varepsilon_i) = a^2\, n_i \cancel{\sigma^2} = \cancel{\sigma^2}$$

$$a^2 n_i = 1 \iff a = \frac{1}{\sqrt{n_i}}$$

ou seja, de acordo com o <u>nosso modelo</u> para variância, $Var(\varepsilon_i) = n_i\,\sigma^2$, temos que, se dividirmos as variáveis pela raiz quadrada da população do município, teremos erros homocedásticos no modelo. Sendo $\tilde{x}_i = \frac{x_i}{\sqrt{n_i}}$

$$\widetilde{Buchanan}_i = \beta_0 + \beta_1\, \widetilde{Gore}_i + \varepsilon_i$$ tem erros homocedásticos.

f) Os resultados dos itens (b), (d) e (e) dependem se o nosso modelo de Heterocedasticidade está correto ou não.

Podemos testar Heterocedasticidade no modelo do item (e) com o teste BP da forma:

$$\hat{\tilde{\varepsilon}}_i = \delta_0 + \delta_1 n_i + \delta_2 \frac{1}{n_i} + \epsilon_i$$

no caso teremos que $LM = 12{,}76 \rightarrow p\text{-valor} = 0{,}0017$.

ou seja, rejeitamos H_0 de que os erros são homocedásticos. Como isso ainda ocorre com a transformação $\frac{1}{\sqrt{n_i}}$, temos que o nosso modelo de Heterocedasticidade pode estar incorreto.

Outra fraqueza do modelo é que ele assume que a número de votos de Buchanan e Gore em cada município é proporcional, e ~~pode~~ pode muito bem ser o caso que o município de Palm Beach tenha características que de fato dão a ele uma proporção maior de votos.

Ou seja, o próximo passo seria encontrar variáveis socio-políticas capazes de justificar essa diferença.

6. a) Aumento de 1% no preço acarreta um aumento esperado de 1,6% na quantidade demandada, ceteris paribus

b) $H_0: \beta_3 = 0$ (modelo 1 é melhor)
$H_A: \beta_3 \neq 0$ (modelo 2 é melhor)

Sob H_0

$$F = \frac{(R^2_{ie} - R^2_R)/g}{(1 - R^2_{ie})m-k-1} \sim F[g; m-k-1]$$

Modelo irrestrito: modelo 2
Modelo restrito: modelo 1

$$F_{obs} = \frac{0,95 - 0,63/1}{(1 - 0,95)/60-3} = 364.8$$

$F_{crítico} = 2,79 \quad RC = [2,79; +\infty)$

$F_{obs} \in RC \rightarrow$ há evidências p/ rejeitar H_0.
Modelo 2 é melhor!

c) $H_0: \delta_2 = \delta_3 = \delta_4 = 0$ (erros são homocedásticos)
H_A: pelo menos 1 diferente de zero (erros são heterocedásticos)

Após estimar o modelo 2, estimar:

$$\hat{\varepsilon}_i^2 = \delta_1 + \delta_2 \cdot \ln(P_i) + \delta_3 \cdot d_i + \delta_4 \cdot [\ln(P_i)]^2 + v_i$$

$LM = n \cdot R^2_{\hat{\varepsilon}^2} \sim \chi^2_{(3)}$

$LM = 16,89 \quad \chi^2_{crit} = 7,81 \quad LM_{obs} > \chi^2_{crit}$
rejeitar H_0!

erros são heterocedásticos c/ 95% de confiança

d) $\Omega_{60\times 60} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{60}^2 \end{pmatrix}$

Uma vez constatada a heterocedasticidade, é preciso utilizar estimador robusto à heterocedasticidade, caso contrário, a variância dos estimadores será viesada, afetando o erro-padrão dos βs, e prejudicando o teste t e F.

Com heterocedasticidade, MQO deixa de ser BLUE e assintoticamente eficiente.

Estimador robusto de White leva heterocedasticidade em consideração e calcula uma variância adequada para os estimadores.

e) A suposição de homocedasticidade dos erros é essencial para confiarmos nos resultados do teste F.

Em caso de heterocedasticidade, é possível que as conclusões obtidas via teste F não sejam corretos.

O melhor seria refazer ~~o~~ o teste, estimando os modelos utilizando estimadores robustos.

f) Uma possível explicação é a possibilidade de a variável relacionada ao preço ser endógena, ou seja, correlacionada com o erro.

Para corrigir esse problema, poderia, por exemplo utilizar variáveis instrumentais que sejam exógenas, suficientemente correlacionados com X. Um exemplo de instrumento seria preço dos insumos.

7. Com as informações das saídas do E-views, é possível realizar um teste BP p/ heterocedasticidade.

Ho: $\delta_2 = \delta_3 = \delta_4 = 0$ (erros são homocedásticos)

HA: pelo menos 1 diferente de zero (erros são heterocedásticos)

Sob Ho

$LM = n \cdot R^2_{\hat{\varepsilon}} \sim \chi^2_{(3)}$

$LM = 172 \cdot 0{,}3497 = 60{,}15$

$\chi^2_{(3)\,crítico} = 6{,}25$

$RC = [6{,}25;\ +\infty)$

$LM_{obs} \in RC \rightarrow$ rejeitar Ho

Há evidências, com 90% de confiança que os erros não sejam homocedásticos.

Apesar de não causar viés nas estimativas dos parâmetros, a violação da suposição de homocedasticidade prejudica a inferência uma vez que distorce os estimativas dos erros-padrão dos estimadores