Resolução Lista de Exercícios 4 de Econometria

1)
1a) Coeficientes
α estimado= 3,6 (2,0902)
β estimado = 0,75 (0,2557)

Interpretação: O α é o intercepto da equação, e representa o fato que se nenhuma hora for trabalhada (X=0), haverá uma quantidade produzida de 3,6 unidades. Provavelmente o chefe da empresa terá que fazer o trabalho.
O β indica o quanto a produção muda(Y),a medida que se amplia as horas trabalhadas (X) na firma. Neste caso para cada hora adicional de trabalho representa um incremento de 0,75 unidades na produção.

1b) Considerando um nível de significância de 5% e gl de 8, tem-se que o t tabelado é = 2,306

O t calc α estimado = 3,6/2,0902 = 1,72 < t tab (2,306) → não significante
O t calc β = 0,75/0,2557 = 2,93 > t tab (2,306) → significante

1c) R^2 = 0,5181
O coeficiente de determinação informa o quanto as variáveis independentes explicam as var. dependentes. Neste caso temos que apenas 51,81% da variação de y é explicada pelas horas trabalhadas. Se olharmos apenas para o R^2 percebe-se que há um baixo poder de explicação, que pode ser justificado pelo baixo tamanho de nossa amostra.

1d) Teste F = 8,60
O valor de F tabelado (considerando g.l.1,8 gl. numerador e denominador, respectivamente) = 5,32.
Como F calc (8,60) > F tab (5,32) = podemos rejeitar a Hipótese de que os coeficientes sejam todos iguais, ou considerar que os coeficientes são estatisticamente significativos.

1e) Em virtude de que o termo α não ser significante, bem como o fato de que para alguns modelos (amparados pela teoria) impõe que o intercepto esteja ausente do modelo e neste caso o modelo sem intercepto é mais adequado (melhora a robustez do modelo, por exemplo; e possibilita estimar o coeficiente de inclinação com maior precisão), podemos fazer a regressão sem o intercepto (regressão pela origem).
Veja que, em nosso modelo no qual visa avaliar o impacto das horas de trabalho sobre a produção, o coeficiente α não fez muito sentido em nossa análise.

1f) Coeficientes
β = 1,18 (1,493)
Com a regressão pela origem o coeficiente ampliou para 1,18, o que representa que para cada hora adicional de trabalho a produção amplia em 1,18 unidades.

1g) Considerando um nível de significância de 5% e gl de 9, tem-se que o t tabelado é = 2,262

O t calc β estimado = 1,18 /1,493 = 0,79 < t tab (2,262) → não significante

2)

Soma de Quadrados	gl	Quadrados médios	Teste F
SQE =2575580,98	2	1287790,49	14,84964
SQR = 173444,02	9	86722,01	
SQT = 2749025	11	1374512,5	

Teste F → F calc 4,26 (5% significância e gl = 2 e 9)
Como F calc (14,85) > F tab (4,26) → os coeficientes são significantes

R2 = SQE/SQT = 0,9369
Com coeficiente de determinação de 0,9369, temos que 93,7% da variação da variável dependente é explicada pela (s) variável (eis) independente (s). Analisando esse indicador de forma isolada demonstra um alto poder de explicação do modelo.

3)

	Coeficiente	**Variância**	**Erro padrão**	**t calc**
β1	3,3324	0,0276	0,1663	20,0437
β2	-0,2180	0,0011	0,0325	-6,70391
				T tab = 3,182
ANOVA	Soma dos quadr	gl	Quadr médios	
SQT	0,492	4	0,123	**F** = 44,94234
SQE	0,46	1	0,461213	**R^2** = 0,9374
SQR	0,03	3	0,010262	

Análise: Os coeficientes são estatisticamente significativos (t calc > t tba) e (F calc > F tab – 10,1), bem como elevado R^2.
Considerando os resultados obtidos pode-se apontar que há uma relação negativa entre X e Y, isto é, a medida que X varia em 1 unidade, Y reduz em 0,218 unidades.
Apesar dos resultados, deve-se fazer a ressalva que o tamanho da amostra é muito pequeno, o que pode gerar vieses.

4) Seguiu-se aqui metodologia proposta por Sartoris (cap 7)
Dados:
Dados do problema: n = 20, distr normal, $\mu = 5$; $\sigma^2 = 9$ ($\sigma = 3$)

4a)
i) IC 95% (95/2 = 0,475 → tab distr normal, Z = 1,96)
ii) calcula σ (na qual a fórmula é = (σ^2/n)^0,5 → (9/20)^0,5 = 0,6708
iii) substitui valores na fórmula da distr. normal padronizada, Z = [(X_ - μ)/ σ]
tem-se: (5 – μ)/0,6708= 1,96 → 5 – μ = 1,3148
Assim tem-se o IC (95%) = 3,6852 < μ < 6,3148

(Em 95% das vezes em que realizaríamos este experimento, o valor verdadeiro da média pop. cairia neste intervalo.)

4b)
Como o IC calculado na questão anterior (4a) representa a região de aceitação para H0 (considerando $\mu = 5$), tem-se que 7 está fora da região de aceitação e portanto rejeitamos H0.

4c)
i) IC 95% (50 +45, como 0,45 = 1,645)
ii) calcula σ (na qual a fórmula é = (σ^2/n)^0,5 → (9/20)^0,5 = 0,6708
iii) substitui valores na fórmula da distr. normal padronizada, $Z = [(X\_ - \mu)/ \sigma]$
tem-se: $(7 - \mu)/0,6708= 1,645$ → $7 - \mu = 1,1035$
Assim tem-se o IC (95%) = $-\text{inf} < \mu < 8,1035$ e $\mu=7$ está dentro da região de aceitação, neste caso aceitamos H0.

5) Seguiu-se aqui metodologia proposta por Sartoris (cap 7)
Dados: $n = 50$; $\mu = 75$; $\sigma = 10$

5a)
i) IC 95% (95/2 = 0,475 → tab distr normal, Z = 1,96)
ii) calcula σ (na qual a fórmula é = (σ^2/n)^0,5 → (100/50)^0,5 = 1,4142
iii) substitui valores na fórmula da distr. normal padronizada, $Z = [(X\_ - \mu)/ \sigma]$
tem-se: $(75 - \mu)/1,4142= 1,96$ → $75 - \mu = 2,772$
Assim tem-se o IC (95%) = $72,23 < \mu < 77,77$
(Em 95% das vezes em que realizaríamos este experimento, o valor verdadeiro da média pop. cairia neste intervalo.)

5b)
i) IC 90% (90/2 = 0,45 → tab distr normal, Z = 1,645)
tem-se: $(75 - \mu)/1,4142= 1,645$ → $75 - \mu = 2,3264$
Assim tem-se o IC (90%) = $72,67 < \mu < 77,32$
(Em 90% das vezes em que realizaríamos este experimento, o valor verdadeiro da média pop. cairia neste intervalo.)

O exercício demonstra que quanto maior é o grau de confiança (IC), amplia-se o tamanho do intervalo (ou a margem de erro). Esta situação acaba sendo um *trade-off* para o pesquisador, pois quanto mais confiança ele busca, maior será o seu intervalo/margem de erro.
Para resolver as duas questões conjuntamente é necessário ampliar o tamanho da amostra.

6) N=122
Exame = 0, 29 - 649, 6*RAP* R2 = 0, 4366
(0,48) (15,5) = d-p
Onde Exame= pontuação obtida nos exames ao longo do ano;
RAP= Razão aluno-professor.

6a) Os resultados apontam uma relação negativa entre RAP e resultados dos exames. Neste caso pode-se inferir que a medida que amplia o número de alunos por professor a qualidade das aulas caem o que é refletido nos exames. Neste caso a cada aluno adicional por professor a pontuação cai em 649,6 pontos. O coeficiente β1 indica que caso RAP for zero o resultado é de apenas 0,29 (um resultado não muito lógico).

6b) t tab (nível de significância de 5% e gl = 120) = 1,98

O t calc β1 = 0,29/0,48 = 0,6042 < t tab (1,98) → não significante
O t calc β2 = 649,6/15,5 = 41,91 > t tab (1,98) → significante

6c)
Sim. Uma das possibilidades poderia-se rodar o modelo pela origem (sem intercepto). Uma segunda maneira seria regredir o modelo considerando uma outra forma funcional, como por exemplo aplicar log nas duas variáveis (modelo log-linear). Ou ainda, poderia-se verificar a possibilidade de incluir outras variáveis explicativas, como por exemplo educação dos pais e/ou anos de escolaridade do professor (titulação).

6d)
Sim. Alguns modelos impõem que o intercepto esteja ausente do modelo e neste caso o modelo sem intercepto é mais adequado (melhora a robustez do modelo, por exemplo; e possibilita estimar o coeficiente de inclinação com maior precisão), podemos fazer a regressão sem o intercepto (regressão pela origem).

6e) Não é possível rodar o modelo sem os valores da amostra.

6f) R^2 = 0,4366
F = [R2 /(k)]/[(1-R2)/(n-k-1)]
F = [0,4366/1]/[(1-0,4366)/120] = 92,99
Como F tab = 3,92 < F calc (92,99) = coefic são significativos.

7) Seguiu-se aqui metodologia proposta por Gujarati (2000) (cap 5)
IC → β +- t tab*d-p

n = 100 dados sobre as casas vendidas:
$\hat{}$*preço* = 6, 6 + 0, 69*Tamanho*
(0,39) (0,054) =d-p
R^2 = 0, 71

7a)
Seja: t tab = 1,98 (nível de significância de 5% e gl = 98)
β_1
6,6 – 1,98*0,39 < β_1 < 6,6 + 1,98*0,39)
IC $_{\beta1(95\%)}$ = 5,83 < β_1 < 7,37
β_2
0,69 – 1,98*0,054 < β_2 < 0,69 + 1,98*0,054)
IC $_{\beta2(95\%)}$ = 0,5831 < β_2 < 0,7969

7b)

Como o IC calculado na questão anterior (7a) representa a região de aceitação para H0 de β1 ou alfa, tem-se que 7 está dentro da região de aceitação e portanto não podemos rejeitar H0.

7c)
Como o IC calculado na questão anterior (7a) representa a região de aceitação para H0 de β2 ou β, tem-se que 0,5 está fora da região de aceitação e portanto rejeitamos H0.

7d e e) Deve ser usada a fórmula:

$$\Pr\left[\hat{\beta}_{\mathrm{i}} - \mathrm{tc}\,\sigma_{\hat{\beta}_i} \leq \beta_{\mathrm{i}} \leq \hat{\beta}_{\mathrm{i}} + \mathrm{tc}\,\sigma_{\hat{\beta}_i}\right] = 1 - \alpha$$

Mas monocaudal temos:

$$\Pr\left[\beta \leq \hat{\beta}_{\mathrm{i}} + \mathrm{tc}\,\sigma_{\hat{\beta}_i}\right] = 1 - 0{,}05$$

ou

$$\Pr\left[\hat{\beta}_{\mathrm{i}} - \mathrm{tc}\,\sigma_{\hat{\beta}_i} \leq \beta_{\mathrm{i}}\right] = 1 - 0{,}05$$

E basta construir o intervalo usando o valor t tabelado monocaudal.

8) Taxa de crescimento anual médio
modelo semilog (log-lin) - na forma ln Y = β1 + β2t+ε

Resultados
β_1 estimado = 6,77 (0,0685)
β_2 estimado= 0,2073 (0,008)

taxa de anual de crescimento é de 100* β_2 (=20,73%). As vendas tiveram um crescimento de 20,73% a.a. no período de 1986 a 1999.