CONCORDIA UNIVERSITY
Department of Economics

ECON 222/4 SECTIONS A, B and BB
STATISTICAL METHODS II
WINTER 2016 – MIDTERM 2 (SOLUTIONS)
Sunday, March 20, 1:00pm – 3:00pm

1. (**34 marks**) A population is assumed to follow the behaviour $y_i = \beta_0 + \beta_1 x_i + e_i$, where β_0 and β_1 are unknown population parameters and $e \sim iid\, N\left(0, \sigma^2\right)$. Use the following sample of five observations to answer the following questions.

x	1	2	3	4	5
y	3	7	5	11	14

a. (**2 marks**) Calculate b_0.

$$b_0 = \bar{y} - b_1 \bar{x} = 8 - 2.6 \cdot 3 = 0.2$$

b. (**2 marks**) Calculate b_1.

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{1\cdot 3 + 2\cdot 7 + 3\cdot 5 + 4\cdot 11 + 5\cdot 14 - 5\cdot 3\cdot 8}{1^2 + 2^2 + 3^2 + 4^2 + 5^2 - 5\cdot 3^2} = \frac{146 - 120}{55 - 45} = 2.6$$

c. (**2 marks**) ***<u>Briefly</u>*** interpret b_0.

When x = 0, y should equal 0.2.

d. (**2 marks**) ***<u>Briefly</u>*** interpret b_1.

When x changes by one unit, y should change by 2.6 units.

e. (**2 marks**) Calculate SST.

$$SST = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = 3^2 + 7^2 + 5^2 + 11^2 + 14^2 - 5\cdot 8^2 = 400 - 320 = 80$$

f. (**2 marks**) Calculate SSR.

$$SSR = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2 = \sum_{i=1}^{n} \hat{y}_i^2 - n\bar{y}^2 = 2.8^2 + 5.4^2 + 8^2 + 10.6^2 + 13.2^2 - 5\cdot 8^2 = 387.6 - 320 = 67.6$$

g. (**2 marks**) Calculate SSE.

$$SSE=\sum_{i=1}^{n}\hat{u}_i^2=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2=\left(3-2.8\right)^2+\left(7-5.4\right)^2+\left(5-8\right)^2+\left(11-10.6\right)^2+\left(14-13.2\right)^2=12.4$$

h. (**2 marks**) Calculate the coefficient of determination, R^2.

$$R^2=\frac{SSR}{SST}=\frac{67.6}{80}=0.845$$

i. (**2 marks**) ***Briefly*** interpret R^2.

84.5 percent of the variation in y can be explained by the estimated regression.

j. (**2 marks**) Calculate the correlation coefficient, r.

$$r=\left(\text{sign of } b_1\right)\sqrt{R^2}=\sqrt{0.845}=0.9192$$

k. (**2 marks**) ***Briefly*** interpret r.

There is an almost perfect positive linear relationship between x and y.

l. (**2 marks**) Calculate the standard error of the estimate, $\hat{\sigma}$.

$$\hat{\sigma}=\sqrt{\frac{SSE}{n-k-1}}=\sqrt{\frac{12.4}{5-1-1}}=2.0331$$

m. (**2 marks**) Calculate $\text{var}\left(b_1\right)$.

$$\text{var}\left(b_1\right)=\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\frac{4.1\dot{3}}{10}=0.41\dot{3}$$

n. (**2 marks**) Calculate $\text{var}\left(b_0\right)$.

$$\text{var}\left(b_0\right)=\frac{\hat{\sigma}^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\frac{4.1\dot{3}\cdot 55}{5\cdot 10}=4.54\dot{6}$$

o. (**2 marks**) Calculate $\text{cov}(b_0,b_1)$.

$$\text{cov}(b_0,b_1)=\frac{-\bar{x}\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{-3\cdot 4.1\dot{3}}{100}=-1.24$$

p. (**2 marks**) Construct a 99-percent confidence interval for β_0, using $t_{0.005,3}=\pm 5.841$.

$$b_0\pm t_C\cdot se(b_0)=0.2\pm 5.841\cdot\sqrt{4.54\dot{6}}\Rightarrow[-12.2547,12.6547]$$

q. (**2 marks**) Test, at the 1-pecent level of significance, whether $\beta_0=10$. Clearly state the null and alternative hypotheses, the test statistic and your conclusion. Use a critical value of $t_{0.005,3}=\pm 5.841$.

$H_0:\beta_0=10$ and $H_1:\beta_0\neq 10$

Since the t-statistic, $t_S=\dfrac{b_0-\beta_0}{se(b_0)}=\dfrac{0.2-10}{\sqrt{4.54\dot{6}}}=-4.596$, *does not lie beyond the lower critical value,* $t_{0.005,3}=-5.841$, *we cannot reject the null hypothesis. Therefore, we conclude that there is insufficient evidence to reject that* $\beta_0=10$.

2. (**4 marks**) In the regression, $\hat{y}_i=\beta_0+\beta_1 x_i$, use the fact that $b_1=\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ to prove

a. (**2 marks**) b_1 is an unbiased estimator of β_1.

$$b_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\beta_0+\beta_1 x_i+e_i)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$=\frac{\beta_0\sum_{i=1}^{n}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}+\frac{\beta_1\sum_{i=1}^{n}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}+\frac{\sum_{i=1}^{n}(x_i-\bar{x})e_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\beta_1+\frac{\sum_{i=1}^{n}(x_i-\bar{x})e_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\Rightarrow E(b_0)=\beta_1+\frac{\sum_{i=1}^{n}(x_i-\bar{x})E(e_i)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\beta_1$$

b. (**2 marks**) $\text{var}\left(b_1\right)=\dfrac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$.

$$\text{var}\left(b_1\right)=\text{var}\left(\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)e_i}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\right)=\frac{\text{var}\left(\sum_{i=1}^{n}\left(x_i-\bar{x}\right)e_i\right)}{\left(\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)^2}=\frac{\left(\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\text{var}\left(e_i\right)\right)}{\left(\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)^2}=\frac{\sigma^2\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{\left(\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)^2}=\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$$

3. (**6 marks**) Price elasticity of demand is given by $\varepsilon=\dfrac{\Delta q}{\Delta p}\dfrac{p}{q}$. In the regression $\hat{q}_i=b_0-b_1p_i$, elasticity can be calculated for any price as $\varepsilon=\dfrac{-b_1p_i}{b_0-b_1p_i}$. Calculate price elasticity of demand from each of the following estimated regressions.

a. (**2 marks**) $\hat{q}_i=b_0-b_1\ln p_i$

$$\frac{\Delta q}{\Delta \ln p}=\frac{\Delta q}{\left(\Delta p/p\right)}=-b_1\Rightarrow\frac{\Delta q}{\Delta p}\frac{p}{q}=\frac{-b_1}{b_0-b_1\ln p_i}$$

b. (**2 marks**) $\widehat{\ln q_1}=b_0-b_1p_i$

$$\frac{\Delta \ln q}{\Delta p}=\frac{\left(\Delta q/q\right)}{\Delta p}=-b_1\Rightarrow\frac{\Delta q}{\Delta p}\frac{p}{q}=\frac{-b_1}{p_i}$$

c. (**2 marks**) $\widehat{\ln q_1}=b_0-b_1\ln p_i$

$$\frac{\Delta \ln q}{\Delta \ln p}=\frac{\left(\Delta q/q\right)}{\left(\Delta p/p\right)}=-b_1$$

4. (**6 marks**) Consider two regression models, $y_i=b_0+b_1x_i+e_i$ and $z_i=a_0+a_1w_i+v_i$, where $z_i=cy_i$ and $w_i=hx_i$.

a. (**2 marks**) Derive expressions that relate b_0 to a_0, b_1 to a_1 and e_i to v_i.

$$z_i = a_0 + a_1 w_i + v_i \Rightarrow c\hat{y}_i = a_0 + a_1 h x_i + v_i \Rightarrow y_i = \frac{a_0}{c} + \frac{ha_1}{c} x_i + \frac{v_i}{c} \Rightarrow b_0 = \frac{a_0}{c}, b_1 = \frac{ha_1}{c}, e_i = \frac{v_i}{c}$$

b. (**2 marks**) Derive expressions that relate $\operatorname{var}(b_0)$ to $\operatorname{var}(a_0)$, $\operatorname{var}(b_1)$ to $\operatorname{var}(a_1)$ and $\operatorname{var}(e_i)$ to $\operatorname{var}(v_i)$.

$$b_0 = \frac{a_0}{c} \Rightarrow \operatorname{var}(b_0) = \frac{\operatorname{var}(a_0)}{c^2}, b_1 = \frac{ha_1}{c} \Rightarrow \operatorname{var}(b_1) = \frac{h^2 \operatorname{var}(a_1)}{c^2}, e_i = \frac{v_i}{c} \Rightarrow \operatorname{var}(e_i) = \frac{\operatorname{var}(v_i)}{c^2}$$

c. (**2 marks**) ***<u>Briefly</u>*** compare the *t*-statistics for a_0 to a_1, $t_{a_0} = \frac{a_0}{se(a_0)}$ and $t_{b_1} = \frac{a_1}{se(a_1)}$, to the *t*-statistics for b_0 to b_1, $t_{b_0} = \frac{b_0}{se(b_0)}$ and $t_{b_1} = \frac{b_1}{se(b_1)}$.

$$t_{a_0} = \frac{a_0}{se(a_0)} = \frac{a_0}{\sqrt{\operatorname{var}(a_0)}} = \frac{cb_0}{\sqrt{c^2 \operatorname{var}(b_0)}} = \frac{b_0}{\sqrt{\operatorname{var}(b_0)}} = \frac{b_0}{se(b_0)} = t_{b_0}$$

$$t_{a_1} = \frac{a_1}{se(a_1)} = \frac{a_0}{\sqrt{\operatorname{var}(a_1)}} = \frac{cb_1/h}{\sqrt{c^2 \operatorname{var}(b_0)/h^2}} = \frac{b_1}{\sqrt{\operatorname{var}(b_1)}} = \frac{b_1}{se(b_1)} = t_{b_1}$$