Solution.

CONCORDIA UNIVERSITY
Department of Economics

ECON 222 SECTIONS C
STATISTICAL METHODS II
Winter 2019 – ASSIGNMENT 2
Due: Wednesday, March 20, before 2:30pm

1. **(28 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 $e \sim iid\,N(0, \sigma^2)$. Use the following sample of five observations to answer the following questions.

x	30	12	75	45	15	50	25
y	120	60	250	180	55	155	140

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

$b_0 = 31.9409$

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

$b_1 = 2.922$

c. **(2 marks)** ***Briefly*** interpret b_1.

one unit increase in x_i would lead to 2.922 unit increase in y_i

$$\frac{d\hat{y}_i}{d x_i} = 2.922.$$

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

$SST = 27892.857$

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

$SSR = 25375.186$

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

$SSE = 2512.903$

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

$$R^2 = \frac{SSR}{SST} = 0.90974$$

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

The model explains about 91% of the total variation in (y)

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

$$r_{xy}^2 = R^2 \implies r_{xy} = +\sqrt{0.90974}$$

$$\implies r_{xy} = 0.9538$$

j. **(2 marks)** Calculate $\hat{\sigma}$.

$$\hat{\sigma}^2 = \frac{SSE}{N-2} = \frac{2512.9034}{5} = 502.58068$$

$$\implies \hat{\sigma} = 22.418$$

k. **(2 marks)** Calculate $\widehat{\text{var}}(b_1)$.

$$\widehat{Var}(b_1) = \frac{\hat{\sigma}^2}{\Sigma(x_i - \bar{x})^2} = 0.1691$$

l. **(2 marks)** Calculate $\widehat{\text{var}}(b_0)$.

$$\widehat{Var}(b_0) = \frac{\hat{\sigma}^2 \Sigma (x_i^2}{N \Sigma (x_i - \bar{x})^2} = 290.7025$$

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

$$\widehat{\text{Cov}(b_0, b_1)} = \frac{-\hat{\sigma}^2 \, \bar{X}}{\Sigma (X_i - \bar{X})^2} = -6.08778$$

n. **(2 marks)** Use the estimated regression line to predict y when $x = 20$.

$$\hat{y} = 31.9409 + (2.922)(20)$$

$$\Rightarrow \hat{y} = 90.3809$$

2. **(8 marks)** Using the regression output for the food expenditure model shown in Figure 2.9, page 65

a. **(2 marks)** Construct a 90 percent interval estimate for β_2 and interpret.

$\alpha = 0.1 \qquad t^c_{(38)} = 1.686$

$$\beta_2 \in b_2 \pm (1.686)(2.093)$$

$$\Rightarrow \beta_2 \in 10.209 \pm (3.5237)$$

$$\Rightarrow \beta_2 \in [6.6853, 13.7327]$$

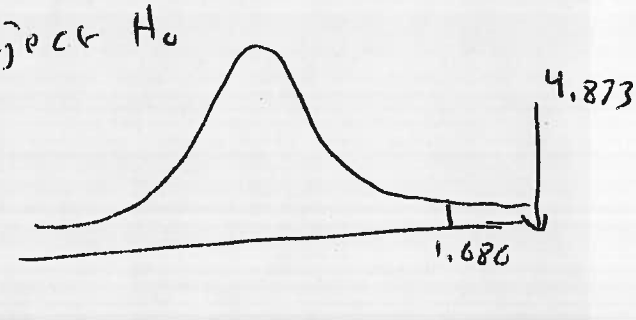
b. **(2 marks)** Test the null hypothesis that β_2 is zero against the alternative that it is not at the 10 percent level of significance without using the reported p-value. What is your conclusion?

$H_0: \beta_2 = 0 \qquad H_1: \beta_2 \neq 0 \Rightarrow$ Two tail test

$$t^c_{(38)} = 1.686$$

$$t_{est} = \frac{10.209 - 0}{2.093} = 4.873$$

$\because t_{est} > t^c_{(38)} \Rightarrow$ we reject H_0



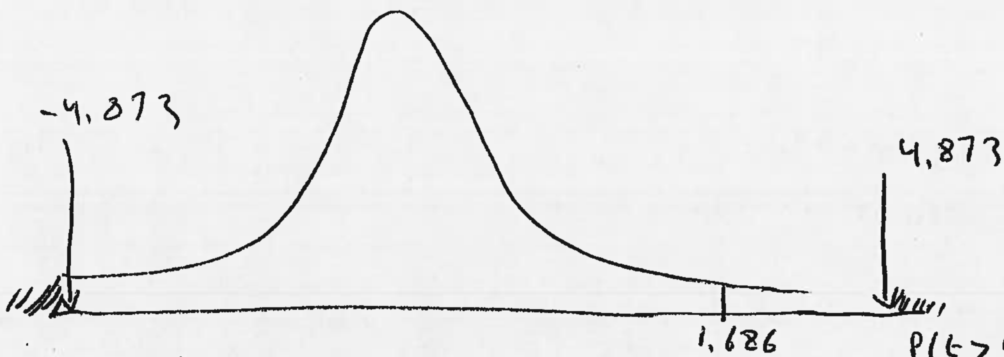
Q1

x	y	x-xbar	(x-xbar)^2	y-ybar	(y-ybar)^2	(x-xbar)(y-ybar)	yhat	(yhat-ybar)^2	ehat
30	120	-6	36	-17.1429	293.8790204	102.8574	119.601	307.7182556	0.399
12	60	-24	576	-77.1429	5951.02702	1851.4296	67.005	4919.325016	-7.005
75	250	39	1521	112.8571	12736.72502	4401.4269	251.091	12984.16949	-1.091
45	180	9	81	42.8571	1836.73102	385.7139	163.431	691.0642016	16.569
15	55	-21	441	-82.1429	6747.45602	1725.0009	75.771	3766.51011	-20.771
50	155	14	196	17.8571	318.8760204	249.9994	178.041	1672.654584	-23.041
25	140	-11	121	2.8571	8.16302041	-31.4281	104.991	1033.744674	35.009
36	137.1429		2972		27892.85714	8685		25375.18633	

b0	31.94097289
b1	2.922274563
SST	27892.85714
SSE	2512.903471
SSR	25375.18633
R-Squared	0.909737794

ehat^2
0.159201
49.07002
1.190281
274.5318
431.4344
530.8877
1225.63
2512.903

c. **(2 marks)** Draw a sketch showing the p-value 0.0000 shown in Figure 2.9, the critical value from the t-distribution used in part (b), and how the p-value could have been used to answer part (b).



$P(t > 4.873) + P(t \leq -4.873)$
$= P\text{-value} = 0$

$\therefore$ P-value $< \alpha \Rightarrow$ we reject H_0

d. **(2 marks)** Test the null hypothesis that β_2 is zero against the alternative that it is positive at the 10 percent level of significance. Draw a sketch of the rejection region and compute the p-value. What is your conclusion?

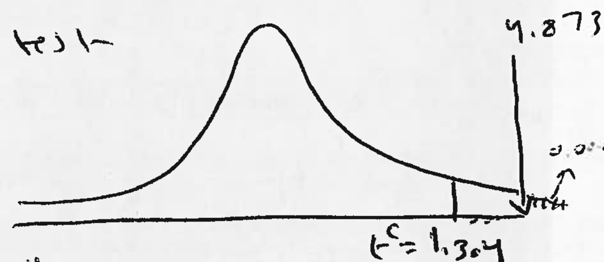
$H_0: \beta_2 = 0$

$H_1: \beta_2 > 0 \Rightarrow$ Right tail test

$t^c_{(38)} = 1.304$

$t_{test} = 4.873$

$\therefore$ Test $> t^c_{(38)} \Rightarrow$ we reject H_0



P-value $= P(t > 4.873) = 0.000$

3. **(6 marks)** Briefly interpret the estimated slope coefficient in each one of the following models. Next, compute. the unit change in (y) as a result of a one unit increase in (x) according to each model given that $y_0 = 154$ *and* $x_0 = 10$.

a. **(2 marks)** $\hat{y}_i = 183 + 120.5 \ln x_i$

- 1% increase in X_i would lead to $\frac{120.5}{100} = 1.205$ unit increase in y_i.

- $\frac{d\hat{y}_i}{d \ln x_i} = 120.5 \Rightarrow \frac{dy_i}{dx_i} = \frac{120.5}{x_i} = \frac{120.5}{10} = 12.05$

b. **(2 marks)** $\widehat{lny_i} = 150 + 0.019x_i$

- one unit increase in x_i would lead to $(0.019)100\%$

$= 1.9\%$ increase in y_i.

- $\frac{d\ln\hat{y}_i}{dx_i} = 0.019 \Rightarrow \frac{dy_i}{dx_i} = 0.019y_i \Rightarrow \frac{dy_i}{dx_i} = 2.926$

c. **(2 marks)** $\widehat{lny_i} = 70 + 3.5\ln x_i$

$\frac{d\,\widehat{\ln y_i}}{d\ln x_i} = 3.5$ "one percent increase in x_i would lead to 3.5% increase in y_i.

$\frac{dy_i}{dx_i} = 3.5\left(\frac{y_i}{x_i}\right) \Rightarrow \frac{dy_i}{dx_i} = 3.5\left(\frac{154}{10}\right) = 53.9$

4. **(12 marks)** Consider the simple linear regression model $y_i = \beta_1 + \beta_2 x_i + e_i$ where all the assumptions SR1-SR5 holds.

a. **(4marks)** Show that the least squares estimator of the slope coefficient is linear and unbiased estimator of β_2.

$$b_2 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$$

$$= \frac{\Sigma (x_i - \bar{x}) y_i}{\Sigma (x_i - \bar{x})^2} - \bar{y} \frac{\Sigma (x_i - \bar{x})}{\Sigma (x_i - \bar{x})^2}$$

$\because \Sigma (x_i - \bar{x}) = 0 \Rightarrow b_2 = \Sigma \left[\frac{x_i - \bar{x}}{\Sigma (x_i - \bar{x})^2}\right] y_i$

Let $w_i = \frac{x_i - \bar{x}}{\Sigma (x_i - \bar{x})^2} \Rightarrow b_2 = \Sigma w_i y_i$

$\Rightarrow b_2$ is linear in y.

$$b_2 = \Sigma w_i y_i$$

$$= \Sigma (w_i (\beta_1 + \beta_2 x_i + e_i))$$

$$= \beta_1 \Sigma w_i + \beta_2 \Sigma w_i x_i + \Sigma w_i e_i$$

$\because \Sigma w_i = 0, \quad \Sigma w_i x_i = 1$

$\Rightarrow b_2 = \beta_2 + \Sigma w_i e_i$

$E(b_2) = \beta_2 + \Sigma w_i E(e_i)$

if $E(e_i) = 0$ "SR2 holds"

$\Rightarrow E(b_2) = \beta_2 \Rightarrow b_2$ is unbiased

b. **(4 marks)** What is the distribution of b_2?

Note that $E(b_2) = \beta_2$ & $Var(b_2) = \frac{\sigma^2}{\Sigma (x_i - \bar{x})^2}$

Since the assumption SR6 is not satisfied, we could not conclude that b_2 is normally distributed.

if SR6 holds, $e_i \sim N(0, \sigma^2) \Rightarrow b_2 \sim N\left(\beta_2, \frac{\sigma^2}{\Sigma (x_i - \bar{x})^2}\right)$

otherwise, we need large sample to invoke the central limit theorem.

c. **(4 marks)** Instead, if we have the model $y_i = \beta_1 + \beta_2 y_i + e_i$ where the dependent variable is the same as the independent variable, what are the values of b_1 and b_2? ~~Briefly interpret the intercept in this model.~~

Recall, $b_2 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$, ~~in~~

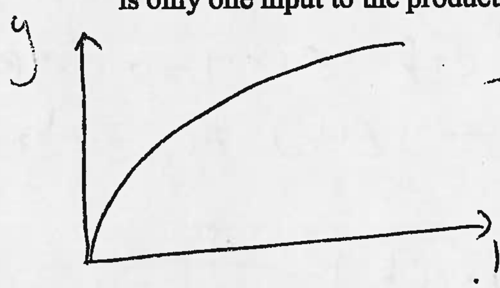
in this particular model, $b_2 = \frac{\Sigma (y_i - \bar{y})(y_i - \bar{y})}{\Sigma (y_i - \bar{y})^2} = 1$

$b_1 = \bar{y} - b_2 \bar{x}$, in this model, $b_1 = \bar{y} - b_2 \bar{x}$

$\Rightarrow \boxed{b_1 = 0}$

5. **(8 marks)** An economist is tasked to estimate the following well-known functions in economic theory. In each case, determine choose an appropriate functional form. Demonstrate your answers graphically and comment on the expected signs of the coefficients.

a. **(4 marks)** The Cobb-Douglas production function with decreasing return to scale. Assume there is only one input to the production function.



Theory.

$\frac{dy}{dk} > 0$

$\frac{d^2y}{dk^2} < 0$

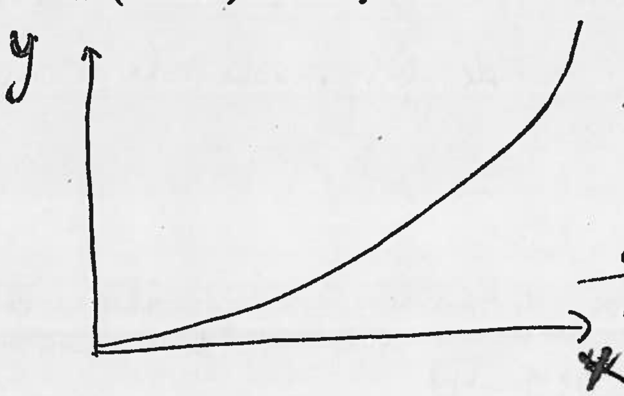
$y = \beta_1 + \beta_2 \ln(X)$

$\frac{dy}{d\ln x} = \beta_2 \Rightarrow \frac{dy}{dx} = \frac{\beta_2}{X}$

Model $\frac{d^2y}{dx^2} = \frac{-\beta_2}{x^2}$

$\Rightarrow \beta_2 > 0$

b. **(4 marks)** The utility function of a risk-lover investor.



$\frac{dy}{dx} > 0$

$\frac{d^2y}{dx^2} > 0$

$\ln y = \beta_1 + \beta_2 X$

$\Rightarrow y = e^{(\beta_1+\beta_2 X)}$

$\cancel{\frac{d\ln y}{dX} = \beta_2}$

$\Rightarrow \frac{dy}{dx} = \beta_2 e^{(\beta_1+\beta_2 X)}$

$\frac{d^2y}{dx^2} = (\beta_2)^2 e^{(\beta_1+\beta_2 X)}$

for $\beta_2 > 0$; $\frac{dy}{dx} > 0$ in the model

and $\frac{d^2y}{dx^2} > 0$.

6. **(8 marks)**Solve the computer exercise question 3.32 in the 5th edition of the textbook.
7. **(8 marks)** Solve the computer exercise question 4.27 (a,b,c,d) in the 5th edition of the textbook.