



--	--	--

**Semester One 2016
Examination Period**

Faculty of Business & Economics

EXAM CODES: ETC1000 / ETW1000

TITLE OF PAPER: Business and Economic Statistics - PAPER 1 of 1

EXAM DURATION: 2 hours writing time

READING TIME: 10 minutes

THIS PAPER IS FOR STUDENTS STUDYING AT:(tick where applicable)

- | | | | | |
|------------------------------------|---|--|--|--|
| ☐ Berwick | ☒ Clayton | ☒ Malaysia | ☐ Off Campus Learning | ☐ Open Learning |
| ☐ Caulfield | ☐ Gippsland | ☐ Peninsula | ☐ Enhancement Studies | ☐ Sth Africa |
| ☐ Parkville | ☐ Other (specify) | | | |

During an exam, you must not have in your possession any item/material that has not been authorised for your exam. This includes books, notes, paper, electronic device/s, mobile phone, smart watch/device, calculator, pencil case, or writing on any part of your body. Any authorised items are listed below. Items/materials on your desk, chair, in your clothing or otherwise on your person will be deemed to be in your possession.

No examination materials are to be removed from the room. This includes retaining, copying, memorising or noting down content of exam material for personal use or to share with any other person by any means following your exam.

Failure to comply with the above instructions, or attempting to cheat or cheating in an exam is a discipline offence under Part 7 of the Monash University (Council) Regulations.

AUTHORISED MATERIALS

OPEN BOOK	☐ YES	☒ NO
CALCULATORS	☐ YES	☒ NO
SPECIFICALLY PERMITTED ITEMS	☐ YES	☒ NO

Candidates must complete this section if required to write answers within this paper

STUDENT ID: _____

DESK NUMBER: _____

INSTRUCTIONS TO CANDIDATES:

Answer ALL questions in this examination paper.

Paper is out of 100 marks

Where you are asked to perform calculations, you should write out the solution as an equation containing the appropriate numerical values from within the question. You do not need to calculate exact values in order to receive full marks for that part of the question.

...

In this paper we will explore some data collected in a survey of households living in a coffee-growing district of Timor-Leste. The last question in this paper asks you to provide a summary of your findings from this analysis.

Question 1 (25 marks)

First let us look at quantity of coffee produced by these households. Below is a table of descriptive statistics for kilograms of coffee produced by households in the last 12 months.

<i>Kilograms of coffee produced</i>	
Mean	763.5037
Standard Error	38.33008
Median	800
Mode	1000
Standard Deviation	447.0017
Sample Variance	199810.5
Kurtosis	1.797512
Skewness	0.837526
Range	2480
Minimum	20
Maximum	2500
Sum	103836.5
Count	136

- (a) Interpret the values for the *Mean*, *Median* and *Mode*. What do these three values tell you about the shape of the distribution for coffee production? (5 marks)
- (b) Interpret the *Standard Deviation*. Would you say this is large? Explain your reasoning. (3 marks)
- (c) There is actually a total of 187 households in this sample, but only 136 of these grow coffee. Those that do not produce coffee have a blank for this variable, and so the Excel output above omits these blank values in the analysis (notice the *Count* is 136). In some data sets, these households may have had a “0” recorded instead of a blank. If this were the case here – that is, we were to include these households with production =0 kilograms into the descriptive statistics – what, if anything, would you expect to see happen to each of the *Mean*, *Median*, *Mode* and *Standard Deviation*? Explain your reasoning. (6 marks)

A common standardisation used in measuring agricultural output is productivity relative to land area used. That is, a household’s productivity in coffee production can be captured by:

$$\text{Yield} = \text{quantity of coffee produced in kilograms} / \text{coffee land area in hectares}$$

- (d) What does this standardisation allow us to compare? Give an example comparing 2 coffee-producing households to illustrate. (3 marks)
- (e) The following regression output estimates mean yield (kilograms per hectare).

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.9359
R Square	0.875909
Adjusted R Square	0.868502
Standard Error	152.4773
Observations	136

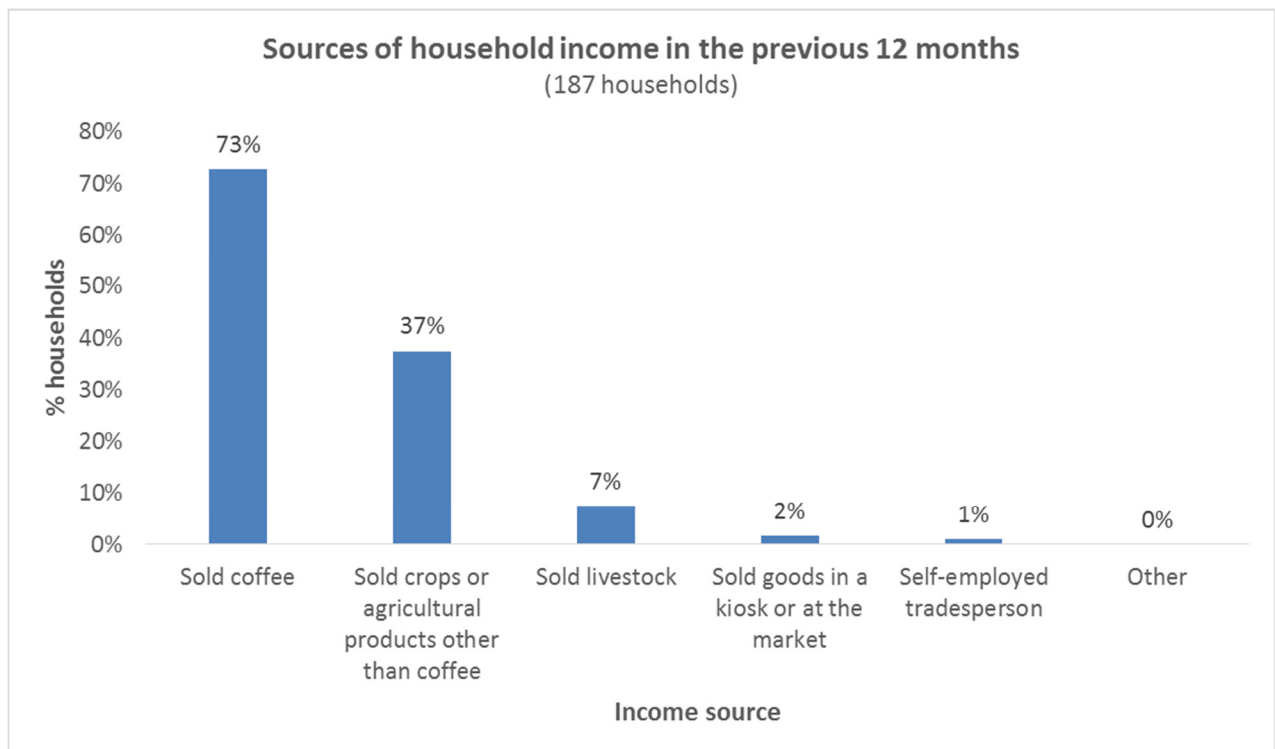
<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	22154640	22154640	952.9151	9.05E-63
Residual	135	3138660	23249.33		
Total	136	25293300			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
Mean	403.6109	13.07482	30.86932	4.92E-63	377.7529	429.4689

- (i) Interpret the values for the Mean under the *Lower 95%* and *Upper 95%* columns.
(3 marks)
- (ii) In neighbouring South-East Asian countries, coffee yield averages around 1000 kilograms per hectare. What does your answer to (i) tell you about the average yield of coffee-producing households in Timor-Leste compared to other countries?
(2 marks)
- (iii) The confidence interval you discussed in (i) is quite a wide interval. Explain intuitively the role of sample size (n) and standard deviation (σ) in determining the width of a confidence interval.
(3 marks)

Question 2 (15 marks)

Surveyed households were asked about their sources of income in the last 12 months. Responses are shown in the bar chart below.



- (a) Provide an interpretation of the 73% and the 37% bars in this chart; that is, explain what these values are measuring. (2 marks)
- (b) What can you say about the income dependency of households in this district on coffee? Explain how you drew your conclusion. (2 marks)
- (c) Explain why a pie chart would be an inappropriate way to display this information. (2 marks)

Many households in Timor-Leste suffer from a shortage of food. It has been argued that growing coffee does not help with this, because instead of growing crops on their land they are growing a crop that is not food. The following table shows the number of households that grow food-crops by coffee-growing status.

	Grows coffee	Does not grow coffee	Total
Grows food crops	70	18	88
Does not grow food crops	66	33	99
Total households	136	51	187

- (d) What is the probability a household does not grow food crops?
(2 marks)
- (e) What is the probability a household does not grow food crops but grows coffee?
(2 marks)
- (f) What is the probability a coffee-growing household grows food crops?
(2 marks)
- (g) What does the table suggest about whether the growing of coffee restricts the growing of food crops? Explain how you drew this conclusion.
(3 marks)

Question 3 (30 marks)

A regression model was estimated to understand why some households have better coffee yields than others. Variables are defined as follows:

Dependent variable: Coffee yield in kilograms per hectare

Explanatory variables:

- Age of trees = age of household's coffee trees, in years
- Maintains trees = 1 if the household regularly prunes and maintains their coffee trees; =0 otherwise
- Zone 1 = 1 if the household is located in zone 1 within the district; =0 otherwise
- Zone 2 = 1 if the household is located in zone 2 within the district; =0 otherwise
- Zone 3 = 1 if the household is located in zone 3 within the district; =0 otherwise

N.B. This coffee-growing district is divided into 3 geographical zones.

Regression output follows.

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.336666
R Square	0.113344
Adjusted R Square	0.08627
Standard Error	145.7519
Observations	136

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	355747.6	88936.89	4.186525	0.003188
Residual	131	2782912	21243.61		
Total	135	3138660			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	336.4716	32.83459	10.24747	1.86E-18	271.5169	401.4263
Age of trees	-0.73468	0.510922	-1.43794	0.152835	-1.7454	0.27605
Maintains trees	77.81872	31.03179	2.50771	0.013373	16.43044	139.207
Zone 2	90.79857	35.06785	2.589225	0.010707	21.426	160.1712
Zone 3	36.37211	29.88624	1.217018	0.225785	-22.75	95.49422

- (a) Interpret the estimated coefficient for the intercept and *Age of trees*. Explain whether these values make sense. (4 marks)
- (b) Consider the coefficient for *Maintains trees*.
- (i) Interpret the estimated coefficient. (2 marks)
- (ii) Perform a hypothesis test to see whether households that maintain their trees experience better yields. Use a critical value approach: the value from the Student's t distribution you need is 1.66. (5 marks)
- (iii) Currently few coffee-growing households in this district prune and maintain their coffee trees (around 20%), and it has been suggested that a program is needed to address this. From a practical point of view, is maintaining trees a key to substantially increased yields? Explain your reasoning. (2 marks)
- (c) The critical value of 1.66 you used in part (b) above is different to the critical value you would have obtained from a Standard Normal distribution.
- (i) Why do we use Student's t critical value? Intuitively, why would you expect the Student's t critical value to be larger than the Normal distribution value? (3 marks)
- (ii) Under what circumstances would values from the Student's t and Normal distributions be virtually the same? (1 mark)
- (iii) For the test in (b), even though we may not know the appropriate critical value for the Normal distribution, we can tell whether the outcome of the test would be any different if the Normal critical value was used. Explain how we can tell in this case, and whether the outcome would change. (2 marks)
- (d) Next consider the Zone coefficients.
- (i) Interpret the estimated coefficients for *Zone 2* and *Zone 3*. (4 marks)
- (ii) What do the p-values for the Zone dummies tell you about differences in coffee yields in different locations? (3 marks)
- (iii) Suggest a reason why we might see differences in coffee yield by location. (2 marks)
- (e) Agricultural scientists assure us that coffee production is strongly related to the age of the tree, however *Age of trees* is not significant in the model. Suggest an possible explanation for why the model does not find this effect. (2 marks)

Question 4 (16 marks)

Coffee is a tree crop that is harvested once per year. This puts a significant labour burden on households in harvest months of the year. It also means that coffee income is only earned for some months of the year.

The survey recorded the coffee harvesting activities of households on a monthly basis over the last 3 years (2012-2015). In any particular month of the last 3 years, therefore, we know the proportion of households in the district that were harvesting coffee.

The regression output below estimates the proportion of households engaged in coffee harvest as a function of the year and month:

Time =1 in January 2012
 =2 in February 2012
 =3 in March 2012
 .
 .
 =36 in December 2015

January =1 if the month is January
February =1 if the month is February
etc.

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.963212
R Square	0.927777
Adjusted R Square	0.890095
Standard Error	0.089332
Observations	36

ANOVA					
	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	12	2.357792	0.196483	24.62138	2.77E-10
Residual	23	0.183544	0.00798		
Total	35	2.541336			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0.033529	0.055229	0.607077	0.549747	-0.08072	0.147779
Time	-0.00256	0.00152	-1.68782	0.104961	-0.00571	0.000579
February	0.003011	0.072955	0.041278	0.967431	-0.14791	0.15393
March	0.008386	0.073002	0.114875	0.909541	-0.14263	0.159403
April	0.020841	0.073081	0.285174	0.778063	-0.13034	0.172021
May	0.115572	0.073192	1.579032	0.127985	-0.03584	0.266981
June	0.332637	0.073334	4.535936	0.000148	0.180935	0.48434
July	0.495202	0.073507	6.736825	7.16E-07	0.343142	0.647262
August	0.704433	0.073711	9.556742	1.79E-09	0.551951	0.856915
September	0.693665	0.073945	9.380796	2.52E-09	0.540697	0.846632
October	0.389563	0.07421	5.249455	2.52E-05	0.236047	0.543078
November	0.132128	0.074505	1.773402	0.089406	-0.022	0.286253
December	0.035359	0.07483	0.472525	0.641005	-0.11944	0.190156

- (a) Interpret the estimated coefficient on *August*. (2 marks)
- (b) In which months does the coffee harvest mainly occur? Approximately how many months of the year would you estimate a typical household spends engaged in harvest? There is no precise answer to this question; just use common sense and explain your reasoning. (3 marks)
- (c) Explain what phenomenon the *Time* variable is capturing. To what extent is this occurring in this district? Explain. (4 marks)
- (d) Forecast the proportion of households engaged in coffee harvest in January 2016 and in August 2016. Remembering the information in the graph in Question 2, what can you deduce about the income situation of households in this district? (4 marks)
- (e) Interpret the *Standard Error* value of 0.089. Would you say this is large? Explain. (3 marks)

Question 5 (14 marks)

A large International NGO plans to implement a program to help farmers improve their coffee yield and hence overall household income. The survey of coffee farmers you have been using formed the baseline stage of their evaluation program.

Your task in this question is to produce a brief report for the NGO, summarising what you have learnt about coffee-farming households and yields to help them design their program.

Use the following headings to guide the content of your report. As a guide to length, you should write about 1 sentence per allocated mark.

(a) Introduction (2 marks)

Introduce the importance of coffee for these households.

(b) Current levels of production and productivity in coffee (5 marks)

Summarise the current situation with respect to how much coffee is being produced by households and how productively.

(c) Potential for improvements in yield (3 marks)

Suggest who/what the NGO's program might target in order to assist farmers in improving their yield.

(d) Recommendations and next steps (4 marks)

Suggest how the NGO might evaluate their program. Draw on the approach to evaluation that was discussed in the lecture on the Quality cycle.