

ST130: Basic Statistics  
Assignment 2, Semester 1, 2021

**Due Date: Friday 20<sup>th</sup> June, 2021, 5pm (Fiji-time)**  
**Weight: 15%**

**Total Marks: 60**

**Instructions:**

1. All questions are compulsory. Use MS excel to complete this assignment.
2. Please complete this assignment on *your own*.
3. Write your information (name, ID, campus, etc.) on the *cover page of the report*.
4. You have to use MS WORD and its math editor to *type your assignment report*. Then convert MS WORD to PDF and upload the PDF document. It must be a single PDF document. You also need to upload the MS excel outputs.
5. All students are to upload the assignment via drop box created on Moodle. Two files are to be uploaded, (1) written report and (2) MS excel output.
6. *Plagiarized assignments* will be given a *mark of 0 (zero)* and will be reported for disciplinary action.

**Access Fiji Farm Survey data set from MOODLE. And use that for all questions.**

Fiji Farm Survey data set is taken from a survey of Sugar cane Farmers in Fiji conducted by a research team from the University of Queensland in 2005.

**Q1. (8 marks)**

- a. Produce a histogram for the variable Age. Take class width as 10, starting from 0. *Give appropriate title, label to axis and it to be solid lines only and no color fill.* **(4 marks)**
- b. Discuss the Histogram result, is there a serious problem in sugar industry regarding farmers age. *If yes then recommend a suitable way to minimize this problem.* **(2 marks)**
- c. “Samu is recommending Pie chart for Farming status” Comment on appropriateness of Samu’s choice. **(2 marks)**

**Q2. (11 marks)**

Create a new column titled “Level of education” with following criteria:

- all the farmers Education(years) less than equal to eight label it as primary
- all the farmers Education(years) greater than 8 and less than equal to 13 label it secondary.
- all the farmers Education(years) greater than 13 label it as tertiary.

**(1 mark)**

Now create a new variable ‘cane output per acre’ in a new column. (Cane output per acre = cane output / cultivated area) **(1 mark)**

- a. Calculate the descriptive statistics using MS-Excel for cane output per acre based for farmers with primary education and for farmers with secondary education. (Hint: ignore non-numeric cells in computing). Interpret each statistic including (Mean, Median, Mode, Standard Deviation, Skewness, Range, and Count) **(7 marks)**
- b. What conclusion can be reach from the comparison of descriptive statistics between cane output per acre for farmers with primary education and for farmers with secondary education? Justify why or why note there is significant difference in output? **(2 marks)**

**Q3. (9 marks)**

Construct a contingency table and relative contingency table (using Pivot table tool in Excel) for farming status in row and Land Owned in column. **(4 marks)**

- a. What is the probability that a randomly selected farmer does not own the land? **(1 mark)**
- b. What is the probability that a randomly selected farmer is working full time and does not own the land? **(1 mark)**
- c. What is the probability that a randomly selected farmer does not own the land given that farming status is full time? **(1 mark)**
- d. What can you conclude from above analysis in regards to land ownership in sugar industry? **(2 marks)**

**Q4. (7 marks)**

- a. Assume that the annual profit from farming is nearly normally distributed. Individual earning less than \$2000 is believed to be in poverty. Calculate what proportion of sugar cane farmers are in poverty? **(2 marks)**
- b. A researcher intends to estimate population proportions of farmers in Fiji are in poverty. The researcher intends to be 99% confident and expects that his estimation to be within 3% of population proportion. How large sample size should be taken to achieve these conditions. **(3 marks)**
- c. Which sampling procedure would you recommend to the researcher in part (b), provide reason for your answer. **(2 marks)**

**Q5. (6 marks)**

- a. Construct a 99% confidence interval for the mean cane output per farmer. Interpret. **(3 marks)**
- b. Construct a 95% confidence interval for population mean can output per acre. Comment on your results. **(3 marks)**

**Q6. (6 marks)**

Test whether the mean annual profit from farming is more than \$2000 or not. Use  $\alpha = 0.05$ .

Show all five steps of Hypothesis testing; excel output and Comment on your results in regards to poverty for sugar cane farmers.

**Q.7 (13 marks)**

It is assumed that annual profit from farming depends on cultivated area.

- a. Draw scatter plot of relationship between annual profit from farming and Cultivated Area, and comment. **(2 marks)**
- b. State and interpret the correlation coefficient. **(2 marks)**
- c. Test whether the coefficient of correlation obtained in (b) is significant at 5% level of significance. Show all five steps of Hypothesis testing. **(6 marks)**
- d. Use simple linear regression to develop regression model. Estimate and interpret your model, intercept and slope. Properly present your model in terms of an equation. **(3 marks)**

**THE END**