

**Assignment 2**  
**(100 Marks)**

**Due Date: Friday, February 28, 2014 at midnight.**

**Late assignments will be accepted with penalty only until midnight on Monday, March 3 2014. Remember to include your Integrity Statement and print or sign your name.**

**General Instructions:**

1. This assignment should be uploaded to Blackboard Learn in ".pdf" format.
2. You may use Minitab for any calculations. However, you must do manual calculations when asked. You may cut and paste your output onto your assignment; however, this output does not replace any of the steps outlined below.
3. If you are performing a hypothesis test, make sure you state the hypotheses, the level of significance (**if not specified, you may assume a default value of 5%**), the decision rule in terms of the critical value, the test statistic or *p*-value, your decision (whether to reject or not to reject the null hypothesis), and your conclusion with a managerial statement. Incomplete tests will not be given full marks.

**Question#1 (30 marks)** Do this question manually, but you may obtain the basic statistics and check the calculations by using Minitab.

In this Question, the column "National Beer: Cals" contains the Calorie values of randomly selected types of beers produced by National brands and "Regional Beer: Cals" contains the Calorie values of randomly selected types of beers produced by Regional brands, also known as micro breweries. For a fair comparison, all the beers are sold in the same size containers. All the values are in Calories with an upper case "C". The relevant data for Question#1 is given in the Minitab file named "Beer\_Calories.MTP".

- a) Test the hypothesis that there is a difference in the mean values of Calories between the two types of beers.
- b) What is the 95% confidence interval for the difference in the mean Caloric values of the two types of beers? Is it consistent with the conclusion reached in part 'a'?
- c) Based on the confidence interval above, state how much more or less, is the mean Caloric value of the Regional beers compared to mean Caloric value of the National beers.
- d) Now for the same data, test the hypothesis that the mean Caloric value of the Regional beers is more than 5 Calories higher than the mean Caloric value of the National beers.
- e) Calculate the appropriate 95% one-sided or asymmetric confidence interval for the test you performed in part 'd' above. Explain what this asymmetric confidence interval means and if this is consistent is with the conclusion you reached in part 'd'.

**Question#2 (15 marks)**

Using the same data you used in Question#1 above,

- a) Test the hypothesis that there is a difference in the median Caloric values for the two types of beers. Use a Level of Significance of 5%.
- b) State the appropriate 95% Confidence Interval for the hypothesis test you performed in part 'a'.
- c) By using appropriate diagrams, justify if the tests used in Question#1 or Question#2 are more appropriate. Name the tests in Question#1 'a', and Question#2 'a'.

**Question#3 (25 marks)**

The data for this question are given in the file "Organic\_FoodBasket\_Prices.MTP". The prices of items in an organic food basket consisting of 30 items are given. In C1, column 1, the food items are specified. In C2 and C3, the prices are given for SMChain1 (Super Market Chain1) and SMChain2 (Super Market Chain2) respectively.

- a) After carefully observing the data, what can you say about the samples? What type/s of test/s would be appropriate in this situation? Justify your answer.
- b) Test whether mean organic food basket prices differ between the two supermarket chains. Use a level of significance of 5%.
- c) Obtain a 95% confidence interval for the test performed in part 'b' above.
- d) Now test whether the median organic food basket prices differ between the two supermarket chains.
- e) Based on appropriate diagrams/graphs, indicate which test is more appropriate. Is it the test in part 'b' or part 'd'?

**Question#4 (30 marks)**

A survey of randomly selected social media users was conducted to find out the number of users who use a given type of "Social Medium" as their first preference by their "Age Group". The various categories of the two variables are given the data table below. For example, "A1" refers to age group 15 to less than 20 years; and "M3" refers to "Twitter". The other categories are defined similarly. Use a level of significance of 5% throughout this question.

Age Group in Yrs	Social Media Used as a First Preference					Total
	M1:E-mail	M2:FBook	M3:Twitter	M4:IGram	M5: Other	
A1:15- <20	50	45	50	25	27	197
A2:20- <30	60	55	50	30	30	225
A3:30- <40	65	45	30	25	25	190
A4:40- <65	55	30	20	15	26	146
A5:>=65	35	45	35	15	17	147
Total	265	220	185	110	125	905

- Calculate manually the "degrees of freedom" and do the calculations manually for the two cells ( $A_1$  &  $M_1$ ) and ( $A_3$  &  $M_5$ ). This means you must manually calculate  $E_{11}$ ,  $\text{Chisquare}_{11}$ ,  $E_{35}$  and  $\text{Chisquare}_{35}$ . However, you may use Minitab to find the value of the test statistic.
- State the hypothesis to test whether the two variables are associated. Name the test you would perform.
- Calculate the appropriate statistic and perform this test by the critical value approach. What conclusion would you draw?
- Explain how the p-value is calculated and state your conclusion. Is this conclusion the same you have reached in part 'b' above?
- The different  $p_i$ 's are the population proportions for the various  $M_i$ 's. It is claimed that the population proportions are as follows:  $p_1 = 35\%$ ,  $p_2 = 25\%$  and  $p_4 = p_5 = 10\%$ ,  $p_3 =$  the rest. Test this hypothesis at the 5% level of significance using the critical value approach. Also find the p-value and reach the conclusion consistent with the p-value.