**Statistical Analysis (I): Estimation & Testing**

**Batch: CBA 2018 Winter (Hyderabad)**

**Group Assignment**

**Due Date: 24th Dec, 11:55 pm**

**Computer Assignment**

Import the dataset in R. Total there are 1348451 rows and from this data select 10000 rows randomly using “**sample**” function in R without replacement and define this data as new data. And now consider this 10000 observations as sample data.

Based on your sample, carryout the following analysis. Your analysis and the interpretation should be submitted by the due date. Please note that there is no penalty for early submissions!

1. Calculate a 95 percent confidence interval for the “Gross output – Year 3 (Rs)”
2. Define two different measures that you consider most appropriate for measuring the performance of the units. This definition is up to you. These can be the variables that are already in the data or new variables defined based on the existing variables. Please explain in one paragraph why you have selected these two measures. Remaining analysis is to be carried based on these definitions.
3. What is the probability that a firm selected at random is a SSSBE unit?
4. What is the probability that a firm selected at random is GOOD in performance? (Calculate the average of the first performance measure that you had defined in question 2 above. If the firm’s performance is above this average, it considered good. If it is below average, it is considered Bad)
5. What is the probability that a firm selected is a SSSBE Unit and ALSO GOOD in performance?
6. What can you say about the performance of the SSSBE units in terms of GOOD or BAD based on the above?
7. Test the null hypothesis that the population average of the variable “Value of Exports for Year 3” = 87,300. Carry out a one sided test. Clearly state your null and alternate hypotheses.
8. Is there a significant difference in the mean value (µ) of “gross output for Year 3(in Rs)” between SSSBE units and SSI units?
9. Consider two variables, namely “technical know how obtained from (KNOW-HOW)” and “type of organization (ORG\_TYPE)”. Is there any evidence to say that these two variables are independent of each other?
10. Some male chauvinists like to think that the productivity of women employees is significantly less than that of the male employees. The productivity is to be measured based on the two metrics defined by you. Do you agree or disagree with this? Justify your answer with appropriate analysis.