**Programming with R**

**Assignment – 3**

**Max. Marks: 40 Due Date: 16 Mar, 2020**

**Note:**

Assignment to be submitted in Blue Book only

Mention the hypotheses for the test where applicable

Except for Q1, all questions carry equal marks

**Answer all the questions**

1. Create a subset of the placement dataset to answer all the remaining questions. Set a seed of 420 and randomly select 100 observations. Name the file *myassign*. **[1]**
2. Show the summary of mba\_p. What functions will you use to separately calculate all the values shown by the summary?
3. Calculate various measures of variation of mba\_p.
4. Create a two-way table *mytab* between specialization and status using the table() function and with() function. Add table margins to *mytab*. How will you show proportions over rows?
5. Construct a three-way crosstab involving specialization, status and gender.
6. Calculate correlation between all numeric variables in *myassign*.
7. Assess the normality of degree\_p using a density plot, histogram, and a Q-Q plot.
8. Use the Shapiro-Wilk’s test to check for normality of degree\_p and mba\_p.
9. Use an appropriate test to check whether there is significant correlation between degree\_p and mba\_p.
10. Use an appropriate test to check whether the mean mba\_p differs from 55%. How will you set-up the test if you had to prove that the average mba\_p is greater than 60%?
11. How will you test whether there is a significant difference in mba\_p based on specialisation?
12. Using ANOVA check whether stream in HSC (hsc\_s) had any significant impact on the average Degree percentage (degree\_p)?
13. Check whether the specialisation and placement status are independent.
14. Check whether there is any significant difference in salary between the levels of specialisation.