****

Symbiosis Centre for Management and Human Resource Development

MBA-BA (2018 - 2020) Sem –I

MID TERM

Subject: Analytics Foundations

Date: 22/08/2018 **Marks: 30** **Duration: 60 Min.**

****

Please write all your answers in a script file. Save the script file on the server following instructions given to you.

1) The MASS library has a dataset called **Cars93**. To load the dataset type: **library(MASS)** and then **data(Cars93)**. The dataset is now available to you for analysis.Type **names(Cars93)** to see what the columns of this dataframe are. Now answer the following questions:

a) Find the Pearson correlation between city mileage and highway mileage (variables: **MPG.city**and **MPG.highway**).

b) What are the different types of cars in the data (variable **Type**) and how many of each type are there in the data? Find the mean weight of the cars of each type.

c) Test the hypothesis that the mean mileage for city driving is 24.

d) Restrict to four passenger cars (variable: **Passengers**). How many such cars does the dataset have? Test the hypothesis that the mean highway mileage for four passenger cars is 29.

e) Restrict the four passenger cars to those whose price is more than 19 (variable: **Price**).

i) How many such cars?

ii) Produce a boxplot for the price for these cars

iii) Test the hypothesis that the mean price for these cars is 26

2) Incomes of people in a population are normally distributed with *mu* = 90 and *sigma* = 22

a) The percentage of people with incomes between 82 and 98 is _____

b) 30% of the population has income more than _____

c) 60% of the population has income less than _____

d) The percentage of people with income more than 110 is _____

3) A manufacturer claims that his packets of peanuts have a mean weight of 200 gm. To test this, an inspector picks a random sample of 60 packets and finds that the mean weight of a packet in the sample is 195 gm with a standard deviation of 15 gm. Do you think that the manufacturer's claim is valid?

4) While R has a built-in function to calculate the standard deviation, it has no built-in function to calculate the standard error. Write such a function. Your function should take a vector of numeric data as input and print the standard error.

==========