The Analytics Edge

Fall 2018

Concept Questions

1. Suppose you want to create a vector of the numbers 100 to 1, where 100 is written 100 times, 99 is written 99 times and so on till 1 is written 1 time, namely $100, 100, \ldots, 3, 3, 3, 2, 2, 1$. Provide a R command that can help you do this. Hint: Check the **rep** command in R. rep([100:1], [100:1])
2. Suppose you have a vector A <− c(1,2,0,4) and a vector B <− c(3,6), then what is the result of A∗B in R? (3,12,0,24)
3. Run the following R commands and explain the numbers that appear, each time we run table(gender).

   > gender <− factor(c(rep("female", 91), rep("male", 92)))
   > table(gender)
   > gender <− factor(gender, levels=c("male", "female"))
   > table(gender)
   > gender <− factor(gender, levels=c("Male", "female"))
   > table(gender)

4. Suppose we want to convert a factor variable to a numeric variable in, how do we do so? For example, convert the factor variable below to numeric.
   > X <− factor(c(4, 5, 6, 6, 4)) as.numeric(X)
5. Load the dataframe mtcars available with the base R installation in data. The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models).
   Do a ggplot() of the weight versus miles per gallon and comment on the relationship. Add to this plot, a coloring of the points based on the number of cylinders, and the sizing of the points based on the displacement (volume) of the car.
   ggplot(mtcars, aes(mtcars$mpg, mtcars$wt))+geom_point()
6. Use the tapply() function to compute the standard deviation of the mpg for groups with the same number of cylinders.
   tapply(mtcars$mpg, mtcars$cyl, sd)