PESIT Department of Computer Science and Engineering

Course: Data Analysis
Semester: 2014 Spring (January – May)
Instructor: BNR (Dr. B. Narsing Rao)

Assignment: 04
Topic: Sampling
Due by: Midnight on Tuesday, February 4, 2014
Method: Send zip archive (.zip, .rar, etc.) by email to bnrao@pes.edu
The name of the zip archive should be: DA-A04-your USN-your name
(USN must be upper case and your name should be in mixed case)
The zip archive should contain the following (see below for details):

1. PDF report including the output (see below) along with your observations
2. Source file containing R functions (named DA-A04-USN-Name.R) used

Distribution of the Sample Mean

Use the supplied file bank-data.csv for this assignment. This file contains information about 600 customers of a bank. However, for the purpose of this assignment, only the income attribute is relevant; therefore you should read in the data and extract the income attribute into a numeric vector. For our purpose, this vector will be the **population**.

The purpose of this assignment is to observe the distribution of the **sample mean** for various sample sizes. You will need to use the sample function in R with the default option which does sampling without replacement.

Write an R function that generates 30 samples of progressively increasing sizes and generates a distribution of the sample mean.

The report that you submit should be in the format shown below:

1. Histogram and summary measures of the original data (i.e. 600 incomes)
2. Table using the format shown below (note that each time you use the sample function, you will get a different sample)

Sample size	Histogram of 30 sample means	Mean of 30 sample means	Standard Deviation of 30 sample means	Observations on the distribution of sample means
5				
15				
30				
50				