jignesh.bhatt@iiitvadodara.ac.in

IIIT Vadodara
Autumn 2020-21
CS605 Data Analytics
Lab-1 | July 30, 2020.
Introduction to R and Graphical Statistics

Follow the link to download R: `https://cran.r-project.org/`

Q. 1: The numbers of blocked intrusion attempts on each day during the first two weeks of the month were
56, 47, 49, 37, 38, 60, 50, 43, 43, 59, 50, 56, 54, 58.
After the change of firewall settings, the numbers of blocked intrusions during the next 20 days were
53, 21, 32, 49, 45, 38, 44, 33, 32, 43, 53, 46, 36, 48, 39, 35, 37, 36, 39, 45.
Comparing the number of blocked intrusions before and after the change, (a) construct side-by-side stem-and-leaf plots; (b) compute the five-point summaries and construct parallel boxplots; (c) comment on your findings.

Q. 2: A network provider investigates the load of its network. The number of concurrent users is recorded at fifty locations (thousands of people),
17.2 22.1 18.5 17.2 18.6 14.8 21.7 15.8 16.3 22.8 24.1 13.3 16.2 17.5 19.0 23.9 14.8 22.2 21.7 20.7 13.5 15.8 13.1 16.1 21.9 23.9 19.3 12.0 19.9 19.4 15.4 16.7 19.5 16.2 16.9 17.1 20.2 13.4 19.8 17.7 19.7 18.7 17.6 15.9 15.2 17.1 15.0 18.8 21.6 11.9
(a) Compute the sample mean, variance, and standard deviation of the number of concurrent users. (b) Estimate the standard error of the sample mean. (c) Compute the five-point summary and construct a boxplot. (d) Compute the interquartile range. Are there any outliers? (e) It is reported that the number of concurrent users follows approximately Normal distribution. Does the histogram support this claim?

Q. 3: Consider three data sets.
(1) 19, 24, 12, 19, 18, 24, 8, 5, 9, 20, 13, 11, 1, 12, 11, 10, 22, 21, 7, 16, 15, 15, 26, 16, 1, 13, 21, 21, 20, 19
(2) 17, 24, 21, 22, 26, 22, 19, 21, 23, 11, 19, 14, 23, 25, 26, 15, 17, 26, 21, 18, 19, 21, 24, 18, 16, 20, 21, 20, 23, 33
(3) 56, 52, 13, 34, 33, 18, 44, 41, 48, 75, 24, 19, 35, 27, 46, 62, 71, 24, 66, 94, 40, 18, 15, 39, 53, 23, 41, 78, 15, 35.
(a) For each data set, draw a histogram and determine whether the distribution is rightskewed, left-skewed, or symmetric. (b) Compute sample means and sample medians. Do they support your findings about skewness and symmetry? How?

Q. 4: The following data set represents the number of new computer accounts registered during ten consecutive days.
43, 37, 50, 51, 58, 105, 52, 45, 45, 10.
(a) Compute the mean, median, quartiles, and standard deviation. (b) Check for outliers using the 1.5(IQR) rule. (c) Delete the detected outliers and compute the mean, median, quartiles, and standard deviation again. (d) Make a conclusion about the effect of outliers on basic descriptive statistics.

"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to the reality." - ***Albert Einstein***.