

# Big Data Analytics

## Lab 9

### 1 Description

In this lab we are going to focus on estimations (part A) and hypothesis testing (part B). Complete part B after today's lecture.

### 2 Estimation - Part A

1. Download `data.csv` file from Moodle. You can load it using `read.csv` command.
2. This file contains some data generated using some normal distribution.
3. Manually calculate the confidence intervals for the confidence level of 95%.
4. Use `t.test` function for the same.
5. The standard deviation for the data is 2. Manually calculate the confidence intervals.
6. Using the `qqnorm` function examine the normality of the data.
7. Download from Moodle and load `data2.csv` file.
8. Was the data set generated using a normal distribution?

### 3 Hypothesis Testing - Part B

1. Suppose a car manufacturer claims a model gets 25 mpg. A consumer group asks 35 owners of this model to calculate their mpg and the mean value was 22 with a standard deviation of 1.5. Is the manufacturer's claim supported?
2. Suppose that we want to test the hypothesis with a significance level of 0.05 that the climate has changed since industrialization. Suppose that the mean temperature throughout history is 50 degrees. During the last 40 years, the mean temperature has been 51 degrees and suppose the population standard deviation is 2 degrees. What can we conclude?

3. An engineer measured the Brinell hardness of 25 pieces of ductile iron that were subcritically annealed. The resulting data were: 170, 167, 174, 179, 179, 156, 163, 156, 187, 156, 183, 179, 174, 179, 170, 156, 187, 179, 183, 174, 187, 167, 159, 170, 179. The engineer hypothesized that the mean Brinell hardness of all such ductile iron pieces is greater than 170.
4. A biologist was interested in determining whether sunflower seedlings treated with an extract from *Vinca minor* roots resulted in a lower average height of sunflower seedlings than the standard height of 15.7 cm. The biologist treated a random sample of  $n = 33$  seedlings with the extract and subsequently obtained the following heights: 11.5, 11.8, 15.7, 16.1, 14.1, 10.5, 15.2, 19.0, 12.8, 12.4, 19.2, 13.5, 16.5, 13.5, 14.4, 16.7, 10.9, 13.0, 15.1, 17.1, 13.3, 12.4, 8.5, 14.3, 12.9, 11.1, 15.0, 13.3, 15.8, 13.5, 9.3, 12.2, 10.3. The hypothesis is that the mean has decreased.