**­­­**

**Statistical Methods for Data Science (Spring 2017)**

Mini Project 3

Contributing members:

**Akhilesh Kumar Kagalvadi Chinnaswamy (axk167131)**

**Vidya Sri Mani (vxm163230)**

**Contribution of Group Members:**

We worked on the problems together trying to find different options to find the results and worked together to compile the report.

Exercise 1 (10 points)

Consider the dataset stored in the file bp.txt. This dataset contains one measurement of systolic blood pressure (in mmHg) made by each of two methods—a finger method and an arm method—from the same 200 patients.

(a) Perform an exploratory analysis of the data by examining the distributions of the measurements from the two methods using boxplots. Comment on what you see. Do the two distributions seem similar? Justify your answer.

Summary of data set and Box plot:

armsys fingsys

Min. : 79.0 Min. : 60.0

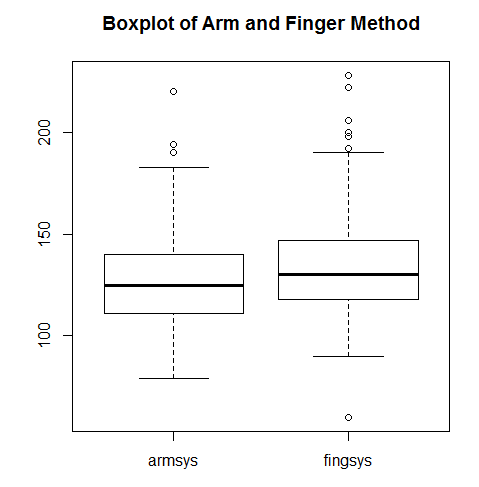
1st Qu.:111.5 1st Qu.:118.0

Median :125.0 Median :130.0

Mean :128.5 Mean :132.8

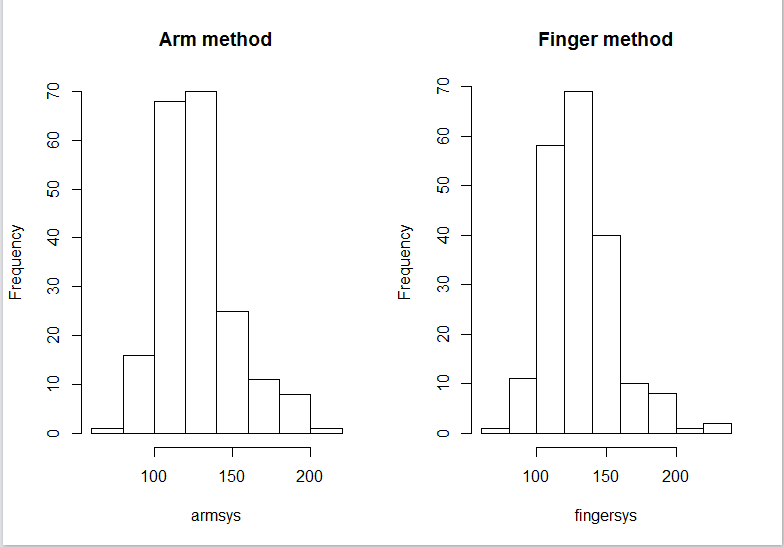
3rd Qu.:140.0 3rd Qu.:146.5

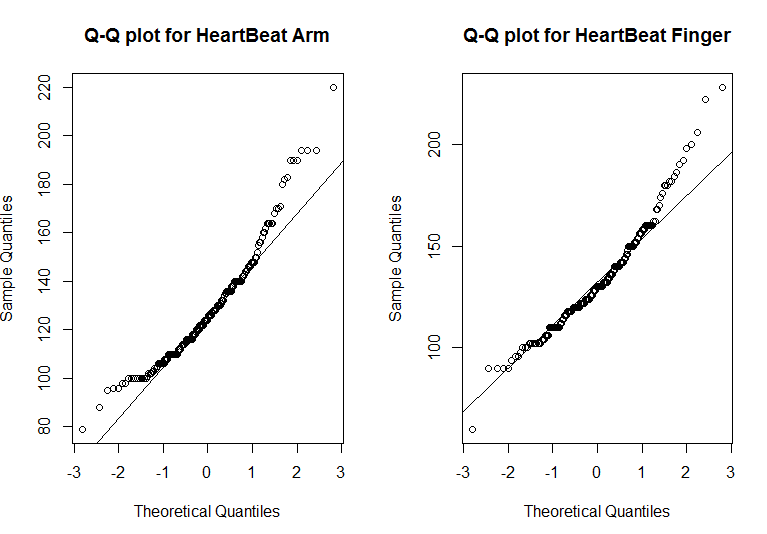
Max. :220.0 Max. :228.0



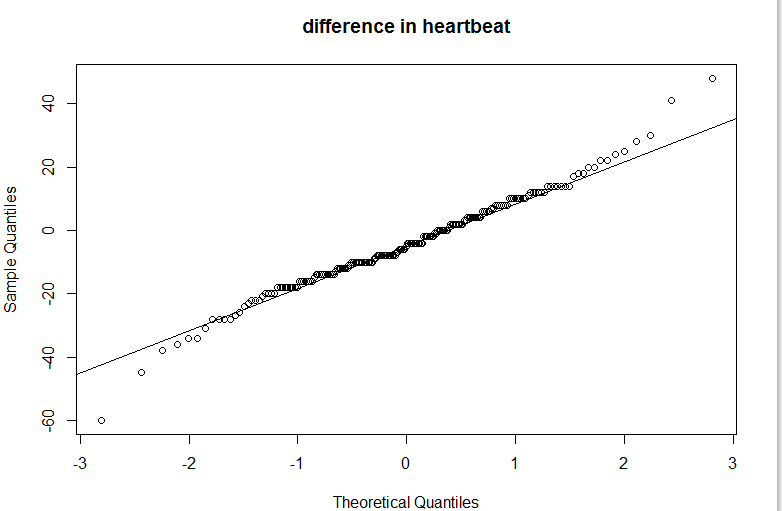
From the boxplot, both the data have some outliers and they appear to be little right skewed. However, the distributions seem similar.

(b) Use histograms and QQ plots to examine the shapes of the two distributions. Comment on what you see. Does the assumption of normality seem reasonable? Justify your answer.





(c) Construct an appropriate 95% confidence interval for the difference in the means of the two methods. Interpret your results. Can we conclude that the two methods have identical means? What assumptions, if any, did you make to construct the interval? Do the assumptions seem to hold? Justify all your answers.



The difference is very normal when plotted.

The confidence interval comes out to be

#[1] -6.316529 -2.273471

From the data since 0 is not present in the range, the two methods do not have identical means.

R Code: part-1

