

## STAT2053 – Introduction to Probability and Statistics Course

*Marmara University, Istanbul, Turkey*

*March 07, 2020*

### Homework 1

Blood pressure is the pressure that the blood exerts against the walls of the arteries. When physicians or nurses measure your blood pressure, they take two readings. The systolic blood pressure is the pressure when the heart is contracting and therefore pumping. The diastolic blood pressure is the pressure in the arteries when the heart is relaxing. The diastolic blood pressure is always the lower of the two readings. Blood pressure varies from one person to another. It will also vary for a single individual from day to day and even within a given day.

If your blood pressure is too high, it can lead to a stroke or a heart attack. If it is too low, blood will not get to your extremities and you may feel dizzy. Low blood pressure is usually not serious.

So, what should your blood pressure be? A systolic blood pressure of 120 would be considered normal. One of 150 would be high. But since blood pressure varies with gender and increases with age, a better gauge of the relative standing of your blood pressure would be obtained by comparing it with the population of blood pressures of all persons of your gender and age in the country. Rather than supplying the whole data set, a very large sample selected from it has been provided you in the course webpage. The blood pressure data on 200 persons, 100 men and 100 women between the ages of 15 and 20, are found at the course webpage (use your Student ID to generate the data specific for your homework). The data are part of a health survey conducted by the National Institutes of Health (NIH), in US. Entries for each person include that person's age and systolic and diastolic blood pressures at the time the blood pressure was recorded.

1. Describe the variables that have been measured in this survey. Are the variables quantitative or qualitative? Discrete or continuous? Are the data univariate, bivariate, or multivariate?
2. What types of graphical methods are available for describing this data set? What types of questions could be answered using various types of graphical techniques?
3. Find the followings for men (denoted with gender value 0) and women (denoted with gender value 1) both for the systolic blood pressure and the diastolic blood pressure.
  - a.) Calculate the sample mean,  $\bar{x}$ .
  - b.) Calculate the sample variance,  $s^2$ .
  - c.) Find the sample standard deviation,  $s$ .
  - d.) Find the lower and upper quartiles.
  - e.) What are the maximum and minimum values?
  - f.) Calculate the range.
  - g.) Compare the range and the standard deviation. The range is approximately how many standard deviations?

- h.) What is the median?
- i.) Calculate the IQR.
- j.) Calculate five-number summaries.
- k.) Construct box plot for the data set. Are there any outliers? What does the box plot tell you about the shape of the distribution?
- l.) Construct and interpret stem-and-leaf plot
- m.) Construct and interpret histograms. In this solution, you will use different number of subintervals to construct the histograms and then compare the results
  - 1) use 5 subintervals to construct the histograms and then interpret the graph
  - 2) use more subintervals (more than 5 subintervals) to construct the graph and then interpret the graph
  - 3) compare two results above and discuss which one is better and why.
- n.) Draw a dotplot of this data set. Are the data mound shaped?
- o.) Can you use Tchebysheff's Theorem to describe this data set? Why or why not?
- p.) Can you use the Empirical Rule to describe this data set? Why or why not?
- q.) Given data sets of two groups, construct and interpret a comparative dotplot to compare the groups
- r.) Construct a relative frequency histogram for these data. How would you describe the shape of the distribution?
- s.) Find the z-scores for the largest and smallest observations. Would you consider them to be outliers? Why or why not?

#### **Deliveries:**

1. The **R code** you developed (*RStudio*). For each question, you have to provide the corresponding R code (except m and n).
2. A **report** that includes the answers to the questions above.

This is an individual homework for students. Collaboration and cooperation between students are not allowed.

Due date is March 23, 2020, until 23:59.

Please post your deliveries by your *TurnItIn* account.

Ask any unclear matter to the lecturer. Good luck...

Mujdat Soy Turk, Ph.D.  
Associate Professor