

Problem Set 1

Statistics - NYU, Summer 2016
Ercan Karadas

- [1] A random sample of customers was asked to select their favorite soft drink from a list of five brands. The results showed that 30 preferred brand A, 50 preferred B, 46 preferred C, 100 preferred D, and 14 preferred brand E.
- a) Construct a pie chart.
 - b) Construct a bar chart.
- [2] For this exercise download the data file `Price-Quantity` from the Data folder on NYU-Classes.
- a) Group the variable `Quantity` into five classes, and construct a relative frequency distribution, cumulative frequency distribution in the same table.
 - b) Using the classes in the first part, construct a histogram for the variable `Quantity`.
 - c) Use the `Price` and `Quantity` variables to obtain a scatter plot of the data.
 - d) Does the scatter plot have the pattern that economic theory predicts between these two variables?
- [3] Suppose that you have data on Education and annual Family Income of 10 individuals (only part (a) should be done in Excel, in other parts show your work/algebra):

Individual #	1	2	3	4	5	6	7	8	9	10
Education	18	14	15	11	16	14	17	16	12	18
Family Income	110	80	75	60	90	70	120	150	60	70

- a) Enter the data to an excel sheet and save as `Education-Income.xlsx` (We will use this file again in Problem Set 2).
 - b) Obtain a scatter-plot.
 - c) Calculate the median for both `Education` and `Family Income` variables.
- [4] For this exercise download the data file `Marathon.xlsx` from the Data folder on NYU-Classes. In the data file, finishing times for male and female winners of the New York Marathon between 1970 and 1999 are given.

- a) Draw a histogram of these data and describe the distribution. For the histogram, use the following six grouping intervals:

$$2 - 2.2, \quad 2.2 - 2.4, \quad 2.4 - 2.6, \quad 2.6 - 2.8, \quad 2.8 - 3.0, \quad 3.0 - 3.2$$

- b) What features of the distribution are apparent in the histogram?
c) What may be the reason for the bimodal distribution? Explain.
d) Now by using data handling features in Excel, separate the data for male and female winners.
e) Draw a histogram of these data again for the two series and compare them with the previous case.
f) Draw the time series plot for both male and female winners on the same diagram.
g) Describe what is visible in this plot but not in the histogram.

- [5] For this exercise download the data file **Economic Activity 2015** from the Data folder on NYU-Classes. Suppose you would like to get a quick idea about how the composition of personal consumption expenditures has evolved in the last fifty years.

- a) Plot gross domestic product, expenditures on goods and services on the same chart after obtaining annual data for these three variables.
b) Plot the share of expenditures on goods and on services as percentages of gross domestic product.
c) Discuss what can you conclude regarding the evolution of personal expenditures over the last fifty years. Is there any pattern in the data?

- [6] For this exercise download the data file **College Data** from the Data folder on NYU-Classes. To address each of the following questions, discuss what kind of graphical tools would be more effective and then obtain the respective chart or table.

- a) You would like to know the percentage of students with GPA below 3.0, 3.1, 3.2, etc.? Obtain a frequency distribution table (with column for cumulative distribution) and a histogram.
b) Someone claims that male students show higher variation in terms of their GPA compare to females. Can you evaluate this claim by using a scatter plot?