ASDS - Applied Statistics with R

Fall 2018, YSU

Homework No. 02

Due time/date: 21:20, 05 October, 2018

Note: Please use R only in the case the statement of the problem contains (R) at the beginning. Otherwise, show your calculations on the paper. Supplementary Problems will not be graded, but you are very advised to solve them and to discuss later with TA or Instructor.

Problem 1. Last Fall I was teaching Financial Math at YSU. Here are student's final grades for that course[1]

$$8, 19, 8, 17, 19, 18, 14, 11, 11, 20, 20, 0, 10, 14, 19, 19, 0, 18, 8, 8, 4, 0, 13, 9,$$

$$4, 6, 4, 3, 14, 9, 0, 17, 9, 18, 9, 16, 12, 0, 1, 2, 10, 4, 0, 9, 7, 15, 13, 12, 0, 11, 0, 14, 2, 15, 16, 17, 4$$

a. Construct the Frequency and Relative Frequency table for FM Grades data;

b. At YSU, we use the following convention:

excellent - grade is 18, 19, or 20;
good - grade is from 13 (inclusive) to 18 (exclusive);
satisfactory - grade is from 8 (inclusive) to 13 (exclusive);
unsatisfactory - grade is below 8.

Using these class intervals, construct the Frequency Histogram, Relative Frequency Histogram, and the Density Histogram for this FM Grades data.
What can be said about my course grades distribution?

Problem 2. As a continuation of the previous problem, Fig. 1 shows my FinMath courses grades (grouped by *excellent*, *good*, *satisfactory* and *unsatisfactory* classes) for the last year and the previous year. Can you compare the results? Try to explain it either from the point of view of my teaching style (say, I have changed my teaching style, and it gave nice/bad etc. results) or from the point of view of students (say, my 2016 class students were studying better/worse etc.)

Problem 3. I have the following Stem-and-Leaf plot in **R**:

[1] At YSU, the lowest grade is 0, and the highest is 20, all grades are integers.

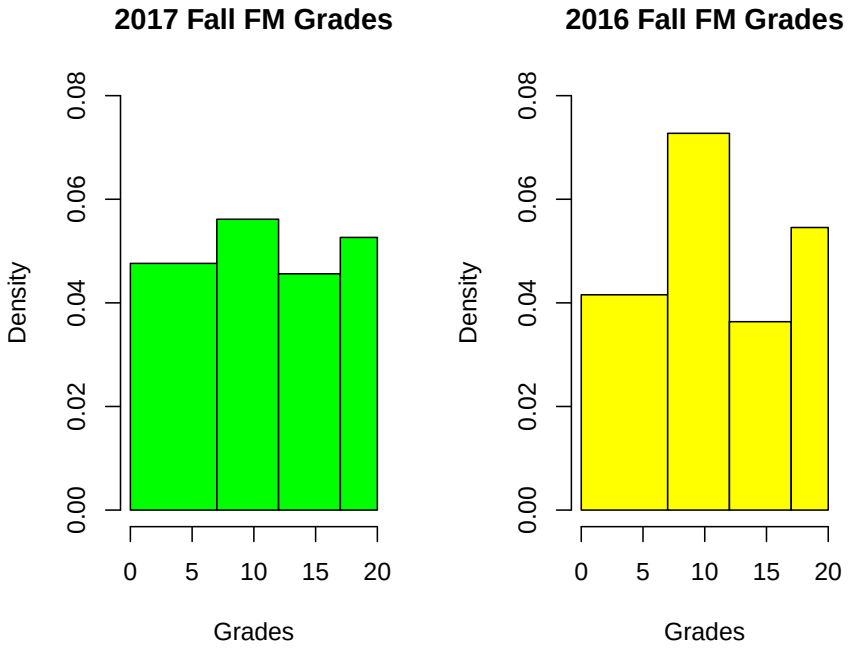

Figure 1: Histograms of FM Grades.

-1	4
-0	66
-0	421
0	0122334
0	779999
1	24
1	566
2	2
2	6

R is giving me the following info: "The decimal point is 2 digit(s) to the left of the |". Reconstruct my data, assuming no rounding was made in my data.

Problem 4. I have obtained the histogram shown in Fig. 2 in **R**:

Describe my data as good as you can (say, describe, if possible, the number of data points, min, max, mode, median, mean, gaps, symmetries, modality (unimoda, multimodal),...). My data points were integers only.

Problem 5. (**R**) Go to Yahoo Finance Page, https://finance.yahoo.com, navigate to the Dow Jones Industrial Average **(Dow 30)** Index webpage, https://finance.yahoo.com/quote/^DJI?p=^DJI, then choose Historical Data, 1 year time period, and weekly frequency. Download that Data.

a. Using Microsoft Excel, R or any other software, using the Adjusted Close Prices

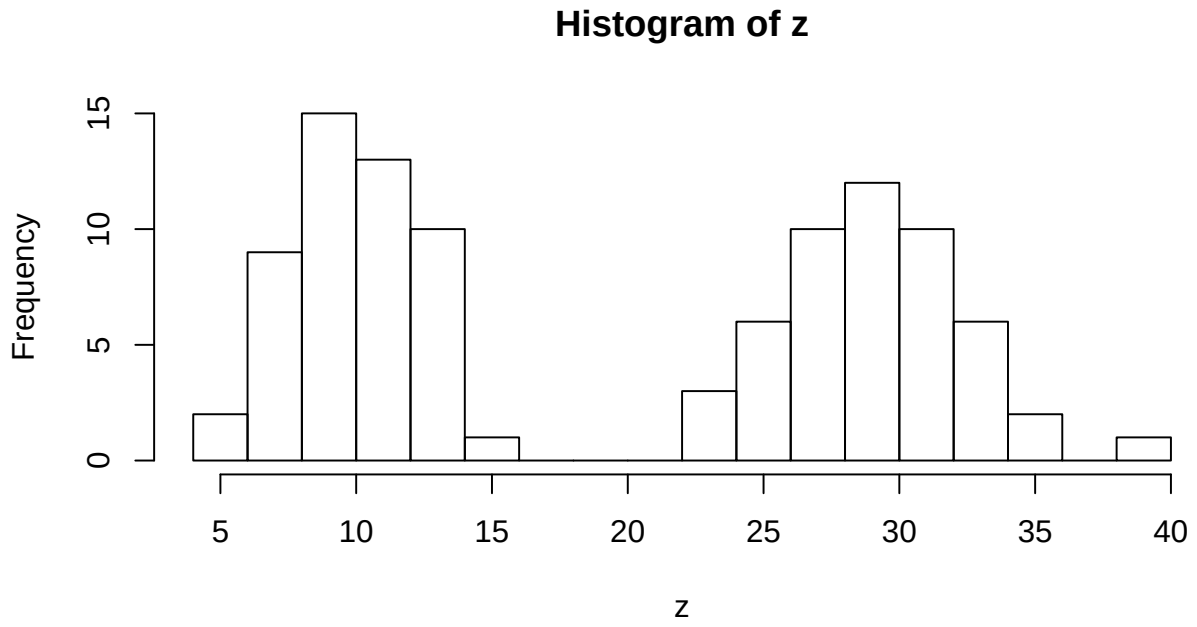

Figure 2: Histogram.

("Adj Close" column), calculate weekly returns of the Dow Jones Index[2].

b. Plot the histogram of weekly returns;

c. Plot the Stem-and-Leaf plot of weekly returns.

Describe the results.

Problem 6. The following is the part of my Financial Math grades data from the Problem 1 (this is our sample):

$$8, 19, 8, 17, 19, 18, 14, 11, 11, 20, 20, 0, 10, 14, 19, 19, 0, 18, 8, 8, 4, 0, 13, 9.$$

a. Calculate the sample mean grade from this data;

b. Find the 5-th and 10-th order statistics;

c. Calculate the sample median grade from this data;

d. Calculate the weighted sample mean grade from this data, taking weights as 1:2:3:4 for, respectively, unsatisfactory, satisfactory, good and excellent grades.

[2] A return for some time period is the ratio

$$\text{Return} = \frac{\text{Last Price - First Price}}{\text{First Price}},$$

where the Last Price is the price at the end of the period, and the First Price is the price at the beginning of the period.

So the return shows the percentage change during that period:

$$\text{Last Price} = \text{First Price} \cdot (1 + \text{Return}).$$

e. Calculate the trimmed sample mean grade from this data by leaving out first and last 2 observations;

Problem 7. We are given a frequency table for some data:

Value	Frequency	Value	Frequency
1	0	10	5
2	4	11	3
3	3	12	8
4	7	13	1
5	2	14	0
6	3	15	0
7	7	16	6
8	7	17	7
9	4	18	8

a. Construct the histogram for this data using bins of width 2, starting from 1 (i.e., $[1,3)$, $[3,5)$,...);

b. Find the mean of this data;

c. Find the median of this data;

d. Find the mode of this data;

Problem 8. Assume we have a dataset $x_1, x_2, ..., x_n$, and we make a transformation

$$y_k = \alpha \cdot x_k + \beta, \qquad k = 1, 2, ..., n,$$

where α, β are some real numbers, not necessarily positive. Express the mean and the median of y in terms of the mean and median of x, respectively.

Problem 9. We are given a dataset

$$2, 2, 2, 5, 3, 2, 0, 0, 3, 5.$$

a. Construct the Empirical CDF for this dataset.

b. (**R**) Construct the Density Histogram for this dataset.

c. (**R**) Construct the Kernel Density Estimator graph for this dataset on the same plot, over the Histogram.

Note: Take care of the alignment of the x-axis values and y-axis values.

Problem 10. Fig. 3 gives the histogram for the number of yearly discoveries:

Calculate approximately, as good as you can, the mode, median and mean for this dataset.

Problem 11. We consider the following dataset:

$$4, 3, 1, 0, -2, 3, 2$$

a. Find the range of this dataset;

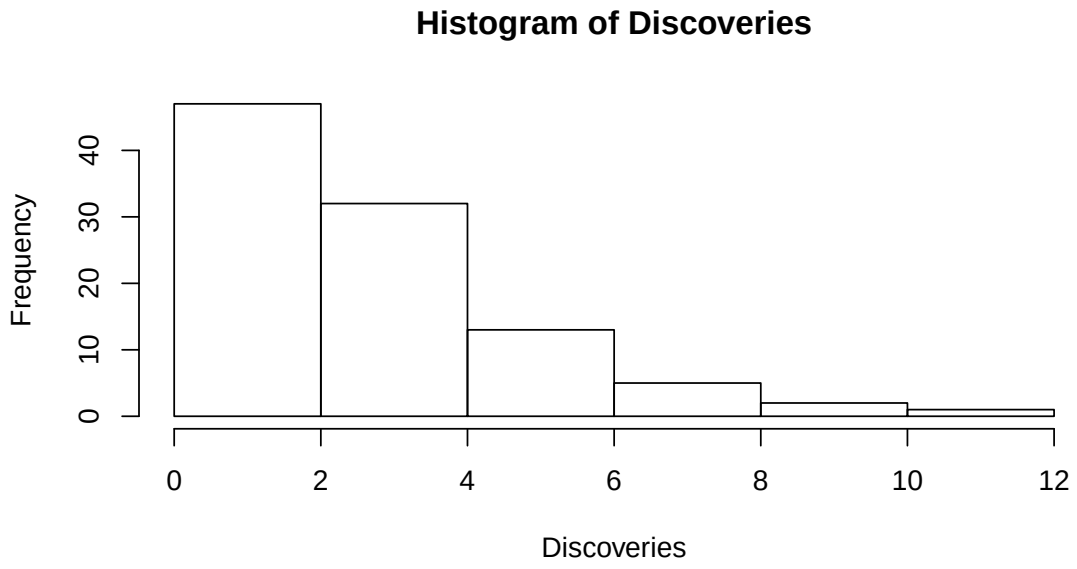

Figure 3: Histogram

b. Find the sample variance (using $n-1$ in the denominator);

c. Find the sample standard deviation;

d. Find the Mean Absolute Deviation from the Mean.

Problem 12. (Supplementary) (**R**) Write an **R** function to calculate all modes of a given dataset.

Problem 13. (Supplementary)

a. Give an example of a dataset x satisfying

$$\bar{x} < Median(x) < Mode(x)$$

b. Give an example of a dataset x satisfying

$$Mode(x) < Median(x) < \bar{x}.$$

c. Assume for a dataset x,

$$\bar{x} = Median(x).$$

Is it true that x is symmetric about the value $\bar{x} = Median(x)$? Prove or give a counterexample.