

### Homework Assignment 3

NOTE: You may use R to verify or check any part of this homework assignment.

1) (10 points) For the following set of 12 scores which represent discrete quantities with no round-off error, compute:

- The range
- The semi-interquartile range
- The variance
- The standard deviation

List them from smallest to largest. Show that both the raw score computational formula and the deviation score formula give the same result by computing the variance **using both formulas**. Show your work!

59, 46, 43, 65, 34, 52, 52, 41, 57, 55, 40, 68
--

2) (10 points) Assume you are working as a research assistant in a lab and your professor gives you the following set of scores from a child development measure of intellectual performance administered to children aged between 3 and 5.

She asks you to enter them into a file and report the mean and standard deviation to her so that she can take a look at them later for further analyses.

Imagine you have computed mean and standard deviation (they are given in the table below).

a) Construct a histogram plot with 9 or 10 categories of the data. You may do this with R if you wish. What do you see? Be sure to include your histogram in your homework handed in.

b) Do you think mean and standard deviation are representing the data well? Or is there more of a story?

c) What could be the reasons for a pattern like the one you see?

29, 19, 17, 27, 10, 26, 20, 37, 16, 26, 24, 15, 21, 31, 8, 22, 25, 17, 18, 18, 15, 17, 32, 29, 32, 66, 46, 73, 59, 57, 59, 69, 68, 56, 64, 67, 77, 75, 50, 76, 66, 65, 85, 66, 64, 65, 71, 75, 76, 66
---

Mean = 44.24
--------------

Standard Deviation $\approx$ 23.81
------------------------------------

3. (10 points) Suppose you have 24 numbers with a mean of 14, and you add an additional number to the group. This new number is 64. What will be the mean of your data?

4. (20 points) Compute the mean *and variance* from the following frequency distribution. Show all work. Hint: Computing the variance requires both the sum of the  $X$  values and the sum of the squares of the  $X$  values. You should be able to figure out how to get the latter, even though we did not explicitly discuss this in class.

$X$	$f$
20	5
19	6
18	13
17	5
16	1

5. (15 points) Compute an *estimate* of the mean from the following *grouped* frequency distribution.

$X$	$f$
36–40	3
31–35	11
26–30	15
21–25	10
16–20	1

6. (20 points) An online data file *homeruns.txt* has the R commands for entering data into a variable called *homeruns*. These data are the numbers of home runs hit by the leading home run hitter in the National League for each of the last 84 years. Download the file or examine it online and copy its contents to your clipboard, then start up R and enter the data. Then use R to do the following:

- Compute the mean.
- Compute the standard deviation.
- Construct a histogram. Print it and include it in your homework. Describe what you see.
- Construct a boxplot. Does the boxplot indicate any positive outliers? What are the values?
- (optional). Do you know anything about the history of baseball that would help illuminate what you see in these results?

7. (15 points). Use R to rescale the homerun variable so that it has a mean of 0 and a standard deviation of 1. What is the largest score in the transformed file (There is a single R command you can use to get this)? Hand in the precise R commands you used to do this.