

Please submit your annotated R code in a file titled hw02.R. Give an example showing the output from your functions where it is possible to include in R.

1. The file `samplegrades.csv` contains simulated course grade data for a large section of introductory statistics.

The format of the file is Comma Separated Values (CSV), which is the text-based format for Excel spreadsheets. R has a specific function for reading in such files. Download the file, put it in your working directory, and then execute the command

```
grade.data <- read.csv("samplegrades.csv", header=TRUE)
```

This will create a data frame named `grade.data`. Use this to answer the following questions about students in this course.

- (a) It is generally thought that students who score well on homework will also do well in the course. Compare the course averages of students who had the maximum possible homework score with those whose homework score was in the bottom 10%. You can find a rough confidence interval for both groups by using the formula

$$\bar{x} \pm 2 \frac{s}{\sqrt{n}}$$

where  $s$  is the sample standard deviation for your set (function `sd` in R) and  $n$  is the size of the set.

- (b) It is speculated that consistent students do better in courses. Compare the course averages of those students who had midterm and final scores within 2% of each other with those students whose exams scores were at least 10% apart. Find approximate confidence intervals for each using the above.
- (c) The course instructor realizes that he forgot to incorporate the promised 5% class participation into the course average. Since he also neglected to keep track of participation, he decides to give everyone 100s on this part of the course average, and to reweight the rest of the components to now make up 95% of the course average.
  - i. Determine the new course average for each student.
  - ii. If a student's letter grade changes, the instructor will need to submit a special grade change form. Determine the number of students whose course grade changes, assuming the grade scale given in Homework 1 shown below.

Course Average $x$	$90 \leq x \leq 100$	$80 \leq x < 90$	$70 \leq x < 80$	$60 \leq x < 70$	$x < 60$
Letter Grade	A	B	C	D	F

2. Write a function `my.summary` that takes a vector as input, and returns a vector containing the minimum value, median, and maximum value.
3. Write a function `my.even` that takes a real number as input and produces the nearest even integer as output. Test your function on inputs 2.8, 1.3, and -5.2.
4. Write a function `my.hist1` that takes an integer as input, generates a random list of  $n$  numbers on the interval  $[0, 10]$ , and produces a histogram of these numbers.
5. Write a function `my.hist2` that takes two integers  $m$  and  $n$  as input. It should produce  $m$  independent random sequences of  $n$  numbers on the interval  $[0, 10]$ , and find the maximum value for each sequence. It should then produce a histogram of the  $m$  maximum values.

6. Write a function `my.centers` that takes a vector as input, and returns each of the following:

- (a) The median.
- (b) The mode, but only if there is at least one repeated value.
- (c) The mean, but only if all values are within three standard deviations of the mean.

A vector of the appropriate values should be returned.