

MATH 324 COMPUTER HOMEWORK ASSIGNMENT 1

DUE FEBRUARY 26, 2024

Complete this assignment using R Markdown. You should knit it as a pdf or html. Save the file as a pdf and submit it here.

Exercise 1: (15 pts) Batting statistics are kept on all major league baseball players. The number of times at bat, number of hits, doubles, triples, home runs, runs batted in, and batting averages are all recorded. The file `mlbstats.csv` contains the most current information for this season to date. You will be investigating the data for home runs (HR) hit by each player.

- (a) Save file “`mlbstats.csv`” at your computer.
- (b) Set the working directory in Rmarkdown to match the directory for the file.
- (c) Read the file using function `read.csv(“mlbstats.csv” , header=T)`. (Make sure to assign the table to a variable of your choosing).
- (d) Construct a stem-and-leaf display of the home run data using `stem ()`.
- (e) Does the distribution of values appear to be reasonably symmetric? If not, how would you describe the departure from symmetry?
- (f) Would you describe any observation as being far from the rest of the data (an outlier)? If so, which one?
- (g) Calculate the mean and standard deviation using `mean()` and `sd()` respectively.
- (h) Calculate the first and third quartiles using `quantile(..., 0.25)` and `quantile(..., 0.75)` respectively.
- (i) Construct several histograms for this dataset choosing different classes. Vary both the quantity of classes and the limits of your classes. Give your histograms a title, x-label, and try changing the colors of the bars. Make comments about the shapes.

Exercise 2: (15 pts) *Consumer Reports* provided overall customer satisfaction scores for AT&T, Sprint, T-Mobile, and Verizon cell phone services in major metropolitan areas throughout the United States. The rating for each service reflects the overall customer satisfaction considering a variety of factors such as cost, connectivity problems, dropped calls, static interference, and customer support. A satisfaction scale from 0 to 100 was used with 0 indicating completely dissatisfied and 100 indicating completely satisfied. The ratings for the four cell phone services in 20 metropolitan areas are in the file `cellservice.csv`.

- (a) Save file “`cellservice.csv`” at your computer in the same directory used in Exercise 1.
- (b) Make a table by reading the file using the function `read.csv(“cellservice.csv” , header=T)`. Assign this table to a variable name of your choosing.
- (c) Calculate the summary statistics for each of the four companies using `summary()`.
- (d) Make comparative boxplots for the four companies, putting them on the same graph using `boxplot(company 1, company 2, company 3, company 4)`. Add labels for each company on the boxplot. You can do this by adding on “`names=c(“company name 1”, “company name 2”, ...)`” as a modifier for `boxplot`.
- (e) Comment on what the boxplots show about the customer satisfaction for each company. Which company has the least favorable customer satisfaction? Which company do you think Consumer Reports recommended? How did this company compare to the other three?

Exercise 3: (10 pts) A survey was conducted in a class where 30 students are enrolled. The survey question is how many courses is the student currently taking this semester. The response data are shown below.

4	3	3	3	1	5	4	2	2	4
5	6	4	3	3	4	4	5	6	1
2	2	3	4	2	1	5	2	1	3

- (a) Use `c()` to generate the data.
- (b) Construct the frequency table using `table()`.
- (c) Construct a pie chart of your table using `pie()`. Make sure you are making a pie chart of your table, not the original data set. Add colors and a title to the chart.
- (d) Construct a bar-plot of your table using `barplot()`. Add color and a title to the plot.
- (e) Use R functions to count how many students are taking more than three courses.

Exercise 4: (10 pts)

- (a) Use `seq()` to generate the sequence 2, 6, 10, ..., 42.
- (b) Use `log()` to generate a new sequence where each element is log-transformed from the sequence in (a).
- (c) Remove the third to seventh elements in the resulting sequence in (b).
- (d) Use `length()` to obtain the length of the resulting sequence in (c).
- (e) Sort the resulting sequence in (c) from high to low using `sort()`.