

Attach all of your R code at the end of your solution file or in a separate text file.

Problem 1 (16 points, 4 points for each part)

The data file used for this problem is `texas-house.csv`. Here we will consider two qualitative variables: the exterior type (“`exter`”) and the last digit of the zip code (“`zip`”).

- (1) Construct a two-dimensional frequency data with the values of “`exter`” as the row and the values of zip code as the column.
- (2) Construct a percentage table based on the table from (1). The percentage is calculated based on each column.
- (3) Construct a bar plot based on the table from (2). Use appropriate main title and labels for the x-axis and y-axis.
- (4) Comment on features of plot from (3).

Problem 2 (10 points)

Someone wants to know whether the direction of price movements of the general stock market, as measured by the New York Stock Exchange (NYSE) Composite Index, can be predicted by directional price movements of the New York Futures Contract for the next month. Data on these variables have been collected for a 46-day period. The data file is: `nyse.csv`. The variables in this data set are:

`index`: the percentage change in the NYSE composite index for a one-day period.

`future`: the percentage change in the NYSE futures contract for a one-day period.

Construct a scatterplot relating these variables and comments on your findings. You need to decide which variable should be used for the x-axis and which variable should be used for y-axis. Use appropriate main title and labels for the x-axis and y-axis.

Problem 3 (15 points)

The use of placement exams in elementary statistics courses has been a controversial topic in recent times. Some researchers think that the use of a placement exam can help determine whether a student will successfully complete a course (or program). A recent study in a large university resulted in the data listed in the data file `exam.csv`. The placement test administered was an in-house written general mathematics test. The course was Elementary Statistics. The students were

told that the test would not affect their course grade. After the semester was over, students were classified according to their status. The variables are:

score: the students' scores on the placement test (from 0 to 100)

status: the status of the student (coded as 0 = passed the course, 1 = failed the course, and 2 = dropped out before the semester was over)

Construct a boxplot stratified by the status of the students. First discuss features of each boxplot then comment on your findings by comparing those boxplots. Three boxplots should be in one single plot. Use appropriate main title and labels for the x-axis and y-axis.