

TEST 1
CHAPTERS 1-3
MAT 160/BUS 232
KOZAK SPRING 2018

Use R for all calculations and graphs. Show what you typed into R for all problems.

- 1.) (16 pts) The csv file at marijuana<-
read.delim("https://krkozak.github.io/MAT160/marijuana_use.txt") is a dataset
containing the legality of marijuana in each state in the U.S.

Hand in:

- a. A statement the individual, variable, and population.
- b. The clearly labeled pie chart.
- c. A frequency table containing the categories and frequencies.
- d. A concise written summary of the data.

- 2.) (24 pts) The csv file at temperature<-
read.delim("https://krkozak.github.io/MAT160/Normtemp.txt") contains
temperature data for males and for females. The units are degrees F.

Hand in:

- a. A statement of the individuals, variables, populations, and samples. (Note there are two of each.)
- b. Two clearly labeled histograms, one for each gender.
- c. A frequency distribution containing the classes, frequencies, relative frequencies, and cumulative frequencies, and stating how many classes you choose.
- d. A table summarizing the mean, median, mode, range, standard deviation, variance, and five-number summary for the temperature for each gender.
- e. Two box and whiskers plots.
- f. A written summary of the data sets and a comparison of the two data sets.

- 3.) (16 pts) Construct a stem and leaf plot of the sulfur dioxide (SO₂) data in the csv
sulfur<-read.delim("https://krkozak.github.io/MAT160/Pollution.txt") that is the
amount of SO₂ in chosen cities.

Hand in:

- a. A statement the individual, variable, population, and sample.
- b. The clearly labeled stem and leaf plot.
- c. A table summarizing the descriptive statistics of mean, median, mode, range, standard deviation, variance, and five-number summary for the sulfur dioxide levels.
- d. A written summary of the data.

- 4.) (12 pts) Construct a time series graph for the closing stock prices for Amazon (AMZN) and one for Coke-a-Cola (KO) stock. The data is found in prices<-
read.delim("https://krkozak.github.io/MAT160/AMZN-KO.txt"). The variable
time are the days since the first date in the variable date. The units are \$.

Hand in:

- a. A statement of the individuals, variables, populations, and samples. (Note there are two of each.)
- b. The time series graph of each data set.
- c. A written summary of each data set and a comparison of the data sets.

5.) (12 pts) The csv file degrees<-

`read.delim("https://krkozak.github.io/MAT160/Arcades_vs_doctorates.txt")`
contains the dataset of total revenue generated by arcades and the number of computer science doctorates. Create a scatter plot of the data.

Hand in:

- a. The scatter plot.
- b. Description of the relationship.
- c. An interpretation of the statement that more money from arcades causes more computer science doctorates.

6.) (20 pts) Using the data set on unemployment rates for countries around the world (World_Bank_data.xls, in Test 1 Assignment on Canvas) (the dataset was collected from the World Bank), perform the following.

- a. statement of the individual, variable, population, and sample.
- b. Select a simple random sample of 10 unemployment rates. Include a description of the process you went through to actually collect the random sample. Provide the sample you chose.
- c. Calculate of the descriptive statistics, mean, median, range, variance, and standard deviation for this sample using R.
- d. Select a systematic sample of 10 unemployment rates. Include a description of the process you went through to actually collect the sample. Provide the sample you chose.
- e. Calculate the descriptive statistics, mean, median, range, variance, and standard deviation for this sample using R.
- f. The population mean is 8.77728 and the population standard deviation is 5.610733. Compare the two means and standard deviations that you calculated from your samples to the population means and standard deviation. Which sample gives a more accurate answer? Why?

Note: You will need to type the data into R in this problem. To do this you need to type in the name of your variable that you pick. and then `<-c(data separated by commas)`. As an example, to type in the data 3, 4, 5, 6, 7, 8, and call it test, type in `test<-c(3, 4, 5, 6, 7, 8)`