

Homework 1

ME EN 2550

Ryan Dalby u0848407

1.4: a) Cases: 600 asthma patients aged 18-69 who relied on medication for asthma treatment.

b) Variables: If Buteyko method was practiced- Categorical; Quality of life, activity, asthma symptoms, medication reduction- all are numerical discrete; Age- numerical continuous

c) Research Question: "Does the Buteyko Method reduce asthma symptoms in patients aged 18-69 who rely on asthma medication."

1.8: a) A single UK resident who was sampled.

b) 1691 participants.

c) sex- categorical; age- numerical discrete; martial- categorical; grossIncome- categorical ordinal; smoke- categorical; amtWeekends- numerical discrete; amtWeekdays- numerical discrete

1.11: a) Population of interest- asthma patients 18-69 relying on medication. Sample- 600 asthma patients 18-69 relying on medication.

b) This study does not indicate whether the 600 people were sampled randomly from the population thus it is not generalizable to the population. Since treatments were randomly assigned we can draw causal relationships.

1.16: a) Explanatory: Percent with Bachelor's degree; Response Variable: Per capita income

b) There seems to be a positive linear correlation between the two variables above.

c) No, because we do not know about how the data was gotten, and it was likely not from an experiment thus we can only conclude association.

1.19: a) There is a positive association between life expectancy and percentage of internet users.

b) This study is an observational study because it examines data that was not acquired from an experiment.

c) A possible confounding variable would be the human development index ranking of the country.

1.25: a) Non-response bias is a flaw in the sampling method that will likely not allow the school to draw said conclusion about the population.

b) Non-response bias is a flaw in the sampling method of the following up survey that will not allow generalization to the population since those who responded own homes and do not represent all the population.

c) The patients do not represent the whole population and the orthopedist did not conduct an experiment and thus cannot draw causation.

1.28: a) No we cannot conclude that smoking causes dementia because this study was an observational study and the people selected were not randomly selected from the total population.

b) No, because the study conducted was an observational study, not an experiment thus we cannot draw the conclusion that sleep disorders cause bullying in school children.

1.33: a) Exam performance is the response variable.

b) Light- fluorescent, yellow, desk lamp. Noise: no noise, construction noise, human chatter noise.

c) Sex is a blocking variable in this study.

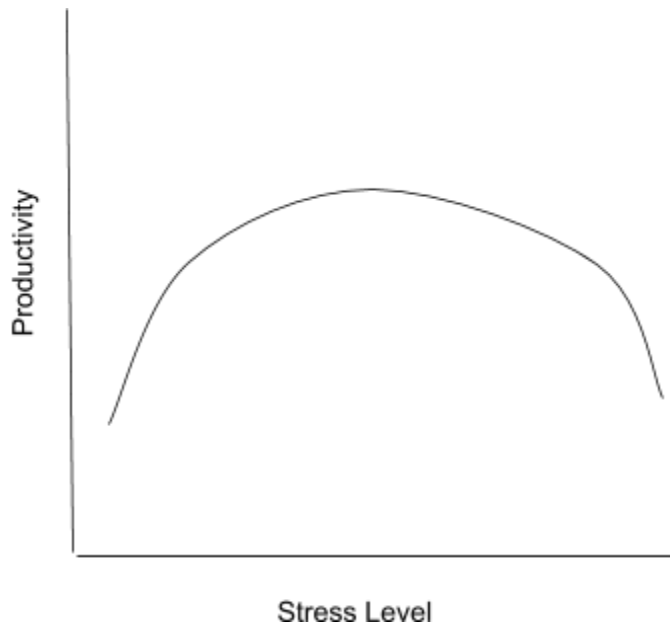
1.37: a) experimental study

b) experimental treatment- 25 grams of chia seeds twice a day; control- placebo

c) Yes blocking was used since men and women were blocked and assigned to the group.

d) No, and yes, blinding was not used for the researchers but it was for the subjects who didn't know what treatment they got.

e) We can make a causal statement since an experiment was conducted with random treatment assignment. We cannot generalize to the population since it was a voluntary study.



1.40:

1.47:

- a) The mean of 2 is higher than the mean of 1. The standard deviation of 2 is bigger than 1.
- b) The mean of 1 is higher than the mean of 2. The standard deviation of 2 is bigger than the 1.
- c) The mean of 2 is higher than the mean of 1. The standard deviation of 1 and 2 are the same.
- d) The mean of 1 and 2 are the same. The standard deviation of 2 is bigger than 1.

1.54: a) The two frequency peaks are more obvious in the histogram. Outliers and skew are more obvious on the box plot.

b) The binomial distribution is likely because the sample includes both men and women, one peak is likely men and the other is likely women. Each group likely has a different mean finish time and distribution.

c) The women's distribution is more skewed than the men's distribution with more outliers. The men's distribution has a much smaller IQR but there still is some skewness.

d) It appears that there was some type of decrease around 1975 in marathon times and then relative plateauing in marathon times.

1.64: a) There are many counties in the US that have a small percentage of Hispanics and there is a very large spread of a smaller number of counties that have a much higher percentage of Hispanics. Using a log-transformed histogram allows us to better look at the distribution of Hispanics in the few higher percentage counties. It gives a relatively normal distribution.

b) From the map, the location of the counties with high percentages of Hispanics is much more clear. From the histogram, we can see more clearly the distribution of high Hispanic percentage counties.

c) It depends on what is trying to be presented. The map is very easy and quick to notice the location and the relative number of counties with high Hispanic percentage populations. The histogram does give more data on the distribution of these high percentage counties.

1.66: a) 40.89%

b) 30.55%

c) 6.26%

d) 15.32% conservatives are in favor of citizenship. 33.06% moderates are in favor of citizenship. 57.71% liberals are in favor of citizenship.

e) No, political ideology and view on immigration do appear to have some type of relationship as we see that a much higher percentage of liberals are in favor of citizenship than conservatives.

Thus political ideology and view on immigration do not appear to be independent.

1.70: a) It does not appear that survival is independent of whether or not the patient got a transplant. There appears to be some type of dependence as the distributions are very different.

b) The box plots suggest that those who got treatment survived much longer than those who did not get treatment.

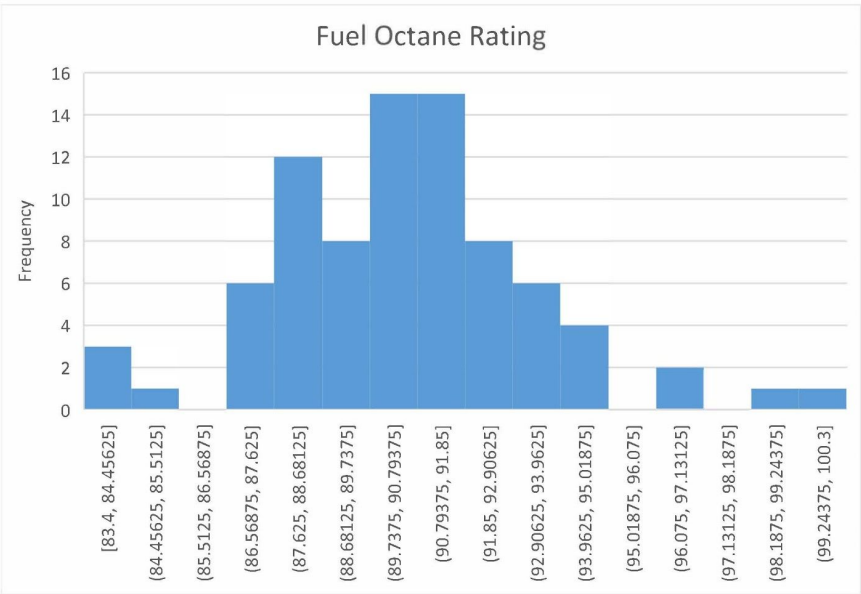
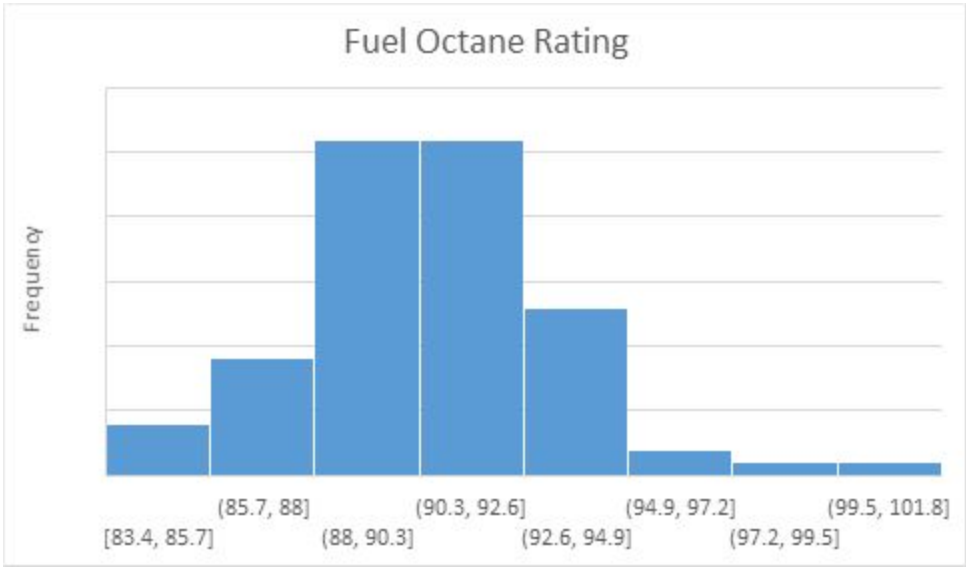
c) Treatment group- 65.22% of people died. Control group- 88.24% of people died.

d) i. Claims being tested: whether an experimental heart transplant program increased lifespan

ii. Fill in blanks(answers separated by commas): 28, 75, 69, 34, 0, lower than -23.02%(The difference between proportion dead in the treatment and control groups of the actual study).

iii. There is some effectiveness of the transplant program since we see a small fraction of the simulated differences in proportions that are smaller than -23.02%.

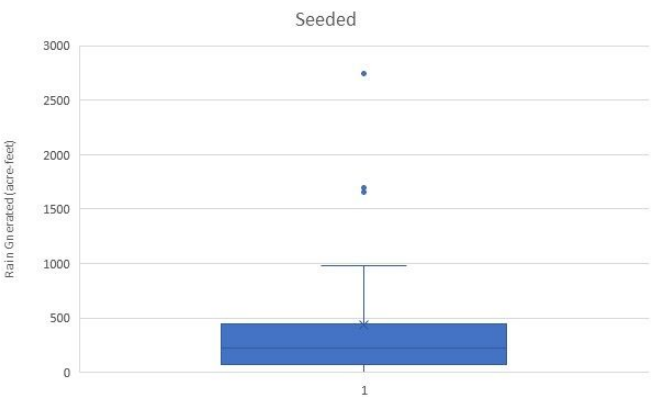
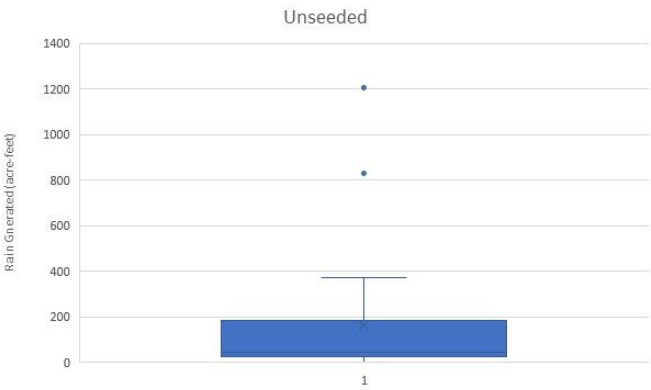
B.1:



B.2:

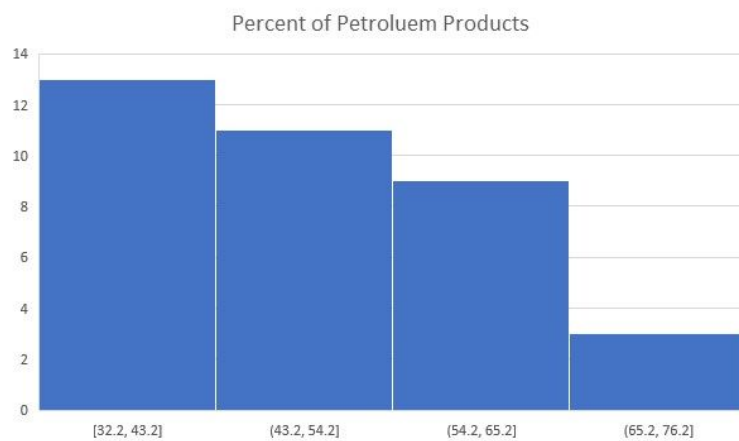
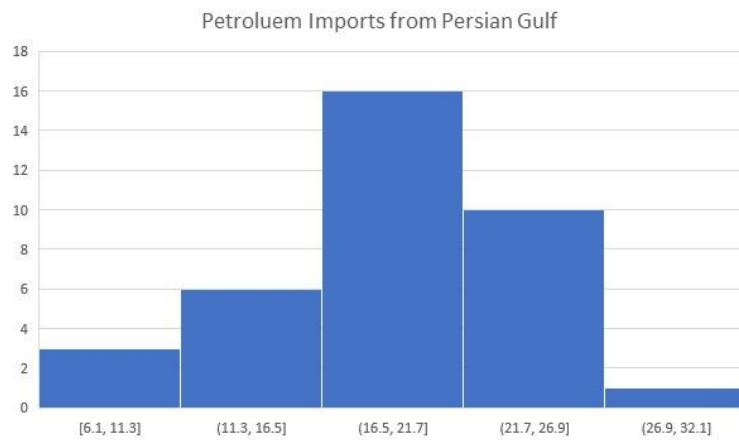
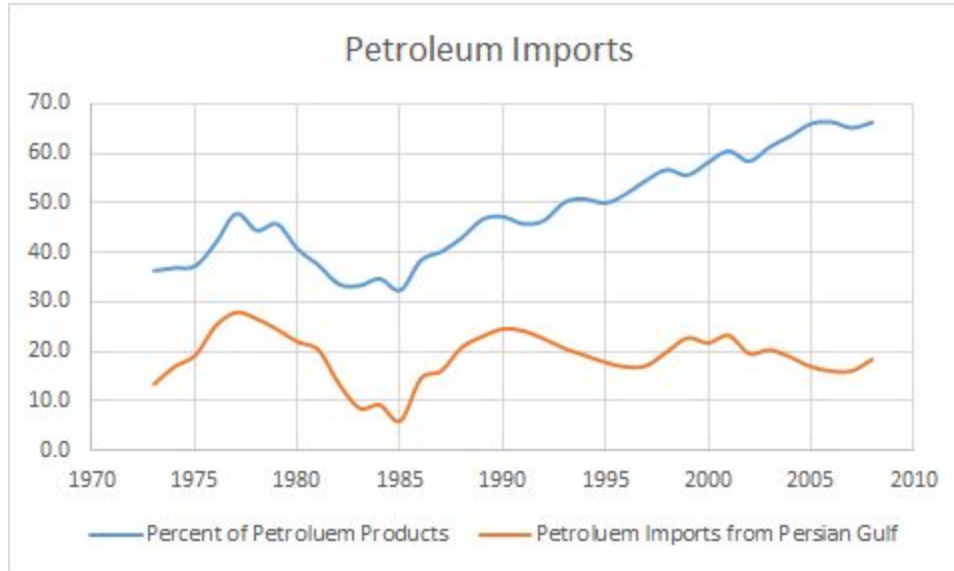
a)

	Unseeded	Seeded
Mean	164.5885	441.9846
Median	44.2	221.6
Standard Deviation	278.4264	650.7872
Range	1201.6	2741.5



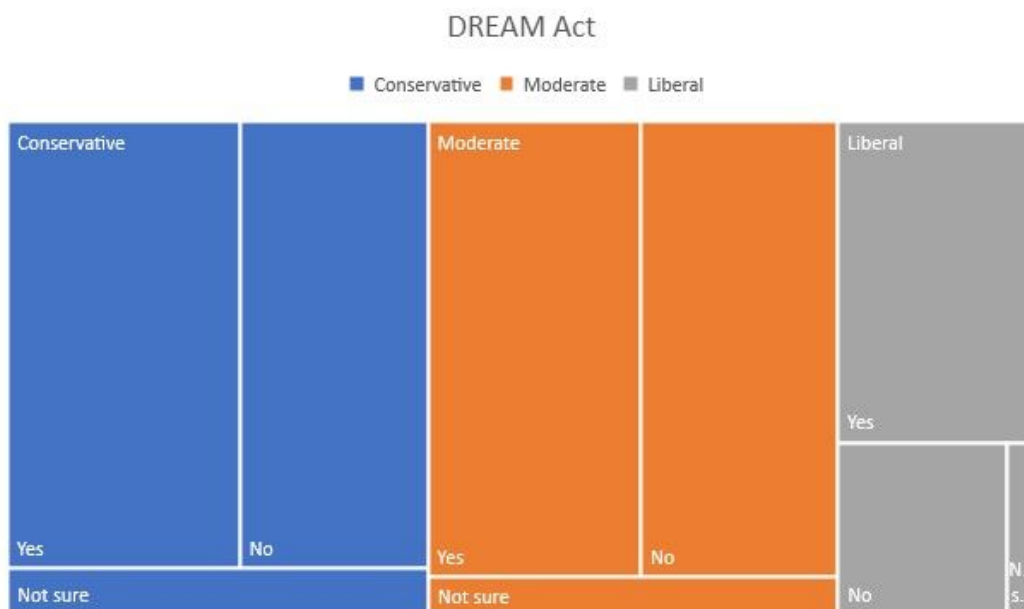
b)

B.3: a)



b) In the time plot, we can see when percent petroleum products and petroleum imports from the Persian Gulf are changing, it gives us a reference of time and from that, we can discern some relationship between the two. In the histograms, we can see the frequency/amount of certain values which tells us how these values are distributed independent of time which can allow us to notice relationships between percent petroleum products and petroleum imports from the Persian Gulf.

B.4: a)



b) The difference between the plots is that the one we created has possible answers of “Yes”, “No” and “Not sure” while the plot in the book has “Support” and “Not Support” and “Not sure”. The tree plot is also formatted slightly differently. From the plot, you can see some dependence between ideology and support of the DREAM act. This is evidenced by the different distribution of responses for liberals compared to moderates and conservatives.