**Name:** _____

**MAT 121**
**Summer 2019**
**Homework 3**

*"Leon, of course it's wrong! The odds against you being right
are... staggering! You have a great advantage, you know the outcome.
You will be wrong. Don't fear it; embrace your wrongness!"*
                              *– Dick Solomon, 3rd Rock from the Sun*

**Problem 1:** Construct a dataset with 6 different positive numbers such that the mean of the dataset is less than the median of the dataset. Then construct a dataset with 6 different positive numbers such that the mean of the dataset is greater than the median of the dataset.

**Problem 2:** Explain what is wrong with the following statement using course concepts: schools with vending machines, especially ones with fatty foods, tend to have higher rates of student obesity. To prevent student obesity, these vending machines should be removed from the schools.

**Problem 3:** Which measures of center are resistant? Which measures of center are not?

**Problem 4:** Explain what a $p$-value is.

**Problem 5:** The total number of ski rentals at a local ski resort $x$ number of weeks after the start of the season appears to follow a linear pattern. The owner creates a linear regression for the data and finds $\hat{y} = 127.1x + 49.3$ with $r = 0.914$. According to the model, how many ski rentals appear to be made each week during the season? What percent of the variability in the data seems to be explained by the model?

**Problem 6:** During their first semester of college, a student receives the following course grades. What is the students GPA? Be sure to show all your computation.

| Course | Credits | Letter Grade |
|---|---|---|
| Vector Calculus | 4 | A |
| Quantum Mechanics I | 3 | A− |
| Women in Music | 3 | B |
| Tree Climbing | 1 | A |
| Marketing & Media | 3 | B+ |
| Russian I | 4 | C− |

**Problem 7:** Consider the following dataset:

$$14, \quad 15, \quad 16, \quad 17, \quad 23, \quad 26, \quad 29, \quad 33, \quad 46, \quad 67, \quad 71$$

What percentile does the data values 17 represent? What is the value of $P_{75}$ for this dataset?

**Problem 8:** Suppose the scores of an exam are normally distributed with mean 81 and standard deviation 3. What proportion of students received between a 78 and a 90? Suppose John received a 67 on this exam. Susan took a similar exam, which also had scores that were normally distributed with mean 68 and standard deviation 6. Susan received a 45. Relative to their exams, who did worse? Explain.

**Problem 9:** In his 1951 paper "The Interpretation of Interaction in Contingency Tables", Edward Simpson described what is now known as Simpson's Paradox. Simpson's Paradox, in essence, is when a trend appears in groups of data but when the data is combined, the trend reverses. Consider the following graph of income and 'happiness.'
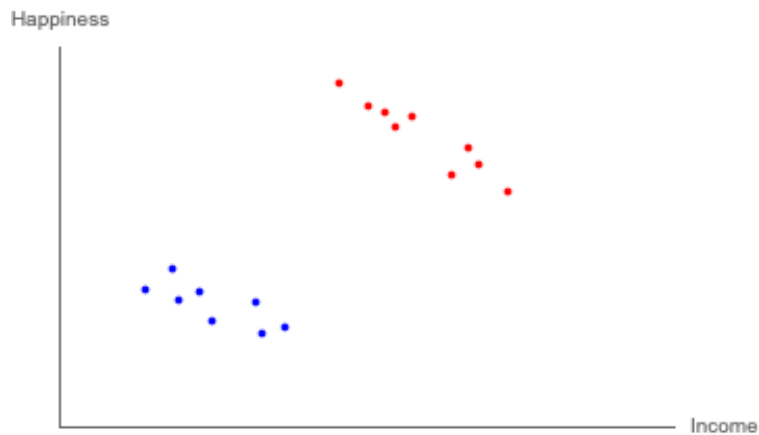


The red dots (the upper collection of dots) were females surveyed while the blue dots (the lower collection of dots) were males surveyed. What trend between income and happiness does there seem to be for females and males, respectively and individually? What does the overall trend between income and happiness appear to be for all people? How does this serve as an example of Simpson's paradox?

**Problem 10:** Continuing the discussion of Simpson's paradox begun in Problem 9, consider the following scenario: a company gives out bonuses to the best sales person. Two people seem to be in contention for the top sales person: Krystina and Ashton. They sell two sizes of cars. Their sales are summarized below:

| Salesperson | Small Cars | Large Cars | Total Cars |
|---|---|---|---|
| Ashton | 90 out of 100 | 16 out of 20 | 106 out of 120 |
| Krystina | 20 out of 20 | 85 out of 100 | 105 out of 120 |

Who had the best overall sales rate? Who had best overall sales rate in each of the categories 'small' and 'large'? Why is this an example of Simpson's paradox?