DMC 106: Homework 1

Due Date: Wednesday, August 27th

Problem 1 (10 Points): Detail the type of variable (*e.g.,* continuous numerical variable) for the following variables in the loan50 dataset (the dataset can be found in Chapter 1 of your textbook).

(a) Population Change

(b) Name

(c) Population

(d) Homeownership

(e) Metro

Problem 2 (10 Points): Researchers studying the effect of antibiotic treatment for acute sinusitis compared to symptomatic treatments randomly assigned 166 adults diagnosed with acute sinusitis to one of two groups: treatment or control. Study participants received either a 10-day course of amoxicillin (an antibiotic) or a placebo similar in appearance and taste. The placebo consisted of symptomatic treatments such as acetaminophen, nasal decongestants, etc. At the end of the 10-day period, patients were asked if they experienced improvement in symptoms. The distribution of responses is summarized as:

		Self-reported improvement in symptoms		
		Yes	No	Total
Group	Treatment	66	19	85
	Control	65	16	81
	Total	131	35	166

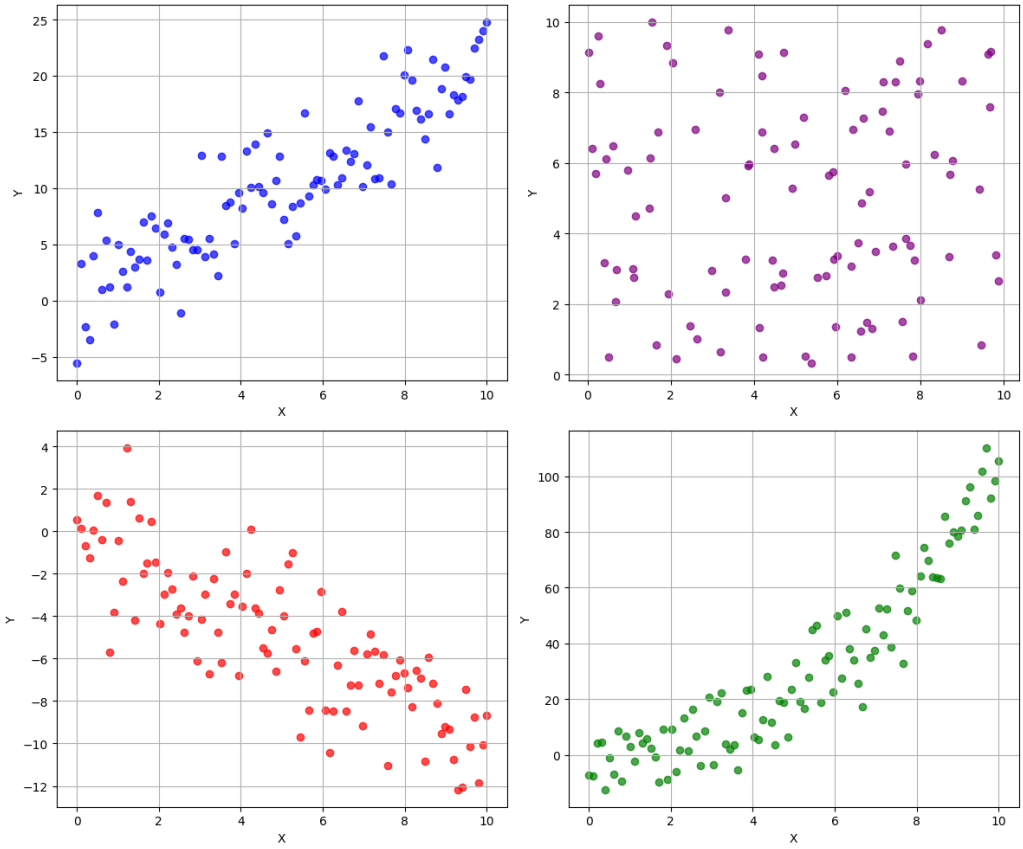
(a) What percent of patients in the treatment group had improvement in symptoms?

(b) What percent experienced improvement in symptoms in the control group?

(c) Which group experienced a greater percentage improvement in symptoms?

(d) Your findings so far might suggest a real difference in effectiveness of antibiotic and placebo treatments. However, this is not the only possible conclusion that can be drawn. What is one other possible explanation for the observed difference between the percentages of patients in the antibiotic and placebo treatment groups that experience improvement in symptoms of sinusitis?

Problem 3 (10 Points): For each of the following plots, determine if the variables are related and, if so, detail in what manner.



Problem 4 (10 Points): For the following questions, what are the explanatory and response variables?

(a) Does a person's height affect their speed?

(b) Do families with more children spend more time playing outdoors?

Now, do the same as above, but determine a possible confounding variable for the question.

(a) Is there a relationship between the number of firefighters at a fire and the amount of property damage caused by the fire?

(b) Is there a link between internet usage and life expectancy?

Problem 5 (15 Points): Create your own fake dataset, detailing what are the (a) data samples, (b) variables, and (c) the type of variable (*e.g.,* continuous numerical). Draw the data matrix along with the explanations for (a), (b), and (c).

Problem 6 (10 Points): A university wants to determine what fraction of its undergraduate student body support a new $25 annual fee to improve the student union. For each proposed sampling method below, describe why the method is reasonable or not.

(a) Survey a simple random sample of 500 students.

(b) Stratify students by their field of study, then sample 10% from each stratum.

(c) Cluster students by their ages (*e.g.* 18 years olds in one cluster, etc.), then randomly sample three clusters and survey all students in those clusters.

Problem 7 (30 Points): There are 20 students in a statistics class. On their first exam, the students achieved the following scores out of one hundred: 85, 92, 78, 88, 95, 80, 75, 90, 82, 98, 65, 99, 74, 55, 70, 82, 80, 91, 100, 76.

(a) Draw a data matrix where the two variables are the scores in both raw and percentage formats and the data samples are the students.

(b) Compute the mean score (show your work!).

(c) Compute the variance and standard deviation of student scores (show your work!).

(d) What is the median score?

(e) If a new student takes the exam and scores a 78, would this:

- Increase or decrease the mean? Why?
- Increase or decrease the standard deviation? Why?
- Increase or decrease the median? Why?

(f) Draw a histogram of the student scores, where the bin size is 10 (*i.e.,* scores will fall between 0-10, 10-20, etc.).

Problem 8 (Fermi, 5 Points): How many total cups of coffee are drank per year by people in Baltimore City? Be sure to show the steps you take towards your solution!