**Directions**

In your group you are expected to collaborate on addressing each question. If you are the first person to post your ideas, do not fear! This is intended to provide you with a safe space to put your ideas on paper.

If you come to the document and there are ideas that you disagree with or would like to revise / clarify, feel free to use comments to discuss with your group members.

My hope is that each group member contributes equally, but that may look different for each person. I expect for each of you to edit and add to posts from the rest of your group. This **is not** an individual assignment, so please don’t write individual responses and fail to engage with your group.

Your final responses are due on Friday, January 8 by 11:59pm. You will be required to download this Google Doc as a Word file to submit in Canvas. Only one submission per group is required.

**Question(s)**

1. What do you believe are similarities and differences between Statistics and Mathematics?

|  |  |
| --- | --- |
| Differences | Similarities |
| * Statistics tends to look at groups of data while math often looks at smaller amounts of data or just a variable * Stats is an application of math * Math is more universal across other fields * Deals with uncertainty and predictions, while mathematics tends to seek concrete solutions * Stats is highly centered around hypotheses, while mathematics is less so * Statistics tends to look at groups in a population and draw real-world conclusions, mathematics is mostly numerically based without room for multifactor analysis * Mathematics is used to solve for a concrete answer, whereas statistics is more open-ended due to its focus on implication and population assumption | * Both use numbers * Analyze numbers * Trend lines used to demonstrate relationships * Both require some calculation * Trends can be deducted |

1. From your perspective, what is Statistics? Is there a difference between “statistics” and “Statistics”? If so, what?

Statistics uses graphs, numbers, and charts that lead to conclusions about things happening all around us. It follows trends and can help us keep track and theorize life. To go a little deeper, Statistics is also about understanding specific data/datasets, analysis, acknowledging randomness within conclusions, and the ability to use math and theory as an explanation within research. One possible difference between “statistics” and “Statistics” could be the attitude surrounding it. Statistics as an area of expertise seems to hold more weight than statistics. Another difference between statistics and Statistics could be that Statistics is the overall term used to describe this field, whereas statistics are the individual measures that make up the data/conclusions. Therefore, Statistics is the collection of statistics.

1. Just because a variable has numeric values, does not mean it is a numeric variable. What are the different types of variables that can appear in a dataset? How can you determine if a variable is numerical versus categorical?

The difference between a numerical and categorical variable is numerical variables would make sense to add, subtract, or apply another mathematical operation on. Categorical variables use numbers as placeholders to identify different categories. There are two types of numerical variables- continuous and discrete. Discrete variables have gaps between the numbers and don’t include a full set of numbers (not always decimals or negative numbers for example). Continuous variables could essentially be any number within reason of the situation.

1. What are the differences between observational studies and experiments?

Observational studies are designed to observe actions and outcomes of participants as they unfold. In observational studies, there is no influence by the observer. Experimental studies have influence from the observer/researcher who changes variables and records outcomes for different people and groups to understand if there is a causal relationship between variables. Since the experimenter can influence those participating, experimental studies assign a treatment group and a control group. The treatment group will be administered a treatment as the independent variable, or explanatory variable, while the control group will either be given a placebo or nothing at all. This way, the researcher can determine if there is or is not a causal effect and see if the explanatory variable might affect the response variable.

1. What are different methods for sampling from a population? Why would a researcher choose a stratified random sample instead of a simple random sample?

There are four different techniques for sampling from a population and those are 1) simple random, 2) stratified random, 3) cluster, and 4) multistage sampling. A simple random sample is used to make chances fair of being selected in a population, and it helps reduce possible bias. Stratified random sampling is used to collect a representative sample from the population. Cluster sampling is when a sample uses different clusters that are formed within a population. Multistage sampling is when certain subgroups in a cluster are used as representation for the sample. A stratified random sample may be used in place of a simple random sample as a way to “divide and conquer” but grouping a population into strata based on similarities and then choosing a random sample from each strata. This can be used as a way to adjust for differences that can be accounted for within each strata before randomly sampling.

1. Describe what you see in the plot below.

Similar to exercise 10 in section 1.2.6 of the book, these plots show voting patterns (depicted by the y-axis) in favor of certain issues (titles of each plot). Each observation point that is plotted is a pair of a country (there are two) and year. The year would be an ordinal categorical variable while the topic of the vote (i.e., colonialism, economic development, human rights, etc.) would be regular categorical variables. This plot shows comparisons of two subjects over the same time period (independent variable) versus differing voting patterns and the country as dependent variables. A least squares regression line is fitted to the data for both subjects.

