**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 |
| * The goal of mathematical operations is to find a single solution to a problem while the goal of a Statistical analysis is to discover evidence for a conclusion/answer to a question * I’ve always thought of Statistics as more a science (a study of data) rather than a form of Mathematics. * With mathematics, there is usually one correct answer, but with statistics, there can be multiple interpretations of the identified data * The goal of mathematics is to use numerical data and find the only valid quantitative answer. Statistics can analyze quantitative data and come up with qualitative results. * Statistics has a huge focus on the context of the problem, why we are trying to solve the calculations and what it means. With mathematics, this is not always the case. | * Both involve calculations and numbers * Make use of data sets and methods of modeling data such as graphs and tables * Both involve calculations and numbers that can determine a specific outcome, whether that be for a statistical study for a health problem or a calculation for a new building project * Each field can be interchangeable (i.e., there are statistics within math and math within statistics) |

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

From my perspective, Statistics is the science of analyzing and studying data in order to answer questions, determine significance, find evidence, and discover relationships between data using tests and other methods of study. There is a definite difference between “statistics” and “Statistics”. “Statistics” is a science that consists of utilizing many different methods and tests in order to interpret and analyze sets of data. “Statistics” can be used to analyze or compute “statistics”, which are singular pieces of data or quantities that represent or tell us something about a set of data such as an average.

“Statistics” can also refer to the all encompassing field, while “statistics” represents the different subfields. Examples of subfields would be biostatistics, epidemiology, financial analysts, etc. All these smaller fields of statistics make up the whole field 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?

Out of all variables, there are two respective types of variables, numerical and categorical. Then, all numerical variables are either continuous or discrete; all categorical variables are either (nominal) regular categorical or ordinal.

Numerical variables are numerical values, where you can add, subtract, or take the mean value, such as population or height. From numerical variables, discrete variables take numerical jumps, for instance, one cannot have a population of 500.2 or 638.77, it must be an integer. However, continuous variables do not follow this pattern, meaning it can describe a person’s height being 5.57 feet or 6.62 feet.

Categorical variables are represented as categories, not numbers, and there are different levels the variables can represent. An ordinal variable is a variable that has levels that already have a natural ordering system. For example, if experimenting with high schoolers, the variable for class standing would be Freshman, Sophomore, Junior, Senior, which are ordinal variables because they are categorical with an ordering system. On the other hand, nominal (regular categorical) variables are variables with no type of special ordering. For example, the U.S. state of residence is a nominal variable.

When determining whether a variable is numerical versus categorical, it is not a guarantee that all numbers are numerical variables. Numerical variables will always be numbers, but numbers will not always be numerical variables. If the number cannot be used for arithmetic, like addition or subtraction, it is not a numerical variable. For example, phone numbers are not numerical variables, although they contain numbers. So to determine whether the variable is numerical or categorical, try to rule out whether it is a numerical variable or not. It is good to keep in mind that categorical variables tend to be able to be grouped or categorized.

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

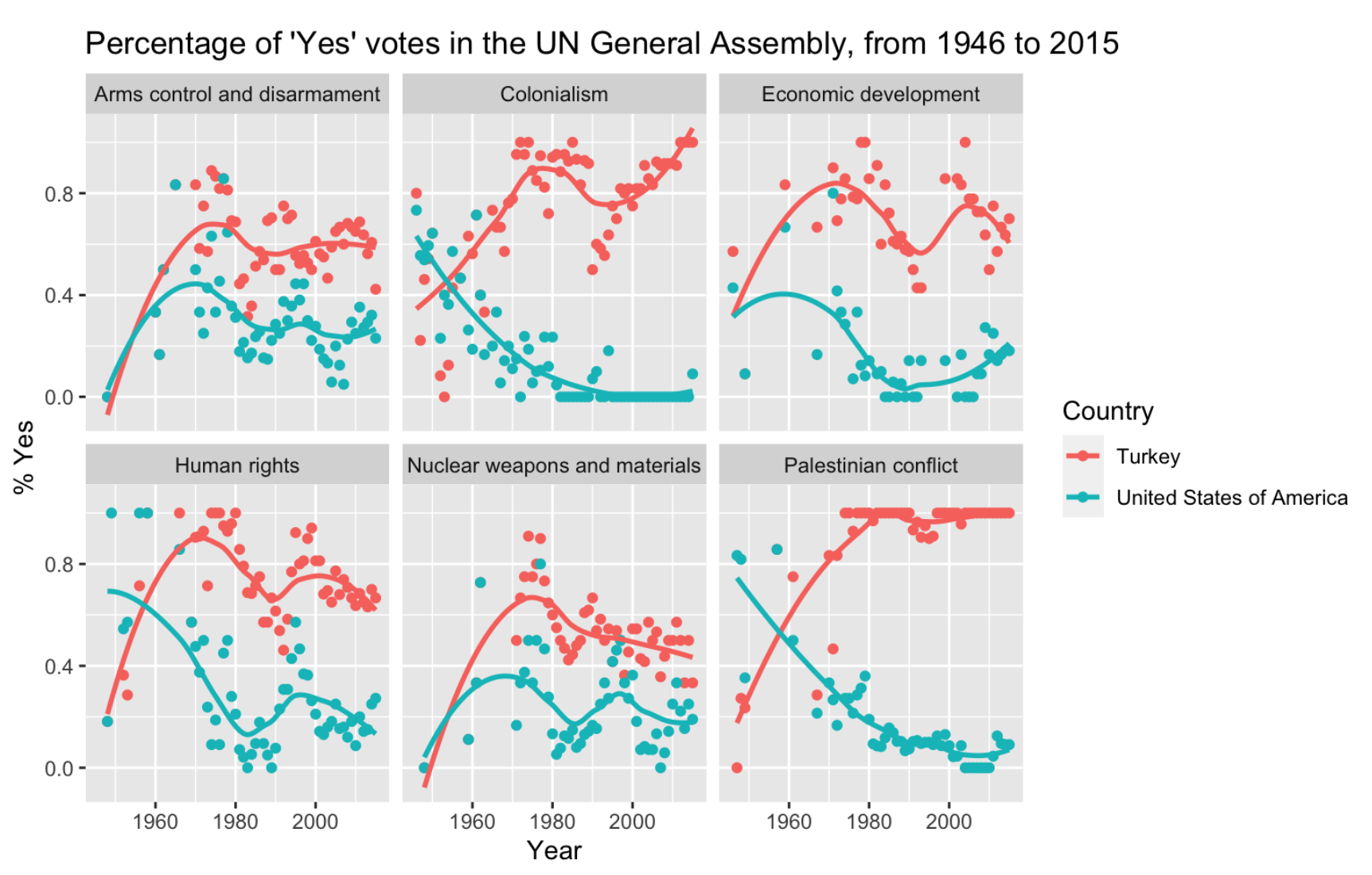
An observational study is when no treatment was applied. The people who are conducting the study want to show evidence of how something naturally occurs between two variables, without any sort of manipulation. An example of an observational study would be the Census. The Census Bureau sends out a survey that people fill out and submit. Information from this data can then be used in observational studies to determine if there’s a connection between two variables.

An experiment is when some sort of treatment is applied in a study, and experiments are very useful to show a causal connection between two variables. This would be to prove that there is a connection between two variables, the explanatory and response variables. An example of this could be an experiment where there’s the treatment group and a control group. The treatment group receives a new heart medication that will help people feel better, while the control group receives a placebo treatment. This study is supposed to prove the new medication is the direct cause of people feeling better, while controlling for all other possible variables.

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?

In order to account for the diversity in a whole population without the ability to take observations from every member of said population, researchers will utilize different methods of sampling such as simple random sampling, stratified sampling, cluster sampling, and multistage sampling. These methods allow researchers to represent the whole population through a small subset of the population. A simple random sample is a small sample where every individual has an equal chance of being selected for the sample. Randomizing the sample helps remove bias that would otherwise occur if researchers chose the participants based on convenience or for some other reason that could skew the results. An example of a simple random sample would be a class of 100 students where every student’s name was put into a bucket and 10 names were chosen out of the bucket. In this scenario, every student has an equal chance of being selected, 1/100 or 0.01.

In a stratified sample, researchers first group participants into “strata”, groups based on some common attribute shared by the participants, before taking a simple random sample from each strata. Researchers may prefer to use this sampling method when there is a lot of diversity in the population, allowing for the diversity to be equally represented in the sample. As an example, researchers may choose to use a stratified sample over a simple random sample if they are studying participants in a geographical location and want equal representation over the whole area. They may stratify participants by zip code first and then take a simple random sample from each zip code to obtain equal representation of the area. Another sampling method is cluster sampling, wherein researchers will group a population into clusters where only a few clusters will be used in the study but all observations in the chosen clusters will be studied. An offset of cluster sampling is multistage sampling, where a population would also be grouped into clusters but only a few random observations from each cluster would be studied.

1. Describe what you see in the plot below. 

Overall

* These graphs are scatter plot graphs with trend lines representing the mean percentages of “Yes” votes of Turkey compared to the United States of America in the Un General Assembly from 1964 to 2015.
* Each graph represents the percentage of “Yes” votes in the UN General Assembly with respect to the topic, and Turkey showcases a higher average percentage of “Yes” votes in each plot as the years go from 1946 to 2015

Arms Control and Disarmament

* Turkey appears to have a constant 50% yes vote going back from 2000 to 2015.
* The USA appears to have a lower rate of yes vote, but with uptake from 20% to about 35%, between 2010 to 2015.

Colonialism

* Turkey’s colonialism yes vote has been on an increasing trend, from 80% to about 95%, between 2000 to 2015.
* The USA appears to have a drastic decrease in the percentage of “Yes” votes, from about 60% in 1946 to nearly 0% between 1990 to 2015.

Economic Development

* The cycles of high and low economic development for Turkey seems to work on a faster timing when compared to the United States’.
* The US looks to be entering a new increase in economic development just as Turkey seems to be entering a falling phase.

Human Rights

* Something must have happened in the late 90’s that encouraged more voters to embrace better human rights.
* The late 50’s offered Turkey with more votes for human rights and the exact opposite for the US.
* Since the 1970’s the trend lines have had very similar movements except with the US trend line consistently sitting much further below.

Nuclear Weapons

* Both are similar shaped except for the small dip in the US around the 90’s. I think this could possibly be due to the Anti-nuclear weapons protests going on in response to the wars and our bomb drop on Japan.

Palestinian Conflict

* Turkey appears to have a near 100% “Yes” vote going between 2000 to 2015.
* The USA appears to have the opposite graph, with a steady decrease between 1960 and 2000
* There was some point in the late 50’s that turned the tables for both Turkey and the US. The US then took a hard “no” stance and Turkey took the opposite “yes” stance.
* I think this opposition as well as the one shown in Colonialism is due to the fact that the US hasn’t experienced a war on their turf in a very long time and tend to be the ones colonizing other countries. Turkey has seen conflict and colonization and therefore could be more willing to support the fight against colonization and conflict.