ACTIVITY 1: Introduction to Least Squares Regression

OVERVIEW: In the first activity, the class will begin with a discussion on least squares regression, aiming to understand its purpose. Students will understand that just because two variables are related does not mean changes in one variable will directly lead to changes in the other variable. Hence correlation does not imply causation. Students will differentiate between independent and dependent variables through interactive questioning. Given a list of example variables, students select potential covariational relationships where the independent variable will potentially cause a change in the dependent variable. Students will use their critical thinking skills to select appropriate variables and weed out inappropriate variables to potentially establish causation. Lastly, students will choose one of the relationships they want to explore further and collect their specific data from the website (gapminder.org) for a random sample of 25 countries that will be provided. They will also create the first slide of their PowerPoint with a description of each of their variables and an explanation as to how and why the independent variable will likely have an impact on the dependent variable.

MATERIALS NEEDED TO CARRY OUT THE LESSON:
- In-Class Activity Worksheet
- Laptop with PowerPoint
- Access to the internet

CONCEPTS TO BE LEARNED/APPLIED:
- Students will understand what least squares regression is, its purpose, and how to select the independent and dependent variables when trying to establish causation.
    - Students will understand that just because two variables are related does not mean that one variable causes a change in the other variable.
    - Students will understand the purpose of least squares regression is to find a linear equation that models the relationship between two quantitative variables.
    - Students will understand the difference between the independent and dependent variables, specifically, that the variables are chosen in such a way that the independent variable causes a change in the dependent variable.

INSTRUCTIONAL PLAN:
Explain to students that statisticians are interested in examining how one quantitative variable impacts another quantitative variable. In this activity, the focus will be on exploring linear relationships. Two variables can be merely correlated meaning as one variable changes, the other changes in sync. Just because two variables are related does not mean changes in one variable will directly lead to changes in the other variable. Hence correlation does not imply causation. When causation can be established, it is important to distinguish between the

independent and dependent variables. Data will be collected on countries to investigate whether there is a linear relationship that can be modeled using a least squares regression line.

Tell them that the goal is to determine whether it is appropriate to use the least squares regression to model the relationship between two quantitative variables enabling prediction. Have the students, in groups of 2-3, start by matching 5 of the variables given together in such a way that the x-variable might result in a change in the y-variable.

Once each group has come up with the 5, have the groups share the variables they chose and why they felt the x-variable might result in a change in the y-variable. Write any new relationships on the board.

Solicit student suggestions to come up with definitions for independent and dependent variables.

The following questions can be used to help the students come up with the definitions.
1. What does it mean when you say "A" depends on "B"?
2. From the relationships that you found, which would be the "A" and which would be the "B"?
   Example: How many babies per woman a country has will depend on the prevalence of contraceptive use among other things. This means that "A" would be the babies per woman and "B" would be the "Contraceptive use".
3. Which variables on the board might be considered "Dependent Variables" and which ones "Independent Variables"?

Below are some possible definitions.

Independent variable (x): The independent variable causes changes in the dependent variable.

Dependent variable (y): The dependent variable measures an outcome in a study and is impacted by the independent variable (i.e. it depends on the independent variable).

Once they are finished, each group should choose one of the relationships to collect data for.

Go over how to collect the required data on gapminder.org. The video below goes over how to collect the data. You can play this video for them or just show them how yourself.

VIDEO LINK:

https://www.youtube.com/watch?v=_T4HyOaevrQ&ab_channel=JoanBrenneman

Explain to them that when they look for the data, they need to do the following:

1. They should avoid using totals. For example, selecting "Total C02 emissions" should be avoided. Instead, they can choose "C02 emissions per person". Ask them why a total should

be avoided. (When they choose a total, the population of the country will greatly impact that variable, thus masking the relationship).

2. They should use the most recent data for each country.
3. If a country has no data for their relationship, they should put an "NA" for each variable.

Emphasize to them that they will be using their data and should bring it back to the next class.

At the end of class, they should start a PowerPoint and create the first slide of the PowerPoint. The slide should include a description of the independent and dependent variables chosen and an explanation as to why they think their independent variable will have an impact on the dependent variable.

MIP COMPONENTS OF INQUIRY

<u>Active Learning</u>

- Students select pairs of variables from a list that they believe have a cause-and-effect covariational relationship. The mental activity entailed in forming this relationship involves the student thinking critically about how one variable might impact another variable.
- Students use their pairs of variables and select which variable likely "depends" on the other variable. After students discuss the difference between the independent and dependent variables, they select a definition and evaluate which variable in each pair is independent and which variable is dependent.

<u>Meaningful Applications</u>

- Students justify the possible cause-and-effect covariational relationships they established with the given variables.
- Students generalize their understanding of dependent and independent variables by coming up with the definitions as a class.
- Students generalize their understanding of dependent and independent variables to real-world data.

<u>Academic Success Skills</u>

- Students engage in critical thinking when selecting possible pairs of variables with a cause-and-effect covariational relationship. Using critical thinking skills to identify possible relationships between two entities is a key mathematical practice and can support students in mathematical success.
- As students generalize the concept of dependent and independent variables to real-world data, they are supported in learning the disciplinary practices that help them see themselves as competent doers of mathematics, helping them form an identity as learners.