Modeling with Two-Variable Data

What am I making?
Create a statistical analysis using global data to show an association between two variables. Then create a workshop to teach others how to make data-based decisions and avoid confirmation bias.

Why am I making it?
To develop mathematical, statistical, and critical thinking skills useful in interpreting and correlating data and in making informed, evidence-based decisions.

How do I make it?
Individually select a global dataset from Gapminder.org and analyze the relationship between two variables using statistical methods. With your group, design and present a workshop teaching others to critically evaluate data and avoid confirmation bias.

Driving Question

How can we analyze global data to uncover relationships between variables, challenge biases in data interpretation, and teach others to make informed decisions using statistics?

The driving question identifies the purpose, product, and audience.

Mission

Your mission is to analyze global data from Gapminder.org to find statistical evidence of an association between two variables. Whether exploring the link between study time and grades or happiness and income, your analysis will strengthen your ability to make data-driven decisions.

Beyond analysis, you'll examine confirmation bias—how personal beliefs can distort data interpretation. To combat this, you will organize a training session for students and adults, teaching them to critically evaluate data. As a team, you'll present your findings and lead a workshop to empower others to make informed, unbiased decisions.

Requirements for Individual Product

Data Selection and Prediction

- Description of the selected topic and data set from Gapminder.org, including an explanation of why you chose it.
- Description of the individuals in the data set (e.g. countries or years) and two variables you believe may be related (e.g., CO_2 emissions and deforestation).
- Definition and explanation of each variable, including what it measures and how the data was collected .
- Identification of your chosen dependent variable (the outcome being predicted) and independent variable (the factor that may influence the dependent variable).
- Initial prediction:
    - An explanation of the expected relationship between the two variables.
    - An estimate of the strength of the relationship.
    - Evidence from the data that supports the prediction.

Each project has an individual product and a group product.

Complete Data Set

- Complete data set with no missing data (e.g., empty cells, placeholders like "N/A").
- Data for all individuals (countries or years) included in the sample across both variables.
- Description of the population from which the data was sampled and an explanation of how it was randomly selected

The teams support the individuals throughout the sprints.

Data Analysis

- A scatterplot that visually represents the relationship between the two variables.

- A calculation of the correlation coefficient, indicating the strength and direction of the relationship.
- A regression model that best fits the trend in the data.
- An interpretation of patterns and insights revealed by the analysis.

Findings and Interpretation

- A description of the discovered relationship between the variables
- An explanation of how the function type (linear, quadratic, or exponential) models the trend
- An interpretation of what the graph reveals
- A discussion of the real-world impact of the findings and a call to action based on the results
- A reflection on the limitations of the model and potential areas for future study

The team product is a culmination of the individuals' learning.

Reflection

- Written reflection on your learning:
    - How has this project helped you think critically about data?
    - In what ways have you improved your mathematical and analytical skills?
    - How will you use these skills to interpret data responsibly in the future?

Requirements for Team Product

Training Session or Educational Resource

Create a training session or educational resource to teach data analysis and critical thinking skills, including:

- A training session or educational resource designed to teach others how to analyze and interpret data.
- An identified audience, which may include classmates, family members, or community members.

Each week, students identify when and how they have met the content expectations.

- An introduction to key concepts such as scatterplots, correlation, and confirmation bias, along with demonstrations of data analysis tools (e.g., spreadsheets, graphing software).
- A focus on empowering others to critically evaluate data and make informed decisions.

Showcase Event

- The project will culminate in:
    - A school-wide showcase where your team will present your findings.
    - An informational poster summarizing key statistical analysis and insights.
    - A detailed presentation plan for engaging an audience of peers, teachers, family, and community members, highlighting your learning and growth.

Content expectations are reinforced through the other activities (5mm, NOWS, Interrupters, ENDS) in the module.

Badge

M112: Modeling with Data: Two-Variable Measurement Data

Learning Goals

In today's data-rich world, the ability to interpret and apply data is essential for making well-informed choices. Real-world situations often require understanding relationships between variables, such as how temperature affects ice cream sales or how economic factors impact quality of life. Developing data analysis skills strengthens problem-solving abilities and supports informed decision-making.

In this project, we will learn to

- Engage in the modeling cycle (112.a)
- Use scatter plots to represent two-variable measurement data (112.b)
- Describe visible patterns between two data sets in a scatter plot (112.c)

- Informally fit a straight line to scatter plots that suggest a linear association (112.d)
- Fit a linear, quadratic, or exponential function to two-variable measurement data (112.e)
- Use functions that have been fitted to two-variable measurement data to answer questions about the relationship being modeled (112.f)
- Assess the fit of a function using the correlation coefficient, residuals, and other tools (112.g)
- Explain insights gained from analyzing two-variable measurement data and limitations of curves fitted to data (112.h)