

Secondary IV Math Disciplinary Task: Correlation

Inquiry Question: How can statistics be used to lead and empower within a society?

Introduction

Political leaders across the world have had to make decisions that affect the daily lives of the people in those countries. In today's world, where knowledge is power, many leaders collect data in order to create positions and influence populations.

By using statistical analysis with different sets of data, we can find trends and/or correlation between quantities. This enables leaders to interpret the data and draw conclusions.

“There are three kinds of lies: lies, damn lies and statistics.”

-Mark Twain

Your task: In a Google document write a report which includes your responses to each of the following tasks. Be certain to include your graphs..

The Data: Get the data you will be using. The following are the instructions

1. You must log in to a Google account
2. Go to <http://msbro.weebly.com>, select 4CST Term 2
3. Click the link to Data for Disciplinary Task
4. Under the “File” menu select “make a copy” to save it on your Google Drive

The definitions for the variables listed are as follows:

- **Human Development Index (HDI):** A composite index measuring average achievement in three basic dimensions of human development—a long and healthy life, knowledge and a decent standard of living.
- **Mean years of schooling:** Average number of years of education received by people ages 25 and older.
- **Life expectancy at birth:** Number of years a newborn infant could expect to live if prevailing patterns of age-specific mortality rates at the time of birth stay the same throughout the infant's life.
- **Seats in national parliament:** Proportion of seats held by women in a lower or single house or an upper house or senate expressed as percentage of total seats.

Remember that Correlation does not imply Causation

- **Gross national income (GNI) per capita**: Aggregate income of an economy generated by its production and its ownership of factors of production, less the incomes paid for the use of factors of production owned by the rest of the world, converted to international dollars using purchasing power parity (PPP) rates, divided by midyear population.
- **Population with at least secondary education/Post-secondary Education**: Percentage of the population ages 25 and older who reached at least a secondary level of education.
- **Overall life satisfaction**: Please imagine a ladder, with steps numbered from zero at the bottom to ten at the top. Suppose we say that the top of the ladder represents the best possible life for you, and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time, assuming that the higher the step the better you feel about your life, and the lower the step the worse you feel about it? Which step comes closest to the way you feel?

Part 1

Prior to doing any calculations, what pairs of variables listed above do you think would have the strongest linear correlation? Give a reason for your answer.

Part 2

Using a spreadsheet (“Google sheets” or Microsoft Excel - the instructions for using Google Sheets are attached to this document), create a scatter plot, a regression line and R^2 (**remember**, you must still calculate R) between *Human Development Index* (y-value) and that of *Life Expectancy at Birth* (x-value) (**include your appropriately formatted graphs in an Appendix – Titles/labelled axes**).

Use your results to respond to each of the following questions (show all of your work):

- What is the correlation coefficient (R) between these 2 variables?
- What is the equation of the line of regression?
- Predict the *Life Expectancy at Birth* if a country had a *Human Development Index* of 0.05? ***Interpret whether this is plausible.***
- Predict the *Life Expectancy at Birth* if a country had a *Human Development Index* of 0.45? ***Interpret whether this is plausible.***
- Predict the *Human Development Index* if a country had a *Life Expectancy at Birth* of 100 years? ***Interpret whether this is plausible.***

Remember that Correlation does not imply Causation

Part 3

Using a spreadsheet (“Google sheets” or Microsoft Excel), create a scatter plot, a regression line and R^2 (**remember**, you must determine R) between **at least 2 other pairs of variables from the data provided** (this means that you will be creating at least 2 more scatter plots!) (**include your appropriately formatted graphs – Titles/labelled axes**).

- a) For each graph, list the variables being compared, the correlation coefficient (R) between the 2 variables as well as the equation of the line of regression. Give a qualitative description of the correlations.
- b) Rank each of the correlations found in Parts 2 and 3 from strongest to weakest. In one paragraph interpret your results. What conclusions can you draw in terms of a cause-and-effect relationship between the variables studied above? Do you believe there may be strong correlations between any of the other pairs of variables? If so, which ones and explain why.
- c) How do your results in (b) above compare to your response to Part 1?

Part 4

You are the leader of a country trying to **convince a population of how their tax dollars** should be spent (you may select any country - clearly state the country you have chosen). You must decide on a **stance** regarding one of the following topics and use statistics (specifically correlation and the regression line) to strengthen your position and influence your populace.

Topics:

- Gender
- Education
- Employment /Unemployment (*Note: You may also choose a topic of your own*)

Use the information relating to your topic from the 2011 list of *Human Development Report* and compare them. Interpret and take a position regarding spending based on one of the topics above – use the correlation and line of regression to support your position and to influence your populace.

Remember that Correlation does not imply Causation

Part 5

Repeat the process from Part 3, however this time use the data to make an **outlandish** claim. (Use the data to say something that you know is not true even though the data says otherwise)

Examples of a Stance/Claim

- More money should be spent on health care because...
- More money should be spent on education because...
- Taxes should be lowered and those who need health care should pay for it themselves because...
- Women's political campaigns should be subsidized by tax dollars because...
- Boys should be allowed to drop out of school at age 12 because...
- Taxes should be lowered and those who want an education should pay for it themselves because...
- Only girls should have their education paid for because...
- Education should be free and mandatory until the age of 21 because...
- More money should be spent to support the unemployed because...

The assessment rubrik and instructions for using Google sheets are both attached.

Criterion C: Communicating

Achievement level	Level descriptor
7–8	The student is able to: <ol style="list-style-type: none"> i. consistently use appropriate mathematical language. ii. use different forms of mathematical representation to consistently present information correctly. iii. move effectively between different forms of mathematical representation. iv. communicate through lines of reasoning that are complete, coherent and concise. v. present work that is consistently organized using a logical structure.
5–6	The student is able to: <ol style="list-style-type: none"> i. use appropriate mathematical language ii. use appropriate forms of mathematical representation to present information correctly iii. usually move between different forms of mathematical representation iv. communicate through lines of reasoning that are complete and coherent v. present work that is usually organized using a logical structure.
3–4	The student is able to: <ol style="list-style-type: none"> i. use some appropriate mathematical language ii. use appropriate forms of mathematical representation to present information adequately iii. communicate through lines of reasoning that are complete iv. adequately organize information using a logical structure. v. adequately organize information using a logical structure.
1–2	The student is able to: <ol style="list-style-type: none"> i. use limited mathematical language ii. use limited forms of mathematical representation to present information iii. communicate through lines of reasoning that are difficult to interpret.
0	The student does not reach a standard described by any of the descriptors above.

Remember that Correlation does not imply Causation

USING “Google Sheets” to make scatter plots, determine correlation coefficient (squared) and equation of the line of regression:

Making the Graphs/ Determining the Correlation Coefficient and Line of Regression

1. Select the data in the column for the x-value (include the “header” row).
2. In the **Insert** menu select **CHART**, **make sure to check the “use row ... as header” box**
3. Next to **DATA** click “select ranges”, **ADD** another range-select the data for the y-value and **OK**
4. Click the **MORE tab**
5. Select the **Scatter Plot** and click the picture on the top right
6. Select the **CUSTOMIZE** tab
7. Enter a title, a label for the x-axis (horizontal) and a label for the y-axis (vertical)
8. Scroll down – under **Trendline** select **LINEAR**
– under **Label** select **EQUATION** / click “Show R²” (you must calculate R)
9. **Click insert**
10. You can click on the arrow at the top right of the chart and select “Move to own sheet” (or copy & paste into a word processing document).

Remember that Correlation does not imply Causation