**Is there a connection between Poverty and Literacy Rate?**

Alex Hagopian and Elisabeth Jacob

Department of Mathematics and Computer Science, Rivier University

MA335: Mathematical Statistics

Dr. Olga Chuyan

March 25, 2024

**Is there a connection between Poverty and Literacy Rate?**

# Introduction

In this project, we examine countries percentage of population in poverty with literacy rate percentage. We explore whether there is a connection between poverty and literacy rate. Our sample pool, as previously mentioned, will be countries. These countries will range from those which are highly developed and wealthy Western nations, to those which are poor and have lower development.

# Data

|  |  |  |
| --- | --- | --- |
| **Country** | **Poverty Rate (%)** | **Literacy Rate (%)** |
| Albania | 1.5 | 97.2 |
| Angola | 52.9 | 66.0 |
| Brazil | 5.3 | 92.0 |
| Bulgaria | 1.4 | 98.4 |
| Burkina Faso | 63.1 | 34.6 |
| Benin | 53.2 | 32.9 |
| Chad | 64.6 | 22.3 |
| Chile | 1.7 | 96.9 |
| Costa Rica | 3.3 | 97.4 |
| Croatia | 0.7 | 99.1 |
| Dominican Republic | 4.3 | 93.8 |

|  |  |  |
| --- | --- | --- |
| Hungary | 0.8 | 99.1 |
| Italy | 1.2 | 98.8 |
| India  Iran | 46.5  5.7 | 74.4  85.5 |
| Malaysia | 0.2 | 95.0 |
| Kazakhstan | 0.5 | 99.8 |
| Luxembourg | 0 | 100.0 |
| Moldova | 0.3 | 99.4 |
| Germany | 0.2 | 99.0 |
| Russia | 0.3 | 99.7 |
| Saudi Arabia | 12.7 | 97.6 |
| Serbia | 2.8 | 99.5 |
| Slovenia | 0 | 99.7 |
| Spain | 1.4 | 98.6 |
| Sri Lanka | 11.3 | 92.3 |
| Syria | 11.9 | 86.4 |
| Turkey | 2.2 | 96.7 |
| United Arab Emirates | 19.5 | 97.6 |
| United States | 0.5 | 86.0 |

0

5

10

15

20

25

0-9.9

10-19.9

20-29.9

30-39.9

40-49.9

50-59.9

60-69.9

More

**Number of Countries**

**Poverty Rate (%)**

0

5

10

15

20

25

20-29.9

30-39.9

40-49.9

50-59.9

60-69.9

70-79.9

80-89.9

90-99.9

100

**Number of Countries**

**Literacy Rate (%)**

**Analysis**

We chose the class sizes of our histograms to be ten percentage points to be convenient to the readers. In our histogram examining the distribution of poverty rate, we found the mean to be 12.3% and the mode of this distribution to fall within 0-9.9, with 22 countries having a poverty rate within that range. The spread of this distribution is from 0-70%. The standard deviation is 20.6, which means that most of our data is 20.6% above or below 12.3%. This explains why this distribution has a single peak to the left (unimodal) and is right-skewed.

The five-number summary for poverty rate is:

Min = 0 ; Q1 = 0.5 ; Median = 1.95 ; Q3 = 12.1 ; Max = 64.6

A graph with numbers and a line

Description automatically generated

In our literacy rate histogram, we found the spread of this distribution is from 20-100%.

The mean is 87.9% and the mode is 21, with these countries having a literacy rate within 9099.9%. The standard deviation is 21.2, which means that most of our data falls within 21.2% above or below 87.9%. This explains why our distribution is unimodal and left-skewed. There were three outliers with extremely low literacy rates: Chad (22.2%), Benin (32.9%), and Burkina Faso (34.6%). These countries are very underdeveloped. They all have very unstable governments and political situations.

The five-number summary for literacy rate is:

Min = 22.3 ; Q1 = 86.3 ; Median = 97.3 ; Q3 = 99.1 ; Max = 100

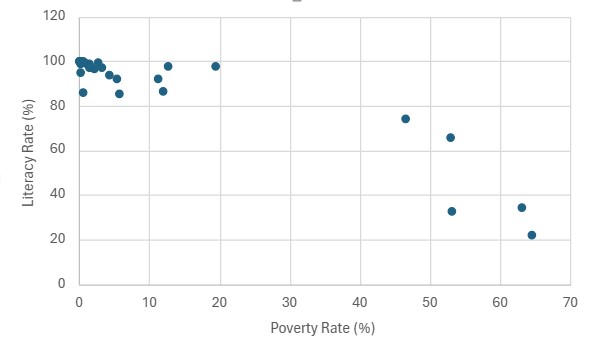
A graph with numbers and a line

Description automatically generated

We can observe that countries which have a lower poverty rate, tend to also have a higher literacy rate (referring to our mean and standard deviation).

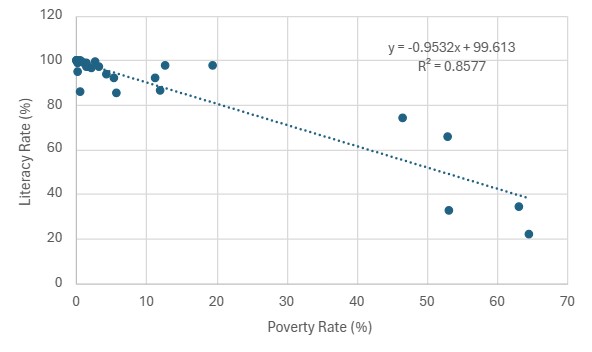
# Connection Between Variables

To confirm that countries with a lower poverty rate indeed have a higher literacy rate, a scatterplot graph was created using Microsoft Excel. The first step in creating the scatterplot involved inputting data including the names of individual countries along with their poverty rate and literacy rate percentages. Once all data was inputted into Excel and a scatterplot was created, the next step was to label the axes with the explanatory and response variables. In statistics, the *explanatory variable* is on the x-axis and is used to predict changes in the response variable. The *response variable* is on the y-axis and can be explained by the explanatory variable. In this scenario, poverty rate percentage is the explanatory variable, while literacy rate percentage is the response variable. In other words, the poverty rate of a country is a contributing factor to whether it will have a high or low literacy rate. Shown in *Figure 1.1* below is a scatterplot created with Excel that compares poverty rate and literacy rate for 30 countries. In this scatterplot, each dot represents an individual country.



*Figure 1.1 - Scatterplot that compares the poverty rates and literacy rates of individual countries.*

Once the data has been graphed using a scatterplot and the variables have been labeled, Excel offers the option to include a trendline and coefficient of determination. In Statistics, a *trendline*(also known as the line of best fit) is a linear representation of the general direction of the data. In this example, the trendline equation is y = -0.95x+99.613, indicating a negative relationship between poverty rate and literacy rate. The reliability of this trendline can be explained using the coefficient of determination (r-squared value) which in this case is 0.86. The *coefficient of* *determination*is always a value between 0 and 1 and measures how reliable a statistical model is in predicting an outcome. Because the coefficient of determination is 0.86, this means that 86% of variation in literacy rate (y-variable) can be explained by the linear regression on poverty rate (x-variable). Overall, this means that the data is a strong fit of the regression model. Shown below in *Figure 1.2* is the scatterplot, this time with both the trendline equation and coefficient of determination (r-squared).



*Figure 1.2 - Scatterplot with both the trendline equation and coefficient of determination (r-squared).*

The coefficient of determination of a data set can be used to calculate correlation. By taking the square root of the coefficient of determination, correlation (R-value) can be calculated. In Statistics, *correlation*measures the relationship between two variables, where |r| is always between –1 and 1. The closer that R is to 0, the weaker the correlation between the variables. On the contrary, if R is close to –1 or 1, then this indicates a strong correlation. A perfect correlation would be exactly 1 or –1, while 0 would indicate that no correlation exists. The sign of the correlation value explains whether there is a positive or negative correlation between the variables. Variables with positive a correlation move in the same direction, while those with a negative correlation move in opposite directions of one another. When the square root of 0.86 was taken, the result was r = -0.93. The trendline, -0.95x+99.613, indicates that there is a negative relationship between poverty rate and literacy rate, so this explains why 0.93 is negative. An r value of –0.93 is very close to –1, which specifies that the correlation is also extremely strong. Therefore, the relationship between poverty rate and literacy rate can be classified as an “extremely strong negative correlation”.

The strong correlation between poverty rate and literacy rate suggests that predictions made are fairly reliable. For a prediction to be considered “fairly reliable”, its R-value must be at least 0.7, indicating strong correlation. In Statistics, an explanatory value (x-value) is typically used to make predictions for a response value (y-value). The x-variable in this scenario is poverty rate, while the y-variable variable is literacy rate. Below in *Figure 1.3* are examples of predictions performed for two countries. The poverty rate of each country is used to predict its literacy rate:

**Guinea:** Poverty rate = 66.2%

* Predicted literacy rate: 0.95(66.2) + 99.613 = **36.732%**

**Canada:** Poverty rate = 7.4%

* Predicted literacy rate: 0.95(7.4) + 99.613 = **92.583%**

*Figure 1.3 - Predicting the literacy rates of two countries based on their poverty rates.*

In *Figure 1.3,* the literacy rates of both Guinea and Canada are predicted using their poverty rates. As evident in the figure above, the country of Guinea has a poverty rate of 66.2% (United Nations Development Programme, n.d.). When 66.2 was plugged into the trendline equation, it was predicted that the literacy rate of Guinea is approximately 36.732%. Although this value is not entirely accurate (the actual literacy rate of Guinea is around 45%), it serves as an estimate. Guinea is small country located on the Western Coast of Africa. As one of the poorest countries in the world, Guinea struggles to provide the necessary educational resources to support students, and this in return contributes to a low literacy rate in Guinea. Although a correlation value of -0.93 meets the requirements for being fairly reliable, this prediction is an example of extrapolation. *Extrapolation*occurs when data falls outside of the data range and is NEVER reliable. Because Guinea’s poverty rate is above the data range at 66.2%, this prediction is not entirely accurate. Therefore, additional factors besides poverty should be considered to help explain Guinea’s low literacy rate.

Contrary to Guinea, Canada is a highly developed western country with one of the largest economies in the world. The estimated literacy rate in Canada using the trendline equation is

92.583%. Although this is just an approximation and the actual literacy rate of Canada is around 99%, it proves that Canada is a very literate country. Canada has one of the strongest educational systems in the world, and it is home to multiples colleges and universities that offer higher educational opportunities for students. This in return leads to a high literacy rate in the country of Canada. Overall, this prediction for Canada is fairly reliable. As aforementioned, the correlation between poverty rate and literacy rate is -0.93, which is extremely strong. In order for a prediction to be considered fairly reliable, correlation must be greater 0.7 and 0.93 clearly meets this criterion. In addition, because Canada’s poverty rate of 7.4% falls within the data range, this is an example of interpolation.

# Conclusion

It can be concluded that there is a strong correlation between poverty rate and literacy rate. It is common for developing countries experiencing poverty to have a low literacy rate. Developing countries are less likely to have access to educational opportunities and resources including books, writing materials, technology, and the internet to name a few. Education is also not always a priority in the developing world. In addition, poverty in developing countries can also be associated with external factors including inadequate nutrition, limited access to healthcare, clean water shortages, and sanitation issues. These poor living conditions can negatively contribute to individual health and ability to learn. Overall, all these factors related to poverty can help to explain the low literacy rate in developing countries.

Unlike developing countries, developed nations often have a high literacy rate. The high literacy rate in developing countries can be explained by low poverty rate and economic stability. Countries with economic stability are industrialized and have the financial means to establish an organized educational system. Due to high standards of living, educational opportunities and resources are readily available. Developed countries are less likely to experience the poor living conditions and external factors evident in the developing world. Lastly, many developed countries have established laws and regulations that guide the educational system.

# References

Adult literacy rate. Our World in Data. (n.d.). https://ourworldindata.org/grapher/literacy-rateadults?tab=chart

*Blueprint for Transformation: the 2023 report of the National Advisory Council on Poverty*. (2023, October 30). Www.canada.ca. https://www.canada.ca/en/employmentsociciteddal-development/programs/poverty-reduction/national-advisorycouncil/reports/2023-annual.html#:~:text=In%202021%2C%20there%20were%20about

*Multidimensional Poverty Index 2023 Unstacking global poverty: data for high impact action Briefing note for countries on the 2023 Multidimensional Poverty Index Guinea*. (n.d.).

<https://hdr.undp.org/sites/default/files/Country-Profiles/MPI/GIN.pdf>

Poverty - World Bank Data explorer. Our World in Data. (n.d.).

https://ourworldindata.org/explorers/povertywb?tab=table&facet=none&Indicator=Share+in+poverty&Poverty+line=%242.15+per+d ay%3A+International+Poverty+Line&Household+survey+data+type=Show+data+from+ both+income+and+consumption+surveys&Show+breaks+between+less+comparable+sur veys=false&country=MOZ~NGA~KEN~BGD~BOL~OWID\_WRL