CS 2704

Final Project

Kyle Allan

3720130

**Introduction**

Does education make a country happier, if yes then does the higher the level of education the stronger the correlation get? This is what this project looks to uncover.

A countries populations happiness is measured in a survey called the Happiness Index with topics like psychological well-being, social support, education, community, and others being measured. This Happiness Index report will be used for the happiness level of a country.

Formal education is any organized structure of education. The data set that will be used for education deals with people aged 15+ who have had any primary, secondary, or tertiary education.

Secondary and Tertiary Education enrollment are anybody who is enrolled expressed as a percent of the population of the official age of that level of education.

These datasets will be used together to uncover if there is a positive correlation between a country's level of education and its happiness index, and if the higher the level of education the stronger the correlation.

The years 2015 and 2020 will be the only years we are concerned with since these are the only two years that exist in all datasets.

**Hypothesis**

The share of a population with some formal education has a positive correlation with the happiness index of a country and the higher the level of education the stronger the correlation.

**Analysis and Implication**

The datasets were taken into Python as data frames. The datasets were then merged on the ‘Entity’ column from the education dataset which was country names and the ‘Country’ column from the happiness index dataset. Once the data frames were created a cleaned-up version was created to have all the relevant info of a country in one entry instead of two separate ones for the years 2015 and 2020.

Once these were created, I ran a Shapiro-Wilk test to test the normality of my data. Out of the education level in the years 2015 and 2020, both were not normally distributed and the happiness index in 2015 was not normally distributed but the happiness index in 2020 was normally distributed. For the level of education, I analyzed secondary education and tertiary education. For secondary enrollment in years 2015 and 2020 it was not normally distributed, and for tertiary enrollment it was normally distributed in 2015 but not in 2020.

* 2015 happiness index P-value: 0.0145 < 0.05
* 2020 happiness index P-value: 0.1128 > 0.05
* 2015 share of the population with formal education: 2.63e-15 < 0.05
* 2020 share of the population with formal education: 5.53e-16 < 0.05
* 2015 Secondary Enrollment: 0.00029 < 0.05
* 2020 Secondary Enrollment: 0.00016 < 0.05
* 2015 Tertiary Enrollment: 0.0794 > 0.05
* 2020 Tertiary Enrollment: 0.015 < 0.05

Since the datasets were majority not normally distributed, I used a Spearman Rank test to test my data for correlation. There was a moderate positive relationship between the happiness index and the share of the population with formal education in both years 2015 and 2020. There was a strong correlation between secondary enrollment and happiness index in 2015 and 2020. There was a moderate positive correlation with tertiary enrollment and happiness index.

* 2015 correlation coefficient between happiness index and formal education is 0.5927 with p-value 5.64e-14
* 2020 correlation coefficient between happiness index and formal education is 0.6023 with p-value 1.72e-14
* 2015 correlation between happiness index and secondary enrollment is 0.662 with p-value 7.427e-11
* 2020 correlation between happiness index and secondary enrollment is 0.7011 with p-value 1.745e-12
* 2015 correlation between happiness index and tertiary enrollment is 0.565 with p-value 1.06e-07
* 2020 correlation between happiness index and tertiary enrollment is 0.527 with p-value 9.527e-07

With this information, basic statistic tables were created for each year, histograms were created to display the distribution of data, and a scatterplot of countries was mapped out.

**Conclusion**

After analyzing the data and performing multiple tests I can conclude that there is a positive correlation between a country's share of population having some formal education and the country's happiness index, and that the higher the level of education there is not a stronger correlation. Through the trials and tribulations, I faced I have now learned a great deal in that the preparation of data is very important before conducting any tests or computations on it as well.

This project was a benefit in understanding real-world data analytics and using statistics to answer questions. I feel more proficient in using Python and have gained a new appreciation for it as it makes dealing with data simple.

**References**

World Bank (2023) – processed by Our World in Data. Gross enrolment ratio in primary education [dataset]. World Bank, World Bank Education Statistics (EdStats) 2023 [original data]. Retrieved December 8, 2023 from <https://ourworldindata.org/grapher/gross-enrollment-ratio-in-primary-education>

World Bank (2023) – processed by Our World in Data. Gross enrolment ratio in secondary education [dataset]. World Bank, World Bank Education Statistics (EdStats) 2023 [original data]. Retrieved December 8, 2023 from <https://ourworldindata.org/grapher/gross-enrollment-ratio-in-secondary-education>

World Bank (2023) – processed by Our World in Data. Gross enrolment ratio in tertiary education [dataset]. World Bank, World Bank Education Statistics (EdStats) 2023 [original data]. Retrieved December 8, 2023 from <https://ourworldindata.org/grapher/gross-enrollment-ratio-in-tertiary-education>

Hannah Ritchie, Veronika Samborska, Natasha Ahuja, Esteban Ortiz-Ospina and Max Roser (2023) - “Global Education” Published online at OurWorldInData.org. Retrieved from: <https://ourworldindata.org/global-education>

Simona ASM. (2023). *World Happiness Index By Reports (2013-2023).* Kaggle. <https://www.kaggle.com/dataset/simonasm/world-happiness-index-by-reports-2013-2023/>