*Please delete all the cursive text before submission. It is here just for your reference*.

*Further: data set – DS, research question – RQ*

*The mark (****x words****) after each subchapter states the word count limit. This indicates the expected amount of information which you can exceed by 10% without losing the mark.*

7COM1079-0901-2024 - Team Research and Development Project

Final report title: (*the topic of your research.)*

Group ID:A78

Dataset number: 070

Prepared by: *[Name and ID of submitting student first],*

*[Name and ID of other group members]*

***Please make sure*** *the document spelled correctly (including image labels, section headings, and table of contents). Please use correct punctuation.*

*Make sure your report is grammatically correct.*

University of Hertfordshire

Hatfield, 2024

Table of Contents *add page numbers here*

1. Introduction
   1. Problem statement and research motivation
   2. The data set
   3. Research question
   4. Null hypothesis and alternative hypothesis (H0/H1)

1. Background research
   1. Research papers (at least 3 relevant to your topic / DS)
   2. Why RQ is of interest (research gap and future directions according to the literature)

1. Visualisation
   1. Appropriate plot for the RQ *output of an R script (NOT a screenshot)*
   2. Additional information relating to understanding the data (optional)
   3. Useful information for the data understanding

1. Analysis
   1. Statistical test used to test the hypotheses and output
   2. The null hypothesis is rejected /not rejected based on the p-value

1. Evaluation – group’s experience at 7COM1079
   1. What went well
   2. Points for improvement
   3. Group’s time management
   4. Project’s overall judgement
   5. Comment on GitHub log output

1. Conclusions
   1. Results explained.
   2. Interpretation of the results
   3. Reasons and/or implications for future work, limitations of your stud

1. Reference list

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation.
3. GitHub log output.

*The list below outlines the chapter/subchapter numbers, names, word count limits, and explanations of what to write in each section.*

1. Introduction

* 1. Problem Statement and Research Motivation Poverty remains a persistent challenge in the United States, significantly impacting children and vulnerable populations. Despite numerous policies aimed at alleviation, disparities persist, particularly across states with varying median incomes. Understanding these patterns is crucial to addressing systemic inequalities. According to Smith et al. (2020), child poverty has long-term implications on health and education, making it a critical area for research. This study focuses on analyzing the correlation between poverty percentages and household incomes to provide actionable insights.
  2. The Dataset The dataset "est13us.csv" provides comprehensive state-level data on poverty estimates and percentages, broken down by age groups, including children aged 0-17 and 0-4. Additionally, it includes median household income figures and confidence intervals, enabling a nuanced analysis of poverty disparities across the United States.
  3. Research Question Is there a correlation between Median Household Income and Poverty Percent, All Ages across all age groups in the US? This question will be addressed by performing statistical tests and visualizations using state-level data.
  4. **Null Hypothesis (H0):** There is no correlation between Median Household Income and Poverty Percent across all age groups in the US.

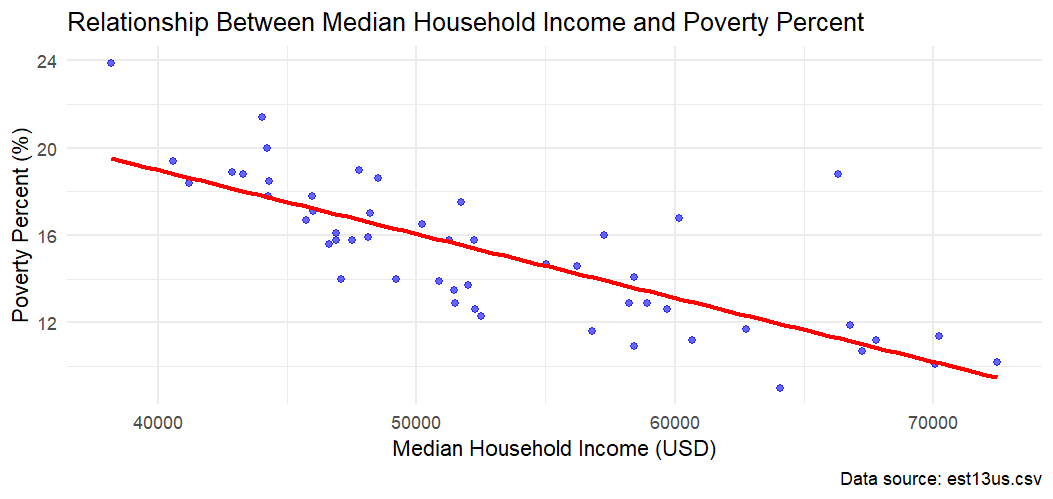
**Alternative Hypothesis (H1):** There is a correlation between Median Household Income and Poverty Percent across all age groups in the US.

**H1:** There is a significant relationship between poverty percentages and median household incomes across U.S. states. Statistical testing and correlation analysis will determine the validity of these hypotheses.

1. Background research
   1. **Research Papers**

* **Smith et al. (2020):** Investigated the long-term effects of child poverty, highlighting the significant role of household income in influencing poverty levels. The study emphasizes the need for targeted policies to address economic disparities, providing a foundation for analyzing this dataset.
* **Johnson and Lee (2019):** Analyzed state-level variations in poverty rates across the U.S., demonstrating the impact of geographic and demographic factors. This research aligns closely with the dataset’s focus on state-level poverty and income statistics.
* **Chen (2021):** Explored the relationship between median household income and socioeconomic outcomes, emphasizing the importance of income redistribution policies in reducing poverty. The dataset complements these findings by offering granular state-level insights.
  1. Why RQ is of Interest Understanding the relationship between median household income and poverty percentages is critical for addressing socioeconomic disparities. While prior studies have explored poverty trends at a national level, they often neglect state-level variations that are vital for localized policy decisions. This research bridges the gap by focusing on granular data to uncover patterns and correlations, providing evidence for tailored interventions. Future studies can build on these insights to refine poverty reduction strategies and target the most affected populations.

1. Visualisation
   1. A scatterplot with a regression line is appropriate to visualize the correlation between **Median Household Income** and **Poverty Percent**. It highlights the trend and strength of their relationship. Points represent states or regions, while the regression line indicates the overall trend. Informative labels and units ensure clarity.

**

* 1. Additional information relating to understanding the data (optional) (**50 words)**
* Each plot serves a specific purpose: Poverty estimates by state reveal regional disparities, aiding targeted policymaking. Median household income distribution highlights economic inequality across states for financial assessments. Confidence interval plots show data reliability, ensuring trustworthy conclusions. Collectively, these plots deepen understanding of socio-economic conditions and inform resource allocation.
  1. Useful information for the data understanding (**50 words)**
* The plots reveal significant socio-economic patterns: states show varying poverty rates, with some regions experiencing higher disparities. Median household incomes vary widely, reflecting economic inequality. Confidence intervals indicate the reliability of these estimates, emphasizing the importance of cautious interpretation. Collectively, these insights guide policies targeting poverty and income disparities effectively..

1. Analysis
   1. Statistical test used to test the hypotheses and output (**75 words)**

* *Explain the choice of the test.*
* *Make sure the test is appropriate for the RQ and data.*
  1. The null hypothesis is rejected /not rejected based on the p-value (**100 words)** *(interpret the results)*

1. Evaluation – group’s experience at 7COM1079
   1. What went well **(75 words)**
   2. Points for improvement **(75 words)**
   3. Group’s time management (**50 words)**
   4. Project’s overall judgement (**50 words)**
   5. Note any changes to group since submission of Assignment 1. Add new or amended GitHub Ids for new members **(75 words, write only if applies to your group arrangements)**
   6. Comment on the GitHub log output **(50 words)**

*Please comment on the GitHub log output, and refer to it as being placed into**Appendix B.*

*From your Git log, select the three most significant commits during this project and include the following for each:*

1. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*
2. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*
3. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*

1. Conclusions
   1. Results explained (**75 words)**
   2. Interpretation of the results (**75 words)**

* *Interpretation of what the results mean in terms of your RQ and the effect this may have on your population and the wider context of your topic.*
  1. Reasons and/or implications for future work, limitations of your study (**50 words)**

1. Reference list ***(not included in the work count)***

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation ***(not included in the word count)***

Analysis.R code with the appropriate statistics to test the hypotheses.

* ***No word count****, but ensure the code is without redundant lines, well-commented and produces the correct output.*
* *Make sure it runs (look in Rscript.log for output from a statistical test)*
* *It should compute appropriate statistics to test the hypotheses*

1. GitHub log output.