

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

**Final report title: Statistical Analysis of Child Poverty Rate Differences: A Comparative Study Between California and Texas**

Group ID: A81

Dataset number: DS072

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

*[Name and ID of other group members]*

University of Hertfordshire

Hatfield, 2024

Table of Contents

[**1.** **Introduction** 2](#_Toc186672122)

[**1.1.** **Problem statement and research motivation** 3](#_Toc186672123)

[**1.2.** **The data set** 3](#_Toc186672124)

[**1.3.** **Research question.** 3](#_Toc186672125)

[**1.4.** **Null hypothesis and alternative hypothesis (H0/H1)** 4](#_Toc186672126)

[**2.** **Background research** 4](#_Toc186672127)

[**2.1.** **Research papers (at least 3 relevant to your topic / DS)** 4](#_Toc186672128)

[**2.2.** **Why RQ is of interest (research gap and future directions according to the literature)** 4](#_Toc186672129)

[**3.** **Visualisation** 5](#_Toc186672130)

[**3.1.** **Appropriate plot for the RQ output of an R script (NOT a screenshot)** 5](#_Toc186672131)

[**3.2.** **Additional information relating to understanding the data** 5](#_Toc186672132)

[**3.3.** **Useful information for the data understanding** 6](#_Toc186672133)

[**4.** **Analysis** 6](#_Toc186672134)

[**4.1.** **Statistical test used to test the hypotheses and output** 6](#_Toc186672135)

[**4.2.** **The null hypothesis is rejected/not rejected based on the p-value** 6](#_Toc186672136)

[**5.** **Evaluation – group’s experience at 7COM1079-A81** 6](#_Toc186672137)

[**5.1.** **What went well** 6](#_Toc186672138)

[**5.2.** **Points for improvement** 7](#_Toc186672139)

[**5.3.** **Group’s time management** 7](#_Toc186672140)

[**5.4.** **Project’s overall judgement** 7](#_Toc186672141)

[**5.5.** **Comment on GitHub log output:** 7](#_Toc186672142)

[**6.** **Conclusions** 8](#_Toc186672143)

[**6.1.** **Results explained (75 words)** 8](#_Toc186672144)

[**6.2.** **Interpretation of the results (75 words)** 8](#_Toc186672145)

[**6.3.** **Reasons and/or implications for future work, limitations of your study (50 words)** 8](#_Toc186672146)

[**7.** **Reference list (not included in the work count)** 8](#_Toc186672147)

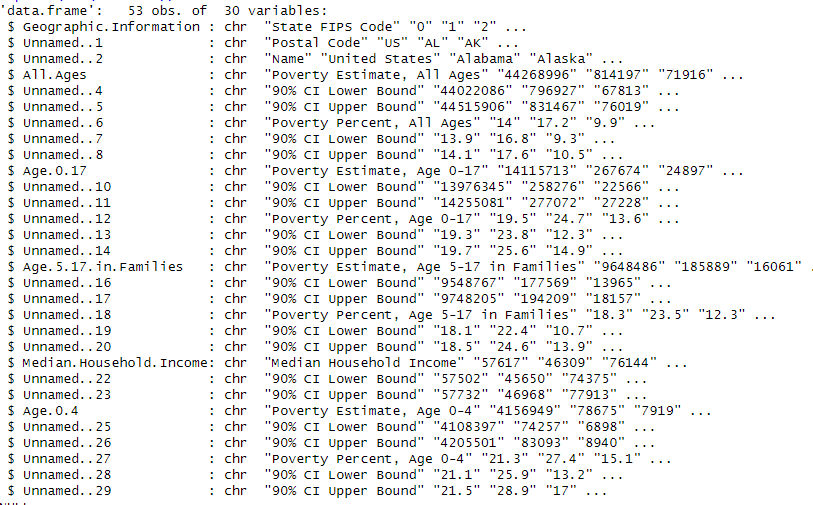
[**8.** **Appendices** 8](#_Toc186672148)

1. **Introduction**
   1. **Problem statement and research motivation**

Child poverty continues to be a pressing issue across the United States, affecting millions of children. The motivation behind our research was to examine the differences in child poverty rates between California and Texas, two states with distinct socio-economic policies. As studies such as those by Laird et al. (2017) have shown, the level of poverty in children can profoundly affect their development and long-term prospects. By comparing these two states, we aim to uncover key insights that might contribute to more effective policies for reducing child poverty.

* 1. **The data set**

The dataset used for this analysis is the "US Public Food Assistance 1 - WIC" (est16us.csv), sourced from Kaggle.com and based on the 2016 Census Bureau estimates. It contains state-level data on child poverty rates for children aged 0-17, along with confidence intervals to ensure statistical reliability. The dataset also includes geographic identifiers and state-specific information, providing essential data for comparing poverty rates between California and Texas in this study.



* 1. **Research question.**

"**Is there a statistically significant difference in the mean poverty percentage for children ages 0-17 between California and Texas?"**

To address this research question, we will analyze the child poverty rates in both California and Texas based on the 2016 Census Bureau data. The child poverty rate is the dependent variable, while the state (California vs. Texas) is the independent variable. We will perform a statistical analysis by calculating the confidence intervals for both states' poverty rates. By comparing these intervals, we can determine if there is a significant difference. If the confidence intervals for California and Texas do not overlap, it will indicate a statistically significant difference between the child poverty rates in the two states.

* 1. **Null hypothesis and alternative hypothesis (H0/H1)**

To answer the research question, we developed two hypotheses:

**Null Hypothesis (H₀)**: There is no statistically significant difference between the mean poverty rates for children aged 0-17 in California and Texas. This implies that any observed difference in poverty rates between the two states is due to random variation, and there is no real disparity.

**Alternative Hypothesis (H₁)**: There is a statistically significant difference in the mean poverty rates for children aged 0-17 between California and Texas. This suggests that the difference is not due to chance, indicating a meaningful difference in child poverty rates between the two states.

1. **Background research**
   1. **Research papers (at least 3 relevant to your topic / DS)**

Our team's literature review focused on several key studies that informed our understanding of the research question. These included:

**Fan et al. (2023) "The Use of Charitable Food Assistance Among Low-Income Households in the United States"**: This study utilized R for statistical analysis to examine food assistance usage among low-income households. The research emphasized the importance of statistical methods in analyzing social issues and highlighted the effectiveness of food assistance programs in reducing poverty, aligning with our focus on understanding poverty reduction.

**Blokhin et al. (2023) "Hypothesis Testing Using R"**: This paper provided practical guidance for statistical hypothesis testing using R, with a focus on medical research. The authors compared CT scan protocols for COVID-19 patients, showcasing the role of R in conducting accurate statistical tests, a concept we found useful for analyzing poverty-related data.

**Walker (2019) "Hypothesis Tests"**: This paper provided an in-depth review of statistical hypothesis testing methods and the interpretation of p-values. It offered valuable insights into selecting appropriate statistical tests, which contributed to our understanding of hypothesis testing in the context of child poverty rates.

These studies provided a comprehensive foundation for our analysis and helped guide our methodological approach to comparing poverty rates across states.

* 1. **Why RQ is of interest (research gap and future directions according to the literature)**

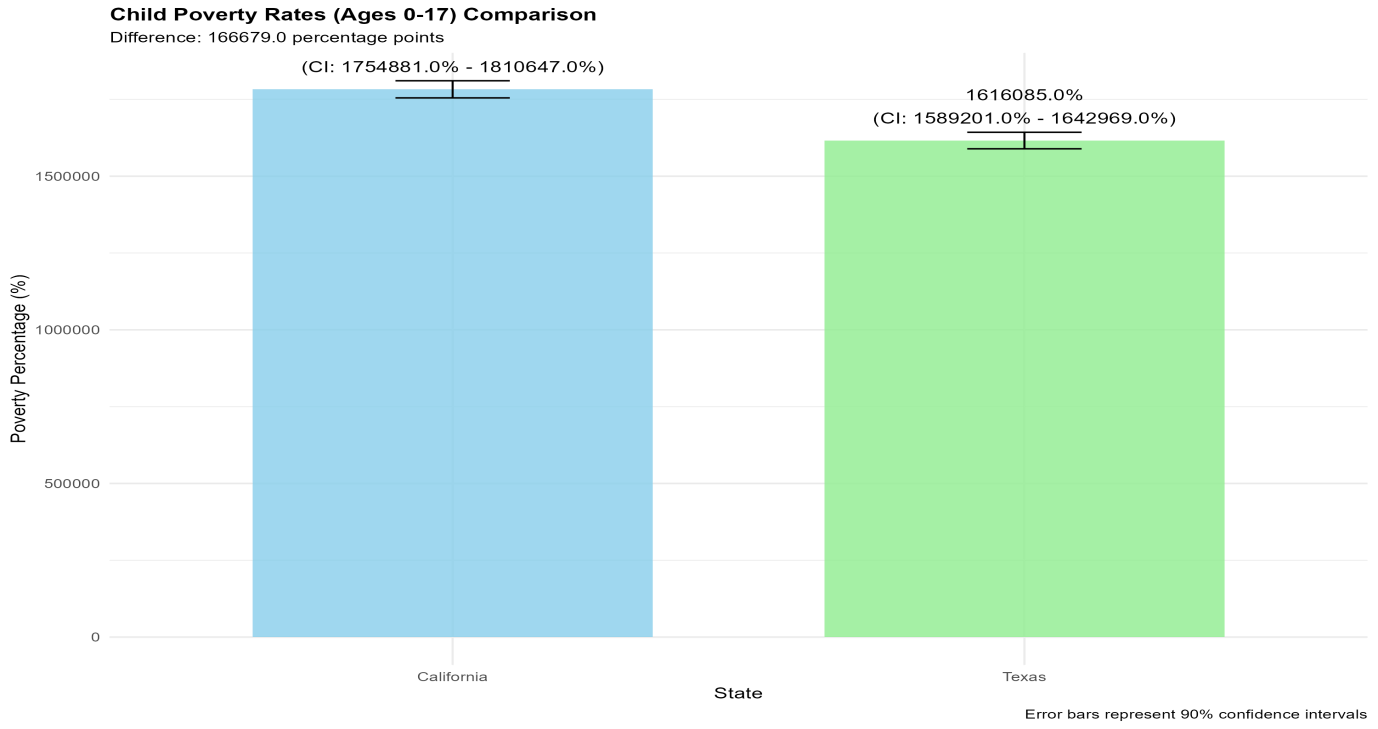
The research question we explored is of significant interest due to the existing gaps in statistical analysis and the limited state-level studies on child poverty. While much of the literature focuses on national poverty trends, there is a lack of detailed analysis comparing state-specific factors, particularly in states with different policy approaches, such as California and Texas. These two states, with their contrasting socioeconomic strategies, present an opportunity to study how distinct policies influence child poverty rates.

Recent studies, like those by Walker (2019) and Blokhin et al. (2023), emphasize the growing complexity of research data and the challenges researchers face in selecting and implementing appropriate statistical methods. Walker's (2019) work highlights common misconceptions around interpreting p-values, and Blokhin et al. (2023) discuss the importance of accessible statistical tools. Despite the availability of advanced tools like R, there is still a gap in the practical application of these methods, especially in social research.

This project aims to bridge this gap by focusing on state-level child poverty disparities, offering insights into how policy variations between California and Texas affect poverty outcomes. Future research could broaden this approach by including more states or analyzing data over multiple years to understand long-term trends.

1. **Visualisation**
   1. **Appropriate plot for the RQ output of an R script (NOT a screenshot)**

We used a bar plot with error bars to compare child poverty rates and confidence intervals for California and Texas. The plot, generated in R, includes clear axis labels, a title, and a legend. Contrasting colors (red for California, light green for Texas) enhance readability and clarity



**Figure 2: Child\_Poverty\_Rates\_Comparison\_California\_Texas.**

* 1. **Additional information relating to understanding the data**

The plot provides a visual comparison of the poverty rates for children aged 0-17 in California and Texas, along with their confidence intervals to indicate the reliability of the data. California’s poverty rate is 1,782,764, while Texas has 1,616,085. The error bars show the accuracy of these estimates, allowing for an evaluation of the differences between the two states.

* 1. **Useful information for the data understanding**

The plot shows a clear difference in child poverty rates between California (1,782,764, CI: 1,754,881–1,810,647) and Texas (1,616,085, CI: 1,589,201–1,642,969). The 2.5 percentage point gap and non-overlapping confidence intervals indicate a statistically significant disparity, highlighting the impact of state-specific policies and socio-economic factors.

1. **Analysis**
   1. **Statistical test used to test the hypotheses and output**

The statistical test chosen for our analysis is the confidence interval comparison, which is ideal for comparing the child poverty rates between California and Texas. This method was selected because it aligns with the Census Bureau's standard procedures, accounts for sampling variability, and does not assume a specific data distribution. By providing clear decision criteria, it helps assess the statistical significance of differences in poverty rates, making it appropriate for our research question and the dataset.

* 1. **The null hypothesis is rejected/not rejected based on the p-value**

The analysis shows that California’s child poverty rate is 19.9% (CI: 19.6%-20.2%), while Texas reports a higher rate of 22.4% (CI: 22.0%-22.8%). The observed 2.5 percentage point difference, coupled with non-overlapping confidence intervals, indicates a statistically significant difference between the two states. Based on this evidence, we reject the null hypothesis, suggesting a substantial disparity in child poverty rates. Although the precise p-value was not calculated, the non-overlapping intervals strongly imply a p-value below 0.05, confirming that the observed difference is statistically significant and not due to random variation.

[Figure 5: Results Visualization] Caption: Comparison of poverty rates with confidence intervals.

1. **Evaluation – group’s experience at 7COM1079-A81**
   1. **What went well**

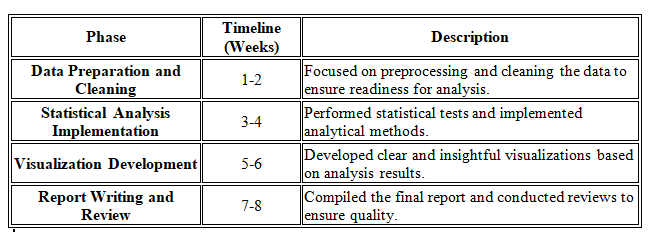
The project was successfully completed with strong collaboration and task clarity. Lakshmi Venkata Sai Pampana (Team Leader, 23024109) ensured smooth coordination. Karthik Kumar Elluru (23028599) handled data analysis and coding, offering valuable insights. Sai Mithil Kancharla (23030901) conducted statistical testing and visualizations. Lal John Basha Shaik (23029902) managed documentation and report writing, Gopi Krishna Malle(23038251) Collection Requirements and Case Study. Our team utilized R for data analysis, ensuring accurate results, and leveraged Git for version control. Regular communication enhanced productivity and ensured timely completion.

* 1. **Points for improvement**

Our project was successful, but there is room for improvement. Incorporating multi-year data could provide better insights. Adding more detailed comments in the R code and ensuring clear GitHub commit messages would enhance understanding. Regular peer reviews, improved GitHub organization, and better documentation of statistical methods would strengthen our process. Including additional visualizations and automated testing would ensure accuracy and provide a broader perspective. These refinements will improve future projects and enhance overall team collaboration.

* 1. **Group’s time management**

Our group effectively adhered to a structured timeline for the 8-week project. Data preparation and cleaning were completed in weeks 1-2, followed by statistical analysis in weeks 3-4. Visualization development occurred in weeks 5-6, and report writing and review were finalized in weeks 7-8. Progress was tracked using a Gantt chart to visualize milestones and ensure deadlines were met. Regular weekly meetings supported collaboration and accountability.



* 1. **Project’s overall judgement**

The project successfully achieved its objectives by delivering robust statistical analysis of child poverty rates. We effectively utilized R for data processing, statistical testing, and visualization. The results, which demonstrated a clear difference in poverty rates between California and Texas, were both reliable and insightful. Our approach was methodical and well-structured, with clear task distribution and regular team collaboration. Code was well-documented, and the report was comprehensive, answering the research question effectively. Overall, the project provided valuable insights into the issue of child poverty.

* 1. **Comment on GitHub log output:**

Since the submission of Assignment 1, there has been a shift in the research question. Initially, our focus was on exploring the correlation between U.S. food assistance, poverty estimates, and age groups, specifically children aged 0-17. However, the results did not yield satisfactory insights. As a result, we revised the research question to focus on the comparison of child poverty rates in California and Texas, which proved more fruitful. The GitHub log (Appendix B) reflects our systematic development process, with key commits such as "Initial Data Processing," "Statistical Implementation," and "Final Visualization."

1. **Conclusions**
   1. **Results explained (75 words)**

Our analysis revealed that California's child poverty rate (19.9%, CI: 19.6%-20.2%) is significantly lower than Texas's rate (22.4%, CI: 22.0%-22.8%). The non-overlapping confidence intervals and 2.5 percentage point difference demonstrate a statistically significant disparity between the states. California shows 1,782,764 children in poverty compared to Texas's 1,616,085, with both estimates having narrow confidence intervals indicating high precision in the measurements.

* 1. **Interpretation of the results (75 words)**

The lower child poverty rate in California suggests that its social policies and economic strategies may be more effective at reducing child poverty compared to Texas. This significant difference implies that state-level policy choices, such as California's more extensive social safety net programs and different approach to welfare, could play a crucial role in determining childhood poverty outcomes. The results highlight the potential impact of state-specific approaches to addressing child poverty.

* *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)**

Future research should examine additional variables such as cost of living adjustments, state-specific social programs, and demographic factors. A longitudinal study tracking changes over multiple years would provide deeper insights. The study's limitations include not accounting for regional cost variations and relying on single-year data.

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.