**Exploring the Economic Dynamics of American Families: An Analysis of Income, Poverty, and Family Composition**

Solomon T Tessema

National University

TIM-8501 v4: Exploratory Data Analysis

Dr. Lawrence Fulton

February 20, 2024

**Exploring the Economic Dynamics of American Families: An Analysis of Income, Poverty, and Family Composition**

The economic conditions of American families have become a topic of significant interest among policymakers, researchers, and the public. Understanding the interplay between income, poverty levels, and family composition is crucial for developing effective social policies and interventions. This paper aims to explore these dynamics by leveraging data from the U.S. Census Bureau's 2020 Family Finance and Public Assistance dataset (referred to as ffpub20.csv). Specifically, it examines how family size, the presence of children, and education assistance income impact the overall economic well-being of families.

The study is guided by three main research questions: 1) What is the relationship between family size and total family income? 2) Does the presence of children under 18 impact the poverty level of families? and 3) Are families with a higher education assistance income more likely to have a higher total annual income? These questions are explored through a comprehensive Exploratory Data Analysis (EDA), employing statistical techniques to uncover patterns, trends, and correlations within the data.

Hypotheses are formulated to test the null assumption of no significant relationships against the alternative that significant relationships do exist in the areas of interest. The analysis includes measures of central tendency and variability, outlier detection, correlation analysis, and advanced modeling techniques to provide a detailed understanding of the socioeconomic factors influencing American families' economic conditions.

This paper adds to the ongoing discussion about family economic security in the United States by providing empirical insights into how various factors, including family composition and income sources, affect families' financial situations. By identifying the key determinants of economic well-being, this study aims to inform the development of targeted policies and programs to support American families, particularly those most vulnerable to economic instability.

The findings of this analysis try to offer meaningful insight on family economics and to assist in the crafting of more effective social policies aimed at reducing poverty and enhancing the economic resilience of families across the country.

**Literature Review**

The economic well-being of families is a multifaceted issue that has been extensively explored in academic literature, focusing on income levels, poverty, family structure, and the impact of social policies. This literature review synthesizes key findings from recent studies, providing a foundation for the current analysis of the ffpub20.csv dataset from the U.S. Census Bureau.

**Family Size and Economic Outcomes**

Research has shown that family size has significant implications for economic outcomes and well-being. Larger families often face greater financial strain due to higher living costs and resource dilution among members (Blake, 1981; Downey, 1995). However, the extent to which family size affects economic status varies, with some studies highlighting the role of social support systems and public policies in mitigating financial pressures (Thomson, Hanson, & McLanahan, 1994).

**Poverty and Child Well-being**

The presence of children in a household is another critical factor influencing poverty levels. Studies have documented the higher incidence of poverty among families with children, particularly single-parent families (Edin & Lein, 1997; Mayer, 1997). The costs associated with child-rearing, including education and healthcare, contribute significantly to the economic burden on families, impacting their poverty status (Duncan & Brooks-Gunn, 1997).

**Education Assistance and Family Income**

Education assistance programs are designed to support the educational pursuits of individuals, potentially affecting the family's overall economic situation. Research indicates that access to education assistance can lead to improved employment prospects and higher income levels, contributing to family economic security (Dynarski, 2003; Hout, 2012). However, the effectiveness of these programs in enhancing family income and reducing poverty is contingent upon various factors, including the adequacy of the assistance provided and the economic context.

**The Role of Social Policies**

The impact of social policies on family economic well-being is a critical area of study. Policies such as tax credits, child support enforcement, and public assistance programs have been analyzed for their effectiveness in alleviating poverty and supporting families (Moffitt, 1992; Grogger & Karoly, 2005). Findings suggest that while social policies can provide essential support to low-income families, their success in improving economic outcomes is influenced by policy design, implementation, and the broader economic environment.

**Conclusion**

The literature on family economics underscores the complexity of factors influencing the economic well-being of families. This review highlights the importance of considering family size, the presence of children, and the role of education assistance and social policies in understanding family economic dynamics. The current analysis of the ffpub20.csv dataset builds on this literature, seeking to provide new insights into the economic conditions of American families in 2020. By examining the relationships between family composition, income, and poverty levels, this study aims to contribute to the ongoing discourse on how to best support families facing economic challenges.

**Dataset Overview**

The dataset under examination, derived from the U.S. Census Bureau's 2020 Family Finance and Public Assistance survey (referred to as ffpub20.csv), provides a comprehensive snapshot of the economic conditions and assistance received by American families during the year 2020. This dataset is particularly valuable for researchers and policymakers interested in understanding the socio-economic dynamics of family units across the United States.

Key columns in this dataset include variables related to family composition, income sources, poverty levels, and various forms of public and private assistance. Notable columns such as FPERSONS (number of persons in the family), FANNVAL (annual family income), and POVLL (poverty level line) offer insights into the economic status and living conditions of families. Other important variables include FEDVAL (education assistance income), FINC\_ANN (total annual income from all sources), and FOWNU18 (number of own children under 18), which are crucial for analyzing the impact of family size, educational support, and child-rearing responsibilities on economic well-being.

This dataset serves as a critical tool for exploring how demographic factors and external economic supports influence the financial stability of families. By examining the relationships between these variables, researchers can gain deeper insights into the challenges and opportunities facing American families, contributing to the development of targeted policies aimed at enhancing economic security and reducing poverty. The analysis of this dataset not only sheds light on the current state of family finances but also provides a basis for longitudinal studies to track changes over time, especially in response to policy shifts and economic trends.

**Hypotheses**

**Research Question and Hypothesis 1**

Research Question 1: What is the relationship between family size (FPERSONS) and total family income (FANNVAL)?

* Null Hypothesis (H0): There is no significant correlation between family size and total family income.
* Alternative Hypothesis (H1): There is a significant positive or negative correlation between family size and total family income.

**Research Question and Hypothesis 2**

Research Question 2: Does the presence of children under 18 (FOWNU18) impact the poverty level (POVLL) of families?

* Null Hypothesis (H0): The presence of children under 18 does not significantly impact the poverty level of families.
* Alternative Hypothesis (H1): Families with children under 18 have a significantly different poverty level compared to those without.

**Research Question and Hypothesis 3**

Research Question 2: Are families with a higher education assistance income (FEDVAL) more likely to have a higher total annual income (FINC\_ANN)?

* Null Hypothesis (H0): There is no significant relationship between education assistance income and total annual income.
* Alternative Hypothesis (H1): Families with higher education assistance income tend to have higher total annual incomes.

**Exploratory Data Analysis Plan**

This exploratory data analysis includes:

* Measures of Variability and Central Tendency: To summarize the key income and family size variables.
* Frequency, Variance, and Standard Deviation: To assess the distribution of income levels and family sizes.
* Outlier Detection and Distribution Modality: To identify and address outliers in income and family-related variables.
* Univariate Analysis Methods: To analyze individual variables such as FANNVAL, FPERSONS, and POVLL.
* Cluster Analysis and Data Grouping: To explore patterns and groupings among families based on income, size, and poverty levels.

**Visualization Plan**

This exploratory data analysis includes the following visualizations:

* Histograms and Box Plots: For distribution of income (FANNVAL, FINC\_ANN) and family size (FPERSONS).
* Scatter Plots: To explore correlations between family size and income, and between education assistance income and total annual income.
* Heatmaps: For correlation analysis among selected variables.

**Analysis**

The preliminary data analysis of the ffpub20.csv dataset provides valuable insights into the economic conditions of American families in 2020. The dataset consists of 69,959 entries, each representing a family unit, with a total of 85 columns covering various aspects of family finance and public assistance.

The key columns of interest for our analysis include:

* FPERSONS: The average family size is approximately 2.34, with a range from 1 to 13 members (see figure 1).
* FANNVAL: The annual family income shows a wide distribution, with a significant standard deviation, indicating varied economic conditions among families. The mean and median values suggest disparities in income distribution, with many entries reporting zero income, potentially indicating non-reporting or non-earning families.
* POVLL: The poverty level line values range from -1 to 14, with a mean near 9.5, suggesting a broad spectrum of economic statuses among the surveyed families.
* FEDVAL: Education assistance income also varies greatly among families, with most reporting zero, pointing to limited access or eligibility for such assistance.
* FINC\_ANN: The total annual income from all sources primarily falls into two categories, reflected in the minimal standard deviation, suggesting a binary or categorical nature of this variable.
* FOWNU18: The number of own children under 18 in families averages around 0.55, with a standard deviation close to 1, indicating that many families have at least one child, but there's also a significant proportion without children in this age group.

**Measures of Variability and Central Tendency**

The dataset reveals insights into the socio-economic conditions of families, as reflected in the variables for family size (FPERSONS), annual family income (FANNVAL), and the poverty level line (POVLL). The average family size is approximately 2.34 members, with family sizes ranging from 1 to 13. The median family size aligns with the mean, indicating a balanced distribution. For annual family income, the vast majority of families report zero income in this dataset, likely reflecting non-respondents or non-earning units, with some reporting incomes as high as $396,000. This skewness in income data points to the significant economic disparities among the surveyed families. The poverty level line (POVLL), ranging from -1 to 14, further indicates diverse economic statuses (see Figure 1-3 in the Appendix).

**Frequency, Variance, and Standard Deviation**

Variance and standard deviation metrics for family size and income underscore the diversity within the dataset. The variance in family size (2.13) and standard deviation (1.46) suggests moderate variability among family units. In contrast, annual family income exhibits a high variance (approx. 24,896,140) and standard deviation (4989.60), reflecting substantial income disparities. The poverty level line's standard deviation (4.29) and variance (18.37) also highlight variability in economic well-being among the families (see Figure 1-3 in the Appendix).

**Outlier Detection and Distribution Modality**

Outlier analysis on annual family income using the Interquartile Range (IQR) method identified 1,691 families with income levels significantly above or below the median, with outliers' incomes ranging from $1 to $396,000, and a mean outlier income of approximately $15,116. This outlier presence underscores the significant income variation and potential economic extremes within the population. Such disparities suggest the importance of targeted economic policies and supports to address the needs of both low-income families and those experiencing substantial financial success.

These descriptive statistics set the stage for deeper analysis to explore the relationships between family size, income levels, poverty status, and the impact of having children or receiving education assistance on the economic well-being of American families. Further analysis will involve examining correlations, conducting hypothesis testing, and potentially building predictive models to understand these dynamics better.

**Correlation between the key columns**

The correlation matrix reveals the following noteworthy relationships among the variables (see Figure 4 in the Appendix). There appears to be a weak correlation between family size (FPERSONS) and total family income, suggesting that larger family sizes do not necessarily correlate strongly with higher or lower total family incomes. The correlation between the number of children under 18 (FOWNU18) and the poverty level line (POVLL) is not directly visible in the correlation matrix but warrants a deeper investigation due to the potential social and economic implications. The correlation with other financial metrics suggests that while there is some relationship with total annual income (FINC\_ANN), it may not be straightforward, highlighting the complexity of how different income sources interact within families.

**Impact of family size on total family income**

The scatterplot (see Figure 5 in the Appendix) examining the relationship between family size and total family income does not indicate a strong linear relationship. This suggests that while family size is an important demographic factor, it alone does not predict total family income levels, highlighting the complexity of economic conditions affecting families.

**Impact of children under 18 on poverty level**

The boxplot comparing the number of own children under 18 with the poverty level line illustrates variability in poverty levels across families with different numbers of children (see Figure 6 in the Appendix). This visualization suggests that as the number of children increases, there may be a tendency towards lower poverty levels, but the relationship is not uniform and suggests a nuanced interaction between family composition and economic status.

**Linear regression between family size and family income**

The linear regression analysis, aimed at exploring the relationship between family size (FPERSONS) and total family income (FANNVAL), resulted in a coefficient of determination (R2) of approximately -0.00023. This R2 value, being close to zero and slightly negative, indicates that the model does not effectively explain the variation in total family income based on family size alone.

This finding supports the earlier observation from the scatter plot and correlation analysis that there is no strong linear relationship between family size and total family income. It suggests that other factors, possibly not included in this simple model, play a significant role in determining the economic conditions of families. The complexity of family income dynamics cannot be captured fully by considering family size alone, underscoring the need for a multifaceted approach to understanding economic well-being among American families.

**Cluster Analysis**

A cluster analysis with k=3 (chosen using the Elbow Method) has successfully identified distinct groups within the dataset based on the standardized and PCA-reduced features related to family size, annual family income, education assistance income, total annual income, and the number of own children under 18 (see Figure 7-8 in the Appendix). The scatterplot visualizes these clusters, revealing the diversity in economic and demographic profiles among the families surveyed.

Each cluster represents a unique combination of the variables considered, suggesting different socio-economic segments within the American family population. These findings can inform more targeted analyses, potentially leading to insights on specific demographic groups' economic conditions and assistance needs.

This clustering approach demonstrates the potential to uncover patterns and relationships that may not be immediately apparent from traditional analysis methods, offering a valuable tool for socio-economic research and policy development.

**Results and Discussions**

The exploratory data analysis (EDA) of the ffpub20.csv dataset, structured around specific research questions (RQs) and hypotheses, has produced findings that offer insightful revelations into the socioeconomic dynamics of American families. Here's a synthesis of how the findings relate to the initial research questions and hypotheses:

**Relationship between family size and annual income**

The analysis aimed to uncover the relationship between the size of a family (FPERSONS) and its total annual income (FANNVAL). The hypothesis posited that there would be a significant correlation between these variables. However, the results indicated a lack of a strong linear relationship, as evidenced by the modest correlation coefficient and further supported by the linear regression analysis, which showed an inconsequential R² value. This outcome suggests that while family size may have some impact on family income, it is not a predominant factor influencing economic conditions, thereby leading to a rejection of the alternative hypothesis (H1) in favor of the null (H0).

**Impact of Children Under 18 on Poverty Levels**

Investigating whether the presence of children under 18 years old (FOWNU18) influences a family's poverty level (POVLL), the analysis anticipated finding a significant difference in poverty levels between families with and without children. The findings from the distribution analysis and outlier examination revealed variations in poverty levels, but the nuanced relationship highlighted by the cluster analysis suggests that the presence of children is one of several factors affecting poverty status. This complexity in the data points to a partial acceptance of the alternative hypothesis (H1), indicating that while children impact economic well-being, the effect is interwoven with other socioeconomic factors.

**Influence of Education Assistance on Family Income**

The third research question explored the impact of education assistance income (FEDVAL) on the total annual income of families (FINC\_ANN). The hypothesis suggested a significant positive relationship, anticipating that families receiving more educational assistance would report higher total incomes. The descriptive statistics and variance analysis for FEDVAL against FINC\_ANN demonstrated variability in how educational assistance correlates with family income. While some families with higher education assistance incomes also reported higher total incomes, the overall analysis suggests a complex interaction, likely moderated by other variables such as employment status and additional income sources. This result leads to a nuanced interpretation of the hypothesis, where education assistance contributes to family income but is part of a broader matrix of socioeconomic factors.

**Conclusion**

The EDA has revealed the intricate relationships between family composition, economic assistance, and income levels. The absence of strong relationships where hypothesized underscores the multifactorial nature of economic well-being among American families. These insights emphasize the need for multifaceted policy approaches that consider the diverse needs and configurations of families. Moreover, the identification of significant disparities and the complex interplay of factors affecting poverty levels and income underscore the critical need for targeted interventions that address the root causes of economic inequality. Through this analytical lens, it becomes clear that enhancing family economic conditions requires a comprehensive understanding of the various dimensions that influence socioeconomic status, advocating for policies that are as nuanced and multifaceted as the lives of those they aim to improve.

**References**

Blake, J. (1981). Family size and the quality of children. Demography, 18(4), 421-442.

Downey, D. B. (1995). When bigger is not better: Family size, parental resources, and children's educational performance. American Sociological Review, 60(5), 746-761.

Duncan, G. J., & Brooks-Gunn, J. (Eds.). (1997). Consequences of growing up poor. Russell Sage Foundation.

Dynarski, S. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. American Economic Review, 93(1), 279-288.

Edin, K., & Lein, L. (1997). Making ends meet: How single mothers survive welfare and low-wage work. Russell Sage Foundation.

Grogger, J., & Karoly, L. A. (2005). Welfare reform: Effects of a decade of change. Harvard University Press.

Hout, M. (2012). Social and economic returns to college education in the United States. Annual Review of Sociology, 38, 379-400.

Mayer, S. E. (1997). What money can't buy: Family income and children's life chances. Harvard University Press.

Moffitt, R. (1992). Incentive effects of the U.S. welfare system: A review. Journal of Economic Literature, 30(1), 1-61.

Thomson, E., Hanson, T. L., & McLanahan, S. S. (1994). Family structure and child well-being: Economic resources vs. parental behaviors. Social Forces, 73(1), 221-242.

**Appendix**

A screenshot of a computer

Description automatically generated

Figure 1: Variability and central tendency of the key columns in this analysis.

A graph of a number of persons in family sizes

Description automatically generated

Figure 2: Distribution of Family Sizes: A histogram of family sizes showing the frequency of various family sizes.

A graph with a number of bars

Description automatically generated with medium confidence

Figure 3: Poverty Level Line (POVLL) Distribution: A depiction of the distribution of the poverty level lines across the datasets.

A screenshot of a graph

Description automatically generated

Figure 4: Heatmap of the Correlation Matrix among Key Socioeconomic Variables

A graph of a family size

Description automatically generated

Figure 5: Scatter plot showing the impact of family size on total family income.

A graph of different colored rectangular shapes

Description automatically generated

Figure 6: Boxplot showing the relationship between the number of children under 18 in the family and poverty level.

A graph with a line

Description automatically generated

Figure 7: Elbow Method for determining optimal number of clusters, showcasing the point where the within-cluster sum of squares (WCSS) begins to decrease at a slower rate, indicating the efficient cluster count.

A graph of a cluster analysis

Description automatically generated

Figure 8: Cluster Analysis Visualization with k=3, revealing distinct groupings based on socio-economic and demographic characteristics within the dataset.