In recent years, affordable childcare has become harder to come by. As childcare costs rise, some parents face an immense financial burden –– at times spending a large portion or most of their yearly salaries on the service. This country’s childcare infrastructure also affects the U.S. economy, costing it roughly $57 billion every year in lost earnings, productivity, and revenue, according to a report published by ReadyNation.[[1]](#footnote-1) The unaffordable nature of childcare has also pushed some parents to rely on programs like food stamps or welfare.[[2]](#footnote-2) And in Maryland alone, the average annual cost of childcare is $15,335, which costs approximately 65.1% more per year than in-state tuition at a public four-year college.[[3]](#footnote-3)

This simple linear regression analysis, which intends to inform parents and policymakers interested in solving the childcare crisis, examines the relationship between a person’s age and the number of children they have. More specifically, I hypothesized that the age (IV) and number of children (DV) variables would have a non-zero correlation. This memo will explain the results of this investigation, showing that while there is a non-zero correlation between these variables, this relationship is not practically significant.

Table

Description automatically generatedThis analysis used a nationally representative dataset extracted from the General Social Survey, which included three variables: year, age, and number of children. The survey responses span from 1972 to 2018. Prior to performing the analysis, additional data preparation was conducted. Text responses such as “No answer” and “Don’t know” were removed alongside missing values. Some responses were also recoded to convert the original character variables into numeric ones. Under the age variable, all responses for “89 or older” were recoded to “89.” Responses for “Eight or more” under the number of children variable were recoded to “8.” After this step, I chose the alpha of 0.01 and began the analysis by observing the data’s summary statistics.

Next, I performed a correlation test, ultimately finding that the age and number of children variables have a weak positive correlation. After this, I went on to perform the linear regression analysis and checked the assumptions. Based on the diagnostic plots alone, I could tell that the data violated some of the assumptions. For example, the Residuals vs. Fitted plot shows that the fitted line is not exactly straight or flat, thus violating the first assumption of linearity. The Normal Q-Q plot also shows that the data is not equally distributed, violating the third assumption.

Table

Description automatically generated After checking each assumption, I further analyzed the linear regression results, R2 = 0.13, F(1, 64400) = 9550, p < 0.01, which showed that age as a predictor variable is statistically significant. The age variable’s p-value is less than the chosen alpha of 0.01, thus allowing me to reject the null hypothesis stating that the variables have a correlation of zero. In addition, I also examined the coefficient, which was very low at 0.04, indicating a weaker relationship between the variables. In other words, a one-unit change in age corresponds to an average 0.04 change in number of children.

Regression Model Equation: ŷ = 0.26 + 0.04(x)

Alongside analyzing the above values, I also examined the model’s r-squared value, which was 0.13. This r-squared value suggest that the predictor variable (age) only explains about 13% of the response variable’s (number of children) variability. Since this analysis suggests that the results are not practically significant, I chose to also test the model’s effect size and power to verify this. The model’s calculated power of 1 shows that the probability of committing a Type II error is very low. The effect size, which was 0.15, shows a small-medium effect size further indicating limited practical applications.

In conclusion, while my findings are statistically significant at the selected alpha level of 0.01, the overall findings of this analysis are not substantively significant, meaning they’re not meaningful for real-world application. Although this analysis did not find a strong relationship between age and number of children, the issue of unaffordable childcare remains and continues to affect parents across the country. The argument can be made that a weaker relationship between these variables indicate that this issue may affect a diverse, wide range of people, making this a more significant problem to solve. A potential weakness in this study is that the large sample size may have influenced the statistical significance. In the future, to better analyze this relationship, I recommend focusing on a specific age group such as, 18-25 or 25-35 or subsetting the data to solely focus on a one or two-year period. Additionally, removing any outliers in the data will also improve the credibility of results overall.

Table

Description automatically generatedDiagram, engineering drawing

Description automatically generated

1. Bishop-Josef, S., Beakey, C., Watson, S., Garrett, T. (2019, January 16). Want to Grow the Economy? Fix the Child Care Crisis. *ReadyNation*. <https://www.strongnation.org/articles/780-want-to-grow-the-economy-fix-the-child-care-crisis> [↑](#footnote-ref-1)
2. Reilly, K. and Luscombe, B. (n.d). The Childcare Crisis. *Time*. Retrieved November 30, 2021, from <https://time.com/child-care-crisis/> [↑](#footnote-ref-2)
3. Economic Policy Institute. (2020 October). *Child care costs in the United States*. [https://www.epi.org/child-care-costs-in-the-united-states/ - /MD](https://www.epi.org/child-care-costs-in-the-united-states/#/MD) [↑](#footnote-ref-3)