**To:** Abdirisak Abdullahi Mohamed, Janzen Consulting Group

**From:** Amanda Hernandez

**Date:** 12/12/21

**Subject:** The Childcare Crisis: Analyzing the Relationship Between Number of Children, Age, Race, Total Family Income & Number of Hours Worked in a Week

In recent years, affordable childcare has become harder to come by — an issue further exacerbated by pandemic-related economic shortcomings and layoffs. Since the COVID-19 pandemic's emergence in 2020, the childcare crisis has forced many people, mostly women, to leave their jobs and contributed to a labor shortage across various industries.[1] The National Partnership for Women and Families also found that fewer women in the workforce costs the U.S. economy approximately $650 billion every year[2] — representing a dire issue requiring careful consideration and thoughtful solutions.

This multiple regression analysis, which intends to inform parents and policymakers interested in solving the childcare crisis, examines the relationship between a person's age, race, total family income, number of hours worked last week (IVs) and the number of children (DV) they have. I hypothesized that the independent variables would influence the dependent variable. This memo will explain the results of this investigation, showing that while the regression analysis indicates a weak relationship between the independent variables and the dependent variable, this relationship is not practically significant.

This analysis used a nationally representative dataset extracted from the General Social Survey, which included six total variables: year, age, race, total family income, number of hours worked last week, and number of children. The survey responses span from 1973 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 any missing values. Some responses were also recoded to convert the original character variables into numeric or factor variables. 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." Text responses that remained under the total family income variable were recoded into income brackets, "$25,000 or more" or "Less than $1000". Lastly, the race and total family income variables were coded into dummy variables to allow for thorough and proper

**Table 1: Quantitative Summary**

| | Freq. | Min | Max | Med. | Mean | SD |
|---|---|---|---|---|---|---|
| Age | 37,304 | 18 | 89 | 39 | 40.78 | 13.1 |
| Number of children | 37,304 | 0 | 8 | 2 | 1.69 | 1.6 |
| Number of hours worked last week | 37,304 | 0 | 89 | 40 | 41.29 | 14.17 |

[1] Ho, S. and Boak, J. (2021, October 27). Exacerbated by pandemic, child care crisis hampers economy. *The Associated Press*. https://apnews.com/article/coronavirus-pandemic-business-pandemics-economy-seattle-17db11434b2bb82a3d7fb3fc82003514

[2] National Partnership for Women & Families. (2021 July). *The Cost of Inaction: How a Lack of Family Care Policies Burdens the U.S. Economy and Families.* https://www.nationalpartnership.org/our-work/resources/economic-justice/other/cost-of-inaction-lack-of-family-care-burdens-families.pdf

analysis. After completing the necessary data preparation, I chose the alpha of 0.01 and began the analysis by observing the data's summary statistics.

Next, I performed the multiple regression analysis and checked the assumptions against the diagnostic plots. The data violated almost every assumption. For example, the Residuals vs. Leverages plot showed points passing beyond the Cook's Distance threshold, indicating influential outliers. After checking the assumptions, I moved on to analyzing the regression results, $R^2 = 0.18$, $F(15, 37288) = 553.3$, $p < 0.01$, which showed that the model as a whole is statistically significant. The p-values for most of the independent variables were less than the chosen alpha, thus allowing me to reject the null hypothesis in favor of the alternative for those variables. The variables that were not found to influence number of children or have meaningful application, however, were number of hours worked last week and the first income bracket, $1000-$2999. The model's equation and interpretations of each independent variable, holding all other variables in the model constant, can be found below.

**Table 2: Categorical Summary**

| **Race** | Freq. | Percent |
|---|---|---|
| White | 29,939 | 80% |
| Black | 5,028 | 13% |
| Other | 2,337 | 6% |
| **Total** | 37,304 | 100% |
| **Total family income** | | |
| Less than $1,000 | 2,232 | 6% |
| $1,000 - $2,999 | 357 | 1% |
| $3,000 - $3,999 | 353 | 1% |
| $4,000 - $4,999 | 338 | 1% |
| $5,000 - $5,999 | 426 | 1% |
| $6,000 - $6,999 | 450 | 1% |
| $7,000 - $7,999 | 502 | 1% |
| $8,000 - $9,999 | 979 | 3% |
| $10,000 - $14,999 | 3,326 | 9% |
| $15,000 - $19,999 | 2,964 | 8% |
| $20,000 - $24,999 | 3,324 | 9% |
| $25,000 or more | 22,053 | 59% |
| **Total** | 37,304 | 100% |

Regression Model Equation: $\hat{y} = -0.41 + 0.05(age) + 0.0(weekly\_hrs) + 0.15(race.b) - 0.27(race.w) + 0.13(income.1) + 0.29(income.2) + 0.30(income.3) + 0.27(income.4) + 0.07(income.5) + 0.32(income.6) + 0.36(income.7) + 0.17(income.8) + 0.35(income.9) + 0.27(income.10) + 0.32(income.11)$

Reference Groups: race.o (other) and income.12 (less than $1,000)

For the following interpretations, model variables not explicitly stated are held constant. A one-unit change in age corresponds to a 0.05 average increase in number of children. A one-unit change in number of hours worked last week, on the other hand, corresponds to a zero average change in number of children. Under the race variable, those who identify as black correspond to an average 0.15 rank increase in number of children compared to those who identify as other. Those who identify as white correspond to an average 0.27 rank decrease in number of children compared to those who identify as other.

Now transitioning into the total family income variable, those who fall under the $1,000-$2,999 income bracket corresponds to an average 0.13 rank increase in number of children compared to those who fall under the less than $1,000 income bracket. Those who fall under the $10,000-$14,999 income bracket corresponds to an average 0.29 rank increase in number of children compared to those who fall under the less than $1,000 income bracket. Those who fall under the $15,000-$19,999 income bracket corresponds to an average 0.30 rank increase in number of children compared to those who fall under the

less than $1000 income bracket. Those who fall under the $20,000-$24,999 income bracket corresponds to an average 0.27 rank increase in number of children compared to those who fall under the less than $1,000 income bracket. Those who fall under the $25,000 or more income bracket correspond to an average 0.07 rank increase in number of children compared to those who fall under the less than $1,000 income bracket. Those who fall under the $3,000-$3,999 income bracket corresponds to an average 0.32 rank increase in number of children compared to those who fall under the less than $1,000 income bracket. Those who fall under the $4,000-$4,999 income bracket corresponds to an average 0.36 rank increase in number of children compared to those who fall under the less than $1,000 income bracket. Those who fall under the $5,000-$5,999 income bracket corresponds to an average 0.17 rank increase in number of children compared to those who fall under the less than $1,000 income bracket. Those who fall under the $6,000-$6,999 income bracket corresponds to an average 0.35 rank increase in number of children compared to those who fall under the less than $1,000 income bracket. Those who fall under the $7,000-$7,999 income bracket corresponds to an average 0.27 rank increase in number of children compared to those who fall under the less than $1,000 income bracket. Those who fall under the $8,000-$9,999 income bracket corresponds to an average 0.32 rank increase in number of children compared to those who fall under the less than $1,000 income bracket.

Although most of the independent variables were found to be statistically significant, the coefficients of every independent variable are very small, suggesting that none are meaningful because they show a very weak relationship with number of children. I also examined the model's adjusted r-squared value, which was 0.18. This r-squared value suggests that the predictor variables only explain about 18% of the response variable's (number of children) variability. Since this analysis suggests that the results are not practically significant, I also tested 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 0.22, indicated a medium effect size and limited practical applications.

In conclusion, while the model is statistically significant at the selected alpha level of 0.01, the findings of this analysis are not substantively significant or meaningful for real-world application. Although this analysis did not find a strong relationship between the independent variables and number of children, it shows that the issue of unaffordable childcare affects people of all backgrounds. A potential weakness in this study is that the large sample size of 37,304 may have influenced the statistical significance. For future analysis, I recommend subsetting the data to a particular year or period of years. Additionally, removing influential outliers in the data will also improve the credibility of results overall.

**Multiple Linear Regression**

| | *Dependent variable:* |
|---|---|
| | num_children |
| age | 0.052*** |
| | (0.001) |
| weekly_hrs | 0.001 |
| | (0.001) |
| race.b | 0.153*** |
| | (0.036) |
| race.w | -0.267*** |
| | (0.031) |
| income.1 | 0.130 |
| | (0.083) |
| income.2 | 0.286*** |
| | (0.040) |
| income.3 | 0.297*** |
| | (0.041) |
| income.4 | 0.271*** |
| | (0.040) |
| income.5 | 0.074** |
| | (0.032) |
| income.6 | 0.321*** |
| | (0.083) |
| income.7 | 0.355*** |
| | (0.085) |
| income.8 | 0.170** |
| | (0.077) |
| income.9 | 0.350*** |
| | (0.075) |
| income.10 | 0.268*** |
| | (0.072) |
| income.11 | 0.323*** |
| | (0.056) |
| Constant | -0.414*** |
| | (0.053) |
| Observations | 37,304 |
| $R^2$ | 0.182 |
| Adjusted $R^2$ | 0.182 |
| Residual Std. Error | 1.451 (df = 37288) |
| F Statistic | 553.255*** (df = 15; 37288) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |