

**Date:** May 14, 2019

**To:** Leana Wen, President, Planned Parenthood

**From:** Julian Vazquez, Junior Data Analyst, Janzen Consulting Group

**RE:** Effect of few selected socio-economic factors in number of children in the US

According to a report from the Census Bureau, individuals with a higher level of education have fewer children than those with a lower level of education<sup>1</sup>. This is significant, because with more people going to college than ever before, this relationship between education and childbearing could potentially change the traditional family structure in the United States. Similarly, other socio-economic indicators, such as race, political views, and perception of income may be having an affect in the number of children a person ends up having.

Fertility in the United States has been steadily decreasing in the past few decades<sup>2</sup>, and several factors may be playing a role in this trend. I hypothesize that education level, race, political views and perception of one's income can help explain the number of children an individual ends up having. To evaluate if this holds true, I will focus my efforts on studying adults in the United States. Analysis results show that some of these factors can indeed help predict the number of children someone has.

The data for the following analysis comes from the General Social Survey (GSS) dataset, a well-respected survey that provides policy makers with a clear and unbiased perspective of what Americans

	<b>Table 1: Summary Statistics</b>					
	Freq	Min	Max	Med.	Mean	SD
Education	2858	0.0	20.0	13.0	13.7	3.0
No. of Children	2859	0.0	8.0	2.0	1.9	1.7

think and feel about a wide range of issues<sup>3</sup>.

The study will focus on five variables:

education level, race, political views,  
perception of one's income and number of

children. Education level, defined by the GSS as years of education, ranges from 0 to 20, showing spikes in the 12 (high school) and 16 (college) units. Number of children ranges from 0 to "Eight or more", which was redefined as simply "8" for analysis purposes. Very few observations fell in this category, therefore it should not have a significant effect on the findings. Race, which originally included 16 categories, was reduced to 5 broader categories for simplicity: white, black, Hispanic, Asian and other. Political views included: extremely liberal, liberal, slightly liberal, moderate, slightly conservative, conservative and extremely conservative. To measure perception of one's income, people were asked to

---

<sup>1</sup> Wetzstein, C. (2011). Education level inversely related to childbearing. The Washington Times. Retrieved from <https://www.washingtontimes.com/news/2011/may/9/education-level-inversely-related-to-childbearing/>

<sup>2</sup> Chappell, B. (2019). U.S. Births Dip To 30-Year Low; Fertility Rate Sinks Further Below Replacement Level. NPR. Retrieved from <https://www.npr.org/sections/thetwo-way/2018/05/17/611898421/u-s-births-falls-to-30-year-low-sending-fertility-rate-to-a-record-low>

<sup>3</sup> About the GSS (n.d.). The General Social Survey. Retrieved from <http://gss.norc.umd.edu/>

Table 2: Categorical Summary		
Race	Freq.	Percent
White	2088	73.44%
Black	487	17.13%
Hispanic	103	3.62%
Asian	90	3.17%
Other	75	2.64%
Total	2843	100%
High Income		
Strongly agree	53	5.67%
Agree	208	22.27%
Neither	259	27.73%
Disagree	318	34.05%
Strongly disagree	96	10.28%
Total	934	100%
Political Views		
Extremely liberal	136	4.93
Liberal	350	12.70
Slightly liberal	310	11.25
Moderate	1032	37.45
Slightly conservative	382	13.86
Conservative	426	15.46
Extremely conservative	120	4.35
Total	2756	100%

$$\text{No. of Children} = 2.34 - 0.087 * \text{Years of Education} + 0.03 * \text{High Income} + 0.09 * \text{Political Views} + 0.11 * \text{Black} + 0.27 * \text{Hispanic} + 0.29 * \text{Asian} + 0.54 * \text{Other} + \varepsilon$$

mark which option corresponded to their opinion regarding the statement “My income is high”. Response options included: strongly agree, agree, neither, disagree, and strongly disagree. Only 934 out of 2867 survey respondents answered this question.

Multiple linear regression analysis was used to test our hypothesis, with an alpha of 0.05. There were no significant violations of assumptions. However, the observations deviated slightly from the QQ plot line at the tails, which could mean issues with linearity, but tracked along nicely in the middle. Based on the patterns found in the Scale-Location plot there might be some homoscedasticity issues, but these patterns could be attributed to the nature of our categorical variables. The results of the regression indicated that the model explained 4.1% of the variance ( $R^2 = .041$ ,  $F(7, 892) = 6.475$ ,  $p < .001$ ). Given our alpha, the model is statistically significant, which is why we reject the null hypothesis. For the following interpretations it is assumed that all other variables remain constant. Our education coefficient was highly statistically

significant, and estimates that each additional year of education corresponds to an

average decrease of 0.09 in number of children. In other words, every 7.7 additional years of education correspond to one less child, which is meaningful<sup>4</sup>. Our political views coefficient is significant as well, and predicts that every 1-rank change towards extreme conservatism results in an average of 0.09 increase in number of children. Like our education coefficient, this is meaningful but not as much because even a radical change of 6 ranks in political views would not result in one more or less children. Race was not statistically significant. However, the coefficient is meaningful for the “Other” category when measured against the “white” category (reference group), resulting in a 0.54 increase in number of children. Perception of one’s income indicated a 0.03 increase in number of children for each 1-rank

<sup>4</sup> Vazquez, J. (2019). Project 3.

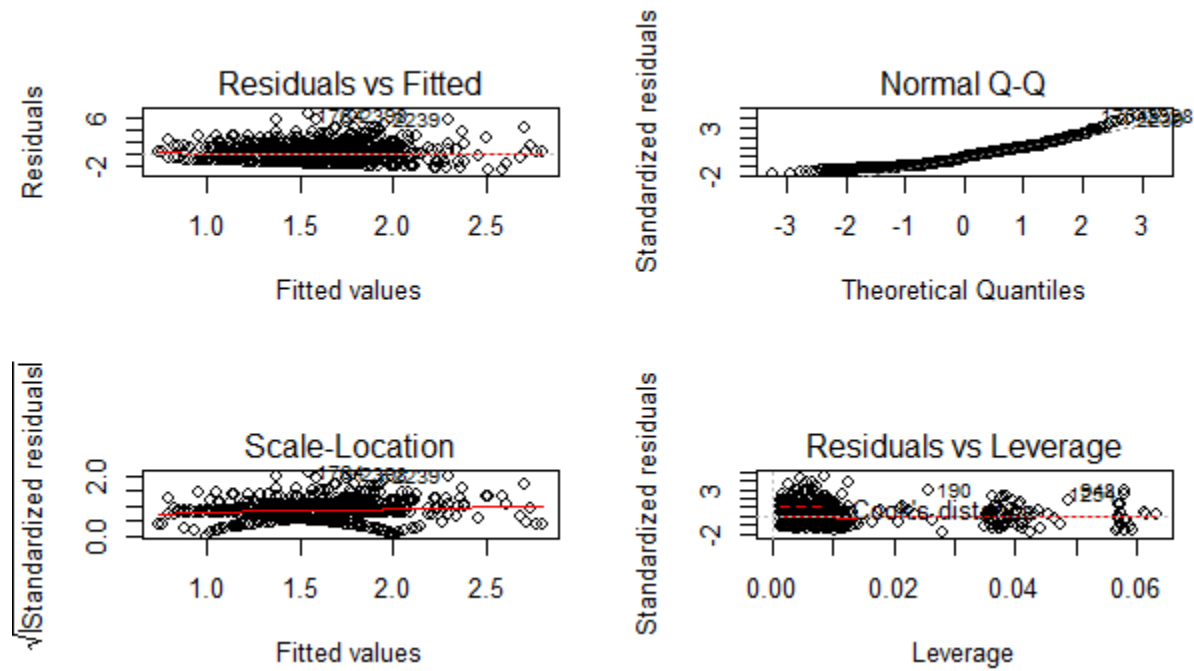
change towards “Strongly disagree” about considering one’s income to be high. But it was not statistically significant either, and it is not meaningful because of its extremely small coefficient. Because our data comes from the GSS dataset and is representative of adults living in the US, we are able to generalize our findings to the population.

One potential weakness of this study is the fact that it does not include individuals who reported having more than eight children. However, since this is such a small portion of the population, I do not expect my results to have deviated significantly if I had included this data. The same applies to the fact that the data set caps at 20 years of education, since very few people go beyond that. Another observation is that even though the General Social Survey (GSS) is meant to be representative of American adults, some racial minorities may be underrepresented. In this case, having broad race categories would make it harder to generalize the results to the population. Lastly, another weakness of this study is the fact that some of the independent variables are not statistically significant, which brings down the coefficient of determination of the model, and thus its ability to predict number of children.

# Multiple Regression Results

	<i>Dependent variable:</i>
	No. of Children
Education	-0.087 <sup>***</sup> (0.017)
High Income	0.025 (0.046)
Political Views	0.089 <sup>***</sup> (0.034)
Black	0.107 (0.131)
Hispanic	0.270 (0.292)
Asian	0.286 (0.280)
Other	0.537 (0.353)
Constant	2.337 <sup>***</sup> (0.350)
Observations	900
R <sup>2</sup>	0.048
Adjusted R <sup>2</sup>	0.041
Residual Std. Error	1.472 (df = 892)
F Statistic	6.475 <sup>***</sup> (df = 7; 892)
<u>Note:</u>	* p<0.1; ** p<0.05; *** p<0.01

## Diagnostic plots



## Closer look at Scale-Location plot

