Evaluation Study on Women, Infant and Children Nutrition Program and Learning Capability

Jie Xu

MSSP897 Applied Linear Modeling

School of Social Policy & Practice

2021 Spring

Evaluation Study on Women, Infant and Children Nutrition Program and Reading Achievement

Abstract: Nutrition and family income are regarded as critical to young children development. Therefore, an evaluation study is designed to examine the effect of *Women, Infant and Children (WIC) Nutrition Program* participation during pregnancy on child reading achievement. More than three thousand samples involving in the study, participation of the program and other demographic variables turned out to have different levels of impact on reading capability. Humanism should play a more important role, rather than economic effectiveness, in the policy development of this social issue.

Key words: nutrition, family income, children development

I. Background Information

Similar as Duncan & Brooks-Gunn (1997) pointed out in *the effects of poverty on children*, it has long been recognized that children living in poverty lag behind other children on a wide range of indicators of physical, mental, academic, and economic well-being. They are more likely to have health, behavioural, learning, and emotional problems, and situation maybe worse for children in families experiencing deep poverty. Shonkoff & Garner (2012) found that under the intervention of strengthening biological systems at early stage, children are more likely to thrive and grow up to be healthy. Sound health in early childhood provides a foundation for the construction of sturdy brain architecture and the achievement of a broad range of skills and learning capacities. Horino & Yang (2020) also suggested that good nutrition during pregnancy is especially important to support foetal development, protect mothers from pregnancy-related risks, and promote children development which may carry over into adulthood.

However, Cameron et al. (2015) pointed out that low-income women and their children may be especially susceptible to the risks of poor nutrition in the early stages of life. A federal program, the *Special Supplemental Nutrition Program for Women, Infants, and Children*, known as the WIC Program, is developed to dealt with this social issue. The WIC program serves to safeguard the health of WIC people by providing nutritious foods and support[1], targeted at the direct help of better nutrition for both the pregnant women and embryo in her abandum. Theoretically, the program could help children build stronger physical condition at the early stage of their life, and make compensation for each other.

II. Descriptive Statistics

Table 1 contains frequency, mean and standard deviation for categorical variables. 829 children of 2042 valid samples are in low birth weight status. 851 people out of 2042 valid data participated in Women, Infant and Children (WIC) Nutrition Program during pregnancy. According to the value definition of 'No' for 0 and 'Yes' for 1, the proportion of positive results can be easily read through mean of the data, that 40.6% of the samples are born with low weight and nearly 41.7% participated in WIC program.

Table 2 contains minimum, maximum, median, mean and standard deviation for quantitative variables. The range of Woodcock-Johnson Revised Mathematics Achievement Test Raw Score is from 0 to 98. The median and mean of the test score is

[1] U.S. Department of Agriculture, *Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)*, retrieved from https://www.fns.usda.gov/wic.

Table 1.

CATEGORY AND FREQUENCY, MEAN AND STANDARD DEVIATION FOR CATEGORICAL VARIABLES

Variables	Frequency		Mean	SD
Low birth weight status of the child	0	1213	0.406	0.4911999
	1	829		
Women, Infant and Children (WIC) Nutrition Program participant during pregnancy	0	1191	0.4167	0.4931412
	1	851		

Table 2.

MINIMUM, MAXIMUM, MEDIAN, MEAN AND STANDARD DEVIATION FOR QUANTITATIVE VARIABLES

Variables	Min	Max	Median	Mean	SD
Woodcock-Johnson Revised Mathematics Achievement Test Raw Score	0.00	98.00	36.00	35.83	22.28801
The child's age in 1997	3.000	13.000	7.000	7.395	2.922137
Total family income in 1997	0	784611	40705	52060	53175.04
A composite total score of the emotional and cognitive stimulation at home	7.90	27.00	20.50	20.19	3.068104

respectively 36 and 35.83, which refers its distribution skewed to the left. The children's age in 1997 ranges from 3 to 13, and its median locates at 7. The mean is 7.395, quite close to the median, and the standard deviation is a small number of 2.922137. The range of total family income in 1997 is from 0 to 784611 dollars, with the respective median and mean of 40705 and 52060. The standard deviation, as large as 53175.04, together with its separated median and mean, show that the data is decentralized to some extent, or maybe contains certain outliers affecting the outcomes. The composite total score of the emotional and cognitive stimulation at home ranges from 7.9 to 27, with the mean of 20.19 and the small standard deviation of 3.068104. The median of

Table 3.

CORRELATION MATRIX FOR ALL INVESTIGATED VARIABLES (N=1848)

Variable	WICpreg	faminc97	RACE	AGE97
WICpreg	—			
faminc97	-0.38632775	—		
RACE	-0.49050419	0.36433193	—	
AGE97	-0.08900043	0.05260234	-0.01611733	—

WICpreg=Women, Infant and Children Nutrition Program participant during pregnancy; faminc97=Total family income in 1997; RACE=Centered Binary Coding of Race; AGE97=The child's age in 1997

the data is 20.5, larger than its mean, showing that its distribution is skewed to the left.

Table 3 presents Pearson correlations among the investigated variables. All the coefficients range approximately from -0.5 to 0.5, showing a low correlation among all the variables. Taking the correlation between *faminc97* and *AGE97* as an example, the family income and the age of the children in 1997 have a positive and relatively stronger relationship.

III. Multiple Regression Analyses

1. Original Regression Model

The original regression model research on child math achievement (mathraw97) over and above the child's age in 1997 (age97), family income (faminc97), low birth weight status (bthwht) and participation in WIC program during pregnancy (WICpreg).

Table 4 provides the total variation of child math achievement accounted for by the model. A total of 2042 observations are used to estimate the model. R reports a result about the F-statistic, 3257 on 4 and 2037 degree of freedom, and a p-value of < 2.2e-16, which is essentially 0. Since the p-value is less than 0.05, we can conclude that at

Table 4.

MULTIPLE REGRESSION RESULT FOR THE ORIGINAL MODEL

Variables	Est	SE	p
(Intercept)	-1.555e+01	5.863e-01	< 2e-16 ***
AGE97	7.013e+00	6.394e-02	< 2e-16 ***
faminc97	3.238e-05	3.707e-06	< 2e-16 ***
bthwht	-2.149e+00	3.826e-01	2.22e-08 ***
WICpreg	-3.113e+00	4.023e-01	1.59e-14 ***
Residual Min/Max	-39.478/28.613	Residual SE	8.204
Multiple R-squared	0.8648	Adjusted R-squared	0.8645
F-statistic	3257	df	4 and 2037
p-value	< 2.2e-16		

*p<0.05 **p<0.01 ***p<0.001

AGE97=The child's age in 1997; faminc97=Total family income in 1997; bthwht=Low birth weight status of the child; WICpreg=Women, Infant and Children (WIC) Nutrition Program participant during pregnancy

least one of the independent variables helps explain the variation in the dependent variable (mathraw97). The two R-squared values tell that about 86.5% of the variance in *mathraw97* is explained by independent variables.

According to the output of coefficients, we can conclude individual effects of each variable. The regression intercept indicates that the average math test score (mathraw97) is -15.6 when child age (AGE97), total family income (faminc97), low birth weight status (bthwht) when WIC program participation during pregnancy (WICpreg) are zero. The standardized parameter estimate for *AGE97* is largest among the four variables, which means the age growth have the strongest effect on math test scores. In detail, a child's math score will increase by 7.0 points on average for every additional year of age, holding all other independent variables constant. A child's math score is estimated to increase 0.03 points for every thousand dollars increase of total family income, holding all other independent variables constant. Children with low birth weight are predicted to have math scores that are 2.15 points lower on average than children born with a normal weight, holding all other independent variables constant. Children participated in WIC Nutrition Program during pregnancy are predicted to have math scores that are 3.11 points lower on average than children not involved in the program, holding all other independent variables constant.

2. Assumption Diagnostics & Violation Corrections

The general linear model assumes linear relationships between its independent and dependent variables. Figure 1 tells the rough linearity correlations of the continuous variables that child age in 1997 (AGE97) and family income in 1997 (faminc97) could be linearly associated with child mathematics achievement test score (mathraw97).

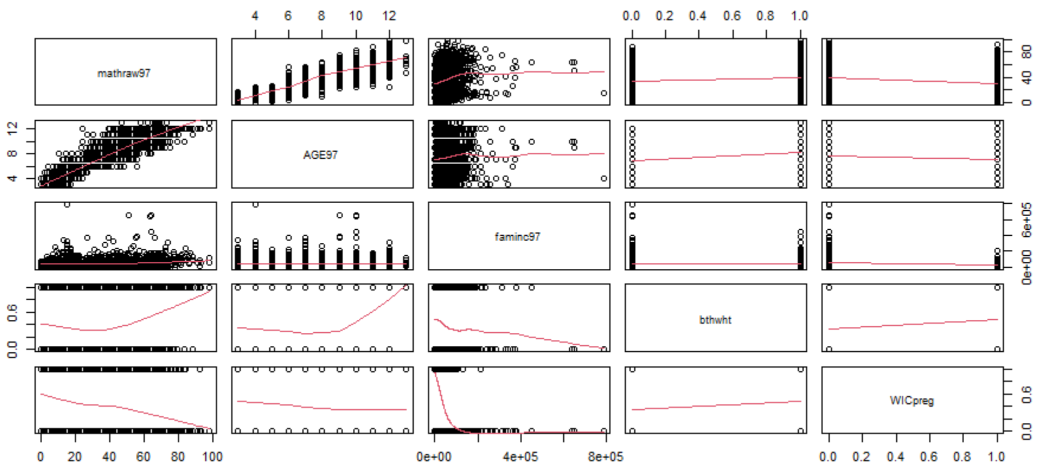

Figure 1. Scatterplot matrix of all variables

Figure 2 shows a more detailed relationship among the variables that neither child age nor family income are exactly linearly associated with the math achievement, both of which fail to meet the linearity assumption.

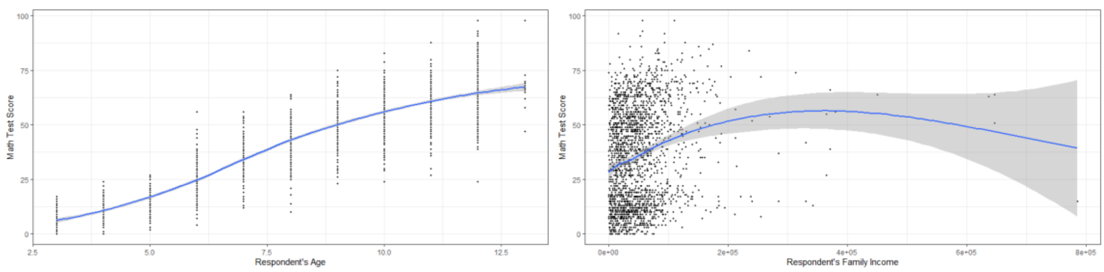

Figure 2. Scatterplots of Math Test Scores against Respondents' Age and Family Income

It is assumed that the variance of the model residuals is constant for each value of the independent variables and the mean value for the residuals is zero. As Figure 3 shows, the residuals of the model could be spread randomly and form an approximate horizontal band around the 0 line. The variance of the model residuals is constant for each independent variable, which is in line with the homo-scedasticity assumption.

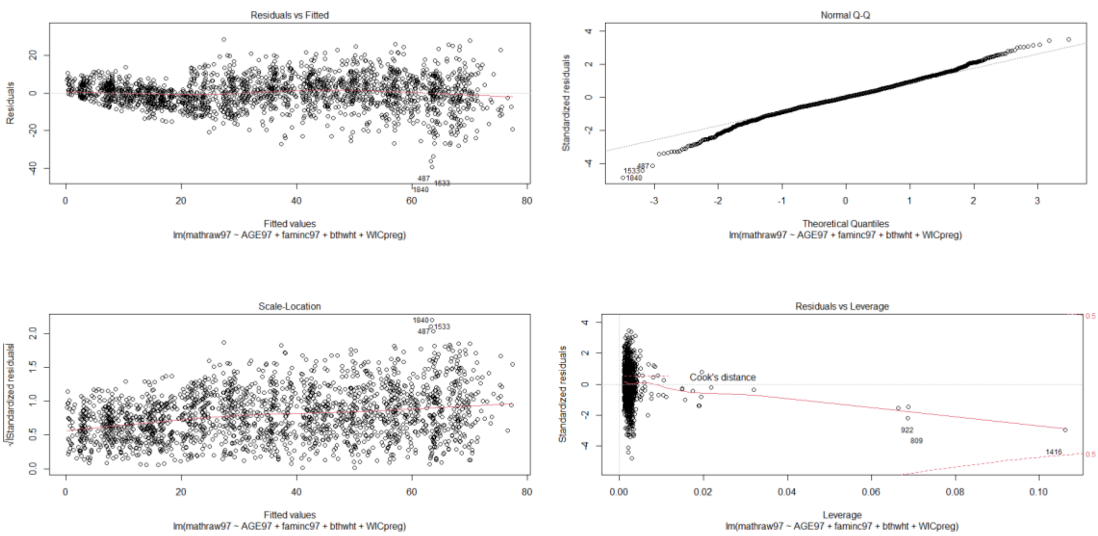

Figure 3. Residual plots of the original linear regression model

Residuals are assumed to be normally distributed, centred around a mean of 0. Figure 4 shows quite ideal distributions that data are normally distributed and centred around 0. Most of the residuals are quite small, and there is little difference between predicted and observed values, which shows the model fits the rule of normality.

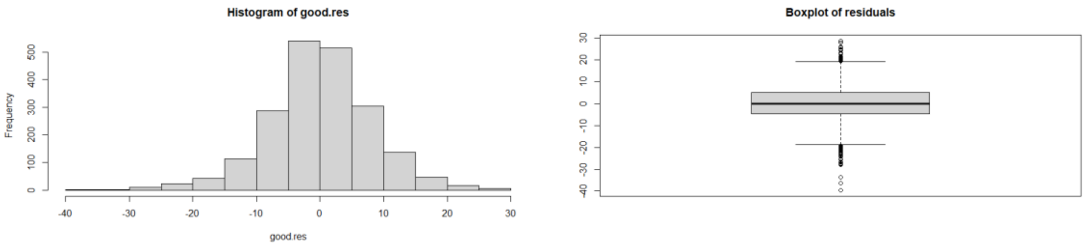

Figure 4. Histogram and boxplot of the original linear regression model residuals

Parenting practices (HOME97) is doubted as a relevant variable for the model. Therefore, two regression models are built for the examination: One regresses *mathraw97* on *AGE97*, *faminc97*, *bthwht* and *WICpreg*, while the other regresses *HOME97* on *AGE97*, *faminc97*, *bthwht* and *WICpreg*. Residuals of both models are plotted as Figure 5, density of *model_home* residual being larger and skewed (to the left). Figure 6 plots the residuals from the initial model against residuals from the model with *HOME97* as the dependent variable. Since the regression and lowess lines are not horizontal, there is reason to suspect that *HOME97* is an omitted variable.

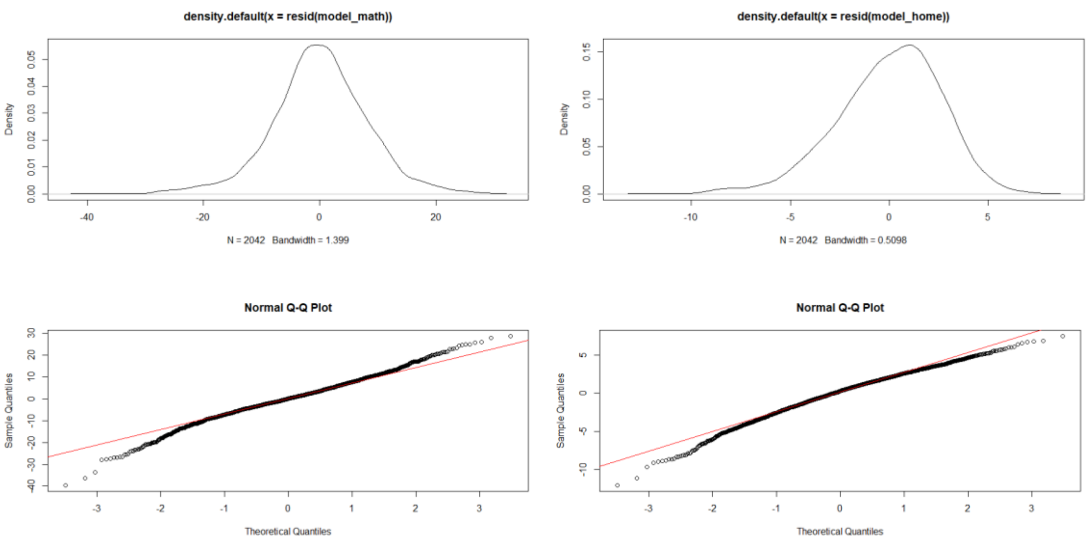

Figure 5. Probability and quantile-quantile plots of *model_math* and *model_home*

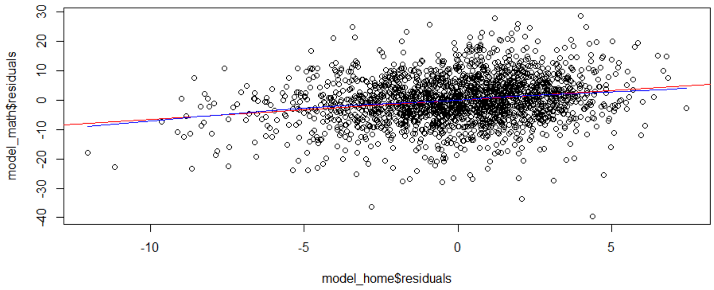

Figure 6. Scatterplot of *model_math* and *model_home* residuals

Based on the previous diagnostics, it looks like there may be problems with the variables *AGE97* and *faminc97*. The problem could be fixed by re-specifying the model: *faminc97* should be log-transformed, while *AGE97* should be centred, subtracting the mean value from each observation so that its new mean is 0, its squared term also included. Figure 7 tells the results of the linearity regression diagnostics, that both the newly involved variable *HOME97* and the log-transformed variable *logfaminc* went nearly horizonal, which shows both variables are qualified for the specified linear model.

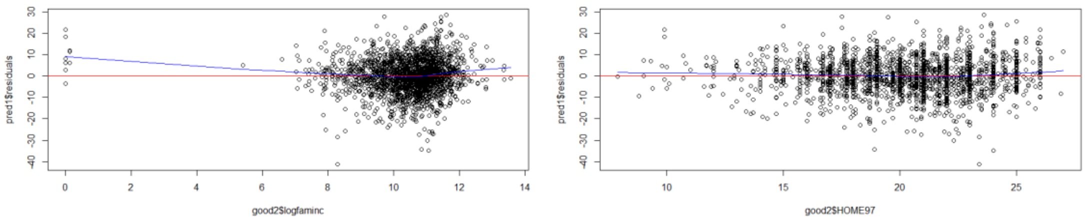

Figure 7. Scatterplots of residuals against independent variables

Since both *AGE97c* and *AGE97c2* are included in the regression model to deal with the potential nonlinearity of age, testing *AGE97c* and *AGE97c2* separately makes no sense. As the coefficients of *AGE97c* and *AGE97c2* are significant in Table 4, both of the variables should be kept, which indicates that there is a curvilinear relationship between age and child math scores.

Table 5.

COEFFICIENT RESULTS FOR MODEL WITH TRANSFORMED DATA

Variables	Est	SE	t	p
(Intercept)	18.14896	2.00751	9.041	< 2e-16 ***
AGE97c	6.90390	0.06400	107.874	< 2e-16 ***
AGE97c2	-0.08536	0.02481	-3.441	0.000592 ***
logfaminc	0.62995	0.17378	3.625	0.000296 ***
bthwht	-1.59559	0.38854	-4.107	4.17e-05 ***
WICpreg	-2.26227	0.41342	-5.472	5.00e-08 ***
HOME97	-2.26227	0.06871	9.685	< 2e-16 ***

*p<0.05 **p<0.01 ***p<0.001

As Figure 8 shows, most of the residuals fall in the range of about -2 to 2, with only a few observations having values lower than -3 or higher than 3. Therefore, observations with standardized residuals greater than +/-3 could be the real outliers in the dataset. Considering predicator and observation numbers, the leverage values where (2*6+2)/2042=0.007 could be the premiums. According to the plot, observations with leverage values greater than 0.01 could be the real outliers. For those observations marked as outliers, their Child ID could be noted as special focus.

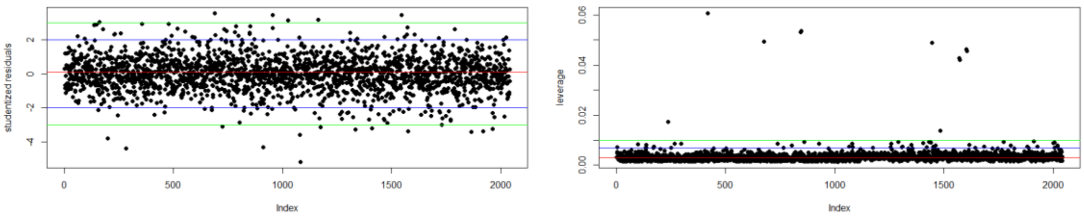

Figure 8. Scatterplot of the studentized residuals and leverage values

Dropping the observations with Cook's D greater than 4/2042, most observations of the new model are pretty close to the critical value of 0.00196. More detailed, as the distribution of extreme values for Cook's D shows, there is a big jump between the 80th and 85th percentile. The 85th percentile value is 0.007008524 which is much larger than the 80th percentile value, which is about 0.005371739. This is a good indicator that observations in the 85th percentile and higher are real outliers needed to be deleted.

3. Respecified Regression Model

The respecified regression model research on child math achievement (mathraw97) over and above the centred and squared age (AGE97c & AGE97c2), log-transformed family income (logfaminc), low birth weight status (bthwht), participation in WIC program during pregnancy (WICpreg) and the newly-added parenting practices (HOME97).

Table 6 provides the total variation of child math achievement accounted for by the respecified model. A total of 2025 observations are used to estimate the model. R reports a result about the F-statistic, 2466 on 6 and 2018 degree of freedom, and a p-value of < 2.2e-16, which is essentially 0. Since the p-value is less than 0.05, we can conclude that at least one of the independent variables helps explain the variation in the dependent variable (mathraw97). The two R-squared values tell that about 88% of the variance in mathraw97 is explained by independent variables.

According to the output of coefficients, we can conclude individual effects of each variable. The regression intercept indicates that the average math test score (mathraw97) is 14.9 when centred child age (AGE97c), log-transformed total family income (logfaminc), low birth weight status (bthwht), WIC program participation during pregnancy (WICpreg) and parenting practices (HOME97) are zero. On average, a child's math score will increase by 6.94 points for every additional year of age above the mean (which is approximately 7.4, years of age), holding all other independent

Table 6.

MULTIPLE REGRESSION RESULT FOR THE RESPECIFIED MODEL

Variables	Est	SE	p
(Intercept)	14.88466	2.10568	2.14e-12 ***
AGE97c	6.93708	0.06172	< 2e-16 ***
AGE97c2	-0.07857	0.02401	0.00108 **
logfaminc	0.90921	0.19071	2.00e-06 ***
bthwht	-1.59216	0.37373	2.14e-05 ***
WICpreg	-1.91186	0.40395	2.37e-06 ***
HOME97	0.67532	0.06688	< 2e-16 ***
Residual Min/Max	-27.4896/27.9359	Residual SE	7.714
Multiple R-squared	0.88	Adjusted R-squared	0.8796
F-statistic	2466	df	6 and 2018
p-value	< 2.2e-16		

*p<0.05 **p<0.01 ***p<0.001

AGE97c=centred AGE97; AGE97c2=squared AGE97c; logfaminc=log-transformed faminc97; bthwht=Low birth weight status of the child; WICpreg=Women, Infant and Children (WIC) Nutrition Program participant during pregnancy; HOME97=A composite total score of the emotional and cognitive stimulation at home

variables constant. A one percent increase in family income is associated with a (0.91/100) = 0.009-point increase in math score, holding all other independent variables constant. Children with low birth weight are predicted to have reading scores that are 1.59 points lower on average than children born with a normal weight, holding all other independent variables constant. Children participated in WIC Nutrition Program during pregnancy are predicted to have reading scores that are 1.91 points lower on average than children not involved in the program, holding all other independent variables constant. A child's math score is estimated to increase 0.68 points for every additional composite score point of the emotional and cognitive stimulation at home, holding all other independent variables constant.

Comparing the original and respecified regression model, the respecified model performs a better display on the relationship between the dependant variable (mathraw97) and independent variables. The data transformation justifies certain biased variables into the ones following a near-Gaussian distribution. Adding the once-omitted variable help to reduce the risk of lurking variable and form out a more reliable linear relationship. Deleting certain outliers makes the model more accurate. The summary outputs tells that the higher Adjusted R-squared value 0.88 of the respecified model, than 0.865 of the original model, represents a better modelling.

IV. Recommendation

The participation in WIC Nutrition Program turns out to have a negative effect on children's mathematic scores. However, the result does not mean that the programs actually have "negative" effect on children learning capability development. There is an underlying assumption that children involved in WIC nutrition program live a harder life than other children. They may not only meet the problem of nutrition and basic allowance, but they may lack sources for growth from many other aspects. From my perspective, even though the effect of participation in the WIC Nutrition program may be subtle in the analysis, they can be of importance to these people in real life. And this can be the reason why WIC program has been implemented till these days.

Appendix. References

Brooks-Gunn J, Duncan G (1997). The effects of poverty on children. *The Future of Children*, 7(2).

Cameron AJ, Spence AC, Laws R, Hesketh KD, Lioret S, Campbell K J (2015). A review of the relationship between socioeconomic position and the early-life predictors of obesity. *Current Obesity Reports*, 4(3):350-362.

Horino M, Yang W (2020). Impact of adverse childhood experiences and fruit and vegetable intake in adulthood. *Public Health Nutrition*, published online April 22.

Shonkoff J, Garner A, Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care, and Section on Developmental and Behavioral Pediatrics (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*, 129(1):e232-246.