MSSP 897

Policy Research Report 1: Multiple Regression

Ziyu Zhao

I. Introduction

This research report aims at exploring the effects of nutrition and family income on children's development by evaluating two social welfare programs: Women, Infant and Children (WIC) Nutrition Program and Aid to Families with Dependent Children (AFDC) program. The dataset is a U.S. national probability sample from the Child Development Supplement to the Panel Study of Income Dynamics and records the mothers' participation in the two programs and their children's reading achievement.

The following parts of this report will answer questions on whether the implementation of the programs has effects or unique contributions on children's reading achievement, which program has a larger effect, and how much variation of children's reading achievement are accounted for by the multiple regression models built to evaluate the size of the programs' effects.

II. Methods

The sample data contains 3563 rows and 112 columns. Each row represents a sample case of a family participated or not in the programs, while the columns include some major binary and continuous variables that reflect households' conditions and evaluate children's development. After cleaning up empty values for each selected variable, there are 1845 rows left for further study.

The method used to fulfill the goal of evaluation is to build multiple regression models to study respectively what the effect of mother's WIC participation and AFDC participation was on child's reading achievement holding constant the child's age, birth weight status, family total income in 1997, and the score of the child's home emotional and cognitive simulation. That is to say, the models to be built will use children's Woodcock-Johnson Revised Reading Achievement Test Age Standardized Score as the dependent variable and whether or not the mothers' participated in WIC

or AFDC during pregnancy as independent variable controlling other variables such as age and family income.

III. Results

a) Sample Descriptives

The following Table 1 gives a brief summary of the binary variables that matter to the analysis. And Table 2 summarizes the continuous variables.

i. Binary Variables

The binary variables, such as mother's WIC participation, AFDC participation during pregnancy and child's low birth weight, are dummy variables that use "1" to label positive responses to the questions and "0" to label negative ones. For example, for the variable of child's low weight status, "YES" means the child was a low birth weight child and "NO" means the child was a non-low birth weight child. And from Table 1 we can read that about 40.11% of the mothers participated in WIC during pregnancy and only 16.26% of them were included in the AFDC program. Moreover, about 35.12% of children were low birth-weight children in the sample.

Table 1

	YES		NO	
	Count	Percentage (%)	Count	Percentage (%)
WIC Participation	740	40.11	1105	59.89
AFDC Participation	300	16.26	1545	83.74
Child's Low Weight Status	648	35.12	1197	64.88

ii. Continuous Variables

The continuous variables included in this research include child's reading achievement score, child's age in 1997, child family's total income in 1997 and the home emotional and cognitive simulation score. The children's age included in the sample range from 3 to 13, and their score of standardized reading achievement test

range from 47.5 points to highest 165.5 points. The average income of the sampled families was 53,559 USD and the average home simulation score was 20.25 points, ranging from 7.9 to 27.

Table 2

	Mean	SD	Min	Max
Child's Reading Achievement Test Score	102.5	16	47.5	165.5
Child's Age in 1997	7.45	2.92	3	13
Total Family Income in 1997	53559	54979.77	0	784611
Home Emotional and Cognitive Simulation	20.25	3.1	7.9	27

b) Regression Analyses

i. P-value and Coefficients

The following Table 3 displays the results of both unstandardized and standardized multiple regression models evaluating the effect of WIC and AFDC on children's reading achievement controlling for age, family income, low birth weight status and home simulation score.

For the unstandardized WIC model (model 1), with adjusted R^2 being 0.1861, about 18.61% of the variation of children's reading achievement can be explained by the independent variables, while accounting for the number of observations used in the model and the number of predictors included. And the p-value of the F-statistic is smaller than 2.2e-16, implying that at least one of the independent variable is significantly related to the dependent variable and in this case, specifically, all of the independent variables are significantly associated with children's reading achievement score. The intercept of the regression on y-axis is about 74.47, i.e. when the values of WIC participation, child's age, low birth weight status, family income and home simulation score are equal to 0s, the value of children's reading achievement is 74.47. Reading from the results, the unstandardized coefficient for WIC participation is about -3.82, which can be interpreted as that, on average, the effect of participating in WIC program is a 3.82-point decrease in children's reading achievement test score.

Table 3

	Dependent variable:				
	readss97			scale(readss97)	
	(1)	(2)	(3)	(4)	(5)
WICpreg	-3.8190***		-3.1226***		
	(0.7828)		(0.8388)		
AFDCpreg		-3.8063***	-2.3973**		
		(0.9779)	(1.0454)		
AGE97	0.4240***	0.4490***	0.4324***		
	(0.1217)	(0.1220)	(0.1217)		
faminc97	0.00005***	0.0001***	0.00005***		
	(0.00001)	(0.00001)	(0.00001)		
bthwht	-2.2771***	-2.3406***	-2.2032***		
	(0.7355)	(0.7370)	(0.7354)		
HOME97	1.2186***	1.2809***	1.1774***		
	(0.1266)	(0.1251)	(0.1277)		
scale(WICpreg)				-0.1601***	
				(0.0234)	
scale(AFDCpreg)					-0.1132***
					(0.0227)
scale(AGE97)				0.0736***	0.0792***
				(0.0224)	(0.0226)
scale(bthwht)				-0.0711***	-0.0751***
				(0.0222)	(0.0223)
scale(HOME97)				0.2836***	0.3118***
				(0.0238)	(0.0232)
Constant	74.4659***	71.8577***	75.3645***	-0.0000	-0.0000
	(2.6335)	(2.4957)	(2.6595)	(0.0213)	(0.0214)
Observations	1,845	1,845	1,845	1,845	1,845
R2	0.1883	0.1845	0.1906	0.1677	0.1579
Adjusted R2	0.1861	0.1823	0.1880	0.1659	0.1561
Residual Std. Error	14.4354 (df = 1839)	14.4690 (df = 1839)	14.4187 (df = 1838)	0.9133 (df = 1840)	0.9186 (df = 1840)
F Statistic	85.3183*** (df = 5; 1839)	83.2138*** (df = 5; 1839)	72.1398*** (df = 6; 1838)	92.6671*** (df = 4; 1840)	86.2761*** (df = 4; 1840)
Note:					*p<0.1; **p<0.05; ***p<0.01

That is to say, averagely, children whose mothers participated in WIC program gained 3.82 points less than those whose mother did not. Given that the p-value of WIC participation variable is way smaller than 0.05, this correlation is significant at 95% level. To briefly sum up, mother's WIC participation is negatively correlated with children's reading score to a comparatively high degree based on the absolute value of its coefficient. Besides WIC participation, children's low birth weight state is also negatively correlated to their reading achievement score, that on average the children who had low birth weight gained 2.28 score less than those who was not born with low weight. Other than these two variables mentioned above, the child's age, family total income, and home simulation scores are all positively correlated. For every unit (year) increase in the child's age, the child's reading score increases by 0.42 points; for every 10,000 USD increase in the child's family total income, the kid scores 0.5 points higher; for every one point increase in the family's home emotional and cognitive simulation score, the kid score 1.22 points higher on average.

Model 2 is the unstandardized multiple-regression model which analyzes the impact of AFDC program. About 18.23% of the variation of children's reading achievement can be explained by this model and all of the independent variables are statistically significantly correlated to children's reading scores according to the model's and their individual p-values. When the independent variables are equal to 0s, the y-axis intercept is 71.86. Mothers' participation in AFDC program may potentially lead to an average 3.81-point decrease in their children's reading performance. Low birthweight kids earn 2.34 points on average less than non-low birthweight ones. For every unit increase in the child's age, family income and home simulation score, the child's reading achievement score rises respectively by 0.45, 0.0001 and 1.28 points. Thus, generally speaking, the mother's participation in AFDC is negatively correlated to child's reading ability development.

The unstandardized model (model 3) taking both WIC and AFDC program participation into account shows a comparison between the two programs in term of the magnitude of influence. Model 3 accounts for 19.06% of the variation of the

children's reading achievements with statistical significance at 99%. In this model, mothers' participation in WIC program leads to a -3.12 change in the children's reading score on average and AFDC participation individually accounts for 2.4 points decrease in the children's reading achievement score. That is to say, generally, the absolute effect of AFDC program on children's reading achievement is about 0.8 points less than WIC program's. Other than these, every unit increase in the variable of child's age, family income, child's low birth weight status and home simulation score correspondingly lead to 0.43, 0.00005, -2.2 and 1.18 points change in the child's reading achievement.

ii. Effect Size

To properly evaluate the effect sizes of the variables, the calculation of the models' partial correlation and semi-partial correlation values is also included in this report.

In Model 1, unstandardized multiple-regression model to evaluate the effect of mothers' WIC participation, the partial correlation for the WIC participation is about 0.0128 and the semi-partial correlation value is about 0.0105. That is to say, WIC participation accounts for 1.28% of the variance of the child's reading achievement score that is not accounted for by the child's age, family income, low birth-weight status and home simulation score. WIC participation makes 1.28% unique contribution to the unexplained variance in reading scores. WIC also accounts for 1.05% of the overall variance of reading score.

Table 4 Model 1 Effect Size

	SSR	df	pEta-sqr	dr-sqr
(Intercept)	166613.854	1	0.303	NA
WICpreg	4959.957	1	0.0128	0.0105
AGE97	2527.405	1	0.0066	0.0054
faminc97	9734.322	1	0.0248	0.0206
bthwht	1997.207	1	0.0052	0.0042
HOME97	19320.172	1	0.048	0.0409
Sum of squared errors (SSE): 383211.2				
Sum of squared total (SST): 472104.5				

In Model 2, the overall variance of reading achievement score accounted for by AFDC participation is 0.67% according to its semi-partial correlation value 0.0067. The unique contribution made by AFDC participation alone to the variance of children's reading achievement test scores that is not accounted by other independent variables is 0.82%.

Table 5 Model 2 Effect Size

	SSR	df	pEta-sqr	dr-sqr
(Intercept)	173550.429	1	0.3107	NA
AFDCpreg	3171.824	1	0.0082	0.0067
AGE97	2834.842	1	0.0073	0.006
faminc97	12543.345	1	0.0316	0.0266
bthwht	2111.412	1	0.0055	0.0045
HOME97	21963.530	1	0.0540	0.0465
Sum of squared errors (SSE): 384999.4				
Sum of squared total (SST): 472104.5				

For Model 3's evaluation of effect size, WIC participation accounts for 0.61% of overall variance of children's reading achievement scores and AFDC participation accounts for 0.23% the independent variable's variance. Comparatively, WIC program participation accounts for more of the overall variance of the children's reading test score than AFDC participation, indicating that, between these two programs, WIC program participation has a larger effect on child reading achievement.

Table 6 Model 3 Effect Size

	SSR	df	pEta-sqr	dr-sqr
(Intercept)	166954.227	1	0.3041	NA
WICpreg	2881.466	1	0.0075	0.0061
AFDCpreg	1093.333	1	0.0029	0.0023
AGE97	2626.044	1	0.0068	0.0056
faminc97	9388.512	1	0.0240	0.0199
bthwht	1866.170	1	0.0049	0.0040
HOME97	17678.749	1	0.0442	0.0374
Sum of squared errors (SSE): 382117.9				
Sum of squared total (SST): 472104.5				

IV. Conclusion

The analyses above show relatively strong negative correlations between mother's WIC and AFDC participation and child reading achievement scores. Given the result of the final Model 3, the WIC participation lead to a 3.12-point decrease in the reading score and the AFDC participation lead to a 2.4-point decrease in the children's reading score on average. According to the semi-partial correlation values of both variables, whether or not the mother's participation in WIC program during pregnancy has a larger effect on child's reading achievement than in AFDC program.

That is to say, the social welfare programs targeting at improving children's nutrition and family income conditions have no direct positive effects on child's intellectual development represented by reading achievement score. But these results are not to deny the value and effectiveness of WIC and AFDC programs because these two programs are not designed inherently for children's grades' improvement. As long as the children's nutrition and living conditions are actually bettered, the programs' purposes are fulfilled. To improve children's reading scores and other cognitive capabilities, projects and programs should be designed directly obeying scientific pedagogical rules.

V. Appendix: Code and Output

a) Pre-processing Data

```
> vars <- c('readss97','WICpreg','AFDCpreg','AGE97','faminc97','bthwht','HOME97')
> good_sub <- na.omit(good[vars])
> dim(good_sub)
[1] 1845    7
```

b) Sample Descriptives

```
> # 3.1 Sample Descriptives
> # Binary Variables
> table(good_sub$WICpreg)
   0    1
1105  740
> table(good_sub$AFDCpreg)
   0    1
1545  300
> table(good_sub$bthwht)
   0    1
1197  648

> # Continuos Variables
> summary(good_sub$readss97)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   47.5    92.5   101.5   102.5   112.5   165.5
> sd(good_sub$readss97)
[1] 16.00069
> summary(good_sub$AGE97)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.000   5.000   7.000   7.449  10.000  13.000
> sd(good_sub$AGE97)
[1] 2.915328
> summary(good_sub$faminc97)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0   20687   43154   53559   69909  784611
> sd(good_sub$faminc97)
[1] 54979.77
> summary(good_sub$HOME97)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   7.90   18.10   20.60   20.25   22.50   27.00
> sd(good_sub$HOME97)
[1] 3.102792
> da <- c(1105,740,1545,300,1197,648)
> for (i in da) {
+   print(round(i/1845,4))
+ }
[1] 0.5989 [1] 0.4011 [1] 0.8374 [1] 0.1626 [1] 0.6488 [1] 0.3512
```

c) WIC Model 1

```
> # 3.2 multiple regression
> wic <- lm(readss97~WICpreg+AGE97+faminc97+bthwht+HOME97, data=good_sub)
> wic_summary <- summary(wic)
> wic_summary

Call:
lm(formula = readss97 ~ WICpreg + AGE97 + faminc97 + bthwht +
    HOME97, data = good_sub)

Residuals:
    Min      1Q  Median      3Q     Max
-52.362  -9.011  -0.594   9.280  56.045

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.447e+01  2.633e+00  28.277  < 2e-16 ***
WICpreg     -3.819e+00  7.828e-01  -4.879 1.16e-06 ***
AGE97        4.240e-01  1.217e-01   3.483 0.000508 ***
faminc97     4.728e-05  6.918e-06   6.835 1.11e-11 ***
bthwht      -2.277e+00  7.355e-01  -3.096 0.001992 **
HOME97       1.219e+00  1.266e-01   9.629  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.44 on 1839 degrees of freedom
Multiple R-squared:  0.1883,    Adjusted R-squared:  0.1861
F-statistic: 85.32 on 5 and 1839 DF,  p-value: < 2.2e-16
```

d) AFDC Model 2

```
> afdc <- lm(readss97~AFDCpreg+AGE97+faminc97+bthwht+HOME97, data=good_sub)
> afdc_summary <- summary(afdc)
> afdc_summary

Call:
lm(formula = readss97 ~ AFDCpreg + AGE97 + faminc97 + bthwht +
    HOME97, data = good_sub)

Residuals:
    Min      1Q  Median      3Q     Max
-55.411  -9.165  -0.348   9.304  56.677

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.186e+01  2.496e+00  28.792  < 2e-16 ***
AFDCpreg    -3.806e+00  9.779e-01  -3.892 0.000103 ***
AGE97        4.490e-01  1.220e-01   3.680 0.000240 ***
faminc97     5.234e-05  6.762e-06   7.740 1.62e-14 ***
bthwht      -2.341e+00  7.370e-01  -3.176 0.001519 **
HOME97       1.281e+00  1.251e-01  10.243  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.47 on 1839 degrees of freedom
Multiple R-squared:  0.1845,    Adjusted R-squared:  0.1823
F-statistic: 83.21 on 5 and 1839 DF,  p-value: < 2.2e-16
```

e) WIC+AFDC Model 3

```
> wic_afdc <- lm(readss97~WICpreg+AFDCpreg+AGE97+faminc97+bthwht+HOME97, data=good_sub)
> wa_sum <- summary(wic_afdc)
> wa_sum

Call:
lm(formula = readss97 ~ WICpreg + AFDCpreg + AGE97 + faminc97 +
    bthwht + HOME97, data = good_sub)

Residuals:
    Min      1Q  Median      3Q     Max
-53.130  -8.980  -0.435   9.113  56.031

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.536e+01  2.659e+00  28.338  < 2e-16 ***
WICpreg     -3.123e+00  8.388e-01  -3.723 0.000203 ***
AFDCpreg    -2.397e+00  1.045e+00  -2.293 0.021946 *
AGE97        4.324e-01  1.217e-01   3.554 0.000389 ***
faminc97     4.649e-05  6.918e-06   6.720 2.42e-11 ***
bthwht      -2.203e+00  7.354e-01  -2.996 0.002772 **
HOME97       1.177e+00  1.277e-01   9.221  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.42 on 1838 degrees of freedom
Multiple R-squared:  0.1906,    Adjusted R-squared:  0.188
F-statistic: 72.14 on 6 and 1838 DF,  p-value: < 2.2e-16
```

f) Standardized Models

```
> wic_scale <- lm(scale(readss97)~scale(WICpreg)+scale(AGE97)+scale(bthwht)+scale(HOME
97),data=good_sub)
> wic_scale
Call:
lm(formula = scale(readss97) ~ scale(WICpreg) + scale(AGE97) +
    scale(bthwht) + scale(HOME97), data = good_sub)

Coefficients:
   (Intercept)  scale(WICpreg)    scale(AGE97)   scale(bthwht)
    -1.435e-16      -1.601e-01       7.357e-02      -7.112e-02
 scale(HOME97)
     2.836e-01
```

```
> afdc_scale <- lm(scale(readss97)~scale(AFDCpreg)+scale(AGE97)+scale(bthwht)+scale(HOM
E97),data=good_sub)
> afdc_scale

Call:
lm(formula = scale(readss97) ~ scale(AFDCpreg) + scale(AGE97) +
    scale(bthwht) + scale(HOME97), data = good_sub)

Coefficients:
    (Intercept)  scale(AFDCpreg)     scale(AGE97)    scale(bthwht)    scale(HOME97)
     -4.358e-16       -1.132e-01        7.924e-02       -7.515e-02        3.118e-01
```

g) Effect Size

```
> library(lmSupport)
> wic_me <- modelEffectSizes(wic)
lm(formula = readss97 ~ WICpreg + AGE97 + faminc97 + bthwht +
    HOME97, data = good_sub)
Coefficients
                SSR df pEta-sqr dR-sqr
(Intercept) 166613.854  1   0.3030     NA
WICpreg       4959.957  1   0.0128 0.0105
AGE97         2527.405  1   0.0066 0.0054
faminc97      9734.322  1   0.0248 0.0206
bthwht        1997.207  1   0.0052 0.0042
HOME97       19320.172  1   0.0480 0.0409
Sum of squared errors (SSE): 383211.2
Sum of squared total  (SST): 472104.5
> afdc_me <- modelEffectSizes(afdc)
lm(formula = readss97 ~ AFDCpreg + AGE97 + faminc97 + bthwht +
    HOME97, data = good_sub)
Coefficients
                SSR df pEta-sqr dR-sqr
(Intercept) 173550.429  1   0.3107     NA
AFDCpreg      3171.824  1   0.0082 0.0067
AGE97         2834.842  1   0.0073 0.0060
faminc97     12543.345  1   0.0316 0.0266
bthwht        2111.412  1   0.0055 0.0045
HOME97       21963.530  1   0.0540 0.0465

Sum of squared errors (SSE): 384999.4
Sum of squared total  (SST): 472104.5
```