

## **The Influences of Caregiver's Aggravation in Parenting and the Caregiver-Child Relationship on Social Behavior and Academic Performance**

Cindy Chiu  
Youngwon Kim

### **1. Summary of Research Questions**

- 1) How do caregiver's (mother/father in a fragile family) aggravation in parenting affect the focal child's academic performance?

In a fragile family, mother's aggravation in parenting did not affect child academic performance, whereas father's aggravation did affect child academic performance positively.

- 1) How does caregiver's (mother/father in a fragile family) aggravation in parenting affect the focal child's delinquent behaviors?

In a fragile family, mother's aggravation in parenting did not have an influence on delinquent behavior. However, father's aggravation had an influence on delinquent behavior negatively.

- 2) Does the regression models from general linear modeling predict the child's academic/delinquent performances better than the machine learning regression models?

The answer is unsure because there is no one answer. Overall, we could see that the bigger sample size we tested, the smaller coefficients difference we could see and the smaller sample size we tested, the smaller RMSE difference we could see.

### **2. Motivation and background**

Families have become more diverse than previous generations. According to the study of McLanahan (2013), about 10 percent of infants were born to unmarried parents in 1970, but about 40 percent of infants were born to unmarried parents in 2006. Thus, there has been increasing interest in fragile families indicting those who include the unmarried or romantically unstable parents and their child(ren).

Most of the parents in the fragile families are likely to be socioeconomically disadvantaged (Carlson & McLanahan, 2010), experience unstable relationships (Durtshi, Soloski, & Kimmes, 2017), vulnerable to psychological and emotional stress and pressures by employment, income or other internal as well as external factors ((Bronte-Tinkew, Horowitz, & Carrano, 2010) in the parent's life. This leads them to experience negative feelings such as anxiety, fears, frustration, or annoyance arising from the demands of being a caregiver or parent

and have a hard time maintaining a positive relationship with their child (Bronte-Tinkew et al., 2010).

The aggravation in parenting and the caregiver-child interaction are important in that previous studies showed that caregiver's aggravation in parenting is associated with child behavior problems (Hofferth, Davis-Kean, Davis, & Finkelstein, 1997) and the quality of the parent-child relationship is related with academic performances among young children and adolescents (Yuan, Weiser, & Fischer, 2016). However, there is a lack of such studies on fragile families because most studies have focused more on general parents and their children. In this context, it is crucial to identify the aggravation in parenting and caregiver-child interaction and their impacts on academic performance. To identify a model to understand the associations between them, we will conduct regression analyses using general linear modeling and machine learning.

### 3. Dataset

The dataset we are using is public data from the Fragile Families and Child Wellbeing study. The study had been conducted six waves of data collecting, spanning from 1998 to 2017. The dataset is consist of 4,989 observations and 16,993 variables. Each wave is consists of core interviews from mother and father. The baseline wave of data collection was taken at the birth of the focal child. Year 3 and Year 5 included the primary caregiver interviews and home visits. In Year 9, the study conducted an interview with the focal child in addition to the recollecting data on the topics from previous years. The Year 15 follow-up included new measures on the focal child's education and school experiences. We will be focusing on the child's academic and delinquent behavior in Year 9, which is taken in 2009.

The detailed documentary of the dataset can be found here: [Fragile Families and Child Wellbeing Study](#). To access the dataset, register, and request for access from [Office of Population Research](#). After getting permission from the Office of Population Research, download the data in SPSS file. We downloaded the data and then converted SPSS file to csv file so we can apply further coding in Python environment. We've extracted the data and converted it to csv, where the TA can access directly from [here](#).

### 4. Methodology

#### 1) Variables

- a) **Academic performance.** Academic performance was measured by Peabody Vocabulary Test (PPVT) to assess the receptive vocabulary and screens for verbal ability. The score of PPVT can be considered as the standardized score ranging from 37 to 159. To get the PPVT score, the interviewer reads a word of an activity book and let a child find the picture related to the word among a set of four pictures.
- b) **Delinquent behavior.** Having similar structures with the National Longitudinal Survey of Youth 1997, the measurement for the delinquent behavior consists of

17 items. A child was asked to report whether he or she committed the acts of delinquency, such as 1) crime against people, 2) theft, 3) vandalism, 4) school delinquency, 5) alcohol use, and 6) drug use. Those variables are answered in “yes/no” format.

- c) **Aggravation in parenting.** This measure was taken from the Child Development Supplement of the Panel Study of Income Dynamics (PSID) to assess the amount of parenting stress and pressures. A mother or father completed 4 items using a 4-point Likert scale from 1 (strongly disagree) to 4 (strongly agree).

## 2) Data Analysis

After cleaning up the dataset, we will provide a few descriptions in order to gain insight from our dataset. To achieve this, we import python packages StatsModels. This process involves examining the average of the academic response and delinquent behavior, the gender ratio of the responses and ratio of missing data. In addition, we will examine the correlation between the explanatory variables (such as father’s aggravation in parenting) and the response variables (academic performance). In order to show those statistics in a concise manner, we will construct correlation plots using Seaborn.

In order to answer the research question regarding how caregiver’s aggravation influence on child’s performances, we will produce a regression model based on father and mother’s response. For the delinquent behavior response variable in particular, the survey is done in a “yes/no” format. In order to make prediction and analysis, we treat the yes as 1, and no as 0, and sum up the values to make it numeric so we can further predict how many delinquent behavior the child will have based on parent’s aggravation. This process involves using the statistical package StatsModels. We will sum up the response of father/mother response to perform as our predict variable. For each question, we will perform the regression on both mother and father, and thus, there will be two regressions for each research questions.

Since we want to analyze whether the regression model makes a good prediction, we will compare it with the result of performing a machine learning algorithm. We will divide the dataset into a training set and testing set, and the proportion will be 80% training and 20% testing. Since we are focusing on the focal child’s academic performances, which are numeric values, we will use regression. In order to compare our result in linear regressions we performed in the previous step, we will implement the model using linear regression model. After getting all the models, we will use mean squared error to compare which model makes a better prediction on the dataset.

The packages we will be using are: pandas, numpy, seaborn, statsmodel, sklearn, scikit-learn.

## 5. Results

### 1) Descriptive Statistics

The function, '*mean\_std\_of\_variables*' provided the means and standard deviations of the focal variables: academic performance ( $M = 92.72$ ,  $SD = 14.95$ ), delinquent behavior ( $M = 1.24$ ,  $SD = 1.77$ ), and mother and father's aggravation in parenting ( $M = 2.98$ ,  $SD = 0.67$ ;  $M = 3.14$ ,  $SD = 0.68$ , respectively).

In Figure 1, you can see distributions of each variable. Mother's and father's aggravations in parenting are negatively skewed and child delinquent behavior is positively skewed. Academic performance is normally distributed because PPVT is the standardized score.

## 2) Correlations

Pairwise pearson correlations and their significant levels among the focal variables such as academic performance, delinquent behavior, and father and mother's aggravation in parenting were given by the function, '*get\_pearson*'. In Figure 1, mother's aggravation in parenting did not have significant relationship with father's aggravation. For academic performance, the mother's aggravation is not significantly associated with child's academic performance ( $r = -.09$ ,  $p > .05$ ), whereas the father's aggravation is significantly correlated with child's academic performance ( $r = .06$ ,  $p = .01$ ). For delinquent behavior, mother's aggravation in parenting had no relationship with child's delinquent behavior ( $r = .00$ ,  $p > .05$ ). However, father's aggravation had negative association with child's delinquent behavior ( $r = -.07$ ,  $p < .01$ ).

## 3) Regressions

There were four results from four regression models using OLS. In terms of academic performance, mother's aggravation in parenting did not have a unique effect on child academic performance ( $b = 1.99$ ,  $SE = 2.25$ ,  $p > .05$ ), but father's aggravation had a unique effect on child academic performance ( $b = 1.22$ ,  $SE = 0.48$ ,  $p = .012$ ). Specifically, there is an estimate mean increase of 1.22 on father's aggravation. In terms of delinquency behavior, mother's aggravation in parenting was not uniquely predictive of child academic behavior ( $b = -.01$ ,  $SE = .27$ ,  $p > .05$ ). However, father's aggravation is uniquely predictive of child academic behavior ( $b = -.17$ ,  $SE = .05$ ,  $p = .001$ ). This indicates that for every unit increase in father's aggravation in parenting we expect a 0.17 decrease in delinquent behavior. See the Figure 2.

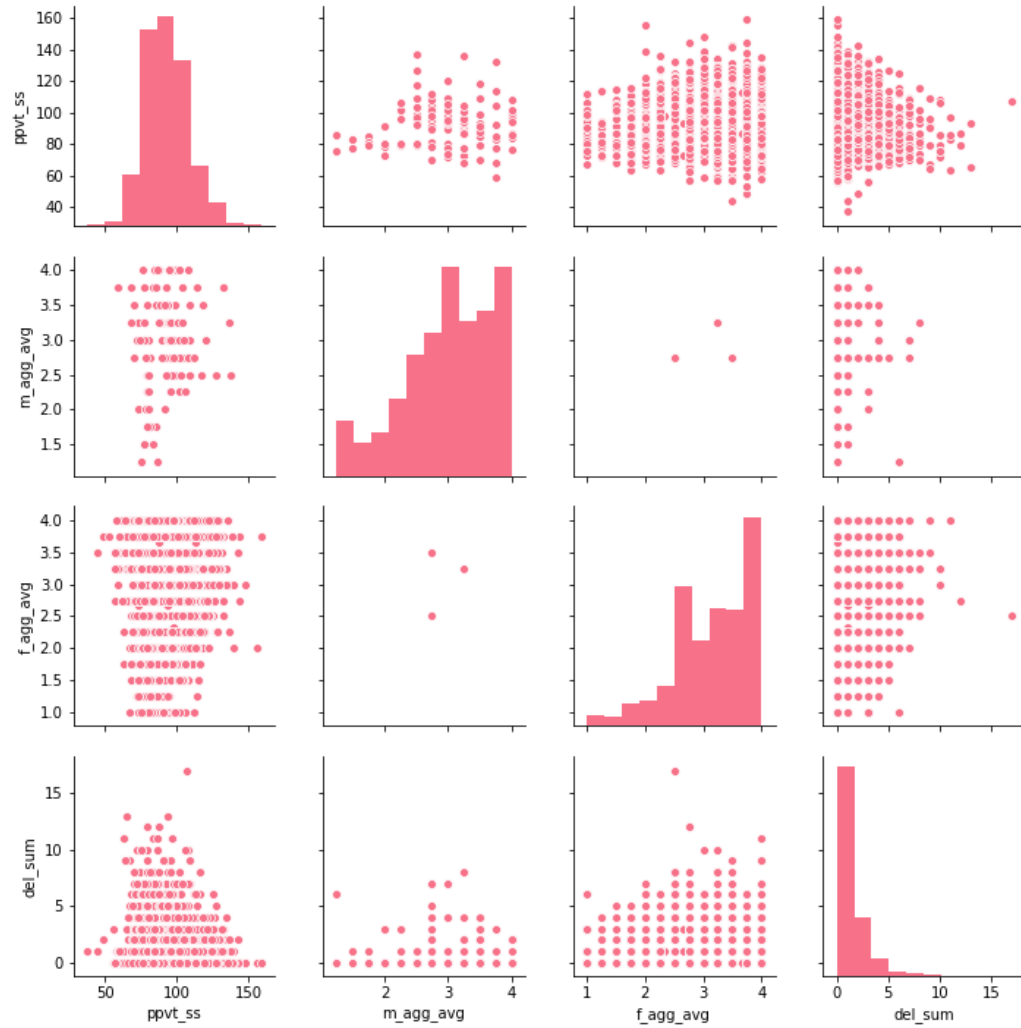


Figure 1. Zero-Order Correlations of Focal Variables.

#### 4) Comparing the Regression Models from General Linear Modeling with Machine Learning Regression Models

The functions ('fit\_and\_predict\_gender' and 'get\_coef\_mse') provided us with coefficients and root mean square errors (rmse) of regressions with machine learning and regressions with OLS.

a) For the regression of mother's aggravation in parenting on the academic performance, the obtained coefficient by OLS is 1.9897 ( $RMSE = 13.1679$ ) and the learned coefficient is 3.0007 ( $RMSE = 14.2050$ ).

b) For the regression of father's aggravation in parenting on the academic performance, the obtained coefficient by OLS is 1.2169 ( $RMSE = 37.9458$ ) and the learned coefficient is 1.2111 ( $RMSE = 15.2732$ ).

c) For the regression of mother's aggravation in parenting on the delinquent behavior, the obtained coefficient by OLS is  $-.0124$  ( $RMSE = 0.0813$ ) and the learned coefficient is  $-.0266$  ( $RMSE = 1.3698$ ).

d) For the regression of father's aggravation in parenting on the delinquent behavior, the obtained coefficient by OLS is  $-.1744$  ( $RMSE = 5.3881$ ) and the learned coefficient is  $-.1655$  ( $RMSE = 1.5400$ ).

After comparing obtained coefficients and RMSE with learned coefficient and RMSE, we could see that the bigger sample size we tested, the smaller coefficients difference we could see and the smaller sample size we tested, the smaller RMSE difference we could see. To see the differences between them, we need more tests and comparisons.

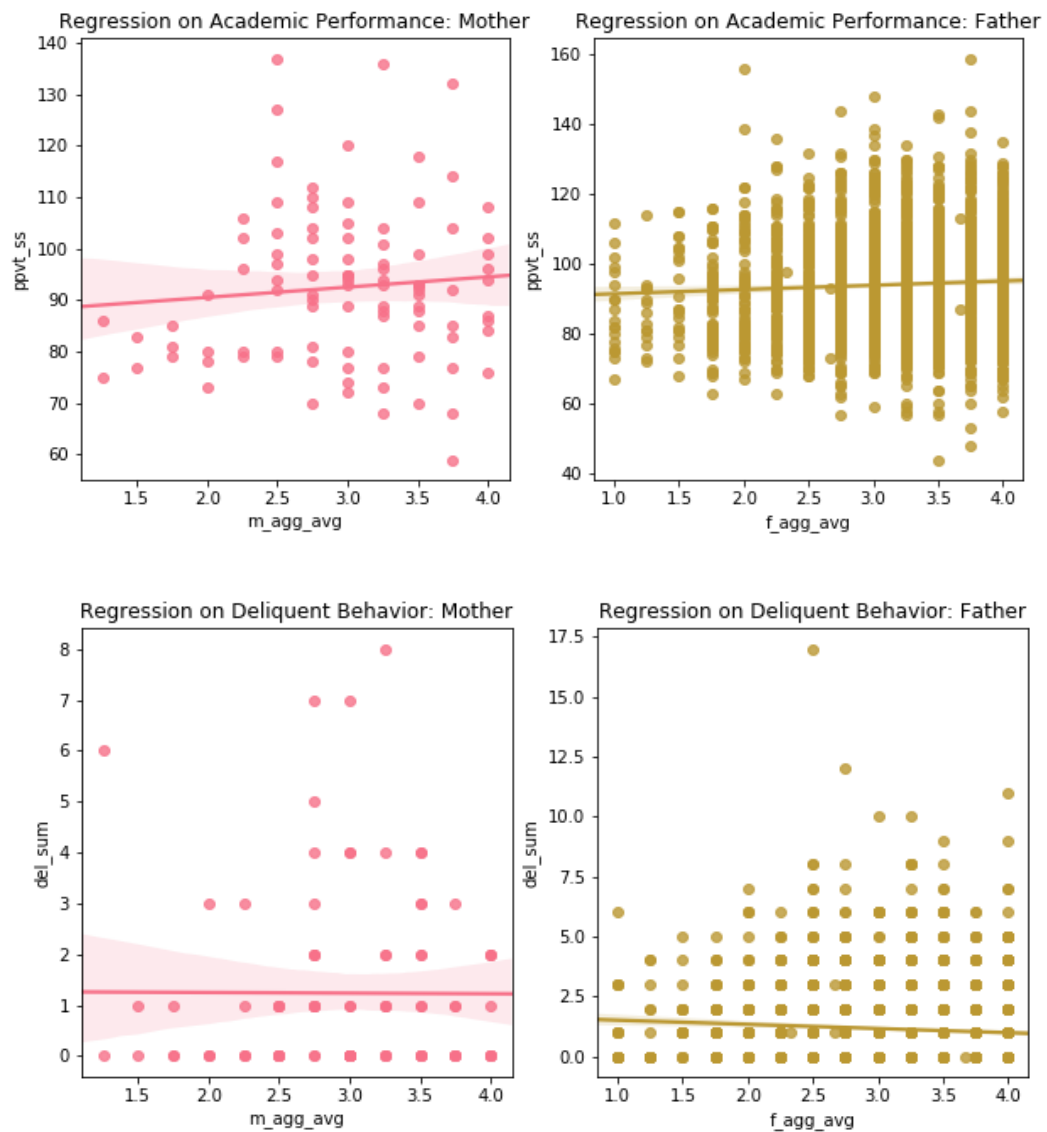


Figure 2. Regressions on Academic Performance and Delinquent Behavior

## 6. Reproducing your results

In order to recreate our results, you need to put the datafile “FF\_allwaves.csv” in the same folder of the python script files. The main program file is called “cse\_finalproject.py”. We’ve also implemented a source file “utils.py” and a class file, “cleaning.py”. We’ve already imported the class object and the functions needed in main program, so you need to make sure that the source file and the class file are in the same folder with the main program.

The main program will first print out the descriptive statistics for the dataset. Then the program should print out a set of lines with correlations and the p-values for each explanatory variable and the two response variables. In addition to the descriptions, it should also save a correlation figure. The main program should also print out the table of summary for the linear regression models using statsmodel package, including the coefficients of the intercept and the explanatory variable. Then it will produce the regression plots for each explanatory variables, which will have two different regression plots, each with both mother’s and father’s response. For the final machine learning part, the main program should print out the coefficients of the four regressions from machine learning model, and print out the RMSE for the test dataset for each regression.

## 7. Work Plan Evaluation

- 1) **Cleaning data:** This part may take up most of our time. We will discuss the procedure of cleaning together, or we might work on this part together. This procedure will take us around 3-5 days.

**Evaluation:** This process takes a little bit less time than we thought. We took around 2 days to finish this part, but we also revise the class object throughout the later development in order to make it load faster.

- 2) **Statistical description/visualization:** We will divide this task into two smaller parts. One of us will take care of the statistical tables with all the descriptive statistics, while the other one will be responsible for producing the visualizations. This procedure will take around 1 or 2 days.

**Evaluation:** This procedure took us a little longer than we thought. We ended up not dividing this task, and let one person handle this part.

- 3) **Regressions:** We will divide this task into several different regressions. We will discuss with the standard procedure we are performing on the data during the meeting, and then one of us will take care of the mother’s regression, while the other one will be responsible for the father’s regression.

**Evaluation:** We also did not divide this tasks into two. The person who develop the regression using statsmodel takes care of this part.

- 4) **Machine learning:** We will look use the regression we developed and apply a machine learning algorithm to predict the academic performance and social behavior in Year 15. This process will take us around 1-3 days.

**Evaluation:** The machine learning part took us around 2 days, including revisiting the class object and modify based upon that.

## **8. Testing**

We set up two smaller testing file. One test file with only 5 entries, and the other test file with 9 entries. We also applied what we've used in the previous homework, the assert statements to check on the output after applying the FF\_wave class functions. For the class we construct, we check its functions by printing out the results for the smaller file, we also check on the shape of the dataframe to make sure it has the desired output. Some functions like regression model are hard to test, so we didn't write asset functions for it. Instead, we use the descriptive statistics we produced and the statistic table to examine whether the results are reliable.

**9. Live Presentation or Video:** We'll do a 4 minute video presentation

**10. Collaboration:** None