Predictors of Child Aggression in a High School Setting

Daniel Pinedo

Psych 308c: Assignment 5

Predictors of Child Aggression in a High School Setting

Aggression and other maladaptive behaviors can be somewhat common and problematic in high school settings. A local intervention program wanted to find the biggest predictors of child aggression in order to target a prevention program. Program facilitators indicated that they observed that when youth had high family adversity and high positive friendships then they don’t act out quite as much. Conversely, facilitators noted that when students had high family and low positive relationships they tended to act out more. The purpose of this study was to determine whether positive peer relationships moderates the relationship between family adversity and child aggression.

**Method**

The present study used a correlational design. Data collection methods included short pen-and-paper survey surveys in two first period classes that collected data on child aggression, family adversity, and positive peer relationships.

**Participants**

Participant observations were 40 youth ages 16 to 18. Demographics included sex (female, *n* = 22; male *n* = 18), class rank (Sophomore, *n* = 16 ; Junior, *n* = 12 ; Senior, *n* = 12), and family income (range was from $43,632 to $72,159).

**Measures**

Each set of participant observations was assessed using the below measures. All measures were scored on a scale of 0 to 20.

**Child Aggression.**Child aggression was assessed using a scale.

**Family Adversity*.*** Family adversity was assessed using a scale.

**Positive Peer Relationships.**Positive Peer relationships was assessed using a scale

**Planned Analysis**

The present study planned to use correlation, simple regression, and multiple regression to assess the relationships between predictors, as well as predictors and the outcome variable.

**Results**

Data analysis is in Appendix A. Observations (*N* = 40) contained no missing parameters in the dataset. Analysis continued with tests of assumptions and inspection of histograms. There were no univariate outliers. One multivariate outlier was determined and removed based on calculations of Cook’s distance (cutoff = .10), a measure of multivariate influence. Descriptive statistics are in Table 1. Data was determined to not meet conditions of normality for child aggression as it was bimodal. Data was verified to otherwise be normally distributed across all variables as evidenced by skew for all variables being within a threshold of  3.00 (Table 1) and kurtosis being within a threshold of  10.00 (Table 1). The homoscedasticity assumption was confirmed using Breusch-Pagan test of non-constant variance, χ2 (1) = 0.68, *p* = .409. When viewing scatterplots with regression lines added and when correlated with child aggression, the assumption of linearity appears to be met for positive peer relationships (*r* = -.72, *p* < .001). However, family adversity (*r* = .39, *p* = .014) appears to not have a linear relationship with child aggression when viewing the scatterplot.

Family adversity and positive peer relationships were both significantly correlated with aggression (Table 2). The relationship between the outcome (child aggression) and potential predictors was further assessed through regression analyses. The best model fit for simple regression was indicated for positive peer relationships (β = -0.72, *p* < .001) which explained 52% of the variance in child aggression, *F*(1, 37) = 40.40, *p* < .001, *R2* = .52 (Table 3, Model 1). Adding family adversity (β = 0.33, *p* = .002) to Model 1 accounted for additional significant variance, *F*(1, 36) = 10.90 , *ΔR2* = .11, *p* = .002 (Table 3, Model 2). Adding positive peer relationships moderating family adversity (β = -0.28, *p* = .095) to Model 2 did not account for additional significant variance, *F*(1, 35) = 2.95 , *ΔR2* = .03, *p* = .095 (Table 3, Model 3). Therefore, Model 2 is determined to be the best fit, indicating that positive peer relationships and family adversity are the best predictors for child aggression.

**Discussion**

Correlation and regression analyses were used to test the hypothesis that positive peer relationships moderated the relationship between family adversity and child aggression. Anecdotal evidence indicated that the moderation relationship may exist, however our study indicated that the moderation relationship did not explain the model over and above that positive peer relationships and family adversity predicted child aggression (Table 3).

Although our hypothesis was not supported by the data, the data itself may prove to be problematic due to an issue of nonlinearity between family adversity and child aggression, and a non-normal bimodal distribution of the outcome variable (child aggression). As such, it is suggested that a more appropriate test be completed for this sample of observations. Notwithstanding the issues with the data set described above, positive peer relationships appears to have a high negative predictive relationship with child aggression, and family adversity has a moderate positive predictive relationship with child aggression . The intervention program should target a program aimed at increasing the quality and quantity of positive peer relationships at school and/or send social work outreach teams to homes identified as high family adversity, and complete a longitudinal study that examines the effects and interactions of these variables of this program.

Table 1

*Descriptive Statistics of Measures*

Variable Mean SD Median Skew Kurtosis

Child Aggression 8.92 5.28 8.00 0.30 -1.25

Family Adversity 11.90 3.71 11.00 -0.04 -0.24

Positive Peer Relationships 9.44 4.63 10.00 0.45 -0.03

Table 2

*Correlation Matrix for Measures Related to Child Aggression*

Variable 1 2 3

1. Child Aggression - .39\* -.72\*\*\*

2. Family Adversity - -.08

4. Positive Peer Relationships -

*Note.* \* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001.

Table 3

*Hierarchical Regression Models Predicting Child Aggression*

Model Variables *B* β SE *R2*

Model 1 Positive Peer Relationships -0.82 -0.72\*\*\* 0.13 .52

(PPR)

Model 2 Positive Peer Relationships -0.79 -0.70\*\*\* 0.12 .63

Family Adversity 0.48 0.33\*\* 0.14

(FA)

Model 3 Positive Peer Relationships -0.54 -0.47\*\* 0.52 .66

Family Adversity 0.47 0.33\*\* 0.14

PPR x FA -0.09 -0.28 0.05

*Note.*\* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001.

Appendix A

**Statistical Analysis in R**

Daniel Pinedo

March 12, 2019

## *Instructions*

You have been hired by a school-based intervention to figure out what predicts child aggression. The school created a short survey to be administered to students. Two classrooms were randomly chosen to participate. The study was conducted through a pen-and-paper survey at the beginning of first period they measured family adversity and positive peer relationships. They aren’t too savvy on the literature, but some of the staff at the program have been noting (anecdotally) an interesting phenomenon which may be impacting that relationship.

Apparently, even when youth have a lot of family adversity, if they also have positive friendships then they don’t act out quite as much. However, this doesn’t seem to be the case for those who also have a lot of adversity in their family but don’t have good peer relationships (they tend to act out a lot).

So, you have been tasked with accessing their data set and analyzing whether **positive peer relationships moderates the relationship between family adversity and child aggression**.

*Three variables for analysis:*

FA: Family adversity on a scale of 0-20.

AGG: Child aggression on a scale of 0-20.

PPR: Positive peer relationships on a scale of 0-20. (outcome)

*Demographics:*

Age: self-report; had to be at least 16 to participate.

Sex: self-report; male, female.

Family Income: $0.00 - $200,000.00.

Class Rank: Sophomore, Junior, Senior.

**Initial Data Diagnosis**

# Descriptives to get an overall view of data  
desc <- descriptives(data = dat,   
 vars = c('FA', 'AGG', 'PPR', 'Age', 'Sex', 'FamIncome', 'Rank'),  
 sd = TRUE,   
 range = TRUE,  
 skew = TRUE,   
 kurt = TRUE,  
 freq = TRUE) # for categorical variables  
desc

##   
## DESCRIPTIVES  
##   
## Descriptives   
## -------------------------------------------------------------------------------------------   
## FA AGG PPR Age Sex FamIncome Rank   
## -------------------------------------------------------------------------------------------   
## N 40 40 40 40 40 40 40   
## Missing 0 0 0 0 0 0 0   
## Mean 11.8 9.05 9.50 16.9 58384   
## Median 11.0 8.50 10.0 17.0 57642   
## Standard deviation 3.82 5.27 4.59 0.736 7457   
## Range 16 17 19 2.00 28527   
## Minimum 4 1 1 16.0 43632   
## Maximum 20 18 20 18.0 72159   
## Skewness -0.0533 0.242 0.411 0.246 0.115   
## Std. error skewness 0.374 0.374 0.374 0.374 0.374   
## Kurtosis -0.353 -1.30 -0.0248 -1.07 -0.904   
## Std. error kurtosis 0.733 0.733 0.733 0.733 0.733   
## -------------------------------------------------------------------------------------------   
##   
##   
## FREQUENCIES  
##   
## Frequencies of Age   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## 16 14 35.0 35.0   
## 17 18 45.0 80.0   
## 18 8 20.0 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of Sex   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## Female 22 55.0 55.0   
## Male 18 45.0 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of Rank   
## ----------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## ----------------------------------------------------   
## Junior 12 30.0 30.0   
## Senior 12 30.0 60.0   
## Sophmore 16 40.0 100.0   
## ----------------------------------------------------

corr.test(dat[2:4]) # Prerequisite: outcome and predictor variables are measured on the continuous level

## Call:corr.test(x = dat[2:4])  
## Correlation matrix   
## FA AGG PPR  
## FA 1.00 0.33 -0.1  
## AGG 0.33 1.00 -0.7  
## PPR -0.10 -0.70 1.0  
## Sample Size   
## [1] 40  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## FA AGG PPR  
## FA 0.00 0.08 0.54  
## AGG 0.04 0.00 0.00  
## PPR 0.54 0.00 0.00  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

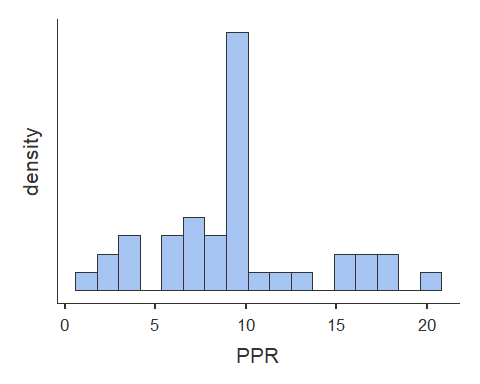
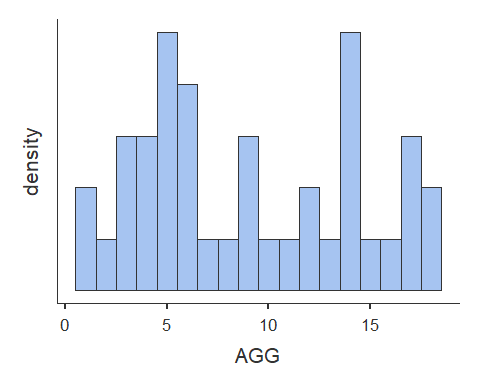
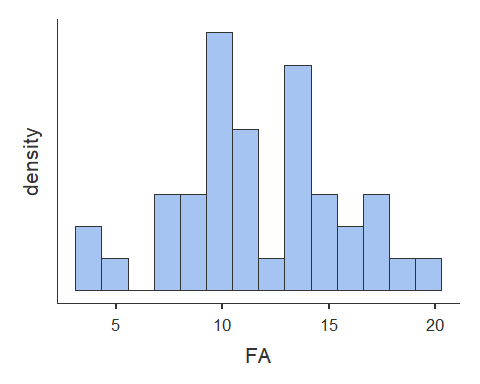
#MISSING DATA --> Different N's and the line that indicates missing items indicates missing cases   
#Running dim(dat) indicates 40 rows/observations  
#Options: (1) delete list-wise (2) impute

*Regression Diagnostics* 1. Missing Data - **NONE** 2. Univariate a. Normality, b. Linearity and c. Outliers 3. Multivariate a. Normality and b.Outliers 4. Heteroscedsticity 5. Multi-collinearity 6. Linearity between outcome and predictor(s)

**2a. Univariate Normality**

#ASSUMPTION: Normal Distribution for continuous variables X and Y (Child Aggression) [i.e. histogram, skew +-3, kurtosis +-10]  
  
desc <- descriptives(data = dat,   
 vars = c('FA', 'AGG', 'PPR'),  
 sd = TRUE,  
 range = TRUE,  
 skew = TRUE,   
 kurt = TRUE,  
 hist = TRUE) # for visual inspection  
desc

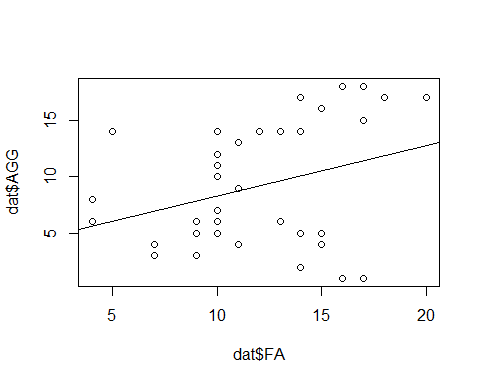
##   
## DESCRIPTIVES  
##   
## Descriptives   
## ------------------------------------------------------   
## FA AGG PPR   
## ------------------------------------------------------   
## N 40 40 40   
## Missing 0 0 0   
## Mean 11.8 9.05 9.50   
## Median 11.0 8.50 10.0   
## Standard deviation 3.82 5.27 4.59   
## Range 16 17 19   
## Minimum 4 1 1   
## Maximum 20 18 20   
## Skewness -0.0533 0.242 0.411   
## Std. error skewness 0.374 0.374 0.374   
## Kurtosis -0.353 -1.30 -0.0248   
## Std. error kurtosis 0.733 0.733 0.733   
## ------------------------------------------------------



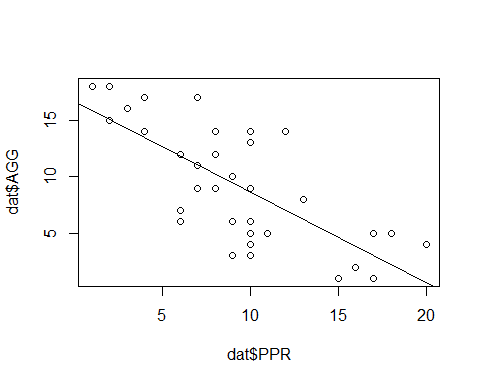
# Histogram for Family Adversity (FA) is normal  
# Histogram for Child Agression (AGG) is bimodal  
# Histogram for Positive Peer Relationships (PPR) is normal  
 # Skewness - ALL PASS  
 # Kurtosis - ALL PASS   
  
#Visual inspection indicates that there are no outliers

**2b. Univariate Linearity**

# Scatterplots [Assumption 2 and 3a]  
plot(dat$FA, dat$AGG, abline(lm(dat$AGG ~ dat$FA)))



plot(dat$PPR, dat$AGG, abline(lm(dat$AGG ~ dat$PPR)))



#visual inspection indicates a likely non-linear relationship between Family Adversity and Child Aggression and is consistent with visual inspection of histograms (step 2a) for outliers  
  
#visual inspection indicates a likely linear relationship between Positive Peer Relationship and Child Aggression and is consistent with visual inspection of histograms (step 2a) for outliers

**2c. Univariate Outliers**

#Identify outliers  
#scale() converts to z scores - "3" refers to standard deviations  
dat[abs(scale(dat$AGG)) > 3, ]

## [1] ID FA AGG PPR Age Sex FamIncome  
## [8] Rank   
## <0 rows> (or 0-length row.names)

dat[abs(scale(dat$FA)) > 3, ]

## [1] ID FA AGG PPR Age Sex FamIncome  
## [8] Rank   
## <0 rows> (or 0-length row.names)

dat[abs(scale(dat$PPR)) > 3, ]

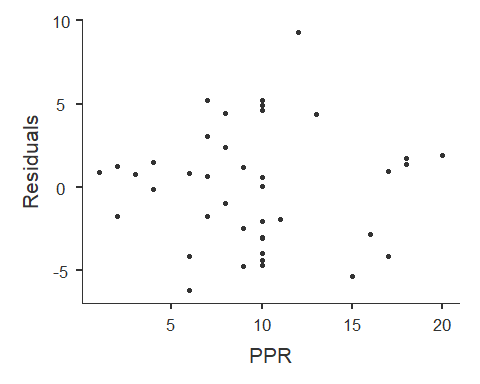
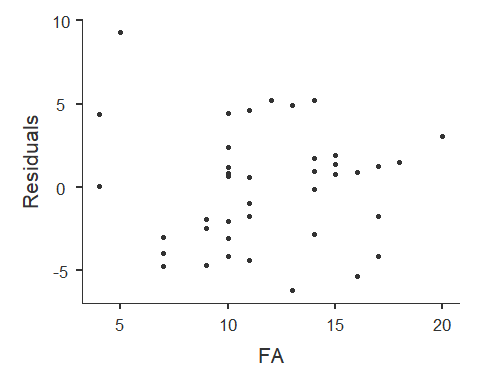
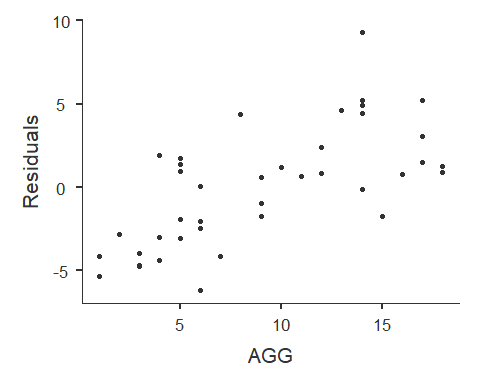
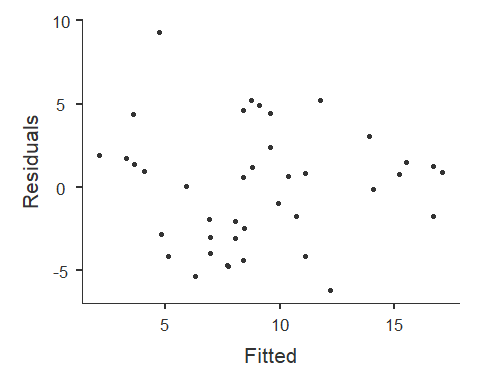
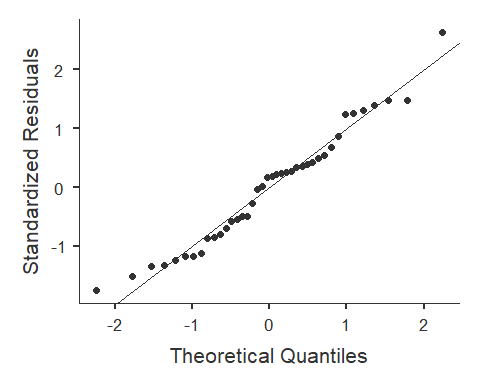
## [1] ID FA AGG PPR Age Sex FamIncome  
## [8] Rank   
## <0 rows> (or 0-length row.names)

#Child Aggression (AGG) has 0 univariate outliers  
#Family Adversity (FA) has 0 univariate outliers  
#Positive Peer Relationships (PPR) has 0 univariate outliers  
  
#There are a total of 0 independent observations that contain outliers

**3a. Multivariate Normality**

#look at residuals and the Q-Q plot  
#Observe Leverage (Mahalanobis' Distance) + Discrepancy (= Influence; Cook's Distance)  
  
model.multi\_norm <- linReg(data = dat,   
 dep = 'AGG',   
 covs = c('FA', 'PPR'),  
 blocks = list(c('FA', 'PPR')),   
 modelTest = TRUE,   
 r2Adj = TRUE,   
 stdEst = TRUE,   
 ciStdEst = TRUE,  
 qqPlot = TRUE, ##QQ plot  
 resPlots = TRUE) ##residuals plot   
  
model.multi\_norm

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## --------------------------------------------------------------------------   
## Model R R² Adjusted R² F df1 df2 p   
## --------------------------------------------------------------------------   
## 1 0.743 0.553 0.529 22.9 2 37 < .001   
## --------------------------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## ---------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## ---------------------------------------------------------------------------------------------   
## Intercept 12.218 2.330 5.24 < .001   
## FA 0.355 0.152 2.33 0.025 0.257 0.0334 0.481   
## PPR -0.773 0.127 -6.08 < .001 -0.672 -0.8959 -0.448   
## ---------------------------------------------------------------------------------------------   
##   
##   
## ASSUMPTION CHECKS



#Alternate not using jvm library  
#model <- lm(Amount ~ Belief + Need, data = dat.no.uni)  
#plot(model)  
  
#inspection of plots of predictors vs residuals indicates possible multivariate normality and possible heteroscadasticity  
#inspection of theoretical quantiles vs standardized residuals indicates a possible problem with multivariate distance and leverage  
#as such, Cook's distance - a measure of influence - will be used to test for multivariate normality  
#for Mahalanobis' Distance (leverage only), see Regression\_Diagnostics.Rmd for how-to

**3b. Multivariate Outliers**

#Check and remove multivariate outliers based on Cook's distance (CD)  
#CD = Influence = Leverage + Discrepancy (Discrepancy = how much an observation deviates from the overall pattern of the model)  
  
#create model  
model.cook <- lm(dat$AGG ~ dat$FA + dat$PPR)  
model.cook

##   
## Call:  
## lm(formula = dat$AGG ~ dat$FA + dat$PPR)  
##   
## Coefficients:  
## (Intercept) dat$FA dat$PPR   
## 12.2177 0.3550 -0.7726

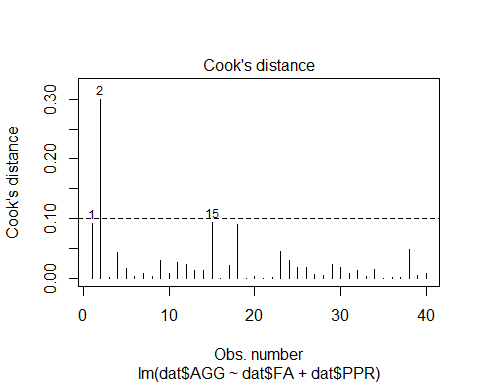
summary(model.cook)

##   
## Call:  
## lm(formula = dat$AGG ~ dat$FA + dat$PPR)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.1978 -2.8647 0.6212 1.7655 9.2779   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 12.2177 2.3304 5.243 6.65e-06 \*\*\*  
## dat$FA 0.3550 0.1525 2.329 0.0255 \*   
## dat$PPR -0.7726 0.1271 -6.080 4.89e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.62 on 37 degrees of freedom  
## Multiple R-squared: 0.5527, Adjusted R-squared: 0.5285   
## F-statistic: 22.86 on 2 and 37 DF, p-value: 3.438e-07

#find cook's distance for that model  
dat$cook <- cooks.distance(model.cook)  
  
#create the cutoff [> 4/N]  
cook.cutoff <- 4/nrow(dat)   
cook.cutoff

## [1] 0.1

# 4/40 --> cutoff = .10  
  
#plot it out  
plot(model.cook, which = 4, cook.levels = cook.cutoff)  
  
#Add a cutoff line  
abline(h = cook.cutoff, lty = 2)



#Show and remove all outliers above your cutoff line  
  
dat[(dat$cook) > cook.cutoff, ]

## ID FA AGG PPR Age Sex FamIncome Rank cook  
## 2 2 5 14 12 17 Male 68026 Junior 0.2988974

dat.final <- dat[!(dat$cook) > cook.cutoff, ]  
  
#N is now 39 after removing 1 multivariate outlier observation(s)  
 #was 40 after removing 0 univariate outlier observation(s)  
 #was 40 after removing 0 observation(s) with missing parameters  
 #was 40 originally (total 1 observation(s) removed from orginal dataset - 3%)

**4. Heteroscedasticity**

#Breusch-Pagan test   
#H0 = no change in variance across residuals.  
model.breusch\_pagan <- lm(dat.final$AGG ~ dat.final$FA + dat.final$PPR)  
ncvTest(model.breusch\_pagan)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 0.6821571, Df = 1, p = 0.40884

#not significant = homoscedastic  
#If violated use Box-cox transformation [boxcox(model)] in library MASS

**5. Multi-collinearity**

#Tolerance = 1 - R squared --> for our purpose < .4 is bad  
#VIF = 1/Tolerance ---> for our purpose > 2.5 is bad  
#Small VIF values (or higher Tolerance values) indicates low correlation among variables under ideal conditions  
  
#Multicollinearity occurs when two or more predictors in the model are correlated and provide redundant information about the response. Multicollinearity is measured by variance inflation factors (VIF) and tolerance. If VIF value exceeds 4.0, or tolerance less than 0.2 then there is a problem with multicollinearity according to Hair et al. (2010).  
  
model.multicoll <- linReg(data = dat.final,   
 dep = 'AGG',   
 cov = c('FA', 'PPR'),  
 blocks = list(c('FA', 'PPR')),   
 modelTest = TRUE,  
 r2Adj = TRUE,  
 stdEst = TRUE,  
 ciStdEst = TRUE,   
 collin = TRUE) #this line does the thing  
model.multicoll

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## --------------------------------------------------------------------------   
## Model R R² Adjusted R² F df1 df2 p   
## --------------------------------------------------------------------------   
## 1 0.796 0.633 0.613 31.1 2 36 < .001   
## --------------------------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## --------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## --------------------------------------------------------------------------------------------   
## Intercept 10.744 2.171 4.95 < .001   
## FA 0.476 0.144 3.30 0.002 0.334 0.129 0.540   
## PPR -0.794 0.116 -6.88 < .001 -0.696 -0.901 -0.491   
## --------------------------------------------------------------------------------------------   
##   
##   
## ASSUMPTION CHECKS  
##   
## Collinearity Statistics   
## ----------------------------   
## VIF Tolerance   
## ----------------------------   
## FA 1.01 0.994   
## PPR 1.01 0.994   
## ----------------------------

#Tolerance for all variables indicates low/no multicollinearity

*Data Analysis* 1. Descriptive Statistics 2. Correlations 3. Center Data (if useful) 4. Simple Regression 5. Hierarchical Model Comparison 6. Visualization

**1. Descriptive Statistics**

#Prerequisite: predictors and outcome all measured on continuous level  
#Assumptions:  
 #1. Normal Distribution for X and Y [i.e. histogram, skew +-3, kurtosis +-10]  
 # Histograms observed are normal  
 # Skewness - ALL PASS  
 # Kurtosis - ALL PASS  
 # N/A Observations with missing parameters were removed (see Diagnostics)  
 # N/A univariate outliers were removed (see Diagnostics)  
 # multivariate outliers were removed (see Diagnostics)  
   
 #2. Linear Relationship beween X and Y  
 #visual inspection indicates a non-linear relationship between Family Adversity and Child Aggression  
 #visual inspection indicates a linear relationship between Positive Peer Relationship and Child Aggression  
 #3. Homoscedasticity - OK (see Diagnostics)  
 #4. Multicollearity - OK (see Diagnostics)  
  
#N is now 39 after removing 1 multivariate outlier observation(s)  
 #was 40 after removing 0 univariate outlier observation(s)  
 #was 40 after removing 0 observation(s) with missing parameters  
 #was 40 originally (total 1 observation(s) removed from orginal dataset - 3%)  
  
desc.final <- descriptives(data = dat.final,   
 vars = c('FA', 'AGG', 'PPR', 'Age', 'Sex', 'FamIncome', 'Rank'),   
 sd = TRUE,   
 range = TRUE,   
 skew = TRUE,   
 kurt = TRUE,  
 freq = TRUE)  
desc.final

##   
## DESCRIPTIVES  
##   
## Descriptives   
## -------------------------------------------------------------------------------------------   
## FA AGG PPR Age Sex FamIncome Rank   
## -------------------------------------------------------------------------------------------   
## N 39 39 39 39 39 39 39   
## Missing 0 0 0 0 0 0 0   
## Mean 11.9 8.92 9.44 16.8 58137   
## Median 11 8 10 17.0 57338   
## Standard deviation 3.71 5.28 4.63 0.745 7387   
## Range 16 17 19 2.00 28527   
## Minimum 4 1 1 16.0 43632   
## Maximum 20 18 20 18.0 72159   
## Skewness -0.0404 0.301 0.450 0.260 0.166   
## Std. error skewness 0.378 0.378 0.378 0.378 0.378   
## Kurtosis -0.239 -1.25 -0.0279 -1.11 -0.819   
## Std. error kurtosis 0.741 0.741 0.741 0.741 0.741   
## -------------------------------------------------------------------------------------------   
##   
##   
## FREQUENCIES  
##   
## Frequencies of Age   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## 16 14 35.9 35.9   
## 17 17 43.6 79.5   
## 18 8 20.5 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of Sex   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## Female 22 56.4 56.4   
## Male 17 43.6 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of Rank   
## ----------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## ----------------------------------------------------   
## Junior 11 28.2 28.2   
## Senior 12 30.8 59.0   
## Sophmore 16 41.0 100.0   
## ----------------------------------------------------

**2. Correlations**

# Correlations of predictor and outcome variables  
cortable <- corrMatrix(data = dat.final,   
 vars = c('AGG', 'FA', 'PPR'),   
 flag = TRUE)  
cortable

##   
## CORRELATION MATRIX  
##   
## Correlation Matrix   
## --------------------------------------------------   
## AGG FA PPR   
## --------------------------------------------------   
## AGG Pearson's r  0.389 -0.723   
## p-value  0.014 < .001   
##   
## FA Pearson's r  -0.079   
## p-value  0.631   
##   
## PPR Pearson's r    
## p-value    
## --------------------------------------------------   
## Note. \* p < .05, \*\* p < .01, \*\*\* p < .001

**3. Center data (if useful)**

# Center only predictor variables  
# centered = x - M  
# Centering only changes the intercept for regression equation  
 # Centering means, on average across all predictor variables Y intercept is [coefficient for X units]  
# Center predictors  
dat.final$FA.centered <- dat.final$FA - mean(dat.final$FA)  
dat.final$PPR.centered <- dat.final$PPR - mean(dat.final$PPR)  
  
#USEFUL - We will center data for models of these predictors, as aggression score across average of all predictors is useful

**4. Simple Regression**

# Simple regression  
# R = correlation between observed scores and predicted scores  
# R squared = percentage of variance explained  
# t = Estimate / SE  
# df1 = k = number of predictors  
# df2 = N - k - 1 [k is number of predictors]  
# H0: B0 = 0; H0; R squared = 0  
  
model.wm <- linReg(data = dat.final,  
 dep = 'AGG',   
 covs = c('FA'),  
 blocks = list('FA'),   
 modelTest = TRUE,   
 stdEst = TRUE,   
 ci = TRUE)  
model.wm #1 fit

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## ----------------------------------------------------------   
## Model R R² F df1 df2 p   
## ----------------------------------------------------------   
## 1 0.389 0.152 6.62 1 37 0.014   
## ----------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## ------------------------------------------------------------------------------------------   
## Predictor Estimate SE Lower Upper t p Stand. Estimate   
## ------------------------------------------------------------------------------------------   
## Intercept 2.314 2.688 -3.132 7.760 0.861 0.395   
## FA 0.554 0.215 0.118 0.991 2.572 0.014 0.389   
## ------------------------------------------------------------------------------------------

model.process <- linReg(data = dat.final,  
 dep = 'AGG',  
 covs = c('PPR'),  
 blocks = list('PPR'),   
 modelTest = TRUE,   
 stdEst = TRUE,   
 ci = TRUE)  
model.process #2 fit

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## -----------------------------------------------------------   
## Model R R² F df1 df2 p   
## -----------------------------------------------------------   
## 1 0.723 0.522 40.4 1 37 < .001   
## -----------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## -------------------------------------------------------------------------------------------   
## Predictor Estimate SE Lower Upper t p Stand. Estimate   
## -------------------------------------------------------------------------------------------   
## Intercept 16.702 1.359 13.95 19.455 12.29 < .001   
## PPR -0.824 0.130 -1.09 -0.562 -6.36 < .001 -0.723   
## -------------------------------------------------------------------------------------------

**5. Hierarchical Model Comparison**

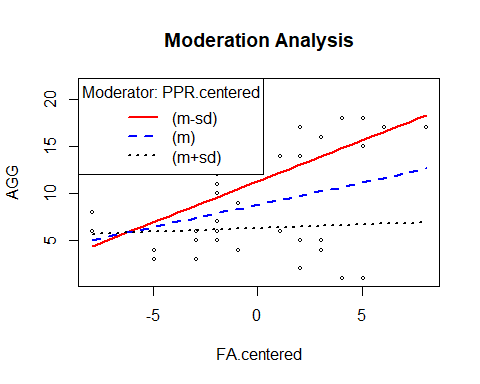
# Model comparison  
# H0 = delta of R squared = 0  
  
# Multiple regression with Moderation  
# REMEMBER: Coefficient estimate for one predictor is slope across average of other predictors.  
#F test tests null hypothesis that null hypothesis is zero (is square of t score of product term)  
#T test test null hypothesis that product term is zero  
  
#create moderator variable  
dat.final$mod.centered <- dat.final$PPR.centered \* dat.final$FA.centered  
  
comparison <- linReg(data = dat.final,   
 dep = 'AGG',   
 covs = c('PPR.centered', 'FA.centered', 'mod.centered'),  
 blocks = list(  
 list('PPR.centered'),   
 list('FA.centered'),   
 list('mod.centered') ),   
 modelTest = TRUE,   
 r2Adj = TRUE,  
 stdEst = TRUE,  
 ciStdEst = TRUE)  
comparison

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## --------------------------------------------------------------------------   
## Model R R² Adjusted R² F df1 df2 p   
## --------------------------------------------------------------------------   
## 1 0.723 0.522 0.509 40.4 1 37 < .001   
## 2 0.796 0.633 0.613 31.1 2 36 < .001   
## 3 0.814 0.662 0.633 22.8 3 35 < .001   
## --------------------------------------------------------------------------   
##   
##   
## Model Comparisons   
## -----------------------------------------------------------------   
## Model Model <U+0394>R² F df1 df2 p   
## -----------------------------------------------------------------   
## 1 - 2 0.1111 10.90 1 36 0.002   
## 2 - 3 0.0285 2.95 1 35 0.095   
## -----------------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## ---------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## ---------------------------------------------------------------------------------------------   
## Intercept 8.923 0.592 15.07 < .001   
## PPR.centered -0.824 0.130 -6.36 < .001 -0.723   
## ---------------------------------------------------------------------------------------------   
##   
##   
## MODEL 2  
##   
## Model Coefficients   
## -----------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## -----------------------------------------------------------------------------------------------   
## Intercept 8.923 0.526 16.97 < .001   
## PPR.centered -0.794 0.116 -6.88 < .001 -0.696 -0.901 -0.491   
## FA.centered 0.476 0.144 3.30 0.002 0.334 0.129 0.540   
## -----------------------------------------------------------------------------------------------   
##   
##   
## MODEL 3  
##   
## Model Coefficients   
## -------------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## -------------------------------------------------------------------------------------------------   
## Intercept 8.8095 0.5164 17.06 < .001   
## PPR.centered -0.5364 0.1875 -2.86 0.007 -0.470 -0.804 -0.1365   
## FA.centered 0.4736 0.1403 3.37 0.002 0.333 0.133 0.5330   
## mod.centered -0.0857 0.0499 -1.72 0.095 -0.282 -0.616 0.0512   
## -------------------------------------------------------------------------------------------------

#model 1 = AGG ~ PPR  
#model 2 = AGG ~ PPR + FA  
#model 3 = AGG ~ PPR + FA + mod

**Moderation Analysis**

# moderator is represented on the plot line that is in the legend  
# high positive peer relationships does not significantly "exaggerate" the relationship between family adversity and child aggression  
plot <- lm(AGG ~ PPR.centered + FA.centered + (PPR.centered \* FA.centered), data = dat.final)  
plotSlopes(plot, plotx = "FA.centered", modx = "PPR.centered", modxVals = "std.dev.", main = "Moderation Analysis", col = c("red", "blue", "black"))



**6. Visualization**

# plotting a regression model based on:   
 # Model 2: AGG ~ PPR.centered + FA.centered  
  
# create linear model  
model.final <- lm(AGG ~ PPR + FA, data = dat.final)  
summary(model.final)

##   
## Call:  
## lm(formula = AGG ~ PPR + FA, data = dat.final)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.1640 -2.4019 0.5021 1.7678 5.6769   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.7438 2.1709 4.949 1.76e-05 \*\*\*  
## PPR -0.7941 0.1155 -6.875 4.78e-08 \*\*\*  
## FA 0.4758 0.1441 3.302 0.00218 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.284 on 36 degrees of freedom  
## Multiple R-squared: 0.6333, Adjusted R-squared: 0.6129   
## F-statistic: 31.08 on 2 and 36 DF, p-value: 1.44e-08

model\_p <- ggpredict(model.final, terms = c('PPR', 'FA'), full.data = TRUE, pretty = TRUE) #for single regression, remove terms = c("v1"", "v2", "vn")  
  
# plot predicted line - for single regression, change to aes(y = VAR, x = VAR)  
plot <- ggplot(model\_p, aes(x, predicted)) +  
 geom\_smooth(method = "lm", se = TRUE, fullrange=TRUE) + xlab("Score") + ggtitle("Model of Peer Relationships and Family Adversity Predicting Child Aggression") + ylab("Child Aggression") +  
 geom\_point() + theme\_minimal()  
  
plot

