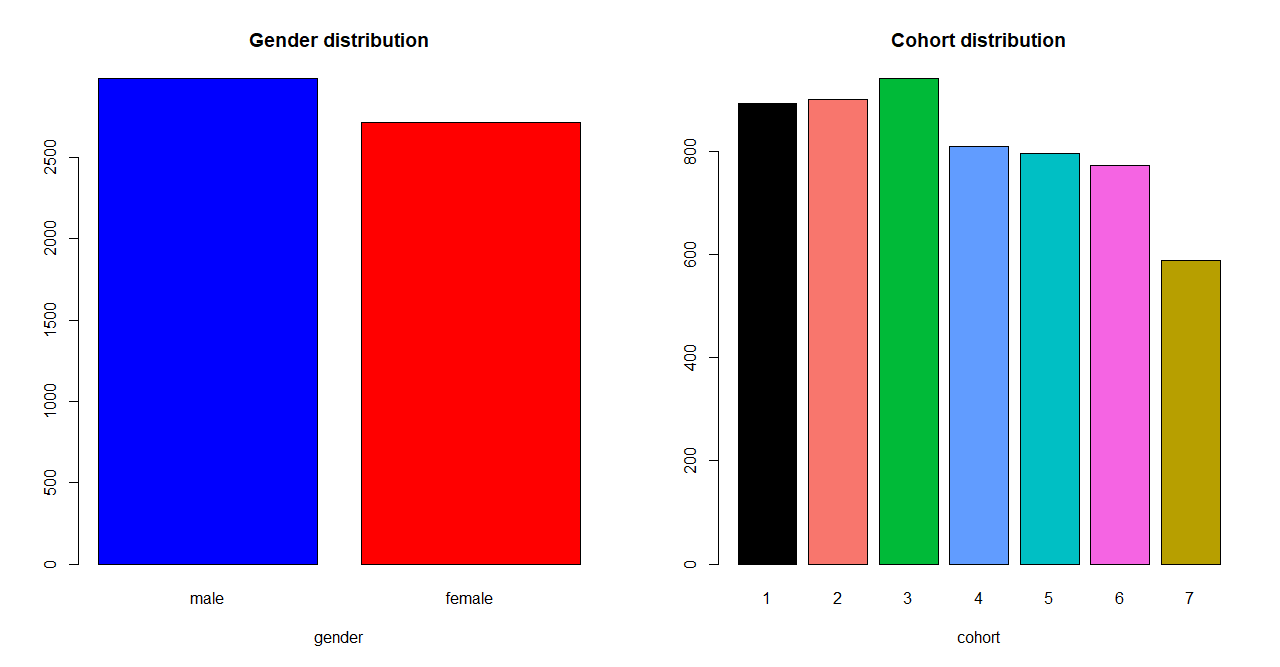
## **A hierarchical Linear Modeling Analysis of the National Youth Longitudinal Study**

Yu Huang

##### **Exploratory Data Analysis**

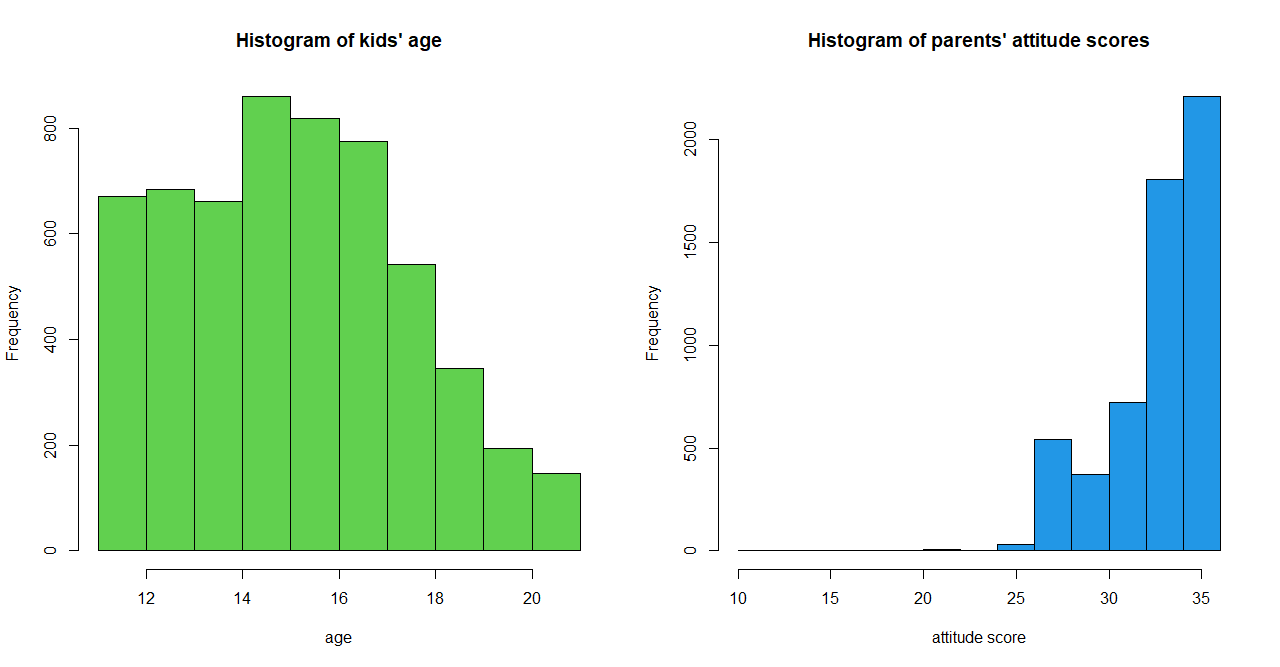
1. Descriptive statistics and distribution graphs

From the summary statistics of the response variable, youth\_dev, across age groups (see table below), we can see a clear decreasing trend – that is, if we pool data across individuals, we can observe that as kids get older, they tend to have lower attitude scores (more approval) towards deviant behaviors. On the other hand, notice that the variance of the response variable also tends to increase with age. Based on distribution plots for multiple independent variables (see figures below) we are interested in, we can see that in this sample (N = 1424 after removing individuals without enough data), there are slightly more males than females; the sample size for each cohort group is uneven (cohort 7 has noticeably fewer individuals); kids’ age distribution is right-skewed with a lot more kids in the middle range (which is a result of the study design); and parents’ attitude scores are very left-skewed: most have high scores, indicating strong disapproval of deviant behaviors.

*Descriptive statistics for youth\_dev as a function of age.*

|  |  |  |  |
| --- | --- | --- | --- |
| age |  | *M* | *SD* |
| 11 |  | 33.60 | 2.43 |
| 12 |  | 33.00 | 2.89 |
| 13 |  | 32.03 | 3.63 |
| 14 |  | 30.96 | 4.16 |
| 15 |  | 30.17 | 4.49 |
| 16 |  | 29.62 | 4.60 |
| 17 |  | 28.99 | 4.49 |
| 18 |  | 28.79 | 4.20 |
| 19 |  | 28.29 | 4.30 |
| 20 |  | 28.16 | 4.62 |
| 21 |  | 28.59 | 4.34 |

*Note.* *M* and *SD* represent mean and standard deviation, respectively.

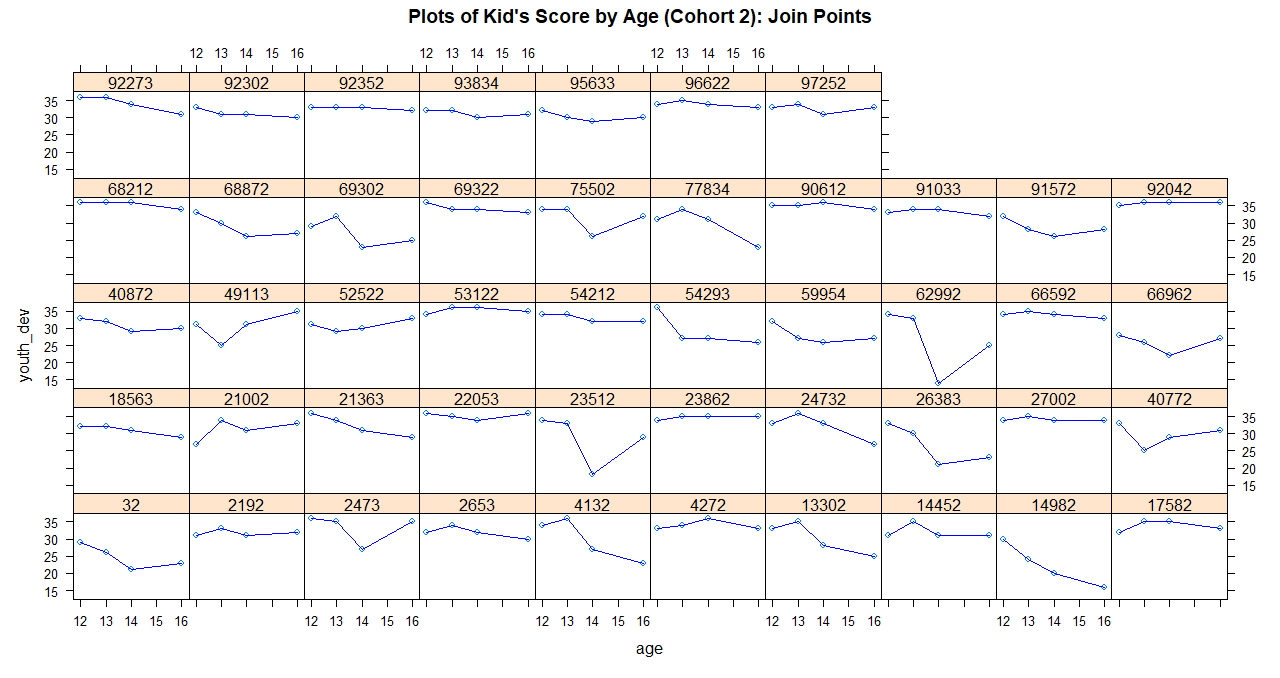
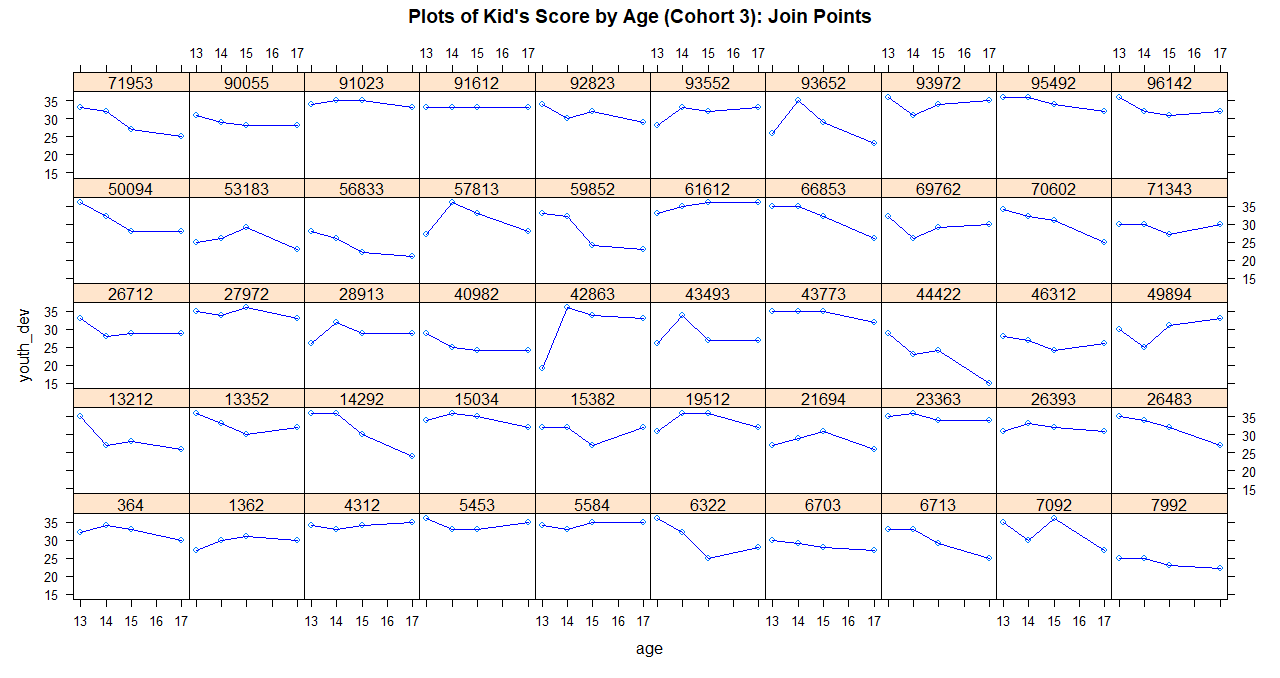
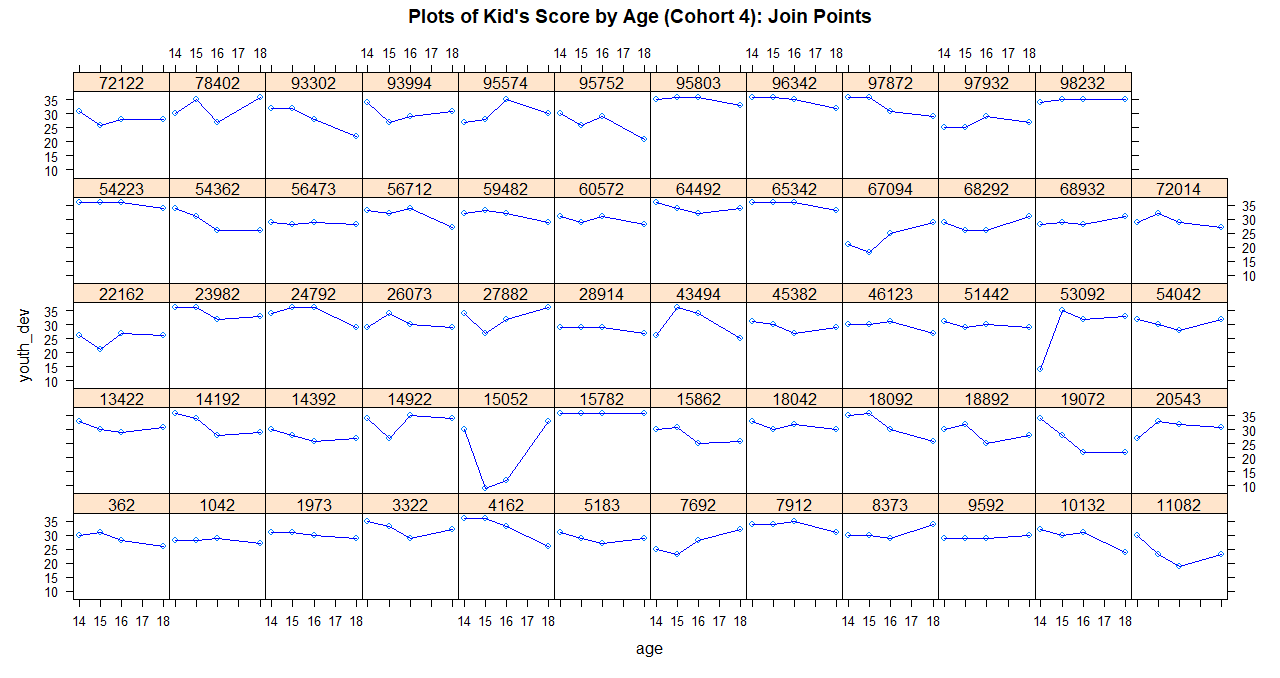
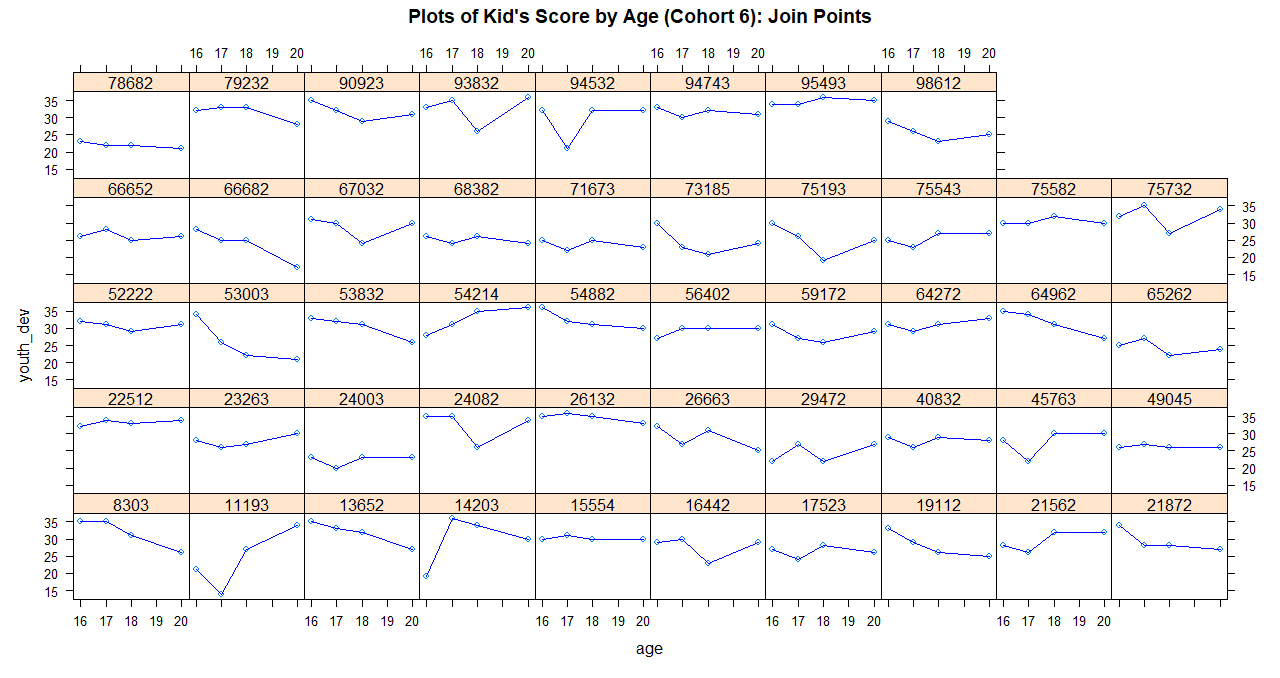


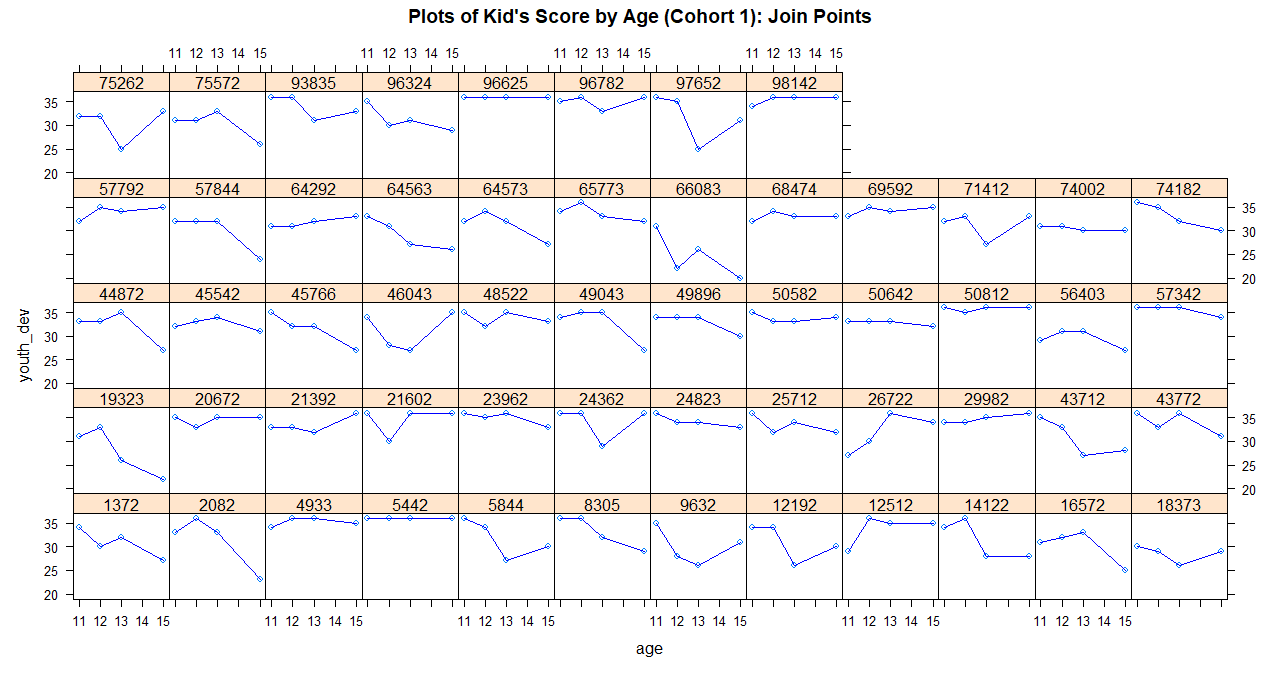
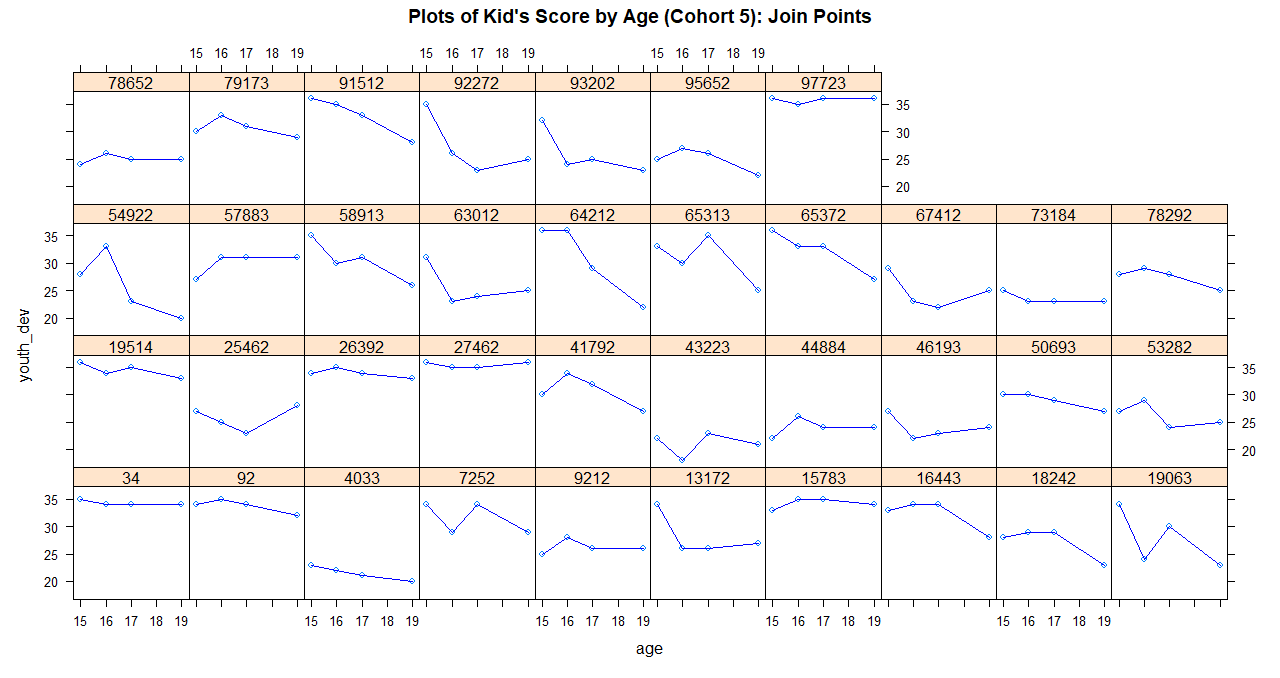
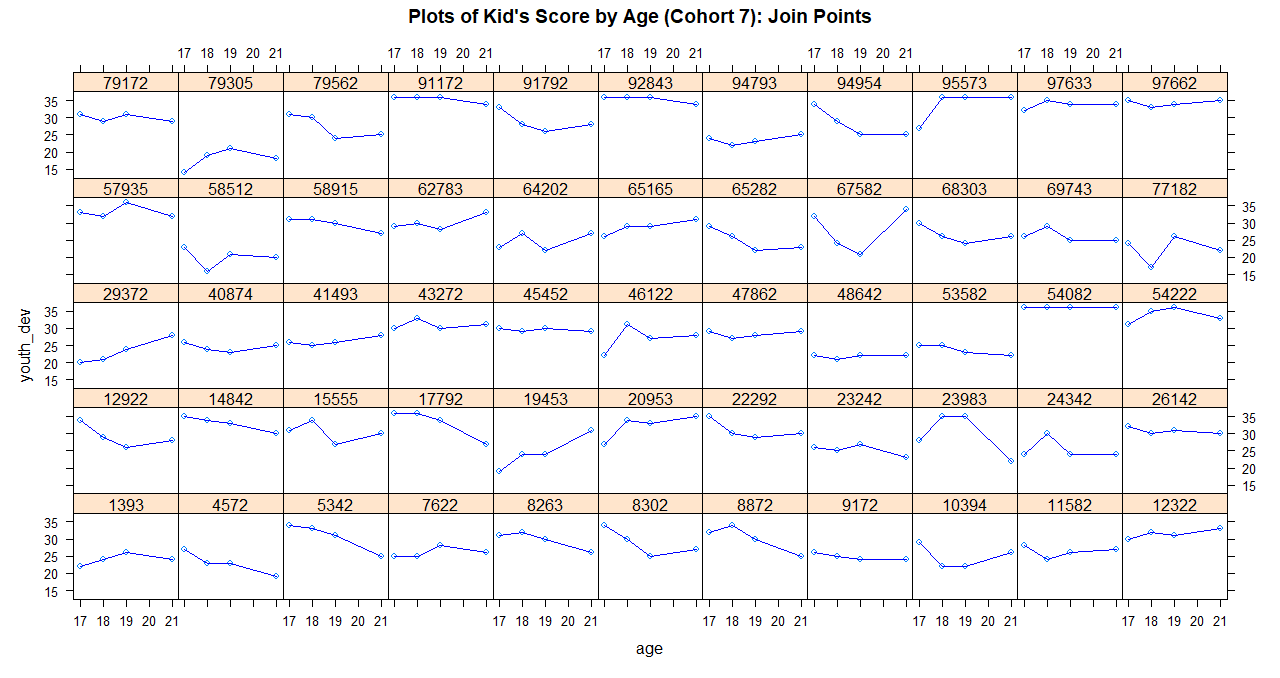
##### 

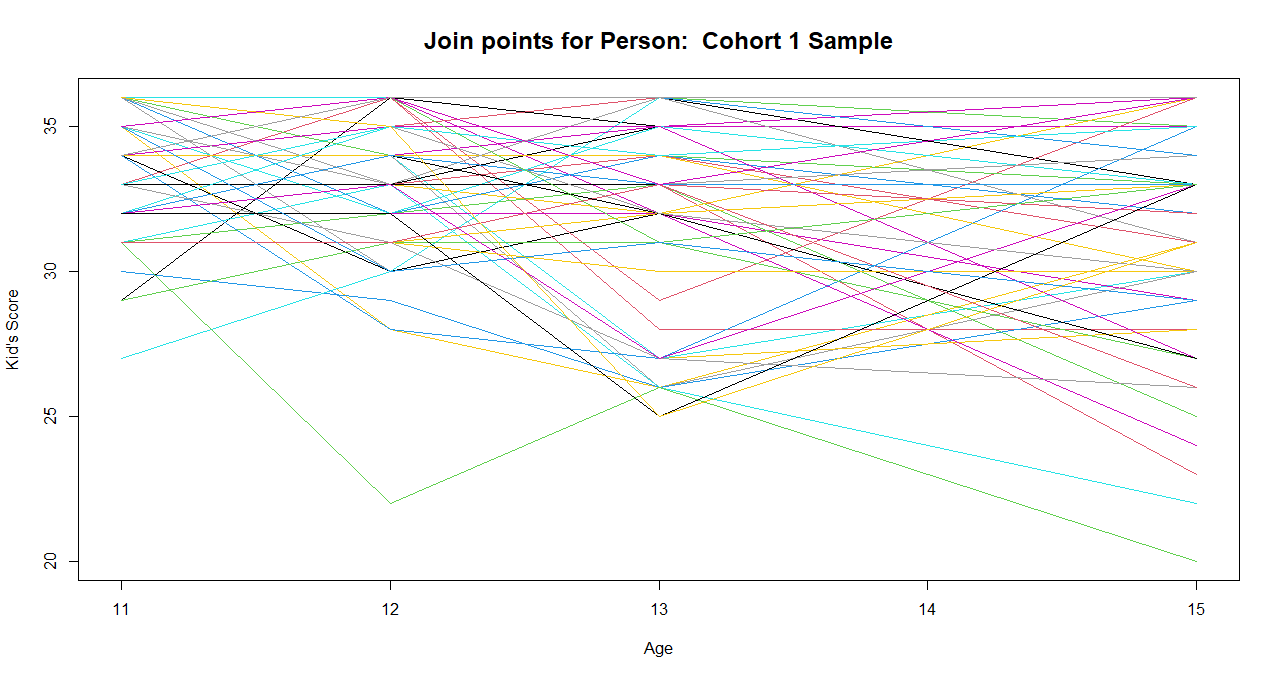
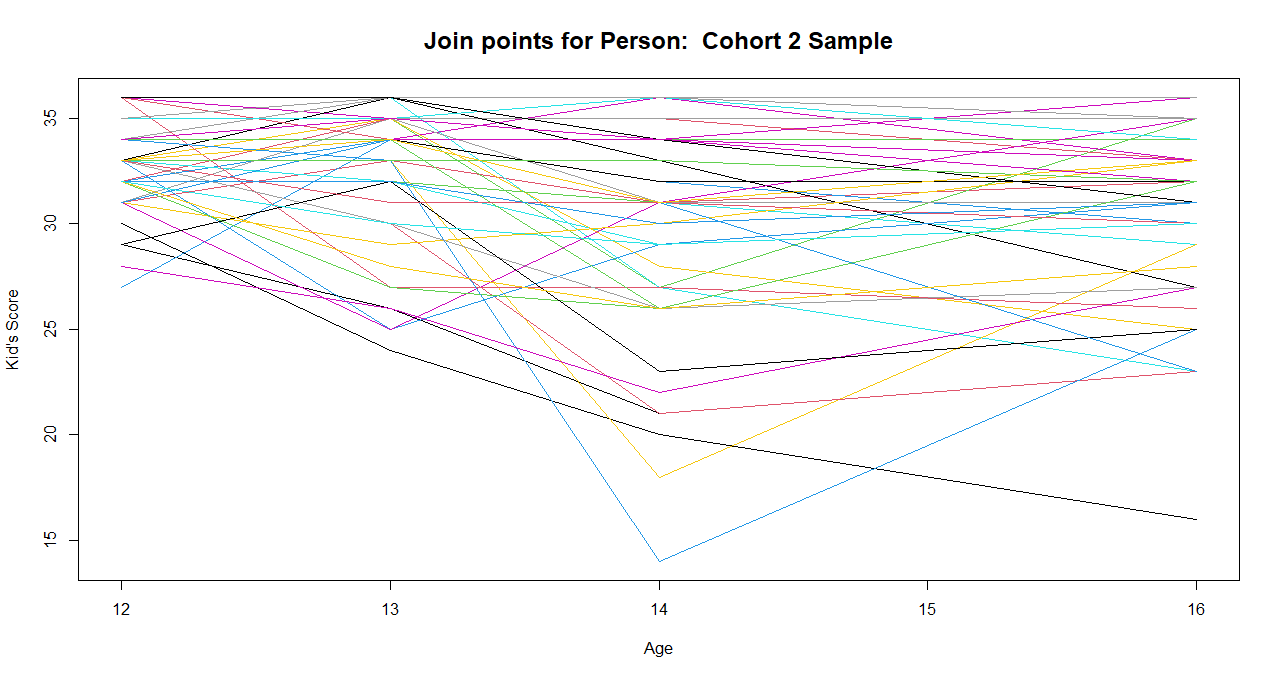
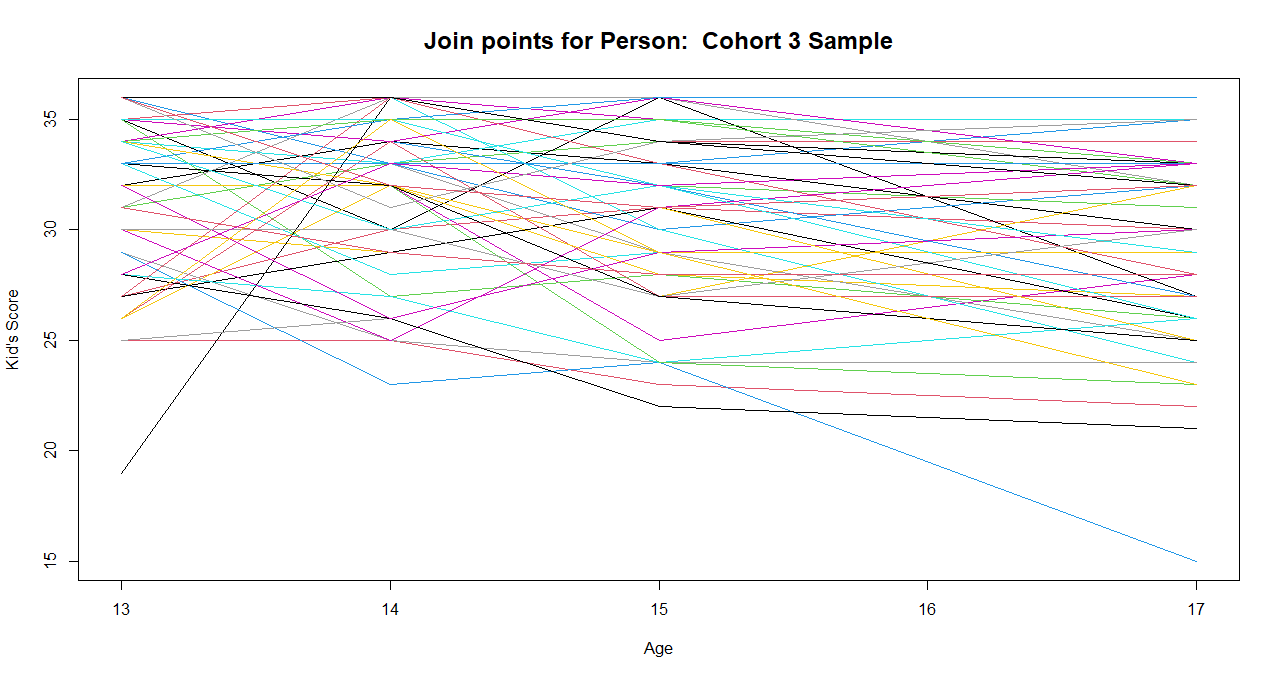
1. Exploring individual structures

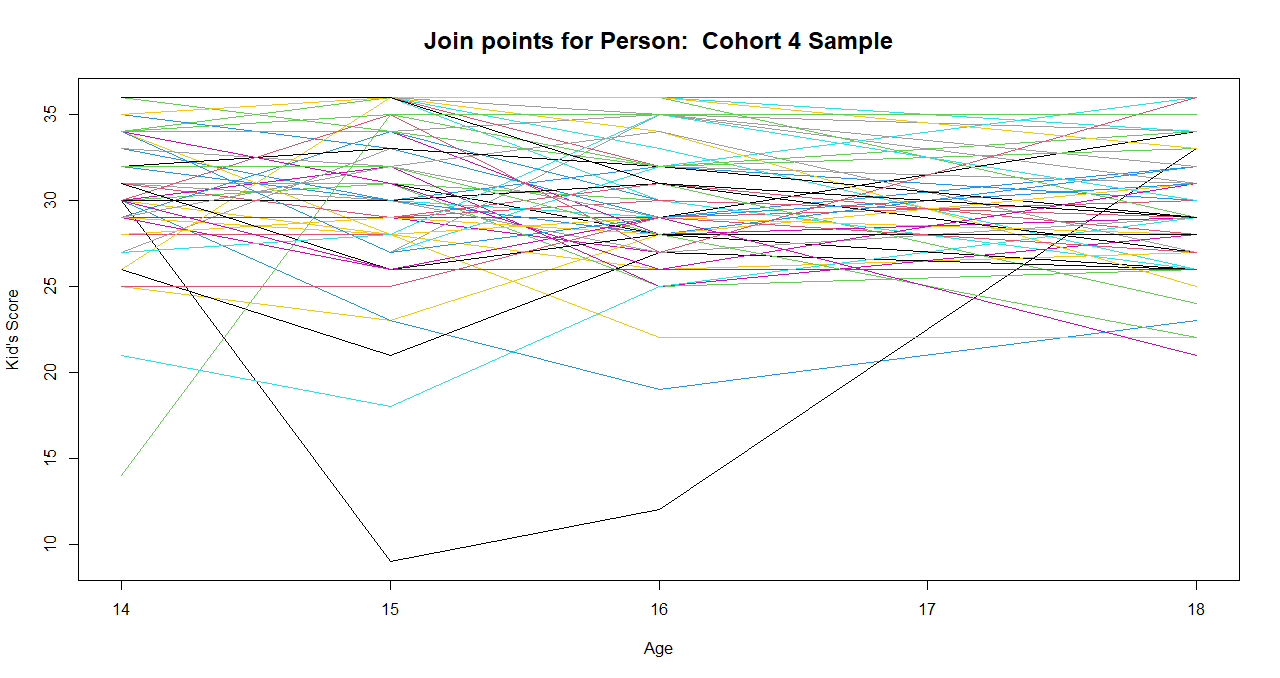
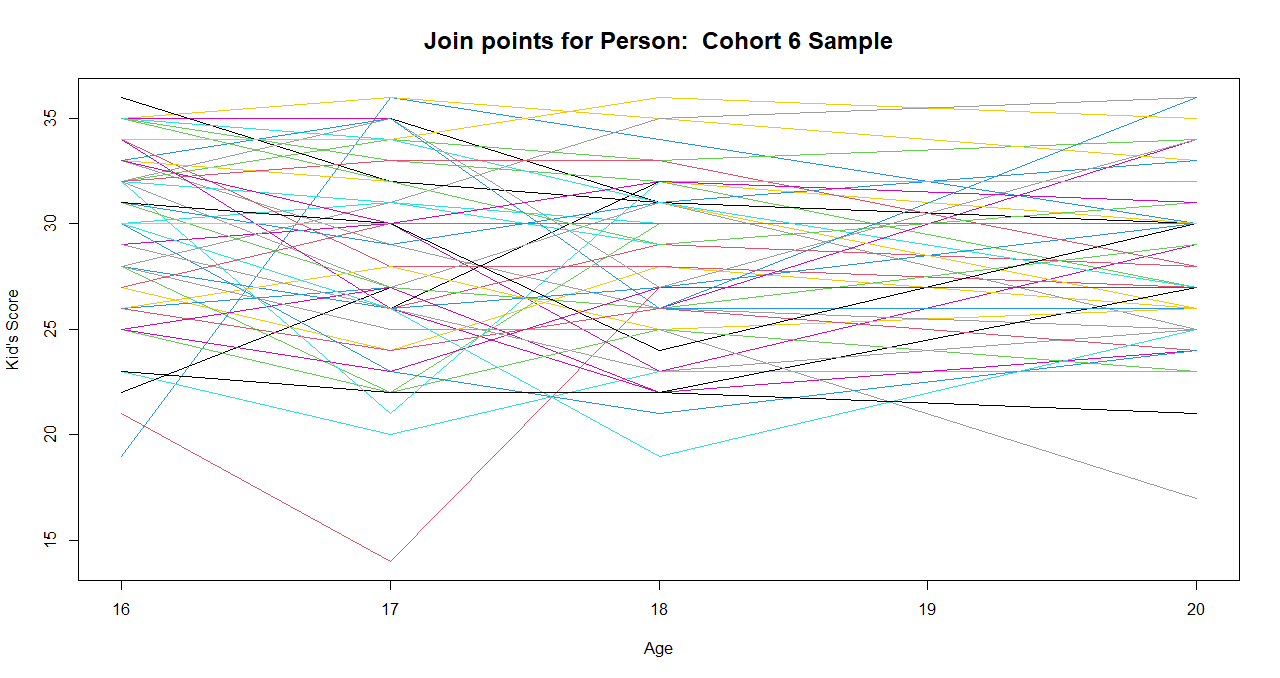
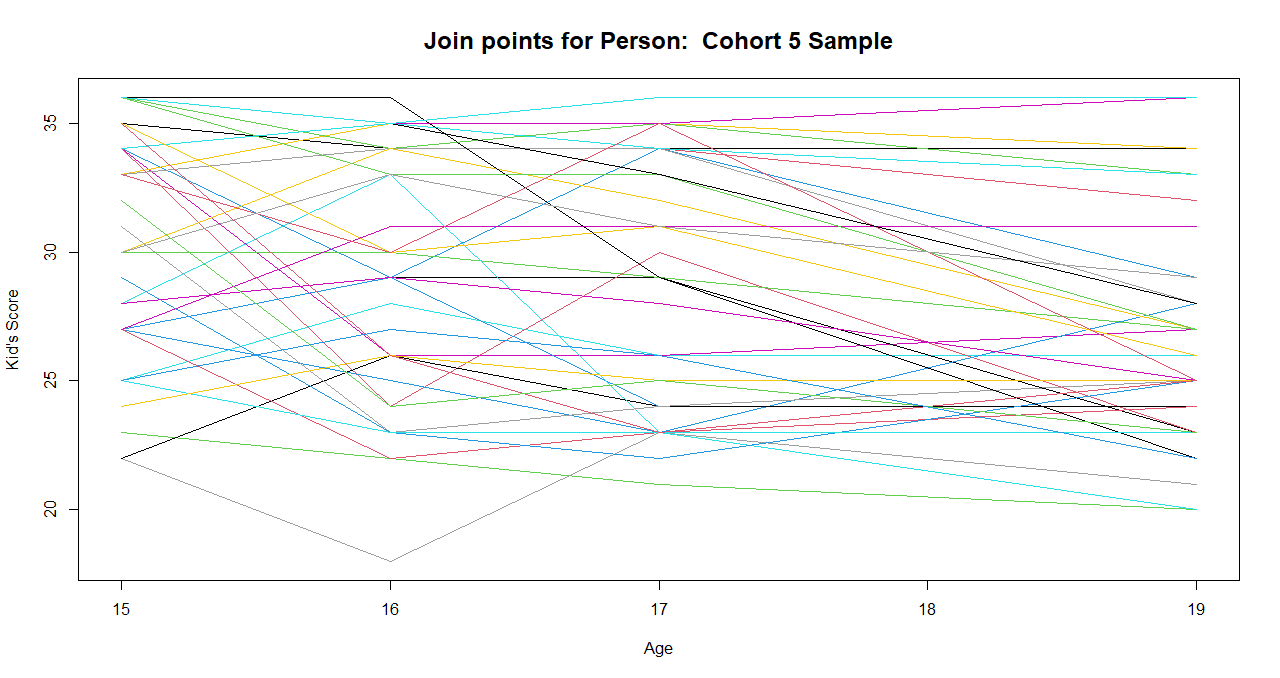
Since the focus of this study is on how kid’s attitude towards deviant behavior changes as they get older, I explored the effects of each covariate on the relationship between age (the main covariate) and youth\_dev (the outcome variable). I first inspected these effects on an individual basis, and then investigated the mean effects across individuals.

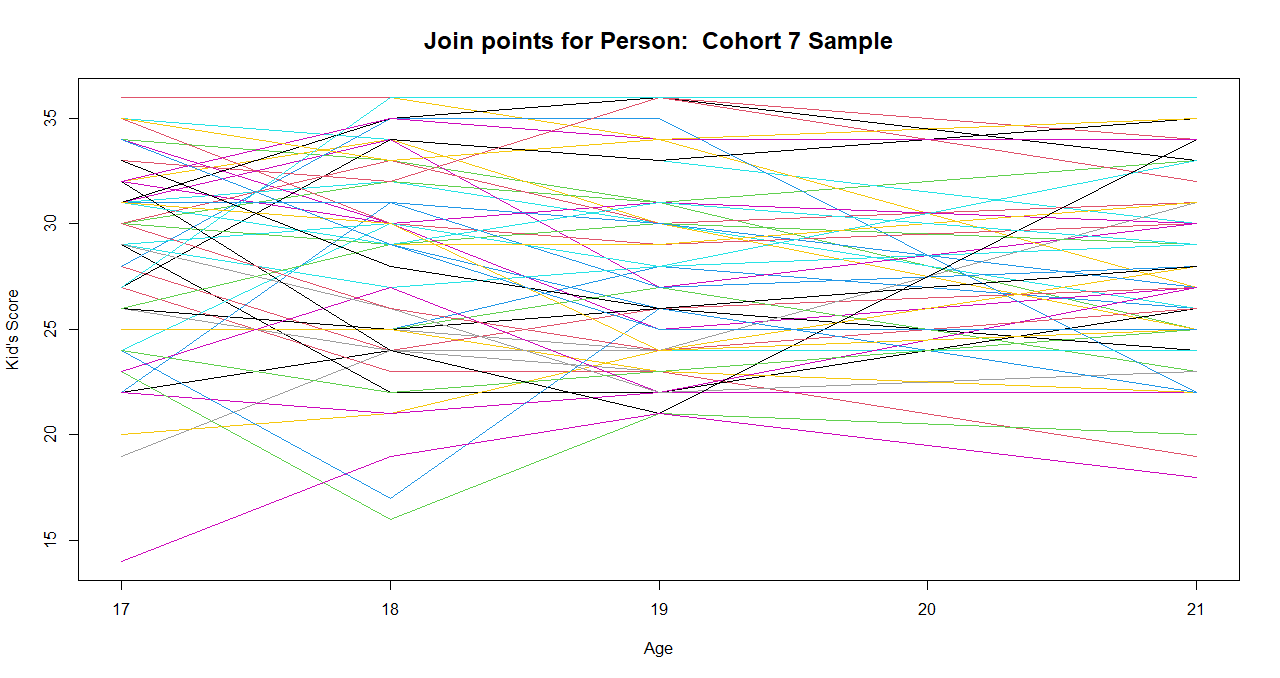
1. Effects of cohort
2. Kid’s attitude score by age: panel plots

First, I split the dataset into 7 separate datasets by cohort. Then, I randomly sampled about 50 subjects per cohort for graphing purposes. Below are 7 panel plots, one for each cohort, for individual kid’s attitude score by age. I plotted both 4 data points (corresponding to 4 collection times) and lines joining those points (see appendix for additional panel plots with linear regressions and splines of 4 data points).

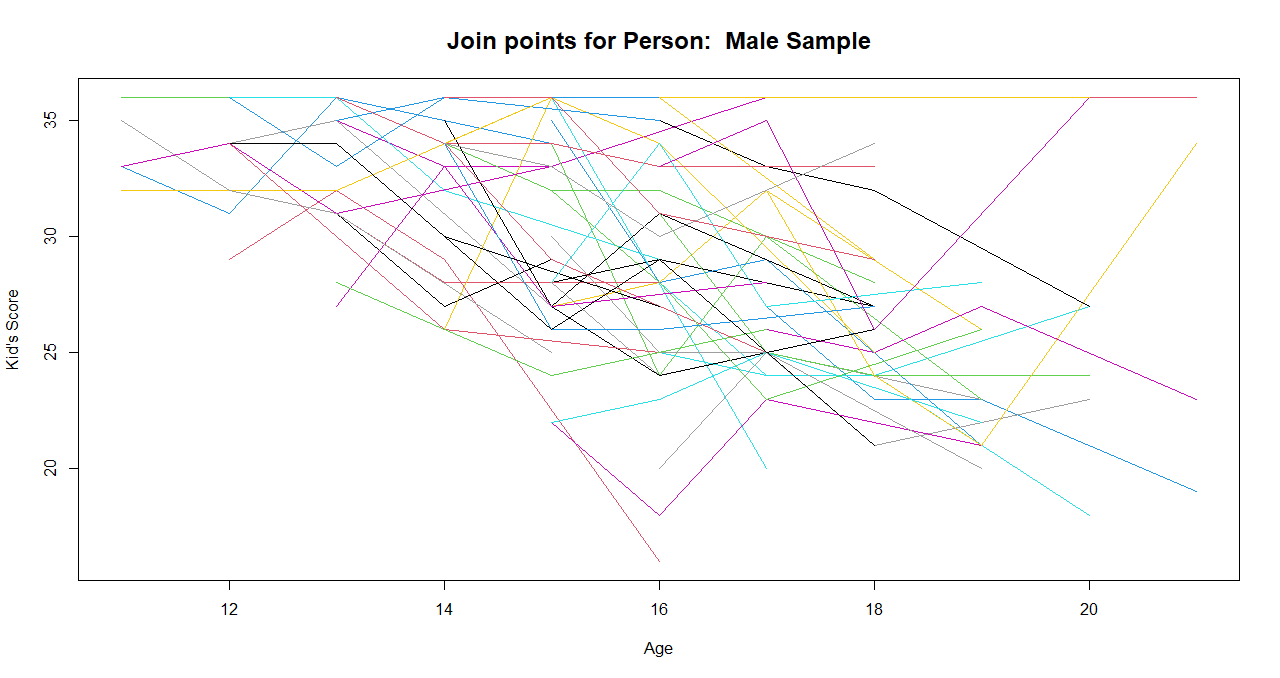


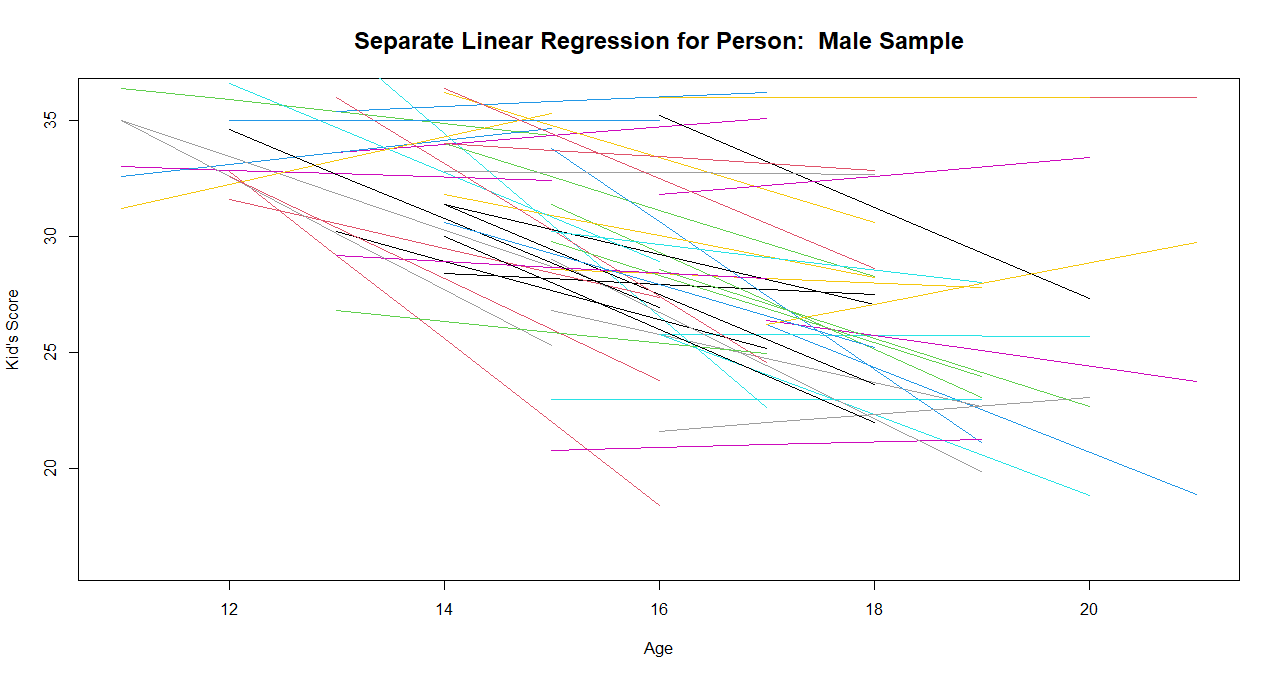
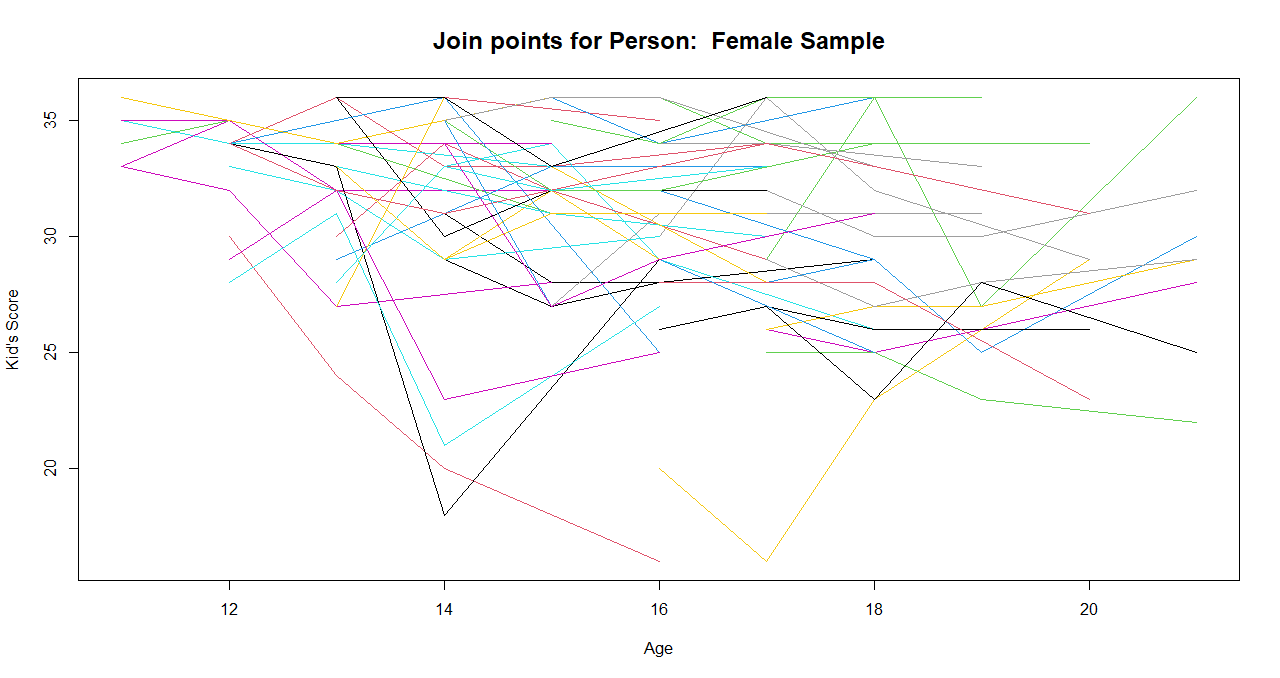
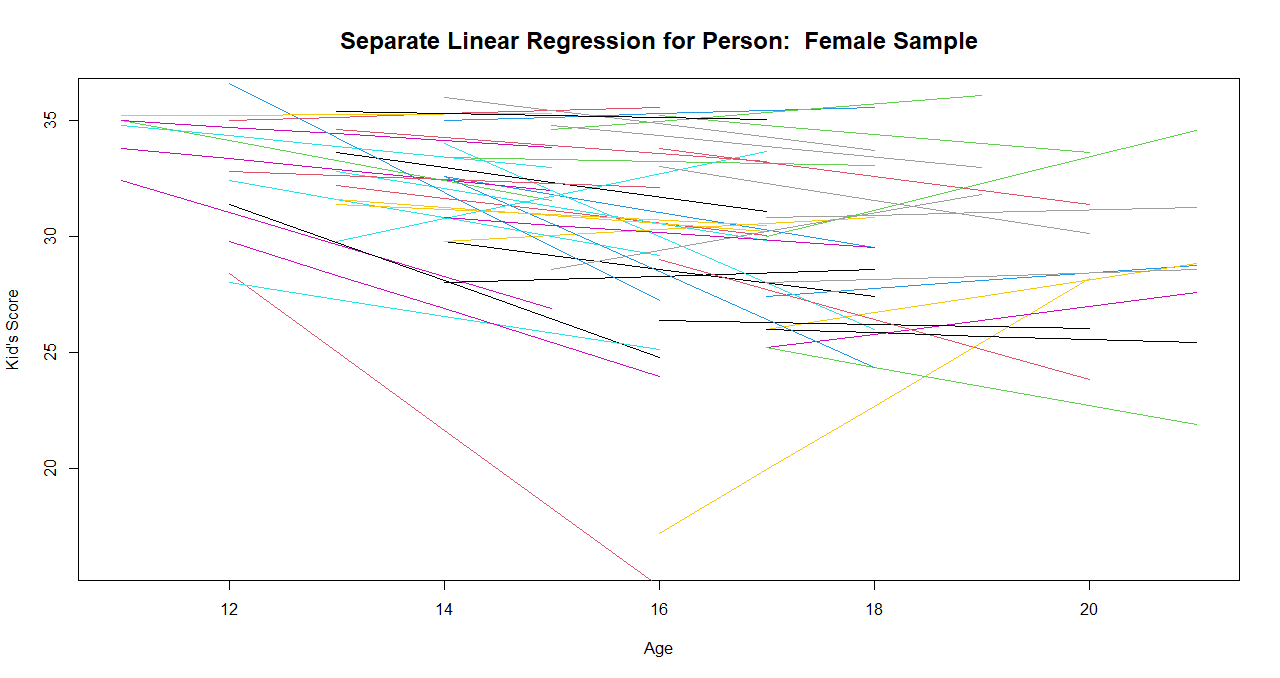
1. Overlay joining points by subject



I also overlayed the joining points for each subject belonging to the same cohort on the same plot to get a better visualization of the general trend of attitude score as a function of age (see appendix for overlay plots of individual regressions with age and regressions with age^2). From the panel plots and the overlay plots for sampled individuals, we can tell that for younger cohorts (cohort 1, 2 and 3), kid’s attitude score generally decreases with age, while older cohorts (cohort 4, 5, 6, and 7) tend to have more variability in trend. Younger cohorts tend to have increasing variance in attitude score as they get older. Within each cohort group, there is also a lot of variability among individuals in the relationship between age and attitude score, suggesting a random effect for age might be appropriate for modeling. Besides, attitude score appears to change in a nonlinear fashion with age, which signals a nonlinear term involving age in the regression model.

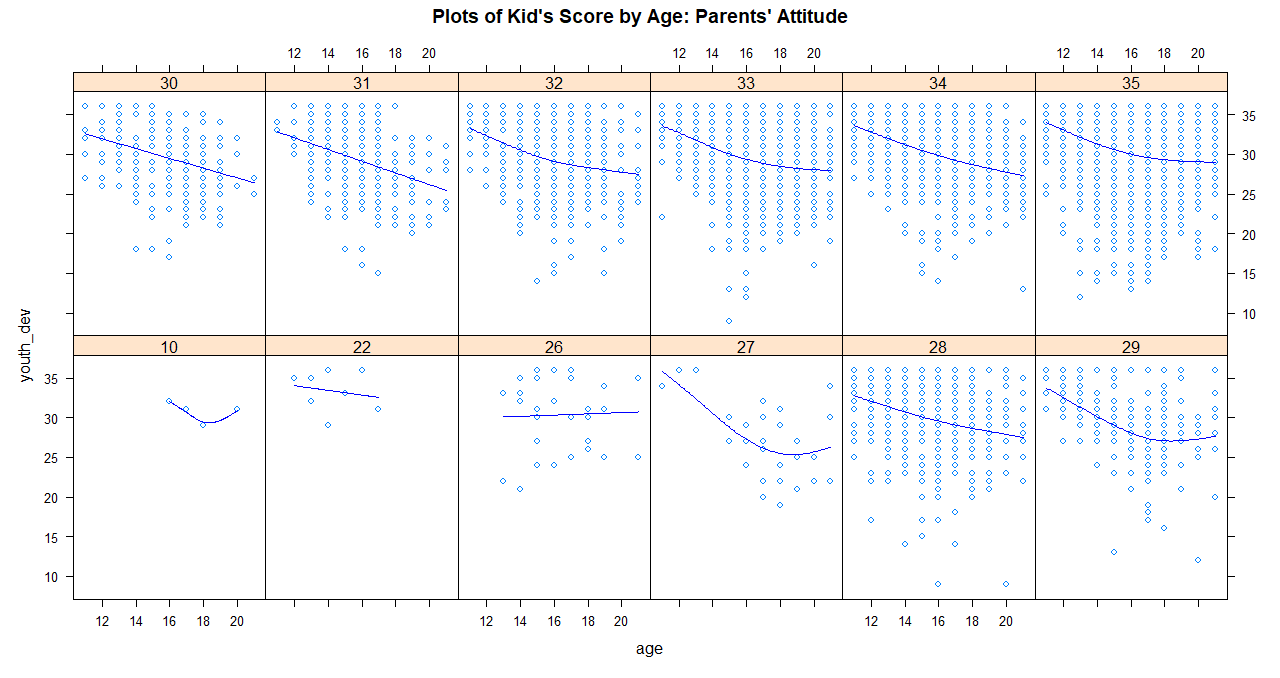
1. Effects of gender

I split the dataset into 2 separate ones by gender. Then, I randomly sampled about 50 subjects per gender for graphing purposes. Similar as above, for each gender, I joined the 4 age points for each of the sampled subject and overlayed them on the same plot. I also included overlay of the individual regression lines with age.



From the plots shown above, we can see that there are nonlinear patterns for both genders, but males tend to have more negative, steeper slopes than females. Clearly, there is a lot of variability in changing patterns within each gender.

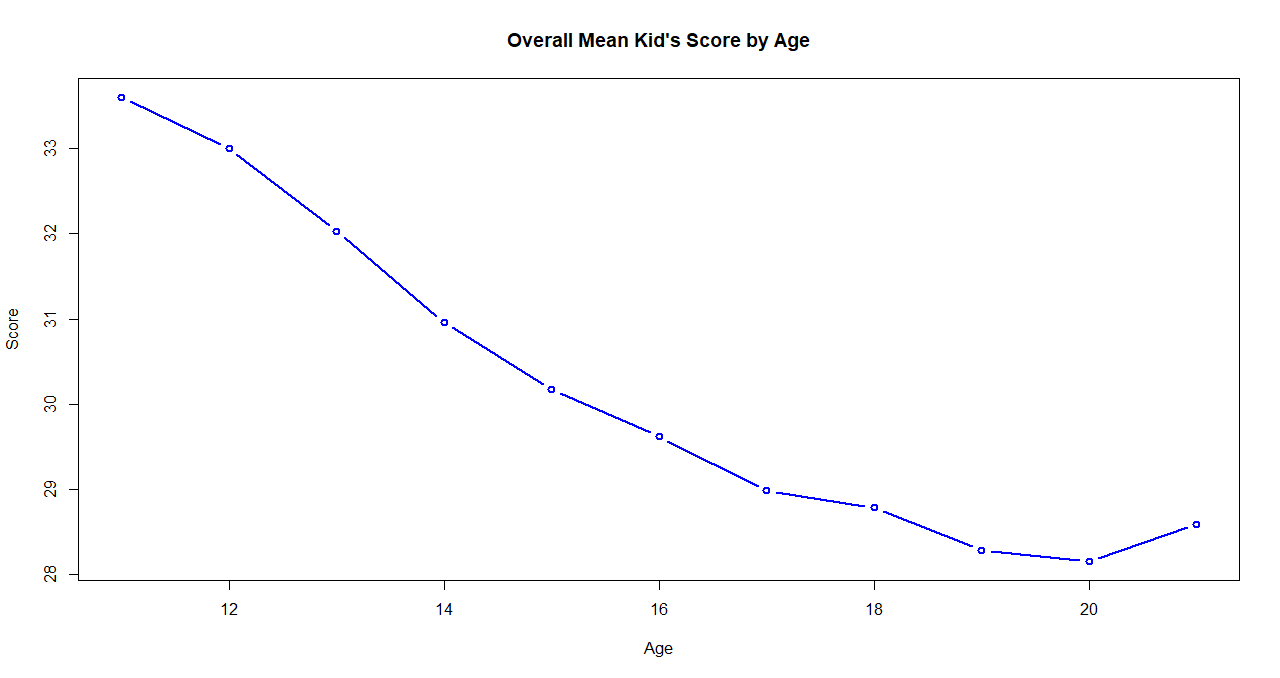
1. Effects of parents’ attitude

Since parents’ attitude scores were only measured once, they are considered invariant in the study. I first grouped kids based on their parents’ attitude scores, and then plotted each kid’s score by age for each group.

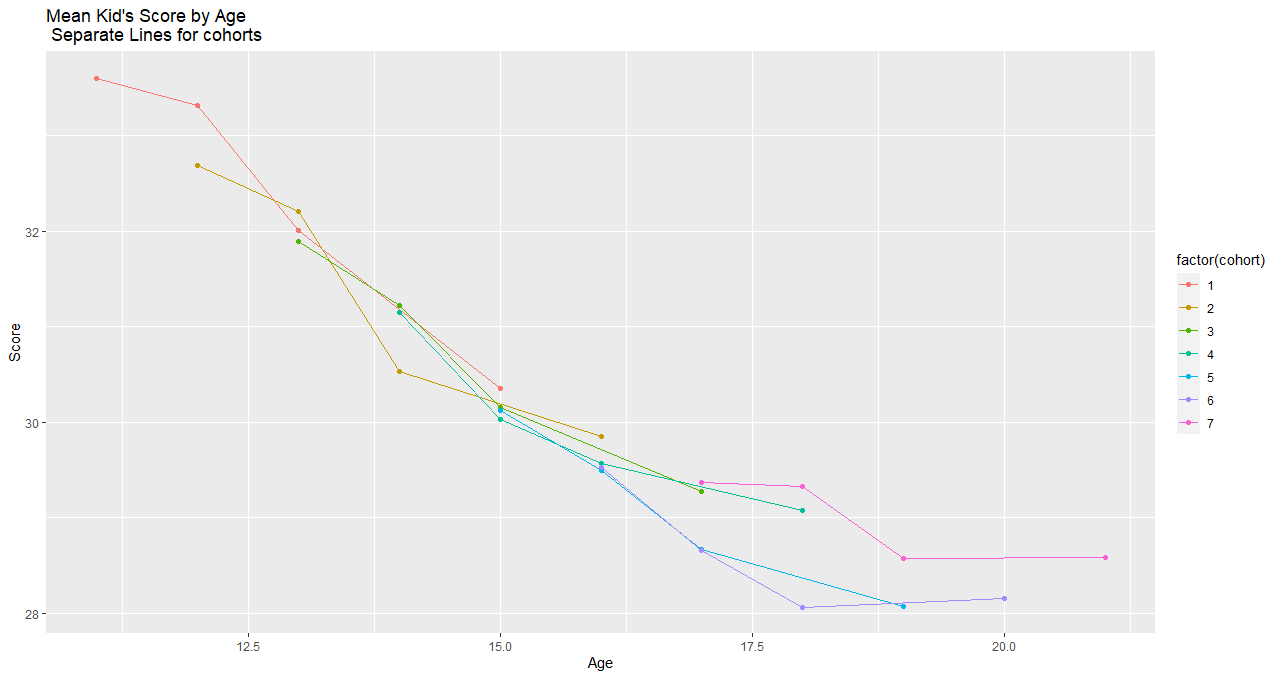
Based on the plots, we can see that kid’s attitude score tends to decrease at a faster rate younger and levels off gradually when getting older if their parents’ attitude score is higher (32-35) (having stronger disapproval of deviant behavior). When parents’ attitude score is lower (30-31), their kid’s attitude score drops more sharply and more linearly with age. Very low parents’ scores are associated with variable trends in terms of their kids’ scores but this may be due to too small sample sizes in those score groups.

1. Exploring mean structures

To better visualize the overall effects of age on kid’s attitude score, as well as the interaction effects of age with other covariates, I created the following plots showing kid’s mean score by age.

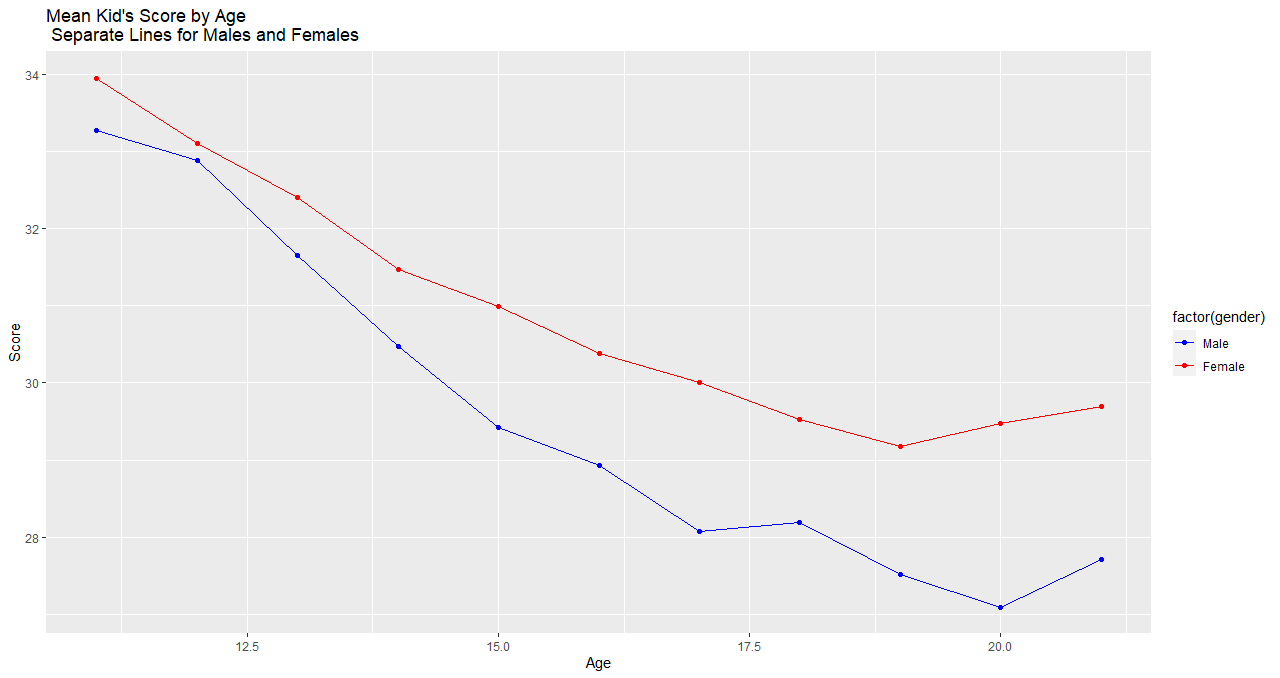
1. Kid’s mean attitude score by age

Overall, there is a nonlinear relationship between age and kid’s attitude score. The mean score decreases first and slightly increases at 20. The decrease is steepest from 12 to 14 and gradually levels off from 17 on.

1. Kid’s mean attitude score by age: separate lines for each cohort

On average, kid’s score decreases with age faster for younger cohorts (cohort 1, 2, and 3) and cohort 5. From cohort 5 on, the mean score decreases slower and eventually flattens around 18 to 19. Cohorts 1-5 have a decreasing value of starting scores and ending scores, while cohorts 6 and 7 have similar starting scores but cohort 7 has a higher ending score than that of cohort 6.

1. Kid’s mean attitude score by age: separate lines for each gender



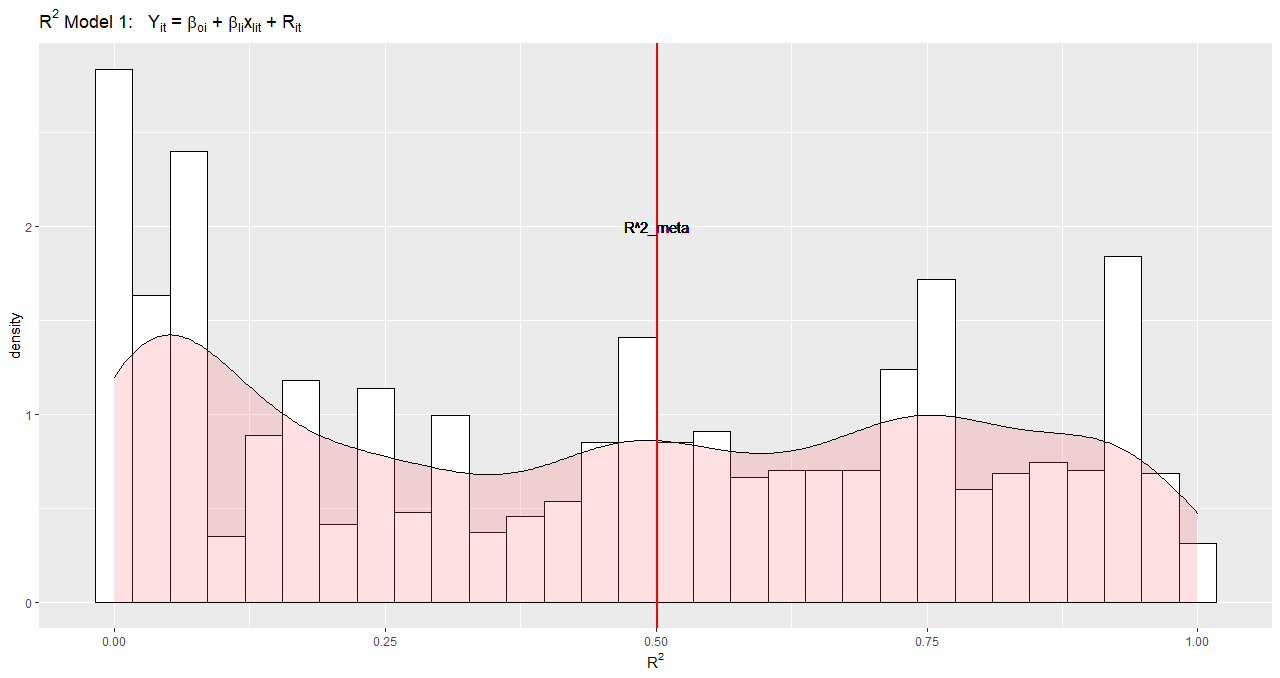
Male kids on average score lower than female ones at any age, and the difference in average scores becomes larger as they become older. Male score decreases more rapidly than female score overall. The mean score starts to rise for both genders after 18, earlier for females.

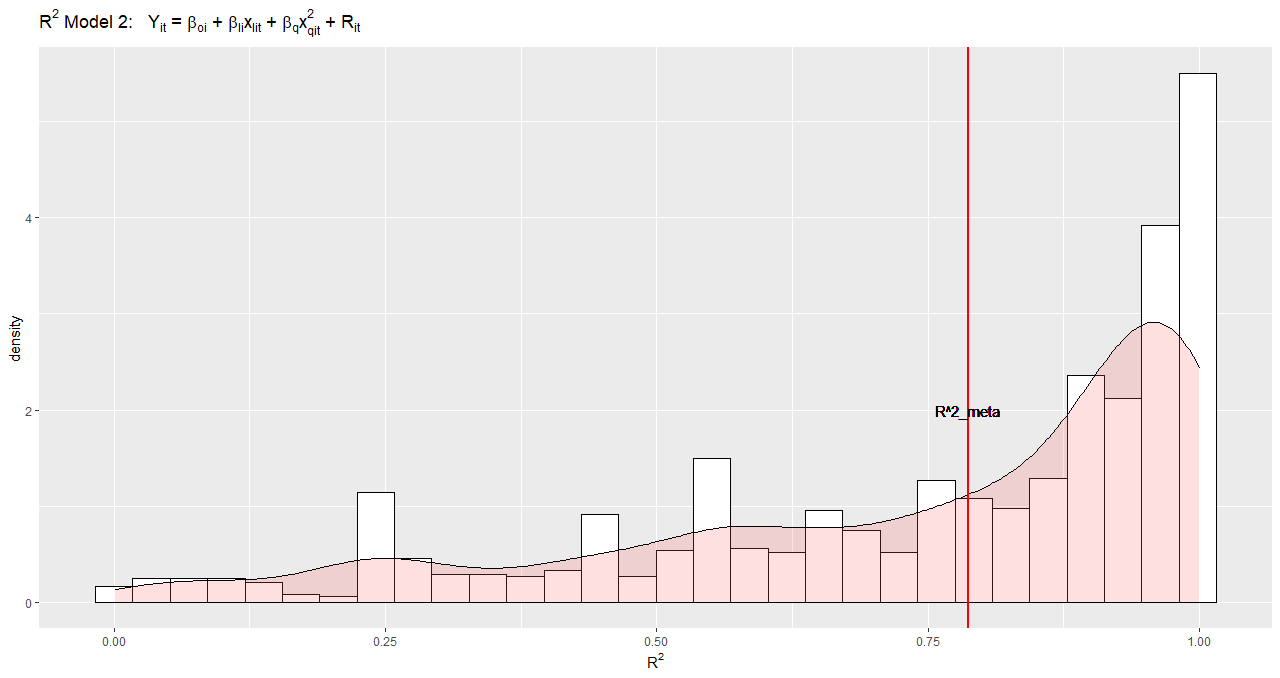
1. Exploring Individual Specific Models

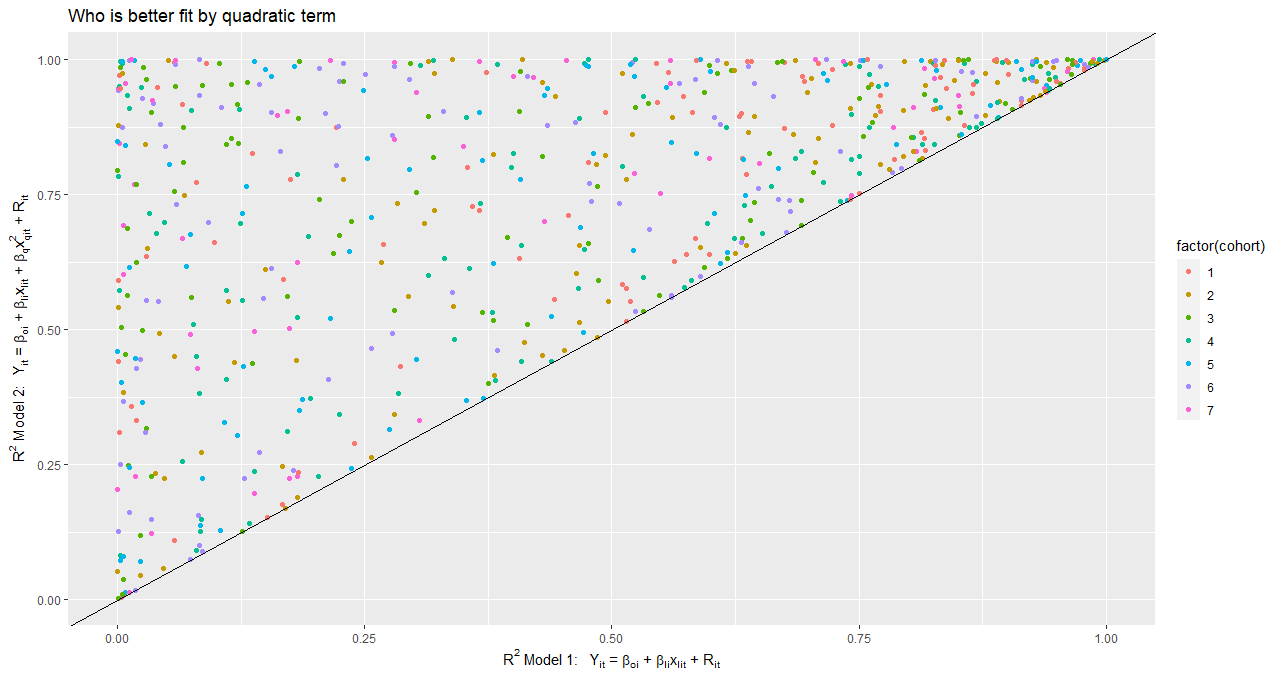
Based on the above exploration of the data, I decided on two plausible linear regression models for level 1:

**Model a**: ,

**Model b**: ,

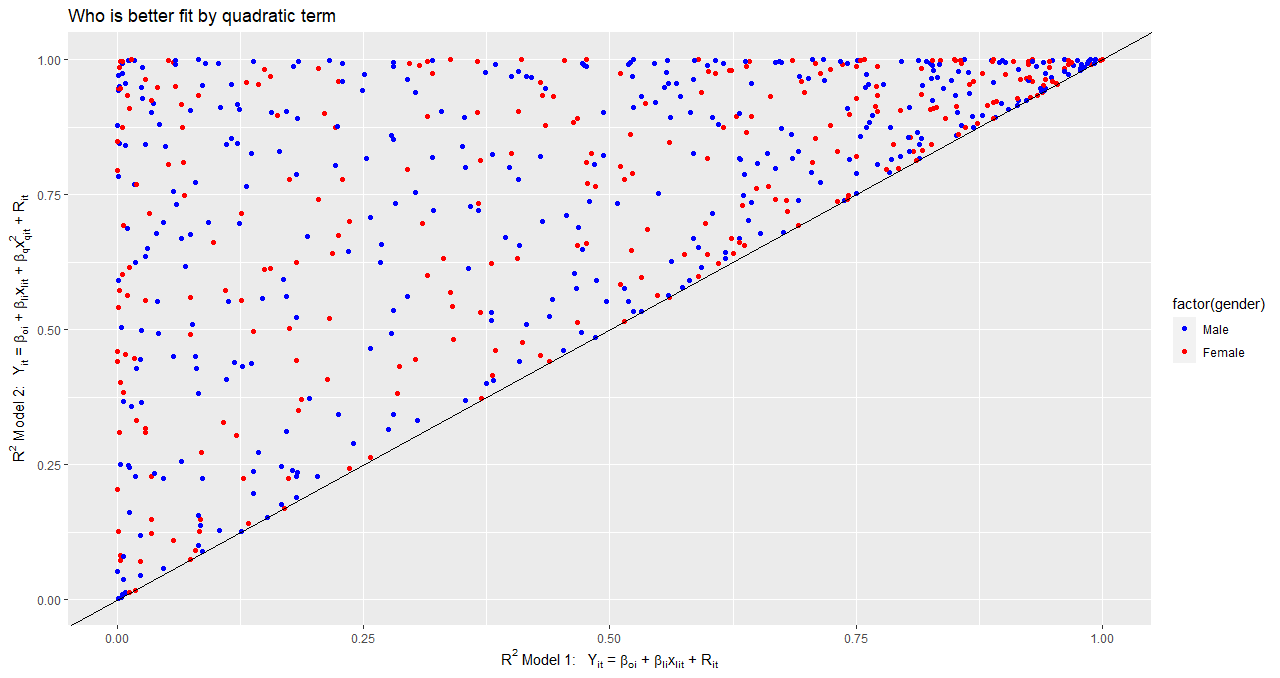
Using ordinary least squares, I fit the two models to each individual’s data separately and computed and to see who improves. I removed individuals with fewer than 4 observations when fitting the models to avoid issues like insufficient data points for fitting as well as overfitting. Below are visualizations of the results.

The above two figures show the distribution of as well as where lies in the distribution for each model. It is evident from the graphs that both and have improved a lot for model b. Thus, I will include the quadratic term for age in the preliminary model. I further investigated who improves in model b by looking at the following figures:

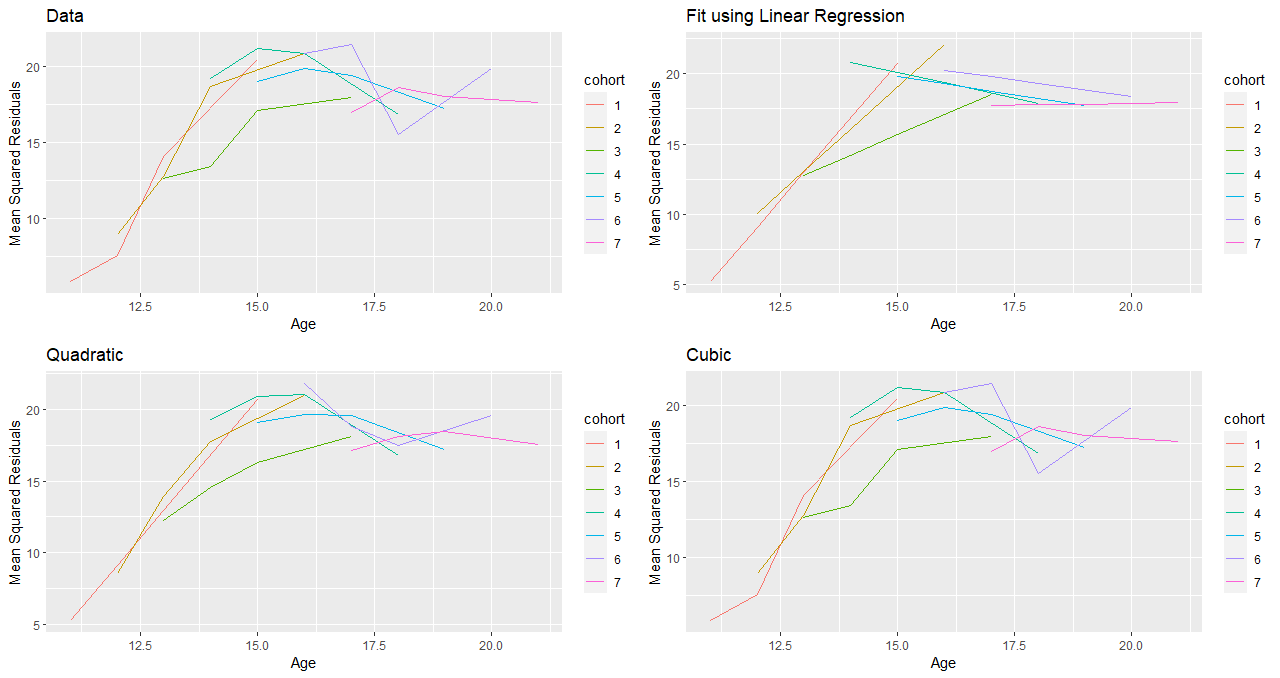
There does not appear to be discernible difference in model performance for individuals in different cohorts or of different genders.

I decided to use the following model as my **Preliminary hierarchical linear model:**

Level 1:

Level 2:

1. Exploring random effects

There is substantial variability between individuals in the relationship between age and kid’s attitude score, so I also considered adding random effects in the model. I looked for possible random effects by investigating the variance functions of Y. To proceed, I computed mean squared residuals by age for each cohort from the preliminary hierarchical linear model and then, fit regression models with linear term of age, quadratic term of age, and cubic term of age, respectively. I plotted the fitted values by age for each cohort, and compared them with the raw data (i.e., mean squared residuals from the preliminary model).

The fitted values for the regression with look exactly the same as the raw data. This is not surprising since there are only 4 data points per cohort, in which case cubic functions should fit the data perfectly. Note also that the regression with also fits well enough with mean squared residuals. Given the preliminary mixed linear model and assuming  and random intercept and slopes for age and , the variance of is a 4th order polynomial function of age. On the other hand, if we assume random intercept and slope for age only, the variance of is a quadratic function of age.

1. Exploring serial correlations

Since the study design is longitudinal in nature, we also need to take into account possible serial correlations between time points. Below are plots of residual correlations between time points for each cohort. We can see that within each cohort, the closer in age, the more correlated in residuals. Also, there are stronger residual correlations among older cohorts.

##### 

##### **Model Fitting**

1. No serial correlation

Built on exploratory analyses, I fit the following hierarchical linear model (my **base model**) with 3 random effects to the data:

Level 1 (within individual):

Level 2 (between individuals):

Below is a comparison table of the base model against the empty model with a random intercept, the model including only main effects of age, age^2, gender, cohort, and parent\_att with a random intercept, the model including both main effects and cross-level interaction effects for age\*gender and age\*cohort with a random intercept, as well as the model including main effects and cross effects plus a random intercept and a random slope for age.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Statistical models | | | | | |
|  | **Null**  **(Model 0)** | **All Main Effects**  **(Model 1)** | **+ Some Cross Effects**  **(Model 2)** | **+ Random Effect for Age**  **(Model 3)** | **+ Random Effect for scaled-Age^2**  **(Base Model)** |
| AIC | **31236.79** | **30460.80** | **30454.05** | **30392.75** | **30227.35** |
| BIC | **31256.74** | **30547.22** | **30587.00** | **30539.00** | **30393.54** |
| Log Likelihood | **-15615.40** | **-15217.40** | **-15207.02** | **-15174.38** | **-15088.68** |
| Var: id (Intercept) | **10.53** | **8.69** | **8.67** | **0.03** | **531.91** |
| Var: Residual | **9.16** | **8.07** | **8.04** | **7.95** | **6.89** |
| Var: id age |  |  |  | **0.04** | **10.88** |
| Cov: id (Intercept) age |  |  |  | **-0.03** | **-75.56** |
| Var: id xagesq |  |  |  |  | **64.65** |
| Cov: id (Intercept) xagesq |  |  |  |  | **179.69** |
| Cov: id age xagesq |  |  |  |  | **-26.28** |

The base model fits the data best as it has the lowest AIC and largest loglikelihood among all models inspected. The residual intra-class correlation or ICC of the null model is 10.53/(10.53+9.16) = 0.53, indicating that over half of the total variance is due to between-individual differences, or the within-individual dependency is over 50%. Then, I performed the following likelihood ratio tests to see whether some random and fixed effects are needed.

To test whether we need a random slope for age^2, I performed a likelihood ratio test for model 3 versus the base model. The test statistic is the difference of the deviance between the full model and the reduced model, which equals 171.3984. The sampling distribution is a mixture of and , and the p-value is 0.5\*(6.349754e-37+6.043678e-38) = 3.477061e-37. So I rejected the null hypothesis and concluded that a random slope for age^2 is needed.

To test whether some fixed effects are needed, I checked both model-based and robust estimation of the standard errors for the fixed effects (see table below).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Fixed Est. | satterthwaite | Model se. | Model t | p-value | Robust se | Robust t | p-value |
| (Intercept) | 55.0856 | 4306.645 | 3.6644 | 15.0326 | 0 | 3.8076 | 14.4674 | 0 |
| age | -3.2704 | 3420.383 | 0.5259 | -6.2191 | 0 | 0.5475 | -5.9731 | 0 |
| xagesq | 6.8243 | 3357.386 | 1.5261 | 4.4718 | 0 | 1.6048 | 4.2523 | 0 |
| gender2 | -1.4654 | 1210.124 | 0.642 | -2.2826 | 0.0226 | 0.6301 | -2.3257 | 0.0202 |
| cohort2 | 0.7377 | 1577.814 | 1.3912 | 0.5303 | 0.596 | 1.4595 | 0.5055 | 0.6133 |
| cohort3 | 2.1198 | 3080.527 | 1.7039 | 1.2441 | 0.2136 | 1.7904 | 1.184 | 0.2365 |
| cohort4 | 1.9709 | 4182.21 | 2.2232 | 0.8865 | 0.3754 | 2.4496 | 0.8046 | 0.4211 |
| cohort5 | 5.0982 | 4243.85 | 2.8107 | 1.8138 | 0.0698 | 2.9929 | 1.7034 | 0.0886 |
| cohort6 | 4.9475 | 4041.912 | 3.4809 | 1.4213 | 0.1553 | 3.6366 | 1.3605 | 0.1738 |
| cohort7 | 6.7535 | 3875.183 | 4.3086 | 1.5675 | 0.1171 | 4.3999 | 1.5349 | 0.1249 |
| parent\_att | 0.1173 | 1343.452 | 0.0331 | 3.5395 | 4.00E-04 | 0.0361 | 3.2518 | 0.0012 |
| age:gender2 | 0.1756 | 1291.217 | 0.0432 | 4.0646 | 1.00E-04 | 0.0425 | 4.1283 | 0 |
| age:cohort2 | -0.0604 | 1998.26 | 0.1098 | -0.5498 | 0.5825 | 0.1163 | -0.5189 | 0.6039 |
| age:cohort3 | -0.1542 | 3197.005 | 0.1264 | -1.2201 | 0.2225 | 0.1333 | -1.1571 | 0.2473 |
| age:cohort4 | -0.1386 | 4121.896 | 0.1555 | -0.8909 | 0.373 | 0.1695 | -0.8176 | 0.4136 |
| age:cohort5 | -0.3526 | 4250.33 | 0.1877 | -1.8786 | 0.0604 | 0.1999 | -1.7636 | 0.0779 |
| age:cohort6 | -0.3394 | 4092.121 | 0.2232 | -1.5203 | 0.1285 | 0.2333 | -1.4546 | 0.1459 |
| age:cohort7 | -0.4033 | 3938.913 | 0.2649 | -1.5225 | 0.128 | 0.2743 | -1.4702 | 0.1416 |

The two estimation methods yield similar p-values and test conclusions for each fixed effect. T-tests indicate that we do not need fixed effects for cohorts or cross-level interaction effects between age and cohorts. As a result, I fit a reduced model that does not include any effect for cohort and carried out a likelihood ratio test for the full model against the reduced model. The test statistic is 30186 – 30177 = 9, with df = 12, and p = 0.7237. So I retained the null hypothesis and concluded that I should drop all effects for cohort. The global fit measures and covariance parameters for the reduced model and the base model are as follows:

|  |  |  |
| --- | --- | --- |
| Statistical models | | |
|  | **Base Model** | **No Cohort**  **(Model 5)** |
| AIC | **30227.35** | **30212.11** |
| BIC | **30393.54** | **30298.53** |
| BIC.new | **30383.84** | **30288.82** |
| Log Likelihood | **-15088.68** | **-15093.05** |
| Var: id (Intercept) | **531.91** | **532.65** |
| Var: id age | **10.88** | **10.88** |
| Var: id xagesq | **64.65** | **64.65** |
| Cov: id (Intercept) age | **-75.56** | **-75.63** |
| Cov: id (Intercept) xagesq | **179.69** | **179.76** |
| Cov: id age xagesq | **-26.28** | **-26.28** |
| Var: Residual | **6.89** | **6.90** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Following are the estimated fixed effects and their standard errors for the reduced model:

|  |  |
| --- | --- |
| Statistical models | |
|  | **No Cohort**  **(Model 5)** |
| (Intercept) | 50.31 (1.98)\*\*\* |
| age | -2.52 (0.21)\*\*\* |
| xagesq | 4.53 (0.52)\*\*\* |
| gender2 | -1.46 (0.64)\* |
| parent\_att | 0.12 (0.03)\*\*\* |
| age:gender2 | 0.17 (0.04)\*\*\* |
| \*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05 |  |
|  |  |
|  |  |
|  |  |

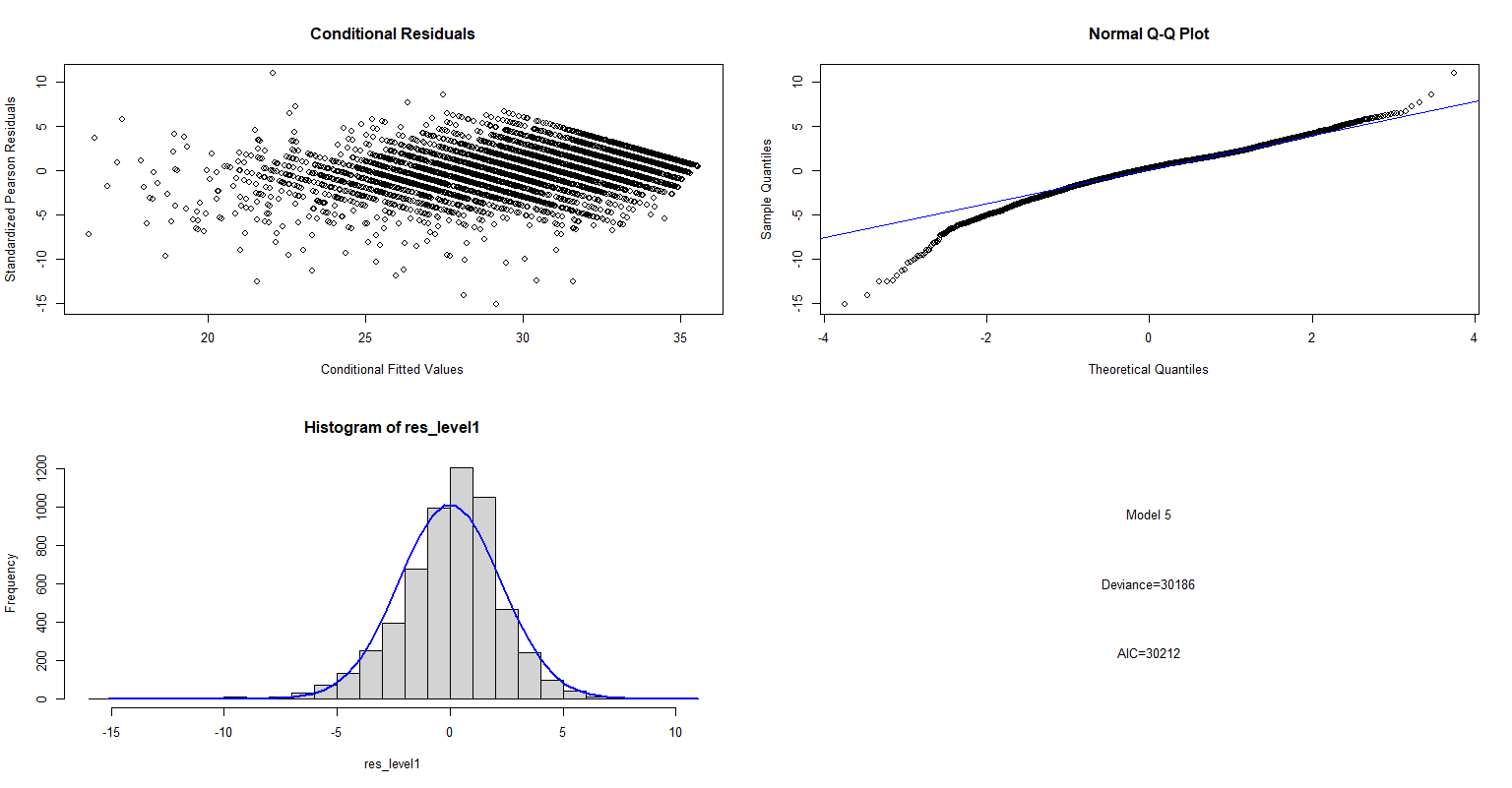
1. With serial correlation

Developed on the final model I refined above, I fit three types of models with autocorrelated errors to account for sources of possible variance within individuals. I tried fitting the linear mixed model with first order autoregressive errors (AR1), first order moving average (MA1), and first order autoregressive-moving average (ARMA(1,1)), respectively. However, none of them converged. After a couple of trials and failures, I eventually succeeded in model convergence by dropping the random slopes for age and age^2 first and fitting the linear mixed model with random intercept only with AR1, MA1, and ARMA(1,1), separately. Then, I attempted to drop the fixed and random effects for age^2, added back the random effect for age, and fit the model with AR1, which managed to converge as well. To test whether a random intercept is needed, I performed a likelihood ratio test of the reduced model versus the full model. I obtained a test statistic of 6.101145, and the p-values for and are 0.04733182 and 0.01350943, respectively, so the resulting p-value is 0.03042063. I decided to reject the null hypothesis and keep the random intercept. Below are global fit statistics of all the models with autocorrelated errors that I fit, the empty model (Model 0), and the final model without any autocorrelations (Model 5).

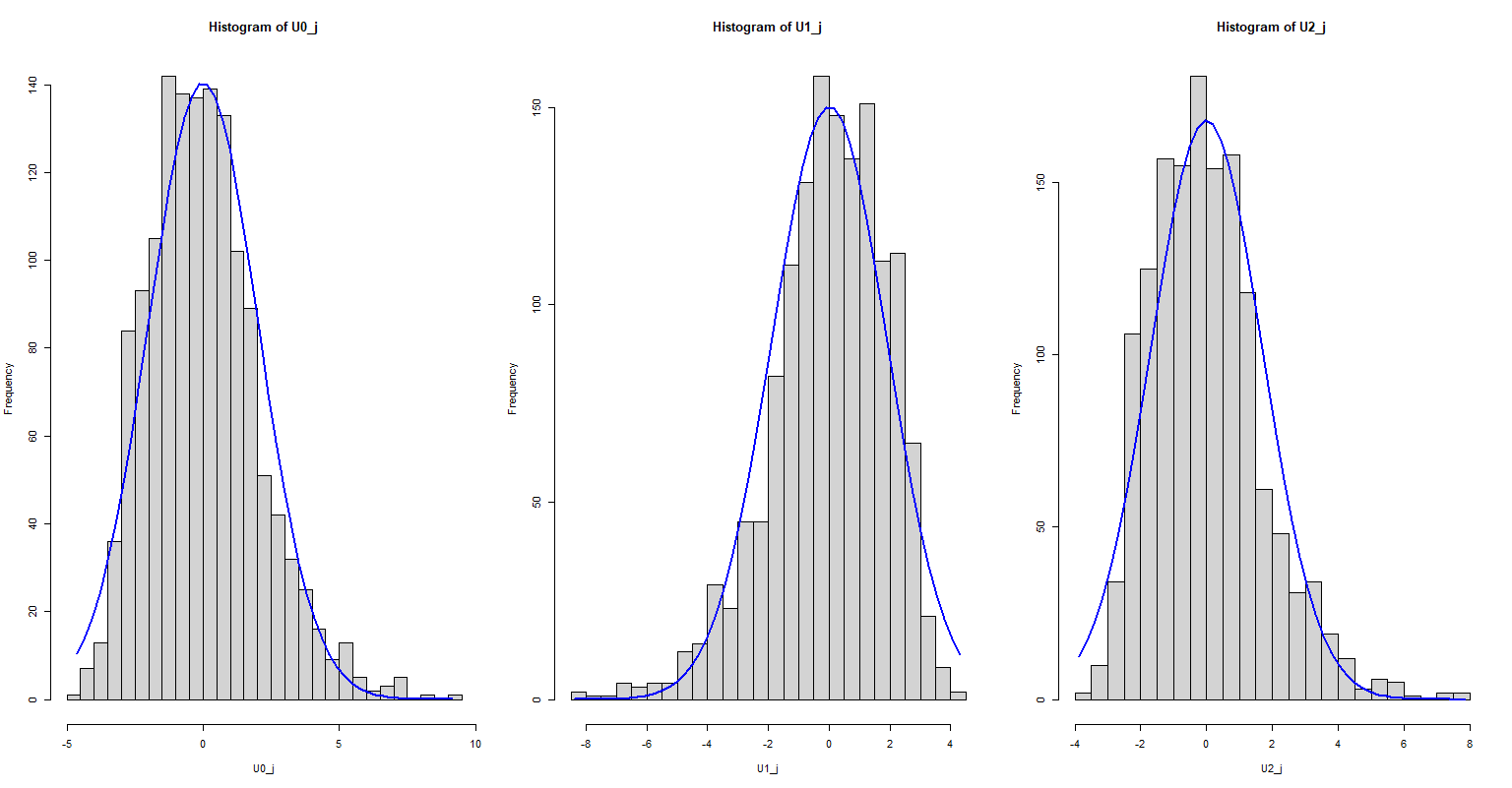
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Statistical models | | | | | | | |
|  | **Null**  **(Model 0)** | **Model 5** | **AR1 w/ U\_0i** | **MA1 w/ U\_0i** | **ARMA(1,1) w/ U\_0i** | **AR1 w/ U\_0i and U1\_i** | **AR1 w/ U\_1i** |
| AIC | 31236.79 | 30212.11 | 30346.61 | 30372.15 | 30334.84 | 30341.27 | 30343.37 |
| BIC | 31256.74 | 30298.53 | 30406.43 | 30431.97 | 30401.31 | 30407.73 | 30396.54 |
| Log Likelihood | -15615.40 | -15093.05 | -15164.31 | -15177.08 | -15157.42 | -15160.63 | -15163.68 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

All global fit measures support the final model without any autocorrelations as the best model. Among models with autocorrelated errors, ARMA(1,1) is the best fitting one. Below are the estimated covariance matrix and correlation matrix for Y based on this model:

1. Model diagnostics

Based on the above model analyses, I decided to choose Model 5 as my final model. Before providing a complete interpretation of it, I conducted some model diagnostics to examine the assumptions of the model. Specifically, I looked at both level 1 residuals and level 2 residuals to see if they follow normal distributions. Below are some graphical displays of the conditional residuals of :

The following are histograms of , with normal curves overlay:



All 4 residuals look to follow normal distributions well.

##### **Model Interpretation**

Parameter estimates from my final model are:

|  |  |
| --- | --- |
|  | Final |
| (Intercept) | 50.31 (1.98)\*\*\* |
| age | -2.52 (0.21)\*\*\* |
| xagesq | 4.53 (0.52)\*\*\* |
| gender2 | -1.46 (0.64)\* |
| parent\_att | 0.12 (0.03)\*\*\* |
| age:gender2 | 0.17 (0.04)\*\*\* |
| AIC | 30212.11 |
| BIC | 30298.53 |
| Log Likelihood | -15093.05 |
| Num. obs. | 5696 |
| Num. groups: id | 1424 |
| Var: id (Intercept) | 532.65 |
| Var: id age | 10.88 |
| Var: id xagesq | 64.65 |
| Cov: id (Intercept) age | -75.63 |
| Cov: id (Intercept) xagesq | 179.76 |
| Cov: id age xagesq | -26.28 |
| Var: Residual | 6.90 |
| \*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05 | |

**Hierarchical model:**

Level 1 (within individual):

Level 2 (between individuals):

**Estimated linear mixed model:**

On average, a kid’s attitude score toward deviant behavior is expected to be

1.46 points higher for males

0.12 points higher for 1 unit increase of parent’s attitude score

2.52 points lower for 1 year increase of age

4.53 points higher for 1 standard deviation increase of age^2.

The effect of age on kid’s attitude score is

That is, for 1 year increase of age, a kid’s attitude score regarding approval of deviant behavior is expected to change by points. Girls have flatter negative slopes – or their scores drop slower than boys as they grow up. Remember that higher attitude scores correspond to stronger disapproval (or less approval) of deviant behavior.

There are large differences between kids in terms of the intercepts and slopes for age and age^2 in the model.

##### **Conclusion and insights**

Through a thorough exploration and a multilevel modeling analysis of the National Youth Longitudinal Study data, I found that a kid’s attitude toward deviant behavior generally becomes more lenient (that is, they tend to tolerate and approve deviant behavior more) as they grow older. However, this change is not linear. The attitude score (higher means stronger disapproval of deviant behavior) tends to drop faster at younger ages, and the decreasing rate slows down as the kid grows up. The attitude eventually stabilizes as the kid reaches adulthood. The change in attitude with age is also affected by the kid’s gender. Boys tend to grow more leniency toward deviant behavior faster than girls as they grow up. Parents’ attitude regarding deviant behavior has a positive impact on their kid’s attitude as well. If they have stronger disapproval of those behaviors, so do their kids. There does not seem to be a statistically significant cohort effect on kid’s attitude. Among all 7 cohorts studied, kids’ attitude changing patterns do not differ much from one another. Nevertheless, there is a lot of variability between kids in terms of their attitude at the beginning of the study and how their attitude changes with their age.