Choir Performances with Conductors' Vocational Details and Demographic Differences

Teaching Fellow - Minh Thu Bui
Supervisor - Professor Masanao Yajima
Team Members - Yan Wang, Mingrui Du, Yueqi(Charlene) Jin

1 Project Background and Objectives

Jacob Wittkopp is a candidate in the Doctor of Musical Arts program at Boston University, Our client's research target is conductors who lead choirs in schools and is interested in studying the affect the conductors' background information such as vocational details and demographic information in the outcomes of choir performances (success or failure).

The research is conducted through a questionnaire of two parts: demographic information and main choir survey. The first part includes participants' educational background, years of experience, classes, choral program size, school stage, school type, ethnicity, race and gender. Open blanks are set in years, classes (number of ensembles), school stage and gender to collect demographic/numeric data or to de-qualify people not teaching at school, such as those who conduct in community or church. The second part consists of two pages, each page has one blank for choir performance description, and a 9-point subscale to rate factors. Two pages are identical except the first asks for the most successful while the latter recalls most unsuccessful. Participants first write about the performances and summarize reasons of success/failure, then rate 12 statements on the scale 1-9 by the extent they believe factually correct. Each scale is designed following a literature in Causal Dimensionality Scale (McAuley et al.). The questions are grouped into four dimensions by summing the scores in those selected questions: Question 1, 6, 9 for locus of causality; 5, 8, 12 for external control; 3, 7, 11 for stability, and 2, 4, 10 for personal control.

For the research, the clients are interested in answering three questions: The first question is to investigate whether there are significant differences in how choral conductors in educational settings attribute their successful and unsuccessful performances, as measured by dimensional scores. The second question is concerned about the attributional score differences correlated with performance outcome (success vs. failure), vocational details, and/or demographic differences? The first two research questions have been completed by the client, and thus, the third question is our primary objective: Can differences in performance outcome, vocational details, and/or demographic differences be used to predict dimensional scores? In other words, our goal is to discover which factors predict success locus score, success internal controllability, success external controlability, and success stability. Then analogous testing for each of the failure subscores.

2 Data Cleaning and Organizing

Questionnaires are posted online through several music organizations/institutes, having reached around 300 participants. However, some of the answers are discovered not working at any level of schools, as a result, 167 observations remained after filtering. The data consists of 167 rows (observations) and 50 columns (variables). The first 24 columns are 9-pt responses from the two subscales ("semantic differential scale" in terminology), values from 1 to 9. Columns 25-32 are scale scores summed into 4 dimensions (locus, internal control, external control, stability). The client removes Stab3 from calculation to improve Cronbach's alpha.

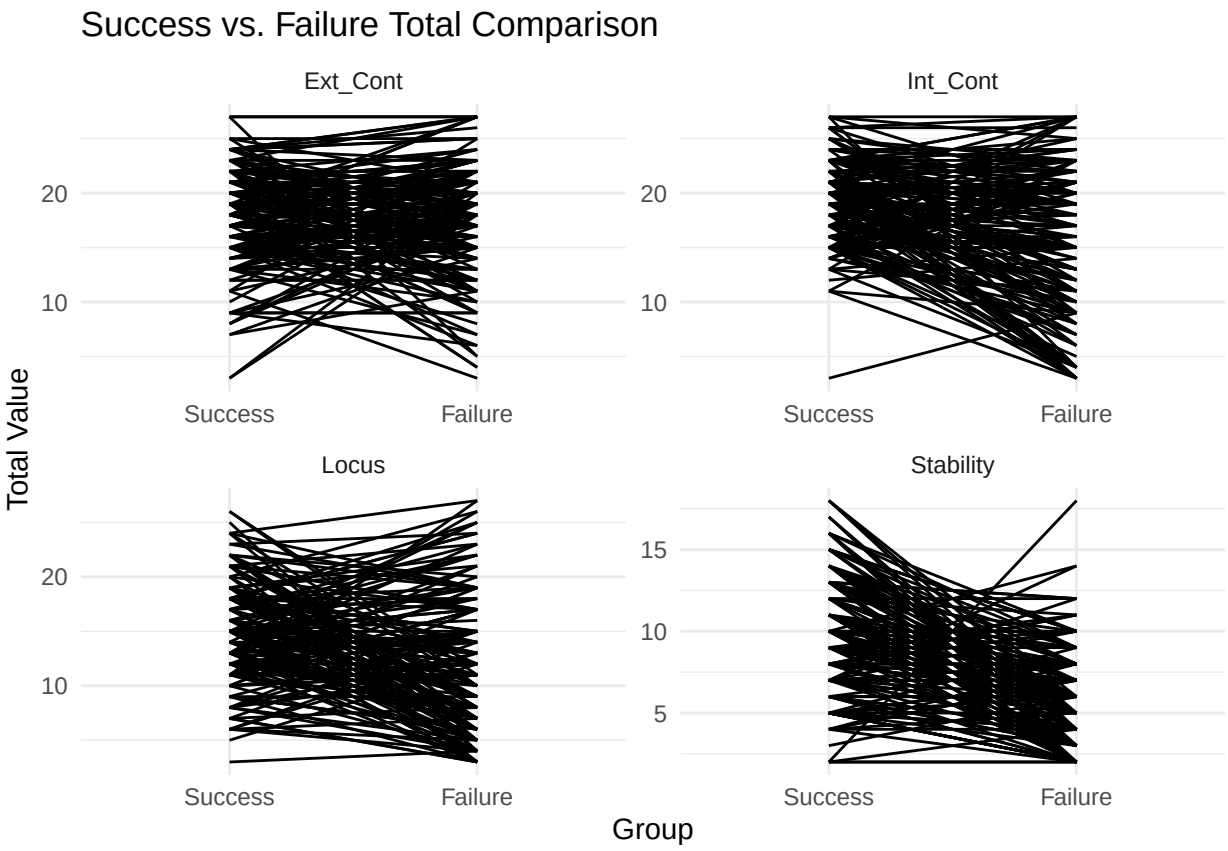
Although seemingly numeric, this part in data should not simply be treated as discrete numbers. Columns 33-40 are independent demographic info or job info, including ordinal variables (size), discrete numeric (years at school, number of ensembles), categorical (degree, school type, gender, race, ethnicity, school level). Column 41-50 are for Z-scores, which will not be considered for our analysis since they are irrelevant to our current objective.

The predictors that we consider in our analysis are as follows. For "Size of Program", the values are ordinal, containing less than 50, 51–100, 100–150, 151+. We also have a column to survey which type of institutions those groups were based from, "Public vs Private", which is binomial categorical. Next is the "Number of Ensembles" for how many choirs participant conducted, presented as discrete numeric values. "Years at School" is a column that tell us how long participant has been at current institution. The next column is "Degree earned", which is nominal categorical that are have some participants marked more than one, and only one marked "non-degree", so these were designated as: Music Education Degree, Music (non-Music Ed) degree, or both. "School level" categorizes the level of education that those choirs were conducted, for example, elementary and middle/junior high to produce K–8 and 2-yr, university, and conservatory are denoted as post-secondary. This is treated non-ordinal categorical. In the "Gender" column, the client left "Prefer not to answer" and "Comment" as options with low responses. All others either male or female. There were a few non-White races represented, and some that marked "White" and one other, so we have two groups, "At least one non-White selected" as n=9 and "Only White selected" as n=155.

3 Exploratory Data Analysis (EDA):

3.1 Comparisons between Success and Failure subscales:

We examined the relationship between success and failure ratings across four dimensions to draw comparisons.



The visualizations show an inverse relationship between success and failure ratings. Specifically, individuals who score highly in areas of perceived success tend to show lower ratings in perceived failures and vice versa, indicating that there might be some intuitive dependencies between the two surveys. Hence, even though we are treating the subscales in success to be independent of that in failure for modeling purposes, the results should be interpreted carefully given the dependence between success and failure in all four dimensions.

4 Modeling:

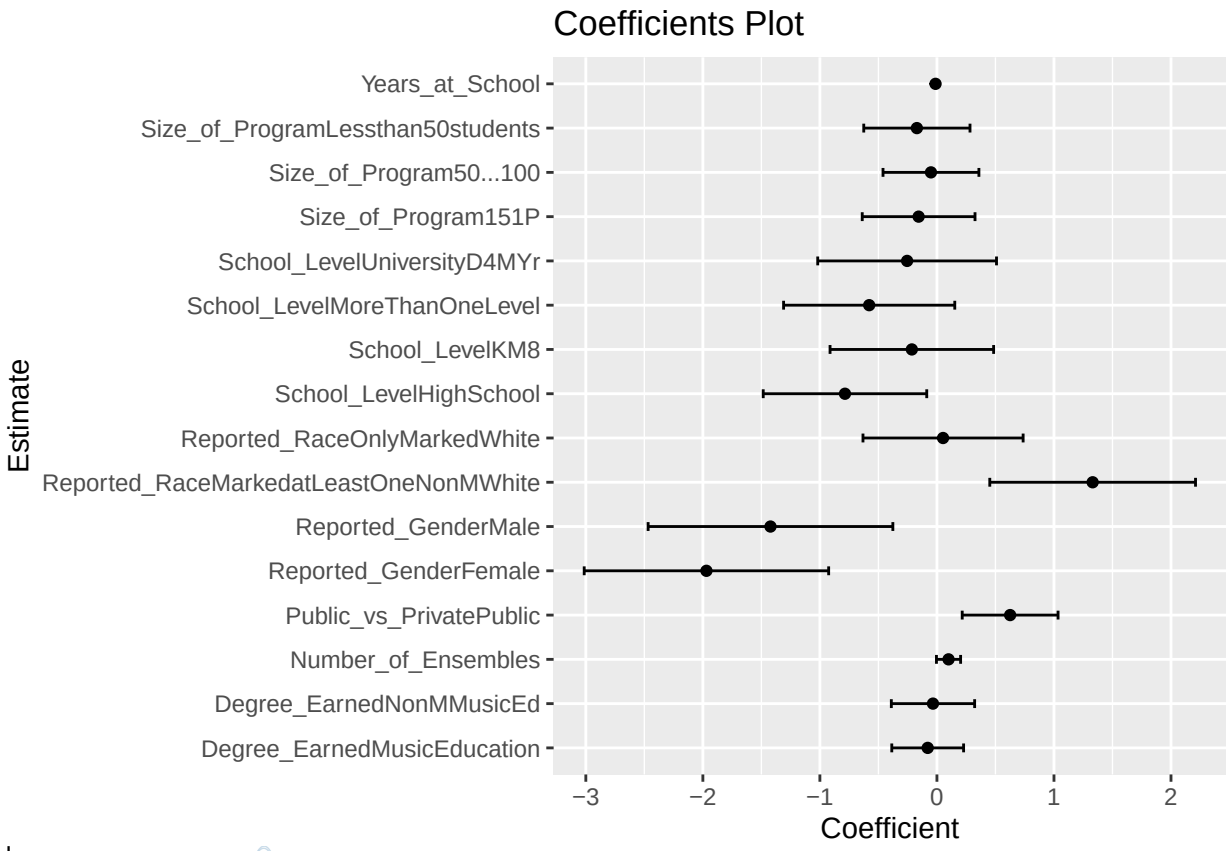
Note that while all participants answered both prompts (success, then failure), the events for each prompt are not presumed to be related to one another. In our analysis, we aim to investigate how demographic factors and vocational details influence the score dimensions (locus of causality, stability, internal control and external control) of success and failure. To achieve this, we employ a mixed effects model that accounts for both fixed and random effects within our dataset. By incorporating demographic information and vocational details as random effects, we can assess their direct impact on the measures of success and failure. Simultaneously, the model's capability to include fixed effects some other predictors, depending on which subgroup, helps us consider the variability attributed to unobserved heterogeneity among subjects. Through this model, we aim to discover which predictors are significant for success and failure subgroups, hence providing insights for the client's research question.

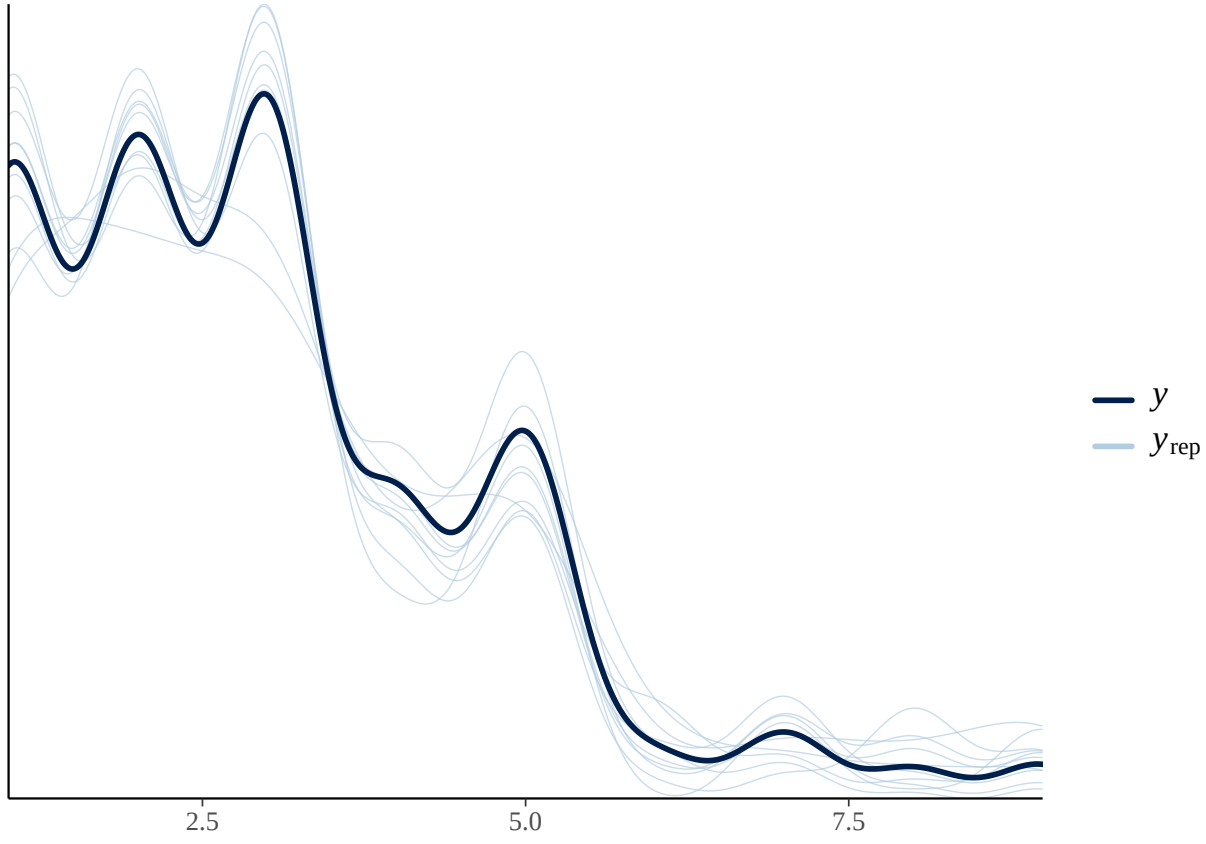
We create a mixed-effect model with fixed effects for the stability score in failure. In R, the package is brm(). It is a function provided by the rstanarm and brms packages in R, which is used for fitting generalized linear mixed-effects models (GLMMs) using the Stan probabilistic programming language. brm() specifically fits GLMMs using a Bayesian framework, which allows for the estimation of posterior distributions of model parameters and the function fits in the model by using Markov chain Monte Carlo (MCMC) sampling methods implemented in Stan to estimate the posterior distribution. We consider logit link in the model and the response variable is "Score", which ranges from 1 to 9. There are 8 regressions in total, corresponding to each of the dimension in the success and failure groups, presented in each subsection below.

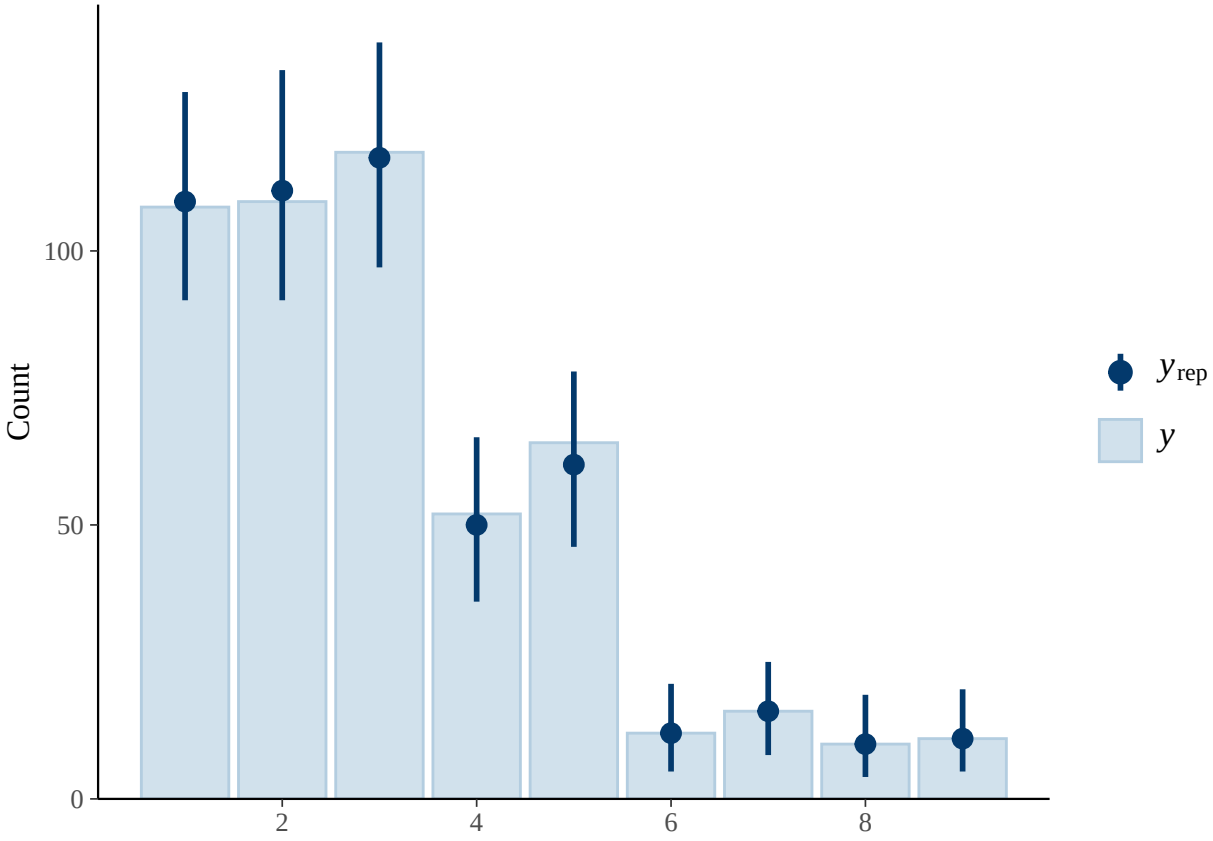
We have specifically modified the data in a way that the model in R accepts. For each dimension, we pivot the table longer with 3 pivoted columns: The first column is the number of participants starting from 1,2,..N = total amount of participants. Then, the second column is the number of questions in the that dimension, and the third column is the rating that the participant chose for that question. The rest of the columns, including the predictors, are pivoted accordingly.

In each of the following subsection, once the model is fitted, we present three plots. The first plot is the coefficient plot with credible intervals. The smaller intervals is for 50% and the wider part are chosen to be 95%. The second plot is a posterior predictive check for the model, which is to compare the observed data to data simulated from the posterior predictive distribution of the model. The third plot is also a visualization of the posterior predictive check that typically displays the observed frequencies of each outcome category alongside the frequencies of those categories in the simulated data. For the second and third graphs, by comparing the observed frequencies to the simulated frequencies, we can assess how well the model's predictions align with the observed data.

4.1 Failure: Stability Subgroup:

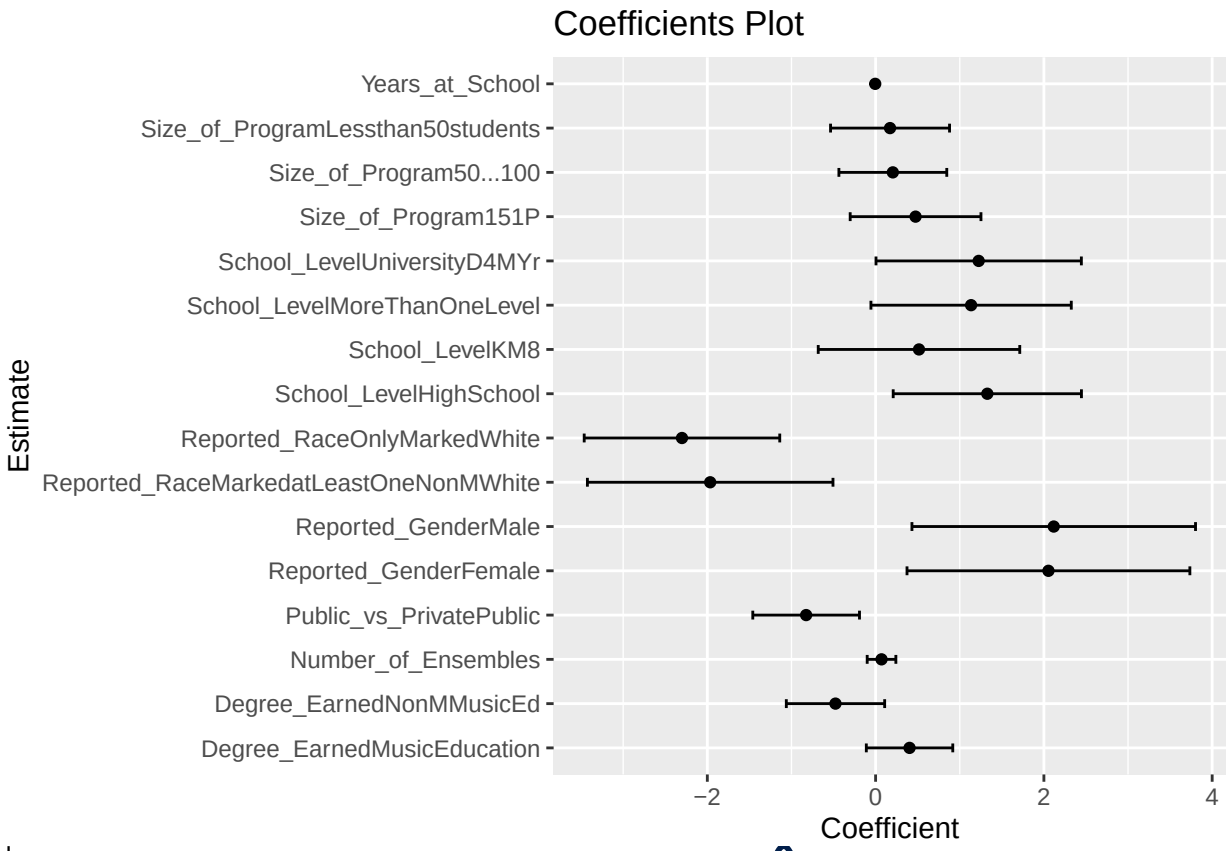


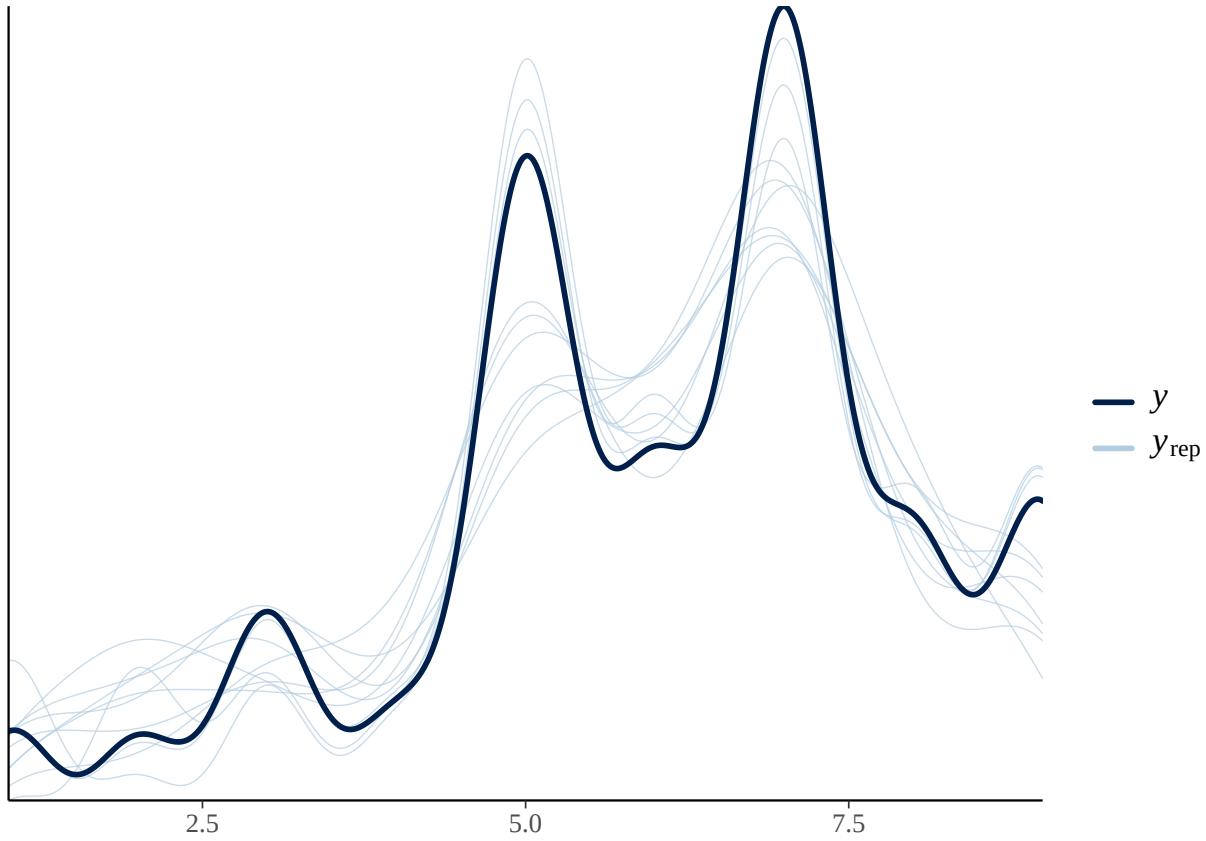


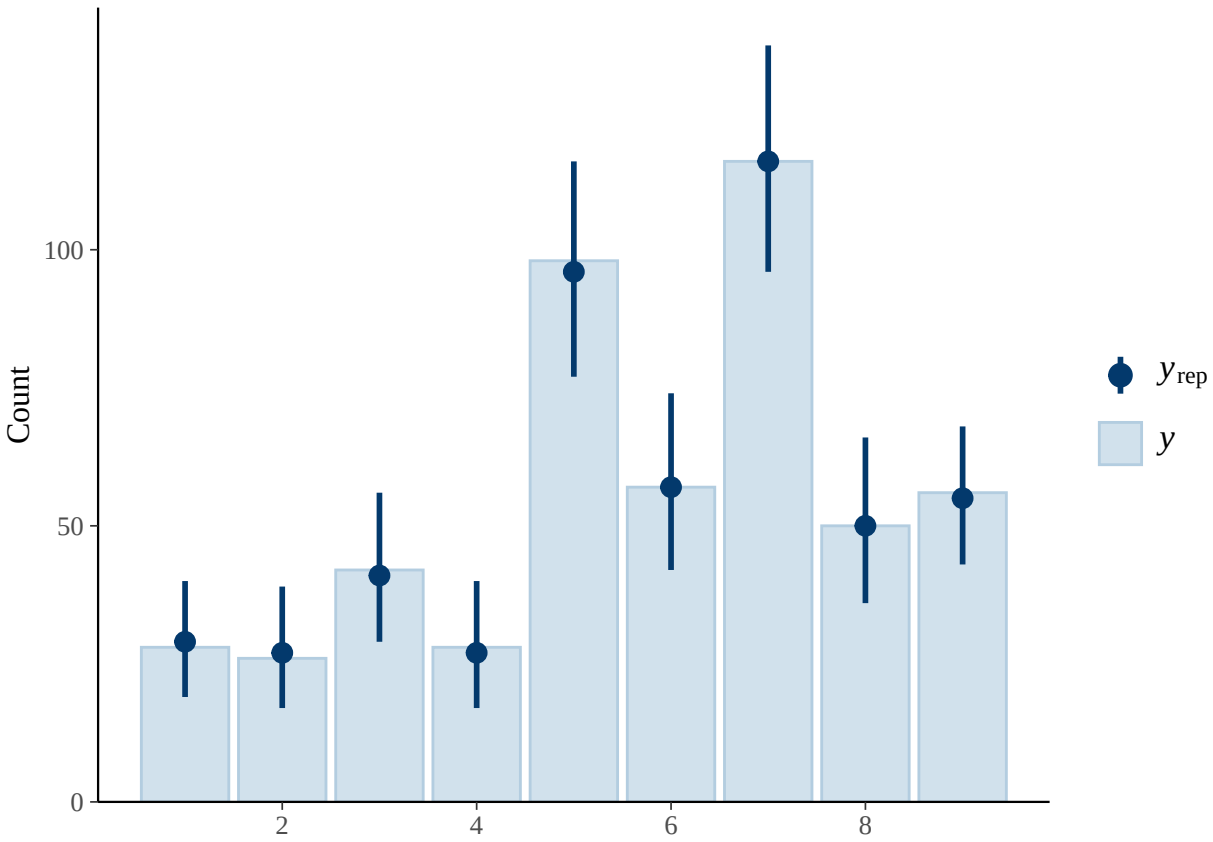


The first graph is a coefficient plot with the dot representing the fitted value and the lines are for the 95% confidence intervals, and whichever variable that crosses the threshold of 0 indicates associations with the outcomes. In this case, "public" institution, both genders "male" and "female", "non-white" race are shown to have associations (crossing 0). When their intervals are in the positive range, it could potentially contribute to failed performance from a stability dimension, and vice versa for tervals in negative range, for which it could signify that the variable can be an indicator for a successful performance. In the second plot, the black line is the real data, and the blue lines are the fitted values, and they seem to fit reasonably close. In the third plot, we can see that the model predicts close enough to the true data distribution.

4.2 Failure: External Control Subgroup:

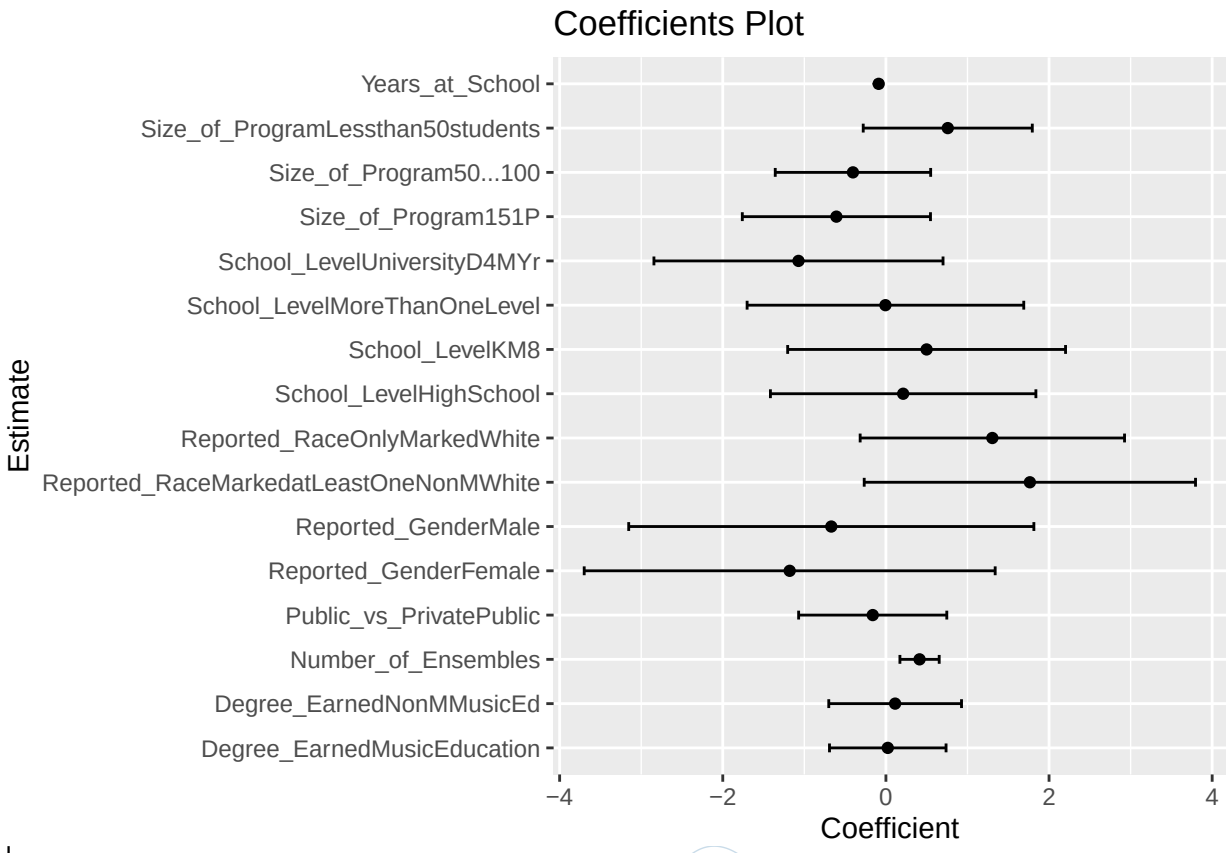


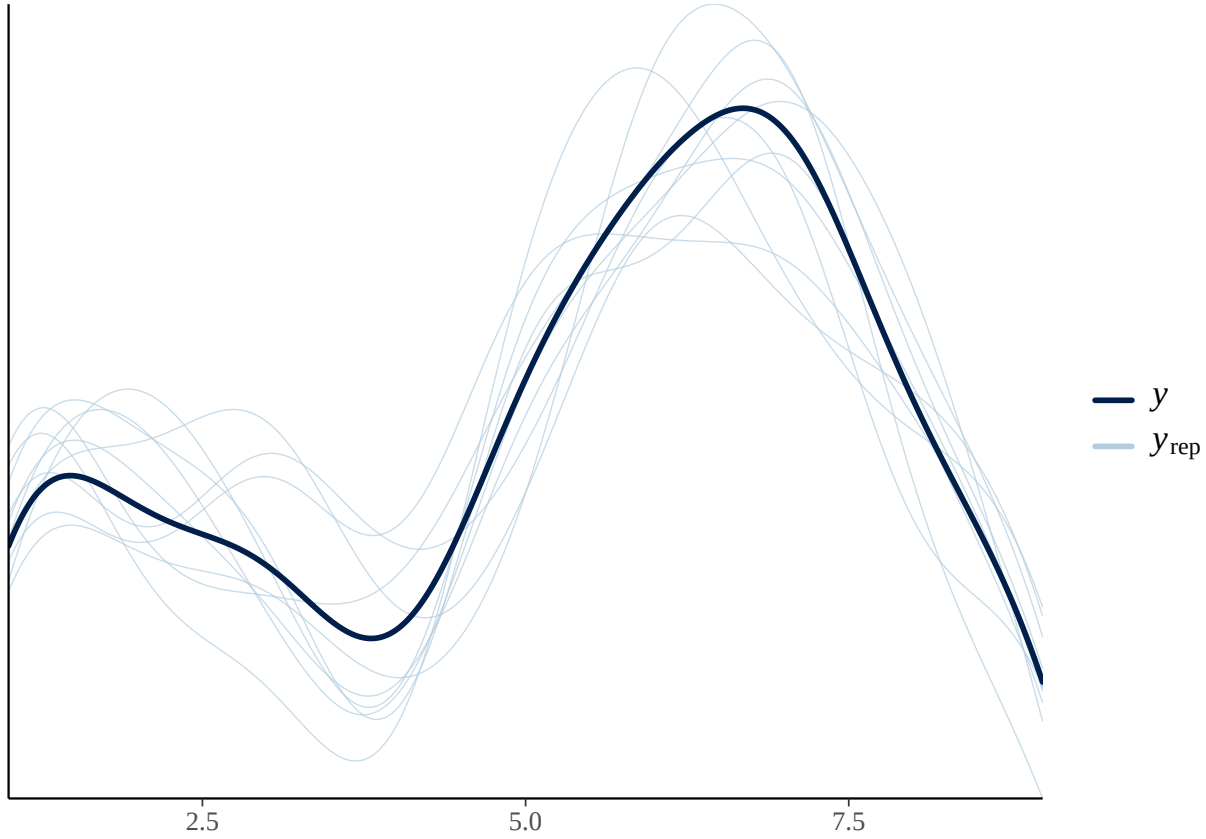


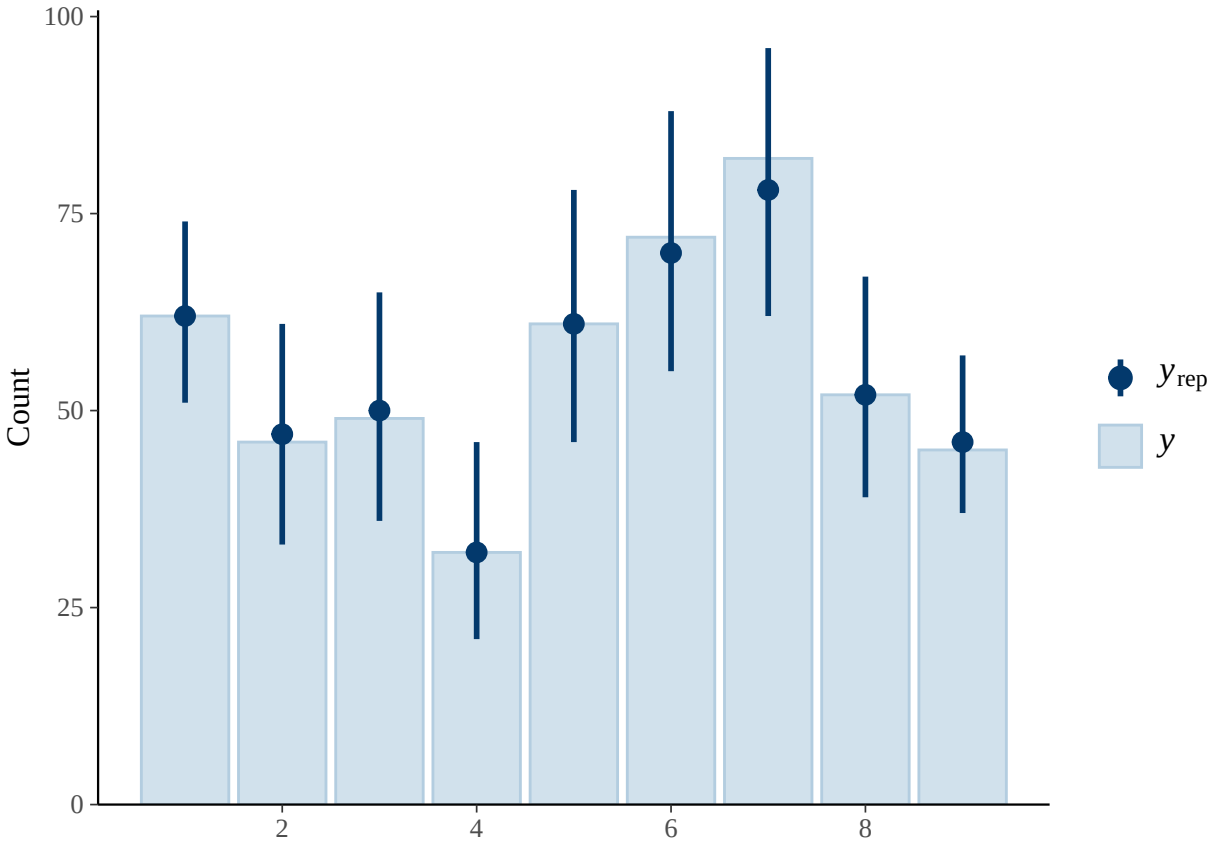


From the first plot, we observe that factors that show associations to the outcome, which cross the threshold of 0, are school level (K-8, More than one level, High School), both races, both genders, public institution. The intepretations are explained in the previous subsection: A variable that has the 95% credible intervals in the positive range correlates to the outcome, and the ones in negative ranges are negatively correlated to the outcome. From the second and third plots, the model fits reasonably well.

4.3 Failure: Internal Control Subgroup:

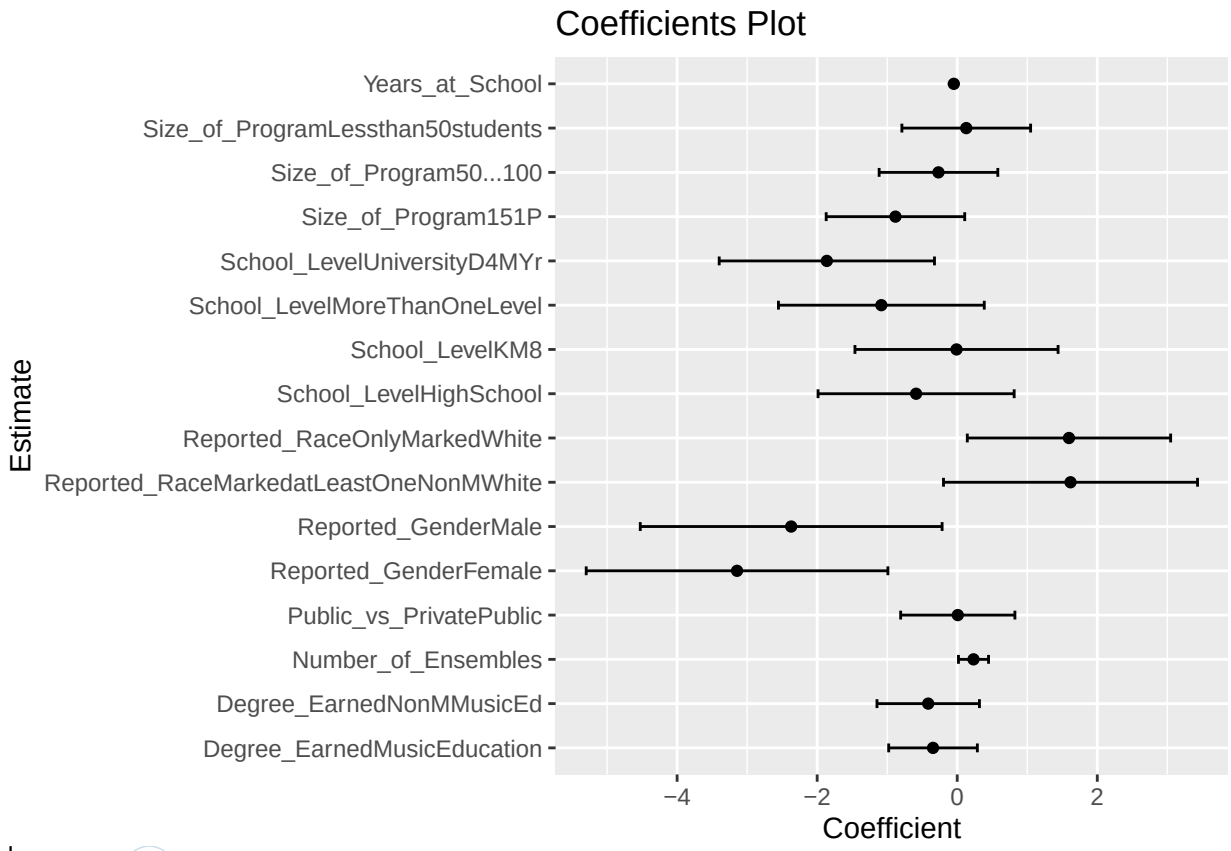


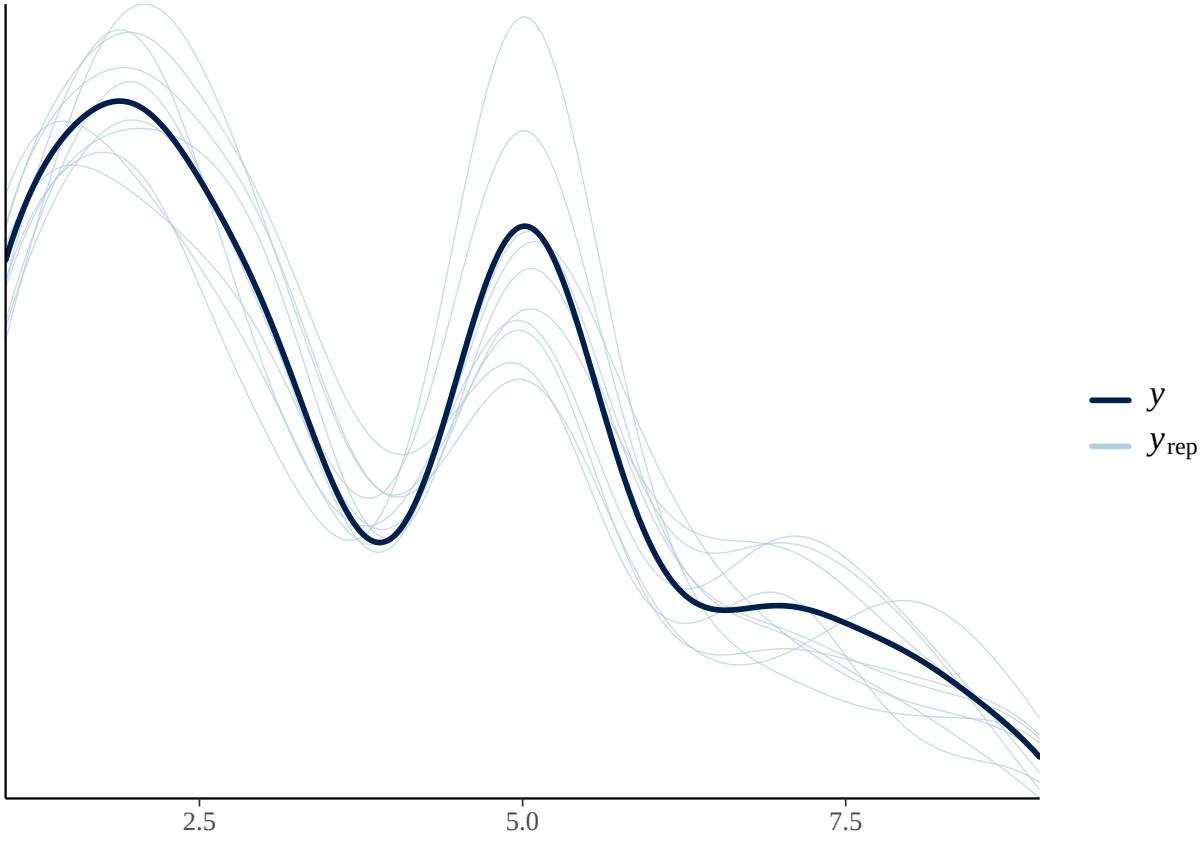


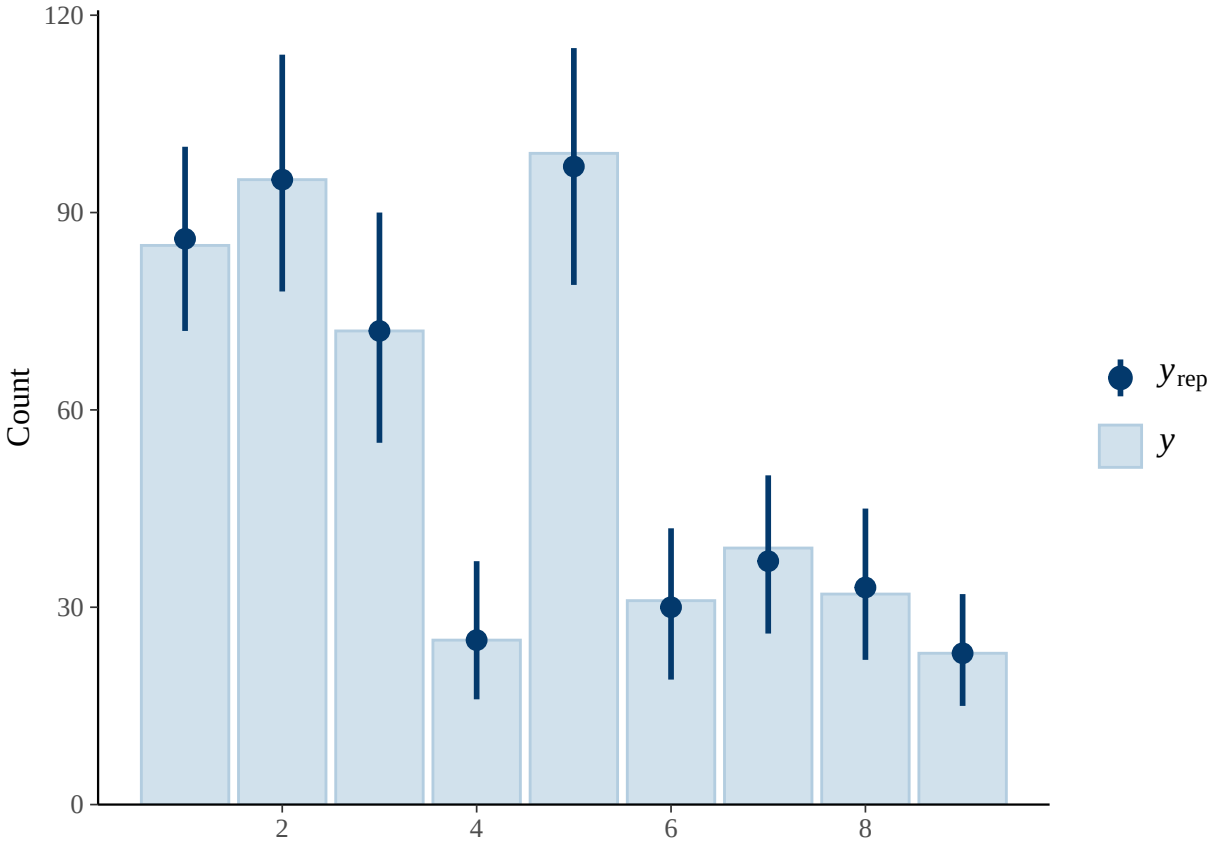


From the second and third plots, we can see that the model fits reasonably well. However, the coefficients plot shows us that a lot of variables have 0 included in their 95% credible intervals, which makes it harder to prove strong association here. We observe that perhaps the number of ensembles might give us the stronger association because it crosses 0 completely, but the other variables are not. It is arguable that years at school might be indicative of the association, but the credible interval cannot be seen clearly, whether it includes 0, from this visualization.

4.4 Failure: Locus of Causality Subgroup:

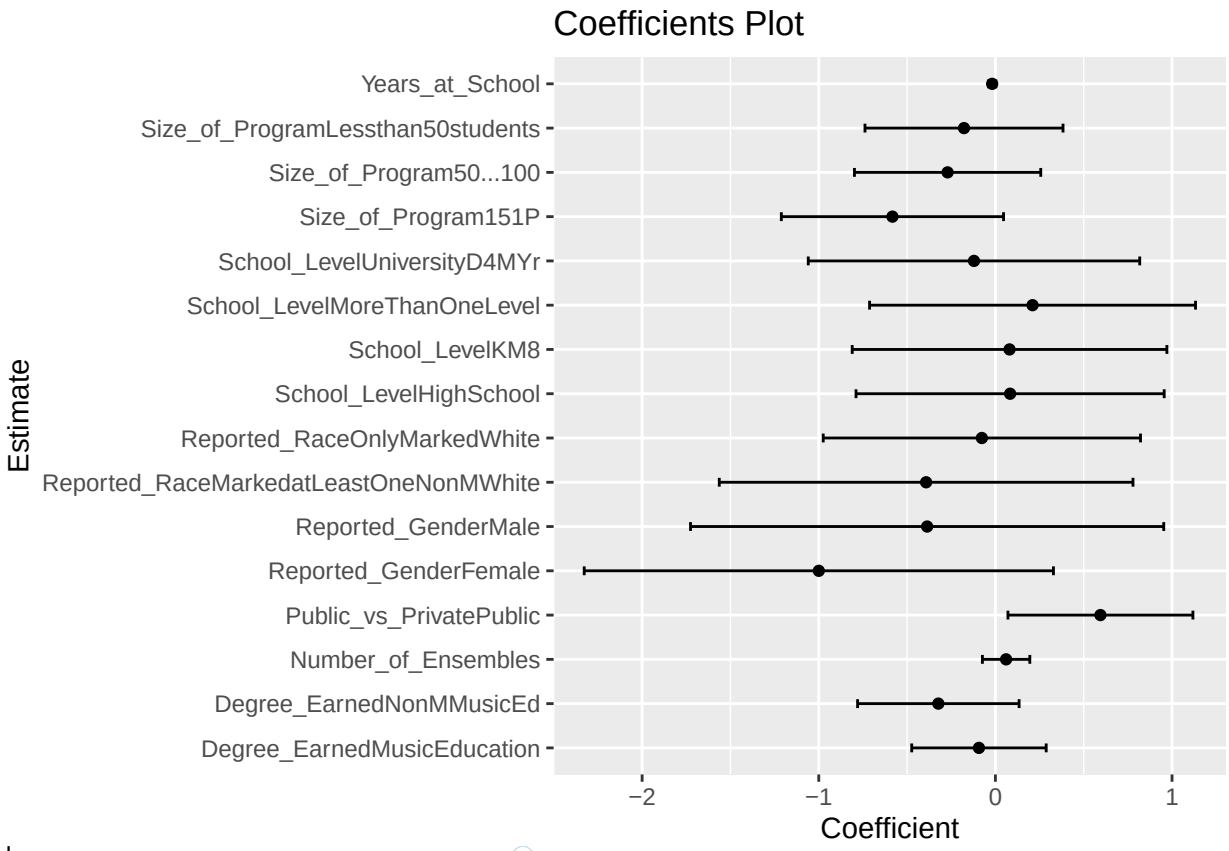


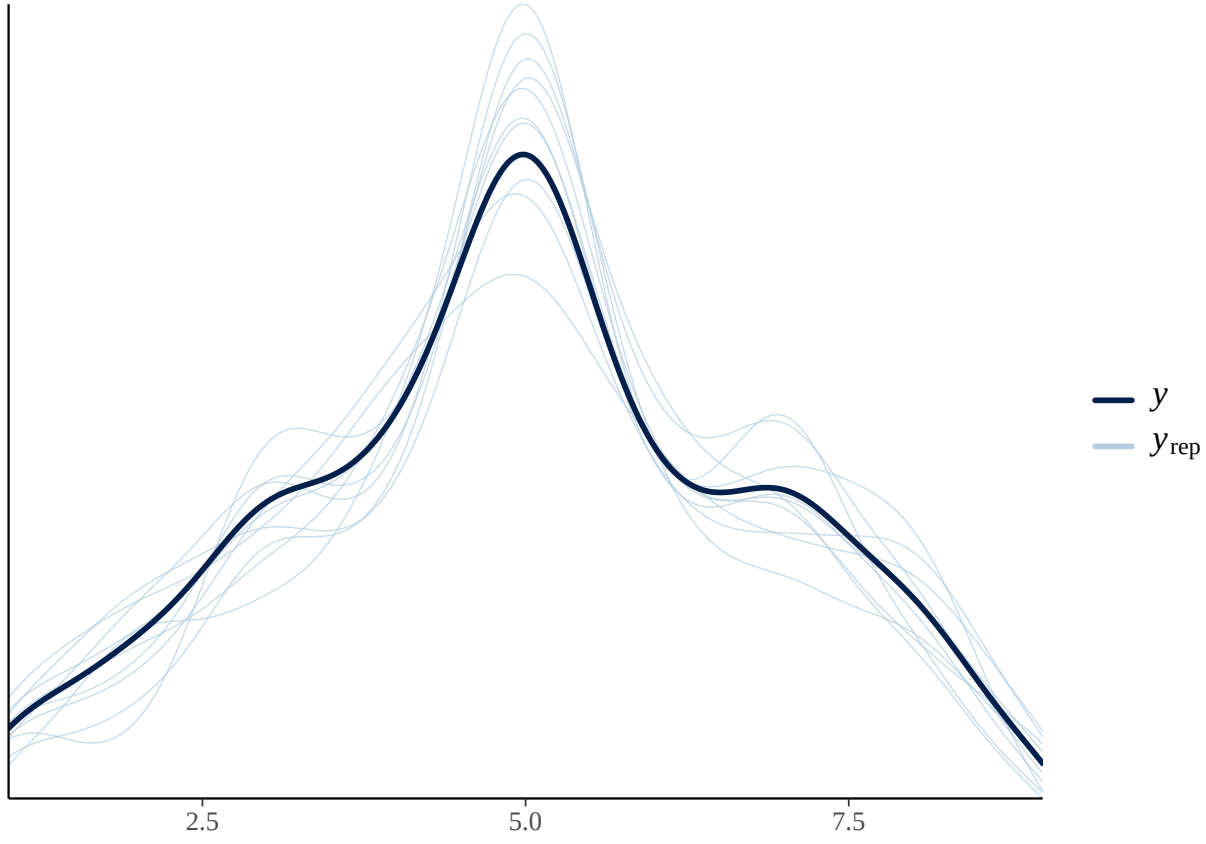


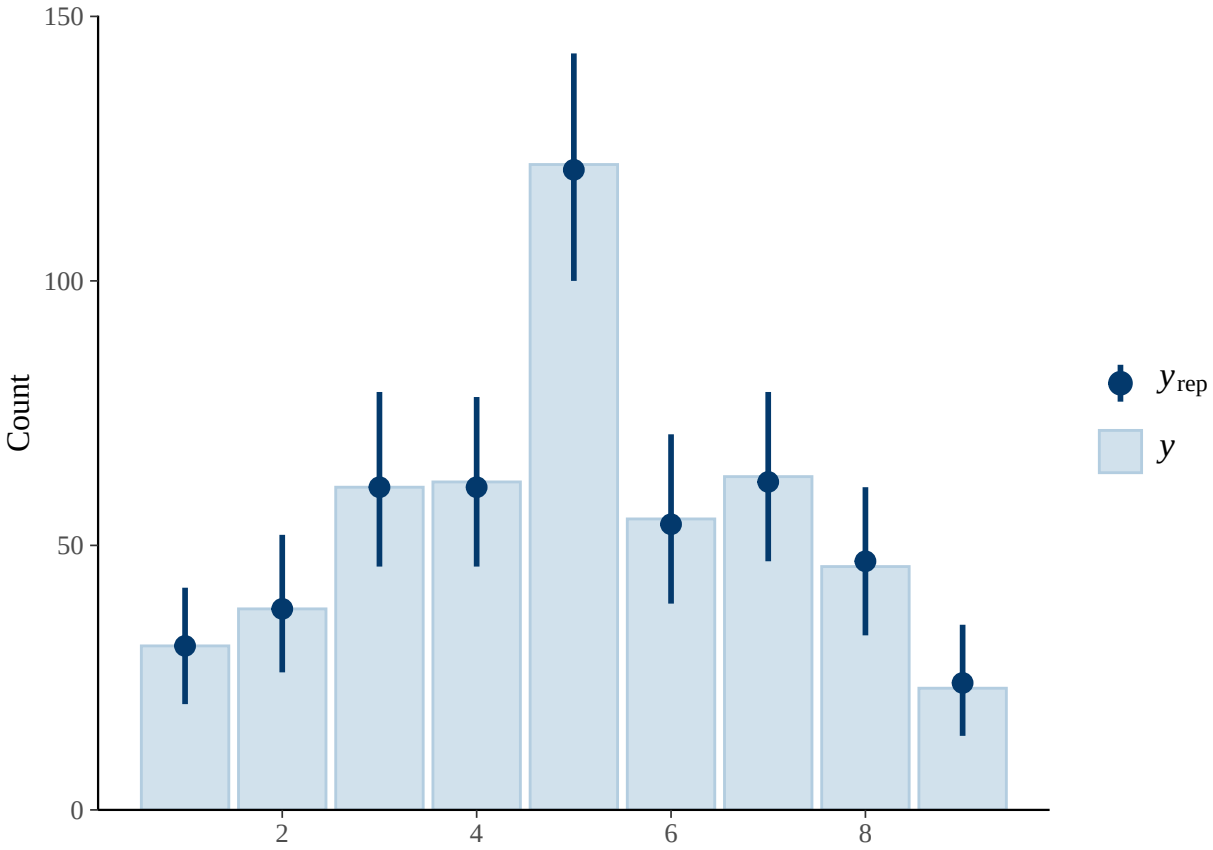


From the first plot, the variables to show associations are: school level (university 4-year), both racial groups, both reported genders, and number of ensembles. Again, the interpretations are associations can be dependent on whether they are in the positive or negative intervals, meaning whether we have a positive or negative associations, respectively. From the second and third plots, the model fits well with the shape of the data over iterations.

4.5 Success: Locus of Causality Subgroup:

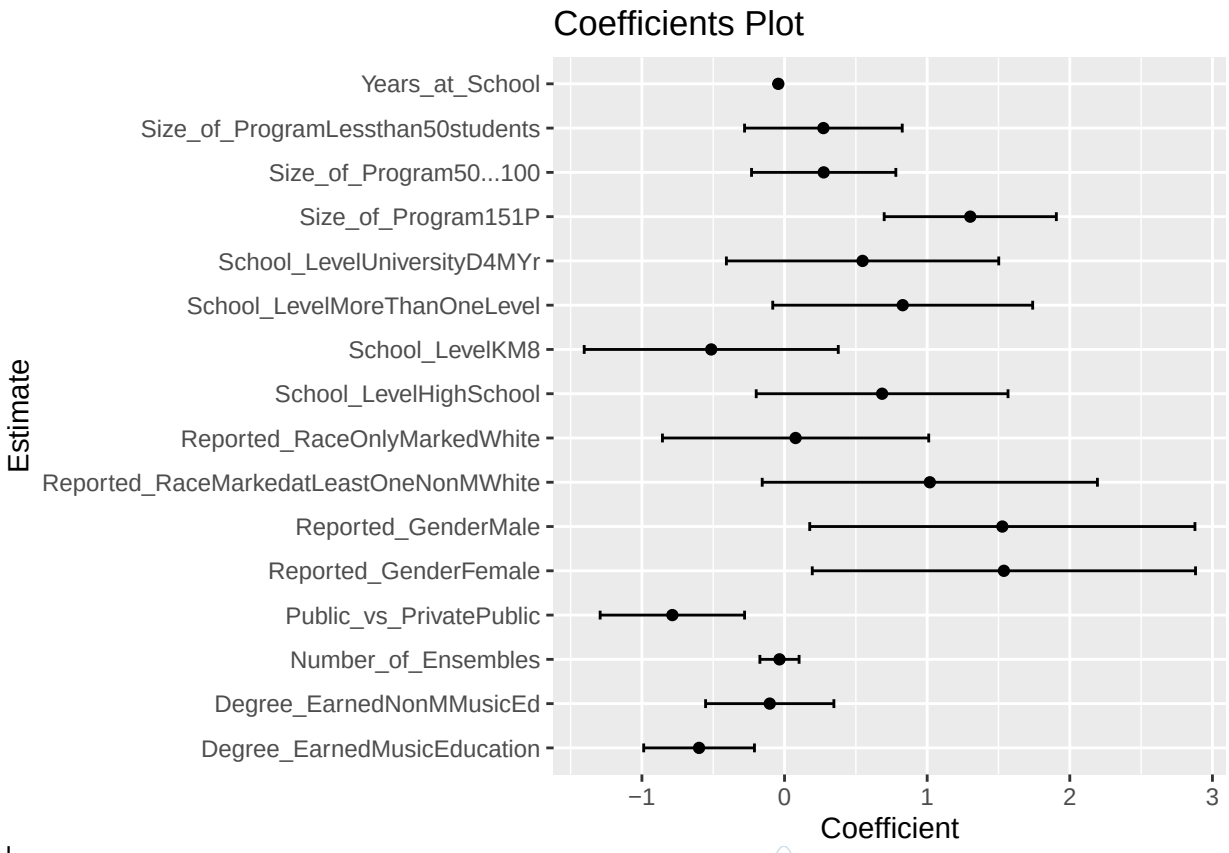


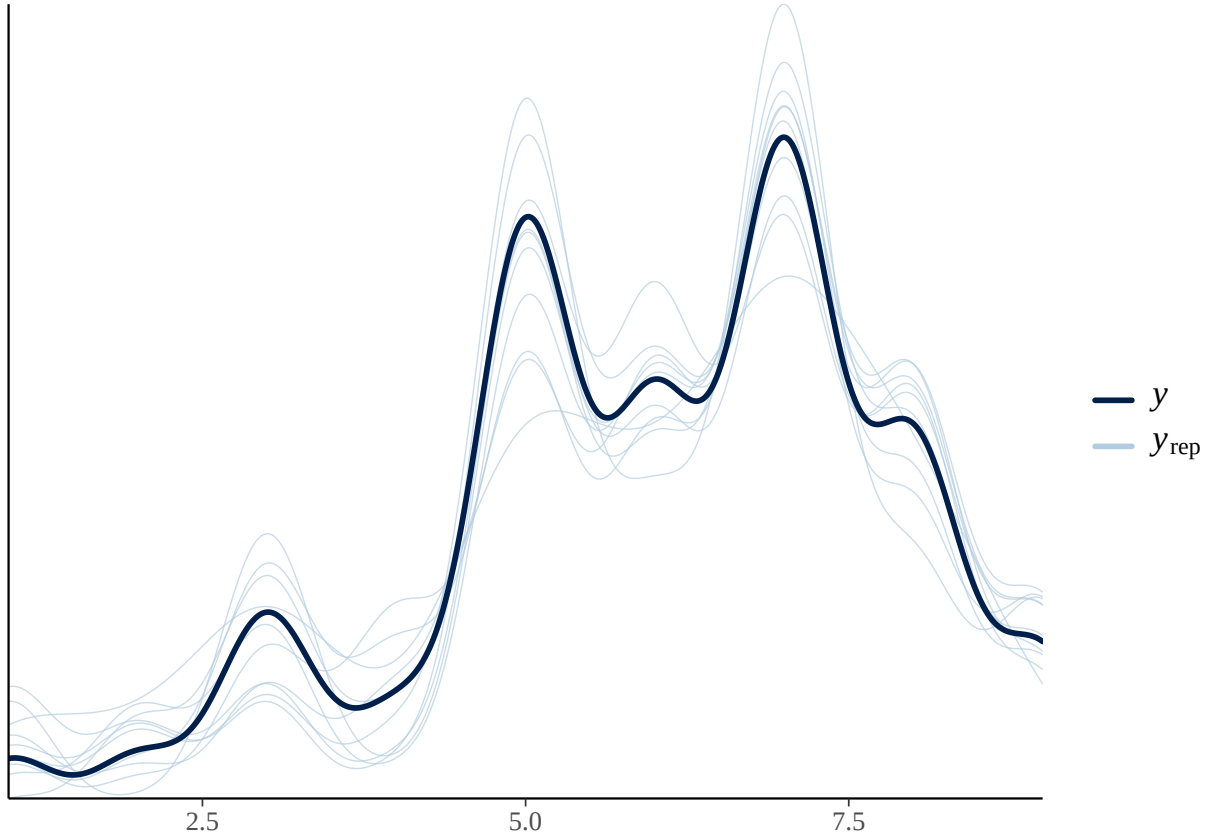


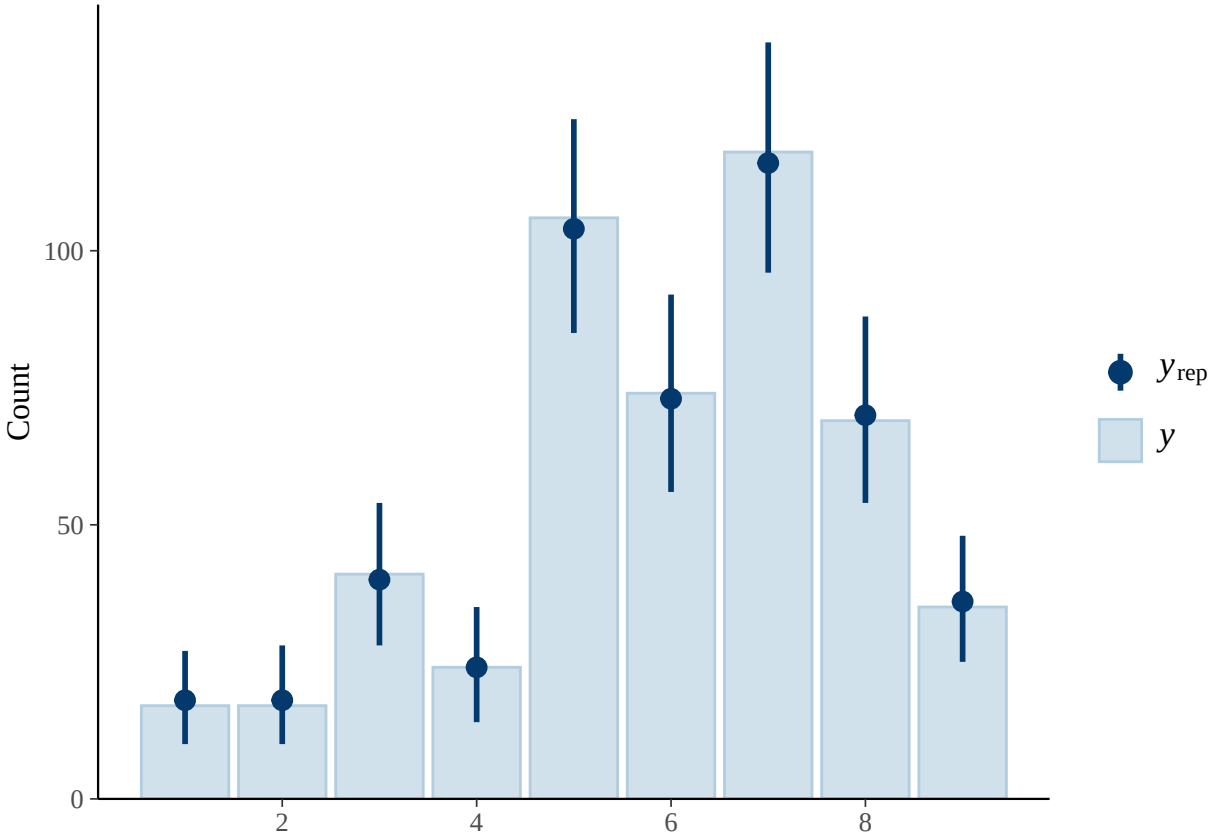


Our model continues to fit well compared to the real data distribution. From the coefficients plot, we see that the associations are shown in only in public institution factor. Years at school can be include, arguably, but we can't see the credible intervals well from this visualization to know if 0 was included in the range.

4.6 Success: External Control Subgroup:

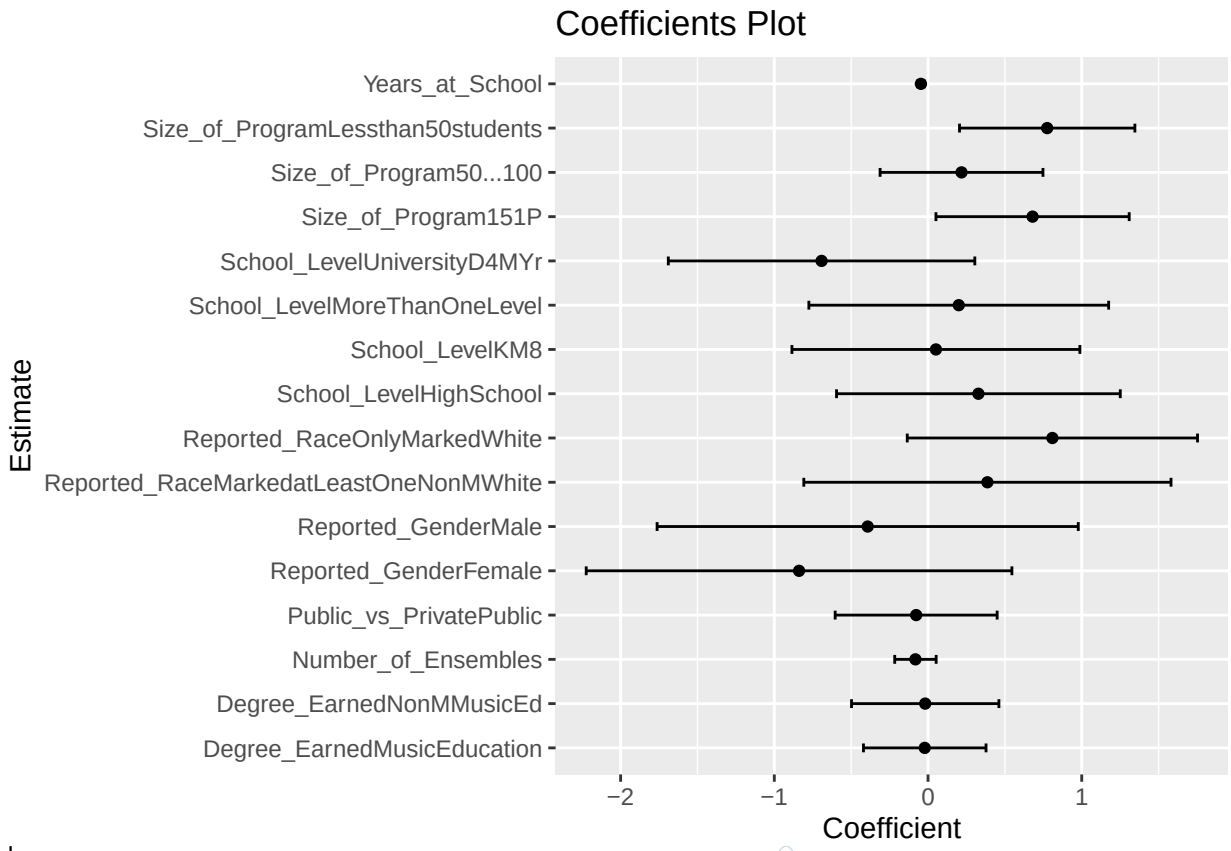


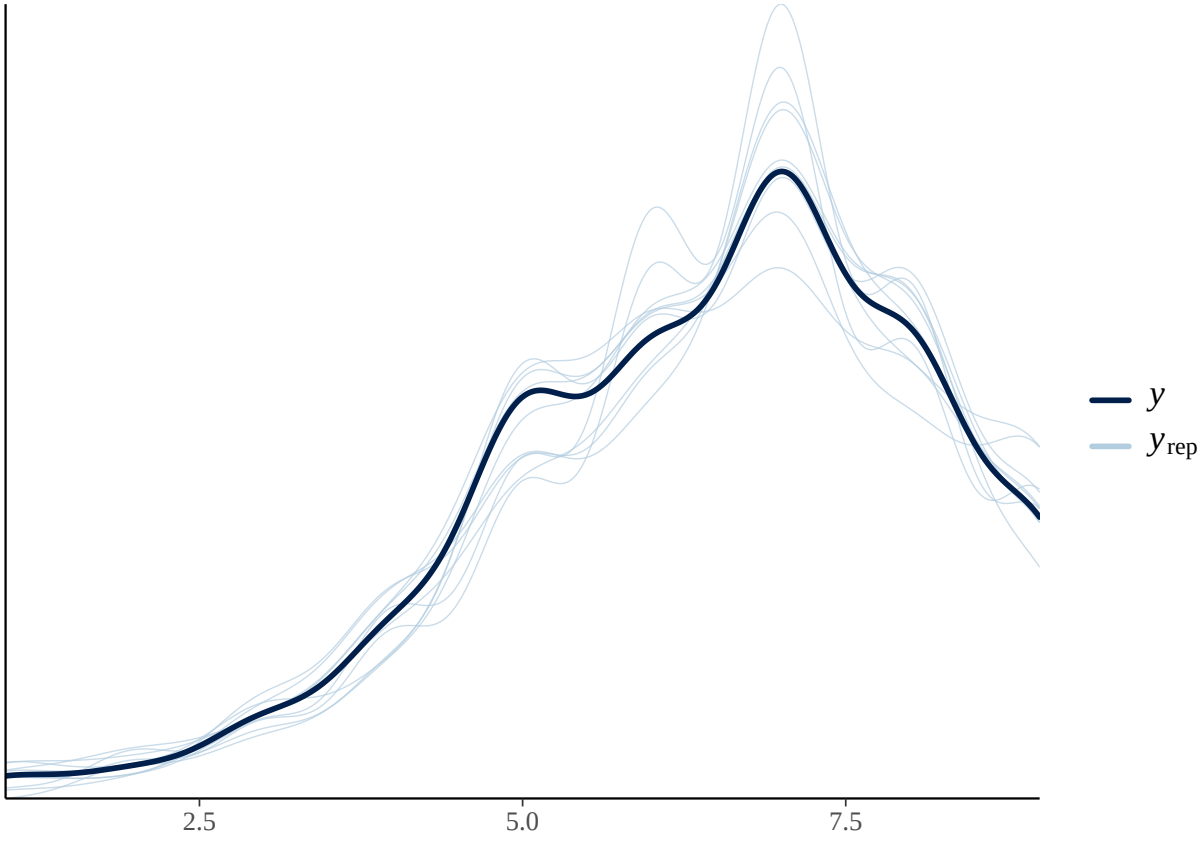


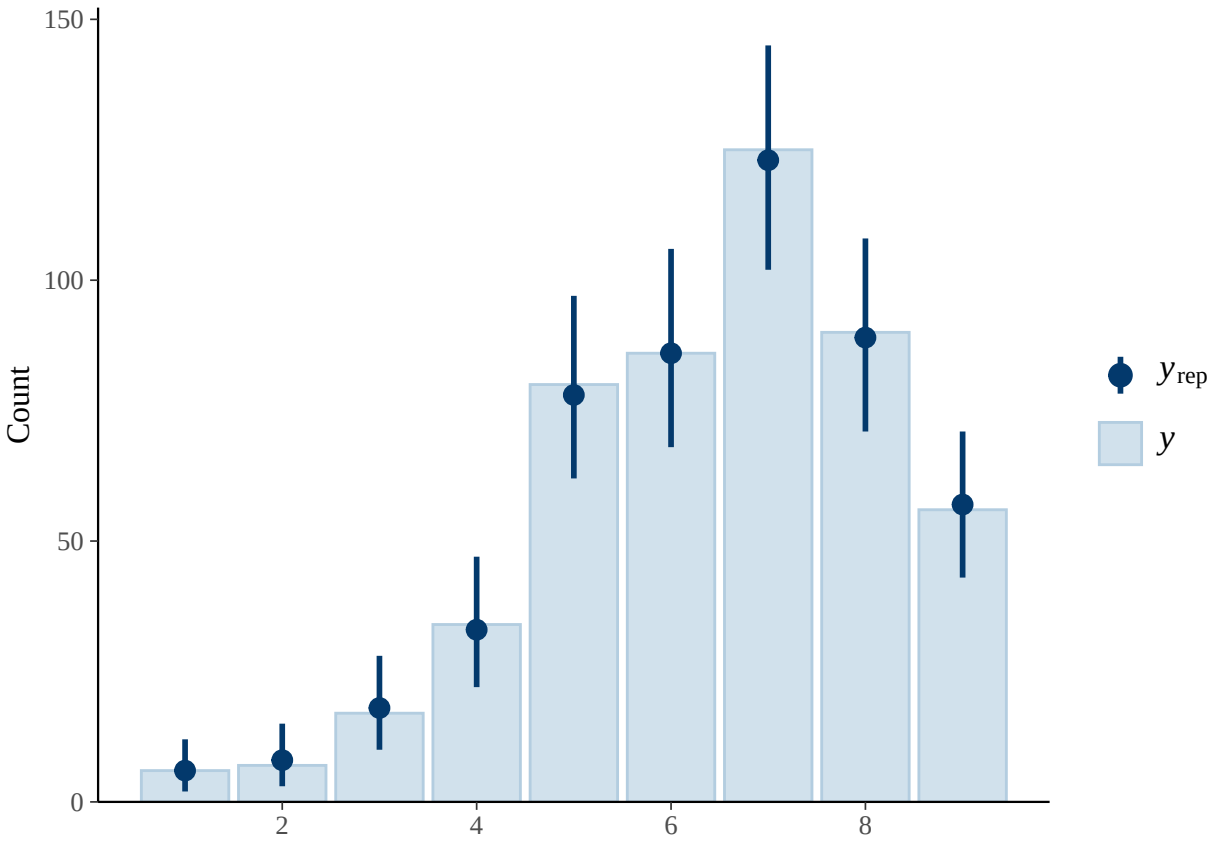


The variables that show associations are size of the program (151+), both reported genders, public institutions, degree earned (music education). We see that the model fits the real data well and as we have mentioned above, the intepretations of the associations must take the sign (positive or negative) of the credible intervals, i.e. showing positive association in success means that the variables might be indicative of a successful performance, but a positive association in failure means that the variables might be indicative of a failed performance.

4.7 Success: Internal Control Subgroup:

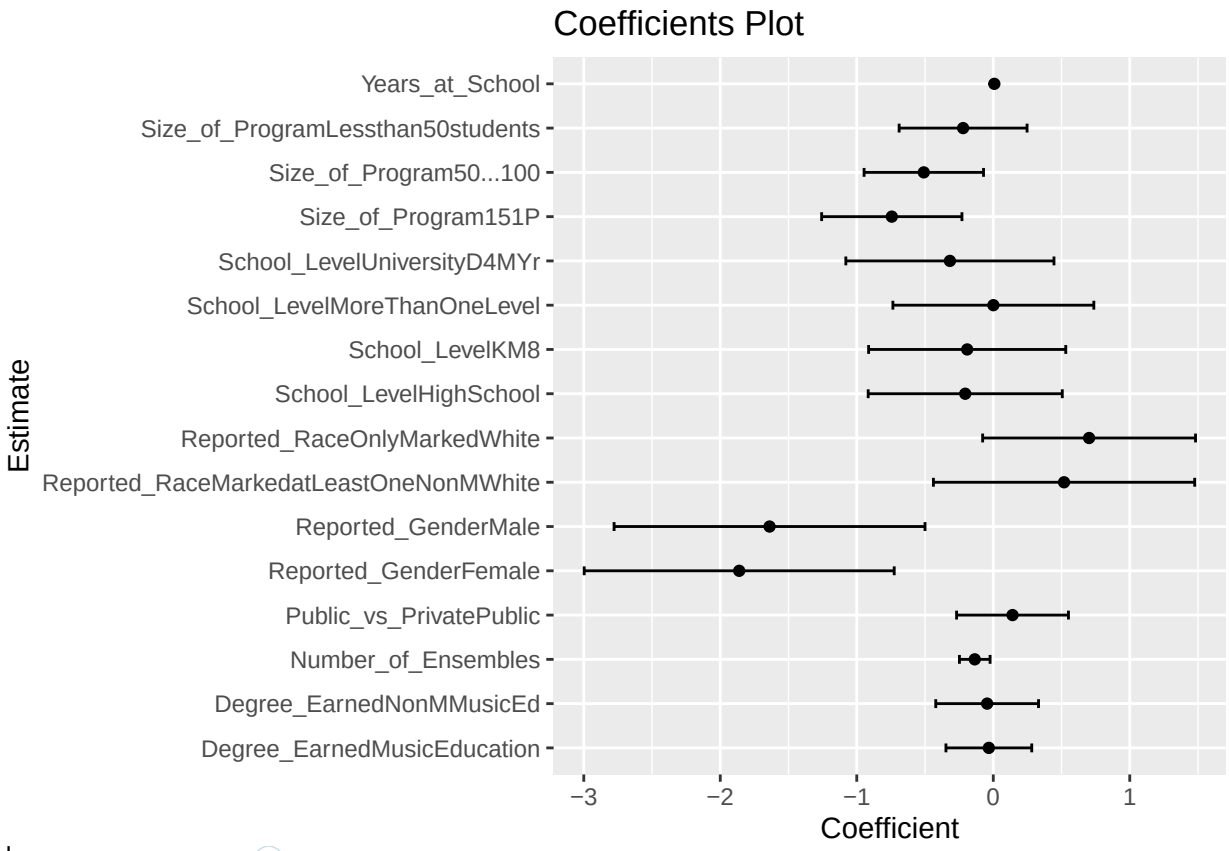


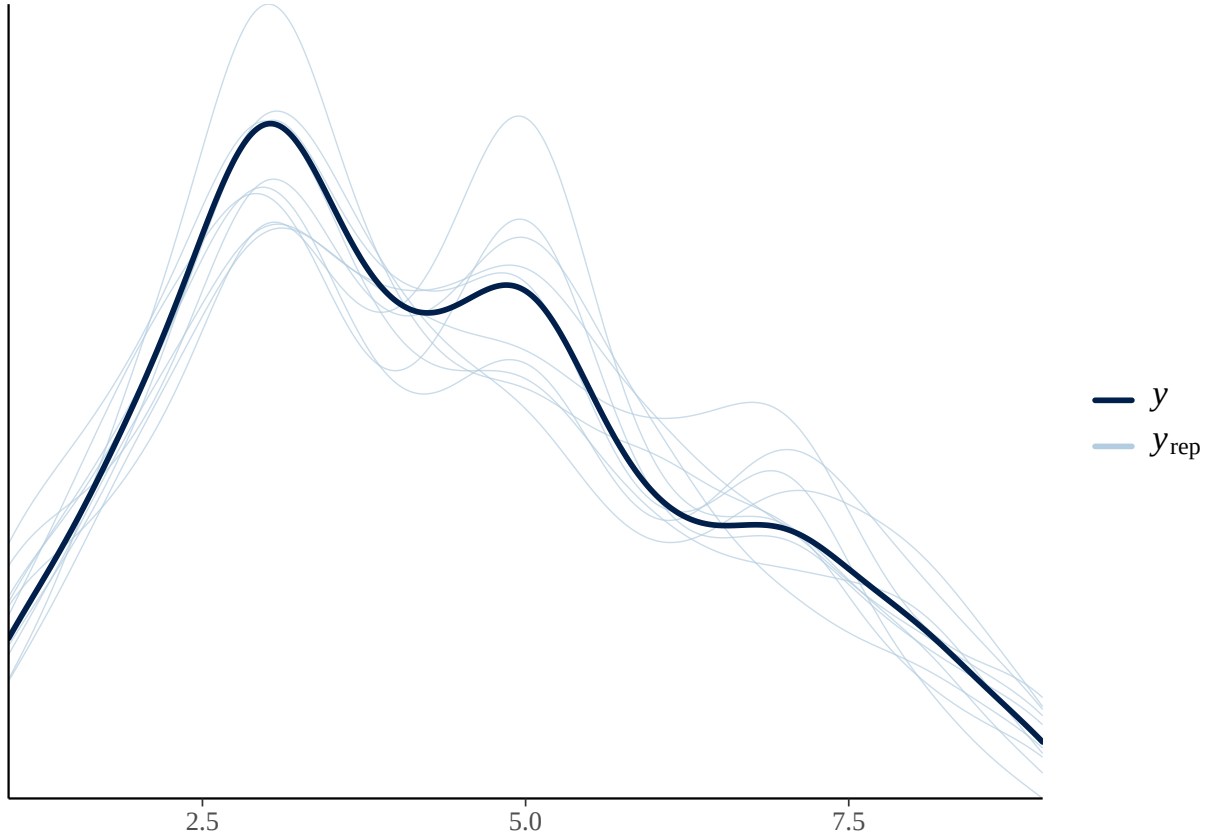


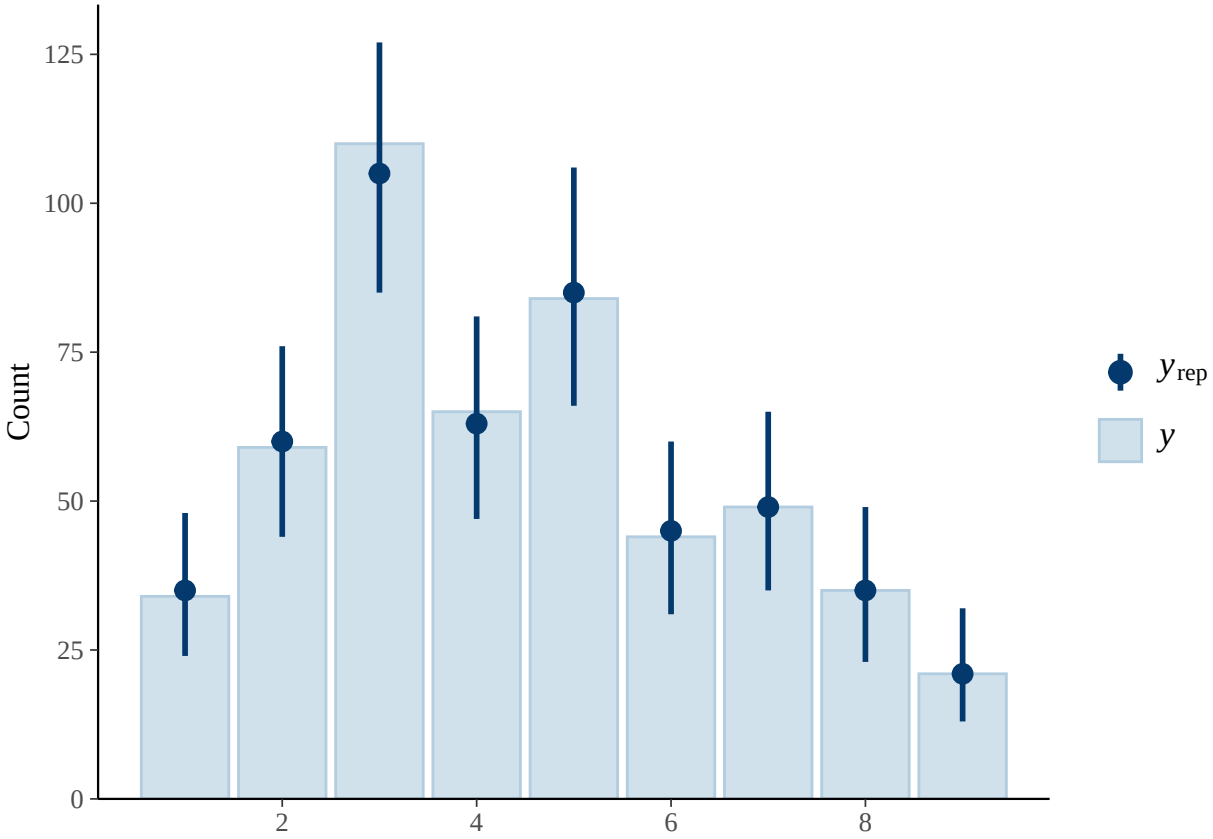


The model still fits the data well. From the first plot, the variables that show associations to the outcome are: size of the program (<50 students, 151+). Years at school can be considered to show association as well, but the credible intervals are not easily readable from this visualization.

4.8 Success: Stability Subgroup:







Similarly, the model fits reasonably well with the real data. A few factors that show good associations with the outcome are: size of the program (151+), number of ensembles, both reported genders. Again, years at school might be arguable for the same reason as mentioned the previous sections.

5 Conclusion:

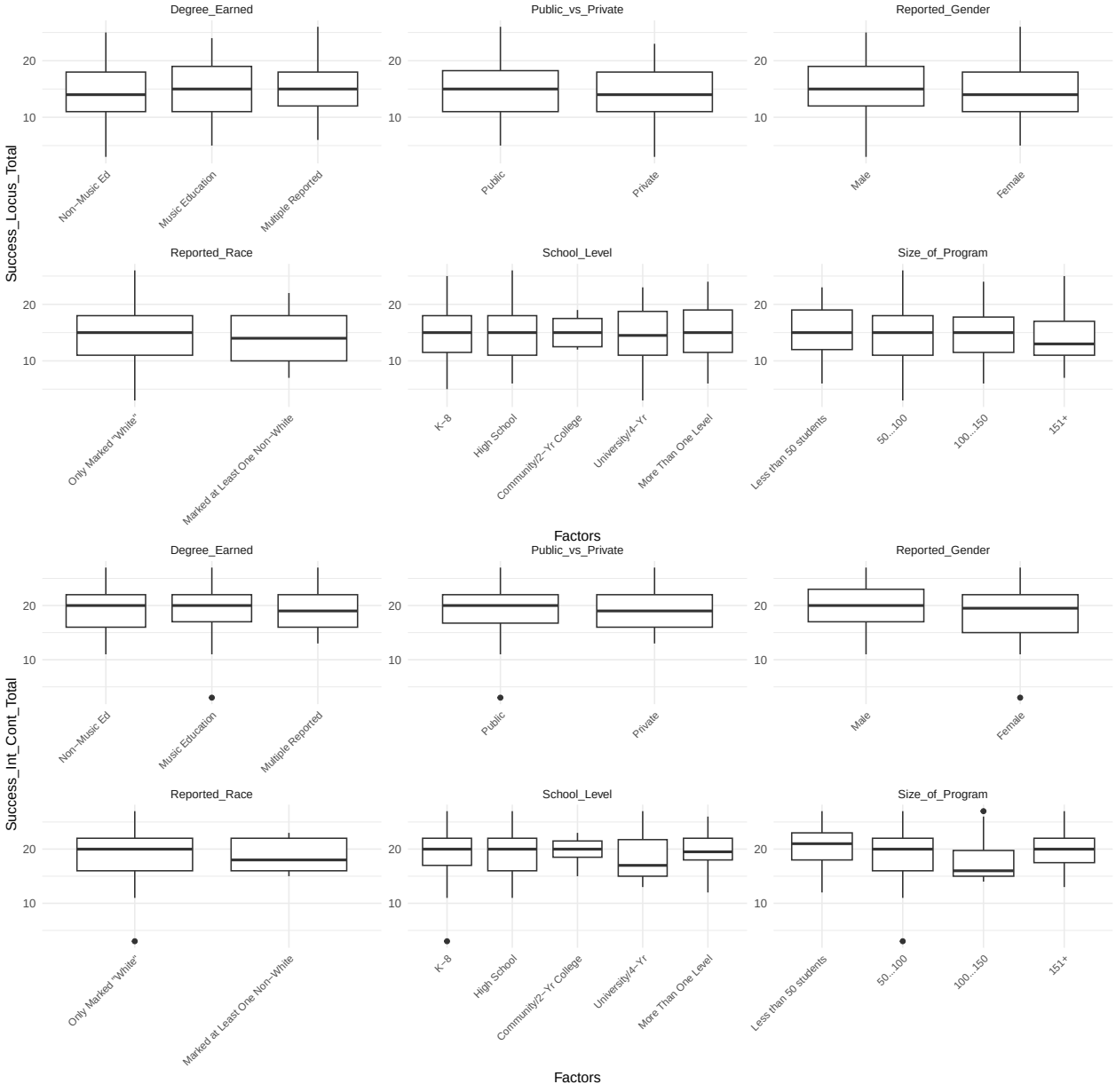
We observe that all of the models fit well with the real data, given that we have all of the predictors as fixed effects and the participants and questions as random effects. The correlated ordinal logit is as close to the "correct" likelihood as we can get. The package we use in R was brm(), which stands for Bayesian Regression Models using Stan. The model we consider is generalized linear mixed models with logit link. Overall, some of the models fit well for the data and are able to show us which variables are significant, but there are not a common trend of significant factors that are consistent throughout all of the groups. There are a few concerns that might make our analysis flawed or invalid: Firstly, the participants are mostly white (>90%) so the sample are not representative of the population, and hence impacts the ability to generalize our results to a larger population. Secondly, there is an inverse relationship between the scores for success and failure for all of the subgroups, meaning that our analysis might need to take this into consideration such as whether a high score in success should be low for failure, and significance in success should indicate less significance in failure and vice versa.

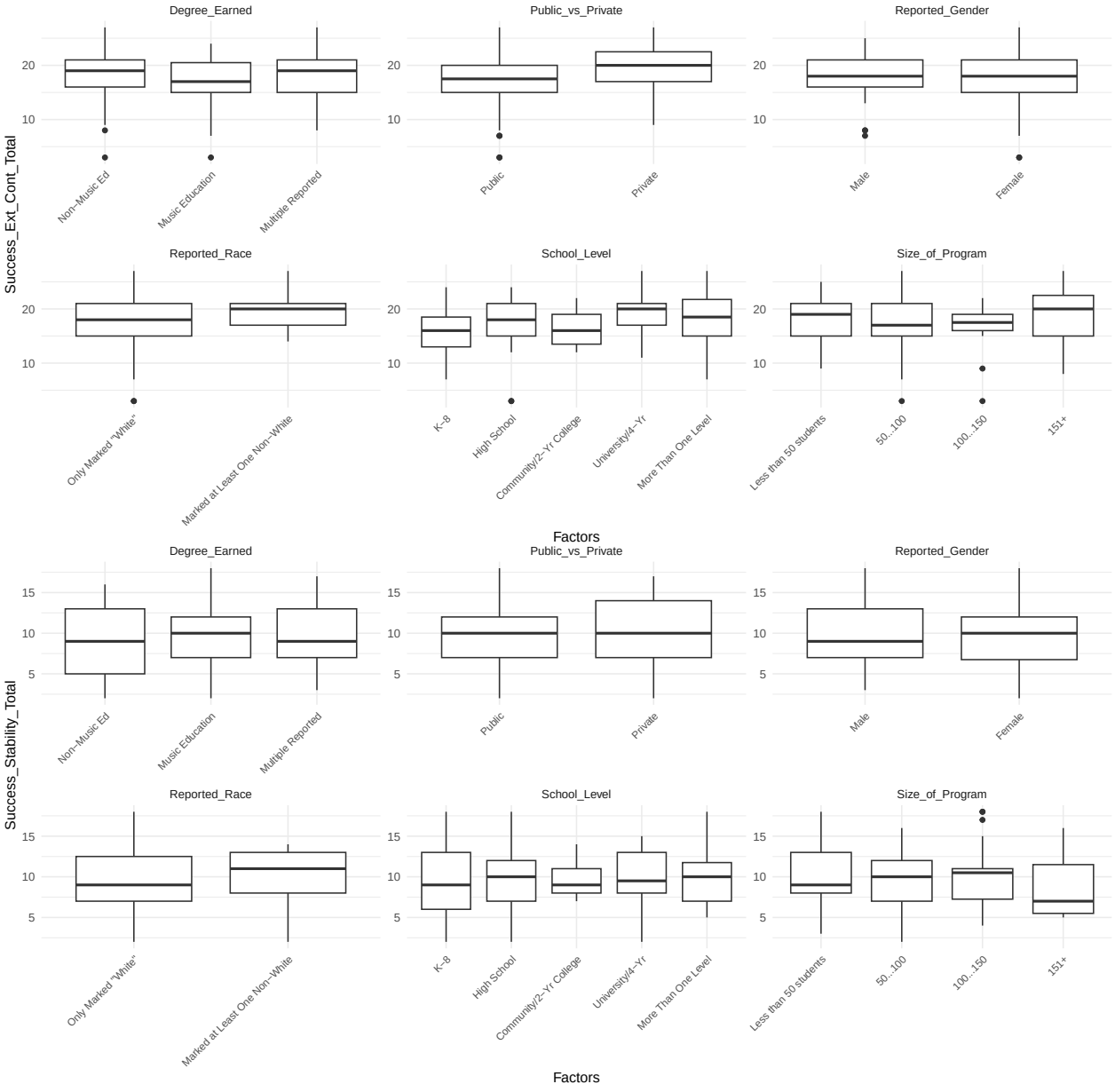
6 Appendix:

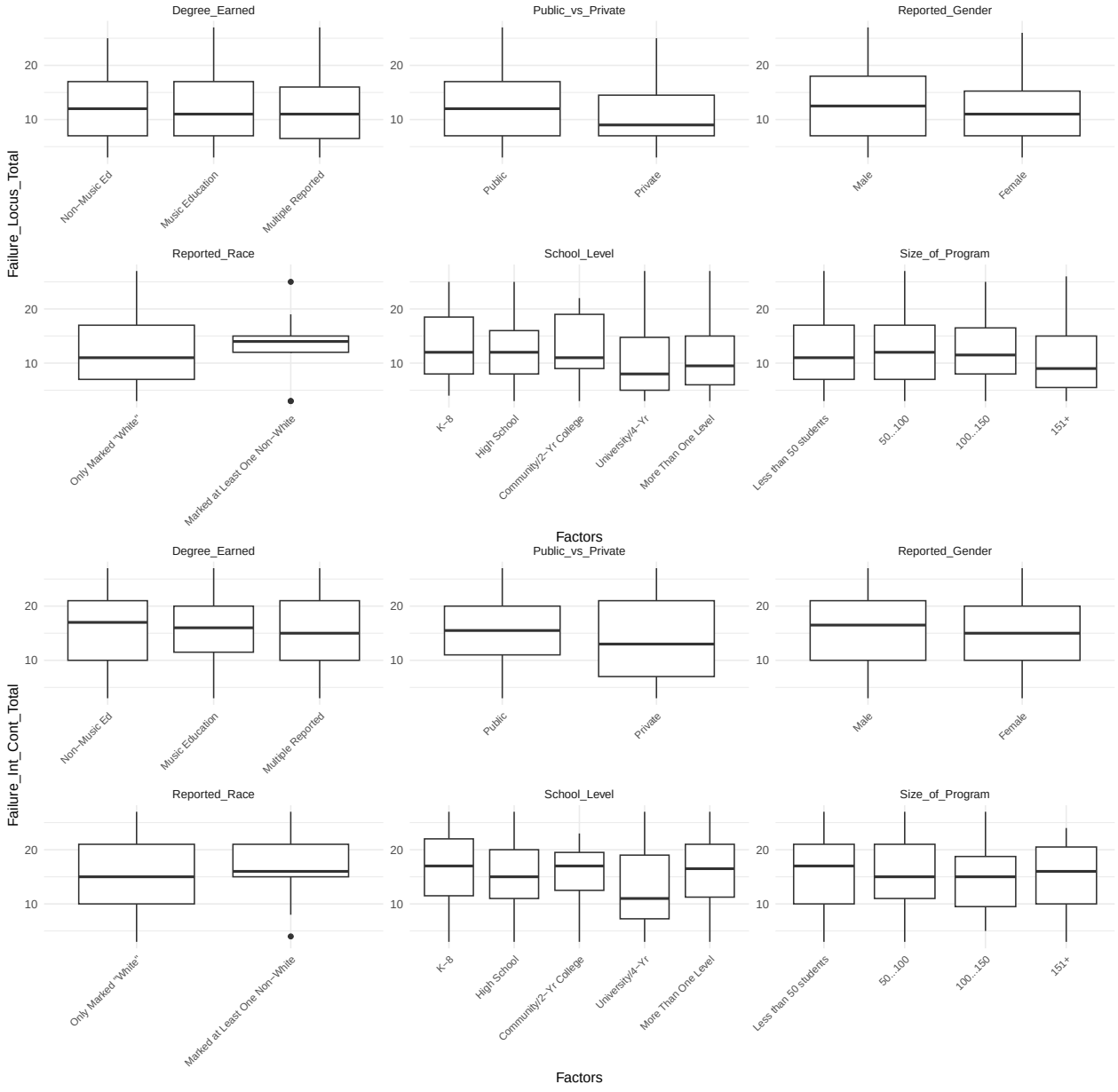
6.1 Score distributions with respect to categorical variables.

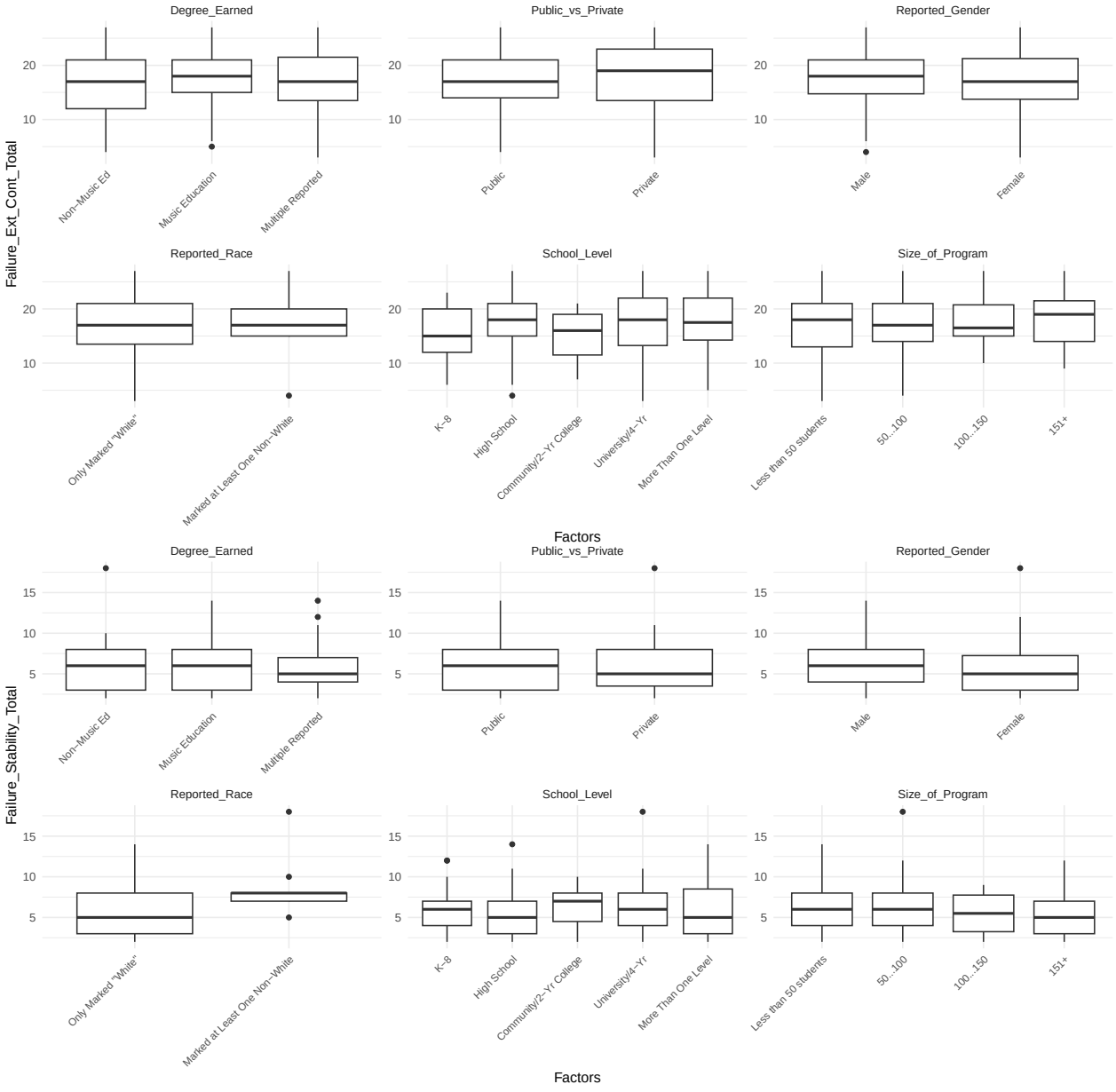
Please note that total scores are treated as numeric here. The boxplots below show the score distributions for each of dimensions for success and failure, and the y-axis denotes which dimension it is. Some variables

have outliers so our analysis of results should be of caution and take outliers into consideration.




Degree_Earned
Public_vs_Private
Reported_Gender
Reported_Race
School_Level
Size_of_Program
Success_Ext_Cont_Total
Non−Music Ed
Music Education
Multiple Reported
Public
Private
Male
Female
Only Marked "White"
Marked at Least One Non−White
K−8
High School
Community/2−Yr College
University/4−Yr
More Than One Level
Less than 50 students
50...100
100...150
151+
Factors
Degree_Earned
Public_vs_Private
Reported_Gender
Reported_Race
School_Level
Size_of_Program
Success_Stability_Total
Non−Music Ed
Music Education
Multiple Reported
Public
Private
Male
Female
Only Marked "White"
Marked at Least One Non−White
K−8
High School
Community/2−Yr College
University/4−Yr
More Than One Level
Less than 50 students
50...100
100...150
151+
Factors


Degree_Earned
Public_vs_Private
Reported_Gender
Reported_Race
School_Level
Size_of_Program
Failure_Locus_Total
Failure_Int_Cont_Total
Factors
Non−Music Ed
Music Education
Multiple Reported
Public
Private
Male
Female
Only Marked "White"
Marked at Least One Non−White
K−8
High School
Community/2−Yr College
University/4−Yr
More Than One Level
Less than 50 students
50...100
100...150
151+
10
20



6.2 Kruskal Test

For most of the relationships, there is little discrepancy within each group. Please note that for this test, we are not considering different dimensions. To convince if the covariates significantly affect response variables (that is, total scores), we use Kruskal Test to detect statistical differences between total scores and all demographic columns.

```
##              Variable       Response_Variable    P_Value
## 1      Size_of_Program  Success_Int_Cont_Total 0.03243838
## 2      Size_of_Program  Success_Ext_Cont_Total 0.24176555
## 3      Size_of_Program Success_Stability_Total 0.54760918
## 4      Size_of_Program     Success_Locus_Total 0.81846068
## 5    Public_vs_Private  Success_Int_Cont_Total 0.39069382
## 6    Public_vs_Private  Success_Ext_Cont_Total 0.01290561
```

```
## 7    Public_vs_Private Success_Stability_Total 0.64305236
## 8    Public_vs_Private     Success_Locus_Total 0.19519094
## 9  Number_of_Ensembles  Success_Int_Cont_Total 0.81397646
## 10 Number_of_Ensembles  Success_Ext_Cont_Total 0.10810472
## 11 Number_of_Ensembles Success_Stability_Total 0.39232564
## 12 Number_of_Ensembles     Success_Locus_Total 0.56994858
## 13     Years_at_School  Success_Int_Cont_Total 0.35026816
## 14     Years_at_School  Success_Ext_Cont_Total 0.05721080
## 15     Years_at_School Success_Stability_Total 0.72201246
## 16     Years_at_School     Success_Locus_Total 0.12977999
## 17       Degree_Earned  Success_Int_Cont_Total 0.71730823
## 18       Degree_Earned  Success_Ext_Cont_Total 0.25332417
## 19       Degree_Earned Success_Stability_Total 0.83566756
## 20       Degree_Earned     Success_Locus_Total 0.56147754
## 21        School_Level  Success_Int_Cont_Total 0.56646139
## 22        School_Level  Success_Ext_Cont_Total 0.01641056
## 23        School_Level Success_Stability_Total 0.95173110
## 24        School_Level     Success_Locus_Total 0.94010740
## 25     Reported_Gender  Success_Int_Cont_Total 0.55394120
## 26     Reported_Gender  Success_Ext_Cont_Total 0.25830352
## 27     Reported_Gender Success_Stability_Total 0.20771608
## 28     Reported_Gender     Success_Locus_Total 0.19957827
## 29       Reported_Race  Success_Int_Cont_Total 0.53590370
## 30       Reported_Race  Success_Ext_Cont_Total 0.16054324
## 31       Reported_Race Success_Stability_Total 0.89828818
## 32       Reported_Race     Success_Locus_Total 0.82678867

##               Variable       Response_Variable     P_Value
## 1      Size_of_Program  Failure_Int_Cont_Total 0.647993800
## 2      Size_of_Program  Failure_Ext_Cont_Total 0.973740766
## 3      Size_of_Program Failure_Stability_Total 0.616254358
## 4      Size_of_Program     Failure_Locus_Total 0.824181820
## 5    Public_vs_Private  Failure_Int_Cont_Total 0.404940395
## 6    Public_vs_Private  Failure_Ext_Cont_Total 0.280249074
## 7    Public_vs_Private Failure_Stability_Total 0.683528296
## 8    Public_vs_Private     Failure_Locus_Total 0.224452023
## 9  Number_of_Ensembles  Failure_Int_Cont_Total 0.337223479
## 10 Number_of_Ensembles  Failure_Ext_Cont_Total 0.550177390
## 11 Number_of_Ensembles Failure_Stability_Total 0.397220223
## 12 Number_of_Ensembles     Failure_Locus_Total 0.769555480
## 13     Years_at_School  Failure_Int_Cont_Total 0.185492862
## 14     Years_at_School  Failure_Ext_Cont_Total 0.161184881
## 15     Years_at_School Failure_Stability_Total 0.473563795
## 16     Years_at_School     Failure_Locus_Total 0.631465942
## 17       Degree_Earned  Failure_Int_Cont_Total 0.746680658
## 18       Degree_Earned  Failure_Ext_Cont_Total 0.692065658
## 19       Degree_Earned Failure_Stability_Total 0.872141543
## 20       Degree_Earned     Failure_Locus_Total 0.996993852
## 21        School_Level  Failure_Int_Cont_Total 0.330401322
## 22        School_Level  Failure_Ext_Cont_Total 0.309425524
## 23        School_Level Failure_Stability_Total 0.798135389
## 24        School_Level     Failure_Locus_Total 0.233683597
## 25     Reported_Gender  Failure_Int_Cont_Total 0.874090255
## 26     Reported_Gender  Failure_Ext_Cont_Total 0.894373324
```

```
## 27     Reported_Gender Failure_Stability_Total 0.003697709
## 28     Reported_Gender     Failure_Locus_Total 0.461450863
## 29       Reported_Race  Failure_Int_Cont_Total 0.825777269
## 30       Reported_Race  Failure_Ext_Cont_Total 0.420034111
## 31       Reported_Race Failure_Stability_Total 0.021382639
## 32       Reported_Race     Failure_Locus_Total 0.670138279
```

For the success group, size of program with respect to internal control, public versus private and school level with respect to external control have p-value < 0.05. Meanwhile, for the failure group, reported gender and race with respect to stability subgroup also have p-value < 0.05.