GLM April 25

Lauren Kohls & Sarah Little

4/26/2021

Section 1.1 NIFA Program Background Information

Healthy Schoolhouse 2.0 is a 5 year long study that is currently in its 3rd year of activity. It's Nutrition Intervention Program is a multicomponent nutrition education program designed to empower teachers to improve nutrition literacy and prevent obesity in elementary school students in Washington, DC. The intention of this program is to study the intervention effects of PD Sessions and classroom implemented nutrition lessons on student knowledge of, and attitude towards, nutrition related concepts.

Section 1.2 Preliminary Analysis

We began our exploratory analysis as we looked at Teacher Level and Student Level data that was collected during the program. Our preliminary data started out in 2 separate SPSS files, one for student data and one for teacher data. The Student dataset contained 330 variables and 1324 observations. The teacher dataset contained 447 variables and 206 observations. Teacher level data contained teacher demographic data, along with health and health education questionnaires grouped by Teacher ID, and spanning each year that they participated in the program. Student level data was structured by a unique student code which contained student information for each year they participated within the program.

The nested and hierarchical structure of the data calls for the utilization of multilevel modeling techniques. While SPSS requires multi-level data to be in wide format to analyze, R requires data to be structured in long format. We therefore melted the data, and merged the two datasets together.The merging of the data was student centered, meaning it contained all student observations from years 1:3, and attached Teacher level data to student information based on TeacherIDs for pre and post tests. It is important to note that while some teachers attended PD Sessions, their students did not participate in the knowledge testing during the program. Those teachers will exist in our teacher_all dataset but will not exist within the full_data/ student and teacher merged dataset. The dataset contains student knowledge and attitude scores and has variables for Program Year[Which Year of the NIFA Program is it?], Student Year [How long has the student participated in the program] and Test Exposure [How many times has the student seen the test prior to this?]

As we began our research, we visualized the data, and utilized t.tests, anovas, and various other methods to assess the relationship between student knowledge scores and the identified interventions. Due to the pre-existing analysis on the data, we had assumed that the data was normally distributed and our methods followed suit of that assumption. We moved forward to utilizing the LMER package to build and assess the significance of multilevel models on our data. After building our MLM and testing assumptions within the model, we realized that our data for all program years was not normally distributed, and therefore violated required assumptions. We returned to the drawing board to analyze data with non-parametric methods.

Our first step in analyzing the data was to identify outliers in Knowledge Percent. Our results indicted that 42 data points within the dataset were outliers. Upon investigation of these observations we concluded that data points for students scoring 1 or 0 on L1Knowledge total, or students scoring 2 or less on L23 knowledge total were being categorized as outliers. As measurement errors and data errors were not present, and due

to low scores of 0-2 being a natural part of the examination process, we decided to keep the outliers in our final data analysis.

Following this, to test the distributions of our the data we employed the Shapiro-Wilk Test. This tests the null hypothesis that a variable is normally distributed. If the p value is less than .05, we reject the null hypothesis and conclude that the data is not normally distributed. We first tested to make sure that we did not affect the distribution of the data when we melted the dataset by year to combine the student and teacher data. We do this by looking at the pre and the post test scores by Program Year, which was the original set up of the data.

Our findings included the following:

In testing Pre Knowledge Percent in Program Year 1, we find that the data is not normally distributed (n = 325, p < .05). In testing Post Knowledge Percent in Program Year 1, we find that the data is not normally distributed (n= 320, p < .05).

The qqplot below demonstrates a stairstep pattern, which demonstrates the discrete nature of the data values. We also see spikes at both the high and low ends of the plot.

Normal Q–Q Plot
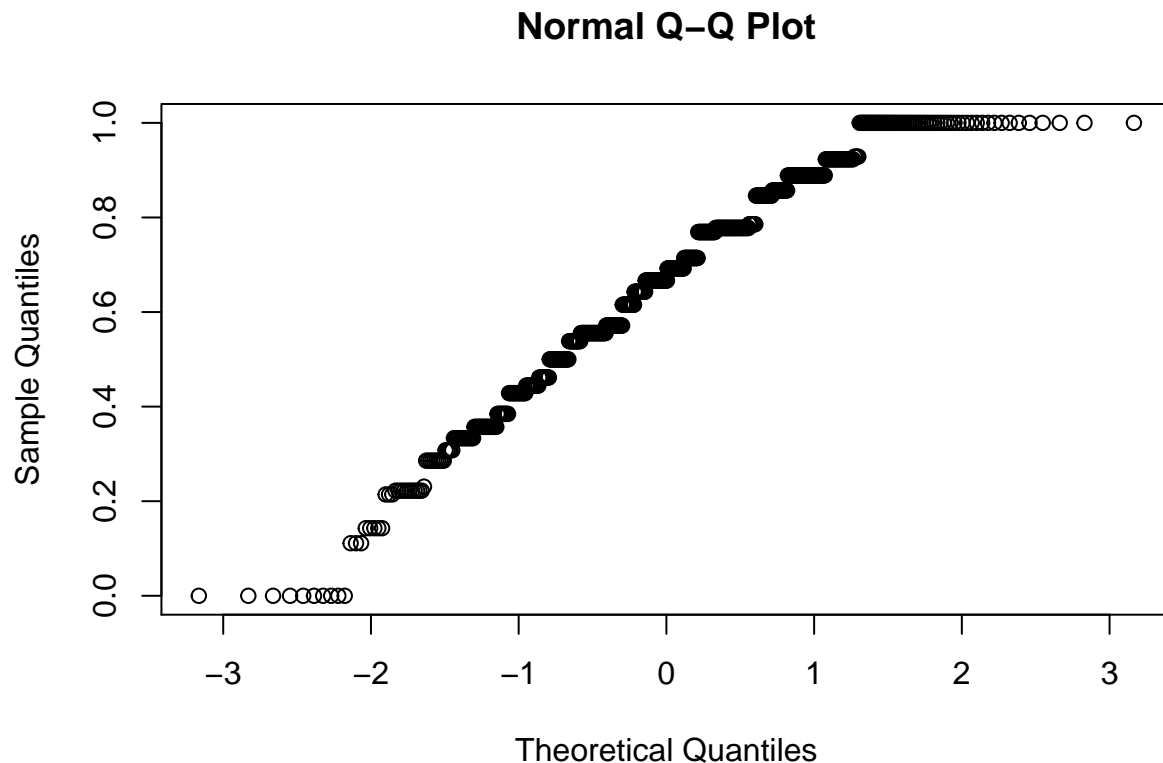
As we saw that the data was not normally distributed, we questioned whether this could be related to the proportional nature of the knowledge percentages. We check to make sure that utilizing Knowledge Total would not not result in a different outcome,and attempted to transform Knowledge Percent to meet normality assumptions.

Despite our efforts, the outcome was the same, displaying a non-normal distribution of our data. There is varying literature regarding the utilization of a T-Test for non-normal data. The central limit theorem notes that the distribution of sample mean values tend to follow a normal distribution regardless of the population distribution if the sample size is large enough. Some suggest that if the sample size per group is large enough, a t-test can be applied without normality test. Others (Kim & Park, 2019), do not recommend

this, and argue that while the "central theorem guarantees the normal distribution of sample means it does not guarantee the normal distribution of samples in the population", and do not support the use of t.tests on non-normal data. Moving forward, we decided to forgo our findings from our t.tests and instead looked at non-parametric tests to analyze our data.

For our tests, we utilized the Wilcoxon Ranked Sum test and the Kruskal-Wallis Test.

Mann-Whitney Test / Wilcoxon Ranked Sum Test: This test is a non-parametric test that is the equivalent of an independent samples t.test. The test is used to compare the differences between independent groups when the dependent variable is ordinal or continuous but is not normally distributed.

In conducting these tests, we found no significant difference in Pre knowledge scores between school types (n = 1001, p = .1). We did find significant differences in Post Knowledge Scores by school type (n = 659, p < .01). We also found significant differences when testing Knowledge Percent Score Changes (n = 659, p<.01). In Student Year 1, Experimental Schools had higher mean and median Post Knowledge Scores and Knowledge Percent Score Changes.

Kruskal-Wallis test: Following these results, we moved onto the Kruskal-Wallis Test. This test is a non-parametric alternative to the one-way ANOVA. It is an extension of the Wilcoxon Test and allows for more than 2 groups.

We conducted the Kruskal-Wallis test of Knowledge Percent by Lessons Received in Experimental Schools in Student Year 1. The results returned significant findings (p = .0003, df = 2, statistic = 16.2, n = 459). A histogram shows the distribution of Knowledge PErcent by the three lesson brackets, being L = Less than 3, I = Intervention of 3 lessons, and M = More than three lessons received.

We also conducted the Kruskal-Wallis test of Lessons Taught by PD Sessions Attended in Experimental Schools. The results returned significant findings (p <.001, df = 5, statistic = 76, n = 218).

Once satisfied with our non-parametric testing, we then moved on to begin our second phase of modeling.

```
## pdf
##   2
```

Knowledge Percent By Lessons Bucket
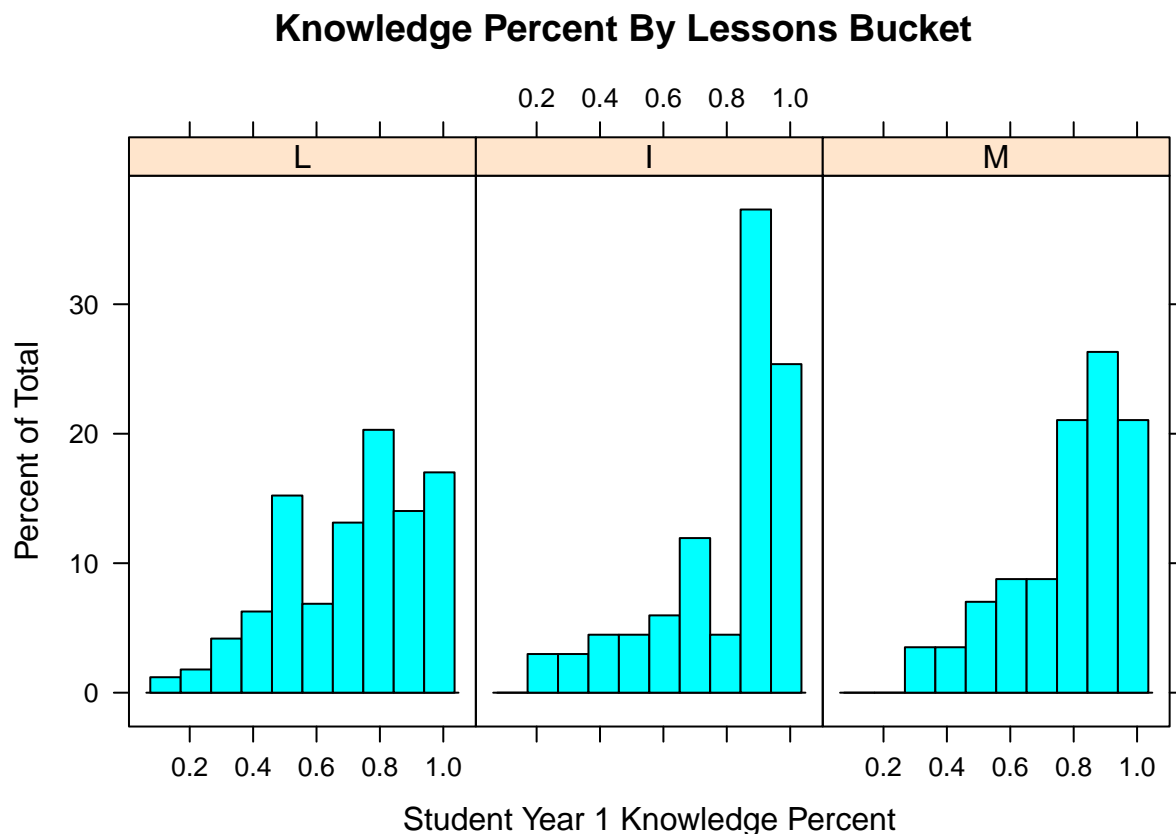
Section 1.3 Ordinal Logistic Regression

As discussed, the data collected by the NIFA project is mostly discrete numbers. This is due to the nature of the pre and post tests given to student and teachers during the program. The Teacher Health Survey consists of 38 questions while the Level 1 student questionnaire consists of a 10 questions, and the Level 2 student questionnaire consists of 18 questions. This defined nature of the test resulted in a discrete number of potential test scores or percentages as outcomes, and therefore violated assumptions of normality necessary to conduct our Multilevel Mixed Linear Model.

We determined that the best possible route to proceed with modeling would be through Ordinal Logistic Regression. An ordinal variable is defined as a variable " with a categorical data scale which describes order, and where the distinct levels of "of such a variable differ in degree of dissimilarity more than in quality (Agresti, 2010)." Our main variable of interest and our dependent variable is this analysis is the Knowledge Percent scores of the students, which by the above definition should be considered an ordinal variable. While some utilize percentages of grade outcomes as continuous or nominal variables, there are several disadvantages that should be considered during modeling (Hoffman & Franke, 1986). Agresti (2010) discusses the disadvantages of using typical regression methods on ordinal data. Some of these disadvantages include results being sensitive to the scores assigned, methods do not allow for the measurement that accounts for the error of replacing ordinal responses with continuous responses, and the models can make predicted values outside of the range of possible values (ie scoring above 100%). Lastly, if typical regression models are applied to ordinal data, the model can lead to misleading results due to "floor and ceiling effects on the dependent variable" (Agresti, 2010, section 1.3.1).

Ordinal logistic regression (OLR) is used to determine the relationship between a set of predictors and an ordered factor dependent variable. This is especially useful when you have test scores data, or proportional data. OLR is more appropriate to use than linear mixed effects models in this case because although these

observations at first seem numeric, these values are inherently categorical. For example, a 3/10 on an exam will always give you the same percentage 33.33% The most common form of an ordinal logistic regression is the "proportional odds model".

To complete our analysis, we utilize the clmm function of the `ordinal` package package in R. CLMM stands for Cumulative Link Mixed Models. The coefficients for OLS are given in ordered logits, or ordered log odds ratios.

It is important to note the difference between probability and odds when utilizing these models. Probabilities are considered proportions or percentages. You can get probabilities by dividing the occurrences of an event by the total number of observations. Odds are the ratio of the probability of one event to the probability of another event, or simplifying as the ratio of the frequency of X to the frequency of Y. For probabilities, if the chances of two events are equal, the probability of either outcome is 0.5, or 50%. Probability ranges from 0 to 1 (0% to 100%). If the odds equal 1, the probabilities of the outcomes are equal. If the odds are lower than 1, the probability of the second event is greater than the first. If odds is higher than 1, the probability of the first event is greater than the second event.

Log odds are logarithmically transformed odds. Log odds are also called logits. Note that logarithmically transformed here means the natural log, not base-10 log. An odds ratio is the ratio of two odds. It tells you if the odds for a particular event is more or less likely in a particular scenario over another. A log odds ratio is the log of the odds ratio. If a log odds ratio is positive, the specified level boosts the chances of a selected outcome. If a log odds ratio is negative, the specified level decreases the chances of a selected outcome. Log odds are centered around 0 (because $\ln(1) = 0$, so when odds are equal, $\ln(\text{odds}) = 0$.

The format of the OLS proportional odds model is as follows. This interpretations will become important as we move into the modeling.

$logit[P(Y \leq j)] = \alpha_j - \beta x, j = 1...J - 1$

As discussed by R. Christensen in Cumulative Link Models for Ordinal Regression with the R Package ordinal, "We interpret this to say: The log odds of the probability of getting a rating less than or equal to J is equal to the equation $\alpha_j - \beta x$, where α_j is the threshold coefficient corresponding to the particular rating, β is the variable coefficient corresponding to a change in a predictor variable, and x is the value of the predictor variable. Note, β is the value given to each coefficient corresponding to a variable, which is similar to a coefficient in a linear model. However, while we have in a linear model the coefficient for a variable in the original units of the response variable, this model gives us the change in log odds. The α value can be considered an intercept of sorts (if comparing to a linear model) - it is the intercept for getting a rating of J or below, again in log odds.

Since we have defined the relationship between probability and log odds as P(X) = exp(X)/(1 + exp(X)) (where X is the log odds ratio), we can extend our definition of the proportional odds model to be:

$P(Y \leq j) = \frac{exp(\alpha_j - \beta x)}{1 + exp(\alpha_j - \beta x)}, j = 1...J - 1$."

We hypothesize our null: There is *no statistical difference* between Knowledge Scores of students who received the intervention (3 lessons or more) and those who did not receive the intervention (2 or less) across the length of the program.

Therefore, our alternative hypothesis is: There *is a statistical difference* between Knowledge Scores of students who received the intervention (3 lessons or more) and those who did not receive the intervention (2 or less) across the length of the program.

Section 2: Data Analysis and Modeling

Section 2.1: Variables of Interest

As we begin our clmm modeling, we take note of specific variables of interest. In our model notation, KP_Bucket (Knowledge Percent as Ordinal Grade Not Integer) is the dependent variable. We have identified

several predictor variables that we will explore as we build and test our models.

These variables include:

Response Variable

KP_brac : This variable was coded by assigning a letter grade based on the percent of correct test answers achieved by the students. It is scaled A-F with A being high, and F being low. Cut off points were: A >= 90, B >= 80, C >= 70, D >= 60, F < 60.

Fixed Effects: (Random Intercepts in the model)

Program_Year: This variable describes the NIFA Program Year Student_Year: This variable describes student participation in the program. lessons_1: This variable is dummy coded with - 1= Student received intervention of: 3 nutrition lessons and - 0= Student received 2 lessons or less. pd_brac (PD Sessions Attended): This variable is coded for "Yes" if the teacher attended PD sessions and "No" if the teacher did not attend pd sessions. test_exp / (Test Exposure): This variable describes how many times the student has seen the test prior to the current attempt. Test: This variable denotes whether the test was a "Pre" test or a "Post" test. Gender_1 : This variable is a dummy coded student gender variable with 1= "Girl" and 0 = "Boy".

Random Effects: (Random Effects allow for random Slopes to be added into the model to account for unobserved heterogenity among groups. As example, allowing our model to account for differences in students within classrooms, or classrooms within schools)

Code: Student ID TeacherID: Teacher ID - in effect setting a classroom for random slope School School_Type: Experimental or Control School Program_Year: Which Year of the program Student_Year: Which year of student participation in the program Level: This variable differentiates between the two levels of the test

Model Assumptions (which will be checked on a couple of the models throughout the code)

The assumptions of the Ordinal Logistic Regression are as follow and should be tested in order: 1. The dependent variable are ordered. 2. One or more of the independent variables are either continuous, categorical or ordinal. 3. No multi-collinearity. 4. Proportional odds

Section 2.2: Data Visualizations

We first look at original scores in percentage form against school type (Experimental vs. Control). We look at student year 1 because this ensures we are looking at the scores of students who have only taken the pre and post test 1 time each. As we can see from the plot, the medians fall around the same area for the pre test scores for each of the school types. But the median post test scores are clearly on different planes, with the experimental score median being slightly higher than the control school. There are 4 outliers we see as well.

```
## pdf
##   2
```

Knowledge Scores Breakdown for Student Year 1
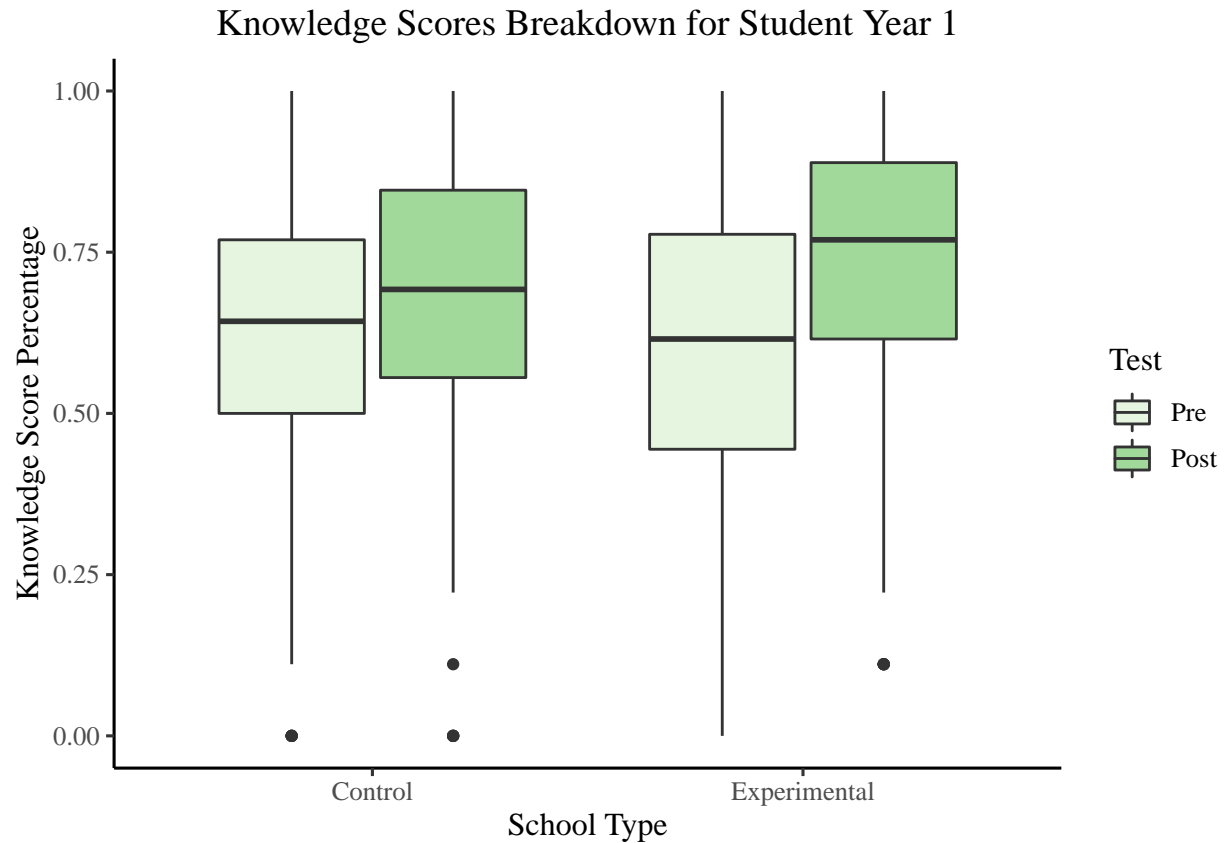
Next, we look at the letter grade scores by school type for Student Year 1. We can see from this histogram that the distributions are fairly even between grades for both school types during the Pre test exam. In the post test, however, we see a large spike in the distribution of letter grade A for the Experimental schools. While both schools appear to have shifted to higher grade scores, the change in the Experimental school is more substantial.

```
## pdf
##   2
```

Breakdown of Student Year 1 Knowledge Grades
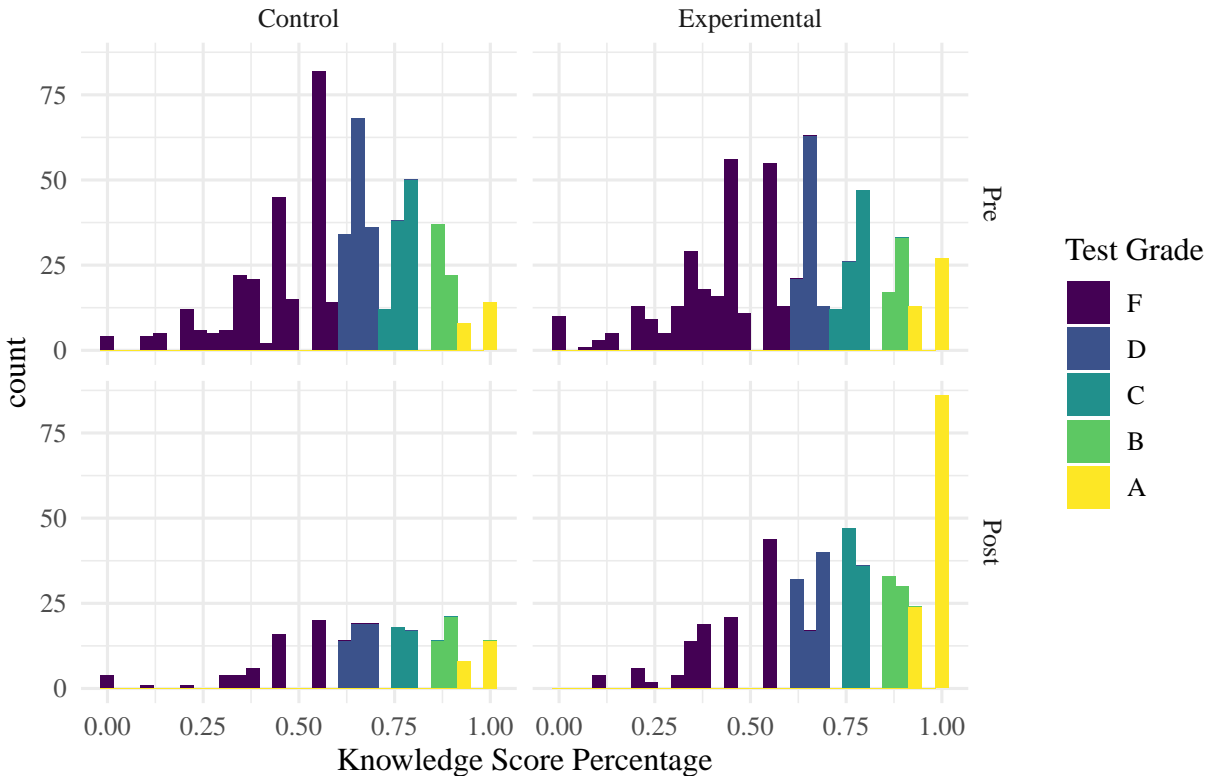
Section 2.3: Modeling

The Null Model:

We begin by creating a null model for our analysis. This model contains KP_brac predicted by only the the intercept. This will be used as a comparison with all other models to assess their significance and determine whether or not added variables increase the accuracy of our model. We will utilize the anova() function to conduct this analysis. The function will take the two models as arguments, and will return an anova test that determines whether or not the more complex model is significantly better at capturing the data than the simpler model. If the p values is less than .05 we will conclude that the complex model is better and will retain it in our analysis.

Random Effects Model

When performing multi level modeling, it is important to look at a model that contains only the random effect variables. We now exclude any fixed effects, and add random effects for Code, TeacherID, and Level within School. Adding these random effects will allow us to account for variance in knowledge between individual students, between different teachers (essential a classroom effect), and for levels nested within schools, since the knowledge test varies between the two levels. We now conduct an anova test to determine whether or not the random effects in this model are necessary. With a p value < .05, we conclude that the random effects added to the model are significant and should be retained when we build the full model.

Fixed Effects Model

Next, we repeat this analysis with a model that contains only fixed effects and no random effects. We will call this the fixed effects model. Our first fixed variables of interest are the research caused interventions, professional development sessions and lessons received. We are also interested in the interaction between these 2 effects. We first analyze the fixed effects model. We conclude that the interaction effect is not significant and we remove it from the model. We rerun the model, with PD Sessions and Lessons Received as fixed effects without the interaction term. We then conduct an anova test to compare the fixed effects model to the null model, and with a p value < .05, we determine that the the more complex fixed effects model is better at capturing the data, and we therefore should retain the fixed effects when building our full model.

Building our full model:

To begin our full model, we combine our fixed effects model and our random effects model. We will call this model fm.1 (full model 1). We include the fixed effects of lessons_1 and pd_brac, and the random effects of Code, TeacherID, and Level nested within school. We then conduct an anova to compare this model with the null model, and conclude that the full model 1 is more significant at the .05 level. *KP_brac ~ lessons_1+pd_brac+ (1|Code) + (1|TeacherID)+(1|Level:School)*

Section 2.4: Checking model assumptions

Like stated earlier, there is a series of assumptions we need to make sure the model follows in order to be complaint with using clmm. The assumptions of the Ordinal Logistic Regression are as follow and should be tested in order: 1. The dependent variable are ordered. 2. One or more of the independent variables are either continuous, categorical or ordinal. 3. No multi-collinearity. 4. Proportional odds

Checking assumptions 1 and 2 –> We know our dataset satisfies assumptions 1 and 2 (see previous analysis)

Checking assumption 3:

From the plot, we can see that lessons received and pdsessions are correlated, they reach a threshold of above 70%. We are concerned about multicolinearity between the two variables. We will continue on to checking the last assumption of proportional odds to determine which variable we should use.

One of the assumptions underlying ordinal logistic and ordinal regression is that the relationship between each pair of outcome groups is the same. In other words, ordinal logistic regression assumes that the coefficients that describe the relationship between the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. This is called the proportional odds assumption or the parallel regression assumption. Because the relationship between all pairs of groups is the same, there is only one set of coefficients.

We can checking this by running the nominal_test() function under the ordinal package. At the time of writing, the nominal_test and scale_test functions don't work with clmm model objects, so we will use the `fixed_effects` model from earlier.

Since we saw potential multicollinearity between pdsessions and lessons received in an earlier assumption, we quickly test to see if there is a difference between a model with just pdsessions and a model with just lessons received. we create the two models and run an anova test to determine which test is more significant.

```
anova(fm.lessons, fm.pd)
```

```
## Likelihood ratio tests of cumulative link models:
##
##              formula: link: threshold:
```

```
## fm.lessons flessons logit flexible
## fm.pd     fpd     logit flexible
##
##           no.par   AIC  logLik LR.stat df Pr(>Chisq)
## fm.lessons     8 6975.8 -3479.9
## fm.pd          8 6961.4 -3472.7  14.348  0
```

From the anova, we can see there is no statistical difference between the model with just pdsession and the model with just lessons received. Since the pvalue for pd_brac in the proportional odds assumption test is slightly more significant than the value for lessons_1 (at the alpha = .05), we choose to only continue with lessons_1 for our remaining models. While this is not a perfect solution and is a limitation of the clmm package, as there does not exist a test of proportional odds for multiple level models. Perhaps in future modeling with the dataset, partial proportional odd models that allow the assumption of proportional odds to be relaxed can be tested.

Section 2.5: Interpreting Full Model 1

Now we will interpret the results of the first model.

clmm(KP__brac ~ lessons__1 + (1|Code) + (1|TeacherID)+(1|Level:School))

A summary of our model provides basic information about the model fit. There are two coefficient tables; one for the regression variables and one for the thresholds or cut-points. "Often the thresholds are not of primary interest, but they are an integral part of the model. It is not relevant to test whether the thresholds are equal to zero, so p-values are not provided for this test. The condition number of the Hessian is a measure of how identifiable the model is; large values, say larger than 1*e4 indicate that the model may be ill defined (Christensen 2015)". From this model it appears that students who received lessons have an estimated coefficient of a 1.733 log odds increase in Grade. But looking at the Hessian score, we have a 1.4e+02 which means our model is ill fitting.

From the model summary, we have 1 variable coefficient and 4 threshold coefficients. Each of the variable coefficients take a beta value in our model specification given below. We can consider the coefficient similarly to coefficients in linear models. That means when we calculate the log odds ratios, whenever we are looking at an observation with lessons_1 = Yes, we include the beta as 1.733, and x = 1. When Lessons_1 == Yes, we expect a 1.733 log odds increase in the expected value of grade. This means that the log odds likelihood of achieving an A vs. a B:F increases by 1.733, the log odds likelihood of a B vs. C:F increases by 1.733, and so on. When we are looking at an observation with lessons_1 = No, we include the beta as 1.733, and x = 0.

`summary(fm.1)`

```
## Cumulative Link Mixed Model fitted with the Laplace approximation
##
## formula:
## KP_brac ~ lessons_1 + (1 | Code) + (1 | TeacherID) + (1 | Level:School)
## data:    data
##
##  link  threshold nobs logLik   AIC     niter     max.grad cond.H
##  logit flexible  2400 -3479.89 6975.77 555(2224) 1.10e-03 1.4e+02
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
```

```
## Code        (Intercept) 1.0100   1.0050
## TeacherID   (Intercept) 0.7027   0.8383
## Level:School (Intercept) 0.1135  0.3369
## Number of groups:  Code 1303,  TeacherID 67,  Level:School 8
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## lessons_1Yes   1.7331     0.2266   7.649 2.03e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##      Estimate Std. Error z value
## F|D  -0.6384     0.1713  -3.726
## D|C   0.5808     0.1715   3.386
## C|B   1.6640     0.1773   9.386
## B|A   2.8109     0.1894  14.838
```

Exponentiating the Coefficients

If we want to look at this on a odds scale, we exponentiate the coefficients. Now, when lessons_1 = Yes, we include the beta as 5.57, and x = 1. When Lessons_1 == Yes, we expect a 5.57 odds increase in the expected value of grade. This means that the likelihood of achieving an A vs. a B:F increases by 5.57, the likelihood of a B vs. C:F increases by 5.57, and so on. When we are looking at an observation with lessons_1 = No, we include the beta as 5.57, and x = 0, at which point beta * 0 would omit lessons_1 from the equation.

If lessons = Yes, the odds ratio of receiving any grade but an A decreases by 5.57. If lessons_1 = "No", the odds ratio of receiving any grade but an A remains the same, at the level of the threshold coefficient.

In this case for exponentiated log odds, we have a threshold coefficient for B|A of 16.62. This threshold coefficient is the liklihood of receiving any grade BUT an A. Therefore, the odds ratio of receiving any grade below an "A" would decrease from 16.62 to 11.05 if the student had received lessons.

```
exp(coef(fm.1))
```

```
##        F|D        D|C        C|B        B|A lessons_1Yes
##  0.5281286  1.7875512  5.2803971 16.6250451    5.6578880
```

Visualizing These Findings:

Next we look to visualize these findings. In the graph below we utilize ggpredict to assess the probabilities of receiving each specific grade. We have successfully accounted for random slopes due to Student Code, Teacher Code, and the nested effect of Level within School. We can see that when a Student received less than 3 lessons, the probability of receiving a Grade of 1 == F was higher than the estimated log odds of receiving a Grade of 1 ==F when a student received 3 or more lessons. On the other hand, we can see that when a student did receive 3 or more lessons, the estimated log odds of relieving a Grade of 5 == A was lower than the estimated log odds of relieving a Grade of 5 ==A when a student received less than 3 lessons We think it is also significant to notice the clear downward sloping trend of the log odds of the predictions without lessons around the Grade 3 == C.

```
graph3
```

Probabilities of Knowledge Grades based on Lessons Received
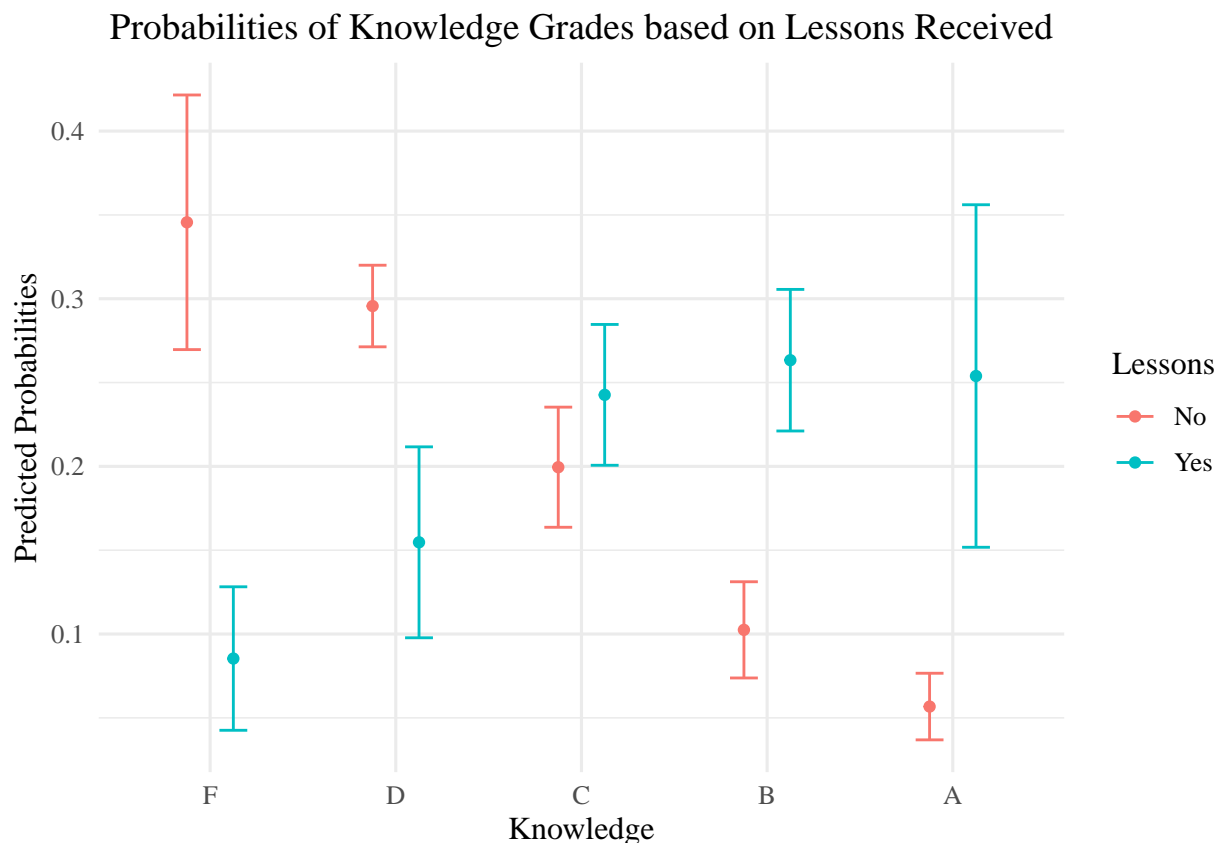
Section 3: Other Models

In this section, we run through a series of different variable combinations. For the sake of saving time, we will not provide output or interpretation of these models in this section. If you are interesting in reviewing our model selection process, please check out or .rmd file.

3.1 Analyzing our final model

After conducting a series of variable selection procedures we arrived at our final model. Our final model consists of the following formula:

clmm(KP_brac ~ lesson_1 + test_exp:Test + test_exp:Program_Year + Gender_1 + (1|Code)+(1|TeacherID)+(1|Level:Sch

All terms in the model are significant (p<.05) and an anova between this final model and the null model conclude that the model is significantly better at capturing the data.

Section 3.2 Checking Assumptions

Checking Assumption 3 Colinearity:

Since assumptions 1 and 2 hold true from earlier analysis, we now conduct a Goodman Kruskal test to check for multicolinearity. We are not concerned with colinearity within this model.

Check assumption 4: proportional odds assumption

To check proportional odds, we must first run a clm while omitting all random effects from our model. Using the alpha = .01 level for a little more leniency in the model, the only term we are concerned that violates the proportional odds assumption is Program_Year. We are still wary we that the nominal_test() would apply properly to our clmm model since that function is currently not available for models that incorporate random effects. This being the case, and since we are failing to account for the random effects in the multilevel model while conducting this test, we choose to retain Program_Year in the final model. *This may be worth revisiting as the clmm package is updated, but it is outside of the scope of this project.*

Section 3.2: Final Model Interpretations

Now that we have checked all the assumptions for the model, we will move into the interpretations and predictions. The summary of our model provides basic information about the model fit. There are two coefficient tables: one for the regression variables and one for the thresholds or cut-points.

Model Summary

From the model summary, we have 7 variable coefficients and 4 threshold coefficients. Each of the variable coefficients take a beta value in our model specification given below. We can consider the coefficient similarly to coefficients in linear models. As example, when we calculate the log odds ratios, whenever we are looking at an observation with lessons_1 = Yes, we include the beta as .95, and x = 1. When Lessons_1 == Yes, we expect a .95 increase in the expected value of grade on the log odds scale. This means that the likelihood of achieving an A vs. a B:F increases by .95, the likelihood of a B vs. C:F increases by .95, and so on. When we are looking at an observation with lessons_1 = No, we include the beta as .95, and x = 0.

Lessons Received:

In terms of its actual meaning in relationship to the variables, we would say that for a one unit increase in lessons_1 (i.e., going from 0 to 1, or No to Yes), we expect a .95 increase in the expected value of grading on the log odds scale, given all of the other variables in the model are held constant. In other words, when going from No to Yes, the likelihood of an A versus a B-F on the grading scale increases by .95 on the log odds scale, the likelihood of a B versus a C-F on the grading scale increases by .95 on the log odds scale, the likelihood of a C versus a D-F on the grading scale increases by .95 on the log odds scale, and the likelihood of a D versus a F on the grading scale increases by .95 on the log odds scale.

Test Exposure

In this case, we say that for each one unit increase in Test Exposure, we expect a 1.24416 increase in the expected value of Grade on the log odds scale, given all of the other variables in the model are held constant. In example the likelihood of a A versus a B-F on the Grading scale, when test exposure = 1 increases by 1.24416 log odds. If Test Exposure = 2, the likelihood of an A versus a B-F on the grading scale would increases by (1.24416 x 2) log odds.

Program_Year

In this case, we say that for a one unit increase in Program Year, we expect a 0.39 increase in the expected value of Grade on the log odds scale, given all of the other variables in the model are held constant. In other words, when Program_Year = 1, the likelihood of an A versus a B-F on the Grading scale increases by 0.39 log odds. When Program Year = 2, the likelihood of an A versus a B-F on the rating scale increases by 0.39 x 2 log odds, and so on.

Test

In terms of its actual meaning in relationship to the variables, we would say that for a one unit increase in Test (i.e., when Test = Post = 1), we expect a 1.14 increase in the expected value of grading on the log odds scale, given all of the other variables in the model are held constant. In other words, when Test = Post, compared to Test = Pre, the likelihood of an A versus a B-F on the grading scale increases by 1.14 log odds. Additionally, the likelihood of a B versus a C-F on the grading scale increases by 1.14 log odds, and so on.

Gender

In terms of its actual meaning in relationship to the variables, we would say that for a one unit increase in Gender (i.e., when Gender = Girl), we expect a .34 increase in the expected value of grading on the log odds scale, given all of the other variables in the model are held constant. In other words, when Gender = Girl, the likelihood of an A versus a B-F on the grading scale increases by .34 log odds. The likelihood of a B versus a C-F on the grading scale increases by .34 log odds and so on.

We will interpret the coefficients and the 2 interaction terms using the odds as well. We find the odds ratio by taking the exponential of the log-odds ratio.

```
exp(coef(fm.6))
```

```
##                   F|D                    D|C                    C|B
##            2.1422876              7.9052421             25.0324612
##                   B|A            lessons_1Yes                test_exp
##           84.9671367              2.5826917              3.4672701
##          Program_Year                TestPost                Gender_1
##            1.4785703              3.1517943              1.4535507
## test_exp:Program_Year      test_exp:TestPost
##            0.6827096              0.6431342
```

Section 3.3 Interpretting the odds ratios

This interpretation is similar to the interpretation above. WE have exponentiated the log odds and will now interpret the odds ratios.

Lessons

Now, when lessons_1 = Yes, the odds ratio of an A versus a B-F on the grading scale is multiplied by 2.1 when going from less than 3 lessons to more than 3 lessons, the odds ratio of a B versus a C-F on the rating scale is multiplied by 2.1 , so on and so forth.

Test Exposure

With 1 unit increase in test exposure the odds ratio of an A versus a B-F on the grading scale is multiplied by 3.5 , and the odds ratio of a B versus a C-F on the rating scale is multiplied by 3.5, so on and soforth.

Program YEar

With 1 unit increase in Program Year the odds ratio of an A versus a B-F on the grading scale is multiplied by .68, and the odds ratio of a B versus a C-F on the rating scale is multiplied by .68, so on and soforth.

Gender

With 1 unit increase in Gender (going from Boy to Girl or 0 to 1) the odds ratio of an A versus a B-F on the grading scale is multiplied by 1.4 , and the odds ratio of a B versus a C-F on the rating scale is multiplied by 1.4, so on and soforth.

Test

Now, when Test goes from Pre to Post, the odds ratio of an A versus a B-F on the grading scale is multiplied by .64 (compared to the likelihood of Test = Post), and the odds ratio of a B versus a C-F on the rating scale is multiplied by .64, so on and soforth.

Test Exposure Program Year Interaction:

The interaction between Test Exposure and Program Year allows us to account for changes in variation among test exposure for students that join during different program years. With the interaction between test exposure and Program year, every unit increase in test exposure multiplied by every unit increase in Program Year, is then multiplied by .68. The odds ratio of an A versus a B-F on the grading scale will then increase by that specified amount. As example, if Program Year == 1 , and Test Exposure == 1, then the increase in the odds ratio of A verses B-F on the grading scale would increase by (1x1x.68).

It may be worth nothing that we should watch this interaction's effects as we anticipate that the effects of the Covid-19 pandemic on student knowledge scores may make it more significant in the future.

Test Exposure and Test

The interaction between test exposure and Test allows us to account for students joining the program in the middle of a program year, and therefore offsetting their test exposure from that of other students. With the interaction between test exposure and Test, every unit increase in test exposure multiplied by every unit increase in Test, which is then multiplied by .64. The odds ratio of an A versus a B-F on the grading scale will then increase by that specified amount. As example, if Test Exposure == 1, and Test == Post (Post == 1), then the increase in the likelihood of A verses B-F on the grading scale would increase by (1x1x.64).

3.4 Threshold Coefficients:

Next we look at the threshold coefficients. These are the coefficients, again in log odds, for receiving a rating of below the Jth rating. They can be considered the "cut points" or thresholds between the two variables. So the coefficient reading F|D is the odds of receiving a D rating as opposed to an F. The coefficient reading for D|C would be the odds of receiving an C as opposed to a D-F. The coefficient C|B is the odds of receiving a B rating as apposed to a C-F. Lastly, the B|A coefficient is the odds of receiving an A as opposed to a B-F.

This is relevant because we can use the formula given above to calculate the log odds, and therefore the probability by extension, of receiving a certain score or below for each value of the predictor variable. We can also use this to get the log odds ratio of receiving an exact Grade.

Section 3.2: Visualizing our final model

We next visualize our final model utilizing GGpredict. GGpredict changes odds ratios into probabilities of an event happening given a certain combination of independent variables. Looking at the graph below, we can see that the probability of a student whose first exam is a pre test, they have less than 3 lessons received, they are a boy, and have test exposure = 0, have a probability of receiving an F relative to all other grades

of around a mean prediction of 49%, with confidence interval at 95% about 38% to 60%. Comparing to the same predictors (boy, pre test, 0 test exposure, less than 3 lessons), the probability of them receiving an A compared to all other grades is around 3% with a confidence interval at 95% of 1% to 4%.

```
graph5
```

Probabilities of Grades by Gender, Lessons, Test Exposure, and Test
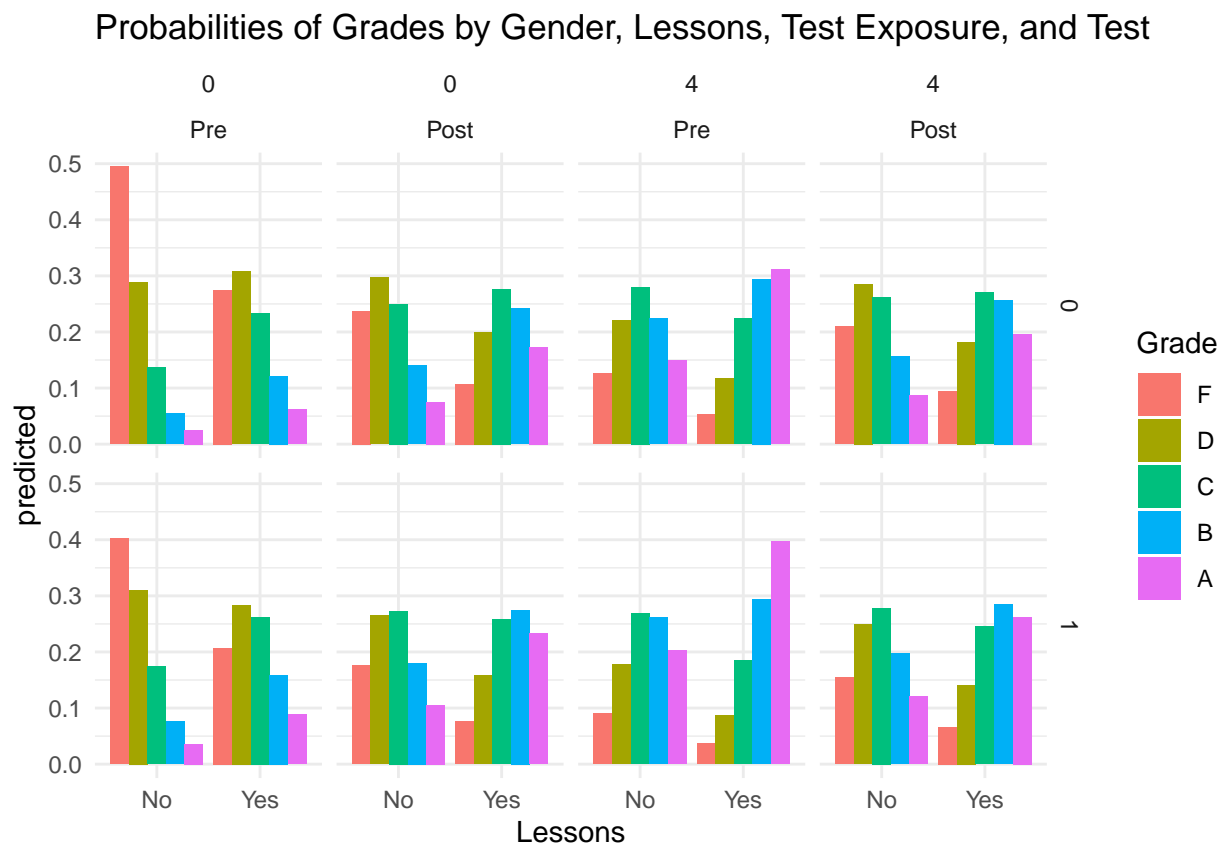
Section 4: Conclusions & Limitations

Section 4.1 Conclusions:

Through our analysis, we found our final model to be Knowledge Percent ~ lessons (Yes/No) + Student Gender + Test Exposure + Type of Test + Program Year + Test:Test Exposure + Program Year: Test Exposure = (1|Code) + (1|Teacher ID) + (1|Level:School) Utilizing a Wald Test, we can see the confidence intervals for our model below. The Wald tests are marginal tests so the test returns predicted confidence levels that measure the effect of fixed variables while controlling for random effects, which in this case includes the random effects of TeacherID, Code, and Level:School.

Nagel Kerkeke Test.

Next, we conduct a Nagel Kerkeke Test to compare this model to the original null model. Looking at the pseudo rsquares, we see much improvement comparing the full model to the null model. The test returns an R2 value of .19. While this number is relatively low, it is important to note that Pseudo R-squared values can not be directly compared to R-squared values for OLS models. They should also not be interpreted

as the proportion of variability explained by the mode. Instead, these measures should be seen as relative measures from which one can gauge how well one model explains the data in comparison to another(R Companion/nagelkerke, 2020).

We consider the pseudo r square value returned form the Nagel Kerkeke Test comparing our first fit (KP_brac ~ lessons_1 + (1|Code) + (1|TeacherID) + (1| level:School)). We have increased our pseudy r2 from .13 to .19, Therefore we do see improvement. As the Nutrition Intervention Program progress through the years and more data is collected, the ability to use other variables in the model may greatly improve, and result in a higher prediction power for these models.

As we breakdown our final model we see significance in the following variables: The intervention of Lessons Received is significant, with the intervention of three or more lessons received resulting in a .95 increase in the odds ratio (p = .00017, Estimate = .94, SE = .25, Z value = 3.75). Test Exposure is significant, with each one unit increase in test exposure resulting in a 1.2 increase in the odds ratio (p = 1.0e-6, Estimate = 1.2, Z = 4.89). There is a significant effect in Program Year, with each one unit increase in program year resulting in .39 increase in the odds ratio (p = .0005). There is also significance in Test = Post (p-value = 8e-13,) with an increase in the odds ratio when Test changes from Pre to Post. There is also significance in the students' gender (p-value = 0.002), with a change from Boy to Girl with an .34 increase in the odds ratio. The interaction of test exposure and program year, and the interaction of test exposure and test were also significant (p-value = 3.5e-5, p-value = 3.7e-05 respectively).

From our analysis, we make a final conclusion. We reject our H0, and we conclude that the Nutrition Intervention Program Lessons have a significant effect on student outcomes in regards to nutrition knowledge and understanding.

Limitations

There are several limitations to this study. The first limitation is the missing data, especially in regard to Teacher Surveys. While we were able to impute on some variables within the dataset, results of modeling can be biased and even inaccurate in the presence of high levels of missing data that have been imputed. On the same note, using mean imputed data can lead to higher biases in the data including elevated correlations and underestimated standard errors. If the probability of missing data for a variable depends on the actual value of the variable, then multiple imputation is inappropriate. Therefore, future analysis could use other variables in the modeling if the amount of missing observations decreases.

Our next limitation of this study is the impact of COVID-19. Students, teachers, and schools have been greatly affected by the pandemic and the associated shifts away from typical learning structures. The impact of the pandemic on students and learning has yet to be fully researched, but we hypothesize that there will be notifiable changes around education. We suspect that recent switches to online learning, and inaccess to school provided lunches and healthy options may impact the results of this program in upcoming years. One way to account for the years of COVID-19 will be to come up with a way to weight the data collected during the rogram years most affected by the COVID-pandemic.

Our final limitation in this study was in the lack of availability of multilevel data analyses packages. While we utilized the clmm package, there have been some lags on development of the package, including in regard to checking the assumptions of clmm models. This required us to utilize a less complex model to check our assumptions, which should not be considered a standard or best practice. That said, with the packages and programming tools at our disposal, we could not find an alternative at this time, and such an endeavor is outside of the scope of this project. In future analyses, it might be useful to use a different package from clmm, such as `mixor`, that provides a function for fitting an ordinal response model when observations are either clustered or collected longitudinally. The `brms` package could also be of use. Brms functions Fit Bayesian generalized (non-)linear multivariate multilevel models using 'Stan' for full Bayesian inference.

Works Cited

###Agresti, A. (2010). Analysis of ordinal categorical data (2nd ed.), Wiley Series in Probability and Statistics Hoboken, New Jersey: Wiley.

###Christensen RHB (2018). "ordinal—Regression Models for Ordinal Data." R package version 2018.8-25, URL http://www.cran.r-project.org/package=ordinal/.

###Hoffman, D. L. , & Franke, G. R. (1986). Correspondence analysis: Graphical representation of categorical data in marketing research. Journal of Marketing Research, 23(3), 213–227.

###Kim, T & Park, J. 2019. "More about the basic assumptions of t-tests: normality and sample size. www.ncbi.nlm.nih.gov/pmc/articles/PMC6676026

###Kwak, S Y Kim, J. 2017. "Central limit theorem: the cornerstone of modern statistics. doi: 10.4097/kjae.2017.70.2.144

###Research attributed to Melissa Hawkins, PhD, MHS1; Erin Watts, MPH1; Sarah Irvine Belson PhD2;Anastasia Snelling, PhD, RD1