**SMALS Research Project 2**

Eric Shin

University of Washington Bothell

March 6th, 2023

**Abstract**

Students in a statistics course made concept models throughout the course. Previous studies suggest that making concept models is a useful way for instructors to assess learning. The growth and changes of these student-constructed models may serve as an indicator of how a student constructs and organizes their cognitive structure. But, little is known about how much impact the Covid-19 disruption has had on a student’s cognitive restructuring in STEM courses. Therefore, we ask the following research question: What factors best explain the changes in the student-made models about statistics as representations of their cognitive structure? To address this research question, we analyzed concept models from three assessment points and two quarters of the same course, one taught before the Covid-19 disruption and the other during the Covid-19 disruption, to estimate the impact that the distributions had on their cognitive structures. After converting the 180 student models into graphs, we calculated the number of concepts in each model to build a linear model to compare a quarter taught before and during the major disruptions. First, we created a box and whisker plot that shows model growth across two different quarters, pre and during, and then we created a model to explain how the disruption term influenced the rate of concepts being added to student models at three different assessment points of the course. Our preliminary results revealed that students in the emergency online environment added concepts to their models at a significantly slower rate. Future directions and limitations will be discussed.

**Methods**

The data set that is used to analyze the changes in the student-made models through our linear model is titled, “NetworkResults.” This data set contains 180 transcribed student concept models collected from UW Bothell. Each individual row represents the individual student’s transcribed model. The data set also contains network measures such as the number of nodes and edges of each student model. The number of nodes represents the number of concepts of each student model and the number of edges represents the number of connections between the concepts of each model. These measures are calculated based on different factors that are separated into different columns. The relevant columns from the data set include the two different quarters (Pre-Covid and During Covid) labeled as, “AcademicTerm,” the concept models at three different assessment points of the quarter labeled as, “ExamType,” as well as the two statistics courses (BIS 215 and BIS 315) labeled as, “Course.” This data is being used to address the research question as a way of analyzing the different factors to see if each one has any influence on the rate at which students are adding concepts to their models. This will indicate which factors are good predictors of changes in student models across the three time points of the quarter.

**Results**

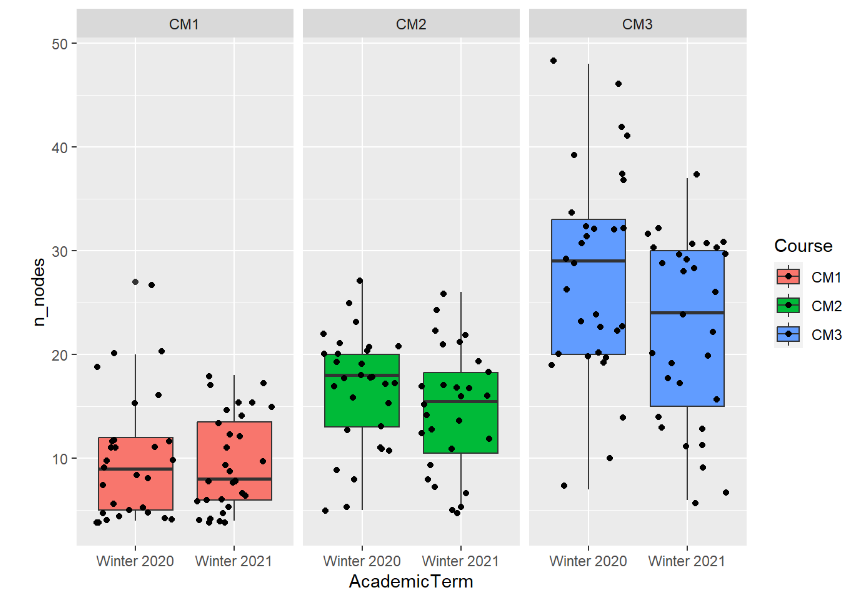


Figure 1: Box and Whisker Plot

We created a box and whisker plot with the ggplot package to visualize whether or not the predictor variable, which consisted of two different quarters, had an influence on the concept models at three different assessment points with the number of nodes as the outcome indicating the number of concepts. The two different quarters were from Winter 2020 (Pre-Covid) and Winter 2021 (During Covid). The box and whisker plot contains the two quarters as the x-value, the number of nodes as the y-value, and the concept models at three different assessment points as the fill value for the box and whisker plots. We can see that the median of the box and whiskers decreased from Winter 2020 to Winter 2021 at all three time points. We can also see that the number of nodes also decreased between the two years at all three time points. This indicates that at all three time points of the course, students are adding concepts at a lower rate during the Covid disruption than the year before. This is the most useful model from the study because the graph interpreted the two different quarters as strong indicators for the changes in student-made models. From this finding, we can infer that the quarters serve as an important predictor from the data set to analyze the changes in the student-made models about statistics as representations of their cognitive structure since we are measuring student model changes both before and during the Covid-19 disruption.

**Discussion**

In regards to the research question that was made as the guide for analyzing the given data, we made a key finding that the two different quarters, pre-Covid disruption and during the Covid disruption, had an influence on the changes in the student concept models at all three assessment points of the quarter, with lower rates of added concepts occurring during the Covid disruption quarter. This implies that switching to the online environment for the statistics course due to the disruption had a negative impact on how much the students were learning. The biggest limitation that we had with this part of the research was the sample size of the data set. Since we were using a data set of only 180 transcribed student models, it may have been harder to have extrapolated findings as well as different variations of findings with the data. The number of different factors that we analyzed as possible predictors of the student model changes was another limitation that we had since we mainly focused on the pre-disruption and during disruption quarters as the predictors. Moving on forward, to further this research, we need to use a data set that has a bigger sample size of the transcribed student concept models collected from UW Bothell to reduce the margins of error as well as have unbiased findings. Next, we need to analyze other factors from the data set that may serve as good predictors for the number of concepts that students are adding to their models at all three assessment points of the quarter. For example, we can conduct research on whether or not the two different quarters are strong predictors for what type of concepts students are understanding more and which concepts have the most connections with other concepts in the model. Finally, we need to consider creating other figures that support the findings from the box and whisker plot that was created for this part of the research. Examples include residuals that help us find the differences between observed and predicted values of the data set we are working with as well as word clouds that help us analyze the frequencies of specific concepts that are appearing in the student models.