

Project: Model & Graphics Plans

Ivan, Hadong, Alex

Question 1

How does the different time variables affect student's confidence for their health assessment?

Description of the graphs we will make in order to show the answer:

We would use a boxplot with Time_Point and Confidence_of_Health_Assessment_num. We would do this as we are working with a qualitative variable (time_point) and we don't really know what Confidence_of_Health_Assessment_num is. We are assuming that this is a quantitative variable. Because of this, the boxplot work well to show how Confidence_of_Health_Assessment_num varies across different Time_Point categories. Boxplots do a good job on showing distributions of numeric data values, especially when you want to compare them between multiple groups.

Question 2:

Does previous VR_Experience affect their confidence for the health assessment?

Description of the graphs we will make in order to show the answer:

This graph would be the same. A boxplot would be a good choice as we use the same variables, one quantitative and one qualitative. Therefore, it will be the same as the above graph to answer our question.

Model plan ideas

As our response variable we can use `Confidence_of_Health_Assessment_num` as a measure of their confidence. We couldn't really find any other measure of confidence and thought this was the best choice. Based off of the number of rows in our data we can only have about 10 betas, this could be an issue.

We are still unsure if our data for `Confidence_of_Health_Assessment_num` is quantitative or qualitative. Even if we changed the data to numeric values, it could still be qualitative. However, we are currently assuming that it is quantitative and so, this is how we are doing our plan:

1st Question Plan

Response: `Confidence_of_Health_Assessment_num`.

Predictor: `Time_Point`

Confounder: no confounders because nothing affects time

Mediation Chain: nothing affects time in our data so we do not have any variables here.

Precision Covariates: `Year_Of_College`, `Actual_Simulation_Expiriences`, `Health_Assessment_num`, `Nursing_Process_num`, `Assesment_Data_Knowledge_num`, `Nursing_Interventions_num`

2nd Question Plan

Response: `Confidence_of_Health_Assessment_num`.

Predictor: `Actual_Simulation_Expiriences`

Confounder: `Year_Of_College`

Mediation Chain: nothing really affects `Actual_Simulation_Expiriences` so there are none.

Precision Covariates: `Health_Assessment_num`, `Nursing_Process_num`, `Assesment_Data_Knowledge_num`, `Nursing_Interventions_num`

Question for prof:

1. What type of data are we working with? Count Data? (why does `nbinom` not work??)
2. If so, how would we implement the $m/15$ rule or $n/15$ rule?

3. Note that fitting a linear regression model with the “numeric” versions of *one* of these questions as the response variable usually leads to many problems with conditions. There are other options - we may need to meet to discuss. One is mentioned below.

Computing Scales?

We might want to sum numeric responses across several similar questions to get a “scale” measuring student responses to a group of like questions.

I do have code to do this, but it’s for the “SET-M” questions which I don’t think are in your dataset (they are arranged in sections predebrieing, learning, confidence, and debrieing. Unless those sound familiar you don’t have them!)

So, if you want/need to do this, then I will need to know *which* questions ought to be grouped together. The choices should be approved by Prof Kunnen before we move too far on it, and I am also happy to meet with your team to discuss this option, and other possible ways of working with this dataset.