**Module 1: Exercise 2**-School District Dataset Analysis

Elizabeth Elizondo

MIS 5375 680

Dr. Vanessa Garza Clark

Texas A&M International University

September 03, 2021

To preface how the data was applied in WarpPLS, the latent variables were applied with the measurement model as *Reflective*. The SEM model settings were: Outer model analysis algorithm ‘Robust Analysis’ for computational efficiency (1)(2), the Inner Model as ‘linear’ since the variables are measured through one single indicator, and the Resampling method ‘Stable3’. The first attempt at SEM analysis did not have the *missing data imputation* settings adjusted and were initially left null (the results appeared to have extreme collinearity along with a very low coefficient of determination and was determined to be valid results and were excluded from the final findings). The second attempt applied the data imputation method with the software’s default *Arithmetic Mean Imputation* and this provided more credible end results to assess and evaluate.

The question states, “What is the order of importance of the predictors with respect to SAT scores?”

Reviewing the SEM model, the predictors in this analysis are: average years of teaching experience by the school district teachers and number of suspensions in the district due to student behavioral problems. Figure A. reveals a p-value (level of significance) is a 0.02 in the correlation between number of student suspensions to SAT scores, this means that the relationship between the variables in the model was significant and had did have an effect. On the other hand, the correlation between the variables of teachers’ years of teaching experience and average SAT score revealed a p-value of <.01, thereby indicating that there is a *greater significance* between these two variables. The coefficient of determination indicates that 16% of the variation of the predictor variables are a result of the linear regression model, the other 94% are due to other predictors (3). The variance in both predictors have a minimal yet significant level of significance which concludes the model can be diagnosed as having limited *collinearity* in which there is a dramatic increase in the p value of one predictor variable like that shown between the relationship of student suspensions and student SAT scores (4).

A picture containing text, necklet, accessory

Description automatically generated

**Figure A.**

Figure B. shows the scatterplot of datapoints from the teacher’s years of experience and its correlation to student SAT scores. The standard deviation of years of experience is at a 1.64, this tells us that the distribution of data points is centered ‘good’ around the mean of 12.41. We can also see the coefficient of determination has a good fit, in that the scatterplot is reasonably near the estimated regression line. We can conclude that schools with teachers that have years of experience of roughly 12 years or so, have students which obtain higher SAT scores.

Chart, scatter chart

Description automatically generated

**Figure B.**

Figure C. shows the correlation between student SAT scores and student suspensions. The scatterplot appears to have a loose fit of points surrounding the estimated regression equation, and has many outliers and appears to have a bulk of the data points closer to the lower number of student suspensions in the school district. It can also be observed that the standard deviation for this particular predictor is very high, thus proving that the number of student suspensions has a lesser correlation to SAT scores than the predictor previously evaluated.

Chart, scatter chart

Description automatically generated

**Figure C.**

In conclusion, to answer the posed question of which level of importance of the predictors with respect to SAT scores, we reviewed Figure A., B., and C., and determined that teachers with a greater number of years in teaching experience has a significant correlation to student having higher SAT scores. The number of student suspensions in this hypothesis as a predictor of SAT scores revealed a very high standard deviation and low coefficient of determination, the result can be concluded that suspensions have a lower level of importance in respect to student SAT scores than years of teaching experience.

References

1. (2017, Aug. 1). Use Data Labels to Better Understand Small Samples in WarpPLS [Video File]. Retrieved from <https://youtu.be/i5-_WIMXVl4>
2. Kock, N. (2020, August). WarpPLS© User Manual: Version 7.0. https://scriptwarp.com/warppls/UserManual\_v\_7\_0.pdf.
3. Wawrdhani, N. W. S., Fernandes, A. A. R., & Nugroho, W. H. (2019, October 30). *Structural Equation Modeling Analysis with Warppls Approach Based on Theory of Planned Behavior*. http://www.hrpub.org. https://www.hrpub.org/journals/article\_info.php?aid=9354.
4. Enders, F. B. (n.d.). *Collinearity*. Encyclopedia Britannica. https://www.britannica.com/topic/collinearity-statistics.