Penguins Team DataFest Document

Peter Kuvshinov, Karim Rizk, Mitchell Dolny

Purpose:

The aim of our analysis was to determine if the student learning experience differed between chapters and to identify educational deficiencies so that appropriate changes to

problematic chapters could be made by course authors.

Analysis:

We decided to measure the educational quality of a chapter by the EOC (End of Chapter) value of every student for every chapter. We chose to ignore data from highschool students as we felt that the textbooks were not written with this demographic in mind. We also ignored the “review” pages as some chapters lacked them and we felt that it was necessary to evaluate chapter quality on the base material alone.We created an adjusted EOC that applies a penalty if a student repeats many questions when writing an end of chapter quiz.

Anova:

We created a model with the mean adjusted EOC of each chapter as the response variable, and the covariates: mean engagement time per section(in minutes), the average number of attempts per question per chapter, an indicator variable based on if there are videos in the chapter, the mean of the proportion of revisited pages in each chapter. Before trying to make any inferences we used VIF() to check for multicollinearity, we found no evidence, thus with this model, we created an ANOVA table.

Using this ANOVA table, we saw that there was no evidence that having a video had any effect on the mean adj EOC, nor did the mean engagement time per section have any effects on it too, (at a 5% significance level), but we noticed that the average number of attempts per question per chapter, and the mean of the proportion of revisited pages in each chapter, showed evidence that they had an effect on the mean adjusted EOC.

We then decided to focus on one of the significant covariates. We ultimately decided to use the mean proportion of revisited pages in each chapter(compl) instead of the proportion of chapter questions redone, as the values in the latter exhibited strange large values whose validity we were unsure of. From there, we created a linear model based on Mean Adjusted EOC & the mean proportion of revisited pages in each chapter which showed a moderate positive correlation. Finally, from there we noticed 2 clusters 4,13,14,15,16 and 2,3,5,...,12 and one outlier; chapter 1.

Conclusions:

In conclusion we have discovered from our adjusted EOC that the students were not as prepared for the EOC questions as originally shown based on the original EOC. We can see there is a moderate positive correlation between Mean Adjusted EOC and mean proportion of revisited pages which may indicate that the more students revisit pages the better their scores. This could show a flaw in the delivery of educational material as students may need to re-read the same material more than once to do better.

One solution would be to implement more in-depth material and mid chapter checkpoints to better improve the understanding of the concepts in each chapter.