![PISA logo]

# 23 Mar 2017 Supervisor Meeting 10

*Created by: Jeremy Ong (23/03/17)*
*Edited and Vetted by: Chermain Ang(23/03/17)*

| DATE | TIME | VENUE |
|---|---|---|
| 23-Mar-17 | 2:00 PM – 3:00 PM | SMU SIS MR 4.6 |
| **Meeting Type** | Supervisor Meeting | |
| **Facilitator** | Chermain Ang | |
| **Note taker** | Jeremy Ong | |
| **Timekeeper** | Gareth Ng | |
| **Attendees** | Chermain Ang (CA), Gareth Ng (GN), Jeremy Ong (JO), Prof Kam Tin Seong (TS) | |
| **Absentees** | NA | |
| **Agenda** | 1. Updates on Multiple Linear Regression analysis. | |

| MEETING ITEM 1: Updates on Multiple Linear Regression analysis. | | Time Allocated 60 mins |
|---|---|---|
| **Name** | **Discussion** | **Follow Up By** |
| CA | a. Showed Prof Kam the steps performed by the group for Multiple Linear Regression (MLR) to check that we are doing it right.<br>b. Showed the list of continuous variables and their correlation matrix and shared how we determine which of the highly correlated variable to drop. | |
| TS | c. Elaborated on the results of MLR in JMP, and how the software treats nominal variables. | |
| JO | d. Asked about how to read the output of stepwise analysis. | |
| TS | e. Shared that the numbering in brackets next to the questions are for nominal variables, telling us what options are considered for the grouping.<br>f. Mentioned that if the response is a range from good to bad, for example, we should change the classification to ordinal.<br>g. Pointed out that when we look at the results, we should refer to the Adjusted R Squared value as well, instead of just comparing only the R Squared results.<br>h. When reporting the results, to use the effect test output – which should not have dummy variables in it. | GN |
| GN | i. Showed the analysis of Score Distribution by schools, and the derived variables from stepwise. | |
| TS | j. Mentioned that if we want to categorize school performance for analysis, to consider segmenting | JO |

| | | |
|---|---|---|
| | them by their score distribution – top 25%, middle 50%, bottom 25%, then perform recursive partitioning. Can be done on students dataset as well. This is a data mining approach of looking at the data.<br>k. Guided the team on using decision tree analysis based on the student performance categories (top, mid, btm).<br>l. Suggested that we can consider analyzing only two groups of student performances. (i.e. top and the rest OR btm and the rest OR mid and the rest).<br>m. Highlighted that for the report, it is up to the team to decide how and what we want to write on the research done, to focus on only the linear regression method, or on recursive partitioning, or both.<br>n. Informed the team that a new folder is created on elearn for the final artefacts, and advised us to upload the files of our analysis done to date (optional). | |
| CA | o. Thanked Prof Kam for his advice and called an end to the supervisor meeting. | |
| **Remarks** | | |
| | | |

## To-do

| No. | Action Items | Person I/C | Deadline | Remarks |
|---|---|---|---|---|
| 1g, h | 1. Reclassification of variables and re-run MLR for school dataset. | GN | 20 Mar 17 | |
| 1k, l, n | 2. Explore decision tree analysis for student dataset.<br>3. Upload current analysis files into elearn | JO | 20 Mar 17 | |