Singapore Management University
ANLY482 Analytics Practicum

Supervisor Minutes 6 as on 27th February 2017

Time Start:	2.30pm
Time End:	3.35pm
Location:	SIS Meeting Room 4-7
Recorded by:	Heng Kok Chin
Vetted By:	Peh Zhan Hao

Attendees:	
Dr Kam Tin Seong	Associate Professor of Information Systems (Practice)
Heng Kok Chin	Undergraduate, Singapore Management University
Peh Zhan Hao	Undergraduate, Singapore Management University
Tan Yong Kiong, Alson	Undergraduate, Singapore Management University

Agenda

1. **Prelim vs. 'O' Levels by Class by Batch**
2. **Multivariate Analysis (Sec 2 vs. 'O' Levels)**
3. **Time-Series Analysis**
4. **Moving Forward**
5. **Team To-Do**

No.	Discussion:	Action by:	Deadline:
1	**Prelim vs. 'O' Levels by Class by Batch**		
	• Go to Table > Stack > Stack Columns (Select the 'select' check box and include the variables) • Need to perform stacking and/or splitting to change the data table into a format such that it is easier to visualize in the manner that we want • Prof. Kam will get back to us on how to obtain the visualization that we need		
2	**Multivariate Analysis (Sec 2 vs. 'O' Levels)**		
	• L1R4 & L1R5 should not be there. It is the response and dependent variable. It should not be there. 'Fit Model' (Multiple Regression Analysis) • L1R5 is the 'Y' • OLVL_Batch is the 'By' • Add in the Secondary 2 subjects at the bottom • Separate the L1R4 and L1R5 data • Cannot have missing data • Right-click -> Add column 'VIF' to check if there is multicollinearity. Variance-Inflated Factor, indicator to detect independent variables correlation with each other • Value greater than 8, confirm is a multicollinearity (remove the variable)		

	• Go to 'Estimates' -> 'Correlation of Estimates' • Click on 'Overall' -> Remove • Maths, Science and English are good predictor (< 0.05 = significant) which makes sense • For L1R4, Maths, Science and English are better predictors • For L1R5, Maths and Science are better predictors • Can add in Secondary 3 and Secondary 4 results for more analysis		
3	**Time-Series Analysis**		
	• Take the 'O' Levels results as dependent • The various subject results and year as independent • Cannot use Panel Data Analysis as the 'O' Levels grade doesn't change in each row (year) of the table • Need to rethink about which analysis to do for the time-series		
4	**Moving Forward**		
	• We want to input Secondary 2 scores in order to predict 'O' Levels score • For R, read the input file, then clean, then use the variables to construct the model • The challenge for R is that there is no user interface • 2 possible approaches: - R Shiny framework - R Markdown framework • Difference between Shiny and Markdown - Markdown is for generating report like PDF (not strong in user interface design) - Shiny has better user interface design • ShowMeShiny (https://www.showmeshiny.com/) – Good website for references to examples that could be useful - Can see their code - Can start with look at the examples building correlation or regression model • Shiny R (https://shiny.rstudio.com/) – Main website for Shiny R - Must do thing is tutorial (teach yourself shiny) - Watch the entire tutorial • Prof. Kam can advise us on which are the libraries to use		
5	**Team To-Do**		
	• Generate graphs using JMP Pro Fit Model • Learn the basics of R Shiny • Look through examples and see if there is any of them that we can use • Think about how the final R interface should look like	Edufy	4th March 2017