Singapore Management University
ANLY482 Analytics Practicum

Supervisor Minutes 8 as on 13th March 2017

Time Start:	2.30pm
Time End:	3.35pm
Location:	SIS Meeting Room 4-7
Recorded by:	Heng Kok Chin
Vetted By:	Tan Yong Kiong, Alson

Attendees:	
Prof. 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. **JMP**
2. **R Model**
3. **Conference Paper**

No.	Discussion:	Action by:	Deadline:
1	**JMP (for the Mixed Model)**		
	• Help -> Book -> Fit Linear Model -> Page 345 • Take a look at the example Cholesterol.jmp (it is under Help -> Sample Data) and its data structure • Data structure: organized in time series of exams • Y is the 'O' Levels score • Need to stack and reorganize the data before running the analysis		
2	**R Model**		
	'O' Levels Estimator • Prof. Kam asked why are there sliders for each of the subjects? • It should be based on the workflow logic, whether does it make sense? • Teachers should be able to import the students or select them straight away, no need to key in • For managing data, can check out the R libraries sqlr & dplyr • It would be better if we had a diagram to explain the logic behind the 'O' Levels Estimator • We can look at the Monte Carlo Simulation • Simulate many times to get an envelope (what is the 95% interval) • So that we are able to say that with 95% confidence, you will get blablabla score • There should be a built-in option to run simulation for R • Monte Carlo Simulation in R (can look at https://web.stanford.edu/class/bios221/labs/simulation/Lab_3_simulation.html and maybe Lab 1 & Lab 2)		

	• For the parameters, we need to enter in the distribution. For scores in our context, normally is normal distribution • Other parameters include mean, standard distribution, seed (if don't know, just put 123456), number of runs (1000) • Can also look at the Monte Carlo book in library <u>Analysis of Performance</u> • Prof. Kam questioned the team on what is the purpose of the average? It should be compare within your batch, not so much with historical data • The team should use boxplot instead of bar charts, boxplot can tell more information (such as 25% quartile, median, mean), can also add in another marker to see the student's position to see the relative position • Instead of using the basic R plot, we can consider ggplot2, a R library which gives more control, or ggvis, newer but less things exposed, mouse over with tooltip capabilities • Prof. Kam suggested that the team can include CA1, SA1, CA2, SA2 using the boxplots • Round the mean score (to like 1 or 2 decimal places) • The current code, do not overwrite it. Can keep it as version 1, then include into the conference paper: 1) Design considerations 2) How the dashboard evolved 3) Best practices • Other alternatives to boxplots can be histogram or bullet chart (which is more often used in business world) – there is a target		
3	**Conference Paper**		
	• Conference paper around 8 – 12 pages • Direction of the paper depends on what we want to focus on; what we want to share • If we are focusing on EDA (or feature engineering), we can make use of what we had in the interim report • For the feature engineering, we can include what have we learnt from this practicum with regards to feature engineering • If we are focusing more on the dashboard stuff, we can talk about the Monte Carlo simulation • For the dashboard, it is traditionally a view used to gain insights and for reporting purposes • We can use Monte Carlo as a use case to describe the transition from descriptive to prescriptive (Davenport, authority of analytics is writing a chapter of his book "How we move from descriptive to prescriptive" • Simulation is a prescriptive model; for example, if you take combined science, what will be the outcome of your grades etc.		