Question 1

a. The research objective is to understand how the choice of universities when respondents are given varying combination of offers (marketing/attributes) mix for full time specific (business analytics may be) master degree (how do you infer this??).

The questionnaire are designed as option cards with universities name, university ranking, course fees and duration of course. Value included in the 4 attributes are:

- 3 universities were considered SMU, NTU and NUS;
- university ranking listed as Tier 1 or Tier 2 where Tier 1 is within top 50 and Tier 2 is outside of Top 50 in Asia;
- Course fees from $40K, $50K and $60K were considered and
- Course duration of 1 year and 1.5years

Target respondents are candidates who can be prospective students in Singapore. Realistically they should be from Singapore (from a specific undergraduate class may be).

Each student will receive option card pack via email consist of 8 (why??) option cards. Based on the scenario given in each card, respondents are to select the university of choice or in other words rank order the options.

b. U_{NTU} = -0.4547 -0.0572Fees - 1.7723Tier - 0.9081Duration
U_{SMU} = -0.7102 - 0.0572Fees - 1.7723Tier - 0.9081Duration
U_{NUS} = -0.0572Fees - 1.7723Tier - 0.9081Duration

The limitation of this output is that it doesn't explicitly give the utility at each attribute level (see slide 68 of Discrete Choice Modeling Note) , instead gives as coefficient of dummy variable regression.

It is possible to comment on the order of importance of the variables looking at the magnitude of the co-efficient.

Extra information:

<u>Sample calculation of market share</u>

Assume all are Tier 2 universities charging 50K for a 1.5 years course

U(NTU) = -.4547-0.0572(50)-1.7723(2)-0.9081(1.5) = -8.22145
U(NUS) = -0.0572(50)-1.7723(2)-0.9081(1.5) = -7.76675
U(SMU) = -.7102-0.0572(50)-1.7723(2)-0.9081(1.5) = -8.4795

Prob(SMU) = $e^{u(smu)} / (e^{u(ntu)} + e^{u(nus)} + e^{u(smu)})$ – Logit Model Rule

Prob (SMU)	Prob (NTU)	Prob (NUS)
23%	30%	47%

https://www.slideshare.net/MinhaHwang/conjoint-analysis-part-33-market-simulator

c. **Tier/ranking** of the university will have most impact to SMU for each unit of change. Given that SMU's rank remained unchanged the share would decrease from **23% to 12%** given that both NUS and NTU were uplifted to Tier 1.

U(NTU) = -.4547-0.0572(50)-1.7723(1)-0.9081(1.5) = -6.4495
U(NUS) = -0.0572(50)-1.7723(1)-0.9081(1.5) = -5.9948
U(SMU) = -.7102-0.0572(40)-1.7723(2)-0.9081(1) = -7.4509

Prob(SMU) = $e^{u(smu)} / (e^{u(ntu)} + e^{u(nus)} + e^{u(smu)})$ = 12.5%

Using $\pi_a = \dfrac{e^{uA}}{e^{uA} + e^{U_B} + e^{U_C}}$,…Share of NTU, SMU and NUS is 34% , 12.5% and 53.5%.

Question 2

a. Any two of the below:

(1) the patient does not experience recurrence before the study ends;
(2) the patient is lost to follow-up during the study period;
(3) the patient person withdraws from the study because of death (if death is not the event of interest) or some other reason (e.g., car accident)

b. The KM curve for the group with hormone therapy is consistently higher than the KM curve for the other group.

This indicates that the treatment group (i.e., with hormone therapy) has better survival prognosis than the other group.

c. The chi-square statistic from the log-rank test is statistically significant at $p < 0.01$.

This suggests that the two groups are statistically different with respect to breast cancer recurrence.

d. The model:

$$h(t,X) = h_0(t)\exp(-0.324\, hormoneYes + 0.007\, size + 0.644\, grade2 + 0.788\, grade3 + 0.049\, nodes - 0.002\, prog\_recp)$$

AIC was used as the criteria of backward stepwise selection. The model building process started from the full model in which all predictor variables were included. Then one variable was removed at each step if such removal results in the minimal AIC of the particular model. Overall three variables are dropped because removing any other variables from the model will increase AIC (how??) .

The three variables removed were all insignificant at the level of 0.05 in the full model. In the final model, except for the variable size, all the other variables are significant because their p values are below 0.05. The size variable is marginally significant with p = 0.06.

Below questions should be answered looking at the exp(coef) column of the output. Please read the theory carefully.

e. hormones: the group without having hormone therapy has 1.38 (1/0.724) (please note that output says hormoneYes) times the hazard of the treatment group while holding other variables constant.

size: the hazard increases by 1.007 times when the size of the tumor increases by 1 mm while holding other variables constant.

grade2: patients with grade2 tumor has 1.904 time the hazard of the patients with grade1 tumor while holding other variables constant.

grade3: patients with grade3 tumor has 2.199 time the hazard of the patients with grade1 tumor while holding other variables constant.

nodes: the hazard increases by 1.05 times when the number of nodes involved increases by 1 while holding other variables constant.

prog_recp: the hazard increases by 1.002 (1/0.998) (please note that this variable has negative coefficient) when the number of progesterone receptors decreases by 1 while holding other variables constant.

f. The most important assumption is the proportional hazard assumption.

It means that the hazard functions for any two individuals at any point in time are proportional. In other words, if an individual has a risk of recurrence at some initial time point that is twice as high as that of another individual, then at all later times the risk of recurrence remains twice as high.