City University of Hong Kong
Department of Management Sciences
MS6221 Predictive Modeling in Marketing
Summer Term 2021
Individual Project

Due on Date: 25th Jul 2021
Total: 50 marks

1 Question

A marketing executive undertook a survey of couples' preferred dining outlets on the Valentine's Day. Thirty couples took part in the survey. Each couple was presented with two choices of cuisine. Let $Y = 0$ be the choice of Chinese cuisine and $Y = 1$ be the choice of non-Chinese cuisine. In addition, the following information on each couple were collected.

X_1	16-70	average age of the couple;
X_2	1	if the couple is married,
	0	otherwise;
X_3	1	if the couple is in a de-facto relationship,
	0	otherwise;
X_4	0-9	the number of children parented by the couple;
X_5	1-45	the number of years the couple has been in the relationship, rounded to the nearest integer;
X_6	1	if at least one partner of the couple is non-Chinese by race,
	0	otherwise;
X_7	1	if at least one partner of the couple lived overseas for more than ten years,
	0	otherwise;
X_8	2000-250000	total monthly income of the couple;
X_9	1	if at least one partner of the couple is a professional by occupation,
	0	otherwise.

The SAS program which includes the data set is

```
DATA choice;
INPUT y x1-x9;
CARDS;
1 17 0 0 0  1 1 1   2000 0
1 35 1 0 0  8 0 0  40000 1
1 28 0 1 0  5 0 0  35000 1
1 31 1 0 0  4 0 0  65000 1
0 65 1 0 3 42 0 1   2000 1
0 32 0 1 0  3 1 1  80000 1
1 23 0 0 0  2 0 0  10000 0
```

```
1 45 1 0 0 25 0 1 120000 1
1 38 1 0 2 19 0 0  90000 1
0 55 0 1 0  2 0 0 200000 0
1 20 0 0 0  2 0 0  22000 0
0 70 1 0 2 45 0 0  20000 0
1 40 1 0 1 13 0 0  75000 1
0 40 1 0 1 10 1 1  80000 1
0 37 0 1 0  3 1 1 120000 1
0 38 0 1 0  2 1 1 250000 1
1 30 1 0 9  7 0 0  60000 0
0 33 0 0 0  3 1 1 100000 1
1 22 0 0 0  4 0 0  30000 0
1 27 1 0 0  5 0 0  70000 1
0 36 0 1 1  7 1 1 150000 0
1 18 0 0 0  1 0 0   3000 0
0 35 1 0 3  8 1 1 190000 1
0 39 1 0 2  7 1 1 170000 1
0 38 1 0 1  8 1 1 150000 1
0 38 1 0 2  8 1 1 150000 1
1 19 0 0 0  1 0 0  10000 0
1 16 0 0 0  1 0 0   2000 0
0 44 1 0 1 16 0 1  90000 0
1 34 1 0 1 13 0 0  70000 0
;
```

2 Requirements

A report consisting of an introduction, data analysis and findings. The report is limited to two pages on 12-point font, 1.5-line spacing and 1-inch margin. An appendix which contains SAS code, SAS output and references is expected. The appendix may also contain supplementary materials such as tables and figures which help explain the report. There is no page limit for the appendix.

The introduction should briefly explains the motivations and include references related to the project question. For example, in Lee et al. (2012, IME, 538-550), "*The Danish fire claim data consists of 2167 observations on fire loss claims in millions of Danish Kroner at 1985 prices between 1980 and 1990. There have been extensive studies on this data set, for example as in Embrechts et al. (1997) and McNeil (1997). It was studied in Wong and Li (2010) to illustrate model fitting for the threshold model (1).*" Each of the citations has to be included in the references.

Data analysis should include but not limited to model selection, parameter significance and model significance. It is not essential to mention each step in the model selection. For example, the statement 'Stepwise selection by eliminating one variance in each step leads to the model' is preferred to 'X_7 is first dropped because the p-value corresponding to β_7 is the highest, followed by X_6...'.

The findings are different from data analysis in the way that clear messages from the resulting model are delivered. For instance, the predicted odds of opting for non-Chinese cuisine for a coupon decreases by 18% for the age increases by one year, holding the marital status and number of children fixed. A suggestion or implication is desirable. For example, if a dining outlet has a target group of young people aged between 18 and 30, non-Chinese cuisine is likely, with probability 0.7 (see Appendix for the calculation), to boost sales.

3 Marking and Grading

The lecturer and the tutor will independently review the individual reports on the following criteria each with a weight

- Appropriate and correct statistical analysis, 50%
- Use of English, 20%
- Clear presentation of results, 30%

It is important to have a coherent analysis of the data, guided by statistical reasoning. Weak reasons or contradictory results are occasional. Selective presentation is sometimes needed. It is also important to write simple sentences with suitable use of words. Check on grammar and sentence structure is essential. Bonus point will be given to appropriate analysis using techniques introduced in the course.

4 Important Dates

- 11st Jul 2021: Submission of draft (optional: no penalty for no submission and no bonus for submission)
- 25th Jul 2021: Submission of individual project