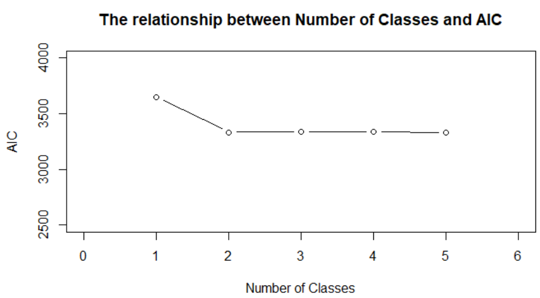
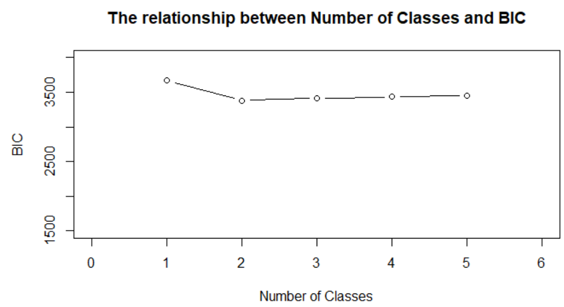
Latent Class Logit Model for Mathematical Models in Marketing
Jin Miao (JMiao18)

1. (a) Model Selection Based on AIC and BIC

number of classe	rate of correct predictions	likelihood	N	k	AIC	BIC
1	0.5839416	3612.147	1918	18	3648.147	3671.238275
2	0.620438	3259.96	1918	37	3333.96	3381.425398
3	0.5912409	3228.94	1918	56	3340.94	3412.779522
4	0.5547445	3187.643	1918	75	3337.643	3433.856645
5	0.6934307	3144.594	1918	94	3332.594	3453.181769





Based on BIC, the model with 2 latent classes is the best.

(b) Preferences from different segments are quite different from each other. There is sufficient evidence to support that different groups have heterogeneous preferences.

2. Interpret the Results

The model has 2 Latent Classes, where the proportions are 54.0% and 46.0%.

- The first segment, taking up 54.0% of the total population, are very sensitive to prices. The estimates for all four parameters related to prices are significantly lower than 0. The effect sizes of these price parameters are also much larger than other parameters.
- The second segment, taking up 46.0% of the total population, are characterized by their preferences for iPad. In addition, this group typically likes large size of Hard Drive and RAM. The parameters for iPad, Hard Drive and RAM are significantly different from zero.

3. Posterior Probabilities of Segment Membership

Most of the individuals have quite clear membership. Only 13 out of the 137 individuals have two segments whose probabilities are more than 20%. The membership probabilities are quite well separated.

ID	Seg 1	Seg 2
1	0.00	1.00
2	0.00	1.00
3	1.00	0.00
4	1.00	0.00
5	1.00	0.00
6	0.97	0.03
7	0.00	1.00
8	0.00	1.00
9	1.00	0.00
10	0.00	1.00
11	1.00	0.00
12	0.01	0.99
13	1.00	0.00
14	1.00	0.00
15	1.00	0.00
16	0.03	0.97
17	1.00	0.00
18	1.00	0.00
19	0.00	1.00
20	1.00	0.00
21	0.99	0.01
22	0.05	0.95
23	1.00	0.00
24	1.00	0.00
25	1.00	0.00
26	1.00	0.00
27	1.00	0.00
28	0.00	1.00
29	1.00	0.00
30	1.00	0.00
31	0.00	1.00
32	0.00	1.00
33	**0.75**	**0.25**

34	1.00	0.00
35	**0.33**	**0.67**
36	**0.71**	**0.29**
37	**0.33**	**0.67**
38	0.99	0.01
39	0.00	1.00
40	1.00	0.00
41	1.00	0.00
42	1.00	0.00
43	0.99	0.01
44	1.00	0.00
45	1.00	0.00
46	1.00	0.00
47	**0.36**	**0.64**
48	0.99	0.01
49	1.00	0.00
50	1.00	0.00
51	1.00	0.00
52	1.00	0.00
53	0.00	1.00
54	0.00	1.00
55	0.00	1.00
56	0.00	1.00
57	1.00	0.00
58	1.00	0.00
59	1.00	0.00
60	1.00	0.00
61	**0.24**	**0.76**
62	0.00	1.00
63	0.00	1.00
64	**0.50**	**0.50**
65	1.00	0.00
66	0.97	0.03
67	**0.31**	**0.69**

68	1.00	0.00
69	1.00	0.00
70	1.00	0.00
71	1.00	0.00
72	1.00	0.00
73	1.00	0.00
74	0.00	1.00
75	1.00	0.00
76	0.88	0.12
77	0.00	1.00
78	1.00	0.00
79	**0.54**	**0.46**
80	0.00	1.00
81	0.97	0.03
82	0.00	1.00
83	1.00	0.00
84	1.00	0.00
85	0.92	0.08
86	0.95	0.05
87	0.08	0.92
88	1.00	0.00
89	0.03	0.97
90	0.00	1.00
91	1.00	0.00
92	0.00	1.00
93	1.00	0.00
94	0.00	1.00
95	0.00	1.00
96	0.00	1.00
97	0.00	1.00
98	0.99	0.01
99	0.98	0.02
100	**0.33**	**0.67**
101	1.00	0.00

Latent Class Logit Model for Mathematical Models in Marketing
Jin Miao (JMiao18)

102	0.00	1.00
103	0.00	1.00
104	0.00	1.00
105	0.00	1.00
106	0.00	1.00
107	0.00	1.00
108	**0.24**	**0.76**
109	0.94	0.06
110	0.00	1.00
111	0.94	0.06
112	0.00	1.00
113	0.00	1.00
114	1.00	0.00

115	0.00	1.00
116	0.13	0.87
117	1.00	0.00
118	1.00	0.00
119	0.00	1.00
120	0.15	0.85
121	0.00	1.00
122	0.00	1.00
123	**0.76**	**0.24**
124	0.00	1.00
125	0.00	1.00
126	0.00	1.00
127	1.00	0.00

128	0.00	1.00
129	0.00	1.00
130	0.00	1.00
131	0.98	0.02
132	0.00	1.00
133	**0.51**	**0.49**
134	0.00	1.00
135	0.00	1.00
136	1.00	0.00
137	0.16	0.84

4. Cross Validation with Hold-out Data

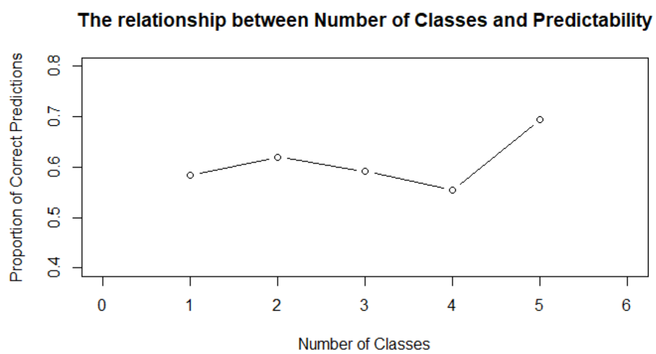
Step 1: randomly select one out of 15 choices from every individual as the holdout set.

Step 2: calculate the probability for each individual i, to choose Option j in Segment s with the Multinomial Logit Model.

Step 3: use Posterior Probabilities of Segment Membership as weights, calculate the Posterior Probability for each Option.

Step 4: for each of the 137 individuals, choose the option with the highest probability as the prediction.

Step 5: calculate the proportion of correct prediction in the holdout set.



Interpretation: With Latent Class Logit Model, we can correctly predict about 65% of consumer choices. This predictability is almost the same for 2, 3, 4 and 5 classes across many experiments.

5. Prediction for new observations

Without knowing individual segment membership probabilities, the overall membership proportions in the training set can be used as the benchmark. The underlying assumption is that the training set should have the same distribution as the test set. Thus, the distribution of membership in the training set is a good predictor for a new observation in the test set. In our example, the probabilities for a new observation to belong to each of the two segments are 54.0% and 46.0%.

Step 1: calculate the probability for each individual i, to choose Option j in Segment s with the Multinomial Logit Model.

Step 2: use Overall Segment Proportions as weights, calculate the Posterior Probability.

Step 3: choose the option with the highest probability as the prediction.

6. Customer Demographics

Customer demographics can be used to predict segment membership. First in the training set, regress multiple segment shares over demographic variables to estimate the parameters. Then in the test phase, predict segment membership from customer demographics and then use segment membership to predict consumer choice. To summarize, demographic variables are partially underlying mechanisms for latent classes and used to construct the proxy for latent classes.