HW 3 Multinomial Choices

Yaxuan Jiao

3/21/2019

Exercise 1 Data Description

- Average and dispersion in product characteristic

I made a table that includes the average of product prices, the range, variance, standard deviation and interquartile range (IQR=Q3-Q1).

According to the table below, PFl_Tub has the highest average price among ten products. In general, tub appears to be more expensive than stick, and the Fleischmann's cheese produce the products with the highest average price in both tub and stick category. Besides, the price of ten products are relative stable, with small variance, standard deviation and price range.

	Average	Range-min	Range-max	Variance	SD	IQR
PPk_Stk	0.5184362	0.19	0.67	0.0226554865	0.15051740	0.12
PBB_Stk	0.5432103	0.19	1.01	0.0144797566	0.12033186	0.11
PFl_Stk	1.0150201	0.95	1.16	0.0018399974	0.04289519	0.09
PHse_Stk	0.4371477	0.19	0.64	0.0141208621	0.11883123	0.28
PGen_Stk	0.3452819	0.25	0.55	0.0012366513	0.03516605	0.03
PImp_Stk	0.7807785	0.33	2.30	0.0131437214	0.11464607	0.16
PSS_Tub	0.8250895	0.50	0.98	0.0037468593	0.06121159	0.05
PPk_Tub	1.0774094	0.98	1.24	0.0008836431	0.02972613	0.02
PFl_Tub	1.1893758	0.69	1.47	0.0001975293	0.01405451	0.00
PHse_Tub	0.5686734	0.33	1.27	0.0052497277	0.07245500	0.03

- Market share, and market share by product characteristics
  - Market share by products

	product sales	market share
PPk_Stk	786.10	0.316273119
PBB_Stk	305.81	0.123037123
PFl_Stk	245.65	0.098832835
Phse_Stk	231.46	0.093123745
PGen_Stk	111.16	0.044723216
PImp_Stk	55.83	0.022462191
PSS_Tub	248.06	0.099802455
PPK_Tub	217.48	0.087499145
PFl_Tub	267.25	0.107523204
PHse_Tub	15.71	0.006320634
M_Industry	2485.51	1.000000000

- Market share by brands

	brand sales	market share
PPk	1003.58	0.40377226
PBB	305.81	0.12303712
PFl	512.90	0.20635604
PHse	247.17	0.09944438
PGen	111.16	0.04472322
PImp	55.83	0.02246219
PSS	248.06	0.09980246
M_Industry	2485.51	1.00000000

- Market share by stick and tub

	type sales	market share
Stick	1736.01	0.6984522
Tub	748.50	0.3011454
M_Industry	2485.51	1.0000000

➢ Mapping between observed attributes and choices

- Frequency of buying each choice for each income level

	PPk_Stk	PBB_Stk	PFl_Stk	PHse_Stk	PGen_Stk	PImp_Stk	PSS_Tub	PPK_Tub	PFl_Tub	PHse_Tub
2.5	19	4	0	2	6	0	16	1	2	0
7.5	117	54	13	34	19	2	27	6	22	1
12.5	196	106	41	44	23	9	40	8	25	3
17.5	318	100	27	111	21	5	54	19	20	2
22.5	292	123	34	154	123	2	41	36	30	8
27.5	195	94	9	67	18	6	24	25	34	4
32.5	209	84	28	64	54	4	49	19	33	5
37.5	132	34	17	29	23	1	15	14	9	5
42.5	125	33	33	23	6	20	27	21	14	1
47.5	83	22	23	16	7	17	6	9	2	3
55	201	201	201	201	201	201	201	201	201	201
67.5	19	4	1	8	6	2	7	3	0	1
87.5	9	10	3	1	0	1	1	0	12	0
130	5	1	3	8	2	2	0	0	5	0

- Frequency of buying each choice for each family size

	PPk_Stk	PBB_Stk	PFl_Stk	PHse_Stk	PGen_Stk	PImp_Stk	PSS_Tub	PPK_Tub	PFl_Tub	PHse_Tub
1	148	49	38	23	10	7	25	18	34	0
2	474	212	123	154	55	26	117	52	112	3
3	400	165	29	119	60	11	77	46	48	3
4	502	195	33	179	127	7	80	76	20	9
5	160	53	20	72	33	23	8	2	11	13
6	76	22	0	33	24	0	12	9	0	5
7	1	1	0	8	2	0	0	0	0	0
8	5	2	0	5	4	0	0	0	0	0

- Frequency of buying each choice for college and non-college

	PPk_Stk	PBB_Stk	PFl_Stk	PHse_Stk	PGen_Stk	PImp_Stk	PSS_Tub	PPK_Tub	PFl_Tub	PHse_Tub
college	561	219	110	174	86	32	103	52	62	15
non-college	1205	480	133	419	229	42	216	151	163	18

- Frequency of buying each choice for white-collar and non-white-collar

	PPk_Stk	PBB_Stk	PFl_Stk	PHse_Stk	PGen_Stk	PImp_Stk	PSS_Tub	PPK_Tub	PFl_Tub	PHse_Tub
white-collar	1007	380	132	351	225	42	184	116	130	31
not white_collar	759	319	111	242	90	32	135	87	95	2

- Frequency of buying each choice for retired and not retired

	PPk_Stk	PBB_Stk	PFl_Stk	PHse_Stk	PGen_Stk	PImp_Stk	PSS_Tub	PPK_Tub	PFl_Tub	PHse_Tub
retired	352	168	129	91	46	28	47	20	81	4
not-retired	1414	531	114	502	269	46	272	183	144	29

Exercise 2 First Model

- Apply conditional logit model to capture the choice's characteristics---the effect of price on demand (since price varies over alternatives)
- Write the likelihood and optimize the model (see code)
- Interpret the coefficient on price
  - optimized coefficients on price: $(\beta, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9)$

```
[1] -6.5744069 -0.9284425  1.8105589 -1.5683915 -2.9928648 -1.0811750 -0.0473792  1.9833614  1.3667610 -4.0970681
```

I think this result may vary with the initial values being assigned to beta and alpha, while in terms of this result, I think the negative sign of the estimated coefficient indicates that holding other product characteristics constant, consumers will be less likely to buy cheese products for an increase in price.

Exercise 3 Second Model

- Apply multinomial logit model to capture the chooser's characteristics---the effect of family income on demand (since family income do not vary over alternatives)
- Write the likelihood and optimize the model (see code)
- Interpret the coefficient on family income
  - optimized coefficient on price:

    $(a_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9, \alpha_{10}; b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10})$

```
 [1] -0.33484595 -0.40301779  0.11254933 -0.11103723 -0.09669469  0.27212609 -0.59622018  0.64979228
 [9] -0.93316507 -0.32059820  0.52318214  0.47499303  0.36289966  0.53117864 -0.90685924  0.49428696
[17]  0.16621994  0.21013966 -0.07102476 -0.30433955
```

This result may vary with the initial values being assigned to beta and alpha, while in terms of this result, for example, the positive sign of b_1 indicates that people are slightly more willing to buy product 1 when their income increase. And for the negative sign of b_5, I think it means that consumers will be less likely to buy this product for an increase in their family income.

Exercise 4 Marginal Effects

- Compute and interpret the marginal effects for the first (conditional logit) model

-1.25230167	0.272940786	0.186743248	0.302065144	0.126273105	0.0520701943	0.105840536	0.154922266	0.0389151021	0.0125312857
0.27294079	-0.724769620	0.086562521	0.138165943	0.059798758	0.0236509132	0.048527743	0.071878788	0.0181193145	0.0051248532
0.18674325	0.086562521	-0.522406287	0.087642757	0.041510825	0.0167747509	0.034332875	0.052048417	0.0129512122	0.0038396807
0.30206514	0.138165943	0.087642757	-0.752221081	0.057883409	0.0258386694	0.049483996	0.068885484	0.0173743741	0.0048813040
0.12627310	0.059798758	0.041510825	0.057883409	-0.366192604	0.0107673726	0.023266028	0.034861125	0.0089063937	0.0029255875
0.05207019	0.023650913	0.016774751	0.025838669	0.010767373	-0.1552569193	0.009028898	0.012958170	0.0032768315	0.0008911187
0.10584054	0.048527743	0.034332875	0.049483996	0.023266028	0.0090288983	-0.310670450	0.030406232	0.0074218131	0.0023623280
0.15492227	0.071878788	0.052048417	0.068885484	0.034861125	0.0129581704	0.030406232	-0.441594139	0.0118989139	0.0037347428
0.03891510	0.018119314	0.012951212	0.017374374	0.008906394	0.0032768315	0.007421813	0.011898914	-0.1198050097	0.0009410547
0.01253129	0.005124853	0.003839681	0.004881304	0.002925587	0.0008911187	0.002362328	0.003734743	0.0009410547	-0.0372319552

As we can perceive from the table, marginal effects on the diagonal is negative, it's consistent with the real life intuition that when the price of the chosen product 1 increase, it's less likely for the person to still choose that product 1; however, when the price of other products increase, it will only increase the probability of choosing product 1 a little bit.

- Compute and interpret the marginal effects for the second (multinomial logit) model

0.04438334	0.06774356	0.008235064	-0.04096157	-0.04422409	-0.02606351	-0.03812152	0.05213143	-0.03652647	0.01340377

According to the results, an increase in family income will have slight effects on the product demand, and I noticed that for products with relatively lower average prices, such as product

4, 5, 6, 7, and 10, they tend to encounter a negative marginal effect from the increase in family income. I think it may because that when people are having higher income, they are more eager to live a higher quality life, so that they may want to buy more expensive products to express their willingness.

Exercise 5 IIA (mixed logit)

- Apply mixed logit model to capture the effect of price and family income on demand
    - Write the likelihood and optimize the model (see code)

      β^f: $(\beta, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9; a_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9, \alpha_{10};$

      $b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10})$(#30)

```
-1.1567912  0.4034614 -0.5625225 -0.6900884 -0.6111647 -0.3756687  0.7226019  0.6933815 -0.3558034  1.7662169 -2.4480849
-0.4400124  0.6841820 -0.8981270 -1.2085059  0.7696783 -1.7617112 -1.7008313 -0.4600140 -0.4798999 -0.8587021  0.5328810
-2.2511447  0.5572762 -1.3739652  0.3890164 -1.2559074 -2.1080558 -0.2466219  0.8046302
```

- Consider an alternative specification by removing data from one choice
    - Write the likelihood and optimize the model (see code)

      β^r: $(\beta, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8; a_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9;$

      $b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9)$(#27)

```
-1.18329116 -1.91046912  0.60753958 -0.01003206 -2.38587740 -0.11738874  1.24997081  1.02316436  0.20579469 -0.15035053
 0.24650469 -0.88196192  1.46644633 -1.12303606  0.42076482  0.91092271 -0.64684237  0.95713140  0.65852651  0.14849119
 0.15826666  0.13738560  0.25704058 -0.03861813 -0.05045913 -1.15337460  0.12661689
```

- Compute the test statistics:

```
> l_betaf <- mixed_logit(mixed_logit_optim)
> l_betar <- nmixed_logit(nmixed_logit_optim)
> MTT <- -2*(l_betaf-l_betar)
> print(MTT)
[1] 56432.06
```

- Conclude on IIA

```
        Chi-squared test for given probabilities

data:  a_betar
X-squared = 14.819, df = 26, p-value = 0.9605
```

we reject the null hypothesis with strong evidence, and able to conclude that the property of independence of irrelevant alternatives holds in the model.