Mingpei Li

ECON613

HW2

Mar 24, 2021

Exercise 1 Data Description

1 Average and dispersion in product characteristics (price)

Average:

PPk_Stk	PBB_Stk	PFl_Stk	PHse_Stk	PGen_Stk	PImp_Stk	PSS_Tub	PPk_Tub	PFl_Tub	PHse_Tub
0.5184362	0.5432103	1.0150201	0.4371477	0.3452819	0.7807785	0.8250895	1.0774094	1.1893758	0.5686734

Dispersion (variance):

PPk_Stk	PBB_Stk	PFl_Stk	PHse_Stk	PGen_Stk	PImp_Stk	PSS_Tub	PPk_Tub	PFl_Tub	PHse_Tub
0.0226554865	0.0144797566	0.0018399974	0.0141208621	0.0012366513	0.0131437214	0.0037468593	0.0008836431	0.0001975293	0.0052497277

Dispersion (standard deviation):

PPk_Stk	PBB_Stk	PFl_Stk	PHse_Stk	PGen_Stk	PImp_Stk	PSS_Tub	PPk_Tub	PFl_Tub	PHse_Tub
0.15051740	0.12033186	0.04289519	0.11883123	0.03516605	0.11464607	0.06121159	0.02972613	0.01405451	0.07245500

2.

Market share (choice frequency) (measured in proportion):

PBB_Stk	PFl_Stk	PFl_Tub	PGen_Stk	PHse_Stk	PHse_Tub	PImp_Stk	PPk_Stk	PPk_Tub	PSS_Tub
0.15637584	0.05436242	0.05033557	0.07046980	0.13266219	0.00738255	0.01655481	0.39507830	0.04541387	0.07136465

Market share by product characteristics:

Market share (over average):

PBB_Stk	PFl_Stk	PFl_Tub	PHse_Stk	PHse_Tub	PImp_Stk	PPk_Stk	PPk_Tub	PSS_Tub
0.119058397	0.110004527	0.101856043	0.051607062	0.008148483	0.033046627	0.356722499	0.091896786	0.127659574

Market share (below average):

PBB_Stk	PGen_Stk	PHse_Stk	PHse_Tub	PImp_Stk	PPk_Stk	PSS_Tub
0.1928350287	0.1393188854	0.2118531623	0.0066342326	0.0004422822	0.4325519682	0.0163644405

3. mapping between observed attributes and choices.

Merged dataset:

	hhid	choice	PPk_Stk	PBB_Stk	PFl_Stk	PHse_Stk	PGen_Stk	PImp_Stk	PSS_Tub	PPk_Tub	PFl_Tub	PHse_Tub	Income	Fs3_4	Fs5.	Fam_Size	college	whtcollar	retired
1	2100016	1	0.66	0.67	1.09	0.57	0.36	0.93	0.85	1.09	1.19	0.33	32.5	0	0	2	1	0	1
2	2100016	1	0.63	0.67	0.99	0.57	0.36	1.03	0.85	1.09	1.19	0.37	32.5	0	0	2	1	0	1
3	2100016	1	0.29	0.50	0.99	0.57	0.36	0.69	0.79	1.09	1.19	0.59	32.5	0	0	2	1	0	1
4	2100016	1	0.62	0.61	0.99	0.57	0.36	0.75	0.85	1.09	1.19	0.59	32.5	0	0	2	1	0	1
5	2100016	1	0.50	0.58	0.99	0.45	0.33	0.72	0.85	1.07	1.19	0.59	32.5	0	0	2	1	0	1
6	2100016	4	0.58	0.45	0.99	0.45	0.33	0.72	0.85	1.07	1.19	0.59	32.5	0	0	2	1	0	1
7	2100016	1	0.29	0.51	0.99	0.29	0.33	0.72	0.85	1.07	1.19	0.59	32.5	0	0	2	1	0	1
8	2100024	1	0.66	0.45	1.08	0.57	0.36	0.93	0.85	1.09	1.19	0.33	17.5	1	0	3	1	1	1
9	2100024	4	0.66	0.59	1.08	0.57	0.36	0.93	0.85	1.09	1.34	0.33	17.5	1	0	3	1	1	1
10	2100024	1	0.66	0.67	1.09	0.57	0.36	0.93	0.85	1.09	1.19	0.33	17.5	1	0	3	1	1	1
11	2100024	4	0.63	0.59	1.08	0.57	0.36	0.93	0.85	1.09	1.19	0.59	17.5	1	0	3	1	1	1
12	2100024	8	0.63	0.59	0.99	0.57	0.36	0.88	0.85	1.09	1.19	0.59	17.5	1	0	3	1	1	1
13	2100024	4	0.62	0.61	0.99	0.49	0.33	0.75	0.85	1.09	1.19	0.59	17.5	1	0	3	1	1	1
14	2100024	3	0.58	0.45	0.99	0.45	0.33	0.72	0.65	1.07	1.19	0.59	17.5	1	0	3	1	1	1
15	2100024	1	0.58	0.58	0.99	0.29	0.34	0.72	0.85	1.07	1.19	0.59	17.5	1	0	3	1	1	1
16	2100024	1	0.29	0.50	0.99	0.33	0.33	0.72	0.85	1.07	1.19	0.59	17.5	1	0	3	1	1	1
17	2100024	1	0.39	0.58	0.99	0.29	0.33	0.69	0.79	1.09	1.19	0.56	17.5	1	0	3	1	1	1
18	2100024	1	0.19	0.58	0.99	0.29	0.34	0.69	0.79	1.09	1.19	0.56	17.5	1	0	3	1	1	1
19	2100495	1	0.25	0.61	0.99	0.45	0.33	0.75	0.85	1.09	1.19	0.59	37.5	0	0	2	0	0	1
20	2100495	1	0.58	0.61	0.99	0.45	0.34	0.75	0.85	1.07	1.19	0.59	37.5	0	0	2	0	0	1

Mean of individual characteristics by choices:

*	choice	Income_mean	mean_Fs3_4	mean_Fs5.	mean_Fam_Size	mean_college	mean_whtcollar	mean_retired
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	26.7	0.511	0.137	3.17	0.318	0.570	0.199
2	2	26.1	0.515	0.112	3.10	0.313	0.544	0.240
3	3	30.7	0.255	0.0823	2.48	0.453	0.543	0.531
4	4	27.6	0.503	0.199	3.47	0.293	0.592	0.153
5	5	26.4	0.594	0.2	3.69	0.273	0.714	0.146
6	6	39.2	0.243	0.311	3.18	0.432	0.568	0.378
7	7	25.3	0.492	0.0627	2.89	0.323	0.577	0.147
8	8	34.2	0.601	0.0542	3.09	0.256	0.571	0.0985
9	9	31.9	0.302	0.0489	2.39	0.276	0.578	0.36
10	10	29.5	0.364	0.545	4.42	0.455	0.939	0.121

Exercise 2 First Model

1. Model specification.

Since price is the product's characteristic, I am going to use conditional logit model.

Individual utility is given by:

$$V_{ij} = \beta_0 + \beta_1 * price_{ij}$$

Probability is given by:

$$p_{ij} = \frac{e^{V_{ij}}}{\sum_{l=1}^{10} e^{V_{il}}} \quad , \qquad j = 1,2, \dots ,10$$

2.

The likelihood is given by:

$$L = \sum_{i=1}^{n} p_i$$

p_i is the probability of the choice made in the i-th observation.

n is the total number of observations.

Result from the optimization (the first 9 parameters is the intercept, setting the first choice as reference. The 10th parameter is the coefficient of price):

```
[1]  -0.9600677   1.3222215  -1.7279904  -2.9213679 -26.3042296   0.2638343   1.4965229   2.3957733  -3.8985959  -6.7208260
```

The coefficient on price is: -6.7208260

3.

This means if a product's price increases, the probability of choosing this product will decrease.

Exercise 3 Second Model

1. model specification.

Since family income is individuals' characteristic, I am going to use multinomial logit model.

Individual utility is given by:

$$V_{ij} = \beta_0 + \beta_1 * Income_j + \beta_2 * Fam\_Size_j + \beta_3 * college_j + \beta_4 * whtcollar_j + \beta_5 * retired_j$$

Where $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ contain 7 parameters in correspondence of product 2-10 respectively.

Probability is given by:

$$p_{ij} = \frac{e^{V_{ij}}}{\sum_{l=1}^{10} e^{V_{il}}} \quad , \qquad j = 1,2, \dots ,10$$

2.

The likelihood is given by:

$$L = \sum_{i=1}^{n} p_i$$

p_i is the probability of the choice made in the i-th observation.

n is the total number of observations.

The result of the optimization is:

```
 [1]   6.53095987   0.89949631  -3.34558420   8.46866950  -4.15368319   6.90813561  -5.09616139  -7.60120684 -14.31199393   1.29235249  -3.66775191
[12]  -1.44677227  -6.81908042  -9.20008164  -5.21845280  28.85858513  -6.86174347 -36.48950900  10.47373301  -9.08459514   0.12157378  -4.69626675
[23]  -3.90655594   0.30835738  -2.29753349  -1.90397441 -17.94714233  -4.82380383   9.84300810   6.14704674   1.06667181   2.92812188  -3.76351387
[34]   1.75247438  -1.40468952  -2.27559122   5.70707959  -0.30017522  -5.12142345   5.30919558  -8.52440240  -3.80626797   1.29216746  -0.08486359
[45]  -8.87566606  -0.28281000  -0.12078170   6.02374813  -6.50700354   6.96784821   7.45890278  -7.24553729  -4.60959310 -13.47756663
```

Here is the position of each coefficient in the vector:

Intercept: 1-9

Coefficients of Income: 10-18

Coefficients of Fam_Size: 19-27

Coefficients of college:28-36

Coefficients of whtcollar:37-45

Coefficients of retired:46-54

Within each group, coefficients are ordered: product 2 to product 10

Coefficients of income are (ordered: product2 to product10):

1.29235249 -3.66775191 -1.44677227 -6.81908042 -9.20008164 -5.21845280 28.85858513 -6.86174347 -36.48950900

3.

If the coefficient on product j is positive, then if the individual's family income increases, then the probability of this individuals to choose product j increases.

If the coefficient on product j is negative, then if the individual's family income increases, then the probability of this individuals to choose product j decreases.

Exercise 4 Marginal Effects

1. first model.

Here are the results of the codes:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	-1.301210e+00	3.068326e-01	1.261713e-01	3.080682e-01	1.630470e-01	6.729329e-13	1.598719e-01	1.038508e-01	1.157786e-01	1.758949e-02
[2,]	3.068326e-01	-7.585947e-01	5.749337e-02	1.392508e-01	7.589836e-02	3.004624e-13	7.198955e-02	4.717455e-02	5.286730e-02	7.088163e-03
[3,]	1.261713e-01	5.749337e-02	-3.459929e-01	5.309441e-02	3.182184e-02	1.293187e-13	3.065564e-02	2.068297e-02	2.287567e-02	3.197740e-03
[4,]	3.080682e-01	1.392508e-01	5.309441e-02	-7.280172e-01	6.716731e-02	2.984872e-13	6.672522e-02	4.122569e-02	4.634566e-02	6.139892e-03
[5,]	1.630470e-01	7.589836e-02	3.182184e-02	6.716731e-02	-4.385412e-01	1.595511e-13	3.968010e-02	2.634196e-02	2.993608e-02	4.648483e-03
[6,]	6.729329e-13	3.004624e-13	1.293187e-13	2.984872e-13	1.595511e-13	-1.941309e-12	1.549790e-13	9.920610e-14	1.117711e-13	1.460003e-14
[7,]	1.598719e-01	7.198955e-02	3.065564e-02	6.672522e-02	3.968010e-02	1.549790e-13	-4.296004e-01	2.703267e-02	2.923114e-02	4.414237e-03
[8,]	1.038508e-01	4.717455e-02	2.068297e-02	4.122569e-02	2.634196e-02	9.920610e-14	2.703267e-02	-2.902171e-01	2.082067e-02	3.087809e-03
[9,]	1.157786e-01	5.286730e-02	2.287567e-02	4.634566e-02	2.993608e-02	1.117711e-13	2.923114e-02	2.082067e-02	-3.213071e-01	3.452053e-03
[10,]	1.758949e-02	7.088163e-03	3.197740e-03	6.139892e-03	4.648483e-03	1.460003e-14	4.414237e-03	3.087809e-03	3.452053e-03	-4.961786e-02

For each element in the matrix, denoted as A[i,j]:

If the price of product i increases by 1 unit, the probability for an individual to choose product j will change by A[i,j] (increase if A[i,j]>0, decrease if A[i,j] <0)

2. second model.

Here are the results of the codes:

```
3.129230e-27  2.036848e-09  1.135617e-34  7.143672e-30 -4.043650e-35  2.143746e-38  6.906165e-26 -2.036848e-09 -3.251129e-40  1.079968e-85
```

For each element in the vector, denoted as B[j] :

If the family income of an individual increase by 1 unit, the probability of this individual to choose product j will change by B[j] (increase if B[j] >0, decrease if B[j]<0)

Exercise 5 IIA

1.

Individual utility is given by:

$$V_{ij} = \beta_0 + \beta_1 * Income_j + \beta_2 * Fam_{Size\,j} + \beta_3 * college_j + \beta_4 * whtcollar_j + \beta_5 * retired_j + \beta_6 * price_{ij}$$

Where $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ contain 7 parameters in correspondence of product 2-10 respectively.

Probability is given by:

$$p_{ij} = \frac{e^{V_{ij}}}{\sum_{l=1}^{10} e^{V_{il}}} \quad , \qquad j = 1,2, \dots ,10$$

The likelihood is given by:

$$L = \sum_{i=1}^{n} p_i$$

p_i is the probability of the choice made in the i-th observation.

n is the total number of observations.

β^f is:

```
-5.2194855  4.1298923 -3.8381049  0.1709513 -8.9670676  1.2913968 -7.5703963  7.8567276 -9.7074549  5.6624221 -8.2007733  0.3837996 -2.3146662
-8.5989501 -3.5871116  3.3699079  8.5280095 -0.5618056 -7.1476931  0.8853951 -6.0765070  7.9716098 -2.2100043 -3.7825844 -6.7994267  7.9237170
-6.6721244  8.0084919 -7.3184361 -7.3677173 -7.8942499  0.2316716 -3.9960189 -9.4656621 -3.8070514  4.8423931 -9.2908655  1.3015222 -4.3948445
-5.9160737 -7.3252220 -3.4863615 -6.8987606 -7.4007572 -1.2893788 -9.2271469  4.2660313 -7.9846192  9.0060988 -7.5636447 -5.6068676  8.2617553
 8.9170624 -4.4168755 -7.5305782
```

2.

I drop the last choice here.

Individual utility is given by:

$$V_{ij} = \beta_0 + \beta_1 * Income_j + \beta_2 * Fam_{Size\,j} + \beta_3 * college_j + \beta_4 * whtcollar_j + \beta_5 * retired_j + \beta_6 * price_{ij}$$

Where $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ contain 6 parameters in correspondence of product 2-9 respectively.

Probability is given by:

$$p_{ij} = \frac{e^{V_{ij}}}{\sum_{l=1}^{10} e^{V_{il}}} \quad , \qquad j = 1,2, \dots ,10$$

The likelihood is given by:

$$L = \sum_{i=1}^{n} p_i$$

p_i is the probability of the choice made in the i-th observation.

n is the total number of observations.

β^r is:

```
 5.94320921  4.88500233  8.31948422  1.97691491  7.87835299 -0.27729185 -4.33081603  3.77858700  0.06510343 -0.06203144 -3.63108530 10.50808825
 0.25982046 -7.45133204 -1.53907653 10.54970876 -0.64415331  8.16229910  1.95486542  0.36993118  6.42843744  0.05499645  9.67269568 -0.56513041
-0.37250103 -2.86025851  2.54955368  1.76562918  1.31936661  9.61573013  1.53625471  1.85331139 -5.42800610 -8.35683875  7.00529818 -3.68013764
 8.78032513  2.03795092  9.97480961 -3.26840600  1.10253262 -1.41112074  1.51755562 -2.08862327 -5.50308472 -8.30030525  2.74596519 -0.64090137
-1.41463718
```

3.

MTT= 7688.999

4.

The MTT is extremely large in chi-square distribution.

We can reject null hypothesis with confidence level 99%.

So, IIA is violated.