BUAN 6337 Homework 5_Multinomial brand choice model_Group 12

This Homework has 3 multiple part questions which are related to each other. You are required to use SAS for answering the questions. Your submission on eLearning must include a pdf/word report which has followed the sample report instructions. You should also upload your SAS code.

The dataset used in this HW is crackers_HW5. This dataset contains store sales data of crackers at a supermarket that carries four brands of crackers. Each observation corresponds to one purchase occasion and provides data on the price, display and feature of each brand as well as which brand was chosen.

1. **OBS** : = Observation number
2. **Private, Keebler, Sunshine, Nabisco** : Indicator variables for which brand was chosen. Value of 1 indicates the brand that was chosen. Other 3 brands will be 0 in that observation.
3. **PricePrivate, PriceNabisco, PriceKeebler and PriceSunshine**: Prices that were offered by
4. each brand for that purchase occasion.
5. **DisplPrivate** : = 1 if **Private** had a store display, =0 if **Private** did not have a store display
6. **DisplKeebler** : = 1 if **Keebler** had a store display, =0 if **Keebler** did not have a store display
7. **DisplSunshin**:= 1 if **Sunshin** had a store display, =0 if **Sunshin** did not have a store display
8. **DisplNabisco**:= 1 if **Nabisco** had a store display, =0 if **Nabisco** did not have a store display
9. **FeatPrivate**: = 1 if **Private** had a store feature, =0 if **Private** did not have a store feature
10. **FeatKeebler**: = 1 if **Keebler** had a store feature, =0 if **Keebler** did not have a store feature
11. **FeatSunshin**: = 1 if **Sunshin** had a store feature, =0 if **Sunshin** did not have a store feature
12. **FeatNabisco**: = 1 if **Nabisco** had a store feature, =0 if **Nabisco** did not have a store feature

Question 1

Apply the following steps and provide a screenshot of the output in your report.

a) Use PROC SURVEYSELECT to sample the original data into training and testing data sets. Use 80% for training and 20% for testing. Use the seed= option to set random seed to a value of 100.

PROC SURVEYSELECT Results:

The SAS System

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	CRACKERS
Random Number Seed	100
Sampling Rate	0.8
Sample Size	2634
Selection Probability	0.800122
Sampling Weight	0
Output Data Set	CRACKERS_SAMPLED

b) The store manager would like to predict the choice probabilities for each brand of crackers depending on the price, display and promotion for all brands. What type of multinomial logit model would you estimate – a model with alternative-specific characteristics or with individual-specific characteristics? Write the general utility model to estimate this logit model.

   a. We need to use a model with alternative-specific characters to take into account the effects of price, display, and promotion (promotion indicated as “Feat” in the data) on a customer’s choice for each brand.

   b. Private brand is taken as reference brand and hence $\alpha_{0,\ \text{Private}=0}$

$V_j = \text{ß}Price_j + \alpha_{1,\ Private}Disp_{Private} * I_{Private} + \alpha_{2,\ Private}Feat_{Private} * I_{Private} +$

$\alpha_{0,\ Keebler}I_{Keebler} + \alpha_{1,\ Keebler}Disp_{Keebler} * I_{Keebler} + \alpha_{2,\ Keebler}Feat_{Keebler} * I_{Keebler} +$

$\alpha_{0,\ Nabisco}I_{Nabisco} + \alpha_{1,\ Nabisco}Disp_{Nabisco} * I_{Nabisco} + \alpha_{2,\ Nabisco}Feat_{Nabisco} * I_{Nabisco} +$

$\alpha_{0,\ Sunshine}I_{Sunshine} + \alpha_{1,\ Sunshine}Disp_{Sunshine} * I_{Sunshine} + \alpha_{2,\ Sunshine}Feat_{Sunshine} * I_{Sunshine}$

c) Is the data formatted as needed to estimate the above multinomial logit model using PROC LOGISTIC? If not, how should the data be formatted? Reformat the data as necessary.

   a. No, the data is not formatted correctly to estimate the above multinomial logit model with alternative-specific characteristics. Each purchase occasion needs to capture the alternative brand choices even if the brand was not selected (e.g. Observation 1 will be shown in 4 rows, one row for each brand choice available). Additionally, the brand name, feature use. and display use needs to be formatted to appear with each brand choice alternative for each of the purchase occasions.

Reformatted Data for PROC LOGISTICS Table:

VIEWTABLE: Work.Crackers_formatted

	OBS	Selection Indicator	Choice	Crackers	Price	Feat	Disp
1	1	1	0	Keebler	0.879999995	0	0
2	1	1	0	Private	0.709999979	0	0
3	1	1	0	Sunshine	0.980000019	0	0
4	1	1	1	Nabisco	1.199999928	0	0
5	2	1	0	Keebler	1.089999914	0	0
6	2	1	0	Private	0.709999979	0	0
7	2	1	0	Sunshine	0.99000001	0	0
8	2	1	1	Nabisco	0.99000001	0	0
9	3	1	0	Keebler	1.089999914	0	0
10	3	1	0	Nabisco	1.089999914	0	0
11	3	1	0	Private	0.779999912	0	0
12	3	1	1	Sunshine	0.49000001	0	1
13	4	1	0	Keebler	1.089999914	0	0
14	4	1	0	Private	0.779999912	0	0
15	4	1	0	Sunshine	1.029999971	0	0
16	4	1	1	Nabisco	0.889999986	0	0
17	5	0	0	Keebler	1.089999914	0	0
18	5	0	0	Private	0.639999986	0	0
19	5	0	0	Sunshine	1.089999914	0	0
20	5	0	1	Nabisco	1.190000057	0	0
21	6	1	0	Keebler	1.089999914	0	0
22	6	1	0	Private	0.839999914	0	0
23	6	1	0	Sunshine	0.889999986	0	0
24	6	1	1	Nabisco	1.190000057	0	0
25	7	1	0	Keebler	1.089999914	0	0
26	7	1	0	Nabisco	1.289999962	0	1
27	7	1	0	Private	0.779999912	0	0

Question 2

a) Estimate the logit model on the training sample using PROC LOGISTIC and report the estimation results (model parameters, significance).

Logit Model PROC LOGISTIC Results:

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Crackers	3	943.1545	<.0001
Price	1	163.8112	<.0001
Disp*Crackers	4	12.8714	0.0119
Feat*Crackers	4	31.0850	<.0001

Analysis of Conditional Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Crackers	Keebler	1	-0.3358	0.1394	5.8063	0.0160
Crackers	Nabisco	1	1.7053	0.1154	218.4802	<.0001
Crackers	Sunshine	1	-0.8046	0.1158	48.3031	<.0001
Crackers	Private	0	0	.	.	.
Price		1	-3.0106	0.2352	163.8112	<.0001
Disp*Crackers	Keebler	1	0.2959	0.2325	1.6201	0.2031
Disp*Crackers	Nabisco	1	0.0974	0.0865	1.2671	0.2603
Disp*Crackers	Sunshine	1	0.5275	0.1814	8.4549	0.0036
Disp*Crackers	Private	1	-0.2172	0.1717	1.6000	0.2059
Feat*Crackers	Keebler	1	0.7262	0.2838	6.5467	0.0105
Feat*Crackers	Nabisco	1	0.6672	0.1600	17.3967	<.0001
Feat*Crackers	Sunshine	1	0.7339	0.2612	7.8940	0.0050
Feat*Crackers	Private	1	-0.0200	0.2357	0.0072	0.9324

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Price	0.049	0.031	0.078

b) Reproduce your results using PROC MDC

Logit Model PROC MDC Results:

The SAS System

The MDC Procedure

Conditional Logit Estimates

Parameter Estimates						
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Parameter Label
CRACKERSKeebler	1	-0.3358	0.1394	-2.41	0.0160	
CRACKERSPrivate	0	0	0			
CRACKERSSunshine	1	-0.8046	0.1158	-6.95	<.0001	
CRACKERSNabisco	1	1.7053	0.1154	14.78	<.0001	
Price	1	-3.0106	0.2352	-12.80	<.0001	
CRACKERSKeeblerDISP	1	0.2959	0.2325	1.27	0.2031	
CRACKERSPrivateDISP	1	-0.2172	0.1717	-1.26	0.2059	
CRACKERSSunshineDISP	1	0.5275	0.1814	2.91	0.0036	
CRACKERSNabiscoDISP	1	0.0974	0.0865	1.13	0.2603	
CRACKERSKeeblerFEAT	1	0.7262	0.2838	2.56	0.0105	
CRACKERSPrivateFEAT	1	-0.0200	0.2357	-0.08	0.9324	
CRACKERSSunshineFEAT	1	0.7339	0.2612	2.81	0.0050	
CRACKERSNabiscoFEAT	1	0.6672	0.1600	4.17	<.0001	
Restrict1	1	-8.02E-11	12.4614	-0.00	1.0000*	Linear EC [1]

* Probability computed using beta distribution.

Linearly Independent Active Linear Constraints							
1	0	=	0	+	1.0000	*	CRACKERSPrivate

c) Use PROC MDC to predict the choice probabilities for the test sample using the estimated model.

PROC MDC Predicted Probabilities Table:

Here, crackers_pred column corresponds to choice probabilities for both training and test based upon model estimated using training samples. Test samples are indicated by . aka missing in Choice_training column.

VIEWTABLE: Work.Crackers_pred_data

	crackers_pred	OBS	Selection Indicator	Choice	Crackers	Price	Feat	Disp	choice_training	CRACKERS = Keebler	CRACKERS = Private	CRACKERS = Sunshine	CRACKERS = Nabisco	CRACKERS = Keebler * DISP	CRACKERS = Private * DISP	CRACKERS = Sunshine * DISP	CRACKERS = Nabisco * DISP	CRACKERS = Keebler * FEAT	CRACKER Private * FE
1	0.1484803567	1	1	0	Keebler	0.879999995	0	0	0	1	0	0	0	0	0	0	0	0	0
2	0.3465575508	1	1	0	Private	0.709999979	0	0	0	0	1	0	0	0	0	0	0	0	0
3	0.0687552954	1	1	0	Sunshine	0.980000019	0	0	0	0	0	1	0	0	0	0	0	0	0
4	0.4362067971	1	1	1	Nabisco	1.199999928	0	0	1	0	0	0	1	0	0	0	0	0	0
5	0.0600930878	2	1	0	Keebler	1.089999914	0	0	0	1	0	0	0	0	0	0	0	0	0
6	0.2639379266	2	1	0	Private	0.709999979	0	0	0	0	1	0	0	0	0	0	0	0	0
7	0.0508110054	2	1	0	Sunshine	0.99000001	0	0	0	0	0	1	0	0	0	0	0	0	0
8	0.6251579802	2	1	1	Nabisco	0.99000001	0	0	1	0	0	0	1	0	0	0	0	0	0
9	0.0534415064	3	1	0	Keebler	1.089999914	0	0	0	1	0	0	0	0	0	0	0	0	0
10	0.4114300136	3	1	0	Nabisco	1.089999914	0	0	0	0	0	0	1	0	0	0	0	0	0
11	0.1901220049	3	1	0	Private	0.779999912	0	0	0	0	1	0	0	0	0	0	0	0	0
12	0.345006475	3	1	1	Sunshine	0.49000001	0	1	1	0	0	1	0	0	0	1	0	0	0
13	0.0516399521	4	1	0	Keebler	1.089999914	0	0	0	1	0	0	0	0	0	0	0	0	0
14	0.1837128458	4	1	0	Private	0.779999912	0	0	0	0	1	0	0	0	0	0	0	0	0
15	0.0387097091	4	1	0	Sunshine	1.029999971	0	0	0	0	0	1	0	0	0	0	0	0	0
16	0.7259374931	4	1	1	Nabisco	0.889999986	0	0	1	0	0	0	1	0	0	0	0	0	0
17	0.0784587962	5	0	0	Keebler	1.089999914	0	0	.	1	0	0	0	0	0	0	0	0	0
18	0.425444179	5	0	0	Private	0.639999986	0	0	.	0	1	0	0	0	0	0	0	0	0
19	0.0490938315	5	0	0	Sunshine	1.089999914	0	0	.	0	0	1	0	0	0	0	0	0	0
20	0.4470031933	5	0	1	Nabisco	1.190000057	0	0	.	0	0	0	1	0	0	0	0	0	0
21	0.0925111149	6	1	0	Keebler	1.089999914	0	0	0	1	0	0	0	0	0	0	0	0	0
22	0.2747253544	6	1	0	Private	0.839999914	0	0	0	0	1	0	0	0	0	0	0	0	0

Question 3

Use the probabilities you predicted in Q2-c, to predict which brand is most likely to be chosen (brand with highest predicted choice probability). Create a 4x4 classification table for actual brand chosen and predicted brand chosen.

Predicted vs Actuals Cross-Tabulation Results:

Frequency
Percent
Row Pct
Col Pct

Table of actual_brand by predicted_brand

	predicted_brand			
actual_brand	Nabisco	Private	Sunshine	Total
Keebler	186 5.65 82.30 6.70	39 1.18 17.26 8.06	1 0.03 0.44 3.33	226 6.87
Nabisco	1618 49.15 90.29 58.24	166 5.04 9.26 34.30	8 0.24 0.45 26.67	1792 54.43
Private	777 23.60 75.07 27.97	249 7.56 24.06 51.45	9 0.27 0.87 30.00	1035 31.44
Sunshine	197 5.98 82.43 7.09	30 0.91 12.55 6.20	12 0.36 5.02 40.00	239 7.26
Total	2778 84.39	484 14.70	30 0.91	3292 100.00

Actual Brands Chosen Classification Table:

The SAS System

The FREQ Procedure

actual_brand	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Keebler	226	6.87	226	6.87
Nabisco	1792	54.43	2018	61.30
Private	1035	31.44	3053	92.74
Sunshine	239	7.26	3292	100.00

Predicted Brands Chosen Classification Table:

The SAS System

The FREQ Procedure

predicted_brand	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Nabisco	2778	84.39	2778	84.39
Private	484	14.70	3262	99.09
Sunshine	30	0.91	3292	100.00