**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.

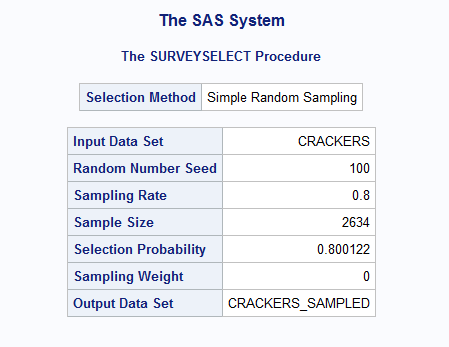
1. **OBS** : = Observation number
2. **Private, Keebler, Sunshine, Nabisco** : Indicator variables for which brand was chosen. Value of 1 indicates the brand that was chosen. Other 3 brands will be 0 in that observation.
3. **PricePrivate, PriceNabisco, PriceKeebler and PriceSunshine**: Prices that were 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.

1. 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:**



1. 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.
   1. 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.
   2. Private brand is taken as reference brand and hence α 0, Private = 0

Vj = ßPricej + α1, PrivateDispPrivate \* IPrivate + α2, PrivateFeatPrivate \* IPrivate +

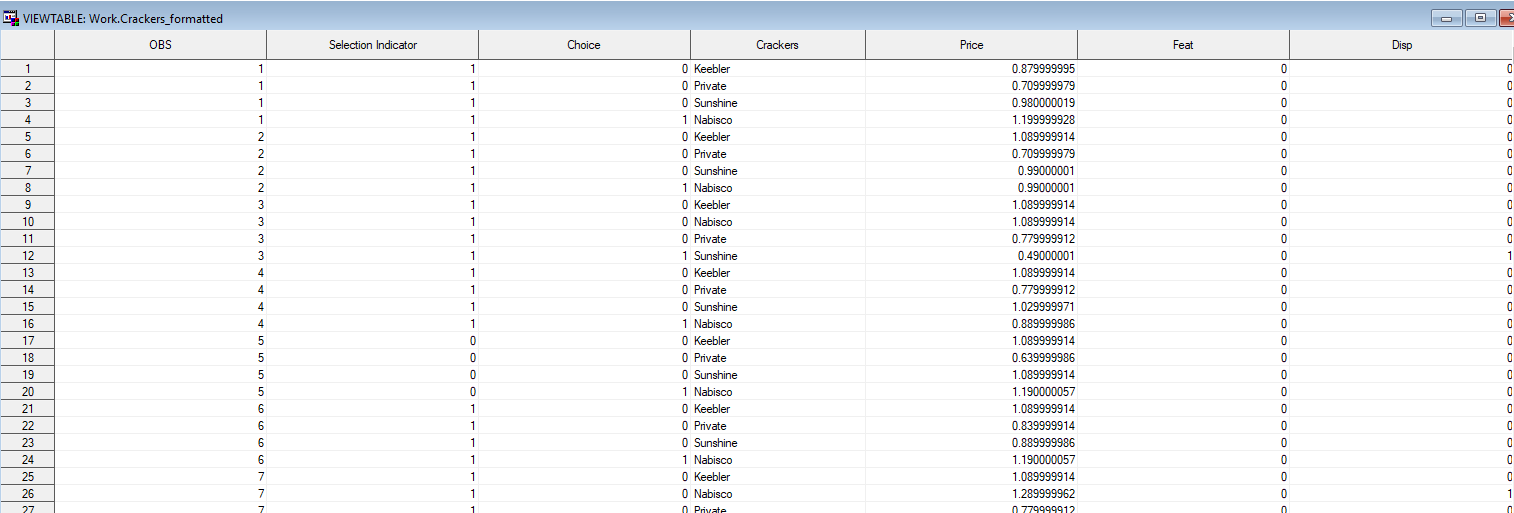
α 0, KeeblerIKeebler + α1, KeeblerDisp Keebler \* IKeebler + α2, KeeblerFeatKeebler \* IKeebler +

α 0, NabiscoINabisco + α1, NabiscoDispNabisco \* INabisco + α2, NabiscoFeatNabisco \* INabisco +

α 0, SunshineISunshine + α1, SunshineDispSunshine \* ISunshine + α2, SunshineFeatSunshine \* ISunshine

1. 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.
   1. 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.

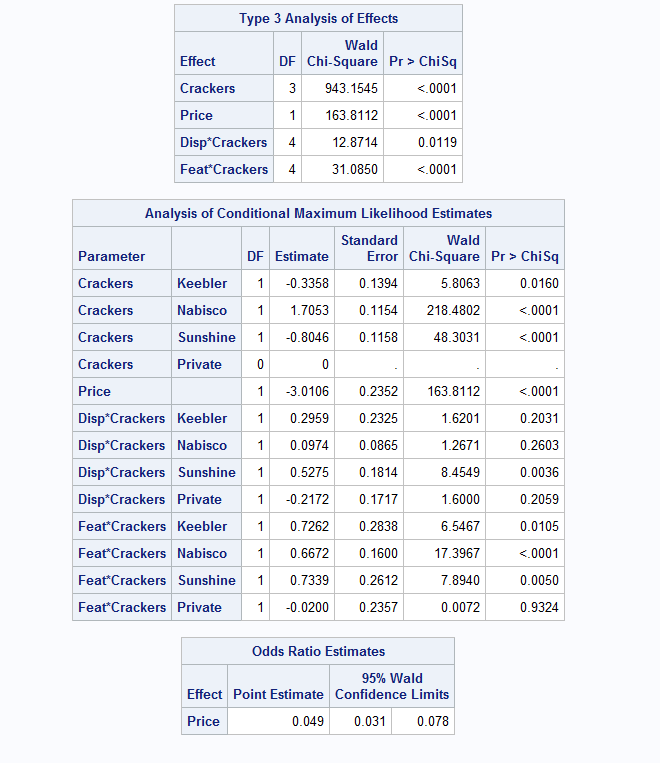
**Reformatted Data for PROC LOGISTICS Table:**



**Question 2**

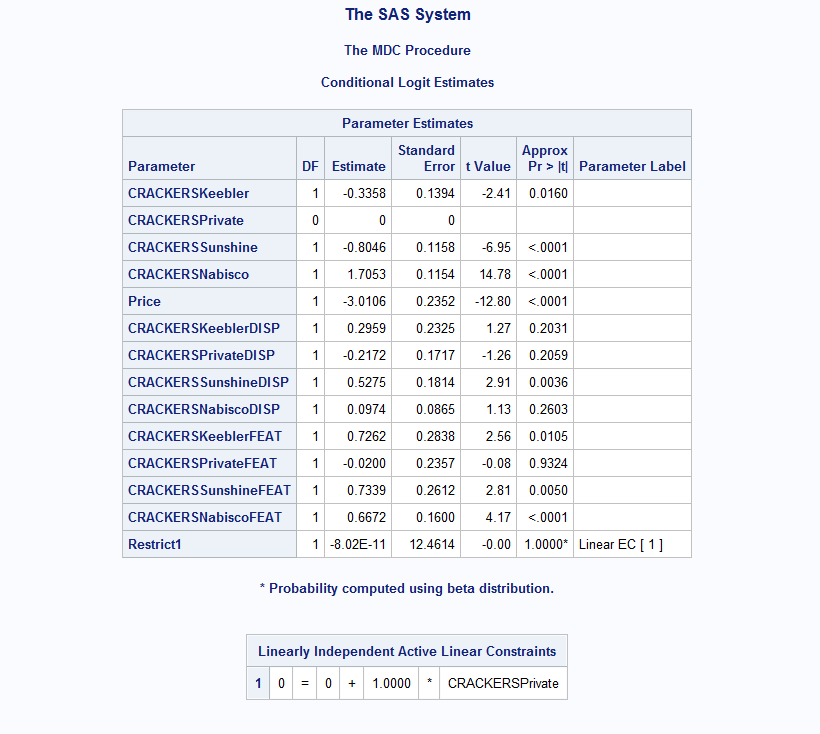
1. Estimate the logit model on the training sample using PROC LOGISTIC and report the estimation results (model parameters, significance).

**Logit Model PROC LOGISTIC Results:**



1. Reproduce your results using PROC MDC

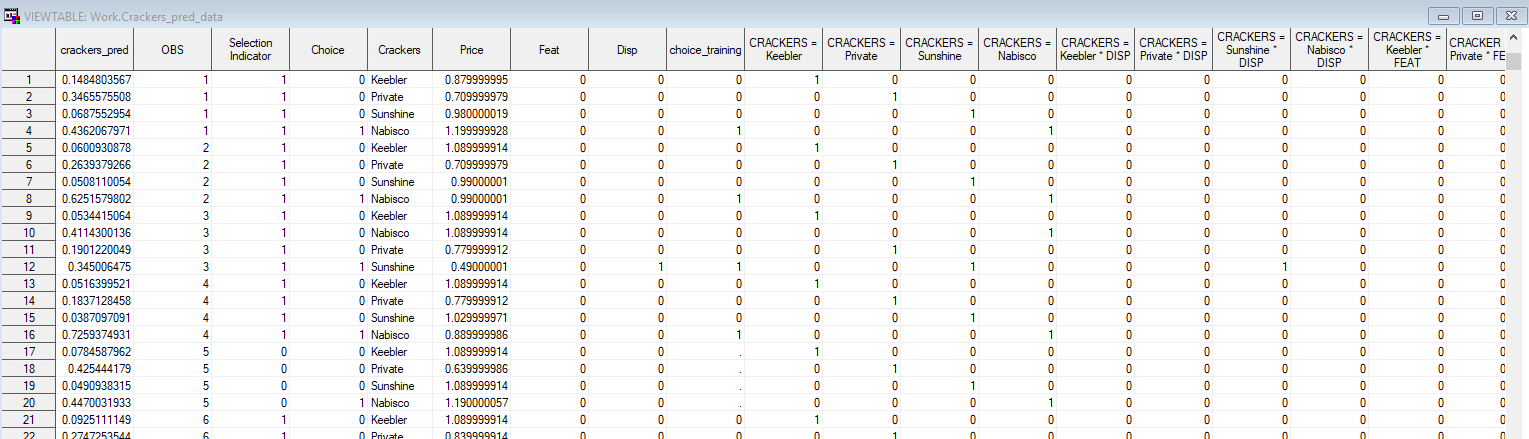
**Logit Model PROC MDC Results:**



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

**PROC MDC Predicted Probabilities Table:**

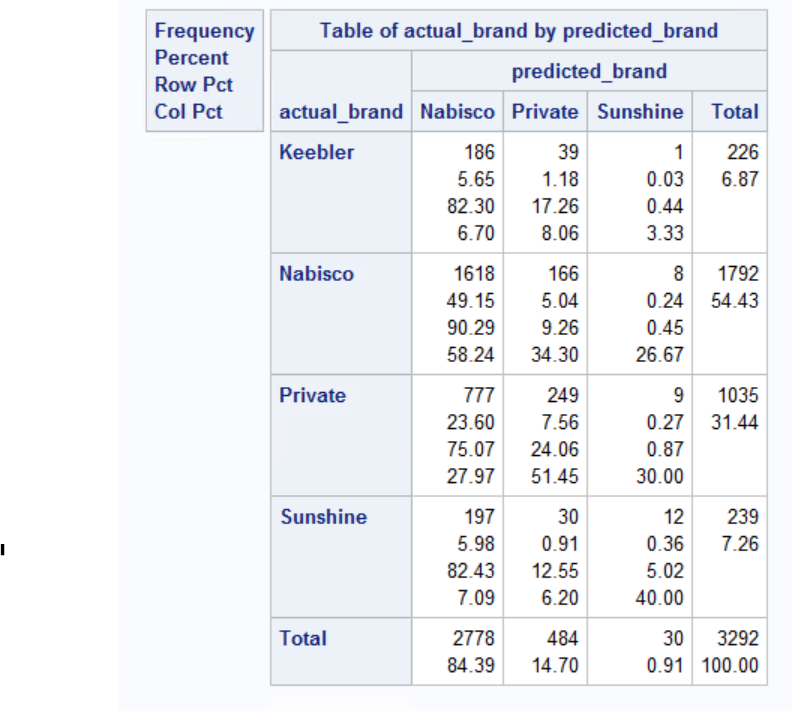
Here, crackers\_pred column corresponds to choice probabilities for both training and test based upon model estimated using training samples. Test samples are indicated by . aka missing in Choice\_training column.



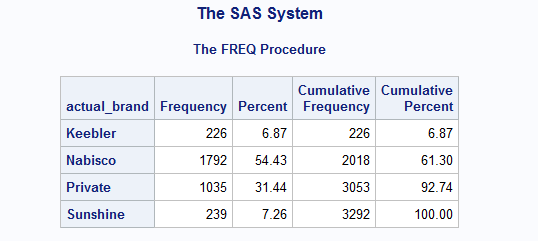
**Question 3**

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:**



**Actual Brands Chosen Classification Table:**



**Predicted Brands Chosen Classification Table:**

