BUAN 6337 Homework 4_Group 12

This Homework has 3 multiple part questions. 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 SAS dataset HeinzHunts has data on grocery store purchases of Hunts and Heinz ketchup. Each observation corresponds to one purchase occasion (of one of these brands) and consists of the following variables:

1. Heinz : =1 if Heinz was purchased, =0 if Hunts was purchased
2. PriceHeinz : Price of Heinz
3. PriceHunts : Price of Hunts
4. DisplHeinz : = 1 if Heinz had a store display, =0 if Heinz did not have a store display
5. DisplHunts : = 1 if Hunts had a store display, =0 if Hunts did not have a store display
6. FeatureHeinz : = 1 if Heinz had a store feature, =0 if Heinz did not have a store feature
7. FeatureHunts : = 1 if Hunts had a store feature, =0 if Hunts did not have a store feature

Question 1

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

a. Create a variable LogPriceRatio = log (PriceHeinz/PriceHunts).

Variable LogPriceRatio Added to Heinzhunts Dataset:

	HEINZ	HUNTS	PRICEHEINZ	PRICEHUNTS	FeatHeinz	FeatHunts	DisplHeinz	DisplHunts	LogPriceRatio
1	1	0	0.052000001	0.034000002	0	0	0	0	0.4248831544
2	1	0	0.052000001	0.044	0	0	0	0	0.1670541039
3	1	0	0.046	0.048	1	0	0	0	-0.042559614
4	1	0	0.052000001	0.034000002	0	0	0	0	0.4248831544
5	1	0	0.046	0.048	1	0	0	0	-0.042559614
6	1	0	0.046999998	0.029999999	0	0	0	0	0.4489502108
7	1	0	0.046	0.041000001	0	1	0	1	0.1150693054
8	1	0	0.046	0.030999999	0	0	0	0	0.3946542243
9	1	0	0.046999998	0.030999999	0	0	0	0	0.4161603869
10	1	0	0.050000001	0.028000001	0	0	0	1	0.5798184795
11	1	0	0.050000001	0.041000001	0	0	0	0	0.1984509343

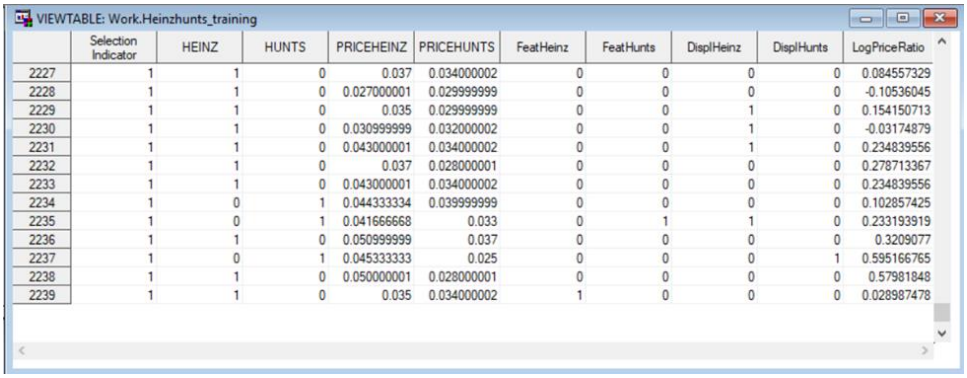
b. Randomly select 80% of the data set as the training sample, remaining 20% as test sample. (set seed=10)

Log Output:

```
11
12
13   /* B */
14   /* Create training and test datasets. 80% of sample in training  */
15   proc surveyselect data=heinzhunts out=heinzhunts_sampled outall samprate=0.8 seed=10;
16   run;

NOTE: The data set WORK.HEINZHUNTS_SAMPLED has 2798 observations and 10 variables.
NOTE: PROCEDURE SURVEYSELECT used (Total process time):
      real time           0.25 seconds
      cpu time            0.12 seconds
```

Test Dataset Showing 80% of the total observations:

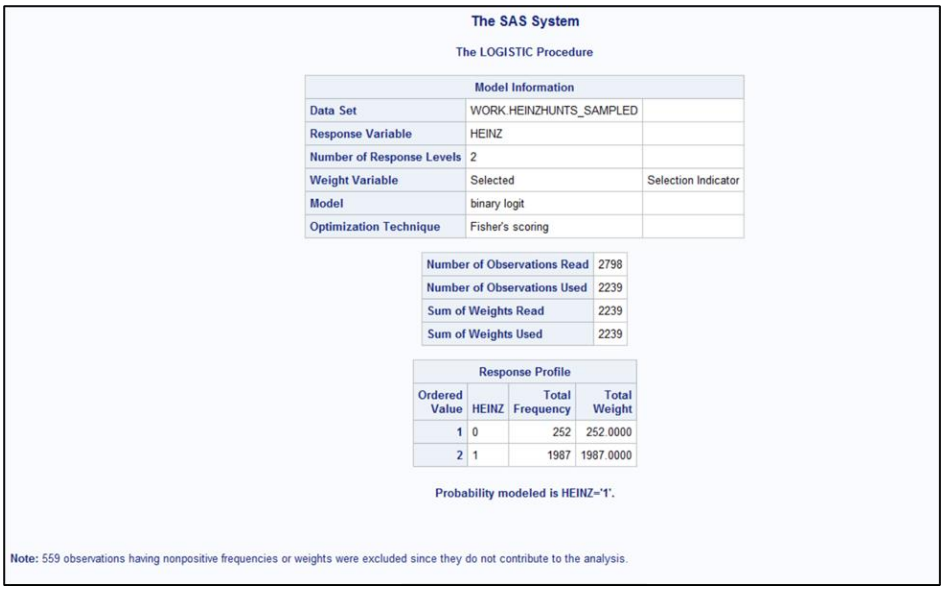


VIEWTABLE: Work.Heinzhunts_training

	Selection Indicator	HEINZ	HUNTS	PRICEHEINZ	PRICEHUNTS	FeatHeinz	FeatHunts	DisplHeinz	DisplHunts	LogPriceRatio
2227	1	1	0	0.037	0.034000002	0	0	0	0	0.084557329
2228	1	1	0	0.027000001	0.029999999	0	0	0	0	-0.10536045
2229	1	1	0	0.035	0.029999999	0	0	1	0	0.154150713
2230	1	1	0	0.030999999	0.032000002	0	0	1	0	-0.03174879
2231	1	1	0	0.043000001	0.034000002	0	0	1	0	0.234839556
2232	1	1	0	0.037	0.028000001	0	0	0	0	0.278713367
2233	1	1	0	0.043000001	0.034000002	0	0	0	0	0.234839556
2234	1	0	1	0.044333334	0.039999999	0	0	0	0	0.102857425
2235	1	0	1	0.041666668	0.033	0	1	1	0	0.233193919
2236	1	1	0	0.050999999	0.037	0	0	0	0	0.3209077
2237	1	0	1	0.045333333	0.025	0	0	0	1	0.595166765
2238	1	1	0	0.050000001	0.028000001	0	0	0	0	0.57981848
2239	1	1	0	0.035	0.034000002	1	0	0	0	0.028987478

c. Estimate a logit probability model for the probability that Heinz is purchased – using LogPriceRatio, DisplHeinz, FeatureHeinz, DisplHunts, FeatureHunts as the explanatory variables. Include interaction terms between display and feature for a particular brand (e.g., DisplHeinz * FeatureHeinz).

Logit Model with Interaction Terms:



The SAS System

The LOGISTIC Procedure

Model Information		
Data Set	WORK.HEINZHUNTS_SAMPLED	
Response Variable	HEINZ	
Number of Response Levels	2	
Weight Variable	Selected	Selection Indicator
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	2798
Number of Observations Used	2239
Sum of Weights Read	2239
Sum of Weights Used	2239

Response Profile

Ordered Value	HEINZ	Total Frequency	Total Weight
1	0	252	252.0000
2	1	1987	1987.0000

Probability modeled is HEINZ='1'.

Note: 559 observations having nonpositive frequencies or weights were excluded since they do not contribute to the analysis.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1577.424	1112.699
SC	1583.138	1158.409
-2 Log L	1575.424	1096.699

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	478.7251	7	<.0001
Score	399.8103	7	<.0001
Wald	261.9204	7	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	3.2142	0.1560	424.4032	<.0001
LogPriceRatio	1	-6.0112	0.4234	201.5199	<.0001
DisplHeinz	1	0.6390	0.2671	5.7238	0.0167
FeatHeinz	1	0.5460	0.3384	2.6032	0.1066
DisplHunts	1	-0.5529	0.2730	4.1019	0.0428
FeatHunts	1	-1.1403	0.3819	8.9140	0.0028
DisplHeinz*FeatHeinz	1	-0.8363	0.6093	1.8843	0.1699
DisplHunts*FeatHunts	1	-0.9322	0.7490	1.5490	0.2133

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
LogPriceRatio	0.002	0.001	0.006
DisplHeinz	1.895	1.122	3.198
FeatHeinz	1.726	0.889	3.351
DisplHunts	0.575	0.337	0.982
FeatHunts	0.320	0.151	0.676
HeinzDispFeat	0.433	0.131	1.430
HuntsDispFeat	0.394	0.091	1.709

d. Interpret the results. What promotional methods (feature / display) are effective for Hunts? For Heinz? How would you interpret the results for the interaction effects?

The log of odds that Heinz is purchased increases by 0.6390 and 0.5460, when Heinz makes use of a store display (vs no display for Heinz) and store feature (vs no feature for Heinz) respectively. Both are statistically significant at alpha of 0.1. Having either a display or feature for Heinz does seem to boost the likelihood of Heinz sales. Overall, Display is most effective for Heinz with an increase of 0.6390 versus an increase of 0.5460 when using a feature.

Whereas log of odds that Heinz is purchased when Hunts makes use of a store display (vs no display for Hunts) and store feature (vs no feature for Hunts) decreases by 0.5529 and 1.1403 respectively (Statistically significant at alpha of 0.05). Here the negative effect of Hunts promotions on Heinz is as expected. Overall, Hunts receives more benefit from the feature with –1.1403 compared to –0.5529 when using a display. The negative number detracting from likelihood of the customer purchasing Heinz which is in Hunts favor.

From interaction terms we infer that using both promotions at the same time does not provide statistically significant results for both brands. Here, the interaction effects for Heinz and Hunts both detract from the probability of a customer purchasing Heinz (-0.8363 for Heinz - Overuse of promotions might create a negative viewpoint in customers' mind that products sold through excessive sales promotions might either be of inadequate quality or that they are being deceived and –0.9322 for Hunts, which would be expected).

Question 2

Based on the estimated model in Question 1, and using the logit probability formula, calculate the change in predicted probability that Heinz is purchased if LogPriceRatio changes from 0.5 to 0.6 and Heinz does not use a feature or display, while Hunts uses a feature and a display.

The change in predicated probability that Heinz is purchased when log price ratio changes from 0.5 to 0.6 is 0.035.

Excel Calculations for Logit Probability Model:

With ineraction effects												
Intercept	DispHeinz	DispHunts	FeatHeinz	FeatHunts	DisplHeinz*FeatHeinz	DisplHunts*FeatHunts	LogPriceRatio					
3.2142	0.639	-0.5529	0.546	-1.1403	-0.8363	-0.9322	-6.0112					
		$Disp_{Heinz}$ & $Feat_{Heinz}$=0, $Disp_{Hunts}$ & $Feat_{Hunts}$=1										
	$Disp_{Heinz}$	**$Disp_{Hunts}$**	**$Feat_{Heinz}$**	**$Feat_{Hunts}$**	**$Displ_{Heinz}$*$Feat_{Heinz}$**	**$Displ_{Hunts}$*$Feat_{Hunts}$**	**LogPriceRatio**	**βX**	**Pr(Heinz)**	**Pr Change**	**Odds(Heinz)**	**Odds Ratio**
	0	1	0	1	0	1	0.5	-2.417	0.082	0.035	0.089	0.548
	0	1	0	1	0	1	0.6	-3.018	0.047		0.049	

Question 3

The above estimated model is to be used for targeting customers for Hunts coupons to build loyalty for the brand. Coupons are to be sent to customers who are likely to buy Hunts, and not to customers who are likely to buy Heinz. Therefore, the coupons should be sent to customers whose predicted probability of buying Heinz is below a certain threshold level that needs to be determined based on the costs of misclassifications (incorrectly sending / not sending a coupon)

The following information about the costs of incorrect classification is available: The cost of incorrectly sending a coupon to a customer who would have bought Heinz is $1 per customer, and the cost of incorrectly failing to send a coupon to a customer who would have bought Hunts is $0.25 per customer.

Based on these costs, what is the optimal threshold probability level that should be used with the estimated model to decide which consumers should receive coupons?

The probability threshold of 0.052318 has the lowest cost of $13.50.

Image of Lowest Least Total Cost Variable and Probability Level:

The SAS System

Least Total Cost	Probability Level
13.5	0.052318