MGMT 30500: Assignment 6 – Solution

PART A. (No problems from the text.)

PART B. Nine questions.

For each question, briefly explain why you chose your answer. Without explanations, there will be no points.

A marketing consultant for a cereal company is investigating the effectiveness of a TV advertisement for a breakfast cereal. The consultant shows the advertisement in a specific community for one month. Then, she randomly samples 100 shoppers as they leave a local supermarket to ask whether they viewed the Ad (Yes or No), whether they had at least a college degree (Yes or No), and whether they purchased the new cereal (1 if purchased, 0 otherwise). The consultant also asks the subjects about their annual household Income ($1000) and number of children. Data is recorded in Minitab (partially shown below) and a binary logistic regression is fit. The output is below (pages 2-3).

↓	C1 ☑		C2	C3	C4-T	C5-T
	Purchase		Income	No. Children	ViewAd	CollegeEdu
1	0		22.5576	2 Yes		Yes
2	0		39.4748	0 No		Yes
3	1		40.1053	2 Yes		Yes
4	1		58.8789	1 No		No
5	0		32.9780	2 Yes		No
6	0		55.5124	1 No		No
7	0		45.6061	1 No		No
8	0		60.2521	1 No		Yes

Binary Logistic Regression: Purchase versus Income, No. Children, ViewAd, CollegeEdu

Method

Link function	Logit
Categorical predictor coding	(1, 0)
Rows used	100

Response Information

Variable	Value	Count	
Purchase	1	42	(Event)
	0	58	
	Total	100	

Regression Equation

$$P(1) = \exp(Y')/(1 + \exp(Y'))$$

ViewAd	CollegeEdu			
No	No	Y'	=	-7.788 + 0.1088 Income + 1.547 No. Children
No	Yes	Y'	=	-8.127 + 0.1088 Income + 1.547 No. Children
Yes	No	Y'	=	-5.108 + 0.1088 Income + 1.547 No. Children
Yes	Yes	Y'	=	-5.447 + 0.1088 Income + 1.547 No. Children

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	-7.79	1.90	-4.09	0.000	
Income	0.1088	0.0323	3.37	0.001	1.35
No. Children	1.547	0.388	3.99	0.000	1.47
ViewAd					
Yes	2.680	0.637	4.20	0.000	1.53
CollegeEdu					
Yes	-0.339	0.574	-0.59	0.554	1.03

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
Income	1.1149	(1.0465, 1.1878)
No. Children	4.6986	(2.1954, 10.0563)

Odds Ratios for Categorical Predictors

Level A	Level B	Odds Ratio	95% CI
ViewAd			
Yes	No	14.5853	(4.1818, 50.8706)
CollegeEdu			
Yes	No	0.7122	(0.2312, 2.1939)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
31.52%	28.58%	103.18	103.82	116.20	0.8543

1. Which of the following model-free predictions (i.e., without using the predictor variable and a model) provides a better result for a new but similar data set?
   A. All 100 shoppers purchased the new cereal. (Error rate is 58%.)
   B. All 100 shoppers did not purchase the new cereal. (This gives a lower prediction error rate of 42%.)

2. Define the event of interest in this case. Also, describe the odds of the event.

   The event is "Purchasing the Cereal." Odds = P(Purchase)/P(No purchase) = P(Purchase)/[1 − P(Purchase)].

3. Write down the prediction equations for predicting the probability of the event.

The coefficients of the prediction equation are on page 3. If we write out the equation for the 4 combinations of levels of the two categorical predictors, we obtain the 4 prediction equations on page 2.

4. Based on this model, for a shopper who has an income of $55,000, no children and a college degree, and who saw the ad, the predicted probability that he/she will buy the cereal is:
   A. 0.631 (Use the last prediction equation with Income = 55, No. Children = 0. Note: ViewAd = 1, CollegeEdu = 1.)
   B. 0.537
   C. 0.989
   D. 0.716
   E. 0.335

5. Controlling for Income, No. Children and CollegeEdu, which of the following is correct?
   A. For a person who viewed the ad, the predicted probability they will buy the cereal increases by a factor of 2.682.
   B. For a person who viewed the ad, the predicted odds they will buy the cereal increase by a factor of 2.682.
   C. For a person who viewed the ad, the predicted odds they will buy the cereal increase by a factor of $\exp(2.682) \approx 14.5853$.
   D. For a person who viewed the ad, the predicted probability they will buy the cereal increases by a factor of $\exp(2.682) \approx 14.5853$.

6. At the 5% level, which of the predictors in this model individually has/have a significant effect on the probability of purchasing the cereal?
   A. All predictors.
   B. Only No. Children and ViewAd.
   C. No predictor is significant.
   D. All predictors except CollegeEdu. (Based on the 95% confidence intervals for the Odd ratios. If the interval does not contain 1, the predictor is significant. We can also find the Z-Values for Regression Coefficients and use the Empirical rule.)
   E. Cannot be determined based on the output.

7. If the odds of an event are 8 to 1 for the event, what is the probability of the event?
   A. 0.900
   B. 0.909
   C. 0.889 ($P(1)/P(0) = P(1)/(1-P(1)) = 8/1 = 8$. Hence, solving for $P(1)$, we obtain $P(1) = 8/9$.)
   D. 0.125

8. Controlling for ViewAd, No. Children, and CollegeEdu, which of the following is correct?

A. The probability of purchasing the cereal is predicted to increase by 0.1088 per unit increase in Income.
B. The odds of purchasing the cereal is predicted to increase by 0.1088 per unit increase in Income.
C. The odds of purchasing the cereal is predicted to decrease by a factor of exp(0.1088) per unit increase in Income.
D. The probability of purchasing the cereal is predicted to increase by a factor of exp(0.1088) per unit increase in Income.
E. The natural log of the odds of purchasing the cereal is predicted to increase by 0.1088 per unit increase in Income. (The standard interpretation of slope. Also, 0.1088 is the predicted change in Y' per unit increase in Income, but Y' = ln(Odds).)

9. Between the random prediction (in Problem 1) and prediction with logistic regression, which is expected to perform better in terms of the prediction accuracy? Why?

The logistic regression where some predictors were significant (useful) predictor(s). In case none of the predictors is significant, the two methods are expected to perform equally well, or poorly, in the long run.