6500: 601: 802 Business Analytics and Information Strategy Lab 3: Multiple Logistic Regression

Description: Terry's Village* is an online retailer that specializes in gifts, outdoor knick-knacks, home décor, and holiday and seasonal items. Terry wants to determine what kinds of customers are likely to purchase more this year than last year, (that is, Increase = 1). She has asked you to run a logistic regression analysis with dependent variable **Increased** and predictors **Age, Gender, OwnHome, Married, Close, Salary, Children, Customer2010, and Catalogs**. She has requested that you use 50% of data (125 cases) for training, and 50% for testing purposes.

Directions: The data are in the *Catalog Sales* file and the codebook is shown below.

Variable	Descriptions
Age	Age of customer
Gender	Gender of customer (0=female, 1=male)
OwnHome	1 if person owns home, 0 if person is renting
Married	marital status (1=married, 0 otherwise)
Close	1 if a customer lives "close" to stores offering similar products, 0 otherwise
Children	Number of children living with a customer
Salary	A customer's annual salary
Customer2010	1 if any purchases were made last year, 0 if none were made last year
Spent2010	Amount that a customer spent in the last year
Spent2011	Amount that a customer spent this year
Catalogs	Number of catalogs sent to a customer this year
Increased	1 if this year spending is greater than last year spending, 0 otherwise

Prepare a report for Terry that shows the results of a multiple linear regression model. In your report, be sure to proceed step-by-step using the example from the Logistic Regression lecture on Springboard as a guide and include:

Task1) Build a Logistic Regression model.

1) In JMP, Analyze → Fit Model → **Increased** to Y and **Age, Gender, OwnHome, Married, Close, Salary, Children, Customer2010, Catalogs** to Construct Model Effects.

Task 2) Interpret the results.

1) Use the Whole Model and Lack of Fit Tests. Identify significant Parameter Estimates and interpret their Odds Ratios.
2) Define most important predictors. What characteristics of customers are most important for predicting the **Increased** variable? Why?

Task 3) Evaluate the model.

1) ROC, LIFT, Confusion Matrix
2) Save "Probability Formula". Using the newly created column "Most Likely Increased", you can create a confusion matrix for training data and a confusion matrix for testing data. In JMP, Fit Y by X → **Increased** to Y and "Most Likely Increased" to X.
3) You can do some trial and error to find a better cutoff probability that increase accuracy in confusion matrix. In JMP, create a new column "Logistic Outcome" with formula "If(Prob[1]>0.3] → 1, else 0". Then create a Confusion Matrix.

Task 4) Explain how Terry can use these results to make a new advertising plan.