**Week 7 – Homework Exercise 1**

The **PVA** table contains data that represents charitable donations made to an American veterans’ association. The data represent the results of a mail campaign to solicit donations. The data set contains the following information:

* a flag to indicate respondents to the appeal and the dollar amount of their donations (**Target Gift Flag** and **Target Gift Amount**)
* respondents’ PVA promotion and giving history
* demographic data of the respondents

#### PVA Metadata Table

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** |  | **Measurement Level** | **Description** |
| **DemAge** |  | Interval | Age |
| **DemCluster** |  | Nominal | Demographic Cluster |
| **DemGender** |  | Nominal | Gender |
| **DemHomeOwner** |  | Binary | Home Owner |
| **DemMedHomeValue** |  | Interval | Median Home Value Region |
| **DemMedIncome** |  | Interval | Median Income Region |
| **DemPctVeterans** |  | Interval | Percent Veterans Region |
| **GiftAvg36** |  | Interval | Gift Amount Average 36 Months |
| **GiftAvgAll** |  | Interval | Gift Amount Average All Months |
| **GiftAvgCard36** |  | Interval | Gift Amount Average Card 36 Months |
| **GiftAvgLast** |  | Interval | Gift Amount Last |
| **GiftCnt36** |  | Interval | Gift Count 36 Months |
| **GiftCntAll** |  | Interval | Gift Count All Months |
| **GiftCntCard36** |  | Interval | Gift Count Card 36 Months |
| **GiftCntCardAll** |  | Interval | Gift Count Card All Months |
| **GiftTimeFirst** |  | Interval | Time Since First Gift |
| **GiftTimeLast** |  | Interval | Time Since Last Gift |
| **ID** |  | Nominal | Control Number |
| **PromCnt12** |  | Interval | Promotion Count 12 Months |
| **PromCnt36** |  | Interval | Promotion Count 36 Months |
| **PromCntAll** |  | Interval | Promotion Count All Months |
| **PromCntCard12** |  | Interval | Promotion Count Card 12 Months |
| **PromCntCard36** |  | Interval | Promotion Count Card 36 Months |
| **PromCntCardAll** |  | Interval | Promotion Count Card All Months |
| **StatusCat96NK** |  | Nominal | Status Category 96NK |
| **StatusCatStarAll** |  | Binary | Status Category Star All Months |
| **TARGET\_B** |  | Binary | Target Gift Flag |
| **TARGET\_D** |  | Interval | Target Gift Amount |

In this exercise, you continue to use the **PVA** data set. You build a logistic regression model to classify those customers who donated.

1. **Building a Logistic Regression in SAS Visual Statistics**
   1. Start Visual Analytics or start a new report. Then select and open the **PVA** data source.
   2. Add a logistic regression to the canvas.
   3. If you did not do so already, in the Measure column, edit **Target Gift Flag**. Select **Category** to create a binary target variable for donations.
   4. Add **Target Gift Flag** as the response.
   5. Add **Gender**, **Home Owner**, and **Status Category 96NK** as classification effects. Then add all 23 variables as continuous effects ***except*** **Target Gift Amount**, **Target Gift Amount with Zero**, and **Median Home Value Region**. (You add 20 columns.)
   6. In the Options pane, select the **Fast Backward** variable selection method. Keep the significance level at **.01**.
   7. Create the logistic model (or review results).
2. Examine the Fit Summary panel. How many of the 23 input variables are not included   
   in this model?

7 variables are not included in this model

1. In this model, are any of the insignificant variables that are not included classification effects?

gender

1. What is the value of the Misclassification Rate (Event) statistic?

0.4265 is misclassification rate

1. **Examining Additional Logistic Regression Results**
   1. Open the details table and click the **Parameter Estimates** tab.
      1. Click the **Estimate** column heading twice to sort the parameter estimates. Then determine which parameter had the largest estimate. What was the value?

The largested estimate is Status Catgegory 96NK E with value of 0.421181.

* + 1. Click the **Response Profile** tab to determine how many of these customers made donations. How many customers did this?

There is 34111 people who made donation.

* 1. Close the details table.
  2. Maximize the assessment plot to gain access to the assessment charts.
     1. Examine the **lift chart** to determine the advantages of using this model for prediction. How does this model compare to the Best model?  
        While best model has big changes between 50 Percent to 60 percent, our model does not have big difference. In about 55%, these two model meet in the intersection and after the intersection, best model dropped under our model.
     2. Select the **ROC** chart and report the KS statistic and the associated cutoff value.
     3. What is the prediction cutoff value that is used in the current logistic regression model?

Cutoff value is 0.54 and 0.1539 of KS statistic.

* + 1. Select the misclassification chart to determine whether this model predicts more true positives or more true negatives. Which is it?

This model has higher true positive than true negative. True positive is the section of people that had prediction of will donate, and actually did. Unlike True positive, True negative is group of people that had prediction of will not donate, and actually did not donate.

* 1. Save your report. You will need to reference your saved work in Homework Exercise 2 below.

**Week 6 – Homework Exercise 2**

In this exercise, you continue to use the **PVA** data set to investigate segmenting a logistic regression model interactively by gender.

1. **Adding an Interactive Group-By Variable in SAS Visual Statistics**
   1. Open your saved report from Exercise 1.
   2. On the Roles option, move the variable **Gender** from the Classification effects role to the Group by role. Do this by dragging and dropping the variable or by right-clicking, removing, and re-adding.
   3. Examine the Fit Summary pane to discover that there are several observations with a value of *U*. Create a report filter to remove this value from the current model.
   4. Maximize the Fit Summary window.
2. Is **Age** a significant variable for both the *Male* and *Female* segmented models?

Yes, I think Age is significant variable for male and female segmented model.

1. Is **Promotion Count Card 36 Months 36 Months** significant for both segmented models?

Promotion count card 36 month is significant for female segmented model but male segment model is no.

* 1. Remove **Gender** from the model. Keep the filter on though.
  2. On the Roles pane, move the variable **Status Category 96NK** from the Classification effects role to the Group by role.

1. Which effects are important to every BY-group level?

There is no BY-Group level.

* 1. Open the details table and click the **Parameter Estimates** tab.

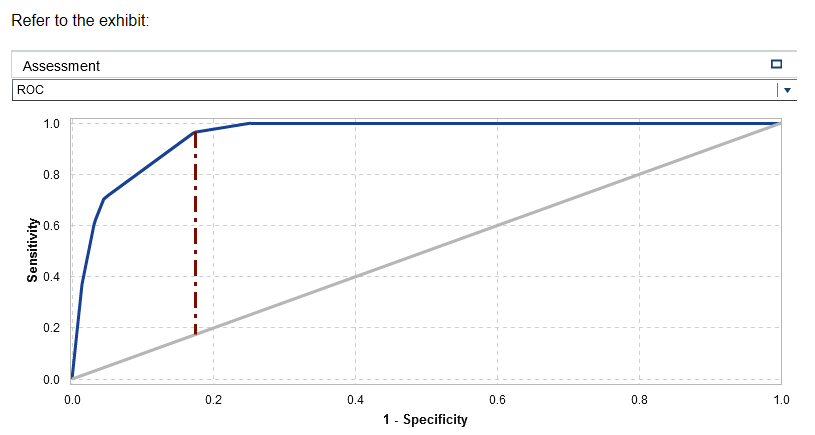
1. How many effects are significant at .01 for the **N** BY-group level?

12 effects

* 1. Save your report.

**Week 6 – Homework Exercise 3**

1. Refer to the exhibit below:

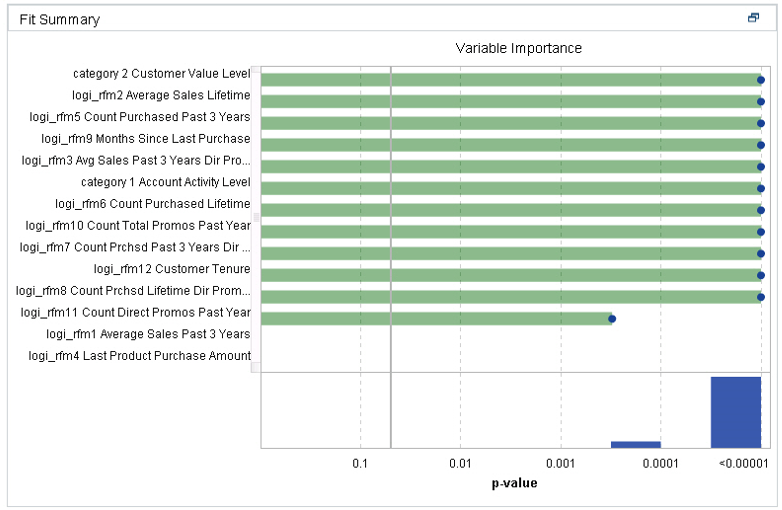


What is the correct interpretation of this Assessment visualization for a logistic regression analysis?

* 1. The model correctly classifies observations at the same rate it incorrectly classifies them
  2. The model classifies more observations incorrectly than correctly
  3. The model classifies more observations correctly than incorrectly
  4. The model classifies all observations correctly

**Answer: C**

1. Refer to the exhibit below:



What can you conclude from the Fit Summary chart of a logistic regression analysis?

* 1. The significance level for variable selection is .05
  2. Count Direct Promos Past Year is almost significant
  3. Two effects (or predictors) do not contribute to this model
  4. The variable selection options was not selected

**Answer: C**

1. If you fit a logistic regression model in SAS Visual Statistics, what is the impact of changing the Prediction cutoff value?
   1. The Lift plot updates automatically
   2. The predicted values are the same as they would be with the default cutoff value
   3. The residuals can become larger or smaller, depending on the change to the cutoff
   4. The Misclassification plot updates automatically

**Answer: d**

1. Which statement is true for the Informative Missingness property in SAS Visual Statistics?
   1. Missing values are imputed, and indicator variables are created for the missing predictors
   2. Missing values are flagged on diagnostic plots
   3. A table of missingness patterns for the predictor variables is displayed in the model output
   4. The user is prompted to fill in a value to impute for missing predictor variables

**Answer: a**