



Practice

In this practice, 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

- a. Start Visual Analytics or start a new report. Then select and open the **PVA** data source.
- b. Add a logistic regression to the canvas.
- c. 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.
- d. Add **Target Gift Flag** as the response.
- e. 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.)
- f. In the Options pane, select the **Fast Backward** variable selection method. Keep the significance level at **.01**.
- g. Create the logistic model.
 - Examine the Fit Summary panel. How many of the 23 input variables are not included in this model?
 - In this model, are any of the insignificant variables that are not included classification effects?
 - What is the value of the AIC statistic?

2. Examining Additional Logistic Regression Results


- a. 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?
 - 2) Click the **Response Profile** tab to determine how many of these customers made donations.
- b. Close the details table.

- c. 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?
 - 2) Select the **ROC** chart and report the KS statistic and the cutoff value.
 - 3) What is the prediction cutoff value that is used in the current logistic regression model?
 - 4) Select the misclassification chart to determine whether this model predicts more true positives or more true negatives.
- d. Save your report as **Practice 6**.

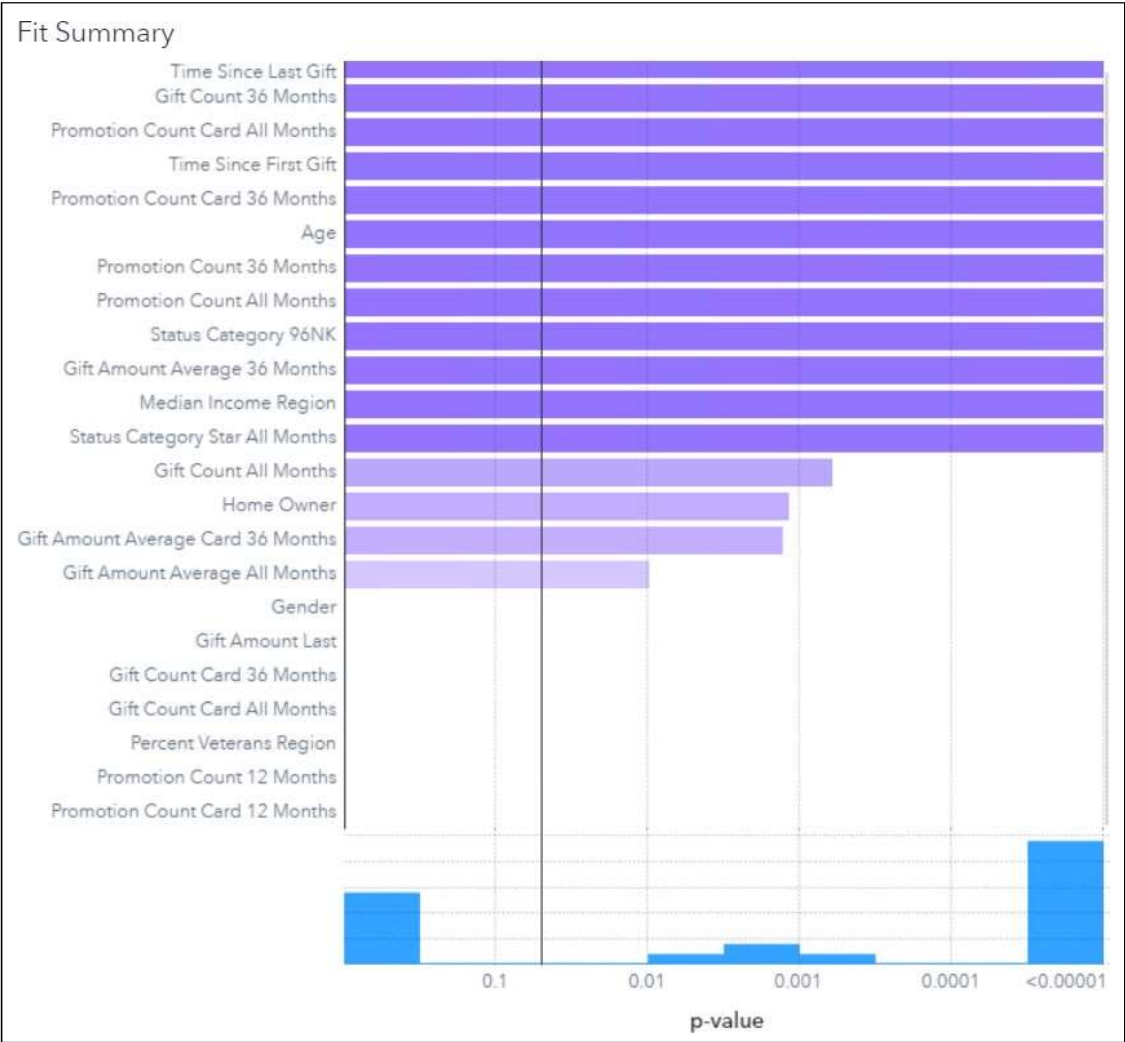
End of Practices

Solutions to Practices

1. Building a Logistic Regression in SAS Visual Statistics


- a. Start Visual Analytics or start a new report. Then select and open the **PVA** data source.
 - 1) Click **Explore and Visualize** in the application shortcut area. Alternatively, in the upper right corner, click  (**Menu**) and select **New**.
 - 2) Select the Data pane. Click **Add data** if necessary. In the Choose Data window, select **PVA** in the Available tab and click **OK**.
- b. Add a logistic regression to the canvas. In the Objects pane, double-click **Logistic Regression**.
- c. 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. Select **Edit properties** on **Target Gift Flag** and select **Category** under Classification.
- d. Add **Target Gift Flag** as the response. In the Roles pane, click **Add** under **Response**.
- e. Add **Gender**, **Home Owner**, and **Status Category 96NK** as classification effects. In the Roles pane, click **Add** under **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.) In the Roles pane, click **Add** under **Continuous effects**.
- f. In the Options pane, select the **Fast Backward** variable selection method. Keep the significance level at **.01**.

- i. Create the logistic model.
- Examine the Fit Summary panel. How many of the 23 input variables are not included in this model? **7**



- In this model, are any of the insignificant variables that are not included classification effects? **Gender**
- What is the value of the AIC statistic? **87,837.47** (Select **R-Square** in the summary bar to change the statistic.)

2. Examining Additional Logistic Regression Results


- a. Open the details table and click the **Parameter Estimates** tab. (On the report, click  (Maximize) on the object toolbar to see the details table.)

- 1) Click the **Estimate** column heading twice to sort the parameter estimates. Then determine which parameter had the largest estimate. What was the value? **Status Category 96NK – E, .4212**

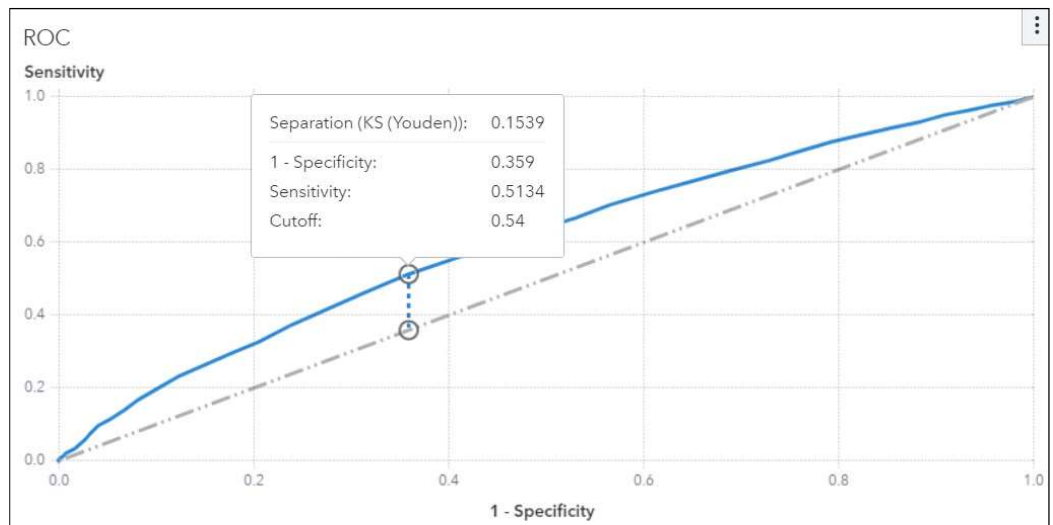
<div> ◀ Convergence Fit Statistics Parameter Estimates Type III Test </div>				
Parameter	Estimate ▼	Standard Error	Chi-Square	Pr > ChiSq
Status Category 96NK E	0.421181	0.091588	21.14774	<0.00001
Intercept	0.229193	0.098056	5.463319	0.01942
Status Category Star All Months	0.11265	0.022303	25.51135	<0.00001
Gift Count 36 Months	0.077338	0.006454	143.5873	<0.00001
Promotion Count Card 36 Months	0.071666	0.00842	72.44742	<0.00001
Home Owner H	0.060707	0.018703	10.5358	0.00117
Status Category 96NK N	0.029824	0.046143	0.417743	0.51807
Promotion Count All Months	0.016059	0.002033	62.4168	<0.00001

- 2) Click the **Response Profile** tab to determine how many of these customers made donations. **34111**

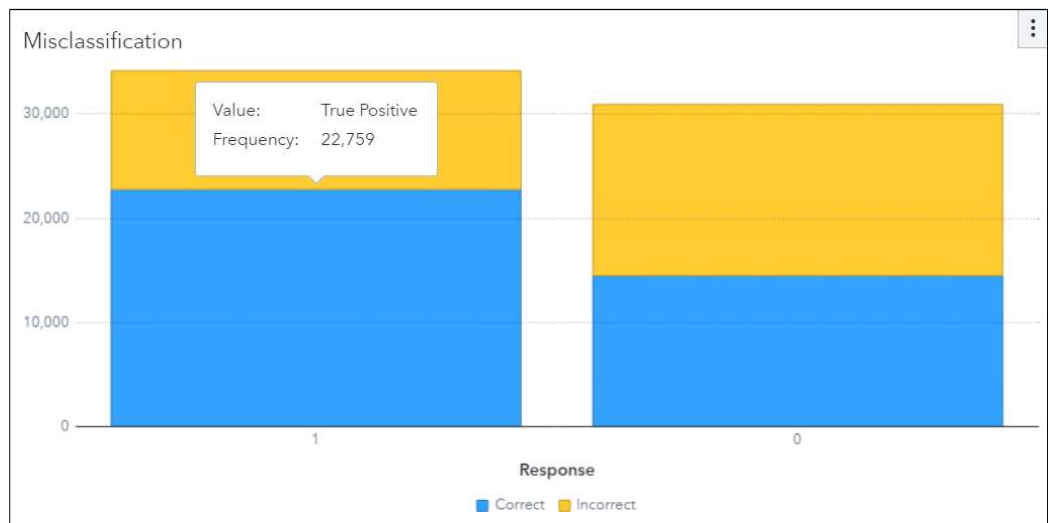
<div> ◀ Dimensions Response Profile </div>		
Ordered Value	Count	Target Gift Flag
1	34111	1
2	30877	0

- b. Close the details table. Click  (Restore) on the object toolbar to exit maximize mode.
- c. Maximize the assessment plot to gain access to the assessment charts. (In the Options pane under Model Display, change the plot layout to **Stack**. Click the **Assessment** tab at the bottom of the results.)
- 1) Examine the lift chart to determine the advantages of using this model for prediction. How does this model compare to the Best model? **When the Model line is closer to the Best line, especially in the lower percentiles, the model is proportionally better. Some of the effects are skewed and some of the effects contain missing values. Performing imputation on the missing values and transforming some of the variables might yield a better model.**

- 2) Select the **ROC** chart and report the KS statistic and the cutoff value. **.154 and .54** (Right-click the chart and select **ROC**.)



- 3) What is the prediction cutoff value used in the current logistic regression model? **Click the Options pane. The default value is specified under the Assessment section as .50.**
- 4) Select the misclassification chart to determine whether this model predicts more true positives or more true negatives. **It predicts true positives.** (Right-click and select **Misclassification**.)



- d. Save your report as **Practice 6**.

Click **⋮ (Menu)** and select **Save As**. Navigate to **Folders** ⇒ **My Folder**. Name the report **Practice 6**.