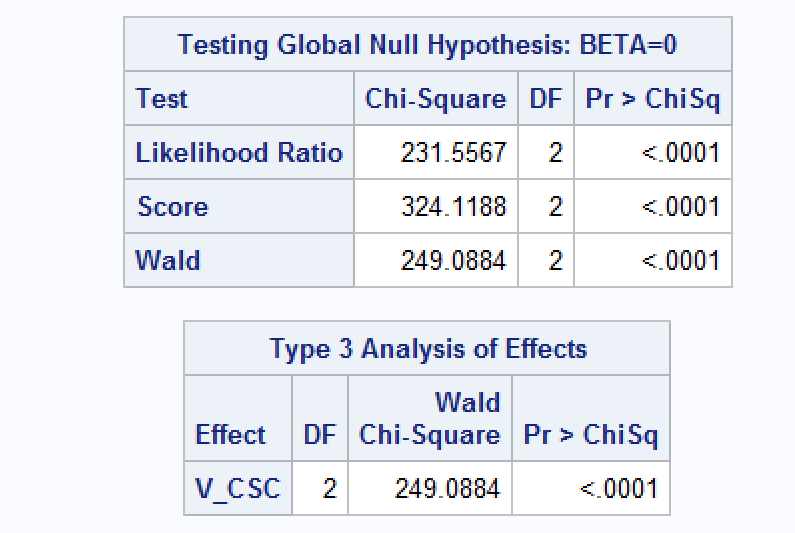
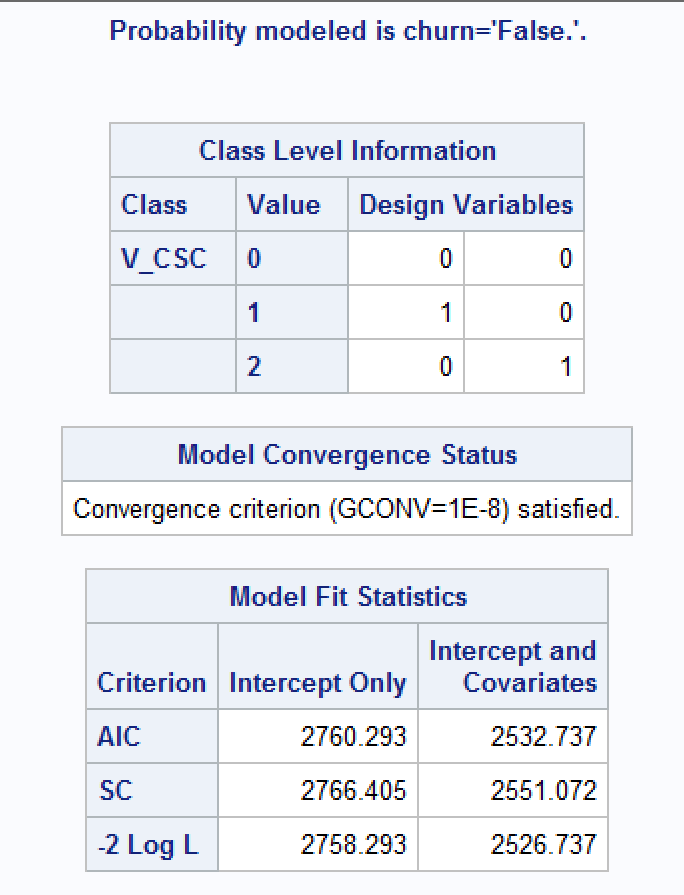
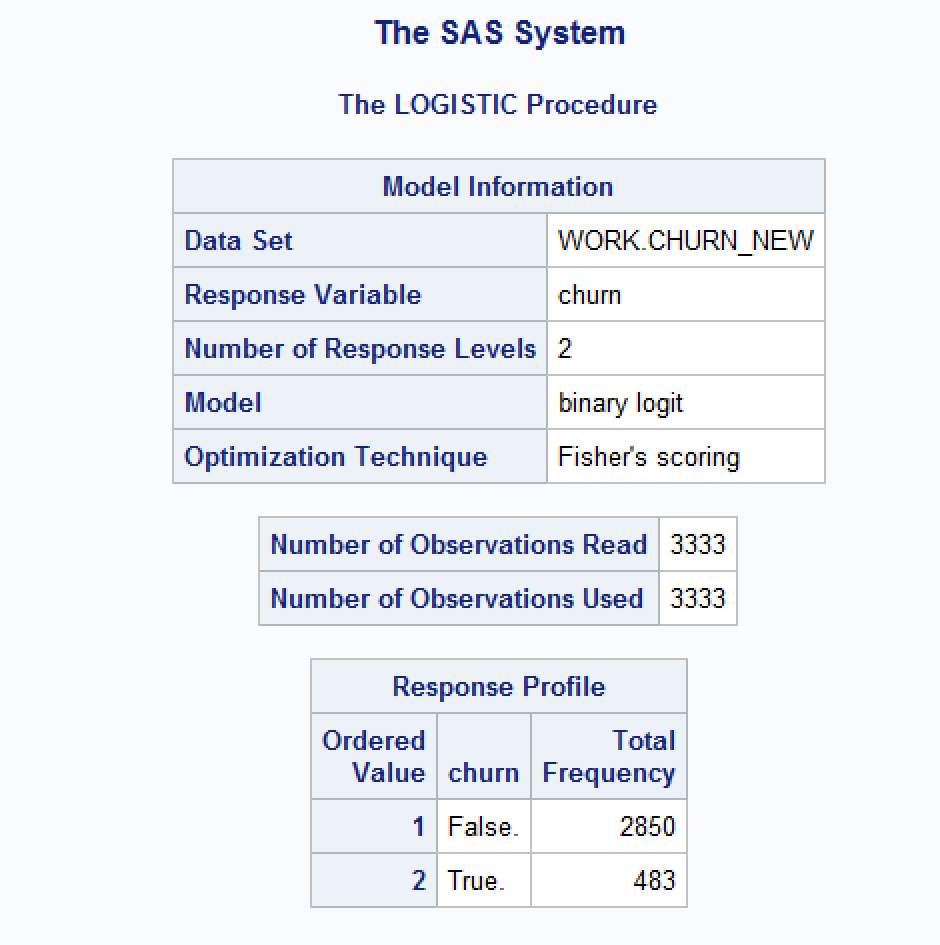
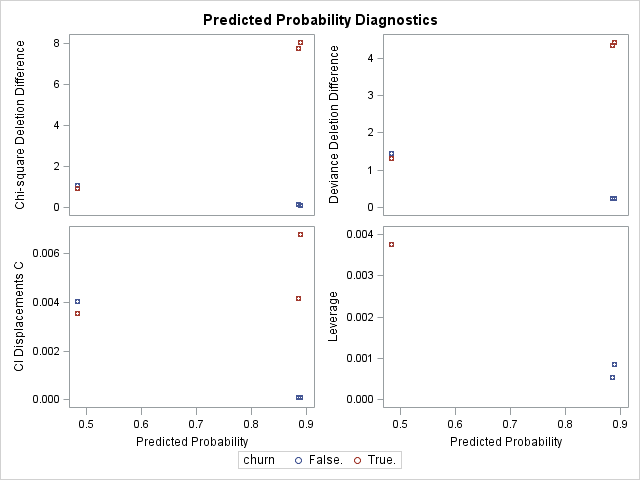
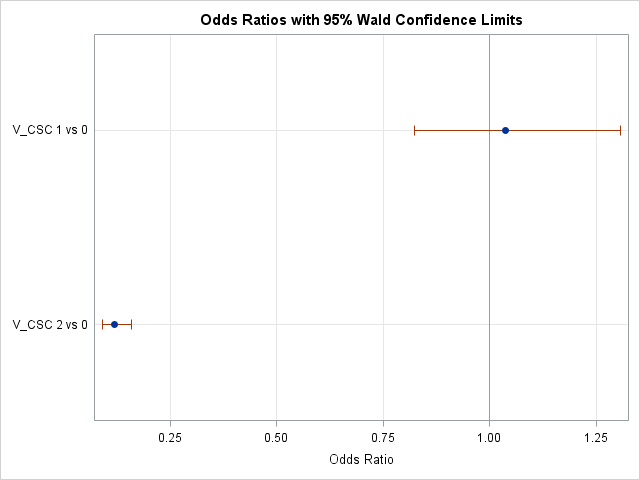
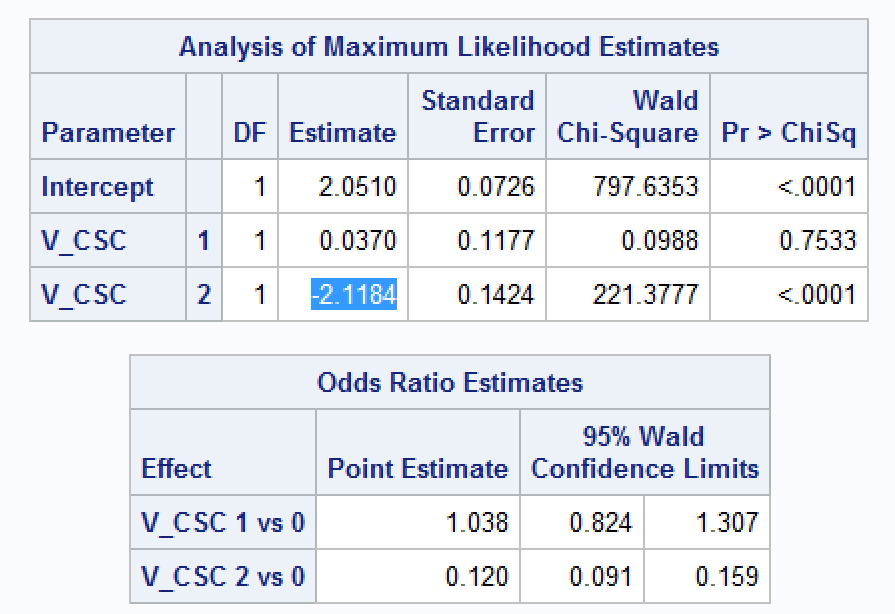
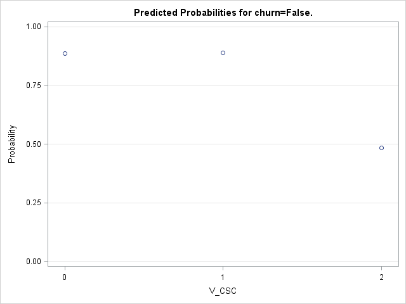
**Question 5.1: Logistic Regression for Churn data**

In this model, we set the base that churn = false and that V\_CSC is 0. While the overall model is valid, the mid-level customer service calls (1) is well beyond the acceptable amount. The maximum likelihood algorithm does converge when using the gradient convergence criterion. The logistic regression is:

**Log [p(churn=1) / (1- p(churn=1))] = 2.0510 + 0.0370\*(V\_CSC1) + -2.1184\*(V\_CSC2).**







**Question 5.2: Logistic Regression for Churn data**

When we split the data into two data sets – odd records and even records, and run a logistic regression of churn on day minutes, we find that the two data sets have disparate logistic regressions so that:

|  |  |
| --- | --- |
| Odd: **Log [p(churn=1) / (1- p(churn=1)) =**  **4.6103 + -0.0147\*(Day\_Mins)** | Even: **Log [p(churn=1) / (1- p(churn=1)) =**  **3.3066 + -0.00807\*(Day\_Mins)** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Question 5.3: Logistic Regression for Churn data**

In order to run a logistic regression on multiple variables in the data set, we need to identify the categorical data and either adding it to class or transforming it into numerical data through creating a new data set. The V\_CSC we create is measuring only the highest category of customer service calls – any customer that had four or more calls. We find that:

**Log [p(V\_churn=1) / (1- p(V\_churn=1))] =**

**-8.1596 + 2.0328\*(International\_Plan\_indicator) + -1.0443\*(Voice\_Plan\_indicator) + 2.6768\*(V\_CSC2) + 0.000801\*(account\_length) + 0.0135\*(day\_mins) + 0.00730\*(eve\_mins) + 0.00424\*(night\_mins) + 0.0853\*(intl\_mins)**

