Incorporating Additional Predictors

On the worksheet labeled “LgstcReg 2 Predictors”, you will find a template for running logistic regression with 2 predictor variables. This spreadsheet follows the same structure as “Logistic Regression Template,” but allows for two predictors to be used in estimating the logistic regression (Predictor 1 and Predictor 2), and two predictors to be used in assessing performance in the validation data (Validation Predictor 1 and Validation Predictor 2).

For demonstration purposes, we will use homeownership and the debt-to-income ratio (dti) to predict whether or not a loan is considered as risky.

In cell B2, reference the recoded value for home ownership on the Calibration Data worksheet in cell R2:

='Calibration Data'!R2

Copy this formula down the worksheet.

In cell C2, reference the dti variable (Column G) on the Calibration Data worksheet in cell G2:

='Calibration Data'!G2

In cell D2, reference the recoded value for the loan status on the Calibration Data worksheet in cell U2:

='Calibration Data'!U2

Copy this formula down the worksheet.

In cell E2, we will enter the probability (based on logistic regression) that the loan is classified as risky. This formula adds an additional term to the expression that we used when we had only a single predictor:

=EXP($J$1+$J$2*B2+$J$3*C2)/(1+EXP($J$1+$J$2*B2+$J$3*C2))

Copy this formula down the worksheet.

In cell F2, we can calculate the probability that the loan is classified as on-time as:

=1-E2

Copy this formula down the worksheet.

In cell G2, we will calculate the log-likelihood of the observed outcome. If the outcome is equal to 1, this is the ln(P(Risky)). If the outcome is equal to 0, this is ln(P(On-time)):

=IF(D2=1,LN(E2),LN(F2))

Copy this formula down the worksheet. In Cell J4, calculate the log-likelihood of the data sample as:

=SUM(G2:G50001)

Next, we'll populate the calculations for the model validation. In M2, reference the recoded homeownership status variable from Validation Data sheet by entering the formula:

='Validation Data'!R2

In cell N2, we will reference the debt-to-income ratio:

='Validation Data'!G2

Copy this formula down the worksheet. Next, in cell O2, we will reference the recoded loan status variable from the Validation data sheet:

='Validation Data'!U2

Copy this formula down the worksheet. In cell P2, we will calculate the probability that the loan is considered risky using the results of the logistic regression. Enter the following formula:

=EXP($J$1+$J$2*M2+$J$3*N2)/(1+EXP($J$1+$J$2*M2+$J$3*N2))

Copy this formula down the worksheet. To calculate the error, we will first calculate a measure of error based on the naïve assumption that all loans have the same likelihood of being risky. In this scenario, we will use the proportion of loans that are risky in our calibration dataset as our naïve estimate. We will calculate the error as |Observed-Predicted|. To calculate the absolute error associated with using this estimate, in cell Q2, enter the formula:

=ABS(O2-AVERAGE($D$2:$D$50001))

In cell R2, we will calculate the absolute error using the results of our logistic regression:

=ABS(O2-P2)

Copy the formulas in Q2 and R2 down the worksheet.

To calculate the mean absolute error under these models, we will average over the errors associated with individual loans. In cell U1, the error can be calculated as:

=AVERAGE(Q2:Q10001)

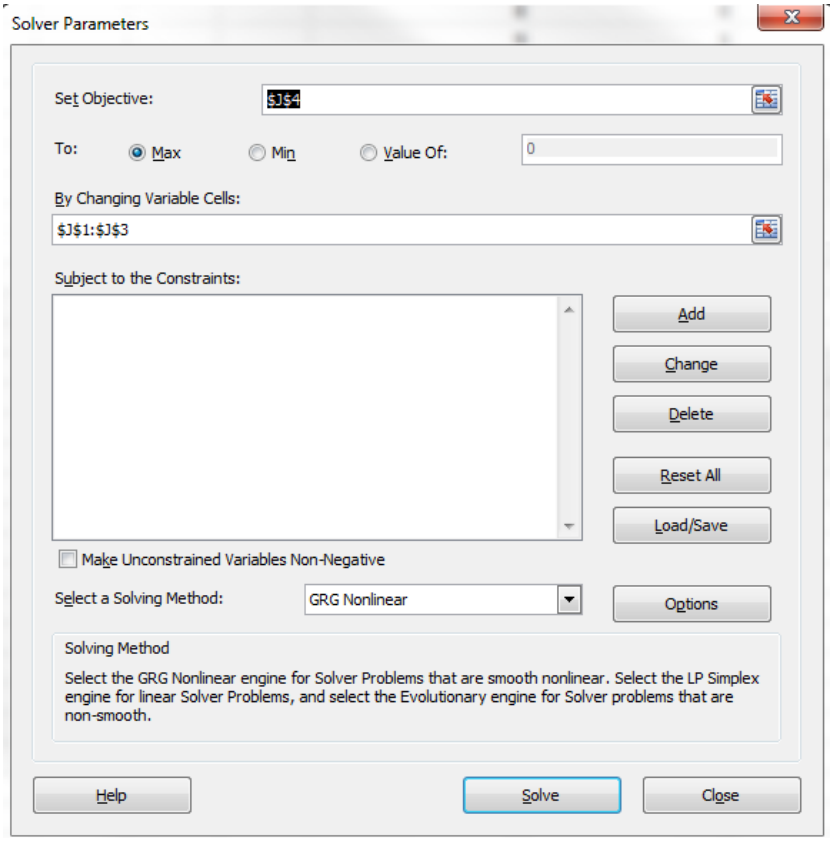
Similarly, in U2, the error from our model can be calculated as:

=AVERAGE(R2:r10001)

In cell U3, the percentage improvement from using our model can be calculated as:

=1-(U2/U1)

Having populated the worksheet, our goal is to now estimate the intercept and slope using Solver. Be sure that you have installed the Solver add-on on the Data tab. Using Solver, set the objective cell to I3. We will maximize this value by changing the values in cells J1, J2 and J3. Be sure to uncheck the box that constrains the coefficients. The Solver dialog box should appear as depicted below:



Click on Solve to estimate the intercept and slope corresponding to this regression.
========