Use the Equity.jmp data from the Sample Data Library for this exercise. A response (or Y) variable is BAD, which is coded as 0 (good credit risk) or 1 (bad credit risk). The other variables are:

LOAN The amount of the loan requested

MORTDUE How much the customer needs to pay on their mortgage

VALUE Assessed valuation

REASON Debt consolidation or home improvement (DebtCon or HomeImp)

JOB Broad job category

YOJ Years on the job

DEROG Number of derogatory reports

DELINQ The number of delinquent trade lines (or credit accounts)

CLAGE Age of oldest trade line (oldest credit account)

NINQ Number of recent credit inquiries

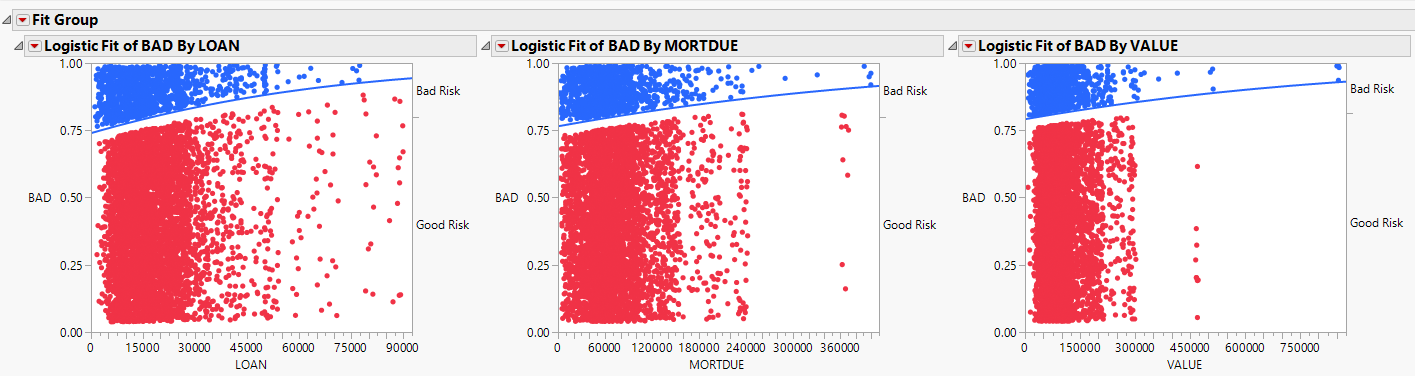
CLNO Number of trade lines

DEBTINC Debt to income ratio

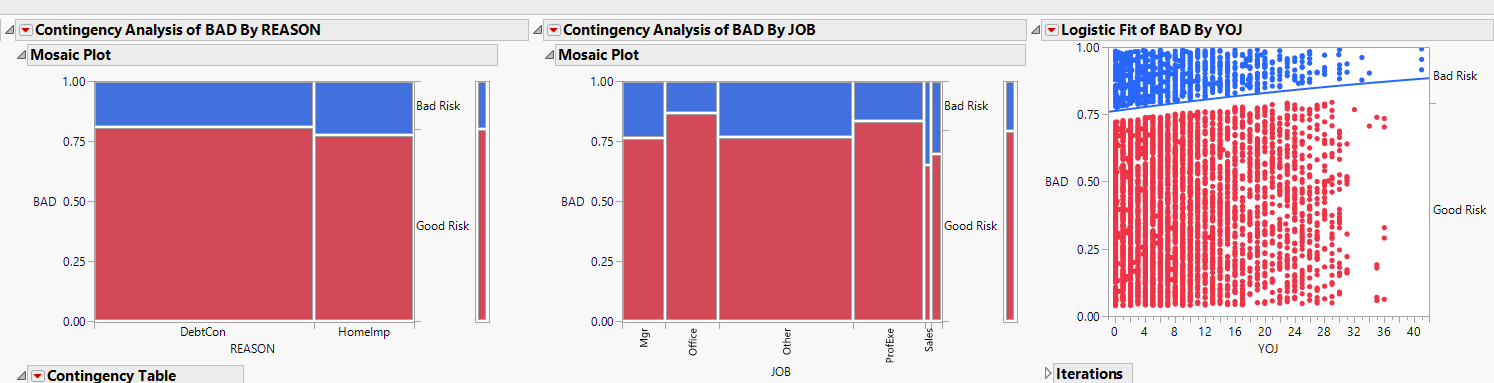
a. Use the Columns Viewer, Distribution and Graph Builder to familiarize yourself with this data. 1. Do any variables appear to be related to BAD? Explain.



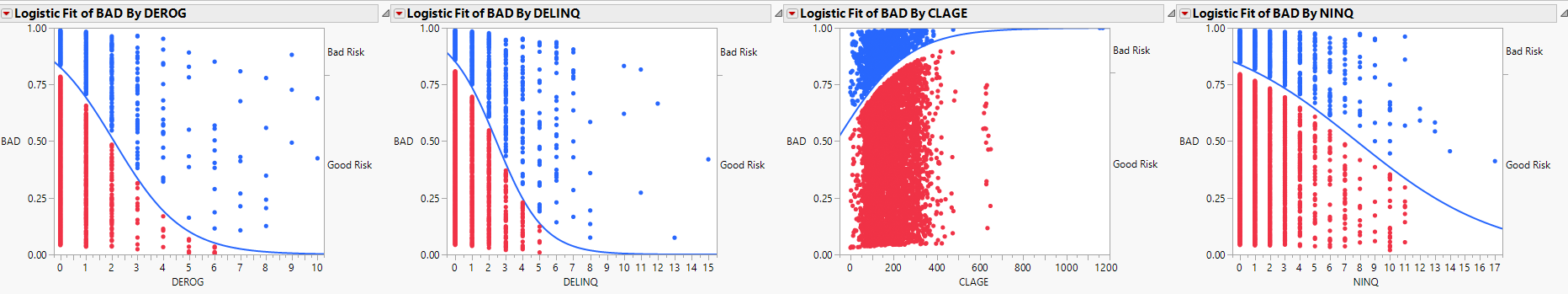
A: From the distributions, we can see that about 80.05% has good credit risk, only 19.95% has bad credit risk.



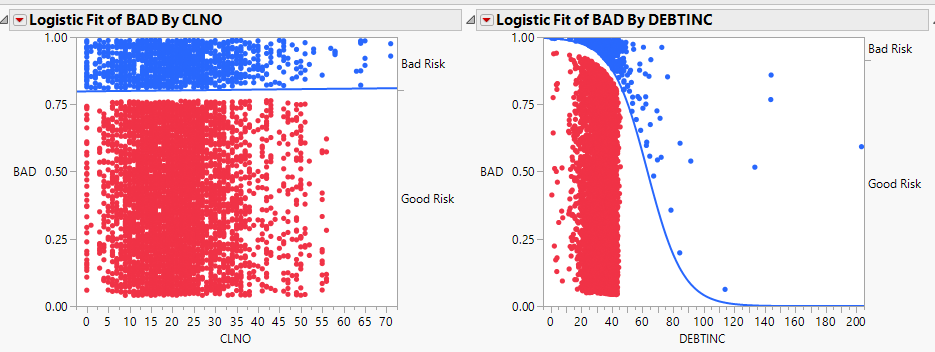
From the graph of Fit Group, we can see that LOAN, MORTDUE and VALUE are related to BAD, the good credit risks increases when these three variables go up.



We can the REASON, JOB and YOJ are also related to BAD. Reason doesn't matter much because the result is roughly the same, but you can see the differences between different jobs. Also, the good credit risk increases when YOJ goes up.



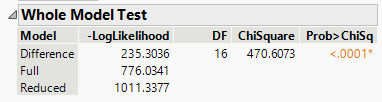
We can see that DEROG, DELINQ, CLAGE and NINQ are related to BAD, it can be seen from the graph that numerical changes have a significant impact on BAD.



From the last graph, we can see that DEBTINC is related to BAD, but CLNO is not related to BAD because the numerical change has no influence on BAD.

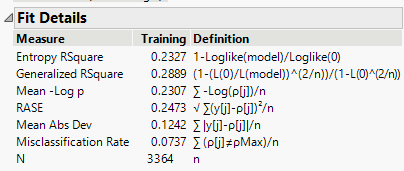
b. Fit a logistic regression model for BAD, including all predictor variables. Do not address data quality issues first (i.e., proceed with the data in its current form).

1. What is the p-value for the model?



A: The p-value for the model is <.0001, which means the model is highly significant.

2. What is the misclassification rate?

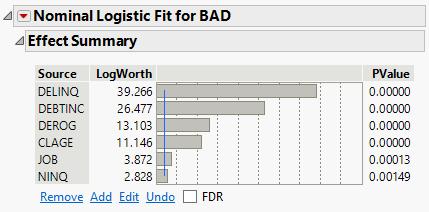


A: The misclassification Rate is 0.0737.

3. What are the two types of misclassification error that can occur in this example? How many misclassifications of each type were made?

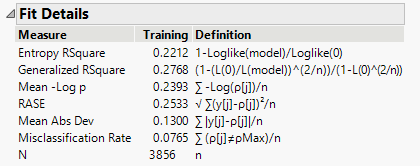
A: The misclassification rate indicates that the model misclassified 7.37% of the REASON or JOB.

4. Use the Effect Summary table to slowly remove non-significant terms from the model. How many terms are in your final model?



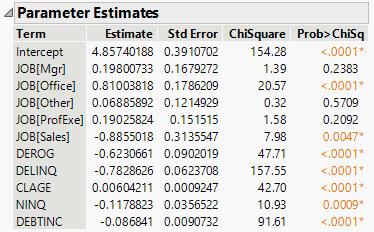
A: There are 6 terms in the final model, they are DELINQ, DEBTINC, DEROG, CLAGE, JOB and NINQ.

5. What is the misclassification rate for this reduced model?

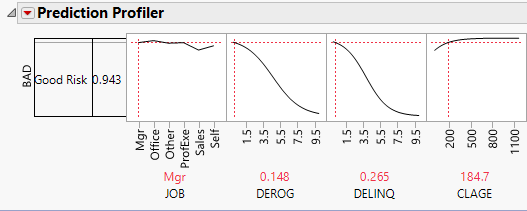


A: The misclassification rate for the new model is 0.0765.

6. What are estimates (coefficients) for DEROG and CLAGE? Open the Prediction Profiler to explore what happens to the predicted probability that BAD=1 as you increase and decrease the values of these two variables.



A: The estimates for DEROG is -0.6230661. Since the DEROG is negative, it indicates that the credit risk will become bad when the number of derogatory reports go up. The estimates for CLAGE is 0.00604211. Since CLAGE is positive, it indicates that older credit account has better credit risk than younger one.



This is more evident in the Prediction Profiler. When DEROG goes up, the possibility of good risk will be drop. When CLAGE goes up, the possibility of good risk will be increase.

7. You need to explain to your manager what coefficients for DEROG and CLAGE represent. Interpret the coefficients for these two variables in non-technical terms. You need to submit (1) your answers in doc and (2) zip file including all JMP results.

A: As I mention above, the credit risk will be bad as people have more derogatory reports. Also, older credit account has better credit risk than younger credit account.