Week 5 Exercises (Include screenshots in your responses when applicable)

**Book: Building Better Models with JMP Pro (Chapter 5)**

Use the **Equity.jmp** data from Blackboard for this exercise. This data set was first introduced in Week 2. Recall that the response variable is the variable **BAD**, where the value **1** indicates that the customer is a bad credit risk.

1. Use functionality like the Columns Viewer, Distribution, Graph Builder, and Multivariate (Correlation) to re-familiarize yourself with this data.
   1. Do any variables appear to be related to **BAD**? Explain in business language (it is not required to be technical in your response).
   2. List any potential data quality issues you observe. There is no requirement to apply fixes. Only share your observations and possible recommendations.
2. Fit a logistic regression model for **BAD**, including all predictor variables.
   1. What is the *p*-value for the whole model test?
   2. What is the misclassification rate?
   3. What are the two types of misclassification error that can occur in this example?

4. How many misclassifications of each type were made?

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

6. What is the misclassification rate for this reduced model (include screenshot)?

7. In the context of this example, define the two types of classification error: false positive and false negative. Which type of classification error occurred more often? What does this mean about our model? Explain.

8. What are the estimates (coefficients) for **DEROG** and **CLAGE** (include screenshot)? Ensure that you have applied Value Ordering to the dependent (target) variable, and the model is predicting the probability of an observation to be a “1”, not a “0”. In addition, open the Prediction Profiler, and explore what happens to the predicted probability that BAD=1 as you increase and decrease the values of these two variables.

9. Continue to leverage the Prediction Profiler. What is the probability of being a BAD Risk individual when DEROG has a value of 2, and CLAGE has a value of 100 (leave all other variables at their original settings).

10. You need to explain what the coefficients for **DEROG** and **CLAGE** represent to your manager. Share your interpretation of the coefficients for these two variables (in non-technical terms).

11. Save the Probability Formula for your model. Return back to your data table and observe the newly generated columns. Now focus on Row 20 in your table.

1. The current prediction for this individual is a Good or Bad risk?
2. Controlling for all other predictors, begin to change the value for DEROG (number of derogatory reports) on Row 20 within the table itself. How many derogatory reports does this individual need before we would predict them to be a Bad risk?