PREDICT 410: Predictive Modeling I Final Exam Study Guide

The final exam for PREDICT 410 will be administered in two parts: (1) a proctored exam administered through Canvas and monitored by ProctorU, and (2) an unproctored take-home exam administered through Canvas.

Exam Dates:

Check the pdf syllabus in your course shell for the exam open date. Both parts of the exam must be completed by the last day of the quarter – a Sunday by 11:55 pm Central Time.

Time Limits:

The proctored exam must be completed in one continuous two-hour period. The unproctored take-home exam must be completed in one continuous one hour period. The exam is designed to take one hour for the proctored exam and thirty minutes for the take-home exam, but we want to give everyone enough time to feel comfortable.

Exam Regulations:

Both parts of the exam are open-book and open-notes. In order to complete the exam you will need a scientific calculator, Excel, Python, or R to perform small computations. Students using eBooks must have the books downloaded to the machine that they are using to take the exam. ProctorU will not allow you to use a reader. It is recommended that you use a text editor such as Notepad to write and save your answers, and then paste your answers into Canvas.

Students should note that even though the test is open book and open notes, if you do not study and rely on looking up each question, then you will not have enough time to complete the exam. Students should prepare as though the test is closed-book and closed-notes.

Exam Format:

The take-home portion of the exam consists of 10 questions worth 5 points each for a total of 50 points. The 10 questions are multiple choice and administered in Canvas in the same format as your quizzes.

The proctored portion of the exam is 4 questions totaling 50 points. The questions are short answer essay and small computation similar to some of our discussion boards and Assignment #4. There is no SAS coding component in either part this exam.

It is recommended that you study for the exam using this study guide, then take the take home part of the final exam first, and then take the proctored portion of the exam. This sequence would make you optimally prepared for the proctored part of the exam.

Here are some topics that you should understand.

1) The core focus of this course consists of linear regression (ordinary least squares regression) and topics from multivariate analysis (factor analysis, principal components analysis, and cluster analysis). You should understand the underlying assumptions of each model or method, how to specify or apply each model or method, and how to interpret the results from each model or method.

2) You should understand the concept of multicollinearity. How do we detect multicollinearity, how does it affect our regression results, and how do we mitigate multicollinearity?

3) When developing predictive models, we typically build and compare alternative models. What statistical metrics can be used to compare the alternative models? What statistical tests can be used to compare models? Which metrics are only valid when comparing nested models and which metrics can be used to compare any set of models? You should understand the underlying assumptions of these metrics and tests and be able to interpret these metrics and tests. Metrics of particular interest include: mean error, mean absolute error, mean square error, R-squared, adjusted R-squared, Mallow's Cp, AIC, and BIC.

4) You should understand the similarities and differences between factor analysis and principal components analysis, and how could you apply them in conjunction with cluster analysis. You should understand how to select the number of components or factors to retain, why we make factor rotations, and the different types of factor rotations available (in particular you should be comfortable with the Varimax rotation).

5) You should understand in complete detail how forward and backward variable selection work in the context of OLS regression. How does each method work at each step and what statistical test is being performed at each step? (In particular you should understand how the F-test is used to test nested models.)

6) You should understand how to define, use, and interpret dummy variables as a means of coding categorical variables for use in a regression model.

7) You need to be able to perform all of the computations in Assignment #4.