Homework 2 – College Football Revenue and Expenses

Data Reading and manipulation

A1. Create two new variables. First, create “total\_enroll” which is equal to male and female enrollment combined. Second, create “percent\_male” which is equal to the percentage of male students (example: 50%=0.5).

A2. What is the mean, median and standard deviation of “total\_enroll” and “percent\_male”

A3. What is the correlation between expenses and revenues?

A4. Create a new variable “percent\_female” equal to 1-“percent\_male”.

A5. Create year dummy variables.

Regression Analytics

B1. What’s the R-squared of a simple regression with total\_expense\_all\_football\_h as the dependent variable (Y) and the lagged expenses as the only independent variable (X)? What does the R-squared statistic mean here? Is the lagged expenses statistically significant? Is there any evidence for a random walk?

B2. Run a simple regression with total\_expense\_all\_football\_h as the dependent variable (Y) and use three independent variables(X): the lagged expenses, “percent\_male” and “efmalecount\_h”. Are the “male” variables statistically significant? What are the “male” coefficients? What problem are you possibly running into and why?

B3. Run a simple regression with total\_expense\_all\_football\_h as the dependent variable (Y) and use total\_revenue\_all\_football\_h as the only independent variable (X). How does the R-squared compare to question B1? Is the coefficient on revenue statistically significant? What problem are you possibly running into and why?

B4. How could you solve the problem in B3 with the data that is already included in the dataset?

B5. Run a simple regression with total\_net\_all\_football\_h as the dependent variable and include the lagged net\_all\_football and “percent\_female” as independent variables. What is the sign and significance of “percent\_female”? What does the coefficient on “percent\_female” imply and what problem are you possibly running into and why?

B6. Run a regression with total\_expense\_all\_football\_h as the dependent variable and use only year dummy variables and conference dummy variables as your independent variables. How does the R-squared compare to question B1? What’s interesting (or not) about this particular regression formation?

Data Mining

C1. Do your best. Forecast total\_expense\_all\_football\_h using any of the information here and any combination/transformation of the data you desire.