WORKSHEET #1: Regression Basics, New Variables

First: Tutorial

Worksheet: Using CollegeFootball_Rev_Exp from Sample Data Sets.

1. Create a new dummy variable “PowerFive” which is equal to one when the program belongs to a Power 5 conference: SEC, Big 10, Big 12, ACC, Pac 10.
2. Run a regression where the Y variable is “total_net_all_football_h” and there are two X variables: “total_net_all_football_l1_h” and “PowerFive”.
3. Describe the results from the regression in #2. What is the R-squared? Does this model fit the data well? Are the coefficients statistically significant? How many millions of dollars in total net revenues is it to be in a Power 5 conference?
4. (Bonus) Run a regression where the Y variable is “total_net_all_football_h” and there are three X variables: “total_net_all_football_l1_h”, “PowerFive” and “year”.
5. (Bonus) Comparing the regression in #2 with #4… what happened to the intercept? Why? Is the coefficient on “year” significant? What does the coefficient suggest is happening over time?

WORKSHEET #2: Regression Basics, Too Many X's

First: Tutorial

Worksheet: Using OremOwlz from Sample Data Sets.

1. Run a regression where the Y variable is "GA/Berm" and there are 26 X variables: All the promos from Pen_Schedule_Giveaway to Food_Bank_Night. What happened and why?
2. Run a regression where the Y variable is "GA/Berm" and there are 7 X variables: All the Days of the Week from Sunday to Saturday. What happened and why?
3. Create a new dummy variable "Weekend" which is equal to 1 when it is Saturday or Sunday.
4. Run a regression where the Y variable is "GA/Berm" and there is one X variable: "Weekend". What does the coefficient suggest? Is it statistically significant?
5. (Bonus) Create a new variable "Weekend_GA" which is equal "GA/Berm" ONLY IF "Weekend" = 1. Create a new variable "Weekday_GA" which is equal to "GA/Berm" ONLY IF "Weekend" = 0.
6. (Bonus) Run a regression where the Y variable is "Total_Concessions" and there are 2 X variables: "GA/Berm" and "Weekend". Interpret your coefficients.
7. (Bonus) Run a regression where the Y variable is "Total_Concessions" and there are 3 X variables: "Weekend_GA", "Weekday_GA", and "Weekend". Interpret your coefficients. What happened and why?

WORKSHEET #3: Regression Basics, Controls

First: Tutorial

Worksheet: Using OremOwlz from Sample Data Sets.

1. Create a new dummy variable “HomeWin” which is equal to 1 when the home team (Orem) wins the game.
2. Run a regression where the Y variable is “HomeWin” and there are 7 X variables: 6 of the 7 opponents from Ogden to Billings (remember to leave a null variable) AND “Announced” Attendance.
3. What role are the Dummy Variables of the opponents playing in this regression?
4. Interpret the coefficient on “Announced”. Is there a home field advantage from fan attendance?
5. (Bonus) Create a new variable “PreGame_Record” which is equal to the current winning percentage of the Owlz. Note that “Overall_Record” includes the current row/game so you’ll need to remove the current row/game to accurately create “PreGame_Record”.
6. (Bonus) Run a regression where the Y variable is “HomeWin” and there are 8 X variables: 6 of the 7 opponents from Ogden to Billings (remember to leave a null variable) AND “Announced” Attendance AND “PreGame_Record”. Interpret your coefficients.

WORKSHEET #4: Regression Basics, Autoregressive Controls

First: Tutorial

Worksheet: Using CollegeFootball_Rev_Exp from Sample Data Sets.

1. Run a regression where the Y variable is “total_net_all_football_h” and there is one variable: “total_net_all_football_l1_h”

2. Run a regression where the Y variable is “total_rev_all_football_h” and there is one variable: “total_rev_all_football_l1_h”

3. Run a regression where the Y variable is “total_exp_all_football_h” and there is one variable: “total_exp_all_football_l1_h”

4. Compare and contrast the three different autoregressive coefficients. Are they significant? What are the R-squareds? Any unit roots?

WORKSHEET #5

Worksheet: Using CF_Line_Outcome from sample Data Sets

1. Create a new variable “off_score_diff” which is equal to fso1/uso1. Create a new variable “def_score_diff” which is equal to fsd1/usd1.
2. Run a regression where the Y variable is “line” and there are six X variables: fso1, uso1, fsd1, usd1, “off_score_diff” and “def_score_diff” What do your results suggest? Are there any possible problems? How does an increase in fso1 impact Y?
3. Run a regression where the Y variable is “line” and there are two X variables: “off_score_diff” and “def_score_diff”. Compare and contrast your results to the previous regression.
4. Run a regression where the Y variable is “line” and there is one X variable: “fminusu”. What do your results suggest? What possible problem is here?

WORKSHEET #6

Worksheet: Using HS_QBs from sample Data Sets

1. Delete all observations where forty2 OR seasons OR att OR ma5_wins_school is missing.
2. Create a new variable “Likely_grad” which is equal to 1 if “seasons” is greater than or equal to 4.
3. Run a regression where the Y variable is “Likely_grad” and there are three X variables: scouts_stars, height2 and forty2. WHEN YOU RUN THE REGRESSION -SELECT “RESIDUALS” What do your results suggest? Interpret your coefficients. What does the Predicted Y mean? What do your Residuals mean? What’s Predicted Y + Residuals equal to?
4. Run a regression where the Y variable is “Likely_grad” and there are six X variables: scouts_stars, height2, forty2, att, transferred_b and ma5_wins_school, . WHEN YOU RUN THE REGRESSION - SELECT “RESIDUALS” What do your results suggest? Interpret your coefficients. What does the Predicted Y mean? What do your Residuals mean? What’s Predicted Y + Residuals equal to?