

OR 603 Sports Analytics

Assignment 6

1. Repeat the full correlation and regression analysis demonstrated in the lesson notes for NFL running and passing. Feel free to copy the script found in the notes. Upload your R script as “NFL <your last name>.r”

2. Use the *cricket batsman stats.csv* file to explore the relationships between a player’s runs scored (Runs) and the number of ducks (0s), 4s, and 6s. Ducks are when a batsman is out before he scores a run. A 4 is similar to a ground rule double in baseball, and counts as 4 runs. A 6 is similar to a home run in baseball and counts as 6 runs.

a. Visualize the relationship between Runs and 4s and between Runs and 6s using the plot function.

HINT: R does not allow the names of vectors to begin with a number. When you upload a data file with column names that start with a number, like 4s, it will append an X to the beginning to make the column name X4s.

b. Generate the correlation matrix for Runs, 0s, 4s and 6s.

c. Create and interpret a linear regression model for Runs as a function of 0s, 4s, and 6s. Which variables are significant? If all you knew about a batsman was that he had 0 ducks, 10 4s, and 5 6s, how many runs would he be expected to have? Write your answers within your script using the # symbol to create comment lines.

Upload your R script as “cricket <your last name>.r”

3. BONUS QUESTION: There are additional variables included in the NFL data file. Explore and model them on your own. How do they relate to winning? What can you learn? Add your scripting to the NFL r script file and embed any insights you can gain with a comment line (beginning with #).

OINTRATE = offensive interception rate (per pass play)

DINTRATE = defensive interception rate (per pass play)

OFUMRATE = offensive fumble rate (per play)

DFUMRATE = defensive fumble rate (per play)

PENRATE = team penalty yards per play