**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Week 10 Lab: CORRELATION & REGRESSION**

When analyzing who to start, and who to bench, fantasy football managers often consider each player’s matchup. For example, a decision between Ahmad Bradshaw & Zac Stacy might depend on the fact that Bradshaw will be playing against the Giants, and Stacy is playing against the 49ers. Differences in the quality of these opponents’ defenses might be a good factor when making a decision about whether to start Bradshaw or Stacy.

|  |  |
| --- | --- |
| opponent | Avg of fantasyPoints |
| ARI | 14.48 |
| ATL | 17.31 |
| BAL | 15.93 |
| BUF | 2.73 |
| CAR | 12.49 |
| CHI | 19.61 |
| CIN | 12.63 |
| CLE | 6.96 |
| DAL | 20.29 |
| DEN | 13.37 |
| DET | 18.35 |
| GB | 15.02 |
| HOU | 16.38 |
| IND | 24.91 |
| JAX | 9.57 |
| KC | 11.35 |
| MIA | 9.77 |
| MIN | 12.68 |
| NE | 14.40 |
| NO | 13.07 |
| NYG | 21.66 |
| NYJ | 24.28 |
| OAK | 15.61 |
| PHI | 21.16 |
| PIT | 22.34 |
| SD | 19.59 |
| SEA | 9.62 |
| SF | 14.31 |
| STL | 23.28 |
| TB | 13.94 |
| TEN | 8.42 |
| WSH | 10.71 |

But there are a lot of ways that you might characterize the quality of a team’s defense. For today’s lab, let’s find the defensive variable that’s most strongly related to a player’s future performance (in fantasy points) against that defense.

Let’s start with Quarterbacks.

First, make a column that represents the average fantasy points scored by Quarterbacks against each team during weeks 7 & 8 of the current NFL season. You can use PivotTables to do this. Your column should look something like the table on the right.

Now create new pivot tables with similar columns-- with variables that might be *related* to QB points against. Here’re three that you must use:

|  |  |  |
| --- | --- | --- |
| # | Variable | r-value |
| 1 | Cumulative fantasy points scored by quarterbacks against each team during weeks 1-6 (combined) of the current season. |  |
| 2 | Cumulative receiving yards against each team during weeks 1-6 (combined) of the current season. |  |
| 3 | Cumulative fantasy points scored by each team’s defense (DST) during weeks 1-6 (combined) of the current season. |  |

And choose one more variable that you think might be related to QB points scored against each team…

|  |  |  |
| --- | --- | --- |
| # | Variable | r-value |
| 4 |  |  |

Use the =CORREL() function in Excel to determine the r-values of each variable’s association with the Weeks 7 & 8 QB points scored against each team, and fill-in the tables accordingly.

Now let’s try some actual modeling.

|  |  |
| --- | --- |
| opponent | Sum of fantasyPoints |
| ARI | 12.10 |
| ATL | 17.90 |
| BAL | 8.30 |
| BUF | 10.10 |
| CAR | 23.50 |
| CHI | 20.40 |
| CIN | 34.30 |
| CLE | 28.90 |
| DAL | 8.40 |
| DEN | 4.30 |
| DET | 11.60 |
| GB | 7.90 |
| HOU | 22.90 |
| IND | 3.60 |
| JAX | 6.80 |
| KC | 8.40 |
| MIA | 22.90 |
| MIN | 15.00 |
| NE | 28.40 |
| NO | 15.90 |
| NYG | 22.90 |
| NYJ | 25.00 |
| OAK | 33.90 |
| PIT | 24.00 |
| SD | 22.90 |
| SEA | 25.40 |
| SF | 26.20 |
| STL | 10.70 |
| TEN | 7.40 |
| WSH | 7.60 |

Big numbers come from running backs who score touchdowns, and a team that allows rushing touchdowns would be a favorable matchup for any RB. We’re going to try to make a regression equation that will predict running back fantasy points using just one predictor variable: the total number of rushing touchdowns allowed by an opponent during the previous two weeks.

First, make a column kinda like the previous exercise, where you list the average running back fantasy points scored against each team in week 7 of the current NFL season. These are the “y” values, the values you’re trying to predict. *Note that there aren't any values for the Eagles and the 49ers, who were on bye.*

Now make a column that represents the total number of rushing yards scored against each team in the first 6 weeks of the season. These are your “x” values, the predictor variable.

You’ll need to enable Excel’s Data Analysis Toolpak to do this exercise… Do a quick Google search (try: enable data analysis toolpak), and follow the instructions. Raise your hand if you have problems with it.

Click on the Data tab, and on the far right, you should now see a button for “Data Analysis”—click on it, then select “Regression” and click “OK.” For the x-values, select the column that represents the total sum of rushing yards scored against each team for the first six weeks. For the y-values, select the running back fantasy points scored against each team from week 7.

Then click OK.

In the output worksheet, identify the following:

1. The correlation coefficient (called   
   “multiple r”) that represents the   
   relationship between these two variables: \_\_\_\_\_\_\_\_\_\_\_\_
2. The p-value of this relationship: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. The equation of the regression line:   
    ŷ = \_\_\_\_\_ + ( \_\_\_\_\_ \* x)