**Offense or Defense?**

Let’s look at offensive and defensive statistics for National Football League [teams from the 2021 season](https://docs.google.com/spreadsheets/d/17jbo_tfe8OCM4VX-4Q0d9-EqEJIGy9nFFjLjc2JtY48/edit?usp=sharing), shown in the table below. What variable does a better job at predicting a team’s winning percentage (PCT): the number of points an offense scores (PF = points for) or the number of points a defense allows (PA = points against)?

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Team** | 49ers | Bears | Bengals | Bills | Broncos | Browns | Buccaneers | Cardinals | Chargers | Chiefs | Colts | Cowboys | Dolphins | Eagles | Falcons | Football Team | Giants |
| **PF** | 427 | 311 | 460 | 483 | 335 | 349 | 511 | 449 | 474 | 480 | 451 | 530 | 341 | 444 | 313 | 335 | 258 |
| **PA** | 365 | 407 | 376 | 289 | 322 | 371 | 353 | 366 | 459 | 364 | 365 | 358 | 373 | 385 | 459 | 434 | 416 |
| **PCT** | 59 | 35 | 59 | 65 | 41 | 47 | 77 | 65 | 53 | 71 | 53 | 71 | 53 | 53 | 41 | 41 | 24 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Team** | Jaguars | Jets | Lions | Packers | Panthers | Patriots | Raiders | Rams | Ravens | Saints | Seahawks | Steelers | Texans | Titans | Vikings |
| **PF** | 253 | 310 | 325 | 450 | 304 | 462 | 374 | 460 | 387 | 364 | 395 | 343 | 280 | 419 | 425 |
| **PA** | 457 | 504 | 467 | 371 | 404 | 303 | 439 | 372 | 392 | 335 | 366 | 398 | 452 | 354 | 426 |
| **PCT** | 18 | 24 | 21 | 77 | 29 | 59 | 59 | 71 | 47 | 53 | 41 | 56 | 24 | 71 | 47 |

1. The scatterplots below show the association between a team’s winning percentage with either points for (PF) or points against (PA). Based on the scatterplots, which explanatory variable – PF or PA – would you guess will do a better job at predicting a team’s winning percentage?



2. On stapplet.com, select the *Multiple Regression* applet. Input PF as the first explanatory variable, PA as the second explanatory variable, and PCT as the response variable. **Be sure that the only box selected with “included in model” is PF**. Write the equation of the LSRL using PF and record the value of R2 and S.

LSRL: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ R2: \_\_\_\_\_\_\_\_\_\_ S: \_\_\_\_\_\_\_\_\_

3. Using the LSRL, calculate the residual for the San Francisco 49ers, with 427 points for (PF) and a winning percentage (PCT) of 59 percent.

4. Go to “edit inputs” and deselect the box next to PF; select the box next to PA (now only PA is “included in model”). Write the equation of the LSRL using PA and state the value of R2 and S.

LSRL: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ R2: \_\_\_\_\_\_\_\_\_\_ S: \_\_\_\_\_\_\_\_\_

5. Using this new LSRL, calculate the residual for the San Francisco 49ers, with 365 points against (PA) and a winning percentage (PCT) of 59 percent.

Rather than using just one explanatory variable at a time, what if we used both PF and PA *in the same model*? Would this improve our predictions? Select “Edit inputs” and click both PF and PA to be included in the model. Begin analysis!

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictor** | **Coef** | **SE Coef** | **T** | **P** |
| Constant |  | 20.618 | 0.799 | 0.431 |
| PF |  | 0.023 | 7.002 | <0.001 |
| PA |  | 0.036 | -2.159 | 0.039 |

6. You should see regression output like the table to the right. Fill in the coefficient boxes, and write the equation of the multiple regression model, in the form:

*Predicted PCT = Constant + (coef) PF + (coef) PA*

7. Using this new multiple regression model, calculate the residual for the 49ers, with a winning percentage 59 percent, 427 points for, and 365 points against.

8. What was the value of R2 and S for this multiple regression model? R2: \_\_\_\_\_\_\_\_ S: \_\_\_\_\_\_\_\_\_

9. Which of the three models did the best at predicting winning percentage among these NFL teams? Explain.

10. What is a variable that may increase the value of R2 in our model? Why do you think so?

11. What is a variable that would not increase the value of R2 in our model? Why do you think so?

Multiple Regression

QuickNotes

Check Your Understanding

Here is a multiple regression model for predicting y = long jump distance (in inches) using x1 = 40-yard dash time (in seconds) and x2 = grade level (input 1 for junior or senior; input 0 for freshmen or sophomore) for a sample of students:

$$\hat{y}=293.56-31.05x\_{1}+42.02x\_{2}$$

1. Predict the long-jump distance for a senior student who had a dash time of 5.41 seconds.
2. The student in part (a) had a long jump distance of 171 inches. Calculate and interpret the residual.