Baseball is a popular American sport played on a diamond-shaped field. Games are 9

innings long and each inning has two halves, the first in which the visiting team bats and the second where the home team bats. Innings end after three outs. An out is when a player from the hitting team is removed from play for the half of the inning due to various reasons. Batters aim to get on base by hitting a ball pitched to them by the pitcher. Batters can get to first, second, or third base depending on how far they hit the ball and how fast they run. If a batter hits the ball past the outfield fences, they, along with any runners on base, automatically score, this is called a home run. Runners can also score if another player hits the ball and then they reach home. The team with the most runs wins the game.

Behind the catcher, at home plate is an official known as the home plate umpire. The umpire's role is to enforce the rules and make decisions during a game. Many of these decisions involve calling balls and strikes. Pitches that are considered strikes are pitches within the strike zone which goes from the batter’s knees to his shoulders. Anything outside of that zone is called a ball. If a batter gets 3 strikes, they are out on a strike out. If the batter gets 4 balls, they get to go to first base on what is called a walk.

Major League Baseball (MLB) is a professional baseball league with 30 teams and a 162-game season. The MLB has 76 umpires in total with four umpires in each game. Umpires are stationed at 1st, 2nd, and 3rd base in addition to home plate, but the home plate umpire is the only one who makes calls on pitches. The <umpires.csv> dataset looks at cumulative data from MLB homeplate umpires dating as far back as 2008. The boost statistics in the dataset investigate how certain umpires compare to the "average" Major League Baseball umpire. The dataset provides insight on if umpires favor defensive players or offensive players more.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| *Games* | The number of games the umpire has umpired since 2008. |
| *k\_pct* |  |
| *bb\_pct* |  |
| *RPG* |  |
| *BA* |  |

1. Fit a model to predict *RPG* with *Games*, *k\_pct*, *BA*, and *bb\_pct*. Write down the model.
2. Predict the *RPG* for an umpire that has umpired 255 games, with a *BA* for hitters of 0.270, a *k\_pct* of 15%, and a *bb\_pct* of 10%.
3. Obtain the p-values for each of the predictors in the model. Are all of the predictors effective? If not, which ones and why?
4. Fill in the ANOVA table for the model and perform a test for the overall fit of the model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **d.f.** | **Sum of Squares** | **Mean Square** | **F-value** | **P-value** |
| **Model** |  |  |  |  |  |
| **Error** |  |  |  |
| **Total** |  |  |

1. Record and comment on the multiple R2 and adjusted R2 values for the model.

**Multiple R2:**

**Adjusted R2:**

1. Drop *bb\_pct* from the model and record the new multiple R2 and adjusted R2 values. Is the model without *bb\_pct* better or worse?

**Multiple R2:**

**Adjusted R2:**

1. Find the VIFs for the original model, are any of the values concerning? Explain why or why not.
2. What would it mean if a predictor had a poor VIF value?
3. What would you do with a predictor with too high a VIF?
4. Find and interpret the 95% confidence interval for the mean *RPG* for umpires that have umpired 400 games, with a *BA* for hitters of 0.215, a *k\_pct* of 18%, and a *bb\_pct* of 9%.
5. Find and interpret the 95% prediction interval for the *RPG* of an umpire that has umpired 400 games, with a *BA* for hitters of 0.215, a *k\_pct* of 18%, and a *bb\_pct* of 9%.