Pitch Effort Injury Analysis

Matthew Rice

DSC 530-T302

Professor Matthew Metzger

**Outcome of your EDA**

The analysis showed confirmed my hypothesis that the amount of effort exerted by Major League Baseball pitchers, measured by how often they throw their fastballs at 95 – 100 percent of their maximum velocity, has a statistically significant impact on the number of days those players missed due to injury.

**What do you feel was missed during the analysis?**

In addition to the velocity, spin rate is a variable that I believe could be used to more accurately depict “effort” from pitchers. The EDA showed many missing spin rate values. This could be because spin rate is a metric that began to be measured in recent years. Equipment to measure spin rate may not be available at all games, making it a difficult variable to obtain robust and accurate information on.

**Were there any variables you felt could have helped in the analysis?**

Many factors can go into how often an athlete finds themselves out of action with an injury. Other variables that could go into building a model to better predict days missed due to injury include: biomechanics (arm angle at release, hip rotation, etc.), number of pitches thrown daily (both in games and in bullpen sessions), a player’s injury history, and environmental factors such as weather at the time of injury.

**Were there any assumptions made you felt were incorrect?**

I assumed that there would be a stronger linear correlation between the “percentage\_above\_95” variable and the amount of time missed. The EDA confirmed a statistically significant correlation between the two, but the analysis showed the relationship is more complex than I expected and more complex models with the additional variables will be necessary for accurate predictions.

**What challenges did you face, what did you not fully understand?**

Visualizing and analyzing different types of correlation is something I could use more practice on. The variables I compared are measured in different units, and Think Stats recommends transforming values to a standard score or to a value’s rank in this situation. Performing those transformations and plotting them was a concept I found challenging.