Anthony Testa and Ian Weihe

April 11th, 2024

## Executive Summary: Examining Eras

Baseball has been defined by seven distinct eras starting in 1901 all the way to the present. Using statistical analysis we examine the differences and trends between eras specifically regarding teams. We also investigate what makes winning teams to see if there are unique factors within the different eras.

Our data was collected from Data.World which contains team data from 1871-2015. We revised our data to just include the years 1901-2015 for two reasons. The first being that the eras started being defined in 1901 with the deadball era. The second is due to some teams playing significantly fewer games in the 1800s and we wanted to have all teams play the same amount of games.

The first statistical analysis we used was a Kruskal Wallis test to examine the differences between eras. We inspected the variables hits/game, runs/game, ERA which all looked extremely similar between the eras. We also examined home runs, steals/game, shutouts/game and strikeouts/game to further understand the differences between eras. When performing Kruskal Wallis we found a significant difference with at least one era being different from the others in each variable. To be more specific we used the Kruskal MC command to examine the differences between each era.

We then examined some trends that aren't linked to specific eras, but have a general trend over time. Strikeouts and saves have gone up over time, while complete games have become less common over time. Additionally, fielding has gotten better over time as evidenced by trends in errors and fielding percentage.

Finally, we used MARS to determine which of our variables most impacted win percentage in each era. Across all eras, Runs and Runs Allowed were the most important factors, which isn't surprising. But, by examining variable importance plots we were able to determine what other factors were the most important by era. No two eras were alike, both in which variables were included and in which variables proved to be the most important. From the deadball era, where park factors were the most important, to the modern era, where many different factors equally affect winning, and every period in between, we were able to distinctly separate the different characteristics of each era in baseball.