**Project 1 – Analyzing World Series Winners**

The purpose of this project was to explore data using major league baseball statistics as a reference. The question to be answered is as follows: is offense, fielding or pitching more important in determining the outcome of a world series winner. To answer this question, the metrics used include wins-above replacement levels to provide a measure of comparison to the rest of the league and between categories. Information was pulled from baseball-reference.com and uses wins-above-replacement (WAR) among world series winners since 1980. The premise behind WAR is calculating the contribution to wins above what a replacement player would have theoretically achieved. It factors in batting, pitching, defense and baserunning statistics and can be aggregated to reflect a team’s WAR broken down as offensive, pitching, or defensive WAR. A formula from fangraphs.com provides offensive WAR calculations as follows: WAR = (Batting Runs + Base Running Runs + Positional Adjustment + League Adjustment +Replacement Runs) / (Runs Per Win). Pitching WAR is defined as WAR = [[([(League “FIP” – “FIP”) / Pitcher Specific Runs Per Win] + Replacement Level) \* (IP/9)] \* Leverage Multiplier for Relievers] + League Correction. Defensive WAR pertains to Fielding Runs above replacement.

The data summary provides the mean, median and standard deviation of all three categories. It also lists the frequency that world series winners appear in the top quartile and top 5 teams in the league. The results are consolidated in Figures 1 and 2. In absolute terms the offensive and pitching WAR statistics are similar. The world series winner appears in the top quartile and among the top 5 teams in terms of offensive WAR 23 and 22 times respectively out of 43 years of data. Pitching was similar with values of 24 and 19 respectively. Less significant were the values for defensive WAR which were 15 and 11 respectively. The data shows there to be a relative importance to pitching and hitting over fielding, but not necessarily one over the other.

Not shown in the data is an overlap that occurred 9 times in which teams were top 5 in both pitching and hitting. Twice were teams top 5 in all three categories (the 2020 Dodgers and 2016 Cubs). General performance and training changes along with the advent of the wild card in 1994 did not seem to affect the data over the course of 43 years. Figure 1 shows a consistent distribution of world series winners ranked in the upper half relative to the rest of the league in all three WAR categories.

Since WAR is an aggregated measurement of many factors, one could attempt to select specific categories in offense, pitching and defense to parse down the specific contributors to world series wins. For example, one could compare slugging percentage against earned runs average. A significant category in fielding might include double plays turned. With only a few years of data, some of these trends could be altered by the inclusion of more teams in the post season.