An Investigation of High School and College Drafted Players Immediate Value in the MLB

Micheas S. Yimam

Math Major at The College of Wooster Class of 2019

Author Note

An extension of a Math Modeling Project Looking at Predicting MVP candidates.

Abstract

Determining a player's value in the Major Leagues is a highly debated topic. Projecting amateur talent is a nearly impossible task, where even top tier prospects fail all the time. This paper will attempt to scratch the surface in connecting Major League success with the Amateur Draft. I will use a project where I created a statistic model alongside Kory Sansom (Class of 2019) in predicting MVP candidates by using nine explanatory variables. Our model returned over eighty percent accuracy. Since this is an extension of the above paper we are only examining position players for this investigation. From there I will extend the project to look at player's immediate

value added to their team. By utilizing my stepwise regression from my old model, I can analyze player's first six seasons in the league to find their predicted MVP votes, which here will be used as a rank value to be compared among other players drafted from college or high school. The five questions that will be the focus of the paper are: Whether players drafted out of high school or college are more productive for Major League teams immediately? Are High School draftees more prone for polarizing results than college players; meaning are these amateurs more likely to be below average players or elite level players and not fall in the middle? Next a look at team tendencies in the draft, are certain teams better at drafting a certain position, or drafting more average major league players and not All-Stars. Fourth, when looking at the Elite players in each draft out of high school or college analyzing scouting reports to find similarities. Finally, looking at below average player's drafted from either high school or college and reporting the similar trends from their scouting reports. I will investigate players drafted from 2000-2012 and use Major league statistics up to the 2017 season. Once my sample size is gathered I will break up the data in three sections: the first being elite players making up the top 20%, the second encompassing the average players which are about 60%, and third which include below average players another 20% of players. From here this grouping will allow for the introduction of ANOVA testing. With the ANOVA testing I can further the analysis of the significance of players being drafted out of college or high school as indicators for performance in the Major Leagues. This investigation will be the first part of a series of analysis to better understand how drafted players will perform in the Major Leagues.