MCS100/Stats50 Final Project - Extended Abstract
Antonino Abundes & Alec Powell
December 2, 2014

The Effect of Rookie Age and College Experience on Major League Baseball Career Performance (1985-2010)

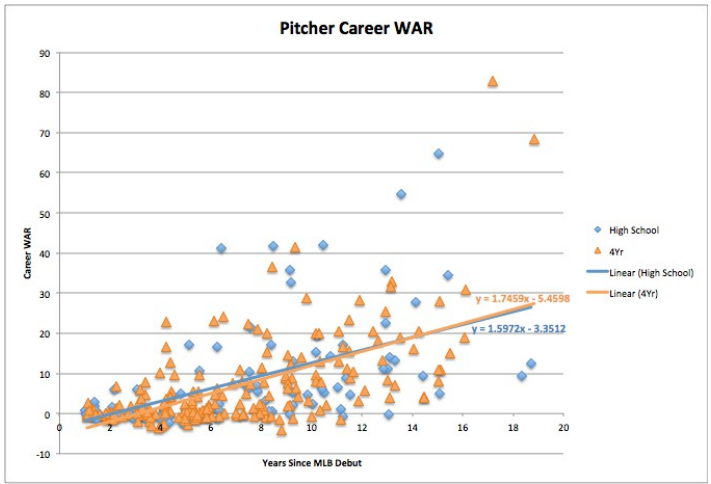
Introduction

Does going to college have an impact on a baseball player's future career performance in the big leagues? Is it worth calling up a player at a younger age so that they can gain experience at the sport's highest level? The answers to these questions could affect managers' decisions about who to draft and when to bring them up to the Show.

Front offices are always confronted with the question of whether their team is better off by drafting players with college experience or drafting players with high upside coming straight out of high school. Excellent players will always have excellent stats in high school, but that cannot be used as a direct correlation with their performance at the professional level. Players that have proven themselves in college as well as high school are more likely to maintain their level of play considering they have played against tougher competition. However, players drafted out of high school are younger and therefore will be exposed to baseball's elite for a longer period of time before hitting their prime.

Procedure

In this project, we analyzed 1st round draft data from the MLB June Amateur Draft for the years 1985-2010 and correlated their career Wins Above Replacement (WAR) to both college experience (whether they were drafted straight out of high school or from a 4-year college) and age at MLB debut, and number of years in the majors. We got these statistics from baseball-reference.com. We graphed several variables with regards to player average WAR, player age, and player debut date to try and calculate the value of a player given their age and experience. Given the vast wealth of different statistics in the data, we decided to compare these different metrics using linear regression.



Results

Figure 1. Pitcher Career WAR vs. Time Since Debut

The trend lines in Figure 1 show little difference in drafting a high school pitcher or a pitcher from a 4-year university over their career. The marginal difference is that college draftees seem to improve more as time progresses while high school draftees start off better. Their career performances eventually converge to similar values. The speculation is that players are drafted out of high school because they have more natural talent, while college players are better coached and therefore are able to improve more. The obvious trend that a player improves as he spends time in the majors is present in the data.

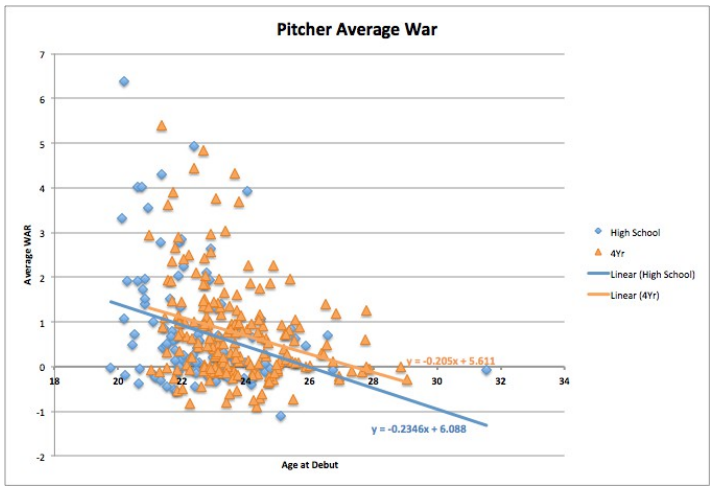


Figure 2. Pitcher Average WAR vs. Age at MLB Debut

The common trend in Figure 2 is that players, regardless of where they were drafted from, tend to perform worse on average over their career if they are debuted at an older age. The trend line slopes show that the average career performance of a college draftee degrades less with age. This would suggest that older prospects in the minor leagues have marginally more promise if they played college ball.

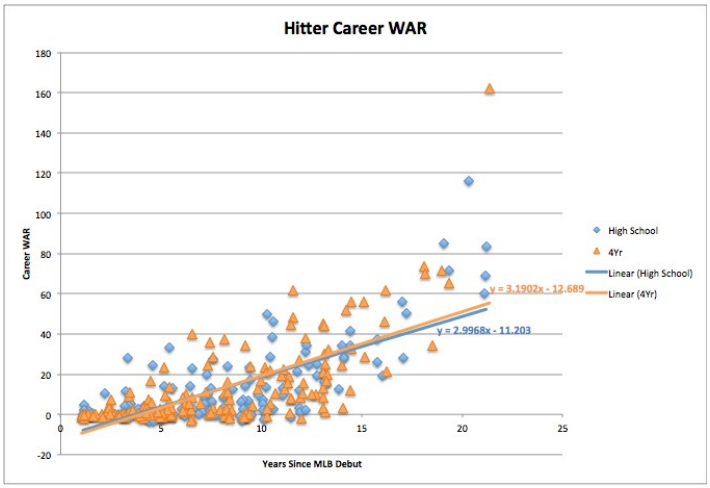


Figure 3. Hitter Career WAR vs. Time Since Debut

Figure 3 shows the trend that hitters perform worse than pitchers but improve more rapidly and are able to contribute much more to a team over their career than pitchers. It also shows that the crossing point where

college players begin to perform better to be earlier in a hitter's career. This would suggest that when considering trades, one might value an older player more if he was drafted out of college because there is marginally more promise.

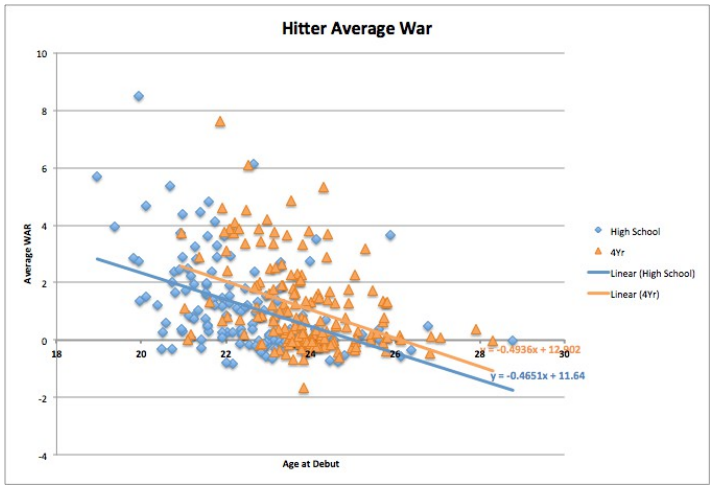


Figure 4. Hitter Average WAR vs. Age at MLB Debut

Hitter performance degradation with age before debut seems to be more significant for college draftees. Despite this trend, both Figure 2 and Figure 4 show better performance for college players across all reasonable debut dates. This is contrary to the result one finds when comparing the average WAR between high school and college draftees. The average of the average WAR of high school players is higher for both pitchers and hitters. An explanation for these differing results is that the trend lines are fit using least squares regression, which is affected by the variance. The variance is higher with college players.

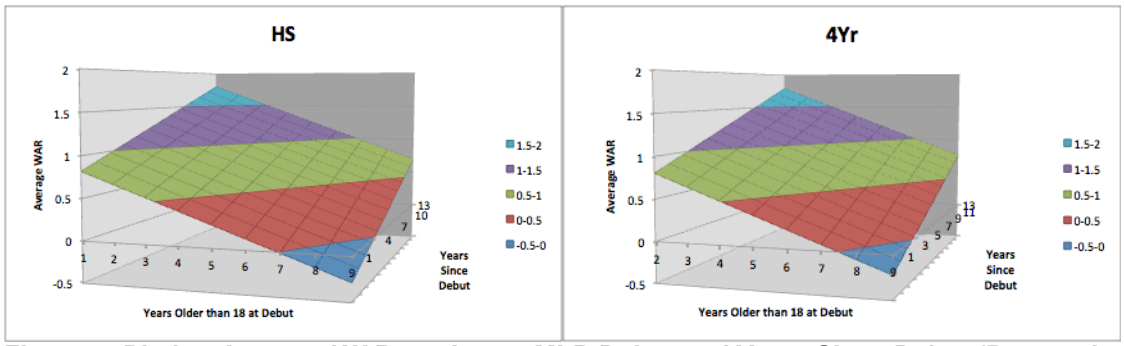


Figure 5. Pitcher Average WAR vs. Age at MLB Debut and Years Since Debut (Regression)

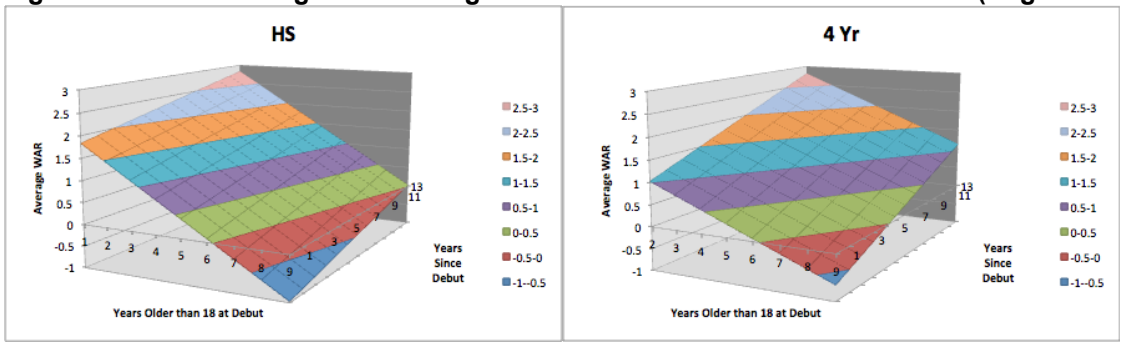


Figure 6. Hitter Average WAR vs. Age at MLB Debut and Years Since Debut (Regression)

Figures 5 and 6 are least square regressions done with average WAR as the dependent variable and age at the time of MLB debut and years since debut as the independent variables. The regression for each combination of HS/4YR and Pitcher/Hitter which resulted in 4 regressions of the form $\alpha_i+\beta_{i,1}*x_{i,1}+\beta_{i,2}*x_{i,2}$. By looking at the graphs for pitchers, it does not seem that there is much of a difference at all with regard to being drafted out of high school or college (although the high school regression did not end up having a p value less than . 05). One will notice that the axes are different due to the fact that college players cannot debut earlier than a certain age given their aging while in college. The graph's similarities exist because this taken into account. The regression does show a large difference for between high school and college hitter draftees. Hitters out of high school start out much stronger if debuted early but only marginally improve after debut. Hitters out of college start out slower than high school draftees, but make up for it in career improvement and eventually close the gap and surpass high school draftees. Both college and high school draftees perform worse the older they are when they debut, but the degradation with age is much higher for high school draftees suggesting a possible quick return when drafting high school hitters but risky with a poor downside.

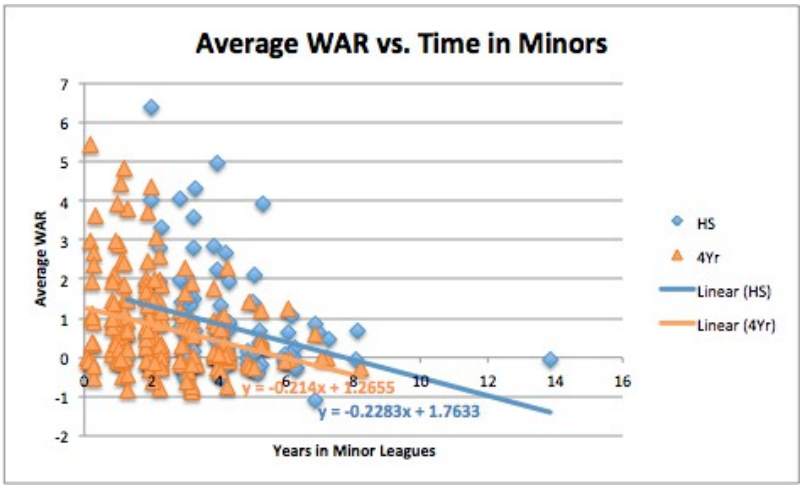


Figure 7. Pitcher Average WAR vs. Time in Minors

Figure 7 shows that on average, for both pitchers drafted out of high school and out of college, their average value decreased with the more time they spent in the minor leagues. This trend adds to the previous idea of diminishing returns on older players that have been in the minors for too long.

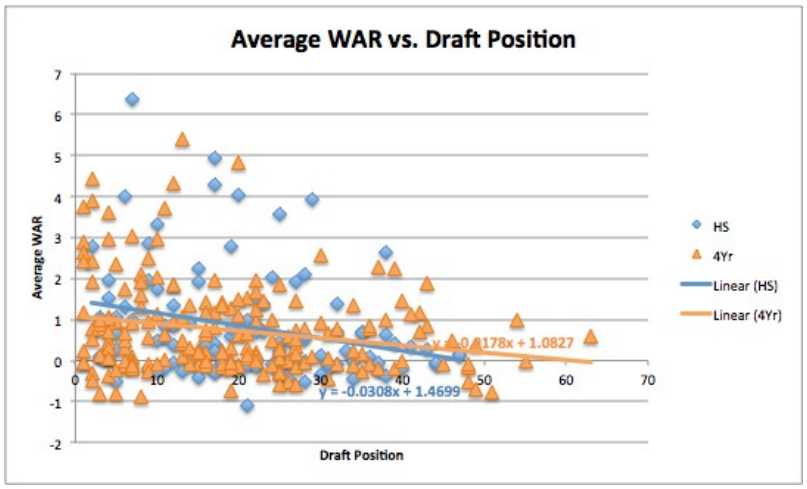


Figure 8. Pitcher Average WAR vs. Draft Position

As expected, pitcher average WAR is negatively correlated with draft position (the draftee's overall pick number in the first round of the draft for year x) in Figure 8. While both high-school and college players drafted later in the first round tend to have lower productivity in terms of WAR, the intersection of the trend lines is noteworthy. The two lines intersect at about pick number 30 - in the earlier parts of the draft, pitchers drafted straight out of high school on average turn out better in the major leagues, while after pick #30 college players are better. This could lead to an interesting conclusion for teams analyzing who to pick in the draft: with higher picks, it is more rewarding to draft a high school pitcher, and for picks after pick 30 it might be a safer bet to go with a pitcher from college.

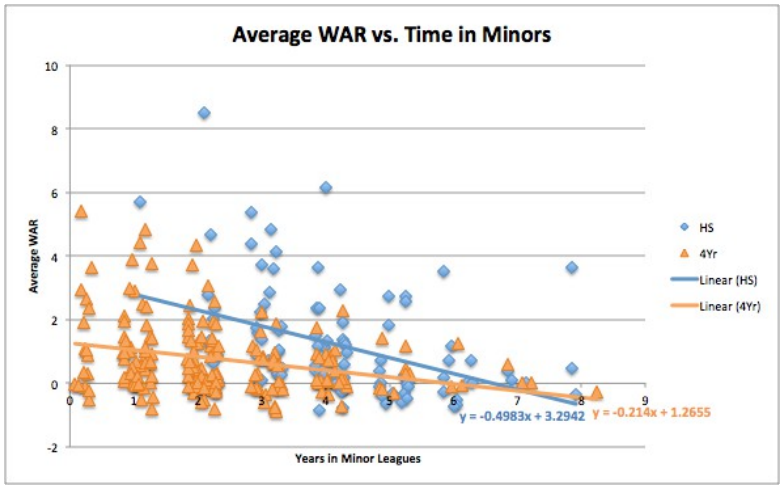


Figure 9. Hitter Average WAR vs. Time in Minors

From Figure 9 we can deduce that high school draftee hitters have more upside than hitters drafted out of college; however, the gap between trend lines seems to decrease in size as minor league experience increases. It seems from this analysis that the hitters with the most potential are highly talented hitters drafted straight from high school who can contribute to their major league team almost immediately, and once the level of minor league experience approaches 6-7

seasons, then they more likely to not be productive if they ever reach the majors. Similar results can be deduced from the trend line for college draftees, however the y-intercept is not as high - so on average the college age player who can immediately make a major league club is not as valuable in terms of average WAR than a high school draftee.

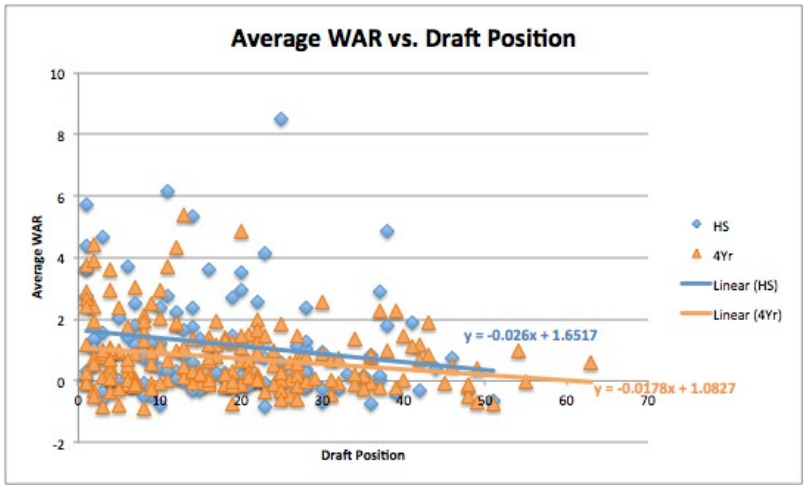


Figure 10. Hitter Average WAR vs. Draft Position

When correlating hitters average WAR to first round draft position in Figure 10, we again see a trend of decreasing as we get further into the draft. However, the slopes of the trend lines when compared to those of Figure 8 (same analysis but for pitchers) are more horizontal. Nevertheless, the high school draftee trend line stays above the college draftee line for the duration of the graph, indicating increase productivity (average WAR) for hitters drafted out of high school.

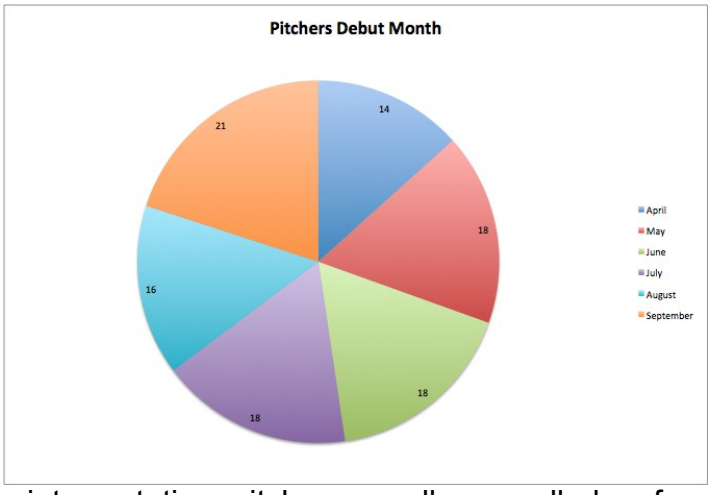


Figure 11. Pitchers Debut Month (%)

This pie chart is an aggregate of both high school and college draftees - we wanted to add some extra analyses to our project and decided to look into debut dates and birth dates for all players we had the data for. Here, it looks like all months are fairly equally represented, and it makes sense with our interpretation: pitchers usually are called up from the minor leagues during all months of the MLB season due to injuries to other starting pitchers, bullpen issues, or just the need for a fresh arm. Therefore, it is pretty easy to understand why a pie chart of aggregated pitcher debut months would be fairly evenly distributed across months.

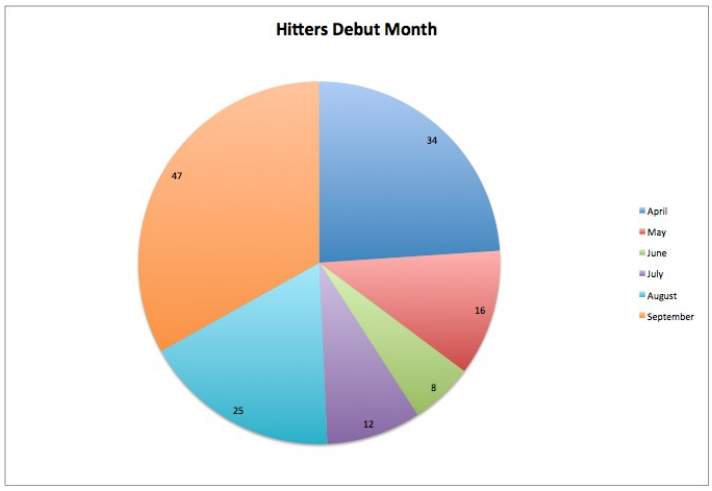


Figure 12. Hitters Debut Month (%)

This pie chart is the same aggregated data across all high school and college draftees, but for hitters. Now we notice a trend towards the earlier and later months of the MLB season - more players make their debuts in April and September than during the so-called “dog days” of summer. This makes sense with our intuition because hitters take more time to develop in the minor leagues and are either granted a promotion to the major leagues either at the very start of the season after spring training, or in September when the roster size expands and teams not in contention want to test out players for the next season.

Conclusion

The summary of our results is as follows:

- On average, players drafted in the first round perform better when drafted out of high school (both pitcher and hitter).
- Hitters picked lower in the draft improve their performance more if they are drafted out of college rather than high school.
- Pitchers performance considering debut age and experience seems to be the same for high school and college players.
- Players perform better over their career when debuted early and players who have spent too much time in the minors lose potential.
- For pitchers, debut age for HS draftees was the *only* variable in our multi regression analysis that was not significant (we tested average WAR vs. age at debut and years since debut). Our interpretation is that high school pitchers simply haven’t had as many years of experience playing organized baseball as college players, so there is more variability and no general trend on how well a high school pitcher is going to perform given their age at debut. This reinforces the notion of the “high risk, high reward” pitcher drafted straight out of high school. On the other hand, a great hitter in high school will perform well against most pitching but a great pitcher in

high school may have not faced the same quality of competition as a college pitcher, and therefore his draft value could be inflated.

- For hitters, all variables were found to be statistically significant - both debut age and years since MLB debut were found to significantly impact player average WAR for both HS and 4-YR draftees, which confirms that high school hitters expected performance degrades more quickly with age before debut and college hitters improve more rapidly.

Further research will need to be done with more years of draft data and more rounds of the draft besides the 1st round in order to definitively filter out any outliers that exist in the top picks. To truly obtain data that a trend can be deduced from, it would have to be a trimmed sample of lower round picks. Lower round picks center more around the mean that the top picks do and would provide a better representation of how high school or college experience factors into the draft.

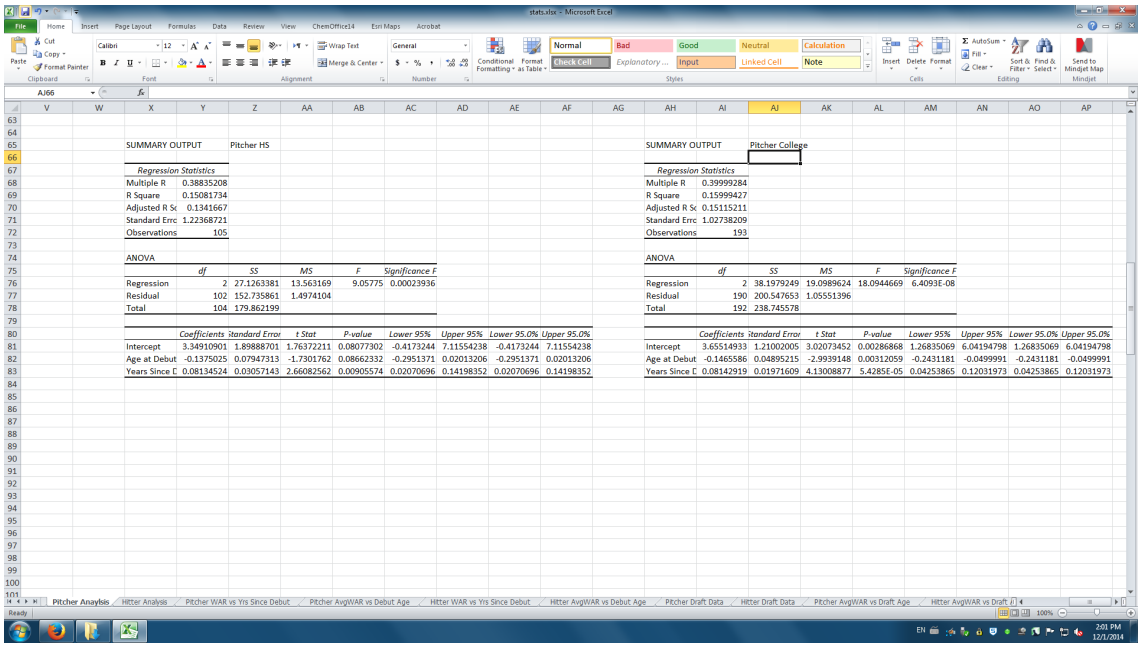
Because our data set was limited, we also had to eliminate several key aspects that we wanted to look into. Players drafted from *2-year* colleges as opposed to high school or 4-year schools were eliminated from our data set, as well as players that had only been in the major leagues for less than a year.

Ultimately, we see this area of research as extremely valuable to the front office of any Major League Baseball team because baseball teams and their scouts are always looking for ways to quantify the best young talent possible and draft as best as they can while trying to predict future performance and maturity of young players. We hope that our analysis is the first step toward a broader trend of baseball statistics projects analyzing the importance of age and pre-draft baseball experience (HS or college) when it comes to seeking the best players for the annual June Amateur Draft.

References

“Doctoring The Numbers: Starting Them Young.” Jazayerli, Rany. 13 October 2011. http://www.baseballprospectus.com/article.php?articleid=15295
“Hurdle Models and Age Effects in the Major League Baseball Draft.” Sims, Justin. 5 May 2014. http://digitalcommons.macalester.edu/cgi/viewcontent.cgi?article=1034&context=mathcs_honors
“MLB June Amateur Draft - Baseball-Reference.com.” Baseball-Reference.com.

Appendix



SUMMARY OUTPUT — Pitcher HS

Regression Statistics	
Multiple R	0.38835208
R Square	0.15081734
Adjusted R Sc	0.1341667
Standard Errc	1.22368721
Observations	105

ANOVA

	df	SS	MS	F	Significance F
Regression	2	27.1263381	13.563169	9.05775	0.00023936
Residual	102	152.735861	1.4974104		
Total	104	179.862199			

	Coefficients	tandard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.34910901	1.89888701	1.76372211	0.08077302	-0.4173244	7.11554238	-0.4173244	7.11554238
Age at Debut	-0.1375025	0.07947313	-1.7301762	0.08662332	-0.2951371	0.02013206	-0.2951371	0.02013206
Years Since D	0.08134524	0.03057143	2.66082562	0.00905574	0.02070696	0.14198352	0.02070696	0.14198352

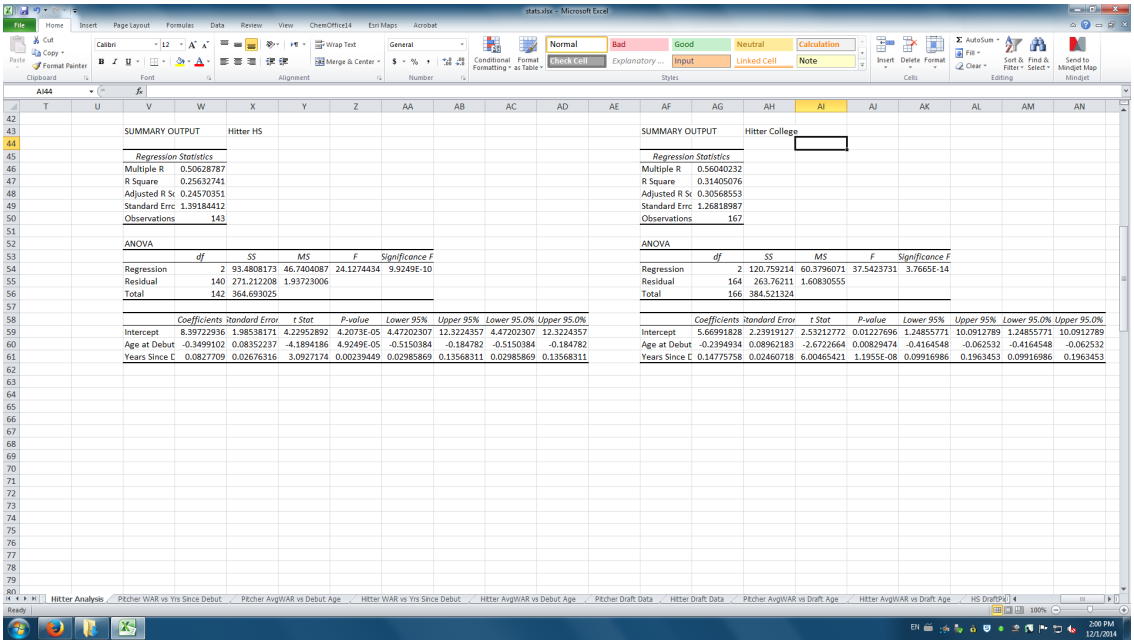
SUMMARY OUTPUT — Pitcher College

Regression Statistics	
Multiple R	0.39999284
R Square	0.15999427
Adjusted R Sc	0.15115211
Standard Errc	1.02738209
Observations	193

ANOVA

	df	SS	MS	F	Significance F
Regression	2	38.1979249	19.0989624	18.0944669	6.4093E-08
Residual	190	200.547653	1.05551396		
Total	192	238.745578			

	Coefficients	tandard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.65514933	1.21002005	3.02073452	0.00286868	1.26835069	6.04194798	1.26835069	6.04194798
Age at Debut	-0.1465586	0.04895215	-2.9939148	0.00312059	-0.2431181	-0.0499991	-0.2431181	-0.0499991
Years Since D	0.08142919	0.01971609	4.13008877	5.4285E-05	0.04253865	0.12031973	0.04253865	0.12031973



SUMMARY OUTPUT — Hitter HS

Regression Statistics	
Multiple R	0.50628787
R Square	0.25632741
Adjusted R Sc	0.24570351
Standard Errc	1.39184412
Observations	143

ANOVA

	df	SS	MS	F	Significance F
Regression	2	93.4808173	46.7404087	24.1274434	9.9249E-10
Residual	140	271.212208	1.93723006		
Total	142	364.693025			

	Coefficients	tandard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	8.39722936	1.98538171	4.22952892	4.2073E-05	4.47202307	12.3224357	4.47202307	12.3224357
Age at Debut	-0.3499102	0.08352237	-4.1894186	4.9249E-05	-0.5150384	-0.184782	-0.5150384	-0.184782
Years Since D	0.0827709	0.02676316	3.0927174	0.00239449	0.02985869	0.13568311	0.02985869	0.13568311

SUMMARY OUTPUT — Hitter College

Regression Statistics	
Multiple R	0.56040232
R Square	0.31405076
Adjusted R Sc	0.30568553
Standard Errc	1.26818987
Observations	167

ANOVA

	df	SS	MS	F	Significance F
Regression	2	120.759214	60.3796071	37.5423731	3.7665E-14
Residual	164	263.76211	1.60830555		
Total	166	384.521324			

	Coefficients	tandard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	5.66991828	2.23919127	2.53212772	0.01227696	1.24855771	10.0912789	1.24855771	10.0912789
Age at Debut	-0.2394934	0.08962183	-2.6722664	0.00829474	-0.4164548	-0.062532	-0.4164548	-0.062532
Years Since D	0.14775758	0.02460718	6.00465421	1.1955E-08	0.09916986	0.1963453	0.09916986	0.1963453