

İSTANBUL
ŞEHİR
ÜNİVERSİTESİ

CS 240 – EXPLORATORY DATA ANALYSIS
FINAL PROJECT REPORT

ANALYSIS ON BASKETBALL STATISTICS

ZEYNELABİDİN ERTÜRK
213910440

ABSTRACT

In this report we are trying to analyse some sort of basketball statistics to answer a selected question from the list of three valid questions.

SECTION 1 - Questioning

1- Is there any relation between the height of player and the performance of him? Is it strong or weak, positive or negative?

2- Is there any relation between coach and winning statistics?

3- Is there any relation between race and personal fouls?

I have chosen first question from this list of questions. Because it's a classic myth that "The tallers plays basketball better." and I want to check that. Also my data is suitable for this analyse that it has height, points, rebounds, succeed field goal, succeed free throw and assist statistics which are enough to measure a player's performance.

Hypothesis

Basketball game is throwing a ball to get it in a basket which is lifting at the top of 3.05 meters stick. Because of this nature of basketball game height of the player plays a significant role on his performance.

SECTION 2 - Data Set Describe

Variables

In basketball game these statistics can show how the player is good:

- Points, rebounds, succeed field goals, succeed free throws, blocks, assists, succeed three goals, steals

But in this analyses only

- points, rebounds, succeed field goal, succeed free throw and assist

statistics used because of the lack of data. And also height of players is another statistic in this analyses. So In the end of data cleaning and merging process the result is like this:

	height	points	rebounds	fg_made	assists	ft_made
player_id						
abdulka01	85.0	14.0	8.0	6.0	3.0	2.0
abdursh01	81.0	9.0	6.0	4.0	0.0	0.0
adamsal01	81.0	4.0	3.0	2.0	0.0	0.0
adamsmi01	70.0	9.0	1.0	4.0	1.0	0.0
aguirma01	78.0	12.0	1.0	4.0	1.0	3.0
alingeda01	76.0	12.0	3.0	4.0	2.0	1.0
allenra02	77.0	15.0	2.0	6.0	2.0	1.0
anthoca01	80.0	24.0	10.0	12.0	2.0	0.0
archina01	73.0	12.0	5.0	4.0	7.0	3.0

Solved Complications

On the merging part we have deal with some complications. First one is duplicated statistics. In the dataset for every player there are several statistics according to years. That

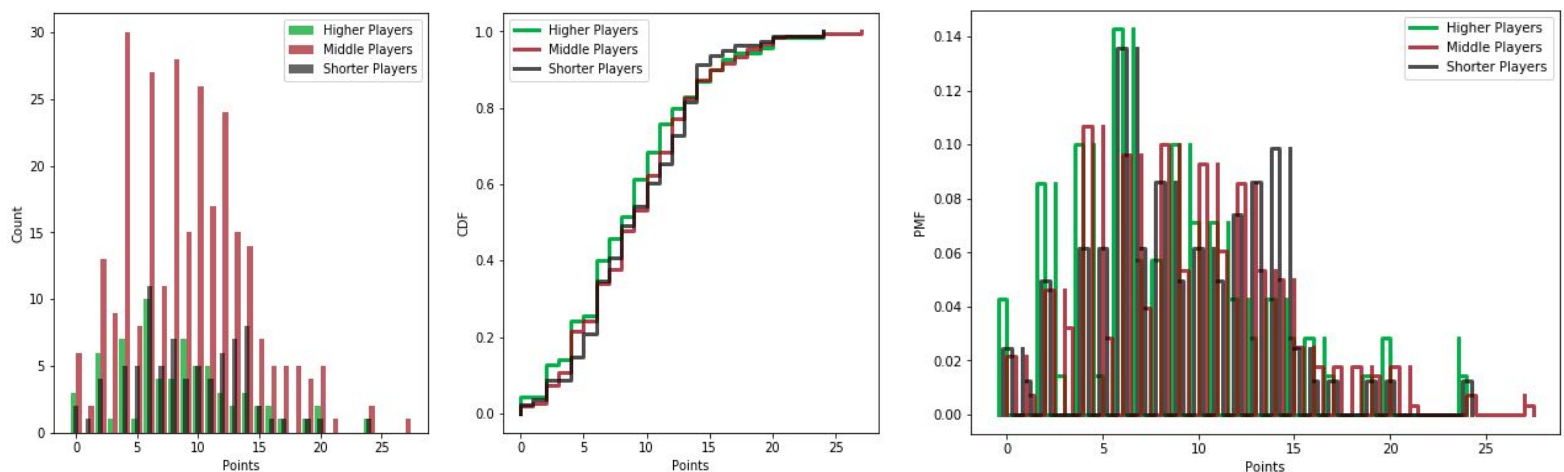
duplication makes the difficult to merge height value, which is in another dataset, via player id. We take the mean of each statistics for each player and also round the result of mean to make graphs much more understandable.

SECTION 3 - Histogram, CDF and PMF Graphs

This section includes the comparison of higher, shorter and middle height players performance in five descriptive statistics: points, rebounds, assists, succeed field goal, succeed free throw. To make the comparison possible we calculated the overall height mean and separated the dataset in three groups:

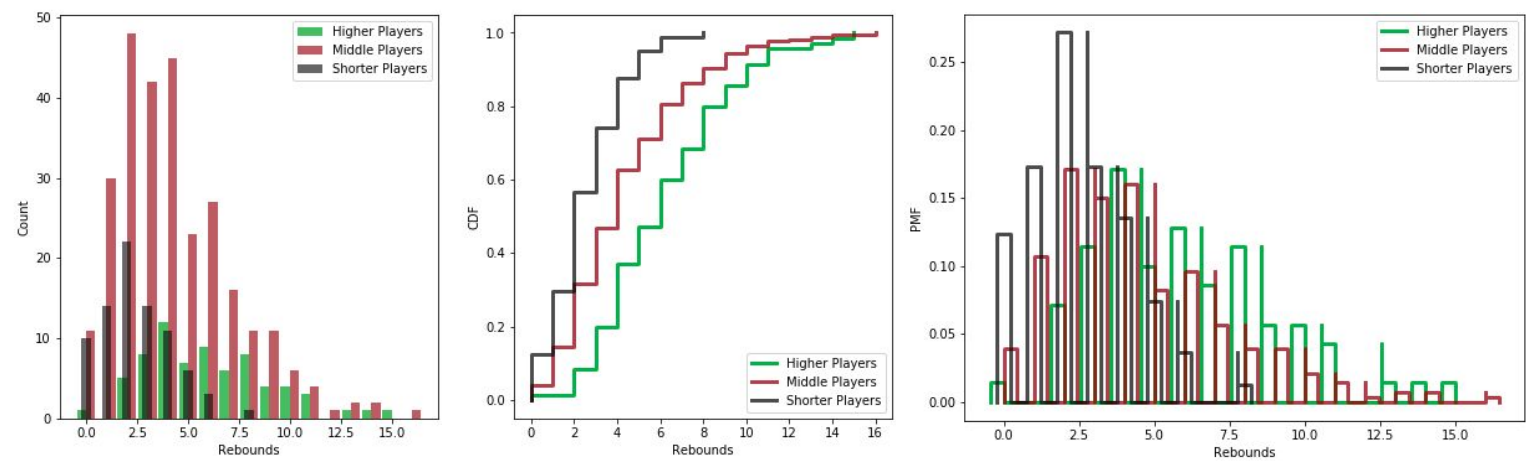
- **Higher Players** > (height mean) + (standard deviation of height mean)
- (height mean) + (standard deviation of height mean) > **Middle Players** > (height mean) - (standard deviation of height mean)
- **Shorter Players** < (height mean) - (standard deviation of height mean)

Points



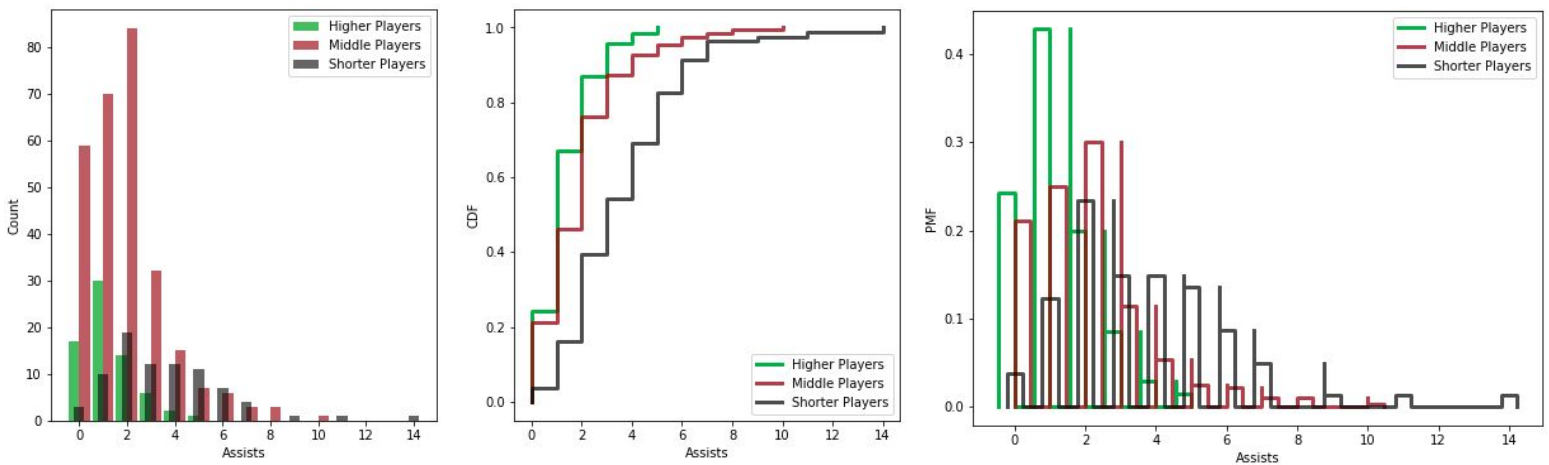
According to the graphs, the higher players has better points generally but at max the middle players are the best.

Rebounds



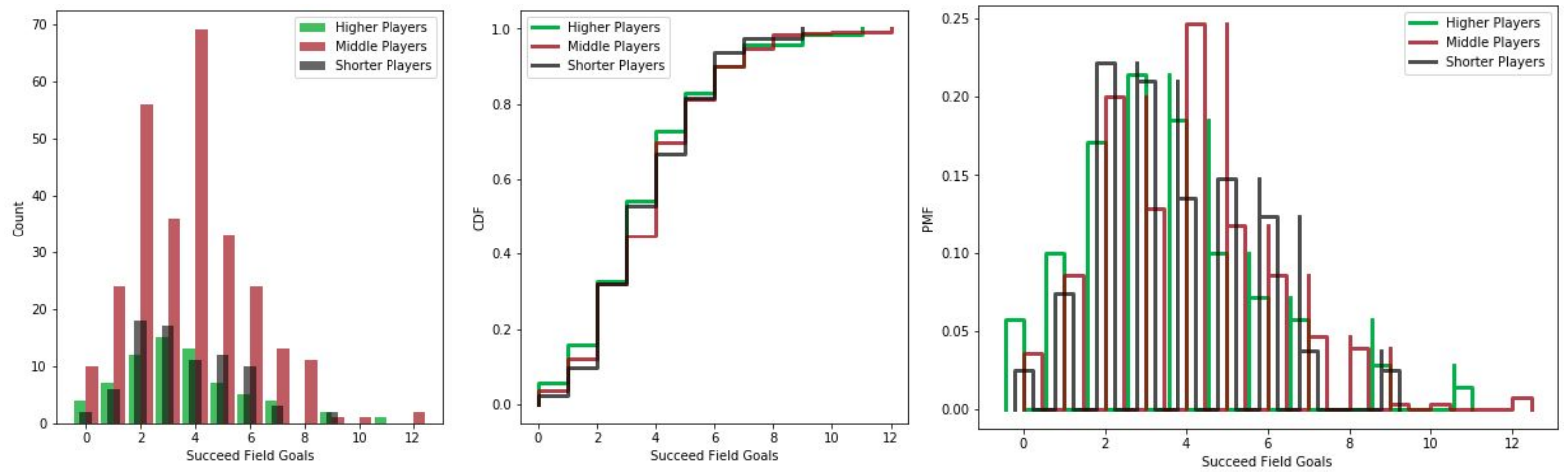
According to the graphs, the higher players clearly has better rebound statistics but at max the middle players are the best.

Assists



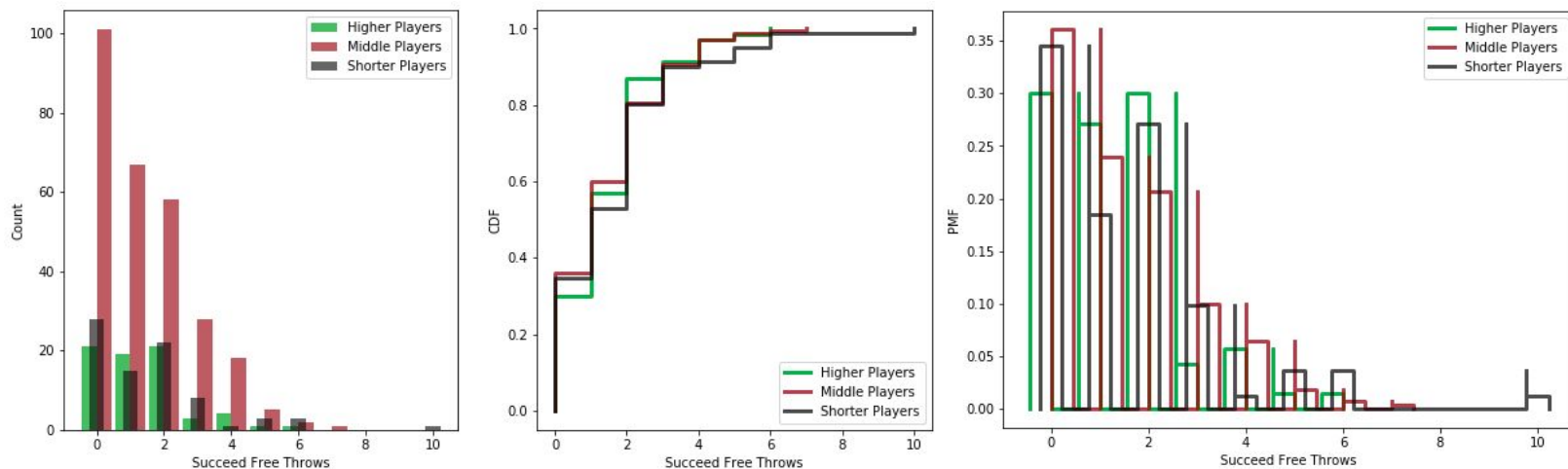
According to the graphs, the shorter players clearly has better assists statistics.

Succeed Field Goals



According to the graphs, three category of players are almost same but at max number of field goal statistics middle player is the best.

Succeed Free Throw



According to the graphs, the shorter players has better free throw statistics.

SECTION 4 - Modelling

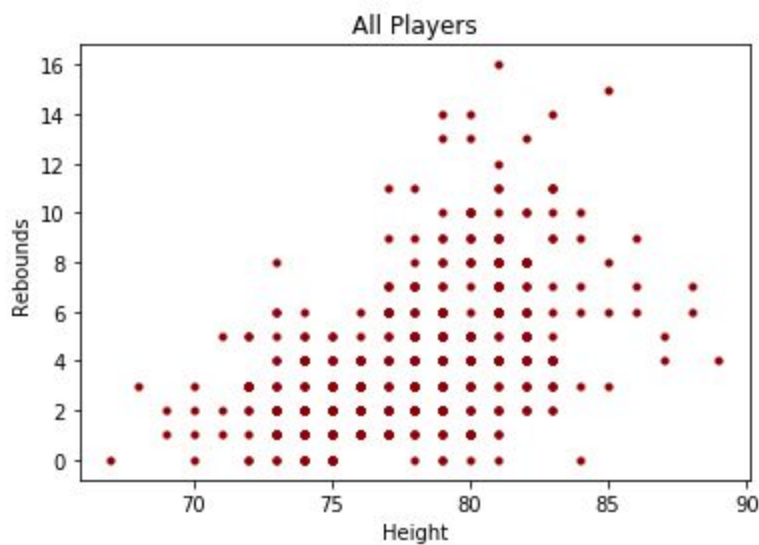
SECTION 5 - Relations

This section includes the analyses the relation between associated variables according to Pearson Correlation Model.

Height - Rebounds Pearson Correlation

According to following graph and Pearson Correlation value we can argue that height and rebound statistics has a positive correlation. We can see on the scatter plot that the dots focused on right side which means higher players.

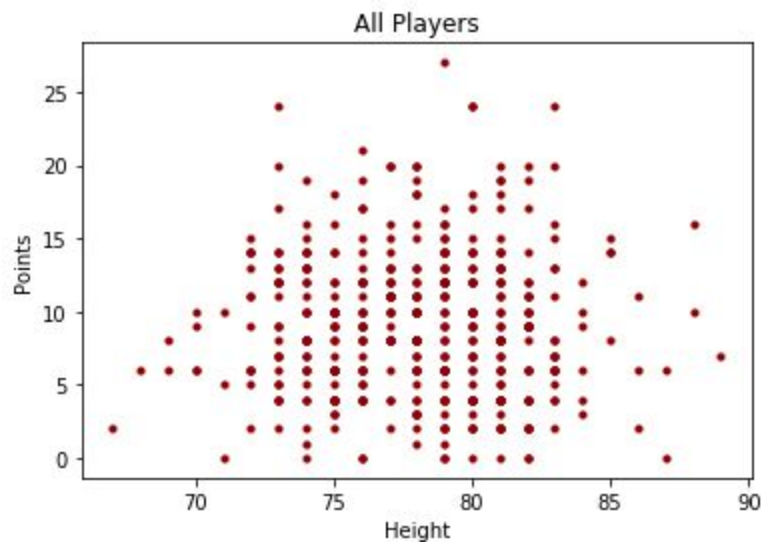
Height - Rebounds Pearson Correlation: 0.4739798204130233



Height - Points Pearson Correlation

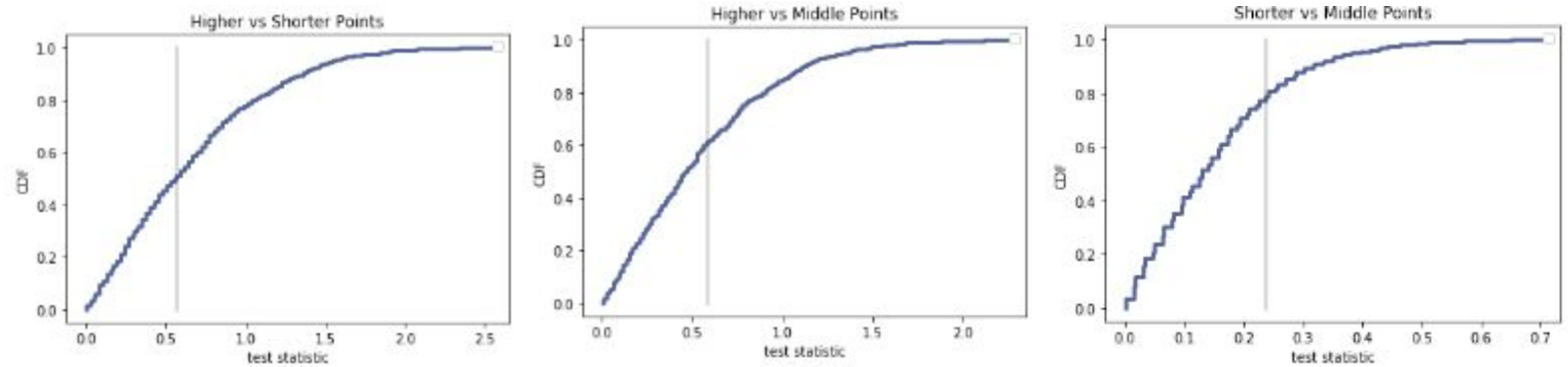
According to following graph and Pearson Correlation value we can argue that height and point statistics has no significant relation. We can see on the scatter plot that the dots are separated almost equally. But just a little it has focussed in the middle that's why the Pearson Correlation value is so close to zero.

Height - Points Pearson Correlation: -0.005517010744692143



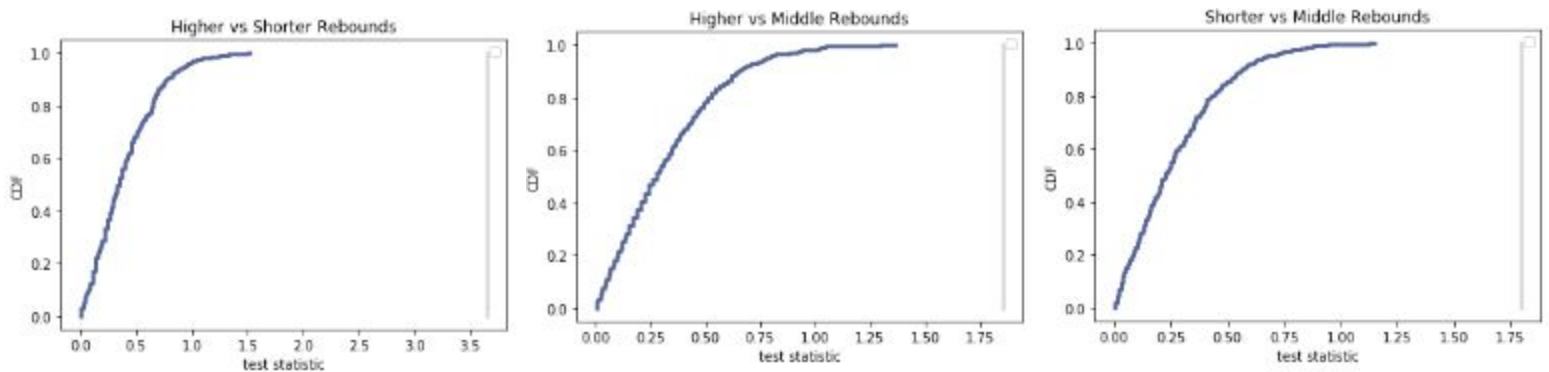
SECTION 6 - Hypothesis Testing

Points



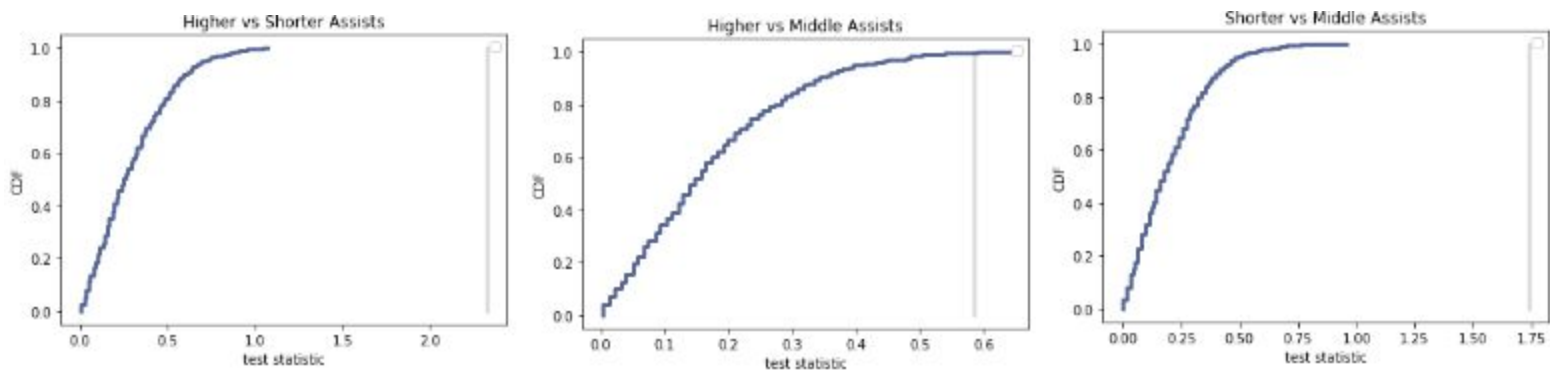
The test results show that our hypothesis failed on points statistics.

Rebounds



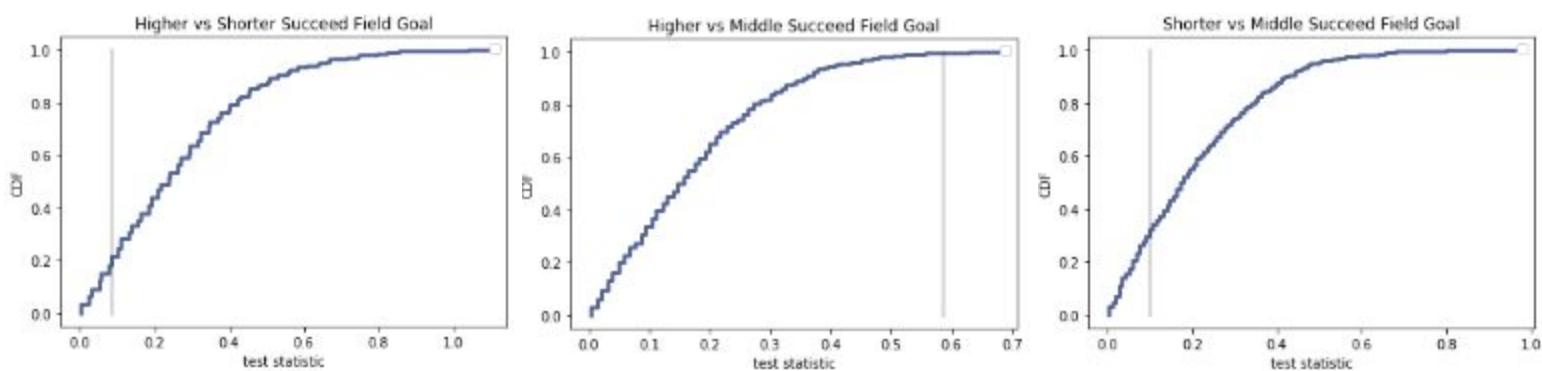
The test results show that our hypothesis succeed on rebounds statistics.

Assists



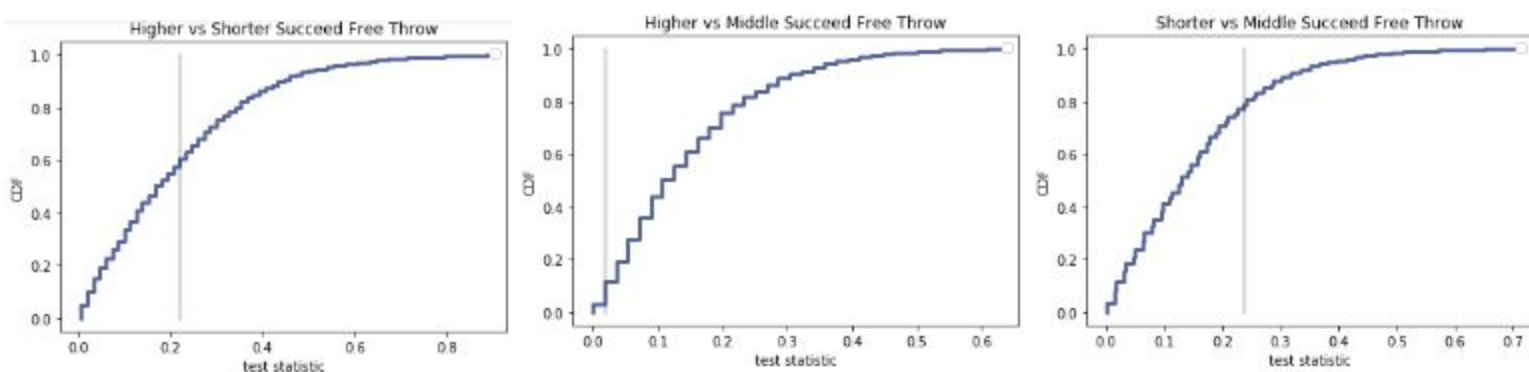
The test results show that our hypothesis succeed on Assists statistics. When we compare higher and middle players the hypothesis tends to fail.

Succeed Field Goal



The test results show that our hypothesis failed on field goal statistics.

Succeed Free Throw



The test results show that our hypothesis failed on free throw statistics.

SECTION 7 - Conclusion

In the conclusion the overall results and tests shows that there are more than one statistics to measure the performance of a basketball player. And there are more than one factors affects that statistics. Also a factor can affect different statistics in different ways.

Height of the player is one of the factor that affects the players performance, we get that result from the test but strongness and the way of affect, negative or positive, is not so clear. Because it depends on the statistics. But as a result we can say that average height is the best for a basketball player.

And of course, "May the best man win!".