



# Academy vs Non-Academy Players: A Statistical Comparison

*Authored By*

Azwad Anjum Islam

Nahian Ashraf

Riddho Ridwanul Haque

Fariha Hossain

Muntasir Wahed

Md. Shahadat Hossain Shahin

*Supervised By*

Dr Moinul Islam Zaber

Assistant Professor

Department of Computer Science and Engineering

University of Dhaka

## Content

1	<i>Abstract</i>	1
2	<i>Collecting Data</i>	2
3	<i>Defining team and individual attributes</i>	3
4	<i>Team and Individual scoring</i>	4
5	<i>Statistical analysis of data</i>	5
6	<i>Hypothesis testing and statistical inference</i>	6
7	<i>Investigating validity</i>	8
8	<i>Reporting and interpreting findings</i>	9
9	<i>Interesting observations</i>	10

*Appendix A: Z-Table*

## Abstract

In team sports like professional association football, the performance of a player depends largely on a variety of factors like his personal relationship with his teammates and managing staff, his professional understanding with the other players, his willingness to contribute to the team's greater good, often sacrificing personal glory in the process, and so on. The player can not always use all of his individual attributes to the full extent, specially when he lacks the ability to fit into his team. There are several factors that may help indicate if a player can better fit into his team. However, one particular factor that we are interested in is the effect of going through a club's youth program and then breaking into the senior team. The focus of this study was to find out if "home-grown" players, who have spent years in their respective club's youth program, have a statistical edge over their transferred "brought in" counterparts. We also seeked to determine if players who stick to their youth academy clubs perform better than those who choose to change clubs, from both an individual and team wise perspective.

The study is based on the data from more than 500 seasonal statistics from players playing in different leagues around Europe, including Spanish La Liga, Barclays Premier League, French Ligue 1, Bundesliga, Eredivisie, Serie A and Primeira Liga. The study includes study about the player's individual attributes including goals scored, clearances, blocked shots, tackles, as well as team attributes including key passes, assists, passing accuracy and so on. To keep it fair, the study uses different scales for measuring the attributes of defenders, midfielders and forwards. We calculated the performance score of the respective players from the collected data of different performance traits and used them to draw to our inferences.

Keywords: Team attributes, individual attributes, academy and non-academy players

### Collecting Data:

Before collecting data, for selecting players we chose teams from different leagues around Europe, including Spanish La Liga, Barclays Premier League, French Ligue 1, Bundesliga, Eredivisie, Serie A and Primeira Liga. From each league, we picked out different teams of varied cultures of successes. Our decision behind choosing players from leagues of same standard was to make sure that the individual standards of players do not vary on a large scale.

The chosen teams were Barcelona, Real Madrid, Real Sociedad, Atletico Madrid and Deportivo La Coruña from Spanish La Liga, Manchester City, Leicester City, Chelsea from Barclays Premier League, Juventus, Udinese from Serie A, Bayern Munich, Borussia Dortmund and Werder Bremen from Bundesliga, Marseille from French Ligue 1 and Tottenham Hotspur from English Premier League.

From the selected teams, academy players and non-academy players were separated and ten players from each team were selected at random making sure enough academy players were chosen for our sample to be valid.

Players were chosen with different national background, playing positions, previous club history and level of experience and physical attributes.

The players chosen were sub-grouped into three groups according to their playing positions as forward, midfielder and defender.

Nationality, physical attributes, club history was noted down with team and individual attributes. Total number of assists, key passes, pass accuracy, total shots, number of clearances were chosen as team attributes and total number of assists, shot accuracy, number of goals, interceptions, blocked shots, duels won, take on, tackles, fouls, red cards and yellow cards as individual attributes.

Data was collected from <http://www.squawka.com/> and <http://www.wikipedia.com/> and occasional help from <https://www.google.com/>.

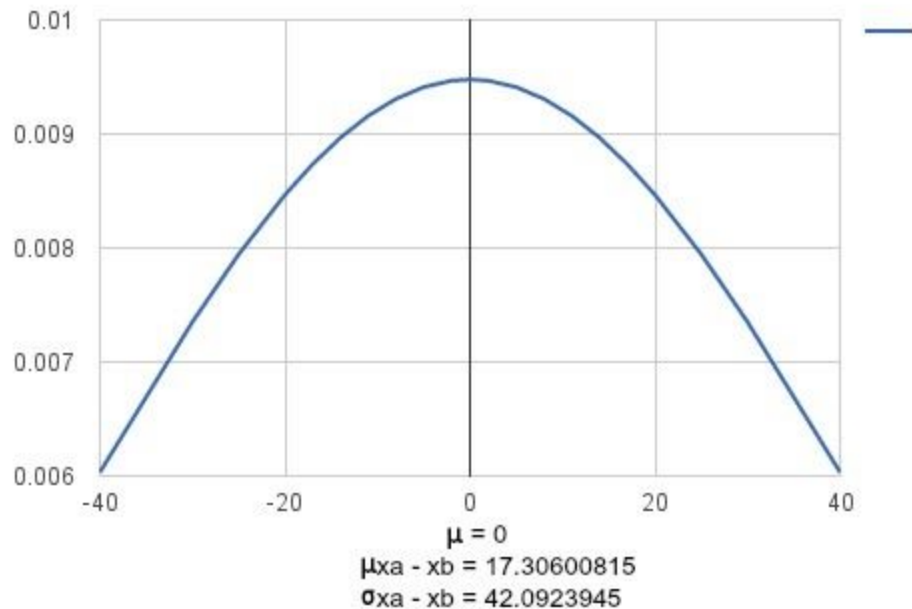
Collected data was stored using Google Spreadsheet.

## Statistical Analysis of Data

Mean Team Score given the player is academic	3205.497438
Mean Individual Score given the player is academic	3382.137734
Mean Total Score given the player is academic	6592.579935
Variance of Team Score given the player is academic	317786.6331
Variance of Individual Score given the player is academic	464136.1612
Variance of Total Score given the player is academic	1350238.824
Standard deviation of Team Score given the player is academic	563.7256718
Standard deviation of Individual Score given the player is academic	681.2753931
Standard deviation of Total Score given the player is academic	1161.997773
Skewness of Team Score given the player is academic	0.5322259935
Skewness of Individual Score given the player is academic	1.195563285
Skewness of Total Score given the player is academic	1.169936947
Kurtosis of Team Score given the player is academic	1.762396509

Kurtosis of Individual Score given the player is academic	4.736964977
Kurtosis of Total Score given the player is academic	3.987094873
Mean Team Score given the player is non-academic	3188.19143
Mean Individual Score given the player is non-academic	3289.146651
Mean Total Score given the player is non-academic	6479.616362
Variance of Team Score given the player is non-academic	289640.7517
Variance of Individual Score given the player is non-academic	528485.8437
Variance of Total Score given the player is non-academic	1202182.243
Standard deviation of Team Score given the player is non-academic	538.1828237
Standard deviation of Individual Score given the player is non-academic	726.9703183
Standard deviation of Total Score given the player is non-academic	1096.440716
Skewness of Team Score given the player is non-academic	-0.3635980033
Skewness of Individual Score given the player is non-academic	0.3379299278
Skewness of Total Score given the player is non-academic	-0.4411258268
Kurtosis of Team Score given the player is non-academic	2.488263565
Kurtosis of Individual Score given the player is non-academic	4.161214713
Kurtosis of Total Score given the player is non-academic	2.703814322

## Hypothesis Testing



Let,  $H_0$  = Playing in the Academy does not matter significantly |  $\mu_{xa} - x_b = 0$

And  $H_1$  = Playing in the Academy matters significantly |  $\mu_{xa} - x_b > 0$

$$\mu_{xa} - x_b = \mu_{xa} - \mu_{xb} = 3205.497438 - 3188.19143 = 17.30600815$$

$$\sigma_{xa} - x_b = \sqrt{\sigma_{xa}^2/n_1 + \sigma_{xb}^2/n_2} = \sqrt{417.7256438^2/116 + 314.6058075^2/370} = 42.0923945$$

$\mu_{xa} - x_b - \mu_0 = 17.30600815/42.0923945 = 0.41$  standard deviation,  
which is far less than the 95% confidence interval.

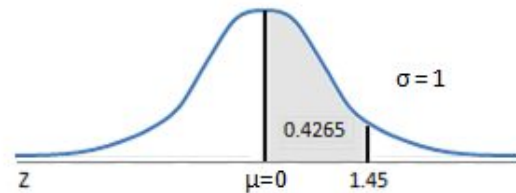
$\Pr(Z) = .2910 = 29.1\%$  (From the Z-Table, appendix A)

Hence, we can not reject the null hypothesis completely.

### Areas Under the One-Tailed Standard Normal Curve

This table provides the area between the mean and some Z score.

For example, when Z score = 1.45  
the area = 0.4265.

[illegible]



## Interesting Observations

Throughout the study, we have made some interesting observations. Some of them are as follows:

1. We defined our own scoring system. After the analysis, we found out that all of our individual, team and total scoring systems generated perfect normal distributions.
2. We have observed that most of our distributions have skewness close to zero.
3. Lionel Messi became the player with the highest individual, team and total score. To show his complete dominance, he maintained a difference of 2000 points with the nearest opponent.
4. We have observed that the total score with respect to years spent in the club generates a regression line which is almost straight. This is an interesting observation, because it is a common belief that players tend to perform better when they spend more time in a club.

