Introduction

The choice of study has been the Italian football premier league Serie A for the season 2013-2014. After the end of each Serie A season, the top 5 football clubs get promoted to next season European Cups while the bottom 3 clubs get relegated to the Serie B league. The larger topic that is investigated is how does a team identify the kinds of performance that most considerably affect their success during the Serie A 2013-2014. Football is a game intended to be ongoing, a continuous flow of action with virtually no down time. There are no "breaks" between plays, no manager conferences, and the game actions only stop at the conclusion of each half or for a serious player's injury. Only three substitutions are allowed per game, no opportunity for the players to gather and discuss strategies between plays.
Data collected by Italian Football Data: Serie A & B (URL: http://datahub.io/dataset/italian-football-data-serie-a-b) for the complete 2013-2014 Italian Serie A season is used to uncover the pitch actions.
The action groups taken into account are league points, goals forward and against, shots taken, corner kicks, fouls committed, yellow cards and betting companies predictions.
The data is organized and tabulated into relevant statistics that can be used to profile team performances.

Research Question and Hypothesis, and Experimental Design

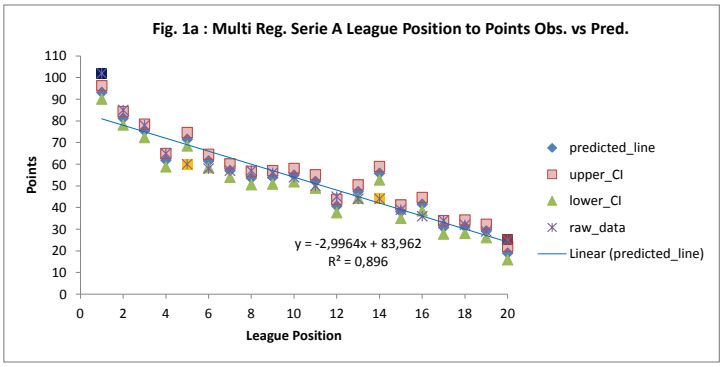
Assumption		On Hypothesis Testing section, when the word significance is used, it favours the alternative hypothesis and when not, it favours the null hypothesis. Our chosen alpha level was 0.05.
Regression	Research Question	What does a team do better on the pitch that differentiates them from the others. Are there specific offensive and defensive pitch actions that are more influential than others in a team's achievement over the season?
	Hypothesis Testing	Is there a significant relationship between the slope coefficient each team's goals against average per game (GA Avg) to y-intercept team league points.
	Outcome	A significant relationship exists between slope regression coefficient GA Avg and y-intercept Points gained throughout the season (p=0,00004). Thus p-value being less than the alpha level, we identify a relationship between those two variables. Conducting the linear multiple regression analysis, we also identified the slope regression coefficient GF to S (% goal forward to shots taken) to be different to zero to y-intercept Points (p=0,003).
Anova	Research Question	Are there any important differences between the best teams (top four), worst teams (bottom 4) and the best of the rest (middle 12 teams) using goals forward to shots (GF to S) taken as an indicator ?
	Hypothesis Testing	Is there at least one pair from those three groups of teams that significantly shoot more or less accurate over another.
	Outcome	Being GF to S factor as our dependent variable, we found out that Goal to Shot accuracy had a significant impact on the ability of teams to gain points and finish higher on the league (p=0,0007). The Bottom 4 teams were shooting less accurate on average than the Middle 12 by approx. 3% and 5% from the Top 4.
t-test	Research Question	Compare Inter, a team that finished 5th (qualified for the European Cups) with Milan that finished 8th (not qualified for the European Cups) in the league, did a goal or otherwise a hypothesized mean difference of 0.51 made the difference for Inter to finish in a higher position ?
	Hypothesis Testing	Is Inter scoring more than Milan on average mean difference of 0.51, that a goal scored by Inter in a game might have made the difference.
	Outcome	Based on our samples, there was no any significant difference that we will predict that Milan scored similar to more amount of goals than Inter (p=0.0982). P value in this case is higher than our alpha level, thus we favour the null hypothesis that Inter was scoring more than Milan
Chi2 Goodness of fit	Research Question	Has Milan's home winning performance changed significantly this year 2013-2014 relative to the previous season 2012-2013 ?
	Hypothesis Testing	If winning generally depends to some amount on the team roster, did Milan had a significant better team roster the previous year 2012-2013
	Outcome	A negative delta difference of 10.53% in Milan's home winning success rate of this year 2013-2014 relative to 2012-2013 proved to be statistically significant (p=0.03).
Chi2 Test of Independence	Research Question	Milan's chances to draw a game at home can be related to the type of betting company and how optimistic is each one of them for the final result ?
	Hypothesis Testing	Does your chances of selecting and trust the right betting company drawing score level amount is significantly related to Milan's tendency to draw or not the game at home.
	Outcome	Our alternative hypothesis presented above was true and indeed the type of betting sport company and Milan drawing the game at home are not independent and a betting company's drawing score level (i.e. low for being optimistic for the result and vice-versa) can have some influence over the result (p=0.0000002).

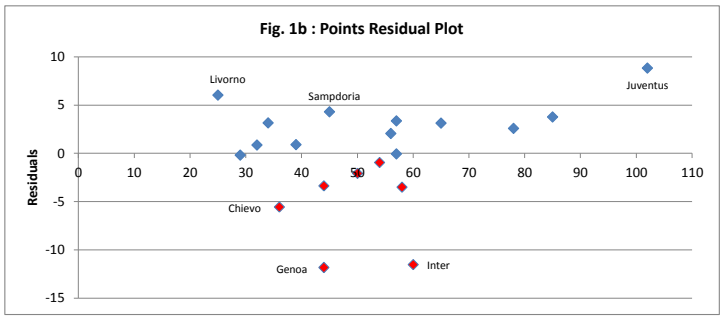
Results

Descriptive, Inferential statistics and visualization that demonstrates the data can be found on the xlsm file attached.

For your information, the type of graphs one can found on the xlsm spreadsheet for each of the statisitcal tests are :

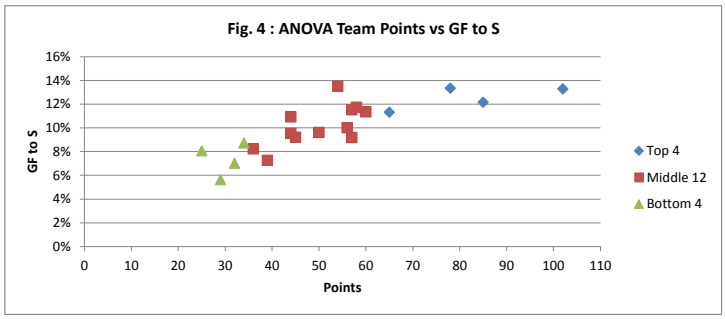
Tests	Graph Type
Regression	Four scatterplots concerning regression are presented on the spreadsheet. Below I would present the first two as an example. scatterplot graph for observed versus expected values of team's League final Position (x-axis) to Points (y-axis) is shown. scatterplot graph for observed versus expected residual values (y-axis) of team's League final Points (x-axis) is shown.





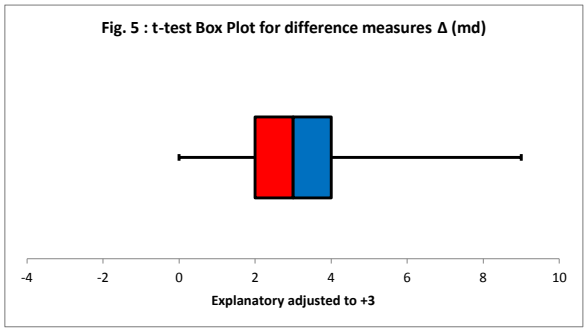
Anova

scatterplot graph for ANOVA Team Points to goals forward to shots (GF to S) percent level is shown.
GF to S percentage level is shown on the Y-axis and league team points are shown on the X-axis.



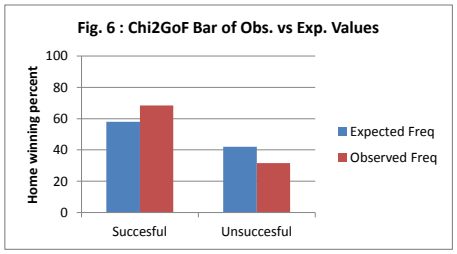
t-test

Box Plot for difference measures Δ (md), in our ex. the delta mean diff. of goals scored by Inter and Milan. Positive delta induces Inter's mean goals difference being greater than Milan's. I had an issue in dealing with negative numbers on Box Plot in Excel and I needed to adjust my values to all positive by a adding value of 3. Therefore, for ex. the real min value from the delta difference series is not 0 as shown in the graph but -3.

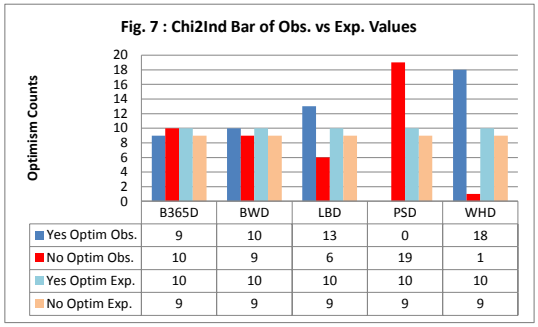


Chi2 Goodness of fit

bar graph for qualitative data for observed versus expected values of Milan's home winning percent level (Successful or UnSuccessful) is shown.
Home winning percentage level is shown on the Y-axis and type of success is shown on the X-axis.



Chi2 Test of Independence bar graph for qualitative data for observed versus expected values of optimism level (Yes or No) is shown. Number of optimism counts are shown on the Y-axis and the type of betting company is shown on the X-axis.



	B365D	BWD	LBD	PSD	WHD
Yes Optim Obs.	9	10	13	0	18
No Optim Obs.	10	9	6	19	1
Yes Optim Exp.	10	10	10	10	10
No Optim Exp.	9	9	9	9	9

Conclusions

Conclusions along with results being meaningful or not can be found on the xlsm file attached.

The Worksheet Names and starting Row, Column number and/or range that present these conclusions are :

Tests	Sheet Name	Row_Col
Regression	CorrelReg	Row = 48 to 77 and 102 to 145, Col = L or 12
Anova	Anova	Row = 76, Col = A or 1
t-test	t-test	Row = 68, Col = F or 6
Chi2 Goodness of fit	Chi2GoF	Row = 39, Col = F or 6
Chi2 Test of Independence	Chi2Ind	Row = 48, Col = R or 18