AN ANALYSIS OF BASKETBALL SCORES STATISTICS

LEE TIM SOON

A project report submitted in partial fulfilment of the requirements for the award of Bachelor of Science (Hons.) Applied Mathematics with Computing

Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

May 2011

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature : ___________________________

Name : Lee Tim Soon

ID No. : 09UEB01634

Date : 19/8/2011

APPROVAL FOR SUBMISSION

I certify that this project report entitled **"AN ANALYSIS OF BASKETBALL SCORES STATISTICS"** was prepared by **LEE TIM SOON** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Science (Hons.) Applied Mathematics with Computing at Universiti Tunku Abdul Rahman.

Approved by,

Signature : ___________________________

Supervisor : Mr Liew Kian Wah

Date : ___________________________

ACKNOWLEDGEMENTS

I would like to thank everyone who had contributed to the successful completion of this project. I would like to express my gratitude to my research supervisor, Mr Liew Kian Wah for his invaluable advice, guidance and his enormous patience throughout the development of the research.

In addition, I would also like to express my gratitude to Nettium Sdn. Bhd. Specially thanks to Mr Lam Mun Choong for giving me a chance to receive this basketball project and apply part of it as my academic project. Also thanks to Mr Alex Morton, my supervisor for his patience and providing assistance when I face problems. Once again, thanks again to Nettium Sdn. Bhd.

AN ANALYSIS OF BASKETBALL SCORES STATISTICS

ABSTRACT

Statistical techniques have been widely used in a variety of disciplines like biostatistics, behavioural science, sports science and more. Sports is an emerging field for applying statistical techniques for development of innovative ideas of dealing with a big pool of data. Furthermore, a lot of money has been invested by sports related industries, providing a lot potential opportunities to everyone. Betting on the outcome of football matches has been a long tradition. Betting on a home win, draw or an away win of a football game is one of the popular and simplest forms of betting. Therefore, a statistical model that can accurately forecast the outcome of a sports game may be a profitable business. In this project, statistical techniques are used to analyse the statistics of basketball scores that leads to development of such model.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

For Association Football it is well known that, on average, there is an general increase in scoring rate as the match progresses and the scoring rate of both teams is dependent on the match situation (Dixon and Robinson, 1998). Moreover, it is clear from analysis of the running ball odds that the bookmakers are fully aware of both phenomena.

In a basketball match, is there some quarters that has significantly more points scored than others? Besides, how are the points scored in a quarter be correlated to the points scored in another quarter? Does the “distance” between the quarters affect the degree of correlation? Then, can we predict the total points using pre-match odds offered by bookmakers using a Linear Regression methodology? Lastly, this project would like to improve the model’s prediction by using the information of current score.

.

1.2 Aims and Objectives

The objective of this project is to investigate whether there is any quarter of a basketball match that has significantly more points than others. Besides, this project also examines whether the points scored by two basketball teams during the match are correlated and whether the degree of correlation depends on the stage of the match. This project also investigates how good the odds in predicting the total points of a basketball game are. Lastly, this project investigates whether the odds can predict the total points better, if given the information about the current score.

1.3 Scope

This project only focuses on the data of the matches in USA's National Basketball Association (NBA) League. NBA 2002/2003 to NBA2009/2010 data is used to data analysis and the NBA 2010/2011 League data is used to perform model validation.

CHAPTER 2

LITERATURE REVIEW

2.1 Literature Review

Dixon and Coles (1997) had come out with a simple bivariate Poisson model for the number of goals scored by each team in football. Their model makes use of the goal scored and time of goal scored as an input. To improve the model, parameters related to past performance were also included.

Dixon and Coles also came out with a betting strategy whereby they bet on all outcomes for which the ratio of the model's probability to the bookmakers' probabilities exceeds certain level. In that paper, it also suggests the possibility of the use of bookmakers' odds along with the model's result to develop a betting strategy based on match scores.

While Dixon and Coles (1997) focuses more on fixed odds betting, Dixon and Robinson (1998) works more on setting prices in the spread betting market. Dixon and Robinson also improved Dixon and Coles's model along with Maher's model. As a result, the resulting model gives a better match outcome estimates than its ancestors. Dixon and Robinson also found out that the prices at that time are inaccurate. Lastly, they have noted that there is a continuously increasing scoring rate as the time progresses.

Thus, this project attempts to do something similar to Dixon and Coles by using the total points scored in each quarters in basketball. This project also tries to look for the possibility of the use of bookmaker's odds along with the model and develop a prediction model based on match scores. Like Dixon and Coles, this project will focus on fixed odds betting.

Besides, Harville(1980) used linear model methodology to produce a predictive model on National Football League (NFL) to forecast the outcome of the game. This is one of the factor this project is applying linear model to predict total points, though is in one of the simplest form, multiple linear regression models.

On the other hand, in Beating the Spread, Zuber, Gandar and Bowers(1985) investigated the efficiency of the gambling market for National Football League (NFL). They've managed to show a profitable gambling opportunity exist within the market, indicating that inefficiencies may appear in the gambling market.

.

CHAPTER 3

METHODOLOGY

3.1 Methodology and Tools

3.1.1 Hypothesis Testing

In statistics, a hypothesis is a claim or statement about a property of a population. Hypothesis testing is a standard process of testing a hypothesis, using data. The main question is that whether the sample data is statistically significant or not, according to a significance level.

A null hypothesis, H_0 is a statistical hypothesis that is assumed to be true until it is rejected. The alternative hypothesis, H_1 is the hypothesis that is contrary to the null hypothesis. Since they contradict each other, one of the two hypotheses must be true.

When testing a hypothesis, the conclusion can never be 100% certain. It is possible only to be confident to a certain confidence level. For example, it is 95% confident that the conclusion drawn is correct. This is called a 95% confidence level or a 5% significance level.

Type I error is the error of rejecting a null hypothesis when it is actually true whereas a Type II error is the error of failing to reject a null hypothesis when it should be rejected.

Steps to perform a hypothesis testing:

1. State the null hypothesis and alternative hypothesis
2. Choose a test statistics and level of significance
3. Determine the rejection region
4. Calculate the value of the test statistics
5. Make a decision whether to reject or do not reject the null hypothesis

We reject the null hypothesis if p-value $\leq \alpha$, else do not reject if p-value $> \alpha$ where α is the significance level of the hypothesis testing.

3.1.2 Correlation

Correlation is a measure of the relation between two or more variables. Correlation coefficients can range from -1.00 to +1.00. Correlation is useful because it can indicate a predictive relationship that may suggests interesting result in this project. Correlation can also suggest possible causal relationship. The figure below illustrates correlation with numerous graphs:

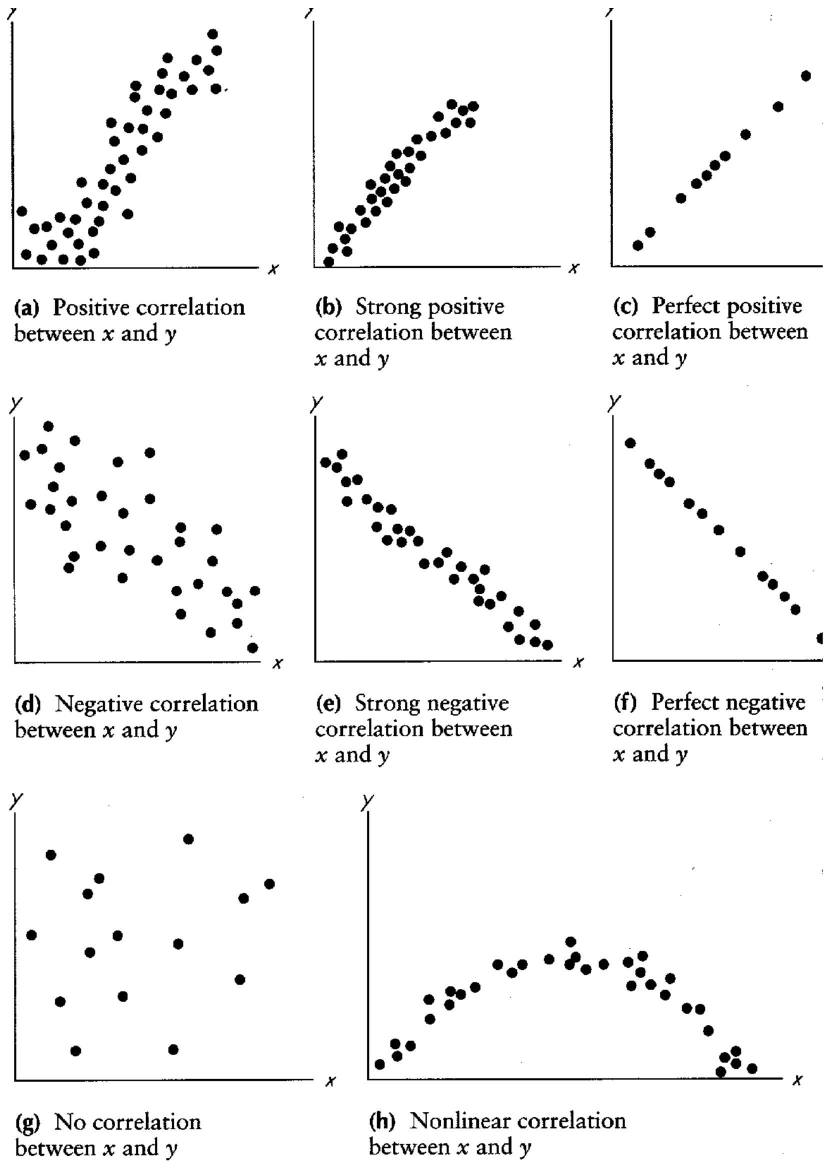


Figure 1 : Data with different correlation

3.1.3 Pearson's Correlation Coefficient

The most widely used measure of linear correlation is the Pearson's Correlation Coefficient, which is defined as the covariance of the two variables divided by the product of their standard deviations:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

The value of $\rho_{X,Y}$ falls between -1.0 and +1.0 all the time.

3.1.4 Shapiro-Wilk Normality test

The Shapiro–Wilk normality test tests the null hypothesis that a sample $x_1, ..., x_n$ came from a normally distributed population. The test statistics is

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

where: a_i is a constant.

x_i is the set of measures to assess, and

$\bar{x}$ is the mean of these measures.

The null hypothesis for this test is that the data are normally distributed. If the chosen alpha level is 0.05 and the p-value is less than 0.05, then the null hypothesis that the data are normally distributed is rejected. If the p-value is greater than 0.05, then the null hypothesis has not been rejected.

One restriction to this test is that, Shapiro-Wilk test does not assure normality but instead gives evidence of non-normality.

3.1.5 Quantile-Quantile Plot (qq-plot)

Qq-plot is a graphical technique for determining if two data sets come from populations with a common distribution.

Qq-plot is a plot of the quantiles of the first data set against the quantiles of the second data set. A 45-degree reference line plotted inside the same graph. If the two data sets originate from a population with the same distribution, the points should fall approximately along this reference line. The bigger the variation from this reference line, the more certain that the two data sets have come from populations with different distributions.

A normal qq-plot is a qq-plot for determining if a dataset comes from a normal population.

3.1.6 Paired Student's t-Test

Hypothesis: $\overline{Difference} = 0$

Given two paired sets X_i and Y_i of n measured values, the paired t-test determines if they significantly differs from each other. Let

$$\hat{X}_i = (X_i - \bar{X})$$

$$\hat{Y}_i = (Y_i - \bar{Y}),$$

$$t = (\bar{X} - \bar{Y})\sqrt{\frac{n(n-1)}{\sum_{i=1}^{n}(\hat{X}_i - \hat{Y}_i)^2}}.$$

with degree of freedom = n-1

3.1.7 Odds and Probability

In statistics, we deal with probability all the time. Probability is a measure of how much an event is likely to occur. Probability ranges from 0 to 1 and the higher the chance of an event to occur, the higher the probability.

The odds that this project is dealing with are all in decimal notation. Decimal odd are commonly used in Europe and are commonly used by online bookmakers. Decimal odds are the amount of pay-out based on one's stakes. In other words, it is the amount one received that includes the initial bet if one wins. For example, odds of 2.0 means that the pay-out is exactly equal to the original stakes if you win. If the odds is less than 2.0, this means that the winnings are less than the stake, which is normally the case when betting of the favourite team. If the odds is more than 2.0, this means that the winnings are more than the stake, which is normally the case when betting on the underdogs.

The formula to convert between odds and probability is

$$Percentage\ Probability = 100\ /\ (Decimal\ Odds)$$

Likewise,

$$Decimal\ Odds = (100\ /\ Percentage\ Probability)$$

3.1.8 Over/Under

Over/Under betting is a type of wagering in which the booksmaker sets a number before the match begins that is the expected total points scored by both teams. Then, people are free to bet on Over if they think the actual total points scored is going to exceed the number or bet on Under if they think the other way.

When placing a Over/Under bet, the only concern is only with the combined scores of each team at the end of the game.

3.1.9 Multiple Linear Regression (MLR)

Multiple linear regression is a technique that is always used to model the linear relationship between a dependent variable and one or more independent variables. The theory behind multiple linear regression is least square approach, which means that the model is a fitted in such a way that the sum of squares of residuals is minimized.

One of the practical applications of multiple linear regression is forecasting. By fitting a linear regression model onto an observed data set of y and X values, a predictive model can be obtained. Then, if we have a new X values, we can use the fitted model to predict the value of y.

A linear regression model assumes a linear relationship between the dependent variable and the vector of independent variables. The model equation is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon_i$$

where y is the dependent variables,

x_i are the independent variables,

β_i are the regression coefficients

ε_i is the error term

The model is estimated using least square approach and then a prediction equation is obtained.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n$$

where the variables with '^' are estimated values.

Multiple linear regression is bounded to several assumptions. Firstly, the model only applies to linear relationships. Then, the error term is normally distributed. Third assumption is that the expected value of the residuals is equal to 0 and the last assumption is that residuals has constant variances.

3.1.10 Coefficient of Determination, R^2

R^2 is usually denoted as the proportion of variance accounted by the regression model. One important point to note is that R^2 does not necessarily imply causation. R^2 is often being treated as a statistics to check model adequacy since it gives some information about the goodness of fit of a model.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

where SSE is the sum of squares of error,
SST is the sum of squares of total, and
SSR is the sum of squares of regression

The value of R^2 ranges from 0 to 1.

3.1.11 Tools used

This project requires some statistical works. Therefore, to ease the work, statistical software called 'R' is brought in. R is an open source software and there are a lot of statistical package available. Thus it is easy to learn and there are many examples around the Internet.

Besides, this project also requires some computer programming skills for designing a web scraping program. Microsoft's C# Programming language and .Net Framework played an important part in the project since it is needed to collect the data. Without the data, this project wouldn't be a success.

Other than that, knowledge in SQL Server and SQL Programming language is also needed. With this knowledge, it is easy to perform queries to extract certain data.

Finally, Microsoft Excel is employed here. This is because Microsoft Excel ease the presentation of data and it is easy to perform calculations on the datasheets.

.

3.2 Data Requirement

A lot of information can be obtained at the end of a basketball match. Total score of both teams for the game can be obtained, as well as the score of both teams for each quarter and an indicator of overtime. Of course the outcome of the game can be affected by other factors like the number of 3 points attempted, the injury of main player in a team, the number of fouls, and the weather condition and so on. Although the extra information is possible to be obtained, it is hard to present. If there are too many variables, it will complicate and burdens the project too much. Besides, qualitative variables like injuries of players and weather are very subjective and difficult to handle. Therefore, this project only makes use of the total and quarters' score.

Each basketball game in NBA consists of 4 quarters of 12 minutes. If the points scored by both teams reach a draw at the end of the 4 quarters, an additional quarter will be played to determine the winner. If both teams still draw after overtime, another additional quarter will be played until there is a winner. Since this is difficult to control, the data is only interested with the scores of the first 4 quarters.

On the other hand, to assist development of a predictive model, odds information is required. The odds this project dealing with is the "Over/Under" odds. However, odds information is available for some of the matches only. For this project, only odds offered by two big Asian bookmakers - 188Bet and SBOBet are captured.

In a nut shell, for each match, we have data for the

- Home and Away team name
- Date and League Period
- Total points of Home and Away team
- Quarters' points of Home and Away
- An indicator for overtime
- Over/Under odds information

3.3 Data Collection

Since NBA official website does not have the complete data for each match throughout all the years, it is needed to find a data source on my own. One extensive archive of match statistics and bookmaker's pre-match odds can be found at http://www.betexplorer.com. BetExplorer records the complete set of scores of each match for several seasons of various leagues including NBA. Besides, they also keep the details of a range of odds including the Over/Under odds for many matches. Thus, BetExplorer is definitely a good source since it has all the desired data this project needs.

Obtaining this information manually, we would need to go to the page of each NBA league, click on links of one of the game and click on the tabs for Over/Under odds and note them down one by one. However, this will cost too much time since there are around ten thousands matches to keep count of. Therefore, a program is designed in this project to capture all these data in a much efficient way.

This program makes use of a technique called web scraping. Web scraping is a computer software technique of extracting information from websites. By observing and matching the common patterns in the source code of pages in the website, the web scraping program can traverse them in some manner and extract out the desired data. In this way, the data collection would be done much faster and minimizes human error. The web scraping program in this project is written using Microsoft's C#.NET Language which is designed to loop through all the matches, search for relevant data and saves them into a SQL Server. Therefore, a decent knowledge of computer programming language of C#, .NET and SQL plays an important role to collect the data.

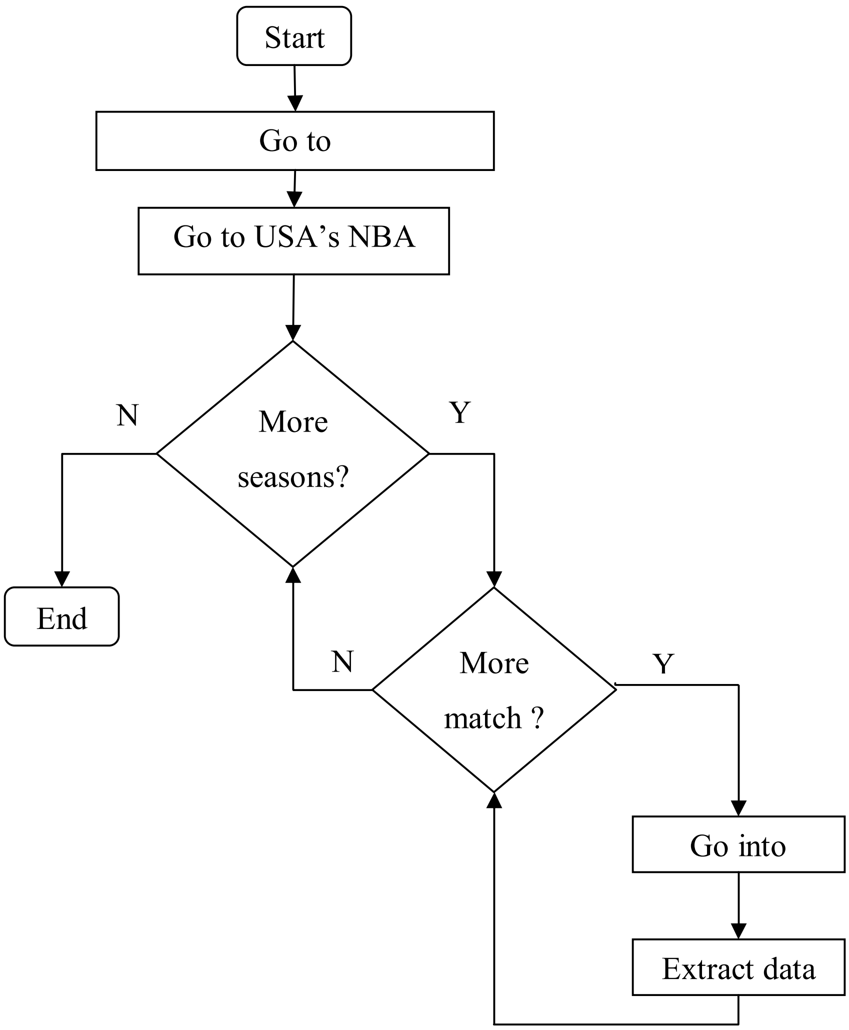


Figure 2 : Flowchart illustrating flow of web scraping

3.4 Data Presentation

A lot of information is being collected into a SQL Server using the web scraping program. Scores of a total of 11347 basketball matches range from NBA 2002/2003 League to mid NBA 2010/2011 game. Besides, all available Over/Under odds of a total of 2305 matches have been gathered. The table below illustrates the distribution of the matches and odds information of each league.

Table 1: Distribution of data grouping by leagues

League	Number of	
	Matches	Over/Under odds
NBA 2002/2003	1260	-
NBA 2003/2004	1268	-
NBA 2004/2005	1311	-
NBA 2005/2006	1319	-
NBA 2006/2007	1309	-
NBA 2007/2008	1316	-
NBA 2008/2009	1315	185
NBA 2009/2010	1312	10352
NBA 2010/2011	937	7819
Total	11347	18356

Please refer Appendix where a snapshot of data is shown there.

3.5 Data Verification and Cleaning

Before the data is ready for analysis and queries, the data needs to be clean and verified first. This is because information in the Internet is posted by human and thus human error may exist inside the data. One of the ways to identify the inconsistencies is to compare the total scores of both teams with the sum of the quarters' score of both teams. To identify potential irregularities in the data, outliers are being checked and validated. After a round of cleaning, the new distribution of matches and odds information in each league is displayed in the table below:

Table 2: Distribution of data grouping by leagues (after data cleaning)

League	Number of	
	Matches	Over/Under odds
NBA 2002/2003	1248	-
NBA 2003/2004	1259	-
NBA 2004/2005	1311	-
NBA 2005/2006	1312	-
NBA 2006/2007	1308	-
NBA 2007/2008	1314	-
NBA 2008/2009	1313	181
NBA 2009/2010	1312	10352
NBA 2010/2011	937	7743
Total	11314	18276

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Preliminary Analysis

Since data for matches in NBA 2010/2011 League will be taken as validation of the upcoming model, it will not be included in the analysis part.

First we look at the mean and standard deviation of numerous variables in the data. The table below shows the mean and standard deviation of numerous variables (without grouping according to league).

Table 3: Mean and Standard Deviation of Various Variables

Variable	Mean	Standard Deviation
Total	195.1881	21.1670
Q1Total	49.1132	8.1886
Q2Total	48.5441	8.1837
Q3Total	48.0221	8.0507
Q4Total	48.0250	8.7642
HomeTotal	99.3024	12.4126
HomeQ1	25.1155	5.5999
HomeQ2	24.7546	5.5290
HomeQ3	24.4762	5.5581
HomeQ4	24.2039	5.6543
AwayTotal	95.8857	12.3257
AwayQ1	23.9978	5.4853
AwayQ2	23.7895	5.4408
AwayQ3	23.5459	5.5457
AwayQ4	23.8211	5.7286

More detailed information about the mean and standard deviation of the variables are attached in the Appendix.

Let q_1 = mean of total points scored in quarter 1

q_2 = mean of total points scored in quarter 2

q_3 = mean of total points scored in quarter 3

q_4 = mean of total points scored in quarter 4

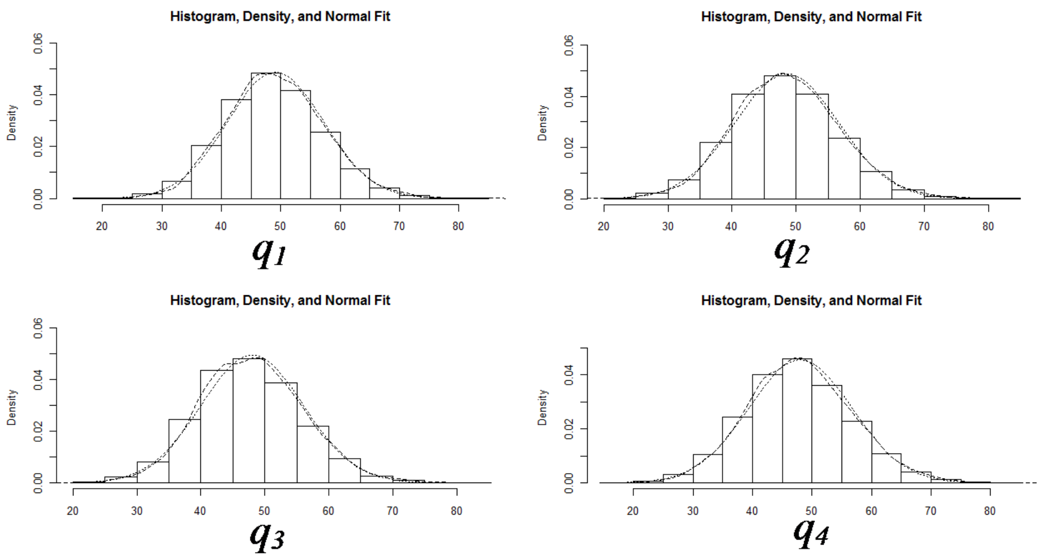


Figure 3: Histogram of q_1, q_2, q_3, q_4

Four histogram of total points scored in each of the 4 quarters has been plotted. Based on the figure above, each histogram does looks similar to normal distribution. An initial impression is that q_1, q_2, q_3, and q_4 behaves normally.

To further investigate this, we would need a formal test of normality:

Shapiro-Wilk test for normality

5000 random data is chosen for this normality test. The data is then tested by Shapiro-Wilk test with significance level $\alpha = 0.05$. The result obtained using R is as below:

	Results
q_1	W = 0.9952, p-value = 8.974e-12
q_2	W = 0.9975, p-value = 2.53e-07
q_3	W = 0.9965, p-value = 1.896e-09
q_4	W = 0.9944, p-value = 5.005e-13

Recalling that the null hypothesis is that the population is normally distributed, if p-value < α then the null hypothesis is rejected; we are forced to conclude that all four data are not from a normally distributed population by Shapiro-Wilk test on $\alpha = 0.05$.

The test exhibits an odd and unusual result, contradicting to the first impression had after looking at the histograms. Further investigation reveals why. This is because Shapiro-Wilk test does not work well when several values in the data is the same.

We proceed and make use of the Central Limit Theorem to assume normality:

If the sample size is sufficiently large, then the mean of a random sample from a population has a sampling distribution that is approximately normal.

We further verify this by plotting Q-Q plots for q_1, q_2, q_3, and q_4 respectively.

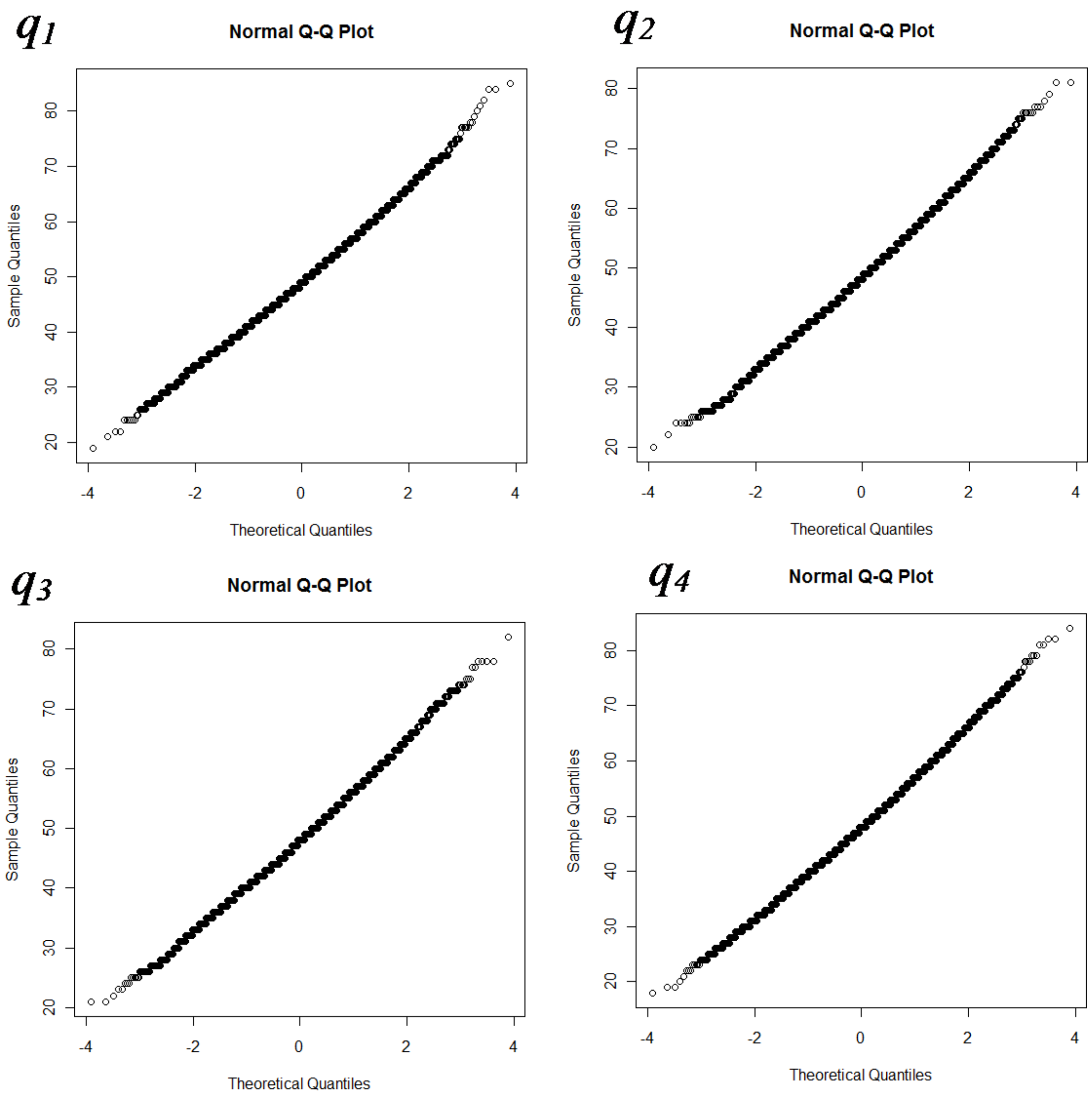


Figure 4 : qq-plot of q_1, q_2, q_3, q_4

The normal Q-Q plots supports the normality assumption as well. Therefore we can now proceed to hypothesis testing on q_1, q_2, q_3, and q_4.

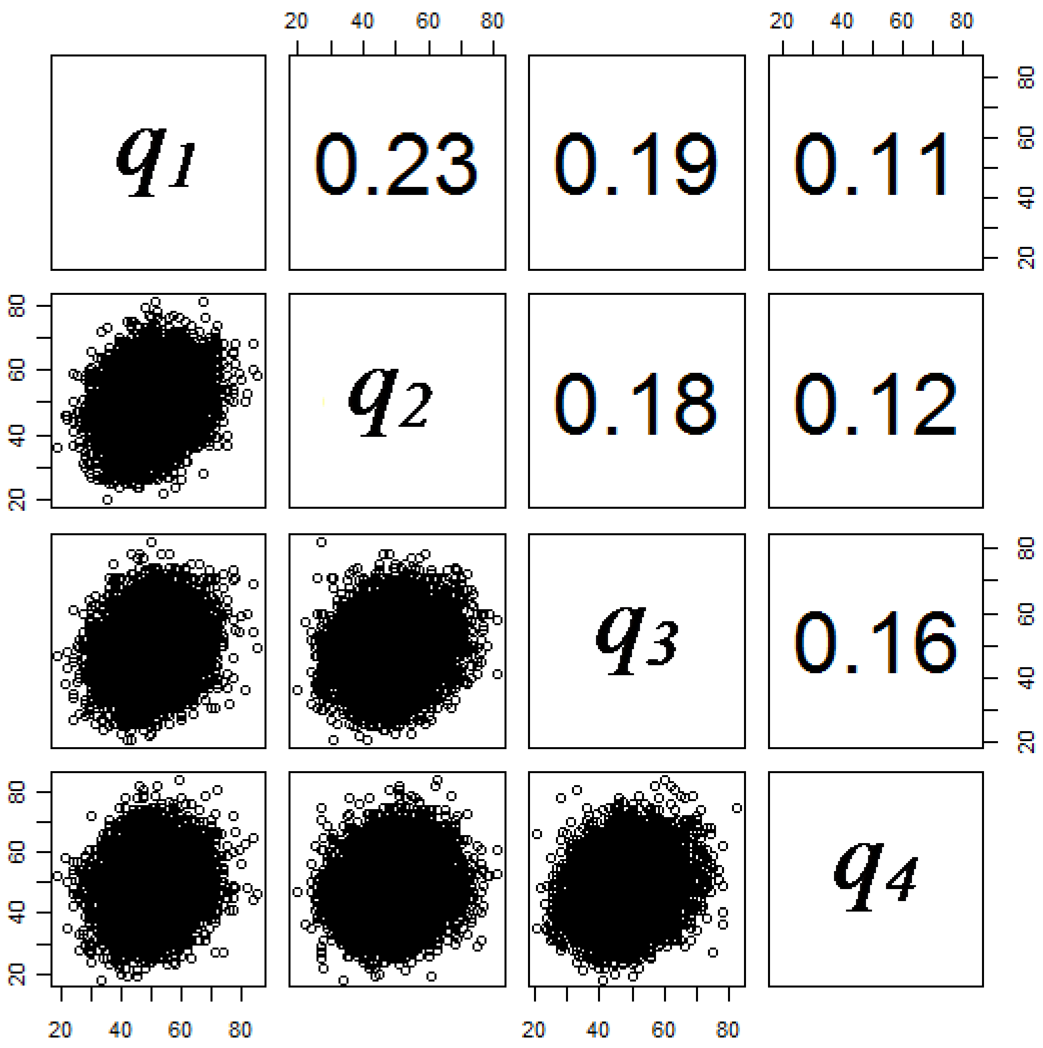


Figure 5: Correlation Coefficient and Matrix plot

Figure above is a scatterplot matrix where the lower triangle consists of scatterplots whereas the upper triangle consists of the Pearson's correlation coefficient, each corresponds to their respective q_i , q_j pairs where $i \neq j$, i,j = 1,2,3,4.

From the scatterplots we can see that the relationship between all the q_i , q_j pairs is not linear. Besides, the scatterplot suggests that the linear correlation between the respective pairs are weak.

On the other hand, looking at the Pearson's correlation coefficients, it proposes that the linear correlation of respective pairs are weak. In fact, the further the quarters are separated from each other in the course of basketball game, the weaker linear correlation it exhibits.

Looking back to Table 3, an initial guess is that there are most points scored in quarter 1 followed by quarter 2, quarter 4 and quarter 3. Thus, we conduct a few hypothesis tests to further verify this.

Paired t-test has been done on each pair of q_i and q_j where $i \neq j$, $i,j=1,2,3,4$. The results is demonstrated below.

Hypothesis testing #1**: $q_1 - q_2 > 0$**

H_0: $q_1 - q_2 \leq 0$

H_1: $q_1 - q_2 > 0$

t = 5.6985, df = 10376, p-value = 6.21e-09

95 percent confidence interval: 0.404846 Inf

mean of the differences: 0.5691433

Conclusion: p-value < $\alpha = 0.05$. Thus, reject H_0. We are 95% confident that in each match, there are significantly more points scored in quarter 1 than quarter 2.

Conducting the rest in a similar way and the results are summarized into Table.

Table 4: Hypothesis Testings on Quarter Scores

	Null Hypothesis, H_0	Alternative Hypothesis, H_1	p-value	Action taken with $\alpha = 0.05$	Conclude on 95% confidence
1	$q_1 - q_2 \leq 0$	$q_1 - q_2 > 0$	6.21e-09	Reject H_0	$q_1 > q_2$
2	$q_1 - q_3 \leq 0$	$q_1 - q_3 > 0$	< 2.2e-16	Reject H_0	$q_1 > q_3$
3	$q_1 - q_4 \leq 0$	$q_1 - q_4 > 0$	< 2.2e-16	Reject H_0	$q_1 > q_4$
4	$q_2 - q_3 \leq 0$	$q_2 - q_3 > 0$	1.685e-07	Reject H_0	$q_2 > q_3$
5	$q_2 - q_4 \leq 0$	$q_2 - q_4 > 0$	1.323e-06	Reject H_0	$q_2 > q_4$
6	$q_3 - q_4 \leq 0$	$q_3 - q_4 > 0$	0.5108	Do not reject H_0	$q_3 \leq q_4$

From the 6 hypothesis testing conducted, we can reasonably conclude at $\alpha = 0.05$ that in a basketball game, the total points scored in quarter 1 is significantly greater than that of quarter 2, 3 and 4.

We may also rationally states that $q_1 > q_2 > q_3$ and $q_1 > q_2 > q_4$ and $q_3 \leq q_4$. The reason is that the p-values obtained from the tests are very small. Restating the inequalities, we eventually get $q_1 > q_2 > q_4 \leq q_3$.

Therefore, we had accomplish one of the objective and conclude that, in NBA basketball league, the 1st quarter has significantly more points than others, followed by the 2nd quarter, then the 4th quarter and lastly the 3rd quarter.

This conclusion is sensible becuase players tends to be fresh, energetic and active in the 1st quarter. Besides, one team will try to score more to be the point lead. Thus, we can see most points scored in the 1st quarter. As time goes, the players tends to be more exhausted and fatigueness affects their shooting rate. Besides, coach of a team would analyze the previous quarter and apply new tactics on the next quarter to overcome another team's strategy. Thus, players tends to score less in the 2nd quarter and lesser in the 3rd quarter. However, at the 4th quarter, the losing team will try to win and the winning team will defend their lead. Thus, more actions is in the 4th quarter and thus, more points is scored.

CHAPTER 5

IMPLEMENTATIONS AND MODEL DEVELOPMENT

5. 1 Model Development

This project also aims to build predictive models to predict total points of basketball matches based on Over/Under odds.

First of all, lets recall that the data contains 18276 Over/Under odds that corresponds to 2305 matches. The data will be used in two parts: one part for model building and another part for model validation. The data for model building consists of data from NBA 2008/2009 - 2009/2010 Leagues and the remaining NBA 2010/2011 League data will be use to validate the model.

Distribution of the data for model building:

Odds Type	Company	Corresponding Matches	Number of Odds
Over/Under	188Bet	1277	6079
	SBOBet	1361	4454
Total			10533

Distribution of the data for model validation:

Odds Type	Company	Corresponding Matches	Number of Odds
Over/Under	188Bet	929	4469
	SBOBet	934	3274
Total			7743

To make model comparison, two sets of predictive model are developed. One set of predictive models is based on the odds offered by 188Bet whereas another set of the predictive models is based on the odds offered by SBOBet. With two sets of predictive model, we can evaluate whether the odds offered by 188Bet or SBOBet better predicts the total points.

5.2 Predictive Model to Predict Total Points based on Over/Under Odds

One of the objectives of this project is to investigate the accuracy of odds in predicting the total points. Therefore, in this project, simple linear regression models are applied to inspect whether there exist a relationship between the expected total points from odds and the actual total points.

The basic predictive model is developed using expected total points and actual total points only. As expected, the results obtained are not very useful. Thus, to improve the prediction of the model, the current score is added into the model as an additional regressor. For example, with the information of the first quarter's score, can the model predict the total point better? This is one of the project objectives. Besides, this project focuses mainly on the multiple linear regression model.

For comparison purpose, two predictive models for total points are developed. One is based on the Over/Under odds from 188Bet and another is based on the Over/Under odds from SBOBet.

Converting Over/Under Odds To Expected Total

MatchID	Total	Over	Under	Company	OddsType	Probability of Over	Probability of Under	Expected Total
938	186.0	1.84	2.04	188Bet	OU	0.5258	0.4742	182.2672
938	186.5	1.90	1.98	188Bet	OU	0.5103	0.4897	187.0066
938	187.0	1.98	1.90	188Bet	OU	0.4897	0.5103	186.4934
938	187.5	2.14	1.76	188Bet	OU	0.4513	0.5487	185.1003
							Average	186.4669

Above is a table showing the details of Over/Under odds by 188Bet corresponding to match#938. The formula below is used to convert O/U odds into probabilities:

$$Probability\ of\ Over = \frac{Odds\ for\ Under}{Odds\ for\ Over + Odds\ for\ Under}$$

Likewise,

$$Probability\ of\ Under = \frac{Odds\ for\ Over}{Odds\ for\ Over + Odds\ for\ Under}$$

To find the Expected Total from the odds, the following process is gone through, demonstrated using data of the first row:

$$P(X < 186.0\ ) = 0.4742$$

$$P\left(\frac{X-\mu_x}{\sigma_x} < \frac{186.0-\mu_x}{\sigma_x}\right) = 0.4742$$

$$P\left(z < \frac{186.0-\mu_x}{\sigma_x}\right) = 0.4742 \qquad (1)$$

$$P\,(z < z_\alpha) = 0.4742 \qquad (2)$$

Comparing (1) and (2):

$$\frac{186.0-\mu_x}{\sigma_x} = z_\alpha$$

$$\mu_x = 186.0 - z_\alpha * \sigma_x \qquad (*)$$

where μ_x is the expected total implied by the odds, z_α is the inverse of the normal cumulative distribution and σ_x is the sample standard deviation of total points.

Lastly, take the average as the expected total for the match, which is 186.4669.

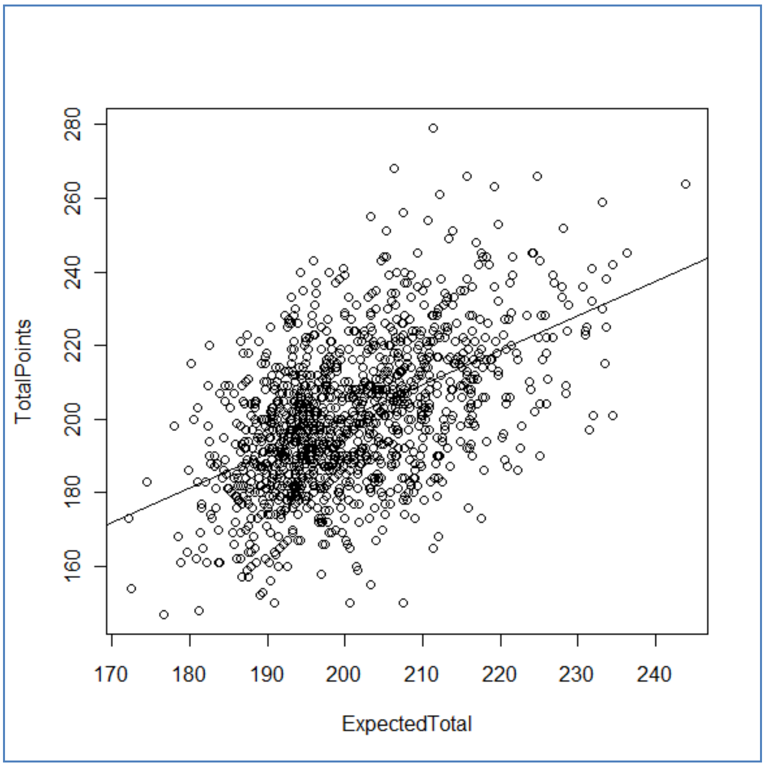
5.2.1 Regression Models using 188Bet Odds

Model Number 1: $TotalPoints = \beta_0 + \beta_1 * ExpectedTotal$

```
lm(formula = data$TotalPoints ~ data$ExpectedTotal)

Residuals:
    Min      1Q  Median      3Q     Max 
-56.980 -11.214  -0.362  10.261  68.386 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)         13.4681     8.8058   1.529    0.126    
data$ExpectedTotal   0.9326     0.0438  21.291   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.84 on 1275 degrees of freedom
Multiple R-squared: 0.2623,     Adjusted R-squared: 0.2617 
F-statistic: 453.3 on 1 and 1275 DF,  p-value: < 2.2e-16
```

Resulting Model:

$TotalPoints = 13.4681 + 0.9326 * ExpectedTotal$



A graph of Total Points versus Expected Total implied by odds is plotted. The slope of the straight line is = 0.9326, which is the coefficient of ExpectedTotal in the model.

Continuing in a similar manner and proceed to the next model.

Model Number 2:

$TotalPoints = \beta_0 + \beta_1 * TotalPointsUpToQ1 + \beta_2 * ExpectedTotal$

```
lm(formula = data$TotalPoints ~ data$Q1Total + data$ExpectedTotal)

Residuals:
    Min      1Q  Median      3Q     Max
-50.208  -9.856  -0.374   9.420  60.963

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        15.81856    7.77571   2.034   0.0421 *
data$Q1Total        1.09524    0.05760  19.015   <2e-16 ***
data$ExpectedTotal  0.64054    0.04161  15.394   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.87 on 1274 degrees of freedom
Multiple R-squared: 0.4254,     Adjusted R-squared: 0.4245
F-statistic: 471.6 on 2 and 1274 DF,  p-value: < 2.2e-16
```

Resulting Model:

$TotalPoints = 15.8186 + 1.09524 * TotalPointsUpToQ1 + 0.6405 * ExpectedTotal$

Model Number 3:

$TotalPoints = \beta_0 + \beta_1 * TotalPointsUpToQ2 + \beta_2 * ExpectedTotal$

```
lm(formula = data$TotalPoints ~ data$UptoQ2Total + data$ExpectedTotal)

Residuals:
    Min      1Q  Median      3Q     Max
-38.420  -8.868  -0.427   7.983  47.860

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        14.59459    6.51759   2.239   0.0253 *
data$UptoQ2Total    1.06364    0.03277  32.457   <2e-16 ***
data$ExpectedTotal  0.38662    0.03652  10.586   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.47 on 1274 degrees of freedom
Multiple R-squared: 0.5962,     Adjusted R-squared: 0.5956
F-statistic: 940.5 on 2 and 1274 DF,  p-value: < 2.2e-16
```

Resulting Model:

$TotalPoints = 14.5946 + 1.0636 * TotalPointsUpToQ2 + 0.3866 * ExpectedTotal$

Model Number 4:

$$TotalPoints = \beta_0 + \beta_1 * TotalPointsUpToQ3 + \beta_2 * ExpectedTotal$$

```
lm(formula = data$TotalPoints ~ data$UptoQ3Total + data$ExpectedTotal)

Residuals:
    Min      1Q  Median      3Q     Max 
-25.524  -6.396  -0.788   5.186  53.753 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        16.40468    5.03095   3.261  0.00114 ** 
data$UptoQ3Total    1.00367    0.01956  51.309  < 2e-16 ***
data$ExpectedTotal  0.16399    0.02916   5.623 2.30e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.622 on 1274 degrees of freedom
Multiple R-squared: 0.7594,     Adjusted R-squared: 0.759 
F-statistic:  2011 on 2 and 1274 DF,  p-value: < 2.2e-16
```

Resulting Model:

$$TotalPoints = 16.4047 + 1.0037 * TotalPointsUpToQ3 + 0.1640 * ExpectedTotal$$

5.2.2 Regression Models using SBOBet Odds

Model Number 5: $TotalPoints = \beta_0 + \beta_1 * ExpectedTotal$

```
lm(formula = data$TotalPoints ~ data$ExpectedTotal)

Residuals:
    Min       1Q   Median       3Q      Max
-56.950 -11.273  -0.413  10.498  67.436

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        12.02657    8.44524   1.424    0.155
data$ExpectedTotal  0.93989    0.04212  22.314   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.93 on 1359 degrees of freedom
Multiple R-squared: 0.2681,     Adjusted R-squared: 0.2676
F-statistic: 497.9 on 1 and 1359 DF,  p-value: < 2.2e-16
```

Resulting Model:

$TotalPoints = 12.0266 + 0.9399 * ExpectedTotal$

Model Number 6:

$TotalPoints = \beta_0 + \beta_1 * TotalPointsUpToQ1 + \beta_2 * ExpectedTotal$

```
lm(formula = data$TotalPoints ~ data$Q1Total + data$ExpectedTotal)

Residuals:
    Min       1Q   Median       3Q      Max
-50.400   -9.911  -0.402   9.413  60.355

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        14.96165    7.49516   1.996   0.0461 *
data$Q1Total        1.07513    0.05604  19.186   <2e-16 ***
data$ExpectedTotal  0.65050    0.04030  16.140   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.02 on 1358 degrees of freedom
Multiple R-squared: 0.4242,     Adjusted R-squared: 0.4234
F-statistic: 500.2 on 2 and 1358 DF,  p-value: < 2.2e-16
```

Resulting Model:

$TotalPoints = 14.9617 + 1.07513 * TotalPointsUpToQ1 + 0.6505 * ExpectedTotal$

Model Number 7:

$$TotalPoints = \beta_0 + \beta_1 * TotalPointsUpToQ2 + \beta_2 * ExpectedTotal$$

```
lm(formula = data$TotalPoints ~ data$UptoQ2Total + data$ExpectedTotal)

Residuals:
    Min      1Q  Median      3Q     Max 
-38.581  -8.861  -0.459   7.757  48.727 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        14.12236    6.31063   2.238   0.0254 *  
data$UptoQ2Total    1.05757    0.03224  32.804   <2e-16 ***
data$ExpectedTotal  0.39246    0.03562  11.017   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 12.65 on 1358 degrees of freedom
Multiple R-squared: 0.5917,     Adjusted R-squared: 0.5911 
F-statistic:   984 on 2 and 1358 DF,  p-value: < 2.2e-16
```

Resulting Model:

$$TotalPoints = 14.1224 + 1.0576*TotalPointsUpToQ2 + 0.3925*ExpectedTotal$$

Model Number 8:

$$TotalPoints = \beta_0 + \beta_1 * TotalPointsUpToQ3 + \beta_2 * ExpectedTotal$$

```
lm(formula = data$TotalPoints ~ data$UptoQ3Total + data$ExpectedTotal)

Residuals:
    Min      1Q  Median      3Q     Max 
-26.465  -6.384  -0.798   5.354  53.745 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        15.89970    4.88401   3.255  0.00116 ** 
data$UptoQ3Total    1.00924    0.01940  52.023  < 2e-16 ***
data$ExpectedTotal  0.16271    0.02857   5.695 1.51e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 9.788 on 1358 degrees of freedom
Multiple R-squared: 0.7555,     Adjusted R-squared: 0.7551 
F-statistic:  2098 on 2 and 1358 DF,  p-value: < 2.2e-16
```

Resulting Model:

$$TotalPoints = 15.8997 + 1.0092\ TotalPointsUpToQ3 + 0.1627*ExpectedTotal$$

5.2.3 Summary of Predictive Model using Over/Under Odds

	R-Squared	
	188Bet	SBOBet
TotalPoints = $\beta_0 + \beta_1$*ExpectedTotal	0.2623	0.2681
TotalPoints = $\beta_0 + \beta_1$*TotalPointsUpToQ1 + β_2*ExpectedTotal	0.4254	0.4242
TotalPoints = $\beta_0 + \beta_1$*TotalPointsUpToQ2 + β_2*ExpectedTotal	0.5962	0.5917
TotalPoints = $\beta_0 + \beta_1$*TotalPointsUpToQ3 + β_2*ExpectedTotal	0.7594	0.7555

The R-squared of a regression can be interpreted as the amount of the variance in the dependent variable that is by the model. For example, if the R-squared value is 1.0, this means that the model's prediction will have perfect accuracy. Though so, an R-squared value of 1.0 is not very likely to happen in the real world.

One distinctive change in the R-Square is observed. As more information about the quarters' scores is provided to the model, the R-Squared value increases. This means that, the prediction of total points using pre-match Over/Under odds may be improved using the current score. To further verify this, the models will go through a validation process.

Looking at the table above, it is found that the R-Squared value for the 4 predictive models based on 188Bet and another 4 based on SBOBet is relatively the same. However, this may not imply that both models will have the same prediction. Therefore, the models are to be validated with a new data, which is discussed in Chapter 5.3.

5.3 Model Validation

8 predictive models have been developed to predict the total points of a basketball game based on pre-match odds. The first 4 models are developed based on 188Bet's Over/Under odds whereas the remaining 4 models are developed based on SBOBet's Over/Under odds.

Model	Model Equation
1	TotalPoints = 13.4681 + 0.9326*ExpectedTotal
2	TotalPoints = 15.8186 + 1.0952*TotalPointsUpToQ1 + 0.6405*ExpectedTotal
3	TotalPoints = 14.5946 + 1.0636*TotalPointsUpToQ2 + 0.3866*ExpectedTotal
4	TotalPoints = 16.4047 + 1.0037*TotalPointsUpToQ3 + 0.1640*ExpectedTotal
5	TotalPoints = 12.0266 + 0.9399*ExpectedTotal
6	TotalPoints = 14.9617 + 1.0751*TotalPointsUpToQ1 + 0.6505*ExpectedTotal
7	TotalPoints = 14.1224 + 1.0576*TotalPointsUpToQ2 + 0.3925*ExpectedTotal
8	TotalPoints = 15.8997 + 1.0092*TotalPointsUpToQ3 + 0.1627*ExpectedTotal

Each model is validated using the remaining NBA 2010/2011 League data. The screenshot below gives a snapshot of the validation process. For each of the odds, the Expected Total is calculated. Then, using the model's equation, the Predicted Total Point is worked out. Then, we check if the predicted total points lies at the same side with the actual total points when compared with "Total", which is part of Over/Under odds. Then, the success rate of prediction is recorded.

	A	B	C	D	E	F	G	H	I	J	K	L
1	TotalPoints = 12.0266 + 0.9399*ExpectedTotal											
2												
3	MatchID	Actual Total Points	Predicted Total Points	Expected Total	Total	Over	Under	Predicted > Total?	Actual > Total?	Success Prediction?		Success Rate
4	1	172	182.3437057	181.207688	181.0	1.96	1.92	TRUE	FALSE	FALSE		52.2149%
5	1	172	182.3437057	181.207688	181.5	2.02	1.85	TRUE	FALSE	FALSE		
6	1	172	182.3437057	181.207688	181.5	1.81	2.05	TRUE	FALSE	FALSE		
7	1	172	182.3437057	181.207688	182.0	1.96	1.92	TRUE	FALSE	FALSE		
8	1	172	182.3437057	181.207688	182.0	2.05	1.83	TRUE	FALSE	FALSE		
9	1	172	182.3437057	181.207688	182.5	2.05	1.83	FALSE	FALSE	TRUE		
10	1	172	182.3437057	181.207688	182.5	2.12	1.77	FALSE	FALSE	TRUE		
11	1	172	182.3437057	181.207688	183.0	2.07	1.82	FALSE	FALSE	TRUE		
12	1	172	182.3437057	181.207688	183.5	2.08	1.78	FALSE	FALSE	TRUE		
13	2	207	200.874743	200.923655	200.0	1.9	1.96	TRUE	TRUE	TRUE		
14	2	207	200.874743	200.923655	201.0	1.9	1.98	FALSE	TRUE	FALSE		
15	2	207	200.874743	200.923655	201.0	1.96	1.92	FALSE	TRUE	FALSE		
16	2	207	200.874743	200.923655	201.5	1.94	1.94	FALSE	TRUE	FALSE		
17	2	207	200.874743	200.923655	201.5	2.02	1.86	FALSE	TRUE	FALSE		

Doing this for each model, the success rate of the 8 models is obtained and tabulated in table below.

	Success Rate	
	188Bet	SBOBet
TotalPoints = $\beta_0 + \beta_1$*ExpectedTotal	52.0212%	52.2149%
TotalPoints = $\beta_0 + \beta_1$*TotalPointsUpToQ1+ β_2*ExpectedTotal	67.1574%	67.3512%
TotalPoints = $\beta_0 + \beta_1$*TotalPointsUpToQ2 + β_2*ExpectedTotal	74.1056%	74.0540%
TotalPoints = $\beta_0 + \beta_1$*TotalPointsUpToQ3 + β_2*ExpectedTotal	80.1111%	80.1756%

Analyzing the table above, it is found that the success rate of predictions of the models that based on 188Bet and SBOBet are similar.

With just the pre-match odds, we can predict the score at a success rate of approximately 52%. If we're given the scores up to the end of first quarter, we can improve the prediction to roughly 67%. If we're given the scores up to the end of second quarter, the prediction is improved to approximately 74%. Then, if we also had the quarter 3 scores, we may predict the total points 80% of the time.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

6. 1 Conclusion

We have achieved the objectives and aims in this project. Firstly, it is found that, in a basketball game, the first quarter has significantly more points than the others, followed by the second quarter, fourth quarter and third quarter. Besides, it is found that there exist weak linear correlations between the pair of total points in two quarters. Furthermore, the further the two quarters is separated, the weaker is the linear correlation.

With the pre-match odds, we can make use of linear models to predict the total points at the end of the match. With only the odds itself, the success rate of prediction is roughly 52%. If the current score is known, it will improve the prediction success rate. If the score up to the end of first quarter is known, the prediction success rate improves by 15% to approximately 67%. If the score up to the end of second quarter is known, the prediction success rate improves by 7% to around 74%. If the scores up to the end of third quarter is known, the prediction success rate increases by 6% to 80%. Therefore, we can conclude that with information about the current score, the odds predicts the total points of a basketball match more accurately.

REFERENCES

Adamantios Diamantopoulos & Bodo B. Schlegelmilch (2006) Taking the fear out of data analysis: a step-by-step approach. Thomson Learning.

Dixon, M. J. & Coles, S. G. (1997) Modelling association football scores and inefficiencies in the football betting market. *Appl. Statist.*, **46,** 265-280

Dixon, M. J. & Robinson, M. E. (1998) A birth process model for association football matches. *The Statistician*, **47**, 523-538

Harville, D (1980) Predictions for National Football League Games Via Linear-Model Methodology. *Journal of the American Statistical Association*, Vol. 75, No. 371, 516-524

Hogg , R. V. & Tanis, E. A. (2000) Probability and Statistical Inference (6^{th} ed). Prentice Hall.

Iversen, G. & Gergen, M.(1997) Statistics: the conceptual approach. Springer.

Zuber, R.A. & Gnadar, J.M. & Bowers, B.D. (1985) Beating the Spread: Testing the Efficiency of the Gambling Market for National FootballLeague Games. *Journal of Political Economy*, Vol. 93, No. 4, 800-806

APPENDICES

APPENDIX A

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ID	HomeTeam	AwayTeam	Date	HomeTotal	AwayTotal	HomeQ1	AwayQ1	HomeQ2	AwayQ2	HomeQ3	AwayQ3	HomeQ4	AwayQ4	OTIndicator	LeagueName
2	1	Milwaukee	Boston Celtics	07.03.2011	83	89	26	27	23	16	16	26	18	20	0	NBA 2010/2011
3	2	Dallas Maver.	Memphis	07.03.2011	103	104	27	21	28	17	23	41	25	25	0	NBA 2010/2011
4	3	Oklahoma	Phoenix Suns	07.03.2011	122	118	25	28	32	33	29	30	23	18	1	NBA 2010/2011
5	4	Atlanta Hawks	New York	07.03.2011	79	92	19	24	23	20	14	18	23	30	0	NBA 2010/2011
6	5	Cleveland	New Orleans	07.03.2011	81	96	24	27	24	24	18	22	15	23	0	NBA 2010/2011
7	6	Philadelphia 76ers	Golden State	07.03.2011	125	117	31	25	33	31	16	24	26	26	1	NBA 2010/2011
8	7	Detroit Pistons	Washington Wiz.	07.03.2011	113	102	29	22	31	29	22	22	31	29	0	NBA 2010/2011
9	8	San Antonio	L.A.Lakers	06.03.2011	83	99	13	34	24	31	15	16	31	18	0	NBA 2010/2011
10	9	Miami	Chicago Bulls	06.03.2011	86	87	22	18	27	22	16	23	21	24	0	NBA 2010/2011
11	10	L.A.Clippers	Denver	06.03.2011	100	94	30	23	18	15	23	27	29	29	0	NBA 2010/2011
12	11	Portland	Charlotte	06.03.2011	93	69	15	16	25	19	25	17	28	17	0	NBA 2010/2011
13	12	Utah	Sacramento	06.03.2011	109	102	30	29	27	20	15	23	23	23	1	NBA 2010/2011
14	13	Houston	Indiana	06.03.2011	112	95	29	18	36	31	29	19	18	27	0	NBA 2010/2011
15	14	Washington Wiz.	Minnesota	06.03.2011	103	96	25	29	26	20	17	23	35	24	0	NBA 2010/2011
16	15	L.A.Lakers	Charlotte	05.03.2011	92	84	29	19	16	16	24	22	23	27	0	NBA 2010/2011
17	16	San Antonio	Miami	05.03.2011	125	95	36	12	26	38	32	22	31	23	0	NBA 2010/2011
18	17	Milwaukee	Phoenix Suns	05.03.2011	88	102	26	22	22	23	15	31	25	26	0	NBA 2010/2011
19	18	Dallas Maver.	Indiana	05.03.2011	116	108	35	33	33	24	22	24	26	27	0	NBA 2010/2011
20	19	Memphis	New Orleans	05.03.2011	91	98	20	20	25	25	27	27	19	26	0	NBA 2010/2011
21	20	New York	Cleveland	05.03.2011	115	119	32	32	32	26	24	27	27	34	0	NBA 2010/2011
22	21	Boston Celtics	Golden State	05.03.2011	107	103	32	27	32	26	24	28	19	22	0	NBA 2010/2011
23	22	Atlanta Hawks	Oklahoma	05.03.2011	104	111	26	31	25	25	22	27	31	28	0	NBA 2010/2011
24	23	Orlando	Chicago Bulls	05.03.2011	81	89	21	21	14	28	22	16	24	24	0	NBA 2010/2011
25	24	Philadelphia 76ers	Minnesota	05.03.2011	111	100	22	31	34	18	29	29	26	22	0	NBA 2010/2011
26	25	New Jersey Nets	Toronto Raptors	04.03.2011	116	103	27	23	20	28	31	30	38	22	0	NBA 2010/2011
27	26	New Jersey Nets	Toronto Raptors	04.03.2011	137	136	23	30	42	32	23	26	22	22	1	NBA 2010/2011

	A	B	C	D	E	F	G	H
1	ID	MatchID	First	Second	Third	OddsCode	Company	Status
2	1	1	181	1.96	1.92	OU	188Bet	1
3	2	1	181.5	2.02	1.85	OU	188Bet	1
4	3	1	181.5	1.81	2.05	OU	Sbobet	1
5	4	1	182	1.96	1.92	OU	188Bet	1
6	5	1	182	2.05	1.83	OU	Sbobet	1
7	6	1	182.5	2.05	1.83	OU	188Bet	1
8	7	1	182.5	2.12	1.77	OU	Sbobet	1
9	8	1	183	2.07	1.82	OU	188Bet	1
10	9	1	183.5	2.08	1.78	OU	188Bet	1
11	15	2	200	1.9	1.96	OU	Sbobet	1
12	16	2	201	1.9	1.98	OU	188Bet	1
13	17	2	201	1.96	1.92	OU	Sbobet	1
14	18	2	201.5	1.94	1.94	OU	188Bet	1
15	19	2	201.5	2.02	1.86	OU	Sbobet	1
16	24	3	210.5	1.78	2.11	OU	Sbobet	1
17	25	3	211	1.84	2.04	OU	188Bet	1
18	26	3	211	1.85	2.03	OU	Sbobet	1
19	27	3	211.5	1.92	1.96	OU	188Bet	1
20	28	3	211.5	1.97	1.91	OU	Sbobet	1
21	29	3	212	1.98	1.9	OU	188Bet	1
22	30	3	212	2.03	1.85	OU	Sbobet	1
23	31	3	212.5	2.06	1.8	OU	188Bet	1
24	32	3	212.5	2.11	1.76	OU	Sbobet	1
25	38	4	202.5	2	1.88	OU	188Bet	1
26	39	4	203	2.05	1.84	OU	188Bet	1
27	40	4	203.5	2.08	1.8	OU	188Bet	1

League	Mean of Total	Q1Total	Q2Total	Q3Total	Q4Total
NBA 2002/2003	190.0473	47.9808	46.9928	46.8422	46.8397
NBA 2003/2004	186.1461	46.8642	46.2025	45.9341	45.8578
NBA 2004/2005	194.3791	48.6850	48.1671	48.0671	47.9283
NBA 2005/2006	194.1707	48.6723	48.1044	47.7073	47.8933
NBA 2006/2007	196.9419	49.2003	48.6865	48.5168	48.7156
NBA 2007/2008	199.1674	50.1903	49.4939	49.1012	49.1400
NBA 2008/2009	199.5050	49.8233	49.9231	49.0518	49.1462
NBA 2009/2010	200.5267	51.3415	50.6098	48.8133	48.5328
On all matches	195.1881	49.1132	48.5441	48.0221	48.0250

League	Standard deviation of Total	Q1Total	Q2Total	Q3Total	Q4Total
NBA 2002/2003	19.9594	7.8608	7.8040	7.8967	8.5650
NBA 2003/2004	20.7800	8.0607	8.2770	7.8577	8.9266
NBA 2004/2005	20.3347	7.9841	7.9350	7.8309	8.8982
NBA 2005/2006	20.4364	8.1059	7.9758	7.9576	8.5660
NBA 2006/2007	21.0867	8.0870	8.1504	8.2179	8.6262
NBA 2007/2008	21.7136	8.6662	8.2301	8.1023	8.8899
NBA 2008/2009	21.2211	8.1372	8.0882	7.9057	8.6146
NBA 2009/2010	19.6036	7.7712	8.0837	8.0921	8.5133
On all matches	21.1670	8.1886	8.1837	8.0507	8.7642

League	Mean of				
	HomeTotal	HomeQ1	HomeQ2	HomeQ3	HomeQ4
NBA 2002/2003	97.0104	24.7147	24.1082	23.9103	23.5649
NBA 2003/2004	94.9627	23.9976	23.6410	23.4170	23.2518
NBA 2004/2005	98.7895	24.8116	24.6690	24.3982	24.1449
NBA 2005/2006	98.8232	24.8003	24.3963	24.4467	24.2782
NBA 2006/2007	100.0176	25.1950	24.8395	24.7378	24.3096
NBA 2007/2008	101.4361	25.6248	25.3029	25.1005	24.7511
NBA 2008/2009	101.4052	25.4882	25.3839	24.8919	24.8507
NBA 2009/2010	101.6845	26.2256	25.6181	24.8361	24.4093
On all matches	99.3024	25.1155	24.7546	24.4762	24.2039

League	Standard deviation of				
	HomeTotal	HomeQ1	HomeQ2	HomeQ3	HomeQ4
NBA 2002/2003	11.8378	5.3733	5.4200	5.5410	5.6296
NBA 2003/2004	11.9978	5.3501	5.5175	5.2673	5.4899
NBA 2004/2005	11.7073	5.4066	5.3506	5.3234	5.6401
NBA 2005/2006	11.7839	5.4854	5.4367	5.6141	5.4618
NBA 2006/2007	12.4148	5.6722	5.4177	5.6267	5.6795
NBA 2007/2008	12.9123	5.9038	5.5805	5.5150	5.8910
NBA 2008/2009	12.6045	5.7662	5.5386	5.6404	5.6452
NBA 2009/2010	12.4010	5.5246	5.6831	5.7293	5.6128
On all matches	12.4126	5.5999	5.5290	5.5581	5.6543

League	Mean of				
	AwayTotal	AwayQ1	AwayQ2	AwayQ3	AwayQ4
NBA 2002/2003	93.0369	23.2660	22.8846	22.9319	23.2748
NBA 2003/2004	91.1835	22.8666	22.5616	22.5171	22.6060
NBA 2004/2005	95.5896	23.8734	23.4981	23.6690	23.7834
NBA 2005/2006	95.3476	23.8720	23.7081	23.2607	23.6151
NBA 2006/2007	96.9243	24.0054	23.8471	23.7791	24.4060
NBA 2007/2008	97.7314	24.5655	24.1910	24.0008	24.3889
NBA 2008/2009	98.0998	24.3351	24.5392	24.1599	24.2955
NBA 2009/2010	98.8422	25.1159	24.9916	23.9771	24.1235
On all matches	95.8857	23.9978	23.7895	23.5459	23.8211

League	Standard deviation of				
	AwayTotal	AwayQ1	AwayQ2	AwayQ3	AwayQ4
NBA 2002/2003	11.8276	5.3188	5.2566	5.4918	5.6048
NBA 2003/2004	12.0103	5.5764	5.4334	5.3977	5.7512
NBA 2004/2005	12.0692	5.3395	5.3951	5.4367	5.8997
NBA 2005/2006	12.1073	5.3986	5.3605	5.4324	5.6387
NBA 2006/2007	12.2106	5.3999	5.4340	5.7010	5.7291
NBA 2007/2008	12.7287	5.6910	5.5264	5.7059	5.7169
NBA 2008/2009	12.4031	5.4583	5.3821	5.3226	5.6896
NBA 2009/2010	11.2447	5.3824	5.3246	5.6625	5.5641
On all matches	12.3257	5.4853	5.4408	5.5457	5.7286