# **Comparing and analysing countries performance with the game League of Legends**

**Introduction**  
This project will be looking into a dataset of statistics collected from the game League of Legends (Riot, 2009). The data includes the top 100 player’s wins, losses and percentages from three different servers. North America, Europe West and Korea. Data which was collected on the 12/12/2016 from lolking.net will be used for this project which will be comparing the performance between servers in League of Legends.

League of Legends  
League of Legends is a very popular MOBA (Massive online battle arena) video game, developed by Riot and made for the PC. The game has one of the biggest player bases in the world, with a very completive scene. The game has a number of different game types, the game type being used for this project is Solo/Duo Que ranked 5s, which is a 5v5 completive game type. There are many aspects of the game that make it fairly complex, however the main objective is to gather gold via killing opponents to buy items to then destroy the opponents base.

Lolking  
Lolking is a website that allows users to view and search for a number of statistics collected from League of Legends. The site has a number of different filters that allows different data to be searched. The most useful features found on the site that were useful to this project were server, game type and top 100 ranking filters (Lolking, 2016).

Hypothesis  
Korea has 477 professional players in Esports which has the most of any country by a significant margin. Does this make them the best players over all? (REF) The goal of this project is to determine if Koreans are in fact the best League of legends players, this will analyse the difference in performance between the top three servers, looking into a number of different factors including: the server that plays the most games, the server with the best win loss ratio, the server that has the most individually skilled players, etc. the findings of the project will in some form determine which server has the best highest ranked players based on the data collected from Lolking. The prediction for this project is that at the highest level of play Koreans do have the best players.

## **Analysis**

In this project the data that has been collected is all numerical showing the amount of wins and losses challenger players have in the game League of Legends. Due to the fact that the data collected is primarily numerical the main method(s) of analysis and evaluation will be done using quantitative data analysis methods. There are a number of different statistical tests that can be run and completed on quantitative data that could be useful in determining an accurate result for this project, some which are quite simple while others may be more complex.

There will be some form of central tendency calculated in the results. Central tendencies provide a single value that can describe a set of data by identifying the central value of the data. The most common forms being Mean, mode and median. For this project only one form of central tendency will be used, most likely mean, as it would be the most accurate as it takes the exact middle value of the dataset even if that value is not present and it doesn’t take into account the popularity of a number. A paper looking at E-sports focusing on DOTA (which is very similar to League of Legends) used mean to compare a team’s win/loss, kills and other related material (Schubert, Drachen and Mahlmann, 2016).  
A measurement of dispersion will be used to find the amount of variability and spread found within a dataset. Standard deviation of the data will be calculated; this will measure how spread out the numbers in the data set are. This is done by taking the square root of the variance, the variance being the average of the squared differences from the mean, essentially this means the difference between each value and the mean value. Each difference value is then squared and the average is the result, the square root of this value is the standard deviation (EduPristine, 2011).

Standard deviation is useful to calculate because it can show the difference and the amount of variation in the data collected. A study completed on evaluating risk in real estate found that standard deviation could show the amount of measurement error from the mean value of internal rate of return (IRR). With the standard deviation calculated it was clear to see the difference between the different IRR values. They stated that with the use of standard deviation they were able to more easily make comparisons within the real estate sector (Wheaton et al.).

An article investigating League of Legends found some interesting statistics on win percentages in some games. The article was looking into the stats of “rager” and “toxic” gamers in league of legends and their impact on games. They state that ragers and toxic players can take your team’s chances of winning from 54% down to 46% (Yin-Poole, 2013). This study took quantitative data gathered from the games, similar to this project, and found out that there was a variable that could change the outcomes of games.

Z-score is the statistical measurement of a scores relationship to mean in a group of scores, telling you how many standard deviations are above or below the mean of the data value is. The z score lets you calculate the probability of a score occurring within the normal distribution and lets you compare multiple scores that are different from the normal distribution (REF). this is a useful value to have in testing as it allows you to find interesting data, for example with in this project you could quickly see a person score is higher than the mean score, and by how much, with the z score you could calculate the percentage of people that score higher than that score by using the value of the z-score using a z-score table taken away from .5000 (the size of half of the measurement grid when displayed visually e.g. as a bell curve in a diagram), from the value given back the percentage can be calculated for people that score higher.

Another form of descriptive quantitative test which was considered was a form of shape test, skewness. This test can show how the data collected is distributed, normally through a histogram. However, because the nature of the data that has been collected is ordered, there is no doubt that the data will return a positively skewed distribution as the values are the top 100 scores in descending order.

## **Methodology**

As previously stated the data was collected from Lolking.net, the website has a number of filters, including most wins, most losses and most ties of individual players and a filter for the server, this data was collected and put into an excel spreadsheet for further use in the project.

With the data collected and processed in an excel spreadsheet. The first task to be completed would be working out the mean value for each column of data which includes; Wins, losses, win percentage for each server. Excels built in functions were used to quickly work out the mean of the data, however another program was used to ensure that the data calculated was accurate, this was IBM SPSS Statistics 22.

The reason that the mean averages were calculated for this project was because of the amount of games played between the different servers varied quite dramatically. With this method the data is put into a format that is clearer to see any distinction. This is incredibly useful as the project goal is to see if Korea is the best performing country, this test makes the data slightly more qualitative as it is independent of the amount of games played and just looks at the ratio of wins.

With the data set that has been collected for the project being quite large, being able to show any form of variance or difference in the values would be useful to see. This project revolves around primarily quantitative data, getting different values could help show or find different conclusions for the project. Standard deviation will help show how varying the data is from the average value. This is extremely useful as it can show which data is more exact and reliable.

## **Results**

Below are the first set of results showing the win loss ratio for the servers on their own, before comparing them against each other.

The charts show a fairly basic result to the question. Currently it would seem that North America is the best performing server with 62%-win rate ad a 38% loss rate, which is a fairly big difference compared to the other servers. This however so far only comparing the mean average value win to loss ratio for each server, this is useful but it needs to be more accurate.

Standard deviation could improve the accuracy of the results collected as it shows how spread out the values collected are from the average. The chart below shows the standard deviation which has been calculated from the variance.

Unfortunately, there is no way to get a 100% accurate answer to this test because of the data collected from each server so different. It is clear from the graph below that EUW has by far played the most games which could affect the data as there is more room for them to have bad games etc. to allow for a more accurate data set the top 10 highest players of each server will be compared which can be seen below. Maybe write was a technique used by someone else.

Below shows the Z-Score in two tables for wins and losses. This is for the top 10 wins and losses for each server.

## **Conclusion**

Can’t do all time as game patches periodically changing the game and there for changing the data so using data from a set period of time to another will mean a more valid data set.

Players in the top 100 are classed as challenger players and make up 0.02% of the player base.

There are other factors to indicate skill such as individual kill/ death ratio, time of the games, champions played etc., however including these into this project will make it too big, perhaps this is something that could be done in future work.

Huge ethical issue comparing races so should be careful work.

The hypothesis was whether Koreans as players perform better compared to other servers, this project to keep things small looked at the top 3 servers which included Korea. There are many factors that can change this result such as including more servers in the study, the fact that the hypothesis was based on e-sports players could mean the top players don’t in fact play this game mode and this is just more of a general which server has the best players not players at an e-sports level. There are also other factors with in the game such as roles, objectives and individual champions that player play as it is a team game wins and losses don’t define the sole skill of a player it just gives a good approximation.

## **References**

(EduPristine, 2011). What is Standard Deviation [Online] Available from: http://www.edupristine.com/blog/what-is-standard-deviation [Accessed December 13, 2016]

Lee, J., An, J. and Lee, S. (2014). Factors Affecting eSports Audience Satisfaction - The case of League of Legends. *Journal of Korea Game Society*, 14(3), pp.35-46.

Lolking.net. (2016). *League of Legends Summoner Stats & Champion Build Guides - LolKing*. [online] Available at: http://lolking.net [Accessed 12 Dec. 2016].

Maniser, A. (2015). *The surprising esports culture of Korea | LoL Esports*. [online] 2015.na.lolesports.com. Available at: http://2015.na.lolesports.com/articles/surprising-esports-culture-korea [Accessed 6 Jan. 2017].

Riot (2009) League of Legends, Video Game, Microsoft Windows, macOS, Riot Games

Schubert, M., Drachen, A. and Mahlmann, T. (2016). Esports Analytics Through Encounter Detection.

Wheaton, W., Torto, R., Sivitanides, P. and Southard, J. (1999). Evaluating Risk in Real Estate.

Zhao, T. (n.d.). The Relationship between Z-Score and Stock Prices. *SSRN Electronic Journal*.