Statistical Analysis Assignment Proposal

Title:

Can a prediction be made for which team will win the game after 10 minutes of play time, and which variable(s) heavily influence the outcome?

Problem outline:

The video game League of Legends (LoL) is played on the PC and has a competitor base of around 115 million active players. The game is team based with a red vs blue theme of 5 player on each team. Depending on your ability you are put into ranks to face other players of a similar skill level. The dataset was found on Kaggle and contains 19 variables for each team, 38 total variables. Each variable, other than “BlueWins”, is an objective in the game which can determine the outcome of the game. The objective is to analyse the data to see which predictor(s) contribute highly to winning the game. A secondary objective could be determining if different predictors affect the outcome of games from a lower skill level, this would require another data set being constructed.

Data:

The data can be found [here](https://www.kaggle.com/bobbyscience/league-of-legends-diamond-ranked-games-10-min) for the “high-rank” games and contains data regarding objective features that can be attributed to the outcome of a game. This will be available in its original form on GitHub for ease of access. The original data contains 40 columns and 9880 rows. It will have to be converted into tidy data and some variables will be removed. The list below shows the variables selected for analysis. They are all repeated for the red team except for “Blue Wins”.

Variables and description:

**Blue wins -**> 1 if blue wins, 0 if blue loses.

**Blue wards placed ->** numerical, number of wards the blue team placed.

**Blue wards destroyed ->** numerical, number of wards the blue team destroyed.

**Blue first blood->** 1 if blue team gets the first kill, 0 if they do not.

**Blue kills->** numerical, how many kills the blue team get.

**Blue deaths->** numerical, how many times the blue team members die.

**Blue assists-**> numerical, how many assists are made by team members.

**Blue elite monsters->** numerical, number of dragons and heralds killed by the blue team.

**Blue towers destroyed->** numerical, number of towers destroyed by the blue team.

**Blue total gold->** numerical, total amount of team gold earned.

**Blue average level->** numerical, average level of whole team.

**Blue total experience->** numerical, team total experience.

**Blue total minions killed->** numerical, total number of minions killed.

**Blue cs per min->** numerical, the number of minions, monsters, and wards that the blue team have destroyed per minute.

**Blue gold per min->** numerical, how much gold the blue team has earned per minute.

Analysis:

I will analyse the data using a supervised learning approach with generalized linear models, linear regressions, and if appropriate, multiple regressions. Models will be selected depending on their validity to the data and then resampling methods, such as bootstrapping and cross validation will be used to determine accuracy of the prediction.

Comments:

Due to LoL games lasting well into 40 minutes a great deal can change throughout the game, other late game factors are not considered here. The total gold variable could be turned into a categorical variable of “Low”, “Medium” “High” and “Very High” which would allow me to run classification analysis on the data.