Introduction of Analysis:

The Artificial intelligence bot was made to scout various maps in StarCraft II. The development of this bot is noted in other documentations and can be referred to for more information. The bot is broken into 3 stages for its analysis

1. No machine learning stage, This stage will be called bot A. This bot is the first 10 times the bot plays the game. In these 10 runs the bot does gain some knowledge, but this will assure that t a minimum amount of machine learning has been performed.
2. The 2nd stage is the majority of the runs, this bot will be called bot B. Bot B is when the bot starts to develop its learning. This is about 30 trials long. The bot repeatedly plays StarCraft II, and uses the knowledge it has to grow. This bot will serve as an intermediate, initially it is expected that this bot shows growth.
3. Bot C is the highest trained bot. It has the maximum amount of machine learning implemented and will find the base the quickest.

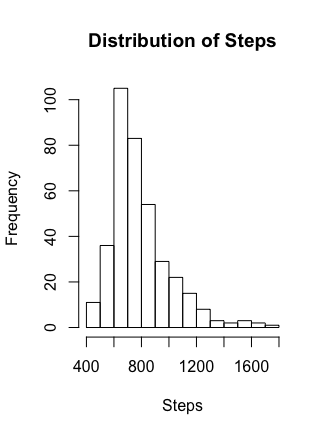
Developing the bot, visually it is hard to decipher how much the bot has improved and is the improvement significant. The analysis done In this report conducts a one-way ANOVA approach to show the bot has improved with runs. The bot will be broken into the three phases explained above.

Ho: The mean steps to find are the same for each bot.

Ha: The mean steps to find are not the same for each bot.

Ho vs Ha, is done in the format of a scientific paper. The number of steps used to find was taken rather than clock time, this was done to data mine easier. Time vs steps should have a linear relationship.

Summary of data:

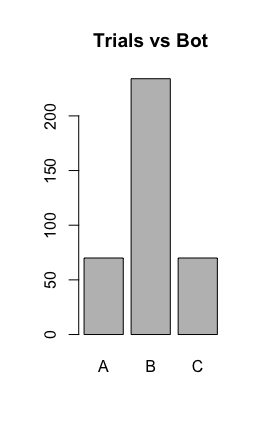


The Histogram to the left, is a projection of the Frequency the bot took to find an die.

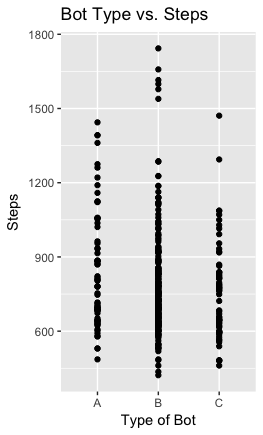
We can see it is not very likely to see a bot take more than approximately 1500 steps. This serves to show how the overall data looks. The data shows the median is 747 which agrees with the Histogram to the left.

Things to note is the majority of the data is consisted in bot B ~ 62.5 %. So in the histogram to left, each bot is not represented equally, but the overall distribution of the data is to be noted.

See (Bot Vs Number of Trials for more information)

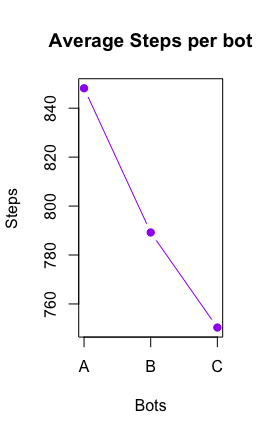


In the figure to the left, this histogram provides a visual to show that the majority of the data is in bot B. As noted, 62% of trials were conducting testing bot B and 19% approximately in each of A and B. This graph helps to show that B has more than 3 times the data of A.



Bot Type vs Steps

We can see that in bot B the highest steps were taken. But also note that this is where 62% of the data is located, so this will be later tested for outliers. This visualization is just to show the distributions of bots related to each other.



Average Steps Per Bot

A = 848.1286

B= 789.2094

C= 750.37

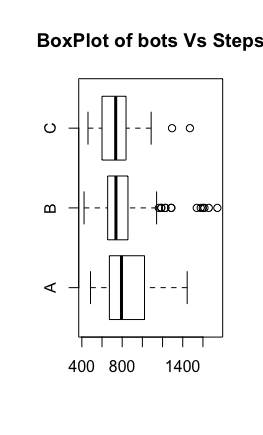
Note: no statistical test has been conducted to show that these numbers have any statistical value. This is stating the averages and to not be interpreted

A B C

Means 848.1286 789.2094 750.3714

Std. Dev 236.1014 212.6714 193.6582

Sample Size 70.0000 234.0000 70.0000

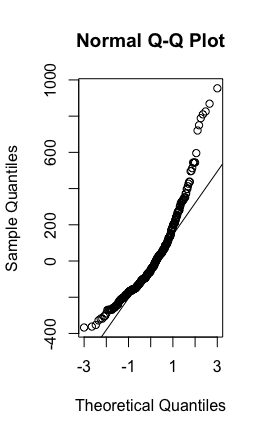


The bot plot to the left shows all the outliers present. It is to be noted that ANOVA assumptions assumes that outliers are not present. So this will be tested to see if the assumptions are still met when conducting the ANOVA test.

Testing for ANOVA Requirements

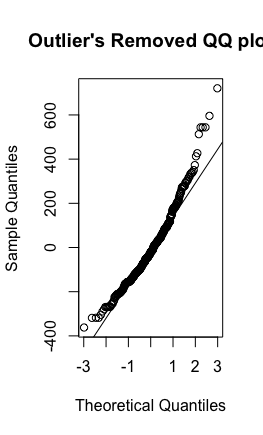
Shapiro Wilks Test:

The SW test is conducted to see if to see if the data provided is normal. If the data is not normal the distribution of the data therefore is not known and methods must be taken to normalize the data. The first approach to normalize data is to remove the outliers that are expected to be seen less than 99% of the time. Meaning if a data point is expected to be seen equal to or less than 1% of the time, then these points can be seen as too rare and therefore it can be removed as an outlier.



As the graph to right shows, this data is not normal and fails the SW test.

This data would be considered to be normal 2.177815e-13 percent of the time. Because this data’s chances of being normal are so low the next approach is to remove the outliers and reevaluate the SW test.



After removing the outliers the data is still not normal. The p value of this is approximately 3.218019e-08. This is a significant improvement but the data is still not normal enough for ANOVA to be used. The last approach to this problem is to force the data to be normalized through using BoxCox methods. BoxCox is used for original data.

Box-Cox transformations:

Results of Box-Cox Transformation

---------------------------------

Objective Name: Shapiro-Wilk

Linear Model: the.model

Sample Size: 374

lambda Shapiro-Wilk

-2.0 0.9578935

-1.5 0.9823425

-1.0 0.9942598

-0.5 0.9917696

0.0 0.9741708

0.5 0.9413683

1.0 0.8946803

1.5 0.8367529

2.0 0.7706757

Results of Box-Cox Transformation

---------------------------------

Objective Name: Log-Likelihood

Data: botdata$Steps

Sample Size: 374

lambda Log-Likelihood

-2.0 -2501.439

-1.5 -2486.727

-1.0 -2479.863

-0.5 -2481.298

0.0 -2491.475

0.5 -2510.800

1.0 -2539.600

1.5 -2578.070

2.0 -2626.21

Results of Box-Cox Transformation

---------------------------------

Objective Name: PPCC

Linear Model: the.model

Sample Size: 374

lambda PPCC

-2.0 0.9784594

-1.5 0.9911270

-1.0 0.9972554

-0.5 0.9960398

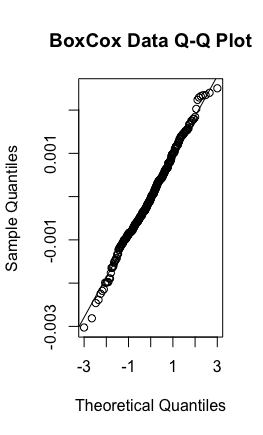
0.0 0.9871012

0.5 0.9701795

1.0 0.9455531

1.5 0.9140423

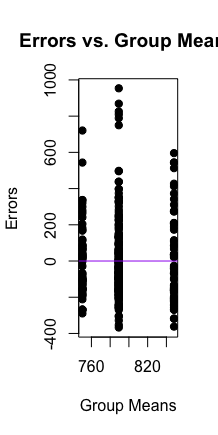
2.0 0.8767403



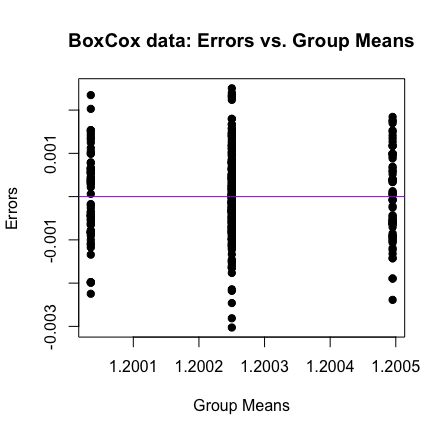
The Q-Q plot to the right, looks as it was normally distributed and has a p-value of .2832. Meaning the data that is observed under any reasonable alpha will pass to being normal.

Brown-Forsythe Test

After establishing that the data obtain can be normalized, the data must be evaluated to see if all the standard errors for each group categorical group is the same. (ie all three of the bots have the same standard error).



The graph to the left is a graph of the original data. Although This data does not meet the requirements of the Brown-Forsythe Test, It is not needed to. The same dataset which was BoxCox above is needed to fit the Brown-Forsythe Test not the orginal data.



This is the BF test for the Box Cox data. Notice that the x and y axis in this graph compared to the above, have changed. This due to the BoxCox.

ANOVA Test for CI intervals

Next, the first thing to check is if the CI will accept or reject the following Hypothesis

H0: all bots find in the same number of steps

Ha: at least one bot does not take the same number of steps to find

95% Bonferroni confidence intervals

Diff Lower Upper Decision Adj. p-value

muA-muB 0.00025 -7e-05 0.00056 FTR H0 0.182286

muA-muC 0.00046 7e-05 0.00085 Reject H0 0.01402

muB-muC 0.00021 -1e-04 0.00053 FTR H0 0.299232

the interpretation of each is such:

1. Given bot A and Bot B, when we compare the mean steps taken to find, the bots do not show any difference in steps. The P value is above .10, a reasonable interpretation of this p value is such: the probability of that this dataset exists and there is a difference between in steps to find for botA and Bot B is 18%. Since 18% is greater than our alpha value of 10%, we conclude there is no significant difference between these two bots. We would expect that this data is natural and will occur like this.
2. Given BotC and Bot B, the same claim as made above can be said but with a p-value of .2992.
3. **Main analysis:** Bot A vs Bot C. This is the two bots that will test if the machine learning implemented has improved the bot. The p-value of this CI is 0.0142. Meaning that if our original data were to be true, it is about 1% likely that this dataset is true. With alpha = .10, we are confident that this data set shows the Machine learning has improved the bots efficeny to find the base, and this dataset could not have been generated by random luck.

Note: these values of upper and lower are based off of BoxCox so they are difficult to interpret (this is the downside of BoxCox) but for generality, take this as a proof that is not quantified.

GROUP EFFECTS:

botdata$Type: A

[1] 52.22544

-----------------------------------------

botdata$Type: B

[1] -6.693732

-----------------------------------------

botdata$Type: C

[1] -45.53171

The overall mean was 792.9679, so a way of interpreting every number above is such:

1. The untrained bot A, increases the mean steps taken to find by approximately 52. This is the effect that is caused by untrained bot A.
2. The effect of the training session is -6.69. When comparing an average bot vs a bot while in training, we would expect the bot in training to take on average approximately 7 less steps than the average bot.
3. The fully trained bot C, will take on average 46 less steps than a mean bot. 46 steps is a significant amount of less steps.

Conclusion:

The bot has shown significant improvement. We have rejected our initial hypothesis which claimed that all three bots are equal in steps to find. The ANOVA test proves these results. In order to conduct the ANOVA test, a BoxCox Transformation of the dataset was used to ensure the SW and BF test passed on the dataset. The original data seemed normal and standard errors seemed equal when during the “summary of data” section.

In terms of difficulty to produce these results, a big challenge that occurred was to find a method which our dataset could meet the requirements for. Various methods originally planned to be used, were discarded to requirements not being met. ANOVA normality test took three different types of tests in itself to pass normality.

At the end, all tests were able to be met by the one-way ANOVA and produce the results that the bot had become better at scouting as time went on.