

STAT 6021: Final Project Report

Star Craft II Total Hours Prediction

Elena Gillis (emg3sc)

Kanika Dawar (kd4hr)

Karan Gadiya (khg8mh)

Varshini Sriram (vs4vx)

December 6, 2018

University of Virginia Data Science Institute

Contents

1	Acknowledgement	3
2	Summary	4
3	Introduction	4
4	Approach	5
4.1	Data Preparation and Initial Exploration	5
4.2	Data Analysis and Statistical Inference	7
5	Conclusion	10
6	References	10
7	Team Contribution	10
8	Appendix	11
8.1	Data Set Description	11
8.2	Data Snapshot	12
8.3	Supplementary Statistics Snapshot	13

1 Acknowledgement

We would like to express our deepest appreciation to all those who provided us with the possibility to complete this report. A special gratitude to the STAT 6021 instructor, Prof. Dan Spitzner, whose contribution in stimulating suggestions and encouragement, helped to coordinate the project, especially in writing this report.

Furthermore, we would also like to acknowledge with much appreciation the crucial role of the TAs, who have always provided guidance and help during the entire tenure of the course. A special thanks goes to all the team mates, who helped each other assemble the parts and contributed in every way to make this project possible. Last but not least, many thanks go to the DSI institute who have invested their full effort in guiding the team in achieving the goal.

2 Summary

The objective of this project is to analyze player on screen behaviour for a game called StarCraft 2. After a gaming company releases a beta version of a newly launched game, they wish to increase the users' playing time by analyzing their behaviour while playing the game. Our primary hypothesis is that a players game statistics can have an influence on the number of hours spent on playing the game. We claim that the total number of hours that each player spends playing the game can be predicted by their on-screen behaviour. We expect that the variables we have will explain a good amount of variation in our response variable. If this proves to be true, the gaming company can identify the most important or challenging portions of their product, improve on product usability and eventually use these findings to improve user-experience, for promotion, and marketing.

3 Introduction

StarCraft 2 (SC2) is a Real-Time Strategy (RTS) video game that is played at extraordinary speeds. All matches start with each player choosing one of three species and controlling one base and 6 worker units. Each of the three species control different buildings and units with unique abilities and upgrades, creating a complex interaction within the games environment. Basically, the three species units play out a dynamic game of rock-paper-scissors. Players must command their units, prepare defences, and build base expansions to gather more resources while building bigger armies to outmaneuver their enemies in what amounts to a high speed game of Go.

To perform this analysis, we use SkillCraft data, which was collected from on-screen movement patterns for players of StarCraft II real-time strategy game. The data was downloaded from UCI Machine Learning repository and was collected in September of 2013 by the De-

partment of Psychology at Simon Fraser University. The objective of the game is to develop a base, battleships, and defeat the opponents by destroying their bases. The bases are created by gathering artifacts and resources. The data published is numbers of different types of the on-screen movements made by individual players in one real-time second.

The final model makes use of 9 variables - LeagueIndex, Age, HoursPerWeek, ActionLatency, AssignToHotkeys, UniqueHotkeys, ActionsInPAC, UniqueUnitsMade, and SelectByHotkeys. LeagueIndex specifies which skill level the player plays at - Bronze, Silver, Gold, Platinum, Diamond, Master, GrandMaster, and Professional leagues (which are ordinally coded as 1-8). Age and HoursPerWeek give the age of the player and the number of hours the player played per week on average. HoursPerWeek is different from TotalHours in a way that some players might play for only a few weeks but have higher average hours per week as compared to other players who have been playing from a long time but play for comparatively fewer hours every week. SelectByHotkeys is the number of unit or building selections made using hotkeys. AssignHotKeys is a continuous value associating the number of units or buildings assigned to hotkeys per timestamp. UniqueHotkeys denotes the number of unique hotkeys used per timestamp while UniqueUnitsMade specifies the number of unique units made per timestamp. ActionLatency denotes the mean latency from the onset of PACs to their first action in milliseconds. And the last variable ActionsInPAC is the number of PACs per timestamp.

4 Approach

4.1 Data Preparation and Initial Exploration

Our original data has 3395 observations and 20 variables (submitted electronically). When performing the exploratory data analysis we noticed that there were missing values in the

Total Hours variable, Hours Per Week, and Age. Since Total Hours is the variable that we are trying to predict, we have removed the rows with missing observations. In addition, we observed that some of the variables in the Total Hours column were filled with a question mark - these rows were also removed from the dataset. This reduced the length of our data to 3338 rows. We converted the variables to their intended types and prepared the data for model building.

To visualize the dataset on a plot and reduce the variables to two dimensions we used the Principle Component Analysis. We plotted the first principle component against the response variable, however no significant patterns were seen in the data due to the wide range of the response. In order to reduce the range, we log-transformed the response variable and were able to observe that there is potentially a linear relationship in the data. This also indicated that we will most likely need to log-transform the response variable when fitting our model.

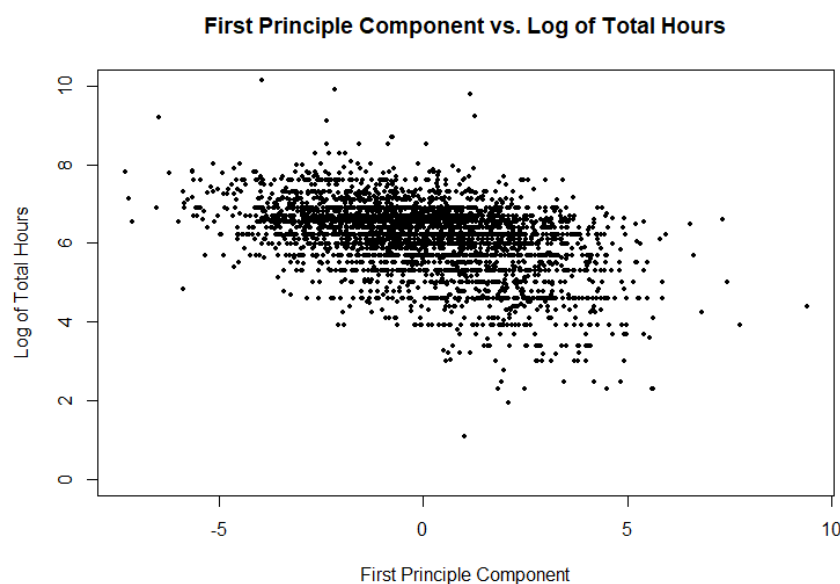


Figure 1: First Principal Component with Log of Total Hours

After plotting a series of histograms for the predictor and the response variables, we noticed that the response variable and some predictor variables were heavily skewed. This could potentially indicate presence of outliers in the data. As a result, we checked for outliers using the hat matrix values and found 213 observations as influential points.

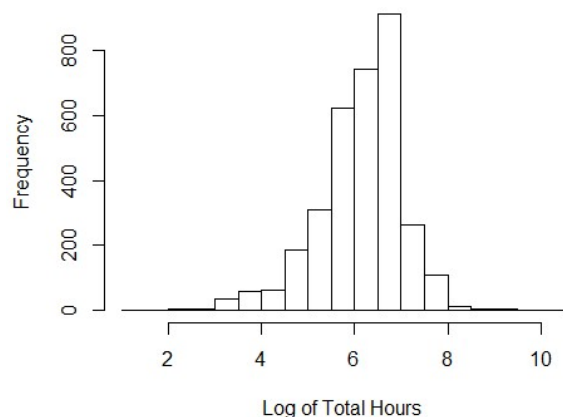


Figure 2: Distribution of Log Transformed Response

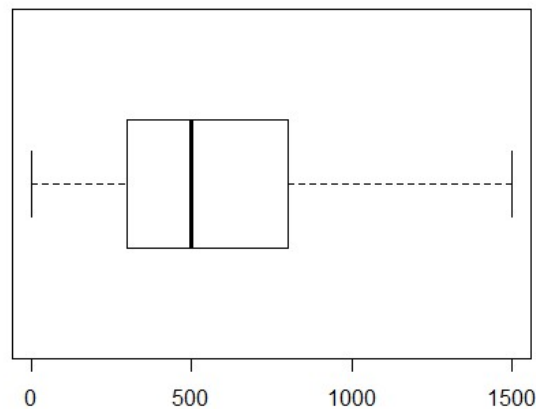


Figure 3: Box Plot of Response without Outliers

4.2 Data Analysis and Statistical Inference

Initially we built a base model with all the predictors and used that model to examine the data for leverage and influence points so that they won't affect the final model. We used several different statistical approaches to identify the outliers and measure their influence - the hat matrix, Cook's distance statistic, DFFITS, and the determinant of the covariance matrix. We identified observations that were classified as outliers by at least three methods, and removed rows with these outliers. In the end, the three methods consistently agreed on one outlier and observation number 1793 was removed from the data.

Next, we scaled the data for performing multicollinearity analysis on the 18 normalized regressor variables using three methods - pairwise correlation, variance inflation factors, and

variance decomposition proportions. Variance decomposition proportions did not detect any multicollinearity. From pairwise correlations, we observed that two variables, SelectBy-Hotkeys and APM had a correlation of 0.8. We then calculated the variance inflation factor for each of the regressors and removed variable APM, since it had the highest VIF value (30).

We then performed residual analysis for the base model by examining residual plots and quantile plots. These plots revealed non constant error variance and non-normality behavior (Figures 4 and 5).

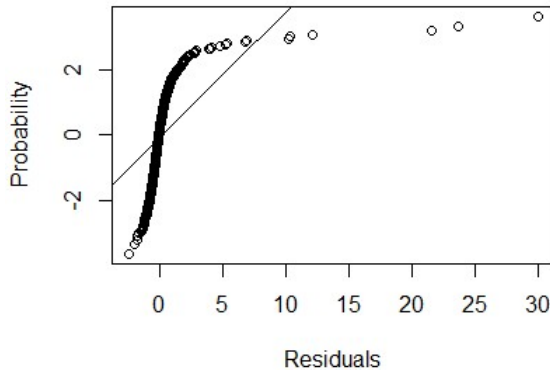


Figure 4: Normal Probability Plot Before Transformation

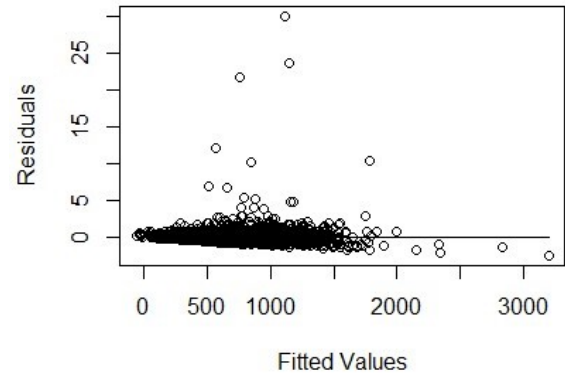


Figure 5: Residual Plot Before Transformation

To correct for the model inadequacies, we performed Variable Transformation. We used the Box-Cox procedure to determine the necessary transformation of the response variable and raised Total Hours to the power of 0.19. We also performed the Box-Tidwell procedure and transformed all the regressors individually. We then tested both transformations separately by performing stepwise variable selection on the data with transformed response. The model with transformed response gave us a higher R-squared value and better residual plots. As a result, in the final model, only transformation on response was used. The residual plots drastically improved and the variance in errors became almost perfectly constant.

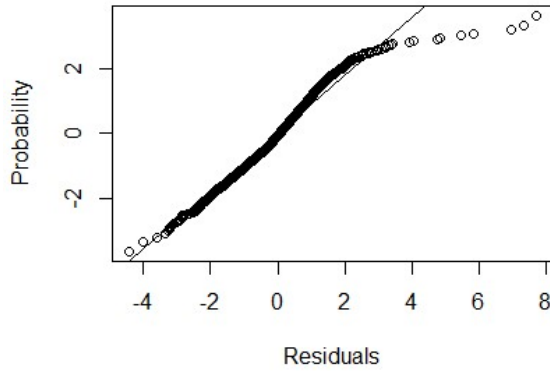


Figure 6: Normal Probability Plot
After Transformation

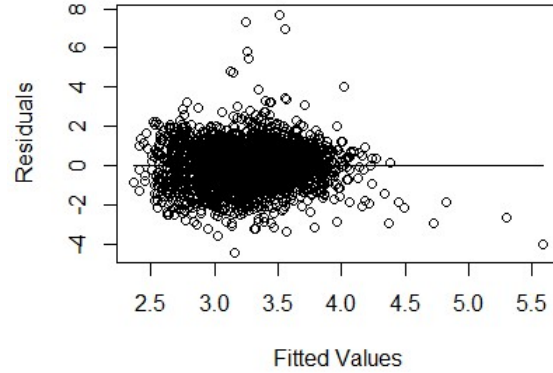


Figure 7: Residual Plot
After Transformation

We still had 17 variables and we needed to know which variables were of significance for the model after transformation and outlier removal. So we performed stepwise selection over all the regressors with an α_{in} value of 0.15 and α_{out} value of 0.10. After performing stepwise regression the number of significant variables reduced to 9.

With everything in place to build the model, we used these 9 variables and regressed them against Total Hours to obtain a linear model. The linear model gave an R squared of 34.68%. It is the percentage of the response variable variation that is explained by a linear model. With a low R squared value, we tested the significance of regression and partial tests of significance for the regressors. For both, the null hypothesis that the regression is not significant and the regressors are insignificant was rejected. Prediction was done on new data (15 test values) and examined for hidden extrapolation which resulted in observations number 3, 4, 8, 13, and 14 to lead to hidden extrapolation even though they fall in the range of the regressor data. Also, we noticed that the prediction intervals are wider than the confidence intervals. It is in line with the concept that Confidence interval tells us about the likely location of the true population parameter whereas prediction intervals must account for both the uncertainty in knowing the value of the population mean, plus data scatter.

5 Conclusion

We realise that 9 variables LeagueIndex, Age, HoursPerWeek, ActionLatency, AssignToHotkeys, UniqueHotkeys, ActionsInPAC, UniqueUnitsMade and SelectByHotkeys are important in terms of predicting power for the total number of hours players spend playing the game. There exists a linear relationship between these 9 regressors and the response variable TotalHours. Although the R squared value is less, our analysis is statistically thorough. The regressors have high p values and residual plots also look normal. This leads us to the conclusion - we can predict the number of total hours players will play the game based on the final variables in the model.

6 References

- [1] Blair, M. (2018). UCI Machine Learning Repository: SkillCraft1
- [2] <http://skillcraft.ca/>
- [3] https://github.com/avinash2692/videogame_analysis

7 Team Contribution

Elena Gillis: Study Planning, Initial Exploration, and Report Compilation

Kanika Dawar: Data Analysis and Statistical Inference

Karan Gadiya: Study Planning, Data Preparation, and Report Compilation

Varshini Sriram: Data Analysis and Statistical Inference

8 Appendix

8.1 Data Set Description

- GameID: Unique ID number for each game (integer)
- LeagueIndex: Bronze, Silver, Gold, Platinum, Diamond, Master, GrandMaster, and Professional leagues coded 1-8 (Ordinal)
- Age: Age of each player (integer)
- HoursPerWeek: Reported hours spent playing per week (integer)
- TotalHours: Reported total hours spent playing (integer)
- APM: Action per minute (continuous)
- SelectByHotkeys: Number of unit or building selections made using hotkeys per timestamp (continuous)
- AssignToHotkeys: Number of units or buildings assigned to hotkeys per timestamp (continuous)
- UniqueHotkeys: Number of unique hotkeys used per timestamp (continuous)
- MinimapAttacks: Number of attack actions on minimap per timestamp (continuous)
- MinimapRightClicks: number of right-clicks on minimap per timestamp (continuous)
- NumberOfPACs: Number of PACs per timestamp (continuous) Perception Action Cycles (PACs) are fixations with at least one action
- GapBetweenPACs: Mean duration in milliseconds between PACs (continuous)
- ActionLatency: Mean latency from the onset of a PACs to their first action in milliseconds (continuous)

- ActionsInPAC: Mean number of actions within each PAC (continuous)
- TotalMapExplored: The number of 24x24 game coordinate grids viewed by the player per timestamp (continuous)
- WorkersMade: Number of SCVs, drones, and probes trained per timestamp (continuous)
- UniqueUnitsMade: Unique unites made per timestamp (continuous)
- ComplexUnitsMade: Number of ghosts, infestors, and high templars trained per timestamp (continuous)
- ComplexAbilitiesUsed: Abilities requiring specific targeting instructions used per timestamp (continuous)

8.2 Data Snapshot

GameID	LeagueIndex	Age	HoursPerWeek	TotalHours	APM	SelectByHotkeys	AssignToHotkeys	UniqueHotkeys	MinimapAttacks
52	5	27	10	3000	143.7180	3.515159e-03	2.196974e-04	7	1.098487e-04
55	5	23	10	5000	129.2322	3.303812e-03	2.594617e-04	4	2.940566e-04
56	4	30	10	200	69.9612	1.101091e-03	3.355705e-04	4	2.936242e-04
57	3	19	20	400	107.6016	1.033542e-03	2.131015e-04	1	5.327537e-05
58	3	32	10	500	122.8908	1.136014e-03	3.273259e-04	2	0.000000e+00
60	2	27	6	70	44.4570	9.783903e-04	2.552323e-04	2	0.000000e+00
61	1	21	8	240	46.9962	8.201141e-04	1.685166e-04	6	0.000000e+00
72	7	17	42	10000	212.6022	9.039739e-03	6.762401e-04	6	1.163531e-03
77	4	20	14	2708	117.4884	2.944275e-03	5.267713e-04	2	1.881326e-05
81	4	18	24	800	155.9856	5.053908e-03	5.241090e-04	8	2.495757e-05
83	3	16	16	6000	153.8010	1.676615e-03	3.185568e-04	4	0.000000e+00
93	4	26	4	190	79.2948	3.785385e-04	2.551020e-04	3	1.645820e-05
97	3	18	12	350	67.4754	4.225216e-04	1.690086e-04	1	2.414409e-05
98	3	38	6	1000	119.4366	4.952043e-03	5.212677e-05	2	8.687795e-05
100	5	16	30	5000	160.4754	4.253860e-03	4.317351e-04	2	7.745835e-04
102	5	17	16	1500	81.7722	2.333484e-03	4.304486e-04	4	0.000000e+00
105	4	28	8	2000	50.8374	6.641086e-04	2.213695e-04	1	1.844746e-05
106	5	20	10	120	160.6464	3.430344e-03	6.336305e-04	7	0.000000e+00

MinimapRightClicks	NumberOPACs	GapBetweenPACs	ActionLatency	ActionsInPAC	TotalMapExplored	WorkersMade	UniqueUnitsMade	ComplexUnitsMade	ComplexAbilitiesUsed
3.923169e-04	0.004849036	32.6677	40.8673	4.7508	28	0.00139660	6	0.000000e+00	0.0000e+00
4.324362e-04	0.004307064	32.9194	42.3454	4.8434	22	0.00119350	5	0.000000e+00	2.0757e-04
4.614094e-04	0.002925755	44.6475	75.3548	4.0430	22	0.00074455	6	0.000000e+00	1.8876e-04
5.434088e-04	0.003782551	29.2203	53.7352	4.9155	19	0.00042620	7	0.000000e+00	3.8358e-04
1.328558e-03	0.002368299	22.6885	62.0813	9.3740	15	0.00117450	4	0.000000e+00	1.9254e-05
0.000000e+00	0.002424707	76.4405	98.7719	3.0965	16	0.00037221	6	0.000000e+00	0.0000e+00
4.493776e-05	0.001988496	94.0227	90.5311	4.1017	15	0.00057296	5	0.000000e+00	0.0000e+00
1.253033e-03	0.004952464	24.6117	41.7671	6.6104	45	0.00227730	9	1.292812e-04	2.4862e-04
4.138917e-04	0.005399406	52.0140	46.4321	3.3746	29	0.00103470	7	2.727923e-04	4.7033e-04
3.993212e-04	0.003568933	24.4632	52.1538	6.5664	27	0.00131030	6	0.000000e+00	0.0000e+00
8.215411e-04	0.003772383	23.4107	48.0711	7.0044	24	0.00159280	7	0.000000e+00	1.6766e-05
1.645820e-04	0.003554970	39.6381	65.5000	4.2269	19	0.00075708	7	1.069783e-04	2.6333e-04
1.448646e-04	0.002885219	42.4370	68.0502	4.3222	16	0.00074947	7	0.000000e+00	4.3455e-04
3.475118e-05	0.002727968	54.8718	79.2102	6.2293	21	0.00149430	5	0.000000e+00	0.0000e+00
4.063389e-04	0.004571312	36.2897	46.8889	5.4361	28	0.00198090	7	0.000000e+00	0.0000e+00
2.492071e-04	0.002899864	45.1654	64.7500	4.5312	15	0.00124600	6	0.000000e+00	2.2655e-04
1.475797e-04	0.002656434	45.7902	76.8889	3.5000	13	0.00066411	6	0.000000e+00	2.9516e-04
9.468042e-05	0.005535163	28.3636	37.7947	4.7671	29	0.00276760	10	8.739731e-05	1.0196e-04

8.3 Supplementary Statistics Snapshot

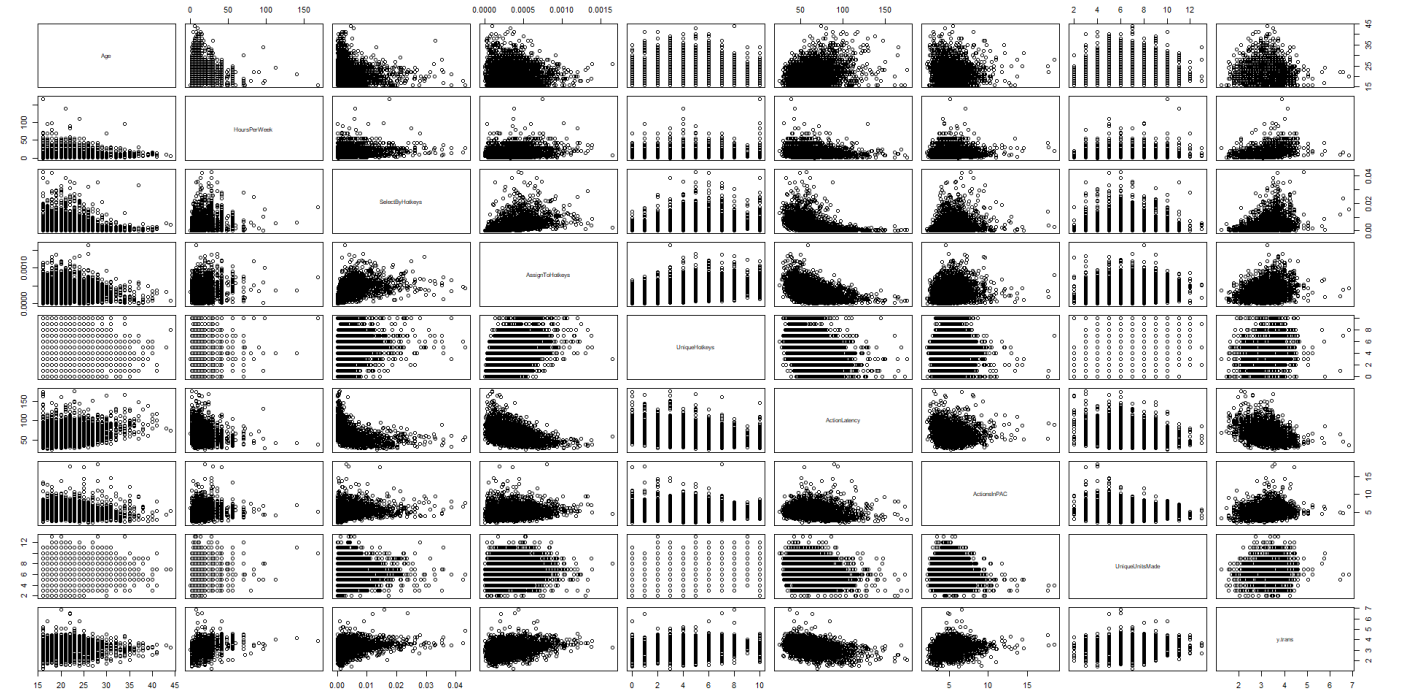


Figure 8: Scatter Plots for the model variables

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
LeagueIndex	-0.3569396	0.16404416	-0.07007395	0.007850680	-0.002480079	-0.06528076	0.16198131	0.127826298
Age	0.1032762	-0.09354108	0.05266932	0.567672022	0.167027418	-0.33441430	0.18201288	0.601114397
HoursPerWeek	-0.1304273	0.11118226	-0.01753090	-0.318272855	0.357372048	0.60314101	-0.06291878	0.519309061
SelectByHotkeys	-0.2460178	0.22136855	-0.10170885	-0.076056955	0.392260274	-0.12201342	-0.25160404	-0.034372483
AssignToHotkeys	-0.3053040	0.10241985	-0.09788260	0.037499749	0.333361947	-0.20126640	-0.09084811	-0.146563322
UniqueHotkeys	-0.2320089	-0.05532867	-0.16295852	0.259370924	0.343563906	-0.19437680	-0.34878877	-0.188408353
MinimapAttacks	-0.1536673	0.07284817	0.14353437	0.426157306	0.207084360	0.37648197	0.40921126	-0.096428213
MinimapRightClicks	-0.1652556	0.08780715	0.39226864	0.302100728	-0.175833293	0.26772293	-0.24576470	-0.288726253
NumberOfPACs	-0.3719621	-0.06886719	-0.32978466	-0.046504802	-0.233899658	-0.00349448	0.06821481	0.053107394
GapBetweenPACs	0.3063609	-0.26529487	-0.03897445	-0.032245930	0.224849021	0.09336744	-0.37314825	0.005951466
ActionLatency	0.3921803	-0.10073118	0.13765101	0.085055747	0.258965851	0.03930074	-0.18127370	-0.006588964
ActionsInPAC	-0.0705906	0.35233537	0.57461076	0.009606908	0.050403243	-0.02343751	-0.08055555	-0.016838200
TotalMapExplored	-0.2361761	-0.40590428	-0.07582779	0.180671653	-0.126439792	0.24723086	-0.16960376	0.014208291
WorkersMade	-0.0919324	0.08416471	0.25951232	-0.050619486	-0.358060331	-0.19449149	-0.44329346	0.437621069
UniqueUnitsMade	-0.1950935	-0.46093164	0.04024308	0.172385537	-0.073184740	0.18818690	-0.13883565	0.044795265
ComplexUnitsMade	-0.1837804	-0.39299302	0.35118193	-0.263543139	0.131608289	-0.17780837	0.12251339	0.007115065
ComplexAbilitiesUsed	-0.1678082	-0.35812217	0.33705085	-0.290076814	0.214019305	-0.19165252	0.26982818	-0.069747520

Figure 9: Loadings - Principal Components

	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17
LeagueIndex	0.068304390	-0.027754722	0.082985627	0.103643472	0.081892198	-0.845367172	0.16583770	-0.146847006	0.019590985
Age	-0.253489675	-0.161234765	0.119502174	0.005118598	-0.050684809	0.056147175	-0.05335798	0.089706437	-0.013558060
HoursPerWeek	-0.224120201	0.174093466	0.116865463	-0.048249170	0.005580833	0.072022740	-0.03511899	-0.018459135	-0.023902420
SelectByHotkeys	0.089696728	-0.635137221	-0.159156864	0.347886439	0.156584747	0.200623252	-0.00496400	-0.130897326	-0.037047124
AssignToHotkeys	0.074114511	-0.034469895	0.106782923	-0.816746969	-0.127981297	0.029595109	-0.01875628	0.019472594	-0.028186815
UniqueHotkeys	-0.130907229	0.637744759	-0.008050968	0.332917502	0.009471986	0.025138335	-0.01532341	0.015330989	-0.019146455
MinimapAttacks	0.604298434	0.082296650	-0.000708408	0.087564991	0.023288456	0.149870914	-0.01224503	0.068625761	-0.003453788
MinimapRightClicks	-0.319999248	-0.187455108	0.560386442	0.014524711	0.125715941	-0.005950612	-0.04495714	-0.047867130	-0.039366481
NumberOfPACs	-0.008254711	-0.063119635	0.196011302	0.078614308	-0.027260262	0.170469660	-0.04916746	0.325369523	0.705015806
GapBetweenPACs	0.290756848	-0.169484623	0.230959759	0.054735205	-0.146449506	-0.302759550	0.06438693	0.587174046	-0.054958770
ActionLatency	0.124370532	0.016724325	0.031807460	-0.094558675	0.098771623	-0.111906616	0.05920749	-0.496963703	0.640231211
ActionsInPAC	-0.161898602	0.034811452	-0.473831953	-0.002679078	-0.256956932	-0.103633387	0.01372068	0.352035325	0.279130169
TotalMapExplored	-0.021943048	-0.167226666	-0.231741460	0.032232163	-0.668295991	-0.067912279	-0.09578639	-0.298835070	-0.045768093
WorkersMade	0.484963532	0.153886330	0.103070918	-0.026864441	0.029811150	0.161238634	0.11107462	-0.123465366	-0.067929147
UniqueUnitsMade	-0.102331042	-0.059459427	-0.427677700	-0.207035913	0.574448058	-0.03320413	0.25974489	0.132934533	-0.008545889
ComplexUnitsMade	0.082169886	0.017719859	0.044078564	0.046843759	0.164539956	-0.098760238	-0.71192782	-0.005329314	0.010052267
ComplexAbilitiesUsed	-0.058436456	0.003263542	0.218799376	0.123073613	-0.175537819	0.165228142	0.59963070	-0.038885528	0.002291009

Figure 10: Loadings - Principal Components

	cookD	flag1	dffits	flag2	covratio	flag3	hatvalue	flag4	flag
1793	1.304431e+00	1	1.075610e+02	1	1.948378e-51	1	0.007427807	0	3
8	1.732030e-04	0	5.735889e-02	0	1.023328e+00	1	0.018184055	1	2
27	1.134402e-04	0	-4.641970e-02	0	1.024319e+00	1	0.018768987	1	2
70	2.891113e-05	0	2.343398e-02	0	1.022481e+00	1	0.016554334	1	2
92	4.204958e-05	0	-2.826151e-02	0	1.021745e+00	1	0.015940565	1	2
94	4.708223e-04	0	-9.457195e-02	0	1.029942e+00	1	0.025412841	1	2
140	1.268018e-07	0	-1.551937e-03	0	1.018047e+00	1	0.012088326	1	2
146	3.009478e-04	0	-7.560945e-02	0	1.023672e+00	1	0.019160299	1	2
162	4.441933e-06	0	9.185393e-03	0	1.019619e+00	1	0.013644060	1	2
198	3.031867e-04	0	7.588734e-02	0	1.134722e+00	1	0.113894423	1	2

Figure 11: Outlier Analysis: 1793 is the outlier

	LeagueIndex	Age	HoursPerWeek	TotalHours	APM	SelectByHotkeys	AssignToHotkeys	UniqueHotkeys	MinimapAttacks	MinimapRightClicks
LeagueIndex	1.0000000	-0.12738133	0.21783352	0.27072864	0.6247014	0.42899836	0.4872432	0.32238327	0.27060609	0.206358
Age	-0.1273813	1.0000000	-0.18428887	-0.03284037	-0.2102248	-0.13038159	-0.1045002	0.01519751	0.04301901	-0.0198571
HoursPerWeek	0.2178335	-0.18428887	1.0000000	0.25513248	0.2466560	0.20549376	0.1580082	0.07020829	0.08417767	0.0494641
TotalHours	0.2707286	-0.03284037	0.25513248	1.0000000	0.2610077	0.21967615	0.1330303	0.08599753	0.13446197	0.0898781
APM	0.6247014	-0.21022484	0.24665603	0.26100770	1.0000000	0.81387965	0.5332839	0.33629205	0.21930903	0.3067531
SelectByHotkeys	0.4289984	-0.13038159	0.20549376	0.21967615	0.8138797	1.0000000	0.4492163	0.27436723	0.13347434	0.1076584
AssignToHotkeys	0.4872432	-0.10450020	0.15800819	0.13303028	0.5332839	0.44921632	1.0000000	0.40263788	0.20577944	0.1549831
UniqueHotkeys	0.3223833	0.01519751	0.07020829	0.08599753	0.3362920	0.27436723	0.4026379	1.0000000	0.15106329	0.1245551
MinimapAttacks	0.2706061	0.04301901	0.08417767	0.13446197	0.2193090	0.13347434	0.2057794	0.15106329	1.0000000	0.2247061
MinimapRightClicks	0.2063587	-0.01985770	0.04946438	0.08987872	0.3067531	0.10765841	0.1549834	0.12455578	0.22470699	1.0000000
NumberOfPACs	0.5891797	-0.19678298	0.17460749	0.18404398	0.6348565	0.35896223	0.4538889	0.35312654	0.13798816	0.1435061
GapBetweenPACs	-0.5374749	0.11193578	-0.13370810	-0.15925859	-0.5676984	-0.27357262	-0.3777413	-0.22370128	-0.21342492	-0.2445224
ActionLatency	-0.6599235	0.23996169	-0.18851008	-0.20619682	-0.7222579	-0.38930849	-0.4610269	-0.30455648	-0.17167095	-0.2166331
ActionsInPAC	0.1402429	-0.04579732	0.09520037	0.08008583	0.4032594	0.16691021	0.0913265	-0.02226228	0.13375652	0.3236941
TotalMapExplored	0.2302213	-0.02395643	0.06472099	0.09082089	0.2388357	0.09651481	0.2014423	0.26789696	0.16422698	0.1734851
WorkersMade	0.3103800	-0.09215769	0.05056916	0.10036330	0.3778258	0.16111687	0.1967931	0.11127569	0.08234538	0.2124441
UniqueUnitsMade	0.1519967	0.02282053	0.03858626	0.05368044	0.1234432	0.02805770	0.1496082	0.23330518	0.12500955	0.1544461
ComplexUnitsMade	0.1713045	-0.08042115	0.05940082	0.04345375	0.1626332	0.06629985	0.1703001	0.12240598	0.05217398	0.0980271
ComplexAbilitiesUsed	0.1561458	-0.06575227	0.07482979	0.05923429	0.1418774	0.06454323	0.1694966	0.11005685	0.04221418	0.0956931

Figure 12: Pairwise Correlations: SelectByHotkeys & APM = 0.81

NumberOfPACs	GapBetweenPACs	ActionLatency	ActionsInPAC	TotalMapExplored	WorkersMade	UniqueUnitsMade	ComplexUnitsMade	ComplexAbilitiesUsed
0.5891797	-0.53747488	-0.6599235	0.14024286	0.23022132	0.31037999	0.15199673	0.17130447	0.15614579
-0.1967830	0.11193578	0.2399617	-0.04579732	-0.02395643	-0.09215769	0.02282053	-0.08042115	-0.06575227
0.1746075	-0.13370810	-0.1885101	0.09520037	0.06472099	0.05056916	0.03858626	0.05940082	0.07482979
0.1840440	-0.15925859	-0.2061968	0.08008583	0.09082089	0.10036330	0.05368044	0.04345375	0.05923429
0.6348565	-0.56769843	-0.7222579	0.40325941	0.23883569	0.37782580	0.12344322	0.16263317	0.14187742
0.3589622	-0.27357262	-0.3893085	0.16691021	0.09651481	0.16111687	0.02805770	0.06629985	0.06454323
0.4538889	-0.37774127	-0.4610269	0.09132650	0.20144235	0.19679312	0.14960819	0.17030006	0.16949658
0.3531265	-0.22370128	-0.3045565	-0.02226228	0.26789696	0.11127569	0.23330518	0.12240598	0.11005685
0.1379882	-0.21342492	-0.1716709	0.13375652	0.16422698	0.08234538	0.12500955	0.05217398	0.04221418
0.1435069	-0.24452243	-0.2166334	0.32369407	0.17348555	0.21244401	0.15446650	0.09802723	0.09569381
1.0000000	-0.49128586	-0.8170293	-0.24289576	0.47074563	0.28204790	0.32248921	0.19665356	0.17806840
-0.4912859	1.0000000	0.6804341	-0.31018824	-0.09450767	-0.23706185	-0.08978067	-0.08305343	-0.09213445
-0.8170293	0.68043412	1.0000000	-0.10636923	-0.34909926	-0.31355756	-0.22339540	-0.19836474	-0.18998773
-0.2428958	-0.31018824	-0.1063692	1.0000000	-0.15681943	0.25368525	-0.13484400	0.05437859	0.05347211
0.4707456	-0.09450767	-0.3490993	-0.15681943	1.0000000	0.13137573	0.57539312	0.30924108	0.25154628
0.2820479	-0.23706185	-0.3135576	0.25368525	0.13137573	1.0000000	0.11244588	0.20338109	0.10505559
0.3224892	-0.08978067	-0.2233954	-0.13484400	0.57539312	0.11244588	1.0000000	0.38219187	0.29437902
0.1966536	-0.08305343	-0.1983647	0.05437859	0.30924108	0.20338109	0.38219187	1.0000000	0.62051867
0.1780684	-0.09213445	-0.1899877	0.05347211	0.25154628	0.10505559	0.29437902	0.62051867	1.0000000

Figure 13: Pairwise Correlations: NumberOfPACs & ActionLatency = -0.81

Index	Variance Decomposition Proportions									
	LeagueIndex	Age	HoursPerWeek	APM	SelectByHotkeys	AssignToHotkeys	uniqueHotkeys	MinimapAttacks	MinimapRightClicks	NumberOfPACs
1	1.000	0.010	0.002	0.003	0.001	0.001	0.009	0.006	0.003	0.002
2	1.605	0.002	0.002	0.003	0.001	0.002	0.001	0.003	0.000	0.001
3	1.878	0.002	0.002	0.000	0.000	0.001	0.004	0.014	0.010	0.005
4	2.055	0.000	0.213	0.067	0.000	0.001	0.001	0.039	0.131	0.000
5	2.258	0.001	0.043	0.057	0.000	0.015	0.058	0.089	0.016	0.004
6	2.332	0.000	0.094	0.399	0.000	0.001	0.011	0.016	0.157	0.000
7	2.500	0.027	0.027	0.000	0.001	0.009	0.001	0.026	0.150	0.000
8	2.641	0.010	0.410	0.322	0.000	0.000	0.015	0.030	0.011	0.000
9	2.687	0.003	0.094	0.047	0.000	0.000	0.010	0.002	0.389	0.000
10	2.731	0.001	0.029	0.070	0.003	0.025	0.039	0.460	0.027	0.000
11	2.993	0.008	0.034	0.007	0.000	0.000	0.035	0.051	0.003	0.004
12	3.233	0.001	0.000	0.000	0.001	0.003	0.758	0.257	0.028	0.002
13	3.559	0.017	0.006	0.000	0.000	0.007	0.034	0.000	0.000	0.000
14	3.686	0.864	0.008	0.017	0.001	0.001	0.012	0.001	0.054	0.006
15	3.877	0.024	0.007	0.003	0.000	0.000	0.001	0.000	0.000	0.000
16	4.596	0.020	0.025	0.001	0.003	0.025	0.004	0.001	0.014	0.028
17	6.948	0.007	0.003	0.003	0.015	0.027	0.002	0.003	0.001	0.029
18	18.379	0.002	0.003	0.001	0.975	0.883	0.002	0.001	0.003	0.718
Index	Variance Decomposition Proportions									
	GapBetweenPACs	ActionLatency	ActionsInPAC	TotalMapExplored	WorkersMade	uniqueUnitsMade	complexUnitsMade	complexAbilitiesUsed		
1	0.007	0.005	0.000	0.004	0.005	0.003	0.002	0.002		
2	0.008	0.000	0.007	0.047	0.001	0.064	0.041	0.037		
3	0.000	0.002	0.026	0.002	0.032	0.001	0.047	0.047		
4	0.001	0.001	0.000	0.012	0.001	0.012	0.032	0.042		
5	0.030	0.014	0.000	0.006	0.083	0.002	0.010	0.025		
6	0.002	0.000	0.000	0.021	0.043	0.013	0.013	0.013		
7	0.084	0.006	0.001	0.044	0.108	0.031	0.014	0.061		
8	0.000	0.000	0.000	0.000	0.199	0.001	0.000	0.004		
9	0.049	0.004	0.004	0.001	0.284	0.011	0.005	0.003		
10	0.005	0.000	0.000	0.027	0.057	0.009	0.000	0.000		
11	0.051	0.000	0.042	0.027	0.006	0.140	0.001	0.037		
12	0.003	0.003	0.002	0.001	0.000	0.090	0.003	0.031		
13	0.019	0.004	0.018	0.547	0.001	0.463	0.035	0.045		
14	0.091	0.004	0.006	0.024	0.045	0.000	0.006	0.027		
15	0.003	0.001	0.000	0.012	0.030	0.108	0.790	0.616		
16	0.618	0.124	0.047	0.219	0.056	0.050	0.000	0.003		
17	0.028	0.823	0.031	0.007	0.035	0.000	0.000	0.000		
18	0.000	0.008	0.816	0.000	0.013	0.003	0.000	0.007		

Figure 14: Variance Decomposition Proportions: No multicollinearity detected

LeagueIndex	2.169426
Age	1.127538
HoursPerWeek	1.107238
APM	36.607390
SelectByHotkeys	12.256610
AssignToHotkeys	1.640089
UniqueHotkeys	1.315145
MinimapAttacks	1.165452
MinimapRightClicks	1.293118
NumberOfPACs	13.802043
GapBetweenPACs	2.281521
ActionLatency	5.517163
ActionsInPAC	8.035799
TotalMapExplored	1.857287
WorkersMade	1.322522
UniqueUnitsMade	1.675188
ComplexUnitsMade	1.820861
ComplexAbilitiesUsed	1.674432

Figure 15: Variance Inflation Factors: APM 36.73, SelectByHotkeys 12.31, NumberOfPACs 13.81

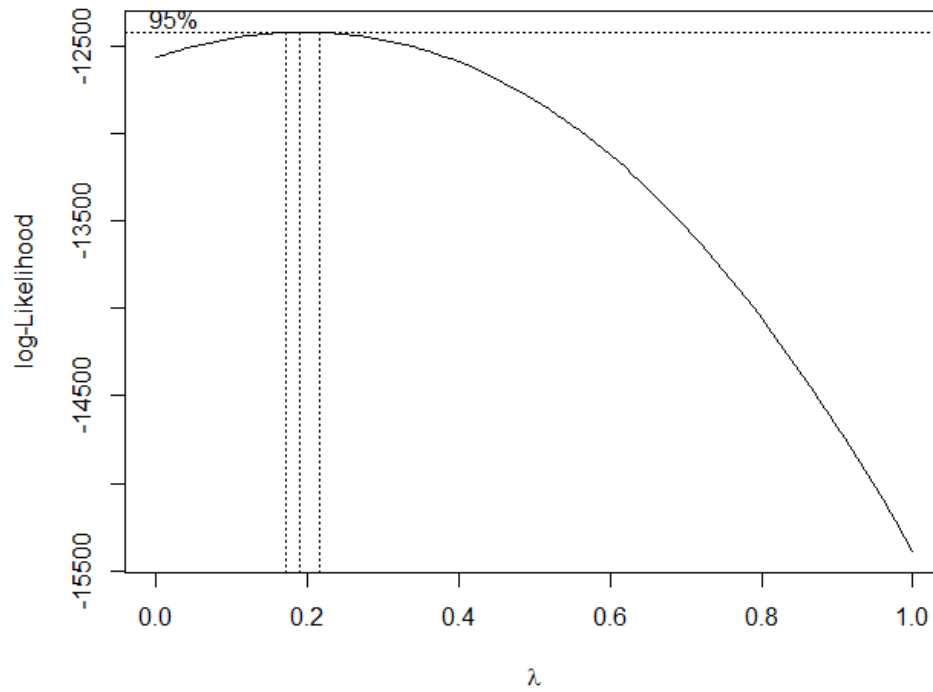


Figure 16: Box-cox Plot: alpha value = 0.19

```
Call:
lm(formula = TotalHours ~ LeagueIndex, data = data.trans)

Residuals:
    Min       1Q   Median       3Q      Max
-1057.9  -293.2   -93.2   147.0 23984.4

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  280.441     26.978   10.39  <2e-16 ***
LeagueIndex   16.827       1.013    16.61  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 823.1 on 3335 degrees of freedom
Multiple R-squared:  0.07645,    Adjusted R-squared:  0.07617
F-statistic:  276 on 1 and 3335 DF,  p-value: < 2.2e-16
```

Figure 17: Stepwise Selection after Box Tidwell Procedure

```

Call:
lm(formula = y.trans ~ LeagueIndex + Age + HoursPerWeek + SelectByHotkeys +
    AssignToHotkeys + UniqueHotkeys + ActionLatency + ActionsInPAC +
    UniqueUnitsMade, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.8880 -0.2552  0.0228  0.2648  3.3218

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.163e+00  8.041e-02  26.898  < 2e-16 ***
LeagueIndex  1.432e-01  7.411e-03  19.324  < 2e-16 ***
Age          1.168e-02  1.887e-03   6.193  6.63e-10 ***
HoursPerWeek 1.321e-02  6.609e-04  19.984  < 2e-16 ***
SelectByHotkeys 7.748e+00  1.908e+00  4.060  5.01e-05 ***
AssignToHotkeys -1.174e+02  4.543e+01 -2.585  0.009777 **
UniqueHotkeys  -7.487e-03  3.679e-03 -2.035  0.041933 *
ActionLatency  -1.943e-03  5.634e-04 -3.450  0.000568 ***
ActionsInPAC    2.301e-02  5.203e-03  4.423  1.01e-05 ***
UniqueUnitsMade 1.248e-02  4.312e-03  2.894  0.003832 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4362 on 3327 degrees of freedom
Multiple R-squared:  0.3448,    Adjusted R-squared:  0.343
F-statistic: 194.5 on 9 and 3327 DF, p-value: < 2.2e-16

```

Figure 18: Stepwise Selection after Box Cox Procedure

```

Analysis of Variance Table

Response: y.trans
          Df Sum Sq Mean Sq F value    Pr(>F)
LeagueIndex  6 238.27  39.712 209.0462 < 2.2e-16 ***
Age           1   0.78   0.781  4.1090 0.0427349 *
HoursPerWeek  1 83.91  83.906 441.6814 < 2.2e-16 ***
SelectByHotkeys  1   3.68   3.679 19.3672 1.112e-05 ***
AssignToHotkeys  1   0.95   0.951  5.0079 0.0252974 *
UniqueHotkeys   1   0.40   0.403  2.1199 0.1454899
ActionLatency   1   2.65   2.650 13.9487 0.0001910 ***
ActionsInPAC    1   2.86   2.859 15.0494 0.0001068 ***
UniqueUnitsMade  1   1.50   1.504  7.9155 0.0049300 **
Residuals     3322 631.08   0.190
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figure 19: ANOVA Table for the final model

```

F-critical 1.694746
F-statistic: 126
p-value: < 2.2e-16
Null Hypothesis is rejected since F > F.crit and
Regression is significant
All individual regressors are also significant (F > F.crit)

```

Figure 20: Test for Significance of Regression

	Coeff. lwr	Coefficients	Coeff. upr
(Intercept)	2.647506e+00	2.720128e+00	2.792751e+00
LeagueIndex.L	6.998200e-01	7.558883e-01	8.119566e-01
LeagueIndex.Q	-9.836887e-02	-5.206672e-02	-5.764578e-03
LeagueIndex.C	-3.249431e-02	4.176597e-03	4.084750e-02
LeagueIndex^4	-2.209763e-03	2.459867e-02	5.140710e-02
LeagueIndex^5	-4.091501e-02	-2.055693e-02	-1.988538e-04
LeagueIndex^6	1.592876e-02	3.276105e-02	4.959333e-02
Age	9.693426e-03	1.158092e-02	1.346842e-02
HoursPerWeek	1.270547e-02	1.337343e-02	1.404139e-02
SelectByHotkeys	6.349790e+00	8.271072e+00	1.019235e+01
AssignToHotkeys	-1.589989e+02	-1.134262e+02	-6.785340e+01
UniqueHotkeys	-1.063948e-02	-6.949921e-03	-3.260360e-03
ActionLatency	-2.413970e-03	-1.846995e-03	-1.280020e-03
ActionsInPAC	1.695023e-02	2.216659e-02	2.738295e-02
UniqueUnitsMade	7.818619e-03	1.213007e-02	1.644152e-02

Figure 21: Confidence Intervals for the model coefficients

	h00	h.max	flag
1	0.02007030	0.05189153	0
2	0.01246045	0.05189153	0
3	0.06552397	0.05189153	1
4	0.06831837	0.05189153	1
5	0.02953396	0.05189153	0
6	0.02475904	0.05189153	0
7	0.01226982	0.05189153	0
8	0.06940504	0.05189153	1
9	0.04039699	0.05189153	0
10	0.01416342	0.05189153	0
11	0.01904981	0.05189153	0
12	0.01248324	0.05189153	0
13	0.06440499	0.05189153	1
14	0.06949328	0.05189153	1
15	0.01136362	0.05189153	0

Figure 22: Extrapolation for prediction data

	fit	lwr	upr
1	2.815520	2.690162	2.940877
2	3.212986	3.112607	3.313365
3	3.327899	3.105885	3.549913
4	2.814307	2.584861	3.043753
5	2.917137	2.760174	3.074099
6	3.282342	3.143730	3.420954
7	3.243037	3.143639	3.342435
8	3.570272	3.341751	3.798794
9	3.584938	3.411598	3.758278
10	3.342382	3.234907	3.449858
11	3.173830	3.052205	3.295455
12	3.221663	3.121076	3.322250
13	3.283883	3.065168	3.502598
14	3.033843	2.802446	3.265240
15	3.290668	3.194039	3.387297

Figure 23: 95% Confidence Interval for new data points

	fit	lwr	upr
1	2.815520	1.951805	3.679234
2	3.212986	2.352542	4.073430
3	3.327899	2.444962	4.210836
4	2.814307	1.929472	3.699142
5	2.917137	2.048273	3.786001
6	3.282342	2.416605	4.148079
7	3.243037	2.382707	4.103367
8	3.570272	2.685677	4.454868
9	3.584938	2.712966	4.456909
10	3.342382	2.481082	4.203683
11	3.173830	2.310649	4.037010
12	3.221663	2.361195	4.082132
13	3.283883	2.401769	4.165996
14	3.033843	2.148500	3.919186
15	3.290668	2.430653	4.150683

Figure 24: 95% Prediction Interval for new data points