Table of Contents

List of Tables

List of Figures

Executive Summary

The analysis of this report is based on a data set containing video game telemetric data from 697 game players. After data exploration (missing values, outliers, etc.), three tasks are analyzed and solved according to the following structure:

Before cluster analysis is conducted, factor analysis is applied to develop the association of the different variables and to explore the hidden factors. By testing 5 and 6 factors with 5 different rotation methods, the best solution is confirmed and interpreted to find the hidden factors related to the characteristics.

There are 8 groups of players identified based on the cluster analysis without outliers after comparing with the result with outliers. To deal with multicollinearity, the clusters are created through 5 factor scores from principal component analysis. The cluster solution is validated stable in 20%-50% subsets from the total dataset.

To explore factors affecting expert level, multiple regression analysis is used. The simple regression model is initially built to illustrate a rough relation between expert level and independent variables. Then, multicollinearity is eliminated by using stepwise regression model and regression model based on factor analysis. Finally, the latter model is preferred through the comparation and five factors are powerful to explain the expert level.

Data Preparation

Missing Values

Descriptive statistics was run over the 19 variables. It was observed that there is only one value missing. In game ID 1064, the value of *TotalHours* is missing.

Internal consistency of data

As descriptive statistics was run on the 19 variables, it was observed that there is difference in scale and magnitudes between variables. While, on the one hand, age varies from 16 to 43 and hours per week varies from 2 to 98, on the other hand total hours varies from 7 to 10000. Variables like *SelectByhotkeys, AssigntoHotkeys, Uniquehotkeys, MinimapAttacks, MinimaprightClicks, NumberofPACs, WorkersMade, ComplexUnitsMade, ComplexAbilitiesUsed* vary in the range from 0 to 1. Therefore, the variables need to be standardized. The descriptive statistics is shown in Table 11.

Outliers

Normal distribution curve over a histogram was prepared for individual variables to explore the spread of every variable are shown in Figure 4.

Box plot of Z-scores of all variables was plotted in Figure 3. As the number of variables is large, it is difficult to predict the outliers using the box plot.

For multivariate outliers Mahalanobi's D^2 statistic was calculated. Thereafter, D^2/(degrees of freedom) was calculated. It is plotted in Figure 5. Degrees of freedom represent how many values involved in a calculation have the freedom to vary. It can be calculated as the number of values in a data set minus 1. Going by the rule of thumb, values > 4 were classified as outliers. These were the following *GameIDs* - 616, 900, 72, 3899, 7258, 83, 318, 5129, 629, 4754, 466, 7415, 1988

Collinearity

Variance Inflation Factor (VIF) > 10 is a good indicator of multicollinearity. Tolerance statistic < 0.1 also indicates a serious problem. Observing the results as observed in Table

12, it was concluded that multicollinearity is present amongst *APM*, *SelectByHotkeys* and *NumberOfPACs*.

Correlations matrix amongst all variables was prepared. As shown in Table 13, it was concluded that collinearity is present between the following pairs:

- *APM* and *SelectbyHotkeys*
- *APM* and *NumberOfPACs*
- *APM* and *ActionLatency*
- *ActionLatency* and *NumberOfPACs*
- *ActionLatency* and *GapBetweenPACs*

Question 3

Data preparation and assumptions

In order to investigate the association of the different variables, factor analysis is applied to analyse the covariance and correlation among variables in factors. Dataset is divided into the training set and test set with 60% (418 observations) and 40% (279 observations) separately using a random sampling method. Furthermore, an academic research is introduced to elaborate whether the dependent variable is included in factor analysis.

> *" When you use exploratory factor analysis the variables are not divided into dependent and independent categories. Instead, all variables are analyzed together to identify underlying patterns or factors. The technique can be used to factor analyze either independent or dependent variables considered separately "* (*Hair, 2015)*

As a result, *LeagueIndex* is kept in order to conduct the factor analysis to identify underlying patterns.

Before conducting factor analysis in the training set, KMO and Bartlett's Test is introduced to measure the correlation, which shows KMO measure is equal to 0.768 greater than 0.5 indicating that sample is adequate, with p-value smaller than 0.05 indicating the correlations among variables are significant to launch factor analysis.

Factor analysis on training dataset with 5 factors

Moreover, the correlation matrix in Table 14 shows that more than 22% of paired variables are correlated representing a substantial number of correlations which is a good sign for factor analysis. Noticeably, the correlations for the pair of "*Age*" and other variables are less than 0.3 indicating the possibility of classifying itself into a factor.

The next stage is to identify the partial correlations through the anti-image matrix. Table 15 displays that only 1 diagonal element of the anti-image matrix as known as Measure of Sampling Accuracy (MSA) is smaller than 0.5 with 3 other partial correlations greater than 0.7, which provides enough information showing the partial correlation is small enough to conduct the factor analysis.

For factor analysis, two possible factor extraction methods, Principal Component Method and Maximum Likelihood Method could be applied.

As shown in Table 16, 5 components are set accounting for 62.698% of the variance with eigenvalues greater than 1, which is enough to represent the dataset. Moreover, the Scree plot graphs (Figure 6) the relation between the eigenvalue and the component number showing that the first component explains around 33% of the variance, 3 times as the second component.

Comparing to the 5 components in the Maximum Likelihood Method as in Table 17, the Principal Component Method is much better with the same number of components, which is 12.803% higher explainability.

Moreover, in the table of communalities (Table 18 & Table 19) as shown, Principal Component Method expounds 6 more variables and higher explainability in each variable than the Maximum Likelihood Method. As a result, the Principal Component Method is adopted to conduct factor analysis.

In order to improve the data classification, the rotation matrix of each method shown in Table 20, Table 21, Table 22, Table 23, Table 24 and Table 25 with 5 different rotation methods applied.

Under the criteria of 0.5, it is clear that all of five rotation methods enhance the classification of components with similar output comparing to the original version. Nevertheless, Equamax and Varimax perform slightly better than others as all variables are divided into exactly one component if the criteria are reduced by 0.01. Eventually, the factor analysis under the training set using the Principal Component Method will rotate the data by the Equamax method and the varimax method to compare the output in test set.

Factor analysis on training dataset with 6 factors

So as to obtain the stabler output of factor analysis, the training set with 6 components is introduced in Table 26. The percentage of variance explanation increases from 62.698% to 68.866%. Moreover, the communalities in Table 27 show that 6 components extract more value in each variable comparing with 5 components.

Comparing 5 factors and 6 factors

Next, 5 rotations methods are employed to select the best output from Table 28, Table 29, Table 30, Table 31, Table 32 and Table 33. Obviously, the rotation matrix with the Promax rotation method includes more variables than others under the 0.5 criteria. However, 5 components using the Equamax or varimax rotation method seem better than 6 components using the Promax rotation method. In order to estimate the stability and performance between them, the test set is brought into test.

Implementation on test data set

Before conducting factor analysis, KMO and Bartlett's Test is used in the test set. The KMO measure, which is 0.742 greater than 0.5, shows that the sample is adequate and p-value smaller than 0.05 indicating the correlations among variables are significant to launch factor analysis.The correlation Matrix as shown in Table 34 provides the information that 24.56% of paired variables are correlated which is sufficient to launch factor analysis.The Anti-image Matrix in Table 35 displays that only 6 partial correlations and 2 diagonal elements fail to fit the standard, which is <0.7 and >0.5 separately, showing that the partial correlations are small enough for the factor analysis.

First of all, Principal Component Method with 5 components using Equamax rotation is applied as shown in Table 36. The output under a small change of criteria (0.483) represents a pretty similar result as the training set except for the MinimapAttacks variable, which might be a potential issue that varies depending on different sample.

Secondly, Principal Component Method with 5 components using varimax rotation is introduced in Table 37. *WorkersMade* variable is not allocated into any factor showing that the stability is worse than Equamax rotation method.

Thirdly, Principal Component Method with 6 components using Promax rotation is conducted as shown in Table 38. It is obvious that the result of the test set is quite different from that of the training set with 1 duplicated output and 3 variables not included in any component.

Finally, factor validation shows that the Principal Component Method with 5 components in the Equamax rotation method is the most stable method comparing to 5 components with varimax rotation and 6 components with Promax rotation. Meanwhile, it also shows that the output of the model for the 5 components explains variables better than other 2 methods.

Effect of outliers

After the best method is identified, populations with and without outliers are introduced to test the influence of outliers.As tables in Table 39 and Table 40, the variable *APM* in the dataset without outliers stands into factor 1 and 3 at the same time, but the value is much larger in factor 1 and the value is barely higher than the criteria by 0.002. Hence, it might be dismissed. Variable *GapBetweenPACs*, however, shows a different group that the value (-0.499) is so close to the threshold in factor 1, but appears in factor 3, which shows that the model might change after the outliers are removed.

Interpretation

Table 1 Rotated Component Matrix (Rotation=Equamax, extract 5 factors, population)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.682	0.192	0.293	0.314	0.176
Age	-0.096	-0.154	-0.057	-0.422	0.585
HoursPerWeek	0.079	0.034	0.018	0.829	-0.053
TotalHours	0.129	0.072	0.128	0.716	0.263
APM	0.744	0.133	0.457	0.3	0.079
SelectByHotkeys	0.69	-0.017	0.138	0.234	0.062
AssignToHotkeys	0.681	0.188	0.16	0.195	0.13
UniqueHotkeys	0.556	0.209	-0.099	0.04	0.34
MinimapAttacks	0.03	0.008	0.249	0.294	0.636
MinimapRightClicks	0.046	0.15	0.619	0.105	0.273
NumberOfPACs	0.781	0.33	0.029	0.307	0.129
GapBetweenPACs	-0.564	-0.054	-0.482	-0.132	-0.011
ActionLatency	-0.723	-0.286	-0.306	-0.264	-0.027
ActionsInPAC	-0.098	-0.076	0.851	0.027	-0.073
TotalMapExplored	0.238	0.565	-0.128	0.188	0.458
WorkersMade	0.336	0.105	0.554	0.079	-0.004
UniqueUnitsMade	0.097	0.648	-0.106	0.119	0.41
ComplexUnitsMade	0.053	0.826	0.156	0.028	-0.079
ComplexAbilitiesUsed	0.039	0.79	0.129	0.045	-0.071

Eventually, in order to interpret these 5 factors, a senior Skill Craft 1 player was interviewed and explained the behaviours and connections behind all the variables. As a result, each factor in the figure of population or training set with 5 components could be interpreted as below:

Factor 1 can be interpreted as "*Proficiency in the Game*" as *LeagueIndex* can represent how good you are in the game, *APM* and *Hotkeys* related variables are interpreted as proficiency of playing, and the rest in factor 1 are related to the reaction time, which is shorter when a player is proficient in the game; Factor 2 represents "*Strategies Used*" because all the variables in factor 2 are related to the strategy a player takes in the game. For example, a player can choose to explore the map to obtain more resources or choose to attack the opposite directly, which connects to *UnitMade* variables and *ComplexAbilitiesUsed*; Factor 3 shows "*Attack Efficiency*" deriving from how many workers are made and how fast they can counterattack the competitor by *minimap*; Factor 4 contains "*Hours Invested*" because all variables are connected to hours a player spends; last but not least, although *MinimapAttacks* might be discussable, Factor 5 connects to "*Age*" because older players tend to drag the map to attack rather than using minimap.

Noticeably, the model might involve some potential issues that although it is already the most stable one, some variables of the output might vary according to different dataset due to a slight correlation. For instance, *MinimapAttacks* and *GapBetweenPACs* might switch between 4 and 5, 1 and 3 respectively. Nevertheless, these issues will not affect the interpretation since both of them involve the meaning of both factors.

Question 2

Data preparation analysis

To identify the relationship of different game players within the dataset, cluster analysis is demonstrated to group observations based on their characteristics. The total dataset has 696 valid observations with one missing case, which is regarded as a large dataset.

Firstly, Outliers in the dataset should be identified as the cluster analysis is sensitive to them; both univariate graphical and multivariate techniques are used to recognize outliers. These box plots for 19 variables show a few observations that are far away from others. The Mahalanobi's D squared statistic is also used to consider all the variables at one time. According to the rule of thumb, D2/df >4 in a large dataset indicates outliers, and there are 11 outliers after filtering with all observations.

Although the number of outliers is less than 5% of total observations, the dendrogram of Hierarchical cluster without outliers is more evenly distributed. Compared with the two different dendrogram figures, the first dendrogram with outliers shows that some of these clusters are not representative because they only have very few observations. However, the under-sampled league index should also be considered. **Outliers cannot be removed directly because some outliers represent an under-sampled segment.**

It can be seen from Table 2 Outliers with LeagueIndex and Table 3 Distribution of Observations, the number of observations that have league index 7 is under-sampled as it only occupies 5% of the total. Therefore, it is necessary to consider whether to remove or retain those outliers with league index 7.

After comparing the cluster analysis results between including outliers and removing outliers, there was a cluster with only two observations before removing outliers. Their *gameID* is 616 and 72 separately. Both observations are outliers in cluster 7 and one of the two has the league index 7. So, we decide to remove all outliers in order to balance cluster groups instead of keeping under-sampled outliers.

Then multicollinearity is confirmed by the Variance Inflation Factor (VIF) and correlations matrix, indicating whether a variable has a strong linear relationship with other variables. There are three variables: *APM*, *SelectByHotkeys* and *NumberOfPACs* which larger than 10, and the pair correlation between *APM* and *SelectByHotkeys*, *NumberofPACs* and *ActionLatency* are also higher than 0.8. To deal with the multicollinearity of variables, we decide to use five factor scores which are created based on Q1 principal component analysis to do cluster analysis instead of manually removing collinearity variables.

Table 2 Outliers with LeagueIndex

GameID	*League Index*	*Mal_df_ratio*
3899	1	8.69
900	2	15.35
83	3	5.53
318	5	5.38
616	5	18.24
629	5	4.85
466	6	4.31
72	7	12.84
4754	7	4.46
5129	7	4.53
7258	7	6.18

Table 3 Distribution of Observations

League Index	*Number of observations*
1	100
2	114
3	119
4	120
5	111
6	98
7	35

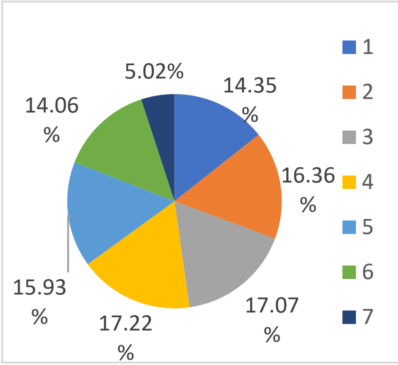


Figure 1 Proportion of League Index

Besides, the table of descriptive statistics shows that standard deviation variance between 19 variables varies greatly, which would affect the cluster analysis results. The variance with the largest variance is *TotalHours* (702.354) while the variance of *ComplexUnitMade* is only 0.0001. The standard deviation of total hours is around 700 times larger than some other variables such as *ActioninPAC*. Due to the sensitivity to differences in scales of distance measures, we decide to standardize the data using Z-score before implementing the hierarchical cluster analysis.

Table 4 Number of Cases in each Cluster without outliers

Cluster	1	180
	2	137
	3	80
	4	21
	5	120
	6	21
	7	83
	8	43
Valid		685
Missing		12

Table 5 Number of Cases in each Cluster with outliers

Cluster	1	170
	2	167
	3	93
	4	100
	5	56
	6	108
	7	2
Valid		696
Missing		

Hierarchical cluster analysis

Table 6 Agglomeration Schedule output from hierarchical cluster analysis using the standardised data, Complete Linkage (furthest neighbour) and Euclidean distance.

Hierarchical clustering analysis has several steps to perform methods, and there are different cluster methods and interval measures to use. Furthest neighbour cluster method performs better than between-groups linkage, and we also take Euclidean distance measure rather than others to produce clustering outputs.

Heterogeneity is the difference between current coefficients and previous coefficients. While selecting a solution, considering the differences among observations is the key factor. A large increase indicates that we should choose the prior cluster number. In table 5, when 12 clusters became 11 clusters, the coefficient value has a big jump (0.346). When the 8 clusters became 7 clusters, the value changes hugely (0.485). Besides, the change decreased from 4 clusters to 2 cluster (0.555,1.254, 0.851). Possible solutions are therefore 8 cluster solutions or 4 clusters.

There is a comparison of whether the outliers of the cluster are included or not. The dendrogram can help us to look up the structure of cluster solutions. From the (Figure 7 Dendrogram of Hierarchical Cluster Analysis Using 5 Factor Scores from Question1 with Outliers, Standardised Data, Complete Linkage, Euclidean Distance.) there are small groups including only one or two observations that is not acceptable. Compared the dendrogram between with and without outliers (Figure 8 Dendrogram of Hierarchical Cluster Analysis Using 5 Factor Scores from Question1 without Outliers, Standardised Data, Complete Linkage, Euclidean Distance.) it is evident that the structure of dendrogram without outliers is balanced. In addition, the sizes of clusters are also in proper difference while first dendrogram has extreme differences in cluster sizes. The 8 cluster solutions without outliers would have clusters of 180, 137, 80, 21, 120, 83, and 43 cases separately. This is better than the 7 clusters solution with outliers which has a tiny size cluster. Thus, we choose 8 clusters solutions without outliers.

Table 6 Agglomeration Schedule

Stage	*Cluster Combined*		*Coefficients*	*Difference*	*Cluster*
	Cluster 1	*Cluster 2*			
671	54	328	4.865	0.212	14
672	3	22	5.034	0.169	13
673	54	572	5.193	0.159	12
674	5	371	5.539	0.346	11
675	1	2	5.835	0.296	10
676	3	102	5.837	0.002	9
677	5	551	6.139	0.302	8
678	201	393	6.624	0.485	7
679	1	54	6.897	0.273	6
680	20	201	6.904	0.007	5
681	3	214	7.117	0.213	4
682	1	3	7.672	0.555	3
683	1	20	8.926	1.254	2
684	1	5	9.777	0.851	1

Non-hierarchical cluster analysis

After deciding on an appropriate number of clusters, non-hierarchical procedures have also been used to cluster total observations into 8 clusters. Table 46 summarizes the results of clusters that each game player is distributed into based on K-Mean clustering.

To sum up, the classification of total observations identifies characteristics from each case, and cluster created method classify similar observations into one group through 5 factors score from the principal component analysis. We use factor scores instead of the original variables because PCA has already identified and incorporated collinear variables. The five factors from 1 to 5 have been interpreted as “proficiency in the game”, “strategies used”, “attack efficiency”, “hours invested” and “age”. The ANOVA in Table 44 indicates that five factors are <0.001 significant to contribute the cluster solution.

Characteristics for each cluster

Table 41 Initial Cluster Centres and Table 42 Final Cluster Centres demonstrate each factor value for 8 clusters. The **cluster 1** players has the lowest value in using strategies (-0.43), and the age of this cluster of players is the youngest (-0.72) among all groups. Players spend less time in the game and have a low level of proficiency in the game. **Cluster 2** has the lowest hours invested value (-0.57), and the player’s attack efficiency is also very low. Their strategies are used less frequently and their proficiency in games is very low. The overall age of this group of players is medium. **Cluster 3** has the highest strategy usage value (1.85). Players in this cluster are younger and spend less time on the game. Their attack speed in the game and their proficiency in the game are moderate. In **cluster 4**, the hours invested value is the highest (2.52), and this cluster of gamers is very young. Young children devote much time to playing this game, their game proficiency is relatively low, and the use of strategy is very small. **Cluster 5** has the highest value (1.31) for proficiency in the game, but its other features are not remarkable. The game player in cluster 5 uses fewer strategies than other clusters. The attack efficiency and hours invested value in this cluster are low. Moreover, players in cluster 5 are medium-aged.

An obvious feature of **cluster 6** is that the players are the oldest (2.11). Players also put in more time than Cluster 5 and the factor 4 value is 1.09, but the value is moderate among all groups. Their attack efficiency or speeds ranks second among all clusters, but there is still a large gap between them and the first-ranked value. Strategies used value in cluster 6 are very low. Although not very high, game players in Cluster 6 have a certain degree of proficiency in the game. Players in **cluster 7** have the highest attack efficiency value (1.42) while proficiency in the game is the lowest (-0.7). They use very few strategies and invest very little time in the game. This group of gamers is of medium age. In **cluster 8**, the player’s attack efficiency is the lowest (-0.89) and a moderate number of strategies is used. Players are moderately proficient in the game, using some strategies but not too much. The player's hours spent in the game is in the middle, and they are in the lower middle age.

Validation

To validate the cluster solution, we select 20%-50% observations randomly from the total dataset as the testing set to run the analysis. There are no more than 20% observations assigned to new clusters, which denotes that the solution is reliable and could represent the general game players to some extent. The solution is more stable in a larger random subset, especially less than 10% cases classified to a new cluster in 50% subset.

Table 7 Validation Results

Random subset	Observation	Ratios of reclassified data
20%	137	20%
30%	219	19%
40%	276	11%
50%	362	8%

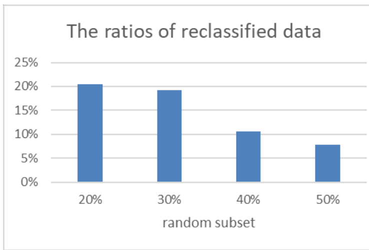


Figure 2 The Ratios of Reclassified Data

Limitation

It is challenging to choose the optimal number of clusters or appropriate k, the difference of coefficients might indicate potential cluster number of k, however, it is still a very subjective problem that should be overcome. The optimization clustering methods are not in a wide degree of acceptance as the number of applications found is far fewer (Everitt, 2011). In this report, the cluster analysis tends to be more stable as it was applied to the total dataset and validated in the subset chosen from the total dataset. The result is uncertain if new data is introduced more in the validation. Different clustering methods using mathematical properties lead to different results when used as the basis of cluster analysis, and it cannot produce interpretable results empirically. Cluster analysis assumes that group structure exists but what if the hypothesis is false. Cluster analysis result is based on summarizing the similarity of observations and should not be generalized.

<u>Question 1</u>

To answer the question given, finding factors which have significant impact on expert level is necessary. Meanwhile, explaining the practical meaning of these factors and interpreting how these factors affect players' level are also important. When the dependent variable has 5 or more categories, it could be regarded as continuous variable(Torra *et al.*, 2006). Therefore, the assumption in this case is that the dependent variable is treated as continuous variable and multiple regression is used.

Data preparation and assumptions

The sample data is sufficient with 697 observations, as well as a reasonable observation to variable ratio which close to 35 to 1. 5% is selected as the criterion for the significance level. Within these 20 variables, two variables would not be considered as independent variables: one is the variable named *GameID*, which is applied to identified players and does not have statistic meaning. Another is the discrete variable *LeagueIndex*, which represents the expert level and works as the dependent variables in the model. Except the overview on sample size, there are other data checking should be applied such as data structure, outliers and missing value checking. By running through the SPSS, basic statistical characteristics of each variable are illustrated by the descriptive statistic in the Appendix A. On the other hand, both outliers and missing value are found. There are more than 60 outliers detecting in univariate detection, whereas in multivariate detection the figure is 12 (see as Appendix A). As extreme values might be the minority group rather than outliers, there is a potential risk that the information might loss as the outliers are deleted directly (Bollen & Jackman, 1985). Thus, these outliers are reserved first for model building. Similarly, the only missing value is ignored during the modelling, because the sample size is larger enough that the impact of a missing value is negligible for the model performance.

Simple regression model

After the data preparation, an entry-force regression which established on the raw data is generated to illustrate the extent of assumptions violation. As the dependent variable is discrete, correlation matrix is preferred to apply on initial detecting of multicollinearity before the model building. The correlation matrix generates by SPSS (See as Appendix L) shown that four pairs independent variables are highly correlated. After model building, the following diagnostic check valid existence of multicollinearity. Despite that, the model result illustrates problems such as autocorrelation and numbers of independent variables are insignificant. Taking these problems into consideration, two methods are selected during the model design: the stepwise regression model and forced entry regression model based on factor analysis of independent variables. The former model could select most powerful variables simultaneously with eliminating multicollinearity by dropping variables from model. Whereas the latter model is expected to remedy multicollinearity by transforming variables and designed to be built basing on new variables which are generated from factor analysis.

Stepwise regression model

For the stepwise model, it applies the backward method by combining both the manual screening and machine selection. There are four pairs high correlated independent variables which include the correlation between *APM*， *NumberOfPACs* and *ActionLatency,* as well as correlations between *SelectByHotkeys* and *APM*. Here in this stage, the variable dropping achieves manually by repeating modelling. Then, as many independent variables as possible are retained when remedy multicollinearity. Thus, the deletion of *APM* becomes the charry pick selection. Whereas the model based on remaining 17 variables (Appendix N) still demonstrate insignificance of several independent variables, and the figure of $\overline{R^2} = 0.648$ is unchanged comparing with the regression based on raw data.

After the manual selecting, the further dropping procedures are generated through SPSS. As a result, the eventual model of stepwise regression keeps eight variables, which includes variables such as *TotalHours* and *AssignToHotkeys* (See as Appendix O). The model fit is satisfactory that all these variables have a significant impact on *LeagueIndex* with the p-value smaller than 5%. Meanwhile, the $\overline{R^2}$ with value reaches 0.65 represents overall fitting is slight improved as well. Following the evaluation of model fit, the diagnostic check illustrates that multiple regression assumptions are nicely followed except the assumption of error term independence. The figures of VIF which are smaller than 10 confirm that multicollinearity does not exist in this model. By graphing the histogram of residuals and p-p plot, the mean of error terms equals to 0. To check the homoscedasticity, partial regression plots between dependent variable and each independent variable are demonstrated in Appendix O. There are no obvious trends about heteroscedasticity. Meanwhile, most of variables show that their residuals are randomly distributed. Also, the model is suspected that contains the autocorrelation and the DW test would be applied for further detecting. The statistic DW that equals 1.123 implies that there is autocorrelation in this stepwise model.

To validate the general fitting of the model, 30% of the data from the data set is selected randomly as the test set. By running the test set through the model above (Appendix P), there is an insignificant variable ($WorkersMade$), and the overall fitting is improved significantly as the $\overline{R^2}$ increases to 0.7. The conclusion of diagnostic check is not change, whereas both the plots of residual normality and partial regression show minor change. Rather than the almost perfect normality in the model above, the normality of test set is weaker than the training set. On the other hand, there is no obvious relationship shown in partial regression plots, but the autocorrelation does not be remedied by using the DW test.

Regression model based on factor analysis

For the regression based on factor analysis (we use PC method and equamax rotation method, and extract 5 factors based on 18 independent variables) (Appendix Q), the initial check illustrates that assumption of no perfect multicollinearity are satisfied. Comparing with the model fit of stepwise regression model, the overall fitting of this model is slightly worse. Even the unified p-value that equals to 0 shows significance of all independent variables. The $\overline{R^2}$ demonstrates a lower overall fitting that 63.3% of *LeagueIndex* change could be explained.

The diagnostic check of this model also illustrates that no assumptions are violated apart from the autocorrelation. Except the calculating the correlation, figures of all tolerance and VIF verify no multicollinearity in the model. Residuals are very close to normal distribution even through there is a valley when standardized residuals ranged from -1 to 0. By analysing partial regression plots, all the plots contain several outliers, but no obvious pattern of heteroscedasticity are found. However, plot of F_1 illustrates a distinct linear relationship between residuals and dependent variable. As the statistic DW of 1.081 is smaller than the lower critical value, autocorrelation is tested as true in this regression model as well.

Validation of model above is generated by using same test set of the stepwise model (Appendix R). The model fitting is considerable that the $\overline{R^2}$ increase to 0.678 and all independent variables are powerful to explain dependent variable. Conclusions of diagnostic check do not change, but there are some minor changes in indicators as well. The first one is correlations between any two independent variables are no longer equals to 0. Another is the points of p-p plot are floating around 45-degree line with a high frequency and low amplitude, which indicates residuals might less normally distribute than the model based on training set.

Table 8 Coefficients

	Reg.(FA)			*Validation (FA)*		
	Std. Coef.	Sig.	VIF	Std. Coef.	Sig.	VIF
FACT1	0.634	0.000	1.000	0.688	0.000	1.029
FACT2	0.198	0.000	1.000	0.196	0.000	1.026
FACT3	0.284	0.000	1.000	0.318	0.000	1.025
FACT4	0.302	0.000	1.000	0.293	0.000	1.040
FACT5	0.148	0.000	1.000	0.152	0.000	1.014
R^2		0.635			0.686	
$\overline{R^2}$		0.633			0.678	

Regression models without outliers

The 12 possible outliers in multivariate detection are also founded in univariate detection and this number does not exceed 5% of the sample size, so these 12 possible outliers are considered to be removed. Then, the model building processes above are repeated without outliers. The statistic improvement of $\overline{R^2}$ is not obvious and validation results also show instability (Appendix S), so the models with original data set are preferred.

Interpretation and conclusion

Table 9 Comparison of Adjusted R square

Model	*Simple Reg.*	*Stepwise Reg.*	*Reg. (FA)*
Adjusted R square	0.648	0.65	0.633

The regression model with 18 original variables shows serious multicollinearity, although it has an adjusted R square with 0.648. Both the regression model with 8 original variables as independent variables and the model with 5 factors generated from factor analysis eliminate multicollinearity, and both of them have adjusted R square over 0.6. The validation result of the regression model with 5 factors demonstrates stability. However, the result in validation for the model with 8 original variables presents that the performance of this model is not stable. Therefore, the model with 5 factors is recommended in this case.

Proficiency in the game (Factor 1) includes cognitive-motor speed and effective operation of players. *Strategies used* (Factor 2) reflects some complex skills which usually bring more information and might increase cognitive load of players. *Attack efficiency* (Factor 3) is related to the attack skills of players, which is important for competitive games. The remaining two factors are *hours used* (Factor 4) and *age* (Factor 5).

Table 10 The Interpretation of 5 Factors

Factor	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*
Interpretation	Proficiency in the game	Strategies used	Attack efficiency	Hours invested	Age

The result shows that all 5 factors could positively affect expert level, which are *proficiency in the game*, *strategies used*, *attack efficiency*, *hours invested* and *age*. By comparing the standardized coefficients (Table 8), *proficiency in the game* has the largest effect on expert level, followed by *hours invested* and *attack efficiency*. Compared with the factors above, *strategies used* and *age* show the least effect on expert level. Therefore, it is necessary for players to learn how to improve response speed and efficient operation as well as how to automatize and train some skills to reduce some cognitive loads and develop other skills, such as attack skills. To improve the expert level, firstly, players could exercise their muscle memory through long hours of practice. Secondly, it is essential to do some targeted training to improve attentional, perceptual and operating speed. Lastly, players could actively learn, analyse and summarize game strategies through training and previous experience, to improve efficient operation and attack skills.

Limitation and further research

All three models above have a problem that the residuals are not uncorrelated. The possible reason is the limitation of multiple regression model for ordinal dependent variable. The multiple regression model does not cover and explain all the information of the dependent variable, so the ordinal part is left in the scatter plot of residuals. Thus, although the recommended model fits well and explains the influence relationship between dependent variable and independent variables, it may still be limited in terms of making prediction.

The further work would focus on in-depth research on two aspects. Firstly, other models for ordinal dependent variables are considered, such as ordinal regression model (Best, Wolf and Long, 2014) or some machine learning models; secondly, multiple regression model is kept trying and possible re-specification of the model is considered, such as using mapping functions to map ordinal variable into a specific contiguous interval (Torra et al., 2006).

References

Bollen, K. A., & Jackman, R. W. (1985). Regression diagnostics: An expository treatment of outliers and influential cases. Sociological Methods & Research, 13(4), 510-542.

Best, H., Wolf, C. and Long, J. scot. (2014) 'Regression Models for Nominal and Ordinal Outcomes', The SAGE Handbook of Regression Analysis and Causal Inference, pp. 173–204. doi: 10.4135/9781446288146.n9.

Everitt,B. and Landau,S.(2011). Cluster Analysis, 5th edition. London: A John Wildy and Sons, ltd.120-142.

Gujarati, D. N., & Porter, D. C. (2010). Essentials of econometrics (Vol. 2). Singapore: Irwin/McGraw-Hill.

Joseph F. Hair et al., The Essentials of Business Research Methods. New York: Routledge, 2015.

Thompson, J. J. et al. (2013) 'Video Game Telemetry as a Critical Tool in the Study of Complex Skill Learning', PLoS ONE, 8(9). doi: 10.1371/journal.pone.0075129.

Torra, V. et al. (2006) 'Regression for ordinal variables without underlying continuous variables', Information Sciences, 176(4), pp. 465–474. doi: 10.1016/j.ins.2005.07.007.

Minutes of Meetings

Discussion 1

- Date and Time of the meeting - Friday, 19th February 2021 3 PM GMT
- Recording URL - https://eu.bbcollab.com/collab/ui/session/playback/load/2ab725a635c04e4b883f8b32bcee7a30
- Duration – 29 mins
- Attendance - Parker Wu, Qianyu Gao, Jieshan Luo, Sisi Wang, Anirudh Batwara, Jiaxi Zeng
- Minutes of discussion – We introduced ourselves and shared our initial thoughts on the data and the 3 questions. An idea to spend the first week in data exploration was put forth. Another idea was proposed which suggested that we straightaway divide ourselves in groups of 2 each and pick a question. We agreed to go with the latter approach. Following groups were created -
    - Question 1 – Jiaxi Zeng and Sisi Wang
    - Question 2 – Qianyu Gao and Jieshan Luo
    - Question 3 - Parker Wu and Anirudh Batwara

  Further, it was discussed that before our meeting next week, all groups will do their independent data exploration and preparation. Agenda of the next week was to collate the thoughts on the data and discuss the next steps in the analysis.

Discussion 2

- Date and Time of the meeting - Friday, 26th February 2021 3 PM GMT
- Recording URL - https://eu.bbcollab.com/collab/ui/session/playback/load/729e2bd598544bc882e933446d986c13
- Duration – 41 mins
- Attendance - Parker Wu, Qianyu Gao, Jieshan Luo, Sisi Wang, Anirudh Batwara, Jiaxi Zeng
- Minutes of discussion – Each group presented their work in SPSS. Main topics discussed by every group included the initial exploration of data. Following characteristics of the data were discussed
    - Need for standardization of data
    - Missing value and its treatment
    - Multicollinearity
    - Treating Categorical variables as ordinal variables
    - Outliers

  It was also discussed about trying different number of factors in factor analysis. Further, it was discussed how to transform the variables for Q1 to remove multicollinearity from the data. We came onto conclusion that Q3 will be an input to Q1 and Q2.
  We also discussed that by our next meeting we should be ready with the first draft of our report. The next meeting was scheduled for 5th March, 2021.

Discussion 3

- Date and Time of the meeting – Friday, 5th March 2021 3 PM GMT

- Recording - https://web.microsoftstream.com/video/416a0a6b-aead-40bb-97d5-1670c508a4e2
- Duration – 27 mins
- Attendance - Parker Wu, Qianyu Gao, Jieshan Luo, Sisi Wang, Anirudh Batwara, Jiaxi Zeng
- Minutes of discussion – We discussed about the research papers and literature review that has to be added in our report. Also, the tentative length of every answer was discussed so that we can adhere to the page limit. Further, we discussed about the outliers and how they are to be treated in each question. Group 1 showed the analysis on the effect of factors. They showed how they are reading more on solving the issues of uncorrelated residuals. Parker showed the rotation analysis of factors. Use of data transformation techniques in Q1 was also discussed. The target before our next meeting was set to add literature review to our answers. The next meeting was scheduled for 12th March 2021.

Discussion 4

- Date and Time of the meeting – Friday, 12th March 2021 3 PM GMT
- Recording - https://web.microsoftstream.com/video/08714075-d56a-4d41-a279-6c88df39a219
- Duration – 22 mins
- Attendance - Parker Wu, Qianyu Gao, Jieshan Luo, Sisi Wang, Anirudh Batwara, Jiaxi Zeng
- Minutes of discussion – During the course of the week, Parker realized that we had not segregated the initial data set into training and test datasets. So he revised the factor analysis by splitting the data into two different datasets. We decided to put the data preparation analysis as the first chapter of the report as it is common to all the three questions. It was also decided that we will put all the SPSS tables and figures in the appendix to give optimum space to our interpretation and analysis. We debated if we should be including the leagueIndex variable in factor analysis or not. The pros and cons of the same were discussed and we agreed to review the results by both including and excluding leagueIndex variable and accordingly update the reports for every questions. It was decided that we will update all our independent reports to the questions by Sunday March 14, 2021 and then prepare a compiled report.

Discussion 5

- Date and Time of the meeting – Sunday, 14th March 2021 2:30 PM GMT
- Recording - https://web.microsoftstream.com/video/416a0a6b-aead-40bb-97d5-1670c508a4e2
- Duration – 53 mins
- Attendance - Parker Wu, Qianyu Gao, Jieshan Luo, Sisi Wang, Anirudh Batwara, Jiaxi Zeng
- Minutes of discussion – Every group presented their respective reports. The feedback on every report was discussed. Jiaxi suggested we should format all our SPSS figures and tables properly. She showed how this can be done and we all agreed that it was a better way to present our findings. Further, it was discussed that

in addition to the answers to the questions, we have to add the minutes of all our meetings and knit everything together in a single document. We agreed the deadline for us to prepare the pre-final draft will be Thursday 18th March 2021.

Discussion 6

- Date and Time of the meeting – Thursday, 18th March 2021 11:00 AM GMT
- Recording - https://my.wbs.ac.uk/-/academic/241039/wbslive/classroom/1205234/#OnlineClassroomSchedule1_C26D10384788B6DC6488612BEAA8C8F1_sessionlist_8
- Duration – 37 mins
- Attendance - Parker Wu, Qianyu Gao, Jieshan Luo, Sisi Wang, Anirudh Batwara, Jiaxi Zeng
- Minutes of discussion – The first draft of the completed version was discussed. The suggestions of the parts were discussed. We further decided to modify more tabled and figures to make them look professionally. The segments of individual answers were marked. We decided to take 2 more days to prepare the final draft.

Appendix A Data Preparation

Table 11 Descriptive Statistics

N=697

	Minimum	*Maximum*	*Mean*	*Std. Deviation*
Age	16.0000	43.0000	21.3917	4.4048
HoursPerWeek	2.0000	98.0000	16.1521	12.3668
TotalHours	7.0000	10000.0000	600.8420	702.3539
APM	24.9042	335.4990	103.7688	49.0019
SelectByHotkeys	0.0000	0.0356	0.0034	0.0043
AssignToHotkeys	0.0000	0.0014	0.0003	0.0002
UniqueHotkeys	0.0000	10.0000	4.0115	2.4003
MinimapAttacks	0.0000	0.0030	0.0001	0.0002
MinimapRightClicks	0.0000	0.0025	0.0004	0.0003
NumberOfPACs	0.0007	0.0080	0.0033	0.0010
GapBetweenPACs	6.6667	237.1429	45.2848	21.1229
ActionLatency	28.7342	173.5556	69.3546	22.3962
ActionsInPAC	2.0389	17.7619	5.0850	1.4671
TotalMapExplored	5.0000	53.0000	21.8321	7.5541
WorkersMade	0.0001	0.0032	0.0010	0.0005
UniqueUnitsMade	2.0000	13.0000	6.4634	1.8721
ComplexUnitsMade	0.0000	0.0007	0.0001	0.0001
ComplexAbilitiesUsed	0.0000	0.0027	0.0001	0.0003

Figure 3 Box Plot for Detecting Univariate Outliers

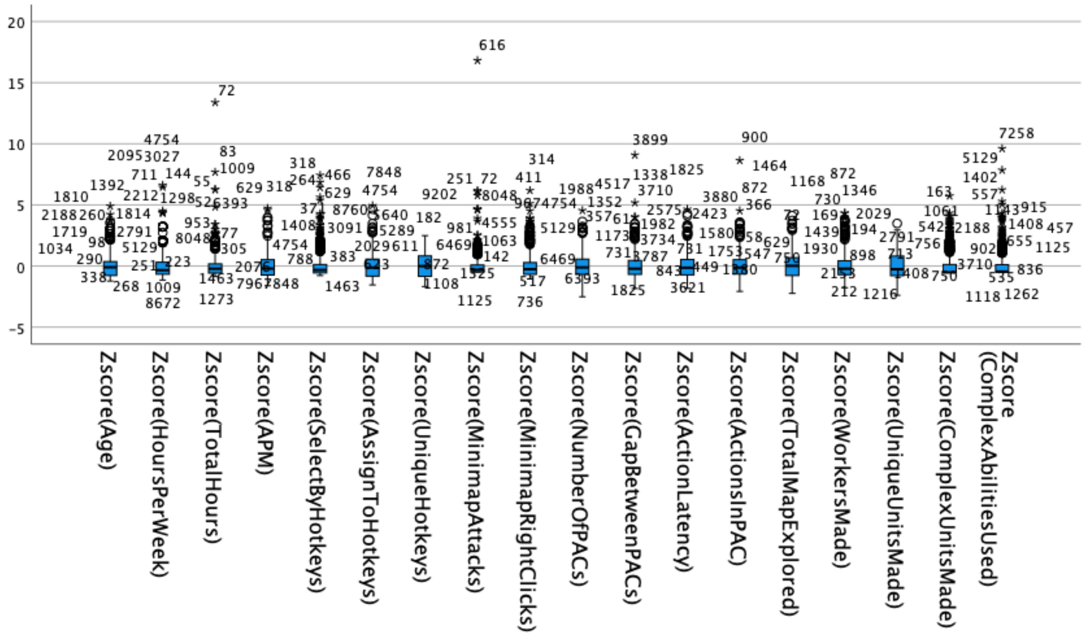
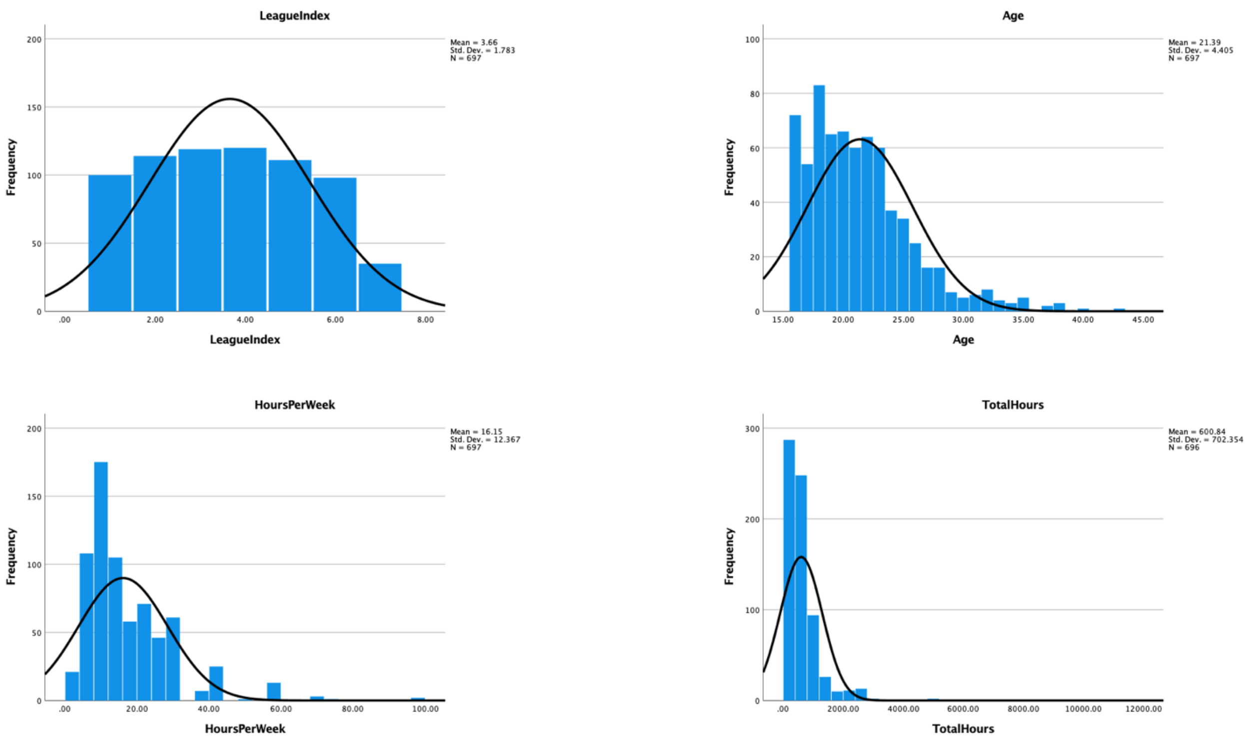
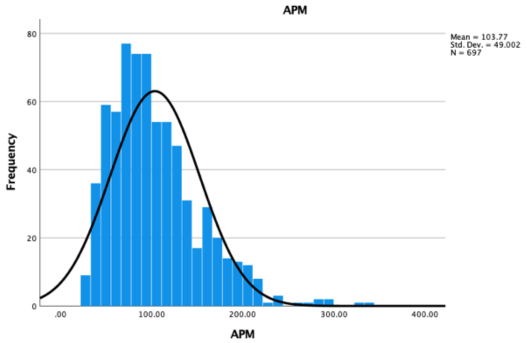
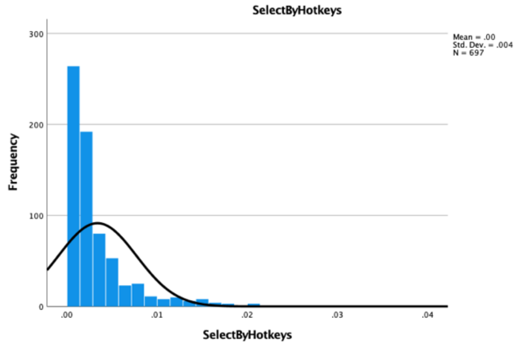
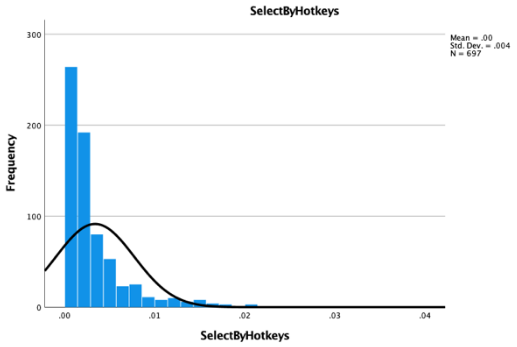


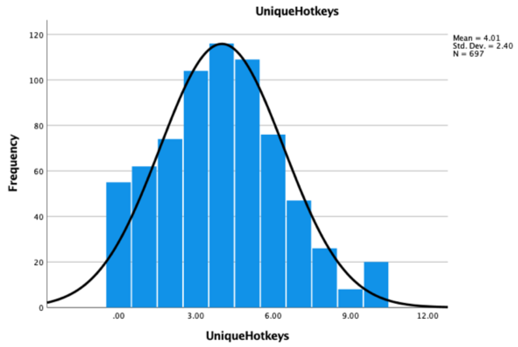
Figure 4 Normal Curve over Frequency Distribution of Individual Variables

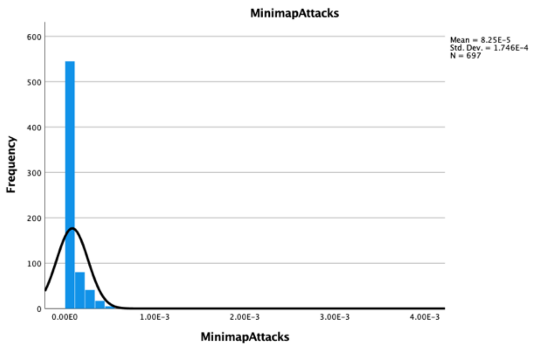


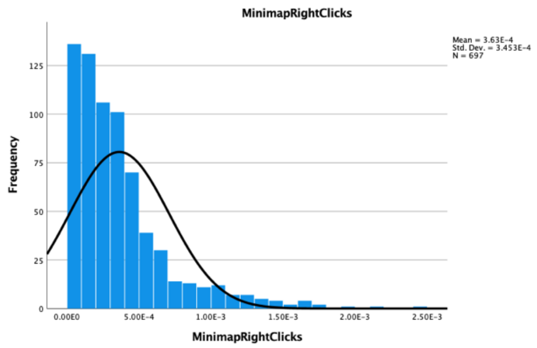

APM
Mean = 103.77
Std. Dev. = 49.002
N = 697
Frequency
APM

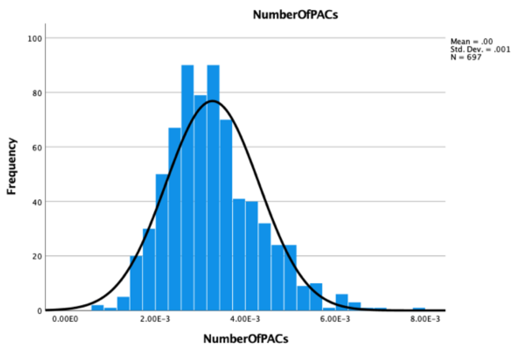

SelectByHotkeys
Mean = .00
Std. Dev. = .004
N = 697
Frequency
SelectByHotkeys

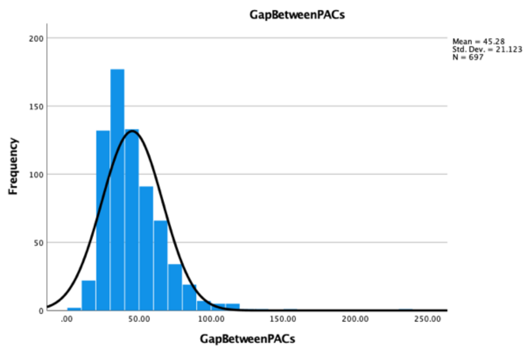

SelectByHotkeys
Mean = .00
Std. Dev. = .004
N = 697
Frequency
SelectByHotkeys

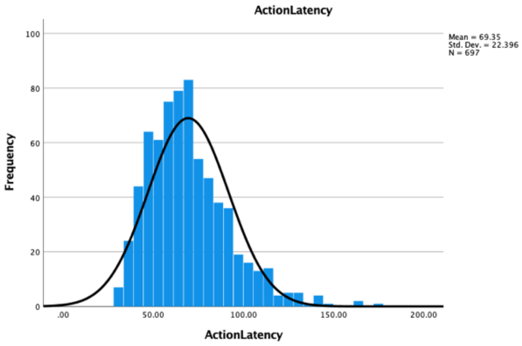

UniqueHotkeys
Mean = 4.01
Std. Dev. = 2.40
N = 697
Frequency
UniqueHotkeys

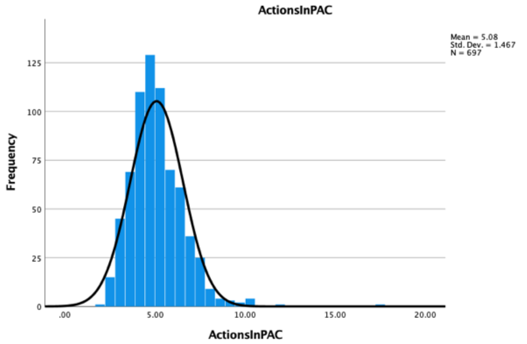

MinimapAttacks
Mean = 8.25E-5
Std. Dev. = 1.746E-4
N = 697
Frequency
MinimapAttacks

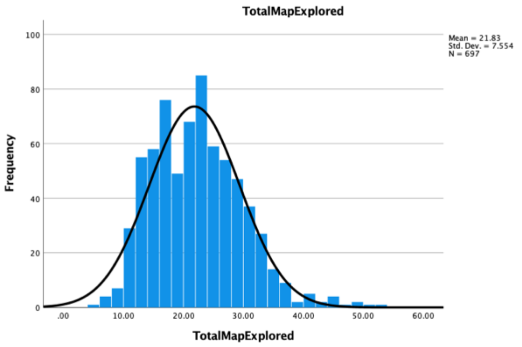

MinimapRightClicks
Mean = 3.63E-4
Std. Dev. = 3.453E-4
N = 697
Frequency
MinimapRightClicks

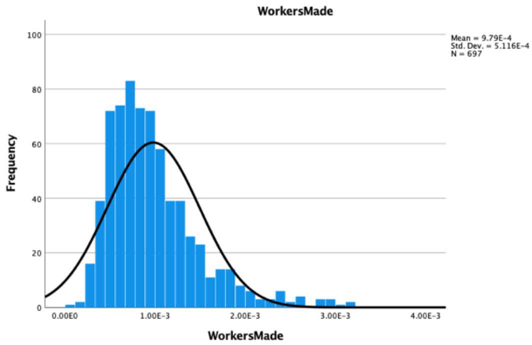

NumberOfPACs
Mean = .00
Std. Dev. = .001
N = 697
Frequency
NumberOfPACs

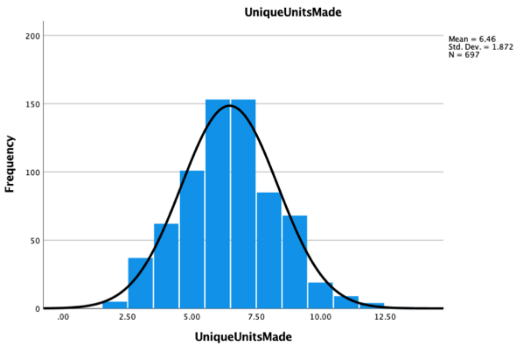

GapBetweenPACs
Mean = 45.28
Std. Dev. = 21.123
N = 697
Frequency
GapBetweenPACs

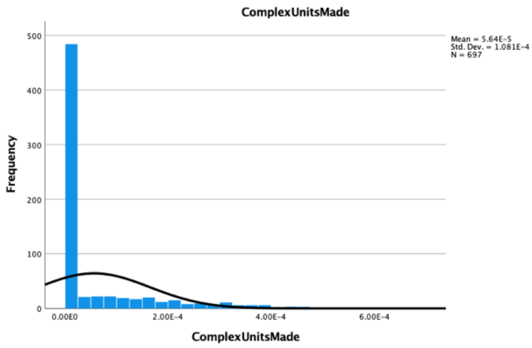

ActionLatency
Mean = 69.35
Std. Dev. = 22.396
N = 697
Frequency
ActionLatency


ActionsInPAC
Mean = 5.08
Std. Dev. = 1.467
N = 697
Frequency
ActionsInPAC


TotalMapExplored
Mean = 21.83
Std. Dev. = 7.554
N = 697
Frequency
TotalMapExplored


WorkersMade
Mean = 9.79E-4
Std. Dev. = 5.116E-4
N = 697
Frequency
WorkersMade


UniqueUnitsMade
Mean = 6.46
Std. Dev. = 1.872
N = 697
Frequency
UniqueUnitsMade


ComplexUnitsMade
Mean = 5.64E-5
Std. Dev. = 1.081E-4
N = 697
Frequency
ComplexUnitsMade

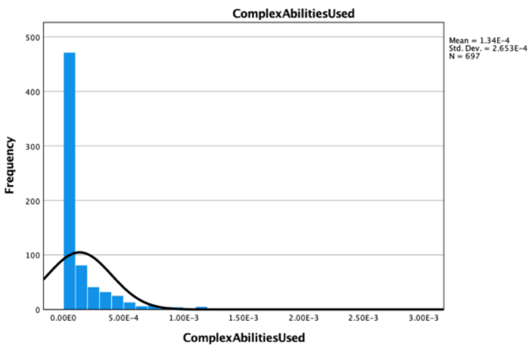

ComplexAbilitiesUsed
Mean = 1.34E-4
Std. Dev. = 2.653E-4
N = 697
Frequency
ComplexAbilitiesUsed

Figure 5 Bar Graph Showing the D2/df Ratio in Descending Rrder for All *GameIDs*. (When D2/df > 4 the *gameID* is considered to be an outlier.)

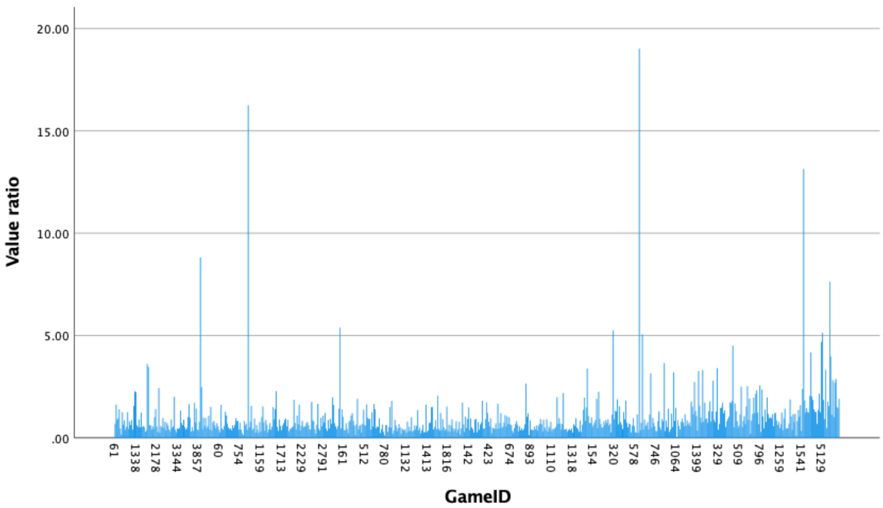


Table 12 Collinearity Statistics

	Unstd. Coef.	*Std. Error*	*Std Coef.*	*t*	*Sig.*	*Tolerance*	*VIF*
(Constant)	2.40E-15	0		0	1		
Zscore: APM	-4.47E-15	0	0	0	1	0.028	35.352
Zscore: SelectByHotkeys	2.40E-15	0	0	0	1	0.094	10.641
Zscore: NumberOfPACs	3.28E-16	0	0	0	1	0.072	13.932
Zscore: GapBetweenPACs	1.60E-16	0	0	0	1	0.428	2.335
Zscore: ActionLatency	-2.12E-15	0	0	0	1	0.174	5.741
Zscore: UniqueUnitsMade	-1.95E-16	0	0	0	1	0.58	1.725
Zscore: ComplexUnitsMade	1.08E-16	0	0	0	1	0.577	1.734
Zscore: Zscore: LeagueIndex	1	0	1	1.2E+09	0	0.343	2.919
Zscore: Zscore: Age	7.94E-18	0	0	0	1	0.883	1.133
Zscore: Zscore: HoursPerWeek	4.14E-17	0	0	0	1	0.752	1.33
Zscore: Zscore: TotalHours	5.17E-17	0	0	0	1	0.688	1.454

Zscore: Zscore: AssignToHotkeys	-9.29E-17	0	0	0	1	0.478	2.094
Zscore: Zscore: UniqueHotkeys	2.02E-16	0	0	0	1	0.673	1.486
Zscore: Zscore: MinimapAttacks	1.20E-16	0	0	0	1	0.812	1.231
Zscore: Zscore: MinimapRightClicks	2.72E-16	0	0	0	1	0.74	1.352
Zscore: Zscore: ActionsInPAC	9.80E-16	0	0	0	1	0.162	6.175
Zscore: Zscore: TotalMapExplored	9.68E-17	0	0	0	1	0.508	1.967
Zscore: Zscore: WorkersMade	2.30E-16	0	0	0	1	0.649	1.541
Zscore: Zscore: ComplexAbilitiesUsed	-1.22E-16	0	0	0	1	0.612	1.634

Table 13 Correlation of Independent Variables

Note: Shaded figures located independent variables which are highly correlated

	Age	*HoursPerWeek*	*TotalHours*	*APM*	*SelectByHotkeys*	*AssignToHotkeys*	*Unique Hotkeys*	*Minimap Attacks*	*Minimap Right Clicks*
Age	1								
HoursPerWeek	-0.202	1							
Total Hours	-0.054	0.415	1						
APM	-0.173	0.321	0.379	1					
SelectByHotkeys	-0.075	0.276	0.258	0.823	1				
AssignToHotkeys	-0.11	0.247	0.289	0.603	0.496	1			
UniqueHotkeys	0.012	0.156	0.186	0.399	0.359	0.482	1		
MinimapAttacks	0.026	0.116	0.267	0.259	0.15	0.251	0.177	1	
MinimapRightClicks	-0.055	0.117	0.159	0.37	0.138	0.207	0.129	0.215	1
NumberOfPACs	-0.175	0.29	0.382	0.735	0.496	0.59	0.445	0.219	0.252
GapBetweenPACs	0.102	-0.178	-0.259	-0.589	-0.314	-0.437	-0.22	-0.211	-0.282
ActionLatency	0.227	-0.255	-0.354	-0.76	-0.448	-0.541	-0.359	-0.208	-0.312
ActionsInPAC	-0.068	0.075	0.075	0.375	0.133	0.084	-0.062	0.108	0.328
TotalMapExplored	-0.022	0.126	0.253	0.319	0.201	0.277	0.338	0.238	0.179
WorkersMade	-0.108	0.118	0.235	0.484	0.24	0.313	0.194	0.124	0.276
UniqueUnitsMade	0.005	0.127	0.168	0.206	0.141	0.184	0.312	0.178	0.149
ComplexUnitsMade	-0.097	0.087	0.136	0.241	0.116	0.253	0.183	0.069	0.121
ComplexAbilitiesUsed	-0.055	0.102	0.145	0.204	0.103	0.245	0.163	0.072	0.127

(Continues)

	Number OfPACs	*Gap Between PACs*	*Action Latency*	*Actions InPAC*	*Total Map Explored*	*Workers Made*	*Unique Units Made*	*Complex Units Made*	*Complex Abilities Used*
Age									
HoursPerWeek									
Total Hours									
APM									
SelectByHotkeys									
AssignToHotkeys									
UniqueHotkeys									
MinimapAttacks									
MinimapRightClicks									
NumberOfPACs	1								
GapBetweenPACs	-0.537	1							
ActionLatency	-0.834	0.695	1						
ActionsInPAC	-0.166	-0.296	-0.143	1					
TotalMapExplored	0.491	-0.112	-0.394	-0.125	1				
WorkersMade	0.357	-0.349	-0.408	0.322	0.106	1			
UniqueUnitsMade	0.343	-0.107	-0.273	-0.109	0.587	0.095	1		
ComplexUnitsMade	0.287	-0.118	-0.277	0.062	0.335	0.185	0.383	1	
ComplexAbilitiesUsed	0.279	-0.151	-0.261	0.036	0.287	0.085	0.333	0.589	1

Appendix B Training Set Data Preparation

Table 14 Correlation Matrix of Training Set

	League Index	*Age*	*Hours Per Week*	*Total Hours*	*APM*	*Select By*	*Assign To Hotkeys*	*Unique Hotkeys*	*Minima p Attacks*
Age	1.000								
HoursPerWeek	-0.136	1.000							
Total Hours	0.216	-0.257	1.000						
APM	0.477	-0.051	0.395	1.000					
SelectByHotkeys	0.740	-0.214	0.255	0.386	1.000				
AssignToHotkeys	0.577	-0.118	0.199	0.285	0.815	1.000			
UniqueHotkeys	0.613	-0.126	0.221	0.295	0.605	0.569	1.000		
MinimapAttacks	0.394	-0.009	0.097	0.203	0.386	0.385	0.498	1.000	
MinimapRightClicks	0.390	0.048	0.059	0.266	0.315	0.233	0.280	0.243	1.000
NumberOfPACs	0.297	-0.096	0.133	0.093	0.393	0.160	0.163	0.098	0.218
GapBetweenPACs	0.699	-0.195	0.255	0.400	0.740	0.515	0.560	0.437	0.285
ActionLatency	-0.527	0.130	-0.167	-0.270	-0.579	-0.316	-0.383	-0.197	-0.233
ActionsInPAC	-0.685	0.263	-0.226	-0.372	-0.770	-0.458	-0.508	-0.337	-0.259
TotalMapExplored	0.125	-0.083	0.068	0.052	0.360	0.116	0.065	-0.100	0.061
WorkersMade	0.341	-0.013	0.110	0.245	0.316	0.232	0.261	0.328	0.278
UniqueUnitsMade	0.422	-0.021	0.090	0.228	0.409	0.230	0.320	0.198	0.153
ComplexUnitsMade	0.236	-0.009	0.076	0.154	0.187	0.116	0.195	0.307	0.174
ComplexAbilitiesUsed	0.242	-0.085	0.043	0.086	0.229	0.097	0.255	0.147	0.088

(Continues)

	Minimap Right Clicks	*Number Of PACs*	*Gap Between PACs*	*Action Latency*	*Actions In PAC*	*Total Map Explored*	*Workers Made*	*Unique Units Made*	*Complex Units Made*
Age									
HoursPerWeek									
Total Hours									
APM									
SelectByHotkeys									
AssignToHotkeys									
UniqueHotkeys									
MinimapAttacks									
MinimapRightClicks									
NumberOfPACs	1								
GapBetweenPACs	0.257	1							
ActionLatency	-0.262	-0.519	1						
ActionsInPAC	-0.326	-0.839	0.679	1					
TotalMapExplored	0.309	-0.183	-0.27	-0.12	1				
WorkersMade	0.163	0.502	-0.092	-0.384	-0.181	1			

UniqueUnitsMade	0.239	0.289	-0.31	-0.35	0.248	0.074	1		
ComplexUnitsMade	0.145	0.348	-0.092	-0.262	-0.149	0.579	0.048	1	
ComplexAbilitiesUsed	0.143	0.265	-0.112	-0.27	0.072	0.301	0.167	0.387	1

Table 15 Anti-image Matrix of Training Set

	League Index	*Age*	*Hours Per Week*	*Total Hours*	*APM*	*Select By*	*Assign To Hotkeys*	*Unique Hotkeys*	*Minimap Attacks*
Age	0.950								
HoursPerWeek	0.004	0.703							
Total Hours	0.078	0.234	0.721						
APM	-0.243	-0.092	-0.354	0.852					
SelectByHotkeys	-0.060	0.066	0.069	-0.017	0.676				
AssignToHotkeys	-0.051	-0.058	-0.075	0.026	-0.883	0.589			
UniqueHotkeys	-0.200	0.029	-0.082	0.050	0.105	-0.214	0.911		
MinimapAttacks	0.004	-0.045	0.030	0.009	0.000	-0.052	-0.278	0.917	
MinimapRightClicks	-0.168	-0.076	0.057	-0.108	-0.063	0.062	-0.060	-0.064	0.888
NumberOfPACs	-0.045	-0.009	-0.094	0.126	-0.227	0.224	0.044	0.015	-0.107
GapBetweenPACs	-0.098	-0.045	-0.119	-0.022	-0.680	0.538	-0.140	-0.047	0.052
ActionLatency	0.089	0.093	0.027	-0.019	0.033	-0.063	0.054	-0.022	0.075
ActionsInPAC	0.054	-0.162	-0.062	0.014	0.197	-0.201	-0.010	-0.018	-0.068
TotalMapExplored	0.003	-0.028	-0.092	-0.014	-0.796	0.647	-0.101	0.039	0.039
WorkersMade	0.006	-0.066	0.006	-0.054	0.029	-0.032	0.046	-0.051	-0.148
UniqueUnitsMade	-0.169	-0.088	0.016	-0.050	-0.148	0.164	-0.089	-0.066	0.045
ComplexUnitsMade	-0.011	-0.011	-0.016	-0.010	0.025	0.003	0.038	-0.151	0.002
ComplexAbilitiesUsed	-0.025	0.042	0.034	0.052	-0.037	0.072	-0.131	0.075	0.040

(Continues)

	Minimap Right Clicks	*Number Of PACs*	*Gap Between PACs*	*Action Latency*	*Actions In PAC*	*Total Map Explored*	*Workers Made*	*Unique Units Made*	*Complex Units Made*
Age									
HoursPerWeek									
Total Hours									
APM									
SelectByHotkeys									
AssignToHotkeys									
UniqueHotkeys									
MinimapAttacks									
MinimapRightClicks									
NumberOfPACs	0.845								
GapBetweenPACs	0.074	0.706							

ActionLatency	-0.027	0.000	0.898						
ActionsInPAC	-0.025	0.341	-0.379	0.906					
TotalMapExplored	0.010	0.785	0.085	0.026	0.237				
WorkersMade	-0.036	-0.150	-0.202	0.074	-0.013	0.840			
UniqueUnitsMade	-0.041	0.080	0.004	0.000	0.025	0.052	0.898		
ComplexUnitsMade	-0.073	-0.021	0.050	-0.016	0.026	-0.424	0.041	0.789	
ComplexAbilitiesUsed	0.017	0.007	-0.093	0.044	-0.031	-0.059	-0.074	-0.262	0.795

Appendix C Factor Analysis On Training Set with 5 Factors

Table 16 Total Variance Explained in Training Set With 5 Factors (Training Set)

Component	*Initial Eigenvalues*			*Extraction Sums of Squared Loadings*		
	Total	*% of Variance*	*Cumulative %*	*Total*	*% of Variance*	*Cumulative %*
1	6.274	33.02	33.02	6.274	33.02	33.02
2	1.943	10.226	43.246	1.943	10.226	43.246
3	1.392	7.327	50.573	1.392	7.327	50.573
4	1.242	6.536	57.109	1.242	6.536	57.109
5	1.062	5.588	62.698	1.062	5.588	62.698
6	0.984	5.179	67.876			
7	0.874	4.602	72.478			
8	0.795	4.185	76.663			
9	0.709	3.73	80.393			
10	0.681	3.583	83.976			
11	0.605	3.183	87.159			
12	0.545	2.867	90.026			
13	0.477	2.512	92.538			
14	0.401	2.11	94.648			
15	0.378	1.991	96.639			
16	0.27	1.422	98.06			
17	0.248	1.307	99.368			
18	0.098	0.516	99.884			
19	0.022	0.116	100			

Figure 6 Scree Plot of Training Set

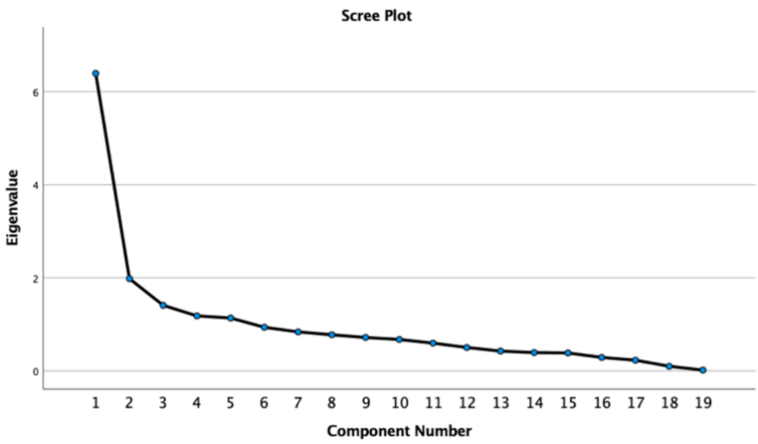


Table 17 Total Variance Explained - Maximum Likelihood

Component	*Initial Eigenvalues*			*Extraction Sums of Squared Loadings*		
	Total	*% of Variance*	*Cumulative %*	*Total*	*% of Variance*	*Cumulative %*
1	6.274	33.02	33.02	5.327	28.038	28.038
2	1.943	10.226	43.246	1.511	7.955	35.992
3	1.392	7.327	50.573	0.914	4.811	40.804
4	1.242	6.536	57.109	1.14	6.001	46.804
5	1.062	5.588	62.698	0.587	3.09	49.895
6	0.984	5.179	67.876			
7	0.874	4.602	72.478			
8	0.795	4.185	76.663			
9	0.709	3.73	80.393			
10	0.681	3.583	83.976			
11	0.605	3.183	87.159			
12	0.545	2.867	90.026			
13	0.477	2.512	92.538			
14	0.401	2.11	94.648			
15	0.378	1.991	96.639			
16	0.27	1.422	98.06			
17	0.248	1.307	99.368			
18	0.098	0.516	99.884			
19	0.022	0.116	100			

Table 18 Communalities in Principal Component Method (Training Set)

	Initial	*Extraction*
LeagueIndex	1	0.742
Age	1	0.645
HoursPerWeek	1	0.679
TotalHours	1	0.609
APM	1	0.883
SelectByHotkeys	1	0.636
AssignToHotkeys	1	0.614
UniqueHotkeys	1	0.505
MinimapAttacks	1	0.541
MinimapRightClicks	1	0.485
NumberOfPACs	1	0.812
GapBetweenPACs	1	0.523
ActionLatency	1	0.751
ActionsInPAC	1	0.72
TotalMapExplored	1	0.662

WorkersMade	1	0.387
UniqueUnitsMade	1	0.646
ComplexUnitsMade	1	0.64
ComplexAbilitiesUsed	1	0.431

Table 19 Communalities in Maximum Likelihood Method (Training Set)

	Initial	*Extraction*
LeagueIndex	0.683	0.741
Age	0.163	0.074
HoursPerWeek	0.247	0.084
TotalHours	0.353	0.256
APM	0.964	0.999
SelectByHotkeys	0.882	0.962
AssignToHotkeys	0.538	0.565
UniqueHotkeys	0.345	0.34
MinimapAttacks	0.225	0.201
MinimapRightClicks	0.282	0.267
NumberOfPACs	0.92	0.989
GapBetweenPACs	0.541	0.502
ActionLatency	0.833	0.816
ActionsInPAC	0.81	0.85
TotalMapExplored	0.488	0.593
WorkersMade	0.271	0.278
UniqueUnitsMade	0.412	0.579
ComplexUnitsMade	0.302	0.262
ComplexAbilitiesUsed	0.169	0.122

Appendix D Rotations in Factor Analysis with 5 Factors of Training Data Set

Table 20 Component Matrix (extract 5 factors, Training Set)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.848	-0.080	-0.071	0.105	0.015
Age	-0.219	0.195	0.039	0.699	0.264
HoursPerWeek	0.337	-0.132	-0.223	-0.600	0.371
TotalHours	0.520	-0.025	-0.241	-0.161	0.503
APM	0.893	-0.264	-0.012	0.056	-0.113
SelectByHotkeys	0.684	-0.150	-0.270	0.155	-0.220
AssignToHotkeys	0.724	0.006	-0.136	0.128	-0.233
UniqueHotkeys	0.536	0.295	-0.196	0.255	-0.166
MinimapAttacks	0.435	0.081	0.002	0.354	0.469
MinimapRightClicks	0.404	-0.220	0.440	0.000	0.282
NumberOfPACs	0.851	0.174	-0.191	-0.084	-0.114
GapBetweenPACs	-0.635	0.335	-0.063	-0.018	0.057
ActionLatency	-0.847	0.079	0.000	0.114	0.119
ActionsInPAC	0.160	-0.647	0.511	0.049	0.116
TotalMapExplored	0.502	0.608	0.035	-0.036	0.195
WorkersMade	0.464	-0.282	0.202	0.224	0.040
UniqueUnitsMade	0.390	0.649	0.216	-0.089	0.137
ComplexUnitsMade	0.365	0.356	0.567	-0.169	-0.172
ComplexAbilitiesUsed	0.263	0.286	0.492	-0.110	-0.161

Table 21 Rotated Component Matrix (Rotation=Varimax, extract 5 factors, Training Set)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.769	0.160	0.218	0.247	0.127
Age	-0.165	-0.075	-0.042	-0.287	0.727
HoursPerWeek	0.112	-0.008	0.057	0.772	-0.259
TotalHours	0.308	0.015	0.081	0.684	0.198
APM	0.843	0.101	0.357	0.182	-0.034
SelectByHotkeys	0.792	-0.062	0.043	0.061	0.008
AssignToHotkeys	0.768	0.146	0.033	0.038	0.021
UniqueHotkeys	0.599	0.216	-0.218	-0.030	0.227
MinimapAttacks	0.270	0.102	0.184	0.272	0.591
MinimapRightClicks	0.131	0.236	0.599	0.199	0.113
NumberOfPACs	0.789	0.308	-0.074	0.296	-0.019
GapBetweenPACs	-0.580	-0.018	-0.404	-0.138	0.071
ActionLatency	-0.748	-0.256	-0.217	-0.253	0.120
ActionsInPAC	0.015	-0.096	0.838	-0.043	-0.080
TotalMapExplored	0.279	0.603	-0.220	0.305	0.283
WorkersMade	0.395	0.034	0.458	-0.005	0.143
UniqueUnitsMade	0.133	0.718	-0.171	0.203	0.207
ComplexUnitsMade	0.127	0.756	0.175	-0.081	-0.123
ComplexAbilitiesUsed	0.078	0.615	0.154	-0.116	-0.096

Table 22 Rotated Component Matrix (Rotation=Direct Oblimin, extract 5 factors, Training Set)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.721	-0.128	0.055	0.102	0.164
Age	-0.166	-0.009	-0.092	0.742	-0.278
HoursPerWeek	-0.032	-0.018	-0.053	-0.262	0.803
TotalHours	0.151	-0.037	-0.079	0.193	0.688
APM	0.827	-0.260	0.010	-0.058	0.085
SelectByHotkeys	0.849	0.041	-0.168	-0.029	-0.033
AssignToHotkeys	0.796	0.056	0.051	-0.016	-0.058
UniqueHotkeys	0.622	0.288	0.118	0.188	-0.107
MinimapAttacks	0.143	-0.164	0.021	0.596	0.256
MinimapRightClicks	-0.006	-0.581	0.244	0.138	0.189
NumberOfPACs	0.750	0.182	0.194	-0.062	0.211
GapBetweenPACs	-0.563	0.340	0.034	0.080	-0.071
ActionLatency	-0.705	0.117	-0.175	0.148	-0.170
ActionsInPAC	-0.032	-0.851	-0.038	-0.040	-0.050
TotalMapExplored	0.135	0.278	0.532	0.261	0.281
WorkersMade	0.368	-0.425	0.004	0.148	-0.057
UniqueUnitsMade	-0.016	0.217	0.685	0.193	0.189
ComplexUnitsMade	0.033	-0.130	0.791	-0.126	-0.116
ComplexAbilitiesUsed	0.009	-0.122	0.650	-0.097	-0.144

Table 23 Rotated Component Matrix (Rotation=Quartimax, extract 5 factors, Training Set)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.831	0.100	0.112	0.142	0.099
Age	-0.181	-0.062	-0.036	-0.271	0.731
HoursPerWeek	0.206	-0.019	0.041	0.754	-0.257
TotalHours	0.409	-0.011	0.028	0.636	0.190
APM	0.901	0.038	0.248	0.070	-0.062
SelectByHotkeys	0.784	-0.125	-0.053	-0.042	-0.025
AssignToHotkeys	0.773	0.084	-0.065	-0.063	-0.013
UniqueHotkeys	0.582	0.164	-0.301	-0.111	0.194
MinimapAttacks	0.354	0.081	0.127	0.227	0.584
MinimapRightClicks	0.252	0.233	0.567	0.176	0.121
NumberOfPACs	0.826	0.242	-0.180	0.190	-0.054
GapBetweenPACs	-0.635	0.023	-0.330	-0.061	0.086
ActionLatency	-0.808	-0.198	-0.118	-0.153	0.147
ActionsInPAC	0.103	-0.084	0.835	-0.045	-0.061
TotalMapExplored	0.343	0.574	-0.279	0.259	0.265
WorkersMade	0.452	0.008	0.401	-0.060	0.135
UniqueUnitsMade	0.199	0.702	-0.212	0.177	0.196
ComplexUnitsMade	0.189	0.746	0.145	-0.101	-0.128
ComplexAbilitiesUsed	0.125	0.609	0.133	-0.129	-0.099

Table 24 Rotated Component Matrix (Rotation=Equamax, extract 5 factors, Training Set, threshold=0.494)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.659	0.180	0.307	0.318	0.283
Age	-0.168	-0.126	-0.055	-0.470	0.614
HoursPerWeek	0.000	-0.014	0.054	0.819	-0.070
TotalHours	0.160	-0.011	0.105	0.655	0.378
APM	0.741	0.141	0.457	0.304	0.117
SelectByHotkeys	0.757	-0.025	0.144	0.166	0.115
AssignToHotkeys	0.725	0.181	0.129	0.141	0.136
UniqueHotkeys	0.583	0.227	-0.142	0.004	0.306
MinimapAttacks	0.137	0.061	0.210	0.163	0.669
MinimapRightClicks	-0.007	0.230	0.602	0.190	0.180
NumberOfPACs	0.713	0.335	0.017	0.403	0.169
GapBetweenPACs	-0.494	-0.050	-0.472	-0.232	-0.032
ActionLatency	-0.654	-0.293	-0.302	-0.380	-0.050
ActionsInPAC	-0.076	-0.080	0.835	-0.018	-0.101
TotalMapExplored	0.190	0.578	-0.199	0.273	0.422
WorkersMade	0.310	0.047	0.505	0.019	0.184
UniqueUnitsMade	0.057	0.696	-0.169	0.174	0.317
ComplexUnitsMade	0.078	0.773	0.180	-0.020	-0.065
ComplexAbilitiesUsed	0.045	0.629	0.156	-0.069	-0.063

Table 25 Rotated Component Matrix (Rotation=Promax, extract 5 factors, Training Set)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.740	0.004	0.122	0.104	0.097
Age	-0.148	-0.103	0.064	-0.339	0.800
HoursPerWeek	-0.140	-0.104	0.003	0.917	-0.321
TotalHours	0.066	-0.154	0.049	0.741	0.162
APM	0.863	-0.034	0.236	0.026	-0.051
SelectByHotkeys	0.936	-0.229	-0.090	-0.091	-0.023
AssignToHotkeys	0.866	0.012	-0.083	-0.140	-0.025
UniqueHotkeys	0.692	0.083	-0.292	-0.205	0.170
MinimapAttacks	0.094	-0.027	0.218	0.223	0.614
MinimapRightClicks	-0.093	0.263	0.634	0.177	0.158
NumberOfPACs	0.776	0.149	-0.198	0.148	-0.110
GapBetweenPACs	-0.579	0.061	-0.321	-0.043	0.062
ActionLatency	-0.718	-0.146	-0.103	-0.117	0.173
ActionsInPAC	-0.079	-0.009	0.868	-0.036	0.030
TotalMapExplored	0.078	0.530	-0.216	0.215	0.193
WorkersMade	0.373	-0.007	0.435	-0.103	0.188
UniqueUnitsMade	-0.087	0.717	-0.140	0.115	0.123
ComplexUnitsMade	-0.015	0.877	0.195	-0.210	-0.170
ComplexAbilitiesUsed	-0.026	0.726	0.177	-0.225	-0.128

Appendix E Factor Analysis of Training Set with 6 Factors

Table 26 Total Variance Explained in Training Set With 6 Factors

Component	*Initial Eigenvalues*			*Extraction Sums of Squared Loadings*		
	Total	*% of Variance*	*Cumulative %*	*Total*	*% of Variance*	*Cumulative %*
1	6.274	33.02	33.02	6.274	33.02	33.02
2	1.943	10.226	43.246	1.943	10.226	43.246
3	1.392	7.327	50.573	1.392	7.327	50.573
4	1.242	6.536	57.109	1.242	6.536	57.109
5	1.062	5.588	62.698	1.062	5.588	62.698
6	0.984	5.179	67.876	0.984	5.179	67.876
7	0.874	4.602	72.478			
8	0.795	4.185	76.663			
9	0.709	3.73	80.393			
10	0.681	3.583	83.976			
11	0.605	3.183	87.159			
12	0.545	2.867	90.026			
13	0.477	2.512	92.538			
14	0.401	2.11	94.648			
15	0.378	1.991	96.639			
16	0.27	1.422	98.06			
17	0.248	1.307	99.368			
18	0.098	0.516	99.884			
19	0.022	0.116	100			

Table 27 Communalities in Principal Component Method with 6 Factors (Training Set)

	Initial	*Extraction*
LeagueIndex	1	0.744
Age	1	0.682
HoursPerWeek	1	0.765
TotalHours	1	0.721
APM	1	0.886
SelectByHotkeys	1	0.648
AssignToHotkeys	1	0.678
UniqueHotkeys	1	0.534
MinimapAttacks	1	0.544
MinimapRightClicks	1	0.62
NumberOfPACs	1	0.848
GapBetweenPACs	1	0.561
ActionLatency	1	0.812
ActionsInPAC	1	0.732
TotalMapExplored	1	0.712
WorkersMade	1	0.411
UniqueUnitsMade	1	0.677
ComplexUnitsMade	1	0.676
ComplexAbilitiesUsed	1	0.646

Appendix F Effect of Rotation on Factor Analysis with 6 Factors

Table 28 Component Matrix (extract 6 factors, Training Set)

	Component					
	1	*2*	*3*	*4*	*5*	*6*
LeagueIndex	0.848	-0.080	-0.071	0.105	0.015	0.037
Age	-0.219	0.195	0.039	0.699	0.264	0.192
HoursPerWeek	0.337	-0.132	-0.223	-0.600	0.371	0.293
TotalHours	0.520	-0.025	-0.241	-0.161	0.503	0.336
APM	0.893	-0.264	-0.012	0.056	-0.113	-0.051
SelectByHotkeys	0.684	-0.150	-0.270	0.155	-0.220	0.111
AssignToHotkeys	0.724	0.006	-0.136	0.128	-0.233	0.253
UniqueHotkeys	0.536	0.295	-0.196	0.255	-0.166	0.169
MinimapAttacks	0.435	0.081	0.002	0.354	0.469	-0.054
MinimapRightClicks	0.404	-0.220	0.440	0.000	0.282	-0.368
NumberOfPACs	0.851	0.174	-0.191	-0.084	-0.114	-0.191
GapBetweenPACs	-0.635	0.335	-0.063	-0.018	0.057	0.194
ActionLatency	-0.847	0.079	0.000	0.114	0.119	0.246
ActionsInPAC	0.160	-0.647	0.511	0.049	0.116	0.108
TotalMapExplored	0.502	0.608	0.035	-0.036	0.195	-0.223
WorkersMade	0.464	-0.282	0.202	0.224	0.040	0.155
UniqueUnitsMade	0.390	0.649	0.216	-0.089	0.137	-0.175
ComplexUnitsMade	0.365	0.356	0.567	-0.169	-0.172	0.188
ComplexAbilitiesUsed	0.263	0.286	0.492	-0.110	-0.161	0.465

Table 29 Rotated Component Matrix (Rotation=Varimax, extract 6 factors, Training Set)

	Component					
	1	*2*	*3*	*4*	*5*	*6*
LeagueIndex	0.763	0.197	0.249	0.226	0.078	0.054
Age	-0.134	-0.013	-0.082	-0.160	-0.014	0.794
HoursPerWeek	0.108	0.018	0.031	0.819	0.035	-0.283
TotalHours	0.313	0.132	0.072	0.759	0.013	0.160
APM	0.838	0.090	0.378	0.139	0.056	-0.101
SelectByHotkeys	0.798	-0.027	0.007	0.105	-0.009	0.008
AssignToHotkeys	0.777	0.041	-0.035	0.131	0.227	0.050
UniqueHotkeys	0.597	0.239	-0.221	0.016	0.160	0.214
MinimapAttacks	0.264	0.327	0.289	0.216	-0.104	0.475
MinimapRightClicks	0.110	0.253	0.736	0.017	0.005	-0.037
NumberOfPACs	0.750	0.475	0.064	0.139	0.009	-0.190
GapBetweenPACs	-0.571	-0.042	-0.453	-0.047	0.054	0.152
ActionLatency	-0.717	-0.323	-0.331	-0.094	-0.028	0.272
ActionsInPAC	0.052	-0.426	0.713	0.060	0.187	0.021
TotalMapExplored	0.227	0.793	-0.005	0.105	0.129	0.066
WorkersMade	0.418	-0.129	0.385	0.082	0.171	0.188
UniqueUnitsMade	0.085	0.763	0.013	0.030	0.294	0.027
ComplexUnitsMade	0.119	0.304	0.142	-0.043	0.733	-0.102
ComplexAbilitiesUsed	0.095	0.075	0.013	0.063	0.792	0.029

Table 30 Rotated Component Matrix (Rotation=Direct Oblimin, extract 6 factors, Training Set)

	Component					
	1	*2*	*3*	*4*	*5*	*6*
LeagueIndex	0.698	0.079	0.031	0.074	0.169	-0.149
Age	-0.137	-0.023	-0.014	0.785	-0.161	0.075
HoursPerWeek	-0.094	-0.032	0.025	-0.243	0.872	0.060
TotalHours	0.099	0.042	-0.015	0.203	0.786	0.034
APM	0.811	-0.034	0.012	-0.082	0.070	-0.269
SelectByHotkeys	0.837	-0.110	-0.061	0.003	0.041	0.115
AssignToHotkeys	0.779	-0.062	0.180	0.041	0.071	0.171
UniqueHotkeys	0.610	0.179	0.115	0.198	-0.037	0.291
MinimapAttacks	0.132	0.251	-0.128	0.513	0.201	-0.284
MinimapRightClicks	-0.007	0.175	0.012	0.012	-0.002	-0.772
NumberOfPACs	0.716	0.405	-0.044	-0.175	0.076	-0.033
GapBetweenPACs	-0.557	0.045	0.081	0.129	0.006	0.396
ActionLatency	-0.681	-0.233	0.012	0.249	-0.031	0.288
ActionsInPAC	-0.012	-0.530	0.211	0.045	0.057	-0.621
TotalMapExplored	0.102	0.761	0.105	0.088	0.084	-0.067
WorkersMade	0.372	-0.240	0.157	0.201	0.047	-0.281
UniqueUnitsMade	-0.043	0.733	0.284	0.043	0.017	-0.091
ComplexUnitsMade	0.023	0.219	0.744	-0.108	-0.061	-0.109
ComplexAbilitiesUsed	0.001	-0.018	0.804	0.013	0.059	0.076

Table 31 Rotated Component Matrix (Rotation=Quartimax, extract 6 factors, Training Set)

	Component					
	1	*2*	*3*	*4*	*5*	*6*
LeagueIndex	0.831	0.113	0.107	0.049	0.147	0.070
Age	-0.172	-0.002	-0.076	-0.005	-0.155	0.789
HoursPerWeek	0.195	0.010	0.020	0.025	0.808	-0.270
TotalHours	0.401	0.098	0.014	-0.003	0.721	0.176
APM	0.909	-0.007	0.221	0.027	0.054	-0.084
SelectByHotkeys	0.784	-0.103	-0.142	-0.036	0.030	0.013
AssignToHotkeys	0.773	-0.029	-0.180	0.200	0.059	0.055
UniqueHotkeys	0.569	0.191	-0.328	0.138	-0.039	0.213
MinimapAttacks	0.353	0.283	0.234	-0.115	0.175	0.491
MinimapRightClicks	0.268	0.207	0.711	-0.001	-0.010	-0.014
NumberOfPACs	0.805	0.400	-0.058	-0.023	0.065	-0.178
GapBetweenPACs	-0.646	0.033	-0.344	0.074	0.014	0.136
ActionLatency	-0.803	-0.240	-0.208	0.000	-0.021	0.255
ActionsInPAC	0.148	-0.462	0.676	0.190	0.046	0.036
TotalMapExplored	0.313	0.768	-0.026	0.113	0.074	0.075
WorkersMade	0.475	-0.186	0.291	0.159	0.035	0.200
UniqueUnitsMade	0.174	0.752	0.017	0.282	0.014	0.034
ComplexUnitsMade	0.194	0.290	0.123	0.725	-0.056	-0.098
ComplexAbilitiesUsed	0.135	0.071	-0.009	0.787	0.055	0.028

Table 32 Rotated Component Matrix (Rotation=Equamax, extract 6 factors, Training Set)

	Component					
	1	*2*	*3*	*4*	*5*	*6*
LeagueIndex	0.624	0.382	0.265	0.347	0.126	-0.042
Age	-0.023	-0.058	-0.002	-0.129	-0.027	0.813
HoursPerWeek	-0.045	-0.009	-0.007	0.806	0.046	-0.332
TotalHours	0.173	0.097	0.141	0.808	0.037	0.088
APM	0.680	0.524	0.166	0.264	0.106	-0.202
SelectByHotkeys	0.750	0.165	0.056	0.222	0.021	-0.076
AssignToHotkeys	0.725	0.112	0.109	0.245	0.258	-0.029
UniqueHotkeys	0.602	-0.095	0.298	0.115	0.188	0.159
MinimapAttacks	0.162	0.345	0.361	0.295	-0.069	0.421
MinimapRightClicks	-0.088	0.732	0.256	0.051	0.045	-0.076
NumberOfPACs	0.614	0.189	0.542	0.245	0.067	-0.278
GapBetweenPACs	-0.425	-0.549	-0.098	-0.132	0.015	0.226
ActionLatency	-0.541	-0.443	-0.384	-0.194	-0.085	0.360
ActionsInPAC	-0.068	0.706	-0.435	0.075	0.187	-0.003
TotalMapExplored	0.115	0.018	0.801	0.152	0.181	0.029
WorkersMade	0.331	0.463	-0.094	0.159	0.190	0.133
UniqueUnitsMade	-0.018	0.004	0.748	0.053	0.338	0.009
ComplexUnitsMade	0.029	0.128	0.270	-0.031	0.757	-0.111
ComplexAbilitiesUsed	0.054	0.001	0.037	0.071	0.798	0.023

Table 33 Rotated Component Matrix (Rotation=Promax, extract 6 factors, Training Set)

	Component					
	1	*2*	*3*	*4*	*5*	*6*
LeagueIndex	0.728	0.065	0.122	0.099	0.003	0.066
Age	-0.074	0.043	-0.008	-0.150	-0.019	0.801
HoursPerWeek	-0.100	-0.028	-0.060	0.873	0.040	-0.309
TotalHours	0.126	0.073	-0.020	0.764	-0.016	0.146
APM	0.832	-0.070	0.244	-0.010	-0.017	-0.083
SelectByHotkeys	0.940	-0.203	-0.155	-0.011	-0.061	0.007
AssignToHotkeys	0.895	-0.117	-0.194	0.027	0.172	0.041
UniqueHotkeys	0.707	0.139	-0.340	-0.069	0.105	0.206
MinimapAttacks	0.093	0.327	0.294	0.150	-0.164	0.500
MinimapRightClicks	-0.164	0.285	0.800	-0.066	-0.046	0.008
NumberOfPACs	0.700	0.357	-0.076	0.004	-0.074	-0.179
GapBetweenPACs	-0.528	0.064	-0.380	0.075	0.108	0.126
ActionLatency	-0.637	-0.206	-0.218	0.050	0.049	0.250
ActionsInPAC	-0.035	-0.448	0.758	0.034	0.191	0.041
TotalMapExplored	0.006	0.831	-0.021	0.039	0.055	0.078
WorkersMade	0.410	-0.206	0.344	0.006	0.136	0.200
UniqueUnitsMade	-0.147	0.833	0.028	-0.015	0.231	0.036
ComplexUnitsMade	0.001	0.333	0.142	-0.071	0.700	-0.109
ComplexAbilitiesUsed	0.054	0.084	-0.005	0.071	0.780	0.005

Appendix G Data Preparation for Factor Analysis of Test Set

Table 34 Correlation Matrix of Test Set

	League Index	*Age*	*Hours Per Week*	*Total Hours*	*APM*	*SelectByHotkeys*	*Assign To Hotkeys*	*Unique Hotkeys*	*Minimap Attacks*
Age	1.000								
HoursPerWeek	-0.119	1.000							
Total Hours	0.356	-0.113	1.000						
APM	0.411	-0.059	0.457	1.000					
SelectByHotkeys	0.692	-0.111	0.404	0.388	1.000				
AssignToHotkeys	0.461	-0.014	0.365	0.241	0.833	1.000			
UniqueHotkeys	0.637	-0.094	0.276	0.295	0.590	0.403	1.000		
MinimapAttacks	0.355	0.048	0.233	0.179	0.412	0.326	0.440	1.000	
MinimapRightClicks	0.272	0.007	0.164	0.256	0.219	0.091	0.234	0.143	1.000
NumberOfPACs	0.303	0.012	0.099	0.229	0.349	0.118	0.285	0.174	0.201
GapBetweenPACs	0.721	-0.143	0.331	0.384	0.725	0.472	0.617	0.451	0.170
ActionLatency	-0.608	0.048	-0.193	-0.276	-0.610	-0.317	-0.510	-0.250	-0.216
ActionsInPAC	-0.742	0.163	-0.295	-0.370	-0.750	-0.438	-0.590	-0.390	-0.173
TotalMapExplored	0.144	-0.041	0.088	0.109	0.406	0.163	0.116	0.000	0.164
WorkersMade	0.364	-0.032	0.141	0.268	0.316	0.156	0.273	0.345	0.219
UniqueUnitsMade	0.375	-0.154	0.060	0.182	0.431	0.219	0.299	0.090	0.094
ComplexUnitsMade	0.225	0.032	0.192	0.183	0.225	0.165	0.157	0.315	0.184
ComplexAbilitiesUsed	0.239	-0.116	0.151	0.188	0.258	0.142	0.247	0.215	0.079

(Continues)

	Minimap Right Clicks	Number Of PACs	Gap Between PACs	Action Latency	Actions In PAC	Total Map Explored	Workers Made	Unique Units Made	Complex Units Made
Age									
HoursPerWeek									
Total Hours									
APM									
SelectByHotkeys									
AssignToHotkeys									
UniqueHotkeys									
MinimapAttacks									
MinimapRightClicks									
NumberOfPACs	1.000								
GapBetweenPACs	0.251	1.000							
ActionLatency	-0.324	-0.568	1.000						
ActionsInPAC	-0.297	-0.828	0.723	1.000					
TotalMapExplored	0.361	-0.138	-0.347	-0.187	1.000				
WorkersMade	0.204	0.470	-0.139	-0.405	-0.034	1.000			
UniqueUnitsMade	0.252	0.349	-0.313	-0.415	0.284	0.030	1.000		
ComplexUnitsMade	0.158	0.329	-0.125	-0.284	-0.043	0.594	0.081	1.000	
ComplexAbilitiesUsed	0.088	0.306	-0.140	-0.280	0.055	0.376	0.113	0.350	1.000

Table 35 Anti-image Matrix of Test Set

	League Index	Age	Hours Per Week	Total Hours	APM	Select ByHotkeys	Assign To Hotkeys	Unique Hotkeys	Minimap Attacks
Age	0.954								
HoursPerWeek	-0.002	0.645							
Total Hours	-0.106	0.105	0.823						
APM	-0.074	-0.045	-0.359	0.872					
SelectByHotkeys	-0.038	0.062	-0.015	-0.074	0.641				
AssignToHotkeys	-0.003	-0.090	-0.042	0.081	-0.948	0.491			
UniqueHotkeys	-0.237	0.027	-0.013	0.015	0.024	-0.044	0.941		
MinimapAttacks	0.047	-0.114	-0.059	0.059	-0.063	0.029	-0.233	0.927	
MinimapRightClicks	-0.112	-0.022	-0.031	-0.112	-0.108	0.123	-0.090	-0.006	0.831
NumberOfPACs	-0.033	-0.069	0.055	-0.079	-0.081	0.116	-0.072	-0.020	-0.026
GapBetweenPACs	-0.041	-0.010	-0.039	0.021	-0.803	0.714	-0.099	-0.011	0.081
ActionLatency	0.137	0.111	-0.097	0.003	0.033	-0.066	0.098	-0.044	0.081
ActionsInPAC	0.176	-0.080	-0.024	0.003	0.051	-0.062	-0.020	0.019	-0.128
TotalMapExplored	0.039	-0.021	-0.038	0.051	-0.865	0.778	-0.028	0.044	0.046
WorkersMade	-0.109	-0.009	0.161	-0.094	-0.057	0.093	0.016	-0.063	-0.089
UniqueUnitsMade	-0.097	0.070	0.165	-0.036	-0.146	0.147	-0.039	0.086	0.022
ComplexUnitsMade	0.064	-0.062	-0.135	0.030	0.178	-0.189	0.115	-0.131	-0.102

ComplexAbilitiesUsed	-0.003	0.087	0.017	-0.079	-0.043	0.047	-0.069	-0.005	0.033

(Continues)

	Minimap Right Clicks	*Number Of PACs*	*Gap Between PACs*	*Action Latency*	*Actions In PAC*	*Total Map Explored*	*Workers Made*	*Unique Units Made*	*Complex Units Made*
Age									
HoursPerWeek									
Total Hours									
APM									
SelectByHotkeys									
AssignToHotkeys									
UniqueHotkeys									
MinimapAttacks									
MinimapRightClicks									
NumberOfPACs	0.924								
GapBetweenPACs	-0.030	0.651							
ActionLatency	0.027	0.066	0.887						
ActionsInPAC	-0.109	0.289	-0.360	0.926					
TotalMapExplored	-0.092	0.884	0.129	0.104	0.235				
WorkersMade	-0.042	-0.079	-0.282	0.095	-0.02	0.775			
UniqueUnitsMade	-0.042	-3.68E-05	-0.163	0.070	-0.014	0.251	0.839		
ComplexUnitsMade	-0.077	-0.110	0.073	0.045	-0.093	-0.469	-0.140	0.723	
ComplexAbilitiesUsed	0.084	0.017	-0.047	0.005	0.010	-0.111	-0.005	-0.151	0.821

Appendix H Implementation of Factor Analysis on Test Data Set

Table 36 Rotated Component Matrix (Rotation=Equamax, extract 5 factors, Test Set, threshold=0.483)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.714	0.190	0.246	0.338	-0.020
Age	-0.032	-0.090	0.004	-0.159	0.737
HoursPerWeek	0.165	0.062	-0.042	0.788	-0.173
TotalHours	0.147	0.109	0.142	0.749	-0.023
APM	0.729	0.151	0.391	0.357	-0.102
SelectByHotkeys	0.609	0.002	0.082	0.338	-0.073
AssignToHotkeys	0.676	0.197	0.206	0.202	0.018
UniqueHotkeys	0.491	0.336	-0.049	0.152	0.307
MinimapAttacks	-0.042	0.109	0.337	0.483	0.336
MinimapRightClicks	0.116	0.167	0.664	0.112	0.224
NumberOfPACs	0.817	0.351	0.002	0.252	-0.025
GapBetweenPACs	-0.634	-0.051	-0.479	-0.103	-0.009
ActionLatency	-0.779	-0.274	-0.274	-0.204	0.078
ActionsInPAC	-0.070	-0.066	0.849	0.090	-0.109
TotalMapExplored	0.192	0.705	-0.023	0.240	0.267
WorkersMade	0.380	0.027	0.483	-0.029	-0.270
UniqueUnitsMade	0.079	0.687	-0.023	0.226	0.285
ComplexUnitsMade	0.055	0.757	0.083	0.075	-0.266
ComplexAbilitiesUsed	0.058	0.685	0.176	-0.096	-0.237

Table 37 Rotated Component Matrix (Rotation=Varimax, extract 5 factors, Test Set, threshold=0.489)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.788	0.152	0.140	0.238	-0.007
Age	-0.079	-0.078	0.018	-0.128	0.740
HoursPerWeek	0.265	0.057	-0.070	0.754	-0.191
TotalHours	0.267	0.107	0.116	0.720	-0.042
APM	0.827	0.111	0.281	0.251	-0.090
SelectByHotkeys	0.654	-0.030	-0.006	0.255	-0.061
AssignToHotkeys	0.728	0.161	0.107	0.110	0.033
UniqueHotkeys	0.502	0.314	-0.118	0.094	0.316
MinimapAttacks	0.064	0.121	0.340	0.489	0.316
MinimapRightClicks	0.223	0.165	0.642	0.092	0.218
NumberOfPACs	0.853	0.306	-0.117	0.141	-0.008
GapBetweenPACs	-0.704	-0.016	-0.385	-0.015	-0.025
ActionLatency	-0.845	-0.231	-0.158	-0.093	0.061
ActionsInPAC	0.061	-0.062	0.849	0.084	-0.118
TotalMapExplored	0.246	0.699	-0.054	0.212	0.257
WorkersMade	0.445	0.001	0.422	-0.092	-0.259
UniqueUnitsMade	0.132	0.688	-0.038	0.213	0.272
ComplexUnitsMade	0.125	0.749	0.064	0.045	-0.276
ComplexAbilitiesUsed	0.114	0.676	0.158	-0.124	-0.242

Table 38 Rotated Component Matrix (Rotation=Promax, extract 6 factors, Test Set)

	Component					
	1	*2*	*3*	*4*	*5*	*6*
LeagueIndex	0.839	-0.091	0.077	-0.038	0.123	-0.084
Age	-0.157	-0.164	-0.151	0.120	0.092	0.885
HoursPerWeek	-0.118	0.033	0.911	-0.008	0.201	-0.125
TotalHours	0.046	-0.085	0.626	0.055	0.523	-0.186
APM	0.690	0.068	0.284	0.241	-0.091	0.079
SelectByHotkeys	0.424	0.011	0.473	0.056	-0.296	0.230
AssignToHotkeys	0.779	-0.028	-0.009	-0.035	0.007	0.019
UniqueHotkeys	0.428	0.137	0.034	-0.165	0.005	0.373
MinimapAttacks	-0.018	-0.160	0.187	0.245	0.771	0.066
MinimapRightClicks	0.212	0.062	-0.168	0.579	0.444	0.145
NumberOfPACs	0.948	0.030	-0.008	-0.325	-0.024	-0.093
GapBetweenPACs	-0.819	0.136	0.151	-0.238	-0.055	-0.013
ActionLatency	-0.945	-0.023	0.071	0.033	-0.005	0.114
ActionsInPAC	-0.122	0.160	0.094	0.958	0.161	0.060
TotalMapExplored	0.182	0.445	-0.052	-0.173	0.400	0.067
WorkersMade	0.588	-0.001	-0.204	0.305	-0.040	-0.282
UniqueUnitsMade	-0.006	0.507	0.013	-0.092	0.368	0.150
ComplexUnitsMade	-0.074	0.887	0.061	0.096	-0.066	-0.211
ComplexAbilitiesUsed	-0.089	0.911	-0.041	0.247	-0.230	-0.054

Table 39 Rotated Component Matrix (Rotation=Equamax, extract 5 factors, population)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.682	0.192	0.293	0.314	0.176
Age	-0.096	-0.154	-0.057	-0.422	0.585
HoursPerWeek	0.079	0.034	0.018	0.829	-0.053
TotalHours	0.129	0.072	0.128	0.716	0.263
APM	0.744	0.133	0.457	0.300	0.079
SelectByHotkeys	0.690	-0.017	0.138	0.234	0.062
AssignToHotkeys	0.681	0.188	0.160	0.195	0.130
UniqueHotkeys	0.556	0.209	-0.099	0.040	0.340
MinimapAttacks	0.030	0.008	0.249	0.294	0.636
MinimapRightClicks	0.046	0.150	0.619	0.105	0.273
NumberOfPACs	0.781	0.330	0.029	0.307	0.129
GapBetweenPACs	-0.564	-0.054	-0.482	-0.132	-0.011
ActionLatency	-0.723	-0.286	-0.306	-0.264	-0.027
ActionsInPAC	-0.098	-0.076	0.851	0.027	-0.073
TotalMapExplored	0.238	0.565	-0.128	0.188	0.458
WorkersMade	0.336	0.105	0.554	0.079	-0.004
UniqueUnitsMade	0.097	0.648	-0.106	0.119	0.410
ComplexUnitsMade	0.053	0.826	0.156	0.028	-0.079
ComplexAbilitiesUsed	0.039	0.790	0.129	0.045	-0.071

Table 40 Rotated Component Matrix (Rotation=Equamax, extract 5 factors, population without outliers)

	Component				
	1	*2*	*3*	*4*	*5*
LeagueIndex	0.660	0.179	0.329	0.357	0.149
Age	-0.140	-0.116	-0.091	-0.257	0.660
HoursPerWeek	0.005	0.017	0.018	0.862	-0.144
TotalHours	0.177	0.077	0.088	0.776	0.171
APM	0.721	0.139	0.502	0.316	0.010
SelectByHotkeys	0.718	-0.002	0.142	0.251	-0.027
AssignToHotkeys	0.706	0.174	0.173	0.176	0.068
UniqueHotkeys	0.583	0.229	-0.098	0.043	0.275
MinimapAttacks	0.145	0.038	0.264	0.254	0.608
MinimapRightClicks	0.003	0.162	0.649	0.072	0.262
NumberOfPACs	0.765	0.329	0.086	0.330	0.090
GapBetweenPACs	-0.499	-0.025	-0.551	-0.188	-0.010
ActionLatency	-0.674	-0.275	-0.362	-0.302	-0.002
ActionsInPAC	-0.106	-0.073	0.854	0.020	-0.074
TotalMapExplored	0.227	0.620	-0.102	0.207	0.391
WorkersMade	0.327	0.089	0.555	0.082	-0.044
UniqueUnitsMade	0.070	0.701	-0.098	0.148	0.315
ComplexUnitsMade	0.064	0.804	0.163	0.033	-0.151
ComplexAbilitiesUsed	0.057	0.779	0.155	-0.032	-0.155

Appendix I Dendogram of Hierarchical Cluster Analysis

Figure 7 Dendrogram of Hierarchical Cluster Analysis Using 5 Factor Scores from Question1 with Outliers, Standardised Data, Complete Linkage, Euclidean Distance.

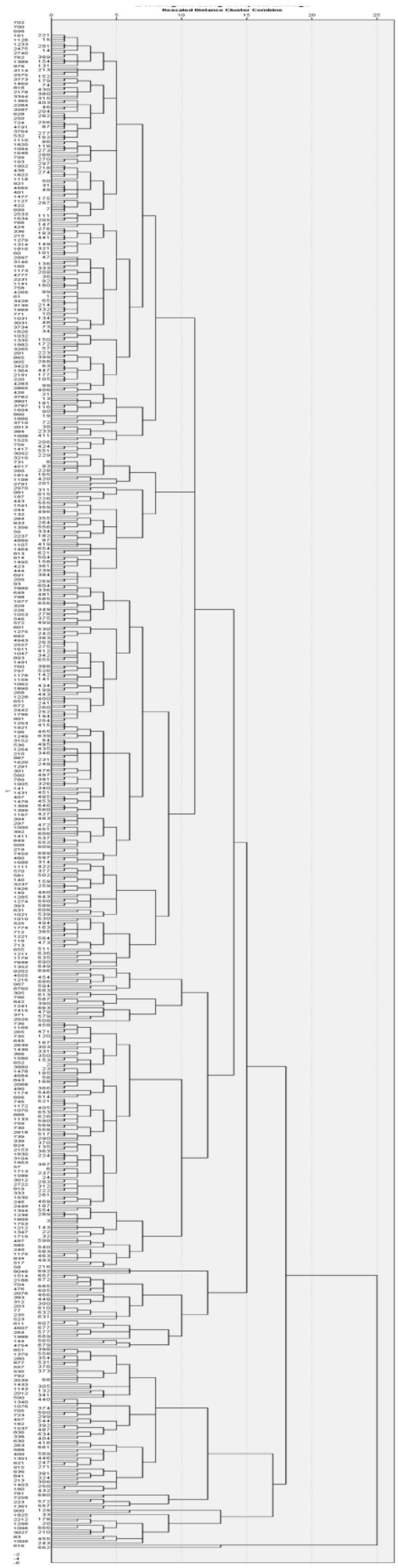
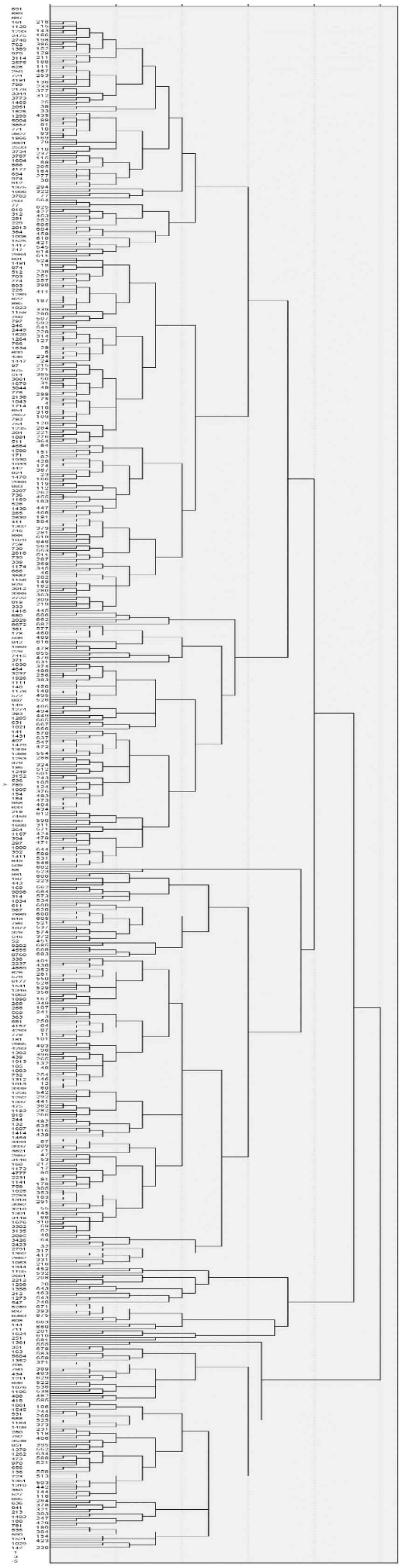


Figure 8 Dendrogram of Hierarchical Cluster Analysis Using 5 Factor Scores from Question1 without Outliers, Standardised Data, Complete Linkage, Euclidean Distance.



Appendix J Cluster Centres

Cluster centres for 8 cluster solution, from Hierarchical cluster analysis using 5 factors score, standardised data, Complete Linkage, Euclidean distance

Table 41 Initial Cluster Centres

	Cluster							
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
FAC1_1	-0.44	-0.15	0.15	-0.62	1.16	-0.75	-0.86	0.99
FAC2_1	-0.32	-0.15	1.68	-0.02	-0.46	0.08	-0.85	0.63
FAC3_1	-0.14	-0.16	0.63	0.41	-0.1	0.46	3.98	-0.89
FAC4_1	0.11	-0.51	-0.43	2.61	-0.07	1.59	-0.56	2.84
FAC5_1	-0.49	0.93	-0.36	-0.49	-0.23	2.98	1.2	-0.15

Table 42 Final Cluster Centres

	Cluster							
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
FAC1_1	-0.48	-0.34	0.11	-0.54	1.31	0.43	-0.6	0.44
FAC2_1	-0.43	-0.04	1.85	-0.32	-0.4	-0.39	-0.39	0.7
FAC3_1	-0.29	-0.83	0.43	0.3	0.17	0.59	1.42	-0.89
FAC4_1	0.03	-0.57	-0.43	2.52	-0.12	1.09	-0.49	1.43
FAC5_1	-0.72	0.77	-0.42	-0.65	-0.08	2.11	0.19	0.09

Table 43 Distances between Final Cluster Centres

Cluster	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1		1.737	2.524	2.561	1.96	3.279	2.005	2.259
2	1.737		2.608	3.596	2.189	2.708	2.363	2.372
3	2.524	2.608		3.722	2.6	3.717	2.621	2.62
4	2.561	3.596	3.722		3.279	3.272	3.32	2.271
5	1.96	2.189	2.6	3.279		2.693	2.332	2.351
6	3.279	2.708	3.717	3.272	2.693		2.825	2.751
7	2.005	2.363	2.621	3.32	2.332	2.825		3.362
8	2.259	2.372	2.62	2.271	2.351	2.751	3.362	

Table 44 ANOVA

	Cluster		Error		F	Sig.
	Mean Square	*df*	*Mean Square*	*df*		
FAC1_1	44.82	7	0.467	677	95.905	0
FAC2_1	52.228	7	0.407	677	128.384	0
FAC3_1	48.272	7	0.468	677	103.196	0
FAC4_1	46.585	7	0.352	677	132.365	0
FAC5_1	41.821	7	0.417	677	100.366	0

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Table 45 Number of Cases in Each Cluster

Cluster	1	180
	2	137
	3	80
	4	21
	5	120
	6	21
	7	83
	8	43
Valid		685
Missing		12

Appendix K GameIDs in Each Cluster

Table 46 *GameID* in Each Cluster

C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*
61	363	590	1298	1030	2970	171	1247
532	731	3539	1156	1275	1624	779	4177
600	1013	263	2212	1611	1810	1228	694
699	1016	701	2661	601	127	1347	2984
766	1141	1029	3027	833	52	1366	1237
771	1173	1433	547	1491	55	1580	1375
1074	1338	1774	1255	1688	981	1719	1886
1128	1604	2012	225	81	251	2095	77
1447	1826	2625	529	141	381	2423	193
1469	1852	209	837	168	513	3265	818
1603	2231	257	1008	184	611	3426	997
1634	2997	396	212	226	750	3880	1004
1648	3135	527	711	284	967	4263	1124
1679	3146	621	759	392	398	4684	1182
1825	3210	630	1009	464	953	181	1417
1830	3237	713	1174	570	4555	187	1454
1902	3302	792	1273	685	4699	645	162
2013	3339	910	383	760	5640	672	281
2051	3423	915	523	784	6469	735	312
2178	3449	1397	1105	893	8048	883	623
2533	3494	1408	1096	1023	9202	1033	655
2677	3621	1521		1047		1212	1274
2988	3782	1848		1167		1380	1281
3031	3787	142		1289		1478	144
3044	4269	213		1431		1809	203
3061	4276	280		1479		1862	235
3104	4283	350		132		1890	476
3152	4510	535		154		2153	608
3287	4517	542		244		2314	638
3344	4777	557		267		2442	704
3389	4889	588		297		2639	756
3428	4943	674		299		2727	892
3580	4960	712		301		58	1125
3587	5004	751		316		98	1285
3710	60	790		320		201	1514
3734	215	851		394		230	1525
3752	220	942		403		512	1463
3764	288	998		407		650	2076
3773	689	1061		525		652	2188
3801	758	1143		536		703	2915
3843	1031	1310		568		736	5289
3857	1071	1340		572		843	6393
3927	1083	118		578		902	7848
4157	1178	182		581		1488	

C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*
4191	1301	229		626		1575	
4565	1312	302		656		1959	
234	1314	408		722		180	
333	1392	434		787		228	
422	1495	435		797		268	
584	1582	457		798		517	
628	1814	458		888		526	
651	1913	576		993		824	
692	1982	632		1021		841	
754	2191	705		1077		874	
776	2228	723		1336		1169	
793	2237	836		1341		1175	
816	2253	877		1346		1208	
819	2662	1076		1356		1325	
891	2729	1106		1399		1351	
901	2791	1361		158		1430	
979	2865	1379		169		139	
1093	2995	1402		219		178	
1127	3037	138		256		265	
1159	3092	163		264		366	
1230	160	223		305		443	
1233	260	351		314		506	
1335	290	400		329		585	
1365	321	418		339		636	
1369	439	531		371		666	
1389	444	970		388		781	
1713	569	1164		393		872	
1714	732	1168		454		875	
1820	765	1179		473		1034	
1876	822	1211		480		1133	
1966	866	1262		497		1344	
2068	905	1277		508		1403	
2138	912	1358		509		194	
2221	995	1537		540		240	
2229	1025	1545		609		361	
2284	1053	5604		631		411	
2449	1193			637		455	
2475	1232			649		710	
2557	1238			680		834	
2575	1265			730			
2618	1279			788			
2657	1292			796			
2722	1382			842			
2740	1595			849			
2974	1607			898			
2978	1676			1000			
2989	1753			1010			

C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*
3012	1816			1221			
3114	1839			1241			
3139	1926			1249			
57	2077			1259			
97	56			1286			
161	93			1309			
175	105			1333			
204	266			1348			
342	272			1352			
378	336			1383			
384	338			1411			
401	389			1439			
426	423			1476			
436	424			1493			
449	475			1541			
604	546			357			
633	691			1216			
654	716			1988			
681	865			2029			
724	966			3091			
774	1062			3125			
780	1063			4607			
958	1107			6177			
974	1108			7415			
975	1111			7459			
1043	1137			7889			
1091	1210			8672			
1094	1248			8760			
1132	1299			9098			
1150	1318						
1154	1319						
1172	1337						
1235	1364						
1240	140						
1253	149						
1339	545						
1413	614						
1477	747						
1588	887						
1620	931						
1626	941						
1692	1256						
1767	1414						
1770	813						
1796	1293						
1822	1464						
1905							

C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*
1921							
1930							
1953							
1969							
210							
442							
490							
511							
514							
519							
591							
607							
683							
762							
789							
799							
921							
928							
961							
987							
1032							
1055							
1110							
1118							
1207							
1264							
1291							
1302							
1416							
196							
246							
250							
438							
560							
664							
720							
739							
746							
1070							
1196							
1390							
643							

Appendix L Correlation of Independent Variables

Table 47 Correlation of Independent Variables

Note: Shaded figures located independent variables which are highly correlated

	Age	*HoursPerWeek*	*TotalHours*	*APM*	*SelectByHotkeys*	*AssignToHotkeys*	*UniqueHotkeys*	*MinimapAttacks*	*Minimap Right Clicks*
Age	1.000								
HoursPerWeek	-0.202	1.000							
Total Hours	-0.054	0.415	1.000						
APM	-0.173	0.321	0.379	1.000					
SelectByHotkeys	-0.075	0.276	0.258	0.823	1.000				
AssignToHotkeys	-0.110	0.247	0.289	0.603	0.496	1.000			
UniqueHotkeys	0.012	0.156	0.186	0.399	0.359	0.482	1.000		
MinimapAttacks	0.026	0.116	0.267	0.259	0.150	0.251	0.177	1.000	
MinimapRightClicks	-0.055	0.117	0.159	0.370	0.138	0.207	0.129	0.215	1.000
NumberOfPACs	-0.175	0.290	0.382	0.735	0.496	0.590	0.445	0.219	0.252
GapBetweenPACs	0.102	-0.178	-0.259	-0.589	-0.314	-0.437	-0.220	-0.211	-0.282
ActionLatency	0.227	-0.255	-0.354	-0.760	-0.448	-0.541	-0.359	-0.208	-0.312
ActionsInPAC	-0.068	0.075	0.075	0.375	0.133	0.084	-0.062	0.108	0.328
TotalMapExplored	-0.022	0.126	0.253	0.319	0.201	0.277	0.338	0.238	0.179
WorkersMade	-0.108	0.118	0.235	0.484	0.240	0.313	0.194	0.124	0.276
UniqueUnitsMade	0.005	0.127	0.168	0.206	0.141	0.184	0.312	0.178	0.149
ComplexUnitsMade	-0.097	0.087	0.136	0.241	0.116	0.253	0.183	0.069	0.121
ComplexAbilitiesUsed	-0.055	0.102	0.145	0.204	0.103	0.245	0.163	0.072	0.127

(Continues)

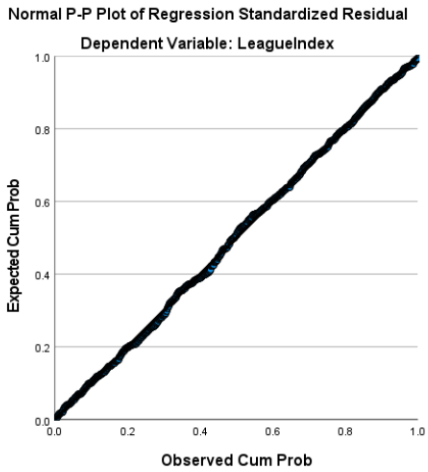
	Number Of PACs	Gap Between PACs	Action Latency	Actions In PAC	Total Map Explored	Workers Made	Unique Units Made	Complex Units Made	Complex Abilities Used
Age									
HoursPerWeek									
Total Hours									
APM									
SelectByHotkeys									
AssignToHotkeys									
UniqueHotkeys									
MinimapAttacks									
MinimapRightClicks									
NumberOfPACs	1.000								
GapBetweenPACs	-0.537	1.000							
ActionLatency	-0.834	0.695	1.000						
ActionsInPAC	-0.166	-0.296	-0.143	1.000					
TotalMapExplored	0.491	-0.112	-0.394	-0.125	1.000				
WorkersMade	0.357	-0.349	-0.408	0.322	0.106	1.000			
UniqueUnitsMade	0.343	-0.107	-0.273	-0.109	0.587	0.095	1.000		
ComplexUnitsMade	0.287	-0.118	-0.277	0.062	0.335	0.185	0.383	1.000	
ComplexAbilitiesUsed	0.279	-0.151	-0.261	0.036	0.287	0.085	0.333	0.589	1.000

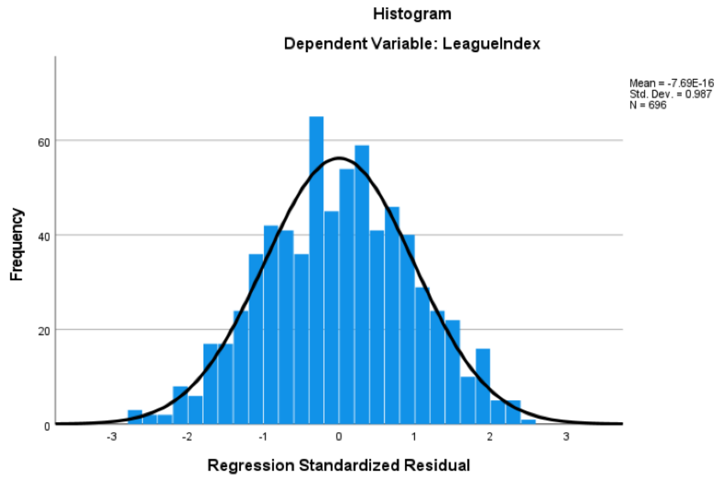
Appendix M Simple Regression Model

Table 48 Model Results of Simple Regression Model

	Coef.	*Std. Errors*	*Std. Coef.*	*t-value*	*p-value*	*Toler-ance*	*VIF*
(Constant)	2.286	0.705		3.243	0.001		
Age	0.004	0.010	0.009	0.389	0.698	0.883	1.133
HoursPerWeek	-0.001	0.004	-0.008	-0.313	0.755	0.752	1.329
TotalHours	0.000	0.000	0.103	3.834	0.000	0.703	1.423
APM	0.006	0.005	0.152	1.136	0.256	0.028	35.285
SelectByHotkeys	24.687	30.117	0.060	0.820	0.413	0.094	10.631
AssignToHotkeys	1632.990	264.571	0.195	6.172	0.000	0.505	1.982
UniqueHotkeys	-0.007	0.020	-0.010	-0.361	0.718	0.673	1.486
MinimapAttacks	889.154	252.526	0.087	3.521	0.000	0.827	1.209
MinimapRightClicks	106.284	135.047	0.021	0.787	0.432	0.740	1.351
NumberOfPACs	282.603	144.506	0.164	1.956	0.051	0.072	13.854
GapBetweenPACs	-0.007	0.003	-0.086	-2.503	0.013	0.432	2.314
ActionLatency	-0.013	0.004	-0.167	-3.115	0.002	0.177	5.660
ActionsInPAC	-0.026	0.068	-0.022	-0.387	0.699	0.162	6.173
TotalMapExplored	0.007	0.007	0.030	0.937	0.349	0.509	1.964
WorkersMade	292.854	96.672	0.084	3.029	0.003	0.658	1.520
UniqueUnitsMade	-0.018	0.028	-0.019	-0.644	0.520	0.580	1.724
ComplexUnitsMade	-149.272	488.314	-0.009	-0.306	0.760	0.577	1.734
ComplexAbilitiesUsed	200.985	193.037	0.030	1.041	0.298	0.613	1.631
	R^2		0.657		$\overline{R^2}$		0.648
					Durbin-Watson		1.123

Figure 9 Residual Distribution of Simple Regression Model





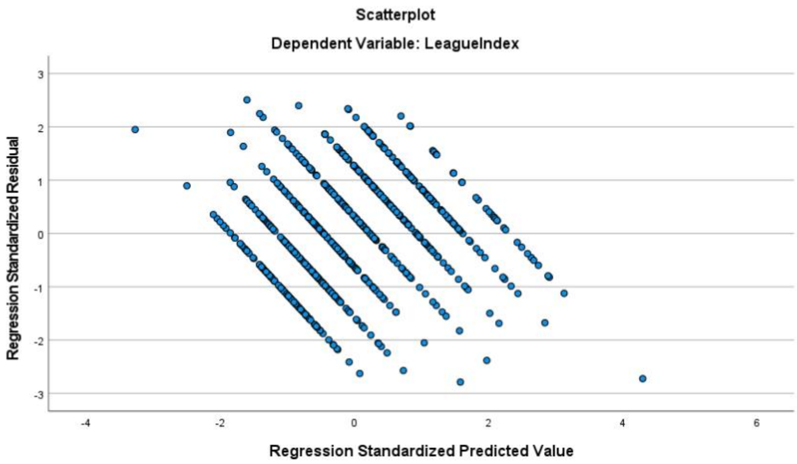
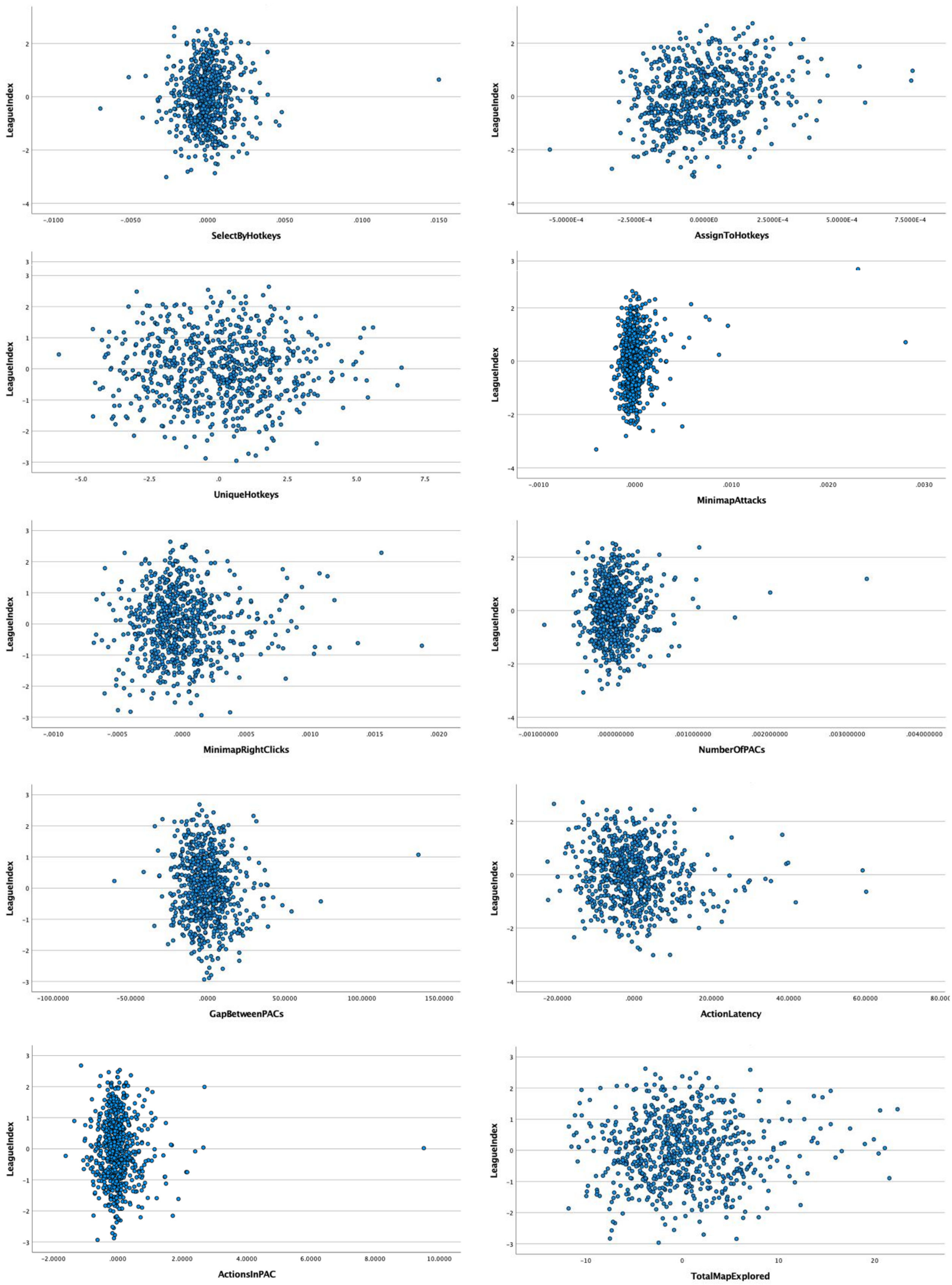
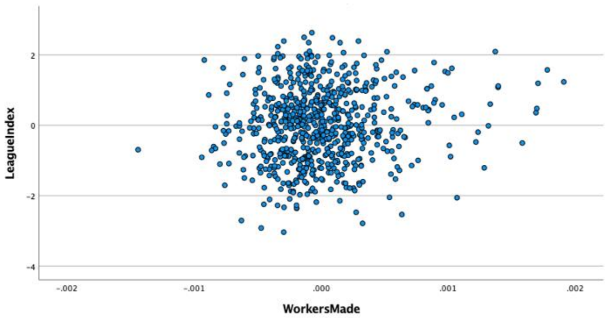
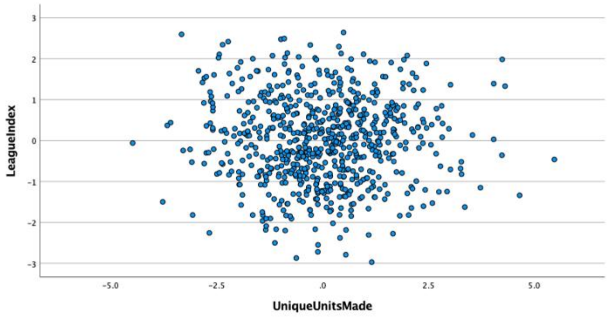
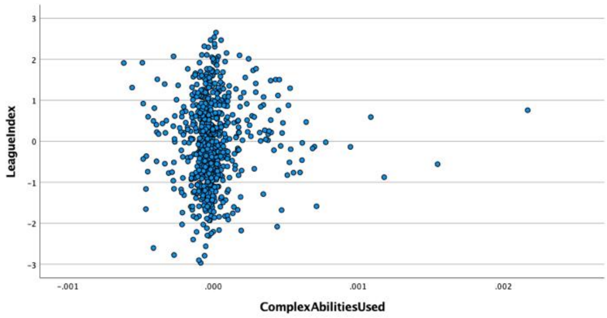


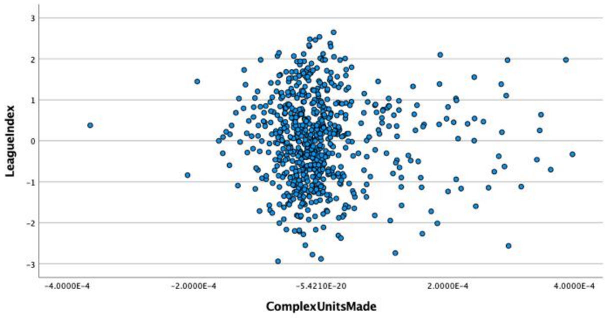
Figure 10 Partial Regression Plots of Simple Regression Model




LeagueIndex
WorkersMade
2
0
-2
-4
-.002
-.001
.000
.001
.002


LeagueIndex
UniqueUnitsMade
3
2
1
0
-1
-2
-3
-5.0
-2.5
.0
2.5
5.0


LeagueIndex
ComplexAbilitiesUsed
3
2
1
0
-1
-2
-3
-.001
.000
.001
.002


LeagueIndex
ComplexUnitsMade
3
2
1
0
-1
-2
-3
-4.0000E-4
-2.0000E-4
-5.4210E-20
2.0000E-4
4.0000E-4

Appendix N Simple Regression Model (Remove *APM*)

Table 49 Model Results of Simple Regression Model (Remove *APM*)

	Coef.	*Std. Errors*	*Std. Coef.*	*t-value*	*p-value*	*Tole-rance*	*VIF*
(Constant)	2.108	0.687		3.067	0.002		
Age	0.003	0.010	0.008	0.319	0.750	0.886	1.128
HoursPerWeek	-0.001	0.004	-0.009	-0.357	0.721	0.753	1.327
TotalHours	0.000	0.000	0.104	3.884	0.000	0.704	1.421
SelectByHotkeys	56.221	11.698	0.137	4.806	0.000	0.624	1.603
AssignToHotkeys	1612.214	263.995	0.193	6.107	0.000	0.507	1.973
UniqueHotkeys	-0.007	0.020	-0.010	-0.347	0.729	0.673	1.486
MinimapAttacks	921.117	251.008	0.090	3.670	0.000	0.838	1.194
MinimapRightClicks	134.944	132.699	0.026	1.017	0.310	0.767	1.304
NumberOfPACs	401.817	99.387	0.233	4.043	0.000	0.153	6.550
GapBetweenPACs	-0.007	0.003	-0.087	-2.556	0.011	0.433	2.309
ActionLatency	-0.014	0.004	-0.177	-3.369	0.001	0.182	5.482
ActionsInPAC	0.036	0.040	0.030	0.912	0.362	0.473	2.114
TotalMapExplored	0.007	0.007	0.030	0.952	0.341	0.509	1.964
WorkersMade	318.433	94.034	0.091	3.386	0.001	0.695	1.438
UniqueUnitsMade	-0.021	0.028	-0.022	-0.734	0.463	0.583	1.714
ComplexUnitsMade	-113.437	487.399	-0.007	-0.233	0.816	0.579	1.726
ComplexAbilitiesUsed	181.181	192.290	0.027	0.942	0.346	0.618	1.618
	R^2		0.657		$\overline{R^2}$		0.648
					Durbin-Watson		1.125

Figure 11 Residual Distribution of Simple Regression Model (Remove *APM*)

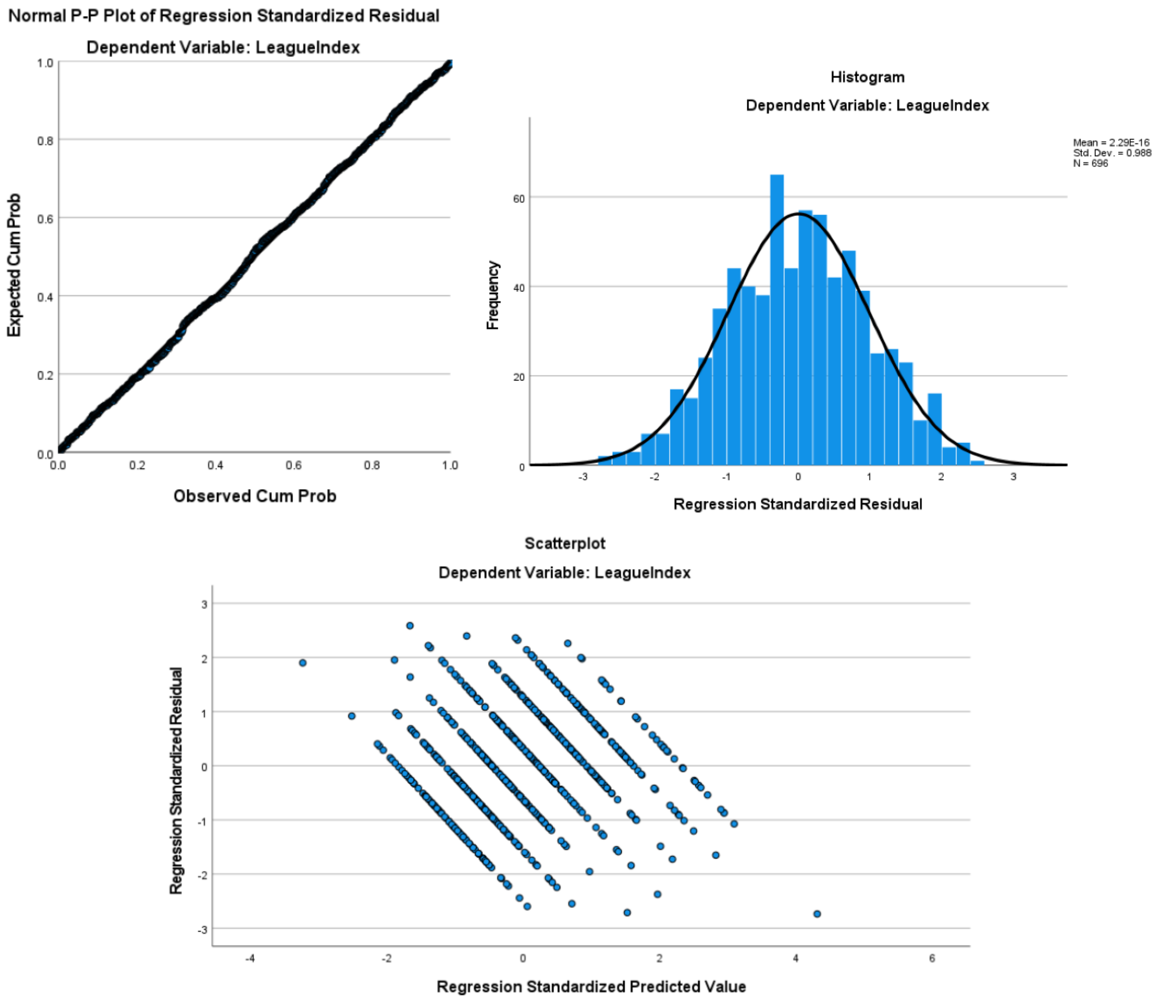
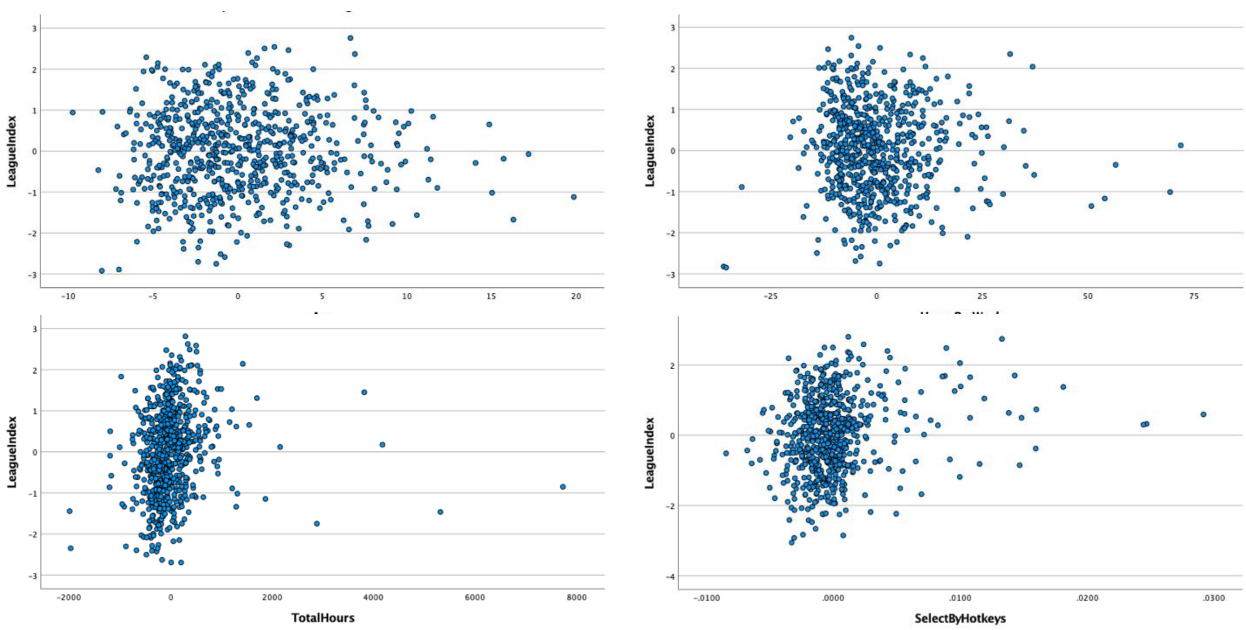
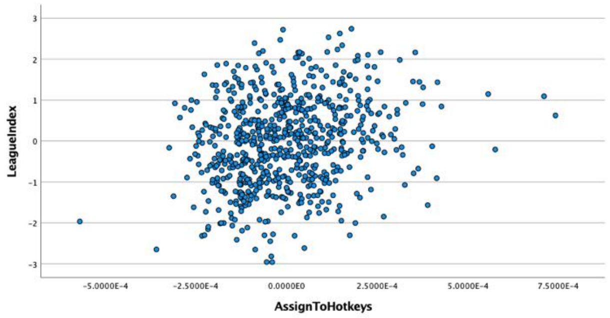
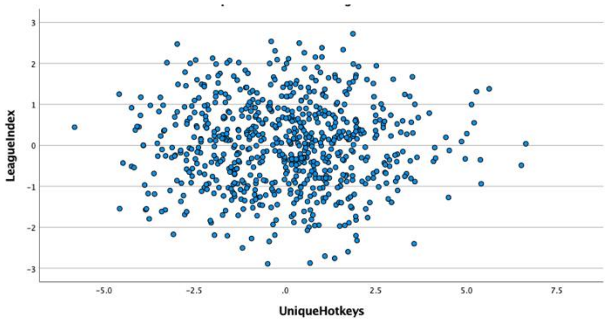
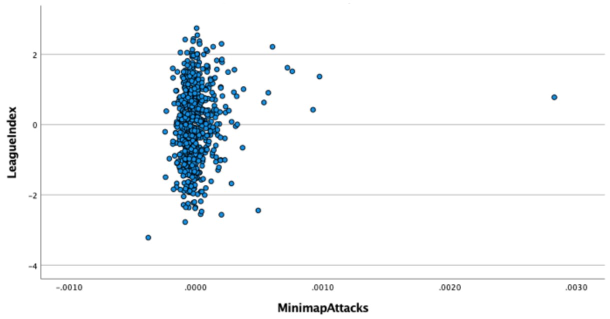


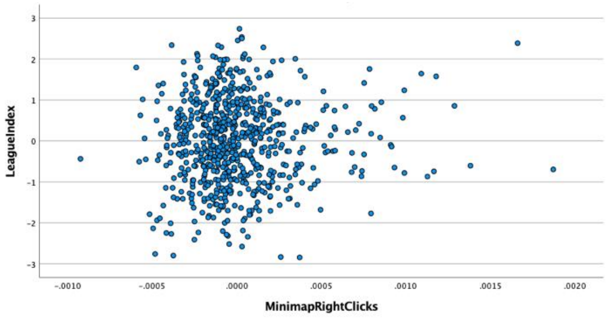
Figure 12 Partial Regression Plots of Simple Regression Model (Remove *APM*)

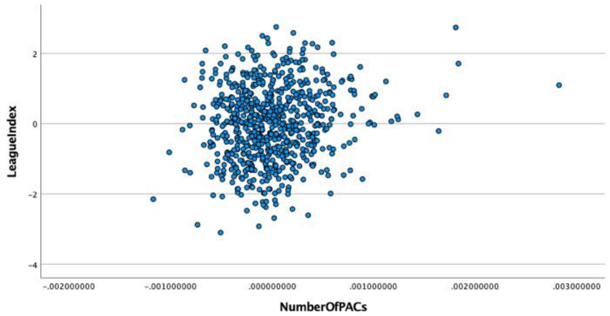


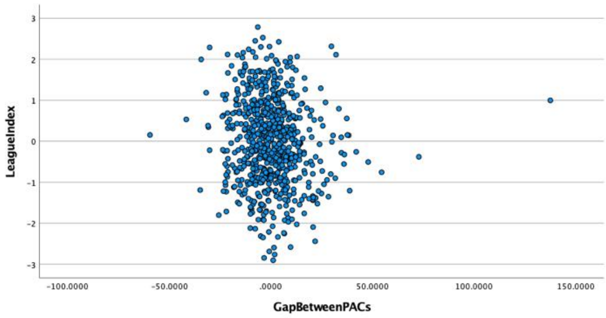

LeagueIndex
AssignToHotkeys
-5.0000E-4
-2.5000E-4
0.0000E0
2.5000E-4
5.0000E-4
7.5000E-4

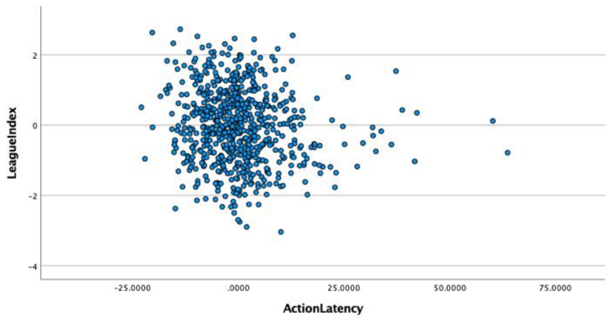

LeagueIndex
UniqueHotkeys
-5.0
-2.5
.0
2.5
5.0
7.5

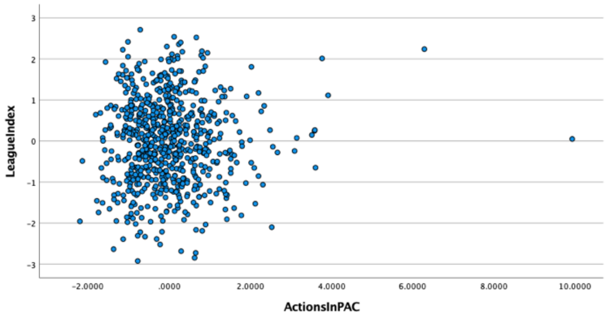

LeagueIndex
MinimapAttacks
-.0010
.0000
.0010
.0020
.0030

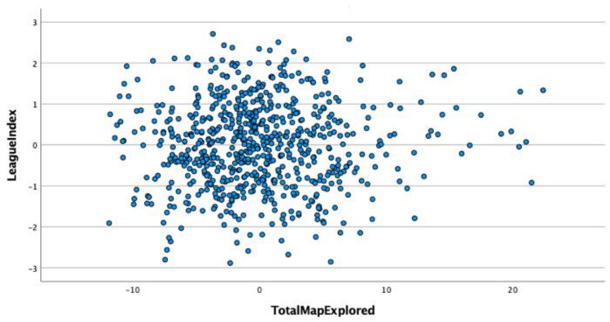

LeagueIndex
MinimapRightClicks
-.0010
-.0005
.0000
.0005
.0010
.0015
.0020

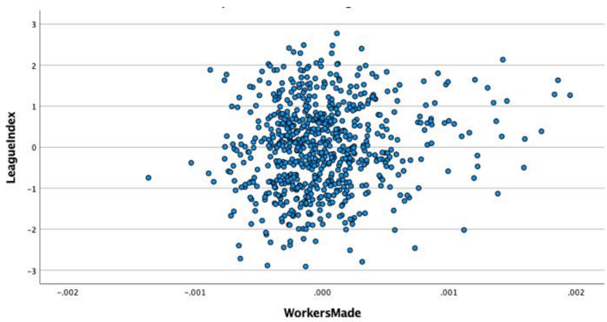

LeagueIndex
NumberOfPACs
-.002000000
-.001000000
.000000000
.001000000
.002000000
.003000000

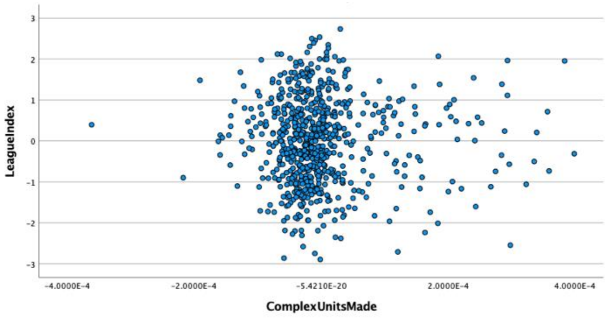

LeagueIndex
GapBetweenPACs
-100.0000
-50.0000
.0000
50.0000
100.0000
150.0000

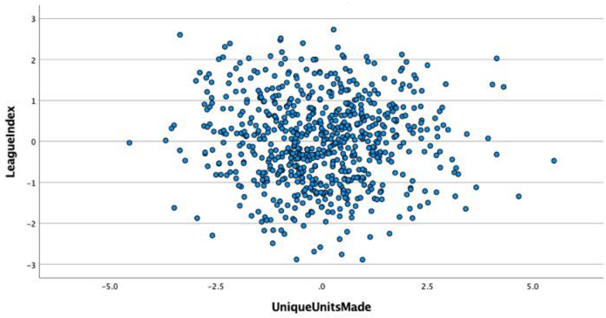

LeagueIndex
ActionLatency
-25.0000
.0000
25.0000
50.0000
75.0000

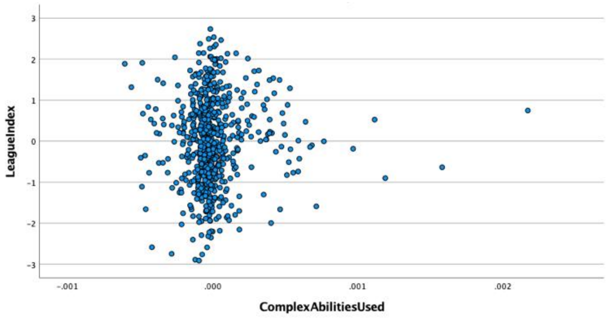

LeagueIndex
ActionsInPAC
-2.0000
.0000
2.0000
4.0000
6.0000
8.0000
10.0000


LeagueIndex
TotalMapExplored
-10
0
10
20


LeagueIndex
WorkersMade
-.002
-.001
.000
.001
.002


LeagueIndex
ComplexUnitsMade


LeagueIndex
UniqueUnitsMade

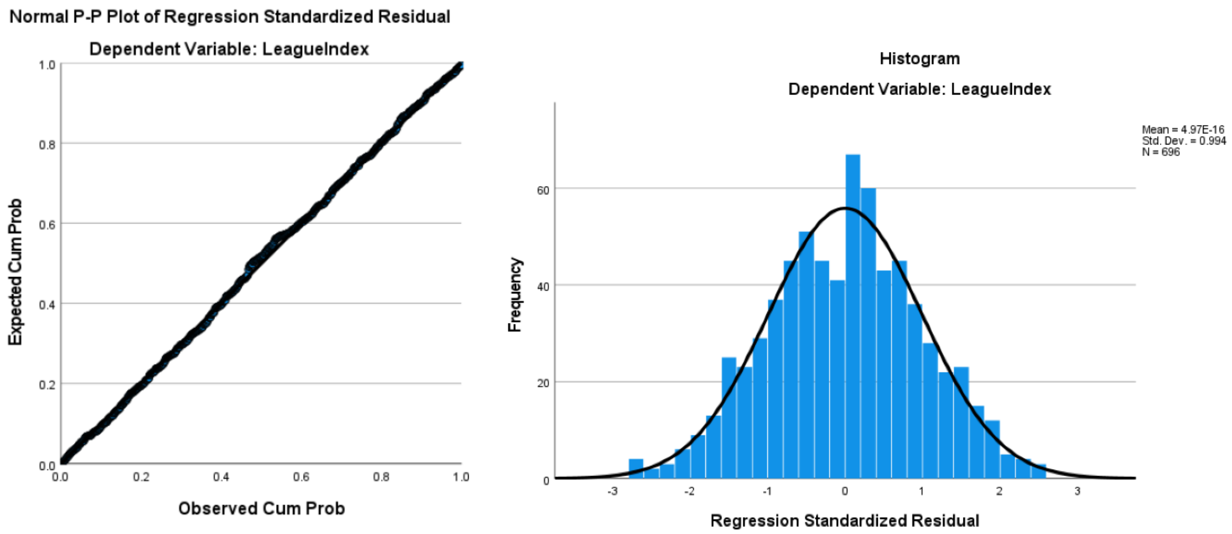

LeagueIndex
ComplexAbilitiesUsed

Appendix O Stepwise Regression Model

Table 50 Model Results of Stepwise Regression Model

	Coef.	*Std. Errors*	*Std. Coef.*	*t-value*	*p-value*	*Tolerance*	*VIF*
(Constant)	2.659	0.456		5.830	0.000		
TotalHours	0.000	0.000	0.105	4.214	0.000	0.804	1.243
SelectByHotkeys	56.312	11.126	0.137	5.061	0.000	0.686	1.458
AssignToHotkeys	1634.398	249.207	0.196	6.558	0.000	0.566	1.767
MinimapAttacks	1002.043	243.100	0.098	4.122	0.000	0.888	1.126
NumberOfPACs	357.259	75.970	0.207	4.703	0.000	0.260	3.847
GapBetweenPACs	-0.008	0.003	-0.092	-2.884	0.004	0.493	2.027
ActionLatency	-0.016	0.004	-0.204	-4.209	0.000	0.214	4.667
WorkersMade	352.221	87.056	0.101	4.046	0.000	0.807	1.239
	R^2		0.654		$\overline{R^2}$		0.650
					Durbin-Watson		1.124

Figure 13 Residual Distribution of Stepwise Regression Model



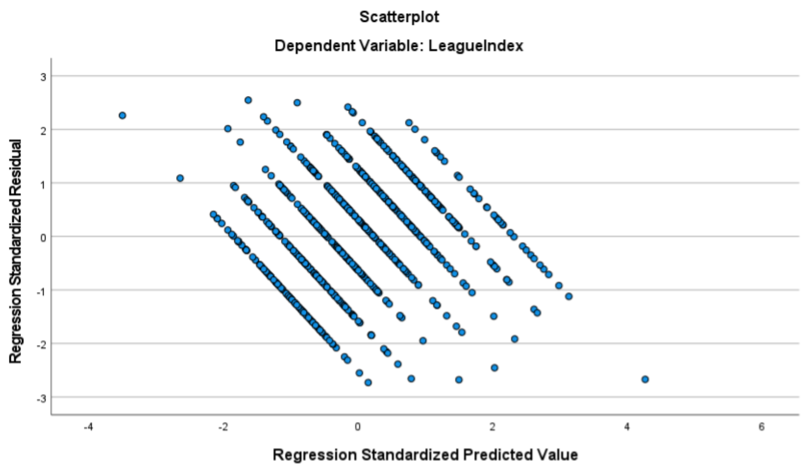
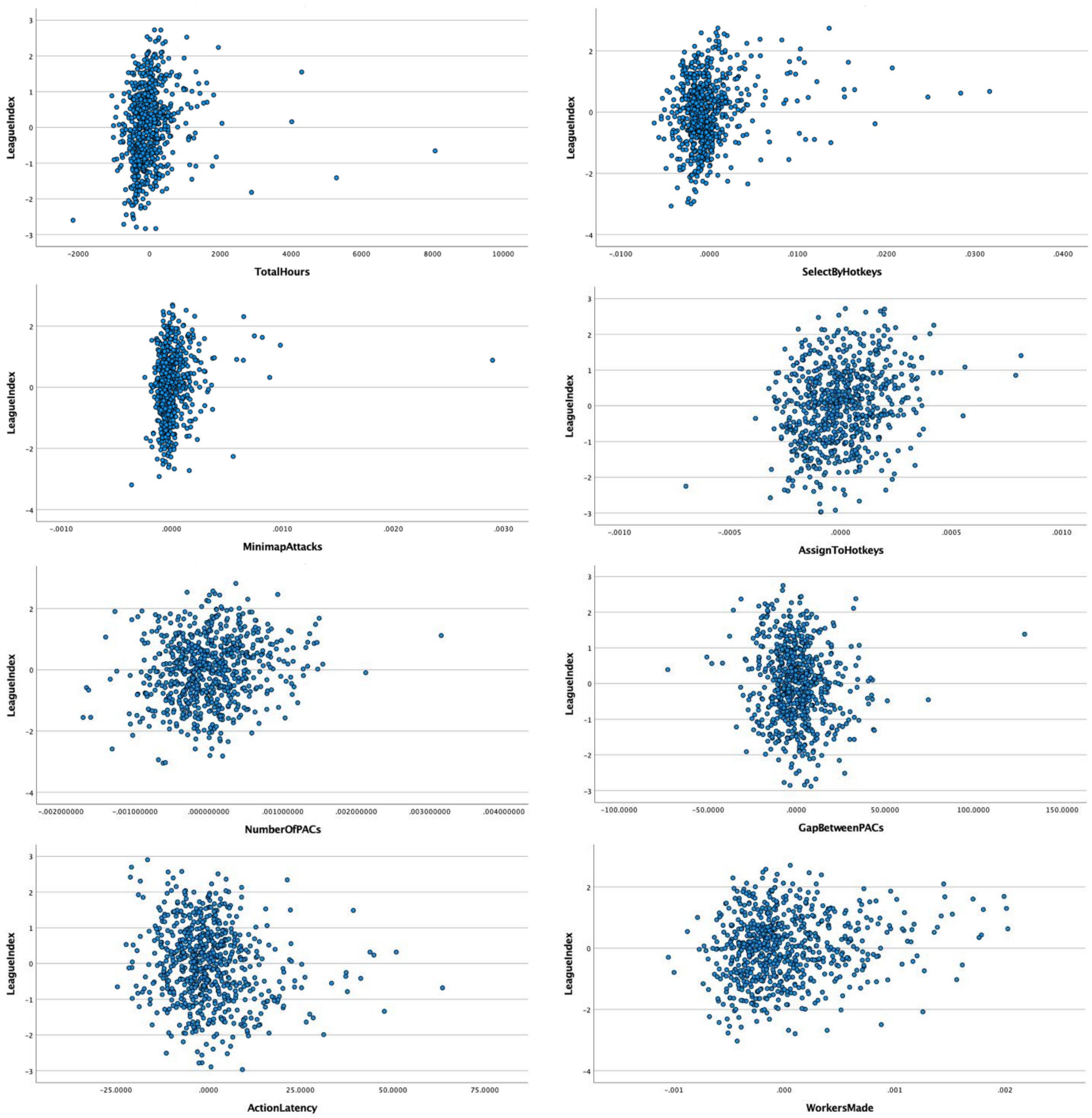


Figure 14 Partial Regression Plots of Stepwise Regression Model

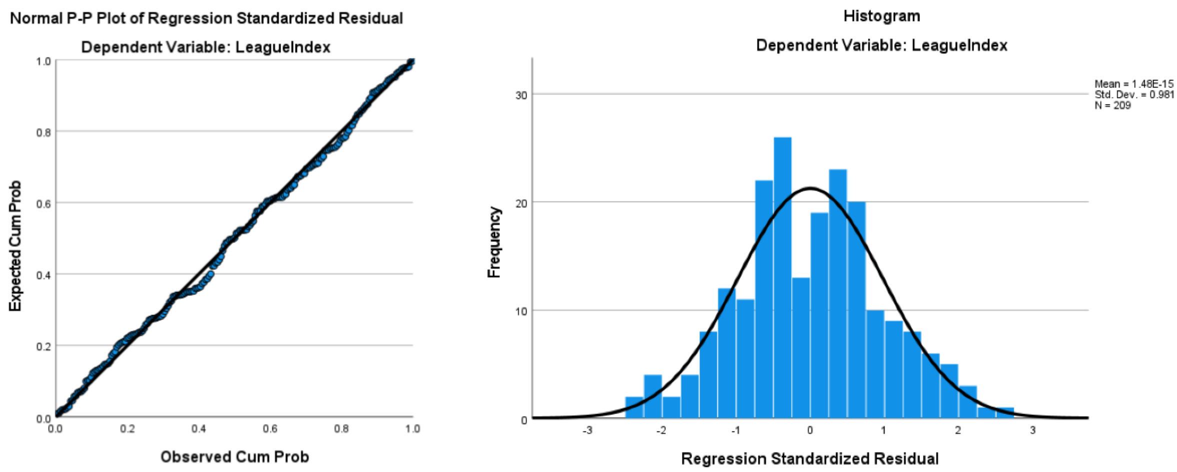


Appendix P Validation : Stepwise Regression Model

Table 51 Model Results of Stepwise Regression Model (Validation)

	Coef.	*Std. Errors*	*Std. Coef.*	*t-value*	*p-value*	*Toler-ance*	*VIF*
(Constant)	3.121	0.717		4.354	0.000		
TotalHours	0.000	0.000	0.125	2.840	0.005	0.745	1.342
SelectByHotkeys	39.487	19.960	0.098	1.978	0.049	0.588	1.700
AssignToHotkeys	1360.340	467.728	0.155	2.908	0.004	0.510	1.961
MinimapAttacks	1613.770	545.972	0.121	2.956	0.003	0.859	1.164
NumberOfPACs	350.279	119.079	0.223	2.942	0.004	0.250	4.001
GapBetweenPACs	-0.013	0.005	-0.149	-2.864	0.005	0.531	1.885
ActionLatency	-0.017	0.006	-0.227	-2.898	0.004	0.235	4.260
WorkersMade	291.441	155.286	0.079	1.877	0.062	0.808	1.238
	R^2		0.712		$\overline{R^2}$		0.700
					Durbin-Watson		1.478

Figure 15 Residual Distribution of Stepwise Regression Model (Validation)



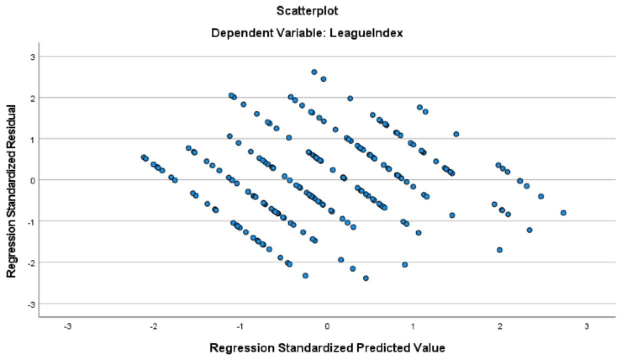
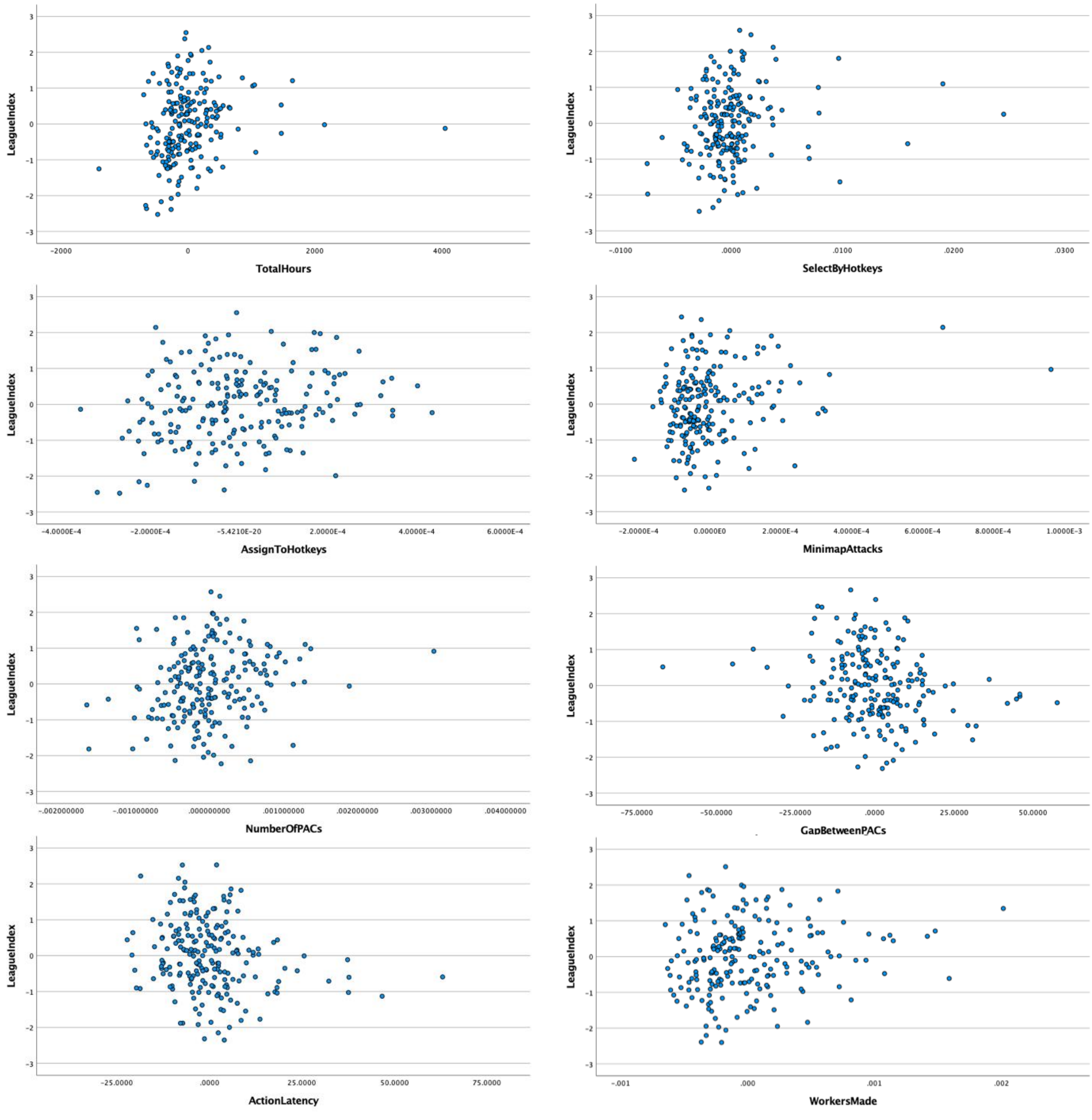


Figure 16 Partial Regression Plots of Stepwise Regression Model (Validation)



Appendix Q Regression Model Based on Factor Analysis

Table 52 Correlation of independent variables

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*
Factor 1	1.00				
Factor 2	0.00	1.00			
Factor 3	0.00	0.00	1.00		
Factor 4	0.00	0.00	0.00	1.00	
Factor 5	0.00	0.00	0.00	0.00	1.00

Table 53 Model Results of Regression Model Based On Factor Analysis

	Coef.	*Std. Errors*	*Std. Coef.*	*t-value*	*p-value*	*Tolerance*	*VIF*
(Constant)	3.661	0.041		89.307	0.000		
Factor 1	1.132	0.041	0.634	27.588	0.000	1.000	1.000
Factor 2	0.352	0.041	0.198	8.590	0.000	1.000	1.000
Factor 3	0.507	0.041	0.284	12.362	0.000	1.000	1.000
Factor 4	0.538	0.041	0.302	13.117	0.000	1.000	1.000
Factor 5	0.264	0.041	0.148	6.445	0.000	1.000	1.000
	R^2		0.635		$\overline{R^2}$		0.633
					Durbin-Watson		1.081

Figure 17 Residual Distribution of Regression Model Based On Factor Analysis

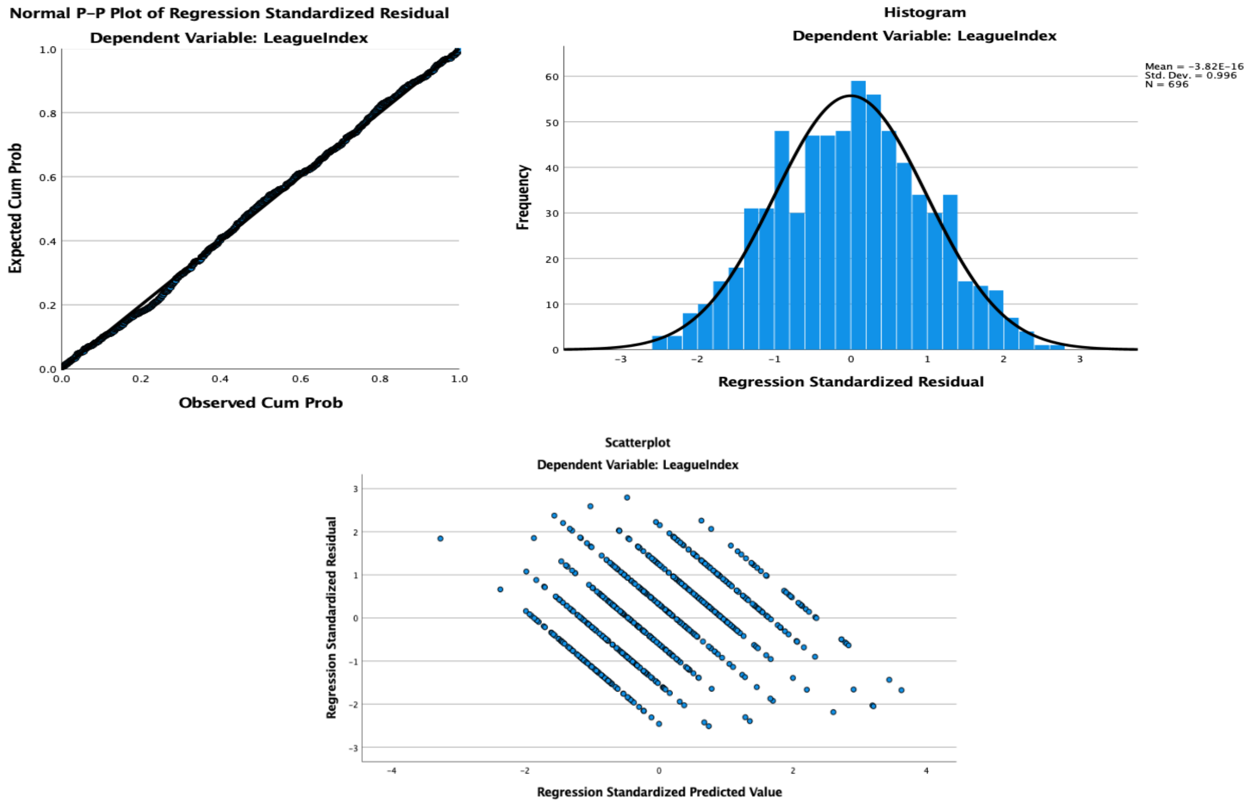
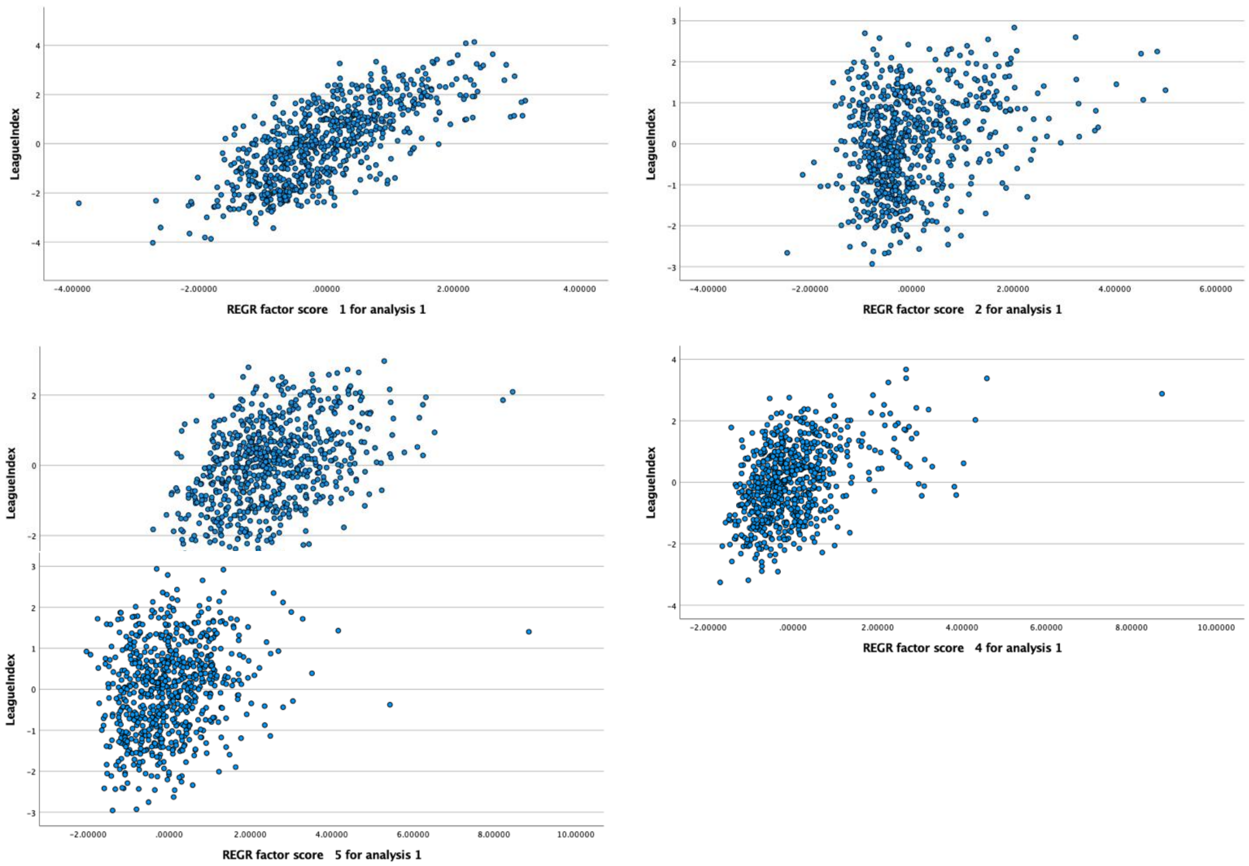


Figure 18 Partial Regression Plots of Regression Model Based On Factor Analysis



Appendix R Validation : Regression Model Based on Factor Analysis

Table 54 Correlation of Independent Variables (Validation)

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*
Factor 1	1.000				
Factor 2	0.696	1.000			
Factor 3	0.185	0.069	1.000		
Factor 4	0.198	-0.117	-0.087	1.000	
Factor 5	0.309	0.102	-0.108	-0.054	1.000

Table 55 Model Results of Regression Model Based On Factor Analysis (Validation)

	Coef.	*Std. Errors*	*Std. Coef.*	*t-value*	*p-value*	*Tolerance*	*VIF*
(Constant)	3.671	0.071		51.503	0.000		
Factor 1	1.160	0.067	0.688	17.249	0.000	0.972	1.029
Factor 2	0.371	0.076	0.196	4.909	0.000	0.974	1.026
Factor 3	0.546	0.068	0.318	7.987	0.000	0.975	1.025
Factor 4	0.566	0.077	0.293	7.311	0.000	0.962	1.040
Factor 5	0.301	0.078	0.152	3.848	0.000	0.986	1.014
	R^2		0.686		$\overline{R^2}$		0.678
					Durbin-Watson		1.289

Figure 19 Residual Distribution of Regression Model Based On Factor Analysis (Validation)

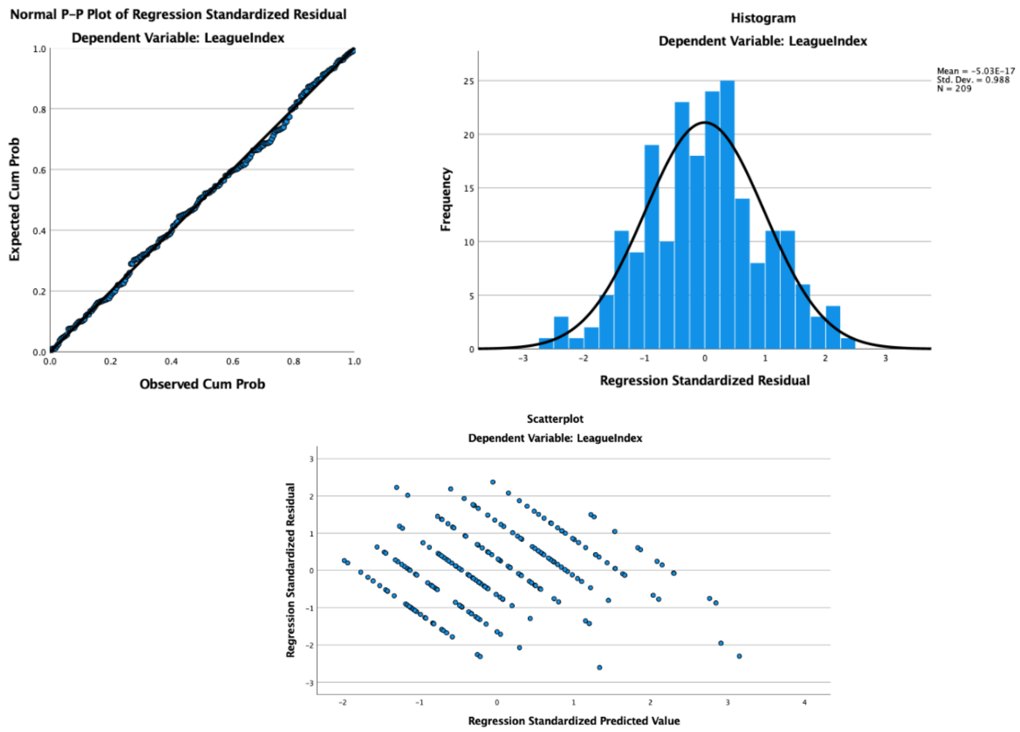
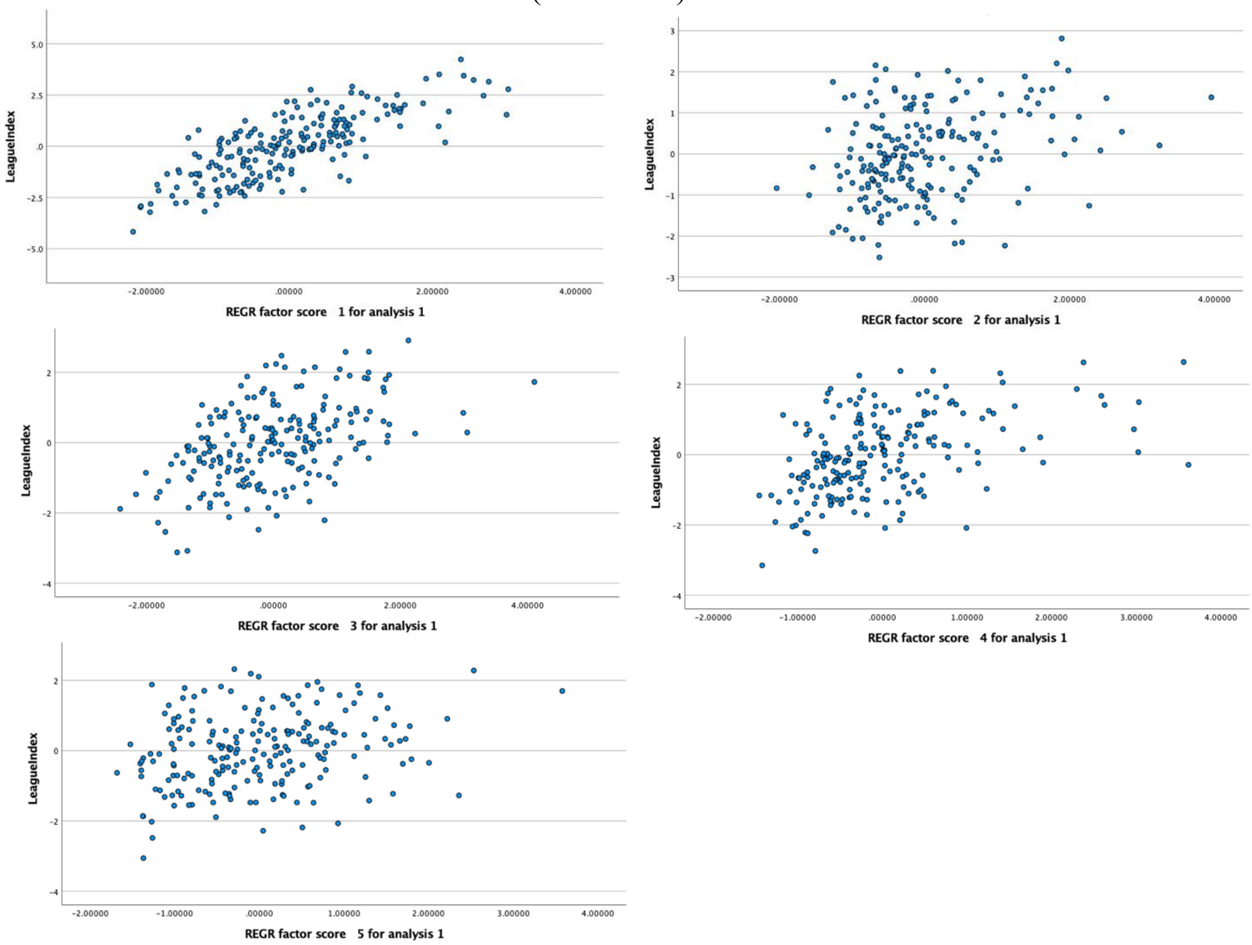


Figure 20 Partial Regression Plots of Regression Model Based On Factor Analysis (Validation)



Appendix S Comparison : Models with Outliers and without Outliers

Table 56 Comparison of Adjusted R Square

	Stepwise regression model		*Regression model based on factor analysis*	
	with possible outliers	*without possible outliers*	*with possible outliers*	*without possible outliers*
Adjusted R square	0.650	0.659	0.633	0.642
Durbin-Watson	1.124	1.138	1.081	1.102

Table 57 Comparison of Results of Validation (Stepwise Regression Models)

	with possible outliers			*without possible outliers*		
	Std. Coef.	*p-value*	*VIF*	*Std. Coef.*	*p-value*	*VIF*
(Constant)		0.000			0.000	
TotalHours	0.129	0.005	1.403	0.125	0.005	1.342
SelectByHotkeys	0.106	0.033	1.678	0.098	0.049	1.700
AssignToHotkeys	0.183	0.000	1.786	0.155	0.004	1.961
MinimapAttacks	0.107	0.012	1.201	0.121	0.003	1.164
NumberOfPACs	0.164	0.059	5.055	0.223	0.004	4.001
GapBetweenPACs	-0.125	0.017	1.848	-0.149	0.005	1.885
ActionLatency	-0.235	0.007	5.167	-0.227	0.004	4.260
WorkersMade	0.140	0.002	1.372	0.079	0.062	1.238
Adjusted R square			0.700			0.701
Durbin-Watson			1.478			1.293

Table 58 Comparison of Results of Validation (Regression Models Based on Factor Analysis)

	with possible outliers			*without possible outliers*		
	Std. Coef.	*p-value*	*VIF*	*Std. Coef.*	*p-value*	*VIF*
(Constant)		0.000			0.000	
Factor 1	0.688	0.000	1.029	0.632	0.000	1.007
Factor 2	0.196	0.000	1.026	0.206	0.000	1.018
Factor 3	0.318	0.000	1.025	0.332	0.000	1.031
Factor 4	0.293	0.000	1.040	0.334	0.000	1.005
Factor 5	0.152	0.000	1.014	0.106	0.008	1.013
Adjusted R square			0.678			0.683
Durbin-Watson			1.289			1.277