

Using interaction data in video game to analyze the relation between reflex, age, and skill level

Joseph Yan

2022/12/18

Introduction

Behaviour in video games is a relatively underresearched field of study compared to many other forms of media. However, as artificial intelligence becomes more prevalent in everyday life, video games have become one of the most effective sources for observing human behaviour. This Report aims to present a linear regression model that can explain the relationship between an individual's age, reflex and skill level in a competitive video game environment, using match data collected across 3395 players ranging from various ages, ranks, and playtime. In order to do so, we will investigate what variables are predictive for our model and how much of an impact they hold on a player's performance.

This research area is still relatively niche due to the field being a fresh cross-section of multiple disciplines. Most of them, like groups from Deepmind (Vinyals, 2019), either focus on using research data to push the boundaries of artificial intelligence or analyze the human decision system from a biological standpoint (Thompson JJ, 2013). In our study, we want to use the constructed model to help train AIs that can more closely imitate human actions. Furthermore, we also want to provide estimations for players who wish to keep track of their improvements and compare them with other players' statistics in the same skill bracket. Compared to the previous two examples of research in the field, this study seeks to explore the middle ground and the gap between AI and human interactions.

Materials and Methods

Data Collection

The Data set used in this study was collected by SFU Cognitive Science Lab and archived by UCI Machine Learning Repository. A few attributes of the data to be noted include the following:

- LeagueIndex: Bronze, Silver, Gold, Platinum, Diamond, Master, GrandMaster(1-7 respectively), indicates expertise of the game.
- Screen movements are aggregated into screen fixations using a Salvucci & Goldberg dispersion-threshold algorithm and defined Perception-Action Cycles (PACs) as fixations with at least one action.
- Perception-Action Cycles are defined as a period of time when players are fixating and acting at a particular location.
- Time is recorded in terms of timestamps in the StarCraft 2 replay file. When the game is played on 'faster', one real-time second equals roughly 88.5 timestamps.
- Hotkeys are keyboard inputs that can replace clicking or combinations of interaction with the screen using the mouse, thus increasing potential actions a player can execute in a given timeframe.

- Action Per Minute(APM) is a measurement of cognitive-motor speed and is often used as a predictor of expertise in the game’s community. (Thompson, 2013)

Other than that, as per the original study’s data collection process, all games are played under the same client version and thus are within a controlled timeframe. Each match data is entirely independent of another. We have cleaned undetermined and missing values and removed some variables from the original data set that do not apply to the study to better fit our goal’s needs in constructing the linear model. Such entries mainly include professional players whose data are vastly different from regular players causing problematic evaluations.

Methods The main question to discuss is the predictive importance of variables to the linear model. To answer this, we first evenly split the data set into training and testing sets. Each included 1668 observations and is compared to ensure the two sets have similar characteristics.[Table 1] Afterwards, we built a linear regression model with all the variables from our cleaned data set.

Table 1: Summary statistics of each variable in the training and test datasets, each of size 1668. All values are summarized in means(s.d). From the table we can see that the two data sets hold similar values.

Variable	Training Set	Test Set
HrsPerWeek	15.640072 (11.2954442)	16.0912365 (12.0474072)
TotalHours	631.815237 (556.2055545)	653.2965186 (630.5991562)
APM	114.1456193 (46.5028393)	114.6525314 (49.2744047)
UniqueHotkeys	4.3233353 (2.2926264)	4.3037215 (2.3692622)
NumberOfPACs	0.0034289 (9.6879382×10^{-4})	0.0034346 (9.6132479×10^{-4})
GapBetweenPACs(ms)	40.5035899 (16.9786379)	40.9494682 (17.1445489)
ActionLatency(ms)	64.0680259 (19.0257941)	64.4109295 (19.0420561)
ActionsInPAC	5.3086187 (1.4780346)	5.2247136 (1.5238703)
UniqueUnitsMade	6.4775045 (1.835672)	6.6044418 (1.8808865)

Methods

The main question to discuss is the predictive importance of variables to the linear model. To answer this, we first evenly split the data set into training and testing sets. Each included 1668 observations and is compared to ensure the two sets have similar characteristics. Afterwards, we built a linear regression model with all the variables from our cleaned data set.

We then proposed our hypothesis that the regression model containing at least one predictor useful in predicting our response variable. That is:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$$

$$H_A : \text{At least one } \beta_j \neq 0 \text{ (for } j = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\text{)}$$

After Hypothesis testing, we found a significant linear relationship between the response and at least one of the predictors present. Thus we reject the null hypothesis.[Figure 1] Then we conducted an ANOVA test to confirm which variables should be dropped. We decided to drop two variables, ActionsInPAC and UniqueUnitsMade, then used the remaining variables to build a reduced model. After an exploratory analysis, we discovered that most variables do not violate assumptions and linearity holds.[] However, a few variables have a heavy right skew in Age and hours-related variables.

This is expected as a limitation in data collection. We used log transformation to correct the skewness of Age. As for the two-hour variables, we chose square root transformation to avoid any values of 0 becoming negative infinity under log transformation.

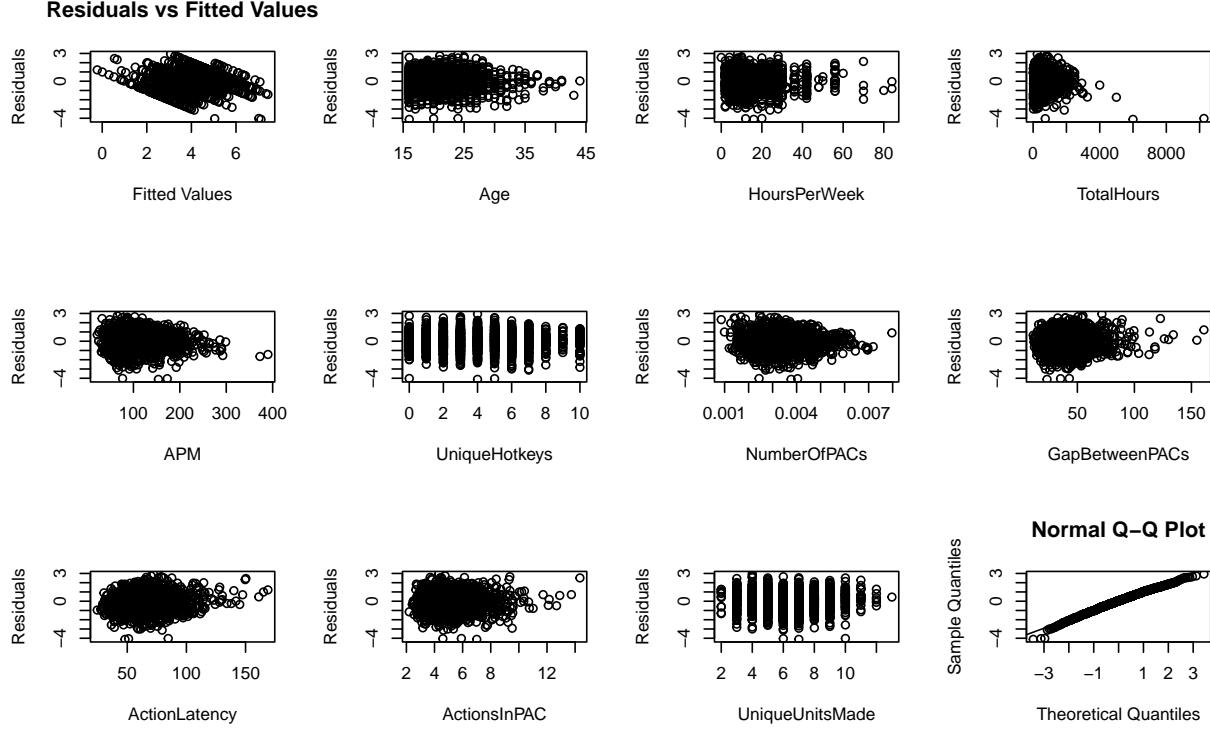


Figure 1: Residual plots for assessing assumptions of the original model

Besides the two variables, our plot shows that the linearity, independence of errors, homoscedasticity, and normality of error distribution assumptions and conditions are all within an acceptable range.

There are a few leverage points but no outstanding outliers in the variables that require transformation. After the transformation, no observations were identified as being influential on the entire regression surface, according to Cook's distance. However, 84 points are identified as influential according to DFFITS, but only 2 by DFBETAS. In our plots, we can observe only a few of them, making them likely only to be influential on their respective predictor variable. There are no entries that are overly common through beta comparisons. Lastly, the QQ plots before and after the transformation look linear without any significant upward or downward trailing on the two edges[Figure 2]. This linearity is partly due to this data set's large observation count.

Finally, we assessed the three models with AIC and BIC and made a table to showcase their change as we operated on the models. Our Final model ended up with 0.5402 adjusted R squared, meaning it explains approximately 54% of the original variation in LeagueIndex, increasing from 52%. The AIC and BIC scores are -56.3692 and 1.8186, respectively.[Table 2]

Table 2: The Adjusted R^2 , AIC and BIC score of the three models in our study.

Model	Adjusted R^2	AIC	BIC
Original Model	0.520196	16.9423615	85.9677321
Reduced Model	0.5204027	14.2362875	72.4240963
Final Model(Transformed)	0.5402918	-56.3692377	1.8185712

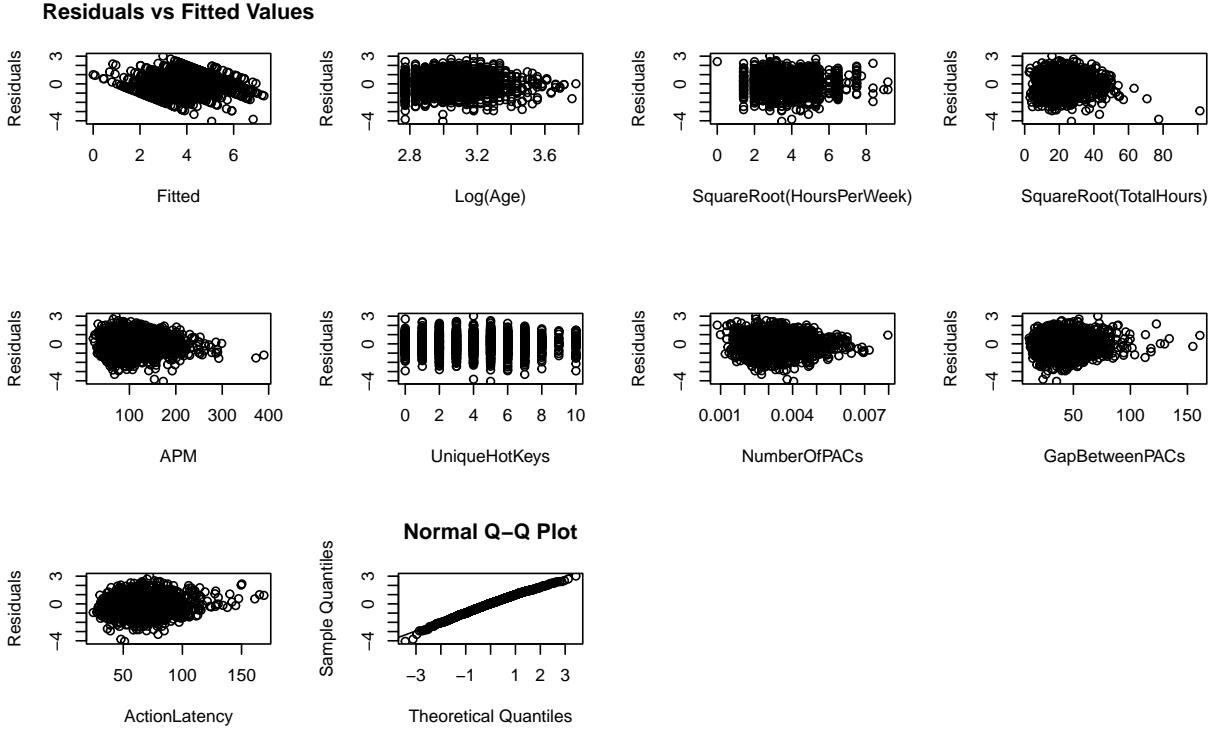


Figure 2: Residual plots for assessing assumptions of the transformed model

Results

With the final model we have decided on, we ended up with variables of the following: Log(Age), SquareRoot(HoursPerWeek), SquareRoot(TotalHours), APM, UniqueHotkeys, NumberOfPACs, GapBetweenPACs, ActionLatency, and predicting for the response variable LeagueIndex, this is reduced from the previous ten predictors and transforming some of the predictors that are have either distribution problem or violate assumptions. Furthermore, in comparing the two split data sets, we found out that the test data set has a lower mean VIF value. This lower score means that the multicollinearity between the predictor variables in that data set is more correlated than the train data set we initially used to construct our model.[Table 3]

Table 3: Summary of characteristics of the final transformed model in the training and test datasets. The model uses Log(Age), SquareRoot(HoursPerWeek), SquareRoot(TotalHours), APM, UniqueHotkeys, NumberOfPACs, GapBetweenPACs, ActionLatency. Response is LeagueIndex. Coefficients are presented as estimate \pm SE (* = significant t-test at $\alpha = 0.05$).

Characteristic	TransformedModel (Train)	TranformedModel (Test)
Mean VIF value	2.2281209	2.171489
# Cook's Distance	0	0
# DFBETAS	28	32
Violations	none	none
Intercept	3.482 ± 0.526 (*)	2.473 ± 0.517 (*)

Characteristic	TransformedModel (Train)	TranformedModel (Test)
Log(Age)	169.161 ± 44.681 (*)	175.001 ± 44.255 (*)
SquareRoot(HoursPerWeek)	-0.01 ± 0.002 (*)	-0.011 ± 0.002 (*)
SquareRoot(TotalHours)	-0.02 ± 0.003 (*)	-0.016 ± 0.003 (*)
APM	0.058 ± 0.137 (*)	0.239 ± 0.138 (*)
UniqueHotKeys	-0.053 ± 0.021 (*)	-0.002 ± 0.02 (*)
NumberOfPACs	0.037 ± 0.003 (*)	0.037 ± 0.003 (*)
GapBetweenPACs	0.006 ± 0.001 (*)	0.006 ± 0.001 (*)
ActionLatency	0.041 ± 0.011 (*)	0.063 ± 0.011 (*)

Discussion

This study, like many others, has several limitations, the first being data collection. Although match records and server data can help validate the accuracy of metadata. Self-reported variables like hours played and age. We are unsure whether the players reported these variables honestly, which is one of the big reasons we had to go through the transformation for said variables in constructing the model. On the other hand, a major hidden confounding factor is that we do not know the gender of the players. Thus, we can not predict whether or not it affects the model's accuracy. The other limitation is the range of prediction this model can provide. In the data collection section, we mentioned that we removed professional-level players' data from the data set due to unrecorded play time and their almost inhuman reflex data in other variables. This is a double-edged sword as it will help with the overall accuracy of the model but limits the ability to predict talented players who have the potential to become professional players, as there is no standard for the model to predict. The following limitation comes with the identified leverage points, as mentioned previously, due to how the data is collected. Some leverage points in variables such as APM are actual data, as they are recorded by computer programs based on interactions with the game. Thus we are not supposed to remove them from the data set as they are not error data.

In light of this study, we want to draw more attention to video games as a field of research in statistics. In today's world, the amount of data generated through computer interactions is uncountable, and video games make up a large chunk of this data. With the variety of types of games and the degree of realism they can simulate. Researchers can find the data they need without needing to simulate them. These data may produce more interesting results compared to real-world counterparts. They can be easily recorded and leveraged to analyze them into trends to help us improve in fields such as cognitive science, economy, and artificial intelligence.

References

1. Vinyals, O., Babuschkin, I., Czarnecki, W.M. et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* 575, 350–354 (2019). <https://doi.org/10.1038/s41586-019-1724-z>
2. Thompson JJ, Blair MR, Chen L, Henrey AJ. Video Game Telemetry as a Critical Tool in the Study of Complex Skill Learning. *PLoS ONE* 8(9): e75129 (2013). <https://doi.org/10.1371/journal.pone.0075129>
3. UCI Machine Learning Repository : <https://archive.ics.uci.edu/ml/index.php>

Appendix

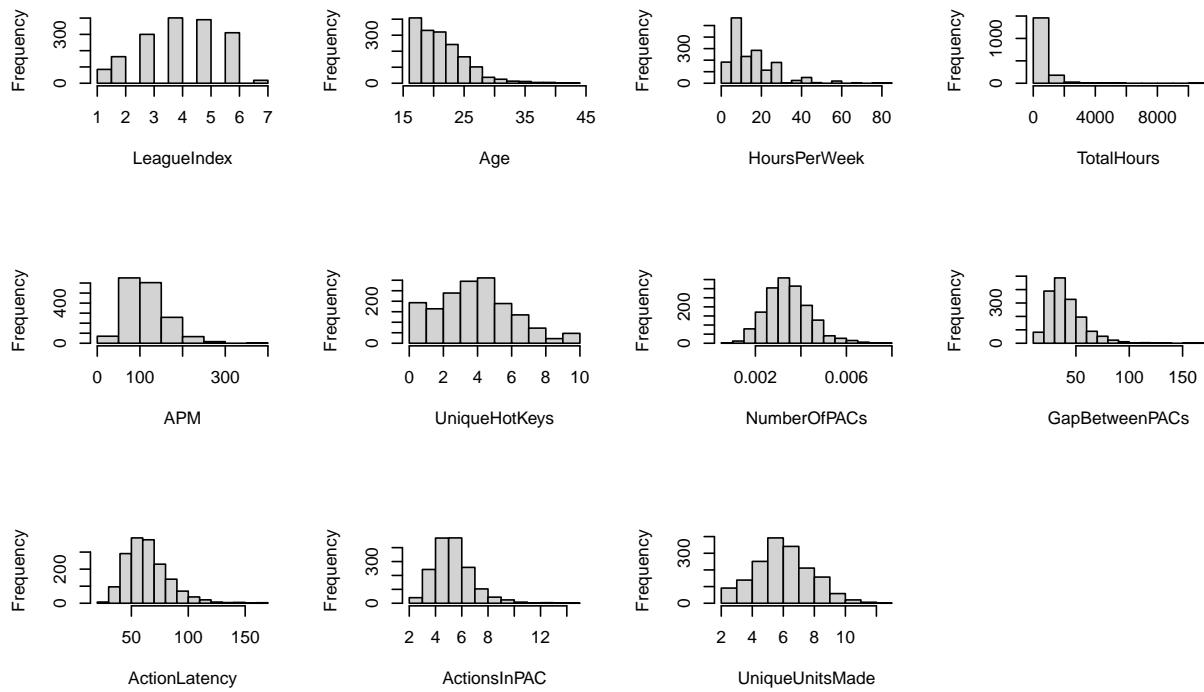


Figure 3: Histograms for assessing distribution of the variables

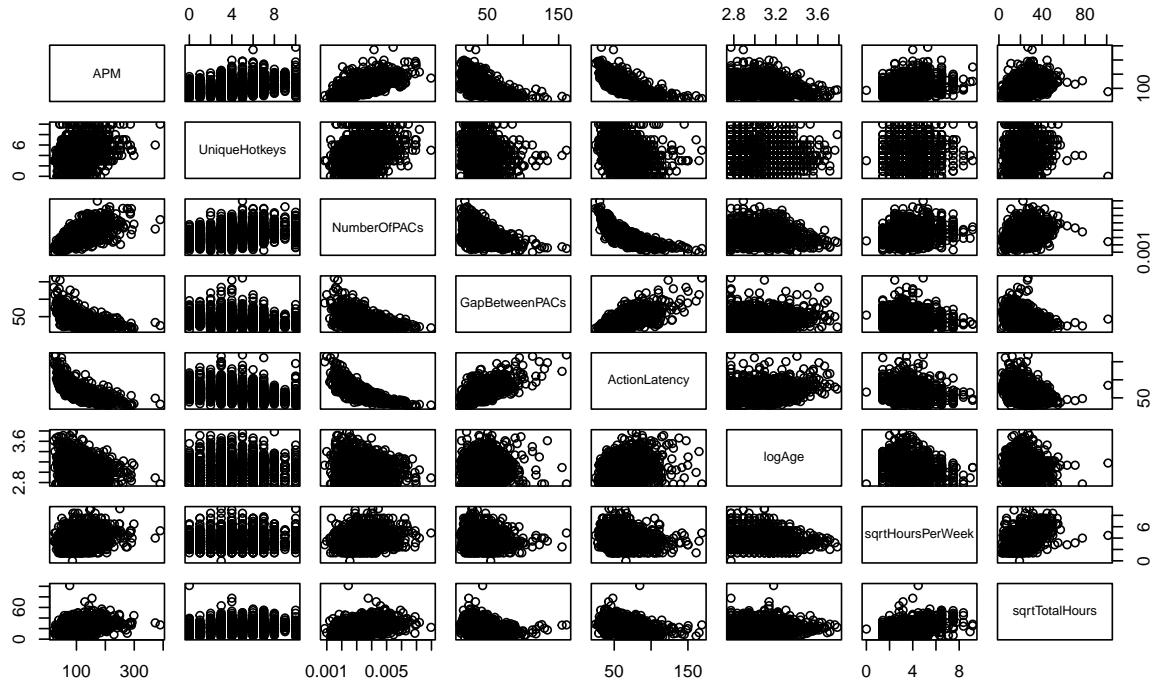


Figure 4: Histograms for assessing correlation of the transformed variables