Evan Nguyen
Professor Kevin Bailey
STA 2260: Prob and Stats for CS Egr
5 May 2025

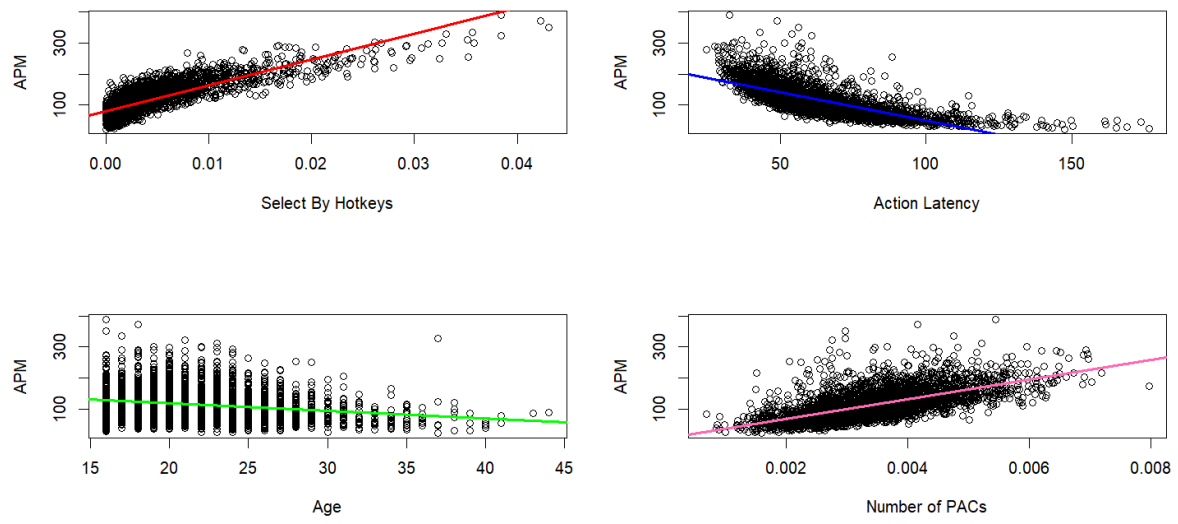
Linear Regression Modeling *StarCraft II* Player Performance Using APM and Hotkey Usage

Introduction

Released by Blizzard Entertainment in 2010 as a sequel to its predecessor, *StarCraft II* quickly became one of the most recognizable and competitive real-time strategy games to date. Players compete by gathering resources, constructing buildings, managing troops, and planning for combat, all in real time. Large amounts of multitasking and quick thinking are needed for competition. As a result, a player's APM, or Actions Per Minute, is widely used as a metric to measure player performance. This project uses linear regression models to predict APM based on a player's in-game behaviors, specifically hotkey usage and reaction speed. From our analysis, we can identify what can influence a player's in-game performance.

Data Summary

The dataset consists of 3,340 players with player details such as age, how many hours of gameplay they have, ranking, and their measured in-game metrics. At first, Total Hours of Gameplay and League Index (Player's Competitive Ladder Ranking) were considered as response variables but were not used as they were both discrete variables with no strong relationships. Instead, APM was chosen as the response because it was continuous and still reflects player performances. The considered predictors for APM as a response were: SelectByHotkeys (number of unit or building selection made using hotkeys per timestamp), ActionLatency (average latency from the onset of a Perception-Action Cycle (PAC) to their first action in milliseconds), Age, and NumberOfPACs. Using AIC and R^2, we can find which predictor has the best relationship with their response.

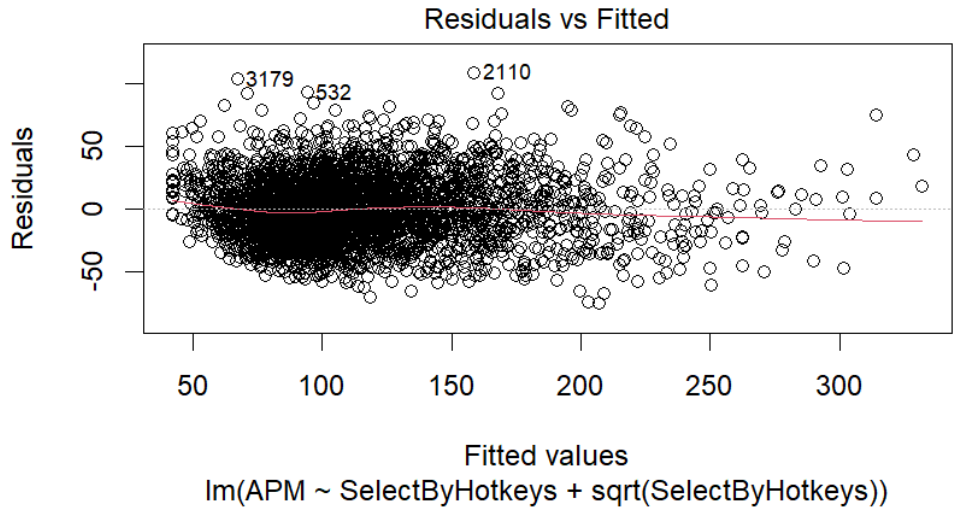


Here, we can see trends of the plots with their predictors and response with a line of best fit (abline()) mapped onto their respective functions. We can see that SelectByHotkeys has a clear positive nonlinear relationship. ActionLatency has a negative trend with APM. Age and NumberOfPACs showed some trends of their own but seemed like weaker predictors due to the nature of their scatter plots. Although not perfect, using the line of best fit, we can infer that either SelectByHotkeys or ActionLatency would be the best to model their relationship with APM.

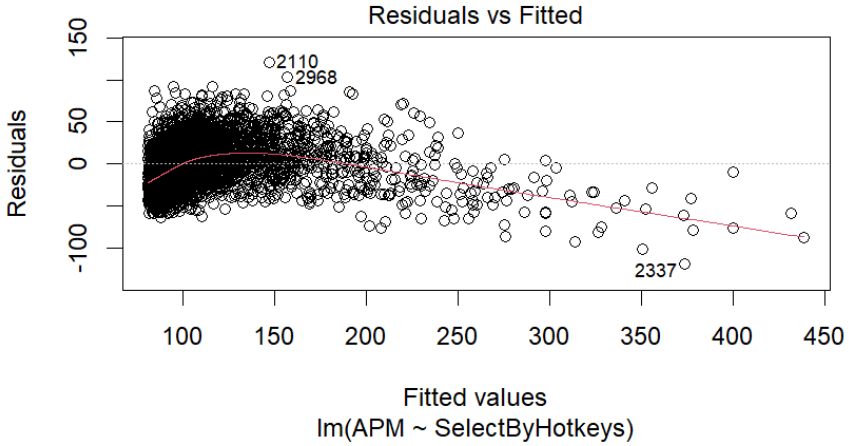
Models	AIC	R^2
m1.SelectByHotkeys	31,700.80	0.6635
m1.ActionLatency	32,876.04	0.5215
m1.Age	35,185.90	0.0444
m1.NumberOfPACs	33,612.63	0.434

The listed AIC and R^2 values of each model confirms our previous suspicion that SelectByHotkeys and ActionLatency were better models with SelectByHotkeys being the best as it has the lowest AIC and highest R^2 values. Therefore, our Model 1, m1, will be m1.SelectByHotkeys. However, a linear relation wouldn't be the best way to model m1 due to its

slight curvature. To improve upon this, a square root term can be added to account for the plateauing returns of hotkey usage.

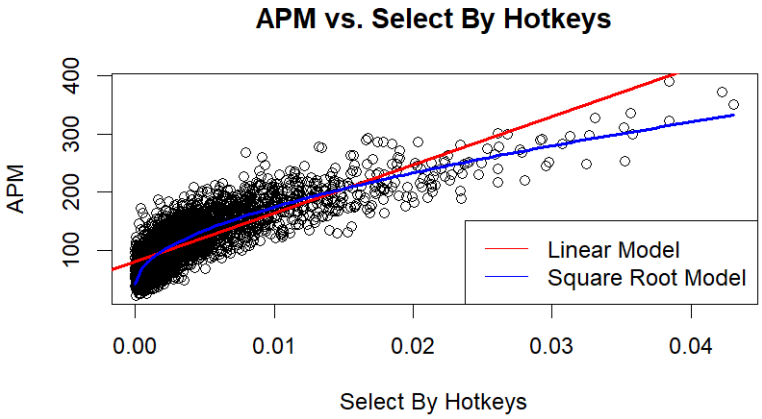


From our plot of Residuals vs. Fitted, the red line is relatively flat throughout, indicating that using a square root term captures the total structure of our data very well. The data doesn't funnel and only has a few outliers that can be seen. This supported the notion that our transformation of SelectByHotkeys was effective in improving the model's fit.



Looking at the residual plot for the linear model, we can see that there is a curve in the red line with the data funneling out, indicating that modeling our plot linearly is suboptimal.

Model 1



Model 1 Equation: $\hat{y} = 42.212 + 724.671(SelectByHotkeys) + 1245.501 * \sqrt{SelectByHotkeys}$

AIC: 31,044.97

R^2: 0.7236

From our new model, we can conclude that players who use hotkeys more often will tend to have a higher APM. The nonlinear factor of the square root more accurately captures the initial increase in APM that later levels off.

To improve our prediction, we can use a multiple regression model. This approach allows us to take a deeper look at the effects of both SelectByHotkeys and ActionLatency on players' APM. Similarly, ActionLatency also had a nonlinear relationship with APM. We can use a square root equation to model, similar to that of Model 1. Using more predictors and a better equation will provide more accuracy, yielding better AIC and R^2 values.

Models	AIC	R^2
m2.SelectByHotkey.ActionLatency	28,856.18	0.8565
m2.SelectByHotkey.ActionLatency.sqrt	28,004.20	0.8889

From our findings, we can see that implementing a square root function in our multiple regression model will lower AIC and increase the R^2. These better results will be our Model 2.

Model 2 Equation:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	435.0158	10.0490	43.289	<2e-16	***
SelectByHotkeys	4335.8713	195.4554	22.183	<2e-16	***
sqrt(SelectByHotkeys)	287.2036	32.4374	8.854	<2e-16	***
ActionLatency	2.8458	0.1378	20.658	<2e-16	***
sqrt(ActionLatency)	-67.6566	2.3304	-29.032	<2e-16	***

Model Comparisons

	AIC	R^2
Model 1	31,044.97	0.7236
Model 2	28,004.20	0.8889

Model 2 outperformed Model 1 on the basis of AIC and R^2 suggesting that a better prediction for what affects a player's APM is the usage of hotkeys as well as their response time.

Conclusion

This project explored using linear regression to take a closer look at player performance in *StarCraft II*. The data gathered from this research found that hotkey usage and the speed at which players react will influence their APM, and that using a nonlinear equation will improve the fit for each of the models. These metrics can be used to assess or train players based upon their in-game habits.