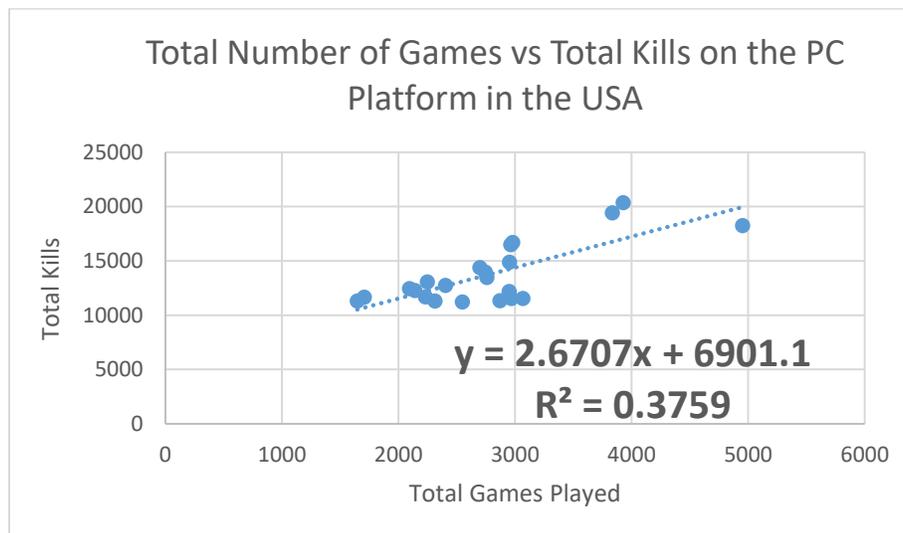


For my linear regression project, I studied the effect of the total number of Fortnite games played on the PC Platform in the USA and the number of total kills. My question of interest was “will the total number of kills increase with the more games that are played for PC players in the USA?” I decided that these two variables had an explanatory and response relationship since I believe that the total number of kills is responding/depending on the number of games played. The explanatory variable is the total number of games played and the response variable is the total number of kills. I found my data using the Fortnite leaderboard in March 2018 where it records all players’ performances on all platforms in all available countries. I decided to narrow my search to only view players on the PC Platform in the USA. Before performing the linear regression, my prediction is that the more games you have played, your total kills should increase.

Below is a scatterplot of the data. Displayed on the scatterplot is the line of best fit and its corresponding r-squared value.



My scatterplot shows a moderate positive correlation with $r = 0.61$. This means that as the total number of games played increases, the total number of kills also increases. The

regression equation for this data is the following: $y = 2.6707x + 6901.1$. The slope of the line is $a = 2.6707$ which means that as your total number of games increases by one, the total number of kills increases by 2.6707 kills or approximately 3 kills since you cannot kill 0.6707 of a player. My y-intercept for this data is $b = 6901.1$. This means if the number of games was equal to zero, then a player would start their total number of kills at 6901.1. This does not make sense in this context because a person cannot start with 6901.1 total kills if they have never played any games in Fortnite.

Using the line of best fit, I can predict the total number of kills based on the total number of games that are played. The last page of this report has a table of predicted values with their residuals. The residuals in this context is the difference between the actual total number of kills and the amount of kills the regression line predicted any player on the PC Platform in the USA would make. If the residual value is negative, then that means the total number of kills by a player is less than the line predicted. If the value is positive, then that means the total number of kills by a player is more than the line predicted. Most of the residuals are between -3500 to 3000. There are very few values that are above or below this range. There were four players that do not fall in this range so I would consider them outliers. Three players had residuals of 6015, 7842, and 4465 which means they killed many more than the line predicted and the other player had a residual of -5846 which means they killed way less than the line predicted.

The r-squared value displayed on my scatterplot is 0.3759. This means that 37.59% of the differences in the total number of kills can be described by my regression line. This is a low percentage which tells us that 62.41% of the total number of kills can be described by lurking variables. Just a few of these lurking variables include if you decided to “camp” in the game, then you will not have as many kills even if you play many games, you could not be very good

and be killed early on even if you play many games, and each game is a different scenario which means you could kill seven people in one game, but then kill one person in the next game.

I do not think that the regression line is a good fit for my data since the r-squared value is low and the percentage of lurking variables is very high. Another indicator that this is not a good fit is that the correlation for these two variables is only moderate. I learned that typically the more games you play, then the total number of kills increase. I was not very surprised by these results since it was close to my prediction. My answer to my question of interest is yes, the total number of kills increase with the more games that are played for PC players in the USA on a moderate level relationship.

X – Total number of games played	Actual Y – Total number of kills	Predicted Y	Residual
4038	23701	17685.39	6015.613
2905	22502	14659.48	7842.517
3588	20949	16483.57	4465.428
3927	20362	17388.94	2973.061
3835	19413	17143.23	2269.766
4954	18239	20131.75	-1892.75
2980	16711	14859.79	1851.214
2963	16491	14814.38	1676.616
5240	15049	20895.57	-5846.57
2952	14871	14785.01	85.9936
2700	14383	14111.99	271.01
2744	13971	14229.5	-258.501
2759	13489	14269.56	-780.561
2248	13064	12904.83	159.1664
2404	12743	13321.46	-578.463
2094	12452	12493.55	-41.5458
2141	12282	12619.07	-337.069
2951	12192	14782.34	-2590.34
2232	11698	12862.1	-1164.1
1708	11658	11462.66	195.3444
2968	11552	14827.74	-3275.74
3067	11535	15092.14	-3557.14
2870	11339	14566.01	-3227.01
2314	11298	13081.1	-1783.1
1646	11298	11297.07	0.9278