Kyle Smith

AP Statistics-5

Mrs. Caso

November 22, 2021

Introduction

As a long time Pokemon fan and absolute nerd when it comes to numbers the stats of pokemon were always a topic of interest for me. Throughout all eight generations and over one thousand pokemon to choose from I was always one to try and find the pokemon with the highest stat possible. Between all of the stats, I was always drawn more towards the basics, attack and defense. I saw those as the most vital to the pokemon, the most make or break per say. The purpose of this study is to uncover any correlation between the attack and defensive stat of all the pokemon. For this study instead of taking a sample of just one generation, I decided to go big and use the entire population of all current pokemon up until the current date of November 21st, 2021. The data provides all of the pokemons names, types, forms and their stats in Attack, Defense, HP, Special Attack, Special Defense and lastly speed but we will only be comparing the Attack and Defensive Stats. All of the data was collected by **https://www.kaggle.com/hamdallak/the-world-of-pokemons** but all of the information was created by the Pokemon Company.

Bias and Weakness

There are not many biases or weaknesses found in the data collection for many reasons which I will present. The first reason being that the study is done on every single Pokemon, it is not picking and choosing between generations, it simply uses every possible pokemon. The bias is also minimized because all of the pokemon are made by the same company and team meaning that it is only one group creating the stats of these given pokemon. In conclusion, the study was made strong and accurate by deciding not to leave any data points out of the sample and because all the data is derived from the same people making the stats of the pokemon.

Results

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	45.01	2.31	19.47	0.000	
Defense	0.4749	0.0286	16.62	0.000	1.00

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	229754	229754	276.36	0.000
Defense	1	229754	229754	276.36	0.000
Error	1043	867120	831		
Lack-of-Fit	112	145024	1295	1.67	0.000
Pure Error	931	722097	776		
Total	1044	1096874			

The regression equation is
Attack = 45.01 + 0.4749 Defense

Model Summary

S	R-sq	R-sq(adj)
28.8335	20.95%	20.87%

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Attack	1045	0	80.47	1.00	32.41	5.00	55.00	77.00	100.00	190.00
Defense	1045	0	74.661	0.966	31.238	5.000	50.000	70.000	90.000	250.000

Conclusion

As we can see from the gathered data, it does not appear as if there is much of a correlation between the Pokemon's attack and defensive stats. With an R squared of 20.95% it does not prove that there is any sort of positive correlation between the two statistics. Only about 20 percent of the points lie suitable to be on or near our line of best fit. Adding on the R value of 0.4577 it shows that our correlation coefficient is low as well, there are too many data points straying away from the line of best fit. The low R value is even more proof that the correlation between the attack and defensive stat is low. All in all, this is a bummer for a pokemon fan like me. I was in high hopes that there would be a strong correlation so that it would be easier to hand craft the most suitable team to go against my opponents. As a whole doing this study shed much light on the topic for me because it is crazy to really put this in perspective. I definitely thought there

would be a much greater correlation between all of the Pokemon but this clearly proves me wrong.

Infographic

https://create.piktochart.com/output/56824772-is-there-a-correlation-between-the-attack-and-defense-stat

Sources

https://www.kaggle.com/hamdallak/the-world-of-pokemons