Linear Regression

hwittlieff

11/20/2021

Introduction

Since 1990 Pokemon has been captivating the hearts and minds of children and adults alike. In the pursuit of Pokemon battle wins, many players of Pokemon often ask themselves what makes up the strongest Pokemon? The complete Pokemon database on Kaggle provides data on over 800 Pokemon from 7 different generations. This data set features base statistics such as: attack, defense, height, weight, and many more. These base statistics can be utilized to evaluate each Pokemon's individual power.

Problem Statement

Do physical characteristics (height, weight, and gender) impact Pokemon base stats?

Addressing the Problem Statement

To address the problem statement I utilized The Complete Pokemon Dataset (Banik 2017). From the database a new statistic was created from the following "power" classified fields:

- attack
- defense
- hp

This new statistic represents the sum of the individual Pokemon's rankings in each of the three categories. For example, a Pokemon that was ranked 5th highest in attack, 20th highest in defense, and 15th in hp would receive an overall ranking of 5+20+15 = 40. This new field is called ranked_power.

A regression model will then be used to attempt to predict this newly created field. This model will use the below independent variables:

- weight_kg
- height_m
- percentage_male

Analysis

1. Create the new ranked_power variable

```
# Create new ranked power field
Pokemon_df$attack_rank <- rank(-Pokemon_df$attack, na.last=TRUE, ties.method="average")
Pokemon_df$defense_rank <- rank(-Pokemon_df$defense, na.last=TRUE, ties.method="average")
Pokemon_df$hp_rank <- rank(-Pokemon_df$hp, na.last=TRUE, ties.method="average")
Pokemon_df$ranked_power <- Pokemon_df$attack_rank + Pokemon_df$defense_rank + Pokemon_df$hp_rank
```

2. Build the regression model

```
# Create linear model
power_lm <- lm(ranked_power ~ height_m + weight_kg + percentage_male, data = Pokemon_df)
summary(power_lm)
```

```
##
## Call:
## lm(formula = ranked_power ~ height_m + weight_kg + percentage_male,
##     data = Pokemon_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -902.97 -334.94  -44.32  341.95 2620.04
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1642.9738    53.3434  30.800  < 2e-16 ***
## height_m        -144.5867    23.4393  -6.169 1.17e-09 ***
## weight_kg         -2.8913     0.3012  -9.599  < 2e-16 ***
## percentage_male   -1.7056     0.8550  -1.995   0.0465 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 458.4 on 699 degrees of freedom
##   (98 observations deleted due to missingness)
## Multiple R-squared:  0.3108, Adjusted R-squared:  0.3078
## F-statistic: 105.1 on 3 and 699 DF,  p-value: < 2.2e-16
```

The fitted model is statistically significant for each independent variable as all have P values lower than 0.05. The R2 value is ~.31 meaning that about 31% of the variability in the ranked_power field can be predicted from the model.

Implications

The results of this model can be used to increase the likelihood of picking the most powerful Pokemon if the Pokemon's height, weight, and percentage male are known. This information will be useful to players early on when playing digital Pokemon games in selecting which Pokemon to train throughout their play through. Additionally, the results from this project could be used to identify gaps in Pokemon characteristics to further improve the balance of the games.

Limitations

This model was only able to account for 31% of the variability in our custom ranked_power field. Adding additional characteristics such as Pokemon type, may be able to further improve the fit of the model. Additionally, Pokemon can be weak or strong against various other types of Pokemon. Investigating these relationships was outside of the scope of this project, but may warrant being included in future analysis.

Concluding Remarks

From this analysis some insight into the characteristics of Pokemon and their impacts on power was gained. However, opportunities still exist to build knowledge on these relationships. This model provides insight as a stepping stone for future analysis.

Bibliography

Banik, Rounak. 2017. *The Complete Pokemon Dataset*. Kaggle. https://www.kaggle.com/rounakbanik/pokemon.