Pokemon Status and Type

Tests of Association between variables Pokemon Status and Type

Jeremy Jian Zhi Chia (s3917515)

14/10/2021

RPubs link information

- You must publish your presentation to RPubs (see here) and add this link to your presentation here.
- Rpubs link comes here: www. . . . .
- This online version of the presentation will be used for marking. Failure to add your link will delay your feedback and risk late penalties.

Introduction

- The purpose of this investigation is to investigate if there are any association between Pokemon Status and Pokemon Type.
- Pokemon Status gives the status of any given Pokemon, namely Normal, Mythical, Sub-Legendary or Legendary.
- This investigation will focus on the Pokemon statuses of Normal, Mythical, Sub-Legendary and Legendary and their associated types. Emphasis will be on Pokemon that are considered "rare" and "common".
- Each Pokemon has a type which determines its weaknesses/resistance to attacks. Some example types include Grass, Fire, Water, Psychic, Electric etc.

Introduction Cont.

- As I grew up watching Pokemon, playing Pokemon video games and collecting Pokemon cards, I thought it would be fun and reminiscent to choose this dataset and be able to apply statistical tests on it to gain some insights!
- It would also allow me to finally know with confidence whether a particular Pokemon type is more likely to be a rare Pokemon!

Problem Statement

- ▶ As all Pokemon are associated with a status and type, I'd like to find out if there exists a relationship between the status and type.
- ▶ Will knowing the Pokemon type allow us to predict the Pokemon status, or will knowing the Pokemon status allow us to predict its type?
- ▶ Does any "rare" Pokemon make up of predominantly a certain Pokemon type?
- ▶ Does any particular Pokemon type have a significant number of Legendary Pokemons?
- ▶ The Chi-square Test of Association is used to test if there is a relationship between the Pokemon status and type.
- ▶ This test will allow us to know if the status and type are independent/dependent on each other and whether knowing the values of Pokemon status helps to predict the value of Pokemon type and vice versa.

Data

- Complete Pokemon Dataset (All 8 generations)
- Source: https://www.kaggle.com/mariotormo/complete-pokemon-dataset-updated-090420
- This dataset contains records of Pokemons across all eight generations.

Data Cont.

- ▶ This dataset contains rich information on each Pokemon, they include Pokemon names in various languages, primary and secondary type, height and weight, special abilities, various battle metrics (Attack, Defense, Speed etc.), egg type, gender distribution and various attack damage against any Pokemon Type.
- ▶ Variables of interest: Status and Type
- ▶ Values of the status variable were conditionally exchanged. With "Normal" Pokemons being classified as "Common" and "Legendary", "Mythical" and "Sub Legendary" Pokemons being classified as "Rare".
- ▶ The status and type variables were set as unordered factors.
- ▶ Unique values of variables of interest were inspected to ensure there are no errors.
- ▶ Existence of null values were also checked.

Descriptive Statistics and Visualisation

Decsriptive Statistics Cont.

Descriptive Statistics Cont.

The barplot displays the proportion of Pokemon status for each Pokemon Type. By observation, all Pokemon types are predominantly made up of "Common" Pokemons. With certain Pokemon types making up a larger proportion of "Rare" Pokemons. The barplot gives an indication as to whether there is an association between the Pokemon type and status. If there was no association between Pokemon type and status, the proportions within the bars of each Pokemon type would be identical. However, the barplot above suggests the possibility of an association between the variables.

Hypothesis Testing

Although the assumption of the Chi-squared test is that at least 75% of the expected value in each cell is greater than 5, there is evidently a clear majority of expected counts over 5. There are also many expected values less than 5 that sit close to the value 5. Therefore, the Chi-squared test will still be used to test for any association between Pokemon status and type.

The critical value was calculated as 27.59. As $89.65 > 27.59$, the null hypothesis is rejected.

Hypthesis Testing Cont.

The reported p-value is 7.062e-12, which is less than the 0.05 significance level. Therefore, the null hypothesis is also rejected using the p-value approach.

▶ Here are the examples of mathematical equations:

H_0 : *There is no association in the population between the categorical varia*

H_A : *There is an association in the population between the categorical varia*

Discussion
Major Findings

Chi-Squared Test

- There are two approaches used to test the null hypothesis within the Chi-squared test. First was the critical value approach and second is the p-value approach.
- The reported X-squared of Pokemon status and type is 89.652 and the reported critical value was found to be 27.587. As X-squared is higher than the critical value, $89.652 > 27.587$, the null hypothesis is rejected.
- The reported p-value was found to be 7.061857e-12 which is less than the significance level of 0.05. Hence, the null hypothesis is rejected with this approach.

Fisher's Exact Test

- The Fisher's Exact Test is used over the Chi-Squared Test when there exists over 20% of cells with expected frequencies of less than 5.

Strengths of Investigation

- The initial Pokemon statuses contained 4 levels, namely "Normal", "Legendary", "Mythical" and "Sub Legendary". However, the levels were combined to increase the number of observations within each subgroup of "Common" and "Rare".
- The gathering of subgroups was still done in a meaningful way. The Chi-Squared Test was able to provide a more accurate result without compromising the insights that can be gained. We can still determine if a particular Pokemon type is considered "Common" or "Rare".
- As this sample size is small and there existed expected frequencies that are below 5, the Fisher's Exact Test was able to provide an accurate result of the p-value.

Limitations of Investigation

▶ Even though gathering these observations increased the accuracy of the Chi-Squared Test, the insights gained were limited to only knowing whether a Pokemon particular Pokemon is "Common" or "Rare".

▶ As a result, we are not able to see into finer granularity as to whether Pokemon type was associated with one of the four statuses.

Considerations for Future Investigations

▶ If there are more generations of Pokemons to come, this will result in a larger sample size. Instead of using a Chi-Squared Test or Fisher's Exact Test, a G-Test of Independence can be used to find any associations between Pokemon status and type.

▶ As this dataset is comprehensive, there are many more analyses that can be conducted. One interesting one would be to use an ANOVA (Analysis of Variance) to test whether a particular Pokemon type have equal heights.

Final Conclusion

The Chi-Squared Test and Fisher's Exact Test of Association were used to test if the association between Pokemon status and type are statistically significant. The results from both tests showed that there exists a significant relationship between the Pokemon status and type. Therefore, the results suggest that some Pokemon types are more likely to be "Rare" Pokemons. Hence, knowing the value of one variable helps predict the value of the other variable.

References

▶ Endod, R. D., 2017. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. [Online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5426219/ [Accessed 12 October 2021].

▶ McDonald, J. H., 2015. Fisher's exact test of independence. [Online] Available at: http://www.biostathandbook.com/fishers.html [Accessed 12 October 2021].

▶ McDonald, J. H., 2015. G–test of independence. [Online] Available at: http://www.biostathandbook.com/gtestind.html [Accessed 12 October 2021].

▶ Soetewey, A., 2020. Chi-square test of independence in R. [Online] Available at: https://statsandr.com/blog/chi-square-test-of-independence-in-r/#conclusion-and-interpretation [Accessed 12 October 2021].

▶ Soetewey, A., 2020. Fisher's exact test in R: independence test for a small sample. [Online] Available at: https://statsandr.com/blog/fisher-s-exact-test-in-r-