

Summary
DSC 530 Term Project
Blandon Lee
03/04/2022

1. Statistical/Hypothetical Question

- Does strength of an athlete play into a sports consideration for being toughest to play?
- What attributes are considered the most important in a sport being labeled as tough to play?
- What attributes are the least desirable to the overall score for a sport?
- Can a single attribute influence the overall ranking? Are multiple attributes necessary?
- Are the toughest sports to play the most popular?
- If profits were considered (variable) would the ranked toughest sports be the most profitable?
- H0 – Popularity does not correlate with the toughness of a sport.
- H1 – Popularity and sports toughness are correlated.

2. Outcome of your EDA

After considering all the variables in the dataset I decided to go with the six that I thought would provide the best possible outcome. I decided to use endurance, durability, strength, analytical aptitude, flexibility, and agility. The reason I chose these variables is because they are great attributes for measuring the toughest sport.

I proceeded to compare the chosen variables with the variable popularity from a different source. Using the popularity variables, I created another called popularity total to measure/test the variable. For the test variable I used the given score divided by total recorded sports in the dataset and produced the mean value. The various actions that were performed using python were for variance, standard deviation, correlation, means test, tails, and p-values.

In addition to the actions listed above I also included visualizations for the main variables used in the project for future reference. The main question that I was trying to answer was does strength of an athlete play into a sports consideration for being toughest to play? In addition to the main question, I was also curious as to what role popularity played into a sports overall score?

Once the analysis was complete, I concluded that strength has a significant impact on determining the toughness of a sport. With the p-value being $6.34554956574914e-38$ between

strength and total and the null hypothesis being rejected indicates that they were in fact significant. This can be verified with the ANOVA hypothesis test.

There was also a multiple regression conducted for strength using total and popularity total as the contributing factors. Having an R^2 of .50898 tells me that strength is a crucial element when looking at ranks. I wanted to see if power was able to produce changes and it did bring the value up to .7568 much closer to 1.

Testing PFM and CDF I was able to discern that I needed at least a value of 5 on the attribute for strength to be a factor in ranking and overall score. Using the CDF I was able to conclude that at the 50th percentile a value of 5.19 was achieved and as the percentile increases the value of strength increases as well. I also used a probability plot against the variable strength, and it returned as expected a normal distribution.

3. What do you feel was missed during the analysis?

Outside of sports popularity I couldn't gather additional and or alternate test data that would have potentially been useful because I simply couldn't find anything else. With there being no real hard scientific research conducted on popularity I feel that it could have created holes in the project. Other than that, I believe that I managed to hit all the aspects of the assignment.

4. Were there any variables you felt could have helped in the analysis?

I did not look too much at nerve, power, or hand-eye coordination. I feel that those attributes could have a more significant role, then I had anticipated when choosing variables to conduct my analysis on.

5. Were there any assumptions made you felt were incorrect?

I think that placing significance on the test variable popularity could have been incorrect because popularity doesn't necessarily mean tough. However, with limited resources available I had little choice but to use it because it was essentially the best option available.

6. What challenges did you face, what did you not fully understand?

The PMF was one area that I was not fully understanding I was constantly looking up and reading various things on its interpretation and use. In the end I am still unsure if I executed it the way it should have been.

Resources:

Biggest Global Sports: A statistics-based analysis of the world's most popular sports. Retrieved from <http://www.biggestglobalsports.com/worlds-biggest-sports/4580873435>

Kriebel, A. (2018). Toughest Sport by Skill. Data.world. Retrieved from <https://data.world/makeovermonday/2018w19-toughest-sport-by-skill>