Name: Kevin Angotti

Course: DSC 530

Title: Toughest Sports by Skill

Date: 30 May 2020

Submit a link to your repository to the assignment link during the final week of class.

1. Statistical/Hypothetical Question

* Does the strength of an athlete contribute to a sports consideration of being the toughest to play?
* What attributes contribute the most to a sport being labeled as tough to play?
* What attribute(s) cause the most harm, or are the least desirable to the overall score?
* Can one attributes influence the overall ranking, or are multiple attributes required to make a sport the tough?
* Are the toughest sports the most popular?
* If there were financial data added in, i.e., have a variable, would the toughest sports by rank also be the most lucrative?
* H0 – Popularity does not correlate with the toughness of a sport.
* H1 – Popularity and sports toughness are correlated.

2. The outcome of your EDA

After looking through the different variables, I focused on six Strength, Endurance, Durability, Flexibility, Analytical Aptitude, and Agility. I compared all of these variables with a variable I created using a second source to calculate for Popularity of a sport[2]. From the variable Popularity, I created a second called Popularity Total to use as a measuring or test variable. I source for sports popularity used a measuring score compiled over four years, for which I took that score and assigned it to the corresponding sport in my original dataset[1]. For the test variable, I simply took the given score of the sport and divided it over the total recorded sports in the dataset for a mean value. Each variable was given different statistical analysis tests conducted in the Python language on the Spyder IDE.

Testing for mean, variance, standard deviation, correlation, covariance, tails, and p-values was conducted. Visual representation of the main variables was also included for reference. The main research question asked was, how strength contributed to a sport being labeled toughest to play? Followed by does Popularity have any influence on the overall score. After conducting my analysis, I was able to conclude that strength does have a significant impact on the toughness of a sport. With a p-value of 6.34554956574914e-38 between strength and total, there is strong evidence against the null hypothesis meaning the two variables are significant. This was verified by the ANOVA hypothesis test, as it was concluded that they are probably the same distribution, i.e., significant. I also conducted multiple regression for strength with total and Popularity total as the contributing factors. An R2 of .50898 tells us that strength has quite a bit of weight when looking at the whole picture for rank. I added in power to see if that changes anything, and sure enough, it brings the value to .7568 even closer to 1.

A test on PMF and CDF concluded that for strength to be a factor in a ranking or overall score, the attribute needed to have a value of at least 5. Looking at CDF, it was concluded that at the 50th percentile, a value of 5.19 was returned. Thus, as percentile rank increased, so does the value for strength with 7.3, 8.1, and 9.2 for the 90th, 95th, and 100th percentiles, respectively. A probability plot was placed against the variable strength, and as expected, the plot shows a normal distribution, an approximately straight line gradually increasing upwards. Being able to place a suggested outcome on my project was delightful to see, and with some of the main summary statistics like Sd and variance producing low values to help confirm was helpful in the process.

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

I think that a big area I missed was finding better test data outside the sport's popularity information I was able to find. There was no real scientific research assigned to the study, so that could be a large hole in the project for the areas relevant.

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?

Putting a high significance on my test variable of popularity total could be somewhat of an oversight. However, I think it was relevant to find out what constitutes a tough sport for everyone.

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

I do not think I was looking at and fully understanding the PMF. I was able to compute one for variables. However, I could not figure out how to calculate a variable such as strength to other data of the same variable. So, what I did for the project may not be relevant.

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

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