

Winning Characteristics in the Olympics

Noah Obuya and Tamya Davidson

Introduction and Data

Research Question

What are the most influential characteristics (between sex, age, height, weight, and country) when it comes to the Summer Olympics, and do these characteristics change over the course of a decade?

We chose Olympics data from TidyTuesday's github repository (<https://github.com/rfordatascience/tidytuesday> 07-27/readme.md). The data were collected by scraping www.sports-reference.com and was created in May 2018. The data contains 271,116 observations of 15 variables. The variables of interest in our research include sex, age, height, weight, noc (country), year, season, and medals (Gold, Silver, Bronze). Based on these variables, we will answer the question of what are the most influential variables that influence an athlete receiving a gold medal, and do these variables change over time. From the data set we will only observe the more recent Olympic games (including the years 2004, 2008, 2012, 2016), and we will analyze our research question through subsets of the data. There are many NA values corresponding to medals, and because this is our variable of interest we will drop all NA values corresponding to medals. After doing this we are left with a case study of 39,783 observations of 15 variables. The motivation behind this project is to analyze what athletes can do to better prepare for the Olympic games, and see which factors are more influential than others.

Variables of Interest

Exploratory Data Analysis

Methodology

Why We Chose an Ordinal Model

Model Selection Process

Variable Interactions and Transformations

Model Conditions and Diagnostics

Results

Final Model

Key Results and Interpretations

Discussion

Revisiting the Research Question

Hypothesis Test

Limitations

Ideas for Future Work

Appendix