Nick Browen

Eric Ortiz

Part 1: Proposal

1. Research Question: How can we best predict an Olympic track athlete’s finishing time (events measured in time i.e. excluding field events)? What is the relationship between an athlete’s finishing time at the Olympics and (at the athlete-level) their sex, age, weight, height, (at the event level) year of competition, and (at the country-level) the athlete’s nationality, their country GDP, and population?

The Olympics always garners so much attention every time this event happens. The whole world watches. We wanted to know what athlete characteristics contribute to their success and how the country they are from affects this. There is so much data readily available about the Olympics, even going back to 1896. We also were excited to be able to combine datasets and information from multiple resources from Olympic data to population and GDP of the countries of the world over time.

We chose to include athlete-level characteristics because the ideal physique for a marathon runner is much different than a 100m sprinter and we think this will be useful in explaining differences in time and are curious what other nuances will be revealed. At the country-level, we chose to include explanatory variables because athletes are sent as a team by their country and so the athlete pool, training resources and even quality of life might reveal some trends in predicting finishing times. I really like how you have athlete level variables and country level variables. But I do wonder if you need to “standardize” the response variable a bit? Nesting within event is good but still difficult to compare the times to each other? Would it be better to pick one event and look at multiple athletes from the same country over time?

1. *Article 1:* Universality, Limits and Predictability of Gold-Medal Performances at the Olympic Games

Radicchi F (2012) Universality, Limits and Predictability of Gold-Medal Performances at the Olympic Games. PLoS ONE 7(7): e40335. https://doi.org/10.1371/journal.pone.0040335

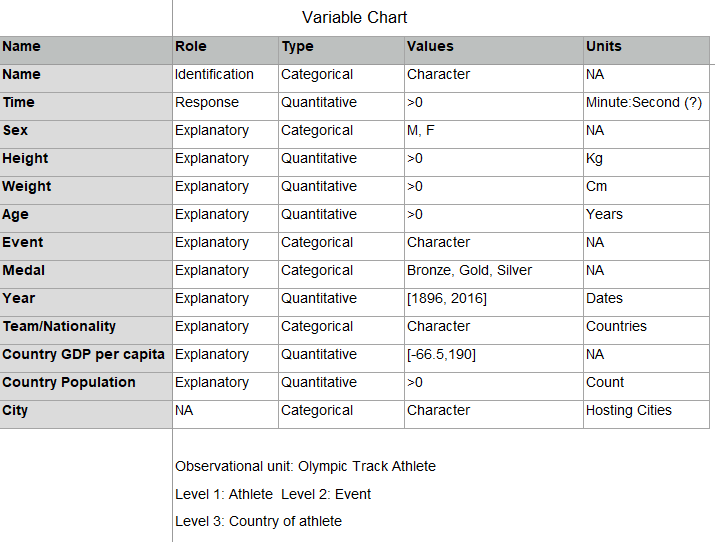
*Article 2:* Predicting Olympic Medal Counts: The Effects of Economic Development on Olympic Performance

Bian, X. 2005. Predicting Olympic Medal Counts: The Effects of Economic Development on Olympic Performance. The Park Place Economist, XIII: 37–44.

Link: <https://pdfs.semanticscholar.org/7293/1ab692bcab9e724b0e5ed4adb53b7ff8097f.pdf>

Article 1 provided a useful primer into all the thought and research that has gone into modeling Olympic performance and medaling. Article focused on all gold medals for all events and attempted to predict the “relative difference between improvements in two different editions of the game.” The authors showed that this statistic, if calculated with respect to an asymptotic performance value (i.e. peak physiological performance time, weight, length, etc. in the competition event, follows a normal distribution. It was an interesting idea to create a meaningful variable to predict Olympic performances. On a side note, the authors also indicated several times that there are some differences in analysis of swimming and all other non-swimming events.

Article 2 was an analysis on country-level characteristics of the Olympics. The author was interested in how socio-economic factors influenced a country’s medal count. He found GDP per capita and population statistically significant in predicting the number of medals won by a country, only including countries in the analysis that won at least one medal. The author said choosing only those countries to include most likely inflated the impact that GDP and population has on medal count. He also discussed addressing the problem of multicollinearity with the variables GDP per capita and population. Also included was a discussion on how GDP per capita and population affect each other claiming “marginal contribution of population growth to the Olympic medal winning process tends to decline as the population size gets bigger”, which gave us the idea to test an interaction with these two variables.



1. We will address this using a hierarchical linear model with level 1 being athlete (random effect) observational unit and level 2 being event (fixed) and level 3 being country (random effect?). do you have level 2 units that are randomly selected from some larger population? Could look at multiple races for the same athlete and athlete within country? Then at least athlete is random?

I’m assuming by even you mean type of race? Or do you mean different track events? How are you incorporating time?

1. We wanted to our project on Olympic data and so we searched around and found the track data on Kaggle and then country characteristics from Gapminder. We also are going to include athlete-level data (height, weight, age, gender) and we found this data on Kaggle.

The link for the data of a country’s population and GDP per capita over time:

<https://www.gapminder.org/data/>

The link for Olympic track results from 1896-2016:

<https://www.kaggle.com/jayrav13/olympic-track-field-results>