STAT3355.002: Project Proposal

Team 5

- Han Nguyen (TXN200004)
- Matthew Bazzell (MXB116930)
- Carlos Flores (CEF200002)

Dataset Information

Olympics 124 years dataset (1896-2020):
https://www.kaggle.com/datasets/nitishsharma01/olympics-124-years-datasettill-2020

Olympic Hosts dataset (1896-2028):
https://www.kaggle.com/datasets/piterfm/olympic-games-hosts

Our World in Data - Population by Age Group
https://ourworldindata.org/grapher/population-by-age-group

Country Area dataset:
https://www.worldometers.info/geography/largest-countries-in-the-world/

Background

This project investigates whether hosting the Olympics provides a measurable home advantage in gymnastics, a sport that relies heavily on subjective judging. While athletes' performances are scored by panels rather than measured by time or distance, subtle judging biases—intentional or not—may favor competitors from the host nation. By analyzing Olympic gymnastics data spanning over a century, we aim to determine whether hosting correlates with a significant increase in medal counts or scores for the host nation's gymnasts—and whether that advantage depends on factors such as first-time hosting, country size, wealth, or demographic structure.

Motivation

Gymnastics is one of the most prestigious and closely watched Olympic sports, but its outcomes depend on human judgment rather than objective measurement. This makes it an ideal case for studying potential bias in subjective evaluation. The Olympics are like a built-in experiment every four years, some countries host, others don't along with measurable results amongst many sports. It's a great setup for stats analysis. Using a long historical dataset, we can identify whether hosting correlates with statistically significant performance boosts, and whether certain

national characteristics amplify that effect. Understanding this relationship is valuable for both statistical modeling and fairness in international competition. If consistent bias is found, it raises questions about the judging system, and transparency in the Olympic Games.

Big Question

Does hosting the Olympics provide a measurable competitive advantage in gymnastics, and what national factors influence the strength of that advantage?

Specific Questions

1. **Does hosting the Olympics provide a measurable competitive advantage in gymnastics?**
   - Do host nations win significantly more gymnastics medals during their hosting year compared to their historical baseline performance?
   - What is the average percentage increase in total medals and gold medals for host countries in gymnastics?
2. **Does the home advantage in gymnastics differ by gender?**
   - Do female gymnasts benefit more from hosting than male gymnasts?
   - Has the gap between men's and women's home advantage changed over time?
3. **Does hosting experience and country size influence the magnitude of home advantage in gymnastics?**
   - Do first-time host nations see a greater competitive advantage in gymnastics than countries that have hosted multiple times?
   - Do geographically smaller countries or countries with smaller populations experience larger percentage increases in gymnastics medal counts compared to larger nations?
   - Does the host country tend to have a relatively larger number of participating gymnasts?
4. **How do economic prosperity and demographic age structure affect the home advantage effect for host countries in gymnastics?**
   - Among host countries, do those with higher GDP per capita see a different magnitude of gymnastics medal increase compared to those with lower GDP per capita?
   - Do host countries with younger median ages or larger working-age populations experience greater competitive advantages in gymnastics when they host?

Why These Questions Are Interesting and Related

These questions explore how and why host countries gain an advantage in gymnastics, a subjectively judged Olympic sport. The first examines if hosting boosts medal counts overall, while the second looks at gender differences to see if men or women benefit more. The third

analyzes whether hosting experience, country size, or participation levels influence that advantage, and the fourth connects it to economic and demographic factors like GDP and age structure. Together, they reveal not only whether home advantage exists, but which nations benefit most and under what conditions.