STAT 630: DAP Proposal

Ashish Ashish Pravin Senkottaiyan Baskaran Arash Ahmadi

Due: Nov 11th, 2024 at 11:59 pm

Introduction

This project investigates the relationship between a country's economic status, represented by its GDP, and its success in winning medals at the Olympics. We aim to determine whether a higher GDP correlates with a greater total medal count, which would suggest that wealthier countries might have better resources to support their athletes. This question is essential to understand how economic factors might influence athletic success on the global stage. However, it is crucial to note potential ethical concerns, such as oversimplifying or overlooking other significant factors like athletes' skills, dedication, and access to training irrespective of GDP, as well as the varied economic circumstances that may affect countries differently.

Data

Link: https://www.kaggle.com/datasets/ernestitus/2024-olympics-medals-vs-gdp?resource=download.

Name: Modified Dataset from '2024 Olympics Medals and Economic status'by Mohamed Yosef.

Reference: Kaggle.com : By Mohamed Yosef.

The dataset includes the 2024 Summer Olympics medal counts for the top-performing countries, showing how many gold, silver, and bronze medals they won, as well as their total medal count and GDP for 2023. The data was put together by combining three sources: Olympic medal data from Kaggle, GDP data from the World Bank, and country codes from IBAN. Country names were used to match the data, and country codes were adjusted to make sure they were in the correct Alpha-3 format. For countries like North Korea that didn't have GDP data, estimates were used. A selective sampling method was used, focusing only on the top-performing countries instead of all the countries that participated in the Olympics. This method was chosen because it helps us focus on countries that are more likely to show a link between economic strength and Olympic success. However, this means the results might not apply to all countries.

Data Analysis Plan

For this analysis, we will use GDP as the independent variable and total medal count as the dependent variable. Our hypotheses are as follows:

- Null Hypothesis (H_0): There is no significant relationship between GDP and total Olympic medals.
- Alternative Hypothesis (H_A): There is a positive relationship between GDP and total Olympic medals.

We plan to use scatter plots to visually assess the relationship between GDP and medal count, followed by a linear regression analysis to test our hypothesis. We will check assumptions such as normality of residuals and homoscedasticity (constant variance) to validate the results. Additionally, we might include population size as a covariate in the analysis, as larger populations could potentially increase the pool of competitive athletes.