

Egypt University of Informatics

Computer and Information Systems

Data Analysis Course

The Analysis of

Olympic Games (1994-2024)

Submitted by: Renad Mohamed, Rawan Ahmed

19/8/2024

# Introduction

The "Olympic Games 1994-2024" dataset of the Summer and Winter Olympics spanning three decades. It includes information on participating countries and medals they gained.

# Research Question

What is the effect of Covid-19 on total number of medals gained before, after and during covid?

# Hypothesis

There is relation between total number of medals and Covid-19 which affect in all fields in the life and world.

# Population of Interest:

Countries participate in Olympics and Medal Outcomes.

# Sampling Method:

Simple Random Sampling: Randomly choosing 4 countries (USA, ITA, FRA, JPN) from the entire list to study their performance over the years (2016,).

# Bias Identification:

The report uses simple random sampling to choose only four countries (USA, ITA, FRA, JPN) to study their Olympic performance. Although random sampling was used, selecting only four countries may introduce bias by not considering a more diverse representation of countries, which could lead to skewed conclusions about the overall impact of COVID-19 on the Olympics.

# Analysis:

To analyse the Olympic Games dataset, we will calculate basic descriptive statistics such as the mean, median, and mode for various metrics

We will also create visual representations of the data using charts and graphs, such as line graphs and pie charts for (USA, ITA, FRA, JPN) through 2016, 2020 and 2020, Box plot for each year (2016, 2020 and 2024). These visualizations will help us identify any significant trends and events and how medal distributions change over different Olympic Games.

|  |
| --- |
|  |
| OLY 2016  Box Plot showing the distribution of a dataset, highlighting the median, the interquartile range (IQR) allowing for a quick comparison of central tendencies and variability across different countries. |

|  |
| --- |
|  |
| OLY 2020  Box Plot showing the distribution of a dataset, highlighting the median, the interquartile range (IQR) allowing for a quick comparison of central tendencies and variability across different countries. |
|  |
|  |
|  |
| OLY 2024  Box Plot showing the distribution of a dataset, highlighting the median, the interquartile range (IQR) allowing for a quick comparison of central tendencies and variability across different countries. |
|  |
| |  | | --- | |  | | OLY 2016  Bar chart showing the proportional distribution of some countries within a dataset. | |
| |  | | --- | |  | | OLY 2020  Pie chart showing the proportional distribution of some countries within a dataset. | |  |  | | --- | |  | | OLY 2024  Pie chart showing the proportional distribution of some countries within a dataset. |  |  | | --- | |  | | Bar chart show total number of medals during Covid-19 (2020), before (2016) and after (2024) |     Bar chart show total count of medals in USA    Bar chart show total count of medals in Italy    Bar chart show total count of medals in France    Bar chart show total count of medals in Japan    This graph shows hypothesis of gold medals in summer and winter Olympics    his graph shows hypothesis of silver medals in summer and winter Olympics    his graph shows hypothesis of bronze medals in summer and winter Olympics  **Hypothesis testing 2 (summer and winter Olympics)**  **Step 1:** Define null and alternative hypothesis.  **Step 2: Choose the Appropriate Test (T-test)**  **Step 3: Calculate the P-Value**  Gold Medals :P-value: 0.62278  Silver Medals :P-value: 0.98413  Bronze Medals :P-value: 0.20074  **Step 4: Determine the Statistical Significance**   * **Gold Medals:**   + P-value = 0.62278 > 0.05   + Fail to reject the null hypothesis (no significant difference in the mean number of gold medals between the groups). * **Silver Medals:**   + P-value = 0.98413 > 0.05   + Fail to reject the null hypothesis (no significant difference in the mean number of silver medals between the groups). * **Bronze Medals:**   + P-value = 0.20074 > 0.05   + Fail to reject the null hypothesis (no significant difference in the mean number of bronze medals between the groups).  |  | | --- | | A graph of a number of people  Description automatically generated with medium confidence | | This graph shows the number of gold medals in summer Olympics in each country |  |  | | --- | | A graph of a person  Description automatically generated | | This graph shows the number of silver medals in summer Olympics in each country |  |  | | --- | | A graph of a number of bars  Description automatically generated with medium confidence | | This graph shows the number of bronze medals in summer Olympics in each country |  |  | | --- | | A graph of a number of bars  Description automatically generated with medium confidence | | This graph shows the number of gold medals in winter Olympics in each  country | | A graph of a number of bars  Description automatically generated with medium confidence | | This graph shows the number of silver medals in winter Olympics in each  country | | A graph of a bar graph  Description automatically generated | | This graph shows the number of bronze medals in winter Olympics in each  country | |  | |

# Hypothesis Testing Steps

* Step 1: Define null and alternative hypothesis.
* Step 2: Choose the appropriate test (T-test).
* Step 3: Calculate the p-value.
* Step 4: Determine the statistical significance.

# Conclusion

The analysis of the Olympic Games data from 1994 to 2024 reveals notable differences in medal distributions across different years, particularly before, during, and after the COVID-19 pandemic. The study focused on the performance of four selected countries (USA, ITA, FRA, JPN) and employed various statistical methods and visualizations to explore trends in medal outcomes.

# Any potential issues

The analysis focuses primarily on the COVID-19 pandemic as the key factor affecting medal outcomes, but other external factors—such as changes in Olympic rules, the introduction of new sports, or political influences—may also have played a significant role. Failing to account for these variables could limit the depth and accuracy of the conclusions.

GitHub links

Rawan : <https://github.com/rawansalman/Project-2.git>

Renad: https://github.com/RenadMuhammed/Olympic-Games-1994---2024-.gits