# AGE ASSOCIATED WITH OLYMPICS MEDALS?

A Hypothesis Testing on Olympics Data


## ABSTRACT

"Faster, Higher, Stronger - Together" – the Olympic motto. At every Olympic Games event, many compete and only a few have the honours of gold, silver, and bronze medals. We saw the performance of athletes of different ages, and we wonder if there is any correlation between age and getting medals. Would a younger age than 25 be an advantage or otherwise? Based on the data we have regarding 261,642 event participations of athletes between 1896-2016 of both summer and winter games, we are unable to prove that age is associated with winning medals. Despite not proving any associated of medal winning, we still have discovered some interesting insights of the Olympians from the dataset.



Steven Leung


DSCI 524 Lab2 (Note: Repo link and persona in Appendix.)

## Introduction

Olympic athletes require to train year after year for their chance of winning a medal for their countries and for themselves. The events are highly competitive and demanding.

We thus want to ask the hypothetic question,

*"Is the proportion of athletes younger than 25 that win a medal greater than the proportion of athletes of age 25 or older that win a medal?"*

In the following section, we will explain how obtain we do hypothesis testing with a dataset.

## Methods

### Hypothesis Test

The hypotheses are defined as follows:

$H_0$ : The proportion of athletes under 25 who won medals is equal to the proportion of athletes 25 and older who won medals.
$H_A$ : the proportion of athletes under 25 that won medals is greater than the proportion of athletes 25 and above who won medals.

We then carry out the following steps:

1. Compute the observed test statistic from the original sample;
2. Use the null model to generate 100 random permuted samples from the original sample and calculate their corresponding r test statistics;
3. Generate the null distribution using these r test statistics;
4. Check if the observed test statistic competed in Step 1 falls on the distribution; and
5. Calculate the p-value to verify the result.

You might wonder about our choice of the age 25 for this hypothesis testing. While it could arbitrarily be any age, the main reason is that we want to pick a number which is close to the centre of the whole distribution. And the mean age in our data (of those observations not missing age in them) is 25.56 so 25 is a reasonable choice. Moreover, U25 (standing for "Under 25") is also a common age division of athletes in many sports. Here is an example: Women's U25 National Team

## Dataset

Our dataset comes from TidyTuesday (https://github.com/rfordatascience/tidytuesday/tree/master/data/2021/2021-07-27). Here are some points to note regarding the data:

1. The data cover the games between years of 1896 and 2016, both inclusive; and
2. We have the data for 261,642 observations and each observation is an athlete-event-games tuple and is NOT a single athlete. Each observation represents 1 participation in

1 event by 1 athlete, where the same athlete might have competed in multiple events in the same games as well as in multiple games in different years.

## Results

Here is a summary of the results:

| Age | Medal | Number of samples | Medal-winning Proportion |
|---|---|---|---|
| **Under 25** | 17,939 | 131,134 | 13.68% |
| **25 or Above** | 21,112 | 130,508 | 16.18% |

We can see from the table above that there were 131,134 athletes under age of 25 and 13.68% of them won medals in the events, while there were 130,508 athletes at the age of 25 or above and 16.18% of them won medals in the events. The p-value from our simulation is 1, which is much larger than the threshold of $\alpha = 0.05$.

To illustrate this point, we, in Figure 1, placed our observed test static in the plot of null distribution (the dark grey histogram). The blue dashed line is the threshold of $\alpha = 0.05$. The red solid line is our p-value of 1. In order to prove that we can statistically say "Yes" to our hypothetical question, the red solid line needs to be on the right side of the blue dashed line. But red solid line is on the far left instead. We therefore we fail to prove "Yes" to our hypothetical question. In other words, there is no statistical significance to say that the medal-winning proportions of "Under 25" and "25 or Above" are different.

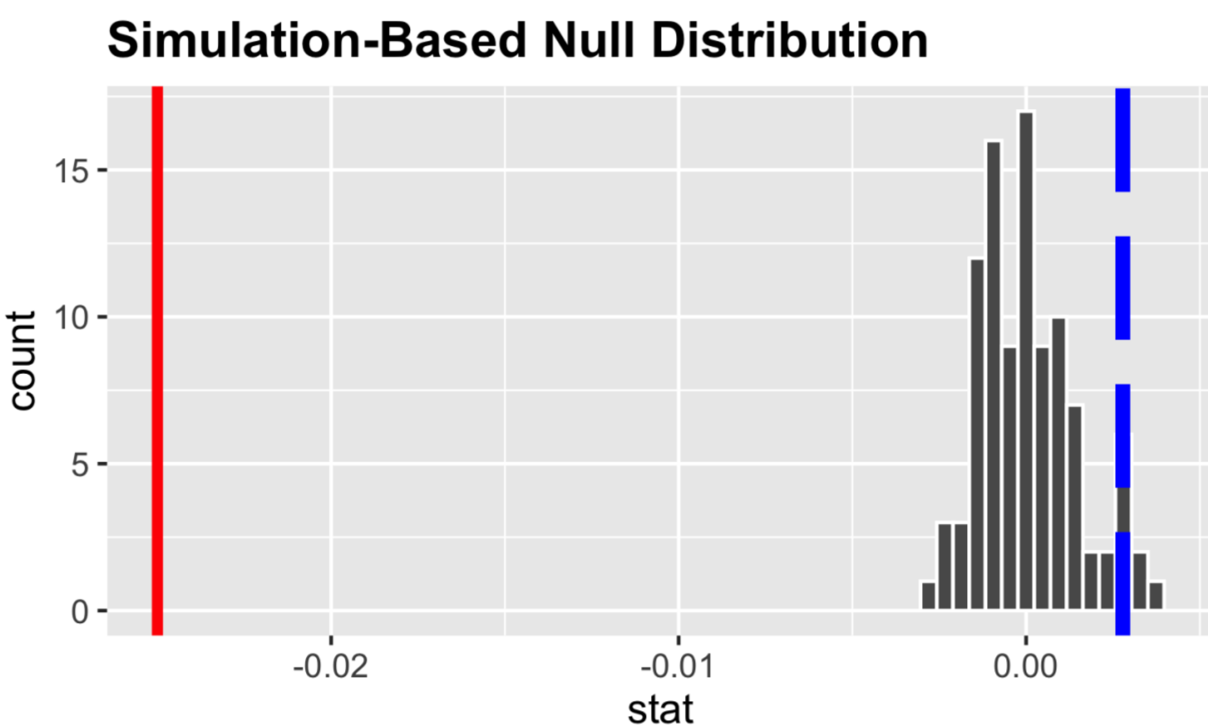


*Figure 1: We fail to reject that "Under 25" and "25 or Above" is different.*

(Please note that technically we cannot say that we have proved that they are the same. If you want to understand more about the nuance between "fail to reject" and "accept" a hypothesis, you can read this blog post at your leisure.)

# Insights of Olympics Athletes

## Age Distribution of Athletes

Let us see what the age distribution looks like for all athletes:

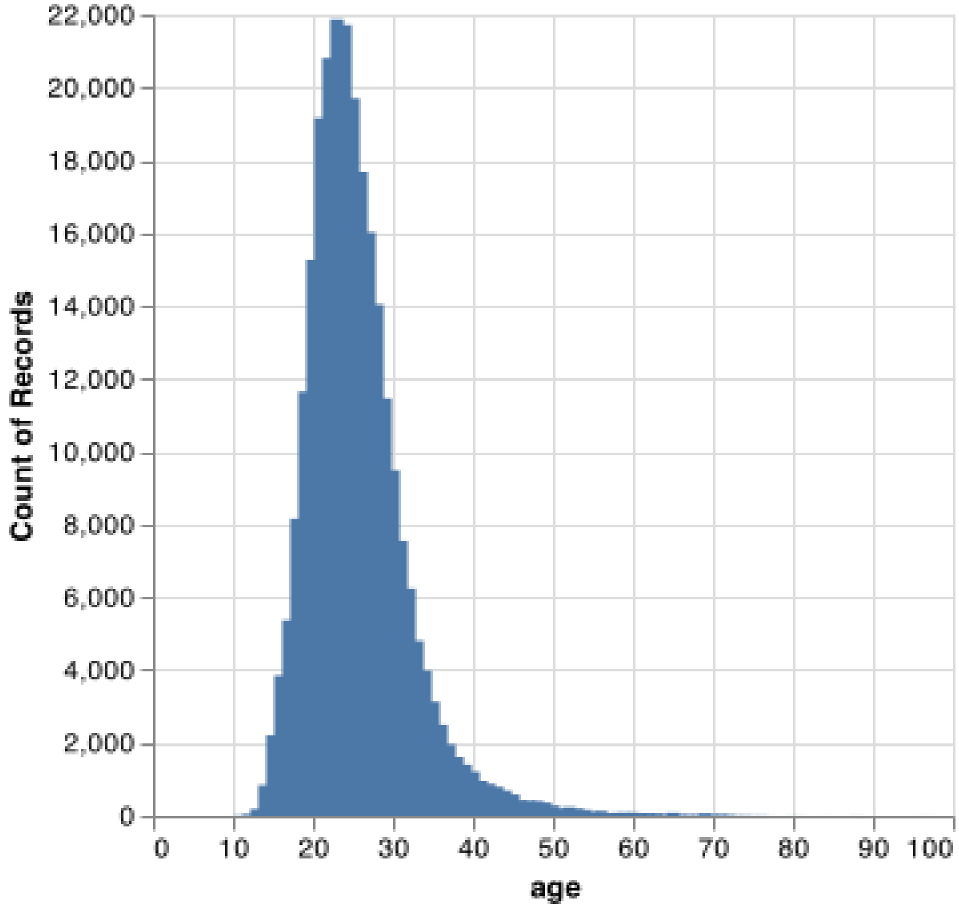


*Figure 2: Age distribution of all Olympic athletes between 1896-2016*

As can be seen above, the age peaks at 23 years old and the distribution is bell-shaped and right-skewed, which means that there are fewer older athletes above 40 years old in the Olympic Games, but the spread can go 70 years old and beyond.

## The Youngest and the Oldest

The Youngest (Figure 3): Dimitrios Loundras (10 years of age at that time) from Greece who won a Bronze Medal in gymnastics Men's Parallel Bars (Teams) in the 1896 Summer Games, the first ever Olympic Games of the modern era.

*Figure 3: Dimitrios Loundras (centre), youngest athlete of Olympic Games of the modern era*

The Oldest: John Quincy Adams Ward (97 years of age at that time) from USA competed in Art Competitions Mixed Sculpturing (Statues) in 1928 Summer Games in Amsterdam. He did not win any medal.

## Any relationship between Height, Weight, Year of the Games and Age?

We now can look at how attributes like height and weight of the medalists and year of the games correlate with age (Figure 4).

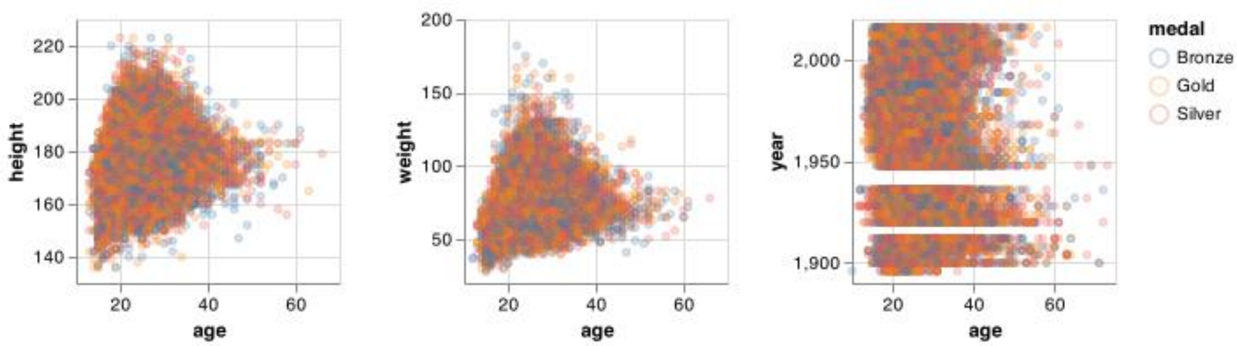


*Figure 4: Height vs Age, Weight vs Age and Year of the Games vs Age*

Some high-level insights:

1. There are some apparent correlations between height and age and between weight and age for the medalists;
2. The maximum age of medalists getting medals seemed to shrink between 1960s and 1980s, and it seemed to increase again till now.

## Age distributions vs Gold, Silver and Bronze?

Please see below the age distributions (Figure 5):

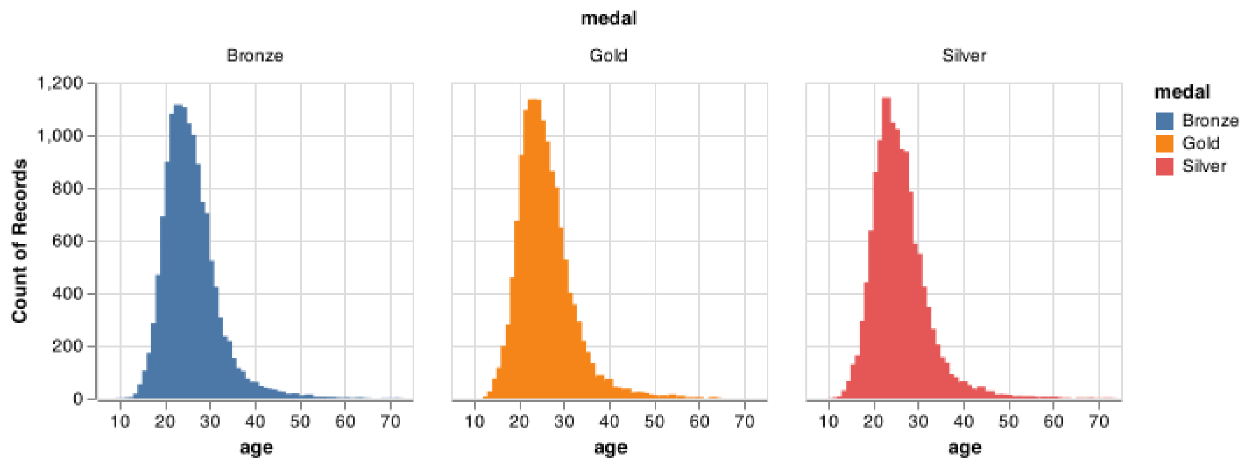

*Figure 5: Different medals, similar distributions vs Age*

The above distributions look very similar. In other words, the age distributions of Gold, Silver and Bronze medalists are very similar.

# Conclusion

As explained above, we did not have enough statistical evident to say that the proportions (Under 25 vs 25 or Above) are different. The results show that athletes under 25 have not been more successful at the Olympics compared to those who are 25 and older.

Despite not being able to say "Yes" to our hypothesis, we have found some additional insights as explained in the Section "Insights of Olympics Athletes" in this report.

## For further study

Winning of medals might be attributed to **experience**:

1. The Olympics have a large variety of events, and these events are constantly changing over the years. Experience, presumably an advantage of older athletes, may be a key advantage in certain sports, while other advantages like physical dexterity might be more important in other sports. Examples would be art-related competitions (such as sculpturing, music, etc. in the 1940's), and precision-related competitions like archery and shooting;
2. We have found a few papers (Singh 2021) and (Elmenshawy, Machin, and Tanaka 2015) that suggest that athlete having more experience is more likely to win a medal.

# References

Singh, Dr Preet Deep. 2021. "Olympic Medals: Matter of Nerves." *Available at SSRN 3901321*.
Elmenshawy, Ahmed R, Daniel R Machin, and Hirofumi Tanaka. 2015. "A Rise in Peak Performance Age in Female Athletes." *Age* 37 (3): 1–8.
AP Photo/Jae C. Hong. 2021. Photo on the cover.
Crispin, Gerard. 2021. "Who is the youngest Olympian ever in the history of the games?" https://www.mediareferee.com/2021/08/05/youngest-olympian-in-history-ever/ for Photo of Dimitrios Loundras.

## Credits

The original study was a group project by Brandon Lam, Sam Quist, Ruben De la Garza and myself for the course DSCI 522 of MDS in Winter Term 1 of 2021 at the University of British Columbia. The writing of this report is based on the version original dated 2021/11/26 and updated on 2021/12-09. Link to the repo of the study: https://github.com/UBC-MDS/olympic_medal_htest

The original dataset can be downloaded here:
https://github.com/rfordatascience/tidytuesday/tree/master/data/2021/2021-07-27

# Appendix

## Author
Steven Leung

## Mechanics
URL to Lab 2 repo: https://github.ubc.ca/mds-2021-22/DSCI_542_lab2_klsleung

## Exercise 1: Persona
KC is the Marketing Lead of my company. She holds a non-technical bachelor's degree and an MBA from a prestige university. She has many years of being a marketing leader for a company, being responsible for devising marketing strategies for the company, defining the target segments for the company's products and services, managing the marketing communications channels including the corporate website, LinkedIn pages, etc. and the generation of contents for those channels and the activities like webinars, attendance for exhibitions, etc.

She has a good understanding of why companies need to transform their businesses by using data-driven methods like data collection, cleaning, exploratory analysis, modeling, visualization and reporting and in general know which companies would find value in those methods. Her motivation to (1) draw customer behavioral insights and (2) learn how to differentiate the different methods and how they should be applied to the business use cases so that she may generate the correct contents to the target segments.

## Peer feedback from KC
*"The logic and structure were clearly laid out and it was easy for readers to follow. It is great if some examples in the reality to further illustrate his conclusion can be listed out."*