Olympic games: Importance of age for winning a gold medal

Daria Borodkina[a], Kseniya Grigoreva[a], Aleksandra Mikhaylova[a], Dariia Shishova[a], Robert Bösch[a]

[a]*HSE University, 20 Myasnitskaya ulitsa, Moscow 101000 Russia*

Abstract

The correlation between the victory (winning a gold medal) at the Olympic Games and the age of athletes is considered. A set of tools of the R programming language is used to check for the presence of a connection. The results are presented in the conclusion: there is a difference in the mean age of the Olympic athletes (gold medal receivers) in 1964 and 2008. The Null Hypothesis of the research was rejected, but the difference in the mean age is not very high.

Keywords: programming language R, Olympic Games, correlation analysis, age, gold medals

1. Intoduction

The Olympic Games are one of the most important events in the world of sports. In this study, we used data that reflects the main participants, their gender, height, weight, age, team, year of the games, as well as the number of gold, silver and bronze medals. Considerable attention was paid to understanding and identifying the dependence of obtaining gold medals on any other aspects of the athlete (age, height, weight). Our hypothesis for this project is “Does the average age of gold medal winners decline?”.

2. Methods

We recruited the data from an open source online, Github.com. The sample includes a lot of data and is divided into several aspects. The data includes 271.116 rows and 15 different columns. The different aspects of the data include the name, sex, age, height, weight, team, nationality, the year of the Olympic Games the person participated in, year, city, sport type, event and the type of medal they won (if they won one). The data was collected through analysis of the previous Olympic Games, beginning from the year 1896 until 2016. The data includes 135.571 different athletes. To test our hypothesis, we made a T-test.

3. Analysis

To test our hypothesis, we need to clarify some details. So, using our data and after analyzing, we got that in total for a given period of time (1896 - 2016) 10148 bronze medals, 9866 silver medals and 10167 gold medals were received.

Code:

```
atheletes_df <- read.csv('athlete_events.csv', header = TRUE, sep = ',') %>%
  na.omit()
kable(atheletes_df %>%
  select(Name,Sex,Age,Height,Weight,Medal,Sport) %>%
```

```
  slice(0:5))
summary(atheletes_df)
kable(atheletes_df %>% summarise(total_records=n()),
  caption = "Total Records")
list_na <- colnames(atheletes_df)[apply(atheletes_df, 2, anyNA)]
list_na
kable(atheletes_df %>%
    group_by(Medal) %>%
    summarise(total_records=n()),caption="Records by Medal Count")
total_records_pie <- c(10148, 10167, 9866)
names(total_records_pie) <- c("Bronze", "Gold", "Silver")
colors <- c('red', 'yellow', 'green')
pie(total_records_pie, labels = names(total_records_pie), col = colors,
main = "Records by Medal Count", radius = -1,
col.main = "darkgreen")
```

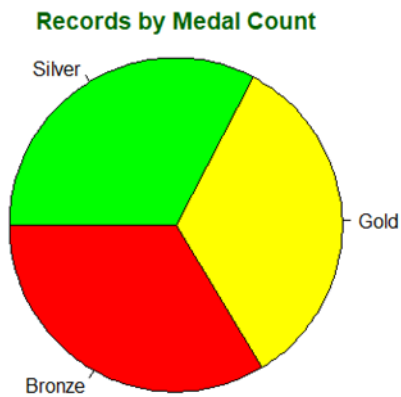

Fig. 1. Records by Medal Count

There is an opinion among researchers that the quality of the national team's performance at the Olympics mainly depends on the welfare of the nation. Extensive research in this area has shown that GDP (per capita) and population have a significant impact on a country's sporting success. In modern research authors use such econometric models as logit models (Andreff W., 2001), probit models (Neville A. and Aktinson G., 2002) and ordered probit models (Johnson D. and Ali A., 2008). We also analyzed and found that most gold medalists come from countries with high GDP as you can see in the graphic down below.

Code:

```
team_champions <- read.csv('athlete_events.csv') %>%
  na.omit() %>%
  select(Team, Medal)
goldmedal <- ifelse(team_champions$Medal == 'Gold', 1,0) #Medal as a dummy variable
team_champions1 <- data.frame(Team = team_champions$Team, Medal = goldmedal) %>%
  group_by(Team) %>%
  filter(Medal == 1) %>%
  count(Medal) %>%
```

```
  arrange(desc(n)) %>%
  select(Team, n)
team_champions1
team_champions2 <- team_champions1[c(1:10), ]
team_champions2
```

Team <chr>	n <int>
United States	2075
Soviet Union	961
Germany	508
East Germany	368
Russia	356
Canada	350
Great Britain	321
Australia	313
China	308
Italy	302

1-10 of 10 rows

Fig. 2. National Teams

Code:

```
barplot(team_champions2$n, names.arg = team_champions2$Team, col = "olivedrab3", horiz=FALSE, cex.names = 0.75, las = 2)
```

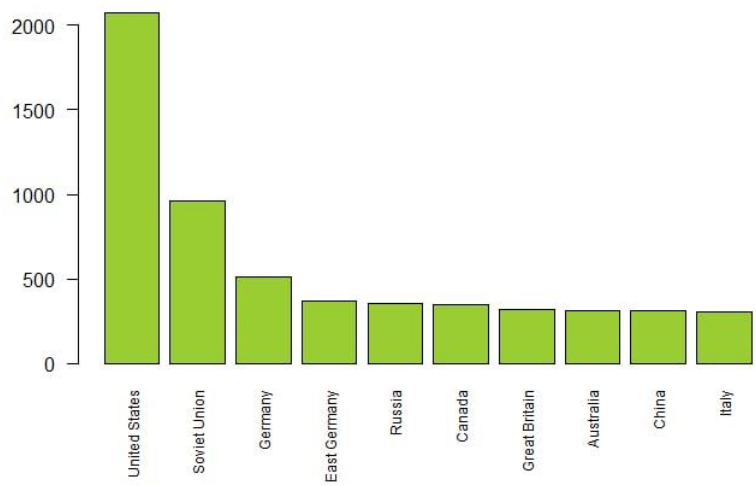

Fig. 3. Barplot of the national Teams

In our opinion, age heavily determines the chance of obtaining a medal. This obviously also depends on the sport itself, but to analyze the influence of age on the chance of winning a medal, we analyzed the age boundaries and calculated the average age of medal winners.

Code:

```
kable (
atheletes_df %>%
summarise(max_age=max(Age), min_age=min(Age), Average_Age=mean(Age)),
caption="Age boundary cases")
```

Age boundary cases

max_age	min_age	Average_Age
66	13	25.42901

Fig. 4. Age boundary cases

According to the data the probability of obtaining a gold medal is more likely in the average age is 25.4 years.

Code:

```
athletes_data1 <- read.csv('athlete_events.csv') %>%
    select(Age, Medal, Year) %>%
  na.omit() %>%
filter(Medal == 'Gold')
histogram(~Age | Medal, data=athletes_data1, type="count", layout=c(1,1), col=c("burlywood", "darkolivegreen"))
```

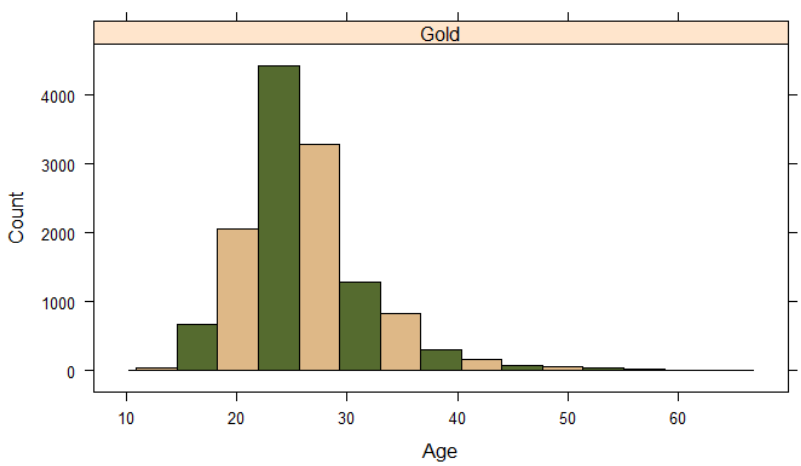

Fig. 5. Histogram of gold medallists by age

In our analysis, we also determined that the majority of Olympic gold medalists received their medals at the age range between 20 and 27 years.

```
Call:
lm(formula = Medal ~ Age + Height + Weight, data = gold_medalists.df1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.3928 -0.3410 -0.3278  0.6537  0.7157

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.957e-01  5.961e-02   3.283  0.00103 **
Age         -1.541e-03  5.467e-04  -2.820  0.00481 **
Height       9.820e-04  4.173e-04   2.353  0.01862 *
Weight       8.005e-05  3.065e-04   0.261  0.79393
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4725 on 30177 degrees of freedom
Multiple R-squared:  0.0008001, Adjusted R-squared:  0.0007007
F-statistic: 8.054 on 3 and 30177 DF,  p-value: 2.321e-05
```

Fig. 6. Result of the test

To check the multicollinearity in the model, in our opinion, it is necessary to insert a correlation matrix. Based on the results of the analysis, the following matrix was derived in R Studio.

Code:

```
gold_medalists.df <-  read.csv('athlete_events.csv') %>%
select(Age, Height, Weight, Medal)  %>%
na.omit()
gold <- ifelse(gold_medalists.df$Medal == 'Gold', 1, 0) #Medal as a dummy variable
gold_medalists.df1 <- data.frame(Age = gold_medalists.df$Age, Height = gold_medalists.df$Height, Weight = gold_medalists.df$Weight, Medal = gold)
model1 <- lm(data = gold_medalists.df1, Medal ~ Age + Height + Weight)
summary(model1)
corrplot(cor(gold_medalists.df1), type = 'upper', method = 'circle')
```

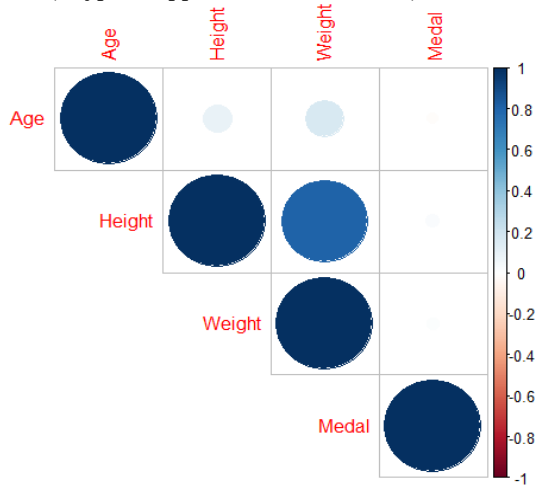

Fig. 7. Correlation plot of gold medallists by age

We can see that there is practically no correlation between the variables, which is a positive factor for the model, it makes it possible to use this model in the T-test. There is a small correlation between weight and height, which equals 0.1-0.2, which is not a significant problem for the model. A slight correlation between height and weight is obvious, since with an increase in height, the weight of a person also increases (axiom) (Colonel Height, 2020).

We wanted to test the hypothesis that over time, the holding of the Olympic Games, the average age of athletes who receive gold medals decreases, as the average age of Olympic participants decreases, since the International Olympic Committee does not officially set any restrictions (International Olympic Committee, (2021). In the theory and practice of sports training, a stable idea has long been formed that achieving high results is possible only at a fairly young age. This was largely facilitated by the work of Soviet scientists V.S. Farfel (Фарфель, 1959, p.67), Z.I. Kuznetsova (Кузнецова, 1975), I.Yankauskas, E. Logvinov (Янкаускас and Логвинов, 1984, с.152). In their publications, the authors, based on a large research material, convincingly showed that various organs and systems of the body are formed, mature and develop non-simultaneously and unevenly throughout a person's life. Also, according to NBC Chicago (2021), there are a variety of sport types where the winning athletes get younger. For example, in diving the age limit is only at 14 years old and often the younger persons close to 14 years win gold (Süddeutsche Zeitung, 2018). Moreover, at the Olympic Games in 2021 in Tokyo, a 12 year old skateboarder won the gold medal (Stuttgarter Nachrichten, 2021)ce . This trend is seen with skeptical eyes since many people think children should not be part of this competition. However, according to Wanner (2021), this trend also has negative effects on the psychological health of these young athletes, since they have more difficulties handling this immense pressure (Volz, 2021). Because of this interesting debate and the general interest in this topic, we wanted to analyze it statistically. For the purpose of this research we observed only age metrics, but height and weight ones are the grounds for future research (Bärtsch, 2021).

For this T-test, we took two years into consideration, the year 1964 and 2008. There is no further reason why we picked exactly these two years, we picked them by coincidence.

Code:

```
t.test(Age~Year, data = athletes_data1964_2008, var.equal=FALSE, alternative  = "less")
```

4. Conclusion

After conducting the T-test we found that the P-value of the test is less than 0.0000000000000002. We can conclude that there is a difference in the mean age of the Olympic athletes (gold medal receivers) in 1964 and 2008. The Hypothesis H0 is rejected, but the difference in the mean age is not very high.

References

Colonel Height. (2020). Does Height Affect Weight? [online] Available at: https://colonelheight.com/does-height-affect-weight/.

Farfel`, V.S. (1959). Razvitie dvizhenij u detej shkol`nogo vozrasta. Moskva: APN RSFSR, s.67.

Kuzneczova, Z.I. (1975). Kriticheskie periody` razvitiya dvigatel`ny`x kachestv shkol`nikov. Fizicheskaya kul`tura v shkole, 1, S.79.

Guzhalovskij, A.A. (1978). Razvitie dvigatel`ny`x kachestv u shkol`nikov. Minsk: Narodnaya asveta, s.88.

Yankauskas, I. and Logvinov, E`. (1984). Motorika rastushhego zhenskogo organizma: ontogenez dvigatel`nogo gomeostaza. Vil`nyus: Moklas, s.152.

Bärtsch, P. (2021). Kommentar: Olympische Spiele sollen keine Kinderspiele sein. [online] Neue Zürcher Zeitung. Available at: https://www.nzz.ch/meinung/kommentar-olympische-spiele-sollen-keine-kinderspiele-sein-ld.1637399 [Accessed 20 Dec. 2022].

International Olympic Committee (2021). What is the age limit for taking part in the Olympic Games. [online] International Olympic Committee. Available at: https://olympics.com/ioc/faq/competing-and-being-part-of-the-games/what-is-the-age-limit-for-taking-part-in-the-olympic-games#:~:text=There%20is%20no%20specific%20age.

NBC Chicago (2021). Is There a Minimum Age for the Olympics? What to Know About Olympic Age Limits. [online] NBC Chicago. Available at: https://www.nbcchicago.com/news/sports/tokyo-summer-olympics/is-there-a-minimum-age-for-the-olympics-what-to-know-about-olympic-age-limits/2581995/.

Stuttgarter Nachrichten (2021). Mindestalter Olympia: Gibt es eine Altersgrenze? [online] stuttgarter-nachrichten.de. Available at: https://www.stuttgarter-nachrichten.de/inhalt.olympische-spiele-2021-wie-alt-muessen-die-olympia-teilnehmer-mindestens-sein.925d6fe4-8a8a-4c18-9ba0-e1fac51884e2.html#:~:text=Generell%20gibt%20es%20bei%20den [Accessed 20 Dec. 2022].

Süddeutsche Zeitung (2018). Olympia-Teilnehmer: Größer, kleiner, älter, jünger. [online] Süddeutsche.de. Available at: https://www.sueddeutsche.de/sport/olympia-olympia-teilnehmer-groesser-kleiner-aelter-juenger-dpa.urn-newsml-dpa-com-20090101-180207-99-968425 [Accessed 20 Dec. 2022].

Wanner, F. (2021). Olympia immer jünger. [online] ArgoviaToday. Available at: https://www.argoviatoday.ch/sport/olympia-immer-juenger-143256438 [Accessed 20 Dec. 2022].