

PORTRAIT ANALYSIS OF AVERAGE SWIMMING ATHELETE

D.A. Zhurman,

Scientific supervisor: E.I. Gubin

Tomsk Polytechnic University

E-mail: daz18@tpu.ru

Introduction

Nowadays swimming is becoming more and more popular. The most famous Olympic champion is also swimmer Michael Phelps. However, there is stereotype that only tall athletes can become Olympic champion in swimming.

That is why the goal of this work is to create portrait of an average swimmer athlete.

To achieve the goal of this work, the following tasks were set:

1. Cleaning and preparing data.
2. Finding correlation between year of Olympic Games and Height (Weight).
3. Finding correlation between distance and Height (Weight).

To achieve the objectives and goals in this work, we use the pandas library for working with tabular data. In addition, we will use numpy library for working with arrays and mathematic computations and matplotlib for plotting graphs. In this work we will use well-know Olympic Dataset and create new dataset with results from official Olympic site [1].

Data preparation

First of all, we checked Olympic dataset. Columns such Games, Team, Sport and Season were dropped. It happened because we will use only swimming results.

Also, we can notice that our dataset contains some experimental distances such as 'men 220 yard freestyle'. We should drop them.

We can see that after this action that dataset has not information about 1904 Olympic Games, because all distancies this year was measured by yards.

Our dataset does not contain information about result. If we want download results from official site of Olympic Games, we should transform our columns for parsing, because links of result use such format: <https://www.olympic.org/beijing-2008/swimming/200m-individual-medley-men>.

We can notice that we have categorical data in our dataset such Medal and Sex. We transformed these columns such way 'Bronze':1, 'Silver': 1, 'Gold': 1 and 'M':1, 'F':0.

After this we splitted data between Male and Female, because they have different values of height and weight.

We can notice that male dataset has 2093 Nan values and female dataset has 1003 Nan values. We replaced these values by median of column.

Finally, we checked dataset for correlations. We create table of correlation, which performed on the figure 1.

	ID	Age	Height	Weight	Year	Medal
ID	1.000000	0.025634	0.016333	0.037056	0.009142	0.020188
Age	0.025634	1.000000	0.214216	0.254992	0.163153	0.017983
Height	0.016333	0.214216	1.000000	0.753113	0.203921	0.094573
Weight	0.037056	0.254992	0.753113	1.000000	0.174285	0.104979
Year	0.009142	0.163153	0.203921	0.174285	1.000000	-0.063270
Medal	0.020188	0.017983	0.094573	0.104979	-0.063270	1.000000

Fig. 1. The table of correlations

From the table we can see that there is only one strong correlation between height and weight. It means that in this work all conclusions for height can be applied to weight.

Correlation between year of Olympic games and Height

In order to visualize and classify data, it was necessary to reduce the dimension. That is, reduce the number of columns from 9083 to 2.

The graph in figure 2 shows that the principal component method is not suitable for reducing the dimension of a given dataset. This is due to the fact that 50 components describe 90% of the variance, that is, in this dataset there are 50 main components.

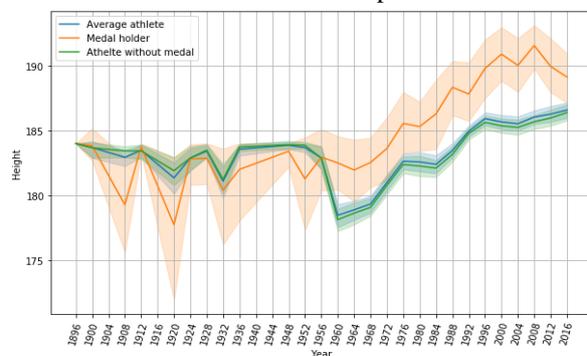


Fig. 2. The correlation between year of Olympic Games and Height for males

We can notice that in 1960 was strange falling of Height (Weight). Let's Figure out. For more detailed analysis, we build graph of correlation between year and number of countries on Olympic Games, which performed on figures 3.

As we can see on figures number of countries constant growth.

After this we checked correlations between country and height in 1956 and 1960 years which performed on figures 4.

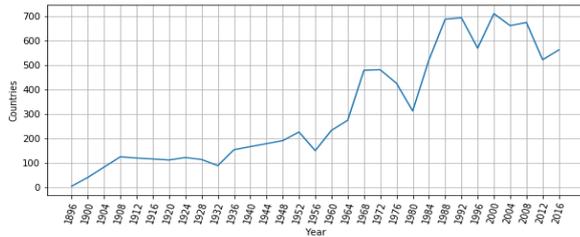


Fig. 3. The correlation between year and number of countries on Olympic Games

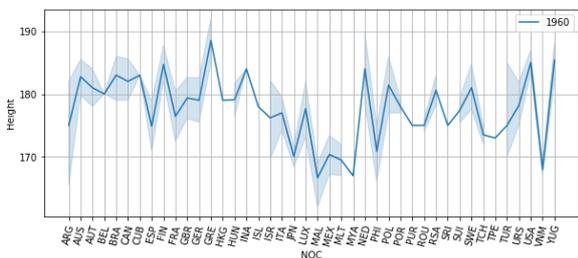
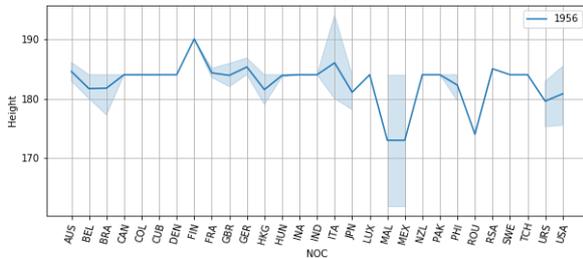


Fig. 4. The correlation between country and height in 1956 and 1960 years

We can see on the figure 4 that fall of height in 1960 connected with increasing of number of countries on Olympic Games. In 1960 Vietnam and Malaysia took part in Olympic games. These countries have low average height.

Also, it connected with growth of numbers of athletes. For example, instead of 2 the most heights athletes in 1956, Finland send 16 athletes which are not so tall.

Correlation between distance and Height

First of all, we create graphs of correlation between year of Olympic games and Height for males, which performed on the figure 5.

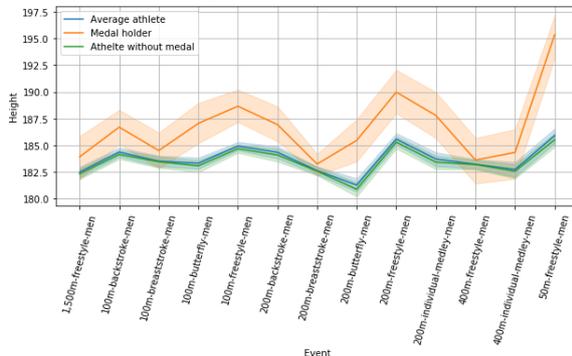


Fig. 5. The correlation between distance and height in for males

We can notice tendency, that than shorter distance then height is more important. Also, we can find out that medal holders in freestyle is higher than medal holders in other styles.

After this, we checked correlation between height and year on the most popular "royal" distance 100m-freestyle, which performed on the figure 6.

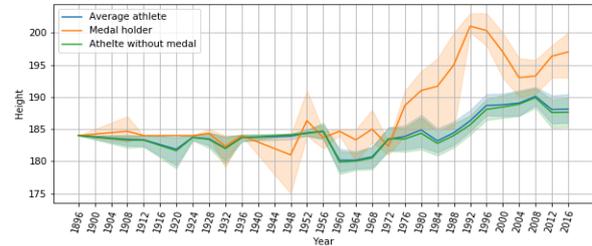


Fig. 6. The correlation between height and year on the distance 100m-freestyle

We can notice that every Olympic Games medal holder becomes higher and higher. The fall of height in 2008 connected with fact that it was the first and last Olympic Games where athletes used swimsuits with new type of material. These swimsuits help decrease correlation between height and medal. After this Olympic Games swimsuits were forbidden [2].

Conclusion

In the course of this work, the following tasks are performed:

1. Cleaning and preparing data. Dataset set of Olympic Games was checked for correlations between columns. Categorical data were replaced by numeric data. Nan values were replaced by median of columns.
2. Finding correlation between year of Olympic games and Height (Weight). We found tendency that every year average medal holder becomes higher and higher.
3. Finding correlation between distance and Height (Weight). We can notice tendency, that than shorter distance then height is more important. Also, we can notice that medal holders in freestyle is higher than medal holders in other styles.

In conclusion, we proved the fact that than higher athlete then more chances he has to become medal holder. Average height of medal holder last Games for male is 186 and 173 for females. Average weight of medal holder for male is 80 and 63 for females.

List of references

1. Kaggle / 120 years of Olympic history: athletes and results// URL: <https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results> – (accessed 14.12.2019).
2. ABC News / Full Body Swimsuit Now Banned for Professional Swimmers // URL: <https://abcnews.go.com/Politics/full-body-swimsuit-now-banned-professional-swimmers/story?id=9437780/>. – (accessed 14.12.2019)