

Report Out: Medal Winners of one decade as a predictor of future participants and medal winners

Contents:

Review of Questions to Answer / Hypotheses / Approach

Discuss Technical Challenges

Detail: Entity Relationship Diagram (ERD)

Initial Findings

Deeper Analysis

Hypotheses Results

Section 1: Questions to Answer

- 1) Medal winners tend to be at the average height, weight, age of the competing group.
- 2) If the medal winners tend to be out of the average (top quartile for example) do you see a subsequent shift of Olympic participants to this new average in future years?
- 3) Are there sports where there is a statistically significant larger standard deviation in age, weight, height than other sports and is this a contributing factor in the shift overtime?

Section 2: Initial Hypothesis

- 1) Do medal winners tend to have the average height, weight, age of the competing group?
- 2) In cases where there is an outlier that outlier's value becomes the new average in future Olympics
- 3) Sports with lower standard deviation will more dramatically show behavior exhibited in 2 then sports with higher standard deviation

Section 3: Data Analysis Approach

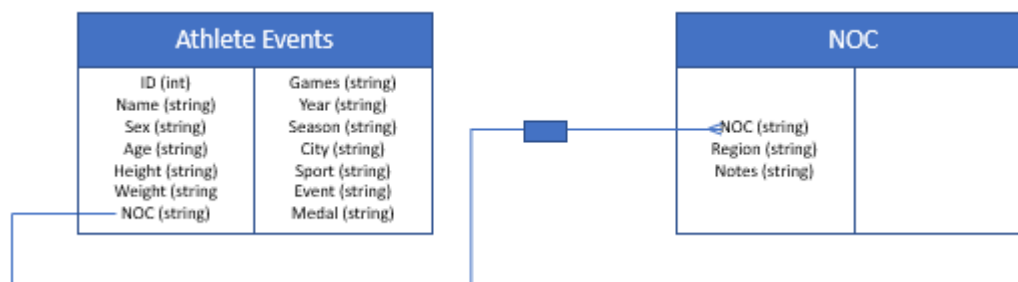
- 1) I will look at the age, weight, and height columns and will determine statistics such as average, median, and standard deviation.

- 2) I will then determine participants and medal winners had the same average metrics in most events in the 90s
- 3) In events where medal winners and participants average were not aligned, I will determine if there is a shift to those averages overtime.

Section 4: Client

- 1) SportsStats (**Olympics Dataset - 120 years of data**) has been selected. Clients for this analysis include coaches and athletes who are trying to qualify for the Olympics

Entity Relationship Diagram



Initial Findings

In my analysis I determined that medalists tended to have age, weight, height ~1% different from the mean of the group, essentially supporting my first hypothesis. The below charts have a sample of the averages for age, height, weight for medalists (columns have M label) and non-medalists from the 1990s

	Age19M	Height19M	Weight19M
Event			
Alpine Skiing Men's Combined	23.89	176.67	83.11
Alpine Skiing Men's Downhill	26.00	181.56	86.56
Alpine Skiing Men's Giant Slalom	25.89	177.67	84.00
Alpine Skiing Men's Slalom	26.00	183.33	89.00
Alpine Skiing Men's Super G	24.67	178.00	84.56
Alpine Skiing Women's Combined	23.33	167.78	63.33
Alpine Skiing Women's Downhill	23.78	170.67	66.56
Alpine Skiing Women's Giant Slalom	24.11	165.00	61.67
Alpine Skiing Women's Slalom	23.67	171.00	66.00
Alpine Skiing Women's Super G	23.11	167.33	65.56

	Age	Height	Weight
Event			
Alpine Skiing Men's Combined	24.03	179.22	82.71
Alpine Skiing Men's Downhill	25.39	179.94	83.82
Alpine Skiing Men's Giant Slalom	23.36	177.88	80.35
Alpine Skiing Men's Slalom	23.83	179.26	81.09
Alpine Skiing Men's Super G	23.71	178.58	80.84
Alpine Skiing Women's Combined	21.56	168.83	64.81
Alpine Skiing Women's Downhill	22.73	168.27	64.56
Alpine Skiing Women's Giant Slalom	22.17	166.46	62.02
Alpine Skiing Women's Slalom	22.06	168.16	63.79
Alpine Skiing Women's Super G	22.37	167.89	63.79

Several events had medalists not clustering towards the mean in the 1990s. Sports like Cycling Women's Points Race and Cycling Men's Sprint both had the mean **weight** for the medalists being closer to the 3rd quartile than the mean. Other examples like Swimming Men's 100M Freestyle and Swimming Men's 50M Freestyle had Medal winners closer to the 3rd quartile of **Height** as opposed to the mean.

	Height_diff_mean	Height_diff_quart	Height_diff_tert		Weight_diff_mean	Weight_diff_quart	Weight_diff_tert
Event				Event			
Swimming Men's 100 metres Freestyle	6.27	9.06	3.44	Shooting Women's Double Trap	22.24	37.93	9.59
Swimming Men's 50 metres Freestyle	4.58	6.97	1.72	Cycling Women's Points Race	13.59	21.84	4.69
Speed Skating Men's 1,500 metres	3.81	5.72	1.15	Cycling Men's Sprint	11.98	19.28	5.16
Speed Skating Men's 5,000 metres	3.61	6.71	0.67	Swimming Men's 100 metres Freestyle	11.67	20.83	4.36
Athletics Women's Triple Jump	3.44	5.34	1.50	Athletics Men's 400 metres	10.83	18.91	3.46
Snowboarding Women's Giant Slalom	3.31	5.98	0.48	Speed Skating Men's 1,500 metres	9.87	16.95	4.12
Shooting Men's Small-Bore Rifle, Prone, 50 metres	3.23	6.01	1.33	Alpine Skiing Men's Slalom	9.75	15.58	2.30
Wrestling Men's Super-Heavyweight, Greco-Roman	3.12	4.23	0.31	Sailing Mixed One Person Dinghy	9.70	14.04	7.33
Swimming Women's 200 metres Freestyle	3.10	5.33	0.65	Swimming Women's 50 metres Freestyle	9.10	19.66	0.58
Cycling Women's Points Race	3.02	5.14	-0.29	Swimming Women's 100 metres Freestyle	8.90	15.33	1.76

Deeper Analysis

The average of Medal Winners in the 1990s is a poorer predictor of the height, weight, and age of participants in the 2000s than the average of all participants in the 1990s.

Correlation 1) For example the average weight of medal winners in the 1990s predicted the weight of participants (4.28kg RSME) worse than the average weight of all participants in the 1990s (2.44kg RSME).

Correlation 2) Another example is that the average height of medal winners in the 1990s predicted the height of participants (3.87cm RSME) worse than the average height of all participants in the 1990s (1.37cm RSME).

Metrics I created include wms which is the degree to which a medal winner weight predicts weight in subsequent years and hms which is the degree to which a medal winner height predicts height in subsequent years. This allows me to determine how well medal winners did

Deeper Analysis (Cont'd)

What jumps is how little medal winner attributes influences future Olympic participants in these outlier cases, which is a repudiation of my hypothesis.

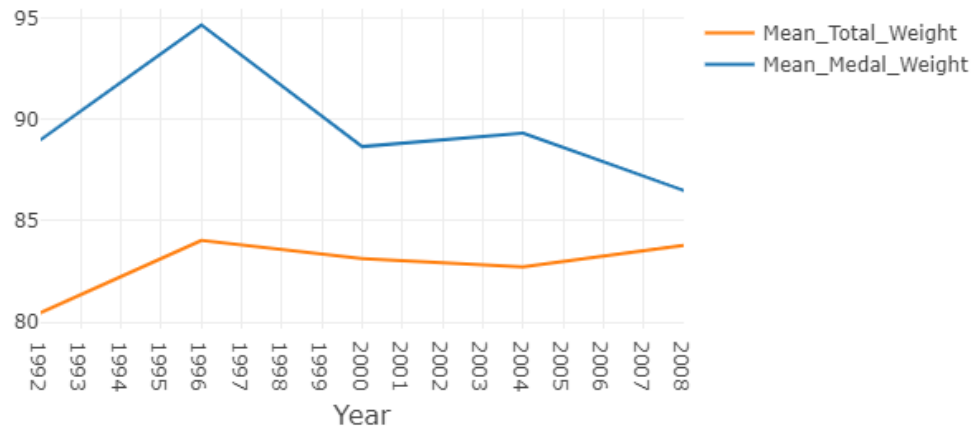
To evaluate why that is I decided to compare how much medal winners in the 1990s predicted the medal winners in 2000s on the key attributes of weight, height and age and how much the average participant in the 1990s predicts the medal winners in the 2000s. My new hypothesis is that outlier medal winners in one decade do not predict subsequent medal winners.

In terms of weight medal winners in the 1990s were a slightly better predictor of the weight of medal winners in the 2000s (4.12kg RSME) than the average participant in the 1990s (4.44 kg RSME). However, in terms of height the medal winners in the 1990s were a poorer predictor of the medalists in the 2000s (3.59cm RSME) than the average participant in the 1990s (2.96cm RSME).

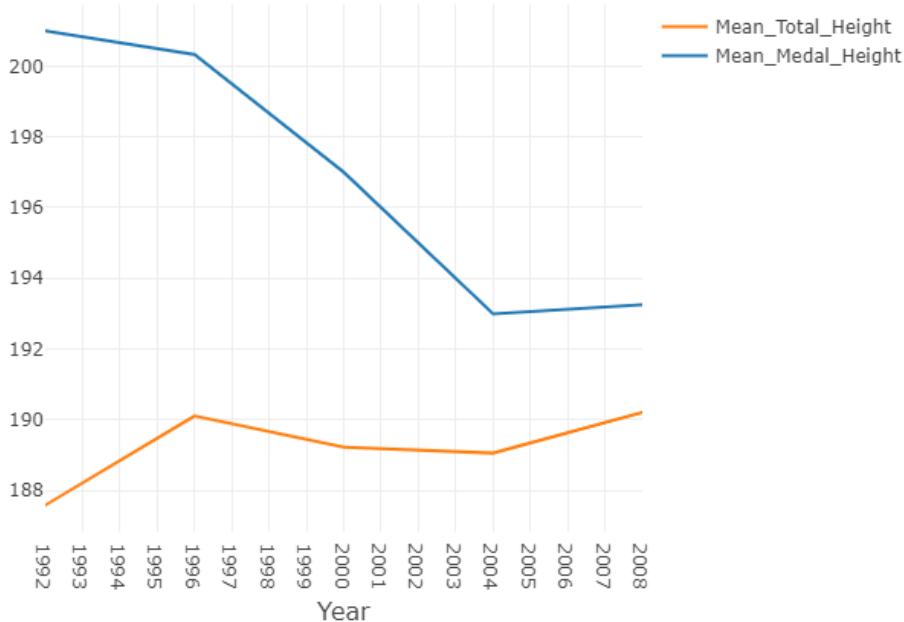
What the above tells me is that when outlier athletes win medals in a decade, that outlier athletes features do not indicate an altered trend in subsequent decades, a confirmation of my new hypothesis.

Deeper Analysis Example (Swimming Men's 100 metres Freestyle Example)

While weight was divergent between medal winners and participants, there was a shift in medal winners towards the total competitor mean in the 2000s



While height was divergent between medal winners and participants, there was a shift in medal winners towards the total competitor mean in the 2000s



Final Findings (Results of Hypothesis)

- 1) Medalists tended to have age, weight, height ~1% different from the mean of the group
- 2) The average of Medal Winners in the 1990s is a poorer predictor of the height, weight, and age of participants in the 2000s than the average of all participants in the 1990s.
- 3) When outlier athletes win medals in a decade, that outlier athletes features do not indicate an altered trend in subsequent decades

The conclusion of my analysis is that athletes and coaches should not necessarily seek to copy outlier Olympic medalists. If for example a significantly heavier athlete than normal won a competition in cycling, athletes and coaches should not prioritize increasing the weight (via muscle building or increased food consumption) to emulate those athletes. Rather they should focus on meeting the weight of the average Olympic participant.