120 Years of Olympic Athlete Data

By Tim and Emefa

Data set

- Olympic Data from Kaggle
- Most olympic games from 1896 to 2016
- Contains fields of: Name, Sex, Age, Height, Weight, Team, Games, Year, City, Event, Medal, and others.
- BMI examined
- Population is known, more or less
- Allows for confirmation of statistics (still learning this material)

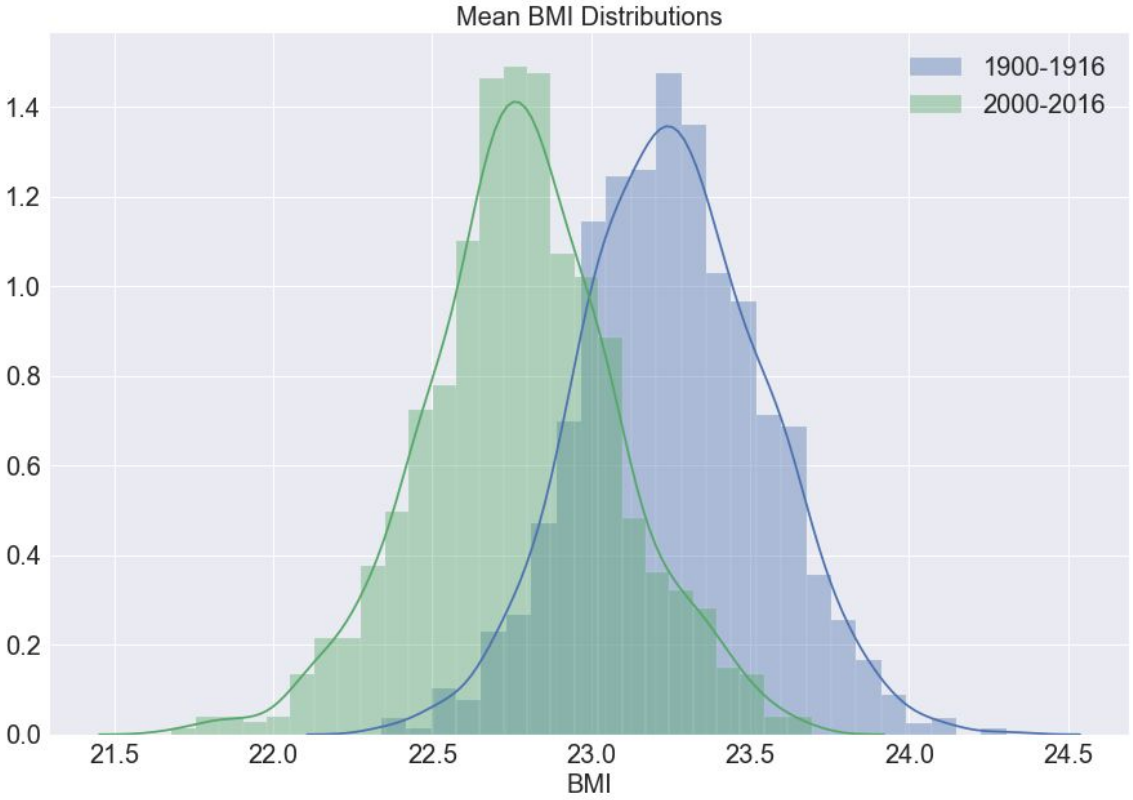
Question 1: Has the Average BMI of Athletes Changed Over Time?

- To observe this, we took mean and standard deviation of population of athletes from 1900-1916 to compare to 1 random sample of 30 athletes from 2000-2016
- H0 = No change in BMI
- H1 = BMI has changed over time
- Alpha value chosen to be standard 0.05 ($\alpha = 0.05$) to reject H0

Results

- P-Value of 0.0456
- Null Hypothesis Rejected
- Alternative Hypothesis: The BMI’s are different!
- Pop mean of 1900’s has BMI around 23.26
- Sample mean of 2000’s has BMI around 22.15
- Did athletes get smaller?

Actual Population Distributions



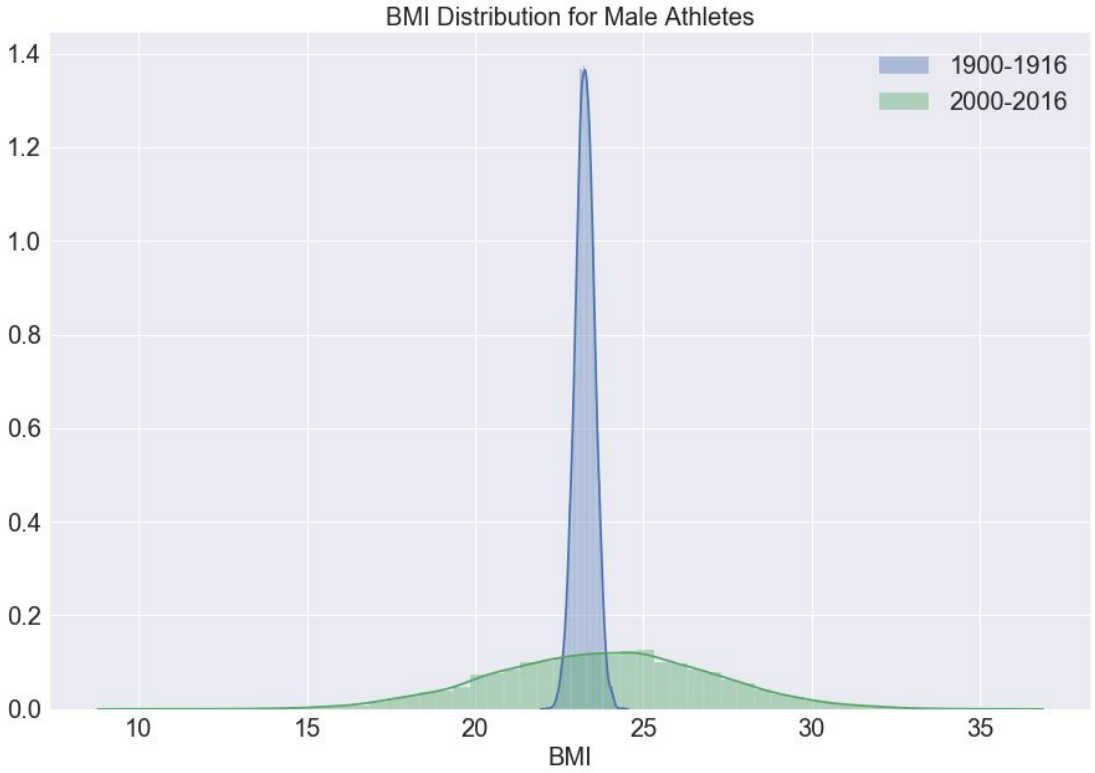
Correction needed

- Unfortunately, 1900’s data has no women in it
- First woman participated in the olympics in 1920 according to this data (1900 according to Wikipedia though)
- Comparing to the 1900-1916 male athletes to 2000-2016 male athletes by the same process with the same test
- H0 = No change in BMI
- H1 = BMI has changed over time
- Alpha value chosen to be standard 0.05 ($\alpha = 0.05$) to reject H0

Results

- P = 0.89
- Fail to reject null hypothesis
- Results suggest BMI of Olympic athletes has not changed over the 100 year time period

Actual Population Data



- Our p-value says these are likely the same distribution
- Or so close it's hard to distinguish samples from either
- Z-test performed 1000 times With sample size of 30 from 2000-2016 data only yielded 3 times in which p value was less than 0.05. That's 0.3% of the time.

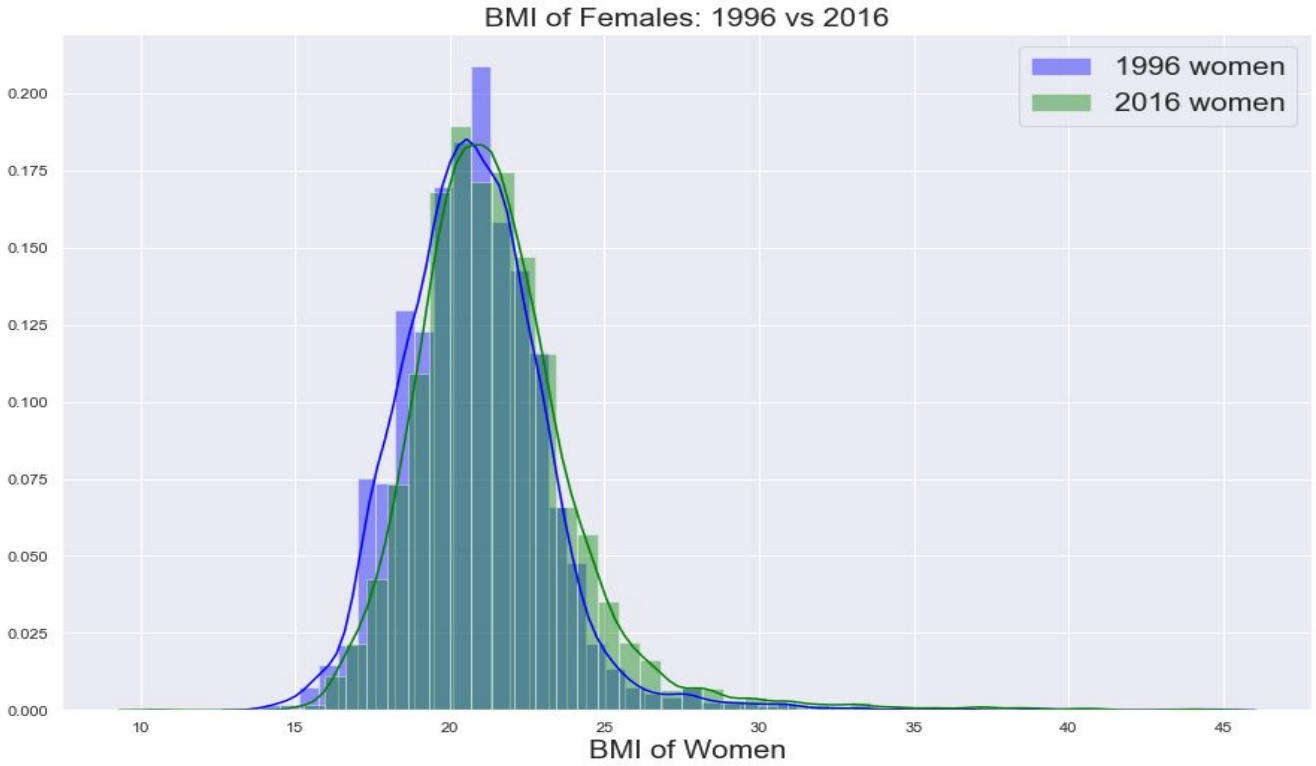
Question 3: Is there a difference in the BMI's of 1996 female participants vs 2016 female participants?

- H0 = No difference in BMI between the two groups
- H1 = There is a difference in BMI between the two groups
- Alpha value chosen to be standard 0.05 ($\alpha = 0.05$) to reject H0
- 1996 Summer Olympics was the Year of the Women..is there a difference then and now?

Results

- P-Value of 0.86
- Fail to reject Null Hypothesis
- Null Hypothesis: The difference in BMI’s are not statistically significant
- Mean BMI of 1996 female participants was around 20.79
- Mean BMI of 2016 female participants was around 21.34

Distribution of Female BMI's: 1996 vs 2016 participants



Question 4: Is there a difference in the age of 1996 female participants vs 2016 female participants?

- H0 = No difference in age between the two groups
- H1 = There is a difference in age between the two groups
- Alpha value chosen to be standard 0.05 ($\alpha = 0.05$) to reject H0

Results

- P-Value of 0.99
- Fail to reject Null Hypothesis
- Null Hypothesis: The difference in Ages are not statistically significant
- Mean Age was around 23.8 for 1996 participants
- Mean Age was around 27.4 for 2016 participants

Distribution of Female Ages: 1996 vs 2016 participants

Does the US perform better when hosting the Olympics or when abroad?

	usa_home	usa_away
Athletics	2.166667	1.370370
Swimming	2.062500	1.740741
Basketball	1.000000	0.592593
Rowing	0.458333	0.240741
Football	0.312500	0.148148

- Columns contain number of gold medals won for each sport divided by number of events competed in
- ANOVA test run to see if there is a difference in performance for the US at home and abroad based on this metric

- H0: No difference between medals won at home and abroad
- Ha: There is a change
- $\alpha = 0.05$

ANOVA Results

- F_critical for this data was 5.79
- Our F stat is 6.29
- Null hypothesis rejected!
- This supports the idea that hosting the olympics may help improve performance in the events

Thank you

Kaggle dataset for anyone interested:

https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results