

COMP3125: Final Project
Boston Marathon Winners Analysis

Mary Witt

December 10, 2023

INTRODUCTION

The Boston Marathon, the world's oldest marathon, is a prominent race that runners from all over the world travel afar to compete in. This 26.2-mile race started in 1897 and has taken place almost every year since then. There are two separate datasets that I found on Kaggle displaying the winners for each of the Boston Marathons. Through my research, I wanted to study what rate the winning times have decreased at for the Boston Marathon. I also was eager to compare the rate of times at which male winners and female winners have decreased. I also wanted to determine whether or not marathon times will continue to decrease.

SELECTION OF DATA

I found two datasets on Kaggle regarding Boston Marathon Winners. One dataset contains the data for all the male winners of the Boston Marathon from 1897 to 2023, and the other dataset contains the data for all the female winners of the Boston Marathon from 1966 to 2023. The datasets include the name of the winner, they year in which he/she won, the country he/she is from, his/her finishing time, and the distance of the race in both miles and kilometers. The data from the datasets can be found on the Boston Athletic Association's website. When examining the data, I decided to remove the rows for 1918 and 2020 due to the lack of information for these years. In 1918, the Boston Marathon was cancelled due to America's involvement in World War I. The Boston Marathon was also cancelled in 2020 due to the COVID-19 pandemic. I also added the men's winner and the women's winner from the 2023 Boston Marathon to the datasets since that data was not already in the datasets.

METHODS

For my research, I wrote python code in Jupyter notebook to create linear regression models to answer these questions. I used SciKit to create Linear Regression models. I used NumPy to add a new axes to arrays as well as to convert 1D arrays into a 2D column vectors. I also used Matplotlib to plot scatterplots. I used Pandas to read the data from the datasets which were csv files.

RESULTS

Question: On average, at what rate has the winning times for the Boston Marathon decreased over time?

Answer: When creating the scatterplot for the winning finishing times for the men's Boston Marathon from 1897 to 2023, the scatterplot displays a somewhat linear relationship. The slope of the line of best fit is around -16.8 seconds. On average, each year, the male winner finishes 16.8 seconds faster than the winner the year prior. The winning times for the Boston Marathon has steadily decreased since the first Boston Marathon in 1897.



Question: Have men's winners or women's winners had more of a significant decrease in marathon times over the years?

Answer: The first female winner's time was recorded in 1966, so I created a scatterplot for the winning times of every female winner from 1966 to 2023. I noticed the scatterplot did not quite have a linear relationship for the entire timeframe. The scatterplot has more of a polynomial regression relationship. The winning times did have a steady decrease from 1966 to around 1984. From 1985 and onward, the winning times remained fairly constant, and there was no longer a steady decrease in winning times. Due to the nonlinear relationship of the women's winning times, a line of best fit is not the best when predicting marathon winning times for future years. However, when calculating a line of best fit, the slope was around -47.4 seconds. If this were a linear model, this would mean that from 1966 to 2023, the female winner finished approximately 47.4 seconds faster the female winner the year prior.



In order to compare the men's winning times to the women's winning times, I created a scatterplot for the men's winning times from 1966 to 2023. The scatterplot ended up having a somewhat linear relationship and was especially more linear than the women's scatterplot. The slope for the men's winning finishing times was around -7.1 seconds. Thus, on average, the male winner finishes around 7.1 seconds faster than the previous year's male winner.

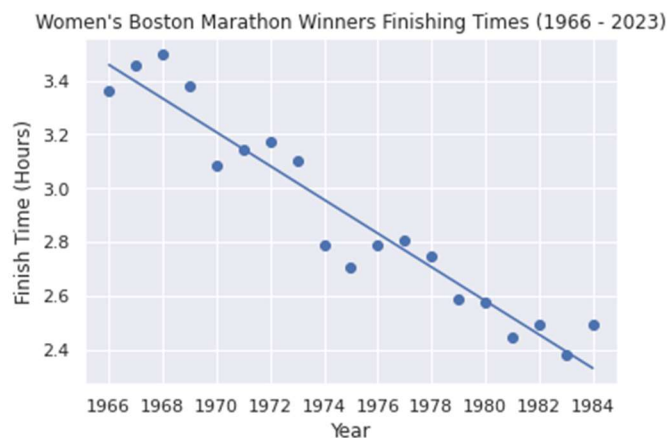


Thus, starting from 1966, the women's winning finishing times had a greater decrease when comparing to the men's winning finisher times.

Question: Will winning times continue to decrease in the future, or will they begin to flatten out and become more uniform?

Answer: For this question, I decided to create linear regression models for two separate timeframes for the women's winning finishing times. I created a scatterplot for the women's time from the years 1966 to 1984, and then a separate scatterplot for the years 1985 to 2023.

The scatterplot displaying the female winning finishing times from 1966 to 1984 has a fairly strong linear relationship. The line of best fit has a slope of -225.6 seconds. Thus, on average, the female winner for each year is approximately 225.6 seconds faster than the winner the year prior.



The scatterplot displaying the female winning finishing times from 1985 to 2023 also has a linear relationship. The line of best fit has a slope of 0.1 seconds. Thus, on average, the female winner for each year from 1985 to 2023 is around 0.1 seconds faster than the previous year's winner.



As shown through the drastic differences in slopes, the winning times have become more constant and uniform since 1985 and will likely not show any significant decrease in the future.

DISCUSSION

The Boston Marathon winning times steadily decreased throughout the early years of the race. Since then, however, winning times had become much faster as runners began to train more seriously and develop the necessary skills on how to run fast for an extended period of time. However, more recent years show that the winning times have remained fairly constant. In the future, the winning times for both males and females will stay fairly constant as it is extremely difficult to run much faster than the recent winning times. Over the past forty years, the men's winners for the Boston Marathon have had times ranging from 123 minutes to 135 minutes. It is extremely difficult to run a marathon in that time range, which explains why the winning times for both men and women will continue to remain steady for future Boston Marathons.