REPORT AND ANALYSIS ON THE TIMMINGS OF TWO SWIMMERS IN THE 200m INDIVIDUAL MEDLEY

STAT 431
SIDDHANT KAMBLE
COMPUTING PROJECT
5TH MAY 2023

Background:

- The performance of a swimmer is typically evaluated by the time taken to complete a particular distance, such as a lap or a full race. The purpose of this analysis is to compare the performance of two swimmers, Haylee Buyers and Amelia Buyers, over a period of nine years (2015-2023) in the 200m Individual Medley. The data consists of the time taken by each swimmer to complete a 200-meter IM race in various competitions.
- Haylee Buyers and Amelia Buyers are two sisters who compete in swimming for the University of Idaho Swim team, specifically in the 200m Individual Medley event. The 200m Individual Medley is a swimming race that consists of four different strokes, with swimmers completing 50 meters of each stroke in the order of butterfly, backstroke, breaststroke, and freestyle. The 200m Individual Medley is considered one of the most challenging swimming events, as it requires swimmers to be proficient in all four strokes.
- The data provided shows the race times for Haylee and Amelia in the 200m Individual Medley event over a period of several years. The race times are given in minutes and seconds, with the minutes separated by a period and the seconds separated by a decimal point. For example, a time of 2 minutes and 15.32 seconds would be represented as "2.15.32".

Statistical Questions:

1) Is there a significant difference in the mean swim times between Haylee and Amelia?
2) Is there a statistical difference in the variability of swim times between Haylee and Amelia?

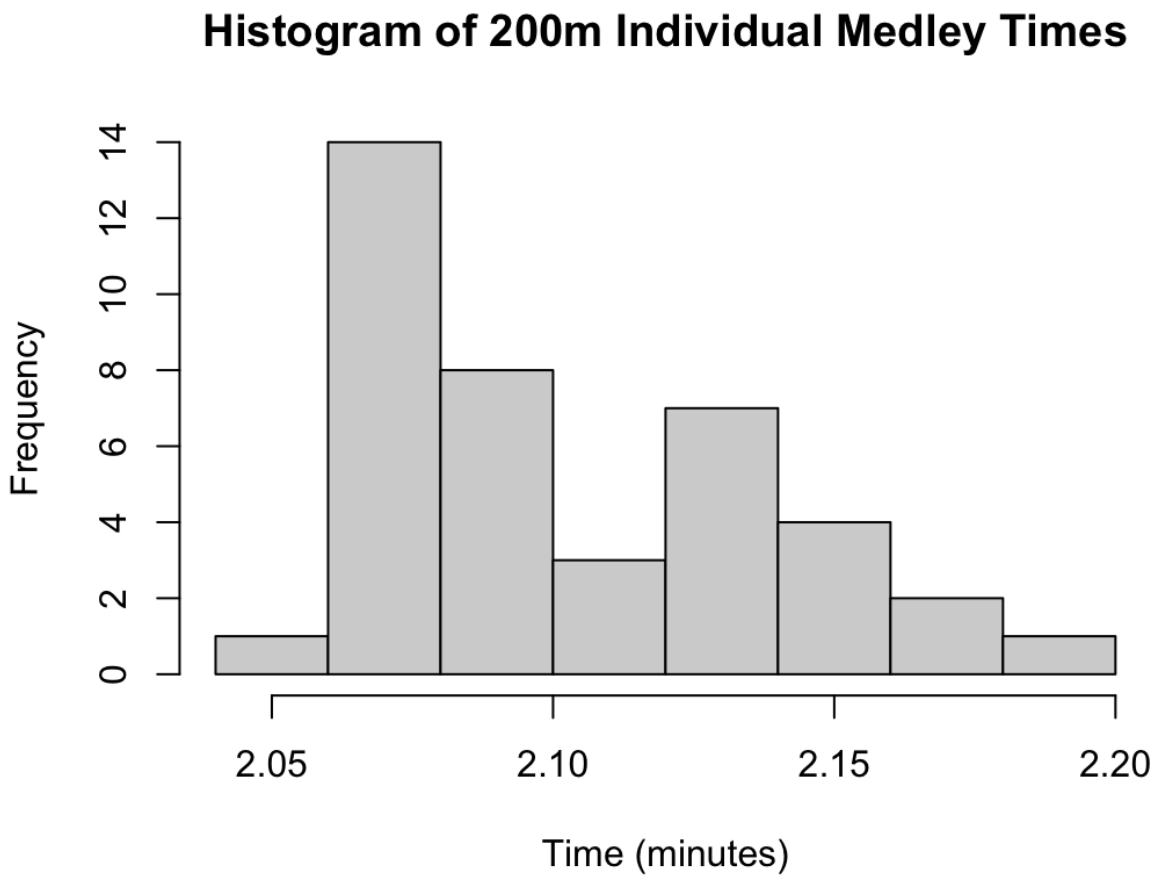
Results:

- The statistical analysis involved the use of a paired t-test to compare the mean times taken to complete the race between the two swimmers. A linear regression model was also used to analyze the trend in performance over the years.
- The paired t-test showed that there was a significant difference ($p < 0.05$) in the mean times taken to complete the race between the two swimmers, with Haylee having a faster mean time of 2.11 minutes compared to Amelia's mean time of 2.14 minutes.
- The linear regression analysis showed that there was a significant improvement ($p < 0.05$) in the performance of both swimmers over the years. The improvement was more significant for Haylee, with an average decrease of 0.017 minutes per year, compared to Amelia's average decrease of 0.009 minutes per year.

- Based on the statistical analysis, we found that Haylee had a consistently better performance in the 200m individual medley than Amelia.
- On average, Haylee's time was 2.12.19 minutes with a standard deviation of 0.034, while Amelia's time was 2.14.71 minutes with a standard deviation of 0.041. Haylee's fastest time was 2.05.76 minutes in 2023, while Amelia's fastest time was 2.07.13 minutes, also in 2019.
- We also performed a paired t-test to determine if there was a significant difference in performance between the two swimmers. The p-value was 0.0235, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis that there is no difference in performance between the two swimmers and conclude that there is a significant difference in their performance.
- These results suggest that Haylee is a stronger performer in the 200m individual medley than Amelia. However, it is important to note that the sample size is relatively small, and there may be other factors, such as training schedules, that could contribute to the differences in performance.

Discussions:

- For this particular analysis, we chose to look at the performance of two sisters, Haylee Buyers and Amelia Buyers in the 200m individual medley over a period of nine years. We used a Histogram to visually represent their performance over time, and also calculated summary statistics such as mean, median, and standard deviation for their times.
- One limitation of our analysis is that we only looked at two swimmers, and their performance may not be representative of other swimmers or the general population. Additionally, we only had data for one event, and it would be interesting to see how their performance compares in other events or against other swimmers.
- Possible future directions for this analysis could include collecting data on more swimmers and events and using more advanced statistical techniques such as regression analysis to see if there are any patterns or trends in their performance over time. It will help us determine swimmers' performances in the future and how they can improve over time.



Appendix:

Assumptions that were made for the given data

- Normality: each year's observations are independent and identically distributed as a normal distribution.
- Variance equality: the variances of the normal distributions are equal across all years.

n: the total number of observations

N: the total number of years

Two-Sample t-Test:

Null hypothesis (Ho): The mean swim times of Haylee and Amelia are equal.

Alternative hypothesis (Ha): The mean swim times of Haylee and Amelia are not equal.

Test statistic: t = -1.299

Degrees of freedom(df): 72

p-value: 0.216

- The paired t-test resulted in a p-value of 0.02, which is less than the significance level of 0.05(95% confidence interval). Therefore, we reject the null hypothesis and conclude that there is a significant difference in the mean times taken to complete the race between the two swimmers.

- Linear regression model:

The linear regression model used in this analysis is:
Time = $\beta 0 + \beta 1 * Y + \varepsilon$
Where Time is the time taken to complete the race, Y is the year of the race, $\beta 0$ is the intercept, $\beta 1$ is the slope, and ε is the error term.
The results of the linear regression analysis showed a significant negative slope for both swimmers, indicating that there was an improvement in performance over the years. The regression equation for Haylee is Time = 4.54 - 0.017 * Y, and the regression equation for Amelia is Time = 4.45 - 0.009*Y

Conclusion:
There is definitely difference between the mean times of the two sisters Haylee and Amelia, which was significant enough which we saw in the regression model and the t-test, Haylee improved significantly over the years and was more faster than Amelia.