Swim Analytica

Providing data-driven insights to enhance swimming performance.

SINGAPORE MANAGEMENT UNIVERSITY

1. INTRODUCTION

The purpose of this application is to enable swimmers and coaches to incorporate a data-driven approach into their training program. Users can utilize analytical tools embedded within the dashboard to gain insights that can help enhance competition performance.

2. METHODOLOGY

The dataset used in the application is taken from swimming events during the *Tokyo Summer Olympics 2020*. It contains *reaction time, speed* and *split times* for each *round of each event,* and swimmer data such as *nationality* and *gender*.

The application visualizes the **exploratory & confirmatory analysis** of speed, reaction time, split times and performance over rounds, and constructs a **regression model** which predicts the athlete swim time in finals.

All the dashboards in the application are interactive and users can adjust statistical models to obtain customized insights. R programming was used through data processing and application development. R packages used include ggiraph, tidyverse, ggplot2, ggstatsplot, performance, gapminder and patchwork.

3. RESULTS

AVERAGE SPEED

The application enables users to explore the swimmer data to find insights, formulate the hypothesis, and perform statistical test to confirm their hypothesis. Here is the hypothesis that could be validate through the app:
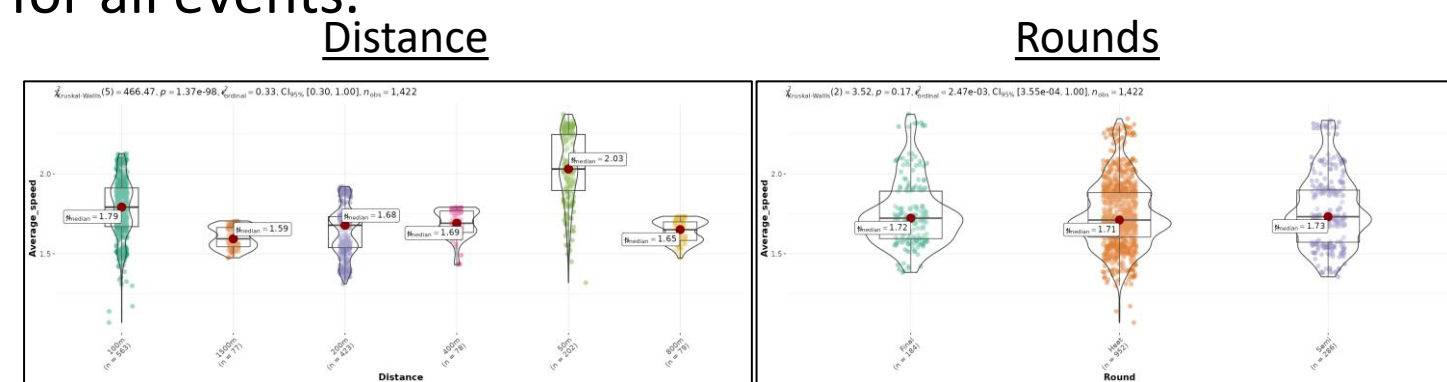
"Is there a significant difference in average speed for different strokes/gender/distance/rounds?"

All events / Mens' Freestyle
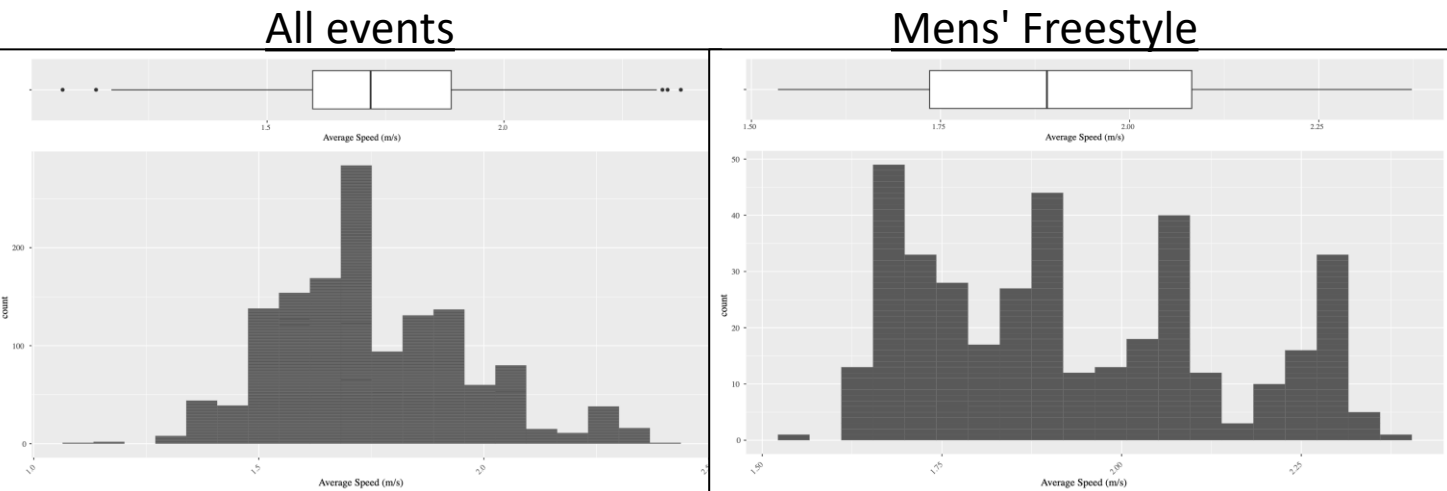
Although the average speed for all events appears to be normally distributed, most of the specific events are either irregular or multimodal, and therefore we choose nonparametric for statistical test.
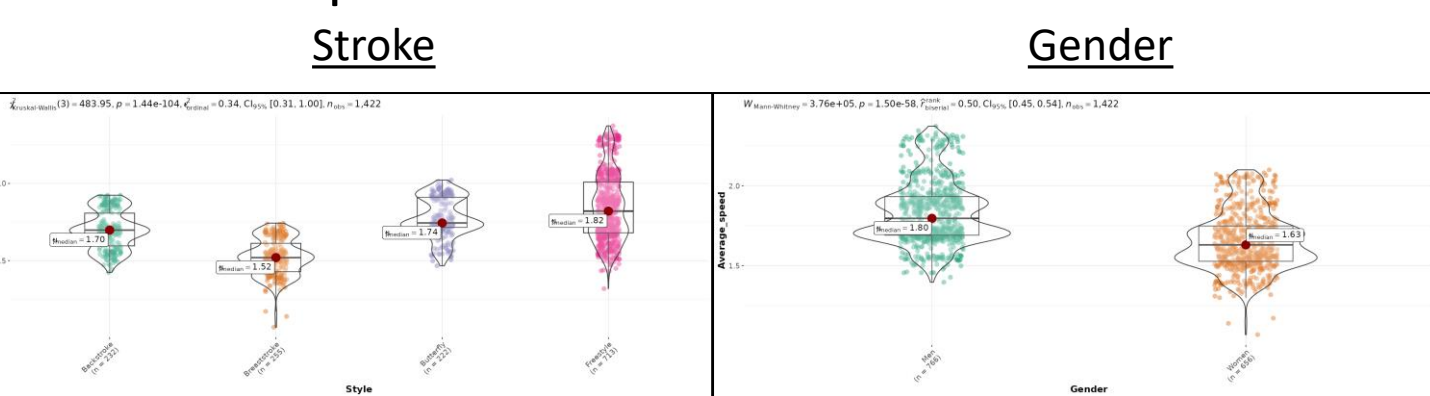
Stroke / Gender

One-way ANOVA testing showed significant difference in the median average speeds for stroke and gender. Freestyle has the highest median swim speed and Breaststroke has the lowest. Male swimmer also has a higher median average speed than women for all events.


Distance / Rounds

There are also significant difference in median average speed for different distance events. Shorter the distance, faster the average speed. On a contrary, there is no significant difference in average swimming speed between different rounds.

REACTION SPEED

The application also enables users to perform similar analysis on Reaction Speed (s). Through the app, the user can confirm the hypothesis on whether different factors has significant impacts to reaction time.

According to the correlation analysis, there is a significant negative relationship between reaction time(s) and average speed(m/s), which means shorter the reaction time, higher the average swimming speed.


Gender / Distance
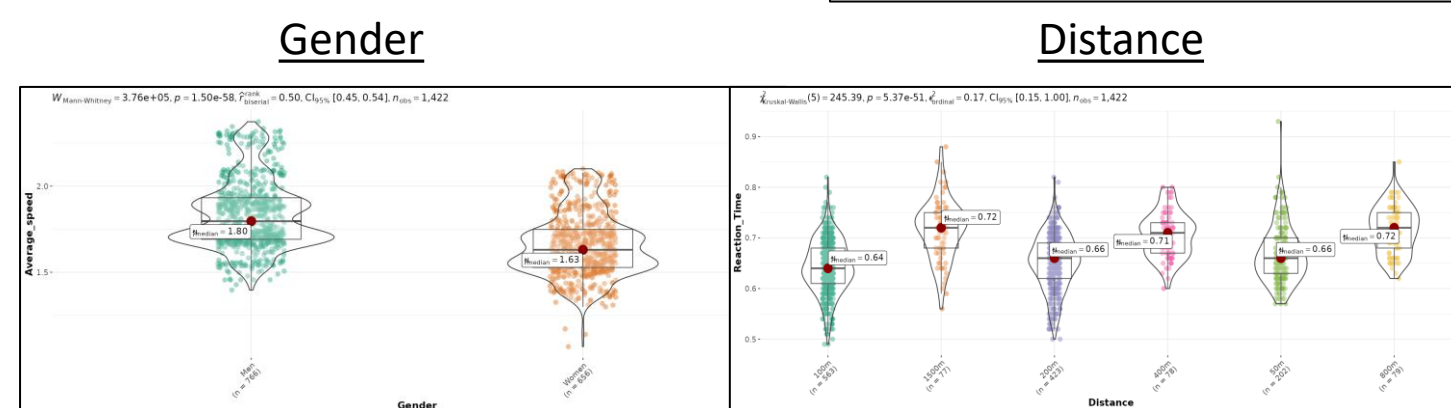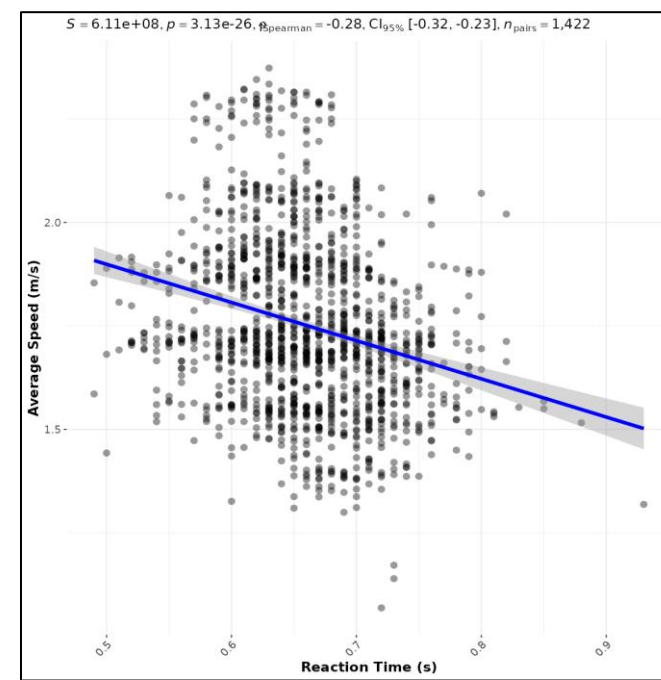
One-way ANOVA testing shows there is a significant difference between male and female swimmers in terms of reaction time, with Men swimmer having a lowest median reaction time. There is also a significant difference in reaction time among difference distances

SPLIT TIMES

"Negative splitting" is an established strategy where the second half of a swim is faster than its first. Going too fast at the start can cause lactate fatigue, which may result in slower performance overall. Coaches closely monitor split times for this reason. This app allows a coach to visualize his swimmers' split performance, examine if negative splitting works, and identify other pacing strategies.
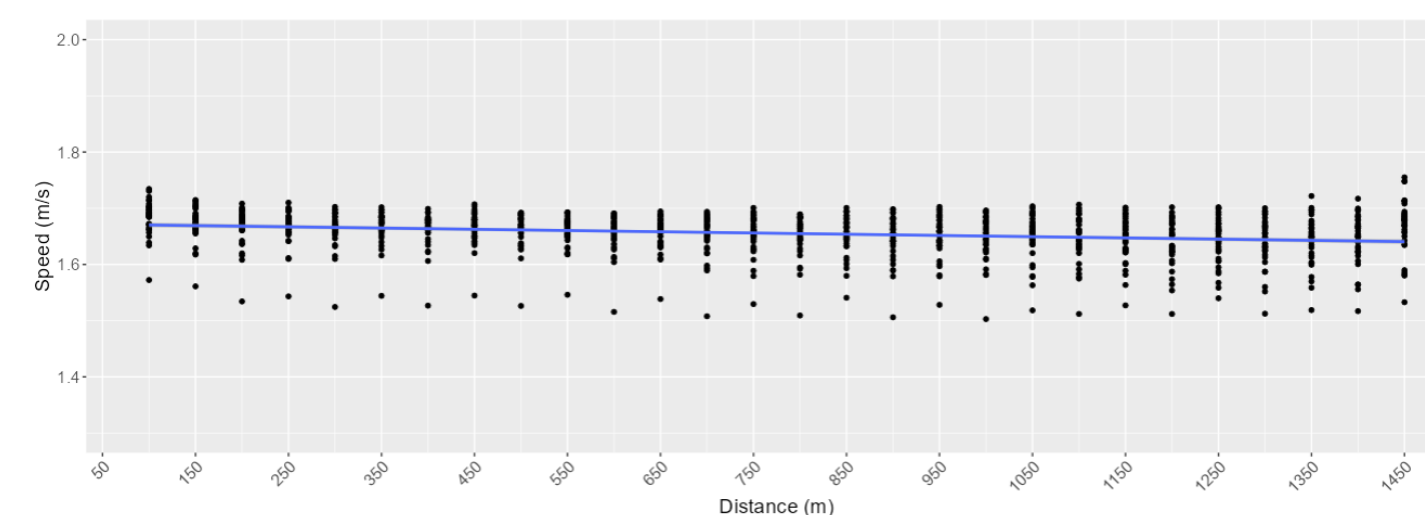
Pearson's correlation test results show very significant but relatively weak negative correlation between speed and distance.
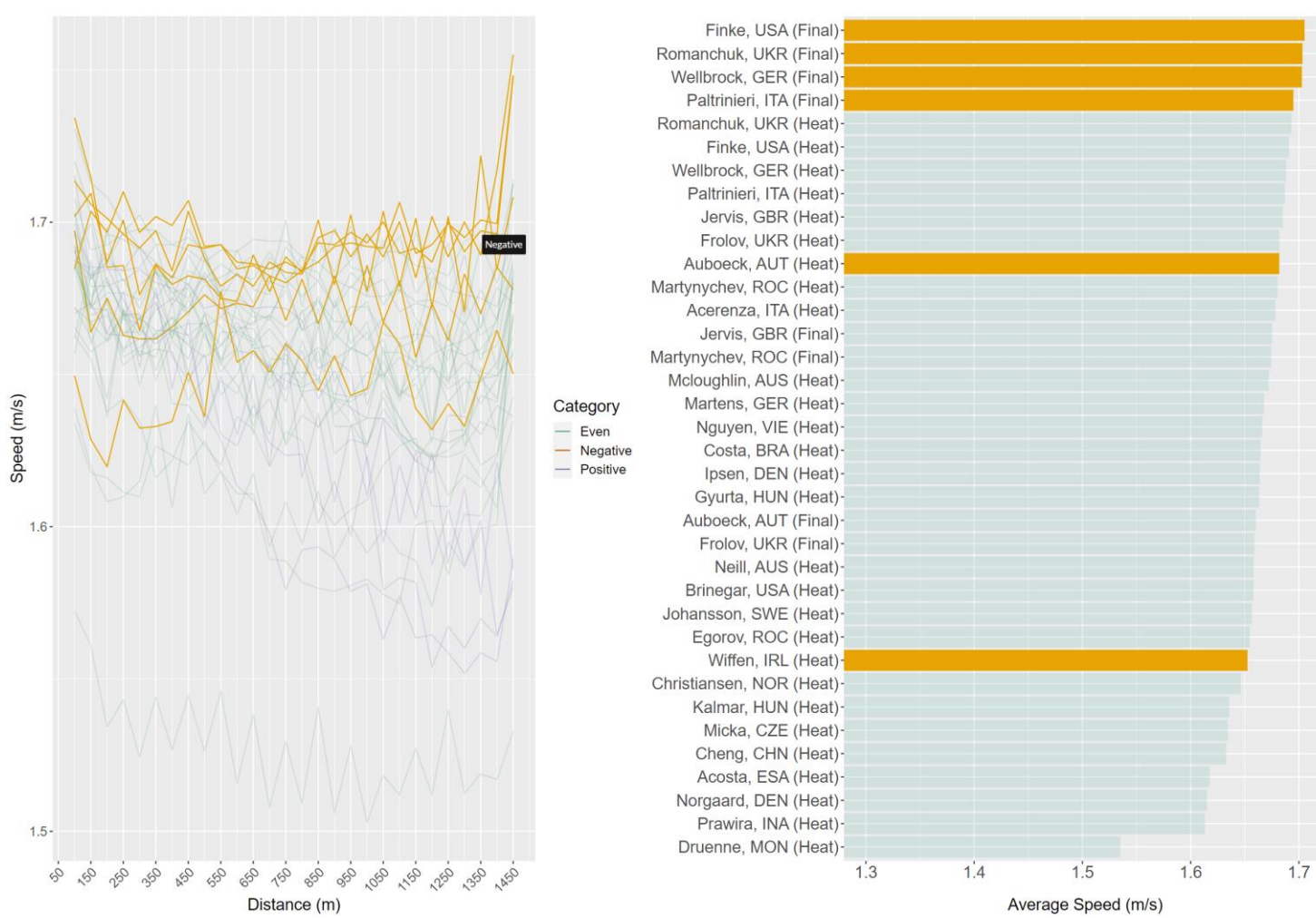

In the 1500m men's freestyle, the 4 swimmers with fastest overall time showed negative splitting. The remaining 6 of the top 10 showed an even pace, and most of the slower swimmers were slower in the second half.

Median average speed by category
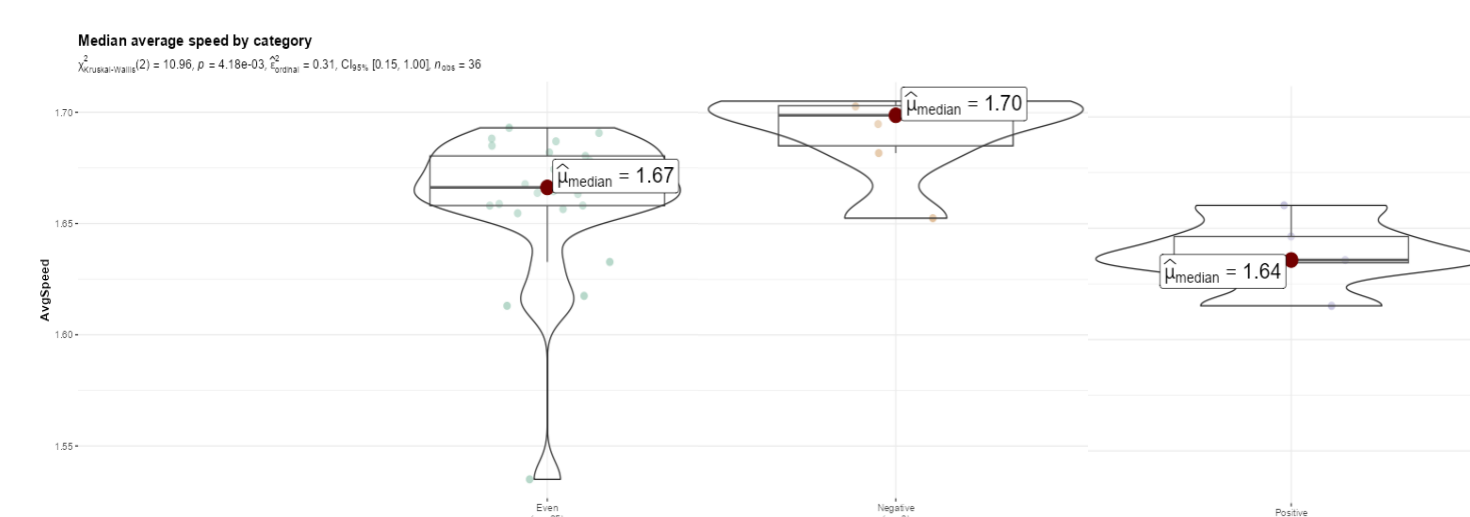
One-way ANOVA testing showed significant difference in the median average speeds for negative splitting (1.7m/s) , even pacing (1.67m/s) and positive splitting (1.64m/s). The use of hierarchical clustering did not reveal alternative pacing strategies with as significant differences in median average speeds.

PERFORMANCE OVER ROUNDS

On average, as indicated by the app's cumulative distribution function plot of the swimmer's relative % improvement between rounds, 75% of swimmers show an improvement in their swim times from the Heats to the Finals, while 61% of swimmers exhibit an improvement in their timings from the Semis to the Finals.
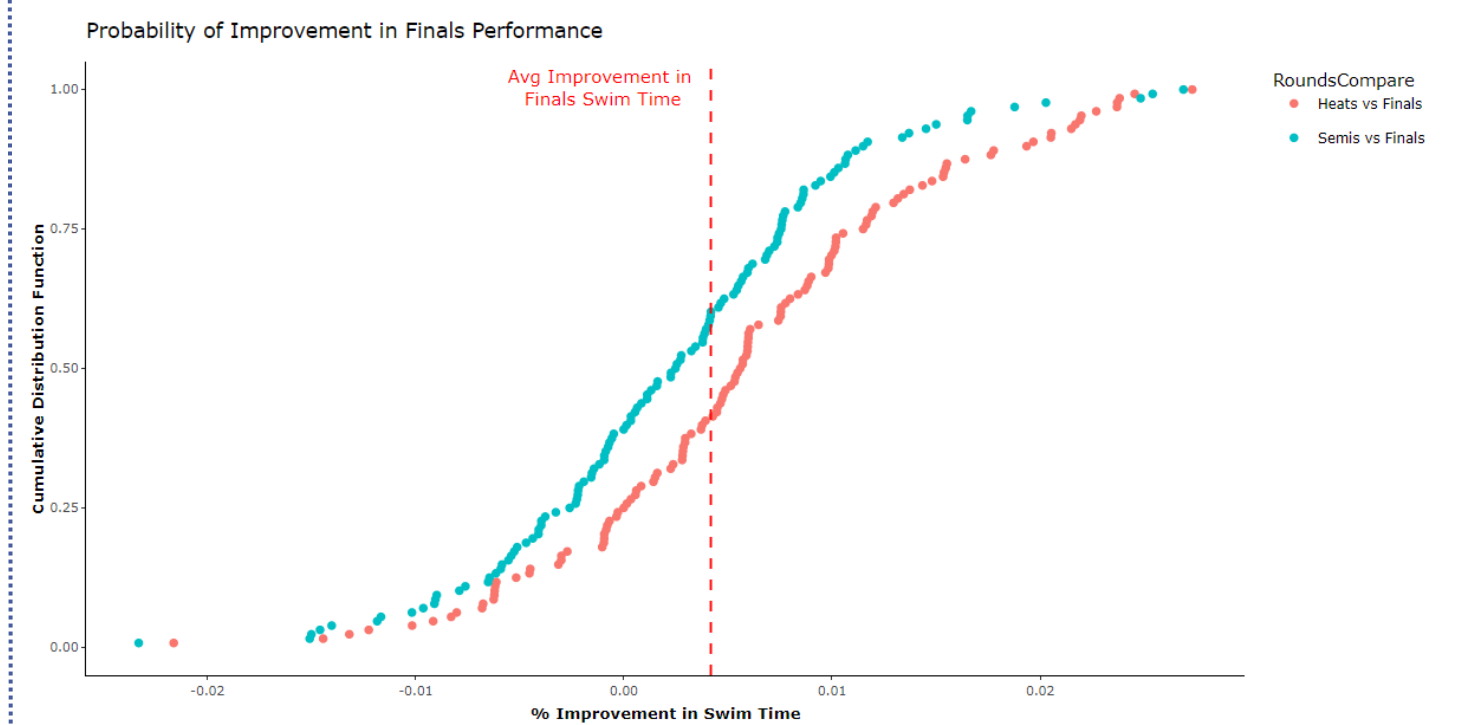

PREDICTORS OF SWIM TIME IN FINALS

The application enables users to construct a regression model that is appropriate for determining the predictors of an athlete's swim time in the finals (the target variable), utilizing the available independent variables. In order to ensure the accuracy of the regression model, certain variables, namely OlypQualTime, OlypRecord, AvgTime, and Heat, have been excluded from consideration due to significant levels of multicollinearity.
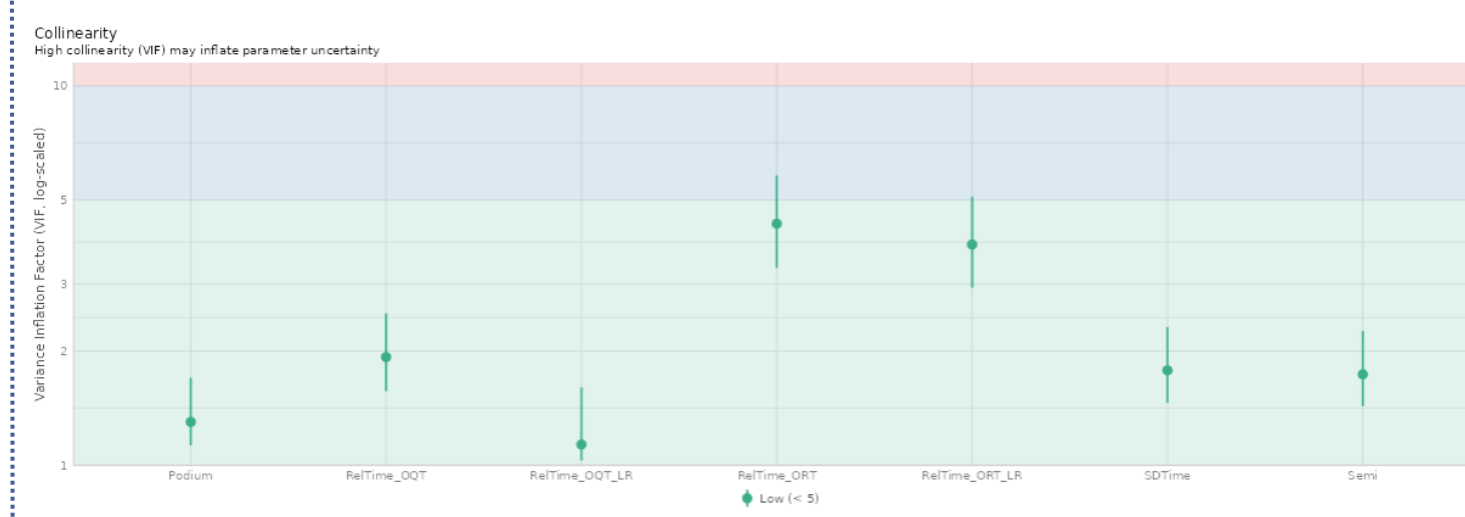
According to the results of the regression model, if all other variables remain constant, the predicted final swim time of a swimmer is more strongly influenced by certain factors, specifically: the swimmer's relative time in comparison to the Olympic Qualifying Time, the swimmer's relative time in comparison to the Olympic Record Time, and whether or not the swimmer has placed in the top three in the given event.

Regression Analysis / Regression Parameters Plot
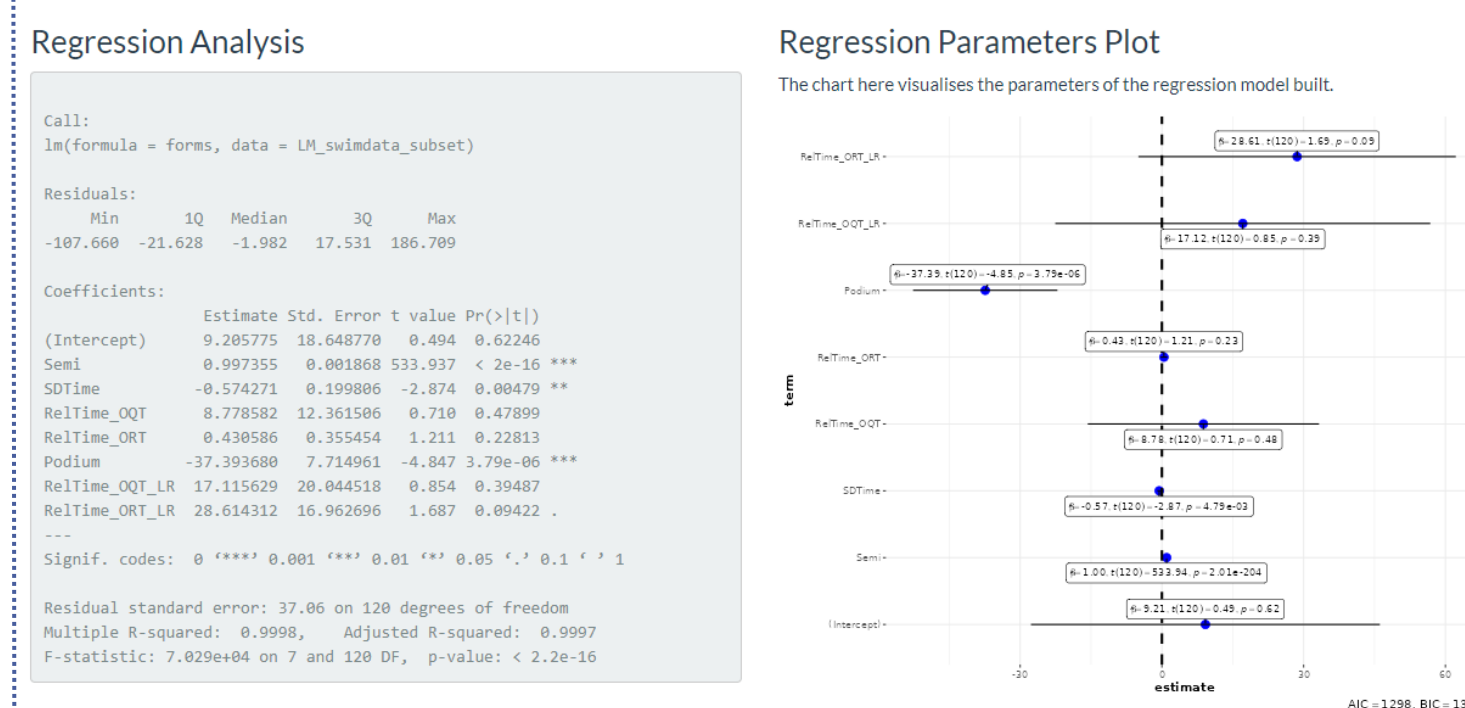
4. CONCLUSION & FUTURE WORK

The results generated from this application are primarily constrained to the available dataset, which includes information on swim times, reaction time, and competition rounds.

To develop a more comprehensive analytical model, it would be beneficial to incorporate additional attributes into the dataset, such as Swimmer characteristics, Swimming strokes, Pool conditions, and performance from previous rounds or training sessions. This approach may reveal more factors affecting swimmer performance and thus improve statistical significance of confirmatory analyses and regression model results.

ISSS608 Group 11:
Farrah Mohd Fadil | Cheryl Jeanne Chiew | Law Man Long
Advisor: Prof KAM Tin Seong