**Analyzing Olympic Swim Times with Boxplots and Confidence Intervals**

The Summer Olympics are traditionally held every four years, with an extensive history that goes back to 1896. Since then, swimming has been an integral part of the events. The sport has attracted thousands of athletes from over 150 countries around the world. The Olympics allows nations to put aside their differences and come together to provide a sense of unity to millions of fans.

We will be investigating the results of the 100-meter men and women races from 1924 to 2020. The dataset includes 606 swimmers and 10 variables.

Our main goals are to find which styles of swimming are the best for the 100m race and whether race times have changed significantly from 1924 to 2020. Each style of swimming has its own advantages and disadvantages. The backstroke makes breathing easier but makes it difficult for the swimmer to know where he is going. The breaststroke allows the swimmer’s head to be out of the water for the longest amount of time, however it requires the most strength and endurance of the four. The butterfly stroke, like the breaststroke, makes breathing easier because the swimmer’s head spends significant time out of the water, however it requires lots of strength and timing. The freestyle stroke uses your full body and is easiest to learn, however the swimmer keeps their head in the water most of the time so it can be difficult to breathe.

The side-by-side boxplots below display race times for different strokes in the 100m event. Use it to answer the questions below.

A diagram of different styles of a race

Description automatically generated with medium confidence

1. Interpret the side-by-side boxplots above. Be sure to compare medians and note any skew, outliers, and other interesting features. What swimming styles seem to be the quickest?

Use the boxplot below to answer the following 2 questions.

A graph with different styles and colors

Description automatically generated with medium confidence

1. How do the older results compare to the more recent results from the 100m event? What changed and what remained similar?
2. How do these results compare to the overall results on the first page?

A graph showing a number of swimming times

Description automatically generated

1. Look at the boxplot above. Brainstorm some ideas on whether the 100m freestyle swim times got significantly faster from the early years of the event (1924-1972) to the recent years (1976-2020).
2. Let’s find confidence intervals to examine these ideas. The mean finish time for the recent Olympics is 52.41 seconds with standard deviation 3.27 and 108 recorded times. For earlier Olympics the mean finish time is 56.28 seconds with standard deviation 2.47 with 42 recorded times. Conduct a difference in means 95% confidence interval to determine how much faster swimmers in recent years are than swimmers in earlier years. You can assume all conditions are met. Write out an interpretation for your interval and if the difference is statistically significant.

1. Now let’s look at the strokes, or styles, of swimming. The two fastest strokes were consistently the butterfly and the freestyle. The mean finishing time for 157 freestyle finishers was 53.38 seconds with standard deviation 3.64. For 93 butterfly swimmers the mean was 55.08 seconds with standard deviation 2.78. Conduct a difference in means confidence interval to determine if there is enough evidence to conclude that Olympic freestyle swimmers are faster than butterfly swimmers in the 100m race. You can assume all conditions are met. Make sure to write out an interpretation of your interval and if the difference is statistically significant.
2. Are there any surprises in your confidence intervals or is this what you expected?