**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 Women’s 100-meter race from 1964 to 2020. The dataset includes 194 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 1964 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 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 Women’s 100m event. Use it to answer the questions below.

A graph with different styles on each side

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

1. How do the older results compare to the more recent results from the 100m event? What changed and what remained similar? The “early years” are from 1964 to 1988 and the “recent years” are from 1992 to 2020. Note that the data only contains the backstroke from 2008 to 2020.

A graph with black lines

Description automatically generated

1. Look at the boxplot above. Brainstorm some ideas on why the 100m freestyle swim times got significantly faster from the early years of the event (1964-1988) to the recent years (1992-2020).
2. Assume this is a random sample. Let’s find confidence intervals to examine these ideas. The mean finish time in the Women’s 100m freestyle event for the recent Olympics is 54.15 seconds with standard deviation 1.1 and 64 recorded times. For earlier Olympics the mean finish time is 57.06 seconds with standard deviation 1.43 with 40 recorded times. Construct a difference in means 95% confidence interval to determine how much faster swimmers in recent years are, on average, 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 are the butterfly and the freestyle. The mean finishing time for 104 freestyle finishers was 55.27 seconds with standard deviation 1.88. For 61 butterfly swimmers the mean was 57.9 seconds with standard deviation 1.24. Construct a difference in means confidence interval to determine how much faster freestyle swimmers are, on average, 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. Why might it be a problem if we did these problems with both the men and women in the data?