**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 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.



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?

The boxplots have approximately no skew, however the freestyle boxplot has a long tail towards the higher times. There are two outliers toward lower times under the backstroke category. The freestyle has the lowest median time of the three styles, with butterfly and backstroke following. Freestyle has the highest range and IQR, while breaststroke has the lowest range and IQR. The freestyle stroke contains the fastest times.

Use the boxplot below to answer the following 2 questions.

The “early years” are from 1964 to 1988 and the “recent years” are from 1992 to 2020. The recorded times for the Woman’s backstroke started in 2008 so there is no boxplot for the “early” years in that event.



1. How do the older results compare to the more recent results from the 100m event? What changed and what remained similar?

The median freestyle time during the early years remained quicker than the median butterfly time during the recent years. Also, the introduction of the backstroke during the more recent years may have affected how the other styles were used and practiced by the Olympians. Although the butterfly stroke is quicker than the backstroke in recent years, in the earlier years it is slower than the backstroke, which may mean there were improvements in the technique over the years. Overall, both the freestyle and butterfly strokes got quicker as the years went on.

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).

Responses will vary. Maybe mention the medians and quartiles and how recent times are quicker.

1. Assume this is a random sample of elite swimmers’ times during Olympic years. 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.

x̄1 = 54.15, S1 = 1.1, n1 = 64,x̄2 = 57.06, S2 = 1.43, n2 = 40

95% CI: $(54.15-57.06)\pm 2.023\*\sqrt{\frac{1.1^{2}}{64}+ \frac{1.43^{2}}{40}}$

95% CI: -2.91 $\pm $ .535

95% CI: (-3.445, -2.375)

Interpretation: With 95% confidence the mean Women’s 100m freestyle time for all recent women Olympic swimmers is between 3.445 and 2.375 seconds faster than all earlier women Olympic swimmers. The difference in means is statistically significant because the interval does not contain 0.

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.

x̄1 = 55.27, S1 = 1.88, n1 = 104,x̄2 = 57.9, S2 = 1.24, n2 = 61

95% CI: $(55.27-57.9)\pm 2\*\sqrt{\frac{1.88^{2}}{104}+ \frac{1.24^{2}}{61}}$

95% CI: -2.63 $\pm $ .487

95% CI: (-3.117, -2.143)

Interpretation: With 95% confidence the mean Women 100m freestyle time for all freestyle women Olympic swimmers is between 3.117 and 2.143 seconds faster than all 100m butterfly women Olympic swimmers. The difference in means is statistically significant because the interval does not contain 0.

1. Why might it be a problem if we did these problems with both the men and women in the data?

Sample Answer: The boxplots don’t show the grouping of the data, just the five number summary and outliers. Since men generally swim faster than women in the Olympics, it creates two different groups of times within the data that the boxplots don’t display which would make them misleading.