Short Track Speed Skating

Short Track Speed Skating is a race conducted on a 111m circuit with 4-6 racers competing for the win. There are several high-profile events during the speed skating season at the World Championships, World Cup, European Championships, and Olympic Games. This data set looks at 500m races where athletes try to complete 4.5 laps ahead of their competitors. The first lap of each race is half of a revolution around the track, followed by 4 full laps. Athletes tend to be crowded together competing for space on the track on every turn leading to exciting races and dangerous crashes.

The data set we will be investigating has 5125 rows where each row is a race performance from an athlete at a high-profile race. In addition to event and personal information, each row contains the splits and placing of the athlete for each lap of the race. We will be looking at the whole race as well as the splits from laps. The full data set can be found at short\_track.csv.

Answer the following questions based on the histogram and table found below

|  |  |
| --- | --- |
| Q1 | 41.56s |
| Median | 42.19s |
| Mean | 43.63s |
| Q3 | 43.14s |
| Min | 39.94s |
| Max | 101.82s |
| SD | 5.76s |
| MAD | 1.076s |



1. Characterize the shape of the histogram and label approximately where the median would fall.
2. a. Find the boundaries of where outliers begin and label the bounds with arrows on the histogram if possible.
3. Are there any outliers that fall below the median time?

Lap Time Summary Statistics

Below are summary statistics for each individual lap in the data set. A 500m speed skating race is 4.5 laps, so lap 1 is only ½ of the distance of the other laps.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Lap | 1 | 2 | 3 | 4 | 5 |
| Q1 | 6.93s | 8.85s | 8.39s | 8.52s | 8.73s |
| Median | 7.09s | 8.98s | 8.52s | 8.67s | 8.90s |
| Q3 | 7.27s | 9.16s | 8.72s | 8.89s | 9.18s |
| Max | 23.87s | 43.03s | 51.87s | 48.72s | 50.68s |
| Min | 6.52s | 8.46s | 8.00s | 8.06s | 8.24s |
| Mean | 7.15s | 9.22s | 8.95s | 9.14s | 9.55s |
| SD | 0.729s | 1.661s | 2.536s | 2.516s | 3.157s |
| MAD | 0.252s | 0.222s | 0.237s | 0.252s | 0.297s |

MAD = Median Absolute Deviation

Use the table above for the following questions:

1. Watch the video below and think about why the maximum lap and total times might be so large. <https://www.youtube.com/watch?v=fAADWfJO2qM&t=109s>
2. Why might the standard deviation increase in later laps? (Note that the first lap is counted from the start to the finish line and is only half of the track).
3. Why are the mean overall and lap times higher than the median overall and lap times?

Notable Performances:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Time | Lap 1 | Lap2 | Lap 3 | Lap4 | Lap 5 |
| JR Celski  (Former WR) | 39.937s | 6.72s | 8.49s | 8.10s | 8.19s | 8.44s |
| Victor An  (Olympic Gold) | 41.312s | 7.00s | 8.76s | 8.51s | 8.53s | 8.52s |
| Shaolin Sandor Liu (Fastest Lap) | 40.523s | 7.21s | 8.65s | 8.00s | 8.22s | 8.46s |
| Wu Dajing  (2014 World Champ) | 40.526s | 6.72s | 8.69s | 8.23s | 8.25s | 8.64s |

A modified z-score can be calculated using the Median Absolute Deviation. This is an alternative to z-score value that is not as affected by outliers as a traditional z-score calculation. MAD is straightforward to find. First you take the median and find the difference between every value in a set of numbers and the median. Then, you find the value that is the median of the new set of numbers you have calculated. Instead of Standard Deviation which finds the average difference between the median and values, MAD finds the middle value, so outliers are not weighted strongly.

Mean Absolute Deviation can be used to calculate the modified z-score with the formula

0.675(x1 – xm) / MAD = Modified z-score

Where MAD is Mean Absolute Deviation, x1 is the value, and xm is the median.

Pick a Notable performance or 2 from above for the following questions:

1. Find the modified z-score for the overall time of your selected athlete if the MAD value is 1.076s.
2. Using modified z-score, what is the relatively strongest lap of your selected athlete from laps 1, 3, and 5?
3. For an athlete with an overall time of 45.25s, and lap 1, 3, and 5 times of 7.72s, 8.54s, 10.49s, are any of these times considered outliers? Why or why not?
4. Despite having the best lap performance in the data set, Shaolin Sandor Liu started and finished his race in 4th place. Based on this, and comparing his time to the other notable performances, which lap appears to hold the most significance for placement? Where does Shaolin Sandor Liu compare to the median time for this lap?