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 approximate where the median would fall?

The shape of the histogram is right-skewed and the arrow on the graph above is the approximate median

1. Find the boundaries of where outliers begin and label with arrows on the histogram if possible.

IQR = 1.58s Q1 = 41.56s Q3 = 43.14s

Lower Bound = 41.56 – 1.5(1.58) = 39.19s

Upper Bound = 43.14 + 1.5(1.58) = 45.51s

Arrows in blue on histogram.

1. Are there any outliers that fall below the lower bound?

No, the minimum time is 39.94s, while the lower bound is 39.19s.

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

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>

Crashes and other incidents lead to skaters falling and losing all momentum. Having to get up and start skating after a fall leads to a much slower time on either the lap or overall time.

1. 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).

2 options, either skaters gain separation from each other as the laps increase leading to more deviation, Or as more laps happen there is more opportunities for crashes which lead to significantly slower times in subsequent laps.

1. Why are the mean overall and lap times higher than the median overall and lap times?

Because there are many outliers and high values that push the mean lap times up more than the median 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.

JR Celski: 0.675(39.937 – 42.19)/1.076 = -1.4134

Victor An: 0.675(41.312­-42.19)/1.076 = -0.5508

Shaolin Sandor Liu: 0.675(40.523-42.19)/1.076 = -1.0457

Wu Dajing: 0.675(40.526 – 42.19)/1.076 = -1.0439

1. Using modified z-score, what is the relatively strongest lap of your selected athlete from laps 1, 3, and 5?

JR Celski: Lap 1: 0.675(6.72 – 7.09)/0.252 = -0.9911, Lap 3: 0.675(8.10 – 8.52)/0.237 = -1.1962

Lap 5: 0.675(8.44 – 8.90)/0.297 = -1.0455, Lap 3 has the lowest/highest z-score

Victor An: Lap 1: 0.675(7.00 – 7.09)/0.252 = -0.2411, Lap 3: 0.675(8.51 – 8.52)/0.237 = -0.0285

Lap 5: 0.675(8.52 – 8.90)/0.297 = -0.8636, Lap 5 has the lowest/highest z-score

Shaolin Sandor Liu: Lap 1: 0.675(7.21 – 7.09)/0.252 = 0.3214, Lap 3: 0.675(8.00 – 8.52)/0.237 = -1.481

Lap 5: 0.675(8.46 – 8.90)/0.297 = -1.0, Lap 3 has the lowest/highest z-score

Wu Dajing: Lap 1: 0.675(6.72 – 7.09)/0.252 = -0.9911, Lap 3: 0.675(8.23 – 8.52)/0.237 = -0.8259

Lap 5: 0.675(8.64 – 8.90)/0.297 = -0.5909, Lap 1 has the lowest/highest z-score

1. 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?

Overall Time: 43.14 + 1.5(1.58) = 45.51s, 45.25s is **not** an outlier

Lap 1: 7.27 + 1.5(0.34) = 7.78s, 7.72s is **not** an outlier

Lap 3: 8.72 + 1.5(0.33) = 9.215s, 8.54s is **not** an outlier

Lap 5: 9.18 + 1.5(0.45) = 9.855s, 10.49s **is** an outlier

The lap 5 time of 10.49s would be considered an outlier as the time is more than 1.5 IQR’s above Q3.

1. 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?

It appears that Shaolin Sandor Liu started slower than the other performances. His lap 1 time is significantly slower than those of the others. His time is the only one in the table that falls above the mean for that lap time. Lap 1 is imperative for skaters to get positioning and the inside of the turns in the 500m races as they are short enough that fatigue is not as much of an issue like in some longer distances. The median time for lap 1 is 7.09 seconds, 0.12 seconds faster than Shaolin Sandor Liu’s 7.21 opening lap.