Data overview

The results from a total of 2370 racers (1105 female, 1265 male) are included in this dataset. The numbers of racers in each division by gender are shown in Fig. 1. Most racers were 20 – 59 years old.

图表, 条形图

描述已自动生成

Age group

t

Count

**Fig. 1. Numbers of racers in each division by gender**

Data preprocessing

1. Remove non-number characters in “Gun Tim” and “Net Tim” and convert to numeric. Then convert them to minutes.
2. Assign divisions according to ages.
3. Create the “Gender” column and concatenate two datasets.
4. Calculate the difference between “Gun Tim” and “Net Tim”.

Calculation & visualization

1. **What are the mean, median, mode, and range of the race results for all racers by gender?**

According to Fig.2 (a), the mean median and mode of female racers are 58.47, 57.85, and 59 minutes respectively. The range is from 32.97 to 106.82 minutes. Note that to calculate the mode, I first rounded the time to whole minutes, since the mode of the original data does not contain as much information due to too scattered data. For males, Fig.2(b) shows the mean median and mode of male racers are 52.13, 51.38, and 50.0 minutes respectively. The range is from 28.78 to 100.63 minutes.

Chart, histogram

Description automatically generated

**Fig. 2(a). Histogram, mean, median, mode, and range of race results for female racers.**

Chart, histogram

Description automatically generated

**Fig. 2(b). Histogram, mean, median, mode, and range of race results for male racers.**

1. **Analyze the difference between gun and net time race results.**

Fig.3(left) shows that for females, the median difference is 3.48 minutes which is slightly longer than 3.08 minutes of males. And more than 75% of the racers’ time difference is less than 4 minutes.

Fig. 3(right) shows that the age group of 60-69 takes a large part of racers with larger than 4 minutes time difference. A strange thing is that for the > 70 age group, the median time difference is only 0.59 minutes. And by looking up the place, most of them are in the place of 600-1300, which is not reasonable for elders to get the small-time differences.

Chart, box and whisker chart

Description automatically generated Chart, box and whisker chart

Description automatically generated

**Fig. 3. Box plots of the time difference by gender (left) and by age group (right).**

Table 1 shows more detailed statistics of the time difference.

**Table 1. Time difference by gender and by age group**

|  |  |  |  |
| --- | --- | --- | --- |
| *Gender* | *Mean* | *Median* | *Std* |
| Female | 3.25 | 3.48 | 1.48 |
| Male | 2.50 | 3.08 | 1.59 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Div.* | *Mean* | *Median* | *Std* |
| 0-14 | 3.29 | 3.67 | 1.49 |
| 15-19 | 1.75 | 1.23 | 1.57 |
| 20-29 | 2.93 | 3.30 | 1.64 |
| 30-39 | 2.91 | 3.28 | 1.54 |
| 40-49 | 2.70 | 3.20 | 1.51 |
| 50-59 | 3.09 | 3.38 | 1.61 |
| 60-69 | 3.25 | 3.38 | 1.67 |
| >70 | 1.50 | 0.59 | 1.84 |

1. **How much time separates Chris Doe from the top 10 percentile of racers of the same division?**

Chris Doe is from the 40-49 division. The top 10 percentile of racers in the 40-49 division is 41.69 minutes. Chris Doe’s net time is 49.72 minutes. Therefore, the time separates are 8.03 minutes.

1. **Compare the race results of each division.**

According to Fig.4, the net time of the 15-19 age group is the shortest whereas the net time of the 70+ age group is the longest but with a large error bar. For the four groups in the middle (29-59), the race results are similar. The mean net time is longer for female racers than male racers for all age groups. The differences between males and females are large in the adults’ groups and small in the juvenile (0-14) and old age (60-69, >70) groups.

Chart, bar chart

Description automatically generated

**Fig.4. Race results for all divisions by group**

Discussion

**How do you think analytics will impact the way companies do business in the next five years?**

In the past few years, big data and analytics are only affordable by large companies. In recent years, with the development of cloud computing, many small and medium-sized enterprises are gradually using data and analytics to obtain business insights. I think in the next five years, big data and analytics will become the most valuable resource for companies. How to analyze the data and extract useful information to promote growth and innovation is the key to the success of the enterprise.

**Data scale**

With the 5g and the Internet of things (IoT), the data scale will be greatly improved. Companies that make good use of these large resources will overperform their competitors.

**Machine Learning**

Machine learning can help to analyze a large amount of data and simplify the tasks of data scientists in the automation process. Companies may be more willing to use machine learning to deal with the large scale of data.

**Data management**

As the company's business relies more and more heavily on data, ensuring the accuracy of data acquisition and the security of data storage also need to be considered by the company.