

Comparing percentile rank

The following example assumes you have already read the standard reading on *Cumulative distribution functions*.

Percentile ranks are useful for comparing measurements across different groups. For example, people who compete in foot races are usually grouped by age and sex. To compare people in different age groups, you can convert race times to percentile ranks.

As an example, suppose a male runner in his forties completes a 10K race in 42:44, placing them 97th in a field of 1633. This means that the runner beat or tied 1537 runners out of 1633, which corresponds to a percentile rank of 94%.

More generally, given position and field size, we can compute percentile rank as follows:

$$100.0 \times \frac{\text{field_size} - \text{position} + 1}{\text{field_size}}$$

The runner belonged to the “male between 40 and 49 years of age” group, and within that grouping came in 26th out of 256. Using the above formula:

```
100.0 * (256 - 26 + 1) / 256
```

```
## [1] 90.23438
```

Thus the runner’s percentile range in this age group was 90%.

If the runner continues to compete in ten years time, then he will be placed into the “male between 50 and 59 years of age” group. We can use the runner’s current percentile rank in the “40 to 49 years of age” group to estimate how he would perform in this new group, everything else remaining equal. The formula is as follows:

$$\text{field_size} - \left(\text{percentile} \times \frac{\text{field_size}}{100.0} \right) + 1$$

There were 171 people in the “50 to 59 years of age” group, so we need to compute:

```
171 - (90.23438 * (171 / 100.0)) + 1
```

```
## [1] 17.69921
```

This means that the runner would have to finish somewhere between 17th and 18th place to maintain the same percentile rank. In this particular race, the finishing time for 17th place in the “50 to 59 years of age” group was 46:05, so that’s the time that the runner needs to train for in order to maintain his percentile rank as he ages.

Credits

This work, *Comparing percentile rank*, is a derivative of [Allen B. Downey, “Chapter 4 Cumulative distribution functions” in *Think Stats: Exploratory Data Analysis*, 2nd ed. \(O’Reilly Media, Sebastopol, CA, 2014\)](#), used under [CC BY-NC-SA 4.0](#). *Cumulative distribution functions* is licensed under [CC BY-NC-SA 4.0](#) by James Glasbrenner.