

z-Scores applied to the 2016 Olympic Heptathlon

Activities 2A and 2B

Quantitative Reasoning

2021-09-06

zhg-scores (*z*-scores, high = good)

Our textbook proposes that the performances of athletes in multi-event disciplines can be compared on the basis of *z*-scores. In this activity, we apply this idea to the 2016 women's Olympic heptathlon.

One subtlety is that a good performance leads to a

- low *z*-score in the 100 m hurdles, 200 m and 800 m runs,
- high *z*-score in the high jump, shot put, long jump and javelin throw.

To make the scores comparable, we'll slightly modify the *z*-scores for some events so that high scores always indicate good performance. Let's call these modified scores '*zhg*-scores'. That label means *z*-scores, high = good.

$$zhg_{\text{event}} = \begin{cases} -z_{\text{event}} & \text{if event is 100 m hurdles, 200 m run or 800 m run,} \\ z_{\text{event}} & \text{otherwise.} \end{cases}$$

Activity 2A

- Import `hept.csv` as a data frame `hept` into R.
- For all seven heptathlon events, append one column to `hept` that contains the *zhg*-score. Name the columns `zhg_hurdles`, `zhg_hj` etc. Hint: Use the `scale()` command to save yourself some typing!
- The sum of all seven *zhg*-scores is a measure of the athletes' overall performance,

$$zhg_{\text{total}} = \sum_{\text{event}} zhg_{\text{event}}.$$

Append the total *zhg*-scores as a column `hept$zhg_total`.

Activity 2B

To compare the official points (`pts_hurdles`, `pts_hj`, ...) with the *zhg*-scores, make scatter plots (e.g., `pts_hurdles` vs. `zhg_hurdles`).

Do so for at least two individual events. Here are some suggested assignments:

- *Team 1*: 800 m run and high jump
- *Team 2*: shot put and 200 m run
- *Team 3*: long jump and javelin throw

Then try the same for the totals (e.g., '`pts_total`' vs. '`zhg_total`').

When you're finished, your instructor will show you how to add a line of best fit to the plots.