**Practice RAT for Week 05, Monday**

**Answers at the end**

|  |  |
| --- | --- |
| 1 | This question refers to the assigned R preparation.  The spreadsheet *hept.csv* on Canvas (under Files > Week04\_Lesson2) contains information about the 2016 Olympic heptathlon. The column hurdles shows the performance of the heptathletes in the 100 m hurdles. Suppose we have imported the spreadsheet into R as a data frame hept. How do we add a column z\_hurdles that contains the *z*-scores in the 100 m hurdles? |
|  |
| (a) | hept$z\_hurdles <-  (hept$hurdles - mean(hept$hurdles)) / sd(hept$hurdles) |
| (b) | hept$z\_hurdles <-  hept$hurdles - mean(hept$hurdles) / sd(hept$hurdles) |
| (c) | hept$z\_hurdles <- scale(hept$hurdles) |
| (d) | hept$z\_hurdles <- z(hept$hurdles) |

|  |  |
| --- | --- |
| 2 | Suppose we have added the *z*-scores for the hurdles (z\_hurdles) and high jump (z\_hj) events using the correct method from Question 1. Here are the *z*-scores for Nafissatou Thiam and Jessica Ennis-Hill.  # By the way, we can subset columns with a vector of column names.  hept[1:2, c("athlete", "z\_hurdles", "z\_hj")]  athlete z\_hurdles z\_hj  1 Nafissatou Thiam -0.02157905 2.263759  2 Jessica Ennis-Hill -1.98058119 1.138347  Which statement below is correct? |
|  |
| (a) | In both events, **Thiam** was the **better** athlete because  in the hurdles event **and** in the high jump. |
| (b) | In the combination of both events, **Thiam** was the **better** athlete because  . |
| (c) | In the combination of both events, **Thiam** was the **worse** athlete because  . |
| (d) | In both events, **Thiam** was the **worse** athlete because  in the hurdles event **and** in the high jump. |

|  |  |
| --- | --- |
| 3 | The Celsius (C) and Kelvin (K) temperature scales have units of the same size but different 0-points (0 K = -273 C). Suppose your friend, a huge nerd, records the classroom temperature each day and gives you the results and summary statistics in K. When you convert to C, which of the following changes would you expect? |
|  |
| (a) | The mean and standard deviation will change, but by different amounts. |
| (b) | The mean and standard deviation will both change by the same amount. |
| (c) | The mean will change, but the standard deviation will stay the same. |
| (d) | It is impossible to say. |

|  |  |
| --- | --- |
| 4 | People with *z*-scores above 2.5 on an IQ test are sometimes classified as geniuses. If IQ scores have a mean of 100 and a standard deviation of 16 points, what IQ score do you need to be considered a genius? |
|  |
| (a) | 160 |
| (b) | 124 |
| (c) | 140 |
| (d) | 250 |

|  |  |
| --- | --- |
| 5 | Here are the summary statistics for the weekly payroll of a small company:   * Lowest salary = $350, * Mean salary = $700, * Standard deviation = $250.   Suppose business has been good and the company gives every employee a 10% raise. Tell the new value of each of the summary statistics. |
|  |
| (a) | * Lowest salary = $420, * Mean salary = $770, * Standard deviation = $320. |
| (b) | * Lowest salary = $420, * Mean salary = $770, * Standard deviation = $250. |
| (c) | * Lowest salary = $385, * Mean salary = $770, * Standard deviation = $275. |
| (d) | * Lowest salary = $385, * Mean salary = $700, * Standard deviation = $250. |

**Answers:**

**1a and 1c.** 1c is a short version of 1a. See ?scale for more information. (Real RATs will always have only one correct answer.)

**2c.** Large z-scores are good in high jump (greater height = better performance), but bad in running events (longer time = worse performance). Ennis was better in the 100 m hurdles, but Thiam was better in the high jump. When we add the *z*-score for the high jump and subtract the *z*-score for the hurdles, Ennis was the better athlete. (See the examples on page 130 of our textbook.)

**3c.** Adding a constant to a distribution changes the mean by that same constant, but it does nothing to the standard deviation. (The data are shifted, but neither stretched nor compressed.) Multiplying the distribution by a constant multiplies the mean and standard deviation–and other summary statistics–by that constant. Other operations (e.g. taking the absolute value or log of the data) will have more complicated effects on the summary statistics.

**4c.** Straightforward application of the formula .

**5c.** From page 133 of our textbook: “When we multiply all the data values by any constant, all measures of position and measures of spread are multiplied by that same constant.”