Chapter 3 – Worked Solutions to Exercise 3J

Learning to solve problems

Like any worthwhile activity, learning to solve mathematics problems is a skill that takes time to develop. Here are some ideas on how to become a great problem solver:

Read each problem carefully	Make sure that you know exactly what information is given, and what you are being asked to find. Often it helps to read the question a few times.
What mathematic topics may be useful?	Problems given in textbooks are based on mathematics that you have learned – either in that chapter, or previously. If you can figure out which mathematics you need to tackle the problem, you are more than half way to its solution.
Revise the examples	Often the mathematics that is needed has been introduced is in the current chapter. Work through the examples of the mathematical procedures that you think will be needed to solve the problem.
Use the hints	At this early stage of learning to solve problems, you may find even after reading the problem and working the examples, you are still not sure how to proceed. At this stage, read the hints, slowly and carefully, and then go back and have another go at the problem.
Look at the solution	It is likely that there will be problems that you are unable to solve, especially the challenging problems at the end of a problem set. For these, read through the solution, then put it aside and try to set out the solution on your own. If you get stuck, look at the solution again. Repeat this until you can do the solution unaided.
Do *all* of the problems[1]	Any complex activity, such as hitting a golf ball or solving a problem, requires practice. You will become a good problem solver by solving lots of problems!
Be persistent	By definition, a problem is a question for which the method of solution is not immediately obvious. If you can't solve a problem after a good attempt, set it aside and come back to it later. Often you will find that you will see a method of solving the problem when you return to it after a break.

1

Chapter 3 – Worked Solutions to Exercise 3L

Hints

1. Make sure you are clear about what the standard deviation measures.
   Does the standard deviation use all of the data values?
   Look at the formula for the standard deviation and ask what happens if a data value equal to the mean is removed from the dataset.

2. a. Which is the relationship between the mean and the median for a skewed dataset?

   b. Read the chapter or your notes to answer this question: under what conditions is the mean the most appropriate measure of the centre?

   c. Which measure of the centre uses the total in its calculation?

   d. The word "drenched" implies an extreme value. Which measure of the centre is most affected by extreme values?

3. a. -

   b. Study the Graphics Calculator Activity on page 130.

4. You might find it useful to verify this with the dataset –3, -2, -1, 0, 1, 2, 3 first. And then use algebra. There are benefits in making *n* the middle member of the dataset – can you see why?

5. Study the Graphics Calculator Activity on page 130.

6. Read the question slowly and carefully. There are many words that have precise definitions and you need to know these. If you don't, look them up in the glossary.

   Make up a suitable dataset (i.e. one with an outlier) and calculate the mid-hinge and mid-range with the outlier included and then with the outlier excluded. Which statistic changed the most? Why?

Chapter 3 – Worked Solutions to Exercise 1J

1. B. The formula for standard deviation is

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}$$

When we remove a data value equal to the mean, the numerator $\sum (x-\bar{x})^2$ doesn't change value while the denominator decreases by 1. Hence the standard deviation will increase.

Here is another way of thinking about it:

> The standard deviation is in some sense a measure of the average deviation of each data value from the mean. By removing a data value that has *no* deviation from the mean, the average deviation of the remaining data value must increase.

Note that simply verifying this for a particular dataset isn't sufficient. You need to show that it is true for *all* datasets that meet the conditions given in the question.

2. a. This implies that the dataset is skewed to the right, or has outliers to the right. Either of these will "pull" the mean further to the right than the median.

b. The median is almost always used for rainfall, because rainfall data is almost always skewed and often has outliers. This is certainly the case here, judging by the large difference in the mean and median rainfall.

c. Since the mean uses the total rainfall in its calculation, we can best estimate total rainfall by adding the mean rainfall for each month.

An estimate for total rainfall = 23 + 24 = … + 19 + 20 = 248 mm.

d. The table shows that the median rainfall is low in these months, implying generally dry conditions, while the mean rainfall is quite high, implying the monthly data contains outliers, in this case storms.

3. The mean = 42 while the standard deviation = 16.

If we add 18 to each member of the dataset, this will raise the average by 18 without affecting the standard deviation. The new set of marks is now:

36 44 52 60 68 76 84

If we wish to reduce the standard deviation from 16 to 15, we might try moving each mark closer to the mean by one unit. This results in this set of marks:

37 45 53 60 67 75 83

This gives us a mean of 60 and a standard deviation of 15.15, which is as close as we can get to $\bar{x} = 60$ and $s = 15$ without using decimal scores.

4. First an algebraic solution.

Let us call the seven numbers

n-3 n-2 n-1 n n+1 n+2 n+3

We first calculate the mean:

$$\begin{aligned}\bar{x} &= \frac{\sum x_i}{n} \\ &= \frac{(n-3)+(n-2)+(n-1)+n+(n+1)+(n+2)+(n+3)}{7} \\ &= \frac{7n}{7} = n\end{aligned}$$

∴ the mean is an integer, and is the middle value

Now we calculate the standard deviation:

$$\begin{aligned}s &= \sqrt{\frac{\sum (x-\bar{x})^2}{n}} \\ &= \sqrt{\frac{[(n-3)-n]^2+[(n-2)-n]^2+\ldots+[(n+3)-n]^2}{7}} \\ &= \sqrt{\frac{9+4+1+0+1+4+9}{7}} \\ &= \sqrt{\frac{28}{7}} = \sqrt{4} = 2\end{aligned}$$

∴ the standard deviation of any dataset consisting of 7 consecutive integers is 2 (which is an integer).

Note that choosing the *middle* value to be *n*, rather than the *first* value, simplified the algebra.

If you want to take a non-algebraic approach, you can solve the problem this way:

Consider the dataset A = { –3, -2, -1, 0, 1, 2, 3}, which has a mean of 0 and a standard deviation of 2.

Now any dataset B that consists of seven consecutive integers can be formed from dataset A by adding *n* to each member of A, where *n* is the middle value of dataset B. The effect of adding *n* to each member of A is to change the mean to *n*. Adding the same number to each member of a dataset doesn't affect the spread, so the standard deviation will still be 2.

Voila! And no algebra, just clear thinking!

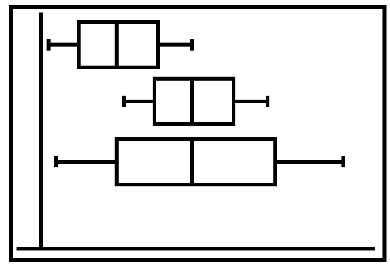
5. Recall that

- if we add a fixed amount *n* to each value in a dataset, the mean increases by *n* but the standard deviation doesn't change.
- if we multiply each value in a dataset by *n*, the mean is the old mean times *n* and the standard deviation is the old standard deviation times *n*.

Hence:

The boxplot for dataset B is congruent to the boxplot for dataset A but shifted horizontally by 50 units.

The boxplot for dataset C is shifted to the right by a factor of 2 and stretched horizontally by a factor of 2.



In the diagram alongside dataset A (on top) consists of 100 random numbers from (5, 10, 15, 90, 95, 100}. Dataset B is in the middle and dataset C on the bottom.

Is it coincidence that the medians of datasets B and C are similar?

6. A robust statistic isn't affected much (or at all) by outliers. Assume that the maximum value in a dataset is increased by 100, making it an outlier. Then the mid-range will increase by 50. The mid-hinge on the other hand will not change at all. The mid-hinge is a more robust statistic than the mid-range as it isn't affected by outliers.