Trimmed Means

The last mean we will look at, but not the last type of mean, is the trimmed mean. It exists because some data sets have **outliers**, numbers that are noticeably different from the rest. Let's look at a case where there is an outlier.

Sophie is taking Science 10 and is doing well. Her first test results as were: 18/20, 17/20, 18/20, 16/20 and 18/20. Unfortunately, she then became sick and missed a week of school. A few days after coming back she wrote a test and only earned 13/20. Let's do some statistics.

Step 1: Order the values
Step 2: Number the values

$13/20_{1},\ 16/20_{2},\ 17/20_{3},\ 18/20_{4},\ 18/20_{5},\ 18/20_{6}$

Step 3: Determine the mode

The mode is **18/20** because it occurs 3 times.

Step 4: Determine the median

Because there is an even number of values (6), we find the average of the two middle values, 3 and 4. We also only need to find the average for the numerators (the tops) because all the denominators (the bottoms) are the same. So, $\frac{17+18}{2} = \frac{35}{2} = 17.5$ The median is 17.5/20.

Step 4: Determine the arithmetic mean

Again, we only need to find the mean of the numerators because all the denominators are the same.

Arithmetic mean = $\frac{13+16+17+18+18+18}{6} = \frac{100}{6} = 16.7$ = 16.7/20

You can see that the arithmetic mean is **skewed** to the lower side because of the low mark after Sophie was sick. The mean is smaller than the mode and the median. Rather than indicating a value close to the centre, it seems closer to the smaller numbers. That's because 13/20 is an **outlier**, an unusually small (or it can be large) value. We can fix this by **trimming** the data set.

To trim properly, you must

- Identify **all** the outliers. There can be more than one. They will always be a lot bigger or smaller than the other numbers.
- Cross out the outliers.
- Cross out an equal number of values on the opposite end of the data set. So, if you have one low outlier, as in the example, then cross out the highest mark. You must always cancel as many low values as high values and vice versa (that is, as many high values as low values).
- Calculate the mean on the number of values left.

Our **trimmed mean** for the example would cross out the 13/20 and one of the 18/20.

Trimmed mean = $\frac{16+17+18+18}{4} = \frac{69}{4} = 17.3$ Note that we **divide by 4** because we are only adding four values. 17.3 seems a better mean than the previous one.

Instructions: Do the following three practices following the example above and referring to the instructions above on how to solve **trimmed mean** problems when you have difficulties. The last problem is your test problem. Send me your answer and I will confirm if it is correct.

1. A data set contains the following numbers:

   125, 164, 183, 154, 132, 129, 524, 476, 165, 133

   a) Calculate the **arithmetic mean**.

   b) Identify any **outliers** if they exist and calculate the **trimmed mean**.[1] (Remember: cross off as many low values as high outliers you cross off.)

[1] 1a) arithmetic mean = 218.5 b) outliers are 476 and 524. Cross out with 125 and 129. Trimmed mean = 155.2

2. Calgary, AB, is known for its chinook winds, warm winds that blow from the Pacific Ocean and that can raise the temperature significantly over a short period of time. The following table shows the daily high temperatures for one week in January, during which there was a chinook wind.

DAILY HIGH TEMPERATURES FOR ONE WEEK IN JANUARY, CALGARY, AB (°C)						
Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.
-12	-14	-5	15	-8	-10	-12

a) What was the mean high temperature for the week? (Make sure to put the **negative sign** in front of the number, if it has one, when adding, such as -12 + -14.)

b) On what day did the chinook occur?

c) What is the trimmed mean of the daily highs?

d) Which is the better indicator of the expected daily high in Calgary for a week in January? Why?[2]

[2] 2a) -6.6^{0}C b) Wednesday c) -9.4^{0}C d) various, something like trimmed is better because they don't have chinooks every week so represents a normal week more accurately.

3. Brian is a real estate salesperson. He recently sold houses for the following prices: $1 356 000.00, $289 000.00, $359 000.00, $376 900.00, $454 900.00, $376 500.00, $425 800.00

   a) Calculate the mean selling price.

   b) Identify any outliers and calculate the trimmed mean.

   c) Which calculation do you think best represents the average price of a house sold by Brian?[3]

[3] 3 a) mean = $519 728.57 b) $1 356 000 is an outlier. Trimmed mean = $398 620.00
c) Answers will vary. As long as the answer is supported by numbers, it is considered valid.

Workplace 10
Statistics and Probability 4

Test Question: Email your answers to this problem at larry.green@sd71.bc.ca and I will confirm your answer. If you do not get this correct, I will email you a second set of weighted mean questions to help you complete your understanding of the second statistics lessons.

Isabelle is a woodworker in Hinton, AB. She specializes in making furniture out of blue pine-wood from trees killed by the mountain pine beetle. Isabelle records how long is takes her to produce each item, in order to help her set the selling price for the furniture. The last 10 items she has made have taken her the following times (in hours):

10, 12, 10, 14, 13, 9, 21, 10, 11, 15

a) What is the mean number of hours Isabelle spent on each piece of furniture?

b) Are there any outliers? If so, calculate the trimmed mean.

c) Which calculation do you think best represents the average number of hours spent on each piece of furniture, and why?