MDM 4U **Module 3.2—Measures of Central Tendency, pp. 152-58**

Read the Introduction provided to familiarize yourself with the learning goals of this module. By the end of the module, return to the goals outlined to guide you in your reflection of what you have learned.

Much like you did in Module 3.1, answer each of the numbered problems, in line with the text already given, as you move through this assignment. Please ensure that I can clearly distinguish your responses from the given problems.

<u>Introduction</u>
In the last module, you took note that there were types of distributions—as described by their shape—symmetric and skewed.

Goal 1: In this section, you will discover that a distribution of data can also be described, in part, by a single value. The value that you calculate will describe the *centre* of the data.

The single values, which can describe the *centre* of a distribution, are the mean, median, and mode.

Goal 2: These values can be labelled on the distributions provided and on those that you can create using technology.

You might also hear of these values each being referred to as the *average*. For instance, the average value for a vehicle might be reported to be $21 500.

Goal 3: Without knowing the type of average being used, the buyer might be misled by the figure being reported by the seller. In fact, *outlying* values may or may not have been considered in the calculation of the average value.

Problem Set

---

1. Define the term *outlier*.
   An outlier is an element of a data set that is very different from the others.

2. Define the term *mean*.
   Mean is a measure of central tendency found by dividing the sum of all the data by the number of pieces of data.

Below is an equation that represents the calculation of the mean. Compare the "parts" of this formula to the definition you wrote in #2.

Sum of values in a set of data

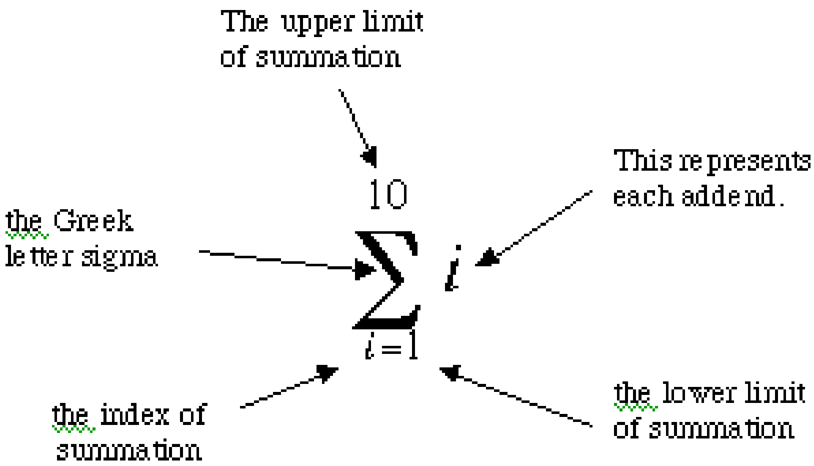
Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Number of values in the set

3. Label the “parts” of this equation, by filling in the textboxes with “parts” of the definition you wrote in #2.

The Greek letter **Σ** (upper-case sigma, σ is lower-case sigma and is used later in this chapter) is used to designate **summation**. We use summation notation to represent series (i.e., sum of terms) in a compact form.

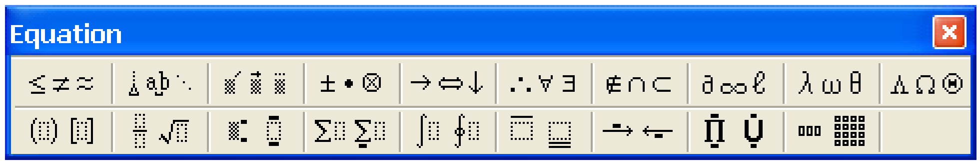


Example 1: To clarify the use of summation notation, consider each of the following.

Notation	Interpretation	Value
a) $\sum_{i=1}^{10} i$	The sum ($\sum$) of the first ten (i.e., values of i are 1 to 10, inclusive) natural numbers, i.	$\sum_{i=1}^{10} i = 1+2+\ldots+9+10$ $= 55$
b) $\bar{x} = \frac{\sum_{i=1}^{10} i}{10}$	The mean, $\bar{x}$ (read “x-bar”), of the first ten natural numbers.	$\bar{x} = \frac{\sum_{i=1}^{10} i}{10} = \frac{1+2\ldots+9+10}{10}$ $= \frac{55}{10}$ $= 5.5$
c) $\bar{x} = \frac{\sum_{i=1}^{10} 2i}{10}$	**The mean of the first 10 natural numbers doubled.**	$\bar{x} = \frac{\sum_{i=1}^{10} i}{10} = \frac{2+4\ldots+18+20}{10}$ $= \frac{110}{10}$ $= 11$

4. Complete the table, above, by typing an interpretation and calculating a value for the equation provided in c).

[**Note:** In the value column, double-click on the equation to edit it. You should get an equation editor (pictured on the next page) that will allow you to show each line in your solution. Once you've finished editing, cursor as far as you can to the right, and then click outside of the editing box. The box will close, and you should have a completed solution.]
Equation Editor:



Example 2: In some cases, data may already be grouped into intervals where the width is not equal to one. The table below exemplifies such a situation.

Table—Age When Car Owners Got Their 1st Vehicle

Age (yrs.)	16-20	21-25	26-30	31-35	36-40
Frequency	10	18	12	8	2

In such a case, we do not have individual ages; rather, we have several classes of age. When this occurs, we must calculate and use the w**eighted mean**. It is a measure of central tendency that reflects the *relative* importance of each data item and is described by the following formula.

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Note that there is a slight difference between this formula and that presented on page 1—you've likely noticed the symbol, w_i. This symbol takes the place of the value that assigns the *relative importance* or *weighting* to a value, x, of a discrete variable X.

In this example, the weightings, w_i, are represented by the *frequencies* for each age group. Each weighting is then multiplied by its corresponding age (value of the discrete variable, *Age*) to account for that interval's relative effect upon the mean. As mentioned above, we have several classes of ages. In these situations, use the *median ages* as the values of x_i.

The process for calculating the weighted mean, in this example, has been summarized in the table below.

Age (yrs.)	16-20	21-25	26-30	31-35	36-40
Midpoint of Age Intervals, *m*	18	23	28	33	38
Frequency, *f*	10	18	12	8	2
$f \times m$	10(18)	18(23)	12(28)	8(33)	2(38)

Using the organized information from the last row of the table, we can now move forward with using the formula for calculating a weighted mean.

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Sum of all weighted values—products, $f \times m$

Sum of all weightings—frequencies, f

simplified

$$\bar{x} = \frac{\sum f \times m}{\sum f}$$

$$= \frac{10(18) + 18(23) + 12(28) + 8(33) + 2(38)}{10 + 18 + 12 + 8 + 2}$$

$$= \frac{180 + 414 + 336 + 264 + 76}{50}$$

$$= \frac{1270}{50}$$

$$= 25.4$$

Therefore, the mean age of when car owners got their first vehicle was about 25 years.

5. The following marks were obtained on a series of mathematics exam. Use the table below, to help organize your calculation of the mean mark obtained and calculate the mean mark. Note: To calculate the mean mark, you're welcome to use the equation editor, or simply record the value you obtained for the weighted mean.

Mark Interval	0-19	20-39	40-59	60-79	80-99
Frequency	1	2	7	23	11

Mark Interval	Frequency, f	Interval midpoint, m	$f \times m$
0-19	**1**	**9.5**	**9.5**
20-39	**2**	**29.5**	**59**
40-59	**7**	**49.5**	**346.5**
60-79	**23**	**69.5**	**1598.5**
80-99	**11**	**89.5**	**984.5**

Weighted mean= $\dfrac{9.5+59+346.5+1598.5+984.5}{1+2+7+23+11}$

$= 68.1$

The weighted mean of the scores from the math exam is 68.1.

Weighted means can also be valuable to students when trying to understand how their teachers calculate report card marks.

6. Read Example 1 (Solution 1) on pp. 153-54. Use this example to solve the following problem. You're welcome to use the equation editor or to type in your solution to the best of your ability.

Problem

A teacher weights student marks in MDM term mark calculations as follows: Knowledge and Understanding, 25%; Application, 25%; Thinking/Inquiry, 30%; and Communication, 20%.

a) A student's marks in these categories are 80, 72, 78, and 86, respectively. Calculate the student's term mark to the nearest percent.

$t = \dfrac{80(0.25) + 72(0.25) + 78(0.3) + 86(0.2)}{0.25 + 0.25 + 0.30 + 0.20}$

$= \dfrac{20 + 18 + 23.4 + 17.2}{1.0}$

$= 78.6$

Therefore, the student's term mark is approximately 79%.

b) If the term mark is worth 70%, and the final exam, 30%, what exam mark, to the nearest percent, must the student achieve to earn a final grade of 80%?

$0.70(78.6) + 0.30E = 80.0$

$E = \dfrac{80 - 0.70(78.6)}{0.3}$

$= 83.3$

Therefore, the student must achieve an 83% on the final exam to achieve a final grade of 80%.

Other Measures of Central Tendency—Median and Mode

7. Define *median*.
   Median is the middle value of an ordered data set.

8. Explain, through the use of an example, how you would determine the median of a data set containing
   a. an even number of data.
      1,2,5,6,8,9
      If these numbers were my set of data, both 5 and 6 are found in the middle of the

data. In order to find the median, you would add 5 and 6 together then divide by 2 to get a value of 5.5.

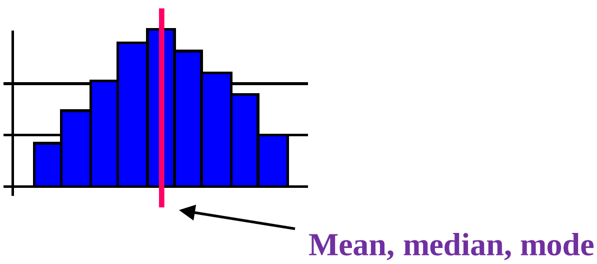
b. an odd number of data.
1,6,8,9,10
If these numbers were my set of data, you would count from both ends to find the middle number, which in this case would be 8.

9. Define *mode*.
Mode is the most frequent value or interval in a set of data.

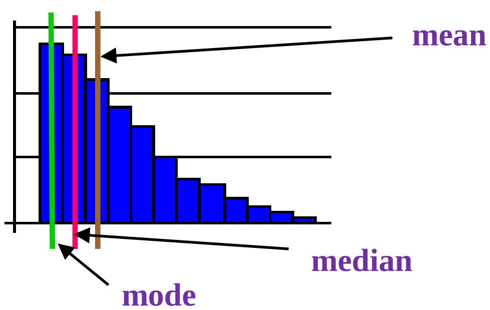
10. Provide your own example of a data set that is *bimodal*.
1,1,1,2,3,4,5,6,6,6,7,8,9,9,10
This data set is bimodal because both number 1 and number 6 appear 3 times, so there are 2 different modes.

11. What measure of central tendency—mean, median, or mode—is the only measure appropriate for *qualitative data*?
Mode is the only measure of central tendency that is appropriate for qualitative data.

12. For the graphical display of the *symmetric* distribution below, label the positions of mean, median, and mode.



13. For the *skewed* distribution below, label the *relative* positions of the mean, median, and mode.



14. Complete each of the following concerning the relationship between the 3 central measures on histogram.

a. For right-skewed distributions, mode < **median** < **mean**.

b. For left-skewed distributions, **mean** < **median** < mode.

14. Which measure of central tendency is most affected by outliers? Instead, what statistic should be used in reporting?
    Mean is the measure of central tendency that is most affected by outliers. Instead, the median should be used to avoid misrepresenting the data.

15. Read Example 6 on p157, and then answer the following questions.

    a. For the month of January, why would either the mean or median income provide a good measure of central tendency?
       Either the mean or median would provide a good measure of central tendency because they are fairly close to each other. Also there is no modal value.

    b. For the month of February, why was the median income value chosen?
       The median value was chosen because there was an outlier (0$) which greatly affected the mean.

    c. Lastly, for the month of March, why was the modal income value chosen?
       The modal income value was chosen because half of the people received this amount, so it the most accurately represented the value.

16. Using your knowledge from this module, thoroughly discuss the following set of yearly salaries with respect to the 3 measures of central tendency. Please ensure that you also select one of the measures as being the most representative measure for the set and justify this choice.

    $16000, $21000, $16000, $20000, and $150 000

    For the values given, the median would be the most representative measure of central tendency. One can see that the $150 000 salary would greatly influence the calculation of the mean.

17. A survey is handed out to 20 people, asking them their favourite colours. Choose a measure that would provide a good measure of central tendency. Briefly explain your choice.
    A good measure of central tendency would be mode because finding the average colour would be somewhat difficult and you can't line up colours in any order. The mode would accurately represent the most common colour. Also, we can think of colour as a qualitative description of an attribute—something that can't be quantified.

Your Homework

For consolidation and further practice, complete p158 #1ace (technology => spreadsheet, Fathom or graphing calculator), 2ace, 3ace, 4, 5, 6aceg, 7ab, 8, 12 to 15.

Note: If there are any concepts that you need explained, try viewing some YouTube videos for assistance.