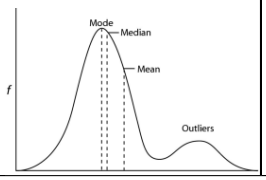
MDM4U: Mathematics of Data Management

Measures of Central Tendency (Part 2)

Grouped Data and Weighted Means

MDM4U: Data Management



Weighted Means

Sometimes, some data are more significant than others.

These data may be given *weights* to indicate relative importance.

For example, a unit test might have a weight of 5, whereas a smaller quiz might only have a weight of 2. This means that questions on the unit test will count more toward a student's mark than those on a quiz.

Weighted Mean of a Data Set

If each datum, x_i, has an associated weight, w_i, then the weighted mean of the data set is given by

$$\overline{x} = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum x_iw_i}{\sum w_i}.$$

Weighted Means

Example

A student earns the following marks in each category of her Data Management class.

Category	Mark (%)	Weight
Knowledge	82	20
Application	73	20
Thinking	69	25
Communication	84	5
Project	85	10
Exam	72	20

Determine the student's final mark in the course.

Weighted Means

Use a table to find $\sum w_i$ and $\sum w_ix_i$.

Mark (x_i)	Weight (w_i)	w_ix_i
82	20	1 640
73	20	1 460
69	25	1 725
84	5	420
85	10	850
72	20	1 440

$\sum w_i = 100$ and $\sum w_ix_i = 7\,535$.

Therefore, the student's final mark is $\overline{x} = \frac{7\,535}{100} = 75.35\%$.

Grouped Data

Sometimes data are grouped together into intervals.

Given the option, calculations should *always* be done using raw data, prior to grouping. This is more accurate and, with software, is not as time-consuming as it once was.

If data are already grouped, however, we can *approximate* the mean, median and mode.

Mean, Median, Mode (Grouped Data)

Approximations of the Mean, Median and Mode for Grouped Data

For grouped data, the mean can be approximated using

$$\overline{x} = \frac{\sum f_im_i}{\sum f_i}$$

where f_i is the frequency for a particular interval and m_i is the midpoint of the interval.

The median is the midpoint of the interval containing the central datum.

The mode is the midpoint of the interval with the highest frequency.

Mean, Median, Mode (Grouped Data)

Example

Surveys about television viewing habits are mailed to 1 000 individuals. Of these, 500 are returned with the following results.

Hours/day	0-1	2-3	4-5	6-7	8-9	10-11	12-13
Frequency	64	92	141	86	71	35	11

Calculate the mean, median and mode of the data, and identify any problems with the data.

Mean, Median, Mode (Grouped Data)

Make a frequency table.

Interval	Midpoint (x_i)	Frequency (f_i)	Cumulative	$f_i x_i$
0-1	0.5	64	64	32
2-3	2.5	92	156	230
4-5	4.5	141	297	634.5
6-7	6.5	86	383	559
8-9	8.5	71	454	603.5
10-11	10.5	35	489	367.5
12-13	12.5	11	500	137.5

Mean, Median, Mode (Grouped Data)

$\sum f_i = 500$ and $\sum f_i x_i = 2\,564$.

$\overline{x} = \dfrac{2\,564}{500} \approx 5.1$ hours/day.

The median is the central datum. Since there are 500 values, the median will be the midpoint of the 250th and 251st data.

Both of these data occur in the 4-5 hour interval, so the median is 4.5 hours/day.

The interval with the highest frequency is also the 4-5 hour interval, so the mode is 4.5 hours/day.

Mean, Median, Mode (Grouped Data)

Example (Continued)

There are a few problems with the data.
The data are grouped, and we do not have access to the raw data. This means that the mean, median and mode are approximations.
The intervals are not well-chosen. For example, there are no intervals for 1-2 hours/day, 3-4 hours/day, etc. This may result in inaccurate data values.
The survey is voluntary, and only half were returned. This opens up the possibility for non-response bias in the survey.

Questions?