Central tendency measures for ungrouped (raw) univariate data

BEA140 Quantitative Methods - Module 2

UNIVERSITY *of* TASMANIA

Central Tendency

In these slides we will look at a number of central tendency measures for ungrouped (raw) univariate data.

In statistics **central tendency measures** attempt to give us an idea of where the *middle* of some data points is.

The **mean** of a data set, which is also commonly known as the **average**, is the sum of the data points divided by the number of data points.

In BEA140 we will use the following notation and formulas for population/sample means:

$$\mu_X = \frac{\Sigma X_i}{N} \text{ (population)}$$

$$\overline{X} = \frac{\Sigma X_i}{n} \text{ (sample)}$$

Both the sample and population mean can be obtained in Excel using the `AVERAGE` function.

Central Tendency - Example Mean Calculation

As an example, we will use the following sample of ungrouped (raw) univariate data consisting of 9 travel times (in minutes):

15	29	8	42	35	21	18	42	26

Note: it is more common to see univariate data in a column, however it will be more space efficient in our slides to place them in a row.

$$\Sigma X_i = 15 + 29 + \ldots + 42 + 26 = 236; \text{ and}$$

$$\overline{X} = \frac{\Sigma X_i}{n} = \frac{236}{9} = 26.22 \text{ (to 2 dp).}$$

I.e. the mean travel time was 26.22 minutes.

Central Tendency - Median

The **median** of an ordered data set, which we will denote as M_d, is:

(i) the middle data point when the number of data points is odd, which occurs in position $\frac{n+1}{2}$ of the ordered data; and

(ii) the mean/average of the two middle data points when the number of data points is even,
which occur in positions $\frac{n}{2}$ and $\frac{n}{2} + 1$ in the ordered data.

The median can be obtained in Excel using the `MEDIAN` function.

Note: that the median is a positional measure, it is appropriate for any numerical or ordinal data set that can be ordered.

However, for an even number of ordinal data points it is not possible to take the mean of the two middle data points, in such a case we will employ the convention of using the ordered data point in position $\frac{n}{2} + 1$.

Central Tendency - Example Median Calculation

Going back to our sample of 9 travel times (in minutes):

15	29	8	42	35	21	18	42	26

To find the median without Excel, first we need to order the data:

8	15	18	21	26	29	35	42	42

Now the median position is $\frac{n+1}{2} = 5$, giving us:

$$\text{median} = M_d = 26 \text{ minutes.}$$

Central Tendency - Mode

The **mode** of a data set is/are the value(s) that appear most often.

Note: A data set with:

(i) two modes is often referred to as **bimodal**; and

(ii) more than two modes is often referred to as **multimodal**.

Example: Going back to our sample of 9 travel times (in minutes):

15	29	8	42	35	21	18	42	26

There is only one travel time which appears twice in the data set, hence:

$$\text{mode} = 42 \text{ minutes.}$$

Excel has a `MODE` function, however it is only really useful for single mode data sets.

Central Tendency - Midrange

The **midrange** of a data set is the mean/average of the maximum and minimum data values.

$$\text{I.e. midrange} = \frac{X_{\max} + X_{\min}}{2}$$

Going back to our sample of 9 travel times (in minutes):

15	29	8	42	35	21	18	42	26

Since $X_{\max} = 42$ and $X_{\min} = 8$, we have:

$$\text{midrange} = \frac{X_{\max} + X_{\min}}{2} = \frac{42 + 8}{2} = 25 \text{ minutes.}$$

. . . that's it for now, thanks for watching!

Don't forget that you can ask questions via:

(i) face-to-face lectures;

(ii) workshops or tutorials;

(iii) consultation hours; or

(iv) email.