

Objectives:

- To learn to calculate certain statistics (**mean**, **median**, and **mode**) for summarizing the center of a distribution of data
- To investigate and discover properties of these summary statistics
- To explore the statistical property of **resistance** as it applies to these statistics
- To develop an awareness of situations in which certain measures are not appropriate
- To recognize that these numerical measures do not summarize a distribution completely
- To acquire the ability to expose faulty conclusions based on misunderstanding of these measures

Three commonly used measures of central tendency (or center of a distribution):

- The **mean** is the ordinary arithmetic average found by adding the value of each observation and divide by the number of observations
- The **median** is the middle observation when arranged in order.
- The **mode** is the most occurring value

For the following formulas and for the remainder of your mathematics you will see the following notation

n = the number of observations in a set

i = When looking at a particular observation in a general formula

x_1 = first term (or observation)

x_2 = second term

x_i = i^{th} term

x_n = final term

Σ = summation, This means we add all the values involved

$\bar{x}$ = **mean**

Calculations for **mean**:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{i.e. the sum of all } n \text{ observations divided by } n.$$

$$\text{This is equivalent to saying } \bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Calculation of **median**, first the data set must be sorted in order:

$$\text{The median of a set of } n \text{ observations is } \begin{cases} x_i | i = \frac{(n+1)}{2} & \text{If } n \text{ is odd} \\ \frac{x_i + x_{i+1}}{2} | i = \frac{n}{2} & \text{If } n \text{ is even} \end{cases}$$

e.g. if there are $n = 11$ data items then x_6 is the median of the set: $i = (11+1)/2 = 6$

if there are $n = 12$ data items then $(x_6 + x_7)/2$ is the median of the set: $i = 12/2 = 6$

Calculation of **mode** is trivial find the frequency of each value possible, The one with the highest frequency is the mode of the set. This is most easily done after the data is sorted.

resistant: when a measure is relatively unaffected by **outliers** in a distribution **mode** is completely unaffected by outliers so it is said to be **resistant**.

Measures of central tendency and their usefulness depending on the type of variable:

Measurement variables:

- Mean useful when the spread is fairly normal

- Median useful when there are extreme outliers

- Mode useful in only certain circumstances

Categorical

- Mean - not calculable

- Median useful as long as there is an ordering to the categories

- Mode very useful

You must remember that though the measures of central tendency are useful in many circumstances - it does not tell you much if anything about the distribution of the data.