Central Tendency

- In statistics, Central tendency is a central or typical value for a probability distribution. It may also be called as centre of location of the distribution.
- The most common measures of central tendency is mean, median and mode.
- A central tendency can be calculated for either a set of finite values or for a theoretical distribution such as normal distribution.

Measures of central tendency

- Measures means methods and central tendency means average value.

- A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within the set of data.

- In statistics, we always find the measure of central tendency because it always make sense to compare individual scores to overall group of scores in order to correctly interpret the results.

- Measures of central tendency are numbers that tend to cluster around the “middle” of a set of values. Three such middle numbers are the **mean,** the **median,** and the **mode**.
- For example, suppose your earnings for the past week were the values shown in Table 1.
- You could express your daily earnings from Table 1 in a number of ways. One way is to use the average, or **mean,** of the data set. The arithmetic mean is the sum of the measures in the set divided by the number of measures in the set. Totaling all the measures and dividing by the number of measures, you get $1,000 ÷ 5 = $200.

Table 1. Earnings for the Past Week

Day	Amount
Monday	$350
Tuesday	$150
Wednesday	$100
Thursday	$350
Friday	$50

- Another measure of central tendency is the **median,** which is defined as the middle value when the numbers are arranged in increasing or decreasing order. When you order the daily earnings shown in Table 1, you get \$50, \$100, \$150, \$350, and \$350. The middle value is \$150; therefore, \$150 is the median.If there is an even number of items in a set, the median is the average of the two middle values. For example, if we had four values—4, 10, 12, and 26—the median would be the average of the two middle values, 10 and 12; in this case, 11 is the median. The median may sometimes be a better indicator of central tendency than the mean, especially when there are **outliers,** or extreme values.

Example

- Given the four annual salaries of a corporation shown in Table 2, determine the mean and the median.
- The mean of these four salaries is $275,000. The median is the average of the middle two salaries, or $40,000.
- In this instance, the median appears to be a better indicator of central tendency because the CEO's salary is an extreme outlier, causing the mean to lie far from the other three salaries.

Table 2. Four Annual Salaries

Position	Salary
CEO	$1,000,000
Manager	$50,000
Administrative assistant	$30,000
Custodian	$20,000

- Another indicator of central tendency is the **mode,** or the value that occurs most often in a set of numbers. In the set of weekly earnings in Table 1, the mode would be $350 because it appears twice and the other values appear only once.

Notation

- The mean of a sample is typically denoted as (read as *x* bar). The mean of a population is typically denoted as μ (pronounced mew). The sum (or total) of measures is typically denoted with a Σ. The formula for a sample mean is:

$$\bar{x} = \frac{\sum x}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

  - where *n* is the number of values.
- The formula for a sample mean for grouped data is:

$$\bar{x} = \frac{\sum fx}{n}$$

  - where *x* is the midpoint of the interval, *f* is the frequency for the interval, *fx* is the product of the midpoint times the frequency, and *n* is the number of values.

- Occasionally, you may have data that consist not of actual values but rather of **grouped measures**.
- For example, you may know that, in a certain working population,
  - 32 percent earn between $25,000 and $29,999;
  - 40 percent earn between $30,000 and $34,999;
  - 27 percent earn between $35,000 and $39,999; and
  - 1 percent earn between $80,000 and $85,000. This type of information is similar to that presented in a frequency table.
- Although you do not have precise individual measures, you still can compute measures for **grouped data,** data presented in a frequency table.

- Substituting into the formula:
- $\bar{x} = \frac{\sum fx}{n} = \frac{486}{32} = 15.19$

  Therefore, the average price of items sold was about $15.
  may not be an
  for the data, because the
  actual values are not always known for grouped data.

Table 3. Distribution of the Prices of Items Sold at a Garage Sale

Class Interval	Frequency (f)	Midpoint (x)	fx
$1.00 to $5.99	8	3	24
$6.00 to $10.99	6	8	48
$11.00 to $15.99	4	13	52
$16.00 to $20.99	2	18	36
$21.00 to $25.99	4	23	92
$26.00 to $30.99	6	28	168
$31.00 to $35.99	2	33	66
	n= 32		Sigma *fx* = 486

Median for grouped data

- As with the mean, the median for grouped data may not necessarily be computed precisely because the actual values of the measurements may not be known. In that case, you can find the particular interval that contains the median and then approximate the median.
- Using Table 3, you can see that there is a total of 32 measures. The median is between the 16th and 17th measure; therefore, the median falls within the \$11.00 to \$15.99 interval.

- The formula for the best approximation of the median for grouped data is: $$\text{median} = L + \frac{w}{f_{\text{med}}}\left(0.5n - \sum f_b\right)$$

- where L is the lower class limit of the interval that contains the median, n is the total number of measurements, w is the class width, f_{med} is the frequency of the class containing the median, and Σf_b is the sum of the frequencies for all classes before the median class.

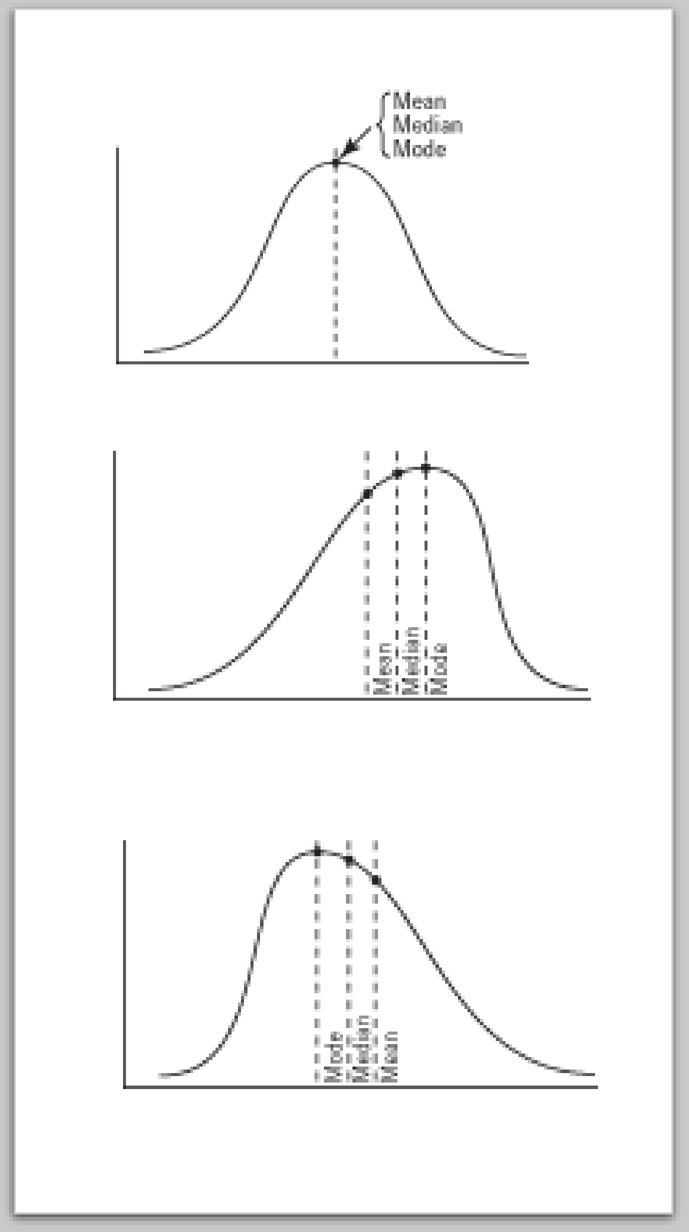
- As we already know, the median is located in class interval \$11.00 to \$15.99. So L = 11, n = 32, w = 4.99, f_{med} = 4, and Σf_b = 14.
- Substituting into the formula:

$$\begin{aligned}
\text{median} &= L + \frac{w}{f_{\text{med}}}\left(0.5n - \sum f_b\right) \\
&= 11 + \frac{4.99}{4}(0.5(32) - 14) \\
&= 11 + \frac{4.99}{4}(16 - 14) \\
&= 11 + \frac{4.99}{4}(2) \\
&= 11 + 2.495 \\
&= 13.495
\end{aligned}$$

Table 4. Distribution of Prices of Items Sold at a Garage Sale

Class Boundaries	Frequency (f)
\$1.00 to \$5.99	8
\$6.00 to \$10.99	6
\$11.00 to \$15.99	4
\$16.00 to \$20.99	2
\$21.00 to \$25.99	4
\$26.00 to \$30.99	6
\$31.00 to \$35.99	2
	n = 32

Symmetric distribution



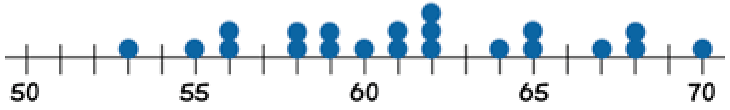
- In a distribution displaying perfect symmetry, the mean, the median, and the mode are all at the same point, as shown in Figure. For a symmetric distribution, mean, median, and mode are equal.
- An outlier can significantly alter the mean of a series of numbers, whereas the median will remain at the center of the series. In such a case, the resulting curve drawn from the values will appear to be skewed, tailing off rapidly to the left or right.
- In the case of negatively skewed or positively skewed curves, the median remains in the center of these three measures.
- A negatively skewed distribution, mean < median < mode.
- A positively skewed distribution, mode < median < mean.

MEAN,MEDIAN MODE FOR UN-GROUPED data

59, 65, 61, 62, 53, 55, 60, 70, 64, 56, 58, 58, 62, 62, 68, 65, 56, 59, 68, 61, 67

$$\textbf{Mean} = \frac{59+65+61+62+53+55+60+70+64+56+58+58+62+62+68+65+56+59+68+61+67}{21}$$

$$= 61.38095...$$



In this case the median is the 11th number:

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, 61, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

Median = 61

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, 61, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

62 appears three times, more often than the other values, so **Mode = 62**

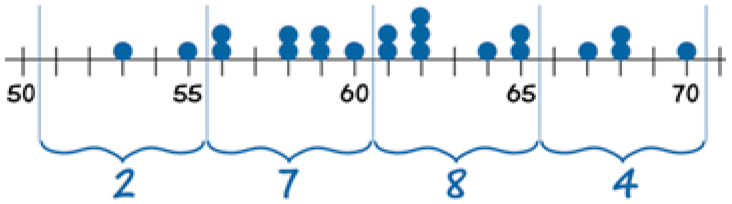
MEAN – 61.3
MEDIAN – 61
MODE - 62

MEAN FOR GROUPED DATA

53, 53, 58, 58, 58, 58, 58, 58, 58, 63, 63, 63, 63, 63, 63, 63, 63, 68, 68, 68, 68

Seconds	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4

Midpoint	Frequency
53	2
58	7
63	8
68	4

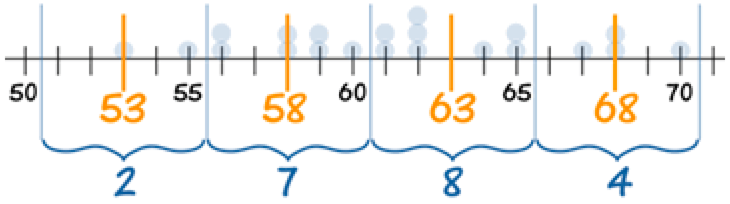


Midpoint x	Frequency f	Midpoint × Frequency fx
53	2	106
58	7	406
63	8	504
68	4	272
Totals:	**21**	**1288**

Estimated Mean $= \frac{1288}{21} =$ **61.333...**

MEAN – 61.3
MEDIAN – 61
MODE - 62

Seconds	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4



The median is the middle value, which in our case is the 11th one, which is in the 61 - 65 group:

We can say "the **median group** is 61 - 65"

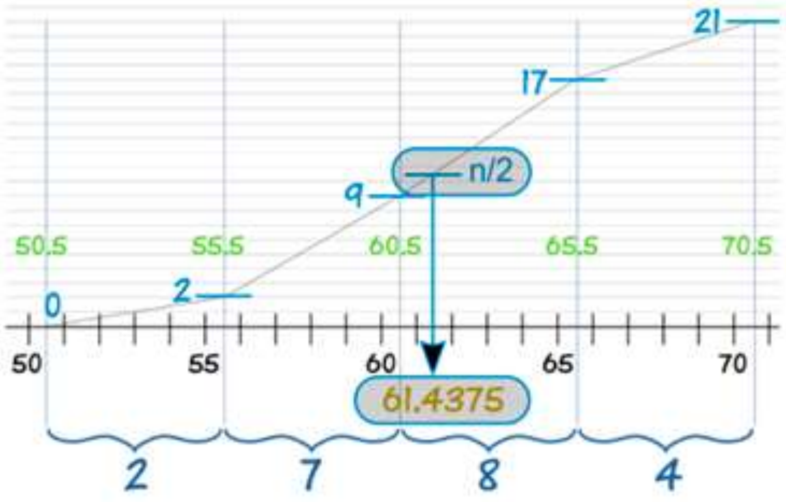
But if we want an estimated **Median value** we need to look more closely at the 61 - 65 group.

- **L** = 60.5
- **n** = 21
- **B** = 2 + 7 = 9
- **G** = 8
- **w** = 5

Estimated Median $= 60.5 + \dfrac{(21/2) - 9}{8} \times 5$

$= 60.5 + 0.9375$

$= \mathbf{61.4375}$



Estimated Median $= L + \dfrac{(n/2) - B}{G} \times w$

- **L** is the lower class boundary of the group containing the median
- **n** is the total number of values
- **B** is the cumulative frequency of the groups before the median group
- **G** is the frequency of the median group
- **w** is the group width

MEAN – 61.3

MEDIAN – 61

MODE - 62

Seconds	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4

We can say "the **modal group** is 61 - 65"

But the actual **Mode** may not even be in that group! Or there may be more than one mode. Without the raw data we don't really know.

$$\textbf{Estimated Mode} = L + \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \times w$$

- L is the lower class boundary of the modal group
- f_{m-1} is the frequency of the group before the modal group
- f_m is the frequency of the modal group
- f_{m+1} is the frequency of the group after the modal group
- w is the group width

- $L = 60.5$
- $f_{m-1} = 7$
- $f_m = 8$
- $f_{m+1} = 4$
- $w = 5$

$$\textbf{Estimated Mode} = 60.5 + \frac{8 - 7}{(8 - 7) + (8 - 4)} \times 5$$

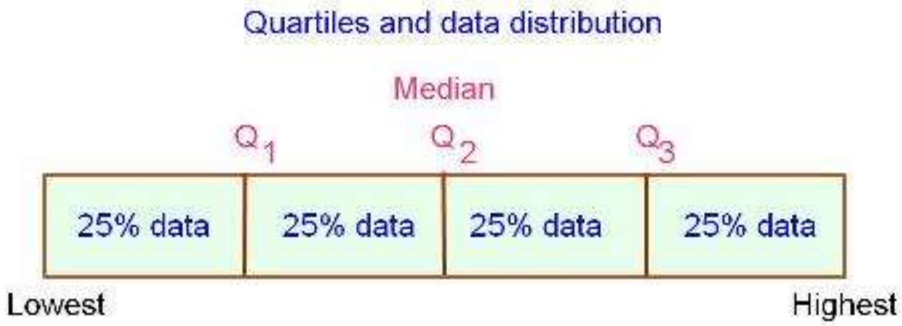
$$= 60.5 + (1/5) \times 5$$

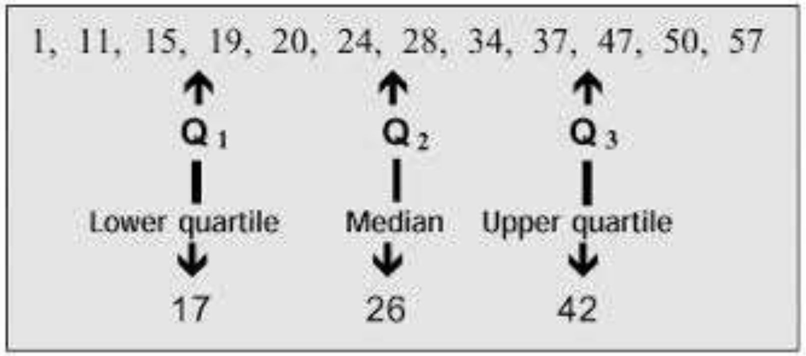
$$= \mathbf{61.5}$$

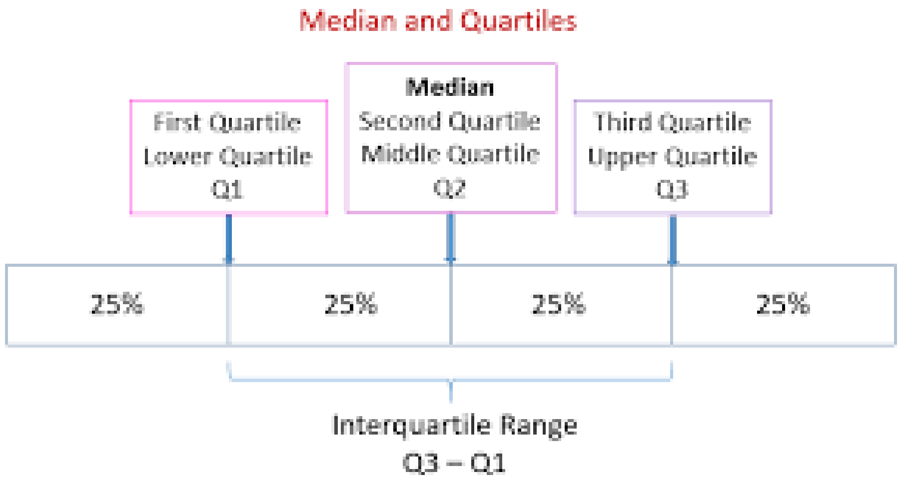
MEAN – 61.3
MEDIAN – 61
MODE - 62

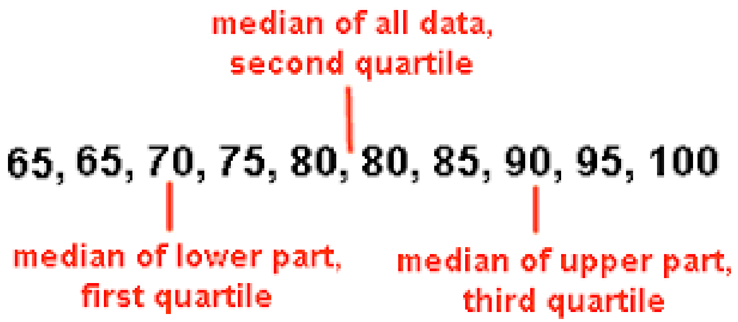
Quartiles

Quartiles in statistics are values that divide your data into quarters

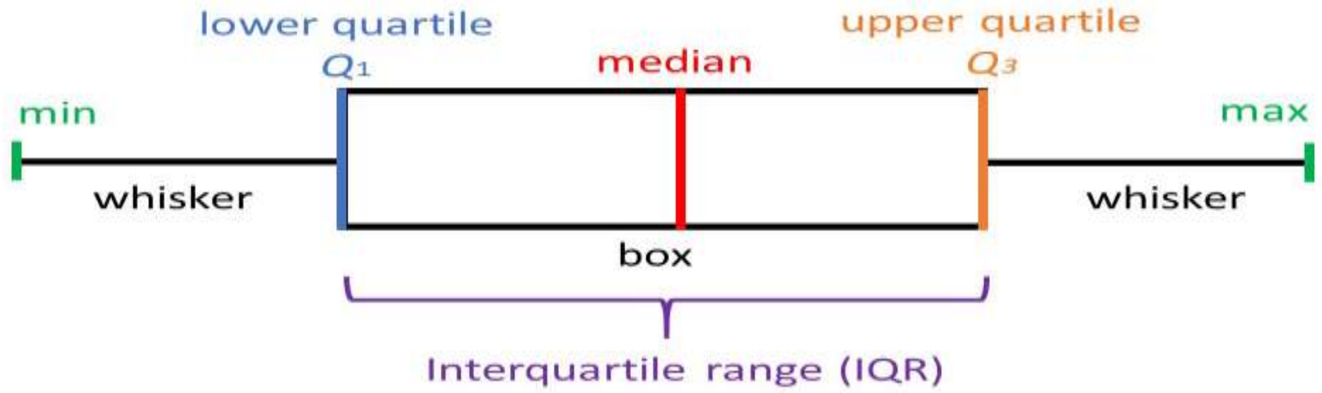








BOX AND WHISKER PLOT



- A boxplot, also called a box and whisker plot, is a way to show the spread and centers of a data set.
- Measures of spread include the interquartile range and the mean of the data set.
- Measures of center include the mean or average and median (the middle of a data set).
- It displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum.

BOX PLOT EXAMPLE

- A sample of 101010 boxes of raisins has these weights (in grams):
- 25, 28, 29, 29, 30, 34, 35, 35, 37, 38
- **Make a box plot of the data.**

Step 1: Order the data from smallest to largest.

Our data is already in order.

25, 28, 29, 29, 30, 34, 35, 35, 37, 38

Step 2: Find the median.

The median is the mean of the middle two numbers:

25, 28, 29, 29, **30**, **34**, 35, 35, 37, 38

$$\frac{30+34}{2}=32$$

Step 3: Find the quartiles.

The first quartile is the median of the data points to the *left* of the median.

25, 28, **29**, 29, 30

$Q_1 = 29$

The third quartile is the median of the data points to the *right* of the median.

34, 35, **35**, 37, 38

$Q_3 = 35$

Step 4: Complete the five-number summary by finding the min and the max.

The min is the smallest data point, which is 25.

The max is the largest data point, which is 38.

The five-number summary is 25, 29, 32, 35, 38.