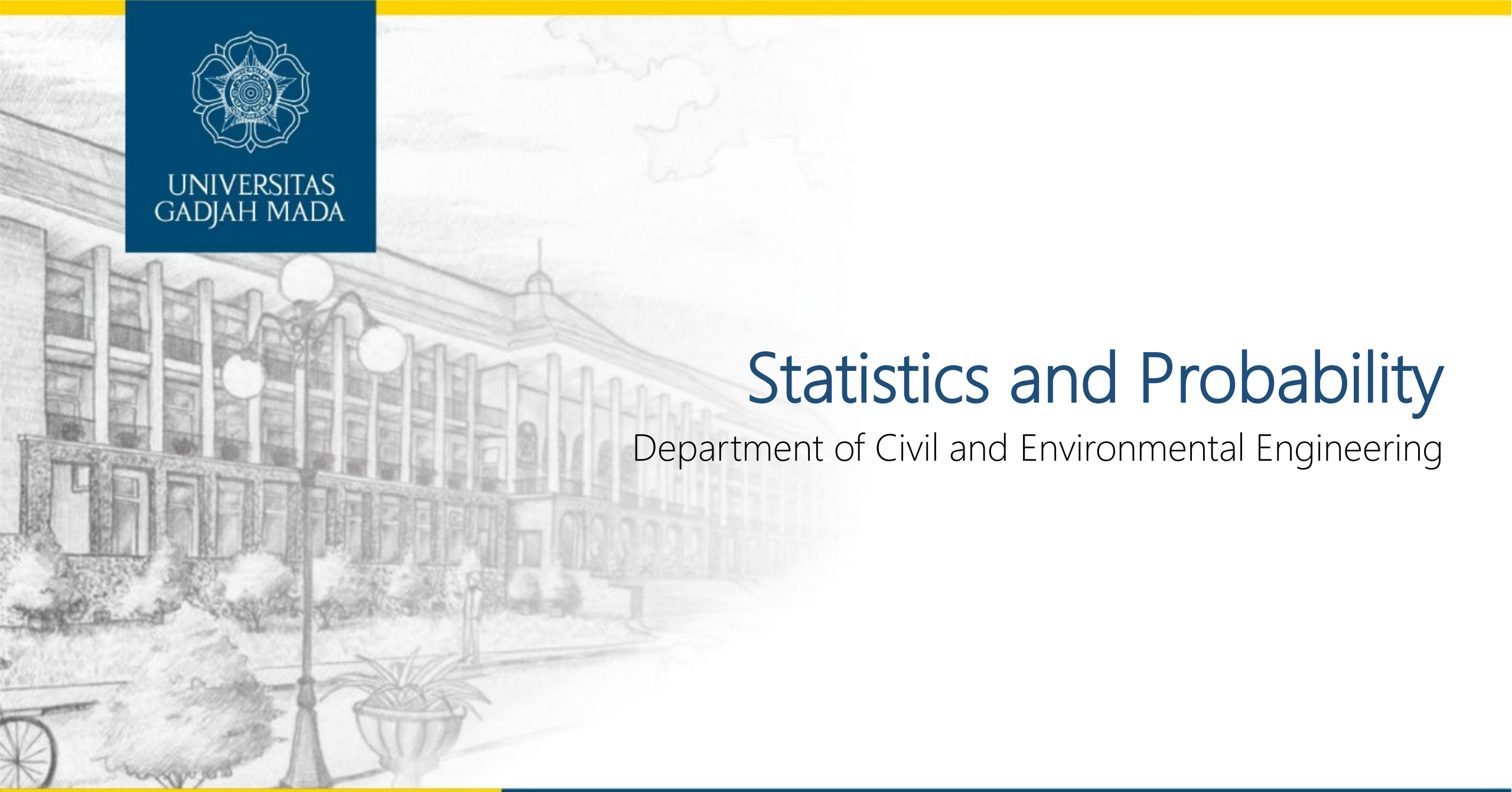


UNIVERSITAS
GADJAH MADA

Statistics and Probability

Department of Civil and Environmental Engineering

Java 46

Outside 6

Partner

Friendzone

Single 52

1st choice 46

2nd or 3rd choice 6

Measure of Central Tendency

- Mean
- Median
- Mode

Measure of Dispersion

- Range
- Variance
- Standard deviation

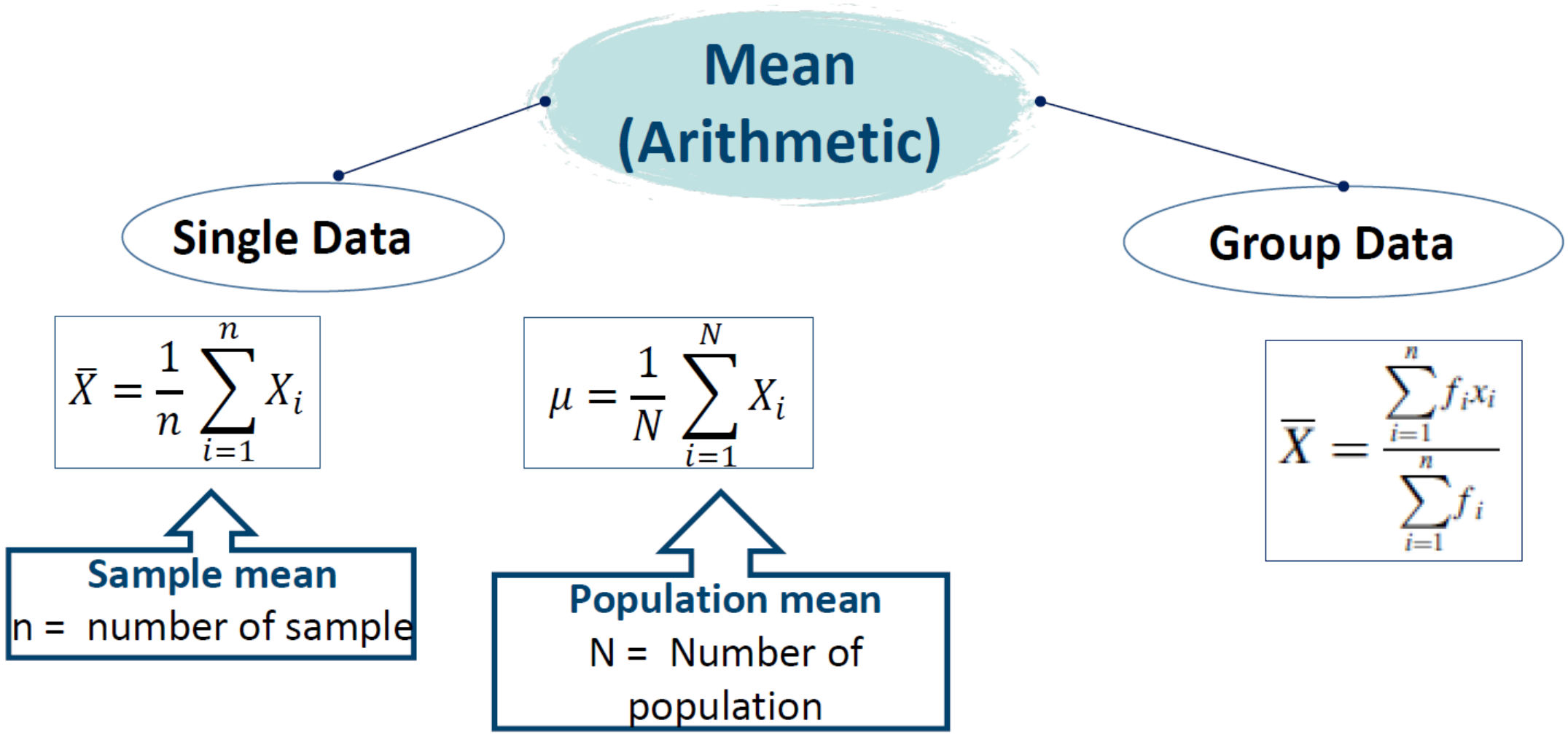
Measure of Individual in a Population

- Z-score
- Percentile rank

Measure of Central Tendency

- Mean
- Median
- Mode

Mean (Arithmetic)



Mean (Arithmetic)

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Data on the number of vehicles queuing up on a road segment in 10 stops (red light):

10, 12, 11, 15, 13, 35, 41, 23, 20

Mean (Arithmetic)

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Data on the age of members of the Engineers organization in Yogyakarta and Central Java Region

Age(in years)	20-30	30-40	40-50	50-60	60-70	70-80	80-90	Total
No. of members	3	61	132	153	140	51	2	542

Weighted Mean

$$\bar{X} = \frac{\sum_{i=1}^{n} (w_i . X_i)}{\sum_{i=1}^{n} w_i}$$

	Population	% not owning a car
Country A	20 million	5%
Country B	500 thousand	30%
Country C	1 million	16%

Geometric Mean

$$\bar{X} = \left(\prod_{i=1}^{n} X_i \right)^{1/n}$$

Harmonic Mean

$$\bar{X} = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

When not to use the mean

The mean has one main disadvantage:

it is particularly susceptible to the influence of outliers. These are values that are unusual compared to the rest of the data set by being especially small or large in numerical value.

Example:

Data on the number of vehicles queuing up on a road segment in 10 stops (red light):

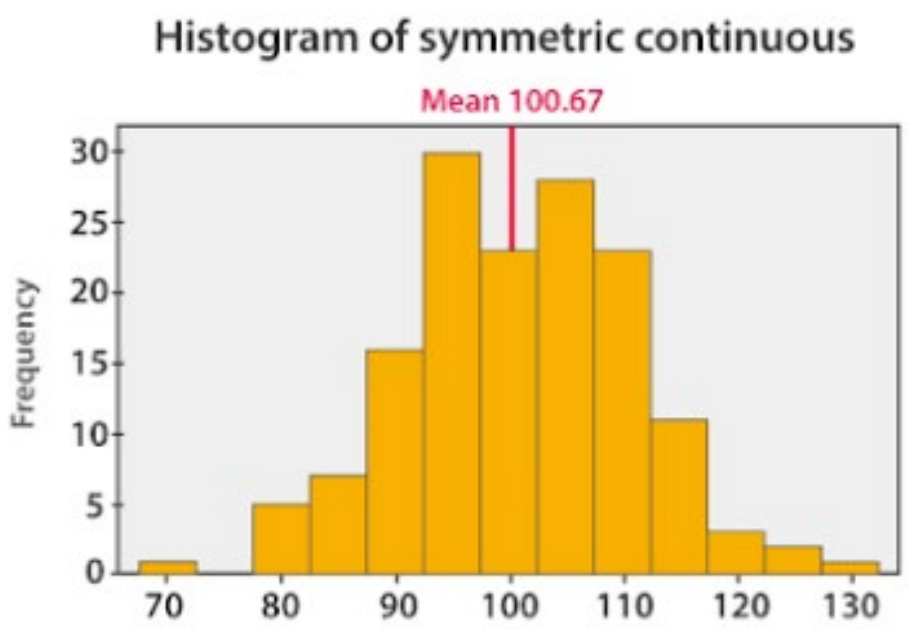
10, 12, 11, 15, 13, 35, 41, 23, 20 , 100

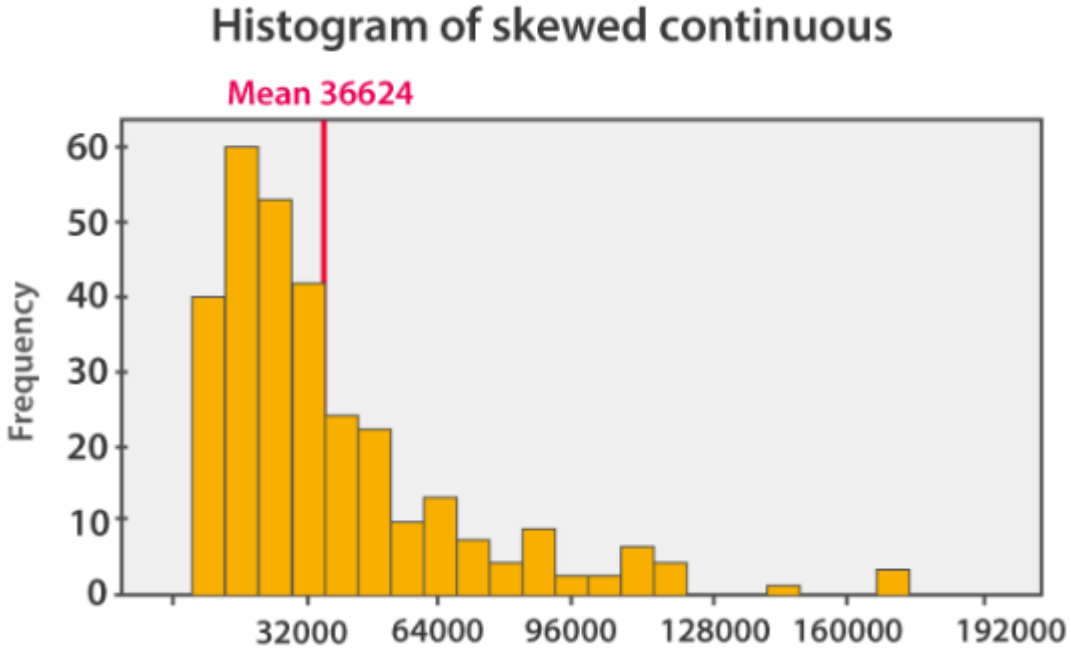
Data on the number of vehicles queuing up on a road segment in 10 stops (red light):

10, 12, 11, 15, 13, 35, 41, 23, 20 , 100

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}{}_i = 20$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}{}_i = 28$$


Histogram of symmetric continuous
Mean 100.67
Frequency
0
5
10
15
20
25
30
70
80
90
100
110
120
130


Histogram of skewed continuous
Mean 36624
Frequency
0
10
20
30
40
50
60
32000
64000
96000
128000
160000
192000

Median

Single Data

Odd

$$Me = x_{\left(\frac{n+1}{2}\right)}$$

Even

$$Me = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

Median (Odd)

It is known that the number of cracks in the concrete wall in 7 simple houses after an earthquake with a magnitude of 6 Mw is as follows:

6, 2, 5, 7, 9, 13, 4

Determine the median of the data

$$Me = x_{\left(\frac{n+1}{2}\right)}$$

2 4 5 6 7 9 13

$$x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{8}{2}\right)} = x_4 = 6$$

Median (Even)

It is known that the number of cracks in the concrete wall in 8 simple houses after an earthquake with a magnitude of 6 Mw is as follows:

6, 2, 5, 7, 9, 13, 4, 11

Determine the median of the data

$$Me = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

2 4 5 6 7 9 11 13

$$\frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} = \frac{x_4 + x_5}{2} = \frac{6+7}{2} = \frac{13}{2} = 6,5$$

Median

Population data

$$Median = Me = Tb + \frac{\frac{n}{2} - F}{f_{me}} c$$

The rainfall data for 29 years is known as follows. Determine the median.

Interval	No. of Observations
38–42	3
42–46	7
46–50	5
50–54	5
54–58	3
58–62	3
62–66	1
66–70	2

Mode

The mode represents the frequently occurring value in the dataset.

Data on the number of vehicles queuing up on a road segment in 11 turns (red lights):

a.	14 20 11 9 8 18 11 25 13 19 21 50	1 Mode
b.	14 20 11 9 8 18 11 25 13 9 21 50	More than 1 mode
c.	14 20 11 9 8 18 17 25 13 19 21 50	There are no mode

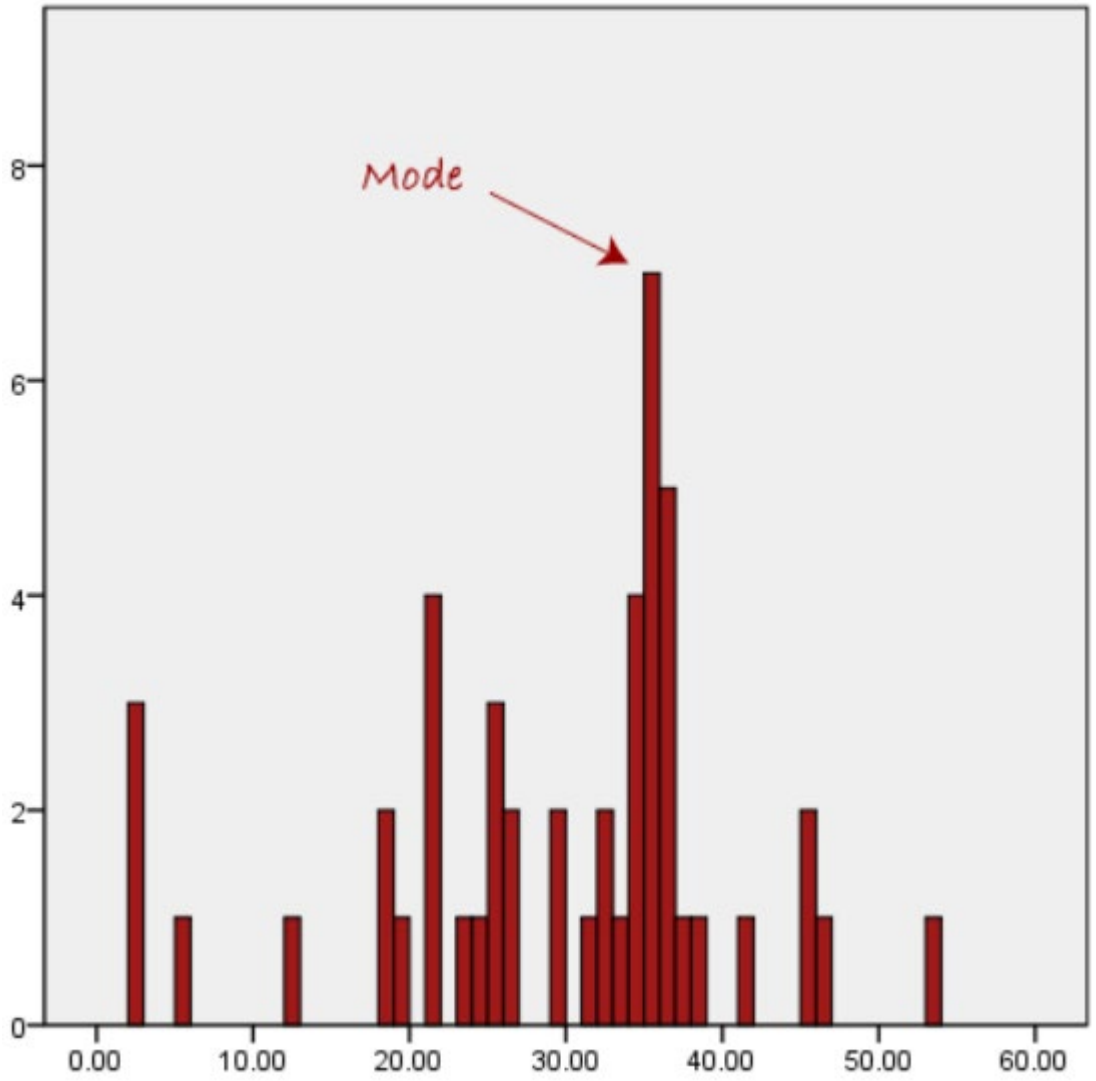
Mode

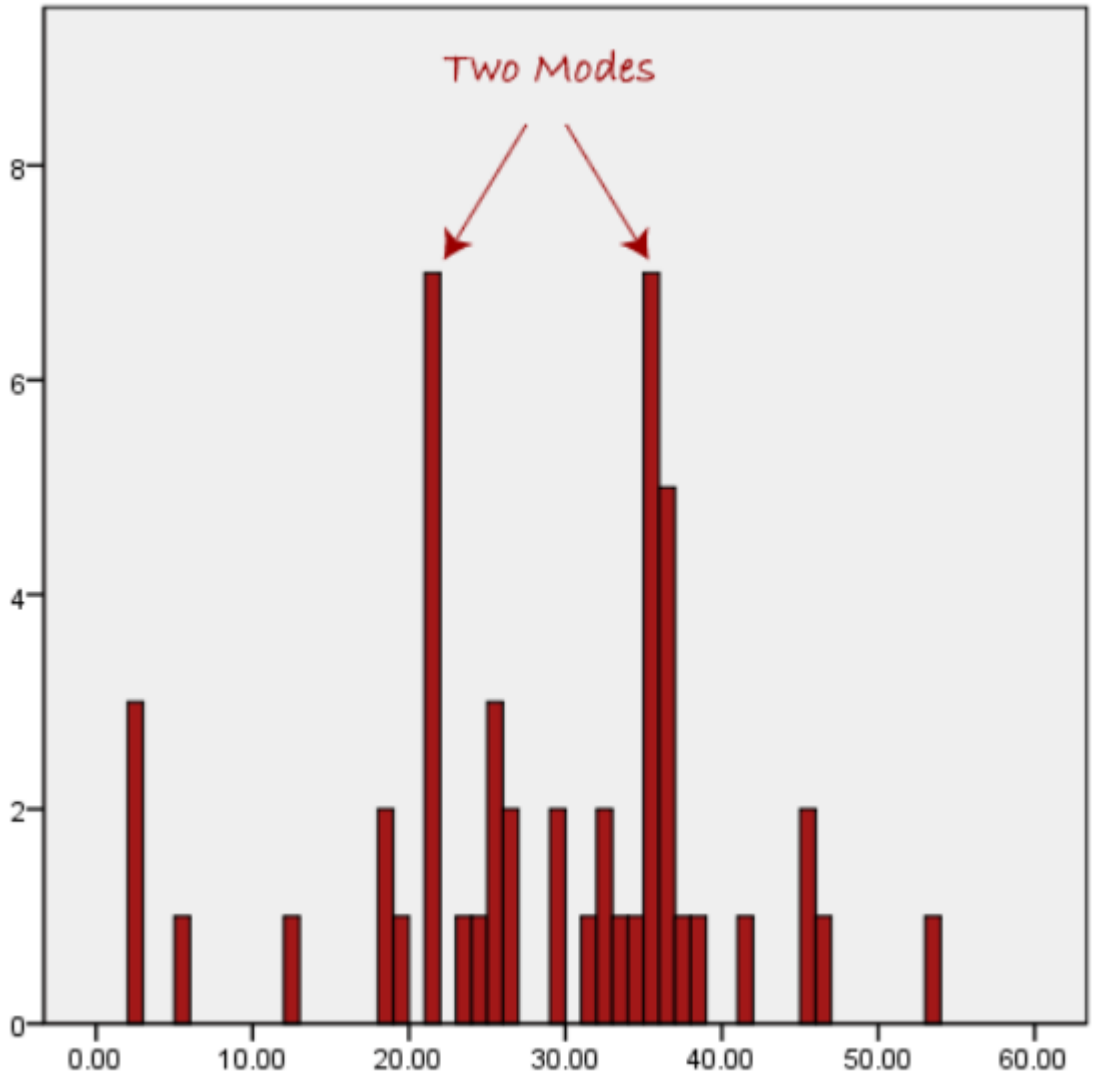
Population data

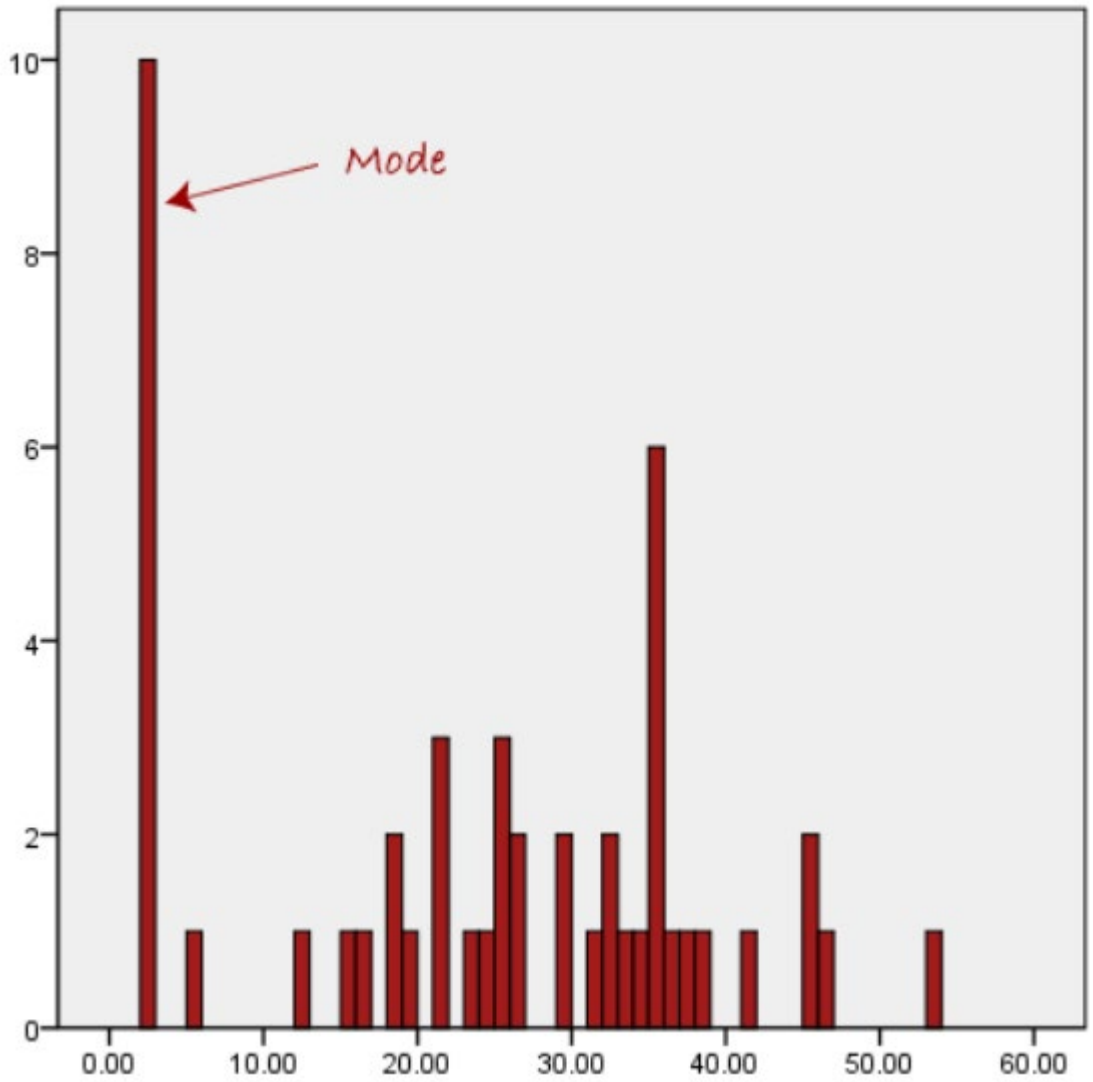
$$Modus = Mo = Tb + \frac{d_1}{d_1 + d_2}\, c$$

The rainfall data for 29 years is known as follows. Determine the mode.

Interval	No. of Observations
38–42	3
42–46	7
46–50	5
50–54	5
54–58	3
58–62	3
62–66	1
66–70	2

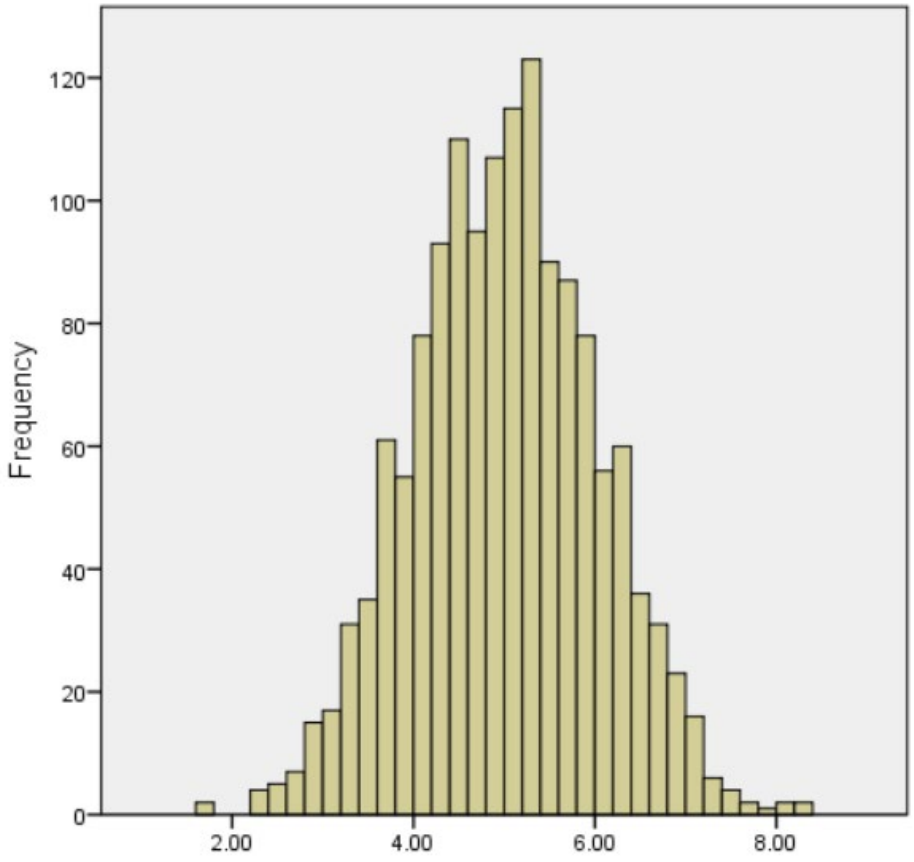

Mode
8
6
4
2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00

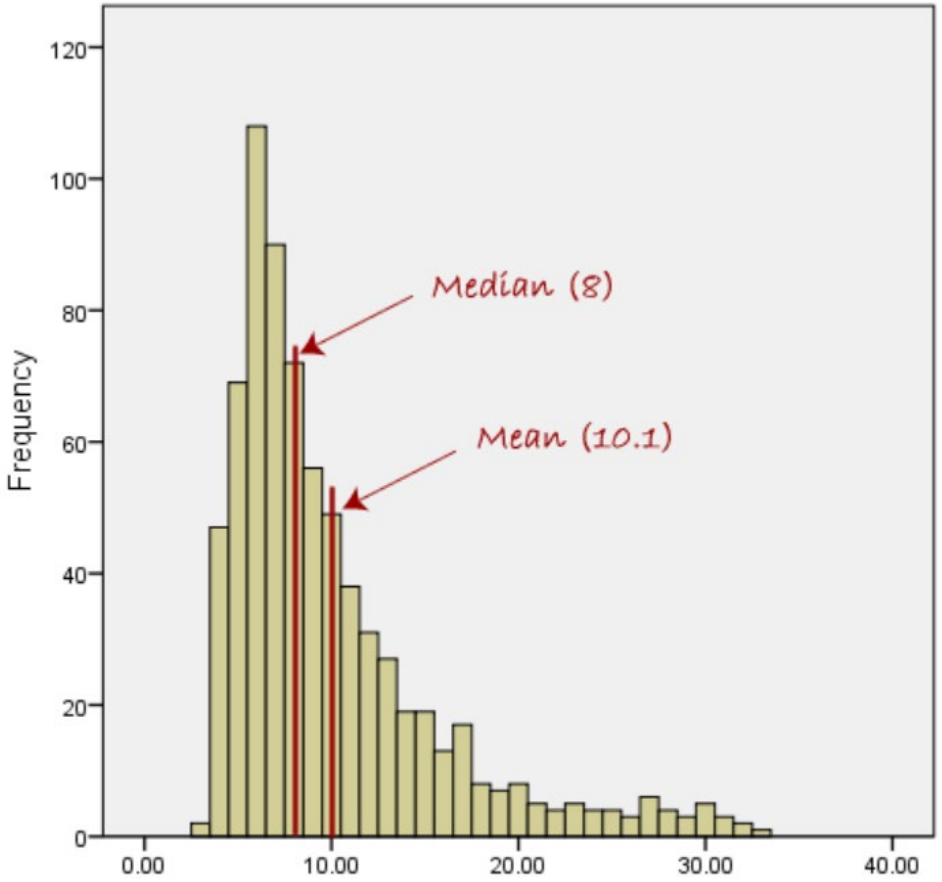

Two Modes
8
6
4
2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00


Mode
10
8
6
4
2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00

Measure of Central Tendency

- Based on the properties of the data, the measures of central tendency are selected.
- If you have a symmetrical distribution of continuous data, all the three measures of central tendency hold good. But most of the times, the analyst uses the mean because it involves all the values in the distribution or dataset.
- If you have skewed distribution, the best measure of finding the central tendency is the median.
- If you have the original data, then both the median and mode are the best choice of measuring the central tendency.
- If you have categorical data, the mode is the best choice to find the central tendency.


Frequency
120
100
80
60
40
20
0
2.00
4.00
6.00
8.00


Frequency
120
100
80
60
40
20
0
Median (8)
Mean (10.1)
0.00
10.00
20.00
30.00
40.00