Discriptive Statistics

Dr. Siddiqua Mazhar

Mid Michigan College

Sep 9, 2021

Charateristics

- Center
- Variation
- Distribution
- Outliers
- Changes over time

Central Tendency

Centre: The Middle of the data set. There are three ways.

Mean: (Arithmetic Average) Add all the values and divide by the number of values you added.

$$Mean = \frac{\sum X}{No.\ of\ Values}$$

Sample Mean

$$\overline{X} = \frac{\sum X}{n}$$

Population Mean

$$\mu = \frac{\sum X}{N}$$

Symbols

1. $\sum$ = Sum
2. n= No. of items in a Sample.
3. N= No. of items in a Population
4. X= Data Value
5. $\overline{X}$= Sample Mean
6. μ= Population Mean

Sample Data: $\{5.40,\ 1.10,\ 0.42,\ 0.73,\ 0.48,\ 1.10\}$

Median: The Middle of the data set.

Note:

- Data values must be in order.
- Find middle value.
  1. If Odd number of values, the measure is the middle number.
  2. If Even number of values, The median is the mean of the two middle numbers.

Examples

1. $\{1,\ 3,\ 4,\ 5,\ 6,\ 7\}$, M$=4.5$
2. $\{8,\ 3,\ 5,\ 11,\ 13,\ 4,\ 6\}$, Find median?
3. $\{3,\ 4,\ 5,\ 6,\ 8,\ 11,\ 13,\ 412\}$, Find median?
4. $\{5.40,\ 1.10,\ 0.42,\ 0.73,\ 0.48,\ 1.10\}$, Find median?

Note: Mean is effected by the outlier value but median does not.

Mode: The most commonly occuring data value.

- $\{5.40,\ 1.10,\ 0.42,\ 0.73,\ 0.48,\ 1.10\}$, Mode $= 1.10$
- $\{27,\ 27,\ 27,\ 55,\ 55,\ 55,\ 88,\ 88,\ \}$ Mode $=$
- $\{1,\ 2,\ ,4,\ 7,\ 9,\ 10,\ 12\}$ Mode $=\phi$

Midrange: Midway between highest and lowest point.

- $\{5.40,\ 1.10,\ 0.42,\ 0.73,\ 0.48,\ 1.10\}$
- $\{27,\ 27,\ 27,\ 55,\ 55,\ 55,\ 88,\ 88,\ \}$
- $\{1,\ 2,\ ,4,\ 7,\ 9,\ 10,\ 12\}$

Expand your knowledge: Grouped Data

When data are grouped, such as frequency table, we can estimate the mean by the following formula.

Sample Mean for a Frequency Distribution

$$\overline{x} = \frac{\sum xf}{n}$$

where
x is the midpoint of a class.

Mean of a Freq. Distribution

Age	f	X	$f.X$
21-30	28	25.5	714
31-40	30	35.5	1065
41-50	12	45.5	546
51-60	2	55.5	111
61-70	2	65.5	131
71-80	2	75.5	151

Weighted Mean

$$Weighted\ \ average = \frac{\sum xw}{\sum w}$$

where x is a data value and w is the weight assigned to that data value. The sum is taken over all data values.

Example

	w	X	$X.w$
H.W.	15 %	70	10.5
T_1	20 %	90	18
T_2	20 %	68	13.6
T_3	20 %	85	17
F	25 %	95	23.75

Practice Question

The following table shows the grouped data, in classes, for the heights of 50 people.

height in cm	f
120-129.	2
130-139	5
140-149	25
150-159	10
160-169	8

Calculate the mean of the height of 50 people.

How to find Central Tendancy on a calculator

Step 1: Press STAT
Step2: Press 1 Edit — *Enter*
Step3: Enter the values of L1
Step 4: Press STAT
Step 5: Select CALC (Moved by the curser sign) and press enter on option 1: 1-Var Stats
Step 6: Hit the Yellow button '2nd' and press 1 to get L1 and then enter

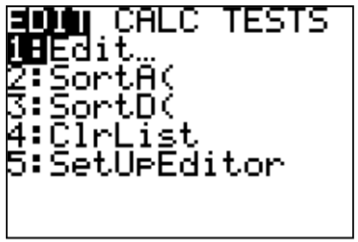
DONE!

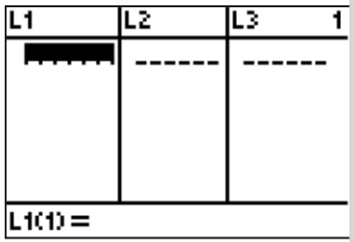
Important Numbers: One Variable

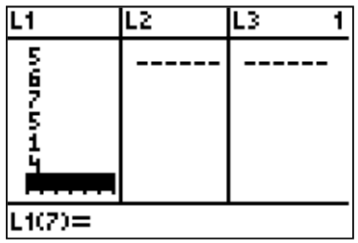
We can find the minimum, maximum, median, quartiles, standard deviation, and mean all in the same place.

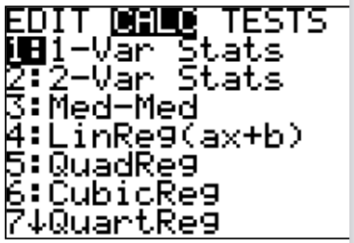
1. Press STAT
2. Press Enter
3. Enter the data into L1. Press enter after each number
4. Press STAT
5. Go over (with the arrow) to Calc
6. Press enter (or hit number 1) for 1-var stats
7. Press Enter again
8. The important numbers appear. To see more, scroll down with the down arrow.

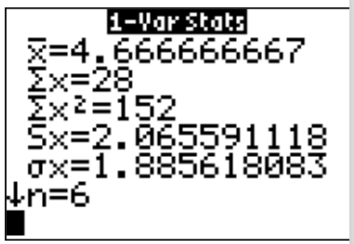
For details on what everything means in 1-var stats, see next page.



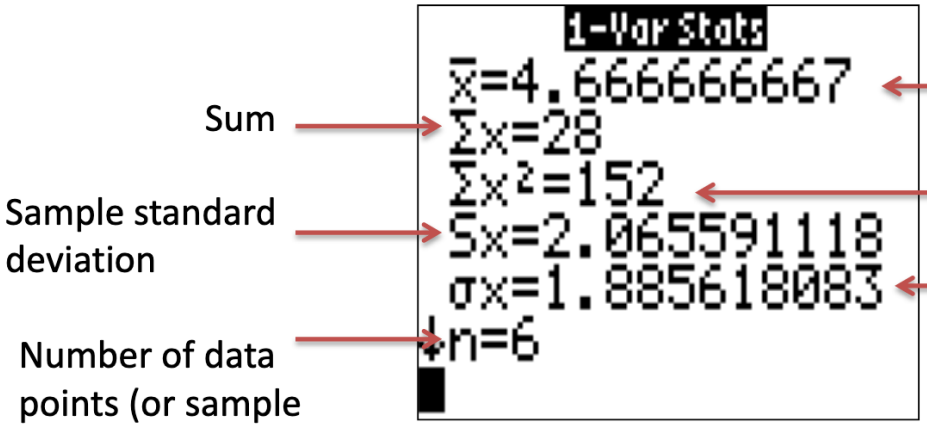








1-Var Stats continued



Sum

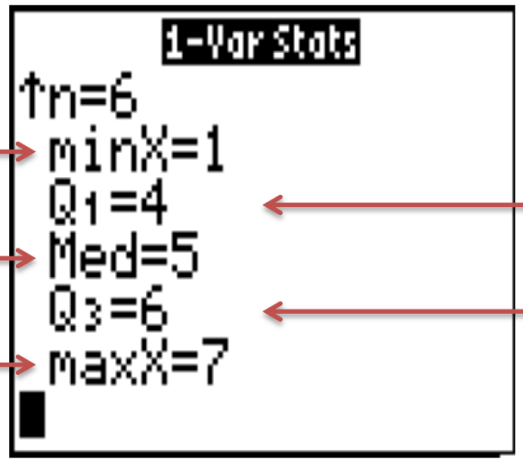
Mean

Sample standard deviation

Sum of squares

Population standard deviation (only use if the data consists of the entire population)

Number of data points (or sample size)



Minimum

First Quartile

Median

Third Quartile

Max

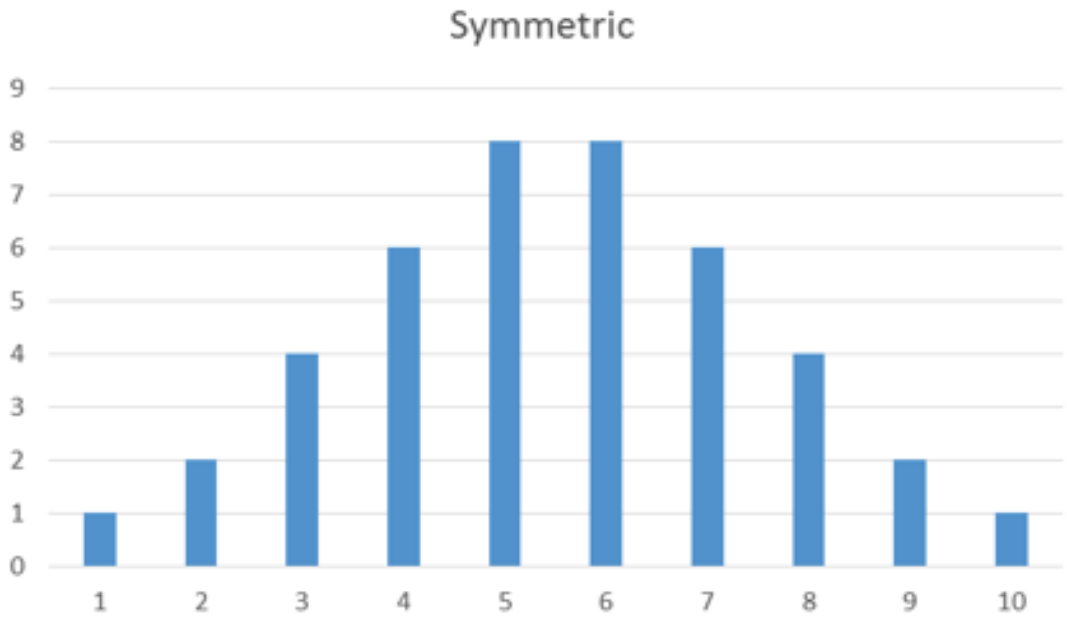
Note: the x means it is for the x variable)

Data Distribution

A histogram can look many ways but the shape of a histogram means something.

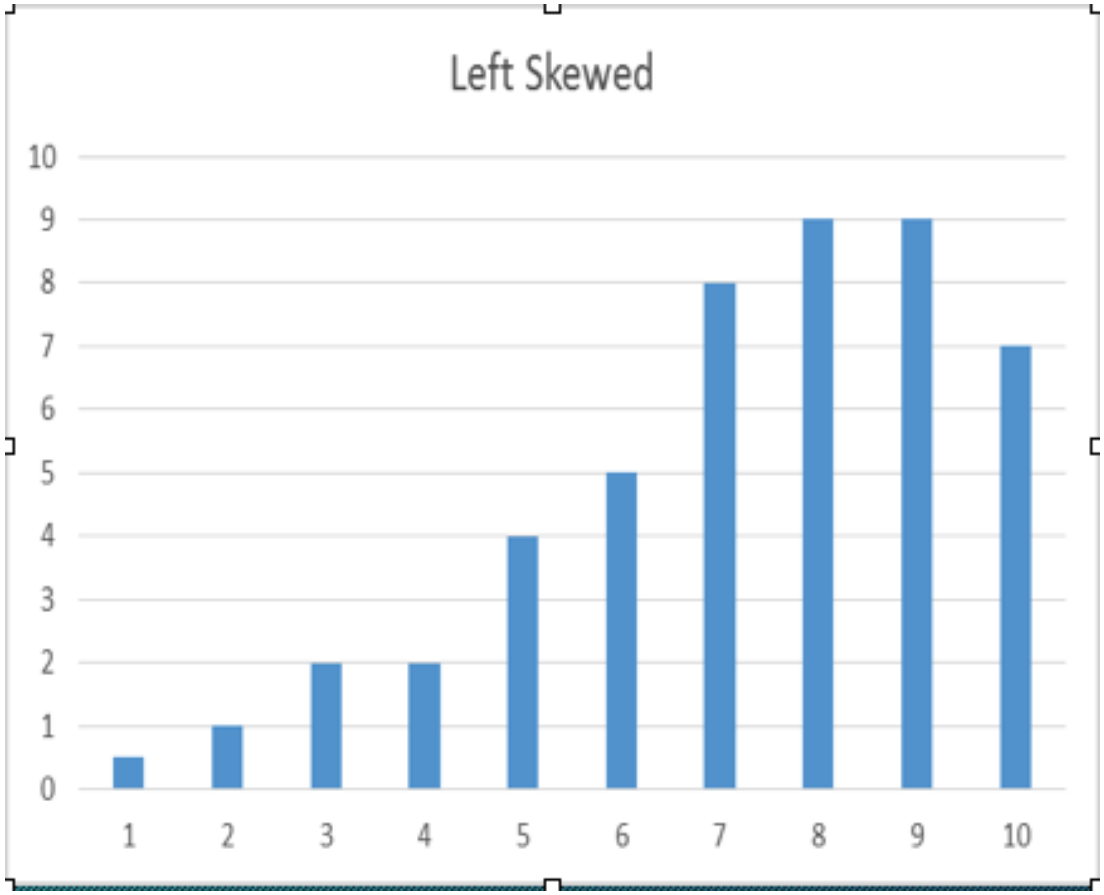
Normal Distribution

Bell curved shape / symmetric . (median = mean)



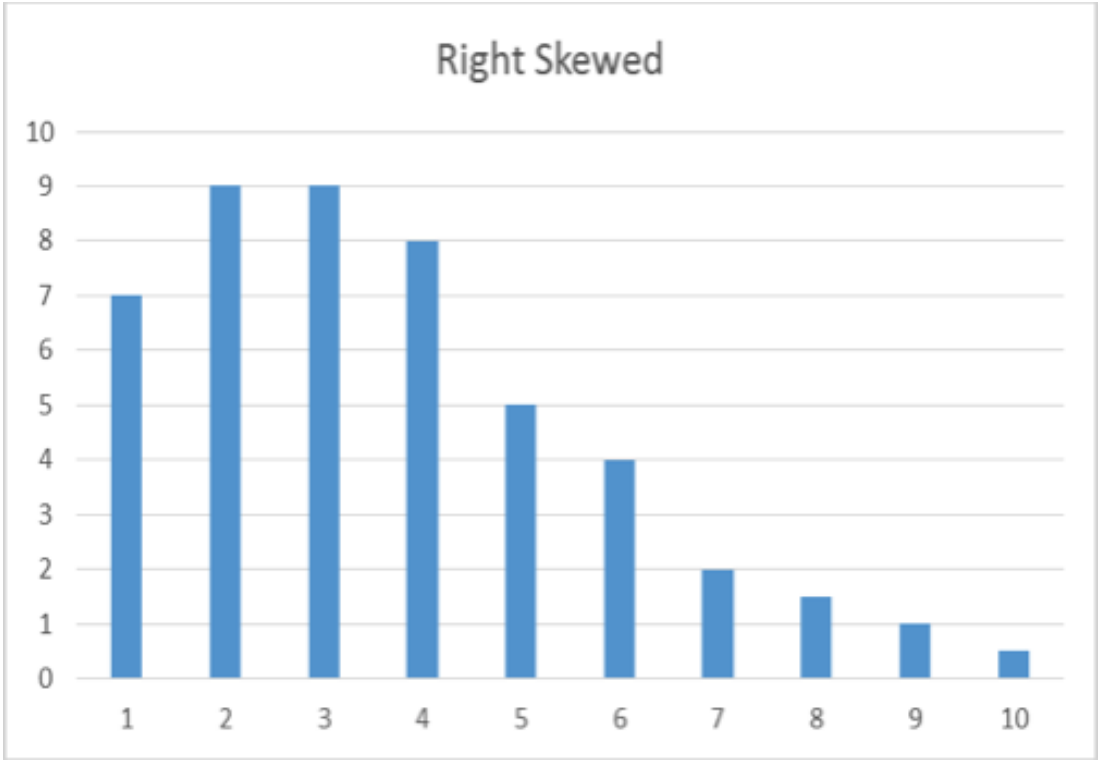
Left skewed

Tail to the left. (mean $<$ median)



Right skewed

Tail to the right. (median $<$ mean)



Variation

Variation means how the data is spread.

The number of customers in the bank waiting in three different lines.

x_1	x_2	x_3	Mean
6	6	6	$\overline{X} =$
4	7	7	$\overline{X} =$
1	3	14	$\overline{X} =$

Range

The is range is the difference between the largest and smallest values of a data distribution.

$$Range = Max\ Value - Min\ Value$$

Note:

- Easy to find.
- Does not consider all values except maximum and minimum value.

Example

For these sets
$S_1 = \{17, 22, 22, 22, 27\}$
$S_2 = \{17, 19, 20, 27, 27\}$
Range is the same $R_1 = 27 - 17 = 10$, and $R_2 = 27 - 17 = 10$,
How do they differ?

Standard Deviation

It measures the average distance your data values are from the mean.

Note:

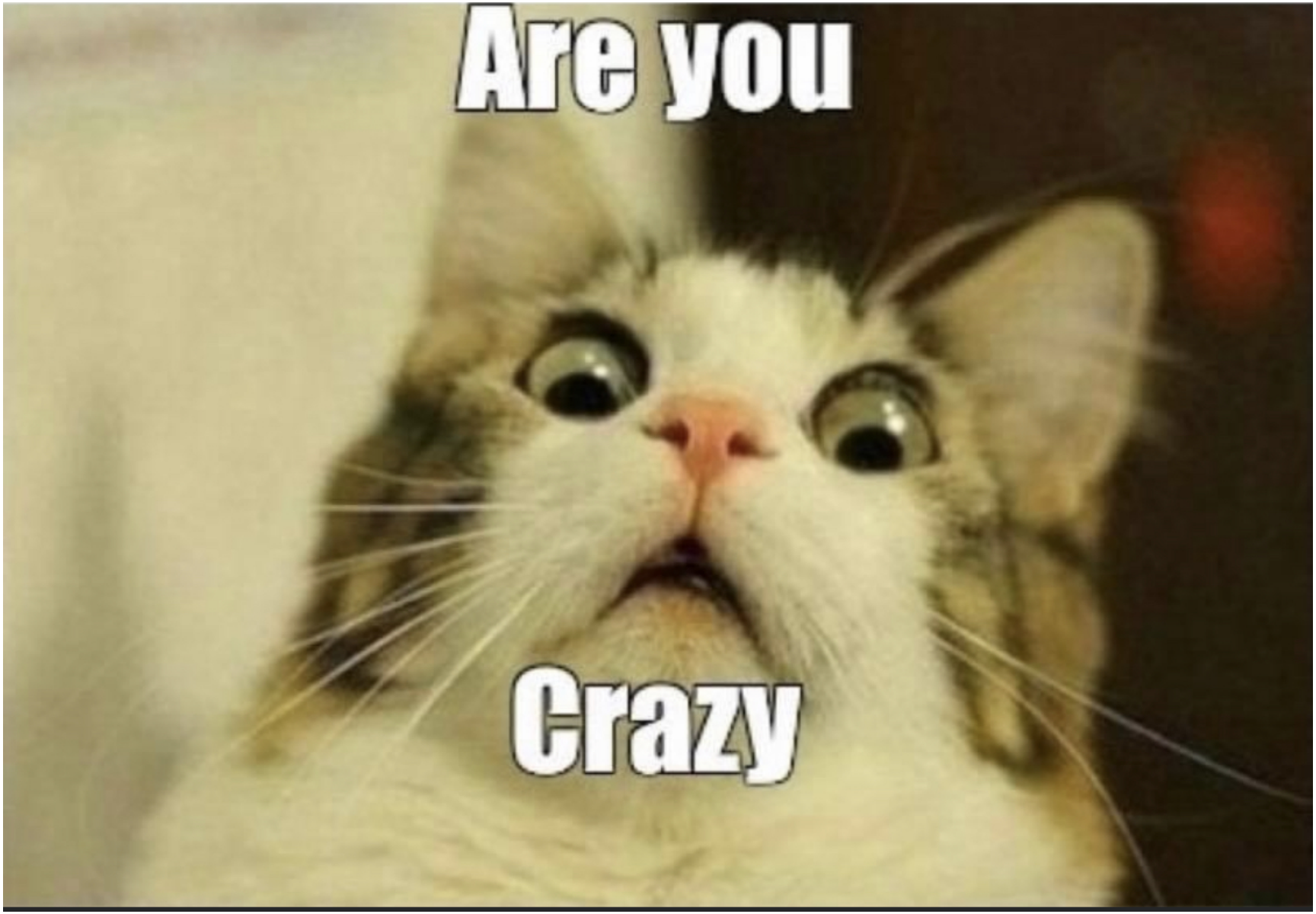
- It can never be negative or Zero, unless all entries are the same.
- Generally effected by Outliers.

Sample standard deviation is denoted by S.

$$S = \sqrt{\frac{\sum(X - \overline{X})^2}{n-1}}$$

OR

$$S = \sqrt{\frac{n\sum(X^2) - (\sum(X))^2}{n(n-1)}}$$


Are you
Crazy

Explanation

Example

Find Standard Deviation of $\{1, 3, 14\}$

X	$X - \overline{X}$	$(X - \overline{X})^2$

Example

Find Standard Deviation of $\{1, 3, 14\}$

X	X^2

Example

Find Standard Deviation of $\{4, 7, 7\}$

X	$X - \overline{X}$	$(X - \overline{X})^2$

Standard Deviation for a population

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

Standard Deviation for a Grouped Data

Sample Standard Deviation for a Frequency Distribution

$$s = \sqrt{\frac{\sum (x - \overline{x})^2 f}{n - 1}}$$

where

x is the midpoint of a class.

Standard Deviation for a Grouped Data

Question

The following table shows the grouped data, in classes, for the heights of 50 people.

height in cm	f
120-129.	2
130-139	5
140-149	25
150-159	10
160-169	8

Calculate the standard deviation of the height of 50 people.

Variance

- Sample Variance: S^2
- Population Variance: σ^2

Note:

- Closely grouped data will have a small standard deviation.
- Spread-out data will have a large standard deviation.

Class Exercise

1. Given the following frequency distribution, find the following things.

Class	frequency
0-4	2
5-9	3
10-14	6
15-19	9
20-24	7
25-29	2
30-24	1

   - Mean
   - Median
   - Mode
   - Standard Deviation
   - Variance