DATA DESCRIPTION

Learning Objective:

- Distinguish different measures of central tendency.
  - Mean, Median, Mode (ungrouped and grouped data)

Parameter Vs. Statistic:

Parameter
- Is a characteristic or measure obtained by using the data values from a specific **population**. - Usually uses Greek letters

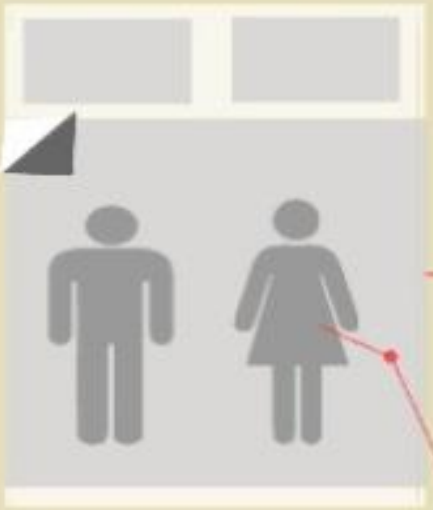
Statistic
- Is a characteristic or measure obtained by using the data values from a **sample.** - Usually uses letters in the English alphabet

MEASURES OF CENTRAL TENDENCY

Mean | Median | Mode

How The Average Person Spends Time in Their Home

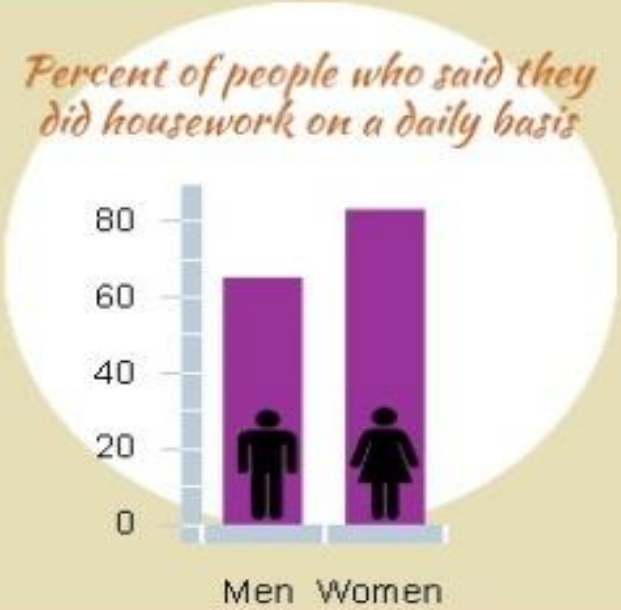
The Average Human American
Aged 15 and Over



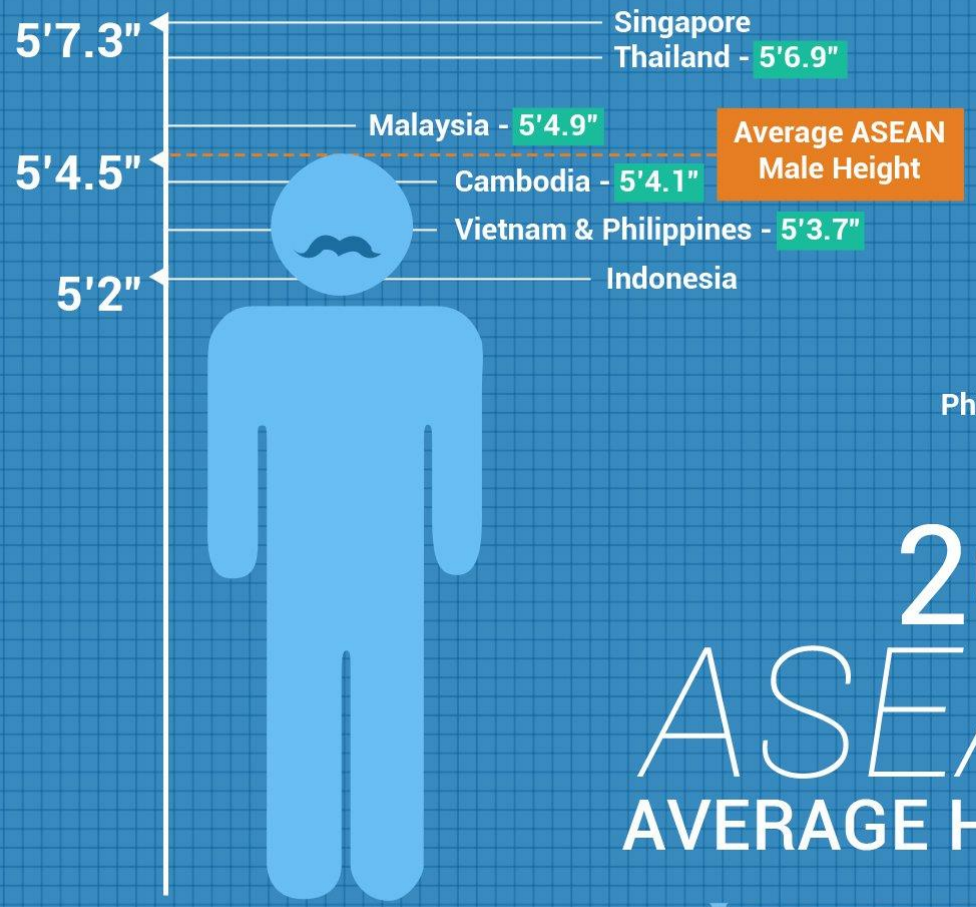
8.6 hours spent sleeping

On average women get an extra 11 minutes sleep

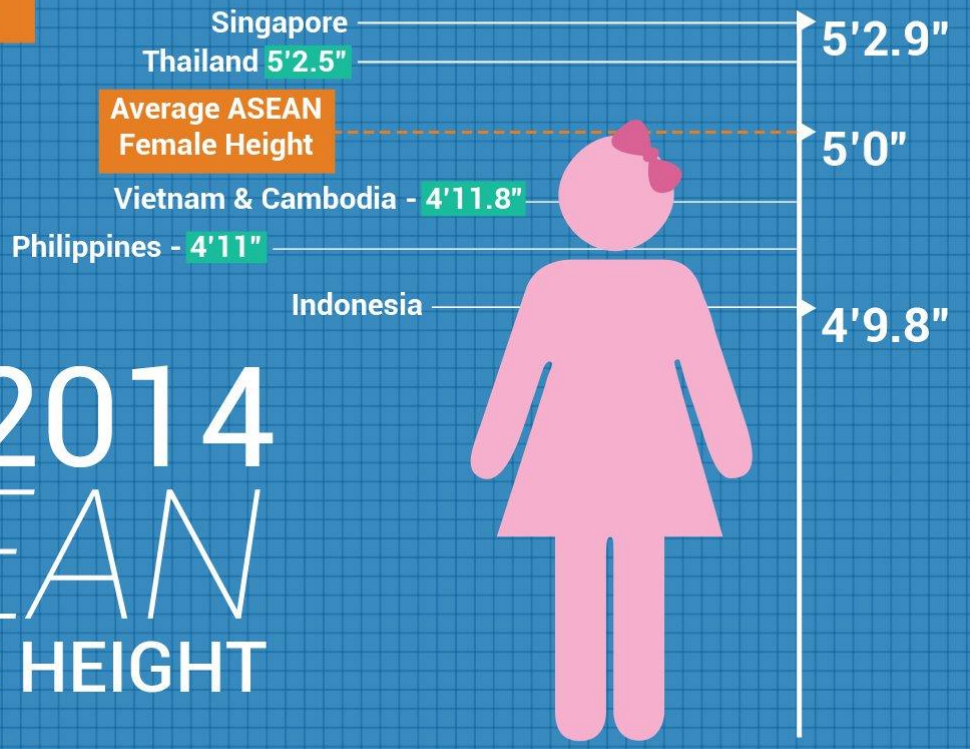
Percent of people who said they did housework on a daily basis



Source: ASEAN DNA | GFX: Mark Diamat



2014 ASEAN AVERAGE HEIGHT



INQUIRER.net

Average Time in WATCHING TV

per peson in ASEAN and OTHERS

SOURCE: ASEANDNA

INFOGRAPHICS by: INQUIRER.net

MEAN

MEAN:

- Arithmetic average
- Sum of the values, divided by the total number of values
- Affected by extremely high or low values called “outliers” and may not be the appropriate average to use in these situations

Population Mean	Sample Mean
$$\mu = \frac{X_1 + X_2 + \ \ldots + X_N}{N} = \frac{\sum X}{N}$$	$$\overline{X} = \frac{X_1 + X_2 + \ \ldots + X_n}{n} = \frac{\sum X}{n}$$
where, μ – population mean X_N– data value N – total number of values in the population	where, $\overline{X}$– sample mean X_n– data value n – total number of values in the sample

MEAN: Example 1

The ages in weeks of a random sample of six kittens at an animal shelter are 3, 8, 5, 12, 14 and 12. Find the average age of this sample.

Sol'n.

$$\bar{X} = \frac{X_1 + X_2 + \ \ldots + X_n}{n}$$

$$= \frac{3 + 8 + 5 + 12 + 14 + 12}{6} = \frac{54}{6} = 9$$

MEAN: Example 2

A small company consists of the owner, the manager, the salesperson, and two technicians. The salaries in pesos are listed as 50 000, 20 000, 12 000, 9 000 and 9 000 respectively. Find the population mean.

(Assume this is the population)

Sol'n.

$$\mu = \frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{\sum X}{N}$$

$$= \frac{\mathbf{50000 + 20000 + 12000 + 9000 + 9000}}{\mathbf{5}}$$

$$= \frac{\mathbf{100000}}{\mathbf{5}} = \mathbf{Php\ 20000}$$

MEAN:

The scores of 25 students on a 4-point seatwork.

Scores	Frequency (*f*)
0	2
1	4
2	12
3	4
4	3

MEAN: FDT

We may use this formula for the mean:

$$\overline{X} = \frac{\sum(f \cdot X)}{n}$$

Where

f – frequency of the corresponding X

n – is the sum of all f

MEAN: FDT

Solution:

Scores	Frequency (*f*)	$f \cdot X$
0	2	0
1	4	4
2	12	24
3	4	12
4	3	12

$$\bar{X} = \frac{\sum(f \cdot X)}{n} = \frac{0 + 4 + 24 + 12 + 12}{25}$$

$$\bar{X} = \frac{52}{25} = 2.08$$

MEAN: Grouped FDT

Given the data below. Find the mean.

Class limits	Frequency (*f*)
16-20	3
21-25	5
26-30	4
31-35	3
36-40	2

MEAN: Grouped FDT

The mean for a grouped frequency distribution is given by:

$$\overline{X} = \frac{\sum(f \cdot X_m)}{n}$$

f – frequency of the corresponding X

X_m – Class Mark / Class middle point

n – is the sum of all f or total number of data values

MEAN: Grouped FDT

Given the data below. Find the mean.

Class limits	f	X_m/CM	$f \cdot X_m$
16-20	3		
21-25	5		
26-30	4		
31-35	3		
36-40	2		

MEAN: Grouped FDT

Given the data below. Find the mean.

Class limits	f	X_m/CM	$f \cdot X_m$
16-20	3	18	
21-25	5	23	
26-30	4	28	
31-35	3	33	
36-40	2	38	

MEAN: Grouped FDT

Given the data below. Find the mean.

Class limits	f	X_m/CM	$f \cdot X_m$
16-20	3	18	54
21-25	5	23	115
26-30	4	28	112
31-35	3	33	99
36-40	2	38	76

MEAN: Grouped FDT

Classes	f	X_m/CM	$f \cdot X_m$
16-20	3	18	54
21-25	5	23	115
26-30	4	28	112
31-35	3	33	99
36-40	2	38	76

$$\bar{X} = \frac{\sum(f \cdot X_m)}{n}$$

$$= \frac{54 + 115 + 112 + 99 + 76}{17}$$

$$\bar{X} = \frac{456}{17} = \mathbf{26.82}$$

Weighted Mean

- It is used when the values are not all equally represented.

$$\bar{X} = \frac{w_1X_1 + w_2X_2 + \ \ldots + w_nX_n}{w_1 + w_2 + \ \ldots + w_n} = \frac{\sum wX}{\sum w}$$

Where,

$w_1 + w_2 + \ \ldots + w_n$ **are the weights and**

$X_1 + X_2 + \ \ldots + X_n$ **are the values.**

Weighted Mean: Example

A student is enrolled in a biology course where the final grade is determined based on the following categories:

tests 40%, final exam 25%, quizzes 25%, and homework 10%. The student has earned the following scores for each category: tests-83, final exam-75, quizzes-90, homework-100. We need to calculate the student's overall grade.

83	40%
75	25%
90	25%
100	10%

Weighted Mean: Example

A student is enrolled in a biology course where the final grade is determined based on the following categories:

tests 40%, final exam 25%, quizzes 25%, and homework 10%. The student has earned the following scores for each category: tests-83, final exam-75, quizzes-90, homework-100. We need to calculate the student's overall grade.

83	40%
75	25%
90	25%
100	10%

$$\bar{X} = \frac{\sum wX}{\sum w}$$

$$= \frac{40(83) + 25(75) + 25(90) + 10(100)}{40 + 25 + 25 + 10}$$

$$= \frac{8445}{100}$$

ANSWER: 84.45 or 84%

MEDIAN

MEDIAN

- Is the middle / halfway point of the data
- Denoted by MD
- The observations must first be arranged in increasing or decreasing order. Then locate the middle value so that half of the observations are less than or equal to that value while the other half are greater than the middle value.

Data Array
When the data set is ordered by ascending and descending order, it is called a data array.

Finding the MEDIAN:

STEP 1. Arrange the data values into a data array.

STEP 2. Determine the no. of values (n) in the data set.

STEP 3.

a. Case 1: **If *n* is odd**, select the middle data.
b. Case 2: **If *n* is even**, find the mean of the two middle values.

MEDIAN: Example 1

The ages of 10 college students are: 18, 24, 20, 35, 19, 23, 26, 23, 19, 20. Find the median.

Solution:

Data array: 18, 19, 19, 20, 20, 23, 23, 24, 26, 35

Since n is even, find the mean of the two middle values.

$$\boldsymbol{MD} = \frac{20 + 23}{2} = \frac{43}{2} = 21.5$$

MEDIAN: Example 2

The scores of 25 students on a 4-point seatwork.

Scores	Frequency (*f*)
0	2
1	4
2	12
3	4
4	3

MEDIAN: Example 2 Solution

The scores of 25 students on a 4-point seatwork.

Scores	Frequency (*f*)	*Cumulative frequency (cf)*
0	2	2
1	4	6
2	12	18
3	4	22
4	3	25

$MD = \frac{25}{2} = 12.5 \approx 13$, find where the 13^{th} value belongs.

MD = 2

MEDIAN: Grouped FDT

The median for a grouped frequency distribution is given by:

$$\hat{x} = LCB_{md} + c\left(\frac{\frac{n}{2} - < cf_b}{f_{md}}\right)$$

Median class: the class interval where <cf is greater than or equal to $\frac{n}{2}$ for the first time.

LCB_{md} – lower class boundary of the median class

c – class size

f_{md} – frequency of the median class

$< cf_b$ – less then cumulative frequency before the median class

n – is the sum of all f or total number of data values

MEDIAN: Grouped FDT

Given the data below. Find the median.

Classes	*f*	LCB	<cf	
10 – 19	5			
20 – 29	8			
30 – 39	11			
40 – 49	12			
50 – 59	6			
total	42			

MEDIAN: Grouped FDT

Given the data below. Find the median.

Classes	*f*	LCB	<cf	
10 – 19	5	9.5		
20 – 29	8	19.5		
30 – 39	11	29.5		
40 – 49	12	39.5		
50 – 59	6	49.5		
total	42			

MEDIAN: Grouped FDT

Given the data below. Find the median.

Classes	*f*	LCB	<cf	
10 – 19	5	9.5	5	
20 – 29	8	19.5	13	
30 – 39	11	29.5	24	median class
40 – 49	12	39.5	36	
50 – 59	6	49.5	42	
total	42			

MEDIAN: Grouped FDT

Classes	f	LCB	<cf
10 – 19	5	9.5	5
20 – 29	8	19.5	13
30 – 39	11	29.5	24
40 – 49	12	39.5	36
50 – 59	6	49.5	42
total	42		

$$\hat{x} = LCB_{md} + c\left(\frac{\frac{n}{2} - {<}cf_b}{f_{md}}\right)$$

$$= 29.5 + 10\left(\frac{\frac{42}{2} - 13}{11}\right)$$

$$\hat{x} = \mathbf{36.77}$$

MODE

MODE

The value that occurs the most in a data set.

Unimodal	When the data set has one mode.
Bimodal	When data set has two modes.
Multimodal	When data set has two or more modes.
No mode	When all frequencies are equal.
Modal class	Is the class with the largest frequency.

MODE: Example 1

The following data represent the duration (in days) of U.S. space shuttle voyages for the years 1992-94. Find the mode.

8, 9, 9, 14, 8, 8, 10, 7, 6, 9, 7, 8, 10, 14, 11, 8, 14, 11.

Solution:

Data array:

6, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 10, 10, 11, 11, 14, 14, 14

Mode = 8

Unimodal

MODE: Example 2

Eleven different automobiles were tested at a speed of 15 mph for stopping distances. The distance, in feet, is given below. Find the mode.

15, 18, 18, 18, 20, 22, 24, 24, 24, 26, 26

Answer:

Mode = 18 and 24

Bimodal

MODE: Example 3

Given the data below. Find the mode.

Scores	Frequency (*f*)
0	2
1	4
2	12
3	4
4	3

Answer: Mode = 2

MODE: Example 4

Given the data below. Find the modal class.

Class limits	f
16-20	3
21-25	5
26-30	4
31-35	3
36-40	2

Answer: Modal Class= 21-25

MODE: Grouped FDT

The mode for a grouped frequency distribution is given by:

$$\boldsymbol{x^{o} = LCB_{mo} + c\left(\frac{f_{mo} - f_{b}}{2f_{mo} - f_{b} - f_{a}}\right)}$$

<u>Modal class:</u> the class interval with the highest frequency

LCB_{mo} – lower class boundary of the modal class

c – class size

f_{mo} – frequency of the modal class

f_b – frequency before the modal class

f_b – frequency after the modal class

MODE: Grouped FDT

Given the data below. Find the mode.

Classes	*f*	LCB	
10 – 19	5		
20 – 29	8		
30 – 39	11		
40 – 49	12		
50 – 59	6		
total	42		

MODE: Grouped FDT

Given the data below. Find the mode.

Classes	*f*	LCB	
10 – 19	5	9.5	
20 – 29	8	19.5	
30 – 39	11	29.5	
40 – 49	12	39.5	modal class
50 – 59	6	49.5	
total	42		

MODE: Grouped FDT

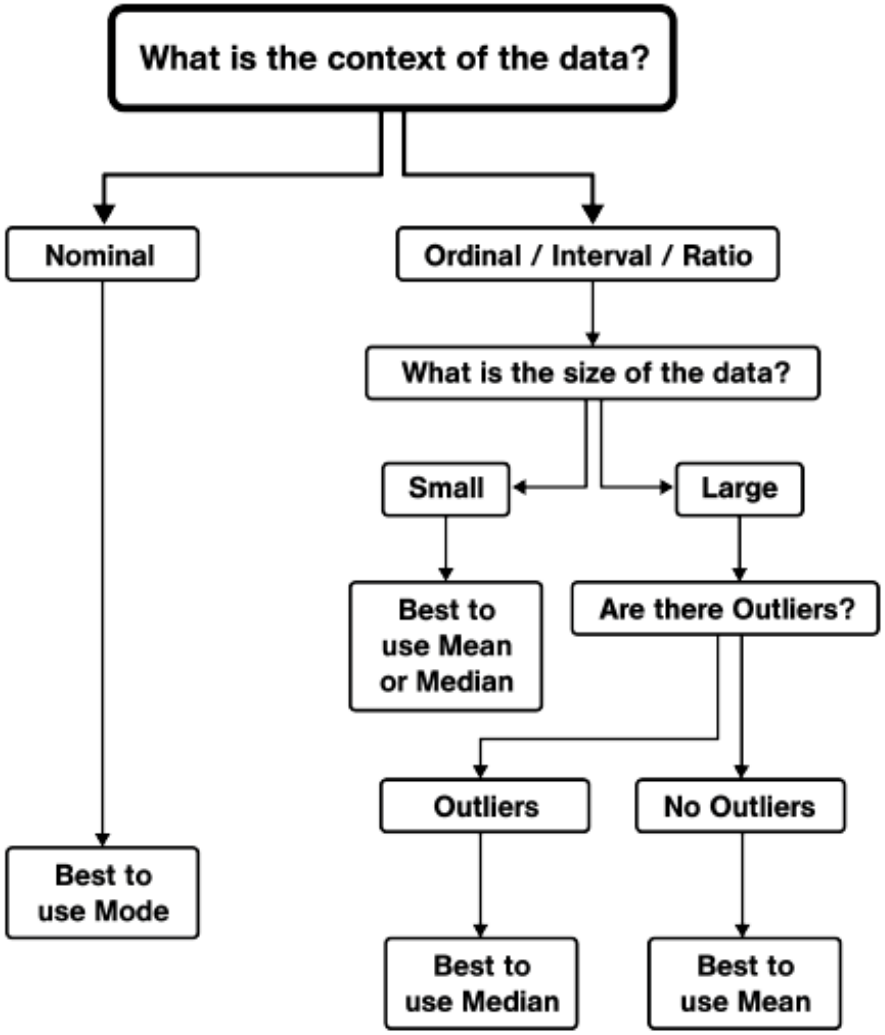
Classes	f	LCB
10 – 19	5	9.5
20 – 29	8	19.5
30 – 39	11	29.5
40 – 49	12	39.5
50 – 59	6	49.5
total	42	

$$x^o = LCB_{mo} + c\left(\frac{f_{mo} - f_b}{2f_{mo} - f_b - f_a}\right)$$

$$= 39.5 + 10\left(\frac{12 - 11}{2(12) - 11 - 6}\right)$$

$$x^o = 40.93$$

When to use mean, median, mode?



When to use mean, median, mode?

Type of Variable	Best measure of central tendency
Nominal	Mode
Ordinal	Median
Interval / ratio (not skewed)	Mean
Interval / ratio (skewed)	Median