Module 1
Statistics

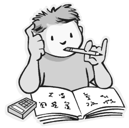
What this module is about

This module deals with the definition of statistics and terms used in the study of statistics. It will also discuss the history and importance of the study of statistics, summation rules, sampling techniques, collection of statistical data and organizing collected data in a table, constructing frequency distribution tables, and finding the measures of central tendency for ungrouped data. As you go over the discussion and exercises, you will appreciate more the importance of statistics in daily life. Enjoy learning this module and go over the discussion and examples if you have not yet mastered a concept.

What you are expected to learn

This module is designed for you to:

1. define statistics, sample, and population.
2. give the history and importance of the study of statistics
3. use the rules of summation to find sums
4. explain the different sampling techniques
5. collect statistical data and organize in a table
6. construct frequency distribution tables
7. find the measures of central tendency using ungrouped data
    - mean
    - median
    - mode



How much do you know

1. It is the branch of mathematics concerned with the techniques by which data are collected, organized, analyzed, and interpreted.

   a. Information Technology
   b. Statistics
   c. Trigonometry
   d. Geometry

2. It is a method of selecting the elements of a sample from the population under consideration.

   a. Sampling
   b. Drawing
   c. Organizing
   d. Collecting

3. It is used to present data in a most systematic and organized manner to make its reading and interpretation simple and easy.

   a. Graph
   b. Table
   c. Drawing
   d. Sampling

3. Mrs. Borromeo wants to study the heights and weights of the students in her class. Which of the following samples is most likely to be a good representation of the whole class?

   a. A sample consisting of all students whose surnames start with E.
   b. A sample consisting of all athletes in the class.
   c. A sample consisting of students whose birthdays are from January to June.
   d. A sample consisting of students whose names were drawn out of a box which contained all the names of the students in the class.

4. Which of the following means $\sum_{i=1}^{5} x_i$?

   a. $1 + 2 + 3 + 4 + 5$
   b. $x + 2x + 3x + 4x + 5x$
   c. $x_1 + x_2 + x_3 + x_4 + x_5$
   d. none of the above

6. The frequency distribution below shows the scores obtained by 300 students in a Mathematics test of 50 items.

Score	Number of Students
45-49	15
40-44	32
35-39	42
30-34	108
25-29	67
20-24	21
15-19	10
10-14	5
Total	300

What interval contains the highest frequency?
   a. 10-14
   b. 45-49
   c. 30-34
   d. 25-29
7. What class size was used in number 6?
   a. 5
   b. 4
   c. 3
   d. 2

5. It is the measure that occurs most in a distribution.
   a. mean
   b. median
   c. mode
   d. class size

6. The following are scores obtained by 10 students in an achievement test.

47 45 35 44 48 39 37 29 28 50

What is the mean?

   a. 28
   b. 35
   c. 40.2
   d. 45

7. What is the median in no, 9

   a. 39
   b. 41.5
   c. 43.5
   d. 44

What you will do

Lesson 1

Definition of Terms Related to Statistics

Statistics is a branch mathematics that deals with the collection, classification, description, and interpretation of data obtained by the conduct of surveys and experiments. Its fundamental purpose is to describe and draw inferences about the numerical properties of a population.

Two important terms that you should understand in studying statistics are population and sample.

In statistics, *population* does not only mean a group of people. *Population* may also mean a defined group or aggregates of objects, animals, materials, measurements, “things”, “events” or “happenings” of any kind. Thus, a sack of rice, a whole pizza pie, or a set of weights and heights are considered population.

Since it would be impractical to study the whole population as in the case of a sack of rice, then it is necessary to just take a sample of the population. Thus, a handful of rice is a sample of the population in a sack of rice. Thus, *sample* is defined as any subgroup of the population drawn by some appropriate method from the population. It should be a representative of the population, that is, the sample will show the properties of the population.

Try this out

Study the following situations. Identify the phrase which represents the sample and which phrase shows the population.

1. Mrs. Jara wants to know the nutritional status of the first year students in her school so she got 150 first year students to represent the year level.
2. When Sandra bought a sack of rice, she examined a handful from the sack to check if it is the variety she wants.
3. A doctor wants to know what causes the infection in a patient so he requested for the patient’s blood examination. The medical technologist extracted only 10 cubic centimeters of blood from the patient for examination.
4. The chef wants to check if the food being cooked tastes as he wants it to be so he tasted a spoonful of it.

5. The school guidance counselor would like to know the course preference of the graduating students in their school so she interviewed 100 of the fourth year students.

Lesson 2

History and Importance of Statistics

Historical records show that since the beginning of civilization simple forms of statistics had already been used. This is manifested by pictorial representations and other symbols used to record numbers of people, animals, and inanimate objects on skins, slabs, sticks of wood, or the walls of caves. Records show that even before 3000 B.C., the Babylonians used small clay tablets to record tabulations of agricultural harvests and of commodities bartered or sold. The Egyptians analyzed the population and material wealth of their country before they begun building the pyramids in the 31st century B.C. Even the biblical books of Numbers and 1 Chronicles show statistical works. Numbers contained two separate censuses of the Israelites while 1 Chronicles described the material wealth of various Jewish tribes. In China, similar numerical records existed before 2000 B.C. As early as 594 B.C., the ancient Greeks held censuses used as bases for taxation

Records also show that the Roman Empire was the first government to gather extensive data about the population, area, and wealth of the territories that it controlled. In Europe, few comprehensive censuses were made during the Middle Ages. in the early 16th century, registration of deaths and births begun in England. Then in 1662 the first noteworthy statistical study of population was made. In 1691, a similar study of mortality made in Breslau, Germany was used by the English astronomer Edmond Halley as a basis for the earliest mortality table. In the 19th century, investigators recognized the need to reduce information to numerical values to avoid the ambiguity of verbal description.

At present, statistics is a reliable means of describing accurately the values of economic, political, social, psychological, biological, and physical data. Statistics serves as a tool to correlate and analyze collected data. It is no longer confined to gathering and tabulating data. Now, it is also a process of interpreting the information that serves as a basis for preparing plans.

Lesson 3

The Summation Process

The study of statistics involves the collection of data or measurement. Thus, there is always a need to add several numbers. The Greek capital letter sigma, Σ is used in the process. The symbol Σ, read as *the sum of* tells you to add certain numerical values.

Example 1: Consider the scores obtained by 10 students in a 50-items mathematics test.

Student No.	Score
1	35
2	40
3	29
4	37
5	25
6	33
7	49
8	47
9	28
10	42

For convenience, variables will be used to present the data.

Let x = score obtained by each student
x_i = different values or observations of x
x_i is read as "x sub *i*" where *i* is a subscript which indicates the position of each value in the series.

In the given data, there are 10 observations denoted as x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}.

Hence, $\sum_{i-1}^{10} x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10}$.

The symbol $\sum_{i-1}^{10} x_i$ is read as "the sum of 10 observations x_1 to x_{10}".

To substitute the data:

$$\sum_{i-1}^{10} x_i = 35 + 40 + 29 + 37 + 25 + 33 + 49 + 47 + 28 + 42$$

$$\sum_{i-1}^{10} x_i = 365$$

For large observations, say 50, the summation will be expressed as:

$$\sum_{i-1}^{50} x_i = x_1 + x_2 + x_3 + \ldots..+x_{50}.$$

In general, $\sum_{i-1}^{n} x_i = x_1 + x_2 + x_3 + \ldots..+x_n.$

If all the given values of a variable are to be used in finding the sum, the limits of the summation are usually omitted, as

$$\sum_{i-1}^{10} x_i = \sum x$$

Example 2: Given are the ages of the first 4 shoppers at a newly opened convenience store in the neighborhood – 12, 24, 30, 45.

1. What will x represent in the information given?
2. What will the subscript *i* represent?
3. Write an expression for the sum.
4. What are the lower and upper limits of the expression?
5. Write the formula for the summation and find the sum of the given information.

Answers:

1. x will represent the ages of the first 4 shoppers in the newly opened convenience store.
2. *I* will represent the first 4 shoppers in the newly opened convenience store.
3. $\sum_{i=1}^{4} x_i$ is the expression for the summation.
4. The lower limit is 1 and the upper limit is 4.
5. $\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$

   $= 12 + 24 + 30 + 45$

   $= 111$

This time, consider 5 observations. If the sum of five observations is written as:

$$\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5;$$

the sum of the squares of the five observations is represented as:

$$\sum_{i=1}^{5} x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2;$$

the sum of the products of pairs of five observations is expressed as:

$$\sum_{i=1}^{5} a_i x_i = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5$$

Example 3: Consider the first four multiples of 2: 2, 4, 6, 8. Use the corresponding summation formula to find the following:

1. the sum of the first four multiples of 2
2. the sum of the squares of the first four multiples of 2
3. the sum of the products of pairs of values consisting of the first four counting numbers and the first four multiples of 2.

Solutions:

1. $\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$

   $= 2 + 4 + 6 + 8$
   $= 20$

2. $\sum_{i=1}^{4} x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$

   $= 2^2 + 4^2 + 6^2 + 8^2$
   $= 4 + 16 + 36 + 64$
   $=120$

3. $\sum_{i=1}^{4} a_i x_i = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$

   $= 1(2) + 2(4) + 3(6) + 4(8)$
   $= 2 + 8 + 18 + 32$
   $= 60$

Example 4: Find 1. $\sum_{i=1}^{6} 3$ 2. $\sum_{i=1}^{6} (-3)$

Solutions:

1. $\sum_{i=1}^{6} 3 = 3 + 3 + 3 + 3 + 3 + 3 = 6(3) = 18$

2. $\sum_{i=1}^{6}(-3)$ = (-3) + (-3) + (-3) + (-3) + (-3) + (-3) = 6(-3) = -18

Observe that in example 4, the summation of a constant *c* is the product of the constant and the number of terms *n* in the summation, that is, $\sum_{i=1}^{n} c = nc$

Try this out

Express each of the following as a sum:

1. $\sum_{i=1}^{7} x_i$
2. $\sum_{j=1}^{5} z_j$
3. $\sum_{k=1}^{3} y_k$
4. $\sum_{j=1}^{5} p$
5. $\sum_{k=1}^{3} a_k y_k$

Express the following sums in summation notation:

6. $x_1 + x_2 + x_3 + \ldots + x_{20}$
7. $x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + x_8^2$
8. $a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + a_4x_4^2$
9. $4x_1^2 + 4x_2^2 + 4x_3^2 + 4x_4^2$
10. $(y_1 + z_1) + (y_2 + z_2) + (y_3 + z_3) + (y_4 + z_4) + (y_5 + z_5)$

Use summation to find the following:

11. the sum of the positive odd integers less than 20
12. the sum of the first ten positive even integers
13. the sum of the squares of the first five positive even integers
14. the sum of the products of the first four counting numbers and the first four multiples of 3.
15. the sum of the products of 5 times the positive odd integers less than 15.

Compute:

16. $\sum_{i=1}^{4} 5$

17. $\sum_{i=1}^{10} (-2)$

18. $\sum_{i=1}^{8} 4$

Lesson 4

Sampling Techniques

The method of drawing samples from a population is of very important. There are several ways of doing this.

One way is by *simple random sampling*. This is a procedure where a sample is selected in such a way that every element is as likely to be selected as any other element in the population.

Another way is by *systematic random sampling.* This method is a sampling procedure with a random start.

Another method is the *stratified random sampling*. This is used when the population can be naturally classified into groups or strata.

Example: The clinic teacher wants to determine the average height and weight of the first year students. How can she randomly select 50 students consisting of 250 male students and 300 female students to represent the population using (a) simple random technique? (b) systematic random technique? (c) stratified random technique?

Answers:

The clinic teacher can randomly select the sample using simple random sampling by following these simple steps:

1. Assign the students with numbers.
2. Write the student number with his/her corresponding height in uniform size slips of paper.

3. Roll the pieces of paper uniformly and place them in a box.
4. Draw a slip of paper at a time, shaking the box after each draw until 50 samples are taken.

The clinic teacher can select the sample using the systematic random sampling using the steps as follows:

1. Using the same data and with the students assigned with numbers, and arranged chronologically, the clinic teacher with eyes closed, points to a number. If the number pointed to is, let us say, 7, student number 7 becomes a part of the sample (sample number 1). This is a "random start".

2. From student 7, count 1 to 7 repeatedly until all 50 samples are taken. Numbers which were previously selected will eliminated in the counting.

The clinic teacher can select the sample using the stratified random sampling by using the following procedures:

1. The data should be classified into two groups, male and female.
2. Get a proportional number of samples from each group or strata. The number of samples from the males will be $\frac{250}{550}$ or $\frac{5}{11}$ of 50 which is 23 and from the females 27.
3. Place the slips of paper, properly filled up in separated boxes for each group.
4. Draw, one at a time, the required number for each group.

Try this out

1. Mrs. Lucas is studying the heights and weights of the students in her class. Which of the following samples is most likely to be a good representation of the whole class? Justify your answer.

   a. A sample consisting of all athletes in the class.
   b. A sample consisting of all students whose surnames start with D.
   c. A sample consisting of students whose names were drawn out of a box which contained all the names of the students in the class.
   d. A sample consisting of students whose birthdays are from January to June.

2. For his report in Social Studies, Dennis wishes to wishes to interview a sample of Metro Manila residents to determine their opinion regarding the economic status of the country today.Tell whether he could find a sample that reflects the entire population being studied at

1) a depressed area in Payatas.
2) a shopping mall in Makati.
3) the Starbucks coffee shop.

3. A researcher wants to know the average age of teachers in a certain community. Fifty teachers from the elementary and 25 teachers from the secondary levels were interviewed for the purpose. How will the researcher choose a sample size of 20 using:

    a. simple random sampling
    b. systematic random sampling
    c. stratified random sampling

Lesson 5

Collecting and Organizing Data in a Table

The study of statistics begins with the collection of data or measurements. Data collected should be organized systematically for easier and faster interpretation. They maybe presented in any of the following forms:

The *textual form* can be used if the data to be presented if few.

The *tabular* and *graphical* forms are used when more detailed information about the data is to be presented.

A table is used when you want to present a data in a systematic and organized manner so that reading and interpretation will be simpler and easier. When a table is used, you must remember the following:

1. The title of the table.
2. Indicate the date of the survey.
3. Arrange the data systematically in columns. The columns must be properly labeled.
4. Identify the source of the data.

Example 1:

Mahusay National High School Enrolment, SY 2005-2006		
Year Level	Male	Female
First	216	267
Second	197	216
Third	187	227
Fourth	176	215
Total	776	925

You will observe that the table above shows clearly the enrolment data in Mahusay National High School for the school year 2005-2006.

Another type of tabular presentation is the frequency table also known as a frequency distribution. It is an arrangement of the data that shows the frequency of occurrence of different values of the variables.

A frequency table is constructed by listing the measurements from highest to lowest, then making tally marks to record how often each number occurs. After tallying, count the marks and record them in the proper column.

Example 2: The scores of 45 students on a 20-point Science quiz are as follows:

17	20	15	18	19	16	11	10	15	16
12	12	13	14	11	10	14	13	12	11
13	15	14	10	15	16	17	17	18	20
20	18	19	19	18	17	16	15	12	12
13	14	15	19	20					

Prepare a frequency table for the set of data.

Solution: To prepare a frequency table for the given set of scores, the scores are listed from highest to lowest, tally marks are made and counted. The counted tally marks will then be recorded under the column frequency. Notice that every 5^{th} tally crosses the first four tallies. This is done to make counting of marks easier especially if the number of cases is rather big.

Score	Tallies	Frequency
20	////	4
19	////	4
18	////	4
17	////	4
16	////	4
15	~~////~~ /	6
14	////	4
13	////	4
12	~~////~~	5
11	///	3
10	///	3
Total		45

Try this out

1. The school budget for Maintenance and Other Operating Expenses of a certain school for Calendar Year 2004 is given below.

Expense Item	Amount (in Pesos)
Power	600 000
Water	95 000
Communication	60 000
Supplies	1 600 000
Repair	920 000
Others	100 000

   a) How much is the total budget of the school for CY 2004?
   b) Which expense item received the biggest allocation? What percent of the total budget was allocated for it?
   c) Which expense item received the least allocation? What percent of the total budget was allocated for it?

2. The following shows the scores of 15 students in mathematics for the second grading period. Prepare a frequency table given the data below.

87	90	89	92	94
88	90	91	88	87
90	94	92	91	90

3. The following are heights of male fourth year students in a school. Prepare a frequency table for this set of data.

1.36	1.51	1.61	1.61	1.62	1.62	1.62	1.59	1.58	1.61
1.38	1.49	1.65	1.63	1.58	1.57	1.61	1.62	1.63	1.65
1.44	1.59	1.57	1.57	1.58	1.60	1.61	1.63	1.64	1.64
1.55	1.58	1.59	1.65	1.66	1.72	1.56	1.68	1.69	1.63

Lesson 6

Frequency Distribution Tables

If the number of measures in consideration is rather big, the presentation of data is further simplified by grouping the measures into class intervals called a *frequency distribution.*

A *frequency distribution* is a distribution of the total number of measures or frequencies over arbitrarily defined *categories* or *classes.* The number of measures falling under a class is called *class frequency.*

Example 1.

The frequency distribution below shows the scores obtained by 300 students in an English test of 50 items.

Score	Number of Students
45-49	15
40-44	32
35-39	42
30-34	108
25-29	67
20-24	21
15-19	10
10-14	5
Total	300

In the example above, the symbol 45-49 and the other symbols which follow up to 10-14 are called *class intervals.* The end numbers are called class limits. For instance in the class interval 45-49, 45 is called the *lower limit* while 49 is called the *upper limit.*

Each class interval has also a lower boundary and a higher boundary. For the class interval 45-49, the lower boundary is 44.5 while the higher boundary is 49.5. Hence, for the class interval 45-49, 44.5 – 49.5 are called the class boundaries.

The size of the class interval, also called *class size* is the difference between the upper boundary and the lower boundary. Hence, the class size in the given example is 5

A class interval has also a *midpoint* or a *class mark*. It is obtained by taking half the sum of the lower and upper class limit. For instance, the midpoint of the class interval 45-49 is $\frac{45+49}{2}$ or 47.

The following are the rules in determining the size of the class interval:

1. The class interval must cover the total range of the observation where the range, $R = H - L$ (H = highest and L = lowest). It is usually between 10 to 20 intervals.

2. Select class intervals with a range of 1, 3, 5, 10, or 20 points since these will meet the requirements of most set of data.

3. Start the class interval at a value which is a multiple of the size of the interval. For example, with a class interval of 3, the intervals should start with the values 3, 6, 9, etc.

4. Arrange the class intervals according to the order of magnitude of the observations they include. The class interval containing the largest observations is usually placed at the top.

Example 2: The following are scores obtained by a group of 50 students on their English IV examinations. Prepare a frequency distribution for these data using a class interval of 5.

39	93	80	49	41	85	75	59	62	68
34	49	50	46	69	72	73	76	77	54
95	63	66	64	88	90	51	53	56	79
70	70	78	85	86	59	66	72	77	76
71	79	70	65	40	57	82	75	89	82

Solution:

Since the class interval is already given, and the lowest score is 34 then the class interval containing the lowest score should be 30-34 since the rule states that the class interval should start with a number which is divisible by the class size. After arranging the class intervals, tally the scores to determine the frequency. Look at the obtained frequency distribution below.

Scores	Tallies	Frequency
95-99	/	1
90-94	//	2
85-89	~~////~~	5
80-84	///	3
74-79	/~~////~~////	9
70-74	~~////~~ //	7
65-69	~~////~~	5
60-64	///	3
55-59	////	4
50-54	////	4
45-49	///	3
40-44	//	2
35-39	/	1
30-34	/	1
Total		50

Try this out

1. A sample of fifty shoppers at a newly opened convenience store has been randomly selected. The following data show the shoppers' ages. Determine the appropriate class interval to use then prepare a frequency distribution for the data.

12	20	17	19	23	32	15	45	60	65
18	22	27	35	37	57	47	38	40	28
13	10	19	24	29	28	38	47	48	57
27	29	33	34	49	76	55	65	37	39
40	14	17	20	32	33	60	65	62	57

2. The following are the weights of randomly selected 1st year students in kilograms. Prepare a frequency distribution for this set of data.

37	35	40	42	36	57	38	44	60	45
52	64	38	39	40	42	50	56	45	43
38	39	50	41	42	56	57	54	55	60
35	38	40	40	42	53	47	48	39	50
35	37	39	39	50					

Lesson 7

Measures of Central Tendency for Ungrouped Data

Aside from tables and graphs, another way of describing a set of data is by stating a single numerical value associated with it. This value is where all the other values in a distribution tend to cluster. It is called the *average* or measure of *central tendency*. There are *three kinds of average*: the *mean*, the *median*, and the *mode*.

The Mean

The *mean* (also known as the *arithmetic mean*) is the most commonly used measure of central position. It is the sum of measures divided by the number of measures in a variable. It is symbolized as $\bar{x}$ (read as x bar).

The mean is used to describe a set of data where the measures cluster or concentrate at a point. As the measures cluster around each other, a single value appears to represent distinctively the total measures. It is, however, affected by extreme measures, that is, very high or very low measures can easily change the value of the mean.

To find the mean of ungrouped data, use the formula

$$\bar{x} = \frac{\Sigma x}{N}$$

where $\sum x$ = the summation of x (sum of the measures)

N = number of values of x

Example: The grades in Chemistry of 10 students are 87, 84, 85, 85, 86, 90, 79, 82, 78, 76. What is the average grade of the 10 students?

Solution:

$$\text{Mean} = \frac{87+84+85+85+86+90+79+82+78+76}{10} = \frac{832}{10} = 83.2$$

The Median

The *median* is the middle entry or term in a set of data arranged in either increasing or decreasing order.

The median is a positional measure. Thus the values of the individual measures in a set of data do not affect it. It is affected by the number of measures and not by the size of the extreme values.

To find the median of a given set of data, take note of the following:

1. Arrange the data in either increasing or decreasing order.
2. Locate the middle value. If the number of cases is odd, the middle value is the median. If the number of cases is even, take the arithmetic mean of the two middle measures.

Example 1: The number of books borrowed in the library from Monday to Friday last week were 58, 60, 54, 35, and 97 respectively. Find the median.

Solution: Arrange the number of books borrowed in increasing order.

35, 54, 58, 60, 97

The median is 58.

Example 2: Cora's quizzes for the second quarter are 8, 7, 6, 10, 9, 5, 9, 6, 10, and 7. Find the median.

Solution: Arrange the scores in increasing order.

5, 6, 6, 7, 7, 8, 9, 9, 10, 10

Since the number of measures is even, then the median is the average of the two middle scores.

$$\text{Md} = \frac{7+8}{2} = 7.5$$

The Mode

The mode is another measure of position. The mode is the measure or value which occurs most frequently in a set of data. It is the value with the greatest frequency. To find the mode for a set of data –

1. select the measure that appears most often in the set;

2. if two or more measures appear the same number of times, and the frequency they appear is greater than any other measures, then each of these values is a mode;

3. if every measure appears the same number of times, then the set of data has no mode.

Example 1: The shoe sizes of 10 randomly selected students in a class are 6, 5, 4, 6, $4\frac{1}{2}$, 5, 6, 7, 7 and 6. What is the mode?

Answer: The mode is 6 since it is the shoe size that occurred the most number of times.

Example 2: The sizes of 9 classes in a certain school are 50, 52, 55, 50, 51, 54, 55, 53 and 54.

Answer: The modes are 54 and 55 since the two measures occurred the same number of times. The distribution is *bimodal.*

Try this out

Find the mean, median, and mode (modes) of each of the following sets of data.

.

1. 29, 34, 37, 22, 15, 38, 40

2. 5, 6, 7, 7, 9, 9, 9, 10, 14, 16, 20

3. 82, 61, 93, 56, 34, 57, 92, 53, 57

4. 26, 32, 12, 18, 11, 12, 15, 18, 21

5. The scores of 20 students in a Physics quiz are as follows:

25	33	35	45	34
26	29	35	38	40
45	38	28	29	25
39	32	37	47	45

Let's Summarize

Statistics is a branch mathematics that deals with the collection, classification, description, and interpretation of data obtained by the conduct of surveys and experiments.

Population is a defined group or aggregates of objects, animals, materials, measurements, “things”, “events” or “happenings” of any kind.

Sample is defined as any subgroup of the population drawn by some appropriate method from the population.

Sampling is the process of selecting the elements of a sample from the population being studied. The methods of sampling include *simple random sampling, systematic random sampling*, and *stratified random sampling.*

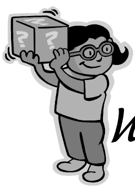
A table is used to present a data in a systematic and organized manner to make its reading and interpretation simple and easy.

A *frequency distribution* is a distribution of the total number of measures or frequencies over arbitrarily defined *categories* or *classes.* The number of measures falling under a class is called *class frequency.*

The *mean* (also known as the *arithmetic mean*) is the most commonly used measure of central position. It is the sum of measures divided by the number of measures in a variable. It is symbolized as $\bar{x}$ (read as x bar). It is used to describe a set of data where the measures cluster or concentrate at a point.

The *median* is the middle entry or term in a set of data arranged in either increasing or decreasing order. It is a positional measure. The values of the individual measures in a set of data do not affect the median. It is affected by the number of measures and not by the size of the extreme values.

The mode is the value which occurs most frequently in a set of data.



What have you learned

1. What meaning of this symbol $\sum$?

   a. sum
   b. difference
   c. product
   d. quotient

2. It a procedure where a sample is selected in such a way that every element is likely to be selected as any other element of a population.

   a. Sampling
   b. Simple random sampling
   c. Systematic random sampling
   d. Stratified random sampling

3. In the expression $\sum_{i=1}^{10} x_i$, what is the upper limit?

   a. i
   b. x
   c. 1
   d. 10

4. The enrolment of a certain school for school year 2005-2006 is shown below.

Enrolment, SY 2005-2006		
Year Level	Male	Female
First	456	497
Second	427	456
Third	487	467
Fourth	356	425

How many students enrolled for the school year?

a. 953
b. 1726
c. 1845
d. 3571

5. What is $x_1 + x_2 + x_{3} + \ldots + x_{20}$ using summation notation?

a. $\sum_{i=1}^{20} x_i$
b. $\sum_{i=1}^{20} x_{20}$
c. $\sum_{i=1}^{20} x$
d. none of the above

6. The frequency distribution below shows the scores obtained by 60 students in a Science.

Score	Number of Students
80-89	4
70-79	6
60-69	12
50-59	10
40-49	8
30-39	10
20-29	7
10-19	3
Total	60

What interval contains the lowest frequency?

a. 10-19
b. 20-29
c. 30-39
d. 45-49

7. What class size was used in number 6?

a. 5
b. 8
c. 9
d. 10

8. It is the middle most measure that occurs in a distribution.

   a. mean
   b. mode
   c. median
   d. class size

9. The following are scores obtained by a student in 10 quizzes.

7	5	5	4	8	9	7	9	8	5

   What is the mode?

a. 8 c. 9
b. 5 d. 7

10. What is the median in no. 9?

a. 5 c. 7
b. 9 d. d. 8

Answer key

How much do you know

1. b
2. a
3. d
4. c
5. c
6. c
7. a
8. c
9. c
10. b

Try this out

Lesson 1

	Sample	Population
1	150 first year student	First year students
2	Handful of rice	One sack of rice
3	10 cubic centimeters of blood	Blood
4	Food cooked	Spoonful of it
5	100 fourth year students	Graduating students

Lesson 3

1. $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7$
2. $z_1 + z_2 + z_3 + z_4 + z_5$
3. $y_1 + y_2 + y_3$
4. $p_1 + p_2 + p_3 + p_4 + p_5$
5. $a_1y_1 + a_2y_2 + a_3y_3$
6. $\sum_{i=1}^{20} x_i$
7. $\sum_{i=1}^{8} x_i^2$
8. $\sum_{i=1}^{4} a_i x_i^{\;2}$
9. $4\sum_{i=1}^{4} x_i^{\;2}$
10. $\sum_{i=1}^{5} (x_i + z_i)$
11. 100
12. 120
13. 55
14. 165
15. 245

16. 20
17. -20
18. 32

Lesson 4

1. c
2. a
3. Using simple random sampling, the following steps should be done:

    a. Write the names of the teachers in uniform sized slips of paper.
    b. Roll the slips of paper.
    c. Put the rolled slips of paper in a box.
    d. Draw at random 20 slips of paper

    By systematic random sampling, the following steps should be done:

    a. Assign to each teacher a number.
    b. Arrange the numbers in order
    c. With eyes closed, point to one of the numbers at random. The number pointed to will be the random start.
    d. Using the number drawn as a random start, count repeatedly until 20 samples are drawn.

    By stratified random sampling, the following should be done:

    a. Classify the teachers into two groups, one for elementary and another for secondary.
    b. Get a proportional number of samples from each group.

       Elementary : $\left(\frac{50}{75}\right)20 = 13$

       Secondary : $\left(\frac{25}{75}\right)20 = 7$

    c. Place the slips of paper properly filled up in separate boxes for each group.

    d. Draw one slip of paper one at a time until 20 samples are drawn.

Lesson 5

1. a. Php 3, 375, 000
   b. supplies – 47.4%
   c. communication – 1.78%

2.

Score	Frequency
94	2
92	2
91	2
90	4
89	1
88	2
87	2
Total	15

3.

Height	Frequency
1.72	1
1.69	1
1.68	1
1.66	1
1.65	3
1.64	2
1.63	4
1.62	4
1.61	5
1.60	1
1.59	3
1.58	4
1.57	3
1.56	1
1.55	1
1.51	1
1.49	1
1.44	1
1.38	1
1.36	1
Total	40

Lesson 6

1.

Ages	Frequency
75-79	1
70-74	0
65-69	3
60-64	3
55-59	4
50-54	0
45-49	5
40-44	2
35-39	6
30-34	5
25-29	6
20-24	5
15-19	6
10-14	4
Total	50

2.

Weight	Frequency
63-65	1
60-62	2
57-59	2
54-56	4
51-53	2
48-50	5
45-47	3
52-44	6
39-41	10
36-38	7
33-35	3
Total	45

Lesson 7

	Mean	Median	Mode
1	30.71	34	None
2	10.18	9	9
3	65	57	57
4	18.33	18	12,18 Bimodal
5	35.25	35	45

What have you learned

1. a
2. a
3. d
4. d
5. a
6. d
7. d
8. c
9. b
10. c