Mathematics for Biology

MAT1142

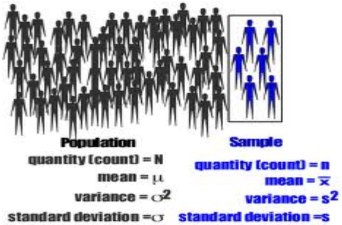
Department of Mathematics
University of Ruhuna

A.W.L. Pubudu Thilan

Statistics

Difference Between Population and Sample

- The term **population** is used in statistics to represent all possible measurements that are of interest to us in a particular study.
- A **sample** is a part of the population of interest.



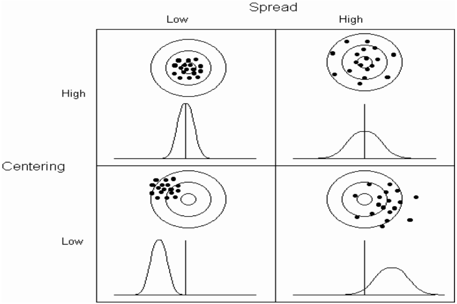
Main Steps in Statistical Survey

- Identification of data
- Collecting data
- Summarizing data
- Organizing data
- Analyzing data
- Present information based on data
- Recommendation/Conclusions

Measures of Central Tendency

Measures of central tendency are measures of the location of the middle or the center of a distribution. Most commonly used measure of central tendency are:

- Mean
- Median
- Mode



- The **mean** is the mathematical average of a set of numbers.
- The average is calculated by adding up two or more scores and dividing the total by the number of scores.

$$\text{Population mean} = \mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\text{Sample mean} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Find the mean of the following data:
5, 15, 10, 15, 5, 10, 10, 20, 25, 15.

Solution

$$\begin{aligned}\text{Mean} &= \frac{\text{Sum of data}}{\text{Number of data}}\\ &= \frac{5+15+10+...+15}{10}\\ &= \frac{130}{10}\\ &= 13\end{aligned}$$

The middle value, or the mean of the middle two values, when the data is arranged in numerical order is called as **median**.

Find the median for the following list of values:
13, 18, 13, 14, 13, 16, 14, 21, 13

Solution

First organize the data in increasing order:

13, 13, 13, 13, 14, 14, 16, 18, 21

There are nine numbers in the list.

So the middle one will be the $(9 + 1) \div 2 = 10 \div 2 = 5$th number.

13, 13, 13, 13, **14**, 14, 16, 18, 21

So the median is 14.

Find the Median of the following data:
5, 15, 10, 15, 5, 10, 10, 20, 25, 15.

Solution

First organize the data in increasing order

5, 5, 10, 10, 10, 15, 15, 15, 20, 25

There are ten numbers in the list.

The middle one will be the $(10 + 1) \div 2 = 11 \div 2 = 5.5$th number.

The numbers 10 and 15 both fall in the middle.

Average these two numbers to get the median$=\frac{10+15}{2} = 12.5$

The number that appears the most time within a data set is called as **mode**.

Find the mode of the following data sets.

(i) 13, 18, 13, 14, 13, 16, 14, 21, 13

(ii) 1, 2, 4, 7

(iii) 5, 15, 10, 15, 5, 10, 10, 20, 25, 15

Solution

(i) Mode is 13

(ii) No mode

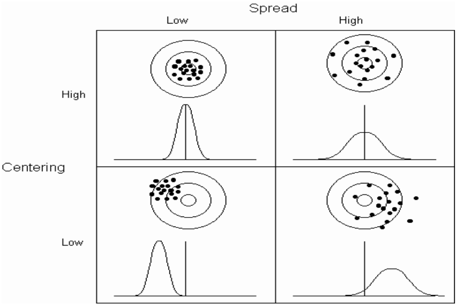
(iii) We have two modes. One is 10 and the other one is 15

- It is possible to have more than one mode.
- Also it is possible to have no mode.
- If there is no mode, write down it as **no mode**, do not write mode is zero.

Measures of Variability

A dispersion is a real number that is zero if all the data are identical, and increases as the data becomes more diverse. The most common measures of variability are:

- Range
- Variance
- Standard deviation



The **range** is the difference between the largest and smallest values in a data set.

Find the range of following data set:
$1, 3, 4, 5, 5, 6, 7, 11.$

Solution

Largest =11

Smallest values=1

Range=11-1=10

In a **population**, variance is the average squared deviation from the population mean, as defined by the following formula:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N},$$

where σ^2 is the **population variance**, μ is the population mean, x_i is the ith element from the population, and N is the number of elements in the population.

The variance of a **sample**, is defined by slightly different formula, and uses a slightly different notation:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{(n-1)},$$

where s^2 is the **sample variance**, $\overline{x}$ is the sample mean, x_i is the ith element from the sample, and n is the number of elements in the sample.

- Population variance is an unknown quantity in most of the situations.
- So, sample variance can be used as an estimate for population variance.

The standard deviation is the square root of the variance. Thus, the standard deviation of a population is:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}},$$

where σ is the **population standard deviation**.

The standard deviation is the square root of the variance. Thus, the standard deviation of a sample is:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{(n-1)}},$$

where s is the **sample standard deviation**.

Example 1

A **population** consists of four observations: $1, 3, 5, 7$. What is the variance?

Example 1
Solution

First, we need to compute the population mean.
$\mu = (1 + 3 + 5 + 7)/4 = 4$

Then we plug all of the known values into formula for the variance of a population, as shown below:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

$\sigma^2 = [(1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2]/4$
$\sigma^2 = [(-3)^2 + (-1)^2 + (1)^2 + (3)^2]/4$
$\sigma^2 = [9 + 1 + 1 + 9]/4 = 20/4 = 5$

Example 2

A **sample** consists of four observations: $1, 3, 5, 7$. What is the standard deviation?

Example 2
Solution

First, we need to compute the sample mean.
$\overline{x} = (1 + 3 + 5 + 7)/4 = 4$

Then we plug all of the known values into formula for the standard deviation of a sample, as shown below:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{(n-1)}}$$

$$s = \sqrt{[(1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2]/(4-1)}$$
$$s = \sqrt{[(-3)^2 + (-1)^2 + (1)^2 + (3)^2]/3}$$
$$s = \sqrt{[9+1+1+9]/3} = \sqrt{(20/3)} = \sqrt{(6.67)} = 2.58$$

Example 3

The mean marks of 8 students is 65. If marks of 7 students are 60, 70, 55, 50, 60, 65 and 85. Then find the marks of 8th student.

Example 3
Solution

$$
\begin{aligned}
\frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8}{8} &= 65 \\
\frac{60 + 70 + 55 + 50 + 60 + 65 + 85 + x_8}{8} &= 65 \\
\frac{445 + x_8}{8} &= 65 \\
445 + x_8 &= 658 \\
x_8 &= 520 - 445 \\
x_8 &= 75
\end{aligned}
$$

So the marks of the 8th student is 75.

Example 4

The mean of the marks obtained by 30 students in the subject of Mathematics is 45. The mean of the top 5 of them was found to be 70 and the mean of the last 10 of them was known to be 20. Find the mean of the remaining 15 students.

Example 4
Solution

$$
\begin{aligned}
\frac{1}{30}\sum_{i=1}^{30} x_i &= 45 \\
\sum_{i=1}^{30} x_i &= 30 * 45 = 1350 \\
\sum_{i=1}^{5} x_i + \sum_{i=6}^{20} x_i + \sum_{i=21}^{30} x_i &= 1350 \qquad (1)
\end{aligned}
$$

Example 4
Solution⇒Cont...

$$\begin{aligned}
\frac{1}{5}\sum_{i=1}^{5} x_i &= 70 \\
\sum_{i=1}^{5} x_i &= 70 * 5 = 350 \qquad (2) \\
\frac{1}{10}\sum_{i=21}^{30} &= 20 \\
\sum_{i=21}^{30} &= 20 * 10 = 200 \qquad (3)
\end{aligned}$$

Example 4
Solution⇒Cont...

Substitute in (1) from (2) and (3),

$$
\begin{aligned}
350 + \sum_{i=6}^{20} x_i + 200 &= 1350 \\
\sum_{i=6}^{20} x_i &= 1350 - 550 \\
\sum_{i=6}^{20} x_i &= 800 \\
\frac{1}{15}\sum_{i=6}^{20} x_i &= \frac{800}{15}
\end{aligned}
$$

So the mean mark of the remaining 15 students is 53.33.

Example 5

Consider the frequency distribution shown in scores of 20 students in a science test. Find the mean marks of the students?

Marks(x)	Frequency(f)
40	1
50	2
60	4
70	3
80	5
90	2
100	3
Total	20

Example 5
Solution

Marks(x)	Frequency(f)	fx
40	1	40
50	2	100
60	4	240
70	3	210
80	5	400
90	2	180
100	3	300
Total	20	1470

$$\text{Mean marks} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i = \frac{1470}{20} = 73.5$$

Example 6

The number of telephone calls received by a company switchboard over 5 minute intervals is given in the below table.

Number of Calls (x)	0	1	2	3	4	5
Frequency (f)	7	10	15	29	13	6

Find the mean number of telephone calls received by the company switchboard over 5 minute intervals.

Example 6
Solution

Number of Calls (x)	Frequency (f)	fx
0	7	7
1	10	10
2	15	30
3	29	87
4	13	52
5	6	30
Total	80	209

$$\text{Mean number of calls} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i = \frac{209}{80} = 2.6125$$

Exercise 1

The number of faulty components from a production line, over 30 days is given in the below table. Calculate the mean number of faulty components over 30 days.

No. faulty	0	1	2	3	4	5
Frequency	9	12	5	3	0	1

Answer is 1.2

Exercise 2

The number of people in each car passing a checkpoint is given in the below table. Calculate the mean number of people in a car.

No. people	1	2	3	4	5
Frequency	37	13	6	3	1

Answer is 1.63

Thank You