

RATU NAVULA COLLEGE**Y11 LIFE MATHEMATICS HOMEKIT 10 QP****STRAND 5 STATISTICS****SUBSTRAND 5.1.1 DATA COLLECTION****LESSON 65**

LO: Study various types of sampling methods in data collection.

- Statistics is the collection, display and analysis of data (information).
- Raw data are gathered using surveys, experiments and so on.
- The aim is to collect unbiased data, which are truly representative of the whole set of data from which they are drawn.
- Statistics is mainly used to provide evidence of decision making.

Collecting Data

- If an investigation is to be taken, every member of the population should be ‘measured’. Practically, this is often impossible.
- Hence a sample (portion of population) is investigated. Statistical analysis of the sample enables conclusions to be reached about the population as a whole.
- The sample chosen should be random and should reflect the characteristics of the population.
- If this does not occur, then sample is biased.

1. **Stratified Samples**

- When a population is made up of a number of distinct layers, a number of names are chosen at random from each layer. The number chosen from each group is in proportion to the numbers in each layer of the population.

Eg:

A 1000 strong school student body consists of 200 Y9 students, 250 Y10 students, 250 Y11 students, 200 Y12 students and 100 Y13 students. A random sample of 40 is to be chosen for student body.

A stratified random sample would require that:

$$200/1000 \times 40 = 8 \text{ Y9 students be chosen}$$

Advantages

- Highest precision
- Captures specific groups
- Disproportionate sampling possible.
- Ensures high degree of representatives of all the layers in population.

Disadvantages

- Requires advance knowledge of population
- Time consuming and tedious (more complex to analyze data and compute sampling errors)

2. Cluster Sampling

- Researcher creates multiple clusters of people from population where they are indicative of homogeneous characteristics and have equal chance of being a part of the sample.

Example

If an organization is looking to survey the performance of smartphones across a country, they can divide entire country's population into cities (clusters) and select further towns with the highest population and also filter those using mobile devices.

Advantages

- Very efficient. Consumes less time and cost. (requires fewer resources)
- Convenient access
- Data accuracy (high reliability)
- Ease of implementation (more feasible)

Disadvantages

- Prone to biases. (inferences about entire population would be biased as well)
- High sampling error
- Difficult to implement
- Difficult to analyse
- Lowest precision. (Imprecise results with improper clusters.)

3. Systematic Sampling

- Researcher chooses elements from a target population by selecting a random starting point and selects sample members after a fixed 'sampling interval'.

Eg: In school, while selecting the captain of a sports team, most of our coaches asked us to call out numbers such as (1-n) and the students with a random number is decided by coach.

Advantages

- High precision
- Easy to analyze data and compute sampling errors
- No need to use a table of random numbers.

Disadvantages

- Less random.
- May not be efficient
- May not capture specific groups
- Ordering of elements in sampling frame can create biases

4. Random Sampling

- Where every member of the population has the same, chance of being chosen for the sample (equal probability).

*Note: when whole population is surveyed, the survey is known as **census**. In Fiji, census is conducted every 10 years.

Advantages

- Ensures high degree of representatives.
- Requires little advance knowledge of population

Disadvantages

- Time consuming and tedious.
- May not be efficient
- May not capture specific groups

Activity

Determine the type of sampling used in the following scenario:

1. Divide the users of the Internet into different age groups and then select a random sample from each age group to survey about the amount of time they spend on the Internet each month.
2. Ms. Prasad samples her class by selecting every third person on her class list.
3. Ms. Mala samples her class by selecting all students sitting at group 1 and group 5 in her classroom.
4. Ms. Naicker samples her class by picking 10 numbers from her hat and each number is assigned to a student.

LESSON 66

LO: Select a sample from a sampling frame.

- The larger the sample is (when compared to the population) the more reliable the results.
Eg: A sample of 20 is ample for a population of 300 but not for a population of 3 million.
- Factors such as cost and practicality often determine the size of a sample.
- A very small sample, although randomly selected, may not necessarily be a representative of the population.

Activity

1. Kevin conducted a survey investigating the distance travelled by students to Ratu Navula College. He stood at the back gate and questioned the 25 students passing through the gate before the start of school.
 - a. Is this method of choosing the sample a random selection?
 - b. Could the sample be biased? Explain.

2. An owner of a supermarket has been granted a license to sell liquor. She decides to conduct a survey to discover what brands of beer she should stock. She decides to interview every fiftieth customer who enters the supermarket starting from 9am on a Monday. She continues until she has surveyed 100 customers.
 - a. Is the owner's sample a random sample?
 - b. Would the survey be biased? Explain.
 - c. Suggest ways in which the owner may obtain the information she wants.

LESSON 67

LO: Organize and display data using tables.

1. Frequency Tables
 - The frequency table lists scores and the frequency of each of those scores.

Example:

The maximum temperatures (in degrees Celsius) and the relative humidity (in percent) for London for the month of July of a particular year, reported by the Meteorological Department are given below.

Maximum temperatures (in degrees Celsius)

32.5, 30.3, 33.8, 31.0, 28.6, 33.9, 33.3, 32.4, 30.4, 32.6, 34.7, 34.9, 31.6, 35.2, 35.3, 35.5, 36.4, 36.6, 37.0, 34.3, 32.5, 31.4, 34.4, 35.6, 37.3, 37.5, 36.9, 37.0, 36.3, 36.9, 36.7

Maximum Temperatures (in degrees)	Tally marks	Frequency
28 - 30		1
30 - 32		5
32 - 34		8
34 - 36		7
36 - 38		10

Relative humidity (in percent)

90,97,92,95,93,95,93,85,83,85,83,77,83,77,74,60,71,65,74,80,87,82,81,76,61,63,58,58,56,57,54

Relative Humidity (in Per cent)	Tally marks	Frequency
50 - 60		5
60 - 70		4
70 - 80		6
80 - 90		9
90 - 100		7

Activity

The data below shows the mass of 40 students in a class. The measurement is to the nearest kg.

55	70	57	73	55	59	64	72
60	48	58	54	69	51	63	78
75	64	65	57	71	78	76	62
49	66	62	76	61	63	63	76
52	76	71	61	53	56	67	71

Complete the frequency table for the data using an appropriate scale.

Mass (kg)	Frequency
45 – 49	
50 – 54	
55 – 59	
60 – 64	
65 – 69	
70 – 74	
75 – 79	

LESSON 68

LO: Organize and display data using line graphs

What Is A Line Graph?

We may represent data using a line graph. A line graph is formed by joining the points given by the data with straight lines.

A **line graph** is usually used to show the **change of information over a period of time**. This means that the horizontal axis is usually a **time scale**, for example minutes, hours, days, months or years.

How to draw a line graph?

1. Decide on a scale and interval.
2. Graph pairs of data and draw a line to connect each point.

How to interpret line graphs?

By observing the upward or downward slant of the lines connecting the points, you can describe the trends in the data and predict future events.

Example 1:

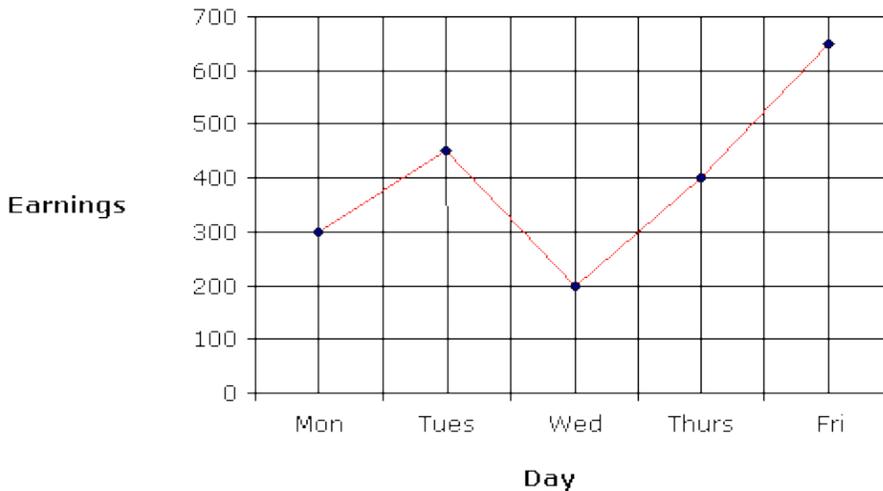
The table shows the daily earnings of a store for five days.

Day	Mon	Tues	Wed	Thurs	Fri
Earnings	300	450	200	400	650

- a) Construct a line graph for the frequency table.
 b) On which days were the earnings above \$ 400

Solution:

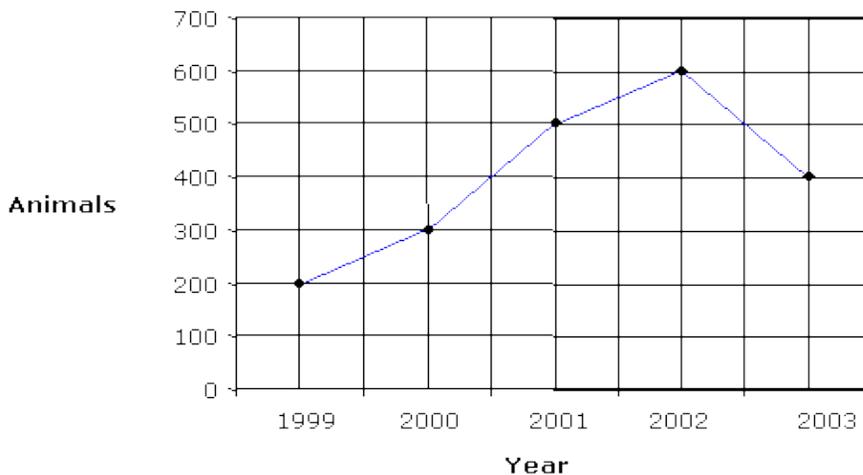
a)



- b) The earnings were above \$ 400 on Tuesday and Friday.

Example 2:

The following line graph shows the total number of animals in a zoo.



- a) In which year did the zoo have the largest number of animals?
 b) What is the percentage increase of animals in the zoo from 1999 to 2001?

Solution:

- a) The zoo had the largest number of animals in 2002

b) The percentage increase of animals in the zoo from 1999 to 2001 is

$$\frac{500 - 200}{200} \times 100\% = 150\%$$

Multiple sets of related data can also be represented by using multiple lines on a line graph.

Activity

The table shows the daily sales in RM of different categories of items for five days.

Day	Mon	Tues	Wed	Thurs	Fri
Drinks	300	450	150	400	650
Food	400	500	350	300	500

- Construct a line graph for the frequency table.
- On what days were the sales for drinks better than the sales for food?
- What is the total earnings for food and drinks on Wednesday?

LESSON 69

LO: Organize and display data using Bar Graphs

A **bar graph** represents the data as vertical bars. The length of each bar is proportional to the amount that it represents.

When constructing a bar chart it is important to choose a suitable scale to represent the frequency.

Example:

The following table shows the number of visitors to a park for the months January to March.

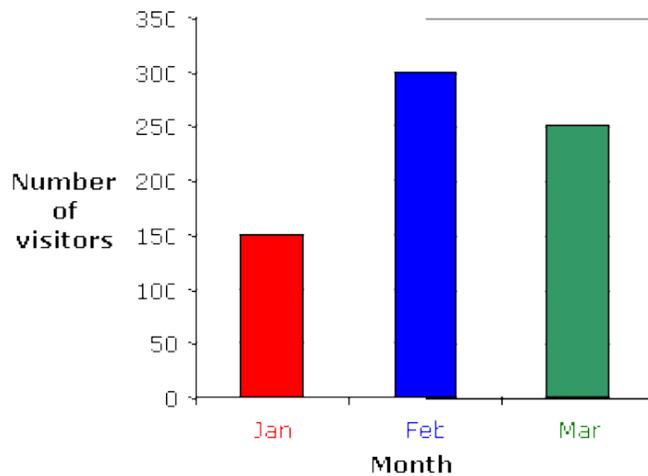
Month	January	February	March
Number of visitors	150	300	250

- Construct a Bar Graph for the table.
- What is the percentage of increase of visitors to the park in March compared to January?
- What percentage of visitors came in February compared with total number of visitors over the three months?

Solution:

- If we choose a scale of 1:50 for the frequency then the bar graph will be:

Vertical bar chart



b) Increase in March compared to January is

$$\frac{250 - 150}{150} \times 100\% = 66.67\%$$

c) Percentage of visitors in February compared to the total number of visitors is

$$\frac{300}{150 + 300 + 250} \times 100\% = 42.86\%$$

Double Bar Chart

The **double bar chart** is used when we want to represent two sets of data on the same chart. We can put the bars side by side or we may put the bars of one set of data on top of the bars of the other set of data.

A side by side chart is more useful when we compare the two sets of data (example: the number of adult visitors as compared to the number of child visitors); whereas the stacked chart emphasizes the totals of the two sets of data (example: total number of visitors).

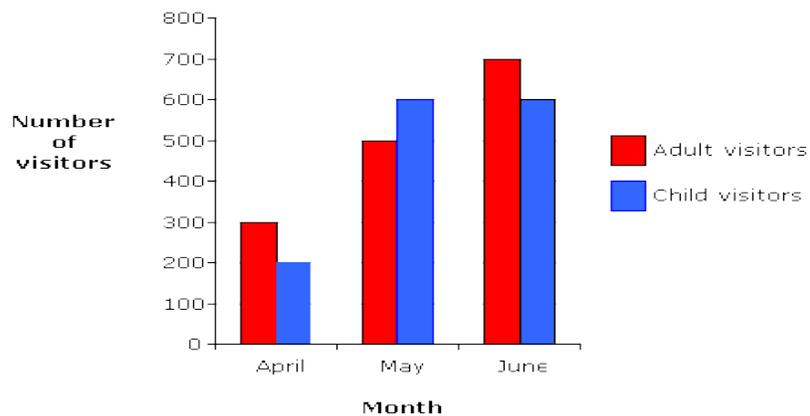
Example:

The following frequency graph shows the number of adult visitors and child visitors to a park. Construct a side by side double bar chart and a stacked double bar chart for the frequency table.

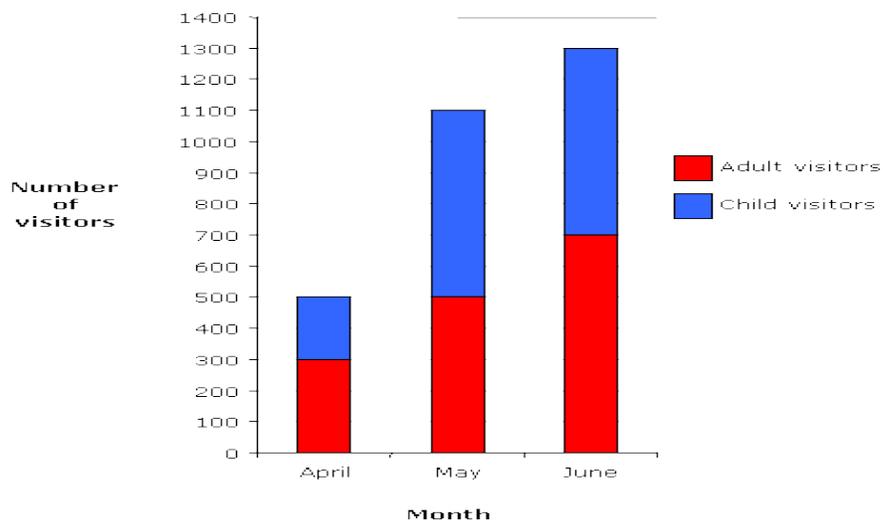
Month	April	May	June
Number of adult visitors	300	500	700
Number of child visitors	200	600	600

Solution:

Side By Side Double Bar Chart



Stacked Double Bar Chart



Activity

1. The weekly sale of pencil boxes in a stationary shop is given in the table below. Using a suitable scale, represent the given information on a bar graph.

Days
Mon, Tues, Wed, Thurs, Fri, Sat
Pencil Boxes Sold
10, 25, 30, 40, 50, 10

- (i) Draw and Write the title of the bar graph.
- (ii) On which day were the maximum pencil boxes sold?
- (iii) If the shopkeeper decides to close his shop for one more day each week, selection of which days would lead to minimum loss of sale and maximum loss of sale?

2. A survey of 36 students of a class was done to find out the mode of transport used by them while commuting to the school. The collected data is shown in the table given below. Represent the data in the form of a bar graph.

Mode of Transport	Number of Students
Cycle	6
School Bus	16
Walking	10
Car	4

- (i) Draw and give a suitable title for the graph.
- (ii) How many students walk to school?
- (iii) Which mode of transport is used by most of the students?

RATU NAVULA COLLEGE
Y11 LIFE MATHEMATICS WORKSHEET 10 QP

1. A simple random sample is one in which
 - A. From a random starting point, every n th unit from the sampling frame is selected
 - B. A non-probability strategy is used, making the results difficult to generalize
 - C. The researcher has a certain quota of respondents to fill for various social groups
 - D. Every unit of the population has an equal chance of being selected
2. It is helpful to use a multi-stage cluster sample when
 - A. The population is widely dispersed geographically
 - B. You have limited time and money available for travelling
 - C. You want to use a probability sample in order to generalise the results
 - D. All of the above
3. What effect does increasing the sample size have upon the sampling error?
 - A. It reduces the sampling error
 - B. It increases the sampling error
 - C. It has no effect on the sampling error
 - D. None of the above
4. The following data shows the test marks obtained by a group of students.

Draw a frequency table for the data and find the mode.

6 7 7 1 3 2 8 6 8 2
 4 4 9 10 2 6 3 1 6 6
 9 8 7 5 7 10 8 1 5 8

5. The number of trees planted by Eco-club of a school in different years is given below.

Year	2005	2006	2007	2008	2009	2010
Number of Trees to be Planted	150	220	350	400	300	380

Draw the bar graph to represent the data.