# Use and Report Health Information

**UNIT 3**

## Introduction

*Data only becomes really useful when they have been processed and analysed. Health data may come from routine information systems, surveillance systems, investigation of epidemics or special surveys.*

(Vaughan & Morrow, 1989: 99)

Unit 3 is a practical guide to using and reporting on epidemiological data. Study Session 1 is an introduction to some important terms and tools used in the analysis of this data. The Study Session deals with preparing tables for data analysis and applying the statistical tools of range, mode, median and mean. Study Session 2 uses the results of this analysis to prepare graphical illustrations to be included in research reports or presentations.

In Study Session 3, you will be taken step-by-step through the process of using EXCEL to represent health information. This includes making spreadsheets and converting these into graphs of various types.

Study Sessions 4 and 5 are designed to assist you in producing Assignment 2. Initially, you will develop a suitable set of critical questions related to your dataset. After this, you will be guided through the process of preparing, drafting, editing and proofreading a well-constructed epidemiological report.

There are five Study Sessions in Unit 3.

Study Session 1: Make Sense of the Data Study Session 2: Represent Health Information

Study Session 3: Represent Health Information using EXCEL

Study Session 4: Develop Critical Questions for your Dataset and Prepare for Assignment 2

Study Session 5: Final Report (Assignment 2)

# Unit 3 - Study Session 1 Make Sense of the Data

## Introduction

In the module on *Health Systems Research*, you explored different methods of data collection and analysis. At this stage, a far simpler approach will be taken, using only the most basic statistical tools. Whether you go on to actively engage in research, carry out routine data collection as part of a disease surveillance system or simply have to make sense of the information presented to you by someone else in a report or journal article, you will need tools to help you analyse the information and make sense of it.

This Study Session introduces a few simple techniques to assist you with the process of summarising, analysing and interpreting a set of epidemiological data. The first involves counting the frequency of occurrence of categorical variables, and testing for association between two categorical variables. The second approach involves calculating the mean, median, mode and range for continuous variables.

## Contents

1. Learning outcomes of this session
2. Readings
3. Prepare tables for data analysis
4. Use simple statistical tools
5. Session summary

## Timing of this session

This Study Session contains one reading and three tasks. It should take you no more than two hours to complete.

## LEARNING OUTCOMES OF THIS SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Review an epidemiological report. * Determine the Public Health implications of a given report or set of data. | ***Academic Learning Outcomes***   * Apply a systematic approach to the analysis of raw data. * Employ simple statistical techniques. |

1. **READINGS**

There is one reading to which you will be referred in the course of the Study Session.

|  |  |
| --- | --- |
| **Author/s** | **Publication Details** |
| Vaughan, J. P. & Morrow, R. H. | (1989). Ch 10 – Data Processing and Analysis. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 99–112. |

## PREPARE TABLES FOR DATA ANALYSIS

Without getting too far into the complexities of statistical analysis, you can learn a lot about the raw dataset by counting the frequency for each variable. In other words, how often the variable occurs within a given interval. The tasks that follow are based on Chapter 10 of Vaughan & Morrow (1989).

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 10 – Data Processing and Analysis. In

*Manual of Epidemiology for District Health Management.* Geneva: WHO: 99–112.

For the tasks you are to assume that you, or one of your colleagues, carried out a small cross-sectional survey of 100 randomly selected villagers to determine the prevalence of hookworm infection and anaemia. The resulting data is contained in the above reading on pages 102–105.

You believe that there is an association between hookworm infestation and the levels of haemoglobin in the villagers and you want to test this theory.

The tasks will guide you through this process.

Before proceeding with the task, it would be useful to read section 10.3 of Vaughan & Morrow (1989) on pages 101–102 for helpful advice on processing data.

Vaughan and Morrow (1989) make the following suggestions:

* When processing data, each variable can be counted and these counts can be summarised in tables. This enables one to draw graphs and diagrams.
* When preparing tables:
  + All data should be in a form that can be classified into categories.
  + All table categories must be mutually exclusive.
  + Each table should include *all the raw data*.

**TASK 1 - Prepare a frequency distribution for the variables in the study**

1. Study the data supplied in Vaughan & Morrow (1989), Chapter 10, pages 102–

105. Note that the study provides information on four variables – age, sex, haemoglobin levels and presence/absence of hookworm infection.

1. Create four separate tables to summarise each of the variables.

The sex and hookworm data is discrete; it is M or F, + or −. It is much easier to simply count up the number of Ms, Fs, +s and −s. Note that the age and haemoglobin data is continuous, i.e. the range of possible values is infinite, or in other words, the numbers possible could range from any number between the smallest fraction (e.g. 0.0000124), to the biggest number (e.g. 135 367) and anywhere in between. Age ranges from 0 to 71 in this group. Haemoglobin ranges from 6.2 to 13.8.

For the continuous data you need to select appropriate *intervals.* This involves grouping certain ages together (e.g. 0–4, 5–14, 15–44, 45+), and the same process of grouping certain haemoglobin levels together (e.g. 6.0–6.9, 7.0–7.9 etc). Next you count how many people are to be found in each of those intervals, and how many people have haemoglobin levels that fall within each of those intervals.

#### FEEDBACK

1. Constructing a frequency count of data in this way is the simplest and most commonly used method of summarising a set of raw data.
2. Compare your own tables with those displayed in Vaughan & Morrow (1989) pages 106–107.

The main question we wish to answer using the hookworm study data is:

*Is the hookworm infection associated with low levels of haemoglobin?*

The next task shows how the data can be organised to make it possible to see if this association exists. It starts by formatting the data into intervals and constructing a two by two table.

|  |  |  |  |
| --- | --- | --- | --- |
| Haemoglobin level | Hookworm infection | | Total |
| Present | Absent |
| Anaemia  (Less than 10 g/100 ml) | ***A***  Number of people with Hk  & with anaemia = | ***B***  Number of people without Hk  & with anaemia = | 46 |
| No anaemia  (10 g/100 ml or more) | ***C***  Number of people with Hk  & without anaemia = | ***D***  Number of people without Hk  & without anaemia = | 54 |
| Total | 59 | 41 | 100 |

#### FEEDBACK

**TASK 2 - Test for an association between hookworm infection and anaemia**

For most people, normal levels of haemoglobin range from 12 to about 16 grams per 100 ml of blood. Based on these figures for normal levels, anaemia is defined as a haemoglobin level of less than 10 g per 100 ml of blood.

1. Using the data from the previous task (Vaughan & Morrow, 1989, 102–105), re-group it into the four groups represented in the shaded blocks of the table below.

**Distribution of people with hookworm infection and anaemia**

2. Finally, compare the data for the two variables by asking the question:

*Does hookworm infection appear to be strongly associated with the presence of anaemia or not?*

If the numbers of people in blocks **A** and **D** are high, then it means that hookworm is often present in the same people who have anaemia, and that hookworm is generally absent in the people who have no anaemia. This will mean that anaemia is strongly associated with the presence of hookworm infection. See Vaughan & Morrow (1989) page 108 for comments.

## USE SIMPLE STATISTICAL TOOLS

It is possible to learn a lot about the raw dataset by counting the frequencies for each variable, as you did for the hookworm example. It is sometimes helpful to express such information as a rate or percentage. For example, 59% of the subjects in the previous example had a hookworm infection and 76.1% had anaemia.

Refer back to Unit 1, Study Session 3, Section 4, Tasks 3–7 for how to calculate rates. Note that a percentage is the same as rate per 100.

Other commonly used methods describe the variables in terms of their range, mode, median or mean (average).

* + The *range* is the difference between the minimum and maximum values recorded.
  + The *mode* is the most commonly occurring value. It is more useful when applied to discrete variables or categories.
  + The *mean* is the same as the average. It is calculated by adding all the values together and dividing the total by the number of people in the sample.
  + The *median* is the value of the person in the middle of the sample once you have arranged the variables from smallest to biggest.

These methods are mainly applicable to continuous data such as age and haemoglobin levels in the previous example.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 10 – Data Processing and Analysis. In

*Manual of Epidemiology for District Health Management.* Geneva: WHO: 99–112.

Read Section 10.6 of Vaughan & Morrow (1989) for a few simple illustrations of these statistical tools.

**TASK 3 – Calculate characteristics of continuous variables**

Go back to the data in the previous example and calculate the characteristics of the continuous variables as indicated below:

1. Range
2. Mode
3. Mean
4. Median

#### FEEDBACK

1. *The range* is the difference between the minimum and maximum values recorded. For age: 71 − 0 = 71 years. For haemoglobin levels: 13.8 − 6.2 = 7.6 g/100ml.
2. *The mode* is the most commonly occurring value. It is more useful when applied to discrete variables or categories. For age it is 4 years of age. There are 6 people of this age in the sample. For haemoglobin level it appears to be 10.9, which also occurs 6 times. You could also say that the interval with the most values in it is the 10–10.9 interval in which there are 24 people. So this interval could also be regarded as the mode of the distribution.
3. *The mean* is the same as the average. It is calculated by adding all the values together and dividing the total by the number of people in the sample. For age, the total of all the ages is 2463. There are 100 people therefore the mean age is

2463 divided by 100 which is 24.6 years. For haemoglobin, the total is 994.2 and the mean is therefore 9.9 g/100ml. The mean tells you that most of the other values are close to this mean value.

1. *The median* is the value of the person in the middle of the sample once you have arranged the variables from smallest to biggest. If you arrange the ages in this way, the 50th person in the sample is 20 years of age and the 51st person is 21 years old. This means the age at the middle of this sample of 100 people is 20.5 years. If you do the same thing with the haemoglobin levels, the median turns out to be 10.6 g/100ml. The median tells you where the middle of the distribution of values is located. It can be close to the mean in a very evenly spread out distribution.

It is a lot easier to sort and rank variables from smallest to biggest if they are entered into a computerised spreadsheet, database or table. If the data is in a database file, you can usually get a statistical programme such as *Epi Info* to do all these calculations for you. These simple statistical tools can be used to do a preliminary analysis of almost any set of descriptive data.

## SESSION SUMMARY

This Study Session has introduced some important terms and tools for the analysis of epidemiological data. In the next Study Session, we use the results of this analysis to prepare graphical illustrations. These can add emphasis and clarity to a research report.



Raw data have to be processed to provide useful information, such as:

* spatial trend: District 1 has a 2-fold HIV incidence as compared to District 2.
* temporal evolution: tuberculosis incidence decreased by 25% between 2005 and 2008.
* associations between indicators: TB incidence and HIV incidence are co-evolving.

**KEY POINTS IN UNIT 3 SESSION 1**

**Making Sense of the Data**

# Unit 3 – Study Session 2 Represent Health Information

## Introduction

So far in this module you have read and interpreted graphical information in both your Module Guide and in the Readings. You have constructed line graphs, histograms and tables to make it easier to understand raw data. The translation of raw data into graphic form can add substantially to the impact it has on the reader. Such graphics can really help you to emphasise the important features of the data. They are an essential component of any audio-visual reporting session and certainly liven up a printed report for those who need to read it and be persuaded by it.

Remember that doing the research is just the first step in the process of making people sit up and take note of your research findings. Your report needs to be persuasive. It must convince people to take it seriously and act to address the health problems you have highlighted.

In this Study Session, we look at the different kinds of graphical illustrations you could use. We briefly discuss how to decide what to illustrate and how.

## Contents

1. Learning outcomes of this session
2. Readings
3. Interpret graphs and tables
4. Use graphical representation to illustrate data
5. Session summary

## Timing of this session

This Study Session contains three readings and four tasks. This should take you about an hour and a half to complete.

## LEARNING OUTCOMES OF THIS SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Assess a variety of formats for representing different health data. * Interpret health data presented in various graphical and numerical formats. | ***Academic Learning Outcomes***   * Illustrate a set of data using simple graphical representation techniques. |

1. **READINGS**

There are three readings to which you will be directed in the course of the Study Session.

|  |  |
| --- | --- |
| **Author/s** | **Publication Details** |
| Bonita, R., Beaglehole, R. & Kjellstrom, T. | (2006). Ch 1 – What is Epidemiology? In *Basic Epidemiology.* 2nd Ed.Geneva: WHO: 1–6, 10. |
| Vaughan, J. P. & Morrow, R. H. | (1989). Ch 11 – Presenting Health Information. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 113–124. |
| Joubert, G., Ehrlich, R.,  Katzenellenbogen,  J. M., & Abdool Karim, S. S. | (2007). Ch 11 – Analysing and Interpreting epidemiologic data. In *Epidemiology: A Manual for South Africa.2nd Ed.* Cape Town: Oxford University Press: 141–154. |

## INTERPRET GRAPHS AND TABLES

Epidemiological texts often make use of tables, graphs and other illustrations to present information. Bonita, Beaglehole & Kjellstrom (2006) include several graphs and tables that support the text. You need to be able to read and interpret graphs and tables with ease.

Here are some examples to check your skills and to practise on. If you have difficulty, ask a colleague to help you to develop a simple strategy for reading them.

Make sure that you understand the concept of *ratio* and *rate* before you start (Joubert

*et al*., 2007: 19–23).

We usually want to compare disease rates across different communities or areas. However, each community has a different population and there is a difference in the number of people who are ill. A percentage is an example of a rate multiplied by 100. When we deal with population figures that are fairly large, we usually calculate rates out of 1 000 or 100 000 because the figures are then easier to grasp.

#### Example

Community A has 250 cases of people with TB. Community B has only 100 cases of TB. Seen on its own, this information seems to suggest that TB is a bigger problem in Community A.

If Community A has a population of 123 500, it means that 250 out of the 123 500 people living there have TB. Use a calculator to divide the number 250 by 123 500 and you will get 0.002. Multiply this by 100 and you will see that 0.2% of the population in Community A have TB. If instead you multiply by 1 000, you will get a prevalence of 2/1 000 cases of TB. Alternatively, you could multiply this by 100 000 to get a rate of 200/100 000.

If Community B has a population of 25 000, this means their TB prevalence is 100/25 000. Divide the 100 cases of TB by the population of 25 000 and you get

* 1. Multiply this by 100 and you get 0.4%. If you multiply this by 1 000 you will get a prevalence of 4/1 000. This is double the prevalence in Community A, so Community B actually has a much bigger TB problem than Community A.

#### TASK 1 - Make sense of a table

Bonita, Beaglehole & Kjellstrom (2006), page 2, Table 1.1, provides summary data on deaths in a cholera outbreak that took place in two districts of London, in 1851.

**READING**

Bonita, R., Beaglehole, R. & Kjellstrom, T. (2006). Ch 1 – What is Epidemiology? In *Basic Epidemiology.* 2nd Ed.Geneva: WHO: 1–14.

* + 1. Compare the number of deaths from cholera in the districts of Southwark and Lambeth.

There are two ways to express this, as a difference and as a ratio.

To calculate the difference, subtract the smaller figure from the larger one, i.e. Lambeth from Southwark.

However, this does not usually tell us how serious the problem is because of the differences in population size in the districts being compared. Calculating the proportion – or ratio – of the population of Southwark that died from cholera, and comparing that to the ratio of cholera deaths in Lambeth is of greater value to Public Health workers.

* + 1. Roughly how many more deaths were there in Southwark than Lambeth? 20 times more ; 10 times more ; 40 times more ; 80 times more 

1. Compare the populations of Southwark and Lambeth. How many times bigger was the Southwark population than the Lambeth population?

20 times bigger ; 8 times bigger ; 2 times bigger ; 50 times bigger 

1. Why is it necessary to calculate the death rate per 1 000 of the population, as shown in the last column? From the figures provided, can you work out how to calculate the death rate?
2. What do these death rates tell us about the risk of getting cholera in these two districts in 1851? Would you rather have lived in Southwark or Lambeth?
3. Which column in the table contains the most important information? Why do you say so?

#### FEEDBACK

1. The difference in the number of deaths in the two areas is 826, but this tells us very little except that there were more deaths in Southwark.
2. To find out how many times more deaths there were in Southwark, you divide the smaller number of deaths (18) into the larger number (844), i.e. how many times does 18 go into 844? You will find that there were over 40 times more cholera cases in Southwark than there were in Lambeth.
3. For the epidemiologist, the number of deaths (on its own) means very little unless we recognise that the population sizes being compared are very different. This is why we use a rate to compare them. Divide Southwark’s population of 167 654 by Lambeth’s population of 19 133 and you get a rough figure of 8. Southwark therefore had a population 8 times greater than Lambeth.
4. The death rates are calculated by dividing No. of Deaths by Population in 1851 and multiplying by 1 000. In Southwark, it would be (844 167 654)  1 000 = 5.0 (or 5 per 1 000), whereas in Lambeth it would be 18  19 133 = 0.9 (or less than 1 per 1 000). The epidemic is therefore much more serious in Southwark.
5. A higher proportion of people got cholera in Southwark. If you lived in Southwark, the probability (risk) that you would be one of the victims of cholera was much higher than if you lived in Lambeth.
6. The main result in this table is to be found in the column entitled *Cholera death rate per 1 000 population.*

The role of water supply in causing cholera and the use of geographical mapping to trace the cause will be explored in Unit 2 Study Session 4, in a description of the 1851 London cholera epidemic by John Snow.

## Make sense of a graph

Now turn to Bonita, Beaglehole & Kjellstrom (2006), page 3, Figure 1.1. The reading illustrates a relationship between deaths from lung cancer amongst British doctors between 1951 and 1961 and their levels of cigarette smoking activity.

Two important elements found in graphs are variables and axes. A variable is a characteristic that can be measured. There are two in this example: the number of people who died of lung cancer, and the number of cigarettes each person smoked per day. In Figure 1.1 the two variables are plotted as follows: the y-axis (vertical) shows how many British doctors died of lung cancer between 1951 and 1961 out of every 1 000 doctors who died during this period. The x-axis (horizontal) indicates the number of cigarettes each of those doctors smoked per day before they died.

#### TASK 2 - Read and interpret a graph

Your task is to interpret the graph and determine what kind of relationship there appears to be between the two variables.

Have a look at Figure 1.1. According to its title, this graph looks at all British doctors who died of lung cancer between 1951 and 1961. It groups these doctors according to how many cigarettes they smoked each day. For convenience, it converts these into a rate, i.e. the number of doctors who died of lung cancer out of every 1 000 doctors who died. For example 1.5 out of every 1 000 doctors who died of lung cancer smoked 20 cigarettes per day.

Here is some guidance on how to read the graph. Focus on the black dots plotted on the graph. Find the dot that represents those doctors who smoked 15 cigarettes per day and experienced a death rate of 1.0. This death rate means that among the doctors who smoked 15 cigarettes per day, one doctor in 1 000 died of lung cancer. The fourth dot from the left represents those doctors who smoked 15 cigarettes per day. A horizontal line drawn from this dot onto the y-axis shows that this group of doctors experienced a death rate of 1.0 per 1 000.

* 1. What is the death rate for doctors who smoked:
     + 10 cigarettes per day?
     + 20 cigarettes per day?
     + 30 cigarettes per day?
     + 40 cigarettes per day?
  2. What do you notice about the association between higher daily cigarette consumption and the lung cancer death rate?
  3. Look at the straight line inserted across the graph. Notice that it is very close to all the dots and almost connects them together. The line was inserted to illustrate the relationship between the information on the x-axis and the information on the y-axis. What can such a straight line tell us about the relationship between the two variables, i.e. death rates and cigarettes per day?
  4. This graph is an example of an x-y scatter plot. What do you think this means? What is the purpose of such a graph?

#### FEEDBACK

1. You should have determined the death rate for doctors who smoked:
   * 10 cigarettes per day as about 0.75 per 1 000
   * 20 cigarettes per day as about 1.6 per 1 000
   * 30 cigarettes per day as about 2.4 per 1 000
   * 40 cigarettes per day as about 3.3 per 1 000
2. Deaths from lung cancer are strongly associated with the number of cigarettes smoked daily. Because the line slopes up toward the right hand side, it means that there are more heavy-smokers who die of lung cancer than light-smokers. Put another way, the more cigarettes you smoke per day, the more likely it is that you will die of lung cancer.
3. The most striking characteristic of this graph is that the dots form an almost completely straight line. The line has been added to the graph to illustrate this. When you see this kind of “straight line relationship” you can usually assume that information on the vertical y-axis is strongly associated with the information on the horizontal x-axis.
4. An x-y scatter plot is a very simple and neat way of showing whether there appears to be a relationship between two variables, the one indicated on the x- axis and the other on the y-axis. If the dots cluster together in some part of the graph, there may be an association of some kind between the variables. If they are spread all over the graph, it is very unlikely that there is any association present.

Now try making sense of another kind of graphical illustration.

**TASK 3 - Read a graphical illustration**

Take a look at Bonita, Beaglehole & Kjellstrom (2006), page 10, Figure 1.8 which illustrates some important features of the AIDS epidemic.

Study this figure and try to identify what it tells you about the AIDS epidemic.

1. What is the main point this illustration is trying to make?
2. Why is HIV/AIDS called the hidden epidemic?
3. Why have the authors chosen to use a pyramid shape to represent what is happening with the HIV/AIDS epidemic?
4. What are the consequences of the hidden nature of HIV/AIDS for the public? In what way does this affect our ability to accurately measure/monitor the size of the epidemic?

#### FEEDBACK

**READING**

Joubert, G., Ehrlich, R., Katzenellenbogen, J. M. & Abdool Karim, S. S. (2007). Ch 11 – Analysing and Interpretating epidemiologic data. In *Epidemiology: A Manual for South Africa.* 2nd Ed. Cape Town: Oxford University Press: 141–54.

1. Figure 1.5 demonstrates how a good choice of graphical image can emphasise the main point the writer is trying to illustrate. The pyramid is a powerful and widely recognised symbol, with a prominent point at the top and a broad solid base below. Others have used the image of an iceberg floating with its tip above water and large bulk below.
2. As the title suggests, the illustration is trying to highlight one main characteristic of the AIDS epidemic, i.e. the fact that most people with AIDS are not identifiable because they are either in the sub-clinical phase of the disease and do not know they have it, or they are not sick enough to report to a clinic or doctor where they can be diagnosed.
3. Like the iceberg or pyramid, there are far more invisible cases (below the shaded plane) than reported, visible cases of AIDS in the population. This makes accurate measurement of the actual extent of the epidemic in the population very difficult.
4. This in turn makes it very difficult to lobby for resources or run programmes to respond to the epidemic. Many people still struggle to believe it is actually a major problem.

## 4 USE GRAPHICAL REPRESENTATION TO ILLUSTRATE DATA

While graphical illustrations must never replace the text, they can be very effective in highlighting or emphasising certain features of the data. You also need to be wary of constructing a table or a graph for every piece of data you have summarised. Many smaller and less complex pieces of information can be clearly and easily stated directly in the text without the aid of graphics or tables.

Study page 113 of the reading below and then proceed to Task 4.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 11 – Presenting Health Information. In

*Manual of Epidemiology for District Health Management.* Geneva: WHO: 113–124.

**TASK 4 - Select effective illustration methods for datasets**

1. Consider which method mentioned in the above reading would most effectively illustrate the dataset that follows.

2. Again, use the dataset below. Decide which variables or relationships between variables you would represent in a table, a histogram, a pie chart, a scatter plot, a line graph or in the text.

A sample of 50 primary school children was surveyed to test the relationship between age (years), gender (M or F), weight (kilograms) and sugar consumption habits (rated on a scale from 0-5, with 5 being the highest level of excess sugar consumption). The following table of raw data was prepared.

#### Sugar Consumption in Primary School Children

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Rec.No.** | **Age** | **Sex** | **Weight** | **Sugar** |
| 1 | 6 | M | 31 | 3 |
| 2 | 14 | M | 60 | 5 |
| 3 | 8 | F | 36 | 1 |
| 4 | 5 | F | 27 | 2 |
| 5 | 10 | M | 42 | 4 |
| 6 | 11 | F | 44 | 3 |
| 7 | 13 | M | 50 | 3 |
| 8 | 7 | M | 34 | 5 |
| 9 | 7 | F | 35 | 2 |
| 10 | 9 | F | 37 | 0 |
| 11 | 6 | M | 32 | 2 |
| 12 | 7 | M | 35 | 3 |
| 13 | 11 | M | 44 | 5 |
| 14 | 10 | F | 40 | 2 |
| 15 | 12 | F | 47 | 2 |
| 16 | 6 | F | 32 | 1 |
| 17 | 10 | M | 40 | 4 |
| 18 | 9 | M | 39 | 2 |
| 19 | 8 | M | 36 | 3 |
| 20 | 11 | F | 44 | 1 |
| 21 | 7 | F | 35 | 1 |
| 22 | 9 | F | 38 | 0 |
| 23 | 6 | M | 29 | 5 |
| 24 | 8 | F | 34 | 4 |
| 25 | 10 | F | 41 | 3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Rec.No.** | **Age** | **Sex** | **Weight** | **Sugar** |
| 26 | 9 | M | 36 | 5 |
| 27 | 9 | F | 38 | 2 |
| 28 | 6 | M | 32 | 4 |
| 29 | 10 | M | 42 | 4 |
| 30 | 7 | F | 30 | 3 |
| 31 | 11 | M | 43 | 3 |
| 32 | 8 | F | 36 | 2 |
| 33 | 12 | F | 46 | 0 |
| 34 | 12 | M | 48 | 4 |
| 35 | 7 | M | 33 | 3 |
| 36 | 6 | F | 30 | 2 |
| 37 | 9 | M | 37 | 1 |
| 38 | 12 | M | 48 | 3 |
| 39 | 11 | F | 45 | 2 |
| 40 | 11 | M | 46 | 3 |
| 41 | 7 | F | 34 | 2 |
| 42 | 10 | F | 40 | 2 |
| 43 | 12 | M | 50 | 4 |
| 44 | 8 | M | 36 | 5 |
| 45 | 8 | F | 37 | 1 |
| 46 | 13 | F | 51 | 3 |
| 47 | 9 | F | 37 | 2 |
| 48 | 11 | F | 41 | 0 |
| 49 | 8 | M | 35 | 4 |
| 50 | 7 | M | 33 | 4 |

**FEEDBACK**

There are no absolute answers to these questions.

There are 50 children, and this could be mentioned in the *text*. Half of them are girls. You could also state in the text that 50% of the sample is male. A *pie chart* is good for illustrating what proportions of the whole are made up by individual groups, but it would be unnecessary to illustrate this simple proportion.

A *frequency distribution table* could be prepared to illustrate the age composition:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Age** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** |
| **Number** | 1 | 6 | 7 | 7 | 7 | 6 | 7 | 5 | 2 | 1 |

This could also be represented on a *histogram*, although this would not really add to the clarity of the data and there is nothing particularly important about age distribution that you would want to emphasise at this stage. A variation on this table would be to extend it to include the breakdown of the sample by gender and age as in the example in Vaughan & Morrow (1989) page 106.

The sugar consumption data could be rather neatly illustrated with a *histogram*. Here too you have the opportunity to split the data by gender.

#### Histogram of Sugar Consumption Levels by Gender

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Frequency** | | | | | | | | | | | | |
|  | **0** | **11** | | | | | | | | **6**  **1** | | **0** |
| **11** | **4 4** | | | |  | **8 8** | | |
| **10** |  |
| **9** |  |
| **8** |  |  | **4** | |
| **7** |  |  |
| **6** |  |  |
| **5** |  |  |
| **4** | **1** | |  | **2** |  |  |  |  |
| **3** |  |  |  |  |  |
| **2** |  |  |  |  |  |  |
| **1** |  |  |  |  |  |  |  |  |  |  |
| **Gender** | **M** | **F** | **M** | **F** | **M** | **F** | **M** | **F** | **M** | **F** | **M** | **F** |
|  | **0** | | **1** | | **2** | | **3** | | **4** | | **5** | |
| **Sugar consumption score** | | | | | | | | | | | |

The *frequency distribution table* would look like this:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Gender** | **Sugar consumption score** | | | | | | |
| **0** | **1** | **2** | **3** | **4** | **5** | **Total** |
| **Male** | 0 | 1 | 2 | 8 | 8 | 6 | 25 |
| **Female** | 4 | 5 | 11 | 4 | 1 | 0 | 25 |
| **Total** | 4 | 6 | 13 | 12 | 9 | 6 | 50 |

Since both age and weight are continuous variables, you could calculate the range, median and mean for each of them and report these in the text of your report.

For the same reason, you could prepare *a scatter plot* (X-Y plot) using the weight and age variables to see if this illustrates some kind of association between them.

***Scatter plot of Age and Weight***

**16**



**14**

**12**

**10**

**Age in years**

**8**

**6**

**4**

**2**

**0**

**0 10 20 30 40 50 60 70**

**Weight in kilograms**

## 5 SESSION SUMMARY

In this Study Session, you have considered a range of methods of representing datasets including tables, graphs, diagrams and descriptions through text. Not every variable requires such representation; it is therefore important to be selective when making graphical representations of data. It has also been noted that graphical representations can make a substantial impact.

In the next Study Session, we explore how to use the computer programme called EXCEL to represent health information and data.



You have covered common ways of representing health information. Try the Self-evaluation task at the end of Session 3.

**KEY POINTS IN UNIT 3 SESSION 2**

**Representing Health Information**

# Unit 3 – Study Session 3 Representing Health Information using EXCEL

## Introduction

In the previous Study Session you learned how to interpret and represent health information. In the last task you would have created some simple figures and graphs by hand. However for presenting your data in a professional scientific report, hand drawn figures and graphs would be problematic. In this Study Session we will learn to create many of the graphs you saw in the previous Study Session using Microsoft EXCEL 2010.

Note if you have Microsoft EXCEL 2007 the basic concepts will be the same however the screens as presented in this Study Session will look slightly different and the location of the pull-down menus will be different. If you have EXCEL 2007 you may want to review the following on-line tutorial from MS Office first before proceeding with this Study Session.

EXCEL 2007 CHARTS

[http://office.microsoft.com/en-za/excel-help/charts-i-how-to-create-a-chart-in-excel-](http://office.microsoft.com/en-za/excel-help/charts-i-how-to-create-a-chart-in-excel-2007-RZ010175754.aspx) [2007-RZ010175754.aspx](http://office.microsoft.com/en-za/excel-help/charts-i-how-to-create-a-chart-in-excel-2007-RZ010175754.aspx)

If you still have Microsoft EXCEL 2003 then please contact the course convenor and we can make an older version of these notes available to you.

Most standard computers have Microsoft EXCEL as part of their Microsoft OFFICE package. If you do not have access to EXCEL then you will not be able to use it for your assignment and you can skip this Study Session. Please notify SOPH Student Administration and Module Tutor that you do not have EXCEL, as the use of EXCEL figures and graphs will be expected in your 2nd Assignment.

If you already know how to create figures and graphs in EXCEL you do not need to complete this Study Session though you might find it a helpful review

## Contents

1. Learning outcomes of this session
2. Readings
3. Get started in EXCEL and make histograms
4. Make a pie chart
5. Make histograms to compare groups
6. Make scatter plots to compare groups
7. Session summary

## Timing of this session

This Study Session contains mostly practical work using a computer, and there are three tasks and no readings. It should take you about an hour and a half to complete.

## LEARNING OUTCOMES OF THIS SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Assess a variety of formats for representing different health data using MS EXCEL. * Interpret health data presented in various graphical and numerical formats using MS EXCEL. | ***Academic Learning Outcomes***   * Illustrate a set of data using simple graphical representation techniques in MS EXCEL. |

* 1. **READINGS**

There are no readings in this Study Session.

## GET STARTED IN EXCEL AND MAKE HISTOGRAMS

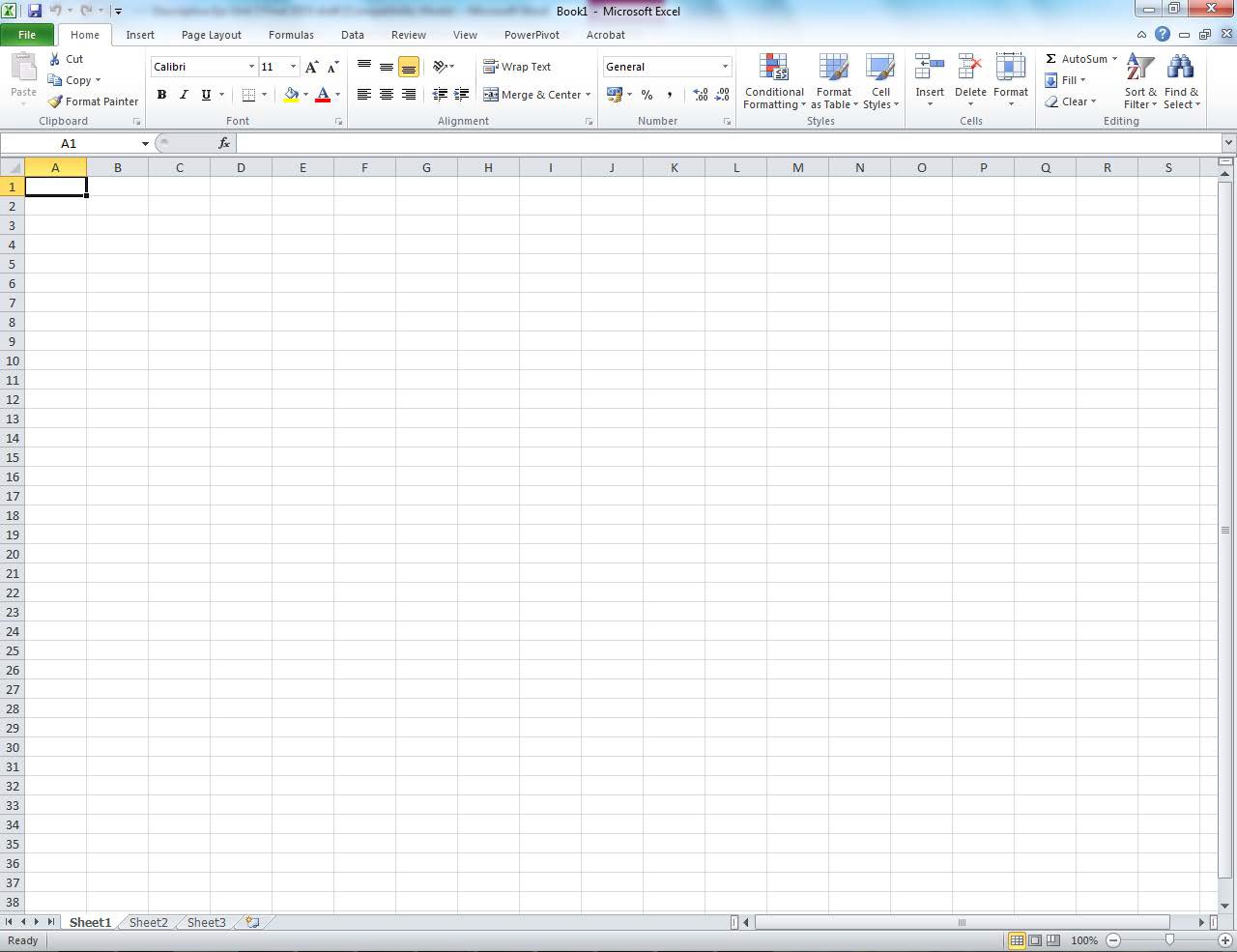
Microsoft EXCEL is a computer software programme that allows you to put information and numbers into spreadsheets. The information and numbers can be sorted or you can apply mathematical formula to them, for example, addition, subtraction, multiplication and percentages. This programme is useful for data management and can also be used to make graphs and other visual representations of the data. EXCEL comes as part of the Microsoft OFFICE package. Other MS Office package components generally include MS Word and MS Powerpoint and sometimes MS Access and other programmes.

For additional information about creating graphs or charts using MS EXCEL 2010 you can also visit: [http://office.microsoft.com/en-za/excel-help/create-a-chart-from-start-to-](http://office.microsoft.com/en-za/excel-help/create-a-chart-from-start-to-finish-HP010342356.aspx) [finish-HP010342356.aspx](http://office.microsoft.com/en-za/excel-help/create-a-chart-from-start-to-finish-HP010342356.aspx)

You should be able to find a green "Microsoft Office EXCEL” icon on your computer desktop. The exact location will vary on each computer. Three common locations are on the desktop, on the lower tool bar and in the START/Programs/Microsoft Office menu.

Once you have located the EXCEL Icon click (or double-click) on the icon to open EXCEL.

You will see the following screen



This is called a spreadsheet with columns labelled alphabetically across the top and rows labelled numerically across the bottom.

For this exercise we are going to use data that has already been summarised. In most cases, you will insert the summarised data into the EXCEL spreadsheet and then use this data to create the graph or figure you require. For some figures however, such as a scatter plot, you will need to insert the original raw data in order to create the graph. You will see how this is done later in this Study Session.

For these exercises, we are going to work with the data from Unit 3 Study Session 2 on the sugar consumption in children. We will attempt to recreate the graphs you did by hand using EXCEL. Start by making a figure for the frequency distribution table of the age data.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Age** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** |
| **Number** | 1 | 6 | 7 | 7 | 7 | 6 | 7 | 5 | 2 | 1 |

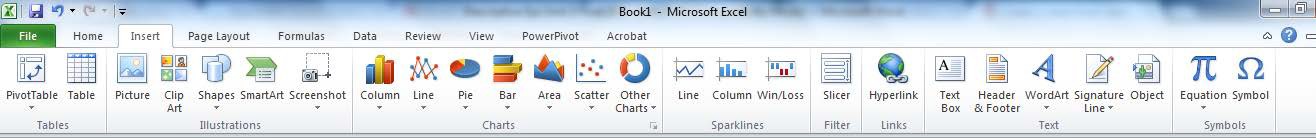
First you will have to enter (type) this summary data into the EXCEL spreadsheet and then highlight the data using your mouse as shown below.

A helpful TIP: Note that I copied the frequency table above except that I changed age from a single numeric to “5yrs” this is because if you leave it as a number EXCEL will see it as data instead of a data label. Your labels need to be text so have at least one alphabetical character or symbol included with any numbers.

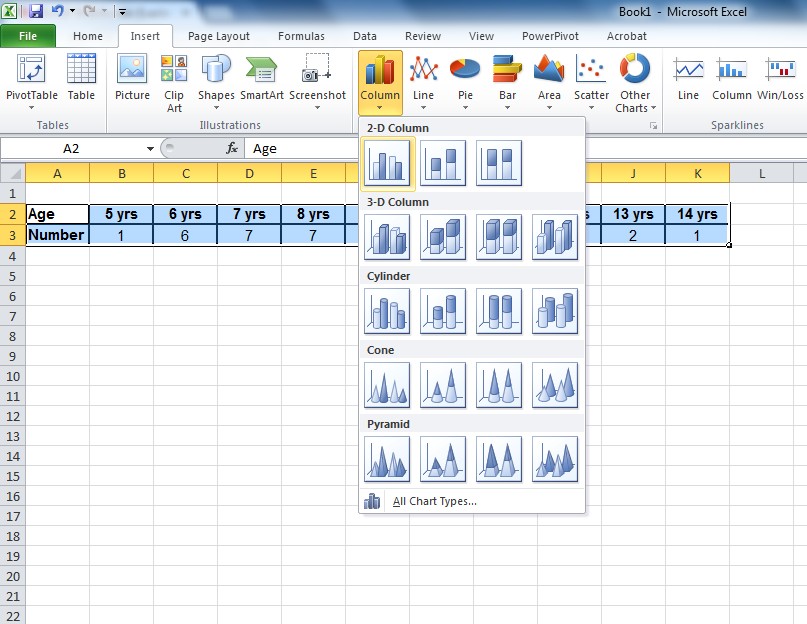
We are going to first create a Histogram or Bar Graph with columns (instead of horizontal bars).

#### Note: Throughout this Study Session the following terms are used interchangeably: Graph, Figure and Chart.

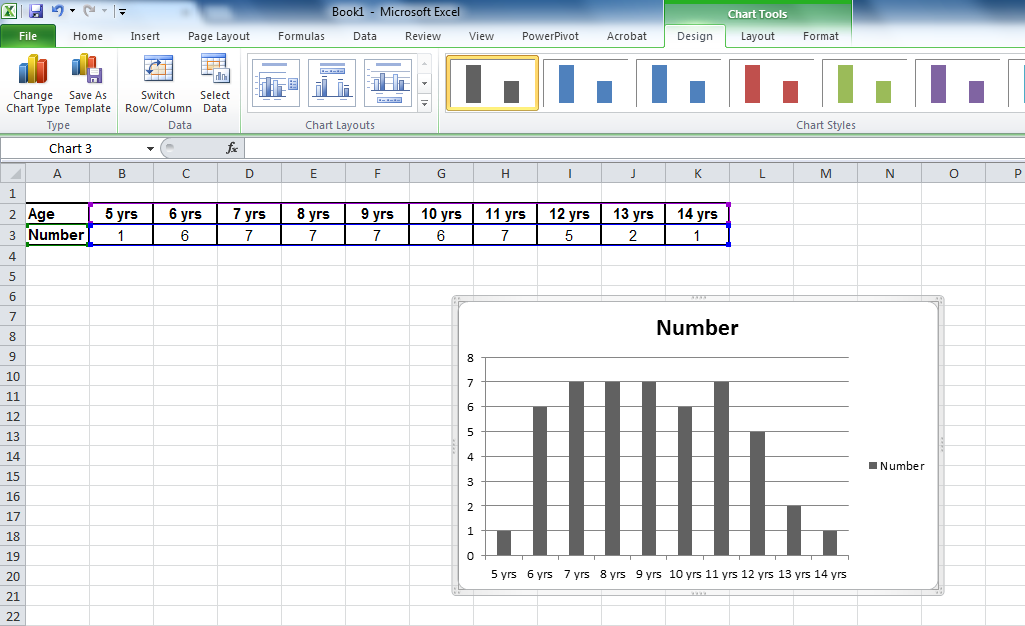
Then click on “Insert” tab on the tool bar at the top of the screen (also called a ribbon) as shown below. This brings up the options below:



You will see that note this ribbon has a chart section from which you can choose the most commonly used charts. So the first thing you need to do is to select the data you want to tell EXCEL what type of chart you want. For this 1st exercise we are going to create a Histogram (Column chart), select the data as below and then you will need to select the type of column chart you would like which is already highlighted so you can now select it:



Look at the chart you have created. Is this what you were hoping for your chart to look like? If not then you might need to reorganise your data in the spreadsheet and start again.



If it looks ok you will see that above the instruction ribbon is a set of tabs under the heading “Chart tools”. You should select the “layout” tab this has several options where you can customise your chart to add labels and other features.

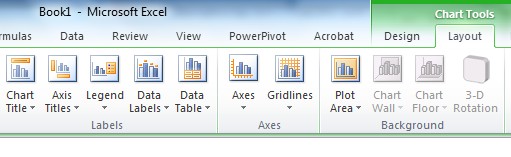
There are 2 primary sections in this tab: Labels and Axes Under labels there are 5 possible options:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chart Titles | Axis Titles | Legend | Data Labels | Data Table |

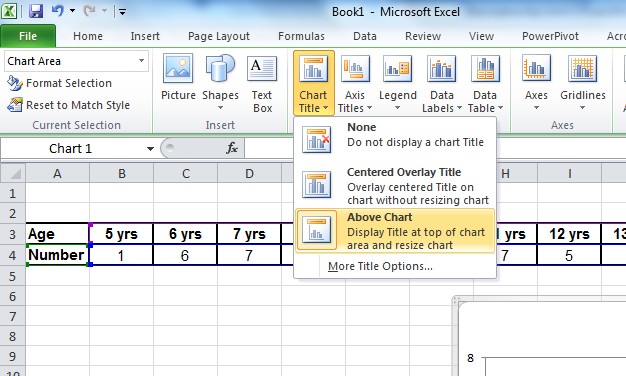
Under Axes there are 2 possible options:

|  |  |
| --- | --- |
| Axes | Gridlines |

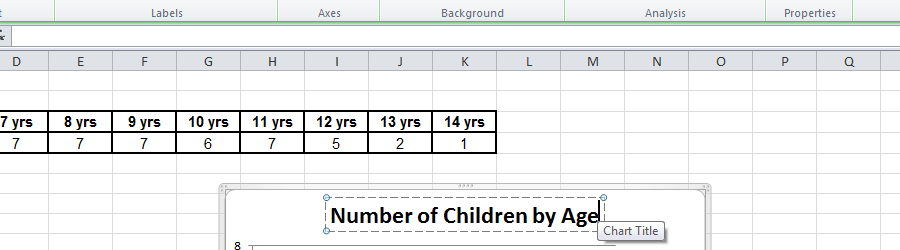
This is what you will see in the ribbon:



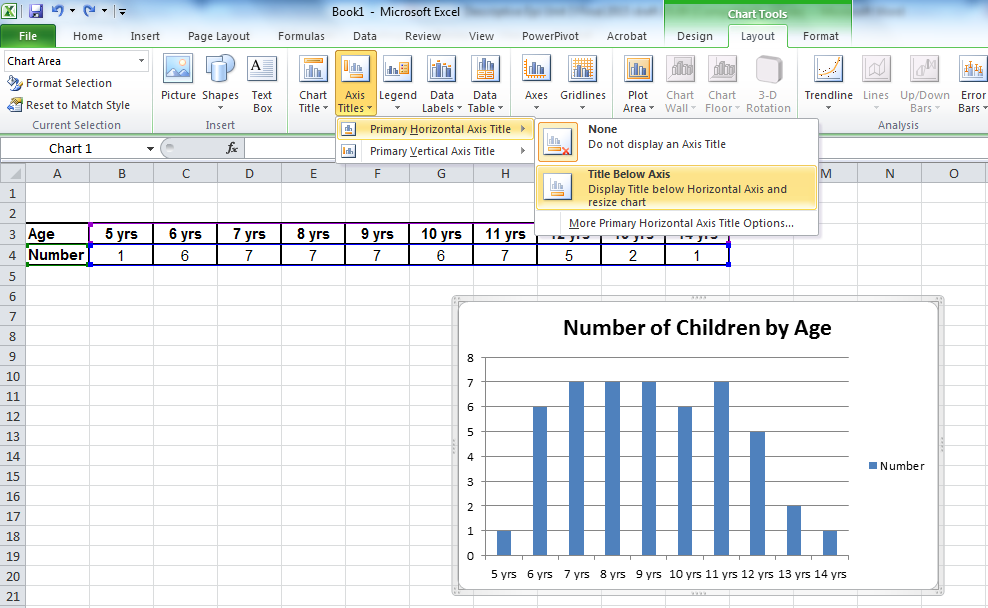
Let us look at several of these features. First two are ‘Titles’. Use the ‘Chart Titles’ option to add the title of the graph.

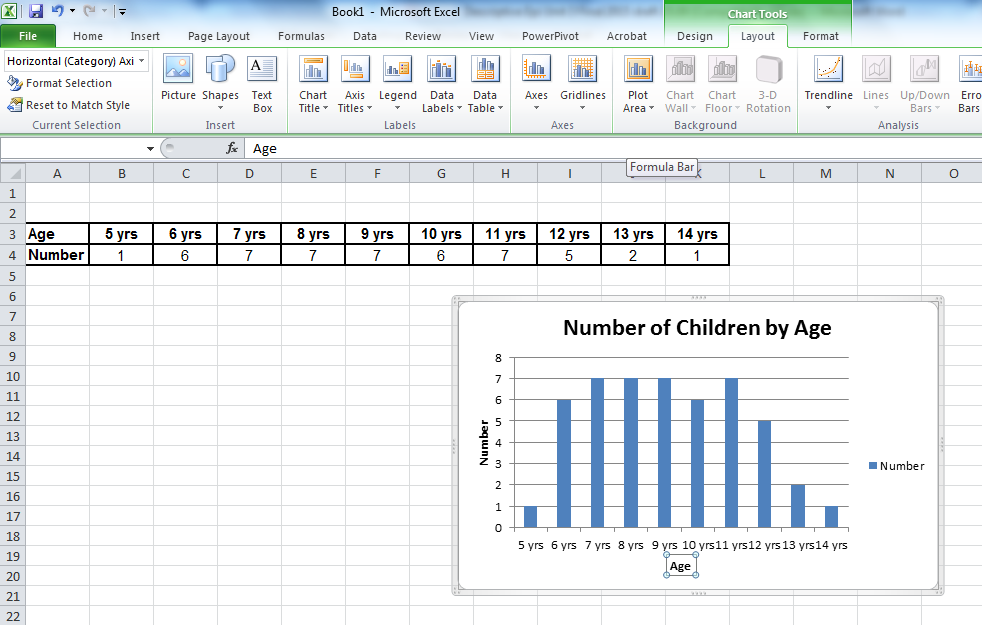


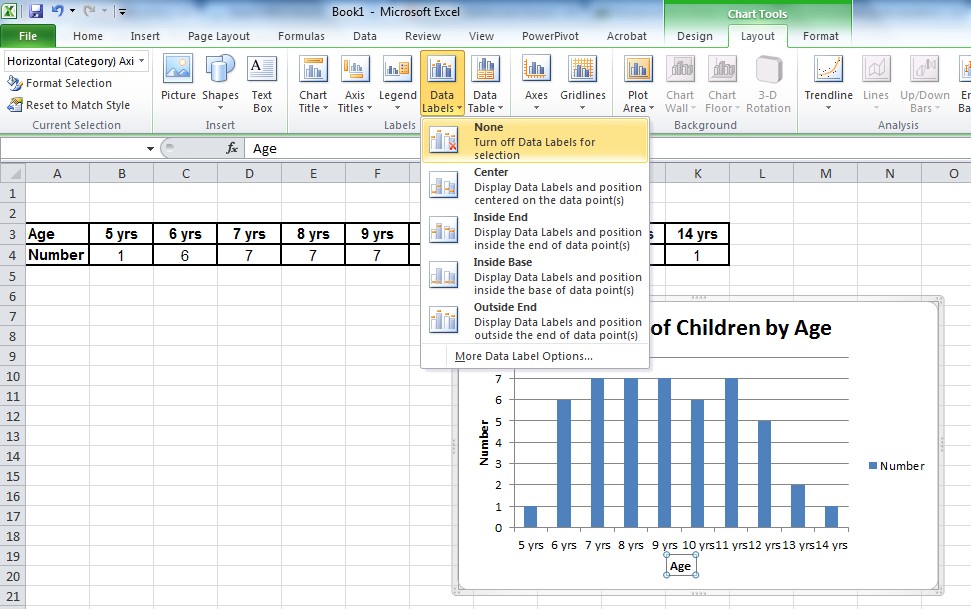
You can then replace the text in the title box with the correct label for the chart. As below:



You can use the ‘Axis Titles’ to add labels both for the horizontal and vertical axes like this:

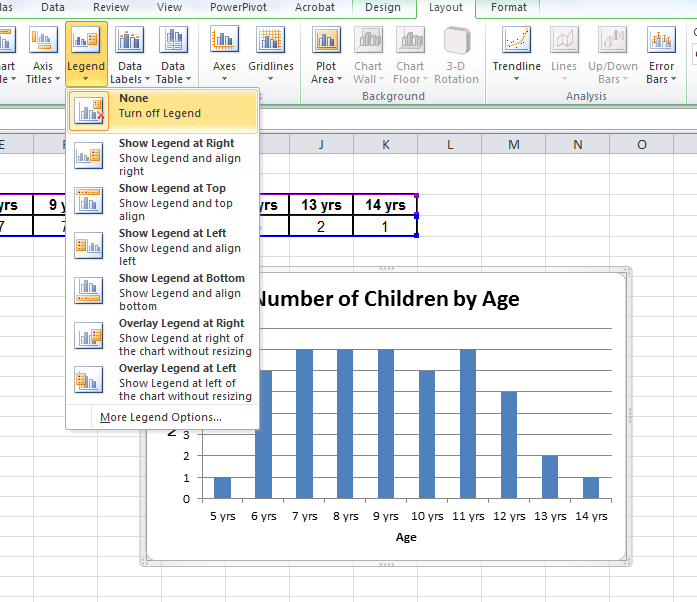






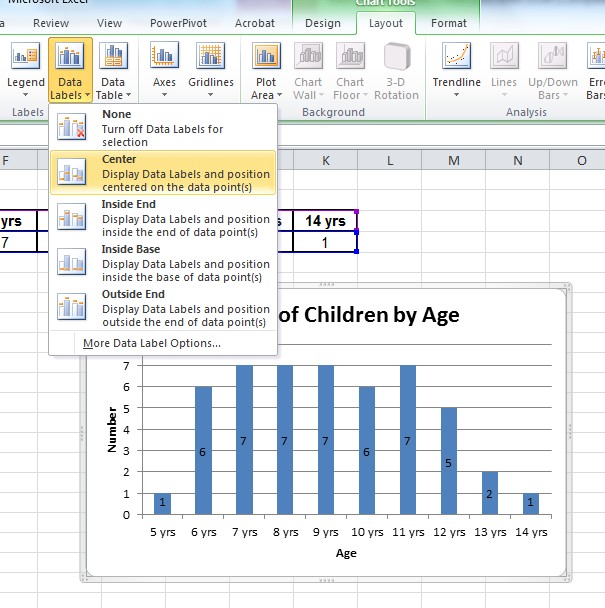
Proper labelling of figures is very important for assisting the reader to understand what they are looking at. Proper labelling of any figures and graphs in your Assignment 2 report will be part of your assignment mark.

Next is the ‘Legend’ tab. The legend is the small box on the right which says “Number” with a blue box. Since there is only one item in the x-axis you do not need the legend in this graph, if you have multiple items in the x-axis then you could label these using this option. Select ‘None’ to turn off and remove the legend for this graph.

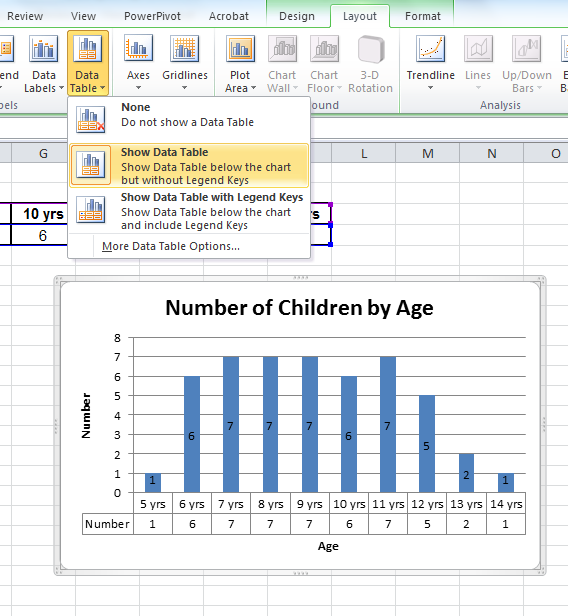


We will see in a later exercise a situation where you will want to keep the legend.

Now click on the “Data Labels” Tab. Here you can select various types of labels for the data. Click on the various options to display information about the value of each bar in a range of positions. See the inside bar example below:



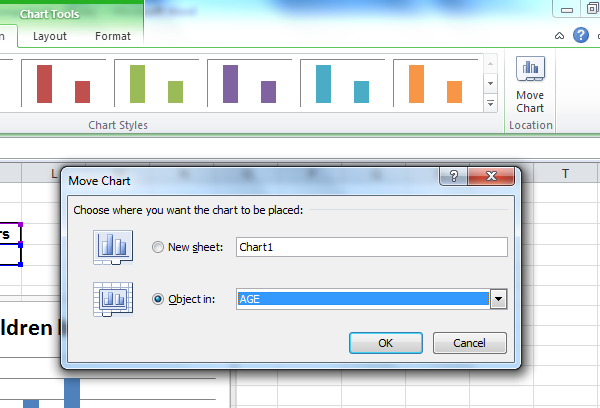
Now click on the “Data Table” tab.



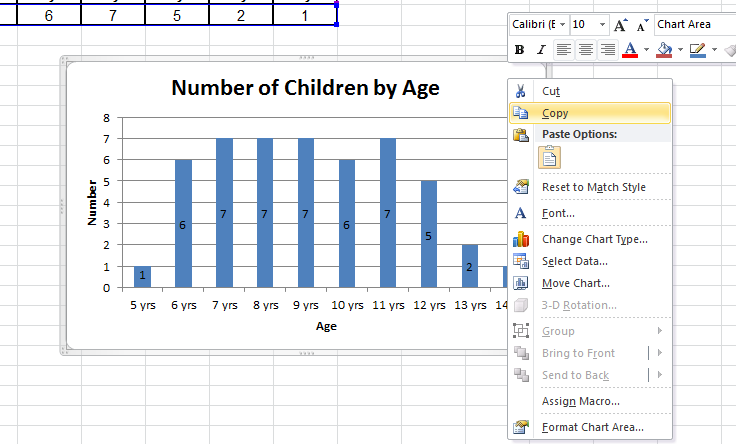
As with the value labels, you can also add the data table to your chart, the presents the value of the bar and the label of the bar in the form of a table below the bars.

In most cases you will not need both the value labels AND data table. You can pick one. For this exercise let’s keep the value labels, so click on the “Data table” tab and select ‘None’ and see the data table disappear and go back to the previous graph with the value labels only.

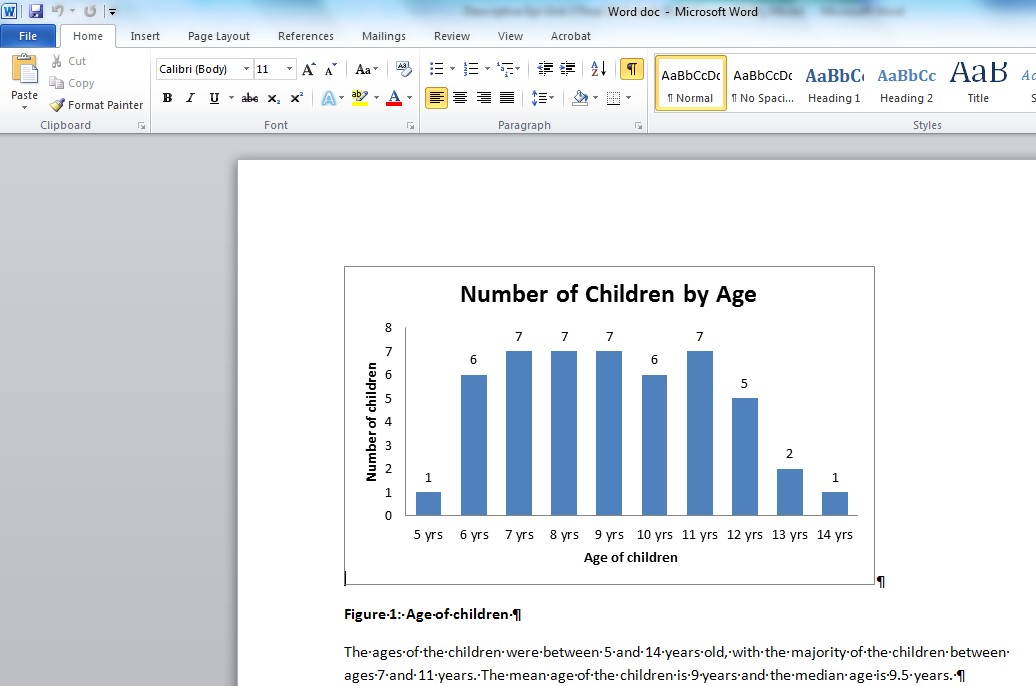
In the top level of the ribbon you can select the design tab, within this tab the ‘Move chart’ icon is situated on the far right hand side. Select this to change the location of your chart. You can choose to have the chart inserted in the current spreadsheet (as ticked above) or you can have it as a separate spreadsheet (the other tick box). I usually use the insert as object in the current spreadsheet. This is the default so go ahead and click “OK”.



Now that we have the graph we will want to insert it into our report. You can do this by first copying the graph. Click anywhere on the graph to highlight it (you can see the little black corners). Then click on “Edit Copy” as shown below:



Then you just open your word processor with your report and paste the figure into the place where you want it. You should then label the figure. Note that for figures/graphs the label goes below the figure. This is different from data tables where the label/title goes above the table. Then you should summarise in text what the figure tells us about the data. See the example below. Also as part of text you can include other calculated data such shown here, e.g. mean and median.



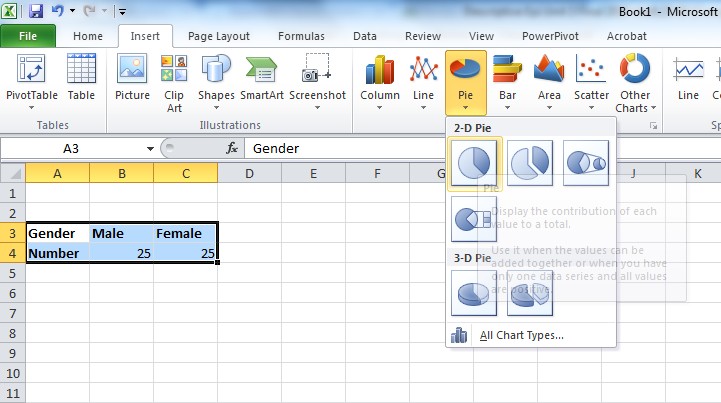
## MAKE A PIE CHART

Now let us try a pie chart. Pie charts are generally good when you have between two to six categories. More than six categories makes a pie chart look confusing as it starts to look like a dart board which is not a helpful presentation of the data.

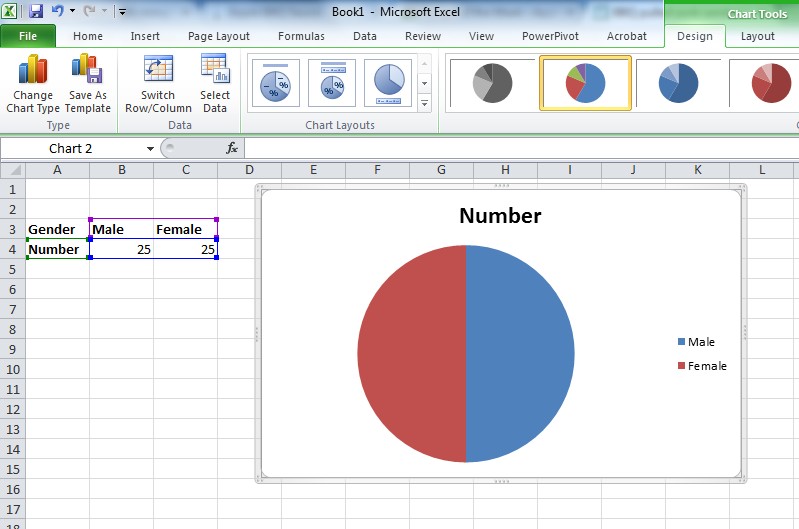
For the pie chart let’s, use child gender which has only two categories. You will recall the gender data from Unit 2: 25 males and 25 females.

To create a new chart you can either open another EXCEL file to enter data or you can actually just move to use Sheet 2 which I have done as shown below. To get to sheet 2 just click on the Sheet 2 tab at the bottom and then a blank spreadsheet will appear.

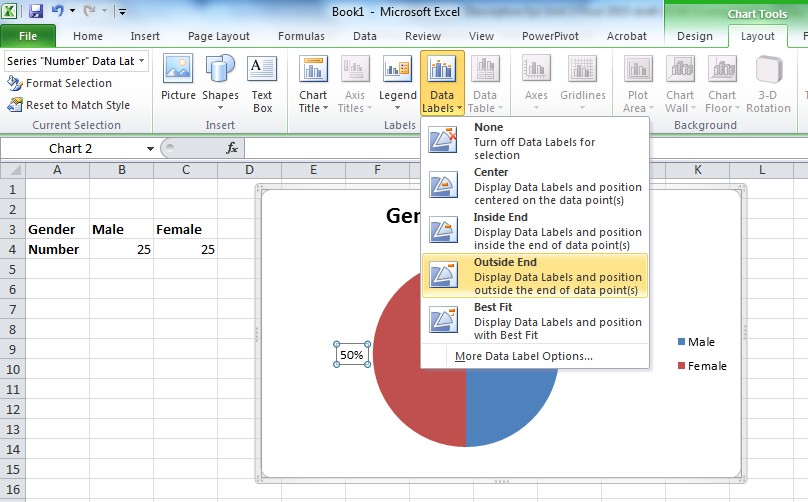
From here enter your summary data, highlight it and then click on “Insert/Chart” as shown previously.



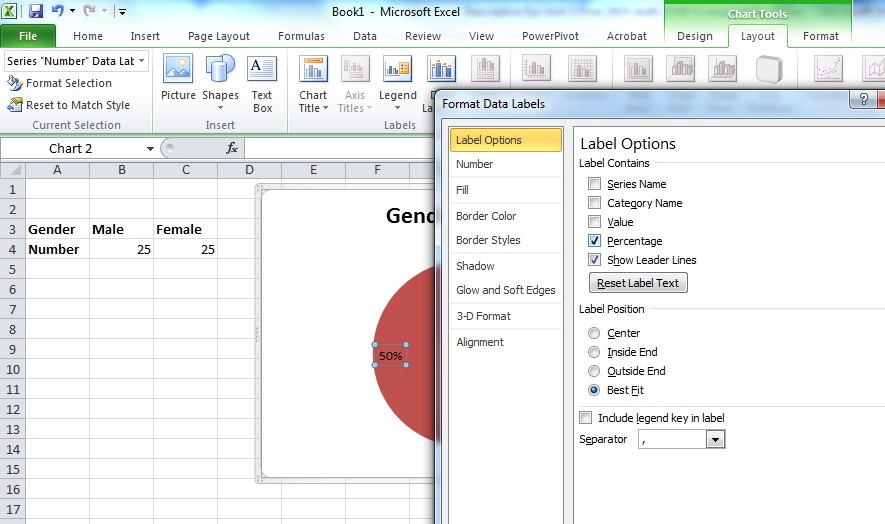
Check to see if you are happy with the way the pie chart looks:



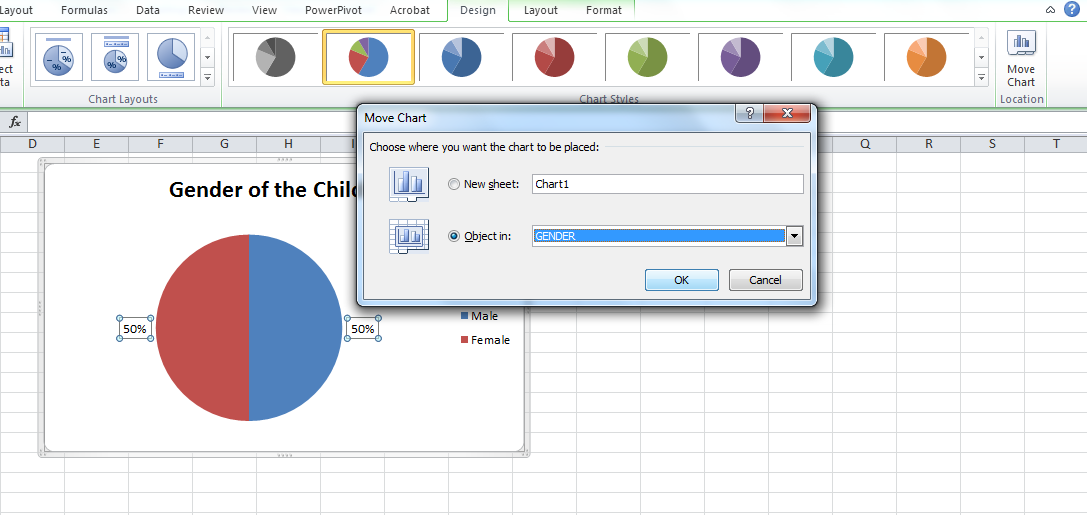
Now you can once again insert Labels as above. For the Pie Chart you will need to keep the Legend as that indicates what the different colours mean. Finally you may want to add data labels, so click on the “Data Labels” tag.



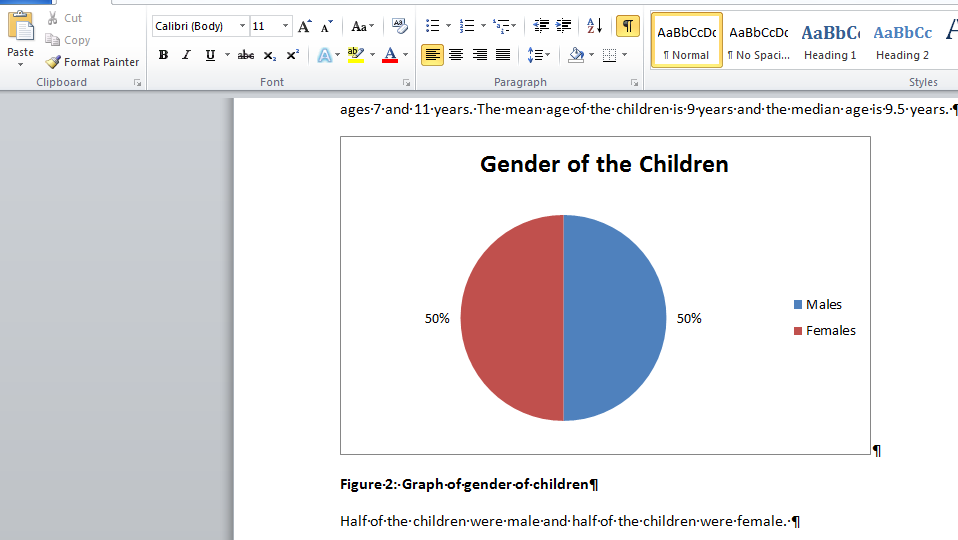
For the data labels you can either choose “Value” or “Percentage”. Do this by clicking on the ‘More Data Label Options’ and then click on the “Percentage” box to select percentage, which then appears in the chart as show above.



As we did before, we will insert the chart into the current Sheet, which is Sheet 2 or ‘GENDER’ as I have called it as shown below.



From here we can copy and paste the chart into our word processing programme and add a label and text explaining the chart, just as we did before.



## MAKE HISTOGRAMS TO COMPARE GROUPS

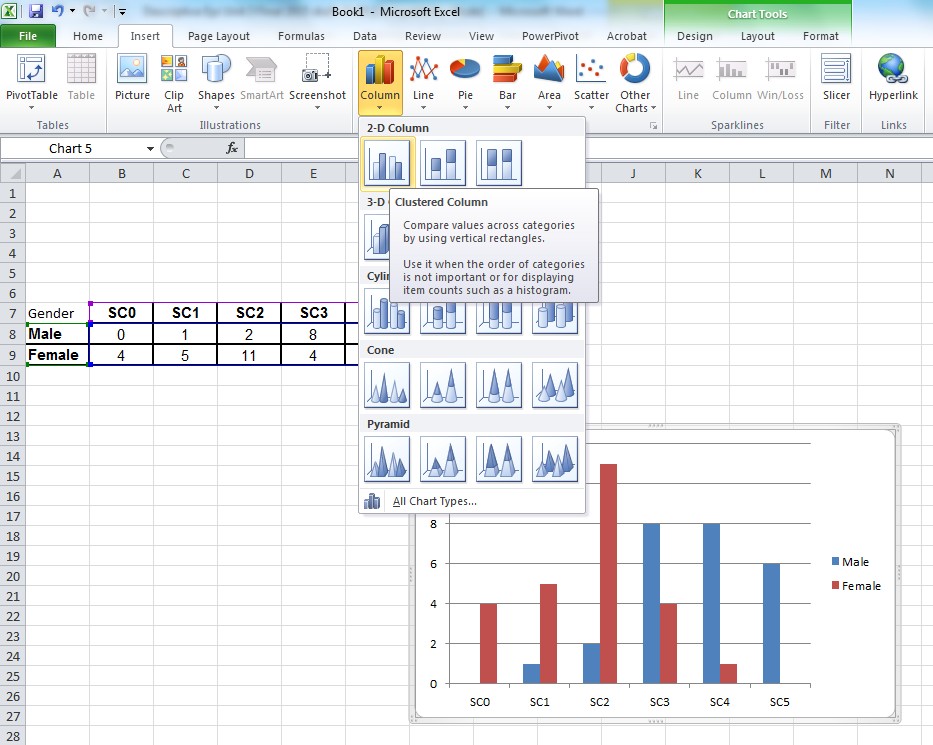
We can also use graphs and figures to compare results across various groups we have studied. For example, do boys and girls have different sugar consumption scores?

Once again open a new sheet by clicking on ‘Sheet 3’ at the bottom of the screen.

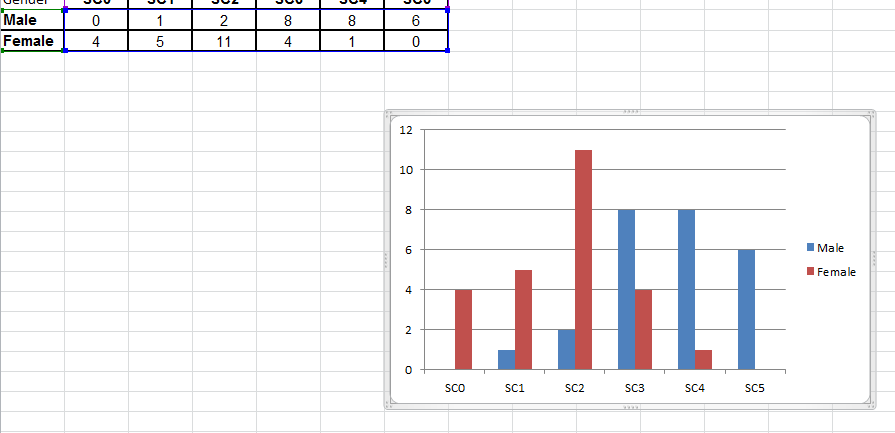
Looking back at Unit 2 we can insert the summary data on gender and sugar consumption from the data table:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Gender** | **Sugar consumption score** | | | | | | |
| **0** | **1** | **2** | **3** | **4** | **5** | **Total** |
| **Male** | 0 | 1 | 2 | 8 | 8 | 6 | 25 |
| **Female** | 4 | 5 | 11 | 4 | 1 | 0 | 25 |
| **Total** | 4 | 6 | 13 | 12 | 9 | 6 | 50 |

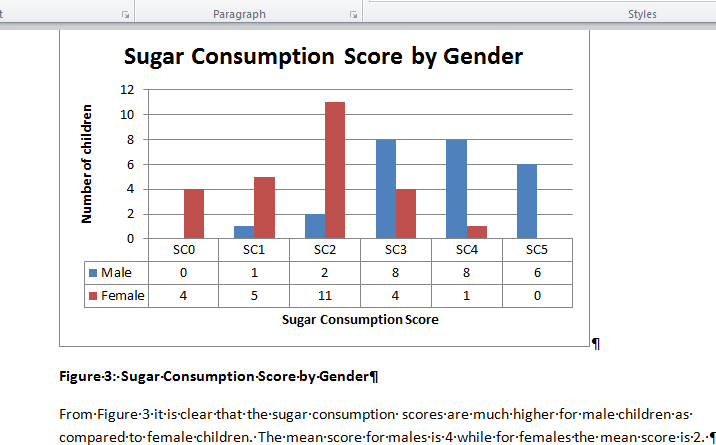
Type in the data as shown above. Select “Column” to get a Histogram. Here because you now have 2 rows of data you will get two bars (columns) for each value of sugar consumption score.



Note how now we have two colours instead of one and the legend indicates which colour is which group. As with the previous examples work through the steps to add labels and the data table and finish by inserting the above graph into your spreadsheet.



Note how this looks like the graph in Unit 2 for this data. Now copy/paste this figure into your report, and add the title and text.



Note how the interpretation of the chart can be expressed using words to describe what the figure is showing the reader.

**TASK 1 – Interpret the graph**

1. What is your interpretation of this graph?

#### FEEDBACK

In this case the graph suggests that boys consume much more sugar than girls. This is supported by also adding the mean consumption for each gender in the text as well.

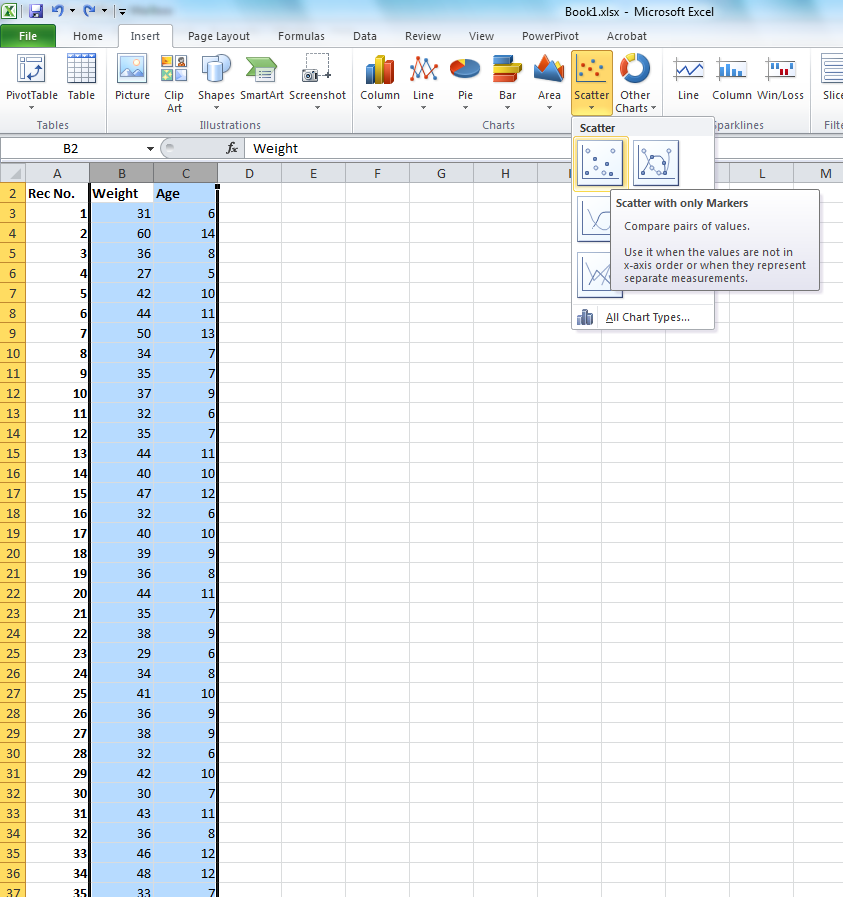
## MAKE SCATTER PLOTS TO COMPARE GROUPS

All of the above charts are used for categorical data. Graphs can also be used to compare continuous data. A common graph to compare continuous data is the scatter plot which was shown at the end of Unit 2. We can now make that same graph using EXCEL.

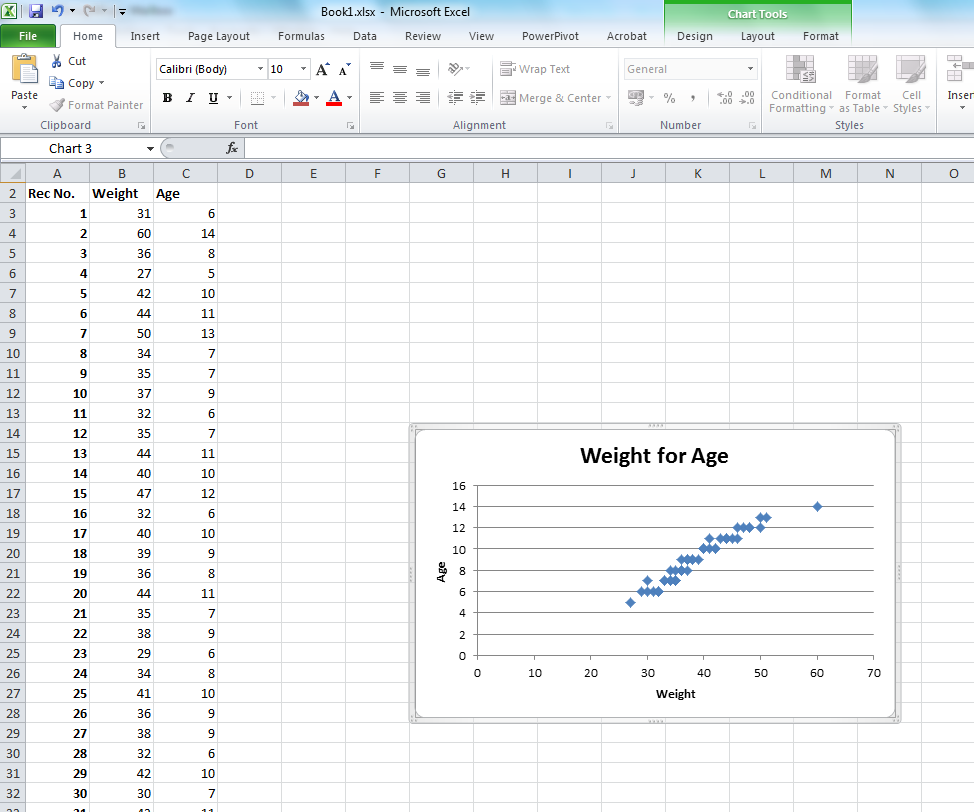
First however you will note that we are out of sheets in our EXCEL file as there are only 3 and we have used them all. So first let’s learn how to add another sheet.

The big difference with this type of continuous data is that we need to insert the raw not the summarized data. This can be a problem in a big data set, but you can often export the data to EXCEL and not have to type it. That is beyond the scope of this module.

For now we will type the data into the new blank sheet, then highlight the data and click “Insert” and then select ‘Scatter’ and the option for ‘Scatter with only Markers’ as we have done below:



Now insert the title and labels as we have done before using the options from the ‘Layout’ tab in the ribbon.



In this type of graph, once again you do not need the legend so click on the ‘Legend’ tab and then click on “Show Legend” to un-tick that feature and the legend will be deleted. You can complete the graph by going to the ‘Design’ tab and selecting ‘Move’ and opting to insert it into the sheet as we have done before above.

**TASK 2 – Analyse a graph**

1. What does this graph show you?

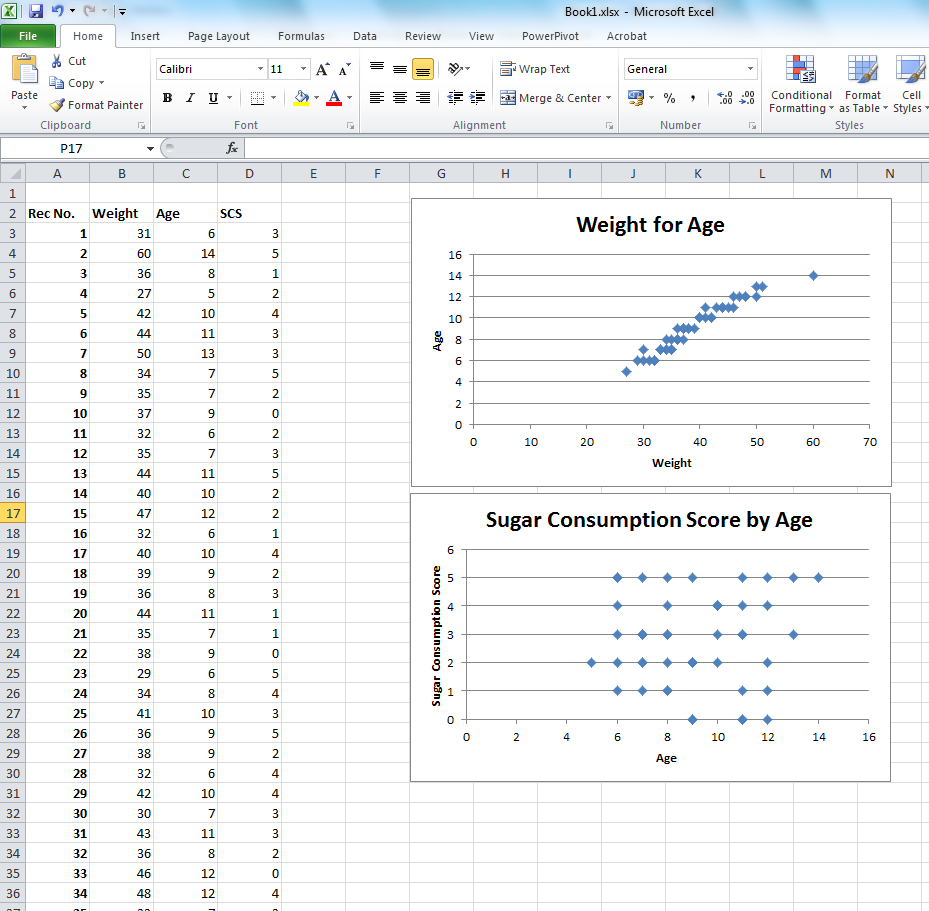
#### FEEDBACK

As age increases weight also increases. This of course is no surprise and you should note this in your text explaining this graph as shown below.

After this, copy/paste the graph into your report, add the title and text.



Now let us examine age and sugar consumption. First let’s add another column of data, i.e. sugar consumption score to the spreadsheet. Then you can highlight age and sugar columns and insert another scatter plot. Now you have two scatter plots.



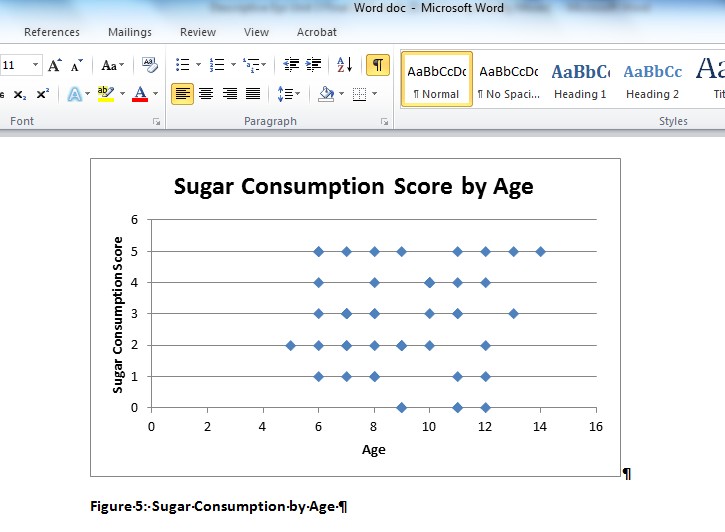
**TASK 3 – Interpret and compare scatter plots**

1. What does the age and sugar consumption scatter plot above suggest?
2. What do you see if you compare the two scatter plots below?

#### FEEDBACK

* + 1. We can see in the age and sugar consumption spreadsheet that there is no clear linear pattern up or down – the data is sort of a box or random. This suggests there is not relationship between age and sugar consumption.
    2. Can you see the difference and understand the different interpretations of the two scatter plots? One clearly shows the data rising, while the other appears quite random.

This can be reflected in your report as show below:



## SESSION SUMMARY

In this Study Session we learned how to create several types of figures and graphs using Microsoft EXCEL. We also reinforced our learning from Unit 3 Study Session 2 on interpreting graphs and figures.

The next Study Session will synthesize the learning from this module. It will start with determining how to ask questions of data that you want to report and interpret and then will provide instructions for your summative assignment which can be found in Study Session 5.

This assignment will take the form of an epidemiologic report on a set of summarised data which will be provided to you separately. You will be asked to review the literature on that topic, summarise and represent the data as part of the report and discuss and make public health recommendations based on that data. You will need the skills from this EXCEL session to improve your representation of the data.



You have covered common ways of representing health information. Try the following Self-evaluation task.

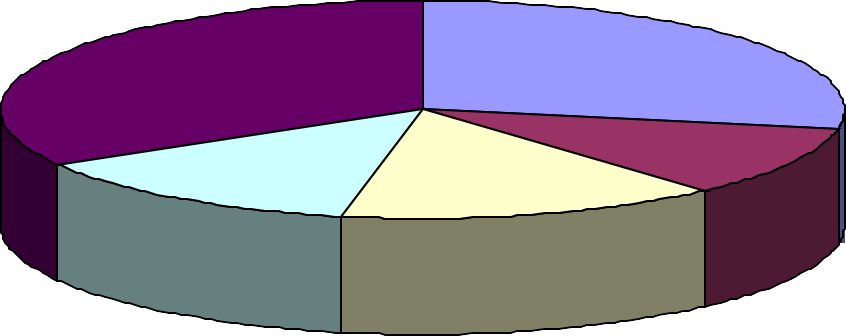
**KEY POINTS IN UNIT 3 SESSION 3**

**Representing Health Information**

**Self-evaluation:** Be aware that some graphic tools can be misused.

Please comment on the graph below, representing the percentages of pregnant women who have been tested for HIV in 5 districts.

**Figure A: The percentages of pregnant women who have been tested for HIV in 5 districts**



78

65

A B C D

E

25

32

34

# Unit 3 - Study Session 4 Develop Critical Questions for your Dataset and Prepare for Assignment 2

## Introduction

In this Study Session you are asked to consider the Public Health problems represented in the dataset provided for Assignment 2. You will find the dataset in the Introduction to this module.

The first tasks guide you through the process of developing a set of relevant study questions, aims and objectives. These critical questions will enable you to proceed with a systematic data analysis that will form the backbone of the report required for Assignment 2. The next tasks involve preparing the structure of your report and organising your information.

Here is an extract from Assignment 2 as a reminder.

**ASSIGNMENT 2 REMINDER - DEVELOP CRITICAL QUESTIONS FOR YOUR DATASET**

Draw on your literature review (from the *Introduction to Public Health* module) and your knowledge of epidemiology to ask critical questions of the dataset. List the most important questions you wish to ask of the dataset. Use these questions and appropriate keywords to prepare subheadings for your report. Organise the material into a logical structure for an outline of the report.

## Contents

1. Learning outcomes of this session
2. Readings
3. Identify questions to resolve the problem
4. How your study questions will guide your report
5. Plan your analysis
6. Session summary

## Timing of this session

There are four tasks in this Study Session, all of which are designed to help you complete part of the assignment.

## LEARNING OUTCOMES OF THE STUDY SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Define a clear set of study questions with which to assess a set of health data. * Develop a structure for your report. | ***Academic Learning Outcomes***   * Develop critical questioning skills for analysing a set of data. * Prepare a structured outline for a report. * Organise information in a logical way. |

1. **READING**

Refer to the readings provided in Study Sessions 1 – 2.

|  |  |
| --- | --- |
| Vaughan, J. P. & Morrow, R. H. | (1989). Ch 12 – Communicating Health Information. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 125–129. |

## IDENTIFY QUESTIONS TO RESOLVE THE PROBLEM

This section takes the process started in Assignment 1 one step further. As with any research task, it is essential that you begin with understanding the nature of the problem. In order to do this, you need to identify relevant questions that will assist you in analysing the problem.

Once again refer to Unit 1, Study Session 2 to assist you in asking relevant questions.

**TASK 1 - Identify questions you need to answer to resolve this problem**

Having defined the problem represented by the data, answer the following questions:

1. What do I need to do about it?
2. What do I need to know in order to respond adequately and with confidence to the problem at hand?

The answers to these questions should help you to structure the different parts of your assignment and even provide a format for your report.

Your list of questions will be part of the draft of Assignment 2.

#### FEEDBACK

Once you have developed your questions, ask yourself whether they will lead to answering the key question posed by the original problem, i.e. What can be done to fix the Public Health problem which you are facing?

Generally, you will identify one or more broad Public Health problems related to your data. These can be restated to be the AIMS of your study. For example the problem might be an increase in HIV/AIDS incidence in your district and you need to understand why this is happening so you can develop programmes to address this increase. This could be restated as the following study aim:

*This study aims to understand the causes of the increase in the incidence of HIV/AIDS in District X and to develop programmes to reduce any further increase.*

However, this is a very general statement, which may not provide enough guidance for you to plan your study and your data analysis. Therefore you will also want to develop a specific list of study questions that you will attempt to answer. These are also referred to as the Study Objectives. Refer again to Unit 1, Study Session 2 and the conceptual framework for identifying a health status problem. Here are examples of some specific study objectives/questions that might be asked:

The objectives of this study are:

* To identify the change in HIV/AIDS incidence in the district over the last two years.
* To identify who is affected and which population sub-groups show the highest increase.
* To identify the likely causes of this increase, e.g. unsafe sexual practices, in- migration of sick individuals, intravenous drug use, lack of knowledge of HIV/AIDS in the population, and so on.
* To identify possible programmes to address HIV/AIDS and make recommendations to the Provincial Government for action to address HIV/AIDS.

Note that the last objective is not generally one that will be obtained from your study data, but will likely come from your literature review, which should include a review of successful programmes discussed in the literature. This last objective is crucial as it will be the component of your report that motivates actions based on your analysis of the data.

**TASK 2 – Develop the aims and objectives of your study**

1. Write a clear statement of the aim/s of your study.
2. List of the objectives of your study. You should choose no more than five study objectives for this assignment.

#### FEEDBACK

Refer to the examples given in the feedback section above.

## HOW YOUR STUDY QUESTIONS WILL GUIDE YOUR REPORT

Now that you have developed your study aims and objectives, you can use them to help you plan your study and your final report.

If you were starting a study from the beginning, the study questions would assist you to develop your study design.

For this assignment, you have already been provided with summary data; therefore you can consider how the questions will guide your plan for analysing and reporting this data.

**TASK 3 – Use the study questions to draft an outline for your report**

1. Now consider whether the set of questions you have developed suggests a structure or a set of sub-headings for your final report.

Use Chapter 12 of Vaughan & Morrow (1989) for support.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 12 – Communicating Health Information. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 125–129.

#### FEEDBACK

Imagine that you have identified three key study questions. You will want to specifically state these questions in the *Aims and Objectives* section of your report. You will then want to use your data to answer these specific study questions. You need to make sure each study question has corresponding data to answer that question, and that data which is not related to any of your study questions is not included in your report.

In addition, you will want to make sure you have information in the literature review relevant to each of these questions. Finally, your discussion section will focus on an interpretation of the data pertaining to these questions, and your conclusion will provide summary answers to these questions. A potential outline of the structure of your report is presented below:

#### A Sample Outline of a Report Structure

1. **Introduction**

Describe the overall study problem

#### Literature Review\*

* 1. Literature related to overall study problem prevalence/incidence
  2. Literature related to Study Question 1
  3. Literature related to Study Question 2
  4. Literature related to Study Question 3

*\**Note that Mouton (2001) and Depoy & Gitlin (1994) describe this as organising your literature review around “Themes”.

#### Aims and Objectives

* 1. Overall Aim
  2. Objectives
     1. Study Question 1
     2. Study Question 2
     3. Study Question 3

#### Study Methods

1. **Results**
   1. Results related to Study Objective 1
   2. Results related to Study Objective 2
   3. Results related to Study Objective 3

#### Discussion

Discuss the results and compare to them to the literature. Review other programmes which have addressed this problem, from the literature or other reports.

* 1. Discussion of data related to Study Objective 1
  2. Discussion of data related to Study Objective 2
  3. Discussion of data related to Study Objective 3

#### Conclusion

* 1. Summary short answer to Study Objective 1
  2. Summary short answer to Study Objective 2
  3. Summary short answer to Study Objective 3
  4. Summary short answer to overall Study Aim

#### Recommendations

Recommendations may address more than one study question and the overall aim, so this section will be organised in terms of the type of programme rather than the study questions.

The above is a very crude outline. For your outline, you would want to use actual key words, topics or themes derived from your study questions / objectives.

## PLAN YOUR ANALYSIS

From the task above, you can see that everything in your report is focused around your study questions or objectives.

Therefore, it is a good idea to keep your study questions in mind as you plan all aspects of your study. This will save you time and will help as you plan how to do your data analysis.

**TASK 4 – Plan your analysis**

You should not approach your analysis and interpretation of your data without a clear path for what you need to do. This path is also defined by your study questions. Based on the study questions/objectives you have developed, on the contents of your literature review and the knowledge you have acquired about epidemiological analysis in this module, carry out these tasks:

1. Check whether any of the literature you have gathered sets out to answer similar study questions. How do they do it? What methods did they use?
2. Apply your questions to the data and make sure that they can be answered.
3. Scrutinise the data and describe how you will carry out your analysis as a series of steps.

#### FEEDBACK

1. What may be helpful here is to look at your literature review references to see how those authors approached their data, and whether there are any similarities between their data and your dataset. This may suggest ways of performing your analysis. It is often, but not always, a good idea to use a similar ways to present your data, so it can be easily compared with the literature.

If you look back to the HIV example introduced earlier, a possible study question is: Study Objective 2: *To identify who is affected or what population sub-groups are showing the highest increase.*

You could start by unpacking this question in relation to your literature:

*First we need to define what we mean by "affected". In this case, we mean people who are HIV+.*

Now we need to clarify what we mean by "sub-groups".

The literature may have suggested that the sub-groups of people affected by HIV/AIDS are often defined according to gender, age, socio-economic status, region where they live, race/ethnicity, etc. We might also have seen these defined as follows from the literature:

*Sex = Male or Female*

*Age = Children 0–12; Teens 13–19; Young Adults 20–35; Older Adults 36+ Socio-economic status = Monthly income*

These are just a few examples, from earlier Study Sessions. You will realise that there are lots of ways to categorise or describe populations. However, through the literature, you should have come to understand which categorisations are most relevant to your problem, in this example HIV/AIDS.

At this point you could refer back to Study Sessions in this module on defining your study questions.

1. You should also make sure that you have linked all the data in your assignment with its corresponding study question or questions. Remember that some data may apply to more than one question.

Consider:

* + Do you have data to answer every question?
  + Does each piece of data relate to a study question?

For example, from the literature, you know that age is a very important grouping for looking at HIV/AIDS. Therefore, if you want to use age as one of your sub- groups, you will need data on the subjects’ age; otherwise you cannot answer this question. While this seems simple, when you start to deal with lots of study questions and lots of data, this kind of simple connection of "question to data" is often overlooked.

In relation to your dataset in the assignment, you need to make sure that you have all the data necessary to answer the study questions you have decided on in the previous task. Check this by linking each piece of study data to a study question. Look to see if you have any questions without data or any data without questions. If you find this, then you may need to revise your study questions.

Finally we have to think about how we would use the data to answer the question. For example, we have defined "affected" as being HIV+. Therefore, if we want to see if certain age sub-groups are *more* affected (that is have more HIV than other age groups), we would have to compare the rate of HIV in each of the sub- groups. Refer back to Unit 2, Study Session 1 for information on "rates". You might plan to look at the data as follows:

#### Age Group HIV+ Rate

Children 0 – 12yrs …………

Teens 13 – 19yrs ………… Young Adults 20 – 35yrs ………… Older Adults 36+yrs …………

Note that I have not filled in the actual HIV+ rates. This is because I will get this from my study data.

1. You will learn more about how to do this in the next unit. At this stage, however, you have an idea of how you might want to use your data to answer one of the study questions in the HIV/AIDS example.

You are now ready to move on to Session 5, where you will be guided to compile your final report.

## SESSION SUMMARY

In this Study Session, you have been guided towards asking questions of your dataset and preparing a structure for your draft report.

In the next Study Session, you will be guided towards compiling a thorough and useful report for the Assignment.

### COMMON KEY POINTS IN UNIT 3 SESSIONS 4 & 5



**Develop Critical Questions and Write a Report**

To develop critical questions for research, one follows the steps below:

* identification of a health-related problem
* decision on the action to be taken for this problem
* selection of what one needs to know to be able to take this action, leading to a set of critical questions to be used to collect data and analyse them (using the conceptual framework introduced to you in Unit 1 Session 2).

These critical questions have to be kept in mind, for the whole duration of the analysis and writing the report, to ensure consistency and to identify the action needed to tackle this problem.

When writing a report, it is important to synthesise information (as readers will not be prepared to read a 50 page report) and to be very systematic and scientific in the presentation of the report.

A mistake commonly made by students is to mix results and discussion. The results section should only consist of description of data from your own dataset, in the text or using tables and graphs. Comparison can also be made between data from the dataset.

The discussion section should include comments, comparison with data in the literature or data from other sources. No result can be discussed in this section unless already presented in the Results section.

**Self-evaluation question:** Below are selected “Results” sections written by previous students, reporting on a dataset of tuberculosis indicators in five districts. Try to comment critically on them:

Example 1: … On analysing analysis, the data shows evidence of high TB incidence in the province of 555.8 per 100 000 population … It is clear that the province has a high incidence of TB across all the districts. Three out of six districts reflect an incidence that is higher that of the African continent incidence at 361 per 100 000 population (WHO, Global TB Report 2009).

District D seems to have enormous TB burden with an incidence of 931.5 per 100 000.

Example 2: … The smear conversion rate is used to monitor treatment success after the intensive phase (7-8 weeks after starting the treatment) specifically for new smear positive patients. Failure to convert from smear positive to smear negative is suggestive of a potential Multi-drug Resistant or Extremely Resistant strain of TB (MDR/XDR–TB). It should be also noted that poor conversion rates have implication for treatment outcomes, i.e. cure rates, treatment success rate and defaulter rate …

# Unit 3 – Study Session 5 Final Report (Assignment 2)

## Introduction

Even the most perfect technical description and analysis of a health or illness event may amount to nothing unless it is effectively communicated to those who have the responsibility to do something about it. The point is that you need to communicate all the information in a way that is not just accurate, but also clear, short, simple and very persuasive. After all, your main objective should be to generate an effective response to the health problem described by your epidemiological investigation. For this reason, the preparation of a well-constructed report is an essential part of the epidemiological process.

This Study Session guides you through the final stages in preparing Assignment 2. This is to write an epidemiological report on the assignment dataset. Remember that the Assignment is outlined in detail in the Introduction to this module.

The report should reflect your grasp of the epidemiological concepts and processes you have encountered during all the preceding Study Sessions of the module. This Study Session provides a generic format to assist you in preparing a report on an epidemiological event. Recognise that your report needs to convince others in the health system that you have an important and valid point to make. The relevant people need to be persuaded to take the report seriously and to act upon your suggestions. This is the real challenge you face as you prepare your report, and one of the criteria whereby it will be evaluated.

## Contents

1. Learning outcomes of this session
2. Readings
3. Review guidelines for report writing
4. Critique a sample report
5. Assignment 2: Final report
6. Analyse and interpret the data
7. Discuss the implications for district health planning
8. Prepare your report
9. Proofread and edit the final version
10. Session summary

## Timing of the session

The Study Session contains fourteen tasks. All these tasks are designed to help you prepare the final version of your assignment.

## LEARNING OUTCOMES OF THIS SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Establish guidelines for effective epidemiological report writing. * Prepare a reader-friendly and convincing epidemiological report. | ***Academic Learning Outcomes***   * Summarise information. * Use simple graphical techniques. * Critically review and edit a report you have written. |

* 1. **READINGS**

There are two readings in this Study Session.

|  |  |
| --- | --- |
| **Author/s** | **Publication Details** |
| Vaughan, J. P. & Morrow, R. H. | (1989). Ch 12 – Communicating Health Information. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 125–129. |
| Community Agency for Social Enquiry. (CASE) | (1999). Introduction & Health Status. In *The Second Kaiser Family Foundation Survey of Health Care in South Africa*. Johannesburg: CASE: 1–7. |

## REVIEW GUIDELINES FOR REPORT WRITING

Before you construct a report or a research article, it would be useful to identify the main components, or sections of such a report and the function that each part of the report is to serve. The tasks in this Study Session help you to do this.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 12 – Communicating Health Information. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 125–129.

Note that you have previously looked at this chapter and considered the structure of a report.

Now, in the final stages of preparing for your report, you will examine these components more closely and from the perspective of the whole report, not just selected parts.

**TASK 1 - Identify the main components of a health report or research article**

1. The above reading lists five points to consider when writing a report. Do you agree with these suggestions or would you alter them in some way?
2. Scan four to five health reports published in any journal. Identify the main components of these articles by writing down the headings or subheadings used. Identify the purpose of each component.

#### FEEDBACK

**Guidelines for preparing your health report or scientific article:**

Most of the scientific articles, research reports or theses you will encounter are fairly rigidly structured around the following subheadings, each subheading representing a major section of the document:

|  |  |  |
| --- | --- | --- |
| Title, Author, and so on | Abstract | Introduction/purpose |
| Literature review | Aims | Objectives |
| Methods | Results | Discussion |
| Conclusion | Recommendations |  |

Each of these sections serves a particular purpose with respect to the overall task of the article or report, which is to communicate the nature of the research and its outcomes. In reports where research is not the primary focus of the report, the literature review may be limited or even fall away. Also see Vaughan & Morrow (1989) pages 126–128. Now consider the objective of each section in more detail.

|  |  |
| --- | --- |
| **TASK 2 - Identify the main objective of each part of a scientific article**  Write down the specific objective that each section of the report must achieve: | |
| **Title, Author and so on** |  |
| **Abstract** |  |
| **Introduction** |  |
| **Literature review** |  |
| **Aims** |  |
| **Objectives** |  |
| **Methods** |  |
| **Results/Findings** |  |
| **Discussion** |  |
| **Conclusion** |  |
| **Recommendations** |  |

#### FEEDBACK

|  |  |
| --- | --- |
| **Article component** | **Objectives that each component of the report must achieve** |
| **Title, Author and so on** | A succinct statement indicating what the study is about, utilising a few key words to catch the reader’s attention. |
| **Abstract** | A representative summary of all the main components of the thesis/report. A line or two per chapter heading and possibly one or two more for the results or major conclusion. |
| **Introduction** | What is the broad purpose for doing the research? How does it fit into the social, political, economic or health care setting/context in which it will eventually be applied? |
| **Literature review** | What is already known about the research problem? What is not known? What methods have been used to research it? |
| **Aims** | Based on what is already known/unknown in the literature, what is the specific research question that you will therefore try to answer through your study? Be very specific about what you will try to prove. |
| **Objectives** | Each objective must identify a step you will take to obtain each main piece of evidence you need to answer the main research question. |
| **Methods** | For each step, there needs to be a method you will use to get the information/evidence you require. This is described in the Methods section of the report. It must include sample size, sample selection method, data collection process, questionnaire design and testing, record verification, examiner variability, how you will analyse the data and so on. |
| **Results/Findings** | This section describes the main features of the data after you have summarised and analysed it, e.g. What is high? What is low? What is unusual? What is statistically significant? What is associated with what? Use a few tables and graphs to illustrate this but make all the important statements in the text. |
| **Discussion** | What does the data mean? This is where you interpret your results and establish what they actually mean. You might compare them with results in the literature. It should include: statement of principal findings; Strengths and weaknesses of your study; strengths and weaknesses in relation to other studies, particularly differences in results; meaning of study mechanisms and implications for policy makers and others involved in setting out protocols or guidelines of some kind; unanswered questions and future research. |

|  |  |
| --- | --- |
| **Conclusion** | What is the final answer to the original research question based on what you discovered through your study? |
| **Recommendations** | If there is a practical suggestion that arises directly out of your research findings, this can be stated here. |

The length and composition of your report can be critical in determining whether your report actually gets read. It is therefore important to think strategically about which parts of your report need to be given greater prominence and more space, and conversely, which parts should be trimmed down to occupy less space. Remember that in most cases, your readers are pressed for time and want to be able to read the important parts quickly and easily.

**TASK 3 - Estimate the approximate length of each section of the report**

Since most people struggle to find the time and energy to read lengthy reports in detail, it is important to keep them short and to the point.

Review the components of a report outline in the table above and suggest an appropriate length for each section.

#### FEEDBACK

Compare your proposed lengths with those on page 129 of the reading below.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 12 – Communicating Health Information. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 125–129.

Now that you have a clear framework for health reports and articles, try your hand at evaluating a report. In other words, put yourself in the position of the person who will mark your health report!

## CRITIQUE A SAMPLE REPORT

Here is a summary report from the Community Agency for Social Enquiry (CASE). We are not going to indicate at this point whether we think it is a good, bad or indifferent report. You will be doing this.

**READING**

Community Agency for Social Enquiry (CASE). (1999). Introduction & Health Status. In *The Second Kaiser Family Foundation Survey of Health Care in South Africa*.

Johannesburg: CASE: 1–7.

**TASK 4 - Read and critique a health survey**

1. Write a short critique of the health report contained in the above reading. Address all the main characteristics including the format, composition, structure, writing style, length, language, and so on.
2. If this was sent to you as part of your preparation for a meeting, would you make the time to read it?
3. Would you be impressed and/or persuaded by its presentation and message?

#### FEEDBACK

The criteria in Vaughan & Morrow, page 126 should have been useful in evaluating this report.

## ASSIGNMENT 2: FINAL REPORT

Prepare a descriptive summary of the dataset. Identify what statistical tools you wish to use. Determine what associations you will test and carry out, and any other analyses that you consider appropriate. Record and illustrate the results with graphs or other kinds of diagrams where appropriate.

Write the final report, complete with summary, references, graphs, tables, and so on. The report must provide a reader-friendly and convincing response to the original problem identified in the dataset. It should provide the basis for a health authority to address the Public Health issues reflected by the data.

A typed report of 6 – 8 pages, and definitely not more than 10, is required.

## ANALYSE AND INTERPRET THE DATA

For this task, your work from Unit 3, Study Session 4, could provide the starting point. As you went through the tasks in that session, you set out a strategy for analysing your data. Before you start, review your study questions and bear them in mind while you do your analysis.

In Task 5 you will test for any associations that may exist between the variables in the assignment dataset. You will need to apply some of the statistical tools introduced earlier in this unit to this dataset. You will also test any of the assertions made in the literature you reviewed earlier; in other words, whether this study confirms or contradicts what the literature says about the same issues.

**TASK 5 - Identify the main results/findings**

There is a lot of information in the dataset but not all of it may tell you something that is essential to know.

In this task:

1. Determine which information is the most important.
2. Determine the main message or result/s emerging from your analysis of the material.

#### FEEDBACK

The main findings are those that are best able to give you an answer to your questions. Remember that your main task is to determine how to address the health problem described. Refer back to the original set of questions you selected during Study Session 4 of Unit 3.

Note that the main findings could also tell you that you have asked the wrong questions, in which case you may need to ask another question. The end result should provide you with information that helps you to understand and/or respond to the health problem more effectively.

Large tables of data similar to the data presented in the assignment dataset are generally difficult to read and interpret. Therefore you will want to break down the data and present sections of it using smaller tables and/or various types of graphs.

**TASK 6 – Use tables or graphs, describe the main findings**

1. Develop a set of tables and graphs that you would use to represent the health information in your assignment dataset.

Refer to your answers in Task 5. Your work from the earlier sessions of this Unit will be of help.

#### FEEDBACK

Revisit Study Session 2 of this Unit for assistance with this task.

Clearly, the most important part of your job is to make sense of the information that has been given to you. You have already carried out a simple statistical analysis of the data and identified its main features (Task 1), and have summarised the main points in easy to read tables or graphs (Task 2). The next step is to determine what this information means.

**TASK 7 - Determine to what extent this is a Public Health concern**

1. What do the results of your analysis tell you about the health problem(s) that the data describes?
2. What are the Public Health implications of this health problem? To what extent *is*

this actually a Public Health problem?

#### FEEDBACK

This task has taken you through the process of data *interpretation*. When interpreting data, the trick is to choose the right questions to ask and to search the material for the answers. Having a thorough understanding of the health problem will guide your recommendations and the decisions you come to on the action that should be taken.

By now you should have a clear idea of the nature of the health problem and of its most important characteristics. Not every health problem in the community can be considered a *Public Health concern*. It would probably be helpful to refer back to your definitions of this concept from the very first Study Sessions of this module. Use this working definition of Public Health to help you decide if the health problem in this study has Public Health implications.

If this is the case, your report will need to persuade your managers that it is important to take some kind of action. If you present sound reasons to demonstrate that this is indeed a Public Health concern, your report will be more persuasive. Your managers will only act if they are convinced that the problem is, in fact, a Public Health concern, and that there is a Public Health intervention that could be made to address it.

Very few health problems are entirely unique. Similar problems have probably occurred somewhere else before. It is very useful to find similar information has been reported in studies from other places or from another time. You can learn a lot from the previous experiences of others and the information to support your own findings and recommendations.

**TASK 8 - Compare your findings to other experiences or studies**

1. Locate similar information reported elsewhere.
2. Compare and contrast your findings with those of other studies.

#### FEEDBACK

In previous Study Sessions you located examples of many different health indicators from a variety of sources. Use the same range of sources and compare your findings with that information.

Existing databases of health indicators such as those of the WHO and UNICEF allow us to compare the rates of particular health problems in countries around the world, or even to compare a district or region with another nearby. Different sources provide a measure of how large or small your own health problem is, when compared to others.

## DISCUSS THE IMPLICATIONS FOR DISTRICT HEALTH PLANNING

This is the final stage in preparing the draft material upon which your final assignment report will be based. This task encourages you to sift through the various points that have emerged from your preceding analysis and interpretation of the findings. It invites you to select the finding that seems to be the most important to you and consider what the health system could do in response to this research conclusion.

This is the recommendation you will need to persuade your managers to act on.

**TASK 9 - Evaluate potential Public Health responses to the findings**

1. Prepare a short list of possible actions to take.
2. Prioritise the most appropriate responses.
3. Present recommendations based on the information in your report.

#### FEEDBACK

The Conclusion and Recommendations sections of your report are the critical sections focused on Public Health *action*. The Conclusion section provides summary short answers to the study objectives/questions, highlighting for the manager the main findings of your report.

These findings should then link directly to your Recommendations. Your recommendations should be realistic and practical (*do-able*). They should also include a list of the resources (money, personnel, transport, training, collaboration) that are necessary to actually achieve your aims. You need to justify and motivate for the expenditure required to address the Public Health problems you have highlighted in your report.

**TASK 10 – Evaluate two sets of recommendations**

Consider how you would respond if you were a manager, and you received the following recommendations for action – the first from your TB coordinator and the second from your MCWH coordinator. Which coordinator is going to get your attention, and which is more likely to get the resources they need in the coming year?

**Recommendation 1 from the TB Coordinator:** This year we need to improve our TB Cure Rates by providing more education and treatment.

**Recommendation 2 from the MCWH Coordinator:** In the coming year we need to increase our immunisation coverage in children under 5 years old by 20%. To do this we recommend that the following be implemented:

1. We need to purchase and provide for running costs of 2 new 4X4 vehicles to increase outreach to remote villages.
2. We need to hire a nursing sister, nursing assistant and driver for each vehicle.
3. We should contract the local School of Public Health to develop and test culturally appropriate community education and awareness materials on childhood immunisation.
4. A pharmacist should be assigned to complete a review of the procurement and transport of vaccines, to identify why vaccines are out of stock or expired in remote sites.

#### FEEDBACK

You will undoubtedly have chosen Recommendation 2 from the MCWH Coordinator, because it is explicit, reasoned and specifies exactly how the recommendations should be carried out.

Note that sometimes, as with Recommendation 2, your recommendations may include the need for further, or specific, studies. Especially when you are dealing with summary data, the reason for the problems you identify may not be available in the data. You may, in fact, have identified a problem but you may not know the cause of the problem, hence the need for further study.

## PREPARE YOUR REPORT

By this stage you will have examined the information and come to some conclusions about it concerning, for example, what it implies, the quality of the dataset and how useable it is, as well as the Public Health issues may need to be addressed. You have summarised and illustrated some of the most prominent features of the data and made some recommendations on how the issue/s should be addressed.

This information needs to be effectively reported to your manager and others involved in the decision-making process of your health authority. The main message you wish to communicate must be clearly laid out and supported by good illustrations and a persuasive argument.

**TASK 11 - Finalise an outline structure for your final report**

In Unit 2, Study Session 5, Task 2, you developed a rough outline for your report. Now you should refine that outline. Remember that you have to catch and hold the attention of your managers (and others) long enough to communicate the most important information, and persuade them to act upon it. For this reason, the report should be clearly and appropriately structured *and not too long!*

1. Select an appropriate structure for your report.

Refer to the guidelines introduced in the previous Study Session and in Chapters 11 and 12 of Vaughan and Morrow (1989).

#### FEEDBACK

You can use the guidelines in Vaughan & Morrow (1989), Chapters 11 and 12, and follow the steps below to structure your report.

#### STEP 1: Select a title and sub-headings

Think of these as road signs that tell the reader where different parts of your report are located. They are short, clear indicators of what issues you are addressing at each point in the report, and also give an indication of the most important issues you will be addressing.

#### STEP 2: Determine the order in which the issues (sections) will appear

The sequence suggested in Chapter 12 of Vaughan and Morrow (1989) might suit your needs, but you should not feel bound by this format. You may wish to add different subheadings and exclude others. Whatever you decide, ensure that the report moves quickly from a brief introduction in which the problem is identified and described, towards a discussion that presents your interpretation and supports your conclusions.

#### STEP 3: Select tables, graphs and other illustrations to include

Tables and graphs carry quite a lot of information, and can be very confusing because of this. Not everything has to be presented in a table or graph. Some facts are easier to report in the text. Carefully select the information that will benefit most from a graphical illustration. You should also not assume that the graph or table speaks on its own. Your text must state clearly what is illustrated in the graph or table. Point out the highlights of the data in this way.

Having finalised the structure and order of your report, it is time to complete the writing and compile your report. You should have already developed most of the sections of your report in rough, draft form. Now you should put them all together.

As you write and review the report section by section, keep in mind the main study objectives of the report and the main findings of the analysis you carried out on the dataset. These should guide you in ensuring that all the sections deal adequately with the objectives and that you do not stray from the key issues. Be selective when

drafting your report.

**TASK 12 - Compile all the information into a report**

Compile a draft version of each section of the report.

#### FEEDBACK

It is useful to begin by writing the parts that you are most sure about. This builds your confidence and focuses your mind for tackling the more difficult sections. Do not assume that your reader knows everything you do. Use short, clear sentences where possible.

## PROOFREAD AND EDIT THE FINAL VERSION

The purpose of this final task is to ensure that you submit a product of the highest possible quality. Your managers in the work context should not be expected to read a report which has not been proofread and checked for accuracy. Your report should also contain proper referencing. A report that looks attractive and is reader-friendly may mean the difference between success and failure, i.e. whether it gets read and taken seriously or not.

**TASK 13 - Check for correctness, logical structure, referencing**

Check the report for clarity, correct spelling, grammar, accuracy, coherence and any other aspects you consider important.

#### FEEDBACK

It can help to get a colleague or friend to read your report as a test of how easy it is to read and how persuasive it is. They may also pick up minor typing errors you missed. Show your report to someone else (not necessarily a health professional) and ask them to give their opinion of your report and what impression it makes upon them.

They may offer some surprising suggestions.

**TASK 14 - Finalise and submit your report**

1. Read the report once more. Ensure that it meets your main objectives. Check it for accuracy, conceptual soundness and fix any typographical errors that may have crept in.
2. Check your references using the guidelines in the *SOPH Academic Handbook*.
3. Once the report is completed to your satisfaction, send it to your lecturer in accordance with the guidelines contained in the Module Introduction and the *SOPH Academic Handbook*.

4. Be sure to keep a duplicate copy of the report and any data you submit. Keep a computer file copy on a separate backup disk and a copy of the final printed version.

#### FEEDBACK

Your lecturer will provide you with feedback, and you will then be able to revise your assignment for final submission.

## SESSION SUMMARY

Hopefully, you have completed your assignment and more importantly, gained some insight and experience of interpreting epidemiological data.

Before putting this module away, please take the time to give us some feedback on your experience of the module and the assignment. It will help us to iron out any confusing aspects of the module, and to improve it as a learning tool. The evaluation form will be sent to you soon. Thank you for your contribution!

### COMMON KEY POINTS IN UNIT 3 SESSION 4 & 5



**Develop Critical Questions and Write a Report**

To develop critical questions for research, one follows the steps below:

* identification of a health-related problem
* decision on the action to be taken for this problem
* selection of what one needs to know to be able to take this action, leading to a set of critical questions to be used to collect data and analyse them (using the conceptual framework introduced to you in Unit 1 Session 2).

These critical questions have to be kept in mind, for the whole duration of the analysis and writing the report, to ensure consistency and to identify the action needed to tackle this problem.

When writing a report, it is important to synthesise information (as readers will not be prepared to read a 50 page report) and to be very systematic and scientific in the presentation of the report.

A mistake commonly made by students is to mix results and discussion. The results section should only consist of description of data from your own dataset, in the text or using tables and graphs. Comparison can also be made between data from the dataset.

The discussion section should include comments, comparison with data in the literature or data from other sources. No result can be discussed in this section unless already presented in the results section.

**Self-evaluation question:** Below are selected “Results” sections written by previous students, reporting on a dataset of tuberculosis indicators in five districts. Try to comment critically on them:

Example 1: … On analysing analysis, the data shows evidence of high TB incidence in the province of 555.8 per 100 000 population … It is clear that the province has a high incidence of TB across all the districts. Three out of six districts reflect an incidence that is higher that of the African continent incidence at 361 per 100 000 population (WHO, Global TB Report 2009).

District D seems to have enormous TB burden with an incidence of 931.5 per 100 000.

Example 2: … The smear conversion rate is used to monitor treatment success after the intensive phase (7-8 weeks after starting the treatment) specifically for new smear positive patients. Failure to convert from smear positive to smear negative is suggestive of a potential Multi-drug Resistant or Extremely Resistant strain of TB (MDR/XDR–TB). It should be also noted that poor conversion rates have implication for treatment outcomes, i.e. cure rates, treatment success rate and defaulter rate …