# TOPIC 2
# VARIABLES

*I have a great subject [statistics] to write upon, but feel keenly my*
*literary incapacity to make it easily intelligible without sacrificing accuracy and thoroughness.*
*Sir Francis Galton*

## Introduction

***concepts and definitions***

In this topic some statistical concepts and definitions and the main types of statistical data are defined and discussed. The objectives are to become aware of the way the concepts and definitions used to collect data can impact on analysis; and to become familiar with the different types of data and the terms and concepts used for defining data.

## Concepts and definitions

***first things to consider***

Before we do any statistical analysis we carefully examine the concepts and definitions used to collect the data. We ask ourselves a number of questions:

- **?** What was the **target population**?
- **?** What was the **survey population**?
- **?** What was the **statistical unit**?
- **?** How were the **observations** made?
- **?** What **standards** or **classifications**, if any, were used? What were the **coding decisions** made?
- **?** What **edits** or consistency checks, if any, were applied?

***number of observations***

Other points to consider include whether there are enough observations for analysis (for example, in some sample surveys detailed statistical analysis of variables might not be possible). Also look at the level of *non-response* to the question – this is discussed in Topic 3.

***definitions***

The **survey population,** or the **coverage** of the data, can have a substantial impact on the data analysis we can and cannot do. Examples of different survey populations include all persons, household heads, visitors, residents or persons over 15 years, employed persons and unemployed persons. Then we have to consider how and why the survey population was defined. Some examples of things to consider are:

? What is a 'visitor' – how long does a person 'visit' before being considered a 'resident' in a household?

? What is 'employed' – does this include both paid and unpaid work (i.e. subsistence labour and paid labour)? How many hours do you have to work before being considered as 'employed'?

? What is 'unemployed' – in some collections to be unemployed the person has to be seeking work and available to work, in others the person simply says they are unemployed.

**caution**

Often the definitions used for survey populations are more complicated that you would expect. The following section looks at the definitions used in labour statistics as a case study.

## Case study – labour force definitions

***labour force***

The definitions used for the labour force are perhaps some of the most complicated for data analysts. This is, in part, because the labour force contains both social and economic concepts (the System of National Accounts (SNA) defines economic activity and hence employment). It is also complicated by different countries having their own standards for defining the labour force and economic activity.

***terms***

The labour force is also a good example of how complicated statistical definitions can be. For example, the following is the International Labour Office (ILO) definition of the **economically active population**:

> *"…comprises all persons of either sex who furnish the supply of labour for the production of goods and services during a specified time-reference period. According to the 1993 version of the SNA, production includes all individual or collective goods or services that are supplied to units other than their producers, or intended to be so supplied, including the production of goods and services used up in the process of producing such goods or services; the production of all goods that are retained by their producers for their own final use; the production of housing services by owner-occupiers and of domestic and personal services produced by employing paid domestic staff."*
>
> *ILO, Yearbook of Labour Statistics, 1994, p. 3*

So the definition of economic activity is related to both the definition of production of goods and services and related to a specific time period.

***employed***

The ILO definition for the **employed population** is even more detailed:

> *"(1) The employed population comprise all persons above a specified age who during a specified brief period, either one week or one day, were in the following categories:*
>
> *"paid employment":*
>
> *(a1) "at work": persons who during the reference period performed some work for wage or salary, in cash or in kind;*
>
> *(a2) "with a job but not at work": persons who, having already worked in their*

*present job, were temporarily not at work during the reference period and had a formal attachment to their job;*

*"self-employment":*

*(b1) "at work": persons who during the reference period performed some work for profit or family gain, in cash or in kind;*

*(b2) "with an enterprise but not at work": persons with an enterprise, which may be a business enterprise, a farm or a service undertaking, who were temporarily not at work during the reference period for any specific reason.*

*(2) For operational purposes, the notion of "some work" may be interpreted as work for at least one hour.*

*(3) Persons temporarily not at work because of illness or injury, holiday or vacation, strike or lock-out, educational or training leave, maternity or parental leave; reduction in economic activity, temporary disorganisation or suspension of work; or other temporary absence with or without leave should be considered as in paid employment provided they had a formal job attachment.*

*(4) Employers, own-account workers and members of producers co-operatives should be considered as in self-employment and classified as "at work" or "not at work:", as the case may be.*

*(5) Unpaid family workers at work should be considered as in self-employment irrespective of the number of hours worked during the reference period.*

*(6) Persons engaged in the production of economic goods and services for own and household consumption should be considered as in self-employment if such production comprises an important contribution to the total consumption of the household.*

*(7) Apprentices who receive pay in cash or in kind should be considered to be in paid employment and classified as "at work" or "not at work" on the same basis as other persons in paid employment.*

*(8) Students, homemakers and others mainly engaged in non economic activities during the reference period, who at the same time were in paid employment or self-employment as defined in subparagraph (1) above should be considered as employed on the same basis as other categories of employed persons and be identified separately, where possible.*

*(9) Members of the armed forces should be included among persons in paid employment*

*ILO, Yearbook of Labour Statistics, 1994, p. 223*

***unemployed***

The definition for the **unemployed** is just as complicated. The main criteria are:

*The unemployed are those persons above a specified age who during the reference period were:*

*"without work" (not in paid employment or self-employed);*

*"currently available for work"; and*

*"seeking work" - had taken specific steps to seek paid employment or self-employment, such as registration at an employment exchange, applied for a job, placed or answered a newspaper advertisement; tried to establish their own business, visited work places looking for work, sought assistance from friends and relatives, or applied for a permit or licence etc.*

*ILO, Yearbook of Labour Statistics, 1994, p. 483*

***seeking work***

The ILO note that in some countries the "seeking work" criteria might need to be relaxed because of the lack of labour market organisation or the limited size of the labour market. If these ILO criteria are applied, a person who buys a newspaper and looks at the 'situations vacant' is NOT considered to be unemployed - he or she would have to have made a job application to be defined as 'unemployed'.

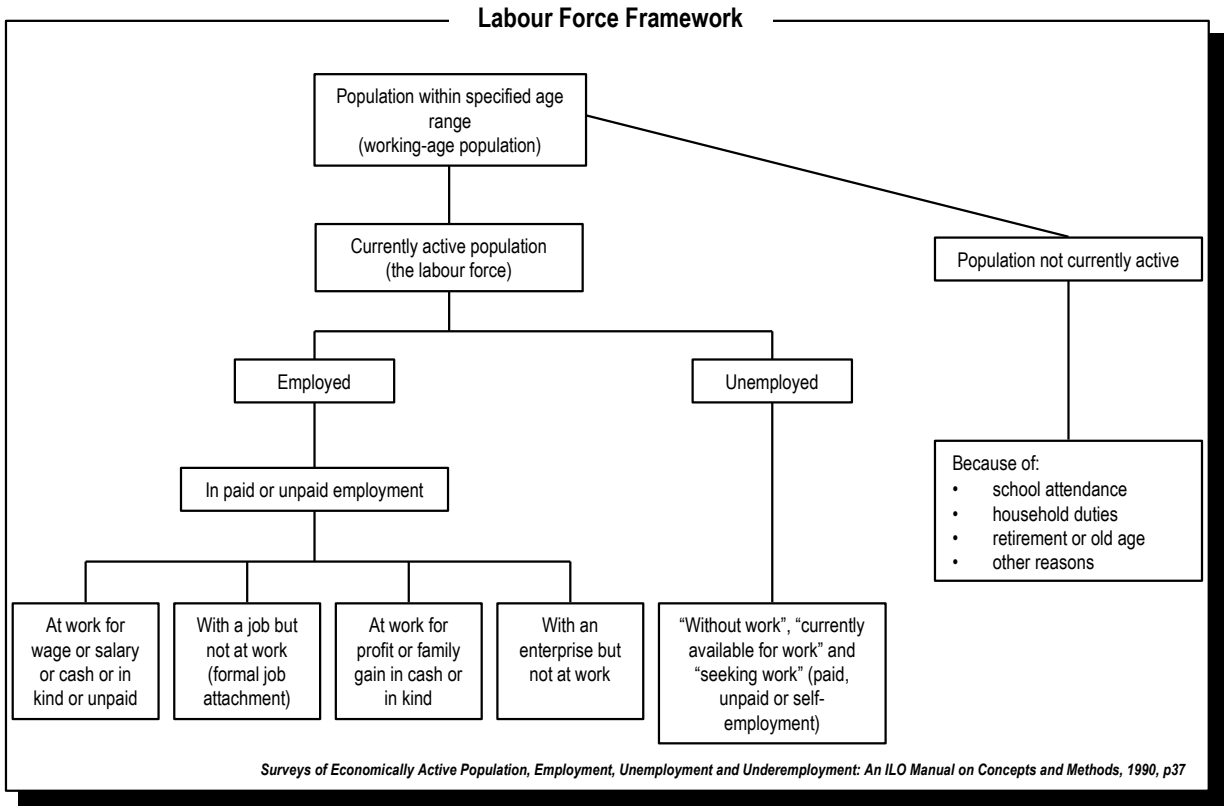
***not in the labour force***

The third main group in an analysis of the economically active population is those who are within the specified age criteria (called the '*working age population*') but who are not classified as employed or unemployed - those who are **not in the labour force**. Therefore:

**Working age population** = Labour force + Not in the labour force

**Labour force** = Employed + Unemployed

These basic population and labour force concepts can be represented diagrammatically:



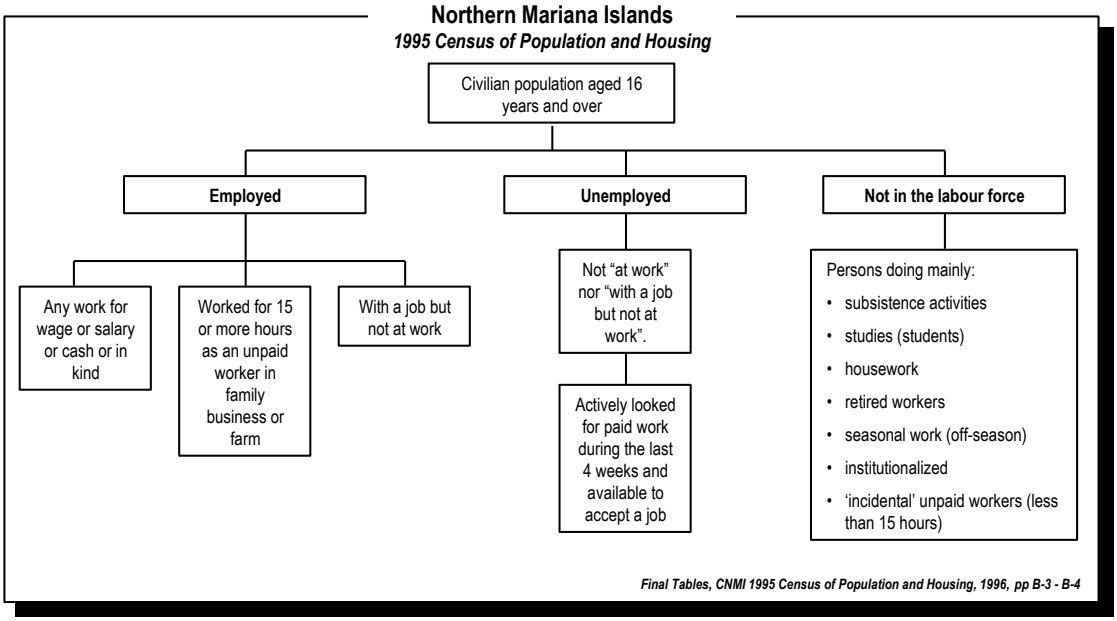

***look past the jargon***

In this summary form the ILO definitions are easier to understand – often the technical language and jargon used in formal definitions make the concepts and definitions difficult to understand.

***check the survey population***

Before doing any analysis of the labour force you have to think about how the population has been divided up into the different survey populations. The labour force is one of the most complicated areas in terms of concepts and definitions. But the same principles apply when you evaluate any statistical variable prior to analysis.

Consider the following example from the Commonwealth of the Northern Mariana Islands (CNMI):




***Northern Marianas***

This diagram was developed based on definitions from the 1995 Census report. We might have to follow up with the Statistics Office to find out more about the survey population – for example what status would be given to a person who was doing mainly housework but also 10 hours of paid work? The Statistics Office would tell you that anyone with a paid job would have been coded as employed, even though the hours spent in the paid job might not have made it their main activity.

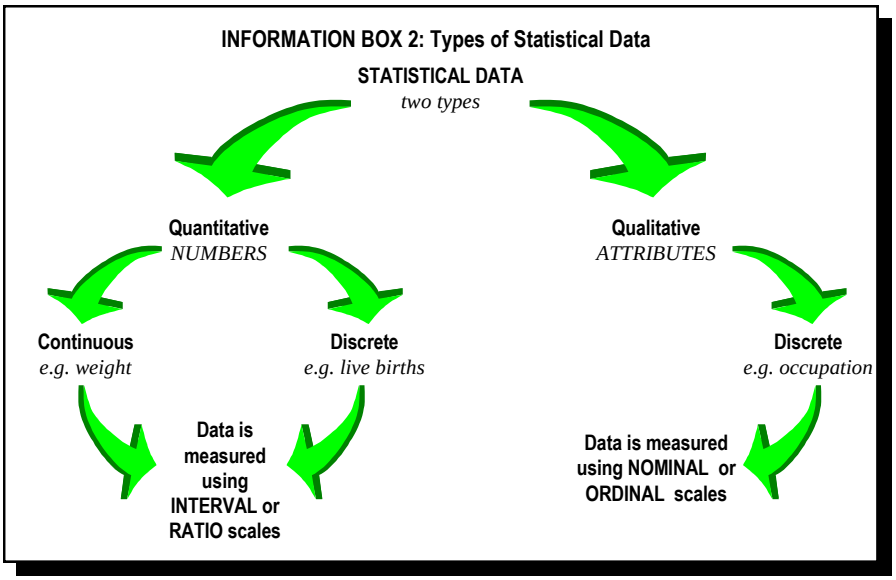
**TIP**

Often a good first place to start to come to grips with the survey population, concepts and definitions is with the Statistics Office or the people who collected the data.

## Types of data

***different types of data***

As well as the concepts and definitions we also determine what type of data has been collected for each variable as this also affects the analysis we can undertake. The following diagram summarises the way we categorise different types of data.




***definitions***

The word **statistics** has three main definitions: the subject, the data and the summary figures derived from the collected data. To help resolve any confusion there may be between these three definitions, we shall use the term **statistical data** to mean the values observed in collecting information.

***two types of statistical data***

Statistical data can either be **quantitative** or **qualitative**, that is, they can be numbers or they may be some attribute of a particular unit. For example, if we were to collect information from the participants of this course on their age and sex, the set of statistical data relating to ages would be quantitative. That is, it would consist of a series of numbers representing the ages of different individuals in years and months. The set of statistical data relating to sex would be qualitative. That is, we would record male or female for each person.

***data collections use both types***

We can observe the values of many variables of a statistical unit. Some of these variables may be quantitative and others qualitative. Thus in a survey of employees in a particular industry, observations for each employee may include age, sex, occupation, educational level, wage or salary amount, and the length of time spent in their job. Of these six characteristics, the first, fifth and sixth are quantitative variables while the second, third and fourth are qualitative variables.

## Types of quantitative and qualitative variables

***continuous variables***

For practical purposes, we can distinguish between two types of quantitative variables and we need to be quite careful to make sure that we know what type we are dealing with in a particular situation. The first type of quantitative variable is one which can take any value within some specified range. We use

the term **continuous** to describe this type of variable. If we were measuring land area of farms, the area is a continuous variable. We would measure the area in hectares (ha) and, in theory, we could have any value as a result within the range of say, 0 – 5,000 ha. Results of 40.63 ha, 7.405 ha and 53 ha would all be valid. If we plotted all the possible areas on an axis, the result would be a continuous line.

***discrete quantitative variables***

The second type of quantitative variable is one that takes values such as the number of workers in an industrial establishment. We describe this type of quantitative variable as **discrete**. If we conduct a household survey and collect information on the number of people who usually live in the household, then we know the answers we get can only be 1, 2, 3, 4, 5, and so on. For this variable, values such as 4.2, 1.3 and 3.6 are invalid. It does not make sense to talk about 4.62 people. If we put all the possible values of a discrete variable on an axis, we would have a series of points. In most practical examples, discrete quantitative variables usually take whole number (or integer) values.

***discrete qualitative variables***

All qualitative variables are discrete. This is because qualitative variables depict attributes which, by their nature, are discrete. In the example above it did not make sense to discuss 4.62 people who usually live in a household. Likewise, a person's eye colour, occupation or address can only be discrete.

### *examples of continuous variables*

Some common examples of continuous variables are:

height of a person
weight of a person
maximum daily temperature
age of a person
length of a road
daily rainfall

How much rain?

### *examples of discrete variables*

Some common examples of discrete variables are:

number of children born to a woman
number of rooms in a house
number of pigs owned by a household
number of employees in a company
number of pupils in a school
number of farms in different area units
number of taxis in a town

*How many rooms?*

***income ?***

An interesting variable to consider is income. Most text books suggest that income is a continuous variable. Take a few moments to think about this variable and discuss this variable during the course.

Income is discrete in terms of the actual amount you get paid – you can count the money and you don't get paid in ½ or ¼ of a cent. But when you look at income in terms of an hourly rate it becomes continuous – if you are on a salary convert it to an hourly rate – now you will have ½ or ¼ or other fractions of cents.

***number of employees ?***

Number of employees is another interesting variable – once again how it is measured determines if it is continuous or discrete. Can you see how this could happen?

Obviously a count of employees gives a discrete number – a company could have 15 full-time employees, five part-time employees and two casual workers. This would give a total of 22 employees, but what if we were interested in full time workers only? 'Full-time' equivalents could be formed – so here there are 15 + $^{5}/_{2}$ = 17.5 full-time equivalent staff. But how will the casual workers be included? We need to know more about the hours that they work to do the calculations to make them 'full-time' equivalents. This could result in a number like 18.25764 full-time equivalent staff – a continuous variable.

***sometimes hard to distinguish***

When we come to look at continuous and discrete variables in real life, the problem is a little more complicated. Although, in theory, the area of a farm may be any value, in practice we may only be able to measure the area to the nearest tenth of a hectare. Therefore, we could have a series of observations such as: 0.6 hectares, 7.4 hectares, 3.4 hectares, 18.9 hectares, 11.4 hectares, 73.6 hectares, 80.0 hectares. At first glance this may seem to look like a discrete variable, since values such as 7.384 do not appear in our results. What has happened, of course, is that these values have been recorded to the nearest tenth of a hectare (for example, 7.38 hectares will be recorded as 7.4 hectares). We know that all values can occur, but we realise that the inaccuracy of our measuring procedure means that values are rounded as they are recorded.

**TIP**

try *'counting'* the variable to see if it is continuous or discrete — you can count each child born (discrete) but could you 'count' height ?? Continuous variables are usually 'measured'

***continuous variables are usually approximated***

Continuous variables are usually approximated and the degree of approximation largely depends on what is being observed, how it is being observed and what precision is required. For example, when height is measured it is usually to the nearest centimetre or inch, not very accurately to say, 10 decimal places.

***but discrete data can be confusing too***

A different confusion can arise when we look at some discrete variables. If we conducted a business survey and recorded the number of employees in different companies we might present the results of the survey as groups of units (for example, that there are 26 companies in the group 50–99 employees). This seems to imply a continuous possible range of values for the companies, but we must realise that this will only contain the values 50, 51, 53 up to 99. We cannot have a firm with 76.38 employees; even though we quote a range of values, the variable is still discrete. We must take care, therefore, not only to look at the actual values in the recorded data, but to examine the variable to determine whether it is continuous or discrete.

## Scales of data measurement

***scales are used to measure data***

Quantitative and qualitative variables are measured in different ways. When we make observations of variables from statistical units, the observed data can be measured on a variety of scales. These scales can be usefully classified as nominal or ordinal which are used for qualitative data; and interval or ratio scales which are used for quantitative data.

### Qualitative data

***nominal scale***

The **nominal** scale is the most basic form of data measurement. It is the result of making observations of variables that are classified as being in a particular category. Examples of variables that have a nominal scale of data measurement are: sex, marital status, employment, industry, religion, country of birth and hair colour. There is no particular order used in the category in nominal data and therefore no category is considered larger or smaller than any other category (for example, there is no order to male and female). It is important that for variables measured using a nominal scale that the categories of the variable are **mutually exclusive** (that is, all values of the variable can only be classified in one of the categories) and **exhaustive** (that is, all possible values of the variable can be classified to one of the categories).

*We use a **Nominal scale** when the observations are mutually exclusive and exhaustive and when there is no order to the observations.*

***ordinal scale***

The **ordinal** scale is measurement that enables the data to be ordered (that is, different values can be identified as being larger or smaller). However, the difference between values cannot be meaningfully measured. Examples of data measured using an ordinal scale are: rating a worker's performance as 'good', 'average' or 'poor', classifying people as 'tall', 'average' or 'short' and rating your personal health as 'above average', 'average' or 'below average'. From these examples, it is clear that the different observations of the variable can be ordered (for example, good is better than average and average is better than poor), but the difference between these values cannot be meaningfully measured. Ordinal scales are most common in attitudinal surveys – such as the evaluation questionnaire for this course.

*We use an **Ordinal scale** when the values of the observations can be ordered or ranked but when we cannot measure the difference between the observations.*

## Quantitative data

***interval scale***

The **interval** scale is when the difference between values can be measured, but there is no relativity between the values (that is, we cannot say one value is so many times larger or smaller than another value). Two examples of data measured using an interval scale are time and temperature.

***temperature***

If the maximum and minimum temperatures today are 15°C and 30°C, then we can say that the maximum temperature is 15°C more than the minimum temperature. However, we cannot say that the maximum temperature is twice as hot as the minimum temperature. The key feature of an interval scale is that the zero point is just another point on the scale. Zero degrees Celsius does not mean the absence of heat. To see this we could convert the temperatures in the above example to the Fahrenheit scale (that is, 59°F and 86°F). We can now see that to say the maximum temperature is twice the minimum temperature is meaningless.

***time***

One group of people started a task at 1pm and it took them 20 minutes. Another group started at 2.10pm and it took them 10 minutes. We can say that the first group was ten minutes slower or twice as slow as the second group. But we cannot say that 1.20pm is ½ as slow as 2.10pm. That is, the difference in the time in minutes can be measured but the time cannot.

*We use an **Interval scale** when we can measure the difference between observations but there is no relativity between the observations and No fixed zero point.*

***ratio scale***

The **ratio** scale is when data is measured so that the relativity between values can be stated meaningfully. As in the interval scale, the observations are ordered and the difference between observations is meaningful. In addition, the ratio scale uses the number zero to indicate the absence of the characteristic being measured — $0 means you have no money. This is the distinguishing feature of the ratio data, the fixed zero point. This enables meaningful relative comparisons to be made between the observations. For example, if you had $5 and your friend had $10, your friend has twice as much money as you. Hence, the ratio between the two observations is meaningful.

There are many examples of ratio data: height, age, distance and income being a few. Not only can we say that someone who is 40 years old is 30 years older than someone who is 10 years old, but we can also say they are four times as old. A zero height, age, weight, distance or income is meaningful and cannot be interpreted in more than one way.

*We use a **Ratio scale** when we can measure the difference between observations and there is relativity between the observations.*

## Univariate, bivariate and multivariate data

***commonly used terms***

The final group of terms used to describe variables relate to how the data is presented or analysed. The terms univariate, bivariate and multivariate are used quite frequently in statistics. It is therefore necessary to know what these terms mean, although for the purposes of this course, we shall confine the discussions to univariate and bivariate data only.

***one variable***

Univariate simply means that we are dealing with one variable. For example, if you had statistical data from a school on the heights of children, then the data would be univariate.

***two variables***

Bivariate means dealing with two variables. Again, if we take the example from a school, we can look at the relationship between the heights and weights of children in the class. We can also look at the relationship of heights and age or the relationship between weight and age.

In everyday life, statisticians are usually involved in analysis of bivariate data. For example, we can plot heights and weights on a graph to see the relationship between height and weight. A scatter diagram is often used to see the relationship between two variables – we shall discuss scatter diagrams in Topic 4.

***more than two variables***

Multivariate means dealing with more than two variables. For example, if we want to know the relationship between height, weight and age, then this would involve a more complicated, multivariate analysis of the data. Multivariate analysis involves some complicated statistical techniques and it is not the intention of this course to go into them. Those having a particular interest in studies involving multivariate data are advised to consult appropriate texts on statistical methods.

# ... Exercises ...

**1. Indicate which of the following are discrete or continuous variables (c for continuous and d for discrete):**

| **Observation** | **Variable** |
|---|---|
| (a) The time taken for you to get to work; | _____ |
| (b) The number of couples who were married last year; | _____ |
| (c) The number of goals scored by a women's basketball team; | _____ |
| (d) The speed of a bicycle; | _____ |
| (e) Your age; and | _____ |
| (f) The number of subjects you can choose to study at university. | _____ |

**2. What is the scale of measurement used in the following table:**

| Behaviour | Number of participants |
|---|---|
| Excellent | 5 |
| Very good | 12 |
| Good | 10 |
| Bad | 2 |
| Very bad | 1 |

Type of scale: _____

# … Self-Review …

**1. Indicate which of the following are quantitative (qn) or qualitative (ql), discrete or continuous variables (c for continuous and d for discrete):**

| Observation | qn/ql | c/d |
|---|---|---|
| (a) The time of a phone call between two people; | | |
| (b) The annual income of an individual; | | |
| (c) The number of people working in the National Statistics Office; | | |
| (d) The number of brothers and sisters you have; | | |
| (e) The distance between your house and work; | | |
| (f) The number of pages in the Data Analysis course notes; and | | |
| (g) Your occupation. | | |

**2. List the possible response categories and type (nominal, ordinal, interval, ratio) for the following:**

| Variable | Categories | Scale of measurement |
|---|---|---|
| (a) sex | | |
| (b) marital status | | |
| (c) attendance at a course | | |
| (d) temperature | | |
| (e) height | | |
| (f) year started work | | |

**3. a) The following is an exert from the Small Holder Agricultural Survey in Vanuatu in 1990. List the variables being collected and what scales of measurement you would use in the table marked 3.a) on the next page.**

**Small Holder Agricultural Survey of 1990, Vanuatu**

**Pat 1 : Kokonas**
***Part 1: Coconuts***

1.1 Yu gat eni kokonas?
*Do you have any coconut palms?*
YES 1 | NO 2 (go long Pat 2)

1.2 Long las 12 manis (Julae '89 kasem Jun '90) yu bin planem eni LOKOL kokonas?
*In the last 12 months (July '89 - June '90) did you plant any local (indigenous species) coconuts?*
YES 1 | NO 2 (go long 1.7)

1.3 Hamas LOKOL navara yu bin planem long las 12 manis?
*How many local coconut palms or shoots did you plant in the last 12 months?*

1.4 Wei we yu planem :
(Namba blong ol navara)
*How did you plant them : (number of coconut palms/shoots)*
Olbaot nomo *Random* | Mekem ol speis i stret *In a grid* (go long 1.5)

1.5 Igat hamas mita long ol speis bitwin blong ol navara :*
*What is the space in metres between the trees?*
Mita *Metres*

1.6 Wanem kaen graon : ( Namba blong navara)
*What type of soil or area did you use (number of coconut palms/shoots)*

| | Namba blong navara *(# palms/shoots)* |
|---|---|
| A. Niu bus we yu kliarem, (i nogat narafala krop) *Newly cleared land (with no other crop)* | |
| B. Niu bus we yu kliarem, (wetem narafala krop) *Newly cleared land (with other crops)* | |
| C. Long wan eria we i gat narafala kokonas i stap finis *In an area that already contained other coconut palms* | |
| D. Long wan eria we i gat ol narafala krop i stap finis. *In an area that already contained other crops.* | |

1.7 Long las 12 manis i jas pas yu bin mekem o salem eni kopra?
*In the last 12 months did you make or sell any copra?*
YES 1 (go long Pat 2) | NO 2

1.8 Talemaot from wanem yu no bin mekem eni kopra long period ia?
*Why didn't you produce any copra in that period?*

| | | | |
|---|---|---|---|
| A. Yu wok ful taem mo ksaem salari from *You had a full-time job with salary* | 1 | F. Ol kokonas oli no karem frut yet *The trees had no fruit yet* | 6 |
| B. Yu bin wok long narafala krop *You worked on other crops* | 2 | G. Yu nogat inaf stamba kokonas *You didn't have enough coconut palms* | 7 |
| C. Yu olfala o yu sik *You were old or sick* | 3 | H. Yu nogat inaf frut kokonas *You did not have enough coconuts* | 8 |
| D. Naturol disasta *Natural disaster* | 4 | I. Narafala rison _____ *Other reason* | 9 |
| *E. I nogat janis blong usum drae* *There was no chance to use the drier* | 5 | | |

* Question 1.5 was added for this exercise.

**3. (a) Variables being collected in the Agricultural Survey in Vanuatu in 1990.**

| | Variable | Scale of measurement |
|---|---|---|
| 1. | | |
| 2. | | |
| 3. | | |
| 4. | | |
| 5. | | |
| 6. | | |
| 7. | | |
| 8. | | |
| 9. | | |
| 5. | | |

**3. (b) Imagine you are deciding on the analysis of the data from this survey. List at least two univariate and bivariate analyses you would plan to do in the following table.**

**Types of data being collected in the Agricultural Survey in Vanuatu in 1990.**

| Data type | Example |
|---|---|
| Univariate | |
| | |
| | |
| | |
| Bivariate | |
| | |
| | |

# Excel – commands

## *What is Microsoft Excel?*

Microsoft Excel is an electronic worksheet that does a lot more than replace a statistical table or accountant's ledger book with its neat columns and rows. Excel can perform just about any mathematical calculation you want, and changes can be made to your numbers and calculations without having to painstakingly rewrite the entire worksheet. A worksheet is essentially a large grid that aligns entries in rows and columns so that it is easy to see which numbers you are working with.

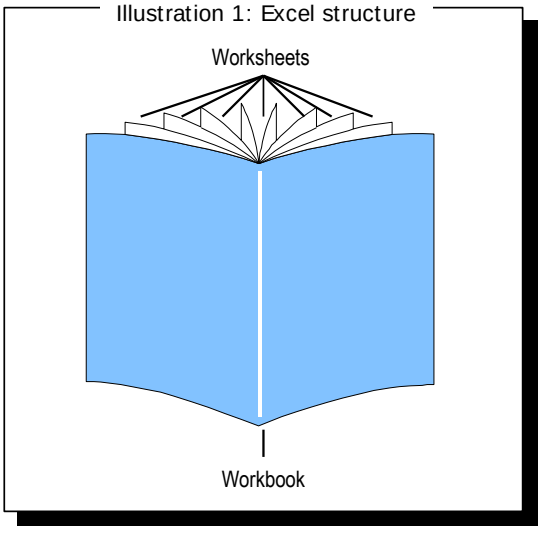
Microsoft Excel can also create charts from your numbers much more easily than you can create them by hand, and can exchange information easily with other programs such as Microsoft Word, Microsoft PowerPoint and Microsoft Access so your information can be presented to others in nearly any form you choose.

With Excel, it is easy to enter information into a worksheet, and then change, delete or add to the information. You can change and update your data easily. You can organise your worksheets in your workbook to have worksheets for different tables, subjects or years (for example, you might place all the worksheets from the Census into one workbook), and then name them so you can locate the information you need quickly.

## *What really happens in Excel …*

The basic idea of Excel is the 'workbook' which contains one or more worksheets.

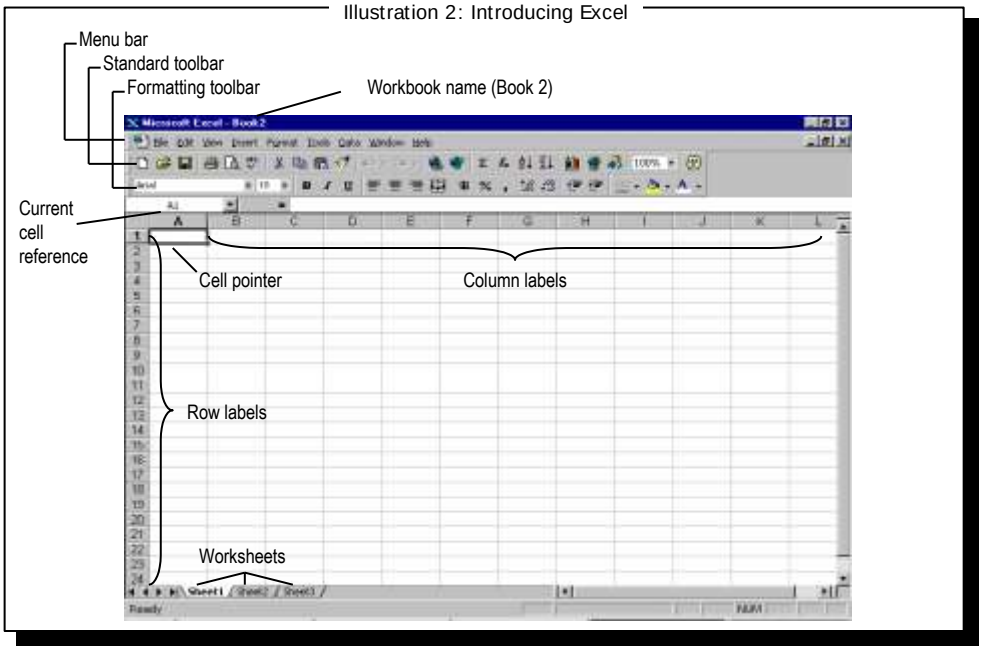
Illustration 1: Excel structure




A worksheet is similar to an accountant's ledger or spreadsheet, with numbers, text, and calculations lined up in columns and rows, and each 'box' in the sheet is a '*cell*'. BUT, unlike an accountant's ledger, when you type numbers into Excel, the program performs the calculations for you. The most important thing about Excel is how it organises the data we put into it by using COLUMNS and ROWS.

As you can see from Illustration 2, Columns are labelled with letters of the alphabet and rows are labelled with numbers. The column and row labels act like street signs in a town – they can help you identify your current position. In Illustration 2 the '*active cell*' is in Column A and Row 1 – so its Cell Reference is A1, which you see in the 'current cell reference' in the Formula bar.

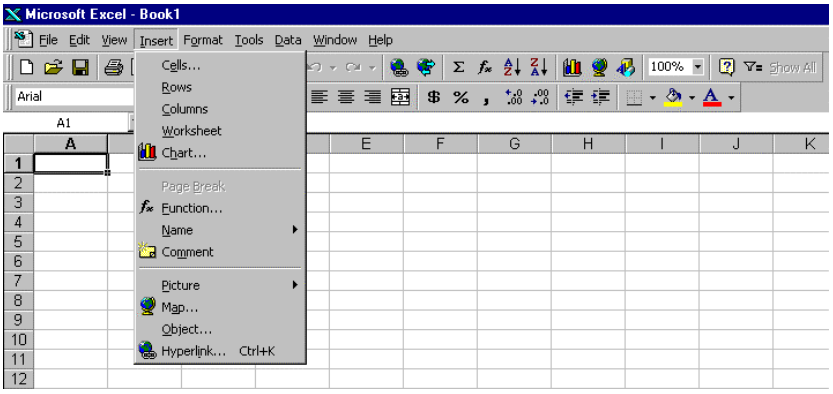
Illustration 2: Introducing Excel




## *The menu bar*

The menu bar contains nine pull down menus, some you will be familiar with from other Microsoft applications. Three commonly used menus that you might not be so familiar with in Excel are Insert, Format and Data. The most frequently used options from these menus are described here.

## *Insert menu key options*




**Cells**

Select this option to insert one or more cells, or an entire row or column.

**Rows**

Select this to insert a row ABOVE the active cell (or selected cells).

**Columns**

Select this to insert a column to the LEFT of the active cell (or selected cells).

**Worksheet**

Select this to insert a blank worksheet BEFORE the active sheet.

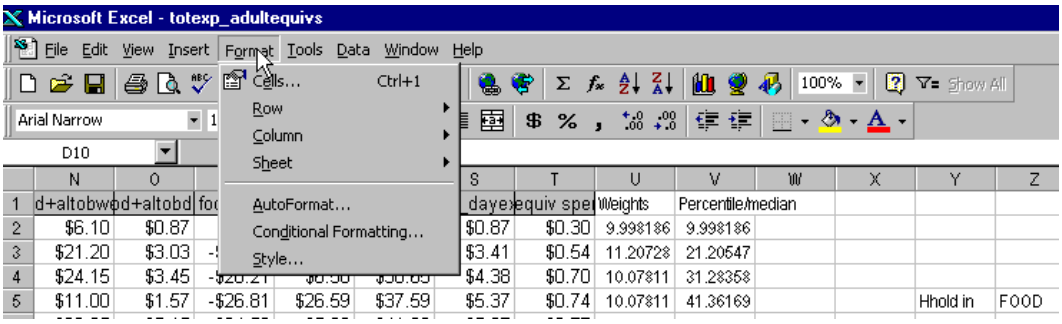
**Chart ( )**

Select this option to start the chart wizard. It is easier to create a chart if you SELECT the cells that you want to chart FIRST.

**Function ( )**

Functions are predefined FORMULAS that perform calculations by using specific values, called arguments. For example, the SUM function adds values or ranges of cells, and the AVERAGE function calculates the arithmetic mean. When you select this option, you then select the function you want to use, and Excel will then prompt you to select the cells you want to perform the function on.

## *Format menu key options*




**Cells**

This is the Excel option you will use the most. Select it to format one cell or a range of cells. These options can also be used with the buttons on the Formatting toolbar:

**Format cells tab**

**number** which is the data type (number, text, currency, date)

**alignment** (to wrap text so it fits in one cell or merge cells)

**font** (type, size, style, colour and effects)

**border** (use preset outlines or click with the mouse in the Border area to add borders to the top, bottom or sides)

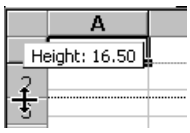
**patterns** (to format the background shade of the cell)

**Formatting toolbar buttons**


**Row**

**Height** – select this option to change the height of one or more rows. You can also use the mouse to do this:

Click on the BOTTOM border of the row label and drag the mouse to the height you want.




**Autofit** – select this option to automatically change the height of one or more rows to 'fit' the data or text you have entered. You can use the mouse to do this:

Position the mouse between two rows and double click to autofit the row height.

**Hide** – select this option to hide the active row. The row is still in your data and has NOT been deleted

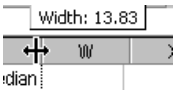
– see how the row numbers 'skip' the hidden row:


**Unhide** – to display hidden rows, select cells in the row above and in the row below the hidden rows and then click on this option.

**Column**

**Width** – select this option to change the width of one or more columns. You can use the mouse to do this:

Click on the RIGHT border of the column label and drag the mouse to the width you want.

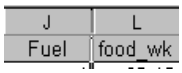



**Autofit selection** – select this option to automatically change the width of one or more columns to 'fit' the data or text you have entered. You can use the mouse to do this:

Position the mouse between two columns and double click to autofit the column width.


**Hide** – select this option to hide the active column. The column is still in your data and has NOT been deleted

– see how the column letters 'skip' the hidden column:




**Unhide** – to display hidden columns, select cells in the column before and in the column after the hidden column and then click on this option.

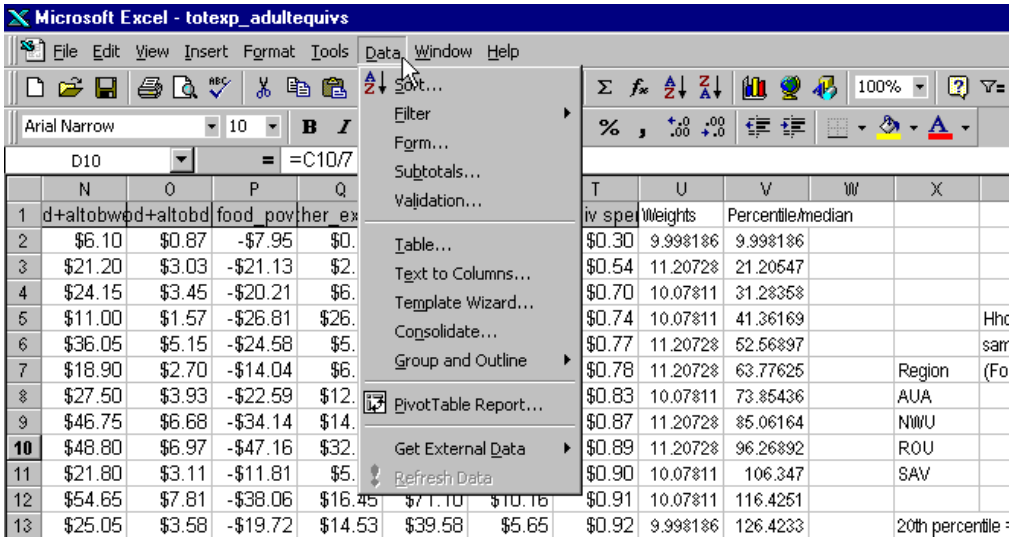
**Sheet**

**Rename** – to change the name of the sheet. You can do this with the mouse by double clicking on the name of the sheet.

**Hide** – hides (does not delete) the sheet.

**Unhide** – if you have hidden one or more sheets use this option to display them again.

## *Data menu key options*

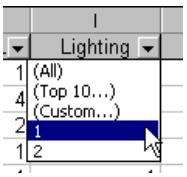



**Sort**

Click in a cell in the column you want to sort by to sort ALL the data in the worksheet by this variable. Be VERY CAREFUL to sort ALL columns and if the SORT WARNING dialog box is displayed make sure you **expand the selection**. Sorting can corrupt your data so be very careful. If you are not sure what you have done use the undo button ( ) to try again. You can use the sort buttons ( ) on the Standard toolbar as well, but be careful because Excel will not give you a warning about expanding the selection if you use the buttons.

**Autofilter**

This is useful when you want to find certain numbers or codes in your data. Click on this option to show arrows beside the column labels in your data, and then select the value you want to display. In the following example, the Lighting column will be filtered by the '1' value.




When you have a filter assigned to a column, the filter arrow changes to BLUE. Click on the arrow and click on the ALL option to remove the filter.

**Form**

You can create a simple data entry form by selecting this option. First make sure you have labelled the data entry columns.

**Subtotal**

This is a quick and easy way to create subtotals, totals or other functions (such as averages) for columns. For statistical tables, it might be easier to create a PivotTable.

**PivotTable Report**

You can create simple or complex statistical tables using this option. It is covered in detail in Topic 3.

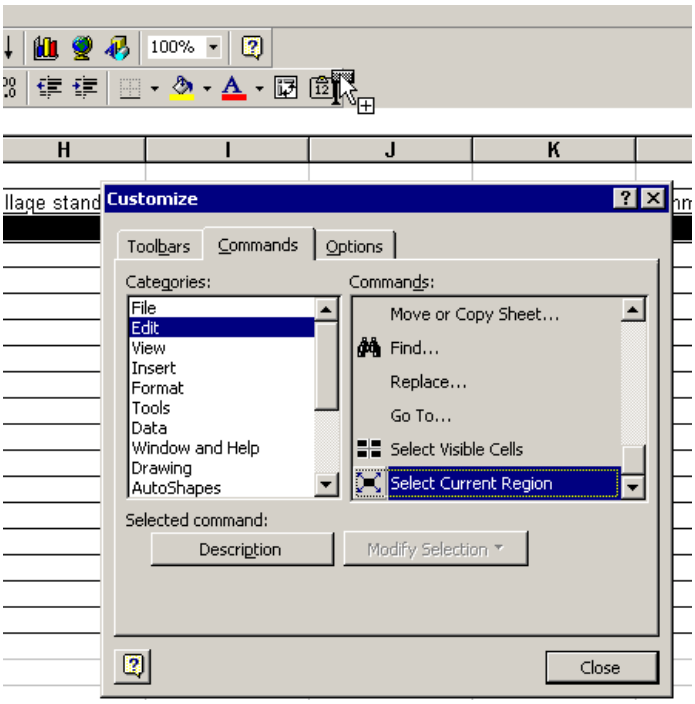
# one step further

## *Selecting data in Excel*

There is a quick and easy way to select a data array in Excel, but its so convenient it is not a standard toolbar feature and you have to customise your toolbar to use it.

1. From the tools menu select the Customize option. The Customize dialog box opens. From the Categories list click on the Edit option.

2. Scroll down the Commands list to the end and you will see the Select current region command. Click on this option and drag the mouse up to the toolbar – do not release the mouse. The mouse pointer will change to a + shape when it is over the toolbar and then you can release the mouse. Close the dialog box.




3. The current region is the cell block including the one you are in and extends up and down until the first empty row and column. So click anywhere in your data and then click on the  button to select all the cells.

You can use the select current region command from the Edit menu if you don't want to customise your toolbar:

1. Click any cell in the range of data.
2. On the Edit menu, click Go To.
3. Click Special.
4. Click Current region.