

2. SORTING OUT YOUR STATISTICS

In this session:

- ▶ measurement and values
- ▶ data and information
- ▶ measurement scales
- ▶ cleaning up data
- ▶ validity and reliability
- ▶ sampling
- ▶ summarising data

Objectives

After this session you will be able to:

- ▶ discuss statistical measurement techniques
- ▶ differentiate between data and information
- ▶ outline the common measurement scales
- ▶ apply the concepts of validity and reliability
- ▶ discuss sampling techniques

Measurement

The basis of all statistics is measurement. All this means is putting a numerical value on something. For example your last electricity bill was expressed in a dollar amount. That is measurement. If the sales of your company rise by a certain percentage, that is measurement.

The important point is that you should worry only about the measurement of things that influence your business decisions. Do not collect statistics on issues that have no relevance to your business situation.

Sometimes measurement may be expressed in other terms – for example – my shirt is blue – where we have used colour as a measuring yardstick. But in a business scenario, while it may be used in conjunction with other measurement methods, numbers are usually what is important.

Note however that the numbers only provide ‘data’. This data is of no use whatever unless we work out what it means to us – thus converting it into ‘information’.

To take a simple example, imagine you are measuring the height of all the students in your class. You might wind up with a list like the one below:

- | | |
|------------------|-----------------|
| ▶ Carlos: 1.80 m | ▶ Joshua: 1.79m |
| ▶ Penny: 1.62 m | ▶ Ting: 1.63m |
| ▶ Marika: 1.77 m | ▶ John: 1.78m |
| ▶ Karen: 1.45m | ▶ Jesse: 1.22 m |

What you have in front of you is a collection of numbers. They may be nice numbers, or interesting numbers, but the key is to work out what they mean and how they can be used.

So to make them meaningful we have to put them in some sort of order that is useful for our purposes. This leads us to the idea of 'types of measurement'.

Types of Measurement

The key concepts you will come across are:

- ▶ nominal scale
- ▶ ordinal scale
- ▶ interval scale

Nominal Scale

In our collection of figures above the heights of all your classmates are written down randomly (probably the sequence in which you measured them). This is said to be 'nominal scale' – that is in no particular order. And therefore probably not very useful when it comes to answering any questions or providing any information.

Ordinal Scale

This is where you take the next step – put them in some sort of ORDER – hence the word 'ordinal'. If you are doing the exercise 'just for the record' you might just put them in alphabetical order. However you may have another purpose say to work out what sizes of clothing you should be ordering for the class sports day. Thus you might arbitrarily group them into:

- ▶ Tall
- ▶ Medium
- ▶ Short

Note that we have simply assigned descriptions to each category. We might say that:

Carlos, Joshua, John and Marika are tall

Penny, Karen and Ting are medium

Jesse is short

Interval Scale

So they are in some sort of useful order but you have not tried to make the table neat and precise by making sure all the categories are of identical mathematical measurement. If you did you might say:

Tall: 1.6m to 1.79m: Carlos, Joshua, John and Marika

Medium: 1.4m to 1.59m: Penny, Karen and Ting

Short: 1.2m to 1.39 m: Jesse

Note that each category has a range of 19 cm. This is called an 'interval' scale because the range – or interval – is the same for each group of numbers.

Errors and Cleaning Them Up

Before you use the statistics to make dramatic decisions you need to make sure they are as accurate as they can possibly be.

Apart from human error in the initial recording of data, the main issues are likely to be missing data or data that does not make sense e.g. the data says some of our customers are 500 years old... Weird figures like this are called 'outliers'.

Most statistical computer analysis packages can deal effectively with these situations

Exercise

Have a look at the website **spss.com**. The statistical analysis packages this firm offers are among the most widely used in industry. The information on the site will give you some idea of the capabilities of the packages and how they can do things like deal with missing data.

Validity and Reliability

The two things that statistics need to have to be useful for business decisions are validity and reliability. It is also 'two ways of being wrong'!

Imagine that you want to find out what your friends think about the taste of Coca-Cola. To get this information you would not ask them to comment on the taste of tea. This is a concept called validity – making sure I am measuring what I am supposed to be measuring.

The other important concept is reliability. That's a word you would know – if you turn up for class every day people say you are 'reliable' – that is you do what you are supposed to do. In statistics the idea is the same – to be of any use we must be sure that the figures would be pretty much the same if we repeated the process several times.

So to summarise the concept:

- ▶ **validity** – the figures I am using are measuring the things I want them to measure
- ▶ **reliability** – if I went through the same measurement process a number of times I would get the pretty much same results

Sampling

In order to get our statistics absolutely correct and therefore have the best chance of making the right business decision it would be ideal if we could get every possible piece of data. For example we could ask every single buyer of our soft drink a detailed series of questions about what they think of it. The total number of buyers is called 'the population'.

Obviously this is not realistically possible for most organisations. The government can do it in the national population census because they have plenty of money and no time or resource restrictions. But even they only do it once every five years.

So we take a 'sample' – that is a group of people we believe would accurately represent the views of the 'population'.

There are a number of methods we could use to pick the best sample. Briefly they are:

- ▶ **comprehensive sample:** you try to get the whole 'population' As mentioned above this is not really possible
- ▶ **random sample:** everyone in the population has an equal chance of being selected. This is usually the best method
- ▶ **convenience sample:** we select the members of the 'population' who are easiest to find. This is not a good idea as you are not likely to select a group that truly represents everyone.
- ▶ **systematic sample:** you select your sample on an organised basis such as every tenth person. This is a technique used widely by telephone researchers who use the 'Rule of Seven' – they contact every seventh person on a list. This method can work if you are sure that everyone on the list is more or less the same.

We will be looking at sampling a bit more later.

