Learning Enhancement Team

University of East Anglia
STUDENT SUPPORT SERVICE

Steps into Statistics

Levels of Data

This guide identifies the different levels (or types) of data used in statistics and helps you assign appropriate levels to your own data.

Introduction

The basis of statistics is the collection and analysis of information in the form of data.[1] Due to the complexity of the world, data come in many different forms and describe many different things. For example you could collect information on an individual's height, eye colour and educational level. But can all data be treated in the same way? The answer is no. It is vital that types of data are identified and categorised into appropriate **levels**. Recognising the levels of data involved in a study affects how the data are presented and also governs choices of **descriptive statistics** and **statistical tests** you may wish to employ. For more information about descriptive statistics you can read the study guides: *The Mode*, *Measurements of Central Tendency*, *Measurements of Spread I: Range and Inter-Quartile Range* and *II: Variance and Standard Deviation*.

According to Stevens,[2] data can be divided into four levels (described below) and every data type can be assigned to one of them. However as areas of research, such as psychology, advance, the boundary between some levels has become blurred. You may like to think of division of data levels into those which are scientific measures or counts of some type (**ratio** and **interval** data have meaningful units assigned to them) and those which are not (**ordinal** and **nominal** data). The blurring between levels is seen between some ordinal and interval data types which sometimes may be assigned to either. You should consult your particular school of study, lecturer or academic literature if you are unsure how to categorise your data.

---

[1] Note that "data" is the plural of the noun "datum" and you must take care when writing about statistics to use the singular or plural word correctly.
[2] S.S. Stevens, Science, **103**, 667 (1946).

Ratio data

The top level of data is ratio data and has a constant, ordered scale.

Ratio data has a defined zero and so direct comparisons made between values in terms of ratios have a legitimate meaning. This means that every type of statistical test and descriptive statistic are available to ratio data. Some examples of ratio data are age, height and temperature (in kelvin, K). The absolute scale of temperature is the kelvin scale. In science 0 K is the coldest temperature possible (it is known as **absolute zero**) and so you *can* say that 10 K is twice as hot as 5 K. You can convert kelvin to the well-known Celsius scale of temperature by subtracting 273 and so 5 K is the same as -268 °C and 10 K is the same as -263 °C. It is obvious that -263 is not twice -268 and so changing scale affects the statements that can be made about data. In fact the Celsius scale is an interval data scale and is explained in more detail in the next section.

Interval data

Interval data has a constant, ordered scale but has an arbitrary zero point.

Although the difference between values is consistent throughout the scale, the arbitrary zero means that comparisons between data, in terms of ratios, are meaningless. Calculations of standard deviations and Pearson Correlation coefficients (amongst many others) are possible with this level of data. Examples of interval data are temperatures (measured in Fahrenheit or centigrade) or dates. For temperature, you may hear on a weather forecast that a certain day has a maximum temperature of 20°C and another of 10°C, it seems that one is twice as hot as the other but it is not. This is because, even though there is a temperature of 0°C, it is arbitrarily chosen (as the freezing point of water) and so does not indicate the coldest that temperature can be.

Ordinal data

When categorising data without meaningful units into groups or sets, if the *order* of the categorisation *is* important, then the level of data is ordinal.

The most common examples of ordinal data are Likert scales. These are associated with questions which begin “On a scale of 1 to 5…”. Of course it does not have to be scale of 1 to 5, it could easily be a list of ordered phrases such as “very unhappy, unhappy, neither unhappy or happy, happy, very happy”. Here responses are subjective

but ordered. There is thought to be overlap between the ordinal and interval data levels within certain subject areas. However you should take care when interpreting Likert scale results as pure numbers (in which case they become interval or ratio data) – take advice from your school or lecturer.

Another example could be if you are recording the educational level of participants in a study the data recorded could be categorised as follows: no education, up to GCSE, up to A-level and so on. Someone educated up to A-level is more highly educated than someone with no qualifications. *The order of the categories is important* and so the data is ordinal. You may label the categories with a number; 1 = Ph.D, 2 = Masters, 3 = Degree, 4 = A-Level and so on. The numbers associated with each category are not true numbers. Does it make sense to calculate the mean or the standard deviation of education level? You should use medians, ranges and inter-quartile ranges for describing ordinal data.

Nominal data

When categorising data into groups or sets if the *order* of the categorisation is *arbitrary* then the level of data is called nominal.

In the nominal data level only the *name* of the category is important. For example if you are recording data relating to individuals eye colour; you may record this data as the actual eye colour (blue, brown, hazel, green and so on) but you may also label each eye colour with a number; 1 = blue, 2 = brown, 3 = hazel and so on. Both the numbers given to each category and their order are unimportant (as blue eyes are no more or less important than brown eyes for instance) so this is nominal data regardless of whether you use the colours or the numbers. Some other examples of nominal data are gender, breed, place of birth and favourite fruit. Only the mode is an appropriate descriptive statistic for nominal data.

Discrete and continuous data

The ratio and interval levels of data are often divided again into measures which can take any value and measures that can only take certain values. If a value can take any value (such as height, weight and so on) then the data are described as **continuous**. If a value can only take certain values (such as number of children, count and so on) then the data are described as **discrete**.

Presenting levels of data

There are a variety of ways of presenting data in charts and graphs, not all are suitable for each level of data (see table below). You can find out more about charts from the study guides: *Pie Charts*, *Bar Charts*, *Box-and-Whisker Plots* and *Histograms*.

	Level of Data		
	Nominal	Ordinal	Interval/Ratio
Pie Chart	Yes	Yes	OK*
Bar Chart	Yes	Yes	OK
Histogram	No	No	Yes
Box Plot	No	Yes	Yes
Line Graph	No	OK	Yes
Area Graph	No	OK	Yes

*OK – You may use this chart-type but there are normally better alternatives.

Want to know more?

If you have any further questions about this topic you can make an appointment to see a **Learning Enhancement Tutor** in the **Student Support Service**, as well as speaking to your lecturer or adviser.

Call: 01603 592761
Ask: ask.let@uea.ac.uk
Click: https://portal.uea.ac.uk/student-support-service/learning-enhancement

There are many other resources to help you with your studies on our website.
For this topic there is a webcast.

Your comments or suggestions about our resources are very welcome.

Scan the QR-code with a smartphone app for a webcast of this study guide.