

# What kind of data do I have?

C. Gissane

*School of Sport, Health and Applied Science, St Mary's University College, Twickenham, Middlesex, TW1 4SX, UK  
E-mail: conor.gissane@smuc.ac.uk.*

One of the first concepts taught in a statistics class is the fact that there are different types or categories of data. An accurate classification of data is important for both the description of data and any hypothesis tests that may be carried out [1]. The premise is that data in a particular category has certain properties, which permit certain mathematical operations and make certain statistical tests valid. Whether or not the data properties should influence the type of statistical treatment is the subject of some dispute [2], with some claiming that certain techniques work with certain types of data [3] and others saying that it doesn't make a difference [4]. Nevertheless, the process of categorising data is a prominent feature in many statistical texts, and some statistical packages require the user to select the measurement level of the variables entered.

There are several typologies which can be used to classify data [4]. The simplest of these typologies is to distinguish between qualitative and quantitative variables. A qualitative variable is a variable that can be recorded in labelled categories [5]. For example, gender (male or female), profession (physiotherapist, doctor, nurse, dentist), or grade of disability (moderate, low, none, severe), are examples of qualitative scales. Quantitative scales are those that use true numerical quantities [5]. For example, a person's body mass, or the range of motion about a particular joint are examples of quantitative data. Many texts then seek to break down qualitative and quantitative variables into subtypes [5–8].

Perhaps the most widely used is the method proposed by Stevens [9]. He proposed that as different types of data had different properties and were therefore, at different levels of measurement; Nominal, Ordinal, Interval and Ratio, which are defined in

Table 1. It has been proposed that Stevens' levels of measurement [9] should be embraced by physical therapy [10]. Some authors [11, 12] do report levels of measurement and the Task Force on Standards for Measurement in Physical Therapy [13] requires knowledge of the four levels of measurement as a standard. Yet, some authors who write for therapists chose to curtail the list [14].

Another term used to describe the characteristics of variables is data type [6–8, 15]. A list of definitions data types and levels of measurement can be seen in Table 1. There are more definitions of data types than there are levels of measurement. Also, the definitions do overlap and some terms are inter-changeable.

In addition, when some authors report about data types or levels of measurement, their listing is not as complete as it might be. For example, Hicks [14], Fisher and Marshall [3], McCrum-Gardner [16] and Neutens [17] each define three, as does SPSSv18 [18]. Once again, they are not uniform, so can they be linked?

Each [3, 14, 16, 17] use the terms nominal and ordinal, the problems come with the definitions of their last level. The definitions Interval/Ratio [14], interval [16, 17], continuous [3], and scale [18] are all used. Two of these were not defined in Table 1, but interval/ratio is used because it has been suggested that both levels of measurement receive the same statistical analysis [5]. If this is true, the term interval is an attempt to cover both, the definition of continuous in Table 1 will also cover interval and ratio. For interval and ratio SPSS18 use the term scale, which is a synonym for continuous. It is also sometimes permissible to treat discrete data as continuous. Yes, by definition it is restricted to certain values, but the type of analysis is the same. It is something that is routinely carried out. Age is a continuous

Table 1  
Definitions of data categories/types used in the literature

Level of measurement or data type	Definition	Synonym
Nominal	Classification into unordered qualitative categories [19].	Polytomous [19], polytomy [19], categorical [6–8,19]
Ordinal	Classification into ordered qualitative categories, where the values have a distinct order but their categories are qualitative in that there is no natural (numerical) distance between their possible values [19].	Ranking scale [19]
Interval	An (equal) interval involves assignment of values with a natural distance between them, so that a particular distance between two values in one region of the scale meaningfully represents the same distance between two values in another region of the scale [19].	
Ratio	A ratio is an interval scale with a true zero point, so that the ratios between the values are meaningfully defined. With this scale a given value can be spoken of as being so many times greater than or less than another value [19].	
Dichotomous	One that arranges items into either of two mutually exclusive categories [19].	Binary [6, 19], attribute [6–8, 20]
Discrete	Data that can be arranged into a naturally or arbitrarily selected or sets of values as opposed to data in which there are no naturally occurring breaks in continuity [19].	
Continuous	Data (variable) with a potentially infinite number of possible values along a continuum [19].	Scale [18]
Ordered Nominal	Grouping subjects into several categories which have an order. For example, if there condition is much improved; improved a little; no change; a little worse; much worse [7].	

variable, but custom dictates that age last birthday is used, a discrete variable [6].

A data type that merits special considerations is dichotomous data. A variable that can be separated into only two categories is technically a nominal variable [15]. Yet, if data is recorded as 0 and 1 (absent or present) it presents opportunities for handling the data. For example, the variable could be analysed using a proportion approach which allows the presentation of confidence intervals [5, 6].

Whether a variable fits into either a level of measurement or a specific data type is always going to be debateable. Several papers have been written that argue the pros and cons of treating a likert scale as an interval level variable [21–23]. Debates like this will continue, and will involve variables and measurements other than likert scales. But a practitioner will need to be able to justify their own arguments and understanding [24].

In spite of any arguments to the contrary [2, 4], it is important to know the different ways of defining data categories. It doesn't matter if the application of either data types or levels of measurement are wrong, they need to be understood as they dictate what comes later in many texts [25]. A reader who is aware of both data

types and levels of measurement is better equipped to deal with the literature than one who isn't.

## References

- [1] Lesser LM. Using context to classify variables. *Teaching Statistics* 2012;35:29-31.
- [2] Townsend JT, Ashby FG. Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin* 1984;96(2):394-401.
- [3] Fisher MJ, Marshall AP. Understanding descriptive statistics. *Australian Critical Care* 2009;22:93-7.
- [4] Gaito J. Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin* 1980;87:564-7.
- [5] Pearson JGC, Turton A. *Statistical methods for environmental health*. London: Chapman & Hall; 1993.
- [6] Altman DG. *Practical statistical for medical students*. London: Chapman & Hall; 1991.
- [7] Bland MJ. *An introduction to medical statistics*. 3rd ed. Oxford: Oxford University Press; 1995.
- [8] Daley LE, Bourke GJ. *Interpretation and uses of medical statistics*. 5th ed. Oxford: Blackwell Science; 2007.
- [9] Stevens SS. On the theory of scales of measurement. *Science* 1946;103:677-80.
- [10] Michels E. Measurement in Physical Therapy. *Phys Ther* 1983;63(2):209-15.

- [11] Payton OD. *Research: The validation of clinical practice*. Philadelphia: FA Davis & Company; 1994.
- [12] Bork CE. *Research in physical therapy*. Philadelphia: J B Lippincott Company; 1993.
- [13] Task Force on Standards for Measurement in Physical Therapy. Standards for test and measurements in physical therapy practice. *Phys Ther* 1991;71:589-622.
- [14] Hicks CM. *Research methods for clinical therapists: Applied project design and analysis*. 5th ed. London: Churchill Livingstone; 2009.
- [15] Driscoll P, Lecky F, Crosby M. An introduction to everyday statistics-1. *Emergency Medical Journal* 2000;17:205-11.
- [16] McCrum-Gardner E. Which is the correct statistical test to use? *British Journal of Maxillofacial. Surgery* 2008;46:38-41.
- [17] Neutens J, Rubinson L. *Research techniques for health sciences*. 3rd ed. London: Benjamin Cummins; 2001.
- [18] SPSS. SPSS Version 18. Chicago IL: IBM; 2012.
- [19] Porta M, Last J, editors. *A dictionary of epidemiology*. OUP, New York; 2008.
- [20] Sullivan LM. *Essentials of biostatistics in public health*. Sunbury MA: Jones and Bartlett Learning; 2012.
- [21] Jamieson S. Likert scales: How to use and (ab)use them. *Medical Education* 2004;38:1212-8.
- [22] Knapp TR. Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research* 1990;39:121-3.
- [23] Norman G. Likert scales, levels of measurement and the 'laws' of statistics. *Advances in Health Science Education* 2010;15:625-32.
- [24] Greenhalgh P. Statistics for the non-statistician. I: Different types of data need different statistics. *BMJ* 1997;315:364-6.
- [25] Velleman PF, Wilkinson L. Nominal, ordinal, interval and ratio typologies are misleading. *The American Statistician* 1993;47:65-72.