Physiotherapy Practice and Research 34 (2013) 123–125

What kind of data do I have?

Gissane, C.

School of Sport, Health and Applied Science, St Mary's University College, Twickenham, Middlesex, TW1 4SX, UK.

gissanec@smuc.ac.uk

One of the first concepts taught in a statistics class is the fact that there are different types and 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] Perhaps the most widely used is the method proposed by Stevens.[5] 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[5] should embraced by physical therapy.[6] Some authors[7-8] do report levels of measurement and the Task Force on Standards for Measurement in Physical Therapy[9] requires knowledge of the four levels of measurement as a standard. Yet, some authors who write for therapists chose to only report three.[10]

Stevens[5] levels of measurement is one method of approaching the problem, another term used to describe the characteristics of data is data types.[11-14] A list of definitions data types and levels of measurement can be seen in table 1. The definitions do overlap and some terms are inter-changeable.

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.[15]	polytomous,[15] polytomy,[15] categorical[11 14-15]
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.[15]	Ranking scale[15]
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.[15]	
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.[15]	
Dichotomous	One that arranges items into either of two mutually exclusive categories.[15]	Binary,[11 15] attribute[11 14 16]
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.[15]	
Continuous	Data (variable) with a potentially infinite number of possible values along a continuum.[15]	Scale[17]
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.[12]	

Probably the simplest way to define data is the categorisation into Qualitative and Quantitative.

In addition, when some authors report about data type or levels of measurement, their listing is not as complete as it might be. For example, Hicks,[10] Fisher and Marshall[3] and McCrum-Gardner[18] each define three, as does SPSSv18.[17]

Once again, they are not uniform, so can they be linked? Each, use the terms nominal and ordinal, the problems come with the definitions of their last level. The definitions Interval/Ratio,[10] interval,[18] continuous,[3] and scale are all used. Two of these were not defined in table one, but interval/ratio is used because it has been suggested that both levels of measurement receive the same statistical analysis.[19] 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, provided any sign before the number is ignored. For interval and ratio SPSS[17] use the term scale which is a synonym for scale.

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.[13] Yet, if data is recorded as 0 and 1 (absent or present) it presents a host of opportunities for handling the data .

In spite of arguments to the contrary,[2 4] it is important to know the different ways of defining data types. 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.

Levels of measurement have been discussed in many healthcare texts. Yet, the definitions of categories are not uniform across these texts In common with other areas of statistics there are different ways of categorising data.

It was proposed that Stevens' levels of measurement[5] should embraced by physical therapy.[6] Some authors have,[7-8] while other major texts combine some levels.[10]

disagreed and only reported three. Beyond that, authors have used several definitions of data, under the heading of either levels of measurement or, data types (Table 1).

Levels of measurement have been discussed in many healthcare texts. Yet, the definitions of categories are not uniform across these texts In common with other areas of statistics there are

different ways of categorising data. There are times when authors curtail descriptions, and combine typologies. Some writers report three levels of measurement, the majority report four, and some five. There is also the consideration that many medical textbooks do not use the term 'level of measurement' but refer to data types. In short, there are different techniques that overlap. To understand the topic correctly, it is important to know the typologies and how they interlink.

Discrete data - 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.

Continuous data – Data (variable) with a potentially infinite number of possible values along a continuum.

Measurement scale – The range of possible values for a measurement.

Measurement scales can be classified according to the quantitative character of the scale:

1. **Dichotomous Scale** (Syn: Binary scale): One that arranges items into either of two mutually exclusive categories.
2. **Nominal Scale** (Syn, polytomous, polytomy): Classification into unordered qualitative categories.
3. **Ordinal Scale**: 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.
4. **Interval Scale**: 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.
5. **Ratio Scale**: 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.

Ranking Scale (Syn: ordinal scale) A scale that arrays the numbers of a group from high to low according to the magnitude of the observations, assigns ranks and neglects distances between the members of the array.

Porta Dictionary of epidemiology

Tertiary puweyor: any person who teaches or prepares instructional material that describes specific tests or specific uses of measurements; this definition includes, but is not limited to, persons teaching in academic institutions, clinical educators, and continuing educators who are not acting in the role of primary or secondary purveyors (see purveyor, prima ry purveyor, and secondary purveyor)

T2.4. Tertiary purveyors must understand the different levels of measurement (ie, nominal, ordinal, interval, and ratio) and the mathematical operations that are appropriate for each level

Test user: one who chooses tests, interprets test scores, or makes decisions based on test scores (this definition is from Standards for Educational and Psychological Tests. American Psychological Asociation, Washington, DC, 1974, page

U3.4. Test users must understand the different levels of measurement (ie, nominal, ordinal, interval, and ratio) and the mathematical operations that are appropriate for each

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. Stevens SS. On the theory of scales of measurement. *Science* 1946;103:677-80.
6. Michels E. Measurement in Physical Therapy. *Phys Ther* 1983;63(2):209-15.
7. Payton OD. *Research: the validation of clinical practice*. Philadelphia: FA Davis & Company, 1994.
8. Bork CE. *Research in physical therapy*. Philadelphia: J B Lippincott Company, 1993.
9. TFSMPT. Standards for test and measurements in physical therapy practice. *Physical Therapy* 1991;71:589-622.
10. Hicks CM. *Research methods for clinical therapists: applied project design and analysis*. 5th ed. London: Churchill Livingstone, 2009.
11. Altman DG. *Practical statistical for medical students*. London: Chapman & Hall, 1991.
12. Bland MJ. *An introduction to medical statistics*. 3rd ed. Oxford: Oxford Univeristy Press, 1995.
13. Driscoll P, Lecky F, Crosby M. An introduction to everyday statistics-1. *Emergency Medical Journal* 2000;17:205-11.
14. Daley LE, Bourke GJ. *Interpretation and uses of medical statistics*. 5th ed. Oxford: Blackwell Science, 2007.
15. Porta M, Last J, editors. *A dictionary of epidemiology*. New York: OUP, 2008.
16. Sullivan LM. *Essentials of biostatistics in public health*. Sunbury MA: Jones and Bartlett Learning, 2012.
17. SPSS. SPSS Version 18. Chicago IL: IBM, 2012.
18. McCrum-Gardner E. Which is the correct statistical test to use? *British Journal of Maxillofacial Surgery* 2008;46:38-41.
19. Pearson JGC, Turton A. *Statistical methods for environmental health*. London: Chapman & Hall, 1993.