Understanding statistics 2

Elliot Abt

Department of Dentistry, Illinois Masonic Medical Center, Chicago, Illinois, USA

The first article in this series addressed both qualitative and quantitative research. This article will discuss the different types and nature of data used in quantitative research.

One of the simplest ways for readers of the scientific literature to understand statistical methodology is to start with the data used in a study. The nature of the data typically determines which statistical test will be employed by the investigators. We will consider 3 types of data:

Categorical data

   A. Binary data (subset of categorical data)

Ordinal data

Quantitative data

Categorical data

As its name suggests, categorical or nominal data represents different categories or names for data with none being better or worse than the other. Common examples include hair colour, eye colour, religious preference or city of origin. We can assign different numbers to eye colour such as:

- Brown
- Blue
- Green
- Grey

With categorical data, we can only say that a "4" is different from, not better than a "2".

An important subset of categorical data is *binary*, or dichotomous data. Binary data is when only two possibilities exist, such as male/female, life/death, or disease/no disease. Readers of the dental literature are likely to encounter binary data as with success or failure of a dental restoration or prosthesis. Other examples might include the presence or absence of pain or dental pathology. Although binary data has only two possibilities, it is relevant as it allows for some important mathematical manip-

ulations which will be addressed in the next article.

Ordinal data

Ordinal, or rank data, is an extension of categorical data, as it consists of different categories of data with one important distinction: one category is *better* (or worse) than the other. For example, at our institution, we have patients with excellent, good, fair, or poor oral hygiene. We could then assign numbers to these levels of oral hygiene such as:

- Excellent
- Good
- Fair
- Poor

With ordinal data, we can now say that it is better to be a "4" than a "2". What we cannot say is that a "4" is twice as good as a "2" as this is simply a rank assignment of numbers which lacks mathematical precision.

Other examples of ordinal data include stages of disease and Likert scales. Periodontitis is often classified as mild, moderate or severe. Likert scales are commonly used for questionnaires and have categories such as "strongly agree", "agree", "undecided", "disagree" or "strongly disagree". We can say that level of agreement is stronger with one category as compared with another, yet we cannot quantify this difference precisely.

Quantitative data

Quantitative data is an extension of ordinal data with an important caveat: we are dealing with numbers which have real precision, or meaning to them. Common examples

would be height, weight, blood pressure, or age. In the dental world, we might look at periodontal pocket depth or alveolar bone height measured in millimeters. Because quantitative data deals with numbers with precise meaning, we can now say that a 4mm pocket is twice the depth of a 2mm pocket.

Quantitative data is commonly subdivided into continuous or discrete data. Continuous data, such as height or weight, can be subdivided into fractions of whole numbers, whereas discrete data consists only of whole numbers such as number of pregnancies, or number of cigarettes lit in a day.

Independent versus paired data

Another major distinction with data is whether it is independent (unmatched) or paired (matched). A study may examine the heights of two separate groups of individuals. This type of data would be *independent*, as the two groups are unrelated. Another type of study may make several observations on the same individual. For example, all patients in a study may take both drug A and drug B during different time periods. This would be *paired* data, as both measurements are taken on the same individual. A dental study may examine the effects of a periodontal intervention versus placebo on two separate groups of patients (independent data) or it may examine the effects of both intervention and placebo in different areas of the mouth on the same individual (paired data). It is worth noting that not all paired data has to involve within-person differences. For example, if we are comparing

men and women randomly, the data would be independent. However, it is likely that studies comparing siblings (brother/sister) or spouses would be paired data due to the associations within these groups.

Defining outcomes

It is possible for similar outcomes of an investigation to belong to any of the 3 data types described. For example, a periodontal study could have as its outcome:

- disease / no periodontal disease (binary data)
- mild, moderate, severe periodontitis (ordinal data)
- periodontal pocket depth in millimeters (quantitative data)

With practice, the reader of dental literature should be able to define the outcome(s) of a study as either being categorical, ordinal, or quantitative in nature. Additionally, the data should be described as being either independent or paired, which will be immensely helpful in determining which type of statistical test should be employed.

The next article in this series will define P values and confidence intervals as well as describe the different statistical tests used for different types of data.

Evidence-Based Dentistry (2010) **11**, 93-94. doi:10.1038/sj.ebd.6400746.