A considerable portion of your everyday thinking is statistical in nature. When you think about birth weights, driving times between cities, body proportions, prices at restaurants, the weather, your electric bills, or the stock market, you think in terms of averages, deviations, variability, probability, and frequency. These five fundamental concepts are the basis for most of the statistical methods used in research.

Statistics is the collection of methods and procedures for collecting, describing, summarizing, and analyzing data, and for making scientific inferences, decisions, or predictions from such data. The term "statistic" or its plural "statistics" is also used to refer to the summarizing, descriptive values calculated from the data.

Nature of data and levels of measurement

Note the prominence of the word "data" in the definition of statistics. Data result from observations and descriptions of the world and natural phenomena and from experiments in which some variables are systematically varied to determine the effects on other variables.

A variable is a quantity that may assume any one of a set of different possible values. It is the qualitative or quantitative feature of interest, i.e., it is a property of the object or process we are trying to measure or study.

Measurement scales

The values of a variable define the scale or level of measurement used to study the variable. We will consider four scales of measurement: nominal, ordinal, interval, and ratio.

Nominal scale

At the nominal level of measurement, the data consist of discrete categories or classifications. Data or variables of this type can be labeled, identified, or classified according to the presence or absence of some characteristic or attribute.

Example: The variable "color" has possible values that include red, green, blue, white, black.

Example: The variable "make of car" has possible values that include Ford, Fiat, Datsun, Dodge.
Ordinal scale

On an ordinal scale of measurement, the data are ranked, rated, or ordered. With an ordinal scale, one is able to say that one value is greater or less than another, although there is no information about how much greater than or less than.

Example: Ranking several foods in order of preference on a scale from 1 to 5.

Example: Ranking job applicants.

Example: Scoring or rating degree of gingival hyperplasia as none, slight, moderate or severe

Interval scale

An interval scale of measurement is based on ordered intervals that are of equal length, and the zero value is arbitrary (there is no true zero). Addition or subtraction is possible on this scale, but not multiplication or division.

Example: Centigrade and Fahrenheit scales for measuring temperature. Zero degrees is not the same on these scales and there is temperature present at the zero point on both these scales. On each scale, 40 degrees is 30 degrees warmer than 10 degrees, but it is not four times as warm as 10 degrees.

Ratio scale

When a measurement scale has ordered intervals of equal length and a true zero point, it is a ratio scale. We can add, subtract, or multiply any two values of such a variable and get a meaningful answer. For a continuous ratio variable, we also can get a meaningful answer when we divide two values, but not for a discrete ratio variable (e.g., 4/3 = 1.333, not an integer).

Example: Length is a ratio variable. There is a true zero length, and it makes sense to say that 4 feet is twice as much as 2 feet.
Example: Periodontal pocket depth in mm is a ratio variable.

Continuous vs. Discrete variables

A continuous variable has an infinite number of possible values and its basic unit of measurement can be subdivided ad infinitum.

Example: The meter is a unit of length; it can be subdivided into centimeters, millimeters, Angstroms, etc.

A discrete variable has a limited or countable number of values and the basic unit of measurement cannot be meaningfully subdivided.

Example: The number of dental students in a class is a discrete value with possible values 0, 1, 2, etc. It makes no sense to say 14.3 students.
Qualitative vs. Quantitative variables

Qualitative variables have only a discrete number of values which are categories of some attribute. Qualitative variables are measured by nominal and ordinal scales. Many dental indices, e.g., plaque index (with scores of 0, 1, 2, or 3) are qualitative variables.

Quantitative variables always have numerical values, may be either discrete or continuous, and are measured by interval and ratio scales. Examples of quantitative variables are periodontal pocket depth (mm), salivary amylase (µg/mg protein).

Special note on rating scales, preferences, and rankings

Many types of preference tests, surveys, and other studies employ rating scales or ranking for preferences. For example, rate your taste preferences for several kinds of vegetables on a scale from 1 to 5, with 5 denoting the most preferred. With a rating scale or ranking, a number may be written as the result of a single trial, but this number is no better than a word. Any variable whose values are ratings or ranks is still a qualitative variable, because only an ordering is implied by the rating or ranking. Intervals and ratios have no meaning at this measurement level. For example, if you rate black-eye peas as 5 and okra as 2, can you say that you like black-eye peas 2.5 times as much as okra? Or is it sensible to say that the difference between black-eye peas and okra is 3?

The dental research literature contains numerous examples of rankings, ratings, scores, and indices, all based on ordinal scales of measurement. For example, the figure to the right (from C.A. Full, et al., JADA, 94, 111-113, 1977), Criteria for scoring Plaque, illustrates a variable, "Plaque score", which has the possible values 0, 1, 2, 3. Note that in assigning one of these values to the variable based on observations of a given tooth, the dentist is subjectively deciding which of these labels best fits that particular tooth. Also note that it is not meaningful to say that 3 - 1 = 2, since the real difference in plaque surface between the scores 3 and 1 may not be equivalent to the score of 2, i.e., the intervals between these values are not necessarily equal.

Criteria for Scoring Plaque

<table>
<thead>
<tr>
<th></th>
<th>Criteria for Scoring Plaque</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No Plaque</td>
</tr>
<tr>
<td>1</td>
<td>Distinct region of stained plaque at gingival margin</td>
</tr>
<tr>
<td>2</td>
<td>Stained plaque covering one-third of the tooth surface</td>
</tr>
<tr>
<td>3</td>
<td>Stained plaque covering two-thirds or more of the tooth surface</td>
</tr>
</tbody>
</table>
In spite of the fact that such limitations exist with ordinal data they are frequently ignored and many investigators apply various types of parametric (defined below) statistical techniques to preference scales and other subjectively-ranked data. Such procedures are hazardous in that these statistical tests are based on the assumption that the measurement scale is based on equal intervals between variable values, and this is unlikely to hold or even to be verifiable for most ordinal variables.

How to distinguish between qualitative and quantitative variables

Given a discrete number as a value of a variable, how do you decide whether the variable is qualitative or quantitative? Rule 1: If the value recorded on a single trial or as the result of a single measurement is a number, and adding or subtracting two such values yields another meaningful value, the variable is probably quantitative; otherwise it is qualitative. Rule 2: In cases where you aren't certain, ask yourself whether words would work as well as numerical values. Would "bad", "poor", "neutral", "good", and "best" do as well as the numerical ranks of 1, 2, 3, 4, and 5? If words are as suitable as numbers, the variable is qualitative.

Parametric vs. Nonparametric Statistical Tests

Parametric statistical tests are related to certain theoretical distributions of variable values, which are characterized by parameters. The assumptions one must make about the data before using parametric tests usually limit their valid applicability to ratio and interval data.

Nonparametric or distribution-free tests are not related to any particular distribution and are thus applicable to more types of data, although they are most frequently used with nominal and ordinal data.
<table>
<thead>
<tr>
<th>Scale</th>
<th>Definition</th>
<th>Qualitative (Q₁)</th>
<th>Quantitative (Qₜ)</th>
<th>Discrete (D)</th>
<th>Continuous (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Variable values are categories, names, or labels. Numerical values are only labels (e.g., phone numbers).</td>
<td>Q₁</td>
<td></td>
<td></td>
<td>D</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Variable values are ordered or ranked, but can't measure the distance between any 2 values. Can only say greater than or less than.</td>
<td>Q₁</td>
<td></td>
<td></td>
<td>D</td>
</tr>
<tr>
<td>Interval</td>
<td>Based on ordered intervals of equal length, numerical variable values. There is no true zero point.</td>
<td>Qₜ</td>
<td></td>
<td></td>
<td>D,C</td>
</tr>
<tr>
<td>Ratio</td>
<td>Has ordered intervals of equal length, numerical variable values and true zero point.</td>
<td>Qₜ</td>
<td></td>
<td></td>
<td>D,C</td>
</tr>
</tbody>
</table>