PHReNet Primary Health Care Research Network
Research Bites
THE UNIVERSITY OF NEW SOUTH WALES SYDNEY · 2052 · AUSTRALIA

Issue 7, August 2003

Quantitative data analysis

This issue of ***Research Bites*** looks at quantitative data analysis.

Useful resources

- UNSW's NSW Primary Health Care Research Capacity Building Program, 2001, *Understanding data and statistics for division programs: a professional development workbook*.
- Rice Virtual Lab in Statistics http://www.ruf.rice.edu/~lane/rvls.html
- Peat J, 2001, *Health Science Research: a handbook of quantitative methods*, Allen & Unwin, Sydney.

Type of variables

An item of data that can be observed or measured is called a *variable*. There are two main types of variables.

Numerical variables can be:

- *Discrete variables* - values that are separate and distinct e.g. number of GP visits; or
- *Continuous variables* - when all values are possible e.g. blood pressure, weight.

Categorical variables represent membership of a particular category. They can be:

- *Ordinal variables* - several categories where order is relevant e.g. physical activity measured as minimal, moderate or vigorous;
- *Nominal variables* - no natural order e.g. area of residence; or
- *Dichotomous variables* - only two responses e.g. yes/no.

Sometimes variables are converted for analysis e.g. age (numerical) to age groups (categorical ordinal).

Descriptive statistics

These provide basic summaries of individual observations or measures in a sample. The statistical technique used depends on the type of variable.

In summarising numerical data, some common measures are:

- *Mean* - sum of all individual counts or measures divided by the number of individuals;
- *Mode* - most frequently occurring count or measure across a group of individuals;
- *Median* - middle observation in a sample of individuals;
- *Range* - difference between the maximum and minimum observations; and
- *Standard deviation* - a measure of how much the individual data tend to deviate from the mean.

To describe the relationship between numerical variables, a common test used is the *correlation coefficient, r.* Correlations range between –1 and +1 and describe the nature and degree of association between two variables.

It is important to remember that correlations are not concerned with causality. An additional factor may underlie both variables.

In summarising categorical data, *counts* are used. They can also be expressed as *proportions* or *percentages* by dividing the count by the total number of individuals. While categorical variables may be coded using numbers, it is important not to summarise them as numerical data e.g. do not average coded numbers in Likert scales as below:

Very dissatisfied				Very satisfied
1	2	3	4	5

The relationship between categorical variables is usually presented in a contingency table and tested using *chi-square test*.

Statistical tests

There are many statistical tests available. Several authors have published flowcharts for selecting statistical tests (Peat, 2001, pp. 197-199). Two of the most common tests are the *t-test* for numerical data and the *chi-square test* for categorical data.

Inferential statistics

In inferential statistics you are trying to reach conclusions about a population based on a sample of individuals from the population.

In hypothesis testing, statistical methods are used to determine the probability of obtaining the observed effect by chance. The *p*-value of your chosen statistical test is compared to the *level of significance* (usually set at 0.05 or 0.01). For example, a *t-test* with a *p*-value of 0.03 (and *level of significance* of 0.05) indicates that the results are not due to chance and are *statistically significant.*

Confidence intervals can also be used. Based on a sample, they are estimations of a range of values (confidence interval) within which the population parameter is likely to lie. With a *level of significance* set at 95%, you can say that based on your sample, you are 95% confident that the population value lies within your confidence interval.

This issue was prepared by Anita Fletcher,PHReNet South, e-mail: afletcher@idgp.org.au
For further information about, PHReNet contact Suzan Mehmet Ph: 9616 8520,
e-mail: Suzan.Mehmet@swsahs.nsw.gov.au

Next issue: Qualitative data analysis

An initiative of the University of New South Wales as part of the NSW Primary Health Care Research Capacity Building Program, funded by the Commonwealth Department of Health and Ageing