

STATISTICAL BRIEFING: WHICH TEST?

CHRISTOPHER R. LAMB

Which statistical test to use is one of the most common questions asked of statisticians. When we are aiming to compare two groups (or compare one group to reference value), the answer to this question depends on the source and type of data.

Sources of data include independent groups and related groups. Examples of independent data groups include those obtained from males and females or from patients randomly assigned to different treatments. Examples of related (paired) data groups include those obtained from matched subjects (e.g., identical twins) or the same subjects under different conditions, such as measurements made before and after treatment. Types of data include categorical, ordinal, and continuous.¹ For the purposes of most statistical tests, continuous data are divided into those that have a Normal distribution² and those that do not. Table 1 summarizes appropriate statistical tests for various combinations of data source and type.

Continuous data with a normal distribution should be tested using parametric methods, such as the *t*-test.

Even if the data do not appear normally distributed, or there are marked differences in the variance of two groups of data, it may be possible to transform the data by taking the logarithm, square root or reciprocal in such a way as to make the data fit a normal distribution and correct any differences in variance.³ To change the scale by transformation is not cheating. Rather it is a legitimate way of enabling use of parametric tests, which are preferred because they provide results with higher levels of significance than nonparametric tests for the same sample size. When suitable transformation is not possible, the data should be treated like ordinal data and tested using nonparametric methods, such as the Mann–Whitney test.

Of the various mistakes that can occur when selecting a statistical test, it appears that applying a parametric test to ordinal data is one of the most common.^{4,5} This mistake can potentially lead to the erroneous conclusion that a significant difference exists (type 1 error), although that outcome is infrequent.⁵

TABLE 1. Appropriate Uses of Selected Statistical Tests

	Type of Data		
	Nominal	Ordinal Not normally distributed continuous	Normally Distributed Continuous
Aim			
Compare one group with a hypothetical value	χ^2 test	Sign test Wilcoxon signed rank test	One-sample <i>t</i> -test <i>z</i> -test
Compare two independent groups	χ^2 test Fisher's exact test	Mann–Whitney test	Unpaired <i>t</i> -test
Compare two paired groups	McNemar's test	Sign test Wilcoxon matched pairs signed rank sum test	Paired <i>t</i> -test

REFERENCES

- Lamb CR. Statistical briefing: measurement scales. *Vet Radiol Ultrasound* 2008;49:406–407.
- Lamb CR. Statistical briefing: the normal distribution. *Vet Radiol Ultrasound* 2008;49:492–493.
- Bland JM, Altman DG. Statistics notes: transforming data. *BMJ* 1996;312:770.
- Kuzon WM Jr., Urbanchek MG, McCabe S. The seven deadly sins of statistical analysis. *Ann Plast Surg* 1996;37:265–272.
- Plant JD, Giovanini JN, Villarroel A. Frequency of appropriate and inappropriate presentation and analysis methods of ordered categorical data in the veterinary dermatology literature from January 2003 to June 2006. *Vet Dermatol* 2007;18:260–266.

From the Department of Veterinary Clinical Sciences, The Royal Veterinary College, Hawkshead Lane, North Mymms, Hertfordshire, AL9 7TA, UK
Address correspondence and reprint requests to Christopher R. Lamb, at the above address. E-mail: clamb@rvc.ac.uk
doi: 10.1111/j.1740-8261.2009.01565.x