

# Understanding statistics 3

**Elliot Abt**

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

The last article in this series described the different types and nature of data used in quantitative research. This article will define and examine the relationships between P-values, confidence intervals and sample size.

## P-values

Central to the theme of hypothesis testing is the P-value, which is often misunderstood or misinterpreted. While it is often stated as the probability of the data having arisen by chance, the P-value is actually the probability of obtaining the observed effect (or a more extreme value), given the null hypothesis is true.<sup>1</sup> Traditionally, the level of significance is set at 5% (0.05), as at this level we are no longer 'comfortable' that an observed effect is due to chance, but this is arbitrary, and it is occasionally set at 1% or even 10%.

If  $P < 0.05$ , this tells us that either two groups are different, or perhaps that there is a 1 in 20 chance we are making an error. This is referred to as a type I error, or a 'false positive', where we falsely reject the null hypothesis. Conversely, if  $P > 0.05$ , either there are no differences between groups, or we have made a type II error, or falsely accepted the null hypothesis (table 1). This is often due to a study which is underpowered, or lacking enough participants to show a difference between groups if one actually exists. One could question the ethics of conducting a trial which does not have enough subjects to show differences (or associations) between groups. In studies with excessive numbers of participants, very small differences between groups (around 1%) are likely to be statistically significant, although not necessarily clinically relevant. Thus, readers of the literature should look for a power (sample size) calculation to estimate an appropriate number of study participants.

Many practitioners have been trained to see if the P-value is greater or less than 0.05. Technically, P-values of 0.049 and 0.051 would be on opposite sides of statistical significance, when, in fact, they are equivalent, and some

Table 1 Type I and Type II errors		
	Null Hypothesis ( $H_0$ ) is true	Alternative Hypothesis ( $H_1$ ) is true
Fail to Reject Null Hypothesis	Right decision	Wrong decision Type II Error False Negative
Reject Null Hypothesis	Wrong decision Type I Error False Positive	Right decision

have argued against this all or none phenomenon of statistical significance.<sup>2</sup> At the very least, a study should report a specific P-value (i.e.  $p=0.03$ ) as this will provide a better estimate of the strength of the evidence against the null hypothesis. Unfortunately, what P-values don't tell us is the direction of the effect (is drug A better than B, or vice versa) and little about the magnitude of the effect. For that, confidence intervals (CIs) are required.

## Intervals and sample size

Many are familiar with *descriptive* statistics; the presentation, organization, and summarization of data such as simple percentages. *Inferential* statistics involve extrapolating data from a sample to a population, and CIs give readers a range of values where we are confident the true population lies.<sup>3</sup> CIs are derived from the standard error (SE), which, in turn is derived from the sample size. That is, the larger the study sample, the smaller the SE, and the narrower the CI. This may be somewhat intuitive, that the more participants, the more confident we are in the study results. Thus, as sample size increases, CIs narrow, providing a more precise estimate of effect. The relationship between sample size and CIs can be illustrated in the following two (basic) equations:

$$SE = s \div \sqrt{n}$$

$$95\% \text{ CI} = \bar{X} \pm 1.96(SE)$$

SE = standard error - which measures the amount of variability in the sample mean and indicates how accurately the sample mean represents the mean of a population  
 $s$  = standard deviation - describes the

variability in a sample, and generally does not change as the sample size increases  
 $n$  = sample size  
 CI = confidence interval  
 $\bar{X}$  = mean or point estimate

This is the equation for a 95% CI which corresponds to setting the level of statistical significance at 5%. Other CIs can be calculated for different levels of statistical significance. These are basic equations which change with increasing complexity of statistical tests.

## P-values, confidence intervals, and the null value

A relationship exists between P-values, confidence intervals and the null value. If the 95% CI includes the null value, by definition, then  $P > 0.05$ . If the 95% CI does not include the null value, then  $P < 0.05$ . It should be noted that the null value is not always "0". It varies depending on what type of data are used in a study.

The next article in this series will begin to explore basic statistical tests and give examples of dental studies using different data types. It will provide assistance with statistical analysis, including interpretation of CIs and their effect on clinical relevance.

1. Altman DG. Practical Statistics for Medical Research. Section 8.5; Chapman and Hall/CRC.
2. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, and Walter S. Basic statistics for clinicians: 1. Hypothesis testing. CMAJ 1995;152: 27-32.
3. Kirkwood BR, and Sterne JAC. Essential Medical Statistics. 2nd Edition. Ch. 8; Blackwell Science

*Evidence-Based Dentistry* (2010) **11**, 118.  
 doi:10.1038/sj.ebd.6400761