Overview of Statistics used in Dentistry

[1]Dr. Ramya Shenoy, [2]Dr.Harsh Priya

ABSTRACT

The present paper introduces the general philosophy behind hypothesis testing, importance of P values and confidence intervals with the use of published article. This article provides a brief overview of the type of statistical tests that available to analyze research data and the most commonly used statistical tests are presented in detail.

Key words: Hypothesis, P value, Confidence interval, Statistical Tests

INTRODUCTION

This article introduces the concept of testing for statistical significance along with a brief overview of some of the statistical techniques that are available to analyze dental research data.

Hypothesis testing

Scientific study is frequently based around the concept of testing hypotheses. A ***hypothesis*** is a theory or statement of belief about the population of interest (e.g. that there is a difference in the mean caries experience between all five-year-old children living in Urban and Rural area). In order to see how likely it is that the theory is true, we use the sample data to test what is called the ***null hypothesis***, i.e no difference between caries experience of all five-year-old children living in Urban and Rural area. So, instead of trying to show that the caries prevalence of 5-year-olds in Urban and Rural is different, we see whether we have enough evidence in the sample to disprove the null hypothesis that there is no difference in caries prevalence in this population of children. However, there will always be a degree of uncertainty associated with the inferences we draw. If we had examined a different sample of children from urban and rural area we might have come to a different conclusion. To overcome this and to decide whether we have enough evidence to reject the null hypothesis is by considering what is called the ***P*-value. This is a probability and its lies between 0 and 1. An event cannot occur if** $P=0$ and it must occur if $P=1$. The *P*-value represents the probability of getting the observed results (or more extreme results) if the null hypothesis is true. A small *P*-value indicates that the results in the sample would be unlikely to arise if the null hypothesis were true; this implies that the null hypothesis is not true, and there is evidence to reject it. A P- value of 0.05 implies that 1 in 20 times we will incorrectly say there is a difference between the groups when in fact there is no difference between the groups. This is known as type 1 error. Usually a cut-off value of $P<0.05$ is selected, so that any value of $P<0.05$ is called “small” and leads to rejection of the null hypothesis. If we reject the null hypothesis, we describe the result as **statistically significant**.[1] However; the P value measures the strength of evidence against the null hypothesis. But the interpretation of P value is not always straight forward. P values do not give any indication to the clinical importance of an observed effect. For example, suppose a new drug for lowering blood sugar is tested against standard treatment, and resulting P value is extremely small. This indicates that the difference is unlikely due to chance, but decision to prescribe new drug depends on side-effects, cost, contra-indications and so on. In particular, just as small study may fail to detect a genuine effect ,a very large study may result in a very small P value based on a small difference of effect that is unlikely to be important when translated into clinical practice.[2]

P values and Confidence intervals

Although P values provide a measure of the strength of association, there is great deal of additional information to be obtained from confidence intervals. Confidence interval gives a range of values within population value lies. Consider P values and confidence intervals given in Table 1. The odds ratio for the Chiche study is 0.33, suggesting that effect of intravenous nitrate is to reduce mortality by two thirds. However confidence interval indicates that the true effect is likely to be between a reduction of 91% and an increase of 13%.The results from the study shows that there may be a substantial reduction in mortality due to intravenous nitrate, but equally it

[1]*Reader,* [2]*Assistant Professor, Dept. Public Health Dentistry, Manipal College of Dental Sciences, Mangalore*

is not possible to rule out an important increase in mortality. Then it would be extremely dangerous to administer intravenous nitrate to patients with AMI. So it is better to present exact P value along with confidence intervals while writing the report.

Table 1. Results of clinical trial of intravenous nitrates in acute myocardial infarction patients (Chiche) [2]

Intravenous nitrate	Control	Odds ratio	95% confidence interval	P value
3/50	8/45	0.33	(0.09.1.13)	0.08

Parametric and Non-parametric Tests

Large set of research data often follow a particular distribution e.g. the normal distribution. Many statistical techniques rely on these theoretical distributions in order to work correctly. Statisticians describe these techniques as **parametric**. **Non-parametric** (also called distribution-free) alternatives exist that can be used for data that do not follow normal distributions. The non-parametric techniques are usually slightly less powerful (i.e. less able to detect a true difference when it exists) and more limited in their scope.[2, 3]

Paired and non-paired tests

It is also necessary to consider whether the data that we have collected arise from independent observations – unpaired tests or pre and post observations – paired observations.[1, 2, 4]

Some commonly encountered statistical techniques

Parametric Tests

Two sample (unpaired) Student *t*-test

Use: To test for equality of means in two groups of data which are independent (i.e. unpaired).

Example: "Is there any evidence that the mean dmft of 5- year old children is different in urban and rural India?"

Null hypothesis to be tested: "There is no difference in the mean dmft of 5- year old children is different in urban and rural India"

Assumptions

1. The two-sample *t*-test is based on the assumption that each set of observations is sampled from a population with a Normal distribution. The test is able to cope with some degree of departure from Normality, but it becomes less reliable as the distribution of the data becomes more non-Normal. An even more important assumption is that the variances are similar in the two groups. Variance is a measure of the spread or dispersion of the data.
2. The *t*-test is usually inappropriate for the analysis of data that are not quantitative (numerical). A common mistake is to treat an index as being quantitative when it is in fact categorical. An example would be the dental health component of the index of orthodontic treatment need (IOTN). We know that a score of '5' on the scale is worse than a score of '3', but we cannot say by how much. Similarly, we could not say that two people with scores of '3' and '5' had the same overall treatment need as two people who both had a score of '4'.

Non-parametric tests

Mann-Whitney U test

Use: The tests are non-parametric equivalents of the unpaired *t*-test. They can be viewed as tests for equality of medians rather than equality of means. They are used when the assumptions of the unpaired *t*-test are not fulfilled.

Example: "Is there any evidence that patients who attend the dentist at least annually are more/less satisfied with their dental care than patients who attend less often (where satisfaction is scored on a 5 point scale such that:

1 = completely dissatisfied

2 = moderately dissatisfied

3 = indifferent

4 = moderately satisfied

5 = completely satisfied.

Null hypothesis: "There is no difference in the median level of satisfaction in the population of patients who attend the dentist at least annually and in the population of patients who attend less often".

Assumptions

This test is based on the assumption that each set of observations is sampled from a population with a Non-normal distribution.

A non-parametric test equivalent to the paired t-test is the Wilcoxon Signed Rank test.

Tests of significance for different situations

Comparisons	Hypothesis tested	Parametrictes	Hypothesis tested	Non parametric tests
Single group	Sample Mean not different from population mean	One sample test	Sample median not different from the population median	Sign test
Two independent samples	The two population means are equal	Unpaired t test	Two population medians are equal	Mann Whitney test
Two related samples or paired sample	Mean difference is zero	Paired t test	Median difference is zero	Wilcoxon's signed rank test
Three or more samples	All the populationmeans are equal	ANOVA	All the population medians are equal	Kruskal Wallis test
Relation between two continuous variables	For normal data Correlation coefficient equals zero	Persons's Correlation coefficient	For non normal data of or dinal data	Spearman's correlation coefficient.

Chi-square test

Use: Chi-squared test is often used to compare the association between two or more qualitative data, unlike the unpaired *t*-test which compares two means. The test is used in the analysis of data when each individual belongs to one of a number of categories of a variable. The test is most commonly used on data that are presented in a 2×2 table of frequencies (a table with two rows representing the categories of one variable, two columns representing the categories of the other variable, and four 'cells' or entries) to determine if there is any evidence of a difference between two proportions, or equivalently to investigate an association between the two categorical variables.

Example: "When asking children whether they are frightened of the dentist (Yes/No) is there any evidence of a difference in the proportions responding 'Yes' between the genders (Male/Female)?"

An example 2x2 table of frequencies is shown below:

Null hypothesis: "The proportions of boys and girls who are frightened of the dentist are the same in the population"

Assumptions:

1. The Chi-square test does not make distributional assumptions about the data.
2. The variables that define the rows and columns in the table must be categorical.
3. The 2×2 test compares the frequencies observed in a sample with those that would be expected if the null hypothesis were true. The test does not work well if the expected frequencies are very small. In such cases a *Fisher's exact test or Yates correction* will be more appropriate .
4. The test may be extended to analyse frequencies in a tables with any number of rows and columns, but the test becomes less informative as the number of categories increases.
5. A different test (*McNemar's test*) is required if the data are paired.

Limitation of the Chi-square test

This test tells about presence or absence of an association between two events but does not measure the strength of association.

Advanced methods

The basic statistical tests described in this leaflet can only be used for relatively uncomplicated comparisons. More complex techniques often require the help of a statistician since there are many 'perils and pitfalls' that may befall the unwary investigator.[6]

Acknowledgements

Authors thank staff of Department of Statistics, Manipal University.

REFERENCES

1. Petrie A,Bulman JS,Osborn JF.Further statistics in dentistry Part 1: Research designs 1. British Dental Journal 193 **(7)**:377-380.
2. WhitleybE,Ball J.Statistics review 1 : Hypothesis testing and P values. Crit Care ; **6**(3): 66-71.
3. Bulman JS, Osborn JF. Statistics in dentistry. London; British Dental Association, 1989.
4. Bulman JS, Osborn JF. Statistics in dentistry. London; British Dental Association, 1989.
5. Kirkwood BR. Essentials of Medical Statistics. Oxford, UK:Black-well Science Ltd,1988.
6. Petrie A, Bulman JS, Osborn JF. *Further Statistics in Dentistry*. London; British Dental Association, 2002.