

How to select appropriate statistical test?

Selection of appropriate statistical test is very important for analysis of research data. Use of wrong or inappropriate statistical test is a common phenomenon observed in articles published in biomedical journals.^[1-4] Wrong statistical tests can be seen in many conditions like use of paired test for unpaired data or use of parametric statistical tests for the data which does not follow the normal distribution or incompatibility of statistical tests with the type of data, etc.^[5] Because of the availability of different types of statistical software, performing the statistical tests become easy, but selection of appropriate statistical test is still a problem.

Selection of appropriate statistical tests depends on the following three things:

- What kind of data we are dealing with?
- Whether our data follow the normal distribution or not?
- What is the aim of the study?

WHAT KIND OF DATA WE ARE DEALING WITH?

Usually our research data fall in one out of the following four types of data: nominal data, ordinal data, interval data, and ratio data.

Nominal data

In these kinds of data, observations are given a particular name. Like a person is observed to be 'male' or 'female' or Name of drug as generic or brand, etc. Nominal data cannot be measured or ordered but can be counted. These types of data are considered as categorical data but the order of the categories is meaningless. Data that consist of two classes like male/female or dead/alive are called binomial data, and those that consist of more than two classes like tablet/capsule/syrup are known as multinomial data. Data of these types are usually presented in the form of contingency tables like 2×2 tables.

Ordinal data

Ordinal data is also a type of categorical data but in this, categories are ordered logically. These data can be ranked in order of magnitude. One can say definitely that one measurement is equal to, less than, or greater than another. Most of the scores and scales used in research fall under

the ordinal data. For example, rating score/scale for the color, taste, smell, ease of application of products, etc.

Interval data

Interval data has a meaningful order and also has the quality that equal intervals between measurements represent equal changes in the quantity of whatever is being measured. But these types of data have no natural zero. Example is Celsius scale of temperature. In the Celsius scale, there is no natural zero, so we cannot say that 70°C is double than 35°C . In interval scale, zero point can be chosen arbitrarily. IQ Test is also interval data as it has no natural zero.

Ratio data

Ratio data has all the qualities of interval data (natural order, equal intervals) plus a natural zero point. This type of data is observed to be used most frequently.^[4] Example of ratio data is height, weight, length, etc. In this type of data, it can be said meaningfully that 10 m of length is double than 5 m. This ratio hold true regardless of which scale the object is being measured in (e.g., meters or yards). Reason for this is the presence of natural zero.

WHETHER OUR DATA FOLLOW THE NORMAL DISTRIBUTION OR NOT?

This is the second prerequisite for selection of appropriate statistical test. If you know the type of data (nominal, ordinal, interval, and ratio) and distribution of data (normal distribution or not normal distribution), selection of statistical test will be very easy. There is no need to check distribution in the case of ordinal and nominal data. Distribution should only be checked in the case of ratio and interval data. If your data are following the normal distribution, parametric statistical test should be used and nonparametric tests should only be used when normal distribution is not followed.

There are various methods for checking the normal distribution, some of them are plotting histogram, plotting box and whisker plot, plotting Q-Q plot, measuring skewness and kurtosis, using formal statistical test for normality (Kolmogorov-Smirnov test, Shapiro-Wilk test, etc). Formal statistical tests like Kolmogorov-Smirnov and Shapiro-Wilk are used frequently to check the distribution

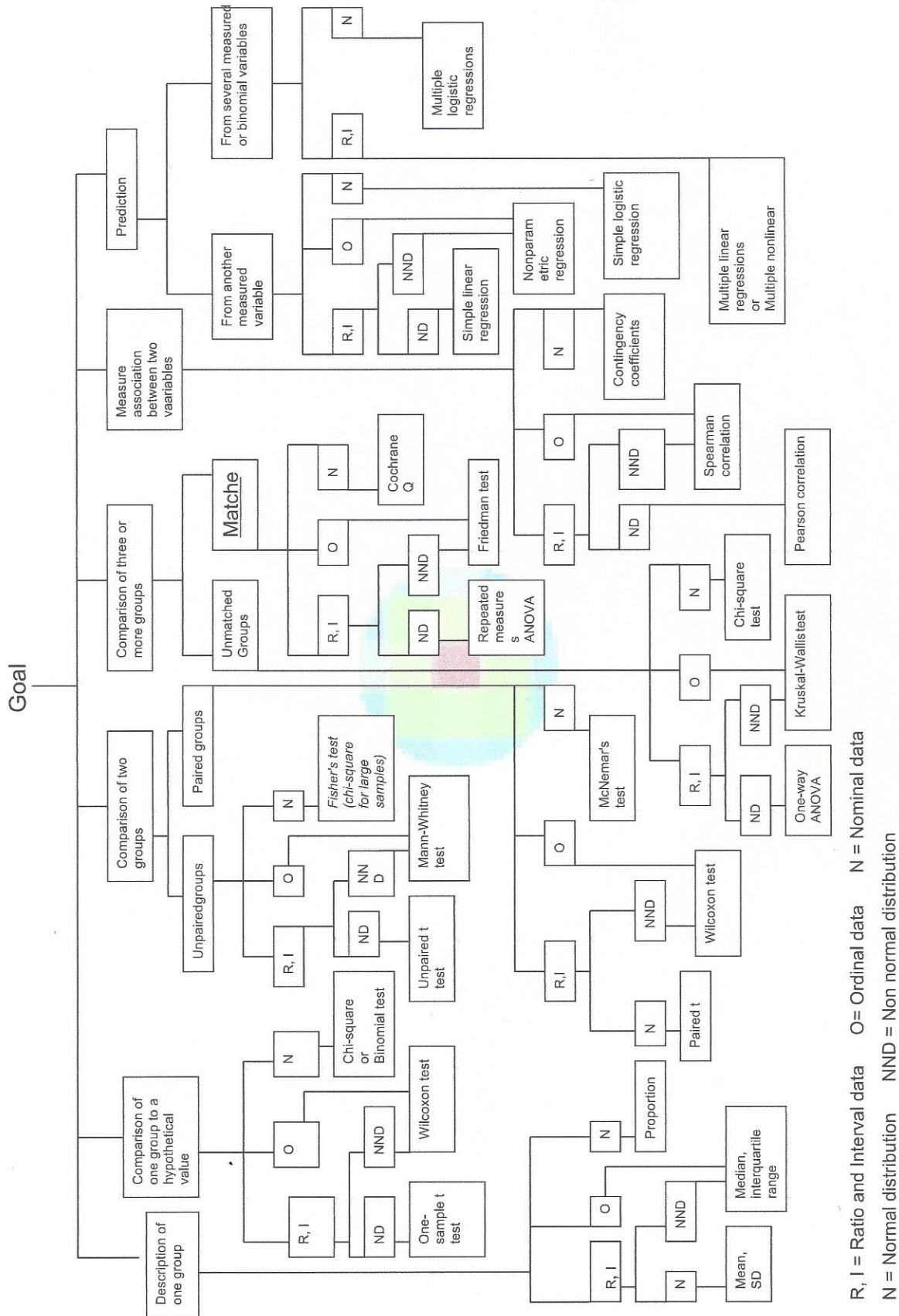


Figure 1: Flow chart for selection of appropriate statistical tests

Technical Notes

of data. All these tests are based on null hypothesis that data are taken from the population which follows the normal distribution. *P* value is determined to see the alpha error. If *P* value is less than 0.05, data is not following the normal distribution and nonparametric test should be used in that kind of data. If the sample size is less, chances of non-normal distribution are increased.

WHAT IS THE AIM OF THE STUDY?

This is the third prerequisite of selection of appropriate statistical test. What we want to compare? Whether we want to compare the drug with placebo? Or we want to compare effect of intervention by comparing preintervention endpoints with postintervention endpoints.

If a researcher is clear about all the three questions mentioned in previous text, appropriate statistical test can be selected from the flowchart [Figure 1]; this flow chart is modified from the table given on a site of software graphpad (www.graphpad.com).

To understand it better, an example has been discussed below.

Incidence of tumors observed in control and treated mice in a preclinical study is as follows:

- Controls - 1 of 14 animals
- Treated - 6 of 14 animals

Use appropriate statistical test to determine if the incidence is significantly different in two groups.

What is the aim of the study?

Our aim is to compare incidence of tumor between two unpaired groups (control *vs* treated).

What kind of data we are dealing with?

In the present study, the data type is nominal data.

Whether our data follow the normal distribution or not?

There is no need to check distribution in nominal data. They follow chi-square distribution.

So by looking at the table, we can say that most appropriate test in this condition will be 'Fishers Test' (Chi-Square if large sample size).

Jaykaran

Department of Pharmacology, Government Medical College, Surat, Gujarat, India


Address for correspondence:

Dr. Jaykaran, Department of Pharmacology, Government Medical College, Surat, Gujarat, India. E-mail: drjaykaran@yahoo.co.in

REFERENCES

1. Lang T. Twenty statistical errors even you can find in biomedical research articles. *Croat Med J* 2004;45:361-70.
2. Jaykaran, Yadav P, Chavda N, Kantharia ND. Some issues related to the reporting of statistics in clinical trials published in Indian medical journals: A survey. *Int J Pharmacol* 2010;6:354-9.
3. Karan J, Kantharia ND, Yadav P, Bhardwaj P. Reporting statistics in clinical trials published in Indian journals: A survey. *Pak J Med Sci* 2010;26:212-6.
4. Karan J, Goyal JP, Bhardwaj P, Yadav P. Statistical reporting in Indian Pediatrics. *Indian Pediatr* 2009;46:811-2.
5. Strasak AM, Zaman Q, Karl PP, Gobel G, Ulmer H. Statistical errors in medical research-a review of common pitfalls. *Swiss MED Wkly* 2007;137:44-9.

Access this article online

Quick Response Code: 	Website: www.pnrjournal.com
	DOI: 10.4103/0976-9234.75708

Staying in touch with the journal

1) Table of Contents (TOC) email alert

Receive an email alert containing the TOC when a new complete issue of the journal is made available online. To register for TOC alerts go to www.pnrjournal.com/signup.asp.

2) RSS feeds

Really Simple Syndication (RSS) helps you to get alerts on new publication right on your desktop without going to the journal's website. You need a software (e.g. RSSReader, Feed Demon, FeedReader, My Yahoo!, NewsGator and NewzCrawler) to get advantage of this tool. RSS feeds can also be read through FireFox or Microsoft Outlook 2007. Once any of these small (and mostly free) software is installed, add www.pnrjournal.com/rssfeed.asp as one of the feeds.