

How to choose and interpret a statistical test? An update for budding researchers

Ahmad Najmi, Balakrishnan Sadasivam, Avik Ray

Department of Pharmacology, All India Institute of Medical Sciences Bhopal, Bhopal, Madhya Pradesh, India

ABSTRACT

Postgraduate medical students are often not able to select and interpret the findings of statistical tests during their thesis or research projects. To go ahead with selection of tests to be performed, researchers need to determine the objectives of study, types of variables, analysis and the study design, number of groups and data sets, and the types of distribution. In this review, we summarize and explain various statistical tests to help postgraduate medical students to select the most appropriate techniques for their thesis and dissertation.

Keywords: Clinical research, biostatistics, primary care physicians, statistical test

Introduction

Postgraduate medical students are often confused in the selection and interpretation of statistical tests during their thesis or research projects. Selection of statistical test is not a rocket science and it is based on some assumptions. We require some basic information for selection of appropriate statistical test such as objectives of the study, type of variables, type of analysis, type of study design, number of groups and data sets, and the type of distribution. In the present article, we will discuss about selection and interpretation of statistical tests.

Types of statistical test

Statistical tests can be broadly classified as parametric^[1] and nonparametric tests. Parametric test is applied when data is normally distributed and not skewed. Normal distribution^[2,3]

is characterized by a smooth bell-shaped symmetrical curve. ± 1 Standard deviation (SD) covers 68% and ± 2 SD covers 95% of the values in the distribution. It is always preferable to use parametric test as these tests are more robust. Sometimes data does not follow normal distribution and is skewed. In such scenarios, data transformation technique^[4] may be applied to convert skewed data into normal data. Only when this transformation is not possible, nonparametric tests should be used. Parametric tests use parameters like mean, SD, and standard error of mean for analysis. The lists of various parametric and nonparametric tests are given in Table 1.

Parametric tests

Student's *t*-test

This is a parametric test which was described by WS Gossett.^[5] He chose his pseudonym as “student” because his company did not allow its scientists to publish confidential data. Therefore, this test is known as the Student's *t*-test. This test is used to compare the two means and is used for small samples ($n < 30$). Paired *t*-test is used when one group serves as its own control, e.g. to compare the blood sugar before and after the administration of a drug. Unpaired *t*-test is used to compare the means of two independent groups, e.g. to compare the blood sugar of two

Address for correspondence: Dr. Avik Ray,
Department of Pharmacology, All India Institute of Medical
Sciences Bhopal, Saket Nagar, Bhopal - 462 020,
Madhya Pradesh, India.
E-mail: avik.jrpharma18@aiimsbhopal.edu.in

Received: 03-03-2021

Revised: 29-03-2021

Accepted: 12-05-2021

Published: 27-08-2021

Access this article online

Quick Response Code:



Website:
www.jfmpc.com

DOI:
10.4103/jfmpc.jfmpc_433_21

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Najmi A, Sadasivam B, Ray A. How to choose and interpret a statistical test? An update for budding researchers. J Family Med Prim Care 2021;10:2763-7.

Table 1: Tests of significance

Parametric tests	Nonparametric tests
Paired t-test	Wilcoxon signed-rank test
Unpaired t-test	Mann-Whitney U test
One-way ANOVA	Kruskal-Wallis test
Multiple measure ANOVA	Friedman's test
Pearson's correlation	Spearman's correlation
Linear regression	Nonparametric regression
Regression analysis	Chi-Square test

independent groups. The data should be normally distributed and quantitative. This test is used when the SD^[6] of two means is almost the same or SD of one group is not twice greater or lesser than that of other.

Analysis of variance (ANOVA) test

This test is used to compare the mean of three or more than three groups.^[7] The data should be normally distributed. One-way ANOVA is used when groups to be compared are defined by just one factor. Repeated measure ANOVA is used when groups to be compared are defined by multiple factors. For example, if we want to evaluate the effect of three different antihypertensive drugs on three different group of human volunteers, then we will use ANOVA test to evaluate about any significant difference between groups. ANOVA test does not indicate which group is significantly different from the others. *Post hoc* tests should be used to know about individual group differences. Various types of *post hoc* tests^[8] are available to know about individual group comparison like Bonferroni, Dunnett's, Tukeys test, etc.

Correlation coefficient test

This parametric test is used to know about the linear relationship^[9] between two variables. For example, if we want to know about any linear relationship between body weight and blood pressure, correlation test will be used. Correlation only shows an association between two variables. It does not show causation. Scatter plot can be used to know about correlation between two variables. Pearson's correlation coefficient test is used for continuous variables, and Spearman's correlation coefficient is used as for categorical variables.

Regression test

This parametric test is used to know about the dependent relationship^[10] between two variables. We can predict the value of dependent variable, based on the value of independent variable. For example, if we draw a curve between time and plasma concentration of a drug, then we can predict a drug concentration at particular time on the basis of time plasma concentration curve. Here, time is the independent variable and plasma concentration is the dependent variable. Dependent variable is plotted on *y*-axis and independent variable is plotted on *x*-axis.

Nonparametric test

These tests are used when the data is not normally distributed (skewed).^[11] Data is usually summarized as median.

Ranks and scores (Apgar scores and visual analogue score) do not follow normal distribution and are summarized as median.

- Wilcoxon test: Wilcoxon signed rank and Mann-Whitney U test are counterparts of paired and unpaired *t*-test for nonparametric test.
- Kruskal-Wallis test: This is counterpart of one-way ANOVA for nonparametric test.
- Friedman's test: This is counterpart of repeated measure ANOVA for nonparametric test.
- Spearman's rank correlation: This test is counterpart of Pearson correlation test for nonparametric test.
- Chi-square test: This nonparametric test is used for binomial or dichotomous data, which is summarized as percentage or proportions. For example, to compare the proportion of death and survival in vaccinated and nonvaccinated children with respiratory tract infections. There is no parametric counterpart for Chi-square test.

Type of variable/data

Variable or data may be numerical or categorical type.^[12,13] Numerical data may be continuous or discrete. Examples of continuous data are blood sugar, blood pressure, weight, height, etc. Examples of discrete data are the number of members in a family, number of persons who attended the outpatient department, number of persons experiencing nausea, etc. Categorical or qualitative data may be nominal or ordinal. Nominal data can be identified by some attributes or names like colour of eyes, names of religion, etc. Ordinal data can be arranged in some meaningful order like stages of cancer, severity of disease in terms of mild, moderate, and severe. Data can be summarized in the form of mean, median, or proportion. Numerical continuous data follows normal distribution and can be summarized as means. Numerical discrete data often follows nonnormal distribution and can be summarized as median. Dichotomous or binomial data^[14] can be defined as those data which have only two outcomes such as yes or no, or male or female. It can be summarized as proportions.

Types of analysis

In statistical terms, analysis may be a comparative analysis, a correlation analysis, or a regression analysis.^[15] Comparative analysis is characterized by comparison of mean or median between groups. Suppose we want to know the relation between two variables, for example, body weight and blood sugar. In such a case, correlation analysis will be used. If we want to predict the value of a second variable based on information about a first variable, regression analysis will be used. For example, if we know the values of body weight and we want to predict the blood sugar of a patient, regression analysis will be used.

Types of study design

In epidemiological studies, there are various type of study design like case control, cohort, and cross-sectional study designs. However, in statistics, there are only two types of study designs. First is paired or matched^[16] study design. Second is unpaired or independent study design. In paired study design, the same group serves as

its own control. For example, we want to evaluate the effect of a new drug on blood pressure in a group of 10 healthy volunteers. If we compare the values of blood pressure in the same group of 10 individuals, before intervention and after intervention, then this is known as paired or matched design. However, if we want to compare the values of blood pressure in two entirely different groups, then this is known as unpaired or independent study design.

Number of groups and data sets

There may be a single group but multiple data sets.^[17] For example, if we want to evaluate the effect of a new drug on heart rate of a single group of individuals, then there may be multiple data sets if we take the reading of heart rate at various time intervals. There may be two groups or two data sets. There may be more than two groups and more than two data sets. Different statistical test is applied for different situations.

Types of distribution

Data can be summarized as means if the variable follows normal distribution. Most of the bodily parameters^[8] like heart rate, blood pressure, blood sugar, serum cholesterol, height, and weight follow normal distribution. Numerical continuous data follows normal distribution and can be summarized as means. Numerical discrete data often follows nonnormal distribution and can be summarized as median. Ranks or scores do not follow normal distribution and can be summarized as median.^[18] Examples are Apgar score and visual analogue scale for pain measurement. Dichotomous data can be summarized as proportions.^[17] There are many statistical tests which are based on the assumption that the data follows normal distribution.

For example, as an investigator, you want to evaluate the melanizing action of three different topical preparations in three different groups of vitiligo patients (10 in each group). The three group of patients will apply either one of the topical preparations and the effect will be measured in scores (0–5, 0—No melanizing action, 5—excellent melanizing action). What will be the most appropriate statistical test?

From the above study, following points can be noted:

Objective: Evaluation of melanizing action of three different topical preparations in three different groups of vitiligo patients.

Type of data—scores summarized as median

Type of distribution—Nonnormal

No. of groups—3

Study design—Unpaired

Type of analysis—Comparison

According to Table 2, the row that matches criteria no. 1 and 2 is row number 2 and the column that matches the criteria no. 3,

4, and 5 is column number 4. The cell where column 4 and row 2 meet indicates Kruskal–Wallis test.

Please note that the list of tests is not comprehensive. It is a simplified table only to crudely demonstrate how to select a test for statistical analysis of data.

Interpretation of P value

The results provided by inferential statistics will be valid provided the selection of subjects, methods, and data collection are correct. For example, if we use *t*-test for highly skewed data, then the results will be invalid. During their thesis, postgraduate medical students are often more concerned about *P* or probability value.^[19] By convention, when *P* value < 0.01, the difference between groups is considered as highly significant. When *P* value is >0.01 but less than 0.05, then difference is considered as just significant. If *P* value is more than 0.05, then one should not immediately declare it as NOT significant. Before declaring it as NOT significant, one should try to know about power of the study. The power of a study is its ability to pick up a difference when it exists. So, power should be calculated especially if *P* value >0.05. The power may be low due to various reasons like small sample size, high dropout rates, noncompliance, etc. If power is <80% and *P* >0.05, then it is judicious to declare that the study has not enough power to detect the difference.

It is not a rule of thumb that a difference between two groups will be considered significant only when *P* value is <0.05. This 5% level is only taken as convention. It can be fixed at 1, 2, or 10% depending upon the study. The *P* value also depends on variance of data. If variance is less, then *P* value will also be less.

There are some situations in which clinical significance overrides statistical significance. Suppose an investigator wants to evaluate a new drug for rabies, he administers the new drug in 10 patients of rabies. In the second group of rabies patients (*n* = 10), standard treatment was given. Two patients survived in the first group and none survived in the second group. The statistical test showed that difference was not significant. What will you do in such situation? Will you dump the study as the results were not significant or evaluate this drug further? We all know that rabies is 100% fatal disease. It would be a miracle even if a single patient survived by a new drug. Therefore, the conclusion should be based on clinical knowledge and experience rather than statistics alone.

Suppose a new antidiabetic drug lowers mean fasting blood sugar by 2 mg% and statistical test concludes that the results are highly significant (*P* < 0.01). This raises an important question—should any physician recommend this new drug to patient of diabetes, which lowers mean fasting blood glucose by just 2 mg%? It is true that the difference between groups is statistically significant, but it is not at all clinically significant. So, practically this new drug is not adding any significant to the armory of medicine against diabetes.

Table 2: Selecting a statistical test

Type of data	Comparison Analysis				Correlation Analysis (Column 5)	Regression Analysis (Column 6)
	2 data sets		More than 2 data sets			
	Paired	Unpaired	Paired	Unpaired		
	Column 1	Column 2	Column 3	Column 4		
Row 1 Normally distributed continuous data (summarized as mean)	Paired <i>t</i> -test	Unpaired <i>t</i> -test	Repeated measures ANOVA	One-way ANOVA	Pearson's correlation	Linear regression
Row 2 Ranks, scores, and non normally distributed continuous data summarized as median	Wilcoxon signed-rank test	Mann-Whitney U test	Friedman's test	Kruskal-Wallis test	Spearman's rank correlation	Nonparametric regression
Row 3 Dichotomous data summarized as proportion	Mc Nemar's test	Fishers test or Chi-square test	Cochrane Q	Chi-square test	Contingency coefficient	Logistic regression

Interpretations of confidence interval

Suppose the mean systolic blood pressure in a sample population is 110 mmHg, and we want to know the population systolic blood pressure mean. Although the exact value cannot be obtained, a range can be calculated within which the true population mean lies. This range is called confidence interval^[20] and is calculated using the sample mean and the standard error (SE). The mean ±1SE and mean ±2 SE will give approximately 68 and 95% confidence interval, respectively. The endpoints of the confidence interval are known as confidence limits. Confidence interval is always mentioned with a particular degree of certainty, e.g. 95%. This is called confidence level and is expressed as percentage. The confidence level which is commonly used is 95%, but 90 and 99% confidence levels can also be calculated.

Confidence intervals should also be mentioned along with *P* value, especially in case of nonsignificant results. Confidence interval indicates the range of likely values of sample means in a population. When two groups are compared, the likely values of difference in means of two population under study can be calculated. For example, the difference in the height of two groups (Asian and European) can be found out and confidence interval for the difference in height can be calculated. The 95% confidence interval for the difference can be calculated using the formula (if *t*-test has been chosen):

$$\text{Upper limit} = \text{mean} + (t_{0.05} \times SE_{\text{diff}})$$

$$\text{Lower limit} = \text{mean} - (t_{0.05} \times SE_{\text{diff}})$$

Point to be noted is that the above formula is used to calculate the confidence interval for the difference between group means and not for individual means. If 95% confidence interval includes a zero value, the difference is not statistically significant at 5% significance level. If *P* value tells us about statistically significant difference, then why do we need to mention the confidence interval? It is because the confidence interval tells us about the precision of the estimate as indicated by the range. If the range is narrow, then it will be more precise. If the range is broad, then it will be less precise. One can get an idea of precision from confidence interval.

Relevance for primary care physicians

In this era of evidence-based medicine, having an in-depth knowledge of biostatistics to analyze health and biomedical research data is of utmost importance. The practice of primary care comes with the privilege of encountering a variety of diseases, both acute and chronic, which comes with their own unique set of statistical parameters, interpretations, and challenges. Choosing a statistical test for significance testing becomes critical if someone wants to analyze and compare the patient characteristics and relevant variables for both internal reporting in institutional assessments and for disseminating their findings to the world in the form of publications. This review would be a quick guide for all primary care physicians to choose the most appropriate statistical test pertaining to their data set and come up with important inferences and propositions.

Conclusion

Although it is difficult to know about the details of every statistical test, a biomedical researcher must have the basic knowledge of inferential statistics. Selection of wrong statistical test can lead to false conclusions which can compromise the quality of research. Similarly, a wrong interpretation will also lead towards a wrong conclusion. The researchers should have a clear idea about the various variable types they are dealing with, their respective distributions, and the kinds of tests they need to apply for analyzing the data set. Both *P* value and confidence interval should be documented for precise results. One may consult standard textbooks of statistics and software tools^[21] for statistical analysis. Various online and offline software like SPSS, Minitab, RStudio, and GraphPad Prism are available for statistical analysis which ease the process of data analysis.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Healey JF. Statistics: A Tool for Social Research. Belmont, CA: Wadsworth; 1993.

2. Kim HY. Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Open lecture on statistics. Restor Dent Endod* 2013;38:52-4.
3. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. *Int J Endocrinol Metab* 2012;10:486-9.
4. Reed JF 3rd, Salen P, Bagher P. Methodological and statistical techniques: What do residents really need to know about statistics? *J Med Syst* 2003;27:233-8.
5. Raju TN. William Sealy Gosset and William A. Silverman: Two "students" of science. *Pediatrics* 2005;116:732-5.
6. Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research: A review of common pitfalls. *Swiss Med Wkly* 2007;137:44-9.
7. Campbell MJ, Swinscow TD. *Statistics at Square One*. 11th ed. Wiley-Blackwell: BMJ Books; 2009.
8. Raveendran R, Gitanjali B, Chapter 8, Statistical Test of Significance and Choosing a Test. Manikandan S. A practical approach to PG Dissertation. 2nd ed. Pharma Med Press; 2012. p. 81-95.
9. Altman DG. *Practical Statistics for Medical Research*. London: Chapman and Hall; 1991.
10. Petrie A, Sabin C. The theory of linear regression and performing a linear regression analysis. *Medical Statistics at a Glance*. 2nd ed. London: Blackwell Publishing; 2005. p. 70-3.
11. Nayak BK, Hazra A. How to choose the right statistical test.? *Indian J Ophthalmol* 2011;59:85-6.
12. Karan J. How to select appropriate statistical test.? *J Pharm Negative Results* 2010;1:61-3.
13. Parikh MN, Hazra A, Mukherjee J, Gogtay N. Hypothesis testing and choice of statistical tests. *Research Methodology Simplified: Every Clinician a Researcher*. New Delhi: Jaypee Brothers; 2010. p. 121-8.
14. Wang D, Clayton T, Bakhai A. Analysis of survival data. In: Wang D, Bakhai A, editors. *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*. London: Remedica; 2006. p. 235-52.
15. Rosnow RL, Rosenthal R. Statistical procedures and the justification of knowledge in psychological science. *Am Psychol* 1989;44:1276-84.
16. Mishra P, Pandey CM, Singh U, Gupta A. Scales of measurement and presentation of statistical data. *Ann Card Anaesth* 2018;21:419-22.
17. Altman DG. *Practical Statistics for Medical Research*. CRC Press; 1990.
18. Barton B, Peat J. *Medical Statistics: A Guide to SPSS, Data Analysis and Clinical Appraisal*. 2nd ed. Wiley Blackwell, BMJ Books; 2014.
19. Dahiru T. *P-value, a true test of statistical significance? A cautionary note*. *Ann Ib Postgrad Med* 2008;6:21-6.
20. Wang EW, Ghogomu N, Voelker CC, Rich JT, Paniello RC, Nussenbaum B, *et al.* A practical guide for understanding confidence intervals and *P* values. *Otolaryngol Head Neck Surg* 2009;140:794-9.
21. Shaikh MA. Use of statistical tests and statistical software choice in 2014: Tale from three Medline indexed Pakistani journals. *J Pak Med Assoc* 2016;66:464-6.