IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

INTERPRETATION OF DIFFERENT PARAMETRIC AND NON-PARAMETRIC STATISTICS AND THEIR IMPORTANCE IN RESEARCH

Dr Rahul Kumar

(JRF, Ph.D.; M. Phil, M.Ed., M.Sc.)

Teacher in Education Department of Rajasthan

ABSTRACT: concept of parametric and non parametric is very complex and after conceptualisation the selection of appropriate techniques is quiet tough task for a researcher.so This section will discuss on how to choose a test?,Definitions of parametric and non-parametric parameters and their importance.

Key words: parametric, non parametric, affect on statistical analyses.

Several fundamental statistical concepts are helpful prerequisite knowledge for fully understanding the terms “parametric” and “nonparametric.” These statistical fundamentals include random variables, probability distributions, parameters, population, sample, sampling distributions and the Central Limit Theorem. I cannot explain these topics in a few paragraphs. Thus, I will limit my explanation to a few helpful topics.

The field of statistics exists because it is usually impossible to collect data from all individuals of interest (population). Our only solution is to collect data from a subset (sample) of the individuals of interest, but our real desire is to know the “truth” about the population. Quantities such as means, standard deviations and proportions are all important values and are called “parameters” when we are talking about a population. Since we usually cannot get data from the whole population, we cannot know the values of the parameters for that population. We can, however, calculate estimates of these quantities for our sample. When they are calculated from sample data, these quantities are called “statistics.” A statistic estimates a parameter. Parametric statistical procedures rely on assumptions about the shape of the distribution (i.e., assume a normal distribution) in the underlying population and about the form or parameters (i.e., means and standard deviations) of the assumed distribution. Nonparametric statistical procedures rely on no or few assumptions about the shape or parameters of the population distribution from which the sample was drawn.

It is valid to use statistical tests on hypotheses suggested by the data, the P values should be used only as guidelines, and the results treated as tentative until confirmed by subsequent studies. A useful guide is to use a Bonferroni correction, which states simply that if one is testing *n* independent hypotheses, one should use a significance level of 0.05/*n*. Thus if there were two independent hypotheses a result would be declared significant only if $P<0.025$. Note that, since tests are rarely independent, this is a very conservative procedure – i.e. one that is unlikely to reject the null hypothesis. The investigator should then ask "are the data independent?" This can be difficult to decide but as a rule of thumb results on the same individual, or from matched individuals, are not independent. Thus results from a crossover trial, or from a case-control study in which the controls were matched to the cases by age, sex and social class, are not independent.

- Analysis should reflect the design, and so a matched design should be followed by a matched analysis.
- Results measured over time require special care. One of the most common mistakes in statistical analysis is to treat correlated variables as if they were independent. For example, suppose we were looking at treatment of leg ulcers, in which some people had an ulcer on each leg. We might have 20 subjects with 30 ulcers but the number of independent pieces of information is 20 because the state of ulcers on each leg for one person may be influenced by the state of health of the person and an analysis that considered ulcers as independent observations would be incorrect. For a correct analysis of mixed paired and unpaired data consult a statistician.

The next question is "what types of data are being measured?" The test used should be determined by the data. The choice of test for matched or paired data is described in Table 1 and for independent data in Table 2.

Table 1 Choice of statistical test from paired or matched observation

Variable	Test
Nominal	McNemar's Test
Ordinal (Ordered categories)	Wilcoxon
Quantitative (Discrete or Non-Normal)	Wilcoxon
Quantitative (Normal*)	Paired *t* test

It is helpful to decide the input variables and the outcome variables. For example, in a clinical trial the input variable is the type of treatment - a nominal variable - and the outcome may be some clinical measure perhaps Normally distributed. The required test is then the *t*-test (Table 2). However, if the input variable is continuous, say a clinical score, and the outcome is nominal, say cured or not cured, logistic regression is the required analysis. A *t*-test in this case may help but would not give us what we require, namely the probability of a cure for a given value of the clinical score. As another example, suppose we have a cross-sectional study in which we ask a random sample of people whether they think their general practitioner is doing a good job, on a five point scale, and we wish to ascertain whether women have a higher opinion of general practitioners than men have. The input variable is gender, which is nominal. The outcome variable is the five point ordinal scale. Each person's opinion is independent of the others, so we have independent data. From Table 2 we should use a χ^2 test for trend, or a Mann-Whitney U test with a correction for ties (N.B. a tie occurs where two or more values are the same, so there is no strictly increasing order of ranks – where this happens, one can average the ranks for tied values). Note, however, if some people share a general practitioner and others do not, then the data are not independent and a more sophisticated analysis is called for. Note that these tables should be considered as guides only, and each case should be considered on its merits.

Table 2 Choice of statistical test for independent observations

		Outcome variable					
		Nominal	Categorical (>2 Categories)	Ordinal	Quantitative Discrete	Quantitative Non-Normal	Quantitative Normal
Input Variable	Nominal	X^2 or Fisher's	X^2	X^2-trend or Mann - Whitney	Mann-Whitney	Mann-Whitney or log-rank[a]	Student's *t* test
	Categorical (2>categories)	X^2	X^2	Kruskal-Wallis[b]	Kruskal-Wallis[b]	Kruskal-Wallis[b]	Analysis of variance[c]
	Ordinal (Ordered categories)	X^2-trend or Mann - Whitney	[e]	Spearman rank	Spearman rank	Spearman rank	Spearman rank or linear regression[d]
	Quantitative Discrete	Logistic regression	[e]	[e]	Spearman rank	Spearman rank	Spearman rank or linear regression[d]
	Quantitative non-Normal	Logistic regression	[e]	[e]	[e]	Plot data and Pearson or Spearman rank	Plot data and Pearson or Spearman rank and linear regression
	Quantitative Normal	Logistic regression	[e]	[e]	[e]	Linear regression[d]	Pearson and linear regression

[a] If data are censored. [b] The Kruskal-Wallis test is used for comparing ordinal or non-Normal variables for more than two groups, and is a generalisation of the Mann-Whitney U test. [c] Analysis of variance is a general technique, and one version (one way analysis of variance) is used to compare Normally distributed variables for more than two groups, and is the parametric equivalent of the Kruskal-Wallistest. [d] If the outcome variable is the dependent variable, then provided the residuals (the differences between the observed values and the predicted responses from regression) are plausibly Normally distributed, then the distribution of the independent variable is not important. [e] There are a number of more advanced techniques, such as Poisson regression, for dealing with these situations. However, they require certain assumptions and it is often easier to either dichotomise the outcome variable or treat it as continuous.

Parametric tests are those that make assumptions about the parameters of the population distribution from which the sample is drawn. This is often the assumption that the population data are normally distributed. Non-parametric tests are "distribution-free" and, as such, can be used for non-Normal variables. Table 3 shows the non-parametric equivalent of a number of parametric tests.

Table 3 Parametric and Non-parametric tests for comparing two or more groups

Parametric test	Non-Parametric equivalent
Paired t-test	Wilcoxon Rank sum Test
Unpaired t-test	Mann-Whitney U test
Pearson correlation	Spearman correlation
One way Analysis of variance	Kruskal Wallis Test

Non-parametric tests are valid for both non-Normally distributed data and Normally distributed data, so why not use them all the time?
It would seem prudent to use non-parametric tests in all cases, which would save one the bother of testing for Normality. Parametric tests are preferred, however, for the following reasons:

1. We are rarely interested in a significance test alone; we would like to say something about the population from which the samples came, and this is best done with estimates of parameters and confidence intervals.

2. It is difficult to do flexible modelling with non-parametric tests, for example allowing for confounding factors using multiple regression.

3. Parametric tests usually have more statistical power than their non-parametric equivalents. In other words, one is more likely to detect significant differences when they truly exist.

Do non-parametric tests compare medians?
It is a commonly held belief that a Mann-Whitney U test is in fact a test for differences in medians. However, two groups could have the same median and yet have a significant Mann-Whitney U test. Consider the following data for two groups, each with 100 observations. Group 1: 98 (0), 1, 2; Group 2: 51 (0), 1, 48 (2). The median in both cases is 0, but from the Mann-Whitney test $P<0.0001$. Only if we are prepared to make the additional assumption that the difference in the two groups is simply a shift in location (that is, the distribution of the data in one group is simply shifted by a fixed amount from the other) can we say that the test is a test of the difference in medians. However, if the groups have the same distribution, then a shift in location will move medians and means by the same amount and so the difference in medians is the same as the difference in means. Thus the Mann-Whitney U test is also a test **for the difference in means. How is the Mann- Whitney U test related to the *t*-test? If one were to input the ranks of the data rather than the data themselves into a two sample *t*-test program, the P value obtained would be very close to that produced by a Mann-Whitney U test**.
Parametric tests and analogous nonparametric procedures As I mentioned, it is sometimes easier to list examples of each type of procedure than to define the terms. Table F contains the names of several statistical procedures you might be familiar with and categorizes each one as parametric or nonparametric. All of the parametric procedures listed in Table F rely on an assumption of approximate normality. TABLE F:

Analysis Type Parametric nonparametric

Procedure Procedure

Compare means Two-sample t-testWilcoxon ranksum test

between two distinct/independent groupCompare two quantitative paired t-Test Wilcoxon signedrank test

measurements taken from the

same individual

Compare means between

three or more distinct/independent groups?Analysis of variance (ANOVA)Kruskal-Wallis test

Estimate the degree of association

between two quantitative variables Pearson Spearman’s rank

coefficient of correlationcorrelation

summary of the major points and how they might affect statistical analyses you perform: • Parametric and nonparametric are two broad classifications of statistical procedures. • Parametric tests are based on assumptions about the distribution of the underlying population from which the sample was taken. The most common parametric assumption is that data are approximately normally distributed. • Nonparametric tests do not rely on assumptions about the shape or parameters of the underlying population distribution. • If the data deviate strongly from the assumptions of a parametric procedure, using the parametric procedure could lead to incorrect conclusions. • You should be aware of the assumptions associated with a parametric procedure and should learn methods to evaluate the validity of those assumptions. • If you determine that the assumptions of the parametric procedure are not valid, use an analogous nonparametric procedure instead. • The parametric assumption of normality is particularly worrisome for small sample sizes ($n < 30$). Nonparametric tests are often a good option for these data. • It can be difficult to decide whether to use a parametric or nonparametric procedure in some cases. Nonparametric procedures generally have less power for the same sample size than the corresponding parametric procedure if the data truly are normal. Interpretation of nonparametric procedures can also be more difficult than for parametric procedures. • Visit with a statistician if you are in doubt about whether parametric or nonparametric procedures are more appropriate for your data. • The book Practical Nonparametric Statistics 2 is an excellent resource for anyone interested in learning about this topic in great detail. More general texts such as Fundamentals of Biostatistics 3 and Intuitive Biostatistics 4 have chapters covering the topic of nonparametric procedures.

Reference;

https://www.mayo.edu/research/documents/parametric-and-nonparametric-demystifying-the-terms/doc-20408960

www.healthknowledge.org.uk/public-health-textbook/research-methods

Campbell MJ and Swinscow TDV. Statistics at Square One 11th ed. Wiley-Blackwell: BMJ Books 2009.

BERD Home page

Walsh, J.E. (1962) Handbook of Nonparametric Statistics, New York: D.V. Nostrand. Conover, W.J. (1980). Practical Nonparametric Statistics, New York: Wiley Motulsky, H. (1995). Intuitive Biostatistics, New York: Oxford University Press.