INVITED FEATURE: STATISTICS PRIMER

Parametric vs nonparametric statistical methods: which is better, and why?

Robin Turner, Ari Samaranayaka, Claire Cameron

The statistical methods used in quantitative research can be described as either parametric or nonparametric. These terms are often misunderstood and thought to be describing the data. They are actually used to describe the assumptions underlying certain statistical methods. Generally speaking, if a statistical method has been developed using properties of a particular distribution, there are resulting assumptions that must be adhered to for the validity of its use. That statistical method is then described as parametric. The term "parametric" refers to the parameters of the underlying statistical distribution. For example, the *t*-test has been developed using normal distribution theory, so it has an underlying assumption that the distribution of the sample *mean* (which is a parameter) is normal. This does not mean that the population data or the sample data need to be normally distributed. We assume that if we took repeated random samples from the population and calculated the mean of each, these means would follow a normal distribution. Hence, the *t*-test is a parametric method.

A common misconception is that "parametric" always refers to the normal distribution assumption. Although it is common for many of the basic statistical procedures to assume normality (of the sampling means), the term parametric also refers to procedures that assume other types of distributions such as the binomial or Poisson distributions.

There are a suite of tests that people often refer to when talking about nonparametric methods (see Table 1). They can also be referred to as *distribution-free* as they assume no underlying formal distribution of the estimates. They are alternatives to many of the standard simple statistical tests. It is worth noting that each test still has some assumptions that need to be checked. For example, the paired sample sign test assumes that the two comparison population groups have the same distribution (without assuming what this distribution is), and the Mann-Whitney *U* test assumes that number of ties is not large. The methods mainly use ranks (where the magnitude of the measures are used to create the ranking) or signs (+/−). Either way, these are an efficient approach if the assumptions of the standard tests are not met, but they have a major disadvantage, with some information about the data being lost by either using the rank or the sign, not the magnitude. Table 1 lists some of the commonly used standard statistical tests, along with the nonparametric equivalent. Some of these nonparametric tests are basically the same test with a different name e.g. the Mann-Whitney *U* test is equivalent to the Wilcoxon rank-sum test and is at times referred to as the Mann-Whitney rank-sum test or the Mann-Whitney-Wilcoxon test. This is not to be confused with the situation for paired data, where the paired sample sign test and the Wilcoxon signed rank test are fundamentally different tests. This can make delving into the world of nonparametric tests more confusing, as different statistical software packages will use different names for tests, so it is important to check what test is actually being used, along with the assumptions of that test.

Table 1. Nonparametric equivalents for some commonly used parametric methods[1]

Standard test	Nonparametric alternative
One sample *t*-test	One sample sign*
Unpaired *t*-test	Mann-Whitney *U*
Unpaired *t*-test	Wilcoxon rank-sum
Paired *t*-test	Paired sample sign*
Paired *t*-test	Wilcoxon signed rank
One way ANOVA	Kruskal-Wallis (generalisation of the Mann-Whitney to more than 2 groups)
Two way ANOVA	Friedman
Pearson's correlation	Spearman rank coefficient
Correlation with more than two variables	Kendall coefficient of concordance

**uses signs, not ranks*

Returning to the example of the *t*-test again: the underlying assumption is not about having normally distributed *data* as is commonly thought, it is about normality in sample *means* from repeated samples. But in practice, we have only one sample mean, therefore we cannot directly assess the validity of this assumption. However, this normality assumption will hold for large samples (usually 30 observations or more) regardless of the distributions of the data or underlying population. This is stated by the Central Limit Theorem. Therefore, when the sample size is large we can use this parametric procedure without worrying about the normality assumption. We are mostly concerned about this assumption when we have a small sample where the central limit theorem no longer applies and we have no reason to assume that the sampling distribution will be normal. If the population, from which the sample is drawn, itself follows a normal distribution then this assumption on the sampling distribution will still hold for small samples, but if we cannot make that very strong assumption about the population then we can no longer use this test.

How can we decide if a parametric or nonparametric procedure is the best for our specific situation? The answer is clear only in some situations. Real world data is much more difficult to handle and to make decisions about than the textbook examples. In practice, if we are concerned about the assumptions, we may run the parametric test first and then run the nonparametric equivalent to see broadly if we get a similar answer. If we do get a similar answer we have some reassurance that the parametric test results are reasonable to report. If we get very different answers and we were concerned about the assumptions, then we would use the nonparametric results as we cannot trust the parametric results.

You may, by now, be wondering why we bother having parametric tests at all: why not just use nonparametric tests all the time, and therefore avoid having to make these distributional assumptions? Nonparametric tests come with a cost: they rely on using ranks (or signs) rather than the actual observations, so information is lost. In the situation where the use of a parametric test is deemed appropriate, the parametric test always has more power than the nonparametric equivalent.[2] Therefore, we prefer parametric tests where we can use them, to maximise power and for ease of interpretation.

There are also some situations where the data dictates what measure (mean, median etc.) is most appropriate for the data and therefore what test should be used. For example, when the research question refers to the "centre" of a highly skewed distribution such as hospital length of stay, where most people have a short length of stay and a small number of people have very long lengths of stay, we would consider that the mean is not a useful measure of the centre of this data and would instead want to report the median. Whilst the parametric tests such as the *t*-test would perform well on this data if the sample size is large, it does not make sense to use this test as it is comparing means which are heavily influenced by the extreme values. We may, by choice in the design, use nonparametric tests such as the Mann-Whitney *U,* which compares medians rather than means. When we report the results of this test we would report the median and the interquartile range as they are the measures that best summarise the data. The point is, the assumptions underlying the statistical procedure are not the only criteria for deciding whether to use parametric or nonparametric procedures.

We have talked about assumptions at length, as knowledge on the specific assumptions underlining the statistical procedure is critical, so that an informed decision can be made regarding which method to use. There are some other ways to deal with assumption issues rather than turning to a nonparametric test. Transformations of the data can help in some instances. For example, one of the assumptions in least squares regression (which is a parametric procedure) is that the residuals are normally distributed. We used this example as there is no obvious nonparametric alternative to this method. So in this situation we need to consider other options. A suitable transformation such as log transformation may fix the non-normality problem if the residuals are skewed. However, these transformations come with an associated cost, in that the results from transformed data are much harder to interpret.

We have also restricted discussion in this paper to the group of nonparametric tests that are either sign tests or ranking tests. There are many other forms of nonparametric methods including bootstrapping, permutation tests, and the very commonly used Kaplan-Meier method for survival data.[3]

Statistical decisions and interpretation are not clear cut and do not follow a series of "easy to apply in all situations" rules. There is a great deal of nuance when analysing and interpreting data and applying statistical tests. It is always good practice to have an experienced biostatistician involved in quantitative research who can advise on these sorts of issues, and has the experience to make informed decisions about the best approaches to use for a particular situation.

References

1. Harraway J. Introductory statistical methods for biological, health, and social sciences. University of Otago Press; 1997.

2. Siegel S. Nonparametric statistics. The American Statistician. 1957 Jun;11(3):13-19.

3. Samaranayaka A, Turner R, Cameron C. Kaplan-Meier plots. New Zealand Medical Student Journal. 2019 Jun;(28):45-46.

About the authors

> Associate Professor Robin Turner, BSc(Hons), MBiostat, PhD, is the Director, Centre for Biostatistics, Division of Health Sciences, University of Otago

> Dr Ari Samaranayaka, BSc, MPhil, PhD, is a Senior Research Fellow and Biostatistician, Centre for Biostatistics, Division of Health Sciences, University of Otago

> Dr Claire Cameron, BSc(Hons), DipGrad, MSc, PhD, is a Senior Research Fellow and Biostatistician, Centre for Biostatistics, Division of Health Sciences, University of Otago

Correspondence

Associate Professor Robin Turner: robin.turner@otago.ac.nz