2010 STAT DAY

Statistics for Quality Improvement

23rd April 2010

Platinum Partner

NSB National Savings Bank

Silver Partners

nielsen

Stax Inc.

Print Media Partner

The Nation

Organized by Stat Circle - University of Colombo

Contrasts between Parametric and Non-parametric tests

by

Ms. H. T. Tharanganie
Assistant Lecturer,
Department of Statistics, University of Colombo

In the development of statistical methods, the techniques of inference which appeared were those that used many assumptions about the nature of the populations from which the observations were drawn, are called **'parametric'**. For example, a technique of inference may be based on the assumption that the data were drawn from a normally distributed population or a technique of inference may be based on the assumption that two sets of data were drawn from populations having the same variance (σ^2). Such techniques produce conclusions which contain qualifiers, e.g. "*If the assumptions regarding the shape of the population distribution(s) are valid, then we may conclude that ...*".

Parametric tests are preferred because, in general, for the same number of observations, they are more likely to lead to the rejection of a false null hypothesis. That is, they have more power.

Non-parametric, or ***distribution-free*** tests are so-called because the assumptions underlying their use are "fewer and weaker than those associated with parametric tests" (Siegel & Castellan, 1988, p. 34). In another way, non-parametric tests require few assumptions about the shapes of the underlying population distributions and they are much weaker than those of parametric tests. For this reason, they are often used in place of parametric tests when one feels that the assumptions of the parametric test have been too grossly violated. In conclusion of the non-parametric tests, we might be able to say, *"Regardless of the shape of the population(s), we may conclude that ..."*.

These non-parametric tests have the obvious advantage of not requiring the assumptions of the distribution and the variances. They compare medians rather than means and, as a result, if the data have few outliers, their influence is negated.

Advantages of non-parametric tests

Non-parametric tests possess several advantages over the parametric tests and few of them are presented. First, if the **sample size is very small**, there may be no appropriate alternative other than using a non-parametric statistical test, unless the nature of the population distribution is known exactly. As the second, since non-parametric tests make **fewer assumptions**, their **applicability is much wider** than the corresponding parametric tests. In particular, they may be applied in situations where less is known about the situation. Third, non-parametric statistical tests are typically **much easier to learn and to apply** than parametric tests. In addition, their **interpretation often is more direct** than those of parametric tests. Another justification for the use of non-parametric tests is **simplicity**. Fourth, non-parametric tests are **less likely to be affected by outliers**. The fifth advantage would be non-parametric methods are **often the only way to analyze nominal or ordinal data**, while parametric methods require the use of interval or ratio scaled data. That is, the researcher may only be able to say of his or her subjects that one has more or less of the characteristic than another, without being able to say how much more or less.

Disadvantages of non-parametric tests

It is firmly needed to point out that non-parametric tests do have at least two major disadvantages in comparison to parametric tests. First, **non-parametric tests are less powerful**. Why? Because, parametric tests use more of the information available in data. Parametric tests make use of information consistent with interval scale measurement; whereas non-parametric tests typically make use of ordinal information only. As Siegel and Castellan (1988) put it, "non-parametric statistical tests are wasteful." Second, **parametric tests are much more flexible, and allow you to test a greater range of hypotheses**. For example, factorial ANOVA designs allow you to test for interactions between variables in a way that it is not possible with non-parametric alternatives. There are non-parametric techniques to test for certain kinds of interactions under certain circumstances, but these are much more limited than the corresponding parametric techniques. Therefore, *when the assumptions for a parametric test are met, it is generally (but not necessarily always) preferable to use the parametric test rather than a non-parametric test.*

Summary of the comparison between parametric vs. non-parametric

The table below illustrates the summary of the comparison between parametric vs. non-parametric tests.

	Parametric	Non-parametric
Distribution	Must be known	Not essential to be known
Variance(s)	Must be known (Homogeneous or Heterogeneous)	Not essential to be known
Data type	Ratio or Interval	Ordinal or Nominal
Data set relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	More powerful, Can draw more conclusions	Simplicity, Less affected by outliers
Tests		
Single group	One sample t-test	Sign test, Wilcoxon's Signed Rank test
Two independent groups	Independent measures t-test	Mann-Whitney test, Tukey's Quick test
Two dependent groups	Matched pair t-test	Matched pair Sign test, Matched pair Wilcoxon's Signed Rank test
More than two independent groups	One-way independent measures ANOVA	Kruskall-Wallis test, Median test
More than two dependent groups	One-way repeated measures ANOVA	Friedman's test
Correlation test	Pearson's Correlation	Spearman's Correlation, Kendall's Rank Correlation

As the table shows, there are different tests for different types of data. There are a wide range of alternatives for different types of groups in non-parametric tests. As mentioned earlier, if all the assumptions of parametric tests hold true, then they are more powerful and produce more accurate and precise estimates than non-parametric tests. However, if those assumptions are incorrect, parametric tests can be very misleading.

Large sample versions of non-parametric tests

You may have noticed that the tables of critical values for many non-parametric statistics only go up to sample sizes of about 25-50. If so, perhaps you have wondered what to do when you have sample sizes larger than that, and want to carry out a non-parametric test. Fortunately, it turns out that the sampling distributions of many non-parametric statistics converge on the normal distribution as sample size increases. Because of that, it is possible to carry out a so-called "large-sample" version of the test (which is really a *z*-test) if you know the mean and variance of the sampling distribution for that particular statistic.

Common structure of all *z*- and *t*-tests

As I have mentioned before, all *z*- and *t*-tests have a common structure. In general terms:

$$\mathrm{z\ (or\ t)} = \frac{\text{statistic} - (\text{parameter} \mid \mathrm{H_0} \text{ is true})}{\text{standard error of the statistic}}$$

When the sampling distribution of the statistic in the numerator is normal, then if the true (population) standard error (SE) of the statistic is known, the computed ratio can be evaluated against the standard normal (*z*) distribution. If the true standard error of the statistic is not known, then it must be estimated from the sample data, and the proper sampling distribution is a *t*-distribution with some number of degrees of freedom.

Reference:
http://www.angelfire.com/wv/bwhomedir/notes/nonpar.pdf