SOME ASPECTS OF THE RANK TRANSFORM IN ANALYSIS OF VARIANCE PROBLEMS

Ronald L. Iman, Sandia National Laboratories

1. INTRODUCTION

A problem that applied statisticians have been confronted with virtually since the inception of parametric statistics is that of fitting real world problems into the framework of normal statistical theory when many of the data they deal with are clearly nonnormal. From such problems have emerged two distinct approaches or schools of thought: (a) transform the data to a form more closely resembling a normal distribution framework or (b) use a distribution free procedure. The first method may include the log transformation, square root transformation, arcsin transformation, and so forth, and may even be broad enough to include robust procedures that tend to give small weights to outliers, that is, to observations that may contribute greatly to the nonnormal form of the data. The second method includes a large body of methods based on the ranks of the data.

There is a way of combining these two methods by presenting many nonparametric methods as parametric methods applied to transformed data. Simply replace the data with their ranks, then apply the usual parametric t test, F test, and so forth, to the ranks. This is called the rank transformation (RT) approach. This approach results in a class of nonparametric methods that includes the Wilcoxon-Mann-Whitney test, the Kruskal-Wallis test, the Wilcoxon signed ranks test, the Friedman test, Spearman's rho, and others. The rank transformation approach also furnishes useful methods in multiple regression, discriminant analysis, cluster analysis, analysis of experimental designs, and multiple comparisons.

Of course, there are several ways in which ranks can be assigned to observations. The following types are suggested by Conover and Iman (1981).

RT-1. The entire set of observations is ranked from smallest to largest, with the smallest observation having rank 1, the second smallest rank 2, and so on. Average ranks are assigned in case of ties.

RT-2. The observations are partitioned into subsets and each subset is ranked within itself independently of the other subsets.

RT-3. This rank transformation is RT-1 applied after some appropriate reexpression of the data.

RT-4. The RT-2 type is applied to some appropriate reexpression of the data.

The rank transformation approach provides a useful pedagogical technique for introducing these nonparametric methods as an integral part of an introductory course in statistics, instead of isolating the methods in a separate unit that may appear to the student to be disconnected from the general flow of the course. Also, it allows the practitioner to make full use of existing statistical packages that may not have suitable nonparametric programs by simply entering the ranks of the data into the programs for the parametric analysis. For example, with SAS one would use the PROC RANK command before using the parametric procedure. And finally, this approach may be viewed as a useful tool for developing new nonparametric methods in situations where satisfactory parametric procedures exist.

2. TWO INDEPENDENT SAMPLES

Let $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_m$ represent two independent random samples. To test the hypothesis that $E(X) = E(Y)$ the parametric procedure employs the two-sample t statistic

$$t = \frac{\overline{X} - \overline{Y}}{\left[(\Sigma(X_i-\overline{X})^2 + \Sigma(Y_i-\overline{Y})^2)\ \dfrac{N}{nm(N-2)}\right]^{1/2}} \qquad (2.1)$$

where $N = n + m$, and compares t with quantiles from the t distribution with $N - 2$ degrees of freedom (df). The nonparametric Wilcoxon-Mann-Whitney two-sample test requires replacing the data by the ranks R_i from 1 to N, and uses the statistic, in its standardized form with the adjustment for ties incorporated,

$$T = \frac{S - n(N+1)/2}{\left[\dfrac{nm}{N(N-1)} \sum_{i=1}^{N} R_i^2 - \dfrac{nm(N+1)^2}{4(N-1)}\right]^{1/2}} \qquad (2.2)$$

where $S = \Sigma_{i=1}^{n} R_i$ is the sum of the ranks of the X's. The statistic T is compared with the standard normal distribution or, if there are no ties and the sample sizes are less than 20, exact tables may be used for S (Conover 1980).

A rank transformation procedure is based on computing t on the ranks R_i to get the statistic

$$T_R = \frac{1}{n} S - \frac{1}{m}\left(\frac{N(N+1)}{2} - S\right) \div \left[\left(\sum_{i=1}^{N} R_i^2 - \frac{1}{n} S^2 - \frac{1}{m}\left(\frac{N(N+1)}{2} - S\right)^2\right)\frac{N}{nm(N-2)}\right]^{1/2} \qquad (2.3)$$

and using the t tables as with (2.1). This is an example of an RT-1 type procedure. A little algebra reveals an important relationship between T_R and T:

$$T_R = \frac{T}{\left[\frac{N-1}{N-2} - \frac{1}{N-2} T^2\right]^{1/2}} , \quad (2.4)$$

which shows that T_R is a monotonically increasing function of T. Since T contains the correction for ties, so does t_R because it is a function of T.

Let us consider the implication of (2.4). When T is in its upper α level tail region, then T_R is in its upper α level tail region also. The same can be said for the lower α level tail regions of each. For example, when n = 14 and m = 18 the upper five percent value for S (Conover 1980, Table A.7) is 274 if there are no ties. Substitution of S into (2.2) and (2.3) reveals the exact upper five percent values for T and T_R to be 1.633 and 1.681, respectively. When T is compared with the upper five percent quantile 1.645 from a standard normal distribution, a slightly conservative test results. The t distribution with 30 df gives an upper five percent critical value of 1.697, also resulting in a slightly conservative test for T_R. Because T_R is a monotonic function of T, the two tests are equivalent when the exact critical values are used. The normal distribution and the t distribution provide two different approximations, which have been compared by Iman (1976). The Wilcoxon-Mann-Whitney test, with all of its good properties, may be performed using T_R as a test statistic instead of T if desired. In fact T_R may be preferred, as computing routines are generally readily available for the t statistic; and also it may be simpler to teach this procedure to someone who understand the t test and transformations, but who does not have a working knowledge of nonparametric statistics.

3. OTHER NONPARAMETRIC PROCEDURES

In a manner similar to that show for two independent samples Conover and Iman (1981) show that the Kruskal-Wallis test for k independent samples is obtained by applying the one-way analysis of variance F-test on ranks to obtain

$$F_R = \frac{\left[\sum_{i=1}^{k} \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4}\right] \Big/ (k-1)}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(R(X_{ij}) - \frac{R_i}{n_i}\right)^2 \Big/ (N-k)} \quad (3.1)$$

as a test statistic. This is another example of an RT-1 type procedure. Elementary algebra reveals F_R is a monotonic function of H (the usual Kruskal-Wallis chi-squared statistic)

$$F_R = (H/(k-1))/((N-1-H)/(N-k)) \quad (3.2)$$

so the two tests are equivalent. The upper α level critical value from the chi-squared distribution, when substituted for H in (3.2), results in a slightly different critical value for F_R than that obtained from the appropriate tables of the F distribution. Both methods, however, merely provide approximations to the true critical value. Iman and Davenport (1976) compare these and other approximations and show that the F approximation should be preferred to the chi-squared in most cases.

For dependent samples the usual parametric procedure is the paired t-test. If this one-sample t-statistic is applied to the signed ranks R_i of $D_i = X_i - Y_i$ where (X_i, Y_i) are matched pairs then Conover and Iman (1981) give the following statistic

$$T_R = \frac{\Sigma R_i}{\left[\frac{n}{n-1} \Sigma R_i^2 - \frac{1}{n-1} (\Sigma R_i)^2\right]^{1/2}} \quad (3.3)$$

which is compared with the t distribution, n - 1 df, as an approximation. Since D_i represents a reexpression of the data (X_i, Y_i) and may be considered to be the product of (sign D_i) and $|X_i - Y_i|$, this is an example of an RT-3 type procedure.

Note that T_R is also expressible as

$$T_R = \frac{T}{\left[\frac{n}{n-1} - \frac{1}{n-1} T^2\right]^{1/2}} , \quad (3.4)$$

where T is the usual Wilcoxon signed ranks test statistic as given in Conover (1980). Since T_R is a monotonic function of T the test that uses T_R is equivalent to the test that uses T. A comparison of the normal approximation, the student's t approximation, and the exact distribution is given by Iman (1974a).

Suppose that the paired t-test is applied directly to RT-1 type ranks; that is, the X's and Y's are replaced by their corresponding ranks 1 to 2n and D_i is the difference in those ranks. This application of the rank transformation approach does not yield the Wilcoxon signed ranks test, but rather introduces a new procedure. This new procedure is asymptotically distribution free by virtue of the central limit theorem. Properties of this test are reported by Iman, Hora, and Conover (1982).

For the analysis of randomized complete block designs the usual nonparametric test involves ranking the observations from 1 to k within each block, making no interblock comparisons. The Friedman test uses the statistic, corrected for ties,

$$T = \frac{(k-1)\sum_{j=1}^{k} [R_j - b(k+1)/2]^2}{\sum_i \sum_j R^2(X_{ij}) - bk(k+1)^2/4} \quad (3.5)$$

where R_j is the sum of the ranks $R(X_{ij})$ for treatment j. The chi-squared distribution with k - 1 df is used as an approximation to the distribution of T.

Conover and Iman (1981) show that another way of considering the Friedman test is to compute the two-way analysis of variance F-statistic on the intrablock ranks that are used in the Friedman test. This is a type RT-2 procedure. The result is a statistic F_R that is a monotonic function of the Friedman statistic

$$F_R = (T/(k-1))/((b(k-1) - T)/(b-1)(k-1)). \quad (3.6)$$

Comparison of F_R with the F distribution provides a more accurate approximation (Iman and Davenport 1980) than the chi-squared approximation used with (3.5).

Suppose F is applied directly to RT-1 type ranks where all of the observations are ranked together, from 1 to bk in this case. This type of ranking takes advantage of both between and within block information. The result is a test which is asymptotically distribution free. This procedure, with the F distribution as an approximation, compares favorably with the RT-2 type Friedman test (Iman, Hora, and Conover 1982), and even Fisher's randomization test (Conover and Iman 1980) in terms of robustness and power.

It is easy to extend the RT-1 type procedure to other experimental designs. This approach is robust and powerful in the two-way layout with interaction (Iman 1974b), in a test for interaction when replication effects are present (Conover and Iman 1976), and in a test for main effects in the presence of replication and interaction effects (Iman and Conover 1976).

The advantage of ranking all of the observations together is that any analysis of variance procedure may be applied to the ranks, with the resulting tests for main effects, interactions, or whatever, following immediately. Other rank tests that involve a separate ranking for each test of hypothesis become difficult or impossible to apply. The same may be said for aligned ranks test, in which the appropriate means are first subtracted from each observation before ranking. This resembles an RT-3 type procedure. In addition, for aligned ranks tests some power may be lost in the process (Conover and Iman 1976).

4. MULTIPLE COMPARISONS

Any of the popular multiple comparisons techniques, including Scheffe's, Tukey's, Duncan's, and Fisher's methods, as well as others, may be applied to the RT-1 type data with good results. The power with normal populations is about the same whether the analysis is done on the data or on the ranks. With nonnormal populations the multiple comparisons procedures are more robust and have more power when rank transformed data are used (Iman and Conover 1980).

5. THE ANALYSIS OF COVARIANCE

A natural extension of the rank transformation approach to the general linear model consists of ranking each quantitative variable separately in the general linear model and applying usual parametric procedures. This is an RT-2 type application. The result is new tests for equal slopes, analysis of covariance (Conover and Iman 1982), and any experimental designs covered by the general linear model. These tests are in general not distribution free, except perhaps in some asymptotic sense, but they may be more robust and powerful than their competitors in nonnormal situations. Each procedure, however, needs to be evaluated on its own merits. Another RT-2 type procedure for analysis of covariance is given by Quade (1967).

6. MYTHS OF PARAMETRIC STATISTICS

Many potential users of nonparametric statistics have avoided their use because of mistaken misconceptions surrounding parametric procedures. Paramount among these misconceptions is the attitude that the F-test is robust against the assumption of normality so there is no need to worry about the assumption. This attitude seems to implicitly imply robustness and power go hand in hand. Nothing could be further from the truth. First of all robustness means than the actual alpha level associated with the test is less than or equal to the alpha level desired by the user. Secondly a test can be robust but still lack power. On the other hand tests based on ranks have less problems with being robust while maintaining power.

Sometimes the distinction between parametric and nonparametric statistics is not clear. Typically a parametric procedure requires more assumptions than a nonparametric procedure. For example, if two independent samples are drawn to test H_0: $\mu_1 = \mu_2$ the assumptions are:

For a nonparametric test (Wilcoxon-Mann-Whitney test)

1. Both samples are random samples.
2. The samples are independent of each other.
3. If the null hypothesis is true, then both samples are from the same distribution function.

For a parametric test (t-test)

1.-3. Same as above.
4. Both populations are normal.

The first two assumptions are usually satisfied because they are part of the sampling procedure incorporated in the experimental design. The third assumption is essentially the assumption that is being tested. The fourth assumption, is neither part of the experimental design, nor being tested. It is usually made only to allow the use of a well known procedure, such as the t-test. What are the effects of making such an assumption?

If assumptions 1.-4. are met, the t-test is well known to have the most power for detecting differences in population means. Surprising to many people are the following facts, however:

1. If the populations are normal, the Wilcoxon-Mann-Whitney test is almost as powerful as the t-test.

2. If the populations are not normal, the Wilcoxon-Mann-Whitney test is often much more powerful than the t-test.

Based on the above comments a second myth can be taken care of, and that is that nonparametric statistics are "shortcut" and "less efficient" or "quick and dirty." However, the effects of the normality assumption show that it is really the parametric procedures that are "quick and dirty" as follows.

1. Quick. When the populations are normal, it is easy to use procedures for hypothesis testing and for estimating parameters. These procedures are well known, they are available in easy-to-use computer packages, and they are readily understood and accepted by referees, editors, project leaders, reviewers, etc.

2. Dirty. If the normality assumption is true the gain in power by using a parametric procedure rather than a nonparametric procedure is usually slight. However the loss in power can be considerable when the distributions are actually skewed, or when outliers are present in the data.

7. DISCUSSION

Nonparametric methods should be among the working tools of any statistician. The rank transformation approach provides a vehicle for presenting both the parametric and nonparametric methods in a unified manner. This should enable novice statisticians to understand the differences and similarities of the two types of analysis. Also the rank transformation approach leads to easier computational methods, since it is often more convenient to enter ranks into a program for parametric analysis than it is to find or write a program for a nonparametric analysis. Most existing programs for nonparametric methods do not incorporate corrections for ties, while rank transformation procedures all automatically make the required corrections for ties. The user merely uses average ranks whenever ties occur.

Other scores may be used instead or ranks, if desired, to obtain nonparametric tests that are equivalent to tests such as the van der Waerden test, Capon test, median test, McNemar test, and others. Our experience indicates that the ranks themselves provide scores that are difficult to improve upon for general all-around use.

Some limitations of the rank transformation approach should be noted here also. The rank transformation procedures lead to distribution free tests in some cases, while in other cases the resulting tests may be only conditionally distribution free, asymptotically distribution free, or neither. An example of the latter case arises when the ratio of two sample variances is computed on RT-1 type data and compared with tables of the F distribution as a test for equal variances in two independent samples. Simulation results indicate a severe lack of robustness for this procedure, even when the population means are equal. This is probably related to the fact that the parametric F test for this problem is notoriously sensitive to normality, and the central limit theorem does not apply in this situation.

REFERENCES

CONOVER, W. J. (1980), Practical Nonparametric Statistics (2nd ed.), New York: John Wiley.

CONOVER, W. J., and IMAN, R. L. (1976), "On Some Alternative Procedures Using Ranks for the Analysis of Experimental Designs," Communications in Statistics, A(5), 1348-1368.

CONOVER, W. J. and IMAN, R. L. (1980), "Small Sample Efficiency of Fisher's Randomization Test when Applied to Experimental Designs," unpublished manuscript presented at the annual meeting of the American Statistical Association, Houston, August 1980.

CONOVER, W. J. and IMAN, R. L. (1981), "Rank transformations as a Bridge Between Parametric and Nonparametric Statistics," The American Statistician, 35(3), 124-133.

CONOVER, W. J. and IMAN, R. L. (1982), "Analysis of Covariance Using the Rank Transformation," Biometrics, 38(3), to appear.

IMAN, R. L. (1974a), "Use of a t-statistic as an Approximation to the Exact Distribution of the Wilcoxon Signed Ranks Test Statistic," Communications in Statistics, 3, 795-806.

IMAN, R. L. (1974b), "A Power Study of a Rank Transform for the Two Way Classification Model When Interaction May Be Present," Canadian Journal of Statistics, 2, 227-239.

IMAN, R. L. (1976), "An Approximation to the Exact Distribution of the Wilcoxon-Mann-Whitney Rank Sum Test Statistic," Communications in Statistics, A(5), 587-598.

IMAN, R. L., and CONOVER, W. J. (1976), "A Comparison of Several Rank Tests for the Two-Way Layout," Technical Report SAND76-0631, Sandia Laboratories, Albuquerque, New Mexico.

IMAN, R. L. and CONOVER, W. J. (1980), "Multiple Comparisons Procedures Based on the Rank Transformation," unpublished manuscript presented at the annual meeting of the American Statistical Association, Houston, August 1980.

IMAN, R. L., and DAVENPORT, J. M. (1976), "New Approximations to the Exact Distribution of the Kruskal-Wallis Test Statistic," Communications in Statistics, A(5), 1335-1348.

IMAN, R. L. and DAVENPORT, J. M. (1980), "Approximations of the Critical Region of the Friedman Statistic," *Communications in Statistics*, A(9), 571-595.

IMAN, R. L., HORA, S. C., and CONOVER, W. J. (1982), "A Comparison of Asymptotically Distribution Free Procedures for the Analysis of Complete Blocks," submitted for publication.

QUADE, D. (1967), "Rank Analysis of Covariance," *Journal of the American Statistical Association*, 62, 1187-1200.