DOCUMENT RESUME

ED 219 407 TM 820 418

AUTHOR Koffler, Stephen L.
TITLE A k-Sample Extension to Fligner's Class of Tests For Scale.
PUB DATE [Feb 82]
NOTE 26p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Hypothesis Testing; *Nonparametric Statistics; Research Problems; Statistical Analysis; Statistical Significance
IDENTIFIERS *Fligner (M A); Normal Distribution; Post Hoc Methods; *Violation of Assumptions

ABSTRACT

Too often researchers rely upon the classical normal theory parametric tests to analyze non-normal data, even though the tests may not be robust to violations of that assumption. Fligner's class of two-sample tests for scale is an important development because the test is distribution-free and has desirable properties. This paper outlines the development of the k-sample extension of the two-sample Fligner class of tests, based upon the generalized Puri model. Assuming rejection of the null hypothesis under test, appropriate post hoc procedures for the test were developed based on the chi-square analogue to the Scheffe theorem. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ERIC

ED219407

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A k-Sample Extension To Fligner's Class Of Tests For Scale

Stephen L. Koffler
N.J. State Department of Education

The parametric tests for equality of variance are well known. The classical F-test is typically used to test the hypothesis of equality of two variances while tests developed by Bartlett(1937), Cochran(1941) and Hartley(1950) are probably the most commonly used for the k-sample hypothesis. These tests assume an underlying normal distribution and are quite sensitive to departures from normality (see Box(1953), for example). Thus, when considering data that are from non-normal distributions, alternative nonparametric tests must be employed.

Fligner(1979) has proposed a class of two-sample distribution-free tests which possesses very desirable properties and is an attractive alternative to other nonparametric tests for scale. The present paper extends the Fligner class of tests to the more general k-sample case. For this general case when the null hypothesis is rejected, it is necessary to apply post hoc multiple comparison procedures to determine specific population differences. Thus, this paper will also consider an appropriate post hoc procedure for the k-sample class.

Notation

Let the k independent random samples (X_{ij}, $i = 1, \ldots, n_j$; $j = 1, \ldots, k$) originate from k populations with absolutely continuous cumulative distribution functions $F_1, \ldots, F_k$, respectively. Let $F_j(X) = F(\theta_j(X - \nu_j))$, where θ_j ($\theta_j > 0$) and ν_j are the scale and location parameter, respectively, for

TM 820 418

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
S. L. Koffler
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

$F_j(X)$. Further, let the following quantities be defined:

1. $Q_1, \ldots, Q_N$ denote the order statistics of the combined sample of the N ($N = \Sigma n_j$) observations;
2. $\xi_j = \xi_j(Q_1, \ldots, Q_N)$, $j = 1, \ldots, k$, are measurable functions of the order statistics from the combined samples;
3. $V_{ij} = h(X_{ij}, \xi_1, \ldots, \xi_k)$, $i = 1, \ldots, n_j$, $j = 1, \ldots, k$, where the only requirement on h is that the V_{ij}'s be measurable;
4. R_m, $m = 1, \ldots, N$, is the rank of V_{ij} in the combined sample of N V_{ij}'s.
5. $Z_i^{(j)}$ is an indicator variable such that $Z_i^{(j)} = 1$ if the ith smallest observation is from the jth population, and 0, otherwise.

The null hypothesis of interest is H_0: $\theta_1 = \ldots = \theta_k$, $\nu_1 = \ldots = \nu_k$, while the alternative hypothesis is H_1: $\theta_i < \theta_j$; $\nu_1 = \ldots = \nu_k$, for some pair (i,j).[1] The hypothesis is stated in terms of the scale parameter since the variances do not always exist. Furthermore, attention is restricted to situations such that $F(0) = \frac{1}{2}$ and $F(\cdot)$ is increasing in some neighborhood about zero; thus, the ν_j's are the unique population medians (Fligner, 1979).[2]

Fligner's Two Sample Class of Tests

Prior to developing the k-sample extension to Fligner's class of statistics, it is necessary to illustrate the two sample case.

---

[1] When testing hypotheses about the scale parameters, the appropriateness of many nonparametric procedures depends upon the assumptions made regarding the location parameter. Many of the usual nonparametric tests make the assumption of equality of the population medians (Fligner, 1979). This requirement assures the consistency of the statistics (Fligner & Killeen, 1976).

[2] There are two general types of alternative hypotheses for nonparametric statistics--the first type of alternative is asymmetric one (with common median). It is this type of alternative for which the usual nonparametric tests are most appropriate. However, those usual tests can be used when the alternatives are asymmetric, so long as $F(0) = \frac{1}{2}$. Nonparametric statistics proposed for alternatives having common median are generally very inefficient for alternatives with mass confined to the positive axis since $F(0) = 0$ in this case and the assumption of common median is violated. Thus, a second general type of scale alternative carries the assumption $F(0) = F_i(0) = 0$, $i = 1, \ldots, k$. For these alternatives a shift in scale parameter causes a change in location parameter. Duran (1976) considers tests for scale for both conditions.

Fligner's (1979) class of distribution-free statistics assumes the location parameters are equal. Let $M_p = M_p(X_{11}, \ldots, X_{n_1 1}, X_{12}, \ldots, X_{n_2 2})$ be the pth combined sample quantile, $0 \le p \le 1$.[3] Further,

$$
\begin{aligned}
r &= Np & & 0 < p < \tfrac{1}{2} \text{ and } Np \text{ a positive integer} \qquad (1)\\
&= \{Np + 1\} & & p = \tfrac{1}{2} \text{ or } Np \text{ not a positive integer,}
\end{aligned}
$$

where $\{\cdot\}$ denotes the greatest integer function.

Define $M_p = Q_r$ and $M_{1-p} = Q_{N+1-r}$ for $0 \le p < \frac{1}{2}$. When $p = \frac{1}{2}$, $M_{\frac{1}{2}} = Q_{(N+1)/2}$ for N odd and $M_{\frac{1}{2}} = (Q_{N/2} + Q_{(N/2)+1})/2$ for N even. This completely defines the pth combined sample quantile for $0 \le p \le 1$.

When $0 \le p \le \frac{1}{2}$, let $V_{ij} = h(X_{ij}, M_p, M_{\frac{1}{2}}, M_{1-p})$ be defined as follows:

$$
\begin{aligned}
V_{ij} &= X_{ij} - M_p & & X_{ij} < M_p \qquad (2)\\
&= |X_{ij} - M_{\frac{1}{2}}| & & M_p \le X_{ij} \le M_{1-p}\\
&= X_{ij} - M_{1-p} & & X_{ij} > M_{1-p}\\
&\quad + \operatorname{Max}((M_{1-p} - M_{\frac{1}{2}}), (M_{\frac{1}{2}} - M_p))
\end{aligned}
$$

The statistic

$$
T_{N,p} = \sum_{i=1}^{N} a_{N,p}(R_i) Z_i^{(1)} = \sum_{i=1}^{n_1} a_{N,p}(R_i) \qquad (3)
$$

where $a_{N,p}(i)$, $i = 1, \ldots, N$, is any vector of scores, is distribution-free under H_0: $\theta_1 = \theta_2$, $\nu_1 = \nu_2$ (Fligner, 1979).

---

[3] The symmetry required in Fligner's definition of M_p and M_{1-p} is not present in the definition of the sample quantiles in many texts (e.g., Gibbons, 1971, p. 41).

For $0 \leq p \leq \frac{1}{2}$ and r given as above, Fligner defines his class of statistics, denoted by $\Pi = (T_{N,p}: 0 \leq p \leq \frac{1}{2})$, by using the following scores:

$$
\begin{aligned}
a_{N,p}(i) &= 2i && i \text{ even and } 1 < i < r && (4)\\
&= 2i-1 && i \text{ odd and } 1 \leq i < r \\
&= N-i+r && r \leq i \leq N+1-r \\
&= 2(N-i)+2 && i \text{ even and } N+1-r < i \leq N \\
&= 2(N-i)+1 && i \text{ odd and } N+1-r < i < N
\end{aligned}
$$

when N is even, and

$$
\begin{aligned}
a_{N,p}(i) &= 2i && i \text{ even and } 1 < i < r && (5)\\
&= 2i-1 && i \text{ odd and } 1 \leq i < r \\
&= N-i+r && r \leq i \leq N+1-r \\
&= 2(N-i)+1 && i \text{ even and } N+1-r < i < N \\
&= 2(N-i)+2 && i \text{ odd and } N+1-r < i \leq N
\end{aligned}
$$

when N is odd.

A series of examples will best illustrate the Fligner class. First, it can easily be shown that the statistic $T_{N,\frac{1}{2}}$ is the Siegel-Tukey (1960) statistic.[4] The following example was reported in Penfield (1972).

Example I

An experimenter wishes to determine whether a special training program will influence the abstract reasoning scores of nine year old mentally retarded females. To test his theories he selects 12 (all that were available) nine

[4] The Siegel-Tukey test replaces the combined samples' data with a reordering of the ranks (based on the original data, not the V_{ij}). To illustrate the ranking procedure, consider the following chart (N is assumed to be an even number).

Ordered Score:	Q_1	Q_2	Q_3	Q_4	...	$Q_{N/2}$	...	Q_{N-3}	Q_{N-2}	Q_{N-1}	Q_N
Siegel-Tukey Ranks	1	4	5	8	...	N	...	7	6	3	2

year old girls who have IQ scores recorded between 65 and 75 on the Stanford Binet. He randomly assigns six of the children to the experimental condition and six to the control. After training the experimental group for a month, the experimenter then gives both groups an abstract reasoning test. He believes that the scores of the group receiving special training will have greater dispersion than those of the control group. Is he justified in making this conjecture ($\alpha = 0.05$)?

For this example $n_1 = n_2 = 6$ and $N = 12$. From equation (1), $r = \{(12)\frac{1}{2} + 1\} = 7$. The original data, the V_{ij}, R_m, and $a_{N,\frac{1}{2}}(R_m)$ are presented in the table below. Applying equation (2), $V_{ij} = X_{ij} - M_{\frac{1}{2}}$, for all X_{ij}, where $M_{\frac{1}{2}} = (Q_6 + Q_7)/2 = 24$. Further, equation (4) is used to determine the scores $a_{N,\frac{1}{2}}(R_m)$ since N is even.

TABLE 1

Pertinent Data For Analysis
Of Example I Based Upon $T_{N,\frac{1}{2}}$

Experimental				Control			
X_{ij}	V_{ij}	R_m	$a_{N,\frac{1}{2}}(R_m)$	X_{ij}	V_{ij}	R_m	$a_{N,\frac{1}{2}}(R_m)$
19	-5	1	1	20	-4	2	4
21	-3	3	5	22	-2	4	8
27	3	9	7	23	-1	5	9
30	6	10	6	23	-1	6	12
31	7	11	3	25	1	7	11
35	11	12	2	26	2	8	10
			$\Sigma = 24$				$\Sigma = 54$

Thus, $T_{12,\frac{1}{2}} = \sum_{i=1}^{12} a_{12,\frac{1}{2}}(R_i)Z_i^{(1)} = 24$. It should be clear that the $a_{12,\frac{1}{2}}(R_i)$ given by equation (4) are the Siegel-Tukey ranks derived according to footnote 4. Further, since the Siegel-Tukey statistic is $T_{ST} = \sum_{i=1}^{12} R_{ST_i}Z_i^{(1)}$ where R_{ST_i} are the Siegel-Tukey ranks, $T_{12,\frac{1}{2}} = T_{ST}$. It can be also easily shown that $T_{N,\frac{1}{2}} = T_{ST}$ for N odd.

The other extreme member of the Fligner class, $T_{N,0}$, is a linear function of the statistic $T_N = \sum_{i=1}^{N} R_iZ_i^{(1)}$, where R_i are the ranks of $V_{ij} = |X_{ij} - M_{\frac{1}{2}}|$ for all X_{ij}.[5]

To determine $T_{N,0}$, $r = \{N \cdot 0 + 1\} = 1$, since $Np = 0$. From equation (2) $V_{ij} = |X_{ij} - M_{\frac{1}{2}}|$ for all X_{ij}. Thus, the V_{ij} for $T_{N,0}$ are equal to $|V_{ij}|$ for $T_{N,\frac{1}{2}}$. For illustrative purposes suppose N is even (the same development holds for N odd). Then from equation (4) $a_{N,0}(i) = N - i + 1$, for all i. Thus,

$$\begin{aligned} T_{N,0} &= \sum_{i=1}^{N} a_{N,0}(R_i)Z_i^{(1)} = \sum_{i=1}^{N} (N - R_i + 1)Z_i^{(1)} \qquad (5) \\ &= n_1(N+1) - \sum_{i=1}^{N} R_iZ_i^{(1)} \\ &= n_1(N+1) - T_N \end{aligned}$$

Examining Table 1 and recomputing the values of R_m corresponding to $|V_{ij}|$, $\sum_{i=1}^{12} R_iZ_i^{(1)} = 9+7+6+10+11+12 = 55$.[6] Thus, $T_{N,0} = 6(12 + 1) - 55 = 23$.

It now remains to consider the non-extreme members of the Fligner class. When $0 < p < \frac{1}{2}$, the vector of scores $a_{N,p}(i)$ agrees with those used in computing $T_{N,\frac{1}{2}}$ for $i \leq r$ and $i > N - r + 1$ from equation (4) or (5). These

[5] T_N was proposed by Fligner and Killeen (1976) and is an appealing statistic if the populations are symmetric. See the Fligner and Killeen (1976) article for a complete description of T_N.

[6] It is assumed that the probability of ties is zero. However, some of the ξ's and h functions create ties in the combined sample of V_{ij}'s. If the method for breaking the ties does not distinguish between the samples, any statistic based on the ranks of the V_{ij} will be distribution-free when $F_1(X) = F_2(X)$. (Fligner, 1979). Thus, tieds to the right of the median were assigned lower ranks than those to the left; ties on the same side of the median were broken based on a random approach.

scores are applied to the outer $2(r-1)$ observations, that is, those less than M_p or greater than M_{1-p}.

Let $T^1_{N,p} = \sum_{i=1}^{r-1} a_{N,\frac{1}{2}}(R_i)Z_i^{(1)} + \sum_{i=N-r+2}^{N} a_{N,\frac{1}{2}}(R_i)Z_i^{(1)}$, the sum of the outer scores for the first sample observations. Then for the remaining $N - 2(r-1)$ observations, R_i is the rank of $V_{ij} = |X_{ij} - M_{\frac{1}{2}}|$ for $M_p \leq X_{ij} \leq M_{1-p}$ from equation (4) or (5).

Thus, $T^2_{N,p} = \sum_{i=r}^{N-r+1} (N - i + r)Z_i^{(1)} = \sum_{i=1}^{N-2(r-1)} (N - i + 1)Z_i^{(1)}$, which is equivalent to the portion of the statistic $T_{N,0}$ corresponding to the central $N - 2(r-1)$ observations (from equation (4)). $T_{N,p} = T^1_{N,p} + T^2_{N,p}$. Therefore, $T_{N,p}$ is computed by performing the Siegel-Tukey statistic ($T_{N,\frac{1}{2}}$) on the outer observations and $T_{N,0}$ on the inner ones. $T_{N,p}$ can be considered a compromise between $T_{N,\frac{1}{2}}$ and $T_{N,0}$.

To illustrate $T_{N,p}$ consider Example I and suppose the interest is in determining the value of $T_{N,\frac{1}{4}}$. Since $p = \frac{1}{4}$, $r = 12 \cdot \frac{1}{4} = 3$ from equation (1). Further, $M_{\frac{1}{4}} = Q_3 = 21$ and $M_{1-\frac{1}{4}} = Q_{10} = 30$. Applying equation (2),

$$
\begin{aligned}
V_{ij} &= X_{ij} - 21 & & X_{ij} < 21 \\
&= |X_{ij} - 24| & & 21 \leq X_{ij} \leq 30 \\
&= X_{ij} - 24 & & X_{ij} > 30
\end{aligned}
\qquad (6)
$$

From equation (4), since N is even,

$$
\begin{aligned}
a_{N,\frac{1}{4}}(i) &= \text{[illegible]} & & i \text{ even and } 1 < i < 3 \\
&= 2i-1 & & i \text{ odd and } 1 \leq i < 3 \\
&= 12-i+3 & & 3 \leq i \leq 10 \\
&= 2(12-i)+2 & & i \text{ even and } 10 < i \leq 12 \\
&= 2(12-i)+1 & & i \text{ odd and } 10 < i < 12
\end{aligned}
\qquad (7)
$$

Table 2 outlines the relationship among the original data from Example I, the V_{ij}, R_m and $a_{12,\frac{1}{4}}(R_m)$.

TABLE 2

Pertinent Data For Analysis

Of Example I Based Upon $T_{N,\frac{1}{4}}$

Experimental				Control			
X_{ij}	V_{ij}	R_m	$a_{N,\frac{1}{4}}(R_m)$	X_{ij}	V_{ij}	R_m	$a_{N,\frac{1}{4}}(R_m)$
19	-2	1	1	20	-1	2	4
21	3	9	6	22	2	7	8
27	5	8	7	23	1	4	11
30	6	10	5	23	1	5	10
31	7	11	3	25	1	3	12
35	11	12	2	26	2	6	9
			$\Sigma = 24$				$\Sigma = 54$

Examining the $a_{12,\frac{1}{4}}(R_m)$, it is evident that for $R_m < 3$ and $R_m > 10$, the $a_{12,\frac{1}{4}}(R_m)$ are identical to the Siegel-Tukey values. For $3 \leq R_m \leq 10$ the values of $a_{12,\frac{1}{4}}(R_m) = N + 1 - R_m'$, where $R_m' = R_m - (r - 1)$ (see the discussion of $T_{N,0}$ for the rationale of this translation). From Table 2

$$T_{12,\frac{1}{4}} = \sum_{i=1}^{12} a_{N,\frac{1}{4}}(R_i)Z_i^{(1)} = 24.$$

Fligner(1979) noted that for each value of p, the vector of scores $(a_{N,p}(1), a_{N,p}(2), \ldots, a_{N,p}(N))$ is a rearrangement of the integers 1, ..., N. Applying a theorem by Fligner, Killeen and Hogg(1976), $T_{N,p}$ has the same distribution, under H_0, as the two sample Wilcoxon(1945) statistic, regardless of p.[7] Thus, H_0 is rejected at level α whenever $T_{N,p} \leq C_{n_1,n_2}(\alpha)$, where $C_{n_1,n_2}(\alpha)$ is the lower αth percentile of the two sample Wilcoxon null distribution.

In Example I $C_{6,6}(.05) = 28$. Therefore, the null hypothesis of equality of the scale parameters would have been rejected based on all three statistics, $T_{N,0}$, $T_{N,\frac{1}{4}}$ and $T_{N,\frac{1}{2}}$.

Clearly, had other valuesof p been selected, the null hypothesis may not have been rejected. Thus, one must determine the conditions under which each member of the Fligner class of statistics should be used. Since there are many alternatives, one should select the particular $T_{N,p}$ which will provide the most powerful test, given the constraints imposed as a result of the shape of the distribution of the underlying population of scores. Thus, F(X) must be known to select the appropriate test.

It is infrequent that F(X) is known. For unknown distributions one may use adaptive procedures to discern the nature of the underlying F(X) from the sample data. Adaptive procedures are techniques which use the sample data to select an appropriate model (i.e., in this case the appropriate value of p) and then to make an inference based on the chosen model.

Fligner(1979) used the procedures of Randles and Hogg(1973) and Hogg, Fisher and Randles(1975) to determine an adaptive test that was distribution-free when the assumptions of equal medians was satisfied and was relatively insensitive to small failures in that assumption. An adaptive test is distribution-free when the preliminary selection of the model is statistically independent of the final test.

---

[7] Lehmann(1975) provides a detailed description of the two sample Wilcoxon statistic.

Fligner(1979) used a Monte Carlo study to investigate the behavior of the test procedures and then used those results to obtain the method for selecting a statistic $T_{N,p}$. He based the selection on the tailweight of the distribution. Let Q be defined as follows:

$$Q \;=\; (Q_N - Q_1)/(2\sum_i |Q_i - M_{\frac{1}{2}}|/N) \qquad \text{for } N \leq 20 \qquad (8)$$

$$\;=\; 10(U_{.05} - L_{.05})/(U_{.50} - L_{.50}) \qquad \text{for } N > 20$$

where $U_\beta(L_\beta)$ is the sum of the largest (smallest) $N\beta$ order statistics with fractional items being used when $N\beta$ is not an integer.

Fligner (1979) asserted that if the statistic Q classifies the distribution as heavy tailed, the test is to be based on $T_{N,\frac{1}{2}}$, for medium tail weights, $T_{N,\frac{1}{4}}$, while for lighter tailed distributions, $T_{N,0}$. Smaller values of Q signify lighter tails. Fligner defined the following selection procedure:

Whenever $Q \leq 2.6$, base the test on $T_{N,0}$; (9)

Whenever $2.6 < Q \leq 3.5$, base the test on $T_{N,\frac{1}{4}}$;

Whenever $Q > 3.5$, base the test on $T_{N,\frac{1}{2}}$.

Examining the data from example I to determine the appropriate test statistic, we obtain the following:

$$Q \doteq (35 - 19)/(2 \cdot 46/12) = 16/7.667 \doteq 2.09 \qquad (10)$$

Hence, from equation (9), the appropriate test statistic to use for those data is $T_{N,0}$.

All members of the Fligner two sample class of statistics, with the exception of the Siegel-Tukey equivalent ($T_{N,\frac{1}{2}}$) cannot be computed from the ranks of the original data alone. Thus, they are not rank statistics. However, they are distribution-free under the null hypothesis (Fligner, 1979). The chief difference among the class members is the manner in which each $T_{N,p}$ uses the dispersion information present in the sample data. As Fligner(1979) notes, each observation's dispersion information can be viewed in terms of the observation's distance from some central value or from where it falls in the ordering of the samples.

Finally, Fligner(1979) also showed that for any p_1 and p_2, $0 < p_1, p_2 < 1$, when testing the null hypothesis, the exact Bahadur(1967) efficiency of the test based on T_{N,p_1} relative to the test based on T_{N,p_2} is one when the populations are symmetric. Fligner noted that the Bahadur efficiency result suggests that for moderate sample sizes, under the assumptions of the null hypothesis, the power properties of the various members of the class should be similar.

k-sample Extension To Fligner's Class

Because of the broad range of use for the Fligner class, it is desirable to extend it to the more general, and probably more frequently occurring, k sample problem. Puri(1964) has developed a generalized k sample testing procedure for considering this problem. Previously, Penfield and Koffler(1979) have derived and compared the k sample analogues to the two sample Mood(1954), Siegel-Tukey(1960) and Klotz(1962) tests based on Puri's methods.

Puri's statistic is defined as follows:

$$L = \sum_{j=1}^{k} n_j (S_{N,j} - \mu_{N,j})^2 / A^2 \qquad (11)$$

where

$$S_{N,j} = \frac{1}{n_j} \sum_{i=1}^{N} E_{N,i} Z_i^{(j)} \qquad (12)$$

$\mu_{N,j}$ and A^2 are normalizing constants that do not depend on j and are equivalent to $E(S_{N,j}|H_0)$ and $Var(S_{N,j}|H_0)$, respectively.

$E_{N,i}$ is a variable which permits the substitution of a variety of statistical quantities. $E_{N,i} = i$ corresponds to the Siegel-Tukey statistic, $E_{N,i} = (i - N+1/2)^2$ to the Mood test and $E_{N,i} = (\Phi^{-1}(i/N+1))^2$, where Φ is the standard normal cumulative distribution function, to Klotz's test. Puri(1968) showed that equation (11) is asymptotically distribution-free when each sample is adjusted for its sample median, provided the populations are symmetric. Using theorems from Chernoff and Savage(1958), Puri(1965) showed that the limiting distribution of (11) is χ^2 with k-1 degrees of freedom, central under H_0 and noncentral under H_1, when the values of $\mu_{N,j}$ and A^2 are $E\ S_{N,j}|H_0)$ and $Var(S_{N,j}|H_0)$, respectively.

The values of $E(S_{N,j}|H_0)$ and $Var(S_{N,j}|H_0)$ can be derived from the methods of Lehmann(1975):

$$E(S_{N,j}|H_0) = \sum_{i=1}^{N} E_{N,i}/N = \bar{E}_{N,i} \tag{13}$$

$$Var(S_{N,j}|H_0) = \frac{N - n_j}{n_j N(N-1)} \sum_{i=1}^{N} (E_{N,i} - \bar{E}_{N,i})^2, \tag{14}$$

To generalize the Fligner class of two sample statistics to the k sample case, $E_{N,i}$ is defined by equation (4) or (5) for N even or odd, respectively. Given the many members of Fligner's class (i.e., $0 \leq p \leq \frac{1}{2}$, for N even or odd, with different values of r depending on the relationship of N, p and V), it might appear that the derivation of equations (13) and (14) would be very cumbersome and complex. This is not so.

As Fligner(1979) has noted, for each value of p (regardless of whether N is even or odd), the vector of scores $a_{N,p}(1), \ldots, a_{N,p}(N)$ is just a rearrangement of the integers (1, ... ,N). Thus, it should be clear that one could determine $E(S_{N,j}|H_0)$ and $Var(S_{N,j}|H_0)$ as follows (Lehmann, 1975):

$$E(S_{N,j}|H_0) = \sum_{i=1}^{N} E_{N,i}/N = \sum_{i=1}^{N} i/N = \frac{N(N+1)}{2N} = \frac{N+1}{2} \qquad (15)$$

$$Var(S_{N,j}|H_0) = \frac{N - n_j}{n_j N(N-1)} \sum_{i=1}^{N} (E_{N,i} - \bar{E}_{N,i})^2$$

$$= \frac{N - n_j}{n_j N(N-1)} \{ \sum_{i=1}^{N} i^2 - \frac{N(N+1)^2}{4} \}$$

$$= \frac{N - n_j}{n_j N(N-1)} \{ \frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{4} \}$$

$$= \frac{(N - n_j)(N + 1)}{12n_j} . \qquad (16)$$

Thus, given equations (11), (15) and (16), the form of the k-sample Fligner test can be represented as

$$L = \frac{\sum_{j=1}^{k} n_j (\sum_{i=1}^{N} a_{N,p}(R_i)Z_i^{(j)}/n_j - \frac{(N+1)}{2})^2}{(N - n_j)(N + 1)/12n_j} \qquad (17)$$

which simplifies to the following:

$$L = \frac{12}{(N+1)} \sum_{j=1}^{k} (a_{N,p}(R_i)Z_i^{(j)} - n_j\frac{(N+1)}{2})^2/(N - n_j)$$

$$= \frac{12}{(N+1)} \sum_{j=1}^{k} (T_{N,p}^{(j)} - n_j\frac{(N+1)}{2})^2/(N - n_j) \qquad (18)$$

where $T_{N,p}^{(j)}$ is the value of $T_{N,p}$ for the jth sample.

Equation (18) is appropriate regardless of the value of p. Further, H_0 is rejected if $L > \chi^2_{k-1}(1-\alpha)$.

It is instructive to illustrate the k-sample Fligner test with an example. We shall simultaneously examine $T_{N,0}$, $T_{N,\frac{1}{4}}$ and $T_{N,\frac{1}{2}}$.

Example II

In a study by Kerst and Levin(1973) imagery and sentence mediators that linked the stimuli and responses of pictorial paired associates were either provided by an experimenter or generated by fourth and fifth-grade students. Subjects were randomly assigned to conditions. The four strategy conditions under study were as follows:

1. subject-generated (sentence)
2. subject-generated (imagery)
3. experimenter-provided (sentence)
4. experimenter-provided (imagery)

Scores represent the number of correct responses to 20 paired-associate learning items. The experimenters note that although the four strategy conditions did not differ among themselves with respect to central tendency, an examination of the variances using a parametric test showed them to be significantly different. The data suggest that the variances of the two experimenter-provided conditions were substantially less than those of the two subject-generated conditions.

To analyze the data using Fligner's k-sample statistic, a sample of ten scores from each condition was selected for illustrative purposes. These data and the V_{ij}, R_k and $a_{N,p}(R_k)$ corresponding to $T_{N,0}$, $T_{N,\frac{1}{4}}$ and $T_{N,\frac{1}{2}}$ are listed in Table 3. Appendix A contains the necessary information for the derivation of the figures in Table 3.

TABLE 3

Pertinent Data For Analysis of Example II

Based On the k-Sample Extension to $T_{N,0}$, $T_{N,\frac{1}{4}}$ and $T_{N,\frac{1}{2}}$

Condition	Score	Q_k	$T_{N,0}$			$T_{N,\frac{1}{4}}$			$T_{N,\frac{1}{2}}$		
			V_{ij}	R_k	$a_{N,0}$	V_{ij}	R_k	$a_{N,\frac{1}{4}}$	V_{ij}	R_k	$a_{N,\frac{1}{2}}$
I	4	3	9	38	3	-4	3	5	-9	3	5
(n_1 = 10)	5	4	8	36	5	-3	4	8	-8	4	8
	5	5	8	37	4	-3	5	9	-8	5	9
	6	8	7	35	6	-2	8	16	-7	8	16
	6	6	7	33	8	-2	6	12	-7	6	12
	16	28	3	17	24	3	26	24	3	28	26
	17	35	4	24	17	4	33	15	4	35	11
	17	33	4	22	19	4	31	19	4	33	15
	19	38	6	29	12	7	38	6	6	38	6
	19	37	6	28	13	7	37	7	6	37	7
				Σ = 111			Σ = 121			Σ = 115	
II	1	1	12	40	1	-7	1	1	-12	1	1
(n_2 = 10)	3	2	10	39	2	-5	2	4	-10	2	4
	6	7	7	34	7	-2	7	13	-7	7	13
	7	9	6	30	11	-1	9	17	-6	9	17
	8	10	5	27	14	5	35	11	-5	10	20
	17	31	4	20	21	4	29	21	4	31	19
	17	34	4	23	18	4	32	18	4	34	14
	18	36	5	26	15	6	36	10	5	36	10
	20	39	7	31	10	8	39	3	7	39	3
	20	40	7	32	9	8	40	2	7	40	2
				Σ = 108			Σ = 100			Σ = 103	
III	9	11	4	25	16	4	34	14	-4	11	21
(n_3 = 10)	12	15	1	5	36	1	14	36	-1	15	29
	13	21	0	2	39	0	12	38	0	21	39
	13	19	0	1	40	0	10	40	0	19	37
	14	22	1	4	37	1	13	37	1	22	38
	15	25	2	11	30	2	20	30	2	25	32
	15	27	2	13	28	2	22	28	2	27	27
	15	26	2	12	29	2	21	29	2	26	29
	16	30	3	19	22	3	28	22	3	30	22
	17	32	4	21	20	4	30	20	4	32	18
				Σ = 297			Σ = 294			Σ = 292	
IV	11	12	2	14	27	2	23	27	-2	12	24
(n_4 = 10)	11	13	2	15	26	2	24	26	-2	13	25
	11	14	2	16	25	2	25	25	-2	14	28
	12	17	1	7	34	1	16	34	-1	17	33
	12	16	1	6	35	1	15	35	-1	16	32
	12	18	1	8	33	1	17	33	-1	18	36
	13	20	0	3	38	0	11	39	0	20	40
	15	24	2	10	31	2	19	31	2	24	34
	15	23	2	9	32	2	18	32	2	23	35
	16	29	3	18	23	3	27	23	3	29	23
				Σ = 304			Σ = 305			Σ = 310	

Ties among the data were broken as follows: ties to the right of the combined sample median were assigned lower ranks than those to the left; ties on the same side of the median were broken based on a random approach.

From equation (18) the following values for L were determined:

$$L_0 = \frac{12}{41}\left\{\frac{\left(111 - \frac{10\cdot41}{2}\right)^2}{30} + \frac{(108 - 205)^2}{30} + \frac{(297 - 205)^2}{30} + \frac{(304 - 205)^2}{30}\right\}$$

$$= 356.20$$

$$L_{\frac{1}{4}} = \frac{12}{41\cdot30}\left\{(121-205)^2 + (100 - 205)^2 + (294 - 205)^2 + (305 - 205)^2\right\}$$

$$= 351.24$$

$$L_{\frac{1}{2}} = \frac{12}{41\cdot30}\left\{(115 - 205)^2 + (103 - 205)^2 + (292 - 205)^2 + (310 - 205)^2\right\}$$

$$= 361.93$$

The null hypothesis was one of no difference in scale among the four types of learning strategies. For $\alpha = .05$, H_0 is rejected if $L > \chi^2_{k-1}(1-\alpha) = \chi^2_3(.95) = 7.81$. Thus, regardless of whether one uses $T_{N,0}$, $T_{N,\frac{1}{4}}$ or $T_{N,\frac{1}{2}}$, the null hypothesis is rejected.

Similar to the two sample case, the question now arises as to which member of the Fligner class is most appropriate for different situations. Puri(1964) noted that in general the efficiency of the L statistic based upon k-samples agrees with the efficiency of the two sample test statistic. Using that information, one could argue that the adaptive procedure derived by Fligner(1979) for the two sample problem could be applied to the k-sample case.[8]

[8] At this point it is conjecture that the two sample adaptive test could be extended to the k-sample case. A Monte Carlo study, similar to the one conducted by Fligner(1979) would provide valuable information in this regard.

Using equation (8),

$$U_{.05} = Q_{40} + Q_{39} = 20 + 20 = 40 \text{ (since } .05 \cdot N = 2);$$

$$L_{.05} = Q_1 + Q_2 = 1 + 3 = 4;$$

$$U_{.50} = \sum_{i=21}^{40} Q_i = 331;$$

$$L_{.50} = \sum_{i=1}^{20} Q_i = 167;$$

$$Q = 10(40 - 4)/(331 - 167) = 2.195.$$

From equation (9) since $Q = 2.195 < 2.6$, the appropriate test is based on $T_{N,0}$.

Post Hoc Procedures

When considering $k > 2$ samples, it is not sufficient to simply reject the null hypothesis of equality of the k scale parameters. Should this hypothesis be rejected, it is necessary to determine the specific reasons for the rejection (i.e., to determine which of the θ_j's are significantly different).

The use of a posteriori or post hoc procedures can be used for such determination. Significant differences among the scale values of the respective populations are determined by using post hoc procedures for testing meaningful contrasts of the θ_j's.

A contrast of the parameters $\theta_1, \ldots, \theta_k$ is a linear combination of the

θ_j's and is defined as $\Psi = \sum_{j=1}^{k} c_j\theta_j$ where the c_j's are known constant coefficients, subject to the restriction that $\sum_{j=1}^{k} c_j = 0$. The sample estimate of the population contrast is $\hat{\Psi} = \sum_{j=1}^{k} c_j\hat{\theta}_j$, where $\hat{\theta}$ is the sample estimate of the scale parameter θ. Furthermore, the variance of $\hat{\Psi}$ for independent random samples is given by the following:

$$\hat{\sigma}^2_{\hat{\Psi}} = \mathrm{Var}(\hat{\Psi}) = \mathrm{Var}\left(\sum_{j=1}^{k} c_j\hat{\theta}_j\right)$$

$$= \sum_{j=1}^{k} c_j^2\mathrm{Var}(\hat{\theta}_j) + \sum_{j=1}^{k}\sum_{\substack{j'=1\\ j\neq j'}}^{k} c_jc_{j'}\mathrm{Cov}(\hat{\theta}_j,\hat{\theta}_{j'}) \qquad (19)$$

When $\hat{\theta}_j$ is defined according to equation (12), $\hat{\sigma}^2_{\hat{\Psi}}$ is derived as follows:

$$\hat{\sigma}^2_{\hat{\Psi}} = \sum_{j=1}^{k} \frac{c_j^2(N - n_j)}{n_jN(N-1)} \sum_{i=1}^{N} (E_{N,i} - \bar{E}_{N,i})^2 + \sum_{j=1}^{k}\sum_{\substack{j'=1\\ j\neq j'}}^{k} \frac{-c_jc_{j'}}{N(N-1)} \sum_{i=1}^{k} (E_{N,i} - \bar{E}_{N,i})^2 \qquad (20)$$

Since $\sum_{j=1}^{k} c_j = 0$, it follows that $\left(\sum_{j=1}^{k} c_j\right)^2 = \sum_{j=1}^{k} c_j^2 + \sum_{j=1}^{k}\sum_{\substack{j'=1\\ j\neq j'}}^{k} c_jc_{j'} = 0$.

Hence, $\sum_{j=1}^{k} c_j^2 = -\sum_{j=1}^{k}\sum_{\substack{j'=1\\ j\neq j'}}^{k} c_jc_{j'}$. Substituting this information into equation (20) and simplifying, we obtain the following:

$$\hat{\sigma}^2_{\hat{\Psi}} = \frac{1}{N - 1} \sum_{j=1}^{k} \frac{c_j^2}{n_j} \sum_{i=1}^{N} (E_{N,i} - \bar{E}_{N,i})^2 \qquad (21)$$

Scheffé(1953) has proposed a method based upon the F distribution for testing contrasts. Scheffé's theorem states that in the limit the probability is 1-α that the values of all contrasts (Ψ) will simultaneously satisfy the inequality

$$\hat{\Psi} - S\sigma_{\hat{\Psi}} \leq \Psi \leq \hat{\Psi} + S\sigma_{\hat{\Psi}} \text{, where } S = (k-1)(F_{k-1,N-k}(1-\alpha)) \quad (22)$$

Gold(1963), Goodman(1964) and Marascuilo(1966) have extended Scheffé's simultaneous confidence interval method to encompass the χ^2 distribution instead of the F distribution.[9] The analogue states that in the limit the probability is 1-α that all linear contrasts of the form $\Psi = \sum_{j=1}^{k} c_j\theta_j$ simultaneously satisfy the inequality given by equation (22) where $S = (\chi^2_{k-1}(1-\alpha))^{\frac{1}{2}}$

If the overall null hypothesis is rejected, there is at least one contrast (Ψ) that is significantly different from zero. Equation (22) can be used to determine the significant contrasts. If the confidence interval does not contain the value zero, one would reject the null hypothesis H_0: $\Psi = 0$ in favor of H_1: $\Psi \neq 0$.

For the Puri generalized k sample statistic, the form of the confidence interval is given by the following:

$$\sum_{j=1}^{k} c_j S_{N,j} - (\chi^2_{k-1}(1-\alpha)\mathrm{Var}(\sum_{j=1}^{k} c_j S_{N,j}))^{\frac{1}{2}} \leq \Psi$$

$$\leq \sum_{j=1}^{k} c_j S_{N,j} + \chi^2_{k-1}(1-\alpha)\mathrm{Var}(\sum_{j=1}^{k} c_j S_{N,j}))^{\frac{1}{2}} \quad (23)$$

---

[9] Marascuilo and McSweeney(1967) present a proof of the χ^2 analogue to Scheffé's theorem.

For the Fligner procedure, equation (23) becomes the following:

$$\frac{\sum_{j=1}^{k} c_j T_{N,p}^{(j)}}{n_j} - \left(\chi^2_{k-1}(1-\alpha)\frac{\{N\cdot(N+1)\}}{12}\frac{\sum_{j=1}^{k} c_j^2}{n_j}\right)^{\frac{1}{2}} \leq \Psi \leq \frac{\sum_{j=1}^{k} c_j T_{N,p}^{(j)}}{n_j}$$

$$+ \left(\chi^2_{k-1}(1-\alpha)\frac{\{N\cdot(N+1)\}}{12}\frac{\sum_{j=1}^{k} c_j^2}{n_j}\right)^{\frac{1}{2}} \qquad (24)$$

In Example II the null hypothesis of equality of the four scale parameters was rejected; thus, it is appropriate to consider post hoc procedures to determine which conditions were yielding significantly different results. All six pairwise contrasts and a complex contrast were considered. The complex contrast examined whether there was a significant difference between conditions 1 and 2 (subject-generated) and 3 and 4 (experimenter-generated). Table 4 presents the relevant information to evaluate the significance of the contrasts.

Table 3

Post Hoc Procedures for Example II

Contrast	Estimated Contrast ($\hat{\Psi}$)	Estimated Variance ($\sigma^2_{\hat{\Psi}}$)	Upper Limit	Lower Limit	Decision
Ψ_1	$S_{N,1} - S_{N,2} = 0.3$	27.3	-14.3	14.9	NS
Ψ_2	$S_{N,1} - S_{N,3} = -18.6$	27.3	-33.2	-4.0	SIG
Ψ_3	$S_{N,1} - S_{N,4} = -19.3$	27.3	-33.9	-4.7	SIG
Ψ_4	$S_{N,2} - S_{N,3} = -18.9$	27.3	-33.5	-4.3	SIG
Ψ_5	$S_{N,2} - S_{N,4} = -19.6$	27.3	-34.2	-5.0	SIG
Ψ_6	$S_{N,3} - S_{N,4} = -0.7$	27.3	-15.3	13.9	NS
Ψ_7	$(S_{N,1}+S_{N,2})/2 - (S_{N,3}+S_{N,4})/2 = -19.1$	13.7	-29.4	-8.8	SIG

From Table 4 it can be seen that the subject-generated conditions (Conditions 1 and 2, Ψ_1) were not significantly different from each other, nor were the experimenter-provided ones (Conditions 3 and 4, Ψ_6). However, the significanceof Ψ_2, Ψ_3, Ψ_4, and Ψ_5 indicates that each of the subject-generated conditions was significantly different from each of the experimenter-provided ones ($p < .05$). Furthermore, the average response to the subject-generated conditions was also significantly different from the average response to the experimenter-provided ones (Ψ_7). Thus, one could conclude that the spread in scores was not affected by which of the two subject-generated conditions nor which experimenter-provided conditions was used. However, the spread of scores was significantly different depending upon whether a subject or experimenter condition was used.

Summary

Behavioral science data are frequently non-normal. However, too often researchers rely upon the classical normal theory parametric tests to analyze such data even though the tests may not be robust to violations of that assumption. Fligner's class of two sample tests for scale is an important development because the test is distribution-free and has desirable properties.

Since researchers typically consider more than two samples, it is equally important to develop similar procedures for the more general k-sample case. This paper outlined the development of the k-sample extension to the two-sample Fligner class of tests, based upon the generalized Puri model. Assuming rejection of the null hypothesis under test, appropriate post hoc procedures for the test were developed based on the chi-square analogue to the Scheffé theorem.

APPENDIX A

	$T_{N,0}$	$T_{N,\frac{1}{4}}$	$T_{N,\frac{1}{2}}$
r	1	10	21
M_p	1	8	13
M_{1-p}	20	17	13
V_{ij}	$\lvert X_{ij} - 13\rvert$ $1\le X_{ij}\le 20$	$X_{ij} - 8$ $X_{ij}<8$ $\lvert X_{ij} - 13\rvert$ $8 \le X_{ij} \le 17$ $X_{ij} - 12$ $17 > X_{ij}$	$X_{ij} - 13$ $1\le X_{ij}\le 20$
$a_{N,p}(i)$	$41 - i$ $1\le i\le 40$	$2i$ i even $1<i<10$ $2i-1$ i odd $1\le i<10$ $50-i$ $10\le i\le 31$ $2(40-i)+2$ i even $31<i\le 40$ $2(40-i)+1$ i odd $31<i<40$	$2i$ i even $1<i<21$ $21-1$ i odd $1\le i<21$ $2(40-i)+2$ i even $20<i\le 40$ $2(40-i)+1$ i odd $20<i<40$

REFERENCES

Bahadur, R.R. (1960). Asymptotic efficiency of tests and estimates. Sankya, 22, 229-252.

Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. Proceedings of the Royal Society A, 160, 268-282.

Bradley, J.V. (1968). Distribution-Free Statistical Tests. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Box, G.E.P. (1953). Non-normality and tests on variances. Biometrika, 40, 318-335.

Chernoff, H. & Savage, J.R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. Annals of Mathematical Statistics, 29, 972-994.

Cochran, W.G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. Annals of Eugenics, 11, 47-52.

Duran, B.S. (1976). A survey of nonparametric tests for scale. Communications in Statistics--Theoretical Methods, A5(14), 1287-1312.

Fligner, M.A. (1979). A class of two-sample distribution-free tests for scale. Journal of the American Statistical Association, 74, 889-893.

Fligner, M.A. & Killeen, T.J. (1976). Distribution-free two sample tests for scale. Journal of the American Statistical Association, 71, 210-213.

Fligner, M.A., Hogg, R.V. & Killeen, T.J. (1976). Some distribution-free rank-like statistics having the Mann-Whitney-Wilcoxon null distribution. Communications in Statistics--Theoretical Methods, A5(4), 373-376.

Gibbons, J.D. (1971). Nonparametric Statistical Inference. New York: McGraw-Hill.

Govindarajulu, Z. (1976). A brief survey of nonparametric statistics. Communications in Statistics--Theoretical Methods, A5(5), 429-453.

Gold, R.Z. (1963). Tests auxiliary to χ^2 tests in the Markov chain. Annals of Mathematical Statistics, 31, 56-74.

Goodman, L.A. (1964). Simultaneous confidence intervals for contrasts among multinomial populations. Annals of Mathematical Statistics, 35, 716-725.

Hajek, J. & Sidak, Z. (1967). Theory of Rank Tests. New York: Academic Press.

Hartley, H.O. (1950). The maximum F-ratio as a short-cut test for heterogeneity of variance. Biometrika, 37, 308-312.

Hogg, R.V., Uthoff, V.A., Randles, R.H. & Davenport, A.S. (1972). On the selection of the underlying distribution and adaptive estimation. Journal of the American Statistical Association, 67, 597-600.

Hogg, R.V., Fisher, D.M. & Randles, R.H. (1975). A two-sample adaptive distribution-free test. Journal of the American Statistical Association, 70, 656-661.

Kerst, S. & Levin, J. (1975). A comparison of experimenter-provided and subject-generated strategies in children's paired-associate learning. Journal of Educational Psychology, 65, 300-303.

Klotz, J.H. (1962). Nonparametric tests for scale. Annals of Mathematical Statistics, 33, 495-512.

Lehmann, E.L. (1975). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day Inc.

Marascuilo, L.A. (1966). Large-sample multiple comparisons. Psychological Bulletin, 65, 280-290.

Marascuilo, L.A. & McSweeney, M. (1967). Multiple contrast methods for analytic surveys. Proceedings of the American Statistical Association, Social Statistics Section, 336-343.

Mood, A.M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. Annals of Mathematical Statistics, 25, 514-522.

Moses, L.E. (1963). Rank tests of dispersion. Annals of Mathematical Statistics, 34, 973-983.

Noether, G.E. (1955). On a theorem of Pitman. Annals of Mathematical Statistics, 26, 64-68.

Penfield, D.A. (1972). A comparison of some nonparametric tests for scale. Paper presented at the annual meeting of the American Educational Research Association.

Penfield, D.A. & Koffler, S.L. (1978). Post hoc nonparametric tests for scale. Journal of Educational Statistics, 3, 265-282.

Penfield, D.A. & Koffler, S.L., (1979) A comparison of some k-sample tests for scale. Journal of Experimental Education, 47, 126-130.

Penfield, D.A. & Koffler, S.L. (1979). Two sample nonparametric test selection based upon tailweight and skewness. Paper presented at the annual meeting of the American Educational Research Association.

Puri, M.L. (1964). Asymptotic efficiency of a class of c-sample tests. Annals of Mathematical Statistics, 35, 102-121.

Puri, M.L. (1965). On some tests of homogeneity of variance. Annals of the Institute of Statistical Mathematics, 17, 323-330.

Puri, M.L. (1968). Multi-sample scale problem: unknown location parameters. Annals of the Institute of Statistical Mathematics, 20, 99-106.

Randles, R.H. & Hogg, R.V. (1973). Adaptive distribution-free tests. Communications in Statistics--Theoretical Methods, 2, 337-356.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. Biometrika, 40, 87-104.

Siegel, S. & Tukey, J.W. (1960). A nonparametric sum of ranks procedure for relative spread in unpaired samples. Journal of the American Statistical Association, 55, 429-445.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics, 1, 80-83.