# Robust Dual Scaling with Tukey's Biweight

John Sachs
Hong Kong

Use of the method of reciprocal biweighted means (MBM) for dealing with the outlier problem in dual scaling compared favorably with other robust estimation procedures, such as the method of trimmed reciprocal averages (MTA). Like the MTA, the MBM was easy to implement and it converged to a stable point when a two-step estimation procedure was used. One advantage of the MBM over the MTA was that it afforded greater control in fine tuning the final solution. Empirical results for four datasets, some containing multiple outliers, are presented. *Index terms: biweight, dual scaling, outliers, reciprocal averages, robust estimation, Tukey's biweight.*

Dual scaling is a technique for the analysis of categorical (non-numerical, qualitative) data (Nishisato, 1979, 1980, 1981, 1988). This technique also has been called correspondence analysis (Greenacre, 1984), the method of reciprocal averages (Horst, 1935), optimal scaling (Bock, 1960), and Guttman weighting (Guttman, 1941).

Essentially, dual scaling involves finding weights for rows and columns of two-dimensional categorical data tables (such as contingency tables or response pattern tables) that maximize Guttman's (1941) criterion of internal consistency. Under the criterion of internal consistency, weights are found for columns (rows) so that scores within rows (columns) are as similar as possible and that scores between rows (columns) are as different as possible. That is, these optimal column (row) weights minimize the within row (column) sum of squares ($SS_w$) and maximize the between row (column) sum of squares ($SS_b$) relative to the total sum of squares ($SS_t$) for the table. Therefore, the dual scaling weights maximize the squared correlation ratio eta-squared ($\eta^2$) (Nishisato, 1980)

$$\eta^2 = \frac{SS_b}{SS_w} \tag{1}$$

and are optimal in the sense that no other set of weights can result in a larger $\eta^2$. This formulation of dual scaling is referred to as the one-way analysis of variance (ANOVA) formulation. An important characteristic of dual scaling illustrated by this formulation is the property of duality, which means that it does not matter whether row weights or column weights are found first because both result in the same maximized value of $\eta^2$ (Nishisato, 1980).

The dual scaling weights that maximize $\eta^2$ also maximize the generalized Kuder-Richardson reliability ($\alpha$) for multiple-choice data (Nishisato, 1980). The relationship between $\alpha$ and $\eta^2$ given by Lord (1958) is

$$\alpha = 1 - \frac{(1-\eta^2)}{\eta^2(n-1)}, \tag{2}$$

where $n$ is the number of items. Detailed mathematical presentations of dual scaling can be found in Nishisato (1980, 1981, 1988).

An important theoretical and applied issue regarding dual scaling weights is the effect of outliers on these weights and how methods for dealing with such outliers affect the optimal solution obtained. Nishisato (1987) characterizes outliers in dual scaling as those observations that result in spuriously large or small item-total correlations and item SS, spurious interitem correlations, and extreme item option weights (too large or too small).

The issue of outliers also has been an important problem in related areas such as principal components analysis, in which the data are continuous. Jolliffe (1986, pp. 195–197) summarized research on robust estimation of principal components, which

*APPLIED PSYCHOLOGICAL MEASUREMENT*
*Vol. 18, No. 4, December 1994, pp. 301–309*

*0146-6216/94/040301-09$1.70*

used among other methods, replacement of weighted means in singular value decomposition with medians, weighted trimmed means, or other robust measures of location. Ammann (1993) and Verboon (1994) also addressed issues of robust estimation in principal components analysis.

In dual scaling, Nishisato (1984, 1987) showed the detrimental effects of outliers on the scaling weights and proposed the following robust quantification methods for dealing with the problem: (1) the method of reciprocal medians (MRM); (2) alternating the MRA and the MRM; (3) the method of trimmed reciprocal averages (MTA), in which the means are replaced with a trimmed mean; (4) the method of generalized forced classification; and (5) the method of projecting onto a subspace. Methods 1 and 2 proved less than optimal and Methods 3, 4, and 5 were promising. However, Methods 4 and 5, although providing results similar to Method 3, were more computationally demanding than Method 3. Therefore, the MTA (Method 3) proposed by Nishisato (1987) is a very practical and computationally simple robust method for dealing with the outlier problem in dual scaling.

However, the trimmed mean is only one type of robust location estimator known as L estimators (Hempel, 1974; Hogg, 1979; Huber, 1972); the median, Winsorized means, and Tukey's trimean are other examples of L estimators. L estimators are weighted linear combinations of the ordered observations. A disadvantage of an L estimator such as the α-trimmed mean, where α indicates the percentage of trim from each end of the ordered observations, is that the trimmed observations are given 0 weight, and those observations retained are given the same weight. This is the case regardless of the influence the retained observations have on the location estimate.

Another class of robust location estimators, called M estimators (Huber, 1972), are more efficient than L estimators. M estimators weight observations by the inverse of their influence on the estimate of location. That is, extreme observations that have a detrimental influence on the estimate of location are given less weight than those with less detrimental influence.

### Tukey's Biweight

One such M estimator is Tukey's biweight (Mosteller & Tukey, 1977), which is defined as:

$$\bar{x}_b = \frac{\sum_{i=1}^{N} W_i X_i}{\sum_{i=1}^{N} W_i}, \tag{3}$$

where

$N$ is the number of observations;

$i$ indicates the $i$th observation, where $i$=1, 2, ..., $N$; and

the weight, $W_i$, is given by

$$W_i = \begin{cases} \left(1 - Z_i^2\right)^2 & |Z_i| \le 1 \\ 0 & otherwise \end{cases}, \tag{4}$$

and $Z_i$, the distance of an observation $x_i$ from an estimate of location, is given by

$$Z_i = \frac{x_i - x^*}{S \times k}, \tag{5}$$

where

$x^*$ indicates some robust location estimate such as the median,

$S$ indicates a robust measure of spread such as the interquartile range, and

$k$ is a multiplicative factor usually fixed at 3 (Mosteller & Tukey, 1977).

Thus, very extreme observations are given 0 weight, and the weight given to other observations is adjusted to reflect their influence. The biweight estimate can be iterated so that the initial estimate is substituted for $x^*$ and the process is repeated until the difference between $x_{b(i+1)}$ and $\bar{x}_{b(i)}$ is some prespecified difference (e.g., .001).

Because the biweight is sensitive to an observation's influence, adjusts the weight to reflect this influence, and is easy to compute, it is an ideal candidate for computing robust dual scaling weights. The purpose of this study, therefore, was to investigate the characteristics of the biweighted dual scaling solution and to compare these results with the MTA.

### Method

Following Nishisato (1984, 1987) for the MRM, the row weights, $y_i$, are the median of the responses weighted by $x_j$ in row $i$ ($i = 1, 2, ..., r$; where $r$ is the number of rows), and the column weights, $x_j$, are the median of the responses weighted by $y_i$ in column $j$ ($j = 1, 2, ..., c$; where $c$ is the number of columns). To obtain the biweight results, the median operation above is replaced by the biweight operation. The biweighted mean is found from Equations 3–5, using the median for $x^*$ and the interquartile range for $S$ in Equation 5. After each iteration, the origins and units of the row weights, $y_i$, and column weights, $x_j$, are adjusted following Nishisato (1984). This method is referred to here as the method of biweighted means (MBM).

Although the MRA always converges to a stable point (see Nishisato, 1980, pp. 60–61) for arbitrary (nonidentical) starting weights (row or column), this is not true of the MRM (Nishisato, 1984). A pilot study conducted using the MBM showed that it converged for most datasets, but there was some oscillation for $k < 3$. By specifying a less stringent convergence criterion, such as .01, this oscillation often was eliminated. Of greater concern was the fact that the choice of initial starting weights sometimes gave slightly different results. Nishisato (1984) reported a similar effect with the MRM.

To obviate this problem, a two-step procedure was used. The constant $k$ in Equation 5 was set to an arbitrarily high value (99). This gave each observation in Equation 3 the same weight of 1 so that the biweighted mean was identical to the mean or average. Therefore, the MBM was identical to the MRA in Step 1 of the two-step procedure, and the optimal weights were obtained. These optimal weights were used as initial weights in Step 2, where $k$ was fixed to values ranging from 1.5 to 5 (values in this range were found effective in a pilot study). Nishisato (1984) suggested using this same two-step procedure for dealing with the problem of initial weights and the problem of oscillation with the MRM.

To evaluate the effectiveness of each method in dealing with outliers, the following summary measures that are used frequently in evaluating dual scaling solutions (Nishisato, 1980) were computed: (1) the item-total correlations (product-moment correlations between the optimal item scores and the optimal person scores), (2) the item SS, and (3) $\eta^2$. Although the internal consistency reliability coefficient, $\alpha$, can be obtained directly from Equation 2, it is also reported, following standard practice in reporting dual scaling results (Nishisato, 1980). The product-moment correlations were used as criteria for evaluating the solutions by examining their contribution to the maximized $\eta^2$, which is equal to the average of the squared correlations between item scores and the total scores,

$$\eta^2 = \sum_{j=1}^{n} \frac{r_{jt}^2}{n}. \qquad (6)$$

Therefore, $r_{jt}^2$ indicates which items contribute the most or least to the total score. Because the item SS for the weighted responses of item $j$ is proportional to $r_{jt}^2$, it also indicates how important an item is to the total scores in dual scaling.

### Data

Four datasets were used. The first dataset was reported in Nishisato (1987). It served as a baseline for comparison with previous research on robust techniques for quantifying categorical data. The other three datasets were simulated. The three simulated datasets contained more than one outlier. Some of these outliers did not appear in the context of other outliers. Only when other more obvious outliers were removed did the detrimental effects of these other outliers on the overall solution appear. Other observations that appeared to be outliers, because that respondent was the only one to select a particular response to an item, were in fact not outliers. That is, they had no detrimental effect (low influence) on the overall solution. Such datasets are probably more typical of those encountered in practice.

*Dataset 1.* 50 students were asked to answer 11 multiple-choice items for a total of 550 responses (Tung, 1980). The item options chosen by each examinee were recorded in a 50 × 11 data matrix. The number of options per item was 2, 3, 4, or 5. One examinee in this group (Examinee 20) had unusual responses to Items 4, 6, and 7. To determine the effect of Examinee 20's responses to Items 4, 6, and 7, the dual scaling solution was obtained first for the complete dataset and then for the same dataset but with Examinee 20's responses to Items 4, 6, and 7 treated as missing (the reduced dataset). Also, biweight results were obtained for three values of $k$ ($k = 3$, $k = 4$, $k = 5$) used in Equation 5. In addition to the biweight results, two sets of results were obtained using the MTA. Because Nishisato (1987) showed empirically that trimming one observation ($q = 1$) from each end of the ordered observations gave good results, this moderate amount of trimming also was used here for comparison purposes along with $q = 2$ (i.e., trimming two observations).

*Datasets 2, 3, and 4.* For Datasets 2 and 3, responses of 40 people to a 10-item four-point Likert scale were simulated. In Dataset 2, six of the simulees showed unusual option choices for one or more items. Dataset 3 contained three outliers. For both these datasets, the unusual responses were simulated so that only those simulees selected option 1 for a particular item.

For Dataset 4, responses of 50 people were simulated to a 15-item five-point scale. Responses were simulated in such a way that only three simulees chose options 4 or 5 for one or more items. Responses to all items were positively skewed so as to model a psychological scale in which low scores were the norm (e.g., no pathology). Following Nishisato (1984, 1987), the detrimental effects of outlying observations were evaluated by (1) large correlations among the items associated with the unusual option choices, (2) large item-total correlations, (3) unusually large and small item SS, and (4) unusually large or small option weights.

## Results

### Dataset 1

*Effect of Examinee 20 on the MBM solution.* As Tables 1 and 2 show, Examinee 20's responses to Items 4, 6, and 7 on the dual scaling solution resulted in spuriously high correlations among these items along with inflated item SS and item-total correlations. Table 1 shows that the correlations among Items 4, 6, and 7 ranged from .67 to .95 for the full dataset, and from .12 to .25 for the reduced dataset. Not only were the correlations among these items much lower in the reduced dataset, but they were also more similar to the correlations among all other items.

Table 2 shows that the item SS and item-total correlations were greatly inflated for the full dataset compared to the reduced dataset. Also, the item option weights for the options selected by Examinee 20 were extremely large in the full dataset. This is because Examinee 20 was the only examinee to select option 4 of Item 4 and option 2 of Item 7, although one other examinee selected option 2 of Item 6. The latter explains the nonzero weight for option 2 of Item 6 in the reduced dataset.

**Table 1**
Dataset 1 Interitem Correlations for the Full (Below the Diagonal) and Reduced (Above the Diagonal) Datasets

| | Item | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | – | .39 | .13 | .04 | −.11 | .37 | .23 | .09 | .39 | .02 | .27 |
| 2 | .13 | – | 0.00 | .30 | .13 | .24 | .18 | .14 | .45 | 0.00 | .53 |
| 3 | .13 | 0.00 | – | .02 | −.08 | .19 | .28 | −.20 | .01 | −.25 | 0.00 |
| 4 | .10 | .09 | .13 | – | .34 | **.15** | **.12** | .31 | .34 | .03 | .42 |
| 5 | .05 | .16 | −.05 | .05 | – | −.04 | .17 | .07 | −.05 | .17 | .11 |
| 6 | .18 | .17 | .22 | **.67** | .14 | – | **.25** | .16 | .18 | −.08 | .17 |
| 7 | .14 | .08 | .19 | **.95** | .13 | **.70** | – | .21 | .05 | −.02 | .02 |
| 8 | .09 | −.04 | −.20 | .06 | .10 | .06 | .07 | – | .15 | −.01 | .25 |
| 9 | .34 | .42 | .03 | .26 | −.01 | .25 | .19 | .17 | – | .14 | .62 |
| 10 | .19 | −.01 | −.22 | .41 | .11 | .28 | .40 | .02 | .20 | – | .22 |
| 11 | .27 | .34 | 0.00 | .21 | −.05 | .18 | .12 | .25 | .60 | .17 | – |

**Table 2**
Biweighted Dual Scaling Option Weights, SS, and Item-Total Correlations for Items 4, 6, and 7

| Dataset and Item | Item Option | Option Weight | SS | Item-Total $r$ |
|---|---|---|---|---|
| Full Dataset | | | | |
| 4 | 1 | .01 | 124.04 | .853 |
| | 2 | −.21 | | |
| | 3 | −1.18 | | |
| | 4 | **10.65** | | |
| | 5 | .27 | | |
| 6 | 1 | −.20 | 102.83 | .776 |
| | 2 | **6.97** | | |
| | 3 | −.67 | | |
| 7 | 1 | −.10 | 119.84 | .838 |
| | 2 | **10.65** | | |
| | 3 | −.93 | | |
| Reduced Dataset | | | | |
| 4 | 1 | .89 | 65.63 | .591 |
| | 2 | −.24 | | |
| | 3 | −2.47 | | |
| | 4 | **0.00** | | |
| | 5 | −1.02 | | |
| 6 | 1 | .32 | 40.30 | .463 |
| | 2 | **2.82** | | |
| | 3 | −1.78 | | |
| 7 | 1 | .25 | 20.52 | .331 |
| | 2 | **0.00** | | |
| | 3 | −1.60 | | |

*MBM versus MTA solutions.* The item option weights, item SS, item-total correlations, and interitem correlations for Items 4, 6, and 7 in Table 3 show that the MTA results for $q = 1$ fell between those obtained with the MBM $k = 4$ and MBM $k = 5$. Table 4 shows that $\eta^2$ for MTA $q = 1$ was only slightly larger than that for MBM $k = 4$ (.263 versus .262); thus, MBM $k = 4$ gave results almost identical to those observed for MTA $q = 1$. For MBM $k = 5$, $\eta^2$ was larger than the optimal value under MRA for the reduced dataset (.270 versus .265). This means that $\eta^2$ for MBM $k = 5$ was closer to the optimum value of $\eta^2$ obtained with the MRA for the full dataset (.293), the theoretical maximum.

### Datasets 2, 3, and 4

The impact of the outliers on the MRA results (full datasets) are seen by examination of the item SS for Datasets 2, 3, and 4 shown in Table 5. Clearly, the effect of outlying responses resulted in spurious item SS. For example, for Items 1 and 3 in Dataset 2 the SS were 108.92 and 1.66, respectively. Whereas in the reduced dataset (outliers removed) all item SS were in the range of approximately 16 to 57. Similar results are shown for Datasets 3 and 4 in Table 5.

**Table 3**
Option Weights and Item Statistics for the MBM and MTA

| Method and Item | Item Option | Option Weight | SS | Item-Total $r$ | Interitem $r$ Item 4 | Interitem $r$ Item 6 |
|---|---|---|---|---|---|---|
| MBM $k = 3$ | | | | | | |
| 4 | 4 | 3.06 | 71.92 | .605 | | |
| 6 | 2 | 2.25 | 59.53 | .550 | .20 | |
| 7 | 2 | 2.97 | 31.11 | .398 | .27 | .37 |
| MBM $k = 4$ | | | | | | |
| 4 | 4 | 3.82 | 76.01 | .631 | | |
| 6 | 2 | 2.77 | 63.40 | .576 | .24 | |
| 7 | 2 | 3.80 | 37.72 | .444 | .35 | .41 |
| MBM $k = 5$ | | | | | | |
| 4 | 4 | 5.60 | 84.76 | .677 | | |
| 6 | 2 | 3.72 | 65.50 | .595 | .34 | |
| 7 | 2 | 5.60 | 49.88 | .519 | .54 | .50 |
| MTA $q = 1$ | | | | | | |
| 4 | 4 | 4.52 | 72.97 | .620 | | |
| 6 | 2 | 3.08 | 61.85 | .571 | .30 | |
| 7 | 2 | 4.56 | 34.15 | .424 | .48 | .46 |
| MTA $q = 2$ | | | | | | |
| 4 | 4 | 2.85 | 52.79 | .510 | | |
| 6 | 2 | 2.18 | 69.65 | .586 | .23 | |
| 7 | 2 | 3.08 | 32.03 | .397 | .35 | .36 |

**Table 4**
$\eta$, $\eta^2$, and $\alpha$ for the Full and Reduced MRA, MTA, and MBM

| | MRA | | MTA | | MBM | | |
|---|---|---|---|---|---|---|---|
| Statistic | Full | Reduced | $q = 1$ | $q = 2$ | $k = 3$ | $k = 4$ | $k = 5$ |
| $\eta$ | .541 | .515 | .513 | .497 | .504 | .512 | .520 |
| $\eta^2$ | .293 | .265 | .263 | .247 | .254 | .262 | .270 |
| $\alpha$ | .759 | .722 | .720 | .694 | .707 | .718 | .730 |

Table 6 shows the item SS for the MTA $q = 1$ and the MBM $k = 3$ for Datasets 2, 3, and 4. Datasets 2 and 4 had item SS that were more similar to those obtained for the optimal solutions on the reduced datasets (outliers eliminated) shown in Table 5; but the item SS for Dataset 3, for both the MTA and the MBM, resembled those for the optimal solution on the full dataset (see Table 5). Obviously, the MTA $q = 1$ and MBM $k = 3$ did not function well for Dataset 3. It appears that for these values of $q$ and $k$, both the trimmed mean and the biweighted mean were being computed on a set of weights that were all large. Thus, a larger value of $q$ for the MTA and smaller values of $k$ for the MBM were required to deal effectively with these large weights. By fixing $q = 2$ and $k = 1.5$, 1.75, and 2.0, more reasonable results were obtained (see Table 7).

**Table 5**
Dual Scaling Item SS From MRA for Full and Reduced Data for Datasets 2 (10 Items), 3 (10 Items), and 4 (15 Items)

| Dataset and Item | Dataset 2 | Dataset 3 | Dataset 4 |
|---|---|---|---|
| Full Dataset | | | |
| 1 | 108.92 | 110.56 | 80.86 |
| 2 | 8.32 | 2.92 | 27.68 |
| 3 | 1.66 | 1.59 | 123.12 |
| 4 | 12.76 | 7.39 | 8.20 |
| 5 | 106.51 | 109.03 | 48.55 |
| 6 | 48.82 | 33.75 | 21.51 |
| 7 | 20.99 | 13.42 | 24.75 |
| 8 | 78.13 | 109.03 | 67.58 |
| 9 | 9.33 | 3.67 | 22.51 |
| 10 | 4.56 | 8.64 | 48.46 |
| 11 | | | 30.48 |
| 12 | | | 42.13 |
| 13 | | | 111.94 |
| 14 | | | 15.60 |
| 15 | | | 76.64 |
| Reduced Dataset With 3 Outliers Eliminated | | | |
| 1 | 33.04 | 38.41 | 42.22 |
| 2 | 42.18 | 40.11 | 57.54 |
| 3 | 57.20 | 49.98 | 54.19 |
| 4 | 39.50 | 49.83 | 10.00 |
| 5 | 16.28 | 12.83 | 58.95 |
| 6 | 45.21 | 50.32 | 51.40 |
| 7 | 40.34 | 35.13 | 26.83 |
| 8 | 30.94 | 36.06 | 12.11 |
| 9 | 27.74 | 33.18 | 38.77 |
| 10 | 37.56 | 24.16 | 14.15 |
| 11 | | | 52.26 |
| 12 | | | 81.55 |
| 13 | | | 21.01 |
| 14 | | | 97.74 |
| 15 | | | 86.29 |

When the squared correlation ratios, $\eta^2$, for the MBM were compared to those for the optimal MRA solution for the reduced datasets (see Table 8), $\eta^2$ for MBM $k = 2$ was larger than the optimal value for Datasets 2 and 3 (.274 for MBM $k = 2$ versus .264 for reduced MRA for Dataset 2; .292 for MBM $k = 2$ and .230 for reduced MRA for Dataset 3) and only slightly less in Dataset 4 (.213 for MBM $k = 2$ and .230 for reduced MRA for Dataset 4). But for MBM $k = 3$, $\eta^2$ was always larger than the optimal MRA value for the reduced datasets (.279 > .264, Dataset 1; .303 > .251, Dataset 2; .231 > .230, Dataset 3). Comparing MBM to MTA (see Table 8), the $\eta^2$ for MTA $q = 1$ was only slightly larger than for MBM $k = 3$ in Datasets 2, 3, and 4. These results were similar to those obtained for Dataset 1.

## Discussion

As these results and those of Nishisato (1984, 1987) have shown, the MRA is not robust in the presence of outliers. Outliers have a detrimental effect on the optimal solution. These effects include (1) spuriously large or small item-total correlations and item SS, (2) spurious interitem correlations, and (3) extreme item option weights (too large or too small). The use of these item statistics for item selection and test development, therefore, is inappropriate when outliers are present.

**Table 6**
The MTA $q = 1$ and the MBM $k = 3$ Item SS for Datasets 2, 3, and 4

| Method and Item | Dataset 2 | Dataset 3 | Dataset 4 |
|---|---|---|---|
| MTA $q = 1$ | | | |
| 1 | 48.79 | 119.67 | 70.73 |
| 2 | 47.04 | 7.70 | 54.96 |
| 3 | 15.99 | 1.50 | 69.39 |
| 4 | 38.46 | 6.47 | 35.72 |
| 5 | 41.28 | 118.49 | 55.21 |
| 6 | 51.95 | 8.46 | 28.53 |
| 7 | 55.43 | 8.29 | 16.84 |
| 8 | 57.17 | 122.94 | 38.59 |
| 9 | 35.10 | 5.03 | 51.67 |
| 10 | 8.80 | 1.45 | 20.26 |
| 11 | | | 59.73 |
| 12 | | | 84.34 |
| 13 | | | 84.15 |
| 14 | | | 30.96 |
| 15 | | | 48.90 |
| MBM $k = 3$ | | | |
| 1 | 43.19 | 131.83 | 57.65 |
| 2 | 58.94 | .94 | 58.47 |
| 3 | 24.49 | .08 | 44.34 |
| 4 | 29.74 | .47 | 32.90 |
| 5 | 32.44 | 131.77 | 74.07 |
| 6 | 62.64 | .88 | 42.23 |
| 7 | 56.07 | .94 | 24.42 |
| 8 | 50.67 | 132.45 | 31.67 |
| 9 | 30.89 | .33 | 35.86 |
| 10 | 10.91 | .30 | 28.64 |
| 11 | | | 78.81 |
| 12 | | | 97.04 |
| 13 | | | 67.36 |
| 14 | | | 29.36 |
| 15 | | | 47.18 |

A simple solution to the outlier problem in dual scaling, when only one outlier is present, is to delete the outlier. Then the MRA can be used to compute the optimal solution on the reduced dataset, as was done in Dataset 1. Often, however, datasets contain more than one outlier. To further complicate matters, some observations only appear as outliers when other more obvious outliers are removed first. This is common in regression analysis. However, in regression analysis, there are well-developed methods for detecting outliers and influential observations (Belsley, Kuh, & Welch, 1980; Rousseeuw & Leroy, 1987).

At present, no such methods exist for identifying and dealing with the outlier problem in dual scaling. Use of the leave-one-out method followed by the computation of the optimal solution using the MRA on the reduced dataset is very inefficient when more than one outlier is present because this process may have to be repeated a large number of times.

**Table 7**
Item SS for Dataset 3 for the MTA $q = 2$ and the MBM ($k = 2$, $k = 1.75$, $k = 1.5$)

| Item | MTA $q = 2$ | MBM $k = 2$ | MBM $k = 1.75$ | MBM $k = 1.5$ |
|---|---|---|---|---|
| 1 | 70.53 | 91.74 | 57.90 | 34.21 |
| 2 | 41.89 | 30.72 | 44.51 | 47.64 |
| 3 | 9.73 | 14.65 | 13.28 | 27.32 |
| 4 | 28.84 | 12.79 | 37.46 | 79.53 |
| 5 | 50.35 | 87.30 | 34.84 | 29.38 |
| 6 | 58.41 | 28.74 | 53.13 | 9.95 |
| 7 | 41.29 | 27.70 | 39.26 | 31.76 |
| 8 | 70.86 | 90.89 | 67.00 | 52.65 |
| 9 | 22.92 | 15.30 | 37.76 | 43.27 |
| 10 | 5.19 | .17 | 14.85 | 44.28 |

Repeating this process a large number of times has three obvious negative consequences: (1) a large amount of data may be lost, (2) removal of observations may actually create more outliers and thus necessitate the removal of even more observations before a stable solution is reached, and (3) options for certain items may appear to have never been selected due to the removal of earlier observations. A consequence of this last point is that these options will be given 0 weights and the item statistics may identify the item as a poor item and, therefore, an item that should be eliminated. This may be the correct decision, but the poor item statistics could be a consequence of having eliminated too many earlier observations. If this situation were to occur for more than one item, then eliminating these items could have an adverse effect on the validity of the instrument.

Clearly, a more satisfactory solution than the leave-one-out method is needed for dealing with the outlier problem in dual scaling. Of the robust methods proposed by Nishisato (1987), the MTA seems to be the best of the simple computational methods. However, this study showed that the MBM competes favorably with the MTA. Unlike the MTA, the MBM adjusts the weight of each observation to reflect its influence on the estimate of location.

**Table 8**
$\eta$, $\eta^2$, and $\alpha$ for the MRA, MTA, and MBM for Datasets 2, 3, and 4

| Method, Dataset, and Statistic | Dataset 2 | Dataset 3 | Dataset 4 |
|---|---|---|---|
| MRA | | | |
| Full | | | |
| $\eta$ | .552 | .586 | .523 |
| $\eta^2$ | .305 | .343 | .273 |
| $\alpha$ | .747 | .788 | .810 |
| Reduced | | | |
| $\eta$ | .514 | .501 | .480 |
| $\eta^2$ | .264 | .251 | .230 |
| $\alpha$ | .691 | .669 | .761 |
| MTA | | | |
| $q = 1$ | | | |
| $\eta$ | .536 | .561 | .496 |
| $\eta^2$ | .287 | .315 | .246 |
| $\alpha$ | .724 | .758 | .781 |
| $q = 2$ | | | |
| $\eta$ | .524 | .536 | .473 |
| $\eta^2$ | .274 | .287 | .224 |
| $\alpha$ | .706 | .724 | .753 |
| MBM | | | |
| $k = 1.5$ | | | |
| $\eta$ | .500 | .561 | .444 |
| $\eta^2$ | .250 | .220 | .197 |
| $\alpha$ | .666 | .606 | .709 |
| $k = 2$ | | | |
| $\eta$ | .514 | .540 | .461 |
| $\eta^2$ | .274 | .292 | .213 |
| $\alpha$ | .706 | .730 | .736 |
| $k = 3$ | | | |
| $\eta$ | .529 | .551 | .481 |
| $\eta^2$ | .279 | .303 | .231 |
| $\alpha$ | .713 | .745 | .762 |

Furthermore, because $k$ in the MBM is a continuous variable and $q$ in the MTA is discrete, the MBM affords more control than the MTA in fine tuning the final solution. This was demonstrated in the case of Dataset 3 (see Table 7).

## References

Ammann, L. P. (1993). Robust singular value decomposition: A new approach to projection pursuit. *Journal of the American Statistical Association, 88,* 505–514.

Belsley, D. A., Kuh, E., & Welch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity.* New York: Wiley.

Bock, R. D. (1960). *Methods and applications of optimal scaling* (Research Memorandum No. 25). Chapel Hill: University of North Carolina, Psychometric Laboratory.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst (Ed.), *The prediction of personal adjustment* (pp. 319–348). New York: Social Science Research Council.

Hempel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69,* 383–393.

Hogg, R. V. (1979). Statistical robustness: One view of its use in applications today. *The American Statistician, 33,* 108–115.

Horst, P. (1935). Measuring complex attitudes. *Journal of Social Psychology, 6,* 369–374.

Huber, P. (1972). Robust statistics: A review. *The Annals of Mathematical Statistics, 43,* 1011–1067.

Jolliffe, I. T. (1986). *Principal component analysis.* Berlin: Springer-Verlag.

Lord, F. M. (1958). Some relations between Guttman's principal components of scale analysis and psychometric theory. *Psychometrika, 23,* 291–296.

Mosteller, F., & Tukey, J. (1977). *Exploratory data analysis and regression.* Reading MA: Addison-Wesley.

Nishisato, S. (1979). Dual scaling and its variants. In R. E. Traub (Ed.), *New directions in testing* (pp. 1–12). San Francisco: Jossey-Bass.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Nishisato, S. (1981). Mathematical expositions of dual scaling. In Y. P. Chaubey & T. D. Dwivedi (Eds.), *Topics in applied statistics* (pp. 629–640). Concordia University Press.

Nishisato, S. (1984). Dual scaling of reciprocal medians. *Proceedings of the 32nd Scientific Conference of the Italian Statistical Society* (Estratto Dagli Atti della XXXII Riunione Scientifica; pp. 141–147), Sorrento, Italy. Rome, Italy: Societa Italiana di Stastistica.

Nishisato, S. (1987). Robust techniques for quantifying categorical data. In I. B. MacNeil & G. J. Umphrey (Eds.), *Foundations of statistical inference* (pp. 209–217). Dordrecht, Holland: D. Reidel Publishing Company.

Nishisato, S. (1988). Dual scaling: Its development and comparisons with other quantification methods. *Proceedings of the Annual Meeting of the German Society of Operations Research* (pp. 376–389). Berlin, Germany: Springer-Verlag.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regres-*

*sion and outlier detection*. New York: Wiley.

Tung, P. (1980). *Examining oral skills: Designing communicative tests in EFL for Hong Kong secondary schools.* Unpublished master's thesis, University of Toronto.

Verboon, P. (1994). *A robust approach to nonlinear multivariate analysis.* Leiden: DSWO Press.

## Acknowledgments

*The author thanks Peter Tung for permission to use his data (Dataset 1) and Raymond Lam for his support. The author also thanks the editor and two anonymous reviewers for their encouragement and helpful editorial and substantive comments.*

## Author's Address

Send requests for reprints or further information to John Sachs C/O Mr. Raymond Lam, Department of Education, University of Hong Kong, Pokfulam Rd., Hong Kong. Internet: hdfelyh@hkucc.hku.hk.