Double weighing procedures

A normal assumption in statistical analysis is that the rows of the data matrix **X** are repeated samples. This means that each row can be viewed as the results of measuring some given characteristics of an object. Many types of data can not be viewed in this way. An example is where judges taste different types of food and gives a rating for the taste. If it were the same judge that repeatedly gave ratings for the food samples, we would have repeated samples. But judges are different. There are needed methods that take into account these types of situations. This can be handled by double weighing procedures. We shall briefly review the methods of double weighing of data.

In double weighing we are looking for a weight vector **w** for columns,

$$\mathbf{t}=\mathbf{X}\mathbf{w}=w_1\mathbf{x}_1+w_2\mathbf{x}_2+\ldots\ w_K\mathbf{x}_K,$$

and a weight vector **v** for rows,

$$\mathbf{p}=\mathbf{X}^T\mathbf{v}=v_1\mathbf{x}^1+v_2\mathbf{x}^2+\ldots\ +v_N\mathbf{x}^N,$$

such that that the score vector **t** and the loading vector **p** have some desirable properties. The weight vectors are usually scaled to have length one, $|\mathbf{w}|= |\mathbf{v}|=1$, although it is not necessary. When the weight vectors have been found, the data matrix **X** is adjusted as

$$\mathbf{X} \leftarrow \mathbf{X} - d\ \mathbf{t}\ \mathbf{p}^T, \qquad \text{with } d=1/\ \mathbf{v}^T\mathbf{X}\mathbf{w}.$$

In Ref 7 it is shown, that **X** has been reduced by rank one. Numerically, there are no restrictions concerning the choice of **w** and **v** as long $\mathbf{v}^T\mathbf{X}\mathbf{w}\neq 0$. The reduced **X**, $\mathbf{X}_{new}$, has an important property given by

$$\mathbf{X}_{new}\ \mathbf{w} = \mathbf{0} \qquad \mathbf{X}_{new}^T\mathbf{v} = \mathbf{0}.$$

These equations can be viewed as follows: we choose **w** and **v** in some way and adjust **X** such that adjusted matrix is orthogonal (independent) of the chosen weights. If this procedure is carried on, we arrive at a decomposition of **X** given by

$$\mathbf{X} = d_1\ \mathbf{t}_1\ \mathbf{p}_1^T + d_2\ \mathbf{t}_2\ \mathbf{p}_2^T + \ldots + d_A\ \mathbf{t}_A\ \mathbf{p}_A^T + \ldots$$

We select the number of terms, *A*, that we judge appropriate, usually until we do not find significant covariance. In (11) the **r**-vectors are found. By symmetry there is similar equation for the **s**-vectors, where **X** is replaced by its transposed. Both equations are shown here for completeness,

$$\mathbf{r}_a = \mathbf{w}_a - ((\mathbf{w}_a^T\mathbf{p}_1)\ \mathbf{r}_1 + \ldots + (\mathbf{w}_a^T\mathbf{p}_{a-1})\ \mathbf{r}_{a-1}\ ), \quad \mathbf{r}_1=\mathbf{w}_1,\ a=2,3,\ldots$$

$$\mathbf{s}_a = \mathbf{v}_a - ((\mathbf{v}_a^T\mathbf{t}_1)\ \mathbf{s}_1 + \ldots + (\mathbf{v}_a^T\mathbf{t}_{a-1})\ \mathbf{s}_{a-1}\ ), \qquad \mathbf{s}_1=\mathbf{v}_1,\ a=2,3,\ldots$$

The vectors $(\mathbf{s}_a,\mathbf{r}_a)$ have the important 'causal' interpretation,

$$\mathbf{t}_a = \mathbf{X}\ \mathbf{r}_a \qquad \text{and} \qquad \mathbf{p}_a = \mathbf{X}^T\ \mathbf{s}_a,$$

where **X** is the original data matrix, while the vectors are the ones obtained at the a^{th} step. The matrices $\mathbf{R}=(\mathbf{r}_1,\mathbf{r}_2,\ldots)$ and $\mathbf{S}=(\mathbf{s}_1,\mathbf{s}_2,\ldots)$ satisfy the properties, $\mathbf{R}^T\mathbf{P}=\mathbf{D}^{-1}$ and $\mathbf{S}^T\mathbf{T}=\mathbf{D}^{-1}$, where **D** is a diagonal matrix with d_i's in the diagonal. A generalized inverse for **X** is computed as,

$$\mathbf{X}^+ = d_1\,\mathbf{r}_1\,\mathbf{s}_1^T + d_2\,\mathbf{r}_2\,\mathbf{s}_2^T + \ldots + d_A\,\mathbf{r}_A\,\mathbf{s}_A^T + \ldots$$

The generalised inverse satisfies $\mathbf{X}\mathbf{X}^+\mathbf{X}=\mathbf{X}$. This also holds for the truncated version of **X** and $\mathbf{X}^+$, where *A* terms are used.

When working with the double weighing procedures, three pairs of vectors are generated. The weight vectors $(\mathbf{w}_a,\mathbf{v}_a)$ reflect how we want to look at the data. They generate a pair of score/ loading vectors $(\mathbf{t}_a,\mathbf{p}_a)$. They again generate the pair $(\mathbf{s}_a,\mathbf{r}_a)$, which shows the precision of the score/ loading vectors found. The scaling constants (d_a) are computed separately in order to secure the numerical precision of the procedure for large data.

Neither $(\mathbf{t}_a)$ nor $(\mathbf{p}_a)$ are mutually orthogonal. In case $\mathbf{v}^T\mathbf{X}\mathbf{w}$ comes close to zero at some step, there may be numerical problems. In this case it may be necessary to revise the choice of **w** or **v**. If it is chosen to revise **v**, it is often chosen as $\mathbf{v}=\mathbf{t}/|\mathbf{t}|$, where **t** comes from **w**. When **w** and **v** have been computed, their significance is studied as explained in section 9.

Normally the data in **X** are scaled before analysis. Separate scaling is used for each type of weighing. The argument for scaling is that the **w**-vectors (**v**-vectors) are weighing the columns (rows). The weights are usually based on the covariances. Therefore each column (row) should get equal opportunity. If scaling is not used, a large column (row) would get relatively large influence. If all values in **X** are approximately of similar sizes, weighing may not be needed. It should be emphasised that the choice of weighing influences on the results obtained by analysis.

The present analysis has been extended to multiple weighing schemes for multi-way data. The algorithms for multi-way data reduce to the above theory, when data are two-way data. Unfortunately, extra notation than that of matrices and operations (e.g. multiplication) is needed in order to present the theory. Therefore, the extension of above theory is not presented here.

Data example

In this example 10 judges (persons) were asked to taste 6 different food samples. They were asked to give each food sample a rating from 0 to 10. All persons were trained judges. Each rating number is an average of different aspects of the food samples. The data are presented in Table 1. The data are of approximately equal size. Thus no scaling has been applied.

	Food type					
Judge	1	2	3	4	5	6
1	7.78	8.74	5.29	6.36	8.24	8.67
2	5.02	5.50	3.25	3.91	5.19	5.45
3	6.05	6.72	4.04	4.85	6.34	6.65
4	5.53	6.15	3.69	4.43	5.80	6.08
5	7.36	8.17	4.90	5.89	7.71	8.08
6	6.53	7.20	4.30	5.16	6.80	7.13
7	5.63	6.31	3.82	4.58	5.95	6.23
8	8.27	9.34	5.69	6.83	8.81	9.26
9	7.29	8.10	4.86	5.84	7.64	8.01
10	5.34	5.93	3.56	4.27	5.59	5.86

Table 1. Rating of 10 judges on 6 food samples.

The task is to find a score vector $\mathbf{t}=\mathbf{X}\mathbf{w}$, such that **t** is good in describing **X**. A good vector is one that maximises the term $|\mathbf{X}^T\mathbf{t}|^2=|\mathbf{X}^T\mathbf{X}\mathbf{w}|^2$. The solution is the eigen vector associated with the largest eigen value of $\mathbf{X}^T\mathbf{X}\mathbf{X}^T\mathbf{X}\mathbf{w}=\lambda\mathbf{w}$. Since $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}^T\mathbf{X}\mathbf{X}^T\mathbf{X}$ have the same set of eigen vectors, it is sufficient to solve $\mathbf{X}^T\mathbf{X}\mathbf{w}=\lambda\mathbf{w}$. When **w** has

	Score vectors		
	t_1	t_2	t_3
1	1.22	-1.01	0,89
2	0.75	1.01	**0,00**
3	0.93	**0.00**	**0,00**
4	0.85	**0.00**	**0,00**
5	1.13	**0.00**	**0,00**
6	0.99	0.60	**0,00**
7	0.88	-0.73	**0,00**
8	1.31	-1.65	0,11
9	1.12	**0.00**	**0,00**
10	0.82	**0.00**	**0,00**

Table 2. Score vectors

	Loading vectors		
	p_1	p_2	p_3
1	6,51	0,14	-0,03
2	7,23	0,08	**0,00**
3	4,34	**0,00**	**0,00**
4	5,21	**0,00**	**0,00**
5	6,82	0,08	**0,00**
6	7,15	0,09	0,03

Table 3. Loading vectors

been found, it is investigated if some of the values in **w** do not differ from zero. If they do not differ significantly from zero, they are set to zero. Same analysis is carried out for finding good loading vector, $\mathbf{p}=\mathbf{X}^T\mathbf{v}$. When score and loading vectors have been found, **X** is adjusted for the results as described above. The same analysis is carried out on the reduced **X**. This is continued as long as significant weights are found.

Three sets of score and loading vectors are found, see Table 2 and 3. The score vectors are scaled such that sum of values in $\mathbf{t}_1$ is 10, the sum of absolute values in $\mathbf{t}_2$ is 5 and the sum is 1 for $\mathbf{t}_3$. This scaling is chosen to simplify the interpretation of the results. Note, that the optimisation tasks above is equivalent to maximising $\mathbf{v}^T\mathbf{X}\mathbf{w}$ subject to $|\mathbf{w}|=|\mathbf{v}|=1$. The weights will be proportional to the score and loading vectors. The interpretation of the results is as follows:

- 5 judges are 'average' judges (t_2- and t_3- values are zeros).
- Judges no 1, 7 and 8 (from $\mathbf{t}_2$) are slightly lower on rating 1, 2, 5 and 6 (from $\mathbf{p}_2$).
- Persons no 2 and 6 (from $\mathbf{t}_2$) are slightly higher on rating 1, 2, 5 and 6 (from $\mathbf{p}_2$).
- The third dimension is a slight adjustment to the judge 1 and 8.