Weighing procedures

Introduction. It is a standard assumption in statistical analysis that the data matrix **X**, an N×K matrix, is obtained in a special way. The columns of **X** are variables or items measured. The rows of **X** are samples or objects and are the results of repeated sampling. In industry the data need not be found in this way. We shall consider closer the Furnace data. In Fig. 1 we show the **X**-data, where a curve is a row in **X**. In the figure to the left all rows are shown, while the three curves in the right figure show the sample associated with the lowest y-value, median and largest y-value.

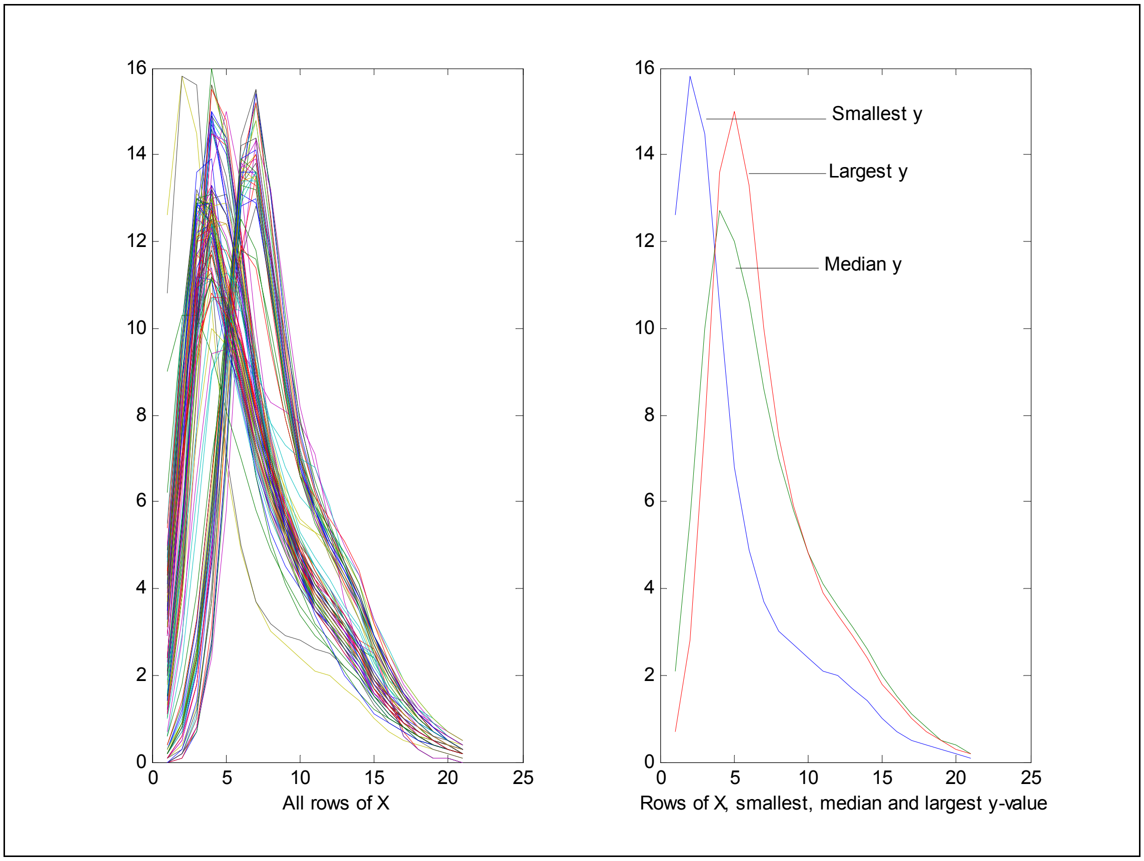

Figure 1. Plot of rows of X

There are 21 variables and therefore the curves have x-axis from 1 to 21. When we study the data, we can see that the values of a sample make a curve that moves to the right the larger the y-value is. On the figure to the left we can see that there are there are two sets of curves, indicating two types of operating conditions. Thus, setting an operating condition defines the values of all 21 variables. Therefore, the company is interested in how all the 21 variable perform in predicting the quality variable, the y-variable.

Double weighing schemes. There is a need for more nuances in the analysis of industrial data. We shall briefly describe how be can use weighing of variables and objects at the same time. The weight vector **w** shows how we weigh the variables, when computing the score vector $\mathbf{t}=\mathbf{X}\mathbf{w}=w_1\mathbf{x}_1+w_2\mathbf{x}_2+\ldots$ Similarly, we can consider a weight vector **v** that reflects how we weigh the samples. It gives the *loading vector* **p**, $\mathbf{p}=\mathbf{X}^T\mathbf{v}=(v_1\mathbf{x}^1+v_2\mathbf{x}^2+\ldots)$. Fig. 2 shows a schematic illustration of the double weighing procedure. The weight vectors are usually scaled to have length one, $|\mathbf{w}|=|\mathbf{v}|=1$, although it is not necessary. Sometimes we only find **w**. The weight vector **v** can then be chosen as $\mathbf{v}=\mathbf{t}/|\mathbf{t}|$. Similarly, if **w** is not found, it is chosen as $\mathbf{w}=\mathbf{p}/|\mathbf{p}|$. An important aspect of the procedures is that there is symmetry between **v** and **w**. Any choice or method to find **w** can be used to find **v** by exchanging rows and columns of **X**, and conversely. **X** is adjusted at each step. There are no restrictions on **w** and **v** except that the value of $\mathbf{v}^T\mathbf{X}\mathbf{w}$ must be different from zero, $\mathbf{v}^T\mathbf{X}\mathbf{w}\neq 0$. We can write $\mathbf{v}^T\mathbf{X}\mathbf{w}=\mathbf{v}^T\mathbf{t}=\mathbf{p}^T\mathbf{w}$. Thus, when we use double weighing schemes, the result of **v**, i.e, **p**, may not be orthogonal to the **w**, that we might choose as weights for variables. It shows that although we in principle can choose **w** and **v** independently of each other, there is in fact some dependence. The data matrix is adjusted as

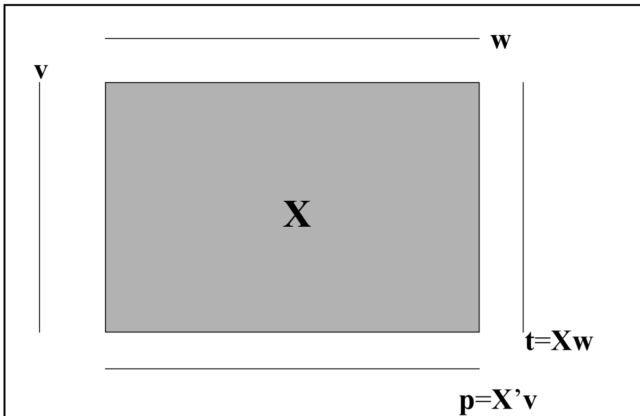

Figure 2 Schematic illustration of double weighing

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

In Ref [??] it is shown, that **X** has been reduced by rank one. 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. For the adjusted matrix we can choose **w** and **v** in any way we want as long $1/d=\mathbf{v}^T\mathbf{X}_{new}\mathbf{w}\neq 0$. 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.

It is shown in Ref. [??] that at each step we can compute the generalized inverse 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 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. The generalised inverse satisfies $\mathbf{X}\mathbf{X}^+\mathbf{X}=\mathbf{X}$. This also holds for the truncated version, where *A* terms are used.

In Ref. [??] it is shown the vectors $\mathbf{r}_a$ and $\mathbf{s}_a$ can be computed from the equations

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

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

These equations show that it is easy to compute the vectors $(\mathbf{r}_a, \mathbf{s}_a)$.

When we work with the double weighing procedures, we generate three pairs of vectors. 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 used in order to secure the numerical precision of the procedure.

When the procedure above is extended to multi-way data, it is important to note that the vectors $\mathbf{r}_a$ and $\mathbf{s}_a$ are generated independently of each other. In computing $\mathbf{r}_a$ only K-vectors are used, and $\mathbf{s}_a$ only N-vectors are used.

In Ref [??] it is shown that if $\mathbf{v}$ is chosen as $\mathbf{v}=\mathbf{t}/|\mathbf{t}|$, the score vectors ($\mathbf{t}_a$) will be orthogonal, ($\mathbf{t}_a^T\mathbf{t}_b$)=0, a≠b. For any other choice of $\mathbf{v}$ the score vectors will not be orthogonal. If neither ($\mathbf{t}_a$) nor ($\mathbf{p}_a$) are orthogonal, which means that we have selected w's and v's in some non-standard way, we may need to be careful in the numerical computations, because $\mathbf{v}^T\mathbf{X}\mathbf{w}$ may become close to zero at some step and possibly create instability.

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 $\mathbf{X}$ is the original data matrix, while the vectors are the ones obtained at the a*th* step. Thus, $\mathbf{r}_a$ tells us how the original variables generate the a*th* score vector and $\mathbf{s}_a$ how the samples generate the a*th* loading vector. We study the pair ($\mathbf{s}_a$,$\mathbf{r}_a$) to see how the results of the a*th* step are found.

When we work with multi-way data, we work in the same or similar way.