The 15th Annual Conference on Statistics & Computer Modeling in Human & Social Sciences
25-26 March 2003

A TRUE POSITION AS A NEW WAY FOR MONITORING DATA

Ezz H. Abdelfattah, Faculty of Science, Helwan University, Cairo, Egypt.[1]

Abstract

In this paper, we introduce a new scale that can identify the true position (TP) of an observation within a group of observations, and can reflect the actual distances between the observations and each other. Through this scale, we will be able to assign different positions for different order statistics of the same size and hence we can monitor different variables measured in a different scales simultaneously.

1. Introduction

Assume that we have a data consisting of n observations, $x_1, x_2, \ldots, x_n$ and that we want to 're-scale' such observations in a way that can keep the 'distances' or the 'relations' between them. Without loss of generality, assume that the data are coming from a continuous distribution (this means that there are no ties between such data). Let $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ represents the order statistics (*OS*) of $x_1, x_2, \ldots, x_n$.

By what we call "a true position" (*TP*), we may be able to assign **different** scaling for **different** OS of the same size. This way may be useful in many applications and it depends on rescaling the data, to reflect a real position for each value. This *TP* may be considered as a level between the ordinal and the scale measurements and it depends on assigning the position 1 for the smallest value $x_{(1)}$, and the position n for the largest value $x_{(n)}$.

2. A True Position

The *TP* procedure goes as follows:

Assign the position 1 for the smallest value $x_{(1)}$, and the position n for the largest value $x_{(n)}$ and let:

[1] Currently, Assistant professor at Dept. of Statistics, King Abdul-Aziz University, Jeddah, Saudi Arabia.

$$D_x = x_{(n)} - x_{(1)} \qquad (2.1)$$

be the range of the n values, and

$$D_R = n - 1 \qquad (2.2)$$

be the range of the n positions. Then the (ascending) TP for $x_{(i)}$ (or x_i); denoted by $TP_a(x_i)$ can be obtained through the equation:

$$\frac{TP_a(x_i) - 1}{x_i - x_{(1)}} = \frac{D_R}{D_x} \cdot \qquad (2.3)$$

Note that equation (2.3) may be considered as the equation of the line passing through the two points $(x_{(1)}, 1)$ and $(x_{(n)}, n)$, and can be written as

$$TP_a(x_i) = \frac{D_R}{D_x}(x_i - x_{(1)}) + 1 \cdot \qquad (2.4)$$

It is interesting to note that the formula (2.4) requires knowing $x_{(1)}$ and $x_{(n)}$ only to obtain $TP_a(x_i)$ for $i=1,2,...,n$. Also, through (2.4), we can obtain the position for each value independently of the position of all the other values.

We can use equation (2.4) when scaling the values in ascending order, while scaling the values in descending order requires using the equation:

$$TP_d(x_i) = \frac{D_R}{D_x}(x_{(n)} - x_i) + 1 \cdot \qquad (2.5)$$

It is clear that (2.5) assigns the position 1 for $x_{(n)}$ and the position n for $x_{(1)}$.

According to what we just introduced, we will assign "same" position for same (repeated) values.

Example(2.1) . Consider the following data (Conover (1980))

X = 22, 9, 4, 5, 1, 16, 15, 26, 47, 8, 31, 7

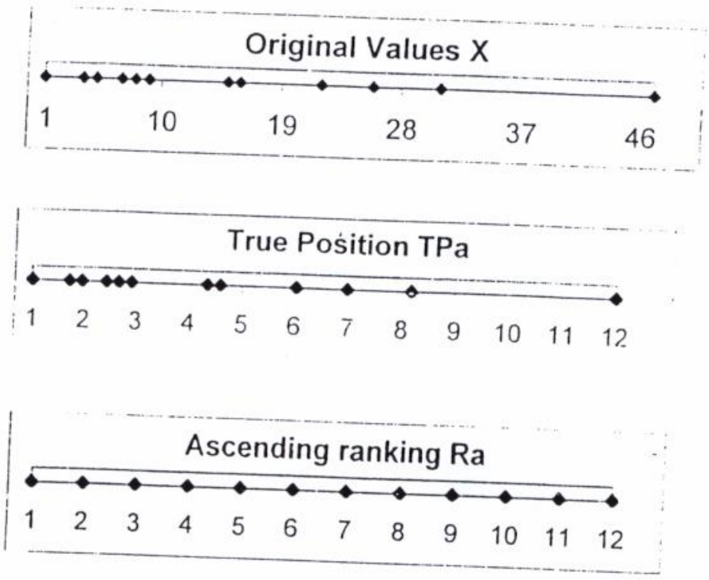
Then the data with the corresponding ascending rank R_a, ascending true position TP_a, descending rank R_d and descending true position TP_d for this data are given by Table 2.1 below.

Figure 2.1 below, shows the representation for the data X_i, i=1,2,…,14; the true position TP_a and the ascending rank R_a. The Figure shows also that the TP reflects the actual position for the original data, but in a different scale depending on the sample size (the proof will be given in the next section).

Table 2.1

i	X_i	R_a	TP_a	R_d	TP_d
1	22	9	6.02	4	6.98
2	9	6	2.91	7	10.09
3	4	2	1.72	11	11.28
4	5	3	1.96	10	11.04
5	1	1	1	12	12
6	16	8	4.59	5	8.41
7	15	7	4.35	6	8.65
8	26	10	6.98	3	6.02
9	47	12	12	1	1
10	8	5	2.67	8	10.33
11	31	11	8.17	2	4.83
12	7	4	2.43	9	10.57

Figure 2.1



3. The relation between TP with the rank R and the original data X

Here, we will discuss the relation between the true position TP and both the ordinary ranking R and the original data X.

3.1 The relation between TP and X

Proposition 3.1: TP reflects the actual position for X.
Proof:

We can see from (2.4) that

$$TP_a(x_i) - TP_a(x_j) = \frac{D_R}{D_x}(x_i - x_j), i, j = 1,2,...,n, \tag{3.1}$$

that is the difference between two values of the TP divided by D_R is exactly the difference between the two corresponding values of the original data divided by D_x.

This show that TP reflects the actual position for the original data, but in a different scale.

Proposition 3.2: TP is equal to the original data X iff $x_{(1)} = 1$ and $x_{(n)} = n$.
Proof:

If $x_{(1)} = 1$ and $x_{(n)} = n$, then $D_R = D_x$, then Substituting in (2.4) we directly obtain $TP_a(x_i) = x_i$ as required. The inverse is also straightforward.

Proposition 3.3: i) The correlation between X and TP_a is +1.00.
ii) The correlation between X and TP_d is -1.00.

Proof:

Note that we can write (2.4) as

$$TP_a(x_i) = A + Bx_i \tag{3.2}$$

where the constants A and B are given by:

$$A = 1 - \frac{D_R}{D_x} x_{(1)}. \tag{3.3}$$

$$B = \frac{D_R}{D_x} \tag{3.4}$$

Since (3.2) describes a linear relation between X and TP_a, we directly obtain the proof of (i).

Similarly, writing (2.5) in the form:

$$TP_d(x_i) = C - Bx_i \tag{3.5}$$

where the constant C is given by

$$C = 1 + \frac{D_R}{D_x} x_{(n)}. \tag{3.6}$$

Since (3.2) describes a linear relation between X and TP_d with opposite sign, we directly obtain the proof of (ii).

It is also clear that the mean and variance of TP can be obtained directly by knowing the mean and variance of X, since they are linearly related.

3.2 The relation between TP and the rank R

Proposition 3.4: TP is equivalent to the rank R iff the observations constitute an arithmetic sequence.

Proof:

Consider at first that the observations constitute the arithmetic sequence:

$$x_{(i)} = a + (i-1)d, \qquad i = 1,2,\dots,n$$

where a is the first term and d is the base. Then $a = x_{(1)}$, $D_R = n-1$ and $D_x = (n-1)d$. Substituting in (2.4) will yield:

$$\begin{aligned} TP_a(x_{(i)}) &= \frac{1}{d}(x_{(i)} - a) + 1 \\ &= \frac{1}{d}((a + (i-1)d - a) + 1 \\ &= i \end{aligned}$$

Conversely, let $TP_a(x_i) = i$ for $i=1,2,\dots,n$. Then substituting in (2.4), will give:

$$i = \frac{n-1}{x_{(n)} - x_{(1)}}(x_{(i)} - x_{(1)}) + 1$$

$$i - 1 = \frac{n-1}{(n-1)d}(x_{(i)} - a),$$

$$i - 1 = \frac{1}{d}(x_{(i)} - a),$$

where $a = x_{(1)}$, $d = (x_{(n)} - x_{(1)})/(n-1)$. Thus $x_{(i)} = a + (i-1)d$, for $i=1,2,\dots,n$. That is the observations constitute an arithmetic sequence. Hence the proof.

It is clear also that the relation between the TP and the rank R will be the same as the relation between the original observations X and the rank R since X and TP are perfectly related.

As an example for the previous relations, the following correlation matrix between the original data X, the ranks R_a, R_d and TR_a, TR_d given in Table 2.1 is shown in Table 3.1:

Table 3.1 The correlation matrix for data given in Table 2.1

	X	R_a	TP_a	R_d	TP_d
X	1	.93	1	-.93	-1
R_a	.93	1	.93	-1	-.93
TP_a	1	.93	1	-.93	.93
R_d	-.93	-1	-.93	1	.93
TP_d	-1	-.93	-1	.93	1

It is clear from Table 3.1 that the correlation between the original data X and the TP is "perfect ."

4. Monitoring different measures simultaneously

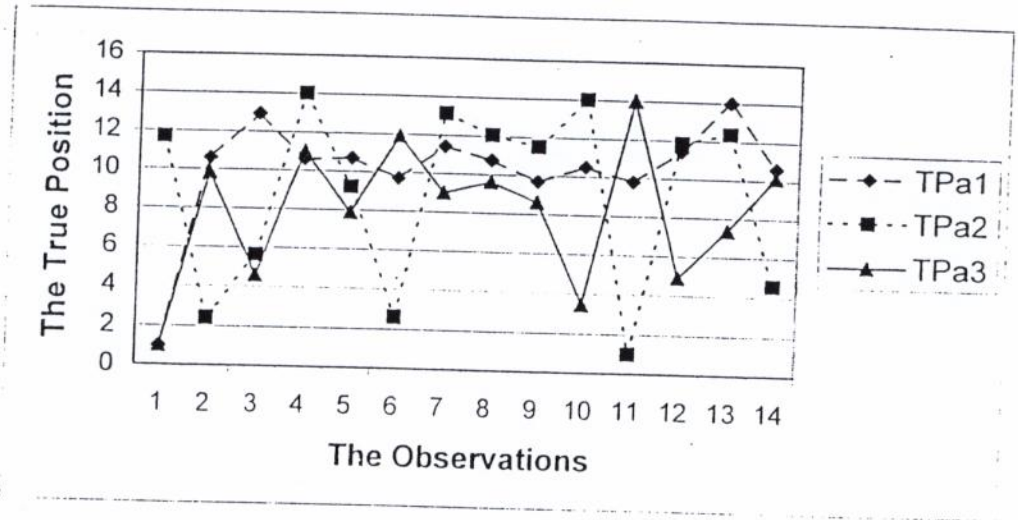
One of the advantages that we can use the TP for, is to monitor more than one variable measured in a different scale. For example the data given by Tracy et.al (1992) represent simultaneous measurement of three variables for a chemical processes : percentage of impurities (X_1) , temperature (X_2) and concentration strength (X_3). It is clear that these measurements are completely different in scale, and hence can not be viewed simultaneously in one figure, but transforming such measurements to the corresponding true positions will enable us to do so. It will also insure that the relations between the data within each variable are still the same.

The data with the corresponding TPa is given in Table 4.1 below:

Table 4.1

X1	TPa1	X2	TPa2	X3	TPa3
14.92	1.00	85.77	11.69	42.26	1.00
16.9	10.60	83.77	2.43	43.44	9.82
17.38	12.93	84.46	5.63	42.74	4.59
16.9	10.60	86.27	14.00	43.6	11.01
16.92	10.70	85.23	9.19	43.18	7.87
16.71	9.68	83.81	2.62	43.72	11.91
17.07	11.43	86.08	13.12	43.33	8.99
16.93	10.75	85.85	12.06	43.41	9.59
16.71	9.68	85.73	11.50	43.28	8.62
16.88	10.51	86.27	14.00	42.59	3.47
16.73	9.78	83.46	1.00	44	14.00
17.07	11.43	85.81	11.87	42.78	4.89
17.6	14.00	85.92	12.38	43.11	7.35
16.9	10.60	84.23	4.56	43.48	10.11

Figure 4.1 represents the true positions for the three variables.



References

Conover, W.J. (1980). "Practical nonparametric statistics". *John Wiley & Sons,Inc.*

Tracy, N.D., Young, J.C. and Mason, R.L. (1992). "Multivariate Control Charts for Individual Observations". *Journal of Quality Technology*, Vol 24, No. 2 , 88-95.

Mailing address:

Ezz Hassan Abdelfattah
Statistics Dept.,
Faculty of Science,
King Abdul-Aziz University,
P.O. Box 80203, Jeddah, 21589
Saudi Arabia

e-mail : ezzhassan@hotmail.com