A General Regression Procedure for Method Transformation

Application of Linear Regression Procedures for Method Comparison Studies in Clinical Chemistry, Part III

By W. Bablok

Biometrie Diagnostica, Boehringer Mannheim GmbH, Mannheim

H. Passing

Abt. Informatik und Kommunikation/Software, Hoechst AG, Frankfurt/Main 80

R. Bender

Techn. Wiss. Informationsverarbeitung, E. Merck, Darmstadt and

B. Schneider

Abteilung Biometrie, Medizinische Hochschule Hannover

(Received March 16, 1987/May 17/August 22, 1988)

Summary: The biometrical treatment of laboratory data may require the estimation of a regression line for the transformation of one set of measurements to another. The regression procedure introduced in part I (1) of our work does not always yield unbiased results in such situations, since its estimators are not scale invariant. In part III we present the parameter estimation of a general regression equation which is scale invariant and retains all properties of the method comparison procedure, in particular its robustness. Its application is demonstrated by several examples, and the results are compared with other robust biometrical procedures. The mathematical aspects are explained in the appendix.

Introduction

In the first part (1) of this series of papers we introduced a new biometrical regression procedure for the evaluation of method comparisons. It is used to determine whether measurements by two methods give the same results apart from random errors, or show a systematic difference. In part II (2) we discussed its properties and compared them with those of other established regression procedures. Based on a linear structural relationship model, the procedure presented there can be used to test the hypotheses $H_0: \beta = 1$ and $\alpha = 0$. Given a significant positive correlation corresponding to a positive value of Kendall’s $\tau$ (3), it produces estimates which are not influenced by the distribution of the data or the presence of biased or extreme data points. In addition, parameter estimation and the statistical test results are independent of the assignment of the methods to the variables.

There are situations in a clinical laboratory or in research which require the transformation of one set of measurements to another. This is the case when one wants to transfer the data from one system or reaction principle to another (e.g. different reaction
conditions, reagents or units, different activities or mass concentrations, etc). As we pointed out previously (2), our regression procedure for method comparison should not be used for this task. The estimators are biased since they are not scale invariant. This is caused by the elimination of slopes with a value of \(-1\) and shifting the median by the number of slopes with a value smaller than \(-1\). In part III we present two versions of a general regression procedure which is scale invariant and possesses all the properties of the method comparison procedure. Obviously, it allows the estimation of a regression line with a negative slope. A detailed description is given in the next section; readers who do not wish to get involved in the mathematical aspects are recommended to proceed to section “Properties of the Regression Procedure”.

Generalisation of the Estimation Procedure

For an experimental design as described in 1. c. (1, 2), we use the linear structural relationship model

\[ y_i^* = \alpha + \beta x_i^* \]

with

\[ x_i = x_i + \xi_i \quad \text{and} \quad y_i = y_i^* + \eta_i \quad (i = 1, \ldots, n). \]

\( x_i^* \) and \( y_i^* \) are the expected values of the \( i \)-th sample of method 1 and method 2, respectively. They have an arbitrary but continuous distribution over all possible samples. \( \xi_i \) and \( \eta_i \) are the measurement errors of the \( i \)-th sample. For their variances the relationship

\[ \frac{\sigma_{\xi_i}^2}{\sigma_{\eta_i}^2} = \beta^2 \quad \text{(Eq. 1)} \]

is assumed. For the derivation of the regression procedure we start with the assumption of a positive correlation, i.e. that Kendall’s \( \tau \) is significantly greater than zero.

Definition under the assumption of \( \tau > 0 \)

Firstly we deal with the estimation \( b \) of \( \beta \). For \( 1 \leq i < j \leq n \) we calculate all slopes \( S_{ij} \) in the following way:

\[ S_{ij} = \begin{cases} 
\frac{y_i - y_j}{x_i - x_j} & \text{if } y_i \neq y_j \text{ and } x_i \neq x_j \\
-\infty & \text{if } y_i < y_j \text{ and } x_i = x_j \\
+\infty & \text{if } y_i > y_j \text{ and } x_i = x_j \\
-0 & \text{if } y_i = y_j \text{ and } x_i < x_j \\
+0 & \text{if } y_i = y_j \text{ and } x_i > x_j \\
\text{not defined} & \text{if } y_i = y_j \text{ and } x_i = x_j
\end{cases} \quad \text{(Eq. 2)} \]

The \( S_{ij} \) are sorted with the convention of \(-0 < +0\) leading to

\[ S_{(1)} \leq S_{(2)} \leq \ldots \leq S_{(N)} \]

where \( N \) denotes all possible slopes according to Eq. 2.

In analogy to part I (1) we define for a suitable \( K \)

\[ b(K) := \begin{cases} 
\frac{S_{(N+1)+K}}{\left(\frac{N}{2} + k\right)} & \text{if } N \text{ odd} \\
\sqrt{\frac{S_{(N+1)+K} \cdot S_{(N+1)+K+1}}{\left(\frac{N}{2} + 1\right)}} & \text{if } N \text{ even}
\end{cases} \quad \text{(Eq. 3)} \]

It follows that \( b(K) \) is the median of the sequence

\[ S_{(K+1)} \leq \ldots \leq S_{(N)}, S_{(1)} \leq \ldots \leq S_{(K)}. \]

When \( N \) is even, the median is defined by the geometric mean, in order to guarantee an exact numerical independence of method assignment. Since slopes with values \(-\infty\) and \(+\infty\) and \(-\infty\) and \(+\infty\) are adjacent in this sequence, the choice of the index for any given data point \((x_i, y_i)\) has no influence on the result. As a consequence, the definition (Eq. 2) of the \( S_{ij} \) is meaningful.

We shall show in the appendix that the offset \( K \) has to meet a simple condition in order to make \( b(K) \) independent of the assignment of the methods to the variables. All the \( K \) meeting this condition define a family of regression procedures.

We distinguish between two alternative definitions of \( K \). For the notation we use square brackets to denote with \([a]\) the greatest integer smaller or equal to a non-negative real number \( a \) and the symbol \# to describe the number of elements in a set.

Regression procedure 1

With \( \text{neg} := \# \{S_{ij} \leq -0\} \) we define

\[ K := \left[ \frac{\text{neg}}{2} \right] \quad \text{(Eq. 4)} \]

and estimate \( b(K) \) by insertion of this \( K \) into Eq. 3.

Obviously, the indices are well defined. It is shown in the appendix that the computation can be simplified when \( \tau > 0 \), provided there are no slopes with a value of \( 0 \) or \( \infty \).

With \( \text{pos} := \# \{S_{ij} \geq +0\} \) \( b \) is estimated by

\[ b = \begin{cases} 
S_{(\text{pos}+1)} & \text{if } \text{pos odd} \\
\sqrt{S_{(\text{pos}+1)} \cdot S_{(\text{pos}+1)+1}} & \text{if } \text{pos even}
\end{cases} \quad \text{(Eq. 5)} \]

provided that only the \( S_{ij} \geq +0 \) are used.
Regression procedure 2

We define

\[ K = \# \{ S(q) < - \beta \} + \left[ \frac{1}{2} \cdot \# \{ S(q) = - \beta \} \right] \]  
(Eq. 6)

The indices are well defined as long as Kendall’s τ is significantly greater than zero (1). Since K in Eq. 6 depends on the unknown \( \beta \), the slope \( b \) cannot be calculated from Eq. 3. Therefore we replace \( \beta \) in Eq. 6 by \( b \) and have

\[ K = \# \{ S(q) < - b \} + \left[ \frac{1}{2} \cdot \# \{ S(q) = - b \} \right] \]  
(Eq. 7)

It follows that \( b \) and \( K \) are implicitly defined as solutions of the equations Eq. 3 and Eq. 7; they have only one solution as we shall show in the appendix. To get to the solution we define an iterative process. It starts with estimating \( b \) as in part I (1) and calculates the sequence \((K_r, b_r)\) from Eq. 7 and Eq. 3:

\[ K_1 = \# \{ S(q) < - 1 \}, \quad b_1 = b(K_1); \]
\[ K_r = \# \{ S(q) < - b_{r-1} \} + \left[ \frac{1}{2} \cdot \# \{ S(q) = - b_{r-1} \} \right], \quad b_r = b(K_r); \]  
(Eq. 8)

for \( r = 2, 3, \ldots \)

Since the set of slopes \( S(q) \) is finite the sequence of the \( b_r \) must be periodic, that is, there must be an \( m \geq 1 \) and an \( r^* \geq 1 \) with \( b_{m+\ast} = b_{r^*} \). If such an \( m \) and \( r^* \) is found the iteration procedure stops and we get

\[ b = \begin{cases} b_{(r^* + m/2)} & \text{if } m \text{ even} \\ \sqrt{b_{(r^* + m - 1)/2} \cdot b_{(r^* + m + 1)/2}} & \text{if } m \text{ odd} \end{cases} \]  
(Eq. 9)

It is easy to show by mathematical induction that either \( m = 1 \) or \( m = 2 \) is true under the assumption of continuously distributed data. With \( b \) defined by Eq. 9, we calculate \( K \) according to Eq. 7.

Estimation of \( \beta \) with \( \tau < 0 \)

For a significant negative correlation we use the transformation

\[ w_i := - y_i, \quad i = 1, 2, \ldots, n. \]

Then the \( x_i \) and \( w_i \) have a positive correlation, and the estimator \( b_w \) can be calculated as in the previous section if we replace \( y_i \) by \( - w_i \).

The estimator \( b \) of the slope between the \( x_i \) and \( y_i \) is defined by

\[ b = - b_w. \]  
(Eq. 10)

Estimation of the intercept \( \alpha \)

As in part I (1), we estimate the intercept in such a way that at least half of the points \((x_i, y_i)\) are located above or on the regression line and at least half of the points below or on the regression line. Therefore the estimator of \( \alpha \) is given by

\[ a = \text{med} \{ y_i - b x_i \}. \]  
(Eq. 11)

Confidence interval for \( \beta \) and \( \alpha \)

Because of equation 3 we can construct a confidence interval for \( \beta \) by

\[ S(M_1 + K) \leq \beta \leq S(M_2 + K) \]  
(Eq. 12)

where \( M_1 \) and \( M_2 \) are defined as in part I (1).

The confidence interval for \( \alpha \) is given in the same way by

\[ \text{med} \{ y_i - S(M_2 + K) \cdot x_i \} \leq \alpha \leq \text{med} \{ y_i - S(M_1 + K) \cdot x_i \}. \]  
(Eq. 13)

Properties of the Regression Procedure

Scale invariance of \( b \)

\( b \) is called scale invariant if, after transforming the data of one method by a factor \( f \), the estimator \( b \) is transformed by the same factor. Equation 4 and 6 show that the value of \( K \) does not change when either the \( x_i \) or \( y_i \) are transformed. Because of equation 3, \( b \) is transformed in accordance with \( x_i \) or \( y_i \). This proves the scale invariance for the estimators of both procedures.

Robustness of the estimators \( b \) and \( a \)

The definition of \( b \) and \( a \) as medians guarantees that there is no serious influence of extreme measurement points on the estimators. This property of the method comparison holds true for the regression procedures without restrictions.

Assignment of method 1 and 2 to variables \( X \) and \( Y \) is interchangeable

In the appendix we show for both procedures that, after changing the assignment of the methods to \( X \)
and Y, b is transformed to 1/b and a to −a/b. This
means that the estimated regression line can be solved
for x, yielding parameters which equal those obtained
for the estimators after interchanging the assignment
of the methods.

Test for linearity
The assumption of a linear structural relationship can
be tested after estimation of β and α by the cusum or
run test. Details are given in 1. c. (1).

Discussion
In the previous section we described the properties of
the two estimation procedures. For all data sets which
we have processed so far with both procedures we
found only negligible differences, if any, between the
resulting estimates. In addition, we performed a sim-
ulation study in analogy to that described in 1. c. (2)
to compare the behaviour of both procedures when
subjected to different data sets. There were only min-
imal differences in the results. We recommend the use
of procedure 1 to estimate the regression parameters,
since it is easier to calculate than procedure 2.

The estimation of α and the test for linearity corre-
sponds directly to what has been outlined in part I
(1). A data set with a negative correlation is treated
by transforming the yj and the resulting b according
to section “Confidence interval for β and a”. There-
fore, the properties of the regression procedure as
claimed in the introduction are proven. It should be
noted that it does not make sense to compute a
regression line for a structural relationship with a
data set which is not significantly correlated. It is
recommended to test the correlation before calculat-
ing the regression line.

The regression procedure is not a special case of the
method comparison procedure, even though it has
been derived therefrom. Therefore the user needs a
criterion for distinguishing between both: if a method
comparison study has to be evaluated, then the pro-
cedure described in part I should be used. It can be
used to test the hypothesis that both methods are
equal. To describe a general structural relationship
between two methods (e. g. a method transformation),
one of the procedures presented here should be ap-
plied. We allow an input of negative data for both
procedures, but for the method comparison one has
to consider whether negative values are meaningful
for the common measurement range of both methods.

When the parameters of a general regression line are
estimated, testing the hypothesis β = 1 does not make
sense. Therefore it should be decided before the eval-
uation which regression problem has to be handled.
The differences in the calculation of a regression line
for the method comparison or for the general case
are shown in figure 1.

Another regression procedure of the structural rela-
tionship model is the standardized principal compo-
nent analysis (4, 5) which also makes it possible to
interchange the assignment of the methods to the
variables. As long as its assumptions regarding the
distribution of the data and the error terms are met,
it will yield reliable results. Since the estimators are
easily calculated its use can be recommended for many
situations. However as discussed in part II (2), ex-
treme data points and skew distributions have a rather
strong influence on the results. The orthogonal regres-
sion which is recommended in the standardization of
coagulation reagents (6) has the same drawbacks. It
is the advantage of our procedure that it produces
robust estimates, especially in situations described
above.

Other approaches for robust regression procedures
are known in the literature. Of interest here is the
concept of Johnstone & Velleman (7) whose resistant
line estimator for the slope b is easier to compute
than the estimator presented in this paper. However,
the independence of the assignment of the methods
to the variables is not guaranteed. This is also the
case for another robust regression procedure: Siegel’s
repeated median estimator (8) is based on a vector of

---

**Fig. 1. Flow chart**
slopes $S_i$ similar to that in our procedure, but it also depends on the assignment of the methods. Both of the above procedures have properties which may be useful in other applications. We consider the interchange of the methods to the variables an essential demand for a suitable regression procedure in method conversion.

The properties of the procedures discussed here become evident for a data sample which has been obtained by measuring a substance with two different chemical reaction principles. It demonstrates how the estimates are effected by the distribution and variation of the data as well as by the assignment of the methods to the variables:

Tab. 1. Estimates of $b$ and $a$ from the regression procedures.

<table>
<thead>
<tr>
<th>Regression procedure</th>
<th>Assignment</th>
<th>$n = 41$</th>
<th></th>
<th>$n = 40$</th>
<th></th>
<th>$n = 39$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$b$</td>
<td>$a$</td>
<td>$b$</td>
<td>$a$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
<tr>
<td>Standardized principal</td>
<td></td>
<td>1.19</td>
<td>1.91</td>
<td>2.41</td>
<td>-2.83</td>
<td>2.06</td>
<td>-1.83</td>
</tr>
<tr>
<td>component</td>
<td>y = f(x)</td>
<td></td>
<td></td>
<td>x = f(y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.84</td>
<td>-1.61</td>
<td>0.42</td>
<td>1.17</td>
<td>0.49</td>
<td>0.89</td>
</tr>
<tr>
<td>Passing &amp; Bablok</td>
<td></td>
<td>2.14</td>
<td>-2.03</td>
<td>2.21</td>
<td>-2.20</td>
<td>2.14</td>
<td>-2.03</td>
</tr>
<tr>
<td></td>
<td>y = f(x)</td>
<td></td>
<td></td>
<td>x = f(y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.47</td>
<td>0.95</td>
<td>0.45</td>
<td>1.00</td>
<td>0.47</td>
<td>0.95</td>
</tr>
<tr>
<td>Johnstone &amp; Velleman</td>
<td></td>
<td>1.96</td>
<td>-1.65</td>
<td>2.07</td>
<td>-1.79</td>
<td>1.94</td>
<td>-1.53</td>
</tr>
<tr>
<td></td>
<td>y = f(x)</td>
<td></td>
<td></td>
<td>x = f(y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.43</td>
<td>1.26</td>
<td>0.36</td>
<td>1.30</td>
<td>0.42</td>
<td>1.29</td>
</tr>
<tr>
<td>Siegel</td>
<td></td>
<td>2.08</td>
<td>-1.93</td>
<td>2.16</td>
<td>-2.07</td>
<td>2.08</td>
<td>-1.93</td>
</tr>
<tr>
<td></td>
<td>y = f(x)</td>
<td></td>
<td></td>
<td>x = f(y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.44</td>
<td>1.14</td>
<td>0.37</td>
<td>1.22</td>
<td>0.44</td>
<td>1.14</td>
</tr>
</tbody>
</table>

Fig. 2. Data sample for application of robust regression.
- Our procedure
- Standardized principal component analysis
Regression lines for
a) $n = 41$
  - Point eliminated in figures 2b and 2c
  - Point eliminated in figure 2c
b) $n = 40$ and
  - Point eliminated in figure 2c
  - Point eliminated in figure 2c
  - Point eliminated in figure 2c
c) $n = 39$
calculated by standardized principal component analysis and our procedure.
The increasing variation of the data in the upper concentration range produces an influence similar to that of extreme data points. If we eliminate the last and the last two data points we get three data sets with varying estimates for the Standardized Principal Component and robust estimates for the other procedures. In contrast, if we change the assignment of the methods, the estimators of the procedures from Johnstone & Velleman and from Siegel differ from those obtained by solving the original equation for $x$. The results are shown in table 1, the graphs of the data in figure 2.

The table shows that the other robust procedures estimate a value for $b$ that cannot be transformed into its reciprocal. For example the Johnstone & Velleman regression produces $b = 1.96$ and $1/b = 0.51$ in contrast to $B = 0.43$ and $1/b = 2.33$, with $B$ as estimator for $x = f(y)$.

### Application in Clinical Chemistry

There are many situations where a general regression procedure is needed in clinical chemistry for research and routine laboratory work. Frequently the linear regression is used without attention to the experimental design and the resulting data from which the regression parameters are calculated. A discussion of the clinical and biochemical conditions, under which a linear transformation based on a structural relationship model may be applied, goes beyond the limits of a statistical paper. However, it is intended to present some proposals elsewhere for further discussion (9).

To demonstrate the properties and the behaviour of our procedure we give a numerical example.

The regression procedure may be of use when two methods with different chemical reagent compositions are compared which measure the same analyte. It is possible to describe the results of one method with respect to the location of the other method. Another application could be in converting the results of an instrument with individual properties to a reference system with known properties. Clearly, any transformation does only affect the location of the measurement points. The original degree of variation is not influenced by the calculation.

### Mathematical Appendix

#### The independence of method assignment

We derive a general condition for $K$ which enables $b(K)$ to be independent of the assignment. First of all we state the following definitions:

$$
K(- \infty) := \# \{S(q) = - \infty\},
K(- 0) := \# \{S(q) = - 0\} \text{ and }
\neg := \# \{S(q) < - 0\}.
$$

For a given $K$ with $K(- \infty) < K < \neg$ we define a partition of the set $\{S(q) \leq - 0\}$ into four subsets:

$$
\{S_{(1)}, \ldots, S_{(K(- \infty))}\},
\{S_{(K(- \infty) + 1)}, \ldots, S_{(K)}\},
\{S_{(K + 1)}, \ldots, S_{(\neg - K(- 0))}\},
\{S_{(\neg - K(- 0) + 1)}, \ldots, S_{(\neg)}\}.
$$

After changing the assignment of the methods to $X$ and $Y$ we get slopes $T_{ij}$ which are calculated analogously to Eq. 2. Sorting the $T_{ij}$ we get

$$
T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(\neg)}.
$$

Defining

$$
K(- \infty) := \{T_{(1)} = - \infty\} = K(- 0) \quad \text{and}
K(- 0) := \{T_{(\neg)} = - 0\} = K(- \infty)
$$

we obtain for $K$ with

$$
k(- \infty) \leq K \leq \neg
$$

which are calculated analogously to Eq. 2. Sorting the $T_{ij}$ we get

$$
T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(\neg)}.
$$

Defining

$$
k(- \infty) := \{T_{(1)} = - \infty\} = K(- 0) \quad \text{and}
k(- 0) := \{T_{(\neg)} = - 0\} = K(- \infty)
$$

we obtain for $K$ with

$$
k(- \infty) \leq K \leq \neg
$$

a partition of the set \( \{T(q) \leq -0\} \) in the same way as for the set \( \{S(q) \leq -0\} \) and an estimation \( B(k) \) according to Eq. 3.

The assignment of the methods to \( X \) and \( Y \) is arbitrary, if and only if \( b = 1/B \) holds. For the proof we assume a positive correlation which can be obtained as outlined in section "Estimation of \( \beta \) with \( \tau < 0 \)" in the case of \( \tau < 0 \). For an odd \( N \) we find

\[
b = S \left( \frac{N + 1}{2} + K \right)
\]

\[
= \frac{1}{T \left( \frac{N + \text{neg} + 1 - \frac{N + 1}{2} - K}{} \right)}
\]

\[
= \frac{1}{T \left( \frac{N + 1}{2} + \text{neg} - K \right)}
\]

and

\[
B = T \left( \frac{N + 1}{2} + K \right).
\]

It follows, that \( b = 1/B \), if

\[
k = \text{neg} - K. \quad \text{(Eq. 13)}
\]

The same is true if \( N \) is even. Equation 13 defines a class of regression procedures. If \( K \) is computed as stated for regression procedure 1, then condition 13 is either exactly or approximately fulfilled. In the latter case, there is a gap of one place in the sorted sequence, this is without importance if \( n \) is not very small.

Computing \( K \) for regression procedure 2 we get from Eq. 8

\[
K + k = \# \{ -\infty \leq S(q) < -b \} + \# \{-b < S(q) \leq 0\} + 2 \cdot [1/2 \cdot \# \{S(q) = -b\}]
\]

\[
= \text{neg} - \varepsilon \quad \text{with} \quad \varepsilon \in \{0, 1\}.
\]

It follows by mathematical induction that the assignment of the methods to \( X \) and \( Y \) is also arbitrary for \( \varepsilon = 0 \). If \( \varepsilon = 1 \), then independence of assignment holds approximately for values of \( n \) that are not too small.

Simplified calculation of \( b \) in regression procedure 1

Since \( b \) is estimated by the median of non-negative slopes, it has to be shown that the indices in Eq. 3 and Eq. 5 belong approximately to the same elements in the sorted sequence of the slopes.

For that we distinguish four cases dependent on an even or odd value of \( \text{neg} \) and \( \text{pos} \) as defined in section "Definition under the assumption \( \tau > 0 \)"

Case 1: \( \text{neg} \) and \( \text{pos} \) are even, it follows for

\[
N/2 + K = \text{neg}/2 + \text{pos}/2 + \text{neg}/2 = \text{neg} + \text{pos}/2 \text{and}
\]

\[
N/2 + K + 1 = \text{neg}/2 + \text{pos}/2 + \text{neg}/2 + 1 = \text{neg} + \text{pos}/2 + 1.
\]

Case 2: \( \text{neg} \) is even and \( \text{pos} \) odd, it follows for

\[
(N + 1)/2 + K = \text{neg}/2 + (\text{pos} + 1)/2 + \text{neg}/2 = \text{neg} + (\text{pos} + 1)/2.
\]

Case 3: \( \text{neg} \) is odd and \( \text{pos} \) even, it follows for

\[
(N + 1)/2 + K = (\text{neg} + 1)/2 + \text{pos}/2 + (\text{neg} - 1)/2 = \text{neg} + \text{pos}/2.
\]

Case 4: \( \text{neg} \) and \( \text{pos} \) are odd, it follows for

\[
N/2 + K = (\text{neg} + 1)/2 + (\text{pos} - 1)/2 + (\text{neg} - 1)/2 = \text{neg} + (\text{pos} - 1)/2 \text{and}
\]

\[
N/2 + 1 + K = (\text{neg} + 1)/2 + (\text{pos} + 1)/2 + (\text{neg} - 1)/2 = \text{neg} + (\text{pos} + 1)/2.
\]

In case 1 and 2 the indices in Eq. 3 and Eq. 5 refer to the same element, in case 3 and 4 there is a gap of about half a place in the sorted sequence of the slopes which has no real consequences on the result. However, there is a problem, when \( S_y \)'s are present with values \( \pm \infty \) or \( \pm 0 \), since the sign depends on the choice of indices for the data points.

Equations 3 and 7 have at most one solution

Let us assume that there are two different solutions \( (K', b') \) and \( (K'', b'') \) of Eq. 3 and Eq. 7 with

\[
K' = \# \{S(q) < -b'\} + [1/2 \cdot \# \{S(q) = -b'\}],
\]

\[
K'' = \# \{S(q) < -b''\} + [1/2 \cdot \# \{S(q) = -b''\}],
\]

and

\[
b' = b(K'), \quad b'' = b(K'').
\]

If \( b' = b'' \) is true then \( K' = K'' \) would follow from Eq. 14 in contradiction to our assumption. Now let \( b' < b'' \), then from Eq. 14 we get \( K' \geq K'' \) and because of Eq. 3 the contradiction \( b' \geq b'' \). The analogous contradiction follows from the assumption \( b' > b'' \) which proves that there is at most one solution.
Is there always an exact solution of equations 3 and 7?

If all \( S_i \geq 0 \) then \( K = 0 \) and \( b(0) = \text{med}\{S_i\} \) satisfy the equations 3 and 7. One can construct examples for which a solution of equations 3 and 7 does not exist even though \( \tau \) is significantly greater than zero. For instance the following data set has this property: 

\[(0,6), (3,1), (5,6), (7,8), (8,15)\]

To explain this peculiarity we have to consider that to every \( K \) an interval of \( b's \) satisfies equation 6, whereas equation 3 defines for every \( K \) exactly one \( b \). If there is a \( K \) with a solution \( b \) for Eq. 3 in an interval according to Eq. 7, then the pair \((K, b)\) is the exact solution.

An MS-DOS program version for the regression procedures is available on request.

References


W. Bablok
c/o Boehringer Mannheim GmbH
Sandhofer Straße 116
D-6800 Mannheim 31