J. Clin. Chem. Clin. Biochem.
Vol. 21, 1983, pp. 709–720

A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods

Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part I

By *H. Passing*

Abt. für Praktische Mathematik, Hoechst AG, Frankfurt/Main 80 and

W. Bablok

Allg. Biometrie, Boehringer Mannheim GmbH, Mannheim

(Received December 20, 1982/June 23, 1983)

Summary: Procedures for the statistical evaluation of method comparisons and instrument tests often have a requirement for distributional properties of the experimental data, but this requirement is frequently not met. In our paper we propose a new linear regression procedure with no special assumptions regarding the distribution of the samples and the measurement errors. The result does not depend on the assignment of the methods (instruments) to X and Y. After testing a linear relationship between X and Y confidence limits are given for the slope β and the intercept α; they are used to determine whether there is only a chance difference between β and 1 and between α and 0. The mathematical background is amplified separately in an appendix.

Ein neues biometrisches Verfahren zur Überprüfung der Gleichheit von Meßwerten von zwei analytischen Methoden

Anwendung von linearen Regressionsverfahren bei Methodenvergleichsstudien in der Klinischen Chemie, Teil I

Zusammenfassung: Bei der statistischen Auswertung von Methodenvergleichen und bei Geräteerprobungen werden in der Regel Verfahren eingesetzt, deren Anforderungen an die Verteilung der experimentiellen Daten häufig nicht erfüllt sind. In unserer Arbeit schlagen wir daher ein neues lineares Regressionsverfahren vor, das keine besonderen Annahmen für die Verteilung der Stichprobe und der Meßfehler voraussetzt. Das Ergebnis ist unabhängig von der Zuordnung der Methoden (Geräte) zu den Variablen X und Y. Nach Prüfung eines linearen Zusammenhanges zwischen X und Y werden Vertrauensgrenzen für die Steigung β und den Achsenabschnitt α angegeben. Mit ihrer Hilfe werden die Hypothesen $\beta = 1$ und $\alpha = 0$ getestet. Die mathematischen Grundlagen werden separat in einem Appendix abgehandelt.

Contents

1. Introduction
2. Method Comparison and Linear Regression
3. A New Regression Procedure
4. Discussion and Examples
5. Appendix: Mathematical Derivations

1. Introduction

Parameter estimation in linear regression models is one of the standard topics in basic statistical textbooks. With the widespread use of pocket calculators the computation of linear regression lines based on least squares has become routine work in scientific research. Unfortunately, many textbooks cover the

model assumptions rather briefly and only rarely the reader is warned of the consequences if those assumptions are violated. However, it is in the very nature of many experiments which call for linear regression that most of the assumptions cannot be observed. For instance, in many situations there is no independent variable which is free of error; but procedures taking this into account (1, 2) are based on more rigid and idealized distributional requirements than can be met in real experiments.

2. Method Comparison and Linear Regression

Clinical chemistry is one of the areas where linear regression models play a major role in the statistical evaluation of experiments. Especially in comparing analytical methods for measuring the same chemical substance or in instrument testing the need for reliable parameter estimation becomes very obvious in the judgement of equality. Clearly a method comparison cannot exclusively be based on the evaluation of a regression model. Many more properties of the methods must be compared; at least accuracy, imprecision, sensitivity, specificity, and range of concentration should be studied. For a detailed discussion see l.c. (3, 4).

The experimental layout can be described as follows: There are 2 different methods (instruments) which measure the same chemical analyte in a given medium (e.g. serum, plasma, urine, . . .). The question is: Do the methods measure the same concentration of the analyte or is there a systematic difference in the measurements? (For simplicity, we only refer to concentrations but our statements are also valid for any other quantity.)

The usual experimental procedure is to draw n independent samples from a population in which a given analyte is to the measured with values x_i and y_i for the i-th sample. These measurements are realisations of a pair of random variables X and Y, where X represents the values of method 1 and Y the values of method 2. For simplicity we also denote method 1 by method X and method 2 by method Y.

Following the statistical model of l.c. (5) each random variable is the sum of two components:

- one variable representing the variation of the expected value of the analyte within the population of all possible samples;
- one variable representing the variation of the measurement error for a given sample.

For the i-th sample this relationship is described by the equations

$$x_i = x_i^* + \xi_i$$

and

$$y_i = y_i^* + \eta_i;$$

x_i^* and y_i^* denote the expected values of this sample and ξ_i and η_i give the measurement errors. In this way each method may have its own expected value for the i-th sample.

If there exists a structural relationship between the two methods it can be described by the linear equation

$$y_i^* = \alpha + \beta x_i^*$$

From the n experimental values (x_i, y_i) the following objectives should be attained:

i) estimation of α and β;

ii) statistical test of the assumption of linearity; and if linearity is given

iii) test of the hypothesis $\beta = 1$;

iv) test of the hypothesis $\alpha = 0$.

If both hypotheses are accepted we can infer $y_i^* = x_i^*$, i.e. the two methods X and Y measure the same concentration within the investigated concentration range.

In practice one of the following four procedures is used:

[1] linear regression $y_i = \alpha + \beta x_i + \eta_i$

[2] linear regression $x_i = A + By_i + \xi_i$

[3] principal component analysis (*Deming*'s procedure) (5)

[4] standardized principal component analysis (5, 6, 7)

All four procedures assume a linear relation between the two methods, however each one has specific theoretical requirements:

- [1] and [2] ask for an error-free independent variable X or Y and normally distributed error terms with constant variance. A statistical test of linearity can be performed only if there exist multiple measurements of the dependent variable[1]). Procedures [1] and [2] are not equivalent and may even give contradictory results.

[1]) Strictly speaking this test should only be used if the independent variable has fixed values, as is assumed in the usual least squares linear regression.

– [3] and [4] assume that the expected values x_i^* and y_i^* come from a normal distribution. The error terms have to be normally distributed with a constant variance σ_ξ^2 and σ_η^2; they follow the restrictions

$$\sigma_\xi = \sigma_\eta \quad \text{for [3]}$$

and

$$\beta = \frac{\sigma_\eta}{\sigma_\xi} \quad \text{for [4].}$$

A statistical test of linearity has not so far been proposed.

However, in method comparison studies we generally find the following situation:

– Neither method X nor method Y is free of random error.

– The distribution of the measurement errors is usually not normal (8).

– The expected values x_i^* and y_i^* are not a random sample from normal distributions, since the methods are compared over a wide concentration range of the analyte which covers values of both healthy and diseased persons.

– Extreme values (outliers) are not necessarily gross measurement errors; they may be caused by different properties of the methods with respect to specificity or susceptibility to interferences. Therefore they should not be removed from the calculation without experimental reason.

– The variance of the measurement errors is not constant over the range of concentrations; in fact the variability increases with the magnitude of the measurements.

Therefore it must be expected that a researcher using any of the above procedures may obtain biased estimations for α and β, and therefore misleading results from the experiment. This situation is equally disappointing for the investigator and the statistician.

In the last 20 years many different proposals have been published for parameter estimation in the linear model, using less stringent distributional assumptions. The estimations were either based on robust procedures [for a detailed discussion with references see l.c. (9) and also l.c. (10)] or on a distribution-free approach (11). We are, however, not aware of a proposal which deals with the problem of a structural relationship.

We now describe a procedure which can achieve all the objectives (i) to (iv) and does not require specific assumptions regarding the distributions of the expected values or the error terms.

3. A New Regression Procedure

On the basis of the structural relationship model as described in chapter 2 we make the following assumptions:

x_i^*, y_i^* are the expected values of random variables from an arbitrary, continuous distribution (i.e. the sampling distribution is arbitrary).

ξ_i, η_i are realisations of random error terms, both coming from the same type of distribution. Their variances σ_ξ^2 and σ_η^2 need not to be constant within the sampling range but should remain proportional, that is

$$\frac{\sigma_\eta^2}{\sigma_\xi^2} = \beta^2.$$

In part II of our paper we shall demonstrate that these rather weak assumptions are sufficient for reliable parameter estimations and hypothesis testing if $\beta \sim 1$. There we shall investigate the influence of the distributions on the result of our procedure.

i) *Estimation of* α *and* β

According to *Theil* (12) the slopes of the straight lines between any two points are employed for the estimation of β. They are given by

$$S_{ij} = \frac{y_i - y_j}{x_i - x_j} \qquad \text{for } 1 \le i < j \le n.$$

There are $\binom{n}{2}$ possible ways to connect any two points.

Identical pairs of measurements with

$$x_i = x_j \text{ and } y_i = y_j.$$

do not contribute to the estimation of β; the corresponding S_{ij} is not defined at this stage. For reasons of symmetry (see appendix) any S_{ij} with a value of -1 is also disregarded.

Furthermore, from $x_i = x_j$ and $y_i \neq y_j$ it follows that $S_{ij} = \pm\infty$, depending on the sign of the difference $y_i - y_j$; from $x_i \neq x_j$ and $y_i = y_j$ it follows that $S_{ij} = 0$. Since (X, Y) is a continuous bivariate variable the occurrence of any of these special cases has a probability of zero (experimental data should exhibit these cases very rarely). In total there are

$$N \le \binom{n}{2}$$

slopes S_{ij}. After sorting the S_{ij} the ranked sequence

$$S_{(1)} \le S_{(2)} \le \ldots . \le S_{(N)}$$

is obtained.

If we substitute the structural relationship in the definition of the S_{ij} we find

$$S_{ij} = \frac{y_i^* - y_j^* + \eta_i - \eta_j}{x_i^* - x_j^* + \xi_i - \xi_j}$$

From $y_i^* = \alpha + \beta x_i^*$ and $d_{ij} = (x_i^* - x_j^*)$ we get

$$S_{ij} = \frac{\beta \cdot d_{ij} + (\eta_i - \eta_j)}{d_{ij} + (\xi_i - \xi_j)}$$

$$= \beta \cdot \frac{d_{ij} + \dfrac{(\eta_i - \eta_j)}{\beta}}{d_{ij} + (\xi - \xi_j)}$$

$$= \beta \cdot \frac{d_{ij} + z_{ij}}{d_{ij} + z'_{ij}},$$

where z_{ij} and z'_{ij} are indepedent and from the same distribution.

Since the values of S_{ij} are not independent it is obvious that their median can be a biased estimator of β. We proceed therefore as follows:

Let K be the number of values of S_{ij} with $S_{ij} < -1$. Then using K as an offset, β is estimated by the shifted median b of the $S_{(i)}$:

$$b = \begin{cases} S_{\left(\frac{N+1}{2}+K\right)} & \text{, if N is odd} \\ \frac{1}{2} \cdot \left(S_{\left(\frac{N}{2}+K\right)} + S_{\left(\frac{N}{2}+1+K\right)}\right) & \text{, if N is even.} \end{cases}$$

For the construction of a two-sided confidence interval for β on the level γ let $w_{\frac{\gamma}{2}}$ denote the $\left(1 - \frac{\gamma}{2}\right)$-quantile of the standardized normal distribution.

With

$$C_\gamma = w_{\frac{\gamma}{2}} \cdot \sqrt{\frac{n(n-1)(2n+5)}{18}}$$

and

$$M_1 = \frac{N - C_\gamma}{2}, \quad M_2 = N - M_1 + 1$$

(M_1 rounded to an integer value)

the confidence interval for β is given by

$$S_{(M_1+K)} \leq \beta \leq S_{(M_2+K)}.$$

The introduction of the offset K is motivated by the request for an arbitrary assignment of the methods to X and Y. The definition of K as the number of values of S_{ij} smaller than -1 corresponds to the null hypothesis $\beta = 1$. It will be demonstrated that, in this case, our b is a good and reliable estimator of β (see appendix and part II).

The estimation of α requires that at least one half of the points is located above or on the regression line and at least one half of the points below or on the line. As (X, Y) is a continuous bivariate variable then an equal number of points lies above and below the regression line with probability 1. A point (x_i, y_i) is located above the line only if $a < y_i - bx_i$. Therefore it can easily be shown that

$$a = \text{med}\,\{y_i - bx_i\}$$

is an estimator of α.

If b_L denotes the lower and b_U the upper limit of the confidence interval for β then the corresponding limits for α are given by

$$a_L = \text{med}\,\{y_i - b_U x_i\}$$

$$a_U = \text{med}\,\{y_i - b_L x_i\}.$$

These limits are conservative.

With the n pairs of measurement (x_i, y_i) one can either calculate

$$y^* = a + bx^* \quad \text{or} \quad x^* = A + By^*.$$

The above estimators for α and β show the following property:

$$B = \frac{1}{b} \quad \text{and} \quad A = -\frac{a}{b};$$

analogous formulas hold for their confidence limits. The proof is given in the appendix. It is therefore irrelevant which one of the two methods is denoted by X.

ii) *Statistical test of the assumption of linearity*

In testing for linearity one has to inspect how the regression line fits the data or how randomly the data scatters about $y^* = a + bx^*$. Naturally the parameters a and b are fixed in this context and our test will be conditional on a and b.

If there is a nonlinear relationship between x^* and y^* one would expect to find too many consecutive measurements either above or below the fitted line. Let l denote the number of points (x_i, y_i) with $y_i > a + bx_i$ and L the number of points with $y_i < a + bx_i$. To every point (x_i, y_i) we assign a score r_i, i.e.

$$r_i = \sqrt{\frac{L}{l}}, \quad \text{if } y_i > a + bx_i,$$

$$r_i = -\sqrt{\frac{l}{L}}, \quad \text{if } y_i < a + bx_i,$$

and

$$r_i = 0, \quad \text{if } y_i = a + bx_i.$$

Unfortunately the sequence of scores depends on the way in which the points (x_i, y_i) are ranked: either by increasing X-values or by increasing Y-values. That is, the result of a test for linearity would depend on which one of the methods is assigned to the X- and which one to the Y-variable.

Both methods can be treated alike by sorting the points (x_i, y_i) along the line $y^* = a + bx^*$. This is achieved by projecting every point (x_i, y_i) on the regression line. The distance between this projection and the y-intercept of the fitted line is given by

$$D_i = \frac{y_i + \frac{1}{b} \cdot x_i - a}{\sqrt{1 + \frac{1}{b^2}}}$$

The scores r_i are sorted according to increasing D_i; this rank order $r_{(i)}$ becomes the basis of the proposed linearity test.

We have considered two possible solutions to such a test. An obvious one would be the employment of a run test which actually would test the randomness of the distribution of scores along the line $y^* = a + bx^*$. The test is the subject of many publications [e.g. l.c. (13)], and its application for testing linearity is discussed in l.c. (14). The other solution which we present is based on a cusum-concept, which is a well known controlling procedure in Clinical Chemistry (15). Consider a coordinate system in which the x-axis represents the ranks of the D_i, i.e. the numbers 1 to n, and the y-axis the cumulative sum of the scores r_i. The sum

$$\text{cusum (i)} = \sum_{k=1}^{i} r_{(k)}$$

denotes the excess of positive or negative scores from point 1 to point i in the sorted sequence of the D_i. A random arrangement of scores as an indication of linearity would result in moderate values of $|\text{cusum (i)}|$, whereas an excess number of consecutive positive or negative scores in a "large" value of $|\text{cusum (i)}|$. Therefore it seems to be appropriate to compare the distribution of the subset of r_i with $r_i > 0$ with the distribution of the subset with $r_i < 0$. Critical values for the cusum statistic can be obtained from the *Kolmogorov-Smirnov* test; the derivation is given in the appendix.

If $|\text{cusum (i)}| \geq h_\gamma \cdot \sqrt{L+1}$ holds for some i (i = 1, . . ., n) a nonrandom arrangement of scores can be concluded and therefore a linear relationship between x* and y* is rejected (h_γ is tabulated in table 1).

It is obvious that the judgement of linearity depends also on the sampling distribution.

Tabl. 1. Critical values of the cusum statistic

γ (%)	h_γ
1	1.63
5	1.36
10	1.22

iii) *Test of the hypothesis* $\beta = 1$

In order to test this hypothesis we make use of the confidence interval for β. The hypothesis is accepted if the value of 1 is enclosed in this interval, otherwise it is rejected. A rejection of $\beta = 1$ demonstrates at least a proportional difference between the two methods. From the theory this test is not independent of the underlying distributions; however, our simulation study shows that in general it gives reliable results (see appendix and part II).

iv) *Test of the hypothesis* $\alpha = 0$

The hypothesis is accepted if the confidence interval for α contains the value of 0. This is a conservative test. If the hypothesis is rejected both methods differ at least by a constant amount (bias).

If we accept both $\beta = 1$ and $\alpha = 0$ we can infer $y^* = x^*$, or, in other words, both methods are identical.

4. Discussion and Examples

The basic concept of our regression procedure is due to *Theil* who developed this idea without reference to the problem of method comparison. His paper assumes x_i to be fixed and restricts itself to the estimation of β alone. Our estimation differs slightly from that of *Theil*, since it employs the offset K, i.e. the number of slopes less than − 1, to ascertain the relationship

$$b = \frac{1}{B}.$$

Consequently, parameter estimation is independent of the assignment of the methods to X and Y.

It is obvious that the estimators a and b are only meaningful if a linear relationship exists between x* and y*. Otherwise a and b cannot be interpreted. Clearly the new procedure takes into account the experimental reality of method comparisons as described in chapter 2.

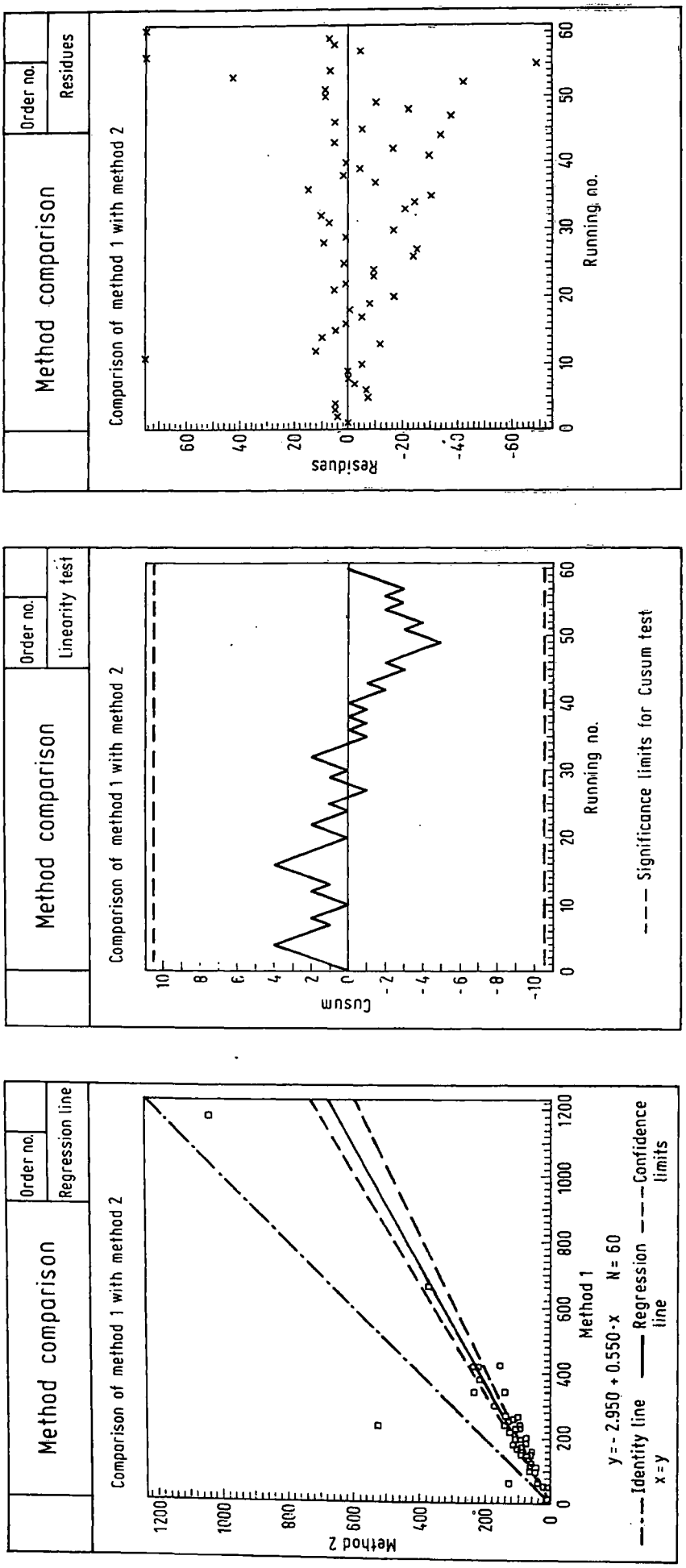


Fig. 1. Graphical description of method comparison.

The figure comprises 3 plots.

Plot 1 shows the observed values and the regression line (solid line) together with its confidence limits (dashed lines) which are obtained from the limits of the slope β and the intercept α. The identity $y = x$ is denoted by a dash-dotted line.

Plot 2 shows cusum as a function of the ranks of the (x_i, y_i) as described in the text. The dashed lines represent the significance limits of cusum.

Plot 3 shows the residual of each point in terms of the orthogonal distance to the fitted line as a function of the ranks of the (x_i, y_i).

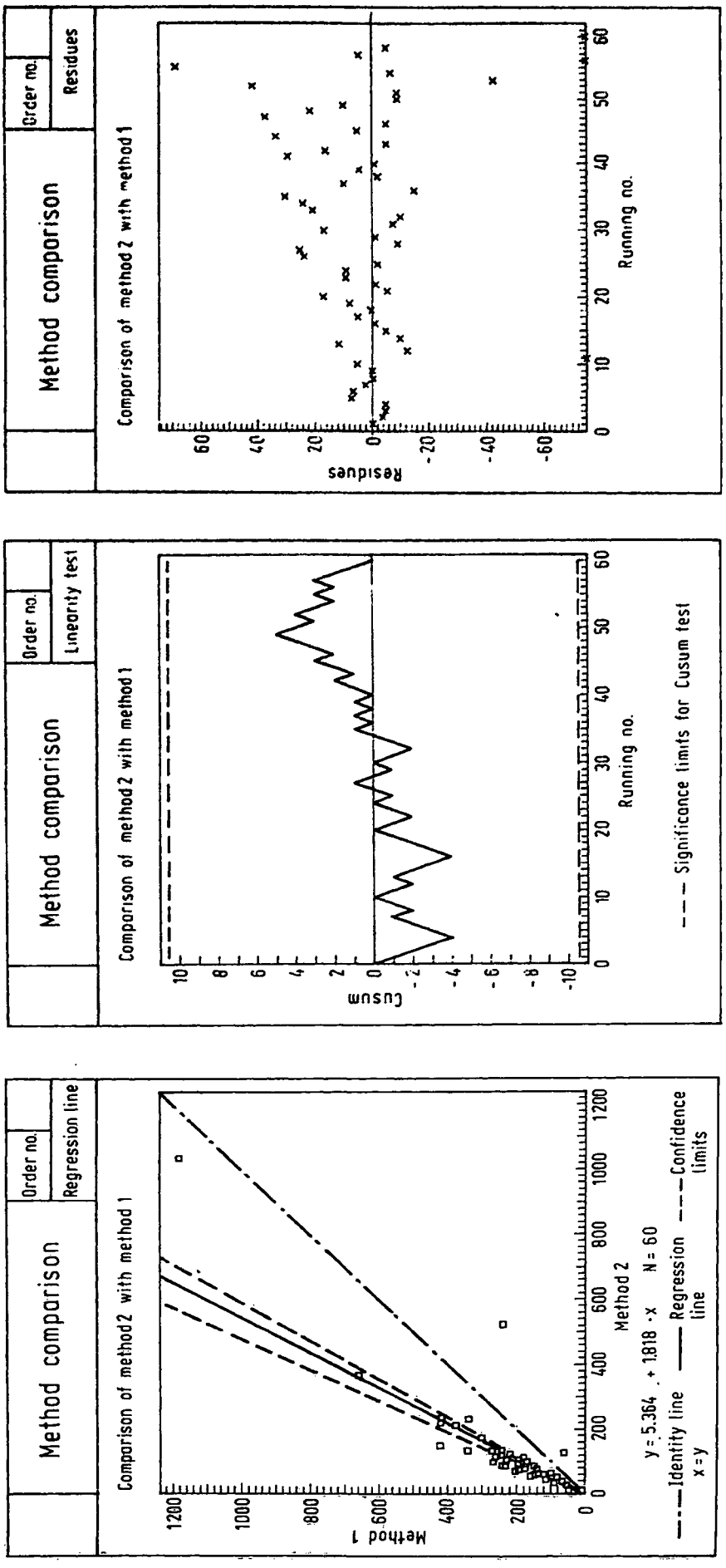


Fig. 2. Graphical description of the same method comparison as in figure 1 except that X and Y are interchanged. All plots can be obtained from the previous one by reflection along the line $y = x$ or along the orientation lines $y = 0$ (plots 2 and 3). This demonstrates that the new method does not depend on the choice of X and Y.

In particular the inconstancy of the variances is the reason for using only signs for the estimation of α and the test for linearity. Moreover, all measurement points (x_i, y_i) have equal weights in the estimation of the regression line; therefore extreme points do not show undue influence on the calculation. The same is true if the range of concentration is rather large (i.e. the range covers several powers of 10).

These theoretical arguments are supported by three figures all based on the same set of samples measured by methods 1, 2 and 3. In figure 1 method 1 is assigned to X and method 2 to Y. In figure 2 this assignment is interchanged. Obviously, all plots within figure 2 are obtained from the corresponding ones in figure 1 by reflection, showing the independence of the assignment to X and Y. We have chosen this example to demonstrate this property of our procedure even though the estimation of β is clearly different from 1.

In figure 3 we demonstrate how one extreme point can influence the outcome of the comparison of method 1 with method 3. First we calculate our procedure and the standardized principal component with the original data set, i.e. including the extreme point (1180, 1398). The estimations of β are 0.998 and 1.165 respectively; the latter is significantly different from 1. We then move the extreme point to the value (1180, 1190) and calculate a second time. The new results are 0.994 and 1.043, which are both not significant.

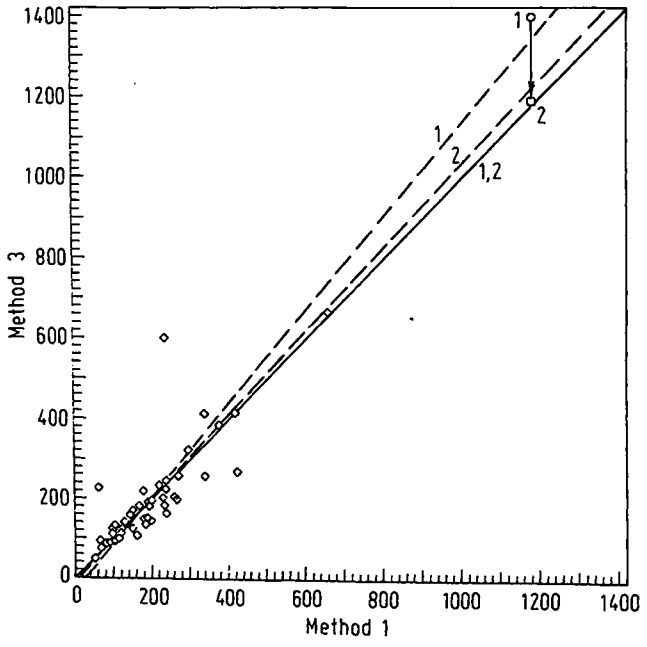


Fig. 3. Demonstration of the different behaviour of the new procedure and the standardized principal component when extreme points are present. Further details are given in the text.

— New procedure
--- Standardized principal component
◇ Basic data set
○ Original point in data set
□ Altered point in data set

It can be argued that this is not a fair comparison with a procedure for which the sampling of the data is not adequate. However, the data comes from a real experiment and the standardized principal component analysis is a recommended and often used evaluation procedure (4).

The procedures [1] to [4] are related to each other. For instance, the slope estimated by procedure [4] is equal to the geometric mean of the slope from procedure [1] and the reciprocal of the slope from procedure [2]. Furthermore, the slope of [4] will always be greater than or equal to that of [1], because $b[4] = b[1]/r$. Since b[4] is just the ratio of two standard deviations, it is independent of any joint function of X and Y. Under the assumption of normally distributed error terms and expected values the ratio

$$\beta = \frac{\sigma_\eta}{\sigma_\xi}$$

is estimated by the ratio of the standard deviations of the measurements. Since the standard deviation depends only on the sum of the variances of the two variation components it cannot reflect any independent change in their distributional properties. In contrast the estimator

$$b = \text{med}\left\{\beta \cdot \frac{d_{ij} + z_{ij}}{d_{ij} + z'_{ij}}\right\}$$

shows that it can respond to changes in both the sampling and the error term distribution.

Contrary to usual practice, we advise against a statement of the magnitude of dispersion or the coefficient of correlation from this experiment. The former adds no further information to the result of the regression procedure. The latter is a measure of association between X and Y and does not describe a functional relationship; besides, it has been shown (16) that its use can lead to erroneous inferences. Any other properties of the methods must be demonstrated by additional experiments using the appropriate statistics.

The computation of the new procedure appears to be rather tedious since the slopes S_{ij} must be sorted. This, however, is required by every statistic which calls for ranking. The calculation can easily be carried out on any computer with at least the size of a mini; several working programs are available at

present for various computers. For small desk computers with a standard memory size we have written a PASCAL program which allows the evaluation of up to 70 samples. With sufficient memory this number can easily be extended. This program is available on request. In addition a BASIC program written for a HP 85 desk computer can be requested; it can easily be adapted to the BASIC-version of other computers.

5. Appendix: Mathematical Derivations

1. What does b estimate?

The values for S_{ij} are identically distributed but not independent. Therefore the sample median of the S_{ij} may give a biased estimation of β. It is plausible that a somehow shifted median would be a better estimator. We cannot prove theoretically that the median shifted by our offset K is unbiased. However, we can demonstrate empirically that our procedure estimates β correctly in the case of the null hypothesis by using the following simulation model.

Let $[c_u, c_o]$ be the common range of concentrations in which both methods are applicable. It is assumed that both methods have constant coefficients of variation CV_ξ and CV_η in $[c_u, c_o]$. Let

$$c := \frac{c_o}{c_u}.$$

The range of both methods is transformed into the interval $\left[\frac{1}{c}, 1\right]$; in doing so CV_ξ and CV_η remain unchanged. On $\left[\frac{1}{c}, 1\right]$ n samples are drawn with "true values" x_i^* and $y_i^* = x_i^*$ for $i = 1, \ldots, n$ from two different distributions respectively: one in which the x_i^* are equidistant on $\left[\frac{1}{c}, 1\right]$, and one where the samples are skewly distributed over $\left[\frac{1}{c}, 1\right]$.

The "true values" x_i^* and y_i^* are distorted by independent "measurement errors" ξ_i and η_i giving "measured values" $x_i = x_i^* + \xi_i$ and $y_i = y_i^* + \eta_i$. Three types of distribution of "measurement errors" are considered: normal distribution, mixture of two normal distributions, and a skew distribution. c is varied between 2 and ∞, n from 40 to 90 and both CV's are varied independently of each other from 1% to 13%. The slope b is calculated for every of 500 data sets which are generated for each choice of parameters and distributions and the median of this 500 slope estimations is computed. The deviation of this median from $\beta = 1$ is an estimate of the bias of b. From the simulation we find that b is unbiased for CV's $\leq 7\%$. The details and the behaviour of our procedure compared with 6 others are given in part II of this paper.

From the above, it follows that a estimates α.

2. The procedure is independent of the assignment to X and Y

For

$$y^* = a + bx^* \text{ and } x^* = A + By^*$$

we show that

$$B = \frac{1}{b} \text{ and } A = -\frac{a}{b}.$$

We define

$$(1)\quad \omega_{ij} = \begin{cases} \operatorname{arctg} S_{ij} \\ \quad \text{if } -1 < S_{ij} \leq \infty \\ \quad (\text{i.e. } -45° < \operatorname{arctg} S_{ij} \leq 90°) \\ \\ \operatorname{arctg} S_{ij} + 180° \\ \quad \text{if } -\infty \leq S_{ij} < -1 \\ \quad (\text{i.e. } -90° \leq \operatorname{arctg} S_{ij} < -45°) \end{cases}$$

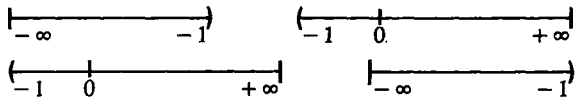
The domain of ω_{ij} with $-45° < \omega_{ij} < 135°$ lies symmetrical to 45° which corresponds to the ideal slope of 1 for a regression line in method comparison. Since $S_{ij} = -1$ cannot uniquely be assigned to ω_{ij}, we have the choice of including these values in both assignments, or of excluding them – as we have done – from the calculation. If we now interchange X and Y, we find that ω_{ij} is transformed to $90° - \omega_{ij}$ and the rank order of the sorted ω_{ij} is reversed, but not changed in the sequence. If the slope conforms to

$$(2)\quad b = \operatorname{tg} \operatorname{med} \{\omega_{ij}\}$$

it follows that

$$B = \operatorname{tg} \operatorname{med} \{90° - \omega_{ij}\}$$
$$= \operatorname{tg} [90° - \operatorname{med} \{\omega_{ij}\}]$$
$$= \frac{1}{\operatorname{tg} \operatorname{med} \{\omega_{ij}\}} = \frac{1}{b}$$

To derive formula (2) let us consider the following two ways of ranking on $R \setminus \{-1\} \cup \{-\infty, \infty\}$ which for simplicity's sake are given in graphical form:



Ranking according to I gives the natural rank order of the S_{ij}; ranking according to II shows the corresponding order of the S_{ij} with $S_{ij} = \text{tg}\,\omega_{ij}$, if the ω_{ij} are ranked in the natural order.

Clearly, the sequence in II is the same as in I for the region $(-1, +\infty]$, only the first K values with $S_{ij} < -1$ are added to the end by a left round-shift. Therefore, if we sort the S_{ij} according to I it is sufficient to use K as an offset for the determination of the median with respect to rank order II.

$$b = \left\{ \begin{array}{ll} S_{\left(\frac{N+1}{2}+K\right)}, & \text{if N odd} \\ \frac{1}{2}\cdot\left[S_{\left(\frac{N}{2}+K\right)} + S_{\left(\frac{N}{2}+1+K\right)}\right], & \text{if N even} \end{array} \right\} \text{(rank order I)},$$

(3)

$$= \text{med}\,\{S_{ij}\} \qquad \text{(rank order II)}$$

$$= \text{med tg}\,\{\omega_{ij}\} \qquad \text{(natural ranking)}$$

$$= \text{tg med}\,\{\omega_{ij}\} \qquad \text{(natural ranking)}$$

The last equality is exact only for odd N's; however even for $N \geq 40$ the difference

$$S_{\left(\frac{N}{2}+1+K\right)} - S_{\left(\frac{N}{2}+K\right)}$$

will be sufficiently small to justify the equal sign.

The limits of the confidence interval for β can be transformed similarly if X and Y are interchanged:

$$S_{(M_1+K)} = \qquad \text{(rank order I)}$$

$$= S_{(M_1)} \qquad \text{(rank order II)}$$

$$= \left(\frac{y_i - y_j}{x_i - x_j}\right)_{(M_1)} \qquad \text{(rank order II)}$$

(4)

$$= \frac{1}{\left(\frac{x_i - x_j}{y_i - y_j}\right)_{(N+1-M_1)}} \qquad \text{(rank order II)}$$

$$= \frac{1}{S_{(N+1-M_1+K)}} \qquad \text{(rank order I, X and Y interchanged)}$$

The result of testing the hypothesis $\beta = 1$ is therefore independent of the assignment of the methods to X and Y.

Analogously we obtain for the intercept after interchanging X and Y:

$$A = \text{med}\,\{x_i - By_i\}$$

$$(5) \qquad = \frac{1}{b}\,\text{med}\,\{bx_i - b\cdot By_i\}$$

$$= -\frac{1}{b}\,\text{med}\,\{y_i - bx_i\} = -\frac{a}{b}.$$

The confidence interval for α can be transformed in the same manner:

$$A_U = \text{med}\,\{x_i - B_L y_i\}$$

$$= \frac{1}{b_U}\,\text{med}\,\{b_U x_i - y_i\}$$

$$= -\frac{a_L}{b_U},$$

and it follows that

$$A_L = -\frac{a_U}{b_L}.$$

The result of testing the hypotheses $\alpha = 0$ is therefore independent of the choice of X.

In the cusum-test the rank order of the D_i and of the r_i remains unchanged if X and Y are interchanged, only the sign of the r_i is reversed. Since the test statistic is $|\,\text{cusum (i)}\,|$ the result is independent of the assignment.

3. *Justification of confidence intervals*

Let

$$K_{(-\infty,-1)} = \#\,\{i \mid S_{(i)} < -1\}$$

$$K_{(-1,0)} = \#\,\{i \mid -1 < S_{(i)} < 0\} + \#\,\{(i,j) \mid y_i = y_j \text{ and } x_i < x_j\}.$$

The last equation in formula (4) is valid if

$$N + 1 - M_1 + K_{(-\infty,-1)} \leq N,$$

that is if $K_{(-\infty,-1)} < M_1$ holds. Moreover, after interchanging X and Y this condition transforms into $K_{(-1,0)} < M_1$. Therefore, the conversion of the limits after interchanging X and Y works if

$$(6) \qquad K_{(-\infty,-1)} < M_1 \text{ and } K_{(-1,0)} < M_1$$

hold.

To justify the formula for M_1 and to give a sufficient condition for formula (6) we proceed as follows. In l.c. (17) it is shown that a confidence interval for β can be constructed by determining all those β's for which

$$x_i \text{ and } R_i := y_i - \beta x_i$$

are not significantly correlated according to *Kendall's* τ. Let

$$(7)\quad \begin{aligned} P(\beta) &= \#\{(i,j) \mid (x_i - x_j)\ (R_i - R_j) > 0\} \\ Q(\beta) &= \#\{(i,j) \mid (x_i - x_j)\ (R_i - R_j) < 0\} \end{aligned}$$

Then $P(\beta)+Q(\beta) = N$ with probability 1.

From

$$(x_i - x_j)\,(R_i - R_j) = (x_i - x_j)^2\,(S_{ij} - \beta)$$

follows that

$$P(\beta) = \#\{(i,j) \mid S_{ij} > \beta\}$$
$$Q(\beta) = \#\{(i,j) \mid S_{ij} < \beta\}.$$

Therefore the condition

$$S_{(M_1+K)} < \beta < S_{(M_2+K)}$$

is equivalent to

$$M_1 + K \le Q(\beta) \quad \text{and} \quad M_1 - K \le P(\beta)$$

and thus to

$$2M_1 - N \le P(\beta) - Q(\beta) + 2K \le N - 2M_1$$

The distribution of $C := P(\beta) - Q(\beta)$ does not depend on the distribution of (X, Y) whereas the distribution of K clearly does so. Therefore, it is impossible to *derive* a formula for M_1 satisfying

$$\begin{aligned} &P\{S_{(M_1+K)} < \beta < S_{(M_2+K)}\} \\ &\quad = P\{2M_1 - N \le C + 2K \le N - 2M_1\} \\ &\quad = 1 - \alpha \end{aligned}$$

completely independent of the distribution of (X, Y). However, C is asymptotically normal distributed with $E(C) = 0$ and

$$\text{Var}\,(C) = \frac{n(n-1)\,(2n+5)}{18}$$

such that $P\{-C_\gamma \le C \le C_\gamma\} = 1-\gamma$ holds, with C_γ defined in chapter 3. Therefore, M_1 is *defined* by

$$N - 2M_1 = C_\gamma$$

or

$$M_1 = \frac{N - C_\gamma}{2}.$$

We studied the properties of the confidence interval for β on the definition of M_1 in the simulation model and obtained the following result: If both methods have the same precision then in all cases the actual confidence level is about 95%; it is never less than 91% or higher than 96%. More details are given in part II of this paper.

Therefore it can be concluded that for method comparisons in clinical chemistry the proposed confidence interval for β has the actual level of about 95%.

The empirical derivation of this statement might seem unsatisfactory. But the same simulation model can also be used to demonstrate the behaviour of the other regression procedures mentioned in chapter 2 under realistic conditions. In our second paper we shall show the favourable properties of our method when compared with the others.

A sufficient condition for (6) is

$$Q(0) = K_{(-\infty,-1)} + K_{(-1,0)} < M_1$$

or

$$C = N - 2Q(0) > N - 2M_1 = C_\gamma;$$

this is true if X and Y show a significant positive correlation according to *Kendall's* τ.

Finally, the actual level of the confidence interval for α is higher than 95%. This is also confirmed from the simulation model.

4. *Test of linearity – Derivation of the cusum statistic*

The cusum-test is conditional on a and b; therefore the D_i are conditionally independent. We divide the D_i into two sets, one with scores $r_i > 0$ and one with $r_i < 0$; their empirical distribution function is denoted by F_1 and G_L respectively. Then, for

$$\lambda \in [D_{(i)}, D_{(i+1)})$$

we get

$$F_1(\lambda) = \frac{1}{l} \cdot \sum_{\substack{k=1 \\ r_{(k)} > 0}}^{i} 1,$$

$$G_L(\lambda) = \frac{1}{L} \cdot \sum_{\substack{k=1 \\ r_{(k)} < 0}}^{i} 1$$

and

$$\begin{aligned} F_1(\lambda) - G_L(\lambda) &= \frac{1}{\sqrt{l \cdot L}} \cdot \sum_{\substack{k=1 \\ r_{(k)} > 0}}^{i} r_{(k)} \\ &\quad + \frac{1}{\sqrt{l \cdot L}} \cdot \sum_{\substack{k=1 \\ r_{(k)} < 0}}^{i} r_{(k)} \\ &= \frac{1}{\sqrt{l \cdot L}} \cdot \sum_{k=1}^{i} r_{(k)} \end{aligned}$$

It follows that

$$T: = \sup_{\lambda \in R} | F_l(\lambda) - G_L(\lambda) |$$

$$= \frac{1}{\sqrt{l \cdot L}} \max_{1 \le i \le n} \sum_{k=1}^{i} r_{(k)}$$

$$= \frac{1}{\sqrt{l \cdot L}} \cdot \max_{1 \le i \le n} | \text{cusum (i)} |,$$

and

$$P\left\{ \max_{1 \le i \le n} | \text{cusum (i)} | < h_\gamma \cdot \sqrt{l + L} \right\}$$

$$= P\left\{ \sqrt{\frac{l \cdot L}{l + L}} \cdot T < h_\gamma \right\} = 1 - \gamma,$$

with h_γ being the critical value of the *Kolmogorov-Smirnov* statistic (18).

References

1. Halperin, M. (1961) J. Amer. Statist. Assoc. *56*, 657–669.
2. Madansky, A. (1959) J. Amer. Statist. Assoc. *54*, 173–205.
3. Stamm, D. (1979) J. Clin. Chem. Clin. Biochem. *17*, 277–279 and 280–282.
4. Haeckel, R. (1982) J. Clin. Chem. Clin. Biochem. *20*, 107–110.
5. Feldmann, U., Schneider, B., Klinkers, H. & Haeckel, R. (1981) J. Clin. Chem. Clin. Biochem. *19*, 121–137.
6. Ricker, W.E. (1973) J. Fish. Res. Board Can. *30*, 409–434.
7. Jolicoeur, P. (1975) J. Fish. Res. Board Can. *32*, 1491–1494.
8. Michotte, Y. (1978) Evaluation of precision and accuracy – comparison of two procedures, In: Evaluation and optimization of laboratory methods and analytical procedures (Massart, D.L., Dijkstra, A. & Kaufmann, L., eds.), Elsevier, Amsterdam, Oxford, New York.
9. Heiler, S. (1980) Robuste Schätzung im Linearen Modell, In: Robuste Verfahren (Nowak, H. & Zentgraf, R., eds.), Springer Verlag, Berlin.
10. Wolf, G.K. (1980) Praktische Erfahrung mit R-robusten Verfahren bei klinischen Versuchen, In: Robuste Verfahren (Nowak, H. & Zentgraf, R., eds.), Springer Verlag, Berlin.
11. Maritz, J.S. (1981) Distribution-Free Statistical Methods, Chapman and Hall, London.
12. Theil, H. (1950) Proc. Kon. Ned. Akad. v. Wetensch. AS3, Part I 386–392, part II 521–525, part III 1397–1412.
13. Bradley, J.V. (1968) Distribution free statistical tests. Prentice Hall, Englewood Cliffs, N.J.
14. Wold, S. & Sjöström, M. (1978) Linear free energy relationship as tools for investigating chemical similarity – Theory and Practice, In: Correlation Analysis in Chemistry, Recent Advances (Chapman, N.B. & Shorter, J., eds.), Plenum Press, New York and London.
15. Van Dobben de Bruyn, C.S. (1968) Cumulative Sum Tests: Theory and Practice, Griffin, London.
16. Cornbleet, P.J. & Shea, M.C. (1978) Clin. Chem. *24*, 857–861.
17. Hollander, M. & Wolfe, D.A. (1973) Nonparametric Statistical Methods, J. Wiley & Sons, New York.
18. Witting, H. & Nölle, G. (1970) Angewandte mathematische Statistik, B.G. Teubner, Stuttgart.

Dr. H. Passing
Abtlg. für Praktische Mathematik, Hoechst AG
D-6230 Frankfurt/Main 80

W. Bablok
Allg. Biometrie, Boehringer Mannheim GmbH
D-6800 Mannheim 31