

Zeitschrift:	Mitteilungen aus dem Gebiete der Lebensmitteluntersuchung und Hygiene = Travaux de chimie alimentaire et d'hygiène
Herausgeber:	Bundesamt für Gesundheit
Band:	88 (1997)
Heft:	1
Artikel:	Laborvergleichsuntersuchungen bei Vorhandensein einer Varianzfunktion vom Horwitz-Typ = A horwitz-like function supports evaluation of a laboratory-performance study
Autor:	Lischer, Peter
DOI:	https://doi.org/10.5169/seals-982316

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 25.01.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Laborvergleichsuntersuchungen bei Vorhandensein einer Varianzfunktion vom Horwitz-Typ*

A Horwitz-like Function Supports Evaluation of a Laboratory-
Performance Study

Key words: Horwitz function, Reproducibility precision, Proficiency test,
Collaborative trial, Analysis of variance, Power transformation, Scoring system

Peter Lischer

Eidgenössische Forschungsanstalt für Agrarökologie und Landbau (FAL),
Institut für Umweltschutz und Landwirtschaft (IUL), Bern

Laborbewertungssysteme

Unter einem *Laborbewertungssystem* ist ein Gesamtkonzept zu verstehen, in dessen Rahmen Laboratorien hinsichtlich ihrer Leistungsfähigkeit kontinuierlich bewertet beziehungsweise bezüglich gegebener Minimalanforderungen überprüft werden (1). Insbesondere dann, wenn die Laboratorien gewisse Freiheiten in der Wahl des Analysenverfahrens haben und wenn nicht jeder einzelne Verfahrensschritt vorgeschrieben wird, ist davon auszugehen, dass ein effizientes und von den Laboratorien akzeptiertes Laborbewertungssystem zu einer erheblichen Verbesserung der analytischen Qualität führen kann. Eine ganze Reihe von Vorschlägen zur Laborbewertung (*Laborvergleichsuntersuchung*) auf Basis der sogenannten *z-Scores* findet sich im AOAC/ISO/IUPAC-Protokoll (2). Stammen die einzelnen Laborresultate x aus einer normalverteilten Grundgesamtheit $N(\mu, \sigma^2_R)$, so drückt der Absolutbetrag der Prüfgrösse $z = (x - \mu) / \sigma_R$ den Abstand des Laborwertes x vom «wahren Wert» μ in Einheiten der Vergleichstandardabweichung σ_R aus und die Wahrscheinlichkeit, dass z eine vorgegebene Schranke übersteigt, kann den entsprechenden statistischen Tabellen entnommen werden. In der Praxis werden deshalb Messungen mit *z-Scores* mit Absolutwerten über 3 als «schlecht» taxiert, zwischen 2 und 3 als «fraglich» und kleiner 2 als «genügend» oder «gut». Eine solche Taxierung muss aber mit Vorsicht angewandt werden, da statt der Werte μ und σ_R i. a. nur Schätzungen m und s_R vorliegen und reale Verteilungen oft

* Ein Teil dieser Arbeit wurde während eines Aufenthaltes an der Bundesanstalt für gesundheitlichen Verbraucherschutz und Veterinärwesen (BgVV) in Berlin verfasst.

beträchtlich von der Normalverteilung abweichen. Es ist darauf hinzuweisen, dass der z-Score sensibel auf Schwankungen der Vergleichstandardabweichung reagiert. Insbesondere wenn diese nur auf der Basis weniger Laboratorien geschätzt wurde, sollte man mit den ermittelten Ergebnissen sehr behutsam umgehen.

Da in Laborvergleichsuntersuchungen in der Regel nicht nur eine, sondern mehrere Proben bezüglich mehrerer Merkmale untersucht werden, stellt sich die Frage, wie aus den z-Scores, die für jede der untersuchten Probe-Merkmal-Kombinationen gebildet werden, Rückschlüsse auf die Laborqualität gezogen werden können. Man benötigt also sogenannte *Kombinationsscores*, welche die Informationen der Scores in einem einzigen Zahlenwert verdichten. Im AOAC/ISO/IUPAC-Protokoll (2) werden mehrere Kombinationsscores diskutiert, so die Summe der Einzelscores sowie die Summe der quadrierten Einzelscores. Eindeutige Empfehlungen werden nicht gegeben. Modifizierte Quadratsummenscores, die eine mögliche Korrelation zwischen den z-Scores berücksichtigen, wurden von *Lischer* (3-5) vorgeschlagen. Bei diesen Scores handelt es sich im wesentlichen um robuste Distanzmasse, die sich von der klassischen Mahalanobis-Distanz (6) nur dadurch unterscheiden, dass die zugehörigen Kovarianzmatrizen robust geschätzt wurden. Formal mathematisch-statistisch handelt es sich bei all diesen Methoden um multivariate Ausreissererkennung. Laboratorien, die Analysenwerte liefern, die sich stark von denjenigen der Mehrheit der anderen Laboratorien unterscheiden, werden als unzuverlässig taxiert. Wie *Uhlig* (1) nachwies, ist dieser Ansatz unbefriedigend. Die Beweislast für die Laborqualität sollte beim Labor, nicht bei den Bewertenden liegen. Die Laboratorien müssten den Nachweis erbringen, dass sie korrekt arbeiten.

Ein weiterer Nachteil all dieser Kombinationsscores besteht darin, dass es sich nur um relative Größen handelt. Anhand eines solchen relativen Massstabes lässt sich nur ermitteln, welche Leistungen das Labor im Vergleich zu den übrigen an den Laborvergleichsuntersuchungen beteiligten Laboratorien erbracht hat. Absolute Massstäbe dagegen ermöglichen die Bewertung der Laboratorien unabhängig von den übrigen Laboratorien. Besonders bei periodisch wiederholten Ringversuchen sind absolute Massstäbe wünschbar, denn die Teilnehmer möchten wissen, ob seit der letzten Durchführung ein Fortschritt stattgefunden hat oder nicht. Ein möglicher absoluter Massstab ist der von *Uhlig* (1) eingeführte mittlere Laborvariationskoeffizient.

Im folgenden wird ein alternatives *absolutes Bewertungskriterium* vorgestellt. Es ist anwendbar, wenn eine einfache funktionale Beziehung zwischen der relativen Vergleichstandardabweichung $RSD_R(\%)$ und der Konzentration c besteht. Durch Vergleich der Ergebnisse von Ringversuchen zur Prüfung verschiedenster Analysenmethoden aus dem Lebensmittel- und Pharmabereich etablierte *Horwitz* (7) empirisch die nach ihm benannte Beziehung zwischen der relativen Vergleichstandardabweichung $RSD_R(\%)$ und der Konzentration c , ausgedrückt als Dezimalbruch (z. B. für g/g, $c = 10^0 = 1$; für ppm, $c = 10^{-6}$). Sie kann durch die Gleichung

$$RSD_R(\%) = 2^{(1-0,5 \log_{10} c)} \approx 2c^{-0,1501} \quad (I)$$

ausgedrückt werden. Äquivalent dazu sind

$$s_R = 0,02 c^{0,8495} \quad (II)$$

und

$$\log_{10} s_R = 0,8495 \log_{10} c - 1,6990 \quad (III)$$

wobei s_R die Vergleichsstandardabweichung bedeutet. Ist beispielsweise $c = 1$, muss mit einem Variationskoeffizienten von $\approx 2\%$ gerechnet werden. Für $c = 10^{-3}$ beträgt er $\approx 5,7\%$, für $c = 10^{-6} \approx 16\%$ und für $c = 10^{-9} \approx 45\%$. Als Horwitz-Bereich wird das Gebiet bezeichnet, das zwischen den Kurven $s_R = 0,01 c^{0,8495}$ und $s_R = 0,04 c^{0,8495}$ liegt. *Thompson* (8) fand in einer Laborvergleichsuntersuchung nahezu die gleiche Beziehung. Auch bei mehreren Laborvergleichsuntersuchungen in der Schweiz konnte ein linearer Zusammenhang zwischen den logarithmierten Größen s_R und c nachgewiesen werden, wobei sich aber die Regressionsparameter in der logarithmierten Form der Gleichung (III) mehr oder weniger stark von denjenigen der Horwitz-Funktion unterscheiden.

Varianzstabilisierende Transformationen

Das Datenmaterial von Laborvergleichsuntersuchungen lässt sich in einer zweifaktoriellen Datenmatrix darstellen. In jeder Zeile stehen die Resultate sämtlicher Probe-Merkmal-Kombinationen der einzelnen Laboratorien. Da meistens die Standardabweichung der Analysenergebnisse mit steigender Konzentration zunimmt, ist eine der Grundvoraussetzungen der klassischen Varianzanalyse, nämlich gleiche Varianzen, offensichtlich nicht erfüllt. Um das Instrumentarium der Varianzanalyse anwenden zu können, ist eine Normierung unumgänglich. Mit der Transformation der Originaldaten in z-Scores können gleiche Varianzen erzwungen werden. Diese Transformation hat aber Nachteile. Sind nur wenige Analysen für die betreffende Probe-Merkmal-Kombination vorhanden, werden die Schätzungen für den Referenzwert, den Skalenparameter und folglich für die betreffenden z-Scores ungenau. Ergibt sich für eine spezielle Probe-Merkmal-Kombination eine sehr grosse Streuung, ist dies aus den abgeleiteten z-Scores nicht mehr ersichtlich, da es sich um normierte Größen handelt.

Eine andere Transformationsmöglichkeit bietet sich an, wenn eine wie in Gleichung (III) beschriebene Beziehung der Form $\log_{10} s_R = a + b \log_{10} c$ existiert. In diesem Fall existiert eine varianzstabilisierende Transformation. Die Daten brauchen nur mit dem Exponenten $(1-b)$ potenziert zu werden. Dann erst ist es zulässig, eine varianzanalytische Zerlegung vorzunehmen, d. h. jedes einzelne transformierte Resultat y_{ij} in 4 Komponenten aufzuspalten:

$$y_{ij} = m + l_i + p_j + r_{ij}, \quad i = 1, 2, \dots, n_L; \quad j = 1, 2, \dots, n_P \quad (IV)$$

wobei

- m allgemeiner Mittelwert der gemessenen Eigenschaft des Testmaterials
- l_i Laboreffekt i (Systematischer Fehler des Laboratoriums);

- p_j Probeneffekt j (Niveau der einzelnen Probe-Merkmal-Kombination);
- r_{ij} Residuum.

Die Anzahl der Laboratorien wird mit n_L und die Anzahl der Probe-Merkmal-Kombinationen mit n_p bezeichnet.

Mit dieser Zerlegung wird versucht, die Gesamtstreuung eines Laboratoriums in eine systematische und in eine zufällige Komponente zu zerlegen.

In der klassischen Statistik wird m durch \bar{y} , $m + l_i$ durch den Zeilenmittelwert \bar{y}_i , und $m + p_j$ durch den Spaltenmittelwert \bar{y}_j geschätzt. Da weder für die Originaldaten noch die transformierten Größen eine exakte Normalverteilung vorausgesetzt werden darf, sind die üblichen klassischen Schätzungen ungeeignet. Insbesondere können extreme Werte eines Laboratoriums auch Residuen der übrigen Laboratorien stark beeinflussen. Dies kann vermieden werden, wenn der arithmetische Mittelwert durch eine robuste Schätzung wie den Median oder den gestutzten Mittelwert ersetzt wird. Für Einzelheiten und Algorithmen sei auf (9) verwiesen.

Ein Beispiel einer Laborvergleichsuntersuchung mit simulierten Daten

20 Laboratorien L_1, L_2, \dots, L_{20} , 8 Proben P_1, P_2, \dots, P_8 . Wir nehmen an, dass die Messwerte normalverteilt sind mit Mittelwerten $c_1 = 0,05, c_2 = 0,2, c_3 = 0,7, c_4 = 1, c_5 = 4, c_6 = 10, c_7 = 60, c_8 = 150$, und einer Varianzfunktion der Form $\sigma = a c^b$ mit $a = 0,15$ und $b = 0,85$.

Abbildung 1 zeigt die untransformierten Daten, in Abbildung 2 sind die Daten nach der varianzstabilisierenden Transformation

$$y = x^{1-b} \quad (V)$$

dargestellt.

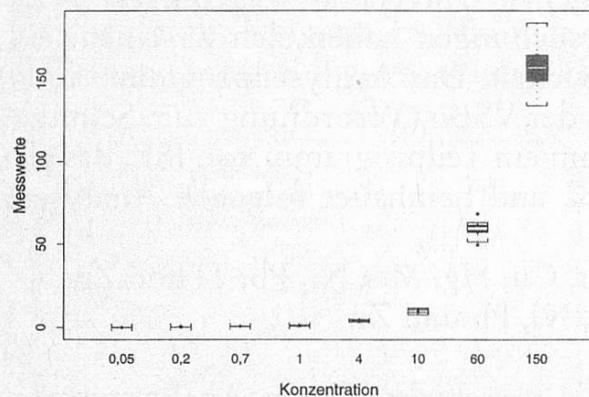


Abb. 1. Boxplots der untransformierten Laborresultate

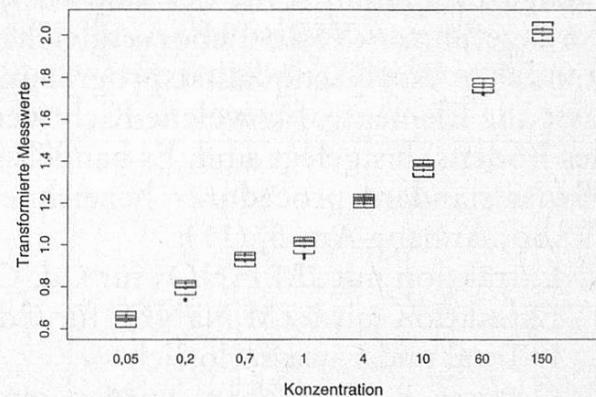


Abb. 2. Boxplots der transformierten Laborresultate

Sowohl die transformierten Messwerte der Probe P_j : $\langle y_{ij}, j \text{ fest}, i = 1, 2, \dots, 20 \rangle$ als auch die transformierten Messwerte von Laboratorium L_i : $\langle y_{ij}, i \text{ fest}, j = 1, 2, \dots, 8 \rangle$ haben alle genähert die gleiche Varianz $a^2 (1-b)^2$ (9). Das gleiche gilt auch

für die aus den transformierten Daten nach der Zerlegung (IV) abgeleiteten Residuen.

In den Abbildungen 3 und 4 sind die mit dem Faktor $f = (1 - b)^{-1}$ multiplizierten Residuen dargestellt. Nach Multiplikation mit f sind die Varianzen der Residuen genähert α^2 , d.h. der Vergleichsvarianz auf dem Konzentrationsniveau $c = 1$. Wegen der Transformation (V) ist die Verteilung der Residuen aber links-schief, was sich dadurch äussert, dass mehr kleine als grosse extreme Abweichungen vorkommen. Quadratsummen solcher Residuen sind demzufolge nur noch approximativ χ^2 – verteilt. Bei praktischen Anwendungen in der analytischen Chemie sind die Ausgangsdaten aber selten exakt normalverteilt, sondern meistens rechts-schief. In diesem Fall stabilisiert die Transformation (V) nicht nur die Varianzen, sondern symmetrisiert auch die Verteilungen.

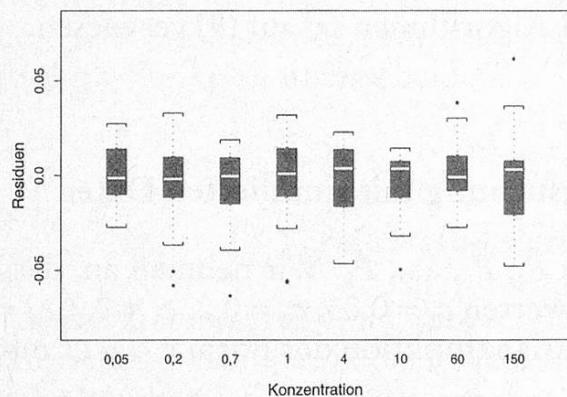


Abb. 3. Boxplots der Residuen, gruppiert nach Proben

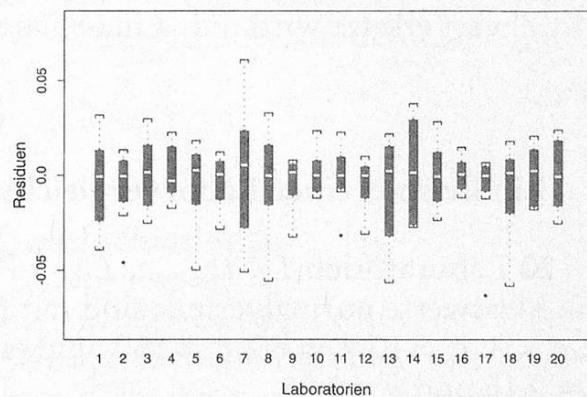


Abb. 4. Boxplots der Residuen, gruppiert nach Laboratorien

Anwendung: Die VSBo-Laborvergleichsuntersuchung

Die seit 1991 in Zusammenarbeit mit dem International Soil-Analytical Exchange Programm (ISE) der landwirtschaftlichen Universität Wageningen (NL) durchgeföhrten VSBo-Laborvergleichsuntersuchungen haben sich zu einem eigentlichen Analysenqualitätsprogramm entwickelt. Das Analysenprogramm umfasst alle Elemente, für welche Richtwerte in der VSBo (Verordnung zum Schutze des Bodens) festgelegt sind. Es handelt sich um ein Teilprogramm von ISE, das als «Swiss standard procedure» bezeichnet wird und beinhaltet folgende Analysen (VSBo, Anhang Art. 5) (11):

- Extraktion mit 2M HNO₃ für Cd, Co, Cr, Cu, Hg, Mo, Ni, Pb, Tl und Zn
- Extraktion mit 0,1M NaNO₃ für Cd, Cu, Ni, Pb und Zn
- F-Total und F-wasserlöslich

Gewisse Bodenproben wurden mehrmals verschickt. Deren Analysenwerte wurden in unserer Auswertung nur das erste Mal berücksichtigt. Für die Analysenwerte jeder einzelnen Probe wurde ein robuster Mittelwert m und eine robuste Standardabweichung s_R berechnet. In den Abbildungen 5 bis 8 sind in logarithmischer Darstellung Mittelwerte der Konzentrationen der einzelnen Proben gegen die zugehörigen Vergleichstreuungen samt Regressionsgeraden für die Jahre 1991

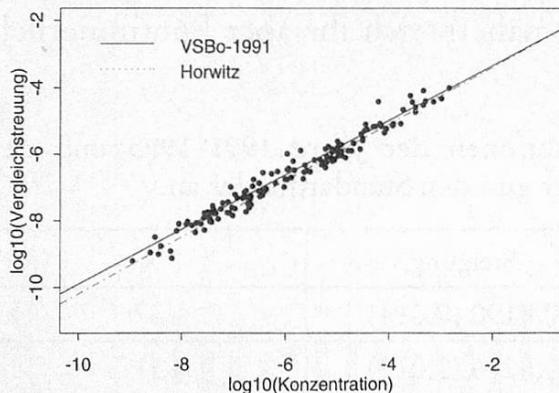


Abb. 5. VSBo-1991, Konzentrationen, Vergleichstreuungen, Regressionsgerade und Horwitzkurve

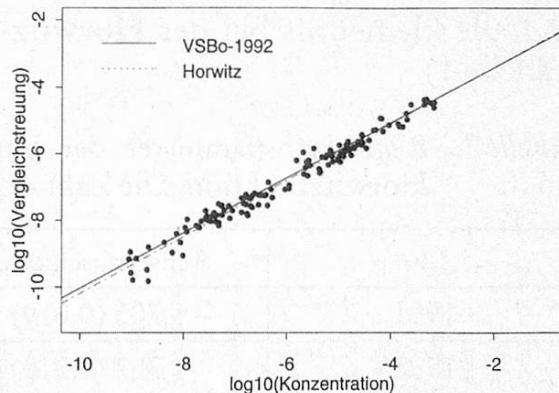


Abb. 6. VSBo-1992, Konzentrationen, Vergleichstreuungen, Regressionsgerade und Horwitzkurve

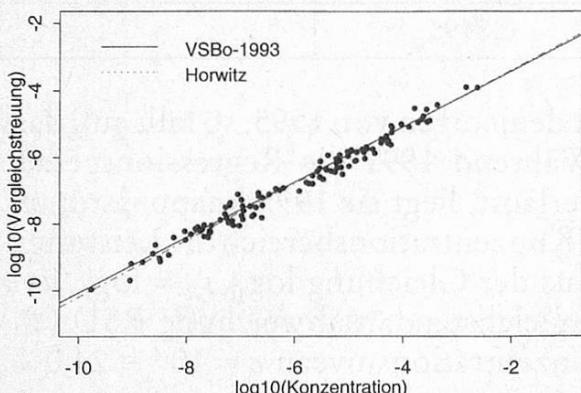


Abb. 7. VSBo-1993, Konzentrationen, Vergleichstreuungen, Regressionsgerade und Horwitzkurve

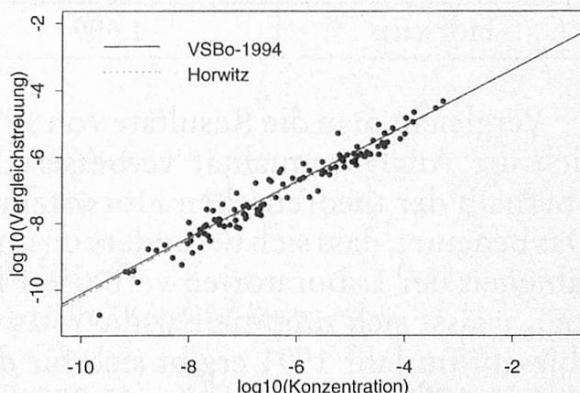


Abb. 8. VSBo-1994, Konzentrationen, Vergleichstreuungen, Regressionsgerade und Horwitzkurve

bis 1994 aufgezeichnet, in Abbildung 9 die entsprechenden Größen für das Jahr 1995. Zusätzlich sind hier noch die Grenzen des Horwitz-Bereiches für die Vergleichstreuung aufgetragen. Es fällt auf, dass die Achsenabschnitte $\log_{10} a$ der einzelnen Geraden $\log_{10} s_R = \log_{10} a + b \log_{10} c$ wenig variieren. Die Steigungen b

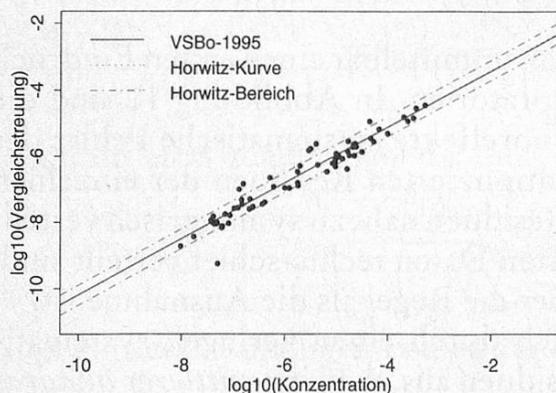


Abb. 9. VSBo-1995, Konzentrationen, Vergleichstreuungen, Regressionsgerade und Horwitzbereich

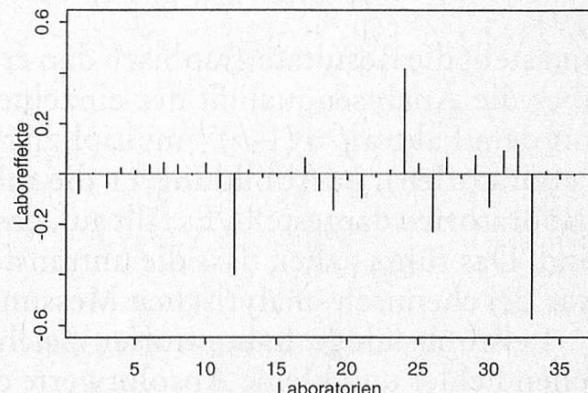


Abb. 10. VSBo-1995, Laboreffekte (Systematische Fehler der einzelnen Laboratorien)

sind alle kleiner als bei der Horwitz-Kurve, nähern sich ihr aber kontinuierlich (Tabelle 1).

Tabelle 1. Regressionsparameter der Varianzfunktionen der Jahre 1991–1995 und der Horwitzfunktion. Die Zahl in Klammer gibt den Standardfehler an.

Jahr	Achsenabschnitt	Steigung	n
1991	– 1,705 (0,019)	0,8190 (0,014)	37
1992	– 1,757 (0,019)	0,8276 (0,012)	31
1993	– 1,778 (0,020)	0,8285 (0,013)	30
1994	– 1,702 (0,026)	0,8456 (0,020)	35
1995	– 1,727 (0,026)	0,8456 (0,020)	36
Horwitz	– 1,699	0,8495	

Vergleicht man die Resultate von 1991 mit denjenigen von 1995, so fällt auf, dass sich die Analysenqualität verbessert hat. Während 1991 die Regressionsgerade oberhalb der theoretischen Horwitzkurve verläuft, liegt sie 1995 knapp darunter. Das bedeutet, dass sich besonders im unteren Konzentrationsbereich die Leistungsfähigkeit der Laboratorien verbessert hat. Aus der Gleichung $\log_{10} s_R = \log_{10} a + b \log_{10} c$ lässt sich unmittelbar die relative Vergleichstandardabweichung RSD (%) ablesen. Im Jahr 1991 ergibt sich für das Konzentrationsniveau $c = 10^{-6} \approx 24,0\%$, im Jahr 1995 aber nur noch $\approx 15,8\%$.

Bewertung der Laboratorien

Potenziert man die Daten von 1995 mit dem Exponenten $(1-b)$ mit $b = 0,8456$ (gemäss Gleichung (5)), unterzieht sie der oben erwähnten varianzanalytischen Zerlegung

$$y_{ij} = m + l_i + p_j + r_{ij}, \quad i = 1, 2, \dots, n_L; \quad j = 1, 2, \dots, n_P$$

und stellt die Resultate graphisch dar, erhält man unmittelbar einen ersten Eindruck über die Analysenqualität der einzelnen Laboratorien. In Abbildung 10 sind die mit dem Faktor $f = (1-b)^{-1}$ multiplizierten Laboreffekte (systematische Fehler der Laboratorien), in Abbildung 11 die mit f multiplizierten Residuen der einzelnen Laboratorien dargestellt. Es fällt auf, dass die Residuen nahezu symmetrisch verteilt sind. Das röhrt daher, dass die untransformierten Daten rechts-schief verteilt sind, was bei chemisch-analytischen Messungen eher die Regel als die Ausnahme ist.

Leistungsfähige Laboratorien zeichnen sich durch einen geringen systematischen Fehler und kleine Absolutwerte der Residuen aus, d. h. ihr *mittlerer quadratischer Fehler* ist klein. Der mittlere quadratische Fehler eines Labors L_i ist zusammengesetzt aus dem quadrierten systematischen Fehler $f^2 l_i^2$ und dem «mittleren zufälligen Fehler» $f^2 \text{Var}\langle r_{ij} \rangle, j = 1, 2, \dots, n_i \rangle$. Da unter den Residuen r_{ij} wie üblich

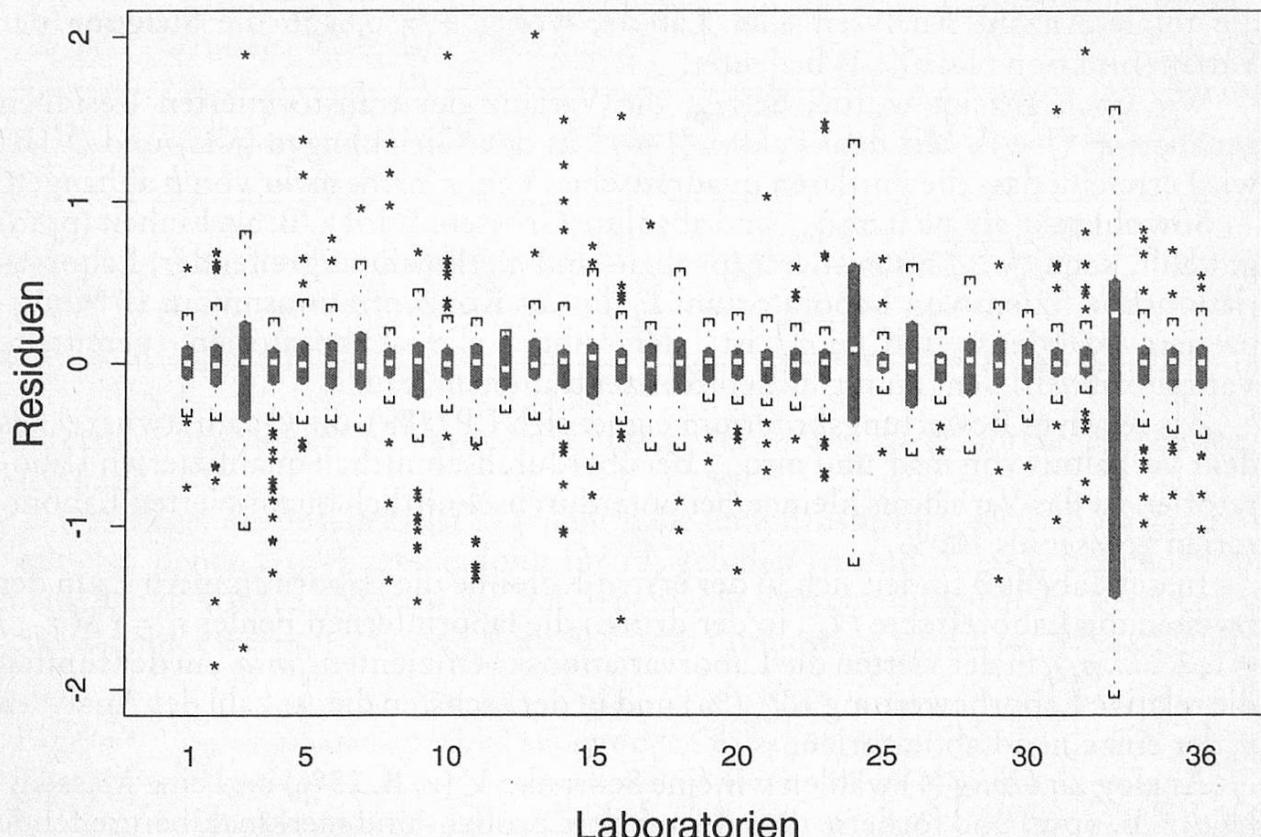


Abb. 11. VSBo-1995, Boxplots der Residuen der transformierten Messwerte

eine gewisse Anzahl extremer Werte auftreten, muss die Schätzung der Varianz entweder robust erfolgen oder es muss ein Kappungsverfahren analog dem AOAC/ISO/IUPAC-Protokoll (2) verwendet werden, bei dem überhöhte z -Scores auf den Wert 3 gestutzt werden. In dieser Arbeit wurde der robuste Skalenschätzer $S \diamond$ aus dem Schweizerischen Lebensmittelbuch (SLB), Kapitel 60A, p. 30–32 (10) verwendet. Dieser Schätzer ist dadurch charakterisiert, dass Residuen mit grossen Absolutbeträgen gestutzt werden. Der so berechnete mittlere quadratische Fehler erlaubt nun eine sinnvolle Klassierung der Laboratorien.

Wir bezeichnen mit

$$msq_i = \frac{1}{(1-b)^2} \left(l_i^2 + S^2 \langle r_{ij}, j=1,2, \dots, n_i \rangle \right) \quad (\text{VII})$$

den mittleren quadratischen Fehler von L_i , mit

$$msq_{tot} = \frac{1}{(1-b)^2} S^2 \langle r_{ij}, i=1, 2, \dots, n_L; j=1, 2, \dots, n_i \rangle \quad (\text{VIII})$$

den mittleren zufälligen Fehler der Laboratorien insgesamt und mit

$$N = \sum_{i=1}^{n_L} n_i \quad (\text{IX})$$

die totale Anzahl Analysen aller Labors, wobei $b = 0,8456$ die Steigung der Varianzfunktion (Tabelle 1) bedeutet.

Wie oben gezeigt wurde, beträgt die Varianz der transformierten Residuen genähert $a^2 (1-b)^2$. Mit dem Faktor $(1-b)^2$ in den Gleichungen (VII) und (VIII) wird erreicht, dass die mittleren quadratischen Fehler nicht mehr von b abhängen.

Sowohl msq_i als auch msq_{tot} sind absolute Größen. Wird z. B. als Einheit (ppm) gewählt, kann $\sqrt{msq_i}$ als mittlerer (proben- und merkmalübergreifender) Laborvariationskoeffizient von Laboratorium L_i für das Konzentrationsniveau 10^{-6} interpretiert werden, und $\sqrt{msq_{tot}}$ ist der über alle Laboratorien gemittelte Variationskoeffizient zum gleichen Konzentrationsniveau.

Als relatives Bewertungskriterium eignet sich $LP_{rel}(\%)$, die Quadratwurzel aus dem Verhältnis von msq_i und msq_{tot} . Bei überdurchschnittlich qualifizierten Laboratorien ist das Verhältnis kleiner, bei unterdurchschnittlich qualifizierten Laboratorien grösser als 100%.

In der Tabelle 3 finden sich in der ersten Kolonne die Labornummern L_i , in der zweiten die Laboreffekte $f l_i$, in der dritten die laborinternen Fehler $s_{ri} = \sqrt{f S \langle r_{ij} \rangle}$, in der vierten die Laborvariationskoeffizienten $\sqrt{msq_i}$, in der fünften die relative Laborbewertung $LP_{rel}(\%)$ und in der sechsten die Anzahl der Analysen n_i der einzelnen Laboratorien.

Analog zu Uhlig (1) wählen wir eine Schranke V_c (z. B. 25%) und eine Masseinheit (z. B. ppm) und fordern, dass der mittlere proben- und merkmalübergreifende Variationskoeffizient des betreffenden Labors kleiner sein soll als V_c :

$$\sqrt{msq_i} < V_c \quad (X)$$

Um die Qualitätsforderung statistisch gesichert nachzuweisen, muss die Nullhypothese

$$H_0: \sqrt{msq_i} \geq V_c$$

gegen die Alternative

$$H_1: \sqrt{msq_i} < V_c$$

getestet werden. Ist

$$\sqrt{msq_i} < V_c \sqrt{\chi^2_{n_i, 0, 10} / n_i} \quad (XI)$$

erfolgt eine Entscheidung für die Alternative H_1 zum Signifikanzniveau 10%, d. h. das Labor hat mit der Irrtumswahrscheinlichkeit 10% nachgewiesen, dass es die statistische Qualitätsforderung erfüllt. Dabei steht $\chi^2_{n_i, 0, 10}$ für das 10%-Quantil der χ^2 -Verteilung mit n_i Freiheitsgraden. Die kritischen Werte der χ^2 -Verteilung sind in Tabelle 2 aufgeführt. Eine Reduzierung des Signifikanzniveaus auf 5% ist wünschenswert, wird jedoch dazu führen, dass eine höhere Zahl von Proben

Tabelle 2. Kritische Werte der χ^2 -Verteilung

n_i	$\sqrt{\chi^2_{n_i,0,5} / n_i}$	$\sqrt{\chi^2_{n_i,0,10} / n_i}$
5	0,479	0,568
10	0,628	0,700
20	0,737	0,789
30	0,785	0,829

untersucht werden muss, bevor die Qualitätsforderung als erfüllt angesehen werden kann.

Das Ergebnis des Tests ist in jedem Falle eine Einstufung der Laboratorien in jene, bei denen eine Entscheidung für H_1 gefallen ist und in jene, bei denen H_0 beibehalten wird: Laboratorien, bei denen eine Entscheidung für H_1 gefallen ist, erfüllen mit einer statistischen Sicherheit von mindestens 90% bzw. 95% die in H_1

Tabelle 3. Qualitätsparameter der Laboratorien (VSBO-1995)

L_i	$f l_i$	s_{r_i}	$\sqrt{msq_i}$	LP _{rel} (%)	n_i	L_i	$f l_i$	s_{r_i}	$\sqrt{msq_i}$	LP _{rel} (%)	n_i
1	-0,077	0,162	0,180	88,33	38	19	-0,143	0,128	0,192	69,77	42
2	0,002	0,165	0,165	89,78	75	20	-0,022	0,118	0,120	64,27	55
3	-0,054	0,455	0,458	247,70	33	21	-0,094	0,155	0,181	84,39	30
4	0,023	0,161	0,162	87,54	85	22	-0,011	0,139	0,139	75,58	75
5	0,011	0,213	0,213	115,94	57	23	0,015	0,161	0,161	87,49	44
6	0,036	0,198	0,201	107,81	70	24	0,417	0,620	0,747	337,79	23
7	0,044	0,210	0,214	114,33	71	25	0,040	0,096	0,104	52,30	65
8	0,036	0,113	0,118	61,33	85	26	-0,229	0,354	0,421	192,63	30
9	-0,003	0,298	0,298	162,40	46	27	0,064	0,136	0,150	73,93	23
10	0,042	0,140	0,146	76,22	85	28	-0,017	0,245	0,245	133,29	30
11	-0,057	0,195	0,204	106,49	70	29	0,074	0,230	0,241	125,27	42
12	-0,399	0,236	0,464	128,67	10	30	-0,130	0,197	0,236	107,55	30
13	0,014	0,185	0,186	100,86	40	31	0,092	0,143	0,170	77,86	65
14	-0,014	0,169	0,170	92,21	85	32	0,116	0,149	0,189	81,33	55
15	-0,007	0,254	0,254	138,53	65	33	-0,573	1,378	1,492	750,54	17
16	0,004	0,175	0,175	95,45	84	34	0,011	0,191	0,191	103,95	75
17	0,063	0,249	0,257	135,76	14	35	-0,017	0,177	0,178	96,33	63
18	-0,066	0,246	0,255	134,26	70	36	-0,005	0,182	0,182	99,14	75

genannte Qualitätsforderung. Bei Laboratorien, bei denen H_0 beibehalten wurde, ist zweifelhaft, jedoch nicht ausgeschlossen, dass sie die Qualitätsforderung erfüllten. Zumindest haben diese Laboratorien den Beweis noch nicht erbracht, den Qualitätsforderungen zu genügen. Wenn diese Laboratorien an weiteren Laborvergleichsuntersuchungen teilnehmen, wird sich herausstellen, ob sie die Qualitätsforderung erfüllen.

Um beispielsweise schon nach 10 untersuchten Proben statistisch nachweisen zu können, dass der proben- und merkmalübergreifende Laborvariationskoeffizient kleiner als 25% ist, muss der Laborvariationskoeffizient des jeweiligen Labors kleiner sein als $0,700 \cdot 25\% = 17,5\%$ (für das Signifikanzniveau 10%). Werden jedoch mehr Proben untersucht, so reduzieren sich die Anforderungen an den Laborvariationskoeffizienten. Relativ «schlechtere» Laboratorien müssen daher mehr Proben untersuchen, bevor sie den Nachweis erbringen können, die statistische Qualitätsforderung zu erfüllen. Wenn diese Laboratorien an weiteren Laborvergleichsuntersuchungen teilnehmen, wird sich herausstellen, ob sie die Qualitätsforderung erfüllen.

Datenanalyse

Aus den transformierten Daten lassen sich ebenfalls noch aufschlussreiche Zusatzinformationen über das Probenmaterial und eventuelle Matrixeffekte gewinnen. Dazu müssen die Residuen statt nach Laboratorien nach Probe-Merkmal-Kombinationen gruppiert werden. Für jede einzelne Probe lässt sich die Streuung der Residuen schätzen. In Tabelle 4 sind die entsprechenden robusten Schätzwerte aufgeführt.

In den Abbildungen 12 bis 16 sind die je 17 Analysen der 5 Einzelproben dargestellt, in der Abbildung 17 die Messwerte der Kontrollprobe 921 im 4. Quartal 1995. Diese Kontrollprobe wird jedes Quartal ein- oder zweimal mit den anderen Proben zusammen verschickt.

Tabelle 4. Standardabweichungen der Residuen

Probenummer	Standardabweichung
944	0,140
946	0,150
947	0,190
948	0,172
953	0,282
921	0,088

Es fällt auf, dass die Analysenresultate von Probe 921 viel weniger streuen als diejenigen der übrigen Proben. Ein möglicher Grund liegt darin, dass die meisten Teilnehmer diese Probe zu identifizieren imstande sind. Rückfragen bei den teilnehmenden Laboratorien haben diesen Verdacht weitgehend bestätigt. In den Abbildungen 12 bis 17 fällt ebenfalls auf, dass einige Elemente aus dem Rahmen fallen, z. B. Mo (Proben 946 und 948), F-löslich (Proben 947 und 948) und dass die

VSBo-1995: Probe 944

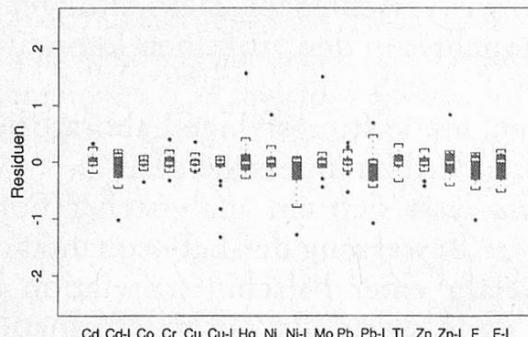


Abb. 12. Probe 944: Boxplots der Residuen der transformierten Laborresultate

VSBo-1995: Probe 946

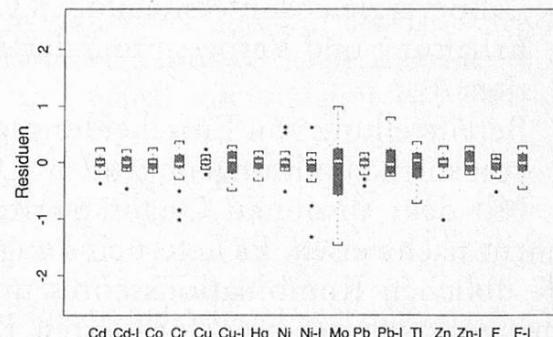


Abb. 13. Probe 946: Boxplots der Residuen der transformierten Messwerte

VSBo-1995: Probe 947

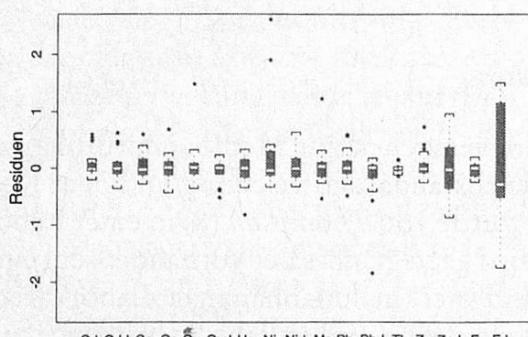


Abb. 14. Probe 947: Boxplots der Residuen der transformierten Messwerte

VSBo-1995: Probe 948

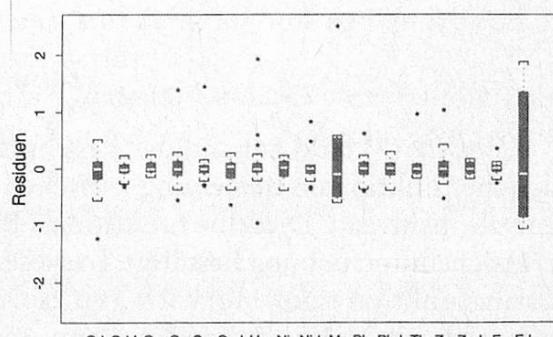


Abb. 15. Probe 948: Boxplots der Residuen der transformierten Messwerte

VSBo-1995: Probe 953

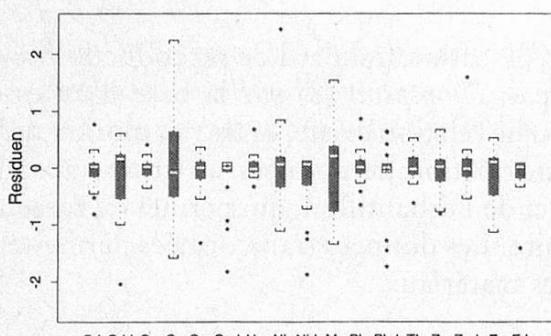


Abb. 16. Probe 953: Boxplots der Residuen der transformierten Messwerte

VSBo-1995: Probe 921

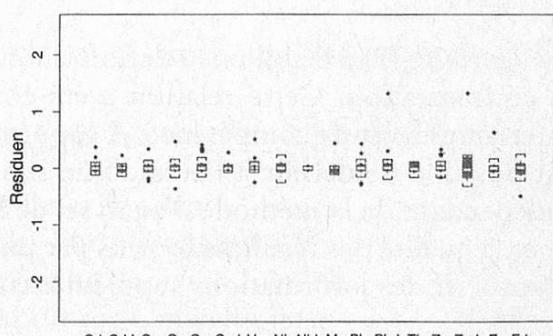


Abb. 17. Probe 921: Boxplots der Residuen der transformierten Messwerte

Resultate bei Probe 953 generell stärker streuen. Es ist an den Organisatoren des Ringversuches zu untersuchen, ob Matrixeffekte vorliegen, die Probe inhomogen ist oder ob andere Gründe das unterschiedliche Verhalten erklären können.

Abschliessende Bemerkungen

- Laborvergleichsuntersuchungen verfolgen zwei verschiedene Zielrichtungen:
- Erhaltung und Verbesserung der Analysenqualität in den einzelnen Laboratorien.
 - Bereitstellung von Entscheidungsgrundlagen, um leistungsfähige Laboratorien von solchen mit ungenügendem Qualitätsstandard zu unterscheiden.

Mit dem absoluten Qualitätskriterium $\sqrt{msq_i}$ lässt sich ein analytischer Fortschritt nachweisen. Es liefert eine angemessene Bewertung der Laborqualität als die üblichen Kombinationsscores und die Gefahr einer Falschinterpretation ist geringer. Aus den transformierten Residuen kann wertvolle Zusatzinformation über das Probenmaterial gewonnen werden. Falls eine Varianzfunktion vom Horwitz-Typ vorliegt und somit eine varianzstabilisierende Transformation möglich ist, liefert die in dieser Arbeit vorgeschlagene Methode eine aussagekräftigere Laborbewertung als das AOAC/ISO/IUPAC-Protokoll (2).

Zusammenfassung

Horwitz (7) fand anhand der Ergebnisse von Ringversuchen zur Methodenprüfung eine einfache funktionale Beziehung zwischen der Vergleichstandardabweichung und dem Konzentrationsniveau. Dieselbe funktionale Beziehung wurde von Thompson (8) in einer Laborvergleichsuntersuchung bestätigt. In dieser Arbeit wird gezeigt, dass bei Vorhandensein einer Varianzfunktion vom Horwitz-Typ ein proben- und merkmalunabhängiger Laborvariationskoeffizient definiert werden kann, mit Hilfe dessen eine einheitliche Laborbewertung möglich wird. Aus den transformierten Daten lassen sich Zusatzinformationen über das Probenmaterial gewinnen.

Résumé

Horwitz (7) a établi une relation simple entre la déviation standard de reproductibilité et la concentration. Cette relation a été confirmée par Thompson (8) sur la base d'un essai interlaboratoire de compétence. A supposer qu'une telle relation existe, ce travail montre qu'il est possible de définir à l'aide d'une simple transformation de données un critère absolu, indépendant de la méthode d'analyse, de l'analyte et de l'échantillon, qui permet de caractériser la qualité des résultats fournis par un laboratoire. Les données transformées permettent d'acquérir des informations supplémentaires sur les matériaux.

Summary

The International Harmonized Protocol for the Proficiency Testing of (Chemical) Analytical Laboratories (2) lays down the minimum requirements for laboratories and other organisations that wish to develop and operate proficiency testing schemes, and recommends statistical treatment of the reported data. This statistical treatment is based on a scoring system, the *z*-scores. *Horwitz* (7) found that a simple relationship described the variation of reproducibility precision in collaborative trials as a fractional power of the concentration of the analyte. A similar relationship has been found in the results of several proficiency tests. If such relationships exist it is possible to transform the data and apply an alternative scoring system which is not only a relative but also an absolute criterion of the performance of the laboratories. The transformed data are examined to get further information about the test material.

Literatur

1. *Uhlig, S.*: Entwicklung und DV-mässige Implementierung eines Programmes zur Auswertung von analytischen Laborvergleichsuntersuchungen gemäss international harmonisierter Protokolle. BgVV Berlin (D), 95 S. (1995).
2. *Thompson, M. and Wood, R.*: International harmonized protocol for proficiency testing of (chemical) analytical laboratories. *J. of AOAC International* **76**, 926–940 (1993).
3. *Lischer, P.*: Sources, propagation and control of errors in soil monitoring. In: *Schulin, R. et al. (eds.)*, *Soil monitoring: Early detection and surveying of soil contamination and degradation*. 277–294. Birkhäuser, Basel, Boston, Berlin 1993.
4. *Lischer, P.*: Ringversuche zur Bestimmung des Qualitätsstandards von Laboratorien. Seminar der Region Österreich-Schweiz der Int. Biometrischen Gesellschaft, Innsbruck, 27. 9.–1. 10. 1993.
5. *Lischer, P.*: Robust statistical methods in interlaboratory analytical studies. *Lecture Notes in Statistics*, Springer-Verlag **109**, 251–264 (1996).
6. *Flury, B. and Riedwyl, H.*: *Angewandte Multivariate Statistik*. Gustav Fischer Verlag, Stuttgart, 1983.
7. *Horwitz, W.*: Evaluation of analytical methods used for regulation of foods and drugs. *Anal. Chem.* **54**, 67A–76A (1982).
8. *Thompson, M. and Lowthian, Ph. J.*: A Horwitz-like function describes precision in a proficiency test. *Analyst* **120**, 271–272 (1995).
9. *Hoaglin, D.C., Mosteller, F. and Tukey, J. W.*: *Understanding robust and exploratory data analysis*. Wiley, New York 1983.
10. *SLB*: Schweizerisches Lebensmittelbuch, Kapitel 60. Statistik und Ringversuche, Eidg. Drucksachen- und Materialzentrale, Bern 1989.
11. Verordnung über Schadstoffe im Boden (VSBo) vom 9. Juni 1986. SR 814.12 Bern.

Dr. Peter Lischer
Eidg. Forschungsanstalt für
Agrarökologie und Landbau
Institut für Umweltschutz
und Landwirtschaft
CH-3003 Bern