

Genauigkeit (Richtigkeit und Präzision) von Messverfahren
und MessergebnissenTeil 5: Alternative Methoden für die Ermittlung der Präzision eines
vereinheitlichten Messverfahrens
(ISO 5725-5:1998)

DIN

ISO 5725-5

ICS 17.020

Accuracy (trueness and precision) of measurement methods and results —
Part 5: Alternative methods for the determination of the precision of a
standard measurement method (ISO 5725-5:1998)

Exactitude (justesse et fidélité) des résultats et méthodes de mesure —
Partie 5: Méthodes alternatives pour la détermination de la fidélité d'une
méthode de mesure normalisée (ISO 5725-5:1998)

Die Internationale Norm ISO 5725-5:1998-07-15 „Accuracy (trueness and precision) of measurement methods and results — Part 5: Alternative methods for the determination of the precision of a standard measurement method“ ist unverändert in diese Deutsche Norm übernommen worden.

Nationales Vorwort

Die Internationale Norm ISO 5725 wurde vom Technischen Komitee 69 „Anwendung statistischer Methoden“ der Internationalen Organisation für Normung (ISO) erarbeitet. Die 2. Ausgabe 1986 wurde in der nachfolgenden 3. Ausgabe 1998 in sechs Teile unterteilt. In diesen sind die Teilaspekte zum Gesamtkomplex „Genauigkeit (Richtigkeit und Präzision) von Messverfahren und Messergebnissen“ einzeln dargestellt. Die deutschen Übersetzungen sind vom Normenausschuss Qualitätsmanagement, Statistik und Zertifizierungsgrundlagen (NQSZ) im DIN erstellt worden. Das Vorwort in jedem der sechs Teile gibt Auskunft über die jeweils behandelten Teilkomplexe.

Die Normen der Reihe DIN ISO 5725 befassen sich mit der Ermittlung und Anwendung von Präzisions- und Richtigkeitsmaßen eines vereinheitlichten Messverfahrens mittels Ringversuchen. Die Genauigkeit eines Messverfahrens wird dadurch beschrieben, dass sowohl ein Richtigkeits- als auch ein Präzisionsmaß gemeinsam angegeben werden. Nach DIN 55350-13 ist Präzision die qualitative Bezeichnung für das Ausmaß der gegenseitigen Annäherung voneinander unabhängiger Ermittlungsergebnisse bei mehrfacher Anwendung eines vereinheitlichten Messverfahrens unter vorgegebenen Bedingungen. Nach derselben Norm ist Richtigkeit die qualitative Bezeichnung für das Ausmaß der Annäherung des Erwartungswertes des Ermittlungsergebnisses an den Bezugswert, wobei dieser je nach Festlegung oder Vereinbarung der wahre oder der richtige Wert sein kann. Der Erwartungswert ist dabei als gedanklich feststellbarer Wert das mittlere Ergebnis bei unablässig wiederholter Anwendung des vereinheitlichten Messverfahrens.

Die Normen der Reihe DIN ISO 5725 sind nur auf kontinuierliche Merkmale anwendbar (siehe 1.2).

Als Maß für die Richtigkeit wird die Differenz zwischen einem Schätzwert für den Erwartungswert und dem gewählten Bezugswert benutzt. Die Differenz zwischen Erwartungswert und Bezugswert wird auch „bias“ genannt. Im Deutschen wird dieses Maß für die Richtigkeit als „systematische Abweichungskomponente“ bezeichnet, um sie von einzelnen systematischen Abweichungen zu unterscheiden.

Fortsetzung Seite 2 bis 123

Normenausschuss Qualitätsmanagement, Statistik und Zertifizierungsgrundlagen (NQSZ)
im DIN Deutsches Institut für Normung e.V.

Maße für die Präzision eines vereinheitlichten Messverfahrens sind im Allgemeinen die Standardabweichungen der unter Wiederholbedingungen, Zwischenbedingungen oder Vergleichbedingungen gewonnenen Messwerte. Aus dem Präzisionsmaß Standardabweichung können auch die früher häufig benutzten speziellen Präzisionsmaße Wiederholgrenze r und Vergleichgrenze R als „kritische Differenzbeträge“ berechnet werden. Diese sind im Vergleich mit den Standardabweichungen um einen angebbaren Faktor größere Beträge, unter denen oder höchstens gleich denen der Betrag der Differenz zweier Einzelergebnisse (gemessen unter den betreffenden Bedingungen) mit einer vorgegebenen Wahrscheinlichkeit P erwartet werden kann. Für diese Wahrscheinlichkeit wird bei Abwesenheit anderer Festlegungen oder Vereinbarungen $P = 95\%$ eingesetzt. Der genannte Faktor hängt von der gewählten oder vorgegebenen Wahrscheinlichkeit ab und hat bei $P = 95\%$ den Zahlenwert 2,8¹⁾. Dem Anwender der Norm ist es nach wie vor prinzipiell freigestellt, ob er als Präzisionsmaße eines vereinheitlichten Messverfahrens

- Wiederholstandardabweichung, Zwischenstandardabweichung und Vergleichstandardabweichung oder
- Wiederholgrenze, Zwischengrenze und Vergleichgrenze

angibt. Empfohlen wird, nur die im vorausgehenden ersten Anstrich genannten Größen zu verwenden.

Nicht jeder für ein betrachtetes vereinheitlichtes Messverfahren durchgeführte Ringversuch führt zu den gleichen Werten der Richtigkeitsmaße bzw. der Präzisionsmaße. Diese Ergebnisse sind nämlich Zufallsergebnisse. Überdies können die Anzahl der beteiligten Labors und die Anzahl der Ermittlungen im Labor unterschiedlich sein. Die ermittelten Werte der Richtigkeitsmaße bzw. der Präzisionsmaße sind also Schätzwerte. Für einen Industrie- oder Wirtschaftssektor oder für ein Fachgebiet können für sie jedoch verbindliche Grenzwerte eingeführt werden, falls es erforderlich ist. Das kann beispielsweise aus juristischen Gründen nötig sein.

In den Normen der Reihe DIN ISO 5725 werden außer den eingangs für Präzision und Richtigkeit genannten die folgenden Begriffe benutzt:

- Ringversuch = Versuch unter Beteiligung mehrerer Labors, z. B. zur Ermittlung der Präzision oder der Richtigkeit eines Messverfahrens nach dieser Normenreihe (siehe auch DIN ISO 5725-1:1997-11, 3.22).

ANMERKUNG Ein Ringversuch kann unterschiedliche Zwecke haben. Beispielsweise kann er nur auf Präzisionsmaße oder nur auf Richtigkeitsmaße oder auf beide ausgerichtet sein. Unterbegriffe zum Ringversuch für dessen unterschiedliche Ziele sind nicht erforderlich.

- Labor = Stelle, die Prüfungen oder nur Ermittlungen ausführt.

ANMERKUNG Gelegentlich wird dieser Begriff auch im Sinn des Untersuchungsorts verwendet.

- Vereinheitlichtes Messverfahren: Siehe DIN ISO 5725-1:1997-11, 4.1.

ANMERKUNG Bisher wurde es „festgelegtes Messverfahren“ genannt.

Im Übrigen wird auf die Normen über Begriffe zu Qualitätsmanagement und Statistik der Normenreihen DIN 55350 und DIN 13303 hingewiesen.

Die vorliegende Normenreihe DIN ISO 5725 ist eine Rahmenregelung für zahlreiche Fachgebiete. Sie bedarf bei der Anwendung gegebenenfalls ergänzender Festlegungen, insbesondere zur Anzahl der am Ringversuch beteiligten Labors und zur Anzahl der Merkmalsniveaus und der Wiederholmessungen. Für beides enthalten die Normen der Reihe Empfehlungen. Branchenspezifisch kann eine feste Anzahl oder Mindestanzahl von Labors, Merkmalsniveaus und von Wiederholmessungen je Merkmalsniveau vorgegeben werden.

Die vorliegende Norm behandelt unter Anwendung der in DIN ISO 5725-1 dargelegten, für alle Normen der Reihe DIN ISO 5725 geltenden allgemeinen Grundlagen und Begriffe die Ermittlung von Präzisionswerten aus voneinander unabhängigen Ermittlungsergebnissen an identischen Untersuchungsobjekten. In Ergänzung zu DIN ISO 5725-2 und DIN ISO 5725-3 stellt sie drei alternative Methoden der Ermittlung dieser Präzisionswerte mit den dazu nötigen Errechnungen vor:

Der erste Fall betrifft auf jedem betrachteten Merkmalsniveau eines Ringversuchs die Benutzung von jeweils zwei Untersuchungsobjekten mit geringfügig unterschiedlichen Werten bei jedem beteiligten Labor. Diese beiden Proben geben der alternativen Methode auch ihren Namen, indem von „Doppelniveauproben“

¹⁾ Dieser Faktor ergibt sich aus dem auf 2,0 gerundeten Quantil $u_{97,5\%} = 1,96$ sowie aus $\sqrt{n}$ mit $n = 2$. Aus deren Multiplikation folgt, für Zwecke der Schätzung hinreichend genau, der Faktor 2,8.

gesprochen wird, während sonst vielfach „Doppelproben auf einheitlichem Merkmalsniveau“ angewendet werden, die man kurz auch „Einheitliche Doppelproben“ nennt. Zweck der Doppelniveauproben ist es, die Verlässlichkeit der Werteangaben durch die Bearbeiter zu erhöhen, weil sie nicht wissen, wie groß die Unterschiede sind.

Der zweite Fall behandelt heterogenes Untersuchungsmaterial wie Sand und Leder.

Schließlich werden robuste Auswertungsmethoden vorgestellt, mit denen die Ergebnisse von Ringversuchen im Hinblick auf Präzisionswerte ohne die für das Ergebnis oft sehr einflussreichen Ermessensentscheidungen zu Statistischen Tests bezüglich Ausreißer und Fastausreißer abgewickelt werden können. Für die Errechnung der Präzisionswerte werden iterative Algorithmen angewendet.

Die drei alternativen Methoden werden anhand von Beispielen ausführlich erläutert.

Die Norm DIN ISO 5725 wurde in DIN ISO 5725-1 bis DIN ISO 5725-6 aufgeteilt und erweitert. Diese Normen enthalten jetzt außer der bisher behandelten Ermittlung von Präzisionsmaßen auch

- die Ermittlung von Richtigkeitsmaßen;
- die Berücksichtigung von Zwischenbedingungen;
- den Vergleich von Labors;
- die Verwendung von Genauigkeitsmaßen in der Praxis;
- alternative Verfahren für die Ermittlung von Präzisionsmaßen.

Nationaler Anhang NA (informativ)

Literaturhinweise

DIN 13303-1:1982-05, *Stochastik — Wahrscheinlichkeitstheorie, Gemeinsame Grundbegriffe der mathematischen und der beschreibenden Statistik — Begriffe und Zeichen.*

DIN 13303-2:1982-11, *Stochastik — Mathematische Statistik — Begriffe und Zeichen.*

DIN 55350-12:1989-03, *Begriffe der Qualitätssicherung und Statistik — Merkmalsbezogene Begriffe.*

DIN 55350-13:1987-07, *Begriffe der Qualitätssicherung und Statistik — Begriffe zur Genauigkeit von Ermittlungsverfahren und Ermittlungsergebnissen.*

DIN 55350-14:1985-12, *Begriffe der Qualitätssicherung und Statistik — Begriffe der Probenahme.*

DIN 55350-15:1986-02, *Begriffe der Qualitätssicherung und Statistik — Begriffe zu Mustern.*

DIN 55350-17:1988-08, *Begriffe der Qualitätssicherung und Statistik — Begriffe der Qualitätsprüfungsarten.*

DIN 55350-21:1982-05, *Begriffe der Qualitätssicherung und Statistik — Begriffe der Statistik — Zufallsgrößen und Wahrscheinlichkeitsverteilungen.*

DIN 55350-22:1987-02, *Begriffe der Qualitätssicherung und Statistik — Begriffe der Statistik — Spezielle Wahrscheinlichkeitsverteilungen.*

DIN 55350-23:1983-04, *Begriffe der Qualitätssicherung und Statistik — Begriffe der Statistik — Beschreibende Statistik.*

DIN 55350-24:1982-11, *Begriffe der Qualitätssicherung und Statistik — Begriffe der Statistik — Schließende Statistik.*

DIN EN ISO 9000:2000-12, *Qualitätsmanagementsysteme – Grundlagen und Begriffe.*

Deutsche Übersetzung

Genauigkeit (Richtigkeit und Präzision) von Messverfahren und Messergebnissen –

Teil 5: Alternative Methoden für die Ermittlung der Präzision eines vereinheitlichten Messverfahrens

Inhalt

	Seite
1 Anwendungsbereich	10
2 Normative Verweisungen	10
3 Begriffe	12
4 Ringversuch mit Doppelniveauuben	12
4.1 Anwendungshinweise zum Ringversuch mit Doppelniveauuben	12
4.2 Anordnung eines Ringversuchs mit Doppelniveauuben	12
4.3 Gestaltung eines Ringversuchs mit Doppelniveauuben	14
4.4 Statistisches Modell	16
4.5 Statistische Auswertung der Ermittlungsergebnisse eines Ringversuchs mit Doppelniveauuben	18
4.6 Prüfung der Ermittlungsergebnisse auf Vereinbarkeit und Ausreißer	20
4.7 Ergebnisbericht zum Ringversuch mit Doppelniveauuben	22
4.8 Beispiel 1: Ein Ringversuch mit Doppelniveauuben – Ermittlung von Protein	22
5 Ringversuch an heterogenem Material	34
5.1 Anwendungshinweise zum Ringversuch an heterogenem Material	34
5.2 Anordnung eines Ringversuchs an heterogenem Material	36
5.3 Gestaltung eines Ringversuchs an heterogenem Material	38
5.4 Statistisches Modell für einen Ringversuch an heterogenem Material	40
5.5 Statistische Auswertung der Ermittlungsergebnisse eines Ringversuchs an heterogenem Material	42
5.6 Prüfung der Ermittlungsergebnisse auf Vereinbarkeit und Ausreißer	48
5.7 Ergebnisbericht zum Ringversuch an heterogenem Material	50
5.8 Beispiel 2: Ein Ringversuch an heterogenem Material	50
5.9 Allgemeine Formeln für Errechnungen bei einem Ringversuch an heterogenem Material	66
5.10 Beispiel 3: Eine Anwendung der allgemeinen Formeln	68
6 Robuste Auswertungsmethoden	74
6.1 Anwendungshinweise zu robusten Auswertungsmethoden	74
6.2 Robuste Auswertungsmethode: Algorithmus A	78
6.3 Robuste Auswertungsmethode: Algorithmus S	82
6.4 Formeln: Robuste Auswertungsmethode für ein einzelnes Merkmalsniveau bei einem Ringversuch mit einheitlichen Doppelproben	84
6.5 Beispiel 4: Robuste Auswertungsmethode für ein einzelnes Merkmalsniveau bei einem Ringversuch mit einheitlichen Doppelproben	86
6.6 Formeln: Robuste Auswertungsmethode für ein einzelnes Merkmalsniveau bei einem Ringversuch mit Doppelniveauuben	94
6.7 Beispiel 5: Robuste Auswertungsmethode für ein einzelnes Merkmalsniveau bei einem Ringversuch mit Doppelniveauuben	94
6.8 Formeln: Robuste Auswertungsmethode für ein einzelnes Merkmalsniveau bei einem Ringversuch an heterogenem Material	100
6.9 Beispiel 6: Robuste Auswertungsmethode für ein einzelnes Merkmalsniveau bei einem Ringversuch an heterogenem Material	100
Anhänge	
Anhang A (normativ) Formelzeichen und Abkürzungen in ISO 5725	110
Anhang B (informativ) Ableitung der in den Algorithmen A und S angewendeten Faktoren	116
Anhang C (informativ) Ableitung der für robuste Auswertung angewendeten Gleichungen	120
Anhang D (informativ) Literaturhinweise	122



Documents

Document Sharing Hub

Contents

	Page
1 Scope	11
2 Normative references	11
3 Definitions	13
4 Split-level design	13
4.1 Applications of the split-level design	13
4.2 Layout of the split-level design	13
4.3 Organization of a split-level experiment	15
4.4 Statistical model	17
4.5 Statistical analysis of the data from a split-level experiment	19
4.6 Scrutiny of the data for consistency and outliers	21
4.7 Reporting the results of a split-level experiment	23
4.8 Example 1: A split-level experiment — Determination of protein	23
5 A design for a heterogeneous material	35
5.1 Applications of the design for a heterogeneous material	35
5.2 Layout of the design for a heterogeneous material	37
5.3 Organization of an experiment with a heterogeneous material	39
5.4 Statistical model for an experiment with a heterogeneous material	41
5.5 Statistical analysis of the data from an experiment with a heterogeneous material	43
5.6 Scrutiny of the data for consistency and outliers	49
5.7 Reporting the results of an experiment on a heterogeneous material	51
5.8 Example 2: An experiment on a heterogeneous material	51
5.9 General formulae for calculations with the design for a heterogeneous material	67
5.10 Example 3: An application of the general formulae	69
6 Robust methods for data analysis	75
6.1 Applications of robust methods of data analysis	75
6.2 Robust analysis: Algorithm A	79
6.3 Robust analysis: Algorithm S	81
6.4 Formulae: Robust analysis for a particular level of a uniform-level design	85
6.5 Example 4: Robust analysis for a particular level of a uniform-level design	85
6.6 Formulae: Robust analysis for a particular level of a split-level design	95
6.7 Example 5: Robust analysis for a particular level of a split-level design	95
6.8 Formulae: Robust analysis for a particular level of an experiment on a heterogeneous material	101
6.9 Example 6: Robust analysis for a particular level of an experiment on a heterogeneous material	101
Annexes	
A (normative) Symbols and abbreviations used in ISO 5725	111
B (informative) Derivation of the factors used in algorithms A and S	117
C (informative) Derivation of equations used for robust analysis	121
D (informative) Bibliography	123

Vorwort

Die ISO (Internationale Organisation für Normung) ist die weltweite Vereinigung nationaler Normungsinstitute (ISO-Mitglieds Körperschaften). Die Erarbeitung Internationaler Normen obliegt den Technischen Komitees der ISO. Jede Mitglieds Körperschaft, die sich für ein Thema interessiert, für das ein Technisches Komitee eingesetzt wurde, ist berechtigt, in diesem Komitee mitzuarbeiten. Internationale (staatliche und nichtstaatliche) Organisationen, die mit der ISO in Verbindung stehen, sind an den Arbeiten ebenfalls beteiligt. Die ISO arbeitet in allen Angelegenheiten der elektrotechnischen Normung eng mit der Internationalen Elektrotechnischen Kommission (IEC) zusammen.

Die von den Technischen Komitees verabschiedeten internationalen Norm-Entwürfe werden den Mitglieds Körperschaften zur Abstimmung vorgelegt. Die Veröffentlichung als Internationale Norm erfordert Zustimmung von mindestens 75 % der abstimmenden Mitglieds Körperschaften.

Die Internationale Norm ISO 5725-5 wurde durch das Technische Komitee ISO/TC 69, *Anwendung statistischer Methoden*, Unterkomitee SC 6, *Messmethoden und Messergebnisse*, erarbeitet.

ISO 5725 besteht unter dem Haupttitel „*Genauigkeit (Richtigkeit und Präzision) von Messverfahren und Messergebnissen*“ aus den folgenden Teilen:

- Teil 1: Allgemeine Grundlagen und Begriffe
- Teil 2: Grundlegende Methode für die Ermittlung der Wiederhol- und Vergleichpräzision eines vereinheitlichten Messverfahrens
- Teil 3: Präzisionsmaße eines vereinheitlichten Messverfahrens unter Zwischenbedingungen
- Teil 4: Grundlegende Methoden für die Ermittlung der Richtigkeit eines vereinheitlichten Messverfahrens
- Teil 5: Alternative Methoden für die Ermittlung der Präzision eines vereinheitlichten Messverfahrens
- Teil 6: Anwendung von Genauigkeitswerten in der Praxis

Die Teile 1 bis 6 von ISO 5725 ersetzen gemeinsam die Norm ISO 5725:1986 und machen diese ungültig, welche erweitert wurde, um Richtigkeit (ergänzend zu Präzision) sowie Zwischenbedingungen (ergänzend zu Wiederhol- und Vergleichbedingungen) abzudecken.

Der Anhang A ist normativer Bestandteil dieses Teils von ISO 5725. Die Anhänge B bis D sind informativ.

Foreword

ISO (the International Organization for Standardization) is a world-wide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organisations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International standard requires approval by at least 75 % of the member bodies casting a vote.

ISO 5725-5 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

ISO 5725 consists of the following parts, under the general title *Accuracy (trueness and precision) of measurement methods and results*:

- *Part 1: General principles and definitions*
- *Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*
- *Part 3: Intermediate measures of the precision of a standard measurement method*
- *Part 4: Basic methods for the determination of the trueness of a standard measurement method*
- *Part 5: Alternative methods for the determination of the precision of a standard measurement method*
- *Part 6: Use in practice of accuracy values*

Parts 1 to 6 of ISO 5725 together cancel and replace ISO 5725:1986, which has been extended to cover trueness (in addition to precision) and intermediate precision conditions (in addition to repeatability conditions and reproducibility conditions).

Annex A forms an integral part of this part of ISO 5725. Annexes B, C and D are for information only.

Einleitung

0.1 ISO 5725 benutzt die beiden Ausdrücke „Richtigkeit“ und „Präzision“, um die Genauigkeit eines Messverfahrens zu beschreiben. „Richtigkeit“ bezieht sich auf das Ausmaß der Annäherung zwischen dem arithmetischen Mittelwert einer großen Anzahl von Ermittlungsergebnissen und dem wahren oder dem richtigen Wert. „Präzision“ bezieht sich auf das Ausmaß der gegenseitigen Annäherung von Ermittlungsergebnissen.

0.2 Allgemeine Überlegungen zu diesen Größen sind in ISO 5725-1 angegeben. Sie wurden deshalb hier nicht wiederholt. Dieser Teil von ISO 5725 sollte zusammen mit ISO 5725-1 gelesen werden, weil dort die zugrunde liegenden Definitionen und allgemeinen Prinzipien angegeben sind.

0.3 ISO 5725-2 betrifft die Schätzung von vereinheitlichten Maßen für die Präzision anhand von Ringversuchen, speziell für die Wiederholstandardabweichung und für die Vergleichstandardabweichung. ISO 5725-2 vermittelt dabei ein grundlegendes Verfahren für einen Ringversuch, das mit Einzelproben auf einheitlichem Merkmalsniveau arbeitet. Dieser Teil von ISO 5725 beschreibt alternative Methoden zu diesem grundlegende Verfahren.

- a) Bei Anwendung des grundlegenden Verfahrens besteht ein Risiko, dass ein Bearbeiter die Beeinflussung eines nachfolgenden Ermittlungsergebnisses an einer anderen Probe des identischen Materials durch das vorausgehende Ermittlungsergebnis hinnehmen könnte. Das würde zur Folge haben, dass die Schätzwerte für die Wiederholstandardabweichung und die Vergleichstandardabweichung eine systematische Abweichungskomponente enthalten. Wenn dieses Risiko als schwerwiegend betrachtet wird, kann der in diesem Teil von ISO 5725 beschriebene Ringversuch mit Doppelniveauproben bevorzugt werden, weil er dieses Risiko verkleinert.
- b) Das grundlegende Verfahren verlangt für die Anwendung im Ringversuch die Bereitstellung einer Anzahl identischer Materialproben. Mit heterogenen Materialien kann das unmöglich werden. Die Anwendung des grundlegenden Verfahrens hätte dann Schätzwerte für die Vergleichstandardabweichung zur Folge, die infolge der Schwankung (der tatsächlichen Werte) zwischen den Proben vergrößert wären. Ein Ringversuch an heterogenem Material, wie er in diesem Teil von ISO 5725 beschrieben ist, liefert Information über die Schwankung zwischen den Proben, wie sie durch das grundlegende Verfahren nicht gewonnen werden kann. Diese Methode kann angewendet werden, um Schätzwerte der Vergleichstandardabweichung zu errechnen, aus denen die Schwankung zwischen den Proben abgetrennt wurde.
- c) Das grundlegende Verfahren verlangt Ausreißertests. Sie müssen angewendet werden, um Ermittlungsergebnisse zu entdecken, die aus der Errechnung der Wiederholstandardabweichung und der Vergleichstandardabweichung ausgeschlossen werden sollten. Dieses Ausschließen von Ausreißern kann manchmal eine große Wirkung auf die Schätzwerte der Wiederholstandardabweichung und der Vergleichstandardabweichung haben. Allerdings muss in der Praxis bei Anwendung der Ausreißertests derjenige, der die Ermittlungsergebnisse analysiert, ggf. eine Ermessensentscheidung treffen, welche Ermittlungsergebnisse auszuschließen sind. Dieser Teil von ISO 5725 beschreibt robuste Auswertungsmethoden, mit denen Wiederholstandardabweichung und Vergleichstandardabweichung ohne Anwendung der Ausreißertests zum ggf. erforderlichen Ausschluss von Ermittlungsergebnissen errechnet werden können. Damit werden die Schätzergebnisse unabhängig von den Beurteilungen derjenigen, welche die Ermittlungsergebnisse analysieren.

Introduction

0.1 This part of ISO 5725 uses two terms *trueness* and *precision* to describe the accuracy of a measurement method. *Trueness* refers to the closeness of agreement between the average value of a large number of test results and the true or accepted reference value. *Precision* refers to the closeness of agreement between test results.

0.2 General consideration of these quantities is given in ISO 5725-1 and so is not repeated here. This part of ISO 5725 should be read in conjunction with ISO 5725-1 because the underlying definitions and general principles are given there.

0.3 ISO 5725-2 is concerned with estimating, by means of interlaboratory experiments, standard measures of precision, namely the repeatability standard deviation and the reproducibility standard deviation. It gives a basic method for doing this using the uniform-level design. This part of ISO 5725 describes alternative methods to this basic method.

- a) With the basic method there is a risk that an operator may allow the result of a measurement on one sample to influence the result of a subsequent measurement on another sample of the same material, causing the estimates of the repeatability and reproducibility standard deviations to be biased. When this risk is considered to be serious, the split-level design described in this part of ISO 5725 may be preferred as it reduces this risk.
- b) The basic method requires the preparation of a number of identical samples of the material for use in the experiment. With heterogeneous materials this may not be possible, so that the use of the basic method then gives estimates of the reproducibility standard deviation that are inflated by the variation between the samples. The design for a heterogeneous material given in this part of ISO 5725 yields information about the variability between samples which is not obtainable from the basic method; it may be used to calculate an estimate of reproducibility from which the between-sample variation has been removed.
- c) The basic method requires tests for outliers to be used to identify data that should be excluded from the calculation of the repeatability and reproducibility standard deviations. Excluding outliers can sometimes have a large effect on the estimates of repeatability and reproducibility standard deviations, but in practice, when applying the outlier tests, the data analyst may have to use judgement to decide which data to exclude. This part of ISO 5725 describes robust methods of data analysis that may be used to calculate repeatability and reproducibility standard deviations from data containing outliers without using tests for outliers to exclude data, so that the results are no longer affected by the data analyst's judgement.

1 Anwendungsbereich

Dieser Teil von ISO 5725

- stellt detaillierte Beschreibungen von Alternativen zum grundlegende Verfahren zur Ermittlung der Wiederholstandardabweichung und der Vergleichstandardabweichung bei der Anwendung eines vereinheitlichten Messverfahrens zur Verfügung, vor allem den Ringversuch mit Doppelniveauproben und einen Ringversuch an heterogenem Material;
- beschreibt die Anwendung robuster Methoden für die Auswertung der Ermittlungsergebnisse von Ringversuchen ohne die Anwendung von Ausreißertests zur gegebenenfalls nötigen Ausschließung von Ermittlungsergebnissen aus den Errechnungen, und insbesondere ausführlich die Anwendung einer solchen Methode.

Dieser Teil von ISO 5725 ergänzt ISO 5725-2 durch Bereitstellung alternativer Ringversuchs-Pläne, die in speziellen Situationen wertvoller sein können als das in ISO 5725-2 beschriebene grundlegende Verfahren, sowie durch Bereitstellung einer robusten Auswertungsmethode, die Schätzwerte für die Wiederholstandardabweichung und die Vergleichstandardabweichung liefert, die weniger von der Entscheidung desjenigen abhängen, der die Ermittlungsergebnisse beurteilt, als bei den in ISO 5725-2 beschriebenen Methoden.

2 Normative Verweisungen

Die folgenden normativen Dokumente enthalten Festlegungen, die durch Verweisung in diesem Text Bestandteil dieses Teils von ISO 5725 sind. Zum Zeitpunkt der Veröffentlichung dieser Internationalen Norm waren die angegebenen Ausgaben gültig. Alle Normen unterliegen der Überarbeitung. Vertragspartner, deren Vereinbarungen auf diesem Teil von ISO 5725 basieren, werden gebeten, die Möglichkeit zu prüfen, ob die jeweils neuesten Ausgaben der im Folgenden genannten Normen angewendet werden können. Die Mitglieder von IEC und ISO führen Verzeichnisse der gegenwärtig gültigen Internationalen Normen.

ISO 3534-1:1993, *Statistik — Begriffe und Formelzeichen — Teil 1: Wahrscheinlichkeit und allgemeine statistische Begriffe.*

ISO 3534-3:1985, *Statistik — Begriffe und Formelzeichen — Teil 3: Versuchsplanung.*

ISO 5725-1:1994, *Genauigkeit (Richtigkeit und Präzision) von Messverfahren und Messergebnissen — Teil 1: Allgemeine Grundlagen und Begriffe.*

ISO 5725-2:1994, *Genauigkeit (Richtigkeit und Präzision) von Messverfahren und Messergebnissen — Teil 2: Grundlegende Methode für die Ermittlung der Wiederhol- und Vergleichpräzision eines vereinheitlichten Messverfahrens.*

Accuracy (trueness and precision) of measurement methods and results —

Part 5:

Alternative methods for the determination of the precision of a standard measurement method

1 Scope

This part of ISO 5725

- provides detailed descriptions of alternatives to the basic method for determining the repeatability and reproducibility standard deviations of a standard measurement method, namely the split-level design and a design for heterogeneous materials;
- describes the use of robust methods for analysing the results of precision experiments without using outlier tests to exclude data from the calculations, and in particular, the detailed use of one such method.

This part of ISO 5725 complements ISO 5725-2 by providing alternative designs that may be of more value in some situations than the basic design given in ISO 5725-2, and by providing a robust method of analysis that gives estimates of the repeatability and reproducibility standard deviations that are less dependent on the data analyst's judgement than those given by the methods described in ISO 5725-2.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this part of ISO 5725. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO 5725 are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 3534-1:1993, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms.*

ISO 3534-3:1985, *Statistics — Vocabulary and symbols — Part 3: Design of experiments.*

ISO 5725-1:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions.*

ISO 5725-2:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.*

3 Begriffe

Für die Anwendung dieses Teils von ISO 5725 gelten die in ISO 3534-1 und ISO 5725-1 angegebenen Begriffe. Die in (den Normen der Reihe) ISO 5725 angewendeten Formelzeichen sind im Anhang A angegeben.

4 Ringversuch mit Doppelniveauproben

4.1 Anwendungshinweise zum Ringversuch mit Doppelniveauproben

4.1.1 Der Ringversuch mit Proben auf einheitlichem Merkmalsniveau, wie er in ISO 5725-2 beschrieben ist, verlangt zwei oder mehr identische Proben des Untersuchungsmaterials in jedem beteiligten Labor und auf jedem Merkmalsniveau des Versuchs. Bei diesem Modell besteht ein Risiko, dass ein Bearbeiter die Beeinflussung eines nachfolgenden Ermittlungsergebnisses an einer anderen Probe des identischen Materials durch das vorausgehende Ermittlungsergebnis hinnehmen könnte. Wenn dies vorkommt, werden die Ergebnisse des Ringversuchs verfälscht: Die Schätzwerte für die Wiederholstandardabweichung σ_r werden verkleinert und die Schätzwerte für die Vergleichstandardabweichung σ_L werden vergrößert sein. Im Ringversuch mit Doppelniveauproben erhält jedes teilnehmende Labor für jedes Merkmalsniveau je eine Probe von zwei ähnlichen Materialien. Den Bearbeitern wird gesagt, dass die Proben nicht identisch sind, aber es wird ihnen nicht gesagt, in welchem Ausmaß sich die Materialien unterscheiden. Der Ringversuch mit Doppelniveauproben stellt demnach eine Methode für die Ermittlung der Wiederholstandardabweichung und der Vergleichstandardabweichung eines vereinheitlichten Messverfahrens bereit, welche das Risiko vermindert, dass beim Ringversuch ein an der einen Probe gewonnenes Ermittlungsergebnis ein Ermittlungsergebnis an einer anderen Probe des Ringversuchs beeinflussen wird.

4.1.2 Die für ein Merkmalsniveau gewonnenen Ermittlungsergebnisse eines Ringversuchs mit Doppelniveauproben können zur Erstellung einer grafischen Darstellung benutzt werden, in der die Ermittlungsergebnisse für das eine Material gegen die des anderen, ähnlichen Materials aufgezeichnet werden. Ein Beispiel ist im Bild 1 gezeigt. Solche grafischen Darstellungen können beim Erkennen solcher Labors hilfreich sein, welche die größten systematischen Abweichungen gegenüber den anderen Labors haben. Es ist nützlich, wenn es möglich ist, die Ursachen für die größten systematischen Abweichungen eines Labors herauszufinden, um Korrekturmaßnahmen zu ergreifen.

4.1.3 Häufig hängen die Wiederholstandardabweichung und die Vergleichstandardabweichung eines (vereinheitlichten) Messverfahrens vom Merkmalsniveau des Materials ab. Ist das Ermittlungsergebnis beispielsweise der Anteil eines Elements, der durch eine chemische Analyse gewonnen wird, steigen Wiederholstandardabweichung und Vergleichstandardabweichung üblicherweise ebenso an wie das Merkmalsniveau des Anteils des Elements ansteigt. Es ist für einen Ringversuch mit Doppelniveauproben erforderlich, dass die für ein Merkmalsniveau des Ringversuchs benutzten beiden ähnlichen Materialien einander so ähnlich sind, dass für sie die gleiche Wiederholstandardabweichung und die gleiche Vergleichstandardabweichung erwartet werden kann. Für die Zwecke des Ringversuchs mit Doppelniveauproben ist es annehmbar, wenn die für ein Merkmalsniveau des Ringversuchs benutzten beiden Materialien im wesentlichen das gleiche Niveau von Ermittlungsergebnissen liefern. Es ist nämlich nichts damit gewonnen, wenn man es so einrichtet, dass sie sich wesentlich unterscheiden.

Bei vielen chemischen Analyseverfahren kann die Matrix, welche den interessierenden Bestandteil enthält, die Präzision beeinflussen. Deshalb sind für einen Ringversuch mit Doppelniveauproben zwei Materialien mit ähnlichen Matrizes auf jedem Merkmalsniveau des Ringversuchs erforderlich. Ein hinreichend ähnliches Material kann vielfach dadurch hergestellt werden, dass man eine kleine Menge des interessierenden Materials zugibt. Wenn das Material ein Naturprodukt oder ein hergestelltes Produkt ist, kann es schwierig sein, zwei Produkte zu finden, die für einen Ringversuch mit Doppelniveauproben hinreichend ähnlich sind: Eine mögliche Lösung kann die Verwendung von zwei Losen des gleichen Produkts sein. Es sollte stets daran gedacht werden: Das Ziel der Materialauswahl für einen Ringversuch mit Doppelniveauproben ist die Bereitstellung von Proben, von denen die Bearbeiter nicht erwarten, dass sie identisch sind.

4.2 Anordnung eines Ringversuchs mit Doppelniveauproben

4.2.1 Die Anordnung eines Ringversuchs mit Doppelniveauproben ist in Tabelle 1 gezeigt.

Die p teilnehmenden Labors untersuchen jeweils zwei Proben auf q Merkmalsniveaus.

Die beiden Proben innerhalb eines Merkmalsniveaus werden mit a und b bezeichnet. Dabei verkörpert a eine Probe des einen Materials, und b verkörpert eine Probe des anderen, ähnlichen Materials.

3 Definitions

For the purposes of this part of ISO 5725, the definitions given in ISO 3534-1 and in ISO 5725-1 apply.

The symbols used in ISO 5725 are given in annex A.

4 Split-level design

4.1 Applications of the split-level design

4.1.1 The uniform level design described in ISO 5725-2 requires two or more identical samples of a material to be tested in each participating laboratory and at each level of the experiment. With this design there is a risk that an operator may allow the result of a measurement on one sample to influence the result of a subsequent measurement on another sample of the same material. If this happens, the results of the precision experiment will be distorted: estimates of the repeatability standard deviation σ_r will be decreased and estimates of the between-laboratory standard deviation σ_L will be increased. In the split-level design, each participating laboratory is provided with a sample of each of two similar materials, at each level of the experiment, and the operators are told that the samples are not identical, but they are not told by how much the materials differ. The split-level design thus provides a method of determining the repeatability and reproducibility standard deviations of a standard measurement method in a way that reduces the risk that a test result obtained on one sample will influence a test result on another sample in the experiment.

4.1.2 The data obtained at a level of a split-level experiment may be used to draw a graph in which the data for one material are plotted against the data for the other, similar, material. An example is given in figure 1. Such graphs can help identify those laboratories that have the largest biases relative to the other laboratories. This is useful when it is possible to investigate the causes of the largest laboratory biases with the aim of taking corrective action.

4.1.3 It is common for the repeatability and reproducibility standard deviations of a measurement method to depend on the level of the material. For example, when the test result is the proportion of an element obtained by chemical analysis, the repeatability and reproducibility standard deviations usually increase as the proportion of the element increases. It is necessary, for a split-level experiment, that the two similar materials used at a level of the experiment are so similar that they can be expected to give the same repeatability and reproducibility standard deviations. For the purposes of the split-level design, it is acceptable if the two materials used for a level of the experiment give almost the same level of measurement results, and nothing is to be gained by arranging that they differ substantially.

In many chemical analysis methods, the matrix containing the constituent of interest can influence the precision, so for a split-level experiment two materials with similar matrices are required at each level of the experiment. A sufficiently similar material can sometimes be prepared by spiking a material with a small addition of the constituent of interest. When the material is a natural or manufactured product, it can be difficult to find two products that are sufficiently similar for the purposes of a split-level experiment: a possible solution may be to use two batches of the same product. It should be remembered that the object of choosing the materials for the split-level design is to provide the operators with samples that they do not expect to be identical.

4.2 Layout of the split-level design

4.2.1 The layout of the split-level design is shown in table 1.

The p participating laboratories each test two samples at q levels.

The two samples within a level are denoted a and b , where a represents a sample of one material, and b represents a sample of the other, similar, material.

4.2.2 Die Ermittlungsergebnisse des Ringversuchs mit Doppelniveauproben werden dargestellt durch:

$$y_{ijk} ,$$

wobei

der Index i die Labors kennzeichnet ($i = 1, 2, \dots, p$);

der Index j das Merkmalsniveau kennzeichnet ($j = 1, 2, \dots, q$);

der Index k die Probe kennzeichnet ($k = a$ oder b).

4.3 Gestaltung eines Ringversuchs mit Doppelniveauproben

4.3.1 Wenn ein Ringversuch mit Doppelniveauproben geplant wird, ist die Anleitung gemäß Abschnitt 6 von ISO 5725-1:1994 zu befolgen.

6.3 von ISO 5725-1:1994 enthält eine Anzahl von Formeln (die eine mit A bezeichnete Größe enthalten), die als Hilfe bei der Entscheidung anzuwenden sind, wie viele Labors in den Ringversuch einzubeziehen sind. Die entsprechenden Formeln für einen Ringversuch mit Doppelniveauproben sind unten angegeben.

ANMERKUNG Diese Formeln sind entwickelt worden gemäß der in ANMERKUNG 24 von ISO 5725-1:1994 beschriebenen Methode.

Zur Bewertung der Unsicherheiten der Wiederholstandardabweichung und der Vergleichstandardabweichung sind die folgenden Größen zu errechnen.

Für die Wiederholpräzision

$$A_r = 1,96 \sqrt{1/[2(p-1)]} \quad (1)$$

Für die Vergleichpräzision

$$A_R = 1,96 \sqrt{\{[1 + 2(\gamma^2 - 1)]^2 + 1\} / [8\gamma^4(p-1)]} \quad (2)$$

mit $\gamma = \sigma_R / \sigma_r$.

Wenn man die Anzahl n von Wiederholproben in den Gleichungen (9) und (10) von ISO 5725-1:1994 mit zwei ansetzt, dann kann man erkennen, dass die Gleichungen (9) und (10) von ISO 5725-1:1994 dieselben sind wie die Gleichungen (1) und (2) oben, nur dass hier manchmal $p-1$ anstelle von p in ISO 5725-1:1994 erscheint. Dies ist ein geringfügiger Unterschied, so dass die Tabelle 1 sowie die Bilder B.1 und B.2 von ISO 5725-1:1994 benutzt werden können, um die Unsicherheit der Schätzwerte für die Wiederholstandardabweichung und die Vergleichstandardabweichung in einem Ringversuch mit Doppelniveauproben zu bewerten.

Zur Bewertung der Unsicherheit des Schätzwertes der systematischen Abweichungskomponente in einem Ringversuch mit Doppelniveauproben ist die Größe A zu errechnen, wie sie durch die Gleichung (13) von ISO 5725-1:1994 mit $n = 2$ festgelegt ist (oder man benutze die Tabelle 2 von ISO 5725-1:1994), und man wende diese Größe A an, wie es in ISO 5725-1 beschrieben ist.

Zur Bewertung der Unsicherheit des Schätzwertes der systematischen Abweichungskomponente eines Labors in einem Ringversuch mit Doppelniveauproben errechne man die Größe A_w , die durch Gleichung (16) von ISO 5725-1:1994 mit $n = 2$ festgelegt ist. Weil die Anzahl der Wiederholproben in einem Ringversuch mit Doppelniveauproben tatsächlich diese Anzahl zwei ist, kann man die Unsicherheit des Schätzwertes für die systematische Abweichungskomponente eines Labors nicht durch eine Erhöhung der Anzahl der Wiederholproben verkleinern. (Wenn es nötig ist, diese Unsicherheit zu verkleinern, sollte statt dessen der Ringversuch mit Proben auf einheitlichem Merkmalsniveau angewendet werden).

4.3.2 Bezüglich der Einzelheiten der Gestaltung eines Ringversuchs mit Doppelniveauproben folge man der Anleitung gemäß den Abschnitten 5 und 6 von ISO 5725-2:1994. Als Anzahl n der Wiederholproben ist bei diesem Rückgriff auf ISO 5725-2 diejenige Anzahl zu wählen, die gleich dem Umfang einer Doppelniveauprobe bei einem Ringversuch mit Doppelniveauproben sein muss, nämlich zwei.

Die a -Proben sollten an die teilnehmenden Labors nach dem Zufallsprinzip verteilt werden, und die b -Proben sollten an die teilnehmenden Labors ebenfalls nach dem Zufallsprinzip, jedoch in einer gesonderten Zufallsoperation verteilt werden.

4.2.2 The data from a split-level experiment are represented by:

y_{ijk}

where

subscript i represents the laboratory ($i = 1, 2, \dots, p$);

subscript j represents the level ($j = 1, 2, \dots, q$);

subscript k represents the sample ($k = a$ or b).

4.3 Organization of a split-level experiment

4.3.1 Follow the guidance given in clause 6 of ISO 5725-1:1994 when planning a split-level experiment.

Subclause 6.3 of ISO 5725-1:1994 contains a number of formulae (involving a quantity denoted generally by A) that are used to help decide how many laboratories to include in the experiment. The corresponding formulae for the split-level experiment are set out below.

NOTE — These formulae have been derived by the method described in NOTE 24 of ISO 5725-1:1994.

To assess the uncertainties of the estimates of the repeatability and reproducibility standard deviations, calculate the following quantities.

For repeatability

$$A_r = 196\sqrt{1/[2(p-1)]} \quad (1)$$

For reproducibility

$$A_R = 196\sqrt{\left\{ \left[1 + 2(\gamma^2 - 1) \right]^2 + 1 \right\} / [8\gamma^4(p-1)]} \quad (2)$$

with $\gamma = \sigma_R/\sigma_r$.

If the number n of replicates is taken as two in equations (9) and (10) of ISO 5725-1:1994, then it can be seen that equations (9) and (10) of ISO 5725-1:1994 are the same as equations (1) and (2) above, except that sometimes $p-1$ appears here in place of p in ISO 5725-1:1994. This is a small difference, so table 1 and figures B.1 and B.2 of ISO 5725-1:1994 may be used to assess the uncertainty of the estimates of the repeatability and reproducibility standard deviations in a split-level experiment.

To assess the uncertainty of the estimate of the bias of the measurement method in a split-level experiment, calculate the quantity A as defined by equation (13) of ISO 5725-1:1994 with $n = 2$ (or use table 2 of ISO 5725-1:1994), and use this quantity as described in ISO 5725-1.

To assess the uncertainty of the estimate of a laboratory bias in a split-level experiment, calculate the quantity A_w as defined by equation (16) of ISO 5725-1:1994 with $n = 2$. Because the number of replicates in a split-level experiment is, in effect, this number of two, it is not possible to reduce the uncertainty of the estimate of laboratory bias by increasing the number of replicates. (If it is necessary to reduce this uncertainty, the uniform-level design should be used instead.)

4.3.2 Follow the guidance given in clauses 5 and 6 of ISO 5725-2:1994 with regard to the details of the organization of a split-level experiment. The number of replicates, n in ISO 5725-2, may be taken to be the number of split-levels in a split-level design, i.e. two.

The a samples should be allocated to the participants at random, and the b samples should also be allocated to the participants at random and in a separate randomization operation.

Die statistische Fachkraft muss in einem Ringversuch mit Doppelniveauproben bei Übermittlung der Ermittlungsergebnisse (von den Labors an sie) in der Lage sein zu sagen, welches Ermittlungsergebnis vom Material *a* und welches vom Material *b* gewonnen worden ist, und zwar für jedes Merkmalsniveau des Ringversuchs. Man kennzeichne die Proben so, dass dies möglich ist, und es ist darauf zu achten, dass diese Information den teilnehmenden Labors nicht bekannt wird.

Tabelle 1 — Empfohlenes Formblatt für die Zusammenstellung der Ermittlungsergebnisse bei einem Ringversuch mit Doppelniveauproben

Labor	Merkmalsniveau									
	1		2			<i>j</i>			<i>q</i>	
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>		<i>a</i>	<i>b</i>		<i>a</i>	<i>b</i>
1										
2										
<i>i</i>										
<i>p</i>										

4.4 Statistisches Modell

4.4.1 Das in diesem Teil von ISO 5725 angewendete Grundmodell ist als Gleichung (1) im Abschnitt 5 von ISO 5725-1:1994 angegeben. Dort ist festgestellt, dass es für das Schätzen der Genauigkeit (Richtigkeit und Präzision) eines Messverfahrens zweckmäßig ist anzunehmen, dass jedes Ermittlungsergebnis die Summe aus drei Komponenten ist:

$$y_{ijk} = m_j + B_{ij} + e_{ijk} \quad (3)$$

wobei für das einzelne untersuchte Material gilt

m_j ist der (arithmetische) Gesamtmittelwert (Erwartungswert) für ein einzelnes Merkmalsniveau $j = 1, \dots, q$;

B_{ij} ist die systematische Abweichungskomponente eines Labors unter Wiederholbedingungen in einem einzelnen Labor $i = 1, \dots, p$ auf dem einzelnen Merkmalsniveau $j = 1, \dots, q$;

e_{ijk} ist die zufällige Abweichung des Ermittlungsergebnisses $k = 1, \dots, n$, gewonnen unter Wiederholbedingungen im Labor i auf dem Merkmalsniveau j .

4.4.2 Für einen Ringversuch mit Doppelniveauproben entsteht daraus das Modell:

$$y_{ijk} = m_{jk} + B_{ij} + e_{ijk} \quad (4)$$

Es unterscheidet sich von Gleichung (3) in 4.4.1 ausschließlich in einem Zeichen, dem (zusätzlichen) Index k in m_{jk} . Er bedeutet entsprechend Gleichung (4), dass der (arithmetische) Gesamtmittelwert beim Merkmalsniveau j nun vom Material *a* oder *b* abhängen kann, mit ($k = a$ oder b bzw.) $k = 1$ oder 2 .

Das Nichtvorhandensein eines Index k in B_{ij} entspricht der Annahme, dass die systematische Abweichungskomponente eines Labors i auf einem Merkmalsniveau nicht vom Material *a* oder *b* abhängt. Deshalb ist es wichtig, dass die beiden Materialien ähnlich sein sollten.

4.4.3 Die Mittelwerte innerhalb einer Zelle lege man fest als

$$y_{ij} = (y_{ija} + y_{ijb}) / 2 \quad (5)$$

und die Differenzen innerhalb einer Zelle als

$$D_{ij} = y_{ija} - y_{ijb}$$

It is necessary in a split-level experiment for the statistical expert to be able to tell, when the data are reported, which result was obtained on material *a* and which on material *b*, at each level of the experiment. Label the samples so that this is possible, and be careful not to disclose this information to the participants.

Table 1 — Recommended form for the collation of data for the split-level design

Laboratory	Level									
	1		2		j		q			
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
1										
2										
<i>i</i>										
<i>p</i>										

4.4 Statistical model

4.4.1 The basic model used in this part of ISO 5725 is given as equation (1) in clause 5 of ISO 5725-1:1994. It is stated there that for estimating the accuracy (trueness and precision) of a measurement method, it is useful to assume that every measurement result is the sum of three components:

$$y_{ijk} = m_j + B_{ij} + e_{ijk} \quad (3)$$

where, for the particular material tested,

m_j represents the general average (expectation) at a particular level $j = 1, \dots, q$;

B_{ij} represents the laboratory component of bias under repeatability conditions in a particular laboratory $i = 1, \dots, p$ at a particular level $j = 1, \dots, q$;

e_{ijk} represents the random error of test result $k = 1, \dots, n$, obtained in laboratory i at level j , under repeatability conditions.

4.4.2 For a split-level experiment, this model becomes:

$$y_{ijk} = m_{jk} + B_{ij} + e_{ijk} \quad (4)$$

This differs from equation (3) in 4.4.1 in only one feature: the subscript k in m_{jk} implies that according to equation (4) the general average may now depend on the material *a* or *b* ($k = 1$ or 2) within the level j .

The lack of a subscript k in B_{ij} implies that it is assumed that the bias associated with a laboratory i does not depend on the material *a* or *b* within a level. This is why it is important that the two materials should be similar.

4.4.3 Define the cell averages as:

$$y_{ij} = (y_{ija} + y_{ijb}) / 2 \quad (5)$$

and the cell differences as:

$$D_{ij} = y_{ija} - y_{ijb} \quad (6)$$

4.4.4 Der (arithmetische) Gesamtmittelwert eines Ringversuchs mit Doppelniveauproben für ein Merkmalsniveau j kann festgelegt werden als

$$m_j = (m_{ja} + m_{jb}) / 2 \quad (7)$$

4.5 Statistische Auswertung der Ermittlungsergebnisse eines Ringversuchs mit Doppelniveauproben

4.5.1 Man stelle die Ermittlungsergebnisse tabellarisch in einem Formblatt zusammen, wie es in Tabelle 1 gezeigt ist. Jede Kombination aus einem Labor und einem Merkmalsniveau ergibt eine „Zelle“ in dieser tabellarischen Aufstellung und enthält zwei Eintragungen von Ermittlungsergebnissen, y_{ija} und y_{ijb} .

Man errechne die Differenzen innerhalb einer Zelle D_{ij} und trage sie in eine tabellarische Aufstellung ein, wie sie in Tabelle 2 gezeigt ist. Die Auswertungsmethode verlangt, dass jede Differenz in derselben Weise als

$$a - b$$

errechnet wird, und dass das Vorzeichen der Differenz erhalten bleibt.

Man errechne die Mittelwerte innerhalb einer Zelle y_{ij} und trage sie in eine tabellarische Darstellung ein, wie sie in Tabelle 3 gezeigt ist.

4.5.2 Wenn eine Zelle in Tabelle 1 nicht zwei Ermittlungsergebnisse enthält (beispielsweise weil Proben verdorben wurden, oder weil Ermittlungsergebnisse infolge der Anwendung der weiter unten beschriebenen Ausreißertests ausgeschlossen wurden), dann bleiben die entsprechenden Zellen in den Tabellen 2 und 3 leer.

4.5.3 Man errechne für jedes Merkmalsniveau j des Ringversuchs den (arithmetischen) Mittelwert D_j und die Standardabweichung s_{Dj} der Differenzen in der Spalte j der Tabelle 2:

$$D_j = \sum D_{ij} / p \quad (8)$$

$$s_{Dj} = \sqrt{\sum (D_{ij} - D_j)^2 / (p - 1)} \quad (9)$$

Die Summation Σ erstreckt sich hierbei über die Labors $i = 1, 2, \dots, p$.

Wenn in der Tabelle 2 leere Zellen vorkommen, ist p die Anzahl von Zellen in der Spalte j der Tabelle 2, die ein Ermittlungsergebnis enthalten, und die Summation erstreckt sich dann über die nicht leeren Zellen.

4.5.4 Für jedes Merkmalsniveau j des Ringversuchs errechne man den (arithmetischen) Mittelwert y_j und die Standardabweichung s_{yj} der (arithmetischen) Mittelwerte in der Spalte j der Tabelle 3, wobei die Gleichungen zu benutzen sind:

$$y_j = \sum y_{ij} / p \quad (10)$$

$$s_{yj} = \sqrt{\sum (y_{ij} - y_j)^2 / (p - 1)} \quad (11)$$

Die Summation Σ erstreckt sich hierbei über die Labors $i = 1, 2, \dots, p$.

Wenn in der Tabelle 3 leere Zellen vorkommen, ist p die Anzahl von Zellen in der Spalte j der Tabelle 3, die ein Ermittlungsergebnis enthalten, und die Summation erstreckt sich dann über die nicht leeren Zellen.

4.5.5 Man benutze die Tabellen 2 und 3 sowie die (Werte der) Kenngrößen, wie sie gemäß 4.5.3 und 4.5.4 errechnet wurden, um die Ermittlungsergebnisse entsprechend 4.6 auf Konsistenz und auf Ausreißer zu prüfen. Wenn Ermittlungsergebnisse rückgewiesen wurden, errechnet man die (Werte der) Kenngrößen noch einmal.

4.4.4 The general average for a level j of a split-level experiment may be defined as:

$$m_j = (m_{ja} + m_{jb}) / 2 \quad (7)$$

4.5 Statistical analysis of the data from a split-level experiment

4.5.1 Assemble the data into a table as shown in table 1. Each combination of a laboratory and a level gives a "cell" in this table, containing two items of data, y_{ija} and y_{ijb} .

Calculate the cell differences D_{ij} and enter them into a table as shown in table 2. The method of analysis requires each difference to be calculated in the same sense

$$a - b$$

and the sign of the difference to be retained.

Calculate the cell averages y_{ij} and enter them into a table as shown in table 3.

4.5.2 If a cell in table 1 does not contain two test results (for example, because samples have been spoilt, or data have been excluded following the application of the outlier tests described later) then the corresponding cells in tables 2 and 3 both remain empty.

4.5.3 For each level j of the experiment, calculate the average D_j and standard deviation s_{Dj} of the differences in column j of table 2:

$$D_j = \sum D_{ij} / p \quad (8)$$

$$s_{Dj} = \sqrt{\sum (D_{ij} - D_j)^2 / (p - 1)} \quad (9)$$

Here, Σ represents summation over the laboratories $i = 1, 2, \dots, p$.

If there are empty cells in table 2, p is now the number of cells in column j of table 2 containing data and the summation is performed over non-empty cells.

4.5.4 For each level j of the experiment, calculate the average y_j and standard deviation s_{yj} of the averages in column j of table 3, using:

$$y_j = \sum y_{ij} / p \quad (10)$$

$$s_{yj} = \sqrt{\sum (y_{ij} - y_j)^2 / (p - 1)} \quad (11)$$

Here, Σ represents summation over the laboratories $i = 1, 2, \dots, p$.

If there are empty cells in table 3, p is now the number of cells in column j of table 3 containing data and the summation is performed over non-empty cells.

4.5.5 Use tables 2 and 3 and the statistics calculated in 4.5.3 and 4.5.4 to examine the data for consistency and outliers, as described in 4.6. If data are rejected, recalculate the statistics.

4.5.6 Man errechne die Wiederholstandardabweichung s_{rj} und die Vergleichstandardabweichung s_{Rj} aus (den Gleichungen):

$$s_{rj} = s_{Dj} / \sqrt{2} \quad (12)$$

$$s_{Rj}^2 = s_{yj}^2 + s_{rj}^2 / 2 \quad (13)$$

4.5.7 Man untersuche, ob s_{rj} und s_{Rj} vom Mittelwert y_j abhängen, und man ermittle, sofern das so ist, die funktionale Abhängigkeit, indem man die in 7.5 von ISO 5725-2:1994 beschriebene Methode anwendet.

Tabelle 2 — Empfohlenes Formblatt für die Zusammenstellung der Differenzen innerhalb einer Zelle bei einem Ringversuch mit Doppelniveauproben

Labor	Merkmalsniveau					
	1	2		j		q
1						
2						
i						
p						

Tabelle 3 — Empfohlenes Formblatt für die Zusammenstellung der Mittelwerte innerhalb einer Zelle bei einem Ringversuch mit Doppelniveauproben

Labor	Merkmalsniveau					
	1	2		j		q
1						
2						
i						
p						

4.6 Prüfung der Ermittlungsergebnisse auf Vereinbarkeit und Ausreißer

4.6.1 Man prüfe die Ermittlungsergebnisse auf Vereinbarkeit mittels der (Mandelschen) h -Statistiken, wie es in 7.3.1 von ISO 5725-2:1994 beschrieben ist.

Zur Prüfung der Vereinbarkeit der Differenzen innerhalb einer Zelle errechne man die (Werte der) h -Statistiken aus

$$h_{ij} = (D_{ij} - D_j) / s_{Dj} \quad (14)$$

4.5.6 Calculate the repeatability standard deviation s_{rj} and the reproducibility standard deviation s_{Rj} from:

$$s_{rj} = s_{Dj} / \sqrt{2} \quad (12)$$

$$s_{Rj}^2 = s_{yj}^2 + s_{rj}^2 / 2 \quad (13)$$

4.5.7 Investigate whether s_{rj} and s_{Rj} depend on the average y_j , and, if so, determine the functional relationships, using the methods described in subclause 7.5 of ISO 5725-2:1994.

Table 2 — Recommended form for tabulation of cell differences for the split-level design

Laboratory	Level					
	1	2		j		q
1						
2						
i						
p						

Table 3 — Recommended form for tabulation of cell averages for the split-level design

Laboratory	Level					
	1	2		j		q
1						
2						
i						
p						

4.6 Scrutiny of the data for consistency and outliers

4.6.1 Examine the data for consistency using the h statistics, described in subclause 7.3.1 of ISO 5725-2:1994.

To check the consistency of the cell differences, calculate the h statistics as:

$$h_{ij} = (D_{ij} - D_j) / s_{Dj} \quad (14)$$

Zur Prüfung der Vereinbarkeit der Mittelwerte innerhalb einer Zelle errechne man die (Werte der) h -Statistiken aus

$$h_{ij} = (y_{ij} - y_j) / s_{yj} \quad (15)$$

Um Labors zu entdecken, die (also deren Werte mit denen der anderen Labors) nicht vereinbar sind, zeichne man beide Wertefolgen dieser Prüfgrößen abhängig vom Merkmalsniveau auf, wie es in den Bildern 2 und 3 gezeigt ist, und zwar geordnet nach den Labors. Die Deutung dieser grafischen Darstellungen ist vollständig besprochen in 7.3.1 von ISO 5725-2:1994. Wenn ein Labor im Vergleich zu den anderen Labors allgemein eine schlechtere Wiederholpräzision erreicht hat, dann wird sich dies dadurch zeigen, dass es eine außergewöhnlich großen Anzahl von großen (Werten der Mandelschen) h -Statistiken in der grafischen Darstellung zu den Differenzen innerhalb einer Zelle hat. Hat ein Labor Ermittlungsergebnisse gewonnen, die im Allgemeinen eine systematische Abweichungskomponente aufweisen, dann wird sich dies dadurch zeigen, dass es (Werte der) h -Statistiken hat, die in der grafischen Darstellung zu den Mittelwerten innerhalb einer Zelle überwiegend auf einer Seite liegen. In beiden Fällen sollte das Labor gebeten werden, dazu eine Untersuchung durchzuführen und über die Ergebnisse an den Organisator des Ringversuchs zu berichten.

4.6.2 Man prüfe die Ermittlungsergebnisse auf Fastausreißer und Ausreißer durch Anwendung des Grubbs-Tests, wie er in 7.3.4 von ISO 5725-2:1994 beschrieben ist.

Zur Prüfung auf Fastausreißer und Ausreißer bei den Differenzen innerhalb einer Zelle wende man Grubbs-Tests der Reihe nach auf die Werte jeder Spalte von Tabelle 2 an.

Zur Prüfung auf Fastausreißer und Ausreißer bei den Mittelwerten innerhalb einer Zelle wende man Grubbs-Tests der Reihe nach auf die Werte jeder Spalte von Tabelle 3 an.

Die Deutung (der Ergebnisse) dieser Prüfungen ist vollständig besprochen in 7.3.2 von ISO 5725-2:1994. Diese Tests werden ausgeführt, um Ermittlungsergebnisse zu erkennen, die mit den übrigen zum Ringversuch berichteten Ermittlungsergebnissen derart unvereinbar sind, dass ihre Einbeziehung in die Errechnung der Wiederholstandardabweichung und der Vergleichstandardabweichung die Werte dieser Kenngrößen beträchtlich verfälschen würde. Üblicherweise werden Ermittlungsergebnisse, die sich als Ausreißer erwiesen haben, von den Errechnungen ausgeschlossen, während Ermittlungsergebnisse, die sich als Fastausreißer gezeigt haben, einbezogen werden; es sei denn, es gibt gute Gründe für eine andere Verfahrensweise. Sofern die Prüfungen zeigen, dass ein Wert in einer der Tabellen 2 oder 3 von den Errechnungen der Wiederholstandardabweichung und der Vergleichstandardabweichung auszuschließen sind, dann sollte der entsprechende Wert in der anderen Tabelle ebenfalls von der Errechnung ausgeschlossen werden.

4.7 Ergebnisbericht zum Ringversuch mit Doppelniveauproben

4.7.1 Eine Empfehlung ist in 7.7 von ISO 5725-2:1994 gegeben bezüglich

- des Ergebnisberichts zur statistischen Auswertung an das Versuchsgremium;
- der durch das Versuchsgremium zu treffenden Entscheidungen; und
- der Vorbereitung des Abschlussberichts.

4.7.2 Empfehlungen zur Form einer bekanntgegebenen Feststellung der Wiederholstandardabweichung und der Vergleichstandardabweichung eines vereinheitlichten Messverfahrens sind in 7.1 von ISO 5725-1:1994 gegeben.

4.8 Beispiel 1: Ein Ringversuch mit Doppelniveauproben – Ermittlung von Protein

4.8.1 Tabelle 4 enthält die Ermittlungsergebnisse eines Ringversuchs^[5], bei dem es um die Ermittlung des Proteingehalts in Futtermitteln durch Verbrennungsanalyse geht. Es gab neun beteiligte Labors. Der Ringversuch enthielt 14 Merkmalsniveaus. Auf jedem Merkmalsniveau wurden zwei Futtermittel mit ähnlichem Massenanteil an Protein im Futtermittel angewendet.

4.8.2 Tabellen 5 und 6 zeigen die Mittelwerte innerhalb einer Zelle und die Differenzen innerhalb einer Zelle, errechnet wie in 4.5.1 beschrieben, und zwar speziell für das Merkmalsniveau 14 ($j = 14$) des Ringversuchs.

To check the consistency of the cell averages, calculate the h statistics as:

$$h_{ij} = (y_{ij} - y_j) / s_{yj} \quad (15)$$

To show up inconsistent laboratories, plot both sets of these statistics in the order of the levels, but grouped by laboratory, as shown in figures 2 and 3. The interpretation of these graphs is discussed fully in subclause 7.3.1 of ISO 5725-2:1994. If a laboratory is achieving generally worse repeatability than the others, then it will show up as having an unusually large number of large h statistics in the graph derived from the cell differences. If a laboratory is achieving results that are generally biased, then it will show up as having h statistics mostly in one direction on the graph derived from the cell averages. In either case, the laboratory should be asked to investigate and report their findings back to the organizer of the experiment.

4.6.2 Examine the data for stragglers and outliers using Grubbs' tests, described in subclause 7.3.4 of ISO 5725-2:1994.

To test for stragglers and outliers in the cell differences, apply Grubbs' tests to the values in each column of table 2 in turn.

To test for stragglers and outliers in the cell averages, apply Grubbs' tests to the values in each column of table 3 in turn.

The interpretation of these tests is discussed fully in subclause 7.3.2 of ISO 5725-2:1994. They are used to identify results that are so inconsistent with the remainder of the data reported in the experiment that their inclusion in the calculation of the repeatability and reproducibility standard deviations would affect the values of these statistics substantially. Usually, data shown to be outliers are excluded from the calculations, and data shown to be stragglers are included, unless there is a good reason for doing otherwise. If the tests show that a value in one of tables 2 or 3 is to be excluded from the calculation of the repeatability and reproducibility standard deviations, then the corresponding value in the other of these tables should also be excluded from the calculation.

4.7 Reporting the results of a split-level experiment

4.7.1 Advice is given in subclause 7.7 of ISO 5725-2:1994 on:

- reporting the results of the statistical analysis to the panel;
- decisions to be made by the panel; and
- the preparation of a full report.

4.7.2 Recommendations on the form of a published statement of the repeatability and reproducibility standard deviations of a standard measurement method are given in subclause 7.1 of ISO 5725-1:1994.

4.8 Example 1: A split-level experiment — Determination of protein

4.8.1 Table 4 contains the data from an experiment [5] which involved the determination by combustion of the content of protein in feeds. There were nine participating laboratories, and the experiment contained 14 levels. Within each level, two feeds were used having similar mass fraction of protein in feed.

4.8.2 Tables 5 and 6 show the cell averages and differences, calculated as described in clause 4.5.1, for just Level 14 ($j = 14$) of the experiment.

Using equations (8) and (9) in 4.5.3, the differences in table 5 give:

$$D_{14} = 8,34 \%$$

$$s_{D14} = 0,436 \%$$

Unter Anwendung der Gleichungen (8) und (9) in 4.5.3 ergeben sich aus den Differenzen in Tabelle 5 (die folgende Werte der Kenngrößen):

$$D_{14} = 8,34 \%$$

$$s_{D14} = 0,436 \text{ 1 } \%$$

Weiter ergeben sich unter Anwendung der Gleichungen (10) und (11) in 4.5.4 auf die Mittelwerte in Tabelle 6 (die folgenden Werte der Kenngrößen):

$$y_{14} = 85,46 \%$$

$$s_{y14} = 0,453 \text{ 4 } \%,$$

so dass sich die Wiederholstandardabweichung und die Vergleichstandardabweichung unter Anwendung der Gleichungen (12) und (13) in 4.5.6 (mit folgenden Werte) ergeben als:

$$s_{r14} = 0,31 \%$$

$$s_{R14} = 0,50 \%$$

Tabelle 7 zeigt die Ergebnisse der Errechnungen für die anderen Merkmalsniveaus.

4.8.3 Bild 1 zeigt für das Merkmalsniveau 14 die Ergebnisse für die Proben *a* aus Tabelle 4, grafisch aufgetragen gegen die entsprechenden Ergebnisse für die Proben *b* in einem „Youden-Diagramm“. Labor 5 liefert einen Punkt in der unteren linken Ecke der Grafik, und das Labor 1 einen Punkt in der oberen rechten Ecke: Das zeigt, dass die Ergebnisse des Labors 5 eine gleichbleibend negative systematische Abweichungskomponente für die Proben *a* und *b* haben, und dass die Ergebnisse vom Labor 1 eine gleichbleibend positive systematische Abweichungskomponente für die beiden Proben haben. Man findet häufig diese Art von Muster, wenn man die Ergebnisse eines Ringversuchs mit Doppelniveauproben aufzeichnet wie im Bild 1. Das Bild zeigt außerdem, dass die Ergebnisse des Labors 4 insofern ungewöhnlich sind, als der Punkt für dieses Labor einen beträchtlichen Abstand von der (diagonalen) Proportionalitätsgeraden (durch den Kreuzungspunkt der Mittelwerte der ähnlichen Proben *a* und *b*) aufweist. Die anderen Labors bilden eine Gruppe in der Mitte der grafischen Darstellung. Dieses Bild liefert daher einen Anlass zur Ermittlung der Ursachen für die systematischen Abweichungskomponenten der drei Labors.

ANMERKUNG Weitere Information zur Deutung von „Youden-Diagrammen“ findet man in den Literaturstellen [7] und [8].

4.8.4 Die Werte der (Mandelschen) *h*-Statistiken, errechnet wie in 4.6.1 beschrieben, sind in den Tabellen 5 und 6 nur für das Merkmalsniveau 14 gezeigt. Die Werte für alle Merkmalsniveaus sind in den Bildern 2 und 3 grafisch dargestellt.

In Bild 3, der Darstellung der *h*-Statistiken für die Mittelwerte innerhalb einer Zelle, zeigt sich, dass Labor 5 auf allen Merkmalsniveaus negative *h*-Statistiken geliefert hat. Das zeigt an, dass dessen Ermittlungsergebnisse eine gleichbleibend negative systematische Abweichungskomponente aufweisen. Für das gleiche Bild 3 haben die Labors 8 und 9 *h*-Statistiken geliefert, die nahezu alle positiv sind. Das zeigt an, dass deren Ermittlungsergebnisse gleichbleibend positive systematische Abweichungskomponenten aufweisen, (allerdings mit kleineren Beträgen als die negative systematische Abweichungskomponente von Labor 5). Zudem zeigen die *h*-Statistiken der Labors 1, 2 und 6 eine systematische Abweichungskomponente, die sich bei jedem dieser Labors (systematisch) mit dem Merkmalsniveau ändert. Solche Wechselwirkungen zwischen den (Werten der) Labors und den Merkmalsniveaus können Anhaltspunkte für die Ursachen der systematischen Abweichungskomponenten eines Labors liefern.

Bild 2 lässt keine beachtenswerten Muster erkennen.

4.8.5 Werte der Prüfgrößen für den Grubbs-Test sind in Tabelle 8 angegeben. Diese Tests zeigen wiederum, dass die Ermittlungsergebnisse des Labors 5 verdächtig sind.

4.8.6 An dieser Stelle der Auswertung sollte die statistische Fachkraft eine Untersuchung der möglichen Ursachen der verdächtigen Ermittlungsergebnisse im Labor 5 in die Wege leiten, bevor sie die Auswertung der Ermittlungsergebnisse weiterführt. Lässt sich die Ursache nicht herausfinden, dann ist das in diesem Beispiel ein Anlass für den Ausschluss aller Ermittlungsergebnisse des Labors 5 aus den Errechnungen der Wiederholstandardabweichung und der Vergleichstandardabweichung. Die Auswertung kann dann mit einer Untersuchung möglicher funktionaler Beziehungen zwischen den Wiederholstandardabweichungen und den Vergleichstandardabweichungen (einerseits) und den Gesamtmittelwerten (auf dem jeweiligen Merkmalsniveau andererseits) fortgeführt werden. Das lässt keine Fragen entstehen, die nicht bereits in ISO 5725-2 erläutert worden wären. Deshalb wird dies hier nicht betrachtet.

and applying equations (10) and (11) in 4.5.4 to the averages in table 6 gives:

$$y_{14} = 85,46 \%$$

$$s_{Y14} = 0,453 \%$$

so the repeatability and reproducibility standard deviations are, using equations (12) and (13) in 4.5.6:

$$s_{r14} = 0,31 \%$$

$$s_{R14} = 0,50 \%$$

Table 7 gives the results of the calculations for the other levels.

4.8.3 Figure 1 shows the results for samples *a* from table 4 plotted against the corresponding results for samples *b*, for Level 14, in a "Youden plot". Laboratory 5 gives a point in the bottom left-hand corner of the graph, and Laboratory 1 gives a point in the top right-hand corner: this indicates that the data from Laboratory 5 have a consistent negative bias over samples *a* and *b*, and that the data from Laboratory 1 have a consistent positive bias over the two samples. It is common to find this sort of pattern when plotting the data from a split-level design as in figure 1. The figure also shows that the results for Laboratory 4 are unusual, as the point for this laboratory is some distance from the line of equality for the two samples. The other laboratories form a group in the middle of the plot. This figure thus provides a case for investigating the causes of the biases at the three laboratories.

NOTE — For further information on the interpretation of "Youden plots", see references [7] and [8].

4.8.4 The values of the *h* statistics, calculated as described in 4.6.1, are shown in tables 5 and 6, for only Level 14. The values for all levels are plotted in figures 2 and 3.

In figure 3, the *h* statistics for cell averages show that Laboratory 5 gave negative *h* statistics at all levels, indicating a consistent negative bias in their data. In the same figure, Laboratories 8 and 9 gave *h* statistics that are nearly all positive, indicating consistent positive biases in their data (but smaller than the negative bias in Laboratory 5). Also, the *h* statistics for Laboratories 1, 2 and 6 indicate a bias that changes with level in each of these laboratories. Such interactions between the laboratories and the levels may provide clues as to the causes of the laboratory biases.

Figure 2 does not reveal any noteworthy pattern.

4.8.5 Values of the Grubbs' statistics are given in table 8. These tests again indicate that the data from Laboratory 5 are suspect.

4.8.6 At this point in the analysis, the statistical expert should initiate an investigation at Laboratory 5 of the possible causes of the suspect data, before proceeding with the analysis of the data. If the cause cannot be identified, there is a case in this instance for excluding all data from Laboratory 5 from the calculation of the repeatability and reproducibility standard deviations. The analysis would then continue with an investigation of possible functional relationships between the repeatability and reproducibility standard deviations and the general average. This does not raise any issues that have not already been covered in ISO 5725-2, so it will not be considered here.

Tabelle 4 — Beispiel 1: Ermittlung des Massenanteils von Protein in Futtermitteln, ausgedrückt als Prozentsatz

Labor	Merkmalsniveau									
	1		2		3		4		5	
	a	b	a	b	a	b	a	b	a	b
1	11,11	10,34	10,91	9,81	13,74	13,48	13,79	13,00	15,89	15,26
2	11,12	9,94	11,38	10,31	14,00	13,12	13,44	13,06	15,69	15,10
3	11,26	10,46	10,95	10,51	13,38	12,70	13,54	13,18	15,83	15,73
4	11,07	10,41	11,66	9,95	13,01	13,16	13,58	12,88	15,08	15,63
5	10,69	10,31	10,98	10,13	13,24	13,33	13,32	12,59	15,02	14,90
6	11,73	11,01	12,31	10,92	14,01	13,66	14,04	13,64	16,43	15,94
7	11,13	10,36	11,38	10,44	12,94	12,44	13,63	13,06	15,75	15,56
8	11,21	10,51	11,32	10,84	13,09	13,76	13,85	13,49	15,98	15,89
9	11,80	11,21	11,35	9,88	13,85	14,46	13,96	13,77	16,51	15,72
Labor	Merkmalsniveau									
	6		7		8		9		10	
	a	b	a	b	a	b	a	b	a	b
1	20,14	19,78	20,33	20,06	46,45	44,42	52,05	49,40	65,84	59,14
2	19,25	20,25	20,36	19,94	46,69	44,62	51,94	48,81	66,31	59,19
3	20,48	19,86	20,56	20,11	46,90	44,56	52,18	48,90	66,06	58,52
4	21,54	20,06	20,64	20,46	47,13	45,29	51,73	48,56	65,93	58,93
5	19,90	19,66	20,56	19,24	45,83	43,73	50,84	47,91	64,19	57,94
6	20,31	20,27	20,85	20,63	46,86	43,96	52,18	49,03	65,73	58,77
7	20,00	20,56	20,25	20,19	46,25	44,31	52,25	49,44	66,06	59,19
8	20,43	20,69	20,85	20,27	47,11	44,40	52,44	48,81	65,66	59,38
9	20,64	21,01	20,78	20,89	47,09	45,15	52,19	48,46	66,33	59,47
Labor	Merkmalsniveau									
	11		12		13		14			
	a	b	a	b	a	b	a	b		
1	84,16	80,86	85,38	81,71	87,64	88,23	90,24	82,10		
2	84,50	81,06	85,56	82,44	88,81	88,38	89,88	81,44		
3	82,26	79,43	85,26	82,15	88,58	88,12	89,48	81,67		
4	84,39	80,08	85,20	81,76	88,47	87,98	90,04	80,73		
5	81,71	79,01	83,58	79,74	86,43	86,19	88,59	80,46		
6	82,85	81,16	84,44	80,90	87,78	86,89	89,40	80,88		
7	86,25	81,00	84,88	81,44	88,06	88,00	89,31	81,38		
8	84,59	81,16	84,96	81,71	88,50	87,98	89,94	81,56		
9	83,05	80,93	84,73	81,94	88,24	88,05	89,75	81,35		

Table 4 — Example 1: Determination of mass fraction of protein in feed, expressed as a percentage

Laboratory	Level									
	1		2		3		4		5	
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
1	11,11	10,34	10,91	9,81	13,74	13,48	13,79	13,00	15,89	15,26
2	11,12	9,94	11,38	10,31	14,00	13,12	13,44	13,06	15,69	15,10
3	11,26	10,46	10,95	10,51	13,38	12,70	13,54	13,18	15,83	15,73
4	11,07	10,41	11,66	9,95	13,01	13,16	13,58	12,88	15,08	15,63
5	10,69	10,31	10,98	10,13	13,24	13,33	13,32	12,59	15,02	14,90
6	11,73	11,01	12,31	10,92	14,01	13,66	14,04	13,64	16,43	15,94
7	11,13	10,36	11,38	10,44	12,94	12,44	13,63	13,06	15,75	15,56
8	11,21	10,51	11,32	10,84	13,09	13,76	13,85	13,49	15,98	15,89
9	11,80	11,21	11,35	9,88	13,85	14,46	13,96	13,77	16,51	15,72
Laboratory	Level									
	6		7		8		9		10	
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
1	20,14	19,78	20,33	20,06	46,45	44,42	52,05	49,40	65,84	59,14
2	19,25	20,25	20,36	19,94	46,69	44,62	51,94	48,81	66,31	59,19
3	20,48	19,86	20,56	20,11	46,90	44,56	52,18	48,90	66,06	58,52
4	21,54	20,06	20,64	20,46	47,13	45,29	51,73	48,56	65,93	58,93
5	19,90	19,66	20,56	19,24	45,83	43,73	50,84	47,91	64,19	57,94
6	20,31	20,27	20,85	20,63	46,86	43,96	52,18	49,03	65,73	58,77
7	20,00	20,56	20,25	20,19	46,25	44,31	52,25	49,44	66,06	59,19
8	20,43	20,69	20,85	20,27	47,11	44,40	52,44	48,81	65,66	59,38
9	20,64	21,01	20,78	20,89	47,09	45,15	52,19	48,46	66,33	59,47
Laboratory	Level									
	11		12		13		14			
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>		
1	84,16	80,86	85,38	81,71	87,64	88,23	90,24	82,10		
2	84,50	81,06	85,56	82,44	88,81	88,38	89,88	81,44		
3	82,26	79,43	85,26	82,15	88,58	88,12	89,48	81,67		
4	84,39	80,08	85,20	81,76	88,47	87,98	90,04	80,73		
5	81,71	79,01	83,58	79,74	86,43	86,19	88,59	80,46		
6	82,85	81,16	84,44	80,90	87,78	86,89	89,40	80,88		
7	86,25	81,00	84,88	81,44	88,06	88,00	89,31	81,38		
8	84,59	81,16	84,96	81,71	88,50	87,98	89,94	81,56		
9	83,05	80,93	84,73	81,94	88,24	88,05	89,75	81,35		

Tabelle 5 — Beispiel 1: Differenzen innerhalb einer Zelle für Merkmalsniveau 14

Labor	Differenz innerhalb einer Zelle in %	<i>h</i> -Statistik
1	8,14	– 0,459
2	8,44	0,229
3	7,81	– 1,215
4	9,31	2,224
5	8,13	– 0,482
6	8,52	0,413
7	7,93	– 0,940
8	8,38	0,092
9	8,40	0,138

Tabelle 6 — Beispiel 1: Mittelwert innerhalb einer Zelle für Merkmalsniveau 14

Labor	Mittelwert innerhalb einer Zelle in %	<i>h</i> -Statistik
1	86,170	1,576
2	85,660	0,451
3	85,575	0,263
4	85,385	– 0,156
5	84,525	– 2,052
6	85,140	– 0,696
7	85,345	– 0,244
8	85,750	0,649
9	85,550	0,208

Tabelle 7 — Beispiel 1: Werte der Gesamtmittelwerte, der mittleren Differenzen innerhalb einer Zelle sowie der Standardabweichungen, errechnet aus den Ermittlungsergebnissen für alle 14 Merkmalsniveaus in Tabelle 4

Merkmals-niveau <i>j</i>	Anzahl der Labors <i>p</i>	Gesamt-mittelwert <i>y_j</i> in %	Mittlere Differenz <i>D_j</i> in %	Standardabweichungen			
				<i>s_{yy}</i> in %	<i>s_{Dj}</i> in %	<i>s_{rj}</i> in %	<i>s_{Rj}</i> in %
1	9	10,87	0,73	0,35	0,21	0,15	0,36
2	9	10,84	1,05	0,36	0,43	0,30	0,42
3	9	13,41	0,13	0,44	0,55	0,39	0,52
4	9	13,43	0,50	0,30	0,21	0,15	0,32
5	9	15,66	0,27	0,39	0,40	0,29	0,44
6	9	20,27	0,06	0,40	0,73	0,52	0,54
7	9	20,39	0,38	0,30	0,41	0,29	0,37
8	9	45,60	2,21	0,44	0,37	0,26	0,47
9	9	50,40	3,16	0,44	0,35	0,25	0,47
10	9	62,37	6,84	0,53	0,40	0,28	0,57
11	9	82,14	3,23	1,01	1,08	0,77	1,15
12	9	83,17	3,45	0,74	0,46	0,33	0,77
13	9	87,91	0,30	0,69	0,41	0,29	0,72
14	9	85,46	8,34	0,45	0,44	0,31	0,50

Table 5 — Example 1: Cell differences for Level 14

Laboratory	Cell difference %	h statistic
1	8,14	− 0,459
2	8,44	0,229
3	7,81	− 1,215
4	9,31	2,224
5	8,13	− 0,482
6	8,52	0,413
7	7,93	− 0,940
8	8,38	0,092
9	8,40	0,138

Table 6 — Example 1: Cell averages for Level 14

Laboratory	Cell difference %	h statistic
1	86,170	1,576
2	85,660	0,451
3	85,575	0,263
4	85,385	− 0,156
5	84,525	− 2,052
6	85,140	− 0,696
7	85,345	− 0,244
8	85,750	0,649
9	85,550	0,208

Table 7 — Example 1: Values of averages, average differences, and standard deviations calculated from the data for all 14 levels in table 4

Level j	Number of laboratories p	General average y_j %	Average difference D_j %	Standard deviations			
				s_{yj} %	s_{Dj} %	s_{rj} %	s_{Rj} %
1	9	10,87	0,73	0,35	0,21	0,15	0,36
2	9	10,84	1,05	0,36	0,43	0,30	0,42
3	9	13,41	0,13	0,44	0,55	0,39	0,52
4	9	13,43	0,50	0,30	0,21	0,15	0,32
5	9	15,66	0,27	0,39	0,40	0,29	0,44
6	9	20,27	0,06	0,40	0,73	0,52	0,54
7	9	20,39	0,38	0,30	0,41	0,29	0,37
8	9	45,60	2,21	0,44	0,37	0,26	0,47
9	9	50,40	3,16	0,44	0,35	0,25	0,47
10	9	62,37	6,84	0,53	0,40	0,28	0,57
11	9	82,14	3,23	1,01	1,08	0,77	1,15
12	9	83,17	3,45	0,74	0,46	0,33	0,77
13	9	87,91	0,30	0,69	0,41	0,29	0,72
14	9	85,46	8,34	0,45	0,44	0,31	0,50

Tabelle 8 — Beispiel 1: Werte der Prüfgrößen nach Grubbs

Werte der Prüfgrößen nach Grubbs für Differenzen innerhalb einer Zelle zur Prüfung				
beim Merkmals-niveau	auf einzelnen Aus-reißer unten	auf zwei Ausreißer unten	auf zwei Ausreißer oben	auf einzelnen Aus-reißer oben
1	1,653	0,5081	0,3139	2,125
2	1,418	0,3945	0,4738	1,535
3	1,462	0,3628	0,5323	1,379
4	1,490	0,5841	0,4771	1,414
5	2,033	0,3485	0,6075	1,289
6	1,456	0,5490	0,3210	1,947
7	1,185	0,6820	0,1712	2,296* (5)
8	0,996	0,7571	0,1418* (6; 8)	1,876
9	1,458	0,5002	0,3092	1,602
10	1,474	0,3360	0,4578	1,737
11	1,422	0,5089	0,2943	1,865
12	1,418	0,6009	0,2899	1,956
13	2,172	0,2325	0,6326	1,444
14	1,215	0,6220	0,2362	2,224* (4)
Werte der Prüfgrößen nach Grubbs für Mittelwerte innerhalb einer Zelle zur Prüfung				
beim Merkmals-niveau	auf einzelnen Aus-reißer unten	auf zwei Ausreißer unten	auf zwei Ausreißer oben	auf einzelnen Aus-reißer oben
1	1,070	0,6607	0,1291* (6; 9)	1,832
2	1,318	0,6288	0,2118	2,165
3	1,621	0,4771	0,4077	1,680
4	1,591	0,5339	0,3807	1,429
5	1,794	0,4018	0,5009	1,333
6	1,291	0,4947	0,4095	1,386
7	1,599	0,5036	0,4391	1,470
8	1,872	0,3753	0,4536	1,404
9	2,328* (5)	0,1317* (4; 5)	0,7417	1,025
10	2,456** (5)	-	-	1,000
11	1,756	0,2469	0,5759	1,472
12	2,037	0,1063* (5; 6)	0,7116	1,130
13	2,308* (5)	0,0733** (5; 6)	0,7777	0,994
14	2,052	0,2781	0,5486	1,576
ANMERKUNG Zahlen in Klammern zeigen die Labors an, die Fastausreißer oder Ausreißer verursachen.				
Die kritischen Werte für die Prüfgrößen des Grubbs-Tests für 9 Labors, sowohl angewendet auf die Differenzen innerhalb einer Zelle als auch auf die Mittelwerte innerhalb einer Zelle, sind die folgenden:				
		Fastausreißer	Ausreißer	
		(*)	(**)	
Kritischer Wert Grubbs-Test für einzelnen Extremwert auf einer Seite		2,215	2,387	
Kritischer Wert Grubbs-Test für zwei Extremwerte auf einer Seite		0,149 2	0,085 1	

Table 8 — Example 1: Values of Grubbs' statistics

Grubbs' statistics for differences				
Level	One smallest	Two smallest	Two largest	One largest
1	1,653	0,5081	0,3139	2,125
2	1,418	0,3945	0,4738	1,535
3	1,462	0,3628	0,5323	1,379
4	1,490	0,5841	0,4771	1,414
5	2,033	0,3485	0,6075	1,289
6	1,456	0,5490	0,3210	1,947
7	1,185	0,6820	0,1712	2,296* (5)
8	0,996	0,7571	0,1418* (6; 8)	1,876
9	1,458	0,5002	0,3092	1,602
10	1,474	0,3360	0,4578	1,737
11	1,422	0,5089	0,2943	1,865
12	1,418	0,6009	0,2899	1,956
13	2,172	0,2325	0,6326	1,444
14	1,215	0,6220	0,2362	2,224* (4)
Grubbs' statistics for cell averages				
Level	One smallest	Two smallest	Two largest	One largest
1	1,070	0,6607	0,1291* (6; 9)	1,832
2	1,318	0,6288	0,2118	2,165
3	1,621	0,4771	0,4077	1,680
4	1,591	0,5339	0,3807	1,429
5	1,794	0,4018	0,5009	1,333
6	1,291	0,4947	0,4095	1,386
7	1,599	0,5036	0,4391	1,470
8	1,872	0,3753	0,4536	1,404
9	2,328* (5)	0,1317* (4; 5)	0,7417	1,025
10	2,456** (5)	-	-	1,000
11	1,756	0,2469	0,5759	1,472
12	2,037	0,1063* (5; 6)	0,7116	1,130
13	2,308* (5)	0,0733** (5; 6)	0,7777	0,994
14	2,052	0,2781	0,5486	1,576
NOTE — Numbers in brackets indicate the laboratories that give rise to the stragglers or outliers.				
The critical values of the Grubbs' test statistics for 9 laboratories, whether applied to the differences or the cell averages, are as follows.				
	Straggler (*)	Outlier (**)		
Grubbs' test for a single outlier	2,215	2,387		
Grubbs' test for a pair of outliers	0,149 2	0,085 1		

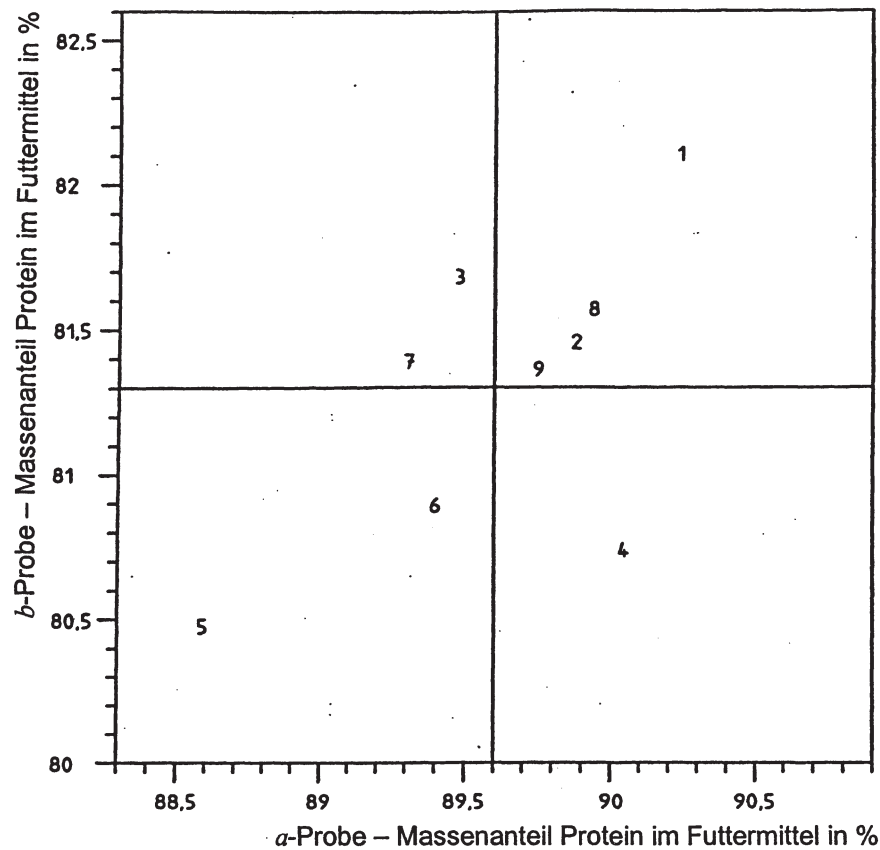


Bild 1 — Beispiel 1: Auf dem Merkmalsniveau 14 gewonnene Ermittlungsergebnisse

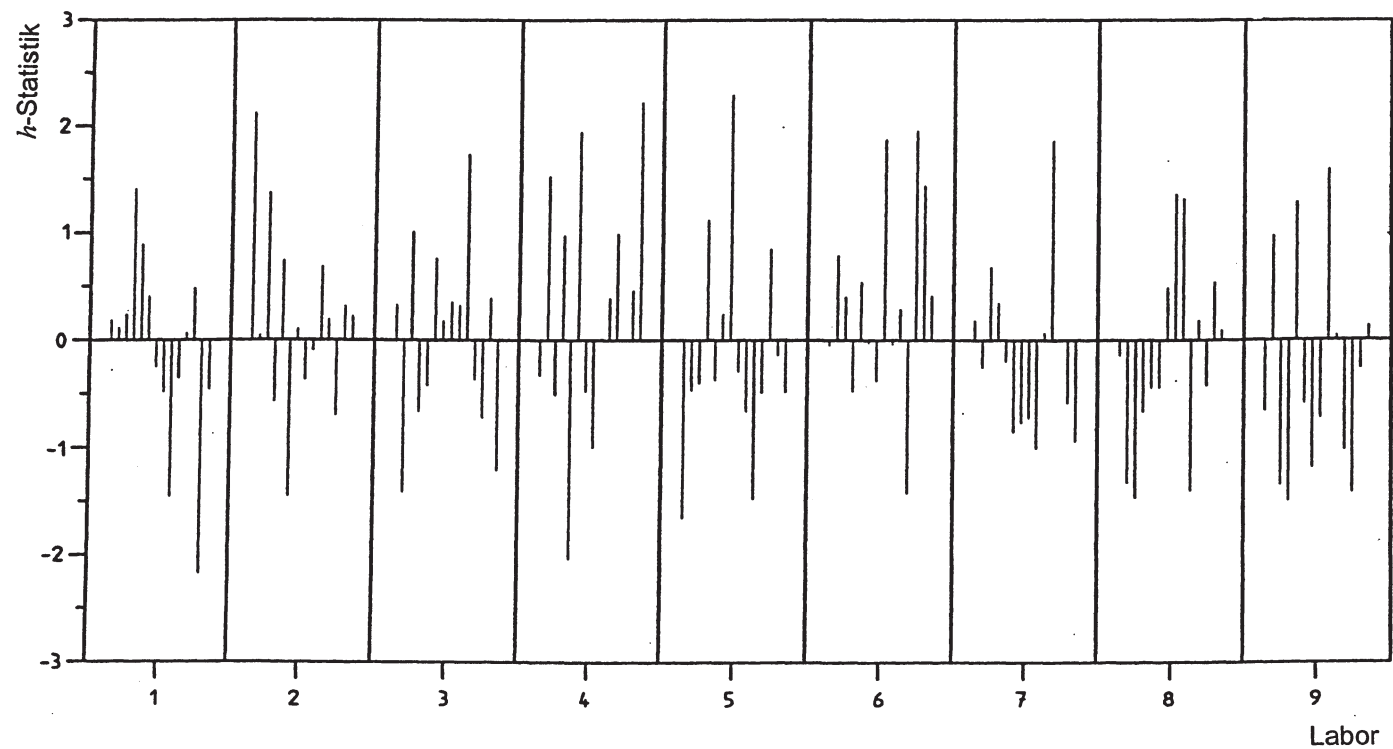


Bild 2 — Beispiel 1: Prüfung der Differenzen innerhalb einer Zelle auf Vereinbarkeit (gruppiert nach Labor)

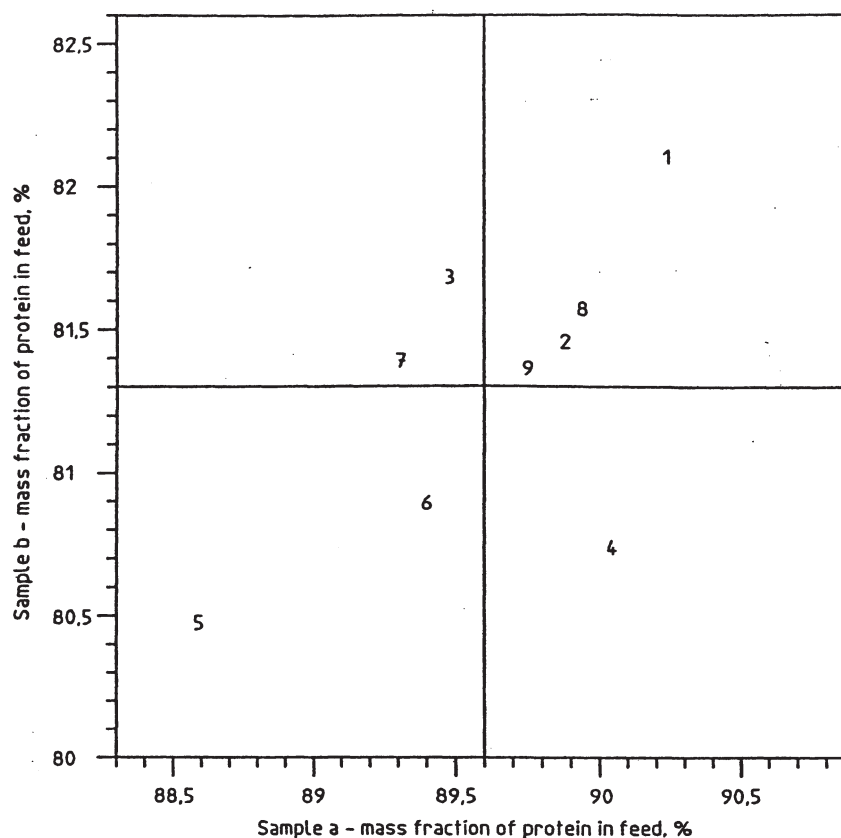


Figure 1 — Example 1: Data obtained at Level 14

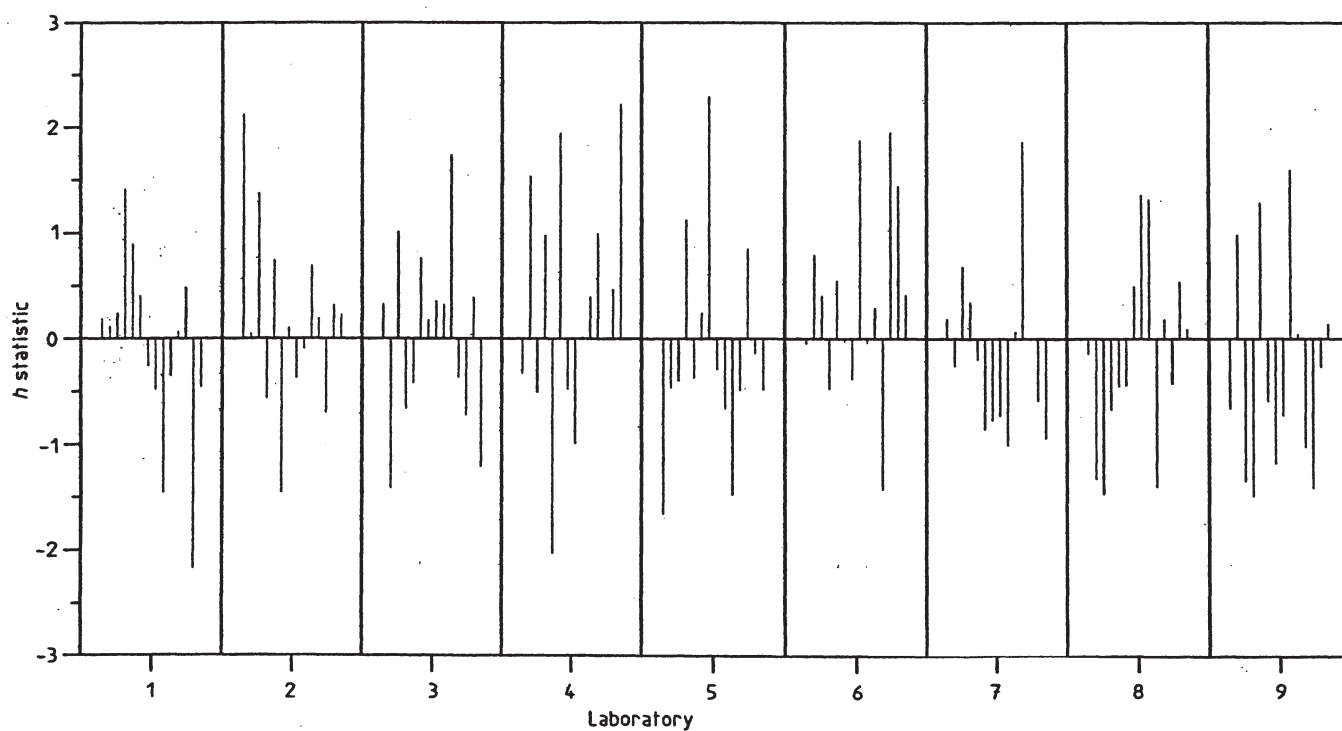


Figure 2 — Example 1: Consistency check on cell differences (grouped by laboratory)

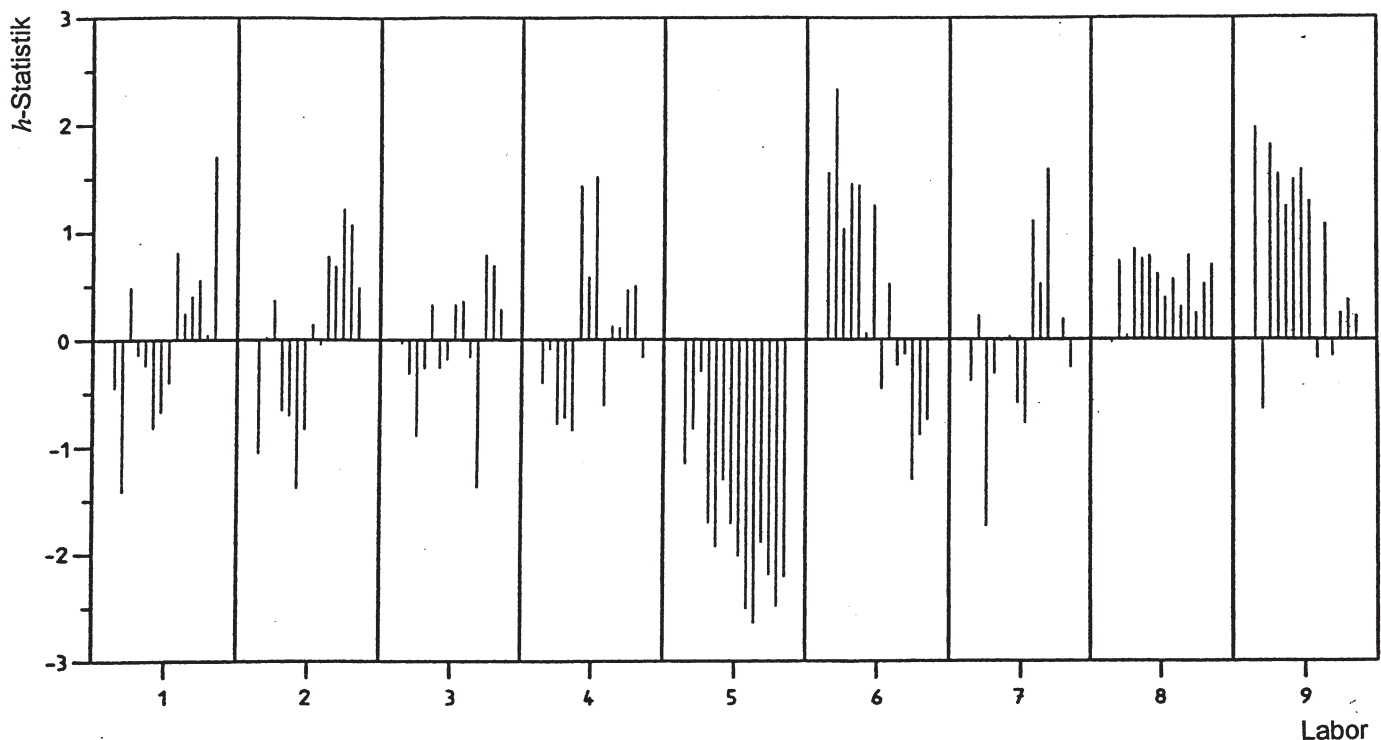


Bild 3 — Beispiel 1: Prüfung der Mittelwerte innerhalb einer Zelle auf Vereinbarkeit (gruppiert nach Labor)

5 Ringversuch an heterogenem Material

5.1 Anwendungshinweise zum Ringversuch an heterogenem Material

5.1.1 Ein Beispiel für heterogenes Material ist Leder. Es gibt keine zwei gleiche Häute, und die Merkmale von Leder schwanken erheblich innerhalb einer Haut. Eine übliche auf Leder angewendete Prüfung ist eine Reißfestigkeitsprüfung gemäß BS 3144 ^[3]. Diese wird ausgeführt an hantelförmigen Prüfstücken (BS 3144 spezifiziert die von einer Haut auszuschneidende Anzahl solcher Prüfstücke samt Ort der Entnahme und Orientierung der Prüfstücke innerhalb der Haut, womit diese naturbezogene Festlegung eine „Stichprobe“ für die Prüfung der ganzen Haut bewirkt). Wenn ein Ringversuch zur Ermittlung der Präzision mit Proben auf einheitlichem Merkmalsniveau ausgeführt wird, wie er in ISO 5725-2 beschrieben ist, bei dem jedem Labor eine einzige Haut für jedes Merkmalsniveau des Ringversuchs gegeben wurde, und zwei Ermittlungsergebnisse von jeder Haut gewonnen werden, wird sich die Schwankung (der Werte der Reißfestigkeit) zwischen den Häuten der Schwankung zwischen den Labors überlagern. Deshalb wird die Vergleichstandardabweichung erhöhen. Wenn allerdings jedem Labor für jedes Merkmalsniveau zwei Häute gegeben wurden, und es werden zwei Ermittlungsergebnisse an jeder Haut gewonnen, dann können die Ermittlungsergebnisse sowohl zur Schätzung der Schwankung zwischen den Häuten als auch zur Errechnung eines Wertes für die Vergleichstandardabweichung des Messverfahrens benutzt werden, aus dem die Schwankung zwischen den Häuten herausgerechnet wurde.

5.1.2 Ein anderes Beispiel eines heterogenen Materials ist Sand (der beispielsweise für die Herstellung von Beton benutzt wird). Dieses Material hat sich unter Einwirkung von Wind oder Wasser in Schichten abgelagert, die stets Abstufungen in der Teilchengröße enthalten. Deshalb ist bei der Anwendung von Sand die Teilchengrößenverteilung stets von Interesse. In der Beton-Technologie wird die Verteilung der Teilchengröße von Sand durch Siebanalyse geprüft (beispielsweise BS 812-103 ^[1]). Um einen Sand zu prüfen, wird eine Schüttgutprobe des Produkts entnommen, aus der anschließend eine oder mehrere Prüfproben hergestellt werden. Typischerweise wird eine Schüttgutprobe eine Masse von etwa 10 kg und die Prüfproben werden etwa 200 g haben. Wegen der natürlichen Schwankung des Materials wird naturgemäß stets eine gewisse Schwankung zwischen Schüttgutproben desselben Produkts bestehen. Also wird in ähnlicher Weise wie bei Leder dann, wenn ein Ringversuch mit Proben auf einheitlichem Merkmalsniveau ausgeführt wird, bei dem jedem Labor eine einzige Schüttgutprobe für jedes Merkmalsniveau (des Ringversuchs) gegeben wurde, die Schwankung zwischen den Schüttgutproben die errechnete Vergleichstandardabweichung des Messverfahrens erhöhen. Wenn aber die Labors zwei Schüttgutproben für jedes Merkmalsniveau erhalten, können Werte für die Vergleichstandardabweichung errechnet werden, welche diese Schwankung unberücksichtigt lassen.

5.1.3 Die obigen Beispiele heben außerdem ein anderes Charakteristikum heterogenen Materials hervor: Wegen der Schwankung des Materials kann die Herstellung von Proben oder Prüfmengen eine bedeutende Quelle für Schwankungen sein. Deshalb kann bei Leder der Prozess des Ausschneidens von Proben aus einer

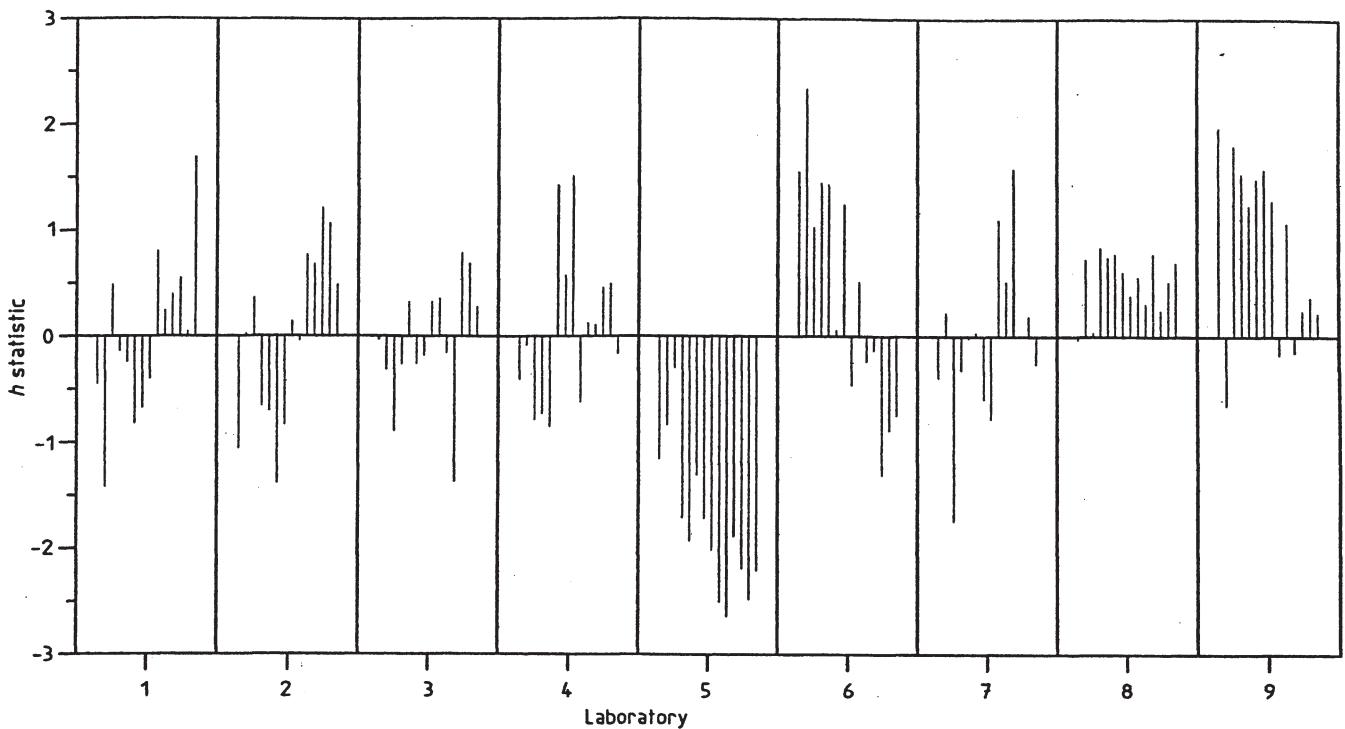


Figure 3 — Example 1: Consistency check on cell averages (grouped by laboratory)

5 A design for a heterogeneous material

5.1 Applications of the design for a heterogeneous material

5.1.1 An example of a heterogeneous material is leather; no two hides are the same, and the properties of leather vary substantially within a hide. A common test that is applied to leather is a tensile strength test in BS 3144 [3]. This is performed on dumbbell-shaped specimens (BS 3144 specifies the number of such specimens to cut from a hide, and also their position and orientation within the hide, so the natural definition of a “sample” to use when testing leather is a complete hide). If a precision experiment is performed using the uniform level design described in ISO 5725-2, in which each laboratory is sent one hide at each level of the experiment, and two test results are obtained on each hide, variation between hides will add to the between-laboratory variation, and so increase the reproducibility standard deviation. However, if each laboratory is sent two hides at each level, and two test results are obtained on each hide, the data can be used to estimate the variation between hides and to calculate a value for the reproducibility standard deviation of the test method from which the variation between-hides has been removed.

5.1.2 Another example of a heterogeneous material is sand (that might be used, for example, for making concrete). This is laid down, by the action of wind or water, in strata that always contain graduations in particle size, so when sand is used the particle size distribution is always of interest. In concrete technology the particle size distribution of sand is measured by sieve testing (e.g. BS 812-103 [1]). To test a sand, a bulk sample is taken from the product, then one or more test portions are produced from the bulk sample. Typically, the bulk sample will be about 10 kg in mass, and the test portions will be about 200 g. Because of the natural variability of the material, there will always be some variability between bulk samples of the same product. Hence, as with leather, if a uniform level experiment is performed in which each laboratory is sent one bulk sample at each level, the variability between the bulk samples will increase the calculated reproducibility standard deviation of the test method, but if laboratories are sent two bulk samples at each level, then values for the reproducibility standard deviation can be calculated that exclude this variation.

5.1.3 The above examples also highlight another characteristic of heterogeneous materials: because of the variability of the material, the specimen or test portion preparation can be an important source of variation. Thus with leather, the process of cutting specimens from a hide can have a large influence on the measured tensile

Haut einen großen Einfluss auf die gemessene Reißfestigkeit haben, und bei Siebanalysen von Sand ist der Prozess der Herstellung der Prüfproben aus den Schüttgutproben üblicherweise die Hauptursache der Schwankungen beim Messverfahren. Wenn Prüfstücke oder Prüfproben für einen Ringversuch in einer Weise vorbereitet werden, die nicht der normalen Praxis entspricht (im Bestreben, identische Proben zu erzeugen), dann werden die durch den Ringversuch gelieferten Werte der Wiederholstandardabweichung und der Vergleichstandardabweichung nicht repräsentativ für die erfahrungsgemäß in der Praxis entstehenden Schwankungen sein. Es gibt Situationen, in denen es wünschenswert sein kann, identische „Stichproben“ durch einen speziellen Prozess zu erzeugen mit dem Ziel, so weit wie möglich die Schwankung des Materials auszuschalten (beispielsweise für eine Leistungsprüfung, oder wenn ein Präzisionsversuch als Teil eines Arbeitsprogramms während der Entwicklung eines Messverfahrens eingesetzt wird). Wenn es aber das Ziel des Präzisionsversuchs ist, die in der Praxis erwartbare Schwankung herauszufinden (beispielsweise wenn Verkäufer und Käufer Proben des gleichen Produkts prüfen), dann ist es bei derjenigen Schwankung, die Folge der Heterogenität des Materials ist, nötig, sie in die Ermittlungen der Präzision des Messverfahrens einzubeziehen.

Sorgfalt sollte auch darauf verwendet werden, sicherzustellen, dass jedes Ermittlungsergebnis bei einer Untersuchung dadurch gewonnen wird, dass das Ermittlungsverfahren unabhängig von anderen Ermittlungen angewendet wird. Das ist nicht der Fall, wenn einige Stufen der Probenvorbereitung für mehrere Proben gemeinsam erfolgen, so dass eine durch die Vorbereitung verursachte systematische Abweichungskomponente oder Abweichung gleichen Einfluss auf die von diesen Proben gewonnenen Ermittlungsergebnisse haben wird.

5.1.4 Der in diesem Abschnitt vorgeschlagene Plan (eines Ringversuchs) an heterogenem Material ergibt für die Schwankung zwischen den Proben eine Information, die mit dem Plan (eines Ringversuchs) für einheitliche Proben, wie er in ISO 5725-2 beschrieben ist, nicht gewonnen werden kann. Unausweichlich sind mit der Gewinnung zusätzlicher Information Kosten verbunden: Der vorgeschlagene Plan (eines Ringversuchs) verlangt mehr zu untersuchende Proben. Diese zusätzliche Information kann wertvoll sein. Bei dem in 5.1.1 besprochenen Leder-Beispiel könnte die Information über die Schwankung zwischen den Häuten für die Entscheidung genutzt werden, wie viele Häute zu untersuchen sind, wenn die Qualität einer Lieferung geprüft wird; oder zu entscheiden zwischen der Untersuchung von mehr Häuten mit weniger Proben je Haut oder der Untersuchung von weniger Häuten mit mehr Proben je Haut. Bei dem in 5.1.2 besprochenen Sand-Beispiel könnte die Information über die Schwankung zwischen den Schüttgutproben für die Entscheidung genutzt werden, ob das Verfahren für die Entnahme der Schüttgutproben zufriedenstellend oder eine Verbesserung nötig ist.

5.1.5 Der in diesem Abschnitt beschriebene Plan (eines Ringversuchs) ist anwendbar auf Versuche mit drei Faktoren, die hierarchisch angeordnet sind: Mit einem Faktor „Labors“ auf der höchsten Stufe der Hierarchie, einem Faktor „Proben innerhalb von Labors“ als nächste Stufe der Hierarchie, und einem Faktor „Ermittlungsergebnisse innerhalb von Proben“ als unterste Stufe der Hierarchie. Ein anderer Fall, den man in der Praxis antreffen kann, hat eine dreistufige Hierarchie mit „Labors“ auf der höchsten Stufe, „Ermittlungsergebnisse innerhalb der Labors“ als die nächste Stufe und „Ermittlungen innerhalb der Ermittlungsergebnisse“ als die unterste Stufe. Dies könnte vorkommen, wenn jedes an einem Ringversuch teilnehmende Labor eine einzelne Probe homogenen Materials erhält und gebeten wird, je Probe zwei (oder vielleicht mehr) Ermittlungen auszuführen, und wenn jede Ermittlung eine Anzahl von Ermittlungen zur Folge hat und das Ermittlungsergebnis als der Mittelwert der Ermittlungen errechnet wird. Die in 5.5, 5.6 und 5.9 angegebenen Formeln können auf in einem solchen Ringversuch gewonnene Ermittlungsergebnisse angewendet werden, aber die Wiederholstandardabweichung und die Vergleichstandardabweichung sind in einer Art und Weise zu errechnen, die sich von der hier gegebenen geringfügig unterscheidet (siehe ANMERKUNG 2 zu 5.5.5). Es ist auch nötig, die Anzahl der Ermittlungen festzulegen, aus denen der Mittelwert zu bilden ist, um das Ermittlungsergebnis zu ergeben, weil dies die Werte der Wiederholstandardabweichung und der Vergleichstandardabweichung beeinflusst.

5.2 Anordnung eines Ringversuchs an heterogenem Material

5.2.1 Die Anordnung eines Ringversuchs an heterogenem Material ist in Tabelle 9 gezeigt.

Jedes der p teilnehmenden Labors wird mit zwei Proben auf q Merkmalsniveaus versorgt und gewinnt zwei Ermittlungsergebnisse an jeder Probe. Demzufolge enthält jede Zelle des Ringversuchs vier Ermittlungsergebnisse (zwei Ermittlungsergebnisse für jede der zwei Proben).

Man kann diesen einfachen Plan (eines Ringversuchs) dadurch verallgemeinern, dass man mehr als zwei Proben je Labor und Merkmalsniveau zulässt, oder mehr als zwei Ermittlungsergebnisse je Probe. Die für den allgemeineren Plan (eines Ringversuchs) nötigen Errechnungen sind erheblich komplizierter als diejenigen für den Fall von zwei Ermittlungsergebnissen je Probe und zwei Proben je Labor und Merkmalsniveau. Allerdings sind die Prinzipien für den allgemeineren Plan (eines Ringversuchs) die gleichen wie für den einfachen Plan. Deshalb werden hier die Errechnungen für den einfachen Plan im Einzelnen dargestellt. Formeln zur Errechnung der Werte der Wiederholstandardabweichung und der Vergleichstandardabweichung sind unten in 5.9 angegeben, und ein Beispiel für ihre Anwendung in 5.10.

strength, and with sieve tests on sand the process of preparing test portions from bulk samples is usually the major source of variability in the test method. If specimens or test portions are prepared for a precision experiment in a way that does not correspond to normal practice (in an attempt to produce identical "samples") then the values of repeatability and reproducibility standard deviations produced by the experiment will not be representative of the variability experienced in practice. There are situations in which it can be desirable to produce identical "samples" by some special process designed to eliminate, as far as possible, the variability of the material (for example, for a proficiency test, or when a precision experiment is used as part of a programme of work during the development of a measurement method). However, when the aim of the precision experiment is to discover the variability that will be experienced in practice (for example, when vendors and purchasers test samples of the same product) then it is necessary for the variability arising as a consequence of the heterogeneity of the material to be included in the measures of the precision of the measurement method.

Care should also be taken to ensure that each test result in an experiment is obtained by carrying out the test procedure independently of other tests. This will not be so if some stages of the specimen preparation are shared by several specimens, so that a bias or deviation introduced by the preparation will have a common influence on the test results derived from these specimens.

5.1.4 The design for heterogeneous materials proposed in this clause yields information about the variability between samples that is not obtainable from the uniform level design described in ISO 5725-2. There is, inevitably, a cost associated with obtaining extra information: the proposed design requires more samples to be tested. This extra information may be valuable. In the leather example discussed in 5.1.1, information about the variability between hides could be used to decide how many hides to test when assessing the quality of a consignment, or to decide between testing more hides with fewer specimens per hide or fewer hides with more specimens per hide. In the sand example discussed in 5.1.2, information about the variability between bulk samples could be used to decide if the procedure for taking bulk samples is satisfactory or in need of improvement.

5.1.5 The design described in this clause is applicable to experiments involving three factors arranged in a hierarchy: with a factor "laboratories" at the highest level in the hierarchy, a factor "samples within laboratories" as the next level in the hierarchy, and a factor "test results within samples" as the lowest level of the hierarchy. Another case that can be encountered in practice is of a three-factor hierarchy with "laboratories" at the highest level, "test results within laboratories" as the next level, and "determinations within test results" as the lowest level. This would arise if the participating laboratories in a precision experiment were each sent a single sample of a homogeneous material, were asked to carry out two (or perhaps more) tests per sample, and if each test involved a number of determinations and the test result is calculated as the average of the determinations. The formulae given in 5.5, 5.6 and 5.9 may be applied to data obtained in such an experiment, but the repeatability and reproducibility standard deviations have to be calculated in a slightly different manner to that given here (see NOTE 2 to 5.5.5). It is also necessary to specify the number of determinations that are to be averaged to give a test result, because this affects the values of the repeatability and reproducibility standard deviations.

5.2 Layout of the design for a heterogeneous material

5.2.1 The layout of the design for a heterogeneous material is shown in table 9.

The p participating laboratories are each provided with two samples at q levels, and obtain two test results on each sample. Thus each cell in the experiment contains four test results (two test results for each of two samples).

It is possible to generalize this simple design, by allowing for more than two samples per laboratory per level, or for more than two test results per sample. The calculations required by the more general design are much more complicated than those required by the case with two test results per sample and two samples per laboratory per level. However, the principles of the more general design are the same as for the simple design, so the calculations will be set out in detail here for the simple design. Formulae for calculating values for the repeatability and reproducibility standard deviations for the general design are given below in 5.9, and an example of their application in 5.10.

5.2.2 The data from the design for a heterogeneous material are represented by:

$$y_{ijk}$$