

Präzision von Meßverfahren
 Ermittlung der Wiederhol- und Vergleichpräzision von
 festgelegten Meßverfahren durch Ringversuche
 Identisch mit ISO 5725 Ausgabe 1986

DIN
ISO 5725

Precision of test methods; Determination of repeatability and reproducibility
 for a Standard Test Method by inter-laboratory tests;
 Identical with ISO 5725 edition 1986

Ersatz für Ausgabe 11.81

Fidélité des méthodes d'essai – Détermination de la répétabilité et de la
 reproductibilité d'une méthode d'essai normalisée par essais interlaboratoires;
 Identique à ISO 5725 édition 1986

Die Internationale Norm ISO 5725, 2. Ausgabe 1986, „Precision of test methods – Determination of repeatability and reproducibility for a standard test method by interlaboratory tests,“ ist unverändert in diese Deutsche Norm übernommen worden.

Nationales Vorwort

Die Internationale Norm ISO 5725 wurde vom Technischen Komitee 69 „Anwendung statistischer Methoden“ der International Organization for Standardization (ISO) erarbeitet. Die deutsche Fassung ist vom Ausschuß Qualitätssicherung und angewandte Statistik (AQS) im DIN erstellt worden.

Die Norm beschreibt die Ermittlung von Maßen für die Wiederholpräzision und die Vergleichpräzision eines festgelegten Meßverfahrens mittels Ringversuchen. Nach DIN 55 350 Teil 13 ist die **Präzision** die qualitative Bezeichnung für das Ausmaß der Übereinstimmung zwischen Ergebnissen, wie sie bei wiederholter Anwendung eines festgelegten Ermittlungsverfahrens gewonnen werden. Als Maße für die Präzision eines festgelegten Meßverfahrens werden im allgemeinen die Standardabweichungen der unter Wiederholbedingungen und unter Vergleichbedingungen gewonnenen Meßwerte verwendet. Aus Wiederhol- und Vergleichstandardabweichung als allgemein bedeutsamen Präzisionsmaßen lassen sich die hier im Vorder-

Fortsetzung Seite 2 bis 44

Zitierte Normen und andere Unterlagen

DIN 55 350 Teil 13 Begriffe der Qualitätssicherung und Statistik; Begriffe zur Genauigkeit von Ermittlungsverfahren und Ermittlungsergebnissen

Weitere zitierte Normen und andere Unterlagen siehe „Schrifttum“

Frühere Ausgaben

DIN ISO 5725: 11.81

Änderungen

Gegenüber der Ausgabe November 1981 wurden folgende Änderungen vorgenommen:

- Die Benennungen wurden der internationalen und der deutschen (DIN 55 350 Teil 13) Normenentwicklung angepaßt, wodurch sich auch der geänderte Titel ergab.
- Die Beispiele wurden weitgehend geändert.
- Die vorliegende 2. Ausgabe dieser Norm eröffnet jetzt auch im eigentlichen Normtext die Möglichkeit, anstelle der Wiederholgrenze r und der Vergleichgrenze R die Schätzwerte s_r für die Wiederholstandardabweichung σ_r und s_R für die Vergleichstandardabweichung σ_R zu verwenden.
- Redaktionelle Änderungen.

Internationale Patentklassifikation

G 01 N

Ausschuß Qualitätssicherung und Angewandte Statistik (AQS) im DIN Deutsches Institut für Normung e.V.

Normenausschuß Materialprüfung (NMP) im DIN



Documents

Document Sharing Hub

grund stehenden und besonders in der Materialprüfung eingeführten spezielleren Maße Wiederholgrenze r und Vergleichsgrenze R als „kritische Differenzbeträge“ berechnen. Sie sind die Beträge, unter denen oder höchstens gleich denen die Beträge der Differenzen zweier Einzelergebnisse (gemessen unter Wiederhol- bzw. Vergleichbedingungen) mit einer vorgegebenen Wahrscheinlichkeit (wenn nichts anderes vereinbart mit 95 %) erwartet werden können. Sie unterscheiden sich durch einen angebbaren – von der zugrunde gelegten Wahrscheinlichkeit abhängigen – Faktor von den entsprechenden Standardabweichungen.*)

Diese Norm ist nur auf quantitative Merkmale anwendbar (siehe Abschnitt 3.2).

Dem Anwender der Norm (z. B. einem Normenausschuß) ist freigestellt, ob er als **Präzisionsmaße** eines festgelegten Meßverfahrens

- **Wiederholstandardabweichung und Vergleichstandardabweichung** oder
 - **Wiederholgrenze und Vergleichsgrenze**
- angibt.

Es ist zu beachten, daß nicht jeder für ein bestimmtes Meßverfahren durchgeführte Ringversuch zu denselben Werten der Präzisionsmaße führt, weil seine Ergebnisse Zufallsergebnisse sind, abgesehen davon, daß unterschiedlich viele Labors beteiligt und die Anzahl der Ermittlungen im Labor verschieden sein kann. Die ermittelten Werte der Präzisionsmaße stellen also Schätzwerte dar. Sie können jedoch branchenspezifisch als verbindliche Grenzwerte eingeführt werden, falls es (z. B. aus juristischen Gründen) erforderlich ist.

Die Verfahren dieser Norm ermöglichen keine Aussage über die **Richtigkeit** (siehe DIN 55 350 Teil 13) eines **Meßverfahrens**. Die Richtigkeit wird quantitativ gekennzeichnet durch die Differenz von Erwartungswert (mittleres Ergebnis bei unablässig wiederholter Anwendung des vorgegebenen Meßverfahrens) und wahren oder richtigem Wert des untersuchten Merkmals (systematische Abweichung, bias) (siehe auch Abschnitt 5.2.2). Präzisions- und Richtigkeitsangaben beschreiben gemeinsam die Genauigkeit (siehe DIN 55 350 Teil 13) des Meßverfahrens.

Bei der bereits begonnenen grundsätzlichen Revision von ISO 5725 soll auch der Aspekt der Richtigkeit von Meßergebnissen mitbehandelt werden. Ferner wird angestrebt, als alleinige Maße der Präzision die Wiederhol- und die Vergleichstandardabweichung zu verwenden.

Die Revision sieht auch vor, die Anwendung der Präzisionsmaße mitzubehandeln.

In der vorliegenden Norm werden Begriffe wie folgt benutzt:

- „**Ringversuch**“ als Versuch unter Beteiligung mehrerer Labors (z. B. zur Bestimmung der Präzision eines Meßverfahrens wie in dieser Norm);
- „**Labor**“ auch im Sinne von Untersuchungsort, Prüfstelle usw.

Im übrigen wird auf die Normen über Begriffe der Qualitätssicherung und der Statistik der Normenreihen mit den Hauptnummern DIN 55 350 und DIN 13 303 hingewiesen (international siehe ISO 3534 [1]).

Die vorliegende Norm stellt eine Rahmennorm für viele Fachgebiete dar. Sie bedarf in speziellen Anwendungsfällen ergänzender Festlegungen, insbesondere über die Anzahl der am Ringversuch zu beteiligenden Labors und die Anzahl der Merkmalsniveaus und der Mehrfachmessungen. Branchenspezifisch kann eine feste Anzahl oder eine Mindestanzahl von Labors, Merkmalsniveaus und von Mehrfachmessungen pro Zelle vorgegeben werden.

*) Ist die Wiederholstandardabweichung σ_r bekannt, dann ergibt sich daraus die Wiederholgrenze r als kritischer Differenzbetrag zur Wahrscheinlichkeit $P=95\%$ mit dem Quantil $u_{97,5\%}=1,96$ der standardisierten Normalverteilung zu $r = 1,96 \sqrt{2} \sigma_r = 2,77 \cdot \sigma_r$. In dieser Norm wird jedoch zur Berechnung von r das Quantil 2,0 anstelle von 1,96 benutzt, so daß sich $r = 2,0 \sqrt{2} \sigma_r = 2,8 \sigma_r$ ergibt (näheres dazu siehe auch Abschnitt 5.5). Dieses ist eine Abrundung des Wertes $2,83 \sigma_r$, der in der vorhergehenden Ausgabe dieser Norm verwendet wurde. Entsprechendes gilt für die Vergleichsgrenze R .

Deutsche Übersetzung

Präzision von Meßverfahren

Ermittlung der Wiederhol- und Vergleichpräzision von festgelegten Meßverfahren durch Ringversuche

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 Mitgliedskörperschaft, 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 von den Technischen Komitees verabschiedeten Entwürfe zu Internationalen Normen werden den Mitgliedskörperschaften zunächst zur Annahme vorgelegt, bevor sie vom Rat der ISO als Internationale Normen bestätigt werden.

Sie werden nach den Verfahrensregeln der ISO angenommen, wenn mindestens 75 % der abstimmenden Mitgliedskörperschaften zugestimmt haben.

Die Internationale Norm ISO 5725 wurde vom Technischen Komitee ISO/ TC 69 „Anwendung statistischer Methoden“ erarbeitet.

Diese zweite Ausgabe hebt die erste Ausgabe (ISO 5725 - 1981) auf und ersetzt sie; sie stellt eine geringfügige Revision dar.

Die Anwender werden darauf hingewiesen, daß alle Internationalen Normen von Zeit zu Zeit überarbeitet werden. Ein in dieser Norm enthaltener Verweis auf eine andere Internationale Norm bezieht sich, sofern nichts anderes angegeben ist, auf die neueste Ausgabe der zitierten Norm.

Inhalt

	Seite		Seite
0 Einleitung	4	15 Aufstellung einer funktionalen Beziehung zwischen r (oder R) und m	23
1 Zweck	4	16 Schrittweise statistische Analyse	25
2 Anwendungsbereich	4	17 Bericht an die Fachgruppe und die Entscheidungen, die von ihr getroffen werden müssen ...	28
Hauptabschnitt eins: Allgemeine Grundsätze	5	Hauptabschnitt vier: Verwendung der Präzisionsdaten	29
3 Quantitative Definitionen von Wiederholpräzision und Vergleichpräzision eines festgelegten Meßverfahrens	5	18 Veröffentlichung von Werten der Wiederholgrenze und Vergleichgrenze	29
4 Praktische Folgerungen aus den Definitionen	7	19 Weitere kritische Differenzbeträge, die aus r und R ableitbar sind	30
5 Statistisches Modell	8	20 Praktische Anwendungen	31
6 Die Planung eines Ringversuchs zur Ermittlung der Präzision	9	Hauptabschnitt fünf: Beispiele	31
7 Analyse der Daten	9	21 Einführung	31
Hauptabschnitt zwei: Organisation eines Ringversuchs zur Ermittlung der Präzision	10	22 Versuch mit einheitlichem Merkmalsniveau ohne das Fehlen von Daten und ohne Ausreißer	31
8 Personalerfordernisse	10	23 Versuch mit einheitlichem Merkmalsniveau mit dem Fehlen von Daten	33
9 Aufgaben und Probleme	10	24 Versuch mit einheitlichem Merkmalsniveau mit Ausreißern	37
10 Bemerkungen zu den Abschnitten 8 und 9	11	25 Versuch mit einem einzigen zweigeteilten Merkmalsniveau	40
Hauptabschnitt drei: Statistische Analyse der Ergebnisse eines Ringversuchs	14	Anhang	
11 Vorbemerkungen	14	A Kritische Werte für den Cochran-Test	42
12 Cochran-Test	17	B Kritische Werte für den Dixon-Test	43
13 Dixon-Test	18	C Symbole und Indizes	44
14 Berechnung des Merkmalsniveaus m , der Wiederholgrenze r und der Vergleichgrenze R	18	Schrifttum	44

0 Einleitung

0.1 Messungen, die an voraussichtlich identischem Material (siehe Abschnitt 4.2) unter voraussichtlich identischen Umständen durchgeführt werden, ergeben im allgemeinen nicht dieselben Resultate. Dies ist auf die unvermeidlichen zufälligen Ergebnisabweichungen zurückzuführen, die jedem Meßverfahren innewohnen; denn nicht alle Faktoren, die das Meßergebnis einer Messung beeinflussen können, kann man vollständig beherrschen. Bei der praktischen Interpretation der Meßdaten müssen diese Zufallsstreuungen berücksichtigt werden. So mag zum Beispiel die Differenz zwischen einem Meßergebnis und einem festgelegten Wert im Bereich der unvermeidbaren zufälligen Ergebnisabweichungen liegen. In einem solchen Fall kann man nicht von einem echten Unterschied gegenüber dem vorgegebenen Wert sprechen. Ebenso wenig wird die Differenz zwischen den Meßergebnissen aus der Messung zweier Materialstichproben einen echten Unterschied anzeigen, wenn diese Differenz auf die dem Meßverfahren innewohnenden Zufallsstreuungen zurückgeführt werden kann.

0.2 Viele verschiedene Faktoren (abgesehen von Streuungen zwischen vermeintlich identischen Exemplaren) können zu den Zufallsstreuungen eines Meßverfahrens beitragen. Zu diesen Faktoren gehören die folgenden:

- a) der Bearbeiter,
- b) die verwendeten Instrumente und Geräte,
- c) die Kalibrierung der Geräte,
- d) die Umgebung (Temperatur, Feuchtigkeit, Luftverschmutzung, usw.).

Die Streuungen zwischen Messungen, die von verschiedenen Bearbeitern und/oder mit verschiedener Geräteausrüstung ausgeführt werden, werden gewöhnlich größer sein als zwischen Messungen, die von einem einzigen Bearbeiter mit derselben Geräteausrüstung ausgeführt werden.

0.3 Präzision ist ein allgemeiner Ausdruck für die Streuung zwischen Ergebnissen wiederholter Messungen. Zwei Arten der Präzision, die Wiederholpräzision und die Vergleichpräzision, haben sich als notwendig und in vielen praktischen Fällen als ausreichend erwiesen, um die Streuung der Ergebnisse eines Meßverfahrens zu beschreiben. Die Wiederholpräzision bezieht sich auf Messungen, die unter möglichst konstanten Bedingungen ausgeführt werden, d. h. in kurzen Zeitabständen (siehe Abschnitt 4.3) am selben Ort (im selben Labor) von demselben Bearbeiter mit derselben Geräteausrüstung. Andererseits bezieht sich die Vergleichpräzision auf Messungen, die unter weithin veränderlichen Bedingungen ausgeführt werden, d. h. in verschiedenen Labors mit verschiedenen Bearbeitern und unterschiedlicher Geräteausrüstung. Unter Wiederholbedingungen werden die Faktoren a) bis d), die in Abschnitt 0.2 aufgeführt sind, als konstant angenommen und tragen nicht zur Streuung der Ergebnisse bei, während unter Vergleichbedingungen diese Faktoren variieren und die Streuung der Meßergebnisse beeinflussen. Also sind die Wiederholpräzision und die Vergleichpräzision zwei Extreme; die erste erfaßt das Minimum und die zweite das Maximum der Streuung der Ergebnisse. Zwischen diesen Extremsituationen sind Zwischenstufen der Streuung denkbar, wie z. B. bei Wiederholung der Messungen innerhalb eines Labors in größeren Zeitabständen oder durch verschiedene Bearbeiter oder unter dem Einfluß einer Neukalibrierung; diese Zwischenstufen sind jedoch nicht in dieser Norm behandelt. Wenn in einer besonderen Situation eine Zwischenstufe nötig sein sollte, muß diese von einer zuständigen Stelle im Zusammenhang mit den Umständen, unter denen sie gilt, und dem Verfahren, nach dem sie bestimmt werden soll, klar definiert werden.

0.4 Die in dieser Internationalen Norm verwendeten Definitionen sind in Abschnitt 3 enthalten und Symbole und Indizes sind in Anhang C angegeben.

Eine Bibliographie der Veröffentlichungen, auf die in dieser Internationalen Norm Bezug genommen wird, ist im Anhang angegeben.

1 Zweck

Diese Internationale Norm führt praktische Definitionen der Wiederholgrenze und Vergleichgrenze ein, die sich für eine numerische Schätzung durch einen Versuch eignen (siehe Abschnitt 3). Sie liefert kein Maß für die Abweichungen, die beim Schätzen der Werte von r und R auftreten.

Diese Internationale Norm führt Grundlagen für den Entwurf, die Organisation und die Analyse von Versuchen ein, die für das Schätzen von r und R vorgesehen sind (siehe Abschnitte 6 bis 17). Diesbezügliche Versuche werden im folgenden als Ringversuche zur Ermittlung der Präzision bezeichnet. Nur der einfachste Versuchstyp für die Schätzung von r und R ist betrachtet, welcher aus Versuchen an identischen Meßobjekten oder an Proben identischen Materials in verschiedenen Labors besteht.

Diese Internationale Norm bietet Regeln für die Deutung und Anwendung dieser Schätzwerte von r und R in der Praxis (siehe Abschnitte 18 bis 20).

Diese Internationale Norm beschäftigt sich nicht mit der Ermittlung der Richtigkeit des Meßverfahrens, deren Maß die Differenz zwischen dem Gesamt-Mittelwert und dem wahren oder dem richtigen Wert ist.

2 Anwendungsbereich

Diese Internationale Norm bezieht sich ausschließlich auf Meßverfahren, welche eine einzige Zahl als Meßergebnis liefern, wobei diese ein Rechenergebnis aus einer Menge von Beobachtungen sein kann.

Das Wesentliche der Ermittlung der Präzisionsmaße ist, daß sie die Fähigkeit eines Meßverfahrens beurteilen, eine gegebene Messung zu wiederholen. Das impliziert, daß genau dieselbe Meßgröße in derselben Art und Weise gemessen wird.

Damit diese Messungen in derselben Art und Weise ausgeführt werden, muß das Meßverfahren festgelegt und in verschiedenen Labors in Gebrauch sein. Alle Versuche, die einen Teil des Ringversuchs bilden, müssen in Übereinstimmung mit dieser Festlegung ausgeführt werden.

Im Idealfall sollten die verschiedenen Messungen unter Benutzung desselben Meßobjekts ausgeführt werden. Leider sind viele Messungen zerstörend (chemische Prüfungen, Festigkeitsprüfungen an Materialien), so daß dasselbe Objekt nicht für weitere Messungen zur Verfügung steht. Unter solchen Umständen müssen verschiedene Objekte benutzt werden, aber, um dem obigen Grundsatz zu genügen, muß jede Anstrengung unternommen werden, um sicherzustellen, daß die Objekte möglichst annähernd identisch sind. Es ist darauf zu achten, daß sie nicht nur identisch sind, wenn die Proben aufbereitet werden, sondern daß sie es auch zum Zeitpunkt der Messung sein sollten.

Aufgrund obiger Grundsätze sollte die Präzision nicht unter Benutzung von Proben ermittelt werden, von denen man weiß oder von denen vermutet wird, daß sie nicht identisch sind. Folglich sollten die Objekte als ähnliche Unterproben einer Hauptprobe behandelt werden und dürfen niemals aus verschiedenen Losen oder Sendungen entnommen werden. Diese Punkte werden weiter in Abschnitt 4.2 behandelt.

Wenn in der Praxis zerstörende Versuche mit einbezogen sind, muß der Beitrag zu der Streuung der Meßergebnisse, der von Unterschieden zwischen den Versuchsobjekten, an denen die Messungen ausgeführt werden, herrührt, entweder

vernachlässigbar verglichen mit der Streuung des Meßverfahrens selbst sein oder anderenfalls einen dem Meßverfahren innewohnenden Teil, also eine Komponente der Präzision (siehe Abschnitt 4.2) bilden.

Das statistische Modell in Abschnitt 5 wird als geeignete Basis für die Deutung und Analyse der Meßergebnisse angesehen, die durch einen Ringversuch zur Ermittlung der Präzision gemäß obigen Grundsätzen gegeben sind.

Hauptabschnitt eins: Allgemeine Grundsätze

3 Quantitative Definitionen von Wiederholpräzision und Vergleichpräzision eines festgelegten Meßverfahrens

3.1 Für praktische Zwecke werden quantitative Definitionen benötigt; die folgenden stimmen mit ISO 3534 [1] überein¹⁾:

3.1.1 Beobachtungswert: Als Ergebnis einer Beobachtung festgestellter Merkmalswert.

Anmerkung: Es ist ein einzelner Wert erhalten aus einer einzelnen Beobachtung

3.1.2 Ermittlungsergebnis: Durch die Anwendung eines festgelegten Ermittlungsverfahrens festgestellter Merkmalswert.

Anmerkung: Das Ermittlungsverfahren kann festlegen, daß eine Anzahl von individuellen Beobachtungen anzustellen und ihr Mittelwert als Ergebnis zu melden ist. Sie kann auch fordern, daß Norm-Korrekturen angewendet werden, wie z. B. die Korrektur eines Gasvolumens bezogen auf Norm-Temperatur und -Druck. Also kann ein einzelnes Ermittlungsergebnis ein aus mehreren Beobachtungswerten berechnetes Ergebnis sein.

3.1.3 Merkmalsniveau: Der Mittelwert der Ermittlungsergebnisse aller Labors für ein bestimmtes untersuchtes Material oder Objekt.

3.1.4 Zelle: Die Ermittlungsergebnisse eines Niveaus erhalten von einem Labor.

3.1.5 Präzision: Das Ausmaß der Übereinstimmung zwischen voneinander unabhängigen Ermittlungsergebnissen erhalten unter festgesetzten Bedingungen.

Anmerkung 1: Die Präzision hängt nur von der Verteilung der zufälligen Abweichungen ab und hat keine Beziehung zu dem wahren, dem richtigen oder einem festgelegten Wert.

Anmerkung 2: Wiederholpräzision und Vergleichpräzision sind Erscheinungsformen der Präzision.

3.1.6 Wiederholpräzision: Das Ausmaß der Übereinstimmung zwischen voneinander unabhängigen unter Wiederholbedingungen erhaltenen Ermittlungsergebnissen.

3.1.7 Wiederholbedingungen: Bedingungen, unter denen voneinander unabhängige Ermittlungsergebnisse mit demselben Verfahren an identischen Objekten am selben Ort (in demselben Labor) durch denselben Bearbeiter mit derselben Geräteausrüstung in kurzen Zeitabständen erhalten werden.

3.1.8 Wiederholstandardabweichung: Standardabweichung der Ermittlungsergebnisse unter Wiederholbedingungen. Sie ist ein Parameter der Streuung der Verteilung der Ermittlungsergebnisse unter Wiederholbedingungen.

Anmerkung: Ähnlich könnten die Wiederholvarianz und der Wiederholvariationskoeffizient definiert und als Parameter der Streuung der Ermittlungsergebnisse unter Wiederholbedingungen benutzt werden.

3.1.9 Wiederholgrenze, r : Betrag, unter dem der Absolutwert der Differenz zwischen zwei einzelnen unter Wiederholbedingungen gewonnenen Ermittlungsergebnissen mit einer Wahrscheinlichkeit von 95 % erwartet werden kann.²⁾

Anmerkung: Zur Abkürzung wird im folgenden die Wiederholgrenze als r bezeichnet, wenn der Zusammenhang eindeutig ist.

3.1.10 Kritischer Wiederholdifferenzbetrag: Betrag, unter dem der Absolutwert der Differenz zwischen zwei einzelnen unter Wiederholbedingungen gewonnenen Ermittlungsergebnissen mit einer vorgegebenen Wahrscheinlichkeit erwartet werden kann.²⁾

Anmerkung 1: Die vorgegebene Wahrscheinlichkeit ist dem Symbol r der Wiederholgrenze als Index anzuhängen, z. B. ist r_{90} % der kritische Wiederholdifferenzbetrag für eine Wahrscheinlichkeit von 90 %.

Anmerkung 2: Die Wiederholgrenze r ist der kritische Wiederholdifferenzbetrag für eine Wahrscheinlichkeit von 95 %, unter Fortlassen des Index in diesem speziellen Fall.

3.1.11 Vergleichpräzision: Das Ausmaß der Übereinstimmung zwischen unter Vergleichbedingungen erhaltenen Ermittlungsergebnissen.

3.1.12 Vergleichbedingungen: Bedingungen, unter denen Ermittlungsergebnisse mit demselben Verfahren an identischem Material an verschiedenen Orten (in verschiedenen Labors) von verschiedenen Bearbeitern mit verschiedener Geräteausrüstung erhalten werden.

3.1.13 Vergleichstandardabweichung: Die Standardabweichung von unter Vergleichbedingungen erhaltenen Ermittlungsergebnissen. Sie ist ein Parameter für die Streuung der Verteilung der Ermittlungsergebnisse unter Vergleichbedingungen.

Anmerkung: Ähnlich könnten die Vergleichvarianz und der Vergleichvariationskoeffizient definiert und als Parameter der Streuung der Ermittlungsergebnisse unter Vergleichbedingungen benutzt werden.

3.1.14 Vergleichgrenze, R : Betrag, unter dem der Absolutwert der Differenz zwischen zwei einzelnen unter Vergleichbedingungen gewonnenen Ermittlungsergebnissen mit einer Wahrscheinlichkeit von 95 % erwartet werden kann.²⁾

¹⁾ Nationale Fußnote: Diese Norm beschäftigt sich mit der Präzision von Meßverfahren und deren Ergebnissen, den Meßergebnissen. In den Abschnitten 3.1 und 3.2 sind jedoch wegen der Allgemeingültigkeit der dort gegebenen Definitionen und Erläuterungen die umfassenderen Begriffe „Ermittlungsverfahren“ und „Ermittlungsergebnis“ verwendet (siehe DIN 55 350 Teil 13).

²⁾ Nationale Fußnote: In der hier vorliegenden Anwendung ist wegen der Stetigkeit der Merkmale (siehe Abschnitt 3.2) die Formulierung „Betrag unter dem ...“ gleichwertig der Formulierung „Betrag unter oder höchstens gleich dem ...“. Die Definition ist in Übereinstimmung mit DIN 55 350 Teil 13 in dem zweiten Sinn aufzufassen.

Anmerkung: Zur Abkürzung wird im folgenden die Vergleichsgrenze als R bezeichnet, wenn der Zusammenhang eindeutig ist.

3.1.15 Kritischer Vergleichsdifferenzbetrag: Betrag, unter dem der Absolutwert der Differenz zwischen zwei einzelnen unter Vergleichsbedingungen gewonnenen Ermittlungsergebnissen mit einer vorgegebenen Wahrscheinlichkeit erwartet werden kann.²⁾

Anmerkung 1: Die vorgegebene Wahrscheinlichkeit ist dem Symbol R der Vergleichsgrenze als Index anzuhängen, z. B. ist $R_{90\%}$ der kritische Vergleichsdifferenzbetrag für eine Wahrscheinlichkeit von 90%.

Anmerkung 2: Die Vergleichsgrenze R ist der kritische Vergleichsdifferenzbetrag für eine Wahrscheinlichkeit von 95%, unter Fortlassen des Index in diesem speziellen Falle.

3.2 Die im Abschnitt 3.1 gegebenen Definitionen beziehen sich auf Ermittlungsergebnisse, die von einem stetigen Merkmal herrühren. Bei abgerundeten Ermittlungsergebnissen oder solchen, die von einem diskreten Merkmal herrühren, sind r und R jeweils das Minimum, gleich oder unterhalb dessen die absolute Differenz zwischen zwei einzelnen Ermittlungsergebnissen mit einer Wahrscheinlichkeit erwartet wird, die nicht unterhalb eines festgelegten Wertes liegt.

3.3 Die Benennungen Wiederholbarkeit und Vergleichbarkeit waren angenommen worden, weil sie jahrelang in allgemeinem Gebrauch waren.³⁾ Die Symbole r und R sind schon für andere Zwecke in allgemeinem Gebrauch; in ISO 3534 ist r für den Korrelationskoeffizienten empfohlen und R (oder w) für die Spannweite einer einzelnen Beobachtungsreihe. Es sollte jedoch keine Verwirrung herrschen, wenn bei jeder Möglichkeit eines Mißverständnisses die vollständigen Wörter Wiederholgrenze und Vergleichsgrenze verwendet werden, besonders wenn r und R in Normen angeführt werden.

3.4 R und r , wie sie in dieser Norm definiert sind, sind vor allem Kriterien, anhand derer zu entscheiden ist, wie weit eine Differenz zwischen zwei einzelnen Meßergebnissen zufälligen Streuungen zugeschrieben werden kann; eine Differenz größer als r oder R ist auffällig und kann den Schluß rechtfertigen, daß eine systematische Differenz besteht, oder sie kann zu zusätzlichen Untersuchungen führen. In diesem Sinne können r und R als kritische Differenzbeträge bezeichnet werden, die auf ein Paar von Meßergebnissen anzuwenden sind, welche unter Wiederholbedingungen bzw. Vergleichsbedingungen erhalten wurden.

3.4.1 Manchmal ist es erforderlich, die Mittelwerte zweier oder mehrerer Versuche miteinander oder den Mittelwert einer Serie mit einem vorgegebenen Wert zu vergleichen. Kritische Differenzbeträge für solche Zwecke können nach den Abschnitten 19.2.1 bis 19.2.4 von r und R abgeleitet werden.

3.4.2 Wie bereits definiert, sind r und R mit einem Wahrscheinlichkeitsniveau von 95% assoziiert. Manchmal können kritische Differenzbeträge mit einem von 95% abweichenden Wahrscheinlichkeitsniveau bevorzugt werden; diese können nach Abschnitt 19.1.1 berechnet werden. In solchen Fällen sollte das Wahrscheinlichkeitsniveau als Index angehängt werden, um Mißdeutungen zu vermeiden, z. B. $r_{99\%}$ oder $R_{90\%}$.

3.4.3 Die Definitionen in den Abschnitten 3.1.9 und 3.1.14 beziehen sich auf theoretische Konstanten, die in Wirklichkeit unbekannt bleiben. Die Werte für r und R , wie sie

tatsächlich nach der Beschreibung in dieser Internationalen Norm durch einen Ringversuch zur Ermittlung der Präzision bestimmt sind, sind im statistischen Sinn Schätzwerte dieser Konstanten und als solche Abweichungen unterworfen. Konsequenterweise wird das mit r und R assoziierte Wahrscheinlichkeitsniveau nicht genau 95%, sondern nur in der Größenordnung von 95% sein, was ebenso für andere davon abgeleitete kritische Differenzbeträge gilt. Das ist unvermeidlich, verringert aber nicht den praktischen Wert der kritischen Differenzbeträge r und R , da sie vornehmlich als Werkzeuge zur Beurteilung vorgesehen sind, ob die Differenz zwischen Meßergebnissen den dem Meßverfahren innewohnenden zufälligen Unsicherheiten zugeschrieben werden kann oder nicht. Differenzen größer als r oder R sind verdächtig.

3.5 Wenn die Erfordernisse dieser Internationalen Norm hinsichtlich der Zahl von Labors, die in den Ringversuch einzubeziehen sind und hinsichtlich der Zahl von Messungen, die jedes Labor ausführen sollte, befolgt sind (siehe Abschnitt 10.1), werden die resultierenden Schätzwerte von r und R ausreichend genau für die Praxis sein, unter der Bedingung, daß die teilnehmenden Labors repräsentativ für alle diejenigen Labors sind, die das festgelegte Verfahren anwenden. Diese hypothetische Gruppe von Labors ist durch Anforderungen ähnlich wie im Abschnitt 10.2 definiert. Wenn es sich zukünftig erweisen sollte, daß diese Bedingung durch den ursprünglichen Ringversuch zur Ermittlung der Präzision nicht erfüllt war oder nicht länger erfüllt ist, kann ein neuer Ringversuch notwendig werden, sofern es nicht möglich ist, r und R neu zu schätzen, um Übereinstimmung mit den geänderten Bedingungen zu erreichen.

3.6 Grundsätzlich kann die Wiederholgrenze r , definiert in Abschnitt 3.1.9, auf jedes Meßverfahren in jedem einzelnen Labor angewendet werden. Eine grundsätzliche Annahme für diese Internationale Norm ist, daß für ein festgelegtes Meßverfahren die Wiederholgrenze zumindest annähernd für alle Labors, die das festgelegte Verfahren anwenden, dieselbe ist, so daß es möglich ist, eine gemeinsame mittlere Wiederholgrenze, anwendbar auf jedes Labor, einzuführen. Jedoch kann jedes Labor durch Ausführung einer Serie von Messungen unter Wiederholbedingungen einen Schätzwert seiner eigenen Wiederholgrenze für das Meßverfahren erlangen und ihn anhand des gemeinsamen Normwertes überprüfen. Ein derartiger Vorgang ist in dieser Internationalen Norm nicht detailliert ausgearbeitet.

3.7 Wenn die Vergleichpräzision als kritischer Differenzbetrag benutzt werden soll, muß das Paar der zu vergleichenden Meßergebnisse von zwei Labors erhalten worden sein, die zufällig aus der gesamten Menge von Labors, welche das festgelegte Meßverfahren anwenden, ausgewählt sind. Wenn Meßergebnisse immer zwischen den zwei selben Labors verglichen worden sind, ist Vorsicht geboten, weil das mit R assoziierte Wahrscheinlichkeitsniveau möglicherweise wegen einer systematischen Differenz zwischen den Ergebnissen dieser zwei bestimmten Labors nicht länger richtig sein könnte. Wenn angenommen wird, daß dieses der Fall sein könnte, sollten die zwei fraglichen Labors einen Ringversuch zwischen sich organisieren, um die Größe dieser systematischen Differenz zu bestimmen.

²⁾ Siehe Seite 5

³⁾ Nationale Fußnote: In dieser Norm werden in Übereinstimmung mit DIN 55 350 Teil 13 nur die Benennungen Wiederholpräzision, Vergleichpräzision sowie Wiederholgrenze und Vergleichsgrenze verwendet.

3.8 Obwohl in diesem Dokument die Wiederholpräzision und die Vergleichpräzision in Form von kritischen Differenzbeträgen behandelt sind, gibt es keinen Grund, die Präzisions-Ergebnisse nicht in Form von Standardabweichungen oder Koeffizienten der Streuung ausdrücken, wenn es für eine besondere Anwendung passender ist.

3.9 Wenn die Werte von r und R einmal bestimmt sind, können sie vielfältig verwendet werden. Beispielsweise können sie dienen:

- für den Beweis, daß eine Versuchstechnik der Norm entspricht (siehe Abschnitt 3.6);
- für das Entwerfen von Verfahren der Qualitätslenkung;
- für das Vergleichen von Meßergebnissen einer Materialcharge mit einer Produktspezifikation;
- für das Entwerfen von Spezifikationen in erster Linie dazu, um sicherzustellen, daß eine Übereinstimmung durch das Meßverfahren beweisbar ist;
- für den Vergleich von Meßergebnissen, die für dieselbe Materialcharge von dem Lieferanten und dem Verbraucher erhalten wurden;
- für die Bewertung der Eignung konkurrierender Meßverfahren.

Bei manchen Anwendungen müssen verschiedene andere Faktoren in die Überlegung miteinbezogen werden, siehe beispielsweise Abschnitt 4.2.6.

4 Praktische Folgerungen aus den Definitionen

4.1 Festgelegtes Meßverfahren

4.1.1 Wie im Abschnitt 2 bestimmt, muß das untersuchte Meßverfahren festgelegt sein. Dies bedeutet, daß es eine Festlegung geben muß, das heißt ein geschriebenes Dokument, das in allen Details festlegt, wie die Messung durchgeführt werden sollte und das am besten eine Beschreibung enthält, wie die Probe zu entnehmen und zu behandeln ist. Bei den Schätzwerten von r und R , die einem solchen Versuch entstammen, sollte jeweils deutlich gemacht werden, daß sie nur gültig sind für Messungen, die gemäß dem festgelegten Verfahren durchgeführt wurden.

4.1.2 Die Existenz einer Festlegung für das Meßverfahren bedeutet auch die Existenz einer Organisation, die für die Erstellung der betrachteten Festlegung verantwortlich ist.

4.1.3 Die Erstellung einer Festlegung eines Meßverfahrens erfordert eine sorgfältige Bewertung des Verfahrens (oder möglicherweise mehrerer alternativer Verfahren) durch Versuche, an denen eine Anzahl von Labors teilnimmt. Ein derartiger Versuch zur Festlegung wird einige vorläufige Informationen hinsichtlich der Werte von r und R liefern. Der wesentliche dem Ringversuch zur Ermittlung von r und R zugrundeliegende Punkt ist, daß gewöhnlich die Zusammenarbeit einer größeren Anzahl von Labors nötig ist, als für einen Versuch zur Festlegung des Meßverfahrens und daß diese Labors aus all denen rekrutiert werden müssen, die die Norm tatsächlich oder wahrscheinlich in ihrer üblichen Arbeit benutzen, und daß sie nicht ausschließlich aus solchen Labors bestehen dürfen, die eine spezielle Erfahrung während des Festlegungsverfahrens erworben haben. Also sollte ein Ringversuch für die Ermittlung von r und R in der Regel nicht organisiert werden, solange die Festlegung für das Meßverfahren noch nicht herausgegeben und in allgemeinem Gebrauch ist. Das bedeutet jedoch nicht, daß irgendwelche Informationen bezüglich der möglichen Werte von r und R , die aus einem Versuch zur Festlegung erhalten werden, wertlos seien, da sie in die Überlegung zur Gestaltung eines Ringversuchs zur Ermittlung der Präzision miteinbezogen werden können.

4.1.4 Ein Ringversuch zur Ermittlung der Präzision kann auch als praktischer Test der Eignung der Festlegung angesehen werden.

Eines der Hauptziele der Festlegung besteht darin, Unterschiede zwischen den Benutzern (Labors) so weit wie möglich zu beseitigen; die aus einem Ringversuch erhaltenen Daten werden zeigen, wie weit dieses Ziel erreicht wurde. Deutliche Unterschiede können zeigen, daß die Festlegung noch nicht ausreichend detailliert ist und möglicherweise verbessert werden kann. In diesem Fall sollte ein Bericht mit der Bitte um weitere Untersuchung an den Ausschuß, der die Festlegung erstellt, gegeben werden (siehe Abschnitte 9.6c), 17.2b) und c) und 17.3).

4.2 Identisches Material

4.2.1 Bei einem Ringversuch zur Ermittlung der Präzision werden Proben eines bestimmten Materials oder Einheiten eines bestimmten Produkts von einem zentralen Punkt aus an eine Anzahl von Labors an verschiedenen Orten, in verschiedenen Ländern oder auch in verschiedenen Kontinenten gesandt. Die Forderung, daß die Messungen in diesen Labors an identischem Material ausgeführt werden sollen, bezieht sich auf den Zeitpunkt, an dem die Messungen tatsächlich stattfinden. Um das zu erreichen, müssen die folgenden zwei verschiedenen Bedingungen erfüllt werden:

- a) Die Proben müssen identisch sein, wenn sie an die Labors abgeschickt werden.
- b) Sie müssen während des Transportes und während der verschiedenen Zeiten, die vor der tatsächlichen Durchführung der Messungen in den teilnehmenden Labors vergehen können, identisch bleiben.

Bei der Organisation eines Ringversuchs zur Ermittlung der Präzision müssen beide Bedingungen sorgfältig beachtet werden.

4.2.2 Eine Flüssigkeit oder ein feines Pulver kann durch Rühren homogenisiert werden, und Proben, die solchen Chargen entnommen werden, können zu dem Zeitpunkt, an dem sie aufbereitet wurden, als identisch angesehen werden. Um sicherzustellen, daß sie bis zu dem Zeitpunkt der Durchführung der Messung identisch bleiben, können zusätzliche Vorkehrungen nötig sein. Besteht das zu messende Material aus einer Mischung von Pulver unterschiedlicher relativer Dichte oder unterschiedlicher Korngröße, so ist Vorsicht geboten, weil durch Schütteln, zum Beispiel während des Transportes, Entmischung auftreten kann. Wenn eine Reaktion mit der Atmosphäre erwartet werden kann, können die Proben in evakuierten oder mit inertem Gas gefüllten Ampullen versiegelt werden. Für verderbliche Materialien, wie z. B. Lebensmittel oder Blutproben, kann es nötig sein, sie den teilnehmenden Labors in tiefgefrorenem Zustand mit detaillierten Auftauanleitungen zu schicken. Jeder Fall muß hinsichtlich seiner Vorteile beurteilt werden.

4.2.3 Wenn die Messungen an einzelnen Meßobjekten durchgeführt werden sollen, die durch das Meßverfahren nicht verändert werden, so könnten die Messungen – zumindest im Prinzip – so ausgeführt werden, daß dieselben Meßobjekte in den verschiedenen Labors verwendet werden. Dies würde jedoch bedingen, daß dieselben Meßobjekte zwischen vielen Labors, die oft weit voneinander entfernt in verschiedenen Ländern oder Kontinenten liegen, zirkulieren müßten, wobei das Risiko eines Verlustes oder einer Veränderung während des Transports beträchtlich wäre.

4.2.4 Wenn die Messungen an festen Materialien durchgeführt werden sollen, die nicht homogenisiert werden können, ist es notwendig, daß die Proben in einer Weise

können (wie etwa Metalle, Gummi oder Textilgewebe) und wenn die Messungen nicht an demselben Meßobjekt wiederholt werden können, stellt die Inhomogenität des Versuchsmaterials eine wesentliche Komponente der Präzision der Messung dar, und die Annahme identischen Materials hat keinen Bestand mehr. Ringversuche können noch ausgeführt werden, aber die Werte für r und R können nur für das spezielle Material gelten und sollten als solche angesehen werden. Ein allgemeiner Gebrauch von r und R wird nur akzeptabel sein, wenn gezeigt werden kann, daß die Werte nicht signifikant zwischen denjenigen Materialien differieren, die zu verschiedenen Zeiten von verschiedenen Herstellern produziert worden sind. Dieser Nachweis würde einen sorgfältigeren Versuch erfordern als er in dieser Internationalen Norm behandelt worden ist.

4.2.5 In den Abschnitten 4.2.1 bis 4.2.4 ist Bezug auf Messungen in verschiedenen Labors genommen worden unter Einbezug des Transports der Meßobjekte zu den Labors. Manche Objekte sind jedoch nicht transportabel, z. B. Öltanks. In diesen Fällen bedingt die Messung durch verschiedene Labors, daß verschiedene Bearbeiter mit ihrer Geräteausrüstung zu dem Ort der Messung geschickt werden. In anderen Fällen kann das gemessene Merkmal veränderlich sein, z. B. die Durchflußmenge eines Flusses, wobei darauf zu achten ist, daß die Messungen unter möglichst gleichen Bedingungen ausgeführt werden. Das Leitprinzip muß immer das Ziel sein, die Fähigkeit zu beurteilen, dieselbe Messung zu wiederholen.

4.2.6 In der Praxis werden r und R sowie andere davon nach den im Abschnitt 19.1.1 und/oder Abschnitt 19.1.2 spezifizierten Verfahren abgeleitete kritische Differenzbeträge häufig benutzt, um Chargen handelsüblichen Materials mit einer Spezifikation oder zwei Chargen miteinander zu vergleichen. Eine Differenz größer als der kritische Differenzbetrag kann dann *unter anderem* durch die handelsübliche Inhomogenität der Materialchargen erklärt werden, sofern es nicht möglich gewesen ist, diesen Mangel an Homogenität in die Bestimmung von r und R einzubeziehen. Die Schwierigkeiten werden in diesen Fällen jedoch dieselben sein, wie im Abschnitt 4.2.4 erwähnt.

4.3 Kurze Zeitintervalle

Nach der Definition aus Abschnitt 3.1.7 müssen Messungen zur Ermittlung der Wiederholpräzision unter konstanten Versuchsbedingungen durchgeführt werden, das heißt, daß während der Dauer der Messungen Faktoren, wie sie im Abschnitt 0.2 benannt sind, konstant sein sollten. Im besonderen sollte die Geräteausrüstung nicht zwischen den Messungen neu kalibriert werden, sofern dieses nicht ein wesentlicher Teil jeder einzelnen Messung ist. In der Praxis sollten Messungen unter Wiederholbedingungen in einem möglichst kurzen Zeitintervall durchgeführt werden, um die Veränderungen in den Faktoren zu minimieren, deren Konstanz nicht immer garantiert werden kann, wie z. B. umweltbedingte Faktoren (siehe Abschnitt 10.4.1c).

5 Statistisches Modell

5.1 Das grundsätzliche Modell

Zur Schätzung der Präzision eines Meßverfahrens ist es nützlich anzunehmen, daß jedes einzelne Meßergebnis, y , die Summe von drei Komponenten ist:

$$y = m + B + e. \quad (1)$$

Dabei bedeutet für das jeweils untersuchte Material:

m Gesamtmittelwert

B Streuung zwischen den Labors

e zufällige, in jedem Versuch auftretende Ergebnisabweichung

Manchmal werden andere Modelle verwendet, aber das obige wird die Mehrheit der praktischen Fälle erfassen (siehe Abschnitt 5.6).

5.2 Gesamtmittelwert m

5.2.1 Der Gesamtmittelwert m des zu messenden Materials wird das Merkmalsniveau des zu messenden Merkmals genannt: Proben verschiedener Reinheit einer Chemikalie oder verschiedener Materialien (z. B. verschiedene Stahlsorten) werden mit verschiedenen Merkmalsniveaus korrespondieren. In vielen technischen Fällen ist das Merkmalsniveau des zu untersuchenden Merkmals ausschließlich durch Meßverfahren definiert, und die Erwähnung eines unabhängigen wahren Wertes ist gegenstandslos. In manchen Fällen kann jedoch das Konzept eines wahren Wertes des zu untersuchenden Merkmals Bestand haben, z. B. die wahre Konzentration einer titrierten Lösung. Das Merkmalsniveau m muß jedoch nicht notwendigerweise mit dem wahren Wert übereinstimmen; wenn eine Differenz ($m - \mu$) existiert, so wird sie systematische Ergebnisabweichung (bias) des Meßverfahrens genannt.

5.2.2 Werden r und R benutzt, um die Differenz zwischen zwei Meßergebnissen zu testen, so ist die systematische Ergebnisabweichung (bias) ohne Einfluß und kann unbeachtet bleiben. Werden sie aber dazu verwendet, Meßergebnisse mit einem vertraglich festgelegten Wert oder einer Festlegung zu vergleichen, so muß man die systematische Ergebnisabweichung (bias) berücksichtigen, wenn sich die vertragliche Spezifizierung auf den wahren Wert μ und nicht auf das Merkmalsniveau m bezieht. Wenn ein wahrer Wert existiert und bekannt ist, so kann die Analyse eines Ringversuchs zur Ermittlung der Präzision die Existenz einer systematischen Ergebnisabweichung (bias) anzeigen (siehe Anmerkung zu Abschnitt 19.2.4).

5.3 Die Komponente B in dem grundsätzlichen Modell

(siehe Abschnitt 5.1)

5.3.1 Die Komponente B wird für jede Serie von Messungen, die unter Wiederholbedingungen durchgeführt wird, als konstant angenommen. Dagegen wird sich diese Komponente in Serien, die unter Vergleichbedingungen durchgeführt werden, wie eine Zufallsgröße verhalten. Die in dieser Internationalen Norm angegebenen Verfahren sind unter der Annahme entwickelt worden, daß diese Zufallsgröße der Ergebnisabweichung annähernd normal verteilt ist, aber in der Praxis sind sie für die meisten Verteilungen anwendbar, sofern diese unimodal sind und sich die kritischen Differenzbeträge auf das 95 %-Niveau beziehen. Die Varianz von B heißt Varianz zwischen den Labors und wird ausgedrückt als:

$$\text{var}(B) = \sigma_L^2$$

σ_L^2 schließt die Streuungen zwischen verschiedenen Bearbeitern und verschiedenen Geräten ein.

5.3.2 Im allgemeinen kann B als die Summe einer zufälligen und einer systematischen Komponente aufgefaßt werden, welche jedoch in dieser Analyse nicht unterschieden sind. Es ist in dieser Internationalen Norm kein Versuch unternommen worden, eine erschöpfende Aufstellung aller Faktoren, die zu B beitragen, zu geben; zu ihnen gehören jedoch verschiedene klimatische Bedingungen, Streuungen der Ausrüstung innerhalb der Herstellungstoleranz und die Techniken, welche die Bearbeiter an verschiedenen Orten gelernt haben.

5.3.3 Wenn zu jeder Zeit wahrscheinlich eine begrenzte Zahl von Labors ein Verfahren benutzt, kann B nur eine

begrenzte Zahl von Werten annehmen, und um für die Praxis von Nutzen zu sein, muß die Vergleichpräzision aus einer Menge von Labors bestimmt werden, die als zufällig ausgewählt aus denjenigen Labors angesehen werden kann, die wahrscheinlich das Verfahren benutzen. Aufmerksamkeit ist nötig, wenn die zu vergleichenden Meßergebnisse immer von denselben Labors erstellt werden. Ein Beispiel für die Art von Problemen, die in diesem Fall auftreten können, ist in Abschnitt 23 angeführt, wo gezeigt wird, daß die Ergebnisse von zwei der Labors (11 und 1) ständig um etwa 4 °C differieren. Wenn regelmäßig nur zwei Labors befaßt sind, sollte die Vergleichpräzision als solche nicht verwendet werden. Stattdessen sollte ein Versuch zur Zusammenarbeit der beiden Labors ausgeführt werden, um ihre relative systematische Ergebnisabweichung und also ihre spezifische Vergleichpräzision zu bestimmen.

5.4 Die Komponente e in dem grundsätzlichen Modell

(siehe Abschnitt 5.1)

5.4.1 Die Komponente e repräsentiert eine zufällige Ergebnisabweichung, die bei jedem einzelnen Meßergebnis auftritt. Die in dieser Internationalen Norm angegebenen Verfahren sind unter der Annahme entwickelt worden, daß diese Komponente annähernd einer Normalverteilung folgt, aber in der Praxis sind sie für die meisten Verteilungen anwendbar, sofern diese unimodal sind, und sich die kritischen Differenzbeträge auf das 95%-Signifikanzniveau beziehen. Die Varianz von e innerhalb eines einzelnen Labors wird laborinterne Varianz genannt und wird ausgedrückt als

$$\text{var}(e) = \sigma_w^2$$

5.4.2 Man darf erwarten, daß σ_w^2 aufgrund von Unterschieden – wie z. B. in den Fähigkeiten der Bearbeiter – in verschiedenen Labors unterschiedliche Werte hat. In dieser Internationalen Norm wird jedoch davon ausgegangen, daß solche Differenzen zwischen den Labors bei einem geeignet festgelegten Meßverfahren klein sein sollten und daß es gerechtfertigt ist, einen gemeinsamen Wert für die laborinterne Varianz anzusetzen, der für alle Labors gilt, die das Meßverfahren benutzen. Dieser gemeinsame Wert, der ein Mittelwert aus den laborinternen Varianzen aller Labors ist, die an dem Ringversuch zur Ermittlung der Präzision teilnehmen, wird die Wiederholvarianz genannt und ausgedrückt als

$$\overline{\text{var}}(e) = \sigma_r^2$$

5.5 Die Beziehung zwischen dem grundsätzlichen Modell und r und R

Wird das grundsätzliche Modell (siehe Abschnitt 5.1) verwendet, hängt die Wiederholgrenze r nur von der Wiederholvarianz ab (siehe Abschnitt 5.4.2), während die Vergleichsgrenze R von der Summe der Wiederholvarianz und der Varianz zwischen den Labors (siehe Abschnitt 5.3.1) abhängt. Es gibt also zwei Größen:

die Wiederholstandardabweichung

$$\sigma_r = \sqrt{\text{var}(e)}$$

und die Vergleichstandardabweichung

$$\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2}$$

sowie daraus

$$\text{Wiederholgrenze } r = f \cdot \sqrt{2} \cdot \sigma_r \quad (2)$$

$$\text{Vergleichsgrenze } R = f \cdot \sqrt{2} \cdot \sigma_R \quad (3)$$

wobei

der Faktor $\sqrt{2}$ auf die Tatsache zurückzuführen ist, daß sich r und R auf die Differenz von zwei einzelnen Meßergebnissen beziehen;

f ein Faktor ist, dessen Wert sowohl von der Anzahl der Meßergebnisse, die für die Schätzung der beiden der Varianzen verfügbar sind, abhängt als auch von der Gestalt der Verteilung der Komponenten B und e (siehe Abschnitt 5.1).

Sind diese Verteilungen jedoch näherungsweise normal und ist die Anzahl der Meßergebnisse nicht zu klein, so wird sich der Faktor f niemals sehr von dem Wert 2 unterscheiden, und deshalb wird in dieser Internationalen Norm empfohlen, letzteren Wert zu verwenden, wobei der Wert $f\sqrt{2}$ zu 2,8 gerundet wird. (Die Berücksichtigung von unterschiedlichen Werten für f würde zu beträchtlichen Komplikationen führen, ohne daß sie wirksam zur praktischen Bedeutung von r und R beitragen würde.)

Da in der Praxis die wahren Werte der Wiederholstandardabweichung und der Vergleichstandardabweichung nicht bekannt sind, werden sie durch ihre Schätzwerte s ersetzt.

Daraus folgt: $r = 2,8 s_r$ (4)

$R = 2,8 s_R$ (5)

5.6 Die Eignung des grundsätzlichen Modells

Es ist klar, daß das Modell nach Abschnitt 5.1 eine Näherung ist, von der aufgrund ausgiebiger Erfahrung bekannt ist, daß sie als Arbeitshypothese für die Versuchsplanung und die Analyse von Daten den praktischen Erfordernissen Genüge leistet. Für die Zwecke dieser Internationalen Norm ist das Modell eine annehmbare Näherung, solange die experimentellen Erfordernisse des Hauptabschnitts zwei erfüllt sind und die statistischen Tests des Hauptabschnitts drei keine signifikanten Ergebnisse liefern, die seine Nichteignung anzeigen. Welche Maßnahmen getroffen werden sollten, wenn die statistischen Tests darauf hindeuten, daß das Modell unangemessen ist, wird in den Abschnitten 16 und 17 diskutiert.

6 Die Planung eines Ringversuchs zur Ermittlung der Präzision

6.1 Ein Plan sieht wie folgt aus: aus q Materialchargen, die q verschiedene Merkmalsniveaus der zu messenden Eigenschaften repräsentieren, werden Proben an p verschiedene Labors gesandt, die alle auf jedem Merkmalsniveau n Messungen unter Wiederholbedingungen durchführen. Diese n Messungen werden also an identischem Material durchgeführt. Dieser Versuchstyp wird Versuch mit einheitlichem Merkmalsniveau genannt.

6.2 Ein alternativer Plan, der unter gewissen Umständen vorgezogen wird (siehe Abschnitt 10.4.2), wird Versuch mit zweigeteiltem Merkmalsniveau genannt; jedes Merkmalsniveau wird in zwei Unterniveaus a und b aufgespalten, die sich nur wenig voneinander unterscheiden. Jedes Labor erhält eine Probe von jedem dieser Unterniveaus zum Messen.

6.3 Vollständige Beispiele werden für beide Alternativen in den Fallstudien im Hauptabschnitt 5 angeführt. Praktische Erwägungen zur Planung und Durchführung werden im Hauptabschnitt 2 angestellt.

7 Analyse der Daten

7.1 Die Analyse der Daten, die sich aus einem Ringversuch zur Ermittlung der Präzision ergeben, sollte als ein statistisches Problem angesehen werden, welches einem

Fachmann der Statistik zu übertragen ist (siehe Abschnitte 8.2 und 9.2). Sie schließt drei aufeinanderfolgende Stufen ein:

- kritische Überprüfung der Daten auf Ausreißer und andere Unregelmäßigkeiten und Entscheidung darüber, wie sie behandelt werden sollen; Überprüfung der Eignung des Modells,
- Berechnung von vorläufigen Werten für r und R , getrennt für jedes Merkmalsniveau,
- Festlegung der endgültigen Werte für diese Parameter einschließlich einer Beziehung zwischen r , R und m , wenn aus der Analyse hervorgeht, daß r oder R vom Merkmalsniveau m abhängen.

7.2 Nach den detaillierten Ausführungen in den Abschnitten 14.7 bis 14.10 berechnet die Analyse eines Ringversuchs

zunächst für jedes Merkmalsniveau getrennt Schätzwerte der Wiederholvarianz s_r^2 , der Varianz zwischen den Labors s_L^2 und der Vergleichvarianz s_R^2 , nach den Definitionen in den Abschnitten 5.3, 5.4 und 5.5, und anschließend die Wiederholgrenze r und die Vergleichgrenze R .

7.3 Zur Analyse, insbesondere zu Stufe a), beschrieben in Abschnitt 7.2, gehört die systematische Anwendung statistischer Tests. Aus der Literatur ist eine Vielzahl von statistischen Tests bekannt, die für die Zwecke dieser Internationalen Norm verwendet werden könnten. Aus praktischen Gründen wurde von diesen Tests nur eine begrenzte Anzahl, die in Hauptabschnitt drei dargelegt wird, in diese Norm aufgenommen.

Hauptabschnitt zwei: Organisation eines Ringversuchs zur Ermittlung der Präzision

Anmerkung: Es ist nicht zu erwarten, daß die Verfahrensweisen der verschiedenen Organisationen identisch sind. Der nachstehende Hauptabschnitt ist deshalb nur als Rahmenanleitung gedacht, die entsprechend modifiziert werden muß, um sie den speziellen Einzelfällen anzupassen.

8 Personalerfordernisse

8.1 Fachgruppe

Die tatsächliche Versuchsplanung sollte in der Hand einer Gruppe von Fachleuten liegen, die mit den Meßverfahren und ihrer Anwendung vertraut sind.

8.2 Fachmann für Statistik

Zumindest ein Mitglied der Fachgruppe sollte in der statistischen Versuchsplanung und Analyse von Versuchen Erfahrung haben.

8.3 Versuchsleiter

Die tatsächliche Organisation des Versuchs sollte einem einzigen Labor übertragen werden, und ein Mitarbeiter dieses Labors muß die volle Verantwortung tragen. Er ist der Versuchsleiter.

8.4 Laborverantwortlicher

Ein Mitarbeiter jedes der teilnehmenden Labors sollte verantwortlich gemacht werden für die Organisation der Meßdurchführung in Einklang mit den Anweisungen des Versuchsleiters und für die Berichterstattung der Meßergebnisse.

8.5 Bearbeiter

In jedem Labor müssen die Messungen von einem einzigen Bearbeiter durchgeführt werden, der repräsentativ ist für all jene, die die Messungen aller Wahrscheinlichkeit nach in üblichen Situationen durchführen. Er sollte von dem Laborverantwortlichen darüber Anweisungen erhalten, welche Daten er liefern soll und in welcher Reihenfolge die Messungen durchzuführen sind. Die Anweisungen sollten jedoch das Meßverfahren selbst nicht näher erläutern.

9 Aufgaben und Probleme

9.1 Die folgenden Fragen sollten von der Fachgruppe besprochen werden:

- Gibt es eine zufriedenstellende Festlegung für das Meßverfahren?
- Welchen Bereich von Merkmalsniveaus trifft man in der Praxis an?
- Wieviele unterschiedliche Merkmalsniveaus sollten bei dem Versuch verwendet werden (siehe Abschnitt 10.1)?
- Welche Materialien sind geeignet, diese Merkmalsniveaus zu repräsentieren?
- Sollte das Material speziell homogenisiert werden, bevor die Proben präpariert werden, oder sollte die Inhomogenität des Materials mit in die Werte von r und R eingehen (siehe Abschnitt 10.3)?
- Wie groß soll die Zahl n der Mehrfachmessungen sein, und welche Probenmengen sollen an die Labors versandt werden (siehe Abschnitt 10.1)?
- Sollten jedem Labor n gesonderte Proben für jedes Merkmalsniveau gesandt werden oder eine einzige Probe für n Mehrfachmessungen (siehe Abschnitt 10.3)? Oder ist ein Experiment mit zweigeteiltem Merkmalsniveau erwünscht (siehe Abschnitt 10.4.2)?
- Sollten an die Labors zusätzliche Proben für praktische Übungen gesandt werden, bevor die offiziellen Messungen durchgeführt werden (siehe Abschnitt 10.5.2)?
- Wieviele Labors sollten für die Mitarbeit an den Versuchen gewonnen werden (siehe Abschnitt 10.1)?
- Wie sollten die Labors ausgewählt werden und welchen Anforderungen sollten sie genügen (siehe Abschnitt 10.2)?

- k) Welche Anweisungen bezüglich der Durchführung der Messungen sollten die Laborverantwortlichen erhalten, und bis zu wieviel signifikante Ziffern sollten die Meßprotokolle enthalten (siehe Abschnitt 10.4.1 und 10.5.1)?
- l) Welche Informationen sollten außer den zahlenmäßigen Meßergebnissen angefordert werden (siehe Abschnitt 10.6)?
- m) Wer sollte als Versuchsleiter bestimmt werden?

9.2 Die Aufgaben des **Fachmannes** für Statistik sind:

- a) sein Spezialwissen zu der Versuchsplanung beizusteuern;
- b) die Daten zu analysieren;
- c) entsprechend den Anweisungen in Hauptabschnitt 3 einen Bericht an die Fachgruppe zu schreiben.

9.3 Die Aufgabe des **Versuchsleiters** besteht in der Organisation des Versuchs, wie er von der Fachgruppe geplant wurde und insbesondere darin,

- a) sich der Mitarbeit der erforderlichen Anzahl von Labors zu vergewissern und dafür zu sorgen, daß Laborverantwortliche ernannt werden;
- b) die Vorbereitung der Materialien und Proben zu organisieren und zu überwachen, und die Proben zu verteilen; zu jedem Merkmalsniveau sollte eine bestimmte Menge des Materials als Reserve zurückbehalten werden;
- c) Anweisungen zu erlassen und sie an die Laborverantwortlichen früh genug zu verteilen, um ihnen die Möglichkeit zu Kommentaren und Rückfragen zu geben;
- d) geeignete Formblätter für die Bearbeiter zu entwerfen, die als Arbeitsprotokoll dienen, und solche für die Laborverantwortlichen, die dem Niederlegen der Meßergebnisse dienen;
- e) die Meßergebnisse zu sammeln und eine Tabelle vorzubereiten, die für die statistische Analyse geeignet ist.

9.4 Die Aufgaben der **Laborverantwortlichen** sind,

- a) die Proben gemäß den Anweisungen des Versuchsleiters an den/die Bearbeiter auszuhändigen;
- b) die Durchführung der Messungen zu überwachen (die Laborverantwortlichen dürfen nicht an der Ausführung der Messungen mitwirken);
- c) die Versuchsergebnisse mit allen Unregelmäßigkeiten oder Schwierigkeiten, die sich ergeben haben, zu sammeln und über sie dem Versuchsleiter zu berichten.

9.5 Die Aufgaben der **Bearbeiter** sind es:

- a) die Messungen entsprechend dem festgelegten Meßverfahren durchzuführen;
- b) über alle Unregelmäßigkeiten und Schwierigkeiten, die sich dabei ergeben haben, zu berichten (siehe Abschnitte 10.5.1 und 10.5.3).

9.6 Die abschließenden Aufgaben der **Fachgruppe** sind es:

- a) den Bericht des Fachmanns für Statistik zu diskutieren;
- b) endgültige Werte für die Wiederholgrenze und Vergleichsgrenze aufzustellen;
- c) zu entscheiden, ob weitere Schritte zur Verbesserung der Festlegung für das Meßverfahren oder im Hinblick auf die Labors nötig sind, die als Ausreißer ausgeschlossen worden sind (siehe Abschnitt 11.2.3 d).

9.7 Da die Abschnitte 9.2 und 9.6 während der abschließenden Stufen der statistischen Analyse betrachtet werden, werden sie weiter in Hauptabschnitt drei behandelt.

10 Bemerkungen zu den Abschnitten 8 und 9

10.1 Anzahl der Labors und Merkmalsniveaus

Es können keine strengen Regeln aufgestellt werden. Die Zahl der Merkmalsniveaus bei einem Versuch zur Ermittlung der Präzision sollte im Verhältnis zu dem Bereich der Merkmalsniveaus, der überdeckt werden soll, ausgewählt werden, wobei die Kosten für die Durchführung der Versuche zu beachten sind.

Ist der Bereich der Merkmalsniveaus sehr groß, so ist zu erwarten, daß r und R von dem Merkmalsniveau m abhängen. Die Verwendung von mindestens 6 Merkmalsniveaus ist wünschenswert, um die Beziehung zwischen diesen Größen in zufriedenstellender Weise herstellen zu können. Andererseits mag man bei dem Beispiel zur Messung der Schmelztemperatur eines Teerproduktes, das in Abschnitt 23 behandelt wird (und bei dem der Meßbereich von 88 °C bis 102 °C reicht), die Verwendung von vier Merkmalsniveaus als mehr als angemessen ansehen.

Die Anzahl der Labors sollte bis zu einem gewissen Grade von der Anzahl der Merkmalsniveaus abhängen. Es wird empfohlen, daß die Zahl der Labors niemals kleiner als 8 sein sollte; eine größere Anzahl von Labors ist vorzuziehen, beispielsweise 15 oder mehr, wenn nur ein einziges Merkmalsniveau von Interesse ist.

Was den Wert von n anbetrifft, so wird $n = 2$ empfohlen, es sei denn, es ist üblich, eine größere Anzahl von Mehrfachmessungen durchzuführen, wie zum Beispiel bei gewissen einfachen physikalischen Messungen.

10.2 Auswahl teilnehmender Labors

10.2.1 Vom statistischen Standpunkt aus sollte man die Labors, die an einem Ringversuch zur Ermittlung der Präzision teilnehmen, nach dem Zufallsprinzip aus allen Labors auswählen, die aller Wahrscheinlichkeit nach das Meßverfahren anwenden. Freiwillige Teilnehmer repräsentieren möglicherweise nicht ein realistisches Gesamtbild der Labors.

In der Praxis können jedoch andere Erwägungen mit diesem Prinzip kollidieren; zum Beispiel kann ein Erfordernis, daß die teilnehmenden Labors gleichmäßig über verschiedene Kontinente oder Klimazonen verteilt sein sollten, die Struktur der Repräsentierung der Labors beeinflussen.

Die Fachgruppe sollte die Auswahlgesichtspunkte und die Anforderungen an die teilnehmenden Labors festlegen.

10.2.2 Bei der Hinzuziehung der erforderlichen Anzahl von Labors für die Mitarbeit sollte deren Verantwortlichkeit genau festgelegt werden. Nachstehend ein Beispiel für einen geeigneten Fragebogen, der diesem Zweck dient:

Fragebogen für Ringversuche	
Benennung des Meßverfahrens (siehe Anlage): _____	
1 Unser Labor wünscht an dem Ringversuch zur Ermittlung der Präzision dieses festgelegten Meßverfahrens teilzunehmen. Ja ☐ Nein ☐ (zutreffendes ankreuzen)	
2 Als Teilnehmer verpflichten wir uns, daß a) alle dafür notwendigen Geräte, Chemikalien und alle weiteren im Verfahren angegebenen Erfordernisse bei Programmbeginn in unserem Labor vorhanden sind; b) spezifische Zeitplanerfordernisse wie Datum des Prüfungsbeginns, Reihenfolge der Proben und Datum des Prüfungsendes des Programms streng eingehalten werden; c) die Methode exakt angewandt wird; d) die Proben gemäß den Anweisungen behandelt werden und e) ein qualifizierter Bearbeiter die Prüfungen durchführt. Wir haben das Prüfverfahren studiert, unsere Fähigkeiten und Möglichkeiten nach bestem Wissen beurteilt und sind der Meinung, daß wir für eine kooperative Prüfung dieses Verfahrens vorbereitet sind.	
3 Bemerkungen _____ _____ _____ _____	
Unterschrift: _____	Institution oder Labor: _____

10.3 Inhomogenität des Materials

Ist das zu messende Material nicht homogen, so ist es wichtig, die Proben in der durch das Verfahren festgesetzten Art und Weise vorzubereiten, wobei man am besten mit einer Probe eines kommerziellen Materials für jedes Merkmalsniveau beginnt. Es mögen einige Modifizierungen nötig sein, um sicherzustellen, daß die Menge des verfügbaren Materials sowohl für den Versuch selbst als auch für eine gewisse Reserve ausreichend ist. Für die Proben auf jedem Merkmalsniveau sollten dann n getrennte Behälter verwendet werden, wenn auch nur die geringste Gefahr besteht, daß das Material verdirbt, wenn der Behälter einmal geöffnet wurde (Zum Beispiel: hygroskopisches Material, Material, das oxidiert oder das flüchtige Komponenten verliert). Bei instabilen Materialien sollten spezielle Anweisungen zur Lagerung und Behandlung festgesetzt werden.

Im allgemeinen, wenn Werte für r und R veröffentlicht werden, wird empfohlen, genau anzugeben, welches Material bei dem Ringversuch zur Ermittlung der Präzision verwendet wurde und auf welche Arten von Materialien sich r und R voraussichtlich beziehen.

10.4 Organisation der Messungen

10.4.1 Bei q Merkmalsniveaus und n Mehrfachmessungen muß jedes teilnehmende Labor qn Messungen ausführen.

Die Durchführung dieser Messungen sollte organisiert und die Bearbeiter sollten wie folgt angewiesen werden:

- a) alle qn Messungen sollten von ein und demselben Bearbeiter ausgeführt werden, wobei immer die gleichen Geräte verwendet werden,
- b) jede Gruppe von n Messungen, die zu einem Merkmalsniveau gehören, muß unter Wiederholbedingungen durchgeführt werden, das heißt, innerhalb einer kurzen Zeitspanne und vom selben Bearbeiter, ohne jegliche Zwischenkalibrierung der Gerätschaft, sofern erstere nicht ein integraler Bestandteil eines Meßvorgangs ist.
- c) Es ist nicht wesentlich, daß alle q Gruppen von jeweils n Messungen strikt in einer kurzen Zeitspanne ausgeführt werden; verschiedene Gruppen von Messungen können an verschiedenen Tagen ausgeführt werden,
- d) sollte im Verlauf der Messungen ein Bearbeiter ausscheiden müssen, so darf ein anderer Bearbeiter die Messungen zu Ende führen, sofern sich dieser Wechsel nicht innerhalb einer Gruppe von n Messungen auf einem Merkmalsniveau, sondern nur zwischen zwei der q Gruppen von Messungen ereignet. Jeder derartige Wechsel muß zusammen mit den Meßergebnissen berichtet werden.
- e) Es ist wesentlich, daß eine Gruppe von n Messungen, die unter Wiederholbedingungen durchgeführt werden,

unabhängig gemacht wird, so als ob es n Messungen an verschiedenen Materialien wären. Es ist jedoch die Regel, daß der Bearbeiter weiß, daß er identisches Material mißt, aber in den ihm gegebenen Anweisungen sollte der Punkt betont werden, daß das übergeordnete Ziel des Versuches darin besteht, die Differenzen zu bestimmen, die in tatsächlichen Messungen zwischen Meßergebnissen auftreten können. Wenn befürchtet wird, daß ungeachtet dieser Warnung, vorhergehende Meßergebnisse die nachfolgenden und folglich auch die Wiederholvarianz beeinflussen können, dann wird ein Versuch mit zweigeteiltem Merkmalsniveau als das richtige Vorgehen angesehen (siehe Abschnitt 10.4.2).

10.4.2 Ist $n = 2$, dann wird oft ein alternatives Verfahren verwendet, bei dem das Merkmalsniveau zweigeteilt ist. Die Anwendung dieses Verfahrens kann erwogen werden, wenn befürchtet wird, daß der Bearbeiter während der Messung fortlaufend identischer Proben von dem Ergebnis der ersten Messung beeinflusst sein könnte. Anstatt 2 Proben zu nehmen, von denen dem Bearbeiter gesagt wurde, daß sie identisch sein sollen, oder anstatt 2 Messungen an dem gleichen Materialstück durchzuführen, werden bei diesem Verfahren 2 Reihen von je p Proben auf geringfügig verschiedenen Merkmalsniveaus m_a und m_b vorbereitet, (wobei $m_a - m_b$ eine kleine Größe ist) und jedes der p Laboratorien erhält eine Probe aus der Serie a und eine Probe aus der Serie b zum Messen. Es muß eindeutig unterschieden werden, welches Meßergebnis zur Reihe a und welches zur Reihe b gehört; sie sind nicht vertauschbar wie die Meßergebnisse an identischem Material. Die Werte für r und R , die man aus einem Versuch mit zweigeteiltem Merkmalsniveau erhält, gelten für das mittlere Merkmalsniveau $m = (m_a + m_b)/2$.

Versuche mit zweigeteiltem Merkmalsniveau erfordern eine geringfügige Änderung bei der statistischen Analyse, die in Hauptabschnitt drei diskutiert wird.

10.4.3 Zusätzliche Aspekte der Organisation der Messungen sind die folgenden:

- Es kann nötig sein, die Zeitdauer zu begrenzen, die zwischen dem Tag, an dem die Proben erhalten wurden, und dem Tag, an dem die Messungen durchgeführt werden, verstreichen darf.
- Jegliche vorherige Überprüfung der Geräte sollte so durchgeführt werden, wie es im festgelegten Verfahren bestimmt ist.
- Alle Proben sollten eindeutig mit der Versuchsbezeichnung und einer Probenbezeichnung gekennzeichnet sein.

10.5 Anweisungen an die Bearbeiter

10.5.1 Vor der Durchführung der Messungen sollten die Bearbeiter keine Anweisungen zusätzlich zu denen erhalten, die in der Festlegung für das Meßverfahren enthalten sind; diese sollten allein ausreichen. Die Bearbeiter sollten jedoch zu Bemerkungen zu der Festlegung ermutigt werden, insbesondere sich zu äußern, ob die darin enthaltenen Anweisungen hinreichend klar und eindeutig sind. Unklarheiten können sich zum Beispiel dann einschleichen, wenn die Festlegung in eine Reihe verschiedener Sprachen

übersetzt worden ist. Es ist jedoch wünschenswert, daß alle teilnehmenden Labors ihre Meßergebnisse mit derselben Anzahl von Dezimalstellen berichten. Die Laborverantwortlichen sollten entsprechend angewiesen werden. In der kommerziellen Praxis können die Meßergebnisse ziemlich grob gerundet sein, und in einem Ringversuch zur Ermittlung der Präzision kann es empfehlenswert sein, eine Dezimalstelle mehr als es üblich oder in der Festlegung vorgesehen ist, zu verwenden. Wenn r oder R von dem Merkmalsniveau m abhängen können, können unterschiedliche Richtlinien zur Rundung der Ergebnisse auf den verschiedenen Merkmalsniveaus erforderlich sein.

10.5.2 Ein Bearbeiter wird vielleicht nicht die übliche Präzision erreichen, wenn er eine Messung zum ersten Mal oder nach einer langen Zeitspanne durchführt. In diesen Fällen kann den Bearbeitern entsprechend der Entscheidung der Fachgruppe oder des Laborverantwortlichen erlaubt werden, einige inoffizielle Messungen durchzuführen, um Erfahrungen mit dem Meßverfahren zu gewinnen, bevor sie mit der Messung der offiziellen Proben des Ringversuchs zur Ermittlung der Präzision beginnen. Solche vorläufigen Messungen zur Erfahrungssammlung dürfen niemals an offiziellen Proben ausgeführt werden, und Material dafür sollte der Versuchsleiter zur Verfügung stellen.

10.5.3 Die Bearbeiter sollten angewiesen werden, in jedem Fall zu berichten, wenn sie entweder ihren Anweisungen nicht folgen können oder sich ungewollt nicht nach ihnen richten. Es muß ihnen auch erklärt werden, daß es besser ist, einen Fehler zu melden als die Ergebnisse zu frisieren. Denn ein oder zwei fehlende Ergebnisse machen den Versuch noch nicht zunichte und können auf eine Schwäche in der Festlegung des Meßverfahrens hindeuten.

10.6 Bericht über die Meßergebnisse

Der Verantwortliche jedes Labors sollte einen ausführlichen Bericht über die Messungen schreiben; dieser Bericht sollte folgendes enthalten:

- die endgültigen Meßergebnisse, wobei besondere Sorgfalt auf die Vermeidung von Übertragungs- und Schreibfehlern verwendet werden sollte, zum Beispiel durch die Verwendung von Fotokopien der Meßergebnisse der Bearbeiter,
- die originalen Beobachtungen oder Ablesungen (falls es solche gibt), von denen die endgültigen Ergebnisse abgeleitet wurden, möglicherweise durch Kopien der Arbeitsbücher der Bearbeiter,
- Bemerkungen des Bearbeiters zur Festlegung des Meßverfahrens,
- Angaben über besondere Vorkommnisse oder Störungen, die während der Messungen aufgetreten sein können, einschließlich jeglichen Wechsels des Bearbeiters zusammen mit einer Erklärung, welche Messungen von welchem Bearbeiter ausgeführt worden sind,
- Zeitpunkt(e), an dem (denen) die Proben eintrafen,
- Zeitpunkt(e), an dem (denen) jede Probe gemessen wurde,
- Angaben über die benutzten Geräte, sofern dies wichtig ist,
- jegliche andere sachdienliche Informationen.

Hauptabschnitt drei: Statistische Analyse der Ergebnisse eines Ringversuchs

11 Vorbemerkungen

11.1 Fachmann für Statistik

Die Analyse der Meßergebnisse, die ein Ringversuch ergeben hat, ist die Aufgabe eines Fachmannes für Statistik, der Mitglied der Fachgruppe ist und an der Versuchsplanung beteiligt war (siehe Abschnitte 8.2 und 9.2).

11.2 Zellen

Jede Kombination aus einem Labor und einem Merkmalsniveau wird eine Zelle des Ringversuchs genannt. Im Idealfall bilden die Ergebnisse eines Versuchs mit p Labors und q Merkmalsniveaus eine Tabelle mit pq Zellen, die jeweils n Mehrfachmessungen enthalten, und sie alle können für die Berechnung der Wiederholgrenze r und der Vergleichsgrenze R verwendet werden. Diese ideale Situation wird jedoch in der Praxis nicht immer erreicht. Abweichungen vom Idealfall entstehen durch überflüssige oder fehlende Daten und durch Ausreißer.

11.2.1 Überflüssige Daten

Manchmal führt ein Labor mehr als die n offiziell festgesetzten Mehrfachmessungen durch und berichtet über diese. In diesem Fall gibt der Laborverantwortliche (siehe Abschnitte 8.4 und 9.4) alle Ergebnisse weiter, oder er ist aufgefordert zu berichten, warum dies geschehen ist und welches die korrekten Meßergebnisse sind. Lautet die Antwort, daß sie alle gleichermaßen gültig sind, so können sie alle entsprechend der Berechnungsmethode nach Abschnitt 14.9 berücksichtigt werden.

11.2.2 Fehlende Daten

In anderen Fällen können Meßergebnisse fehlen, zum Beispiel, weil eine Probe verlorengegangen ist, oder weil ein Fehler bei der Durchführung der Messungen unterlaufen ist. Die Analyse, die in Abschnitt 16 empfohlen wird, ist so gear- tet, daß vollständig leere Zellen einfach ignoriert werden können, während teilweise leere Zellen entsprechend der Berechnungsmethode nach Abschnitt 14.9 berücksichtigt werden können. Die Gründe für fehlende Meßergebnisse sollten in dem Bericht des Laborverantwortlichen enthalten sein.

Anmerkung: Fehlt eines der beiden Ergebnisse in einer Zelle bei dem Versuch mit zweigeteiltem Merkmalsniveau (siehe Abschnitt 10.4.2) so muß das andere Meßergebnis verworfen und die Zelle so behandelt werden, als ob sie leer wäre.

11.2.3 Ausreißer

Ausreißer sind Werte unter den Originalmeßergebnissen oder in den Tabellen, die aus ihnen abgeleitet sind, die so stark von vergleichbaren Werten in derselben Tabelle abweichen, daß sie als unvereinbar mit den anderen Daten angesehen werden müssen. Die Erfahrung hat gelehrt, daß Ausreißer nicht immer vermieden werden können und zu berücksichtigt sind.

Folgendes Verfahren wird im Umgang mit Ausreißern empfohlen:

- Empfohlen werden der Cochran-Test auf Ausreißer (einseitig) (siehe Abschnitt 12) und der Dixon-Test auf Ausreißer (siehe Abschnitt 13) in Verbindung mit dem folgenden Verfahren:

$P \geq 5\%$, d. h. der Wert der Cochran- oder Dixon-Prüfgröße ist **kleiner als** ihr zu einem Signifikanzniveau von 5% ge-

hörender kritischer Wert; der zu testende Wert wird als korrekt angenommen;

$5\% > P > 1\%$, d. h., der Wert der Prüfgröße liegt zwischen ihren zu einem Signifikanzniveau von 5% und 1% gehörenden kritischen Werten: der zu testende Wert wird als möglicher Ausreißer (Straggler) bezeichnet und mit einem einfachen Stern markiert;

$P \leq 1\%$, d. h., der Wert der Prüfgröße ist **größer als** ihr zu einem Signifikanzniveau von 1% gehörender kritischer Wert: der zu testende Wert wird **statistischer Ausreißer** genannt und mit zwei Sternen markiert;

P ist die Wahrscheinlichkeit (unter der Nullhypothese), mit der die Prüfgröße einen Wert gleich dem oder größer als der Beobachtungswert annimmt.⁴⁾

Die zu den Signifikanzniveaus von 5% und 1% gehörenden kritischen Werte für den Cochran- und den Dixon-Test sind in den Anhängen A und B wiedergegeben.

- Manchmal kann von der Anwendung dieser Tests abgesehen oder es können andere statistische Tests gewählt werden, weil ein Fachmann für Statistik schon bei einer flüchtigen Betrachtung der Daten sieht, (zum Beispiel aufgrund einer grafischen Darstellung), daß der Test entweder zu einem nicht-signifikanten oder zu einem hoch-signifikanten Ergebnis führen wird. Im Zweifels- falle jedoch sollte der Test immer durchgeführt werden.
- Als nächstes wird untersucht, ob die möglichen Ausreißer (Straggler) und/oder die statistischen Ausreißer durch irgendeinen technischen Fehler erklärt werden können; zum Beispiel durch einen Fehler bei der Durchführung der Messungen, durch einen Rechenfehler, durch den Fehler beim Übertragen der Meßergebnisse oder durch die Analyse einer falschen Probe. Wenn der Fehler ein Rechen- oder Übertragungsfehler war, sollte das zweifelhafte Meßergebnis durch den korrekten Wert ersetzt werden. Wenn eine falsche Probe analysiert worden ist, sollte das Meßergebnis in die richtige Zelle übertragen werden. Nachdem eine solche Korrektur ausgeführt worden ist, sollte die Überprüfung auf mögliche Ausreißer und statistische Ausreißer wiederholt werden. Wenn es sich durch die Erklärung der technischen Fehler als unmöglich erweist, das zweifelhafte Meßergebnis zu ersetzen, sollte es als wirklicher Ausreißer, der nicht zu dem Versuch gehört, verworfen werden,
- Treten mehrere unerklärliche mögliche Ausreißer (Straggler) und/oder statistische Ausreißer auf verschiedenen Merkmalsniveaus innerhalb desselben Labors auf, so kann das Labor selbst als Ausreißer betrachtet werden, weil es eine zu hohe laborinterne Varianz oder zu große systematische Abweichung im Merkmalsniveau seiner Meßergebnisse hat. Es ist dann vielleicht angebracht, einige oder alle Daten aus einem solchen Labor zu verwerfen. (Diese Internationale Norm sieht keinen statistischen Test vor, solchermaßen unter Verdacht geratene Labors zu beurteilen. Die Hauptentscheidung sollte in der Verantwortlichkeit des Fachmanns für Statistik liegen, jedoch müssen alle ver-

⁴⁾ Nationale Fußnote: Dieser Satz ist in der deutschen Ausgabe redaktionell berichtigt worden.