

DIN ISO 13528

ICS 03.120.30

Einsprüche bis 2007-08-31

Entwurf**Statistische Verfahren für Eignungsprüfungen durch Ringversuche
(ISO 13528:2005); Text Deutsch und Englisch**

Statistical methods for use in proficiency testing by interlaboratory comparisons
(ISO 13528:2005); Text in German and English

Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons
interlaboratoires (ISO 13528:2005); Texte en allemand et anglais

Anwendungswarnvermerk

Dieser Norm-Entwurf wird der Öffentlichkeit zur Prüfung und Stellungnahme vorgelegt.

Weil die beabsichtigte Norm von der vorliegenden Fassung abweichen kann, ist die Anwendung dieses Entwurfes besonders zu vereinbaren.

Stellungnahmen werden erbeten

- vorzugsweise als Datei per E-Mail an nqsz@din.de in Form einer Tabelle. Die Vorlage dieser Tabelle kann im Internet unter www.din.de/stellungnahme abgerufen werden;
- oder in Papierform an den Normenausschuss Qualitätsmanagement, Statistik und Zertifizierungsgrundlagen (NQSZ) im DIN, 10772 Berlin (Hausanschrift: Burggrafenstr. 6, 10787 Berlin).

Die Empfänger dieses Norm-Entwurfs werden gebeten, mit ihren Kommentaren jegliche relevante Patentrechte, die sie kennen, mitzuteilen und unterstützende Dokumentationen zur Verfügung zu stellen.

Gesamtumfang 149 Seiten

Normenausschuss Qualitätsmanagement, Statistik und Zertifizierungsgrundlagen (NQSZ) im DIN
Normenausschuss Wasserwesen (NAW) im DIN

Inhalt

	Seite
Nationales Vorwort	4
0 Einleitung.....	5
0.1 Ziele von Eignungsprüfungen	5
0.2 ISO/IEC Guide 43.....	5
1 Anwendungsbereich	7
2 Normative Verweisungen	7
3 Begriffe	7
4 Statistische Leitlinien für die Planung und die Auswertung von Eignungsprüfungen.....	8
4.1 Eingriffs- und Warnsignale	8
4.2 Leitlinien für das Eingrenzen der Unsicherheit des zugewiesenen Werts	9
4.3 Leitlinien für die Festlegung der Anzahl von Wiederholmessungen	10
4.4 Homogenität und Stabilität von Stichproben	10
4.5 Betriebsspezifische Messverfahren.....	11
4.6 Angabe von Daten	11
4.7 Gültigkeitsdauer der Ergebnisse der Eignungsprüfung	11
5 Bestimmung des zugewiesenen Werts und der zugehörigen Standardunsicherheit	12
5.1 Wahl des Verfahrens zur Ermittlung des zugewiesenen Werts	12
5.2 Stichprobenherstellung	12
5.3 Zertifizierte Referenzwerte	13
5.4 Referenzwerte	13
5.5 Konsenswerte erfahrener Prüfstellen.....	16
5.6 Konsenswert der Teilnehmer	17
5.7 Vergleich des zugewiesenen Werts	22
5.8 Fehlende Werte	22
6 Bestimmung der Standardabweichung für die Eignungsbeurteilung.....	23
6.1 Auswahl des Verfahrens	23
6.2 Vorgabewert	23
6.3 Wahrgenommener Wert	24
6.4 Wert nach einem allgemeinen Modell.....	25
6.5 Wert aus den Ergebnissen eines Versuchs zur Präzision	25
6.6 Wert aus Daten, die während eines Durchlaufs eines Eignungsprüfungsplans ermittelt wurden	26
6.7 Vergleich von Werten für die Präzision aus einer Eignungsprüfung mit festgestellten Werten	27
7 Berechnung von Leistungskenngrößen.....	27
7.1 Schätzwerte für die systematische Abweichungskomponente einer Prüfstelle.....	27
7.2 Prozentuale Differenzen	31
7.3 Rangordnungen und prozentuale Rangordnungen	33
7.4 z-Werte	34
7.5 E_n -Werte	37
7.6 z' -Werte.....	37
7.7 Zeta-Werte (ζ)	38
7.8 E_z -Wert	39
7.9 Beispiel der Auswertung von Daten bei Angabe von Unsicherheiten	40
7.10 Kombiniertes Leistungswert	44
8 Grafische Verfahren zur Kombination von Leistungswerten für mehrere Messgrößen in einem Durchlauf einer Eignungsprüfung.....	45
8.1 Anwendung	45

	Seite
8.2	Histogramme von Leistungswerten 45
8.3	Balken-Diagramme standardisierter systematischer Abweichungskomponenten von Prüfstellen 47
8.4	Balkendiagramme standardisierter Wiederholmessungen.....48
8.5	Youden-Diagramm.....48
8.6	Diagramme von Wiederholstandardabweichungen.....56
8.7	Geteilte Stichproben 58
9	Grafische Verfahren zur Kombination von Leistungswerten über mehrere Durchläufe eines Eignungsprüfungsplans 63
9.1	Anwendungen.....63
9.2	Shewhart-Regelkarten für $\bar{z}$ -Werte 63
9.3	Cusum-Karte für $\bar{z}$ -Werte.....66
9.4	Auftragungen standardisierter systematischer Abweichungskomponenten von Prüfstellen gegen Prüfstellenmittelwerte.....67
9.5	Punktwolke.....68
Anhang A (normativ) Formelzeichen 70	
Anhang B (normativ) Prüfung von Stichproben auf Homogenität und Stabilität 71	
Anhang C (normativ) Robuste Auswertung 75	
Literaturhinweise 77	

Nationales Vorwort

Dieser Norm-Entwurf enthält die deutsche Übersetzung der Internationalen Norm ISO 13528:2005-09, die vom Technischen Komitee ISO/TC 69, *Anwendung statistischer Methoden*, Unterkomitee 6, *Messverfahren und Messergebnisse*, erarbeitet wurde, dessen Sekretariat vom JISC/JSA (Japan) gehalten wird. Der zuständige nationale Normenausschuss ist der NA 147, *Qualitätsmanagement, Statistik und Zertifizierungsgrundlagen* (NQSZ), Arbeitsausschuss *Angewandte Statistik*. Deutscher Mitträger ist der Normenausschuss NA 119, *Wasserwesen* (NAW), Arbeitsausschuss *Wasseruntersuchung*.

Es wird auf die Möglichkeit hingewiesen, dass für Teile dieses Dokuments Patentrechte bestehen. Das DIN übernimmt keine Verantwortung für die Kennzeichnung solcher Rechte.

Für die in diesem Dokument zitierten Internationalen Normen wird im Folgenden auf die entsprechenden Deutschen Normen hingewiesen:

ISO 5725-1	siehe DIN ISO 5725-1
ISO/IEC Guide 43-1:1997	siehe DIN V 55394-1:2000-09

Nationaler Anhang NA (informativ)

Literaturhinweise

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

DIN V 55394-1:2000-09, *Eignungsprüfung durch Vergleiche zwischen Laboratorien — Teil 1: Entwicklung und Durchführung von Programmen für die Eignungsprüfung*

Statische Verfahren für Eignungsprüfungen durch Ringversuche

0 Einleitung

0.1 Ziele von Eignungsprüfungen

Eignungsprüfungen durch Ringversuche werden angewendet, um die Leistungsfähigkeit einzelner Prüfstellen für bestimmte Prüfungen oder Messungen zu bestimmen und die fortgesetzte Leistungsfähigkeit von Prüfstellen zu überwachen. Die Einleitung zum ISO/IEC Guide 43-1:1997 enthält eine umfassende Darstellung der Zwecke der Eignungsprüfung. In der Sprache der Statistik kann die Leistungsfähigkeit einer Prüfstelle durch drei Eigenschaften beschrieben werden: systematische Abweichungskomponente der Prüfstelle, Stabilität und Wiederholpräzision. Die systematische Abweichungskomponente der Prüfstelle und die Wiederholpräzision sind in ISO 3534-1, ISO 3534-2 und ISO 5725-1 definiert. Ein Maß für die Stabilität der Ergebnisse einer Prüfstelle ist die Präzision unter Zwischenbedingungen, wie in ISO 5725-3 definiert.

Die systematische Abweichungskomponente einer Prüfstelle darf anhand von Prüfungen mit Referenzmaterialien mit Hilfe des in ISO 5725-4 beschriebenen Verfahrens beurteilt werden, wenn solche Referenzmaterialien verfügbar sind. Andernfalls liefert die Eignungsprüfung durch Ringversuche ein allgemein verfügbares Mittel, um Angaben über die systematische Abweichungskomponente einer Prüfstelle zu erhalten, und die Verwendung von Daten aus Eignungsprüfungen, um Schätzwerte für die systematische Abweichungskomponente einer Prüfstelle zu erhalten, ist ein wichtiger Aspekt der Auswertung solcher Daten. Jedoch beeinflussen Stabilität und Wiederholpräzision durch Eignungsprüfungen gewonnene Daten. Deshalb ist es möglich, dass eine Prüfstelle während eines Durchlaufs der Eignungsprüfung Daten erhält, die eine systematische Abweichungskomponente aufzeigen, die tatsächlich durch dürftige Stabilität oder dürftige Wiederholpräzision verursacht wird. Es ist deshalb wichtig, dass diese Aspekte der Leistungsfähigkeit von Prüfstellen regelmäßig beurteilt werden.

Stabilität kann durch das wiederholte Prüfen aufbewahrter Stichproben oder durch regelmäßige Messungen an einem Referenzmaterial oder einem prüfstelleneigenen Referenzmaterial (einem Vorrat an Material, der von einer Prüfstelle zur Verwendung als privates Referenzmaterial angelegt wird) beurteilt werden. Diese Techniken werden in ISO 5725-3 beschrieben. Stabilität kann auch beurteilt werden, indem Schätzwerte der systematischen Abweichungskomponente einer Prüfstelle, die aus Eignungsprüfungen hergeleitet wurden, in Qualitätsregelkarten grafisch dargestellt werden. Dies kann Informationen über die Leistungsfähigkeit der Prüfstelle liefern, die bei der Überprüfung der Ergebnisse einzelner Durchläufe von Eignungsprüfungsplänen nicht erkennbar werden, und stellt einen weiteren wichtigen Aspekt der Auswertung solcher Daten dar.

Daten, die zur Beurteilung der Wiederholpräzision geeignet sind, dürfen durch im üblichen Arbeitsverlauf einer Prüfstelle durchgeführte Prüfungen oder anhand von zusätzlichen Prüfungen erzeugt werden, die eigens zum Zweck der Beurteilung der Wiederholpräzision innerhalb einer Prüfstelle durchgeführt werden. Folglich ist die Beurteilung der Wiederholpräzision nicht unbedingt ein wichtiger Aspekt der Eignungsprüfung, obwohl es wichtig ist, dass Prüfstellen ihre Wiederholpräzision auf irgendeine Weise überwachen. Die Wiederholpräzision kann durch die grafische Auftragung von Spannweiten von Doppelmessungen in einer Qualitätsregelkarte beurteilt werden, wie in ISO 5725-6 beschrieben.

Das Ablaufdiagramm (Bild 1) veranschaulicht, wie die in dieser Internationalen Norm beschriebenen Techniken anzuwenden sind.

0.2 ISO/IEC Guide 43

ISO/IEC Guide 43-1 beschreibt verschiedene Arten von Prüfplänen für Eignungsprüfungen und gibt Hinweise hinsichtlich der Organisation und Planung von Plänen für Eignungsprüfungen. ISO/IEC Guide 43-2 gibt Hinweise zur Auswahl und Verwendung von Plänen für Eignungsprüfungen durch Akkreditierungsstellen für Prüfstellen. Detaillierte Angaben hierzu sollten diesen Dokumenten entnommen werden. (Die Information wird hier nicht wiederholt). ISO/IEC Guide 43-1 enthält einen Nachtrag, der kurz die statistischen Verfahren beschreibt, die bei Plänen für Eignungsprüfungen verwendet werden.

Diese Internationale Norm ergänzt ISO/IEC Guide 43, indem sie detaillierte Hinweise zur Anwendung statistischer Verfahren bei Eignungsprüfungen liefert, die im ISO/IEC Guide 43 fehlen. ISO 13528 basiert weitgehend auf einem harmonisierten Protokoll für die Eignungsprüfung von Analyselabors [1], aber ist zur Anwendung bei allen Messverfahren bestimmt.

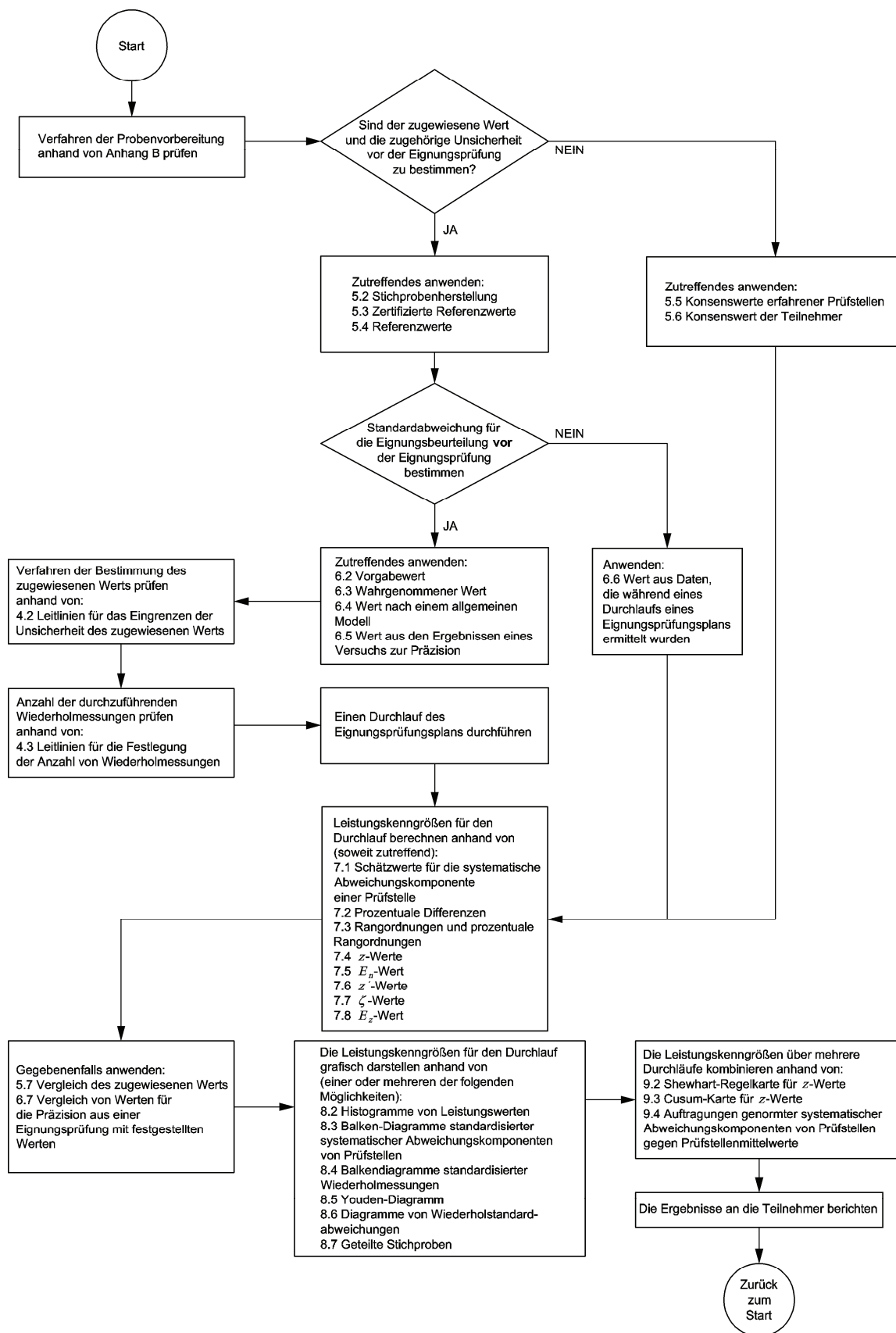


Bild 1 — Ablaufdiagramm der Aktivitäten, die die Verwendung von statistischen Verfahren erfordern, wenn ein Eignungsprüfungsplan ausgeführt wird

1 Anwendungsbereich

Diese Internationale Norm ergänzt ISO Guide 43 (alle Teile), indem sie detailliert zuverlässige statistische Verfahren beschreibt, die von Ausrichtern von Eignungsprüfungsplänen zur Auswertung der erhaltenen Daten anzuwenden sind, und indem sie Empfehlungen zur praktischen Anwendung dieser Verfahren durch Teilnehmer an solchen Prüfungen beziehungsweise durch Akkreditierungsstellen gibt.

Diese Internationale Norm kann angewendet werden, um nachzuweisen, dass die von Prüfstellen erhaltenen Messergebnisse keine unannehmbare systematische Abweichungskomponente aufweisen.

Sie ist auf quantitative, nicht jedoch auf qualitative Daten anwendbar.

2 Normative Verweisungen

Die folgenden in Bezug genommenen Dokumente sind für die Anwendung dieses Dokuments unentbehrlich. Bei datierten Verweisungen gilt die jeweils in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die jeweils letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich Änderungen).

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

ISO 3534-2:—¹⁾, *Statistik — Begriffe und Formelzeichen — Teil 2: Angewandte Statistik*

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

ISO/IEC Guide 43-1:1997, *Eignungsprüfung durch Vergleiche zwischen Laboratorien — Teil 1: Entwicklung und Durchführung von Programmen für die Eignungsprüfung*

3 Begriffe

Für die Anwendung dieses Dokuments gelten die Begriffe nach ISO 3534-1, ISO 3534-2 und ISO 5725-1 und die folgenden Begriffe.

3.1

Ringversuch

Organisation, Durchführung und Auswertung von Prüfungen oder Messungen an denselben oder ähnlichen Prüfgegenständen durch zwei oder mehr Prüfstellen entsprechend vorherbestimmten Bedingungen

ANMERKUNG Angepasst nach ISO/IEC Guide 43-1.

3.2

Eignungsprüfung

Bestimmung der Leistungsfähigkeit einer Prüfstelle durch Ringversuche

3.3

zugewiesener Wert

Wert, welcher einer bestimmten Größe zugeordnet wurde und welcher, mitunter durch Absprache, als mit einer für einen gegebenen Zweck geeigneten Unsicherheit behaftet akzeptiert wird

1) 2006 erschienen.

3.4

Standardabweichung für die Eignungsbeurteilung

Maß für die Streuung, das bei der Beurteilung der Eignung auf der Grundlage der verfügbaren Information verwendet wird

3.5

z-Wert

standardisiertes Maß für die systematische Abweichungskomponente einer Prüfstelle, berechnet unter Verwendung des zugewiesenen Werts und der Standardabweichung für die Eignungsbeurteilung

3.6

Koordinator

Organisation oder Person, die die Verantwortung für die Koordination aller Tätigkeiten trägt, die im Rahmen eines Eignungsprüfungsplans anfallen

4 Statistische Leitlinien für die Planung und die Auswertung von Eignungsprüfungen

(siehe ISO/IEC Guide 43-1:1997, 5.4.2)

4.1 Eingriffs- und Warnsignale

4.1.1 Diese Internationale Norm beschreibt einige einfache numerische oder grafische Kriterien, die auf die im Rahmen einer Eignungsprüfung erhaltenen Daten angewendet werden sollten, um festzustellen, ob Eingriffs- oder Warnsignale vorliegen. Selbst in einer gut geführten Prüfstelle mit erfahrenem Personal können mitunter anomale Ergebnisse ermittelt werden. Ebenso kann ein standardisiertes Messverfahren, auch wenn es in einem Präzisionsversuch validiert wurde, Fehler enthalten, die erst nach mehreren Durchläufen eines Eignungsprüfungsplans auffallen. Die Anweisung für die Eignungsprüfung selbst kann Fehler enthalten. Aus diesen Gründen sollen die hier gegebenen Kriterien nicht verwendet werden, um Prüfstellen als ungeeignet für die Durchführung des betrachteten Messverfahrens zu beurteilen. Wenn die Eignungsprüfung angewendet wird, um Prüfstellen als ungeeignet zu beurteilen, so müssen zwingend entsprechende Kriterien für diesen Zweck entwickelt werden.

4.1.2 Die hier angegebenen Kriterien sind so ausgelegt, dass sie Eingriffssignale geben, wenn die Standardabweichung für die Eignungsbeurteilung auf der (nach einem der in 6.4 bis 6.6 beschriebenen Verfahren) beobachteten Leistungsfähigkeit basiert und die Ergebnisse so aus dem Rahmen fallen, dass eine genauere Prüfung und Korrekturmaßnahmen angemessen erscheinen.

4.1.3 Der Koordinator sollte die wichtigsten Quellen der Variabilität verstehen, die bei Daten der Eignungsprüfung für die betreffende Messung zu erwarten sind. Erster Schritt jeder Auswertung sollte die Untersuchung der Verteilung von Ergebnissen daraufhin sein, ob Hinweise für unvorhergesehene Quellen von Variabilität vorliegen. Zum Beispiel könnte eine bimodale Verteilung ein Hinweis auf eine gemischte Grundgesamtheit von Ergebnissen sein, verursacht durch unterschiedliche Verfahren, kontaminierte Stichproben oder nachlässig formulierte Anleitungen. In dieser Situation sollten solche Bedenken ausgeräumt werden, bevor die Auswertung oder Beurteilung fortgesetzt wird. Akkreditierungsstellen müssen Leitlinien für die Reaktion auf unannehmbar Leistungen bei Eignungsprüfungen haben. Maßnahmen als Folge solcher Fehlleistungen werden anhand dieser Leitlinien oder durch die Qualitätssicherungsverfahren der Prüfstelle bestimmt. Es gibt jedoch allgemein empfohlene Maßnahmen für den Fall, dass eine Prüfstelle bei einer Eignungsprüfung ein unannehmbares Ergebnis produziert. Leitlinien für Maßnahmen von Prüfstellen als Reaktion auf erfolglose Durchführung einer Eignungsprüfung werden in 4.1.4 gegeben.

4.1.4 Bei Plänen, bei denen die Standardabweichung für die Eignungsbeurteilung auf der beobachteten Leistungsfähigkeit basiert, muss die Prüfstelle entscheiden, welche Nachforschungen und Korrekturmaßnahmen angemessen sind, wenn ein Ergebnis ein Eingriffssignal liefert; nötigenfalls ist diese Entscheidung in Abstimmung mit dem Koordinator oder einer Akkreditierungsstelle zu treffen. Spricht kein triftiger Grund dagegen, muss die Prüfstelle ihre Verfahrensweisen prüfen und eine oder mehrere Korrekturmaßnahmen ermitteln, die nach dem Ermessen des Prüfstellenpersonals wahrscheinlich das

erneute Auftreten solcher Ergebnisse verhindern. Die Prüfstelle darf sich hinsichtlich möglicher Ursachen für ihr Problem durch den Koordinator beraten lassen oder den Koordinator bitten, weitere Fachleute zu Rate zu ziehen. Die Prüfstelle muss an weiteren Durchläufen des Eignungsprüfungsplans teilnehmen, um die Wirksamkeit der Korrekturmaßnahmen zu beurteilen. Als geeignete Korrekturmaßnahme kommt eine der folgenden in Frage:

- a) Überprüfung, ob das Personal das Messverfahren versteht und befolgt;
- b) Überprüfung, ob alle Einzelheiten des Messablaufs korrekt sind;
- c) Überprüfung der Kalibrierung der Geräte und der Zusammensetzung von Reagenzien;
- d) Ersatz von fehlerverdächtigen Geräten oder Reagenzien;
- e) vergleichende Prüfung des Personals, der Geräte und/oder Reagenzien mit einer anderen Prüfstelle.

Die Verwendung der Ergebnisse von Eignungsprüfungen durch Akkreditierungsstellen für Prüfstellen ist in ISO/IEC Guide 43-2:1997, Abschnitt 6 beschrieben.

4.2 Leitlinien für das Eingrenzen der Unsicherheit des zugewiesenen Werts

Der zugewiesene Wert X hat eine Standardunsicherheit u_X . Diese Unsicherheit hängt vom Verfahren ab, das verwendet wird, um sie bestimmen. Wenn sie durch Prüfungen in mehreren Prüfstellen bestimmt wird, hängt diese Unsicherheit auch von der Anzahl von Prüfstellen und möglicherweise von weiteren Faktoren ab. Verfahren zur Berechnung der Standardunsicherheit des zugewiesenen Werts werden in Abschnitt 5 angegeben.

Die Standardabweichung für die Eignungsprüfung, $\hat{\sigma}$, wird verwendet, um die Größe von Schätzwerten der in einer Eignungsprüfung gefundenen systematischen Abweichungskomponente einer Prüfstelle zu beurteilen. Verfahren zur Bestimmung der Standardabweichung für die Eignungsprüfung werden in Abschnitt 6 angegeben und Kriterien, die sie mit Schätzwerten der systematischen Abweichungskomponente einer Prüfstelle vergleichen, werden in Abschnitt 7 gegeben.

Wenn die Standardunsicherheit u_X des zugewiesenen Werts zu groß im Vergleich zur Standardabweichung für die Eignungsprüfung, $\hat{\sigma}$, ist, dann besteht ein Risiko, dass einige Prüfstellen Eingriffs- oder Warnsignale erhalten, dies jedoch nicht wegen einer Ursache innerhalb der Prüfstellen, sondern auf Grund von Ungenauigkeiten bei der Bestimmung des zugewiesenen Werts. Daher muss die Standardunsicherheit des zugewiesenen Werts bestimmt und den an Eignungsprüfungsplänen teilnehmenden Prüfstellen mitgeteilt werden (siehe ISO/IEC Guide 43-1:1997, A.1.4 und A.1.6).

Wenn gilt

$$u_X \leq 0,3\hat{\sigma} \quad (1)$$

dann ist die Unsicherheit des zugewiesenen Werts vernachlässigbar und muss nicht in die Auswertung der Ergebnisse der Eignungsprüfung einbezogen werden.

Wenn diesen Leitlinien nicht entsprochen wird, dann muss der Koordinator Folgendes berücksichtigen:

- a) Es ist ein Verfahren zur Bestimmung des zugewiesenen Werts zu suchen, dessen Unsicherheit der oben genannten Leitlinie entspricht.
- b) Es ist bei der Auswertung der Ergebnisse der Eignungsprüfung die Unsicherheit des zugewiesenen Werts zu verwenden (siehe 7.5 bezüglich der E_n -Werte oder 7.6 bezüglich des z' -Werts).
- c) Die Teilnehmer der Eignungsprüfung sind zu informieren, dass die Unsicherheit des zugewiesenen Werts nicht vernachlässigbar ist.

BEISPIEL Angenommen, der zugewiesene Wert X wird als Mittelwert $\bar{x}$ der Prüfungsergebnisse von 11 Prüfstellen bestimmt und die Standardabweichung für die Eignungsprüfung als die Standardabweichung s derselben 11 Ergebnisse, also $\hat{\sigma} \equiv s$. In erster Näherung kann die Standardunsicherheit des zugewiesenen Werts in dieser Situation zu $u_X = s/\sqrt{11} = 0,3 s$ abgeschätzt werden, so dass die Anforderung scheinbar erfüllt ist. Jedoch kann die Anforderung in dieser Situation mit weniger als 11 Prüfstellen nicht erfüllt werden. Weiter wird die Unsicherheit des zugewiesenen Werts größer als $s/\sqrt{11}$ sein, wenn die Stichproben inhomogen oder instabil sind oder wenn es einen Faktor gibt, der eine gemeinsame systematische Abweichungskomponente in den Ergebnissen der Prüfstellen verursacht (zum Beispiel wenn alle dasselbe Referenznormal verwenden).

4.3 Leitlinien für die Festlegung der Anzahl von Wiederholmessungen

Wiederholstreuung trägt zur Streuung in den systematischen Abweichungskomponenten von Prüfstellen bei einer Eignungsprüfung bei. Wenn die Wiederholstreuung zu groß im Vergleich zur Standardabweichung für die Eignungsprüfung ist, dann besteht ein Risiko, dass die Wiederholstreuung zu irregulär streuenden Werten (Ausreißer) in den Ergebnissen der Eignungsprüfung führt. In dieser Situation besteht die Möglichkeit, dass eine Prüfstelle bei einem Durchlauf eine große systematische Abweichungskomponente aufweist, im nächsten Durchlauf jedoch nicht, und die Ursache dafür wird nur schwierig zu ermitteln sein.

Wenn es als wünschenswert betrachtet wird, den Einfluss der Wiederholstreuung einzugrenzen, ist daher die Anzahl n der Wiederholmessungen, die von jeder Prüfstelle bei einer Eignungsprüfung durchzuführen sind, so zu wählen, dass

$$\sigma_r / \sqrt{n} \leq 0,3\hat{\sigma} \quad (2)$$

Dabei ist σ_r die Wiederholstandardabweichung, die bei einem vorangegangenen Ringversuch ermittelt wurde.

Der Faktor von 0,3 ist dadurch gerechtfertigt, dass die Wiederholstandardabweichung nicht mehr als etwa 10% zur Standardabweichung für die Eignungsprüfung beiträgt, wenn dieses Kriterium erfüllt ist.

Weiterhin müssen alle Prüfstellen *dieselbe* Anzahl von Wiederholmessungen durchführen. (Die später in dieser Internationalen Norm angegebenen Auswertungsverfahren basieren auf der Annahme, dass diese Anforderung erfüllt ist.) Wenn die durch Ungleichung (2) definierte Anforderung nicht erfüllt wird, muss die Anzahl der Wiederholmessungen erhöht werden, oder die Ergebnisse der Eignungsprüfung müssen mit Vorsicht interpretiert werden.

Dieser Ansatz geht davon aus, dass Prüfstellen im Allgemeinen ähnliche Wiederholpräzisionen aufweisen. Es können sich Fälle ergeben, bei denen diese Annahme nicht erfüllt ist. Um die Verfahren, die in dieser Internationalen Norm beschrieben werden, anzuwenden, darf in solchen Fällen wie folgt vorgegangen werden. Der Koordinator sollte die Anzahl von Wiederholmessungen n mit Hilfe eines typischen Werts für die Wiederholstandardabweichung festsetzen. Dann sollte jede Prüfstelle überprüfen, ob sie Ungleichung (2) mit ihrer eigenen Wiederholstandardabweichung erfüllt. Wenn dies nicht der Fall ist, dann sollte sie ihr Messverfahren so modifizieren, dass sie ein Prüfergebnis als Mittelwert einer Anzahl von Bestimmungen erhält, die so gewählt wurde, dass Ungleichung (2) erfüllt ist.

4.4 Homogenität und Stabilität von Stichproben (siehe ISO/IEC Guide 43-1:1997, 5.6.2 und 5.6.3)

Anhang B gibt Verfahren an, die geeignet sind zu überprüfen, ob die bei einer Eignungsprüfung zu verwendenden Stichproben hinreichend homogen und stabil sind.

Wenn ein Verfahren zur Stichprobenvorbereitung angewendet wird, so dass das Homogenitätskriterium in Anhang B nicht erfüllt wird, müssen die Teilnehmer Wiederholstichproben prüfen. Alternativ muss die Standardabweichung für die Eignungsprüfung einen Zuschlag für die Heterogenität der Stichproben beinhalten, wie in Anhang B beschrieben.

4.5 Betriebsspezifische Messverfahren

Bei einem betriebsspezifischen Messverfahren ist das Messergebnis durch das Messverfahren bestimmt. Zum Beispiel kann die Größenverteilung eines partikulären Materials mit Hilfe von Sieben mit runden oder quadratischen Löchern bestimmt werden. Es mag keinen Grund geben, warum eine bestimmte Art von Sieb zu bevorzugen ist, wenn aber die Art des Siebs nicht vorgegeben ist, können Prüfstellen, die verschiedene Arten von Sieben verwenden, unterschiedliche Ergebnisse erhalten. Wenn ein Teilnehmer ein anderes Verfahren anwendet als dasjenige, welches angewendet wird, um den zugewiesenen Wert zu bestimmen, dann können seine Ergebnisse selbst dann eine systematische Abweichungskomponente aufweisen, wenn kein Fehler in der Durchführung vorliegt.

Wird den Teilnehmern Wahlfreiheit zwischen betriebsspezifischen Verfahren eingeräumt, kann es sein, dass kein gültiger Konsens unter ihnen ersichtlich ist. Dieses Problem kann auf zweierlei Weise gelöst werden:

- a) Wenn die Teilnehmer routinemäßig ein standardisiertes Verfahren anwenden, wird dieses zur Ermittlung des zugewiesenen Werts angewendet, und die Teilnehmer werden angewiesen, dieses Verfahren in der Eignungsprüfung anzuwenden;
- b) für jedes angewendete Verfahren wird ein eigener zugewiesener Wert bestimmt.

Eine ähnliche Situation ergibt sich, wenn die Messgröße vorgegeben wird, aber nicht das Verfahren, und es muss dieselbe Wahl getroffen werden.

4.6 Angabe von Daten

(siehe ISO/IEC Guide 43-1:1997, 6.2.3)

Für die Zwecke der in Eignungsprüfungen erforderlichen Rechnungen wird empfohlen, dass einzelne Messergebnisse nicht um mehr als $\sigma_r/2$ gerundet werden sollten.

Teilnehmer müssen angehalten werden, ihre tatsächlichen Messergebnisse anzugeben. Messergebnisse dürfen nicht verkürzt als einseitig beschränkte Intervalle angegeben werden (das heißt Ergebnisse dürfen nicht in der Form "< 0,1" oder "unterhalb der Nachweisgrenze" wiedergegeben werden.) Gleichmaßen gilt: Wenn ein negatives Ergebnis beobachtet wird, muss der tatsächliche negative Wert angegeben werden (das heißt Ergebnisse dürfen auch dann nicht als null wiedergegeben werden, wenn das Messergebnis logischerweise nicht negativ sein kann). Die Teilnehmer müssen darüber in Kenntnis gesetzt werden, dass alle Daten für eine Stichprobe von der Auswertung ausgeschlossen werden, wenn für diese Stichprobe verkürzte Ergebnisse oder der Wert Null an Stelle von negativen Ergebnissen angegeben werden. Falls notwendig, darf auf dem zur Angabe der Ergebnisse verwendeten Formular die Möglichkeit vorgesehen werden, dass der Teilnehmer ankreuzt, dass ein Ergebnis unterhalb der Nachweisgrenze liegt.

4.7 Gültigkeitsdauer der Ergebnisse der Eignungsprüfung

Die Gültigkeitsdauer des von einer Prüfstelle in einem einzelnen Durchlauf eines Eignungsprüfungsplans ermittelten Ergebnisses ist auf die Zeit beschränkt, während der die Prüfstelle die Prüfung durchführte. Wenn also eine Prüfstelle bei einem einzelnen Durchlauf ein zufrieden stellendes Ergebnis erzielt, darf dieses Ergebnis nicht als Beleg für die Behauptung verwendet werden, die Prüfstelle hätte bei anderen Gelegenheiten zuverlässige Daten gewonnen.

Eine Prüfstelle, die ein Qualitätssystem anwendet und über eine große Zahl von Durchläufen eines Eignungsprüfungsplans zufrieden stellende Ergebnisse dokumentieren kann, ist berechtigt, diese Ergebnisse als Beleg dafür zu verwenden, dass sie in der Lage ist, konsistent zuverlässige Daten zu erhalten.

5 Bestimmung des zugewiesenen Werts und der zugehörigen Standardunsicherheit

5.1 Wahl des Verfahrens zur Ermittlung des zugewiesenen Werts

Fünf Arten, den zugewiesenen Wert X zu bestimmen, werden in 5.2 bis 5.6. beschrieben. Die Wahlmöglichkeit zwischen diesen Verfahren liegt in der Verantwortung des Koordinators in Abstimmung mit technischen Fachleuten, wie in ISO/IEC Guide 43-1 beschrieben. Die in 5.5 und 5.6 beschriebenen Verfahren sind wahrscheinlich nicht anwendbar, wenn die Anzahl der an einem Plan teilnehmenden Prüfstellen klein ist. Die Verfahren zur Berechnung der Standardunsicherheit u_X des zugewiesenen Werts, die in diesem Abschnitt angegeben werden, werden üblicherweise für diejenigen Anwendungen geeignet sein, für die sie in dieser Internationalen Norm angewendet werden. Andere Verfahren mit gesicherter statistischer Grundlage dürfen angewendet werden; das angewendete Verfahren ist dann im dokumentierten Plan zu beschreiben.

Die Bestimmung des zugewiesenen Werts liegt in der Verantwortung des Koordinators. Der zugewiesene Wert darf den Teilnehmern erst offenbart werden, wenn sie ihre Ergebnisse beim Koordinator gemeldet haben. Der Koordinator muss einen Bericht erstellen, aus dem Einzelheiten der Bestimmung des zugewiesenen Werts, die Identitäten der Prüfstellen, die an seiner Bestimmung beteiligt waren, und Angaben zur Rückverfolgbarkeit und Messunsicherheit des zugewiesenen Werts hervorgehen.

Der *Leitfaden zur Angabe der Unsicherheit beim Messen* gibt Hinweise zur Beurteilung von Messunsicherheiten.

Diese Internationale Norm empfiehlt die Verwendung von robusten statistischen Verfahren, wenn diese als die am besten geeigneten Verfahren angesehen werden (zum Beispiel wie in 5.5 und 5.6). Alternativ dürfen Verfahren zur Erkennung und Beseitigung von Ausreißern zur Anwendung gebracht werden, vorausgesetzt, dass die statistische Grundlage dieser Verfahren abgesichert ist, und dass das angewendete Verfahren dokumentiert wird. Hinweise für die Anwendung von Prüfungen auf Ausreißer werden in ISO 5725-2 angegeben.

5.2 Stichprobenherstellung

[siehe ISO/IEC Guide 43-1:1997, A.1.1, Aufzählung a)]

5.2.1 Allgemeines

Das Prüfmaterial darf durch Mischen von Bestandteilen in festgelegten Anteilen oder durch das Hinzufügen eines festgelegten Anteils eines Bestandteils zu einem Ausgangsstoff hergestellt werden. In diesem Fall wird der zugewiesene Wert X durch Berechnung aus den verwendeten Massen hergeleitet.

Dieser Ansatz ist besonders wertvoll, wenn auf diese Weise Einzelstichproben vorbereitet werden können und die Anteile der Bestandteile oder des Zusatzstoffs zu bestimmen sind: es besteht dann keine Notwendigkeit, eine Sammelstichprobe vorzubereiten und sicherzustellen, dass diese homogen ist. Wenn sich jedoch Stichproben ergeben, bei denen der Zusatzstoff weniger fest gebunden ist als in üblichen Materialien, oder bei denen er in einer unterschiedlichen Form vorliegt, so kann die Anwendung eines anderen Ansatzes sinnvoller sein.

5.2.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn der zugewiesene Wert aus der Zusammensetzung des Prüfmaterials berechnet wird, wird die Standardunsicherheit durch Kombination der Unsicherheiten nach dem im *Leitfaden zur Angabe der Unsicherheit beim Messen* beschriebenen Ansatz abgeschätzt. Zum Beispiel werden die Unsicherheiten in chemischen Analysen normalerweise jene sein, die mit gravimetrischen und volumetrischen Messungen verbunden sind.

Die Einschränkung dieses Verfahrens (bei der chemischen Analyse) besteht darin, dass Sorgfalt nötig ist, um sicherzustellen, dass

- a) der Ausgangsstoff effektiv keinen Anteil des hinzugefügten Bestandteils enthält oder dass der Anteil des hinzugefügten Bestandteils im Ausgangsstoff genau bekannt ist;
- b) die Bestandteile homogen vermischt werden (soweit erforderlich);
- c) alle Fehlerquellen ermittelt werden (zum Beispiel wird nicht immer erkannt, dass Glas Quecksilberverbindungen absorbiert, so dass die Konzentration einer wässrigen Lösung einer Quecksilberverbindung durch den Behälter verändert werden kann);
- d) keine Wechselwirkung zwischen den Bestandteilen und der Stichprobenmatrix stattfindet.

5.2.3 Beispiel: Bestimmung des Zementgehalts von Festbeton

In diesem Fall können Betonstichproben durch das Abwiegen von Mengen der Bestandteile (Zement, Zuschläge und Wasser) und Mischen der Komponenten erstellt werden, um jede einzelne Betonstichprobe zu erzeugen. Der Ansatz ist zufrieden stellend, weil die Genauigkeit, mit der Stichproben erstellt werden können, weit höher ist als bei dem analytischen Verfahren zur Bestimmung des Zementgehalts.

5.3 Zertifizierte Referenzwerte

[siehe ISO/IEC Guide 43-1:1997, A.1.1 Aufzählung b)]

5.3.1 Allgemeines

Wenn das bei einer Eignungsprüfung verwendete Material ein zertifiziertes Referenzmaterial (CRM – certified reference material) ist, wird sein zertifizierter Referenzwert als der zugewiesene Wert X verwendet.

5.3.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn ein zertifiziertes Referenzmaterial als das Prüfmaterial verwendet wird, wird die Standardunsicherheit des zugewiesenen Werts aus den Angaben zur Unsicherheit hergeleitet, die dem Zertifikat zu entnehmen sind.

Die Einschränkung dieses Ansatzes besteht darin, dass die Bereitstellung einer Stichprobe des zertifizierten Referenzmaterials für jeden Teilnehmer einer Eignungsprüfung aufwändig sein kann.

5.3.3 Beispiel: Los-Angeles-Wert von Zuschlagstoffen

Der "Los-Angeles-Wert" ist ein Maß für die mechanische Festigkeit von Zuschlagstoffen, die im Straßenbau verwendet werden, und Ergebnisse der Prüfung werden in "LA-Einheiten" ausgedrückt. Anlässlich einer Übung, die das Ziel verfolgte, ein Referenzmaterial zu zertifizieren, wurde eine große Anzahl von Stichproben eines bestimmten Zuschlagstoffs vorbereitet, und einige dieser Stichproben wurden bei einem Ringversuch verwendet, an dem 28 Prüfstellen beteiligt waren. Dieser Ringversuch gestattete die Bestimmung eines zugewiesenen Werts von $X_{\text{CRM}} = 21,62$ LA-Einheiten mit einer Standardunsicherheit von $u_{X_{\text{CRM}}} = 0,26$ LA-Einheiten. Die verbleibenden Stichproben dieses Zuschlagstoffs konnten bei Eignungsprüfungen verwendet werden.

5.4 Referenzwerte

[siehe ISO/IEC Guide 43-1:1997, A.1.1 Aufzählung c)]

5.4.1 Allgemeines

Bei diesem Ansatz werden zunächst Stichproben des Prüfmaterials, das als Referenzmaterial (RM – reference material) vorgesehen ist, zur Verteilung an die Teilnehmer vorbereitet. Eine Anzahl zufällig

ausgewählter Stichproben wird dann zusammen mit zertifizierten Referenzmaterialien in einer Prüfstelle, die ein geeignetes Messverfahren anwendet, und unter Wiederholbedingungen (wie definiert in ISO 3534-2) geprüft. Der zugewiesene Wert X_{RM} des Prüfmateri als wird dann von einer Kalibrierung gegen die zertifizierten Referenzwerte der CRM abgeleitet.

5.4.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn der zugewiesene Wert eines Prüfmateri als von den Ergebnissen einer Reihe von Prüfungen an diesem Material und am CRM abgeleitet wird, wird die Standardunsicherheit des zugewiesenen Werts aus den Prüfergebnissen und den Unsicherheiten der zertifizierten Referenzwerte des CRMs abgeleitet. Wenn das Prüfmateri al und das CRM (hinsichtlich Matrix, Zusammensetzung und Größenordnung der Ergebnisse) nicht ähnlich sind, dann ist die sich daraus ergebende Unsicherheit ebenfalls einzubeziehen.

Dieses Verfahren gestattet es, den zugewiesenen Wert in einer Weise zu bestimmen, die auf die zertifizierten Werte der CRM rückführbar ist. Seine Standardunsicherheit kann berechnet werden. Das Verfahren vermeidet auch die Kosten für die Verteilung des CRMs an alle Teilnehmer. Dies sind gute Gründe dafür, es anderen Verfahren vorzuziehen. Jedoch geht das Verfahren davon aus, dass es keine Wechselwirkungen zwischen den verwendeten Materialien und den Prüfbedingungen gibt.

Das Beispiel in 5.4.3 veranschaulicht die Berechnung der erforderlichen Unsicherheit für den einfachen Fall, in welchem der zugewiesene Wert eines Prüfmateri als durch direkten Vergleich mit einem einzelnen CRM ermittelt wird.

5.4.3 Beispiel: Los-Angeles-Wert von Zuschlagstoffen

Das im Beispiel in 5.3 beschriebene CRM kann verwendet werden, um den zugewiesenen Wert für ein RM zu bestimmen, wobei es sich um einen anderen, ähnlichen Zuschlagstoff handelt. Diese Bestimmung erfordert die Durchführung einer Reihe von Prüfungen an Stichproben der beiden Zuschlagstoffe in einer Prüfstelle unter Verwendung desselben Messverfahrens und unter Wiederholbedingungen. Es sei

X_{CRM} der zugewiesene Wert für das CRM

X_{RM} der zugewiesene Wert für das RM

D_i die Differenz (RM – CRM) zwischen den gemittelten Ergebnissen für das RM und das CRM bei den Stichproben mit der Zählnummer i

$\bar{D}$ der Mittelwert der Differenzen D_i

Dann gilt

$$X_{\text{RM}} = X_{\text{CRM}} + \bar{D} \quad (3)$$

Die Standardunsicherheit des zugewiesenen Werts des RM kann wie folgt berechnet werden:

$$u_{X;\text{RM}} = \sqrt{u_{X;\text{CRM}}^2 + u_D^2} \quad (4)$$

Tabelle 1 gibt ein Beispiel für Daten, die sich bei einer solchen Reihe von Prüfungen ergeben könnten, und zeigt, wie die Standardunsicherheit u_D der Differenz berechnet wird.

Mit diesen Ergebnissen gilt

$$X_{\text{RM}} = 21,62 + 1,73 = 23,35 \text{ LA-Einheiten} \quad (5)$$

und

$$u_{X, RM} = \sqrt{0,26^2 + 0,24^2} = 0,35 \text{ LA-Einheiten} \quad (6)$$

Dabei ist 0,26 die Standardunsicherheit des zugewiesenen Werts für das (im Beispiel in 5.3 gegebene) CRM, und 0,24 ist die Standardunsicherheit von $\bar{D}$.

Table 1 — Berechnung der mittleren Differenz zwischen einem CRM und einem RM und der Standardunsicherheit dieser Differenz

Stichprobe	RM		CRM		Differenz der Mittelwerte RM – CRM LA-Einheiten
	Prüfung 1 LA-Einheiten	Prüfung 2 LA-Einheiten	Prüfung 1 LA-Einheiten	Prüfung 2 LA-Einheiten	
1	20,5	20,5	19,0	18,0	2,00
2	21,1	20,7	19,8	19,9	1,05
3	21,5	21,5	21,0	21,0	0,50
4	22,3	21,7	21,0	20,8	1,10
5	22,7	22,3	20,5	21,0	1,75
6	23,6	22,4	20,3	20,3	2,70
7	20,9	21,2	21,5	21,8	–0,60
8	21,4	21,5	21,9	21,7	–0,35
9	23,5	23,5	21,0	21,0	2,50
10	22,3	22,9	22,0	21,3	0,95
11	23,5	24,1	20,8	20,6	3,10
12	22,5	23,5	21,0	22,0	1,50
13	22,5	23,5	21,0	21,0	2,00
14	23,4	22,7	22,0	22,0	1,05
15	24,0	24,2	22,1	21,5	2,30
16	24,5	24,4	22,3	22,5	2,05
17	24,8	24,7	22,0	21,9	2,80
18	24,7	25,1	21,9	21,9	3,00
19	24,9	24,4	22,4	22,6	2,15
20	27,2	27,0	24,5	23,7	3,00
Mittlere Differenz, $\bar{D}$					1,73
Standardabweichung					1,07
Standardunsicherheit von $\bar{D}$ (Standardabweichung/ $\sqrt{20}$)					0,24
ANMERKUNG Bei den Daten handelt es sich um Messungen der mechanischen Festigkeit des Zuschlagstoffs nach der Los-Angeles-(LA-)Prüfung.					

5.5 Konsenswerte erfahrener Prüfstellen

[siehe ISO/IEC Guide 43-1:1997, A.1.1 Aufzählung d)]

5.5.1 Allgemeines

Wie beim Ansatz unter Verwendung von *Referenzwerten* (5.4) werden zunächst Stichproben des Prüfmaterials zur Verteilung an die Teilnehmer vorbereitet. Einige dieser Stichproben werden dann von einer Gruppe von erfahrenen Prüfstellen zufällig ausgewählt und analysiert. Alternativ kann die Gruppe von erfahrenen Prüfstellen Teilnehmer einer Runde eines Eignungsprüfungsplans sein, wenn der zugewiesene Wert und seine Unsicherheit nach Beendigung des Durchlaufs abgeleitet werden. Der zugewiesene Wert X wird berechnet als nach Algorithmus A im Anhang C berechneter robuster Mittelwert der Ergebnisse, die von der Gruppe von erfahrenen Prüfstellen angegeben werden.

Anstelle von Algorithmus A dürfen andere Rechenverfahren angewendet werden, vorausgesetzt, dass sie eine verlässliche statistische Grundlage haben und das verwendete Verfahren im Messbericht beschrieben wird.

5.5.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn jede einzelne von p erfahrenen Prüfstellen eine Messung x_i am Prüfmaterial zusammen mit einem Schätzwert u_i der Standardunsicherheit der Messung angibt und der zugewiesene Wert X als robuster Mittelwert nach Algorithmus A berechnet wird, wird die Standardunsicherheit des zugewiesenen Werts X wie folgt abgeschätzt:

$$u_X = \frac{1,25}{p} \times \sqrt{\sum_{i=1}^p u_i^2} \quad (7)$$

Geben die erfahrenen Prüfstellen keine Standardunsicherheiten an oder sind die Unsicherheiten nicht unabhängig validiert (zum Beispiel durch eine Akkreditierungsstelle für Prüfstellen), ist die Standardunsicherheit des zugewiesenen Werts wie in 5.6 beschrieben abzuschätzen.

ANMERKUNG Der Faktor 1,25 bildet das Verhältnis der Standardabweichung des Medians zur Standardabweichung des arithmetischen Mittelwerts für große Stichprobenumfänge ($p > 10$) bei einer Normalverteilung ab. Bei normalverteilten Daten ist die Standardabweichung eines nach dem Algorithmus im Anhang C berechneten robusten Mittelwerts unbekannt; sie wird jedoch zwischen der Standardabweichung des arithmetischen Mittelwerts und der Standardabweichung des Medians liegen, so dass die Gleichung einen konservativen Schätzwert für die Standardunsicherheit u_X ergibt. Für $p < 10$ ist der geeignete Faktor kleiner als 1,25, so dass die Gleichung doppelt konservativ ist.

Die Einschränkungen dieses Ansatzes bestehen darin, dass die Ergebnisse der Gruppe von erfahrenen Prüfstellen eine unbekannte systematische Abweichungskomponente enthalten können und dass die angegebenen Unsicherheiten unzuverlässig sein können.

5.5.3 Beispiel: Petrografische Analyse von Zuschlagstoffen

Dieser Ansatz kann angewendet werden, wenn Stichproben der Zuschlagstoffe verteilt werden sollen und die Teilnehmer die Aufgabe haben, die petrografische Zusammensetzung der Stichproben zu bestimmen. Die Klassifizierung von Zuschlagstoffen erfordert Sachkenntnis und Erfahrung und es gibt keine Referenzmaterialien. Daher kann in diesem Fall der Konsens einer kleinen Gruppe von Fachleuten der am besten geeignete Weg zur Feststellung von zugewiesenen Werten sein.

5.6 Konsenswert der Teilnehmer

[siehe ISO/IEC Guide 43-1:1997, A.1.1 Aufzählung e)]

5.6.1 Allgemeines

Bei diesem Ansatz ist der zugewiesene Wert X für das bei einem Durchlauf eines Eignungsprüfungsplans verwendete Prüfmaterial der robuste Mittelwert der Ergebnisse, die von allen Teilnehmern an diesem Durchlauf angegeben werden; er wird nach Algorithmus A im Anhang C berechnet.

Andere Berechnungsverfahren als Algorithmus A dürfen verwendet werden, vorausgesetzt, dass sie auf einer gesicherten statistischen Grundlage beruhen, und dass der Bericht das zur Anwendung gebrachte Verfahren beschreibt. So kann zum Beispiel die Berechnung in C.1 bei Gleichung (53) angehalten werden, wenn der Median ermittelt wird, und an Stelle von C.2 darf der Median des Betrags der Abweichung verwendet werden.

Dieser Ansatz kann sich bei einem betriebsspezifischen Messverfahren als besonders nützlich erweisen, vorausgesetzt, dieses Verfahren ist standardisiert.

5.6.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn der zugewiesene Wert als robuster Mittelwert nach Algorithmus A hergeleitet wird, wird die Standardunsicherheit des zugewiesenen Werts X wie folgt abgeschätzt:

$$u_X = 1,25 \times s^* / \sqrt{p} \quad (8)$$

Dabei ist s^* die robuste Standardabweichung der Ergebnisse, berechnet nach Algorithmus A im Anhang C. (Hier gilt als „Ergebnis“ für einen Teilnehmer der Mittelwert aller seiner Ergebnisse mit dem Prüfmaterial.)

Die Einschränkungen dieses Ansatzes bestehen darin, dass

- a) möglicherweise kein tatsächlicher Konsens unter den Teilnehmern besteht;
- b) der Konsens auf Grund von allgemein üblichen Verfahrensfehlern eine systematische Abweichungskomponente enthalten kann, die sich nicht in der Standardunsicherheit des zugewiesenen Werts, die wie oben beschrieben berechnet wird, widerspiegelt.

Keine dieser Bedingungen kommt bei der Ermittlung von Spurenbestandteilen selten vor.

5.6.3 Beispiel: Antikörper-Konzentrationen

Tabelle 2 listet Daten eines Durchlaufs einer Eignungsprüfung auf, bei der Konzentrationen von drei allergenspezifischen IgE-Antikörpern (Immunglobuline E) bestimmt wurden. Bild 2 stellt dieselben Daten in Histogrammform dar.

Zur Anwendung von Algorithmus A werden die Daten zunächst in aufsteigender Reihenfolge sortiert und die anfänglichen robusten Schätzwerte für den Mittelwert und die Standardabweichung werden berechnet. Nachfolgend wird das iterative Verfahren des Algorithmus angewendet. Tabelle 3 stellt diese Berechnungen für die Ergebnisse für den allergenspezifischen IgE-Antikörper d1 aus Tabelle 2 dar.

Die im Rahmen der Anwendung von Algorithmus A erforderlichen Berechnungen können wie folgt in einer Tabellenkalkulation durchgeführt werden.

- a) **Schritt 1:** Die Daten werden in aufsteigender Reihenfolge in eine Spalte eingetragen, wie in Tabelle 3 für Iterationsschritt 0 ausgewiesen. Ihr Mittelwert und ihre Standardabweichung werden berechnet (10,91 und 3,13 in Tabelle 3). Anfängliche Werte für den robusten Mittelwert und die robuste Standardabweichung (10,85 und 3,53 in Tabelle 3) werden nach den Gleichungen in C.1 berechnet.
- b) **Schritt 2:** Die Daten werden in die nächste Spalte kopiert, wie in Tabelle 3 für Iterationsschritt 1 ausgewiesen. Die anfänglichen Werte für den robusten Mittelwert und die robuste Standardabweichung werden unter Verwendung der in C.1 angegebenen Gleichungen verwendet, um die Grenzwerte (5,56

und 16,15 in Tabelle 3) zu berechnen. Werte außerhalb der Grenzwerte werden durch die Grenzwerte ersetzt (2,18 wird durch 5,56 ersetzt, 16,30 durch 16,15). Der neue Mittelwert und die neue Standardabweichung der veränderten Daten (11,03 und 2,81 in Tabelle 3) werden berechnet. Nach den Gleichungen im Anhang C ist der robuste Mittelwert mit diesem Mittelwert (11,03) identisch und die robuste Standardabweichung (3,19) errechnet sich durch Multiplikation der Standardabweichung mit 1,134.

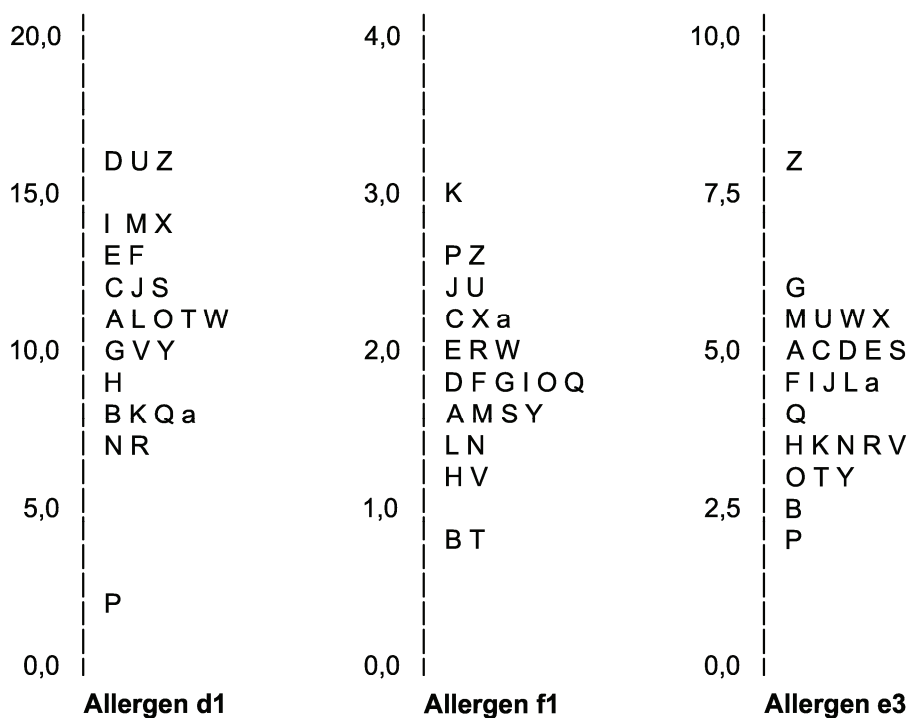
- c) **Schritt 3:** Bei einer Tabellenkalkulation entfällt jetzt die Notwendigkeit, weitere Datenreihen zu erzeugen. Statt dessen wird die Berechnung der Grenzwerte oben in der zweiten Spalte so verändert, dass diese den robusten Mittelwert und die robuste Standardabweichung vom unteren Ende derselben Spalte verwendet. Dies ergibt die Grenzwerte (6,24 und 15,82), die in Tabelle 3 unter dem Iterationsschritt 2 ausgewiesen werden. Die Berechnung kann dann vollendet werden, indem Daten außerhalb der Grenzwerte weiterhin durch die Grenzwerte ersetzt werden, bis die Iterationen konvergieren. Beim Ersetzen von Daten aktualisiert die Tabellenkalkulation automatisch die Mittelwerte und Standardabweichungen sowie Grenzwerte, jedoch werden die Änderungen dieser Werte zunehmend kleiner, bis sie schließlich insignifikant werden.

Die robusten Mittelwerte und Standardabweichungen für die beiden anderen allergenspezifischen IgE-Antikörper werden in ähnlicher Weise berechnet.

Interessanterweise wird das Ergebnis der Prüfstelle P für Antikörper d1 nach dem Grubbschen Test nicht als Ausreißer oder Nachzügler klassifiziert. Werden also bei diesem Beispiel Prüfungen auf Ausreißer nach ISO 5725-2 verwendet, so werden Mittelwert und Standardabweichung aus allen Daten berechnet und es ergeben sich die Werte, die in Tabelle 3 für *Iterationsschritt 0* ausgewiesen sind. Beim robusten Verfahren haben der niedrige Wert der Prüfstelle P und die hohen Werte der Prüfstellen D, U und Z keinen Einfluss auf die Werte der robusten Schätzwerte. Die Balkendiagramme (siehe Bild 9 in 8.3) würden die Werte einiger Prüfstellen als einer weiteren Untersuchung würdig kennzeichnen. So zeigt zum Beispiel Bild 9, dass Prüfstelle Z für alle drei Allergenspiegel hohe positive *z*-Werte aufweist.

**Tabelle 2 — Konzentrationen dreier allergenspezifischer IgE-Antikörper (d1, f1 und e3) —
Daten wie von $p = 27$ Prüfstellen angegeben**

Prüfstelle	Konzentrationen		
	d1 kU/l	f1 kU/l	e3 kU/l
A	11,30	1,69	5,02
B	8,29	0,74	2,52
C	11,90	2,23	5,15
D	15,60	1,76	5,15
E	13,40	1,91	4,84
F	12,50	1,71	4,54
G	10,40	1,88	5,94
H	9,38	1,14	3,50
I	14,20	1,74	4,48
J	12,10	2,39	4,75
K	8,10	3,10	3,70
L	10,80	1,39	4,70
M	13,80	1,52	5,59
N	7,00	1,50	3,40
O	10,85	1,80	2,80
P	2,18	2,52	1,88
Q	8,39	1,83	3,80
R	6,95	1,92	3,52
S	11,80	1,58	4,86
T	10,90	0,80	2,80
U	16,30	2,39	5,60
V	9,71	1,21	3,33
W	10,50	1,93	5,35
X	13,60	2,23	5,53
Y	10,10	1,63	3,18
Z	16,07	2,69	8,22
a	8,47	2,16	4,64
p	27	27	27
Robuster Mittelwert x^*	11,03	1,83	4,35
Robuste Standardabweichung s^*	3,04	0,50	1,25
ANMERKUNG Bei den Daten handelt es sich um Anzahlen von Einheiten (U – Units) in eintausend (k) je Liter (l) Stichprobe; eine Einheit ist definiert als die Konzentration eines internationalen Referenzmaterials.			



ANMERKUNG 1 Bei den Daten handelt es sich um Anzahlen von Einheiten (U - Units) in eintausend (k) je Liter (l) Stichprobe; eine Einheit ist definiert als die Konzentration eines internationalen Referenzmaterials.

ANMERKUNG 2 Die in Tabelle 2 angegebenen Zahlenwerte sind diejenigen, die ermittelt werden, wenn die Berechnung manuell bis auf zwei Dezimalen durchgeführt wird.

**Bild 2 — Konzentrationen dreier allergenspezifischer IgE-Antikörper —
Histogramme für angegebene Daten (Daten aus Tabelle 2)**

**Tabelle 3 — Konzentrationen dreier allergenspezifischer IgE-Antikörper —
Berechnung des robusten Mittelwerts und der robusten Standardabweichung für Antikörper d1**

Iterationsschritt	0	1	2	3	4	5
$\delta = 1,5 s^*$	—	5,30	4,79	4,62	4,58	4,56
$x^* - \delta$	—	5,56	6,24	6,41	6,45	6,47
$x^* + \delta$	—	16,15	15,82	15,65	15,61	15,59
P	2,18	5,56	6,24	6,41	6,45	6,47
R	6,95	6,95	6,95	6,95	6,95	6,95
N	7,00	7,00	7,00	7,00	7,00	7,00
K	8,10	8,10	8,10	8,10	8,10	8,10
B	8,29	8,29	8,29	8,29	8,29	8,29
Q	8,39	8,39	8,39	8,39	8,39	8,39
A	8,47	8,47	8,47	8,47	8,47	8,47
H	9,38	9,38	9,38	9,38	9,38	9,38
V	9,71	9,71	9,71	9,71	9,71	9,71
Y	10,10	10,10	10,10	10,10	10,10	10,10
G	10,40	10,40	10,40	10,40	10,40	10,40
W	10,50	10,50	10,50	10,50	10,50	10,50
L	10,80	10,80	10,80	10,80	10,80	10,80
O	10,85	10,85	10,85	10,85	10,85	10,85
T	10,90	10,90	10,90	10,90	10,90	10,90
A	11,30	11,30	11,30	11,30	11,30	11,30
S	11,80	11,80	11,80	11,80	11,80	11,80
C	11,90	11,90	11,90	11,90	11,90	11,90
J	12,10	12,10	12,10	12,10	12,10	12,10
F	12,50	12,50	12,50	12,50	12,50	12,50
E	13,40	13,40	13,40	13,40	13,40	13,40
X	13,60	13,60	13,60	13,60	13,60	13,60
M	13,80	13,80	13,80	13,80	13,80	13,80
I	14,20	14,20	14,20	14,20	14,20	14,20
D	15,60	15,60	15,60	15,60	15,60	15,59
Z	16,07	16,07	15,82	15,65	15,61	15,59
U	16,30	16,15	15,82	15,65	15,61	15,59
Mittelwert	10,91	11,03	11,03	11,03	11,03	11,03
Standardabweichung	3,13	2,81	2,72	2,69	2,68	2,68
Neuer Wert x^*	10,85	11,03	11,03	11,03	11,03	11,03
Neuer Wert s^*	3,53	3,19	3,08	3,05	3,04	3,04

ANMERKUNG Bei den Daten handelt es sich um Anzahlen von Einheiten (U – Units) in eintausend (k) je Liter (l) Stichprobe; eine Einheit ist definiert als die Konzentration eines internationalen Referenzmaterials.

5.7 Vergleich des zugewiesenen Werts

Werden die in 5.2 bis 5.4 beschriebenen Verfahren zur Bestimmung des zugewiesenen Werts X verwendet, so ist nach jedem Durchlauf eines Eignungsprüfungsplans der robuste Mittelwert x^* , hergeleitet aus den Ergebnissen des Durchlaufs, mit dem zugewiesenen Wert zu vergleichen. Wenn die in 5.5 und 5.6 beschriebenen Verfahren zur Bestimmung des zugewiesenen Werts verwendet werden, muss der Wert, soweit möglich, mit einem Referenzwert verglichen werden, der durch eine sachkundige Prüfstelle ermittelt wurde. Die Standardunsicherheit der Differenz x^*-X ist wie folgt abzuschätzen:

$$\sqrt{\frac{(1,25s^*)^2}{p} + u_X^2}$$

Dabei ist:

s^* die robuste Standardabweichung;

p die Anzahl der Prüfstellen.

Beträgt die Differenz mehr als das Doppelte ihrer Unsicherheit, ist der Grund hierfür festzustellen. Mögliche Gründe sind

- eine systematische Abweichungskomponente im Messverfahren,
- eine gemeinsame systematische Abweichungskomponente in den Ergebnissen der Prüfstellen,
- die Nichtbeachtung der Einschränkungen des Verfahrens bei Anwendung des Verfahrens zur Stichprobenherstellung nach 5.2,
- eine systematische Abweichungskomponente in den Ergebnissen der „erfahrenen Prüfstellen“, wenn der Ansatz des „Konsenswerts erfahrener Prüfstellen“ zu Grunde gelegt wird, und
- Verfahren mit systematischer Abweichungskomponente bei dem oder den Teilnehmern oder mehrere Prüfstellen mit systematischer Abweichungskomponente, wenn der robuste Konsensmittelwert als zugewiesener Wert verwendet wird.

5.8 Fehlende Werte

Wenn die Anzahl der Wiederholmessungen n bei einer Eignungsprüfung zwei oder mehr beträgt, besteht der erste Schritt der Auswertung der Ergebnisse in einer Berechnung des Mittelwerts und der Standardabweichung der Ergebnisse einer jeden Prüfstelle. Die Mittelwerte werden dann beispielsweise benutzt, um die Leistungskenngrößen nach Abschnitt 7 zu bestimmen und um Histogramme oder Balkendiagramme zu erstellen, wie in 8.2 und 8.3 beschrieben. Die Standardabweichungen werden beispielsweise verwendet, um grafische Darstellungen der Maße für die Wiederholpräzision zu erstellen, wie in 8.4 und 8.6 beschrieben.

Wiewohl alle Teilnehmer die Absicht verfolgen mögen, dieselbe Anzahl von Wiederholmessungen zu erzielen, werden sie möglicherweise nicht diese Anzahl von Messungen angeben; dies ist zum Beispiel der Fall, wenn einige Prüfungen misslingen und nicht wiederholt werden können. Wenn diese Situation eintritt, wird die folgende Vorgehensweise empfohlen.

Wenn eine Prüfstelle wenigstens 0,59 n Wiederholmessungen angibt, sind Mittelwert und Standardabweichung ihrer Messungen in die Berechnungen einzubeziehen und so zu behandeln, als wenn n Messungen angegeben worden wären. Der Bericht muss ausweisen, wie viele Messungen diese Prüfstelle angegeben hat.

Gibt eine Prüfstelle weniger als $0,59n$ Wiederholmessungen an, so sind ihre Ergebnisse nicht in die Berechnung der Kenngrößen einzubeziehen, die andere Prüfstellen beeinflussen. Beispielsweise sind ihre Ergebnisse nicht in die Berechnung des zugewiesenen Werts, wie in 5.6 beschrieben, oder die Berechnung der Standardabweichung für die Eignungsbeurteilung, wie in 6.6 beschrieben, einzubeziehen. Ihre Ergebnisse dürfen verwendet werden, um eigene Leistungskenngrößen zu berechnen, wie in Abschnitt 7 beschrieben. Sie dürfen ebenso in grafische Darstellungen, wie in Abschnitt 8 beschrieben, einbezogen werden, doch muss der Bericht ausweisen, wie viele Messungen angegeben wurden und dass diese Anzahl geringer war als die Anzahl, die durch den Prüfplan gefordert wurde.

ANMERKUNG Der Faktor 0,59 wird wie folgt begründet: Die Standardabweichung des Mittelwerts von n Wiederholmessungen beträgt $\sigma_r/\sqrt{n}$. Wird die Anzahl von Wiederholmessungen reduziert, so nimmt diese Standardabweichung zu. Wenn also die tatsächliche Anzahl von Wiederholmessungen von n auf $0,59n$ reduziert wird, nimmt die Standardabweichung um den Faktor 1,3 zu. Dies kann als Grenze für eine annehmbare Zunahme in der Standardabweichung betrachtet werden. Wird daher der Grenzwert von $0,59n$ verwendet, so wird dies eine Zunahme der Standardabweichung über diesen Wert hinaus verhindern. Das Kriterium zur Herleitung dieser Regel beinhaltet offensichtlich ein gewisses Maß an Willkür; ein Koordinator darf es daher in Abstimmung mit den Teilnehmern an dem Plan variieren, wenn der Wunsch besteht.

6 Bestimmung der Standardabweichung für die Eignungsbeurteilung (siehe ISO/IEC Guide 43-1:1997, A.2.1.3)

6.1 Auswahl des Verfahrens

In 6.2 bis 6.6 werden fünf Ansätze zur Bestimmung der Standardabweichung für die Eignungsbeurteilung, $\hat{\sigma}$, beschrieben. Die Wahl zwischen diesen Verfahren liegt in der Verantwortung des Koordinators in Abstimmung mit den Teilnehmern an dem Plan und eventuellen relevanten Akkreditierungsstellen. Zusätzlich sind jeweils anzuwendende Vorschriften zu beachten. Das in 6.6 beschriebene Verfahren lässt sich wahrscheinlich nicht auf Fälle anwenden, in denen die Anzahl der an dem Plan teilnehmenden Prüfstellen gering ist. Die Bestimmung der Standardabweichung $\hat{\sigma}$ obliegt dem Koordinator. Er muss einen Bericht erstellen, aus dem Details der Bestimmung der Standardabweichung hervorgehen.

ANMERKUNG ISO/IEC Guide 43-1 verwendet als Formelzeichen für die Standardabweichung für die Eignungsbeurteilung den Buchstaben s . Dies stimmt mit der üblichen Konvention überein, wenn damit die Standardabweichung einer Stichprobe dargestellt wird. In dieser Internationalen Norm wird die Standardabweichung für die Eignungsbeurteilung mitunter nach anderen Verfahren bestimmt, so dass die Darstellung durch ein anderes Formelzeichen sinnvoll erscheint. Es wurde daher hier $\hat{\sigma}$ gewählt.

6.2 Vorgabewert

6.2.1 Allgemeines

Die Standardabweichung für die Eignungsbeurteilung darf auf einen bestimmten, für eine bestimmte Aufgabe der Datenauswertung notwendigen Wert festgelegt werden; alternativ darf sie aus einer gesetzlich vorgegebenen Anforderung hergeleitet werden.

Dieser Ansatz hat den Vorzug, dass die Standardabweichung für die Eignungsbeurteilung damit direkt mit einer Aussage über die „Gebrauchstauglichkeit“ für das Messverfahren verknüpft ist.

6.2.2 Beispiel: Aflatoxine in Nüssen, Nüsse enthaltenden Produkten, getrockneten Feigen und getrocknete Feigen enthaltenden Produkten

Bestehende Gesetze fordern, dass ein Verfahren zur Prüfung auf Aflatoxine einen Variationskoeffizienten der Vergleichspräzision von nicht mehr als 50 % aufweisen sollte, wenn der vorgeschriebene Grenzwert $10 \mu\text{g/kg}$ beträgt. Wenn also bei einem Eignungsprüfungsplan ein Prüfmaterial mit einem Aflatoxingehalt von $10 \mu\text{g/kg}$ verwendet wird, sollte die Vergleichsstandardabweichung bei diesem Material nach dem Gesetz nicht mehr als

5 µg/kg betragen. In diesem Fall wäre es sinnvoll, die Standardabweichung für die Eignungsbeurteilung ebenfalls auf 5 µg/kg festzusetzen.

6.3 Wahrgenommener Wert

6.3.1 Allgemeines

Die Standardabweichung für die Eignungsbeurteilung könnte auf einen Wert festgesetzt werden, der dem Leistungsfähigkeitsniveau entspricht, das der Koordinator und die Teilnehmer am Plan für ein für Prüfstellen wünschenswertes Ziel halten.

Bei diesem Ansatz ist die Standardabweichung für die Eignungsbeurteilung gleichwertig mit einer Feststellung der „Gebrauchstauglichkeit“ für das Messverfahren.

Wird als Standardabweichung für die Eignungsbeurteilung, $\hat{\sigma}$, ein Vorgabewert oder ein wahrgenommener Wert gewählt, so ist es möglich, dass der gewählte Wert im Hinblick auf die Vergleichpräzision des Messverfahrens unrealistisch ist. Das folgende Verfahren darf angewendet werden, um zu prüfen, ob der gewählte Wert von $\hat{\sigma}$ realistisch ist, vorausgesetzt, dass Angaben zur Wiederholpräzision und Vergleichpräzision des Verfahrens verfügbar sind. Vorgaben:

σ_R ist die Vergleichstandardabweichung und

σ_r ist die Wiederholstandardabweichung.

Die Standardabweichung zwischen Prüfstellen ist wie folgt zu berechnen:

$$\sigma_L = \sqrt{\sigma_R^2 - \sigma_r^2} \quad (9)$$

Hieraus wird dann der Wert des Faktors ϕ errechnet, indem die Werte von σ_L und σ_r und der gewählte Wert von $\hat{\sigma}$ in Gleichung (10) eingesetzt werden.

$$\hat{\sigma} = \sqrt{(\phi \times \sigma_L)^2 + (\sigma_r^2 / n)} \quad (10)$$

Dabei ist n die Anzahl von Wiederholmessungen, die jede Prüfstelle durchzuführen hat.

Wenn der so gefundene Wert für ϕ gering ist (zum Beispiel $\phi < 0,5$), so lässt dies darauf schließen, dass der gewählte Wert von $\hat{\sigma}$ einer Wiederholpräzision entspricht, die von Prüfstellen in der Praxis nicht erreicht werden kann.

6.3.2 Beispiel 1: Messung des Glukosespiegels in menschlichem Serum

Angenommen, es wird akzeptiert, dass medizinische Labors in der Lage sein sollten, den Glukosespiegel im Blut innerhalb von $\pm 10\%$ des zugewiesenen Werts zu bestimmen, obwohl bei extrem niedrigen Konzentrationen (unterhalb von 60 mg/dl) eine Fehlergrenze von ± 6 mg/dl akzeptabel ist. Diese Angabe kann verwendet werden, um wie folgt die Standardabweichung für die Eignungsbeurteilung zu bestimmen:

d) für zugewiesene Werte X unterhalb von 60 mg/dl:

$$\hat{\sigma} = 6,0 / 3,0 = 2,0 \text{ mg/dl,}$$

e) für zugewiesene Werte X oberhalb von 60 mg/dl:

$$\hat{\sigma} = 0,1 X / 3,0 = 0,033 X \text{ mg/dl.}$$

Der hier eingeführte Faktor 3,0 entspricht dem kritischen Wert 3,0, wie er bei der Auswertung der z -Werte verwendet wird (siehe 7.4).

6.3.3 Beispiel 2: Bestimmung des Zementgehalts von ausgehärtetem Beton

Üblicherweise wird der Zementgehalt von Beton als Masse an Zement in Kilogramm je Kubikmeter Beton gemessen (das heißt in kg/m^3). In der Praxis wird Beton in Qualitäten erzeugt, deren Zementgehalte sich um jeweils 25 kg/m^3 unterscheiden, und es ist wünschenswert, dass Prüfstellen die Qualitäten korrekt erkennen können. Aus diesem Grund ist es wünschenswert, dass der gewählte Wert von $\hat{\sigma}$ nicht mehr als die Hälfte von 25 kg/m^3 beträgt. Ein Präzisionsversuch lieferte die folgenden Ergebnisse für einen Beton mit einem mittleren Zementgehalt von 260 kg/m^3 : $\sigma_R = 23,2 \text{ kg/m}^3$ und $\sigma_r = 14,3 \text{ kg/m}^3$.

Damit gilt:

$$\sigma_L = \sqrt{23,2^2 - 14,3^2} = 18,3 \text{ kg/m}^3 \quad (11)$$

Wird n als 2 angenommen und werden $\sigma_L = 18,3 \text{ kg/m}^3$, $\sigma_r = 14,3 \text{ kg/m}^3$ und $\hat{\sigma} = 12,5 \text{ kg/m}^3$ in Gleichung (10) eingesetzt, so ergibt sich:

$$12,5 = \sqrt{(18,3 \cdot \phi)^2 + (14,3^2 / 2)} \quad (12)$$

Hieraus lässt sich $\phi = 0,40$ errechnen. Die Wahl $\hat{\sigma} = 12,5 \text{ kg/m}^3$ lässt also darauf schließen, dass Prüfstellen in der Lage sind, eine Standardabweichung zwischen Prüfstellen zu erzielen, die um den Faktor 0,4 geringer ist als diejenige im Präzisionsversuch. Dies ist eindeutig unrealistisch.

6.4 Wert nach einem allgemeinen Modell

6.4.1 Allgemeines

Der Wert der Standardabweichung für die Eignungsprüfung kann aus einem allgemeinen Modell für die Vergleichpräzision des Messverfahrens abgeleitet werden.

Ein Nachteil dieses Ansatzes liegt darin, dass die wahre Vergleichpräzision eines bestimmten Messverfahrens deutlich von dem nach einem Modell ermittelten Wert abweichen kann, da die Zugrundelegung eines allgemeinen Modells impliziert, dass die Vergleichpräzision ausschließlich von der Höhe der Messgröße abhängt, nicht jedoch von der Messgröße selbst, vom Messverfahren oder vom Stichprobenumfang.

6.4.2 Beispiel: Horwitz-Kurve

Horwitz [3] gibt ein allgemeines Modell für die Vergleichpräzision von Analyseverfahren an, das verwendet werden kann, um den folgenden Ausdruck für die Vergleichsstandardabweichung herzuleiten:

$$\sigma_R = 0,02c^{0,8495} \quad (13)$$

Dabei ist c die Konzentration des zu bestimmenden chemischen Agens in Prozent (Massenanteil).

6.5 Wert aus den Ergebnissen eines Versuchs zur Präzision

6.5.1 Allgemeines

Handelt es sich bei dem für einen Eignungsprüfungsplan anzuwendenden Messverfahren um ein standardisiertes Messverfahren und sind Angaben zur Wiederholpräzision und zur Vergleichpräzision des Verfahrens vorhanden, so kann die Standardabweichung für die Eignungsbeurteilung, $\hat{\sigma}$, auf Grundlage dieser Angaben wie folgt ausgerechnet werden. Vorgaben:

σ_R ist die Vergleichsstandardabweichung und

σ_r ist die Wiederholstandardabweichung.

Die Standardabweichung zwischen Prüfstellen ist wie folgt zu berechnen:

$$\sigma_L = \sqrt{\sigma_R^2 - \sigma_r^2} \quad (14)$$

Dann ist die Standardabweichung für die Eignungsbeurteilung wie folgt zu berechnen:

$$\hat{\sigma} = \sqrt{\sigma_L^2 + (\sigma_r^2 / n)} \quad (15)$$

Dabei ist n die Anzahl von Wiederholmessungen, die jede Prüfstelle während eines Durchlaufs des Plans durchzuführen hat.

Wenn Wiederhol- und Vergleichstandardabweichungen vom Mittelwert der Messergebnisse abhängen, so sind funktionale Zusammenhänge nach den in ISO 5725-2 beschriebenen Verfahren hergeleitet worden. Diese Zusammenhänge sollten dann verwendet werden, um Werte der Wiederhol- und Vergleichstandardabweichungen zu berechnen, die für den für die Eignungsprüfung zu verwendenden zugewiesenen Wert geeignet sind.

6.5.2 Beispiel: Bestimmung des Zementgehalts von ausgehärtetem Beton

Mit denselben Daten wie im Beispiel in 6.3 ergibt Gleichung (15) eine Standardabweichung für die Eignungsprüfung von

$$\hat{\sigma} = \sqrt{18,3^2 + (14,3^2 / 2)} = 20,9 \text{ kg/m}^3 \quad (16)$$

wobei angenommen wurde, dass $n = 2$ Wiederholmessungen durchzuführen sind.

6.6 Wert aus Daten, die während eines Durchlaufs eines Eignungsprüfungsplans ermittelt wurden

6.6.1 Allgemeines

Bei diesem Ansatz wird die Standardabweichung $\hat{\sigma}$, die verwendet wird, um die Eignung von Teilnehmern an einem Durchlauf eines Plans zu beurteilen, aus den Ergebnissen hergeleitet, die die Teilnehmer an eben diesem Durchlauf angeben. Die Standardabweichung muss die robuste Standardabweichung der von allen Teilnehmern angegebenen Ergebnisse sein, berechnet nach Algorithmus A im Anhang C. In diesem Zusammenhang muss das von einem Teilnehmer angegebene Ergebnis der Mittelwert der n Wiederholmessungen sein, den der Teilnehmer während des Durchlaufs erhielt.

An Stelle von Algorithmus A dürfen andere Berechnungsverfahren mit gesicherter statistischer Grundlage angewendet werden; das angewendete Verfahren ist dann im Bericht zu beschreiben.

Ein Nachteil dieses Ansatzes liegt darin, dass der Wert von $\hat{\sigma}$ von Durchlauf zu Durchlauf erheblich schwanken kann; dadurch wird einer Prüfstelle die Verwendung von Werten des z -Werts zur Suche nach Trends, die über mehrere Durchläufe fortbestehen, erschwert. Dieser Nachteil lässt sich bei einem eingeführten Plan dadurch überwinden, dass ein nach Algorithmus S im Anhang C berechneter robuster Poolwert der Standardabweichungen aus einer Anzahl von Durchläufen verwendet wird.

6.6.2 Beispiel: Antikörper-Konzentrationen

Tabellen 2 and 3 liefern ein Beispiel für diesen Ansatz.

6.7 Vergleich von Werten für die Präzision aus einer Eignungsprüfung mit festgestellten Werten

Es wird empfohlen, dass der Koordinator zur Prüfung der Leistungsfähigkeit der Teilnehmer sowie um zu bestimmen, in welchem Maß der Plan den Teilnehmern Nutzen bringt, das folgende Verfahren anwendet. Die während eines Durchlaufs eines Eignungsprüfungsplans erhaltenen Ergebnisse sollten verwendet werden, um Schätzwerte für die Wiederhol- und Vergleichstandardabweichungen des Messverfahrens zu berechnen; hierzu sollten die robusten Verfahren nach ISO 5725-5 angewendet werden. Diese Schätzwerte sollten zusammen mit Werten der Wiederhol- und Vergleichstandardabweichungen, die aus Präzisionsversuchen ermittelt wurden (sofern solche Werte verfügbar sind), als Zeitreihen grafisch aufgetragen werden.

Diese Graphen sollten dann durch den Koordinator geprüft werden. Falls sie aufzeigen, dass die in der Eignungsprüfung ermittelten Präzisionswerte um den Faktor zwei oder mehr von den Werten des Präzisionsversuchs abweichen, sollte der Koordinator den Grund hierfür zu ermitteln suchen. Zeigen sie auf, dass die Präzision des Verfahrens nicht mit der Zeit zunimmt, so deutet dies darauf hin, dass

- die teilnehmenden Prüfstellen die Ursachen für Eingriffs- und Warnsignale nicht untersuchen oder Korrekturmaßnahmen nicht in geeigneter Weise umsetzen;
- die teilnehmenden Prüfstellen nicht in der Lage sind, die Ursachen für Eingriffs- und Warnsignale zu erkennen;
- das Verfahren statistisch beherrscht ist und die nach diesem Verfahren ermittelten Daten zuverlässige Rückschlüsse erlauben.

7 Berechnung von Leistungskenngrößen

7.1 Schätzwerte für die systematische Abweichungskomponente einer Prüfstelle [siehe ISO/IEC Guide 43-1:1997, A.2.1.4 Aufzählung a)]

7.1.1 Allgemeines

Sei x das Ergebnis (oder der Mittelwert der Ergebnisse), wie von einem Teilnehmer für die Messung eines Merkmals des Prüfmaterials bei einem Durchlauf eines Eignungsprüfungsplans angegeben.

Dann lässt sich ein Schätzwert für die systematische Abweichungskomponente D der Prüfstelle bei Messung dieses Merkmals wie folgt berechnen:

$$D = x - X \quad (17)$$

Dabei ist X der zugewiesene Wert.

Leistungskenngrößen, die den Betrag $|D|$ der systematischen Abweichungskomponente der Prüfstelle oder D^2 verwenden, sollten nicht benutzt werden, da sie das Vorzeichen der systematischen Abweichungskomponente nicht erkennen lassen.

7.1.2 Auswertung von systematischen Abweichungskomponenten von Prüfstellen

Wenn ein Teilnehmer ein Ergebnis angibt, auf Grund dessen sich eine systematische Abweichungskomponente der Prüfstelle oberhalb von $3,0 \hat{\sigma}$ oder unterhalb von $-3,0 \hat{\sigma}$, ergibt, so ist das Ergebnis als „Eingriffssignal“ zu werten. In gleicher Weise ist eine systematische Abweichungskomponente einer Prüfstelle oberhalb von $2,0 \hat{\sigma}$ oder unterhalb von $-2,0 \hat{\sigma}$ als „Warnsignal“ zu werten. Ein einzelnes „Eingriffssignal“ während eines Durchlaufs oder zwei „Warnsignale“ bei aufeinander folgenden Durchläufen sind als Beleg dafür zu werten, dass eine Anomalie aufgetreten ist, die untersucht werden muss. Dieses

Kriterium ist mit dem in 7.4 für z -Werte in dem Sinn gleichwertig, dass es dieselben Eingriffs- und Warnsignale ergibt.

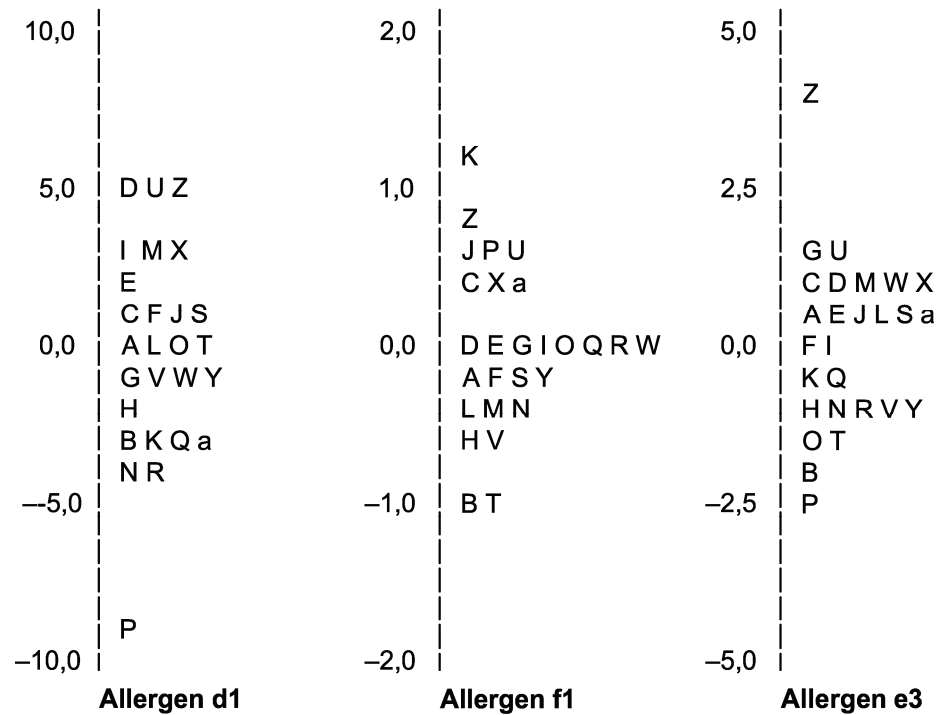
Die Verwendung der Faktoren 2,0 und 3,0 in diesem Zusammenhang (sowie in den ähnlichen, im Folgenden angegebenen Kriterien) wird wie folgt begründet: Sind $\bar{X}$ und $\hat{\sigma}$ gute Schätzwerte für den Mittelwert und die Standardabweichung der Grundgesamtheit, aus der die Werte für x ermittelt werden, und handelt es sich bei der zu Grunde liegenden Verteilung um eine Normalverteilung, so werden die Werte von D näherungsweise mit einer Standardabweichung $\hat{\sigma}$ um den Mittelwert 0 normalverteilt sein. Unter diesen Gegebenheiten sind nur ungefähr 0,3 % der Schätzwerte für die systematischen Abweichungskomponenten der Prüfstellen außerhalb eines Intervalls $-3,0 \hat{\sigma} < D < 3,0 \hat{\sigma}$ zu erwarten, und nur ungefähr 5 % sind außerhalb eines Intervalls $-2,0 \hat{\sigma} < D < 2,0 \hat{\sigma}$ zu erwarten. Bei solch geringen Wahrscheinlichkeiten ist es unwahrscheinlich, dass Eingriffssignale zufällig auftreten, wenn kein tatsächliches Problem vorliegt. Daher besteht eine gute Chance, den Grund für eine Anomalie zu erkennen, wenn ein Eingriffssignal auftritt.

Wird die Standardabweichung für die Eignungsbeurteilung nach einem der Verfahren in 6.2, 6.3 oder 6.4 bestimmt, kann sie erheblich von der Vergleichsstandardabweichung abweichen; die Wahrscheinlichkeiten von 0,3 % und 5,0 % gelten in diesem Fall nicht mehr.

Wenn die Standardabweichung für die Eignungsbeurteilung nach einem der Verfahren in 6.2 oder 6.3 bestimmt wird, kann es für die Teilnehmer hilfreich sein, die geeignete Leistungskenngröße zu verwenden, so dass ein unmittelbarer Vergleich mit der vorgegebenen oder wahrgenommenen Leistungsanforderung möglich ist. So kann man etwa im Beispiel in 6.3, wo der Zielwert für den Fehler in Glukosestichproben ± 10 % des zugewiesenen Werts beträgt, dem Beispiel folgen, eine Standardabweichung für die Eignungsbeurteilung von 3,33 % bestimmen, z -Werte berechnen und 7.4 folgen. Eine Berechnung der systematischen Abweichungskomponenten der Prüfstellen als prozentuale Differenz (wie in 7.2 beschrieben) ist gleichwertig und kann unmittelbar mit dem Zielwert von 10 % verglichen werden.

7.1.3 Beispiel: Antikörper-Konzentrationen

Tabelle 4 stellt die Ergebnisse nach Anwendung dieses Verfahrens auf die Daten aus Tabelle 2 dar und Bild 3 zeigt Histogramme der Schätzwerte der systematischen Abweichungskomponenten der Prüfstellen. Ein Vergleich der Bilder 1 und 2 zeigt, dass den Werten der systematischen Abweichungskomponenten der Prüfstellen (abgesehen von leichten Rundungseffekten) dieselbe Verteilung zu Grunde liegt wie den ursprünglichen Daten, jedoch verteilen sie sich um den Ursprung.



ANMERKUNG Bei den Daten handelt es sich um Anzahlen von Einheiten (U – Units) in eintausend (k) je Liter (l) Stichprobe; eine Einheit ist definiert als die Konzentration eines internationalen Referenzmaterials.

Bild 3 — Konzentrationen dreier allergenspezifischer IgE-Antikörper —
Histogramme von Schätzwerten der systematischen Abweichungskomponenten von Prüfstellen
 (Daten aus Tabelle 4)

Tabelle 4 — Konzentrationen dreier allergenspezifischer IgE-Antikörper (d1, f1 und e3) — Schätzwerte für die systematischen Abweichungskomponenten von Prüfstellen

Prüfstelle	Konzentrationen		
	d1 kU/l	f1 kU/l	e3 kU/l
A	0,27	–0,14	0,67
B	–2,74	–1,09 W	–1,83
C	0,87	0,40	0,80
D	4,57	–0,07	0,80
E	2,37	0,08	0,49
F	1,47	–0,12	0,19
G	–0,63	0,05	1,59
H	–1,65	–0,69	–0,85
I	3,17	–0,09	0,13
J	1,07	0,56	0,40
K	–2,93	1,27 W	–0,65
L	–0,23	–0,44	0,35
M	2,77	–0,31	1,24
N	–4,03	–0,33	–0,95
O	–0,18	–0,03	–1,55
P	–8,85 W	0,69	–2,47
Q	–2,64	0,00	–0,55
R	–4,08	0,09	–0,83
S	0,77	–0,25	0,51
T	–0,13	–1,03 W	–1,55
U	5,27	0,56	1,25
V	–1,32	–0,62	–1,02
W	–0,53	0,10	1,00
X	2,57	0,40	1,18
Y	–0,93	–0,20	–1,17
Z	5,04	0,86	3,87 A
A	–2,56	0,33	0,29
3,0 $\hat{\sigma}$	9,12	1,50	3,75
2,0 $\hat{\sigma}$	6,08	1,00	2,50
–2,0 $\hat{\sigma}$	–6,08	–1,00	–2,50
–3,0 $\hat{\sigma}$	–9,12	–1,50	–3,75

ANMERKUNG 1 Nach Konzentrationsangaben bedeutet A ein Eingriffssignal und W ein Warnsignal.

ANMERKUNG 2 Die systematischen Abweichungskomponenten der Prüfstellen in dieser Tabelle wurden aus den Daten nach Tabelle 2 ermittelt; dabei wurden die robusten Mittelwerte nach Tabelle 2 als zugewiesene Werte für die drei Allergenspiegel verwendet. Die Eingriffs- und Warnsignale unten in der Tabelle wurden unter Verwendung der robusten Standardabweichungen in Tabelle 2 als Standardabweichungen für die Eignungsbeurteilung berechnet.

ANMERKUNG 3 Bei den Daten handelt es sich um Anzahlen von Einheiten (U – Units) in eintausend (k) je Liter (l) Stichprobe; eine Einheit ist definiert als die Konzentration eines internationalen Referenzmaterials.

7.2 Prozentuale Differenzen

[siehe ISO/IEC Guide 43-1:1997, A.2.1.4 Aufzählung b)]

7.2.1 Allgemeines

Bei Verwendung der Notation nach 7.1 werden die prozentualen Differenzen wie folgt berechnet:

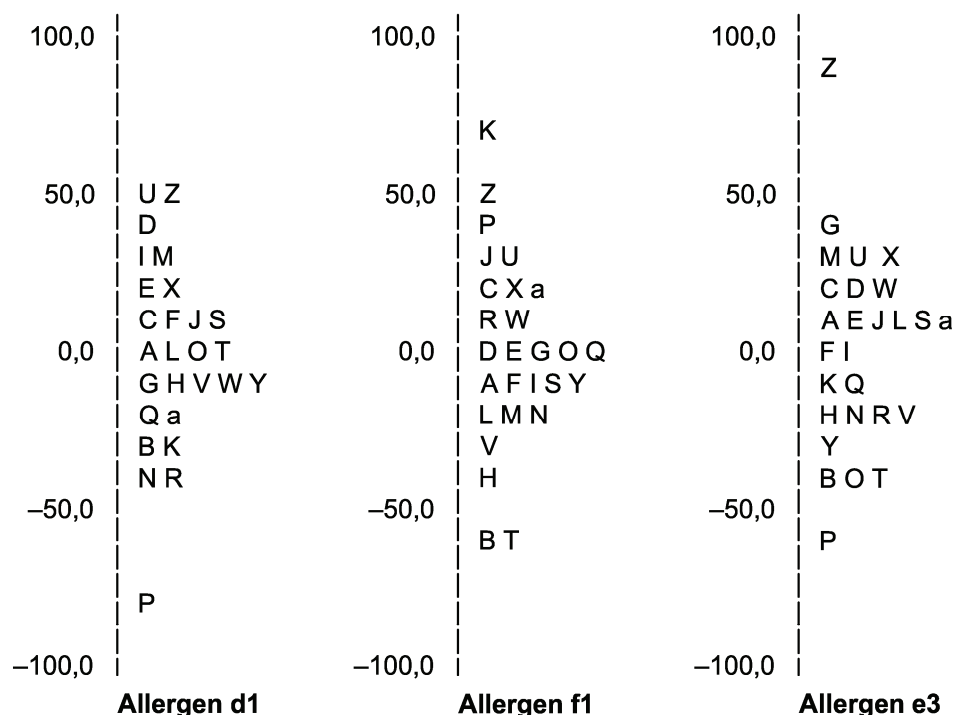
$$D_{\%} = 100 (x - X) / X \quad (18)$$

7.2.2 Auswertung von prozentualen Differenzen

Prozentuale Differenzen sind nach Leitlinien auszuwerten, die gleichwertig zu denjenigen sind, die für systematische Abweichungskomponenten von Prüfstellen in 7.1 angegeben werden. Dies bedeutet: Wenn ein Teilnehmer ein Ergebnis meldet, das zu einer prozentualen Differenz oberhalb von $300 \hat{\sigma}/X$ % oder unterhalb von $-300 \hat{\sigma}/X$ % führt, dann ist dieses Ergebnis als „Eingriffssignal“ zu werten. Gleichmaßen ist eine prozentuale Differenz oberhalb von $200 \hat{\sigma}/X$ % oder unterhalb von $-200 \hat{\sigma}/X$ % als „Warnsignal“ zu werten. Ein einzelnes „Eingriffssignal“ oder aber „Warnsignale“ bei zwei aufeinander folgenden Durchläufen sind als Beleg dafür zu werten, dass eine Anomalie aufgetreten ist, die untersucht werden muss.

7.2.3 Beispiel: Antikörper-Konzentrationen

Tabelle 5 stellt die Ergebnisse nach Anwendung dieses Verfahrens auf die Daten aus Tabelle 2 dar und Bild 4 zeigt Histogramme prozentualer Differenzen. Ein Vergleich der Bilder 1 und 3 zeigt, dass den prozentualen Differenzen (abgesehen von leichten Rundungseffekten) dieselbe Verteilung zu Grunde liegt wie den ursprünglichen Daten, jedoch verteilen sie sich wie die Schätzwerte der systematischen Abweichungskomponenten der Prüfstellen um den Ursprung.



ANMERKUNG Bei den Daten handelt es sich um Anzahlen von Einheiten (U – Units) in eintausend (k) je Liter (l) Stichprobe; eine Einheit ist definiert als die Konzentration eines internationalen Referenzmaterials.

**Bild 4 — Konzentrationen dreier allergenspezifischer IgE-Antikörper —
Histogramme der prozentualen Differenzen (Daten aus Tabelle 5)**

Tabelle 5 — Konzentrationen dreier allergenspezifischer IgE-Antikörper (d1, f1 und e3) — prozentuale Differenzen

Prüfstelle	Prozentuale Differenz		
	d1 %	f1 %	e3 %
A	2	–8	15
B	–25	–60 W	–42
C	8	22	18
D	41	–4	18
E	21	4	11
F	13	–7	4
G	–6	3	37
H	–15	–38	–20
I	29	–5	3
J	10	31	9
K	–27	69 W	–15
L	–2	–24	8
M	25	–17	29
N	–37	–18	–22
O	–2	–2	–36
P	–80 W	38	–57
Q	–24	0	–13
R	–37	5	–19
S	7	–14	12
T	–1	–56	–36
U	48	31	29
V	–12	–34	–23
W	–5	5	23
X	23	22	27
Y	–8	–11	–27
Z	46	47	89 A
a	–23	18	7
300 $\hat{\sigma}/X$	82,7	82,0	86,2
200 $\hat{\sigma}/X$	55,1	54,6	57,5
–200 $\hat{\sigma}/X$	–55,1	–54,6	–57,5
–300 $\hat{\sigma}/X$	–82,7	–82,0	–86,2

ANMERKUNG 1 Nach Konzentrationsangaben bedeutet A ein Eingriffssignal und W ein Warnsignal.

ANMERKUNG 2 Die prozentualen Differenzen in dieser Tabelle wurden aus den Daten nach Tabelle 2 ermittelt; dabei wurden die robusten Mittelwerte nach Tabelle 2 als zugewiesene Werte für die drei Allergenspiegel verwendet. Die Eingriffs- und Warnsignale unten in der Tabelle wurden unter Verwendung der robusten Standardabweichungen in Tabelle 2 als Standardabweichungen für die Eignungsbeurteilung berechnet.

7.3 Rangordnungen und prozentuale Rangordnungen

[siehe ISO/IEC Guide 43-1:1997, A.2.1.4 Aufzählung c)]

7.3.1 Allgemeines

Bei Ergebnissen von p Prüfstellen aus einem Durchlauf einer Eignungsprüfung wird die Rangordnung abgeleitet, indem der Prüfstelle, die den niedrigsten Wert meldet, der Rang 1 zugewiesen wird, derjenigen Prüfstelle, die das zweitniedrigste Ergebnis meldet, Rang 2 usw., bis hin zum Rang p für die Prüfstelle, welche den höchsten Wert meldet. Werden zwei oder mehr identische Ergebnisse gemeldet, so wird diesen derselbe mittlere Rang zugewiesen. So gaben beispielsweise in Tabelle 2 die Prüfstellen C und X beide eine Konzentration von 2,23 für f1 an. Sie teilen sich Rang 21 und 22, und beiden wird in Tabelle 6 der Rang 21,5 zugewiesen. Wenn der Versuch mehrere Messgrößen beinhaltet, werden für jede Messgröße einzeln Rangordnungen zugewiesen.

Werden die Rangordnungen mit $i = 1, 2, \dots, p$ bezeichnet, so werden die prozentualen Rangordnungen als $100(i - 0,5)/p$ % berechnet. Ein Beispiel für die Berechnung von prozentualen Rangordnungen wird in Tabelle 6 dargestellt.

7.3.2 Auswertung von Rangordnungen und prozentualen Rangordnungen

Die Auswertung von Rangordnungen oder prozentualen Rangordnungen basiert nicht auf der Annahme, dass die Daten einer bestimmten Wahrscheinlichkeitsverteilung folgen und ihre Herleitung verwendet nicht den zugewiesenen Wert oder die Standardabweichung für die Eignungsbeurteilung. Daher bieten Rangordnungen und prozentuale Rangordnungen ein einfaches Verfahren zur Erkennung derjenigen Prüfstellen, die die extremen Ergebnisse melden. Als besonders nützlich erweisen sie sich während der frühen Durchläufe eines Eignungsprüfungsplans zur Erkennung derjenigen Prüfstellen, bei denen sich mit der größten Wahrscheinlichkeit Verbesserungen der Leistungsfähigkeit erzielen lassen. Es sollte jedoch der Warnhinweis in ISO Guide 43-1:1997, 6.6.5 beachtet werden: „Die Angabe von Leistungsfähigkeiten durch Anordnung der Prüfstellen in einer nach Leistungsfähigkeit sortierten Tabelle wird bei Eignungsprüfungen nicht empfohlen. Rangordnungen sollten daher nur mit äußerster Vorsicht angewendet werden, da sie irreführend und anfällig für Fehlinterpretationen sein können.“

Tabelle 6 — Konzentrationen dreier allergenspezifischer IgE-Antikörper (d1, f1 und e3) — Rangordnungen und prozentuale Rangordnungen

Prüfstelle	Rangordnung			Prozentuale Rangordnung		
	d1	f1	e3	d1 %	f1 %	e3 %
A	16	10	19	57	35	69
B	5	1	2	17	2	6
C	18	21,5	20,5	65	78	74
D	25	13	20,5	91	46	74
E	21	17	17	76	61	61
F	20	11	13	72	39	46
G	11	16	26	39	57	94
H	8	3	8	28	9	28
I	24	12	12	87	43	43
J	19	23,5	16	69	85	57
K	4	27	10	13	98	35
L	13	5	15	46	17	54
M	23	7	24	83	24	87
N	3	6	7	9	20	24
O	14	14	3,5	50	50	11
P	1	25	1	2	91	2
Q	6	15	11	20	54	39
R	2	18	9	6	65	31
S	17	8	18	61	28	65
T	15	2	3,5	54	6	11
U	27	23,5	25	98	85	91
V	9	4	6	31	13	20
W	12	19	22	43	69	80
X	22	21,5	23	80	78	83
Y	10	9	5	35	31	17
Z	26	26	27	94	94	98
a	7	20	14	24	72	50

7.4 *z*-Werte

[siehe ISO/IEC Guide 43-1:1997, A.2.1.4 Aufzählung d)]

7.4.1 Allgemeines

In der Notation nach 7.2 wird der *z*-Wert wie folgt berechnet:

$$z = (x - \bar{X}) / \hat{\sigma} \quad (19)$$

Dabei ist $\hat{\sigma}$ die Standardabweichung für die Eignungsbeurteilung.

ANMERKUNG ISO/IEC Guide 43 benutzt in der Definition des z -Werts das Formelzeichen s für die Standardabweichung. Dies ist angemessen, wenn diese Größe als Standardabweichung einer Anzahl von Ergebnissen hergeleitet wird, in anderen Fällen jedoch nicht (beispielsweise dann nicht, wenn sie aus den Ergebnissen eines Präzisionsversuchs oder durch Referenznahme auf ein allgemeines Modell, wie in 6.4, berechnet wird).

7.4.2 Auswertung von z -Werten

Wenn ein Teilnehmer ein Ergebnis meldet, auf Grund dessen sich ein z -Wert oberhalb von 3,0 oder unterhalb von -3,0 ergibt, dann ist dieses Ergebnis als „Eingriffssignal“ zu werten. Gleichmaßen ist ein z -Wert oberhalb von 2,0 oder unterhalb von -2,0 als „Warnsignal“ zu werten. Ein einzelnes „Eingriffssignal“ oder aber „Warnsignale“ bei zwei aufeinander folgenden Durchläufen sind als Beleg dafür zu werten, dass eine Anomalie aufgetreten ist, die untersucht werden muss.

Bei Eignungsprüfungsplänen mit sehr großen Anzahlen von Prüfstellen (zum Beispiel mehr als 100 Prüfstellen) dürfen die grafischen Darstellungen der normalverteilten Wahrscheinlichkeiten, wie in 7.9 und/oder Bild 6 dargestellt, benutzt werden, um die Auswertung der z -Werte zu ergänzen. Im anderen Extremfall, wo nur eine kleine Anzahl von Prüfstellen (zum Beispiel weniger als 10 Prüfstellen) teilnimmt, darf kein Signal gegeben werden. In diesem Fall bieten die grafischen Verfahren, welche die Werte über mehrere Durchläufe kombinieren, nutzbringendere Hinweise auf die Leistungsfähigkeit der Prüfstellen als die Ergebnisse einzelner Durchläufe.

7.4.3 Beispiel: Antikörper-Konzentrationen

Tabelle 7 stellt die z -Werte dar, die unter Verwendung der wie in Tabelle 2 berechneten robusten Mittelwerte und Standardabweichungen hergeleitet wurden, und Bild 5 zeigt die entsprechenden Histogramme. Ein Vergleich der Bilder 1 und 4 zeigt, dass den z -Werten (abgesehen von leichten Rundungseffekten) dieselbe Verteilung zu Grunde liegt wie den ursprünglichen Daten, jedoch verteilen sie sich wie die Schätzwerte der systematischen Abweichungskomponenten der Prüfstellen um den Ursprung.

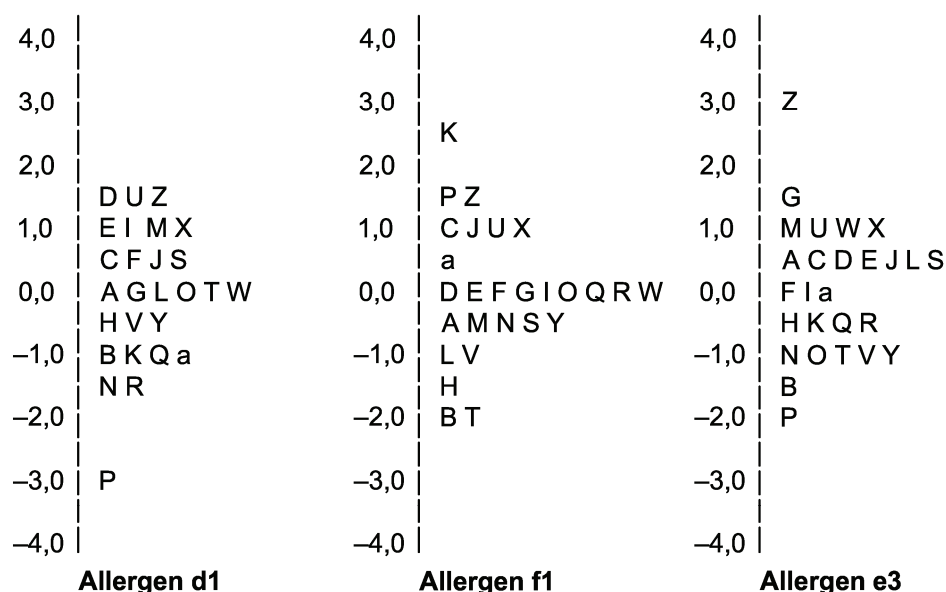


**Bild 5 — Konzentrationen dreier allergenspezifischer IgE-Antikörper —
Histogramme der z -Werte (Daten aus Tabelle 7)**

Tabelle 7 — Konzentrationen dreier allergenspezifischer IgE-Antikörper (d1, f1 und e3) — z-Werte

Prüfstelle	z-Wert		
	d1	f1	e3
A	0,09	−0,28	0,54
B	−0,90	−2,18 W	−1,46
C	0,29	0,80	0,64
D	1,50	−0,14	0,64
E	0,78	0,16	0,39
F	0,48	−0,24	0,15
G	−0,21	0,10	1,27
H	−0,54	−1,38	−0,68
I	1,04	−0,18	0,10
J	0,35	1,12	0,32
K	−0,96	2,54 W	−0,52
L	−0,08	−0,88	0,28
M	0,91	−0,62	0,99
N	−1,33	−0,66	−0,76
O	−0,06	−0,06	−1,24
P	−2,91 W	1,38	−1,98
Q	−0,87	0,00	−0,44
R	−1,34	0,18	−0,66
S	0,25	−0,50	0,41
T	−0,04	−2,06 W	−1,24
U	1,73	1,12	1,00
V	−0,43	−1,24	−0,82
W	−0,17	0,20	0,80
X	0,85	0,80	0,94
Y	−0,31	−0,40	−0,94
Z	1,66	1,72	3,10 A
A	−0,84	0,66	0,23

ANMERKUNG 1 Nach Konzentrationsangaben bedeutet A ein Eingriffssignal und W ein Warnsignal.

ANMERKUNG 2 Die z-Werte in dieser Tabelle wurden aus den Daten nach Tabelle 2 ermittelt; dabei wurden die robusten Mittelwerte nach Tabelle 2 als zugewiesene Werte für die drei Allergenspiegel verwendet und die robusten Standardabweichungen in Tabelle 2 als Standardabweichungen für die Eignungsbeurteilung. Die Gleichung für den z-Wert in diesem Beispiel lautet daher $z = (x - \bar{X}) / \hat{\sigma} = (x - x^*) / s^*$.

7.5 E_n -Werte

[siehe ISO/IEC Guide 43-1:1997, A.2.1.4 Aufzählung e)]

Diese Leistungskenngröße wird wie folgt berechnet:

$$E_n = \frac{x - X}{\sqrt{U_{\text{lab}}^2 + U_{\text{ref}}^2}} \quad (20)$$

Dabei ist:

X der von einer Referenzprüfstelle bestimmte zugewiesene Wert;

U_{ref} die erweiterte Unsicherheit von X ;

U_{lab} die erweiterte Unsicherheit des Ergebnisses x eines Teilnehmers.

Im Gegensatz zu den bei z -Werten verwendeten kritischen Werten von 2,0 und 3,0 ist bei E_n -Werten ein kritischer Wert von 1,0 üblich. Dies liegt daran, dass E_n -Werte unter Verwendung erweiterter Unsicherheiten an Stelle von Standardabweichungen im Nenner berechnet werden.

ANMERKUNG 1 E_n -Werte sollten mit Vorsicht verwendet werden, wenn bei den Teilnehmern möglicherweise nur ein geringes Verständnis ihrer Unsicherheiten vorliegt und Angaben möglicherweise nicht einheitlich erfolgen. Die Einbeziehung von Angaben zur Unsicherheit in die Auswertung der Ergebnisse von Eignungsprüfungen kann jedoch eine wichtige Rolle bei der Verbesserung des Verständnisses für diese schwierige Materie spielen.

Werden die erweiterten Unsicherheiten mit einem Erweiterungsfaktor 2,0 berechnet, so ist der kritische Wert 1,0 für einen E_n -Wert gleichwertig mit dem kritischen Wert 2,0 bei z -Werten.

ANMERKUNG 2 Werden Unsicherheiten im Einklang mit dem *Leitfaden zur Angabe der Unsicherheit beim Messen* (GUM) geschätzt, drücken E_n -Werte die Gültigkeit des zu jedem Ergebnis gehörigen Schätzwerts der erweiterten Unsicherheit aus. Ein Wert mit einem Betrag $|E_n| < 1$ stellt einen objektiven Nachweis dar, dass der Schätzwert der Unsicherheit mit der Definition der erweiterten Unsicherheit im GUM übereinstimmt.

7.6 z' -Werte

7.6.1 Allgemeines

In der Notation nach 7.4 wird der z' -Wert wie folgt berechnet:

$$z' = (x - X) / \sqrt{\hat{\sigma}^2 + u_X^2} \quad (21)$$

Dabei ist u_X die Standardunsicherheit des zugewiesenen Werts X .

Gleichung (21) darf verwendet werden, wenn der zugewiesene Wert nicht aus den von den Teilnehmern gemeldeten Ergebnissen berechnet wird. Sie darf folglich verwendet werden, wenn der zugewiesene Wert nach den in 5.2, 5.3 und 5.4 beschriebenen Verfahren bestimmt wird, sowie dann, wenn das in 5.5 beschriebene Verfahren angewendet wird und die erfahrenen Prüfstellen nicht an der Eignungsprüfung teilnehmen. Wird das in 5.6 beschriebene Verfahren verwendet, besteht eine Korrelation zwischen dem zugewiesenen Wert und den von den Teilnehmern gemeldeten Ergebnissen, so dass die Verwendung von z' -Werten nach der Definition in Gleichung (21) nicht gültig ist.

7.6.2 Auswertung von z' -Werten

z' -Werte sind in derselben Weise auszuwerten wie z -Werte (siehe 7.4); dabei sind dieselben kritischen Werte von 2,0 und 3,0 zu verwenden.

ANMERKUNG Die in 7.1 und 7.2 angegebenen Kriterien für die Auswertung systematischer Abweichungskomponenten von Prüfstellen und prozentualer Differenzen können in ähnlicher Weise modifiziert werden, indem $\hat{\sigma}$ durch $\sqrt{\hat{\sigma}^2 + u_X^2}$ ersetzt wird.

7.6.3 Anwendung von z' -Werten

Ein Vergleich der Gleichungen für den z -Wert und den z' -Wert in 7.4 beziehungsweise 7.6 zeigt, dass die z' -Werte für einen Durchlauf eines Eignungsprüfungsplans sämtlich kleiner sein werden als die entsprechenden z -Werte; zwischen beiden Werten liegt ein konstanter Faktor von

$$\hat{\sigma} / \sqrt{\hat{\sigma}^2 + u_X^2}$$

Wird die Leitlinie für das Eingrenzen der Unsicherheit des zugewiesenen Werts in 4.2 erfüllt, wird dieser Faktor innerhalb folgender Grenzen liegen:

$$0,96 \leq \hat{\sigma} / \sqrt{\hat{\sigma}^2 + u_X^2} \leq 1,00 \quad (22)$$

Daher werden die z' -Werte in diesem Fall nahezu mit den z -Werten übereinstimmen und es darf die Schlussfolgerung gezogen werden, dass die Unsicherheit des zugewiesenen Werts vernachlässigbar ist.

Wird die Leitlinie in 4.2 nicht erfüllt, können die z' -Werte und die z -Werte größenordnungsmäßig so weit voneinander abweichen, dass einige der z -Werte die kritischen Werte von 2,0 oder 3,0 überschreiten und daher „Warnsignale“ beziehungsweise „Eingriffssignale“ ergeben, während die entsprechenden z' -Werte diese kritischen Werte nicht überschreiten und daher auch keine Signale erzeugen.

Bei der Entscheidung, ob entweder die z -Werte oder z' -Werte verwendet werden sollen, muss der Koordinator die folgenden Gesichtspunkte berücksichtigen.

- Entspricht die Unsicherheit des zugewiesenen Werts den Leitlinien in 4.2? Falls ja, ist es unwahrscheinlich, dass die Verwendung von z' -Werten irgendeinen Nutzen bringt.
- Wenn die Leitlinie nicht erfüllt wird, wird trotz der größeren Komplexität und der Schwierigkeiten, diese den Anwendern zu erklären, die Verwendung von z' -Werten empfohlen.
- Wie ernst sind die Konsequenzen für Prüfstellen, wenn deren Ergebnisse Warn- oder Eingriffssignale auslösen? Werden die Ergebnisse verwendet, um Prüfstellen von der Durchführung des Messverfahrens für bestimmte Nutzergruppen auszuschließen?

7.7 Zeta-Werte (ζ)

7.7.1 Allgemeines

In der Notation nach 7.4 wird der ζ -Wert wie folgt berechnet:

$$\zeta = (x - X) / \sqrt{u_x^2 + u_X^2} \quad (23)$$

Dabei ist u_x der prüfstelleneigene Schätzwert der Standardunsicherheit des eigenen Ergebnisses x , und u_X ist die Standardunsicherheit des zugewiesenen Werts X .

Gleichung (23) darf verwendet werden, wenn der zugewiesene Wert nicht aus den von den Teilnehmern gemeldeten Ergebnissen berechnet wird. Sie darf folglich verwendet werden, wenn der zugewiesene Wert nach den in 5.2, 5.3 und 5.4 beschriebenen Verfahren bestimmt wird, sowie dann, wenn das in 5.5 beschriebene Verfahren angewendet wird und die erfahrenen Prüfstellen nicht an der Eignungsprüfung teilnehmen. Wird das in 5.6 beschriebene Verfahren verwendet, besteht eine Korrelation zwischen dem

zugewiesenen Wert und den von den Teilnehmern gemeldeten Ergebnissen, so dass die Verwendung von ζ -Werten nach der Definition in Gleichung (23) nicht gültig ist.

ANMERKUNG 1 ζ -Werte unterscheiden sich von E_n -Werten durch die Verwendung von Standardunsicherheiten $u_{(x)}$ an Stelle von erweiterten Unsicherheiten $U_{(x)}$.

ANMERKUNG 2 Zum gegenwärtigen Zeitpunkt ist die Einbeziehung von Angaben der teilnehmenden Prüfstellen bezüglich der Unsicherheit ihrer eigenen Messungen in die Werte, welche bei Eignungsprüfungsplänen verwendet werden, nicht allgemein üblich. Möglicherweise werden solche Angaben in Zukunft jedoch immer häufiger zur Verfügung gestellt werden. 7.7 wird in diese Internationale Norm aufgenommen, um Koordinatoren eine Leitlinie an die Hand zu geben, in welcher Weise diese Angaben einbezogen werden können, wenn sie vorliegen. Angaben zur Messunsicherheit werden nunmehr durch ISO/IEC 17025 gefordert, so dass Koordinatoren bei Eignungsprüfungsplänen für Prüfstellen, die für sich die Übereinstimmung mit dieser Internationalen Norm in Anspruch nehmen, Leitlinien für die Behandlung solcher Angaben benötigen.

7.7.2 Auswertung von Zeta-Werten

Sofern ein wirksames System zur Validierung der prüfstelleneigenen Schätzwerte für die Standardunsicherheiten der eigenen Ergebnisse betrieben wird, dürfen an Stelle von z -Werten ζ -Werte verwendet werden; diese sind dann in derselben Weise auszuwerten wie z -Werte (siehe 7.4), wobei dieselben kritischen Werte von 2,0 und 3,0 zu verwenden sind.

Wird kein solches System betrieben, sind ζ -Werte nur entsprechend den folgenden Erläuterungen in Verbindung mit z -Werten als Hilfe zur Verbesserung der Leistungsfähigkeit von Prüfstellen zu verwenden. Erhält eine Prüfstelle wiederholt z -Werte oberhalb des kritischen Werts von 3,0, so kann es sich für sie als hilfreich erweisen, ihr Messverfahren Schritt für Schritt zu untersuchen und für dieses Verfahren ein Unsicherheitsbudget zu entwickeln. Das Unsicherheitsbudget wird diejenigen Schritte des Verfahrens kenntlich machen, bei denen die größten Unsicherheiten auftreten, so dass die Prüfstelle in die Lage versetzt wird zu erkennen, wo Anstrengungen unternommen werden sollten, um eine Verbesserung zu erzielen. Wenn ihre ζ -Werte ebenfalls wiederholt den kritischen Wert von 3,0 überschreiten, lässt dies darauf schließen, dass das Unsicherheitsbudget nicht alle signifikanten Unsicherheitsquellen umfasst (mit anderen Worten: dass die Prüfstelle einen wichtigen Faktor übersieht).

Wenn bei einer Prüfstelle eine große systematische Abweichungskomponente auftritt und der zugewiesene Wert außerhalb ihres Unsicherheitsintervalls $X \pm U_x$ liegt, dann wird auch ein großer ζ -Wert oder ein großer E_n -Wert vorliegen.

7.8 E_z -Wert

Der E_z -Wert kann wie folgt definiert werden: $E_{z-} = \frac{x - (X - U_x)}{U_x}$ und $E_{z+} = \frac{x - (X + U_x)}{U_x}$

Dabei ist X der zugewiesene Wert nach 7.4 oder 7.6 oder ein nach 7.5 bestimmter Referenzwert und U_x ist die erweiterte Unsicherheit von X .

Ebenfalls ist x der Wert einer Prüfstelle und U_x ist die erweiterte Unsicherheit von x . U_x kann in 7.4 und 7.6 gleich $2\hat{\sigma}$ oder $3\hat{\sigma}$ sein und gleich U_{lab} in 7.5.

Üblicherweise werden Werte von E_z mit einem kritischen Wert von 1,0 verglichen:

- Wenn sowohl E_{z-} als auch E_{z+} in den Bereich von $-1,0$ bis $1,0$ fallen, ist die Leistungsfähigkeit der Prüfstelle zufrieden stellend;
- liegt einer der Werte E_{z-} oder E_{z+} außerhalb des Bereichs von $-1,0$ bis $1,0$, ist die Leistungsfähigkeit der Prüfstelle fragwürdig;

- c) liegt sowohl E_{z-} als auch E_{z+} unterhalb von $-1,0$ oder liegen beide Werte oberhalb von $1,0$, ist die Leistungsfähigkeit der Prüfstelle nicht zufrieden stellend.

7.9 Beispiel der Auswertung von Daten bei Angabe von Unsicherheiten

7.9.1 Allgemeines

Grafische Darstellungen wie in den Bildern 6 und 7 bieten hilfreiche Zusammenfassungen der Ergebnisse und Unsicherheiten von Prüfstellen. Diejenigen Prüfstellen, bei denen die größten systematischen Abweichungskomponenten auftreten, werden auch die größten z -Werte aufweisen. (Sie werden an den Enden der grafischen Darstellung liegen.)

Tabelle 8 enthält authentische Daten, die während einer Übung ermittelt wurden, bei der 181 Prüfstellen Ergebnisse mit Unsicherheiten für Blei im Wasser meldeten. Die Daten wurden in mol/l angegeben. Die Werte wurden der besseren Handhabung wegen mit einem Faktor 10^{10} multipliziert. Das bedeutet, dass die Ergebnisse entweder nicht dieselbe Unsicherheit oder nicht dieselbe systematische Abweichungskomponente haben. Jedes Ergebnis hat seine eigene systematische Abweichungskomponente D_i , die sich aus der Summe der systematischen Abweichungskomponente des *Verfahrens* und der systematischen Abweichungskomponente der *Prüfstelle* ergibt. Darüber hinaus hat natürlich jedes Verfahren eine ihm eigene Vergleichsstandardabweichung σ_R . Die beteiligten Prüfstellen nutzten verschiedene Verfahren zur Bestimmung der Werte und berechneten die Unsicherheiten selbst. Die Unsicherheiten werden hier als erweiterte Unsicherheiten behandelt. Die Daten werden in Tabelle 8 so wiedergegeben, wie sie gemeldet wurden. Sie wurden lediglich so sortiert, dass die Ergebnisse der Größe nach in aufsteigender Reihenfolge angeordnet sind, und die Prüfstellen wurden in dieser Reihenfolge durchnummeriert. $U = 0$ kann bedeuten, dass keine Unsicherheit angegeben wurde. Die Tabelle beinhaltet mehrere Werte mit zweifelhafter Gültigkeit, die in praktischen Situationen näher untersucht werden sollten, und die Extremwerte können in der Grafik nicht dargestellt werden. Insbesondere werden negative Werte nicht in der Grafik ausgewiesen, jedoch bei den Berechnungen verwendet. Diese Daten liefern ein Beispiel für einen Fall, in dem negative Ergebnisse wie in 4.6 gefordert angegeben wurden, obwohl negative Bleikonzentrationen logischerweise nicht möglich sind.

7.9.2 Zugewiesener Wert und zugehörige Unsicherheit

Der zugewiesene Wert wird entsprechend der Beschreibung in 5.6 als robuster Mittelwert der Ergebnisse berechnet. Hierzu wird Algorithmus A im Anhang C verwendet. Damit ergeben sich der zugewiesene Wert

$$X = x^* = 605 \times 10^{-10} \text{ mol/l} \quad (24)$$

und eine robuste Standardabweichung

$$s^* = 142 \times 10^{-10} \text{ mol/l} \quad (25)$$

Nach 5.6 beträgt die Standardabweichung dieses zugewiesenen Werts

$$u_X = 1,23 \times s^* / \sqrt{181} = 13 \times 10^{-10} \text{ mol/l} \quad (26)$$

7.9.3 Standardabweichung für die Eignungsbeurteilung

Die Standardabweichung für die Eignungsbeurteilung wird nach der Beschreibung in 6.6 als robuste Standardabweichung ermittelt:

$$\hat{\sigma} = s^* = 142 \times 10^{-10} \text{ mol/l} \quad (27)$$

7.9.4 Leitlinien für die Deutung der Unsicherheit des zugewiesenen Werts

Nach den Leitlinien in 4.2 ist die Unsicherheit des zugewiesenen Werts vernachlässigbar, wenn

$$u_X \leq 0,3\hat{\sigma} \quad (28)$$

Mit $u_X = 1,25 \times s^* / \sqrt{p}$ und $\hat{\sigma} = s^*$ wie in diesem Beispiel kann errechnet werden, dass Gleichung (28) für $p > 16$ erfüllt ist. Für $p = 181$ teilnehmende Prüfstellen wird das Kriterium leicht erfüllt. Bei diesen Daten bringt eine Betrachtung der z' -Werte, wie in 7.6 beschrieben, also keinen Nutzen.

7.9.5 Auswertung von Daten für eine große Anzahl von Prüfstellen unter Verwendung von Normal-Wahrscheinlichkeits-Kurven

Bild 6 stellt die Ergebnisse der 181 Prüfstellen als Funktion der prozentualen Rangordnung (berechnet nach 7.3) dar; dabei wird eine Normal-Wahrscheinlichkeits-Skala für die prozentualen Rangordnungen verwendet. Ergebnisse unterhalb von 0 mol/l oder oberhalb von $1\,600 \times 10^{-10}$ mol/l wurden nicht in dieses Bild aufgenommen.

Die z -Werte lassen sich berechnen als $z = (x - 605)/142$. Wenn eine Prüfstelle einen z -Wert oberhalb von 3,0 oder unterhalb von -3,0 erreicht, wird der z -Wert im Bild neben dem entsprechenden Wert dargestellt.

Die Summenhäufigkeitsfunktion der Normalverteilung mit dem Mittelwert von 605×10^{-10} mol/l und der Standardabweichung von 142×10^{-10} mol/l wird im Bild als gestrichelte Gerade ausgewiesen.

Die für den robusten Algorithmus verwendeten "Grenzwerte" (Abschneidewerte) sind

$$x^* - 1,5s^* = 605 - 1,5 \times 142 = 392 \times 10^{-10} \text{ mol/l} \quad (29)$$

und

$$x^* + 1,5s^* = 605 + 1,5 \times 142 = 818 \times 10^{-10} \text{ mol/l} \quad (30)$$

Bild 6 lässt erkennen, dass die Punkte außerhalb dieses Bereichs zunehmend von der gestrichelten Linie abweichen. Dies lässt darauf schließen, dass nicht alle Ergebnisse derselben Normalverteilung entstammen. Punkte, die weiter entfernt von der gestrichelten Linie liegen, entstammen einer Grundgesamtheit mit einer größeren Varianz als solche, die nahe bei der gestrichelten Linie liegen.

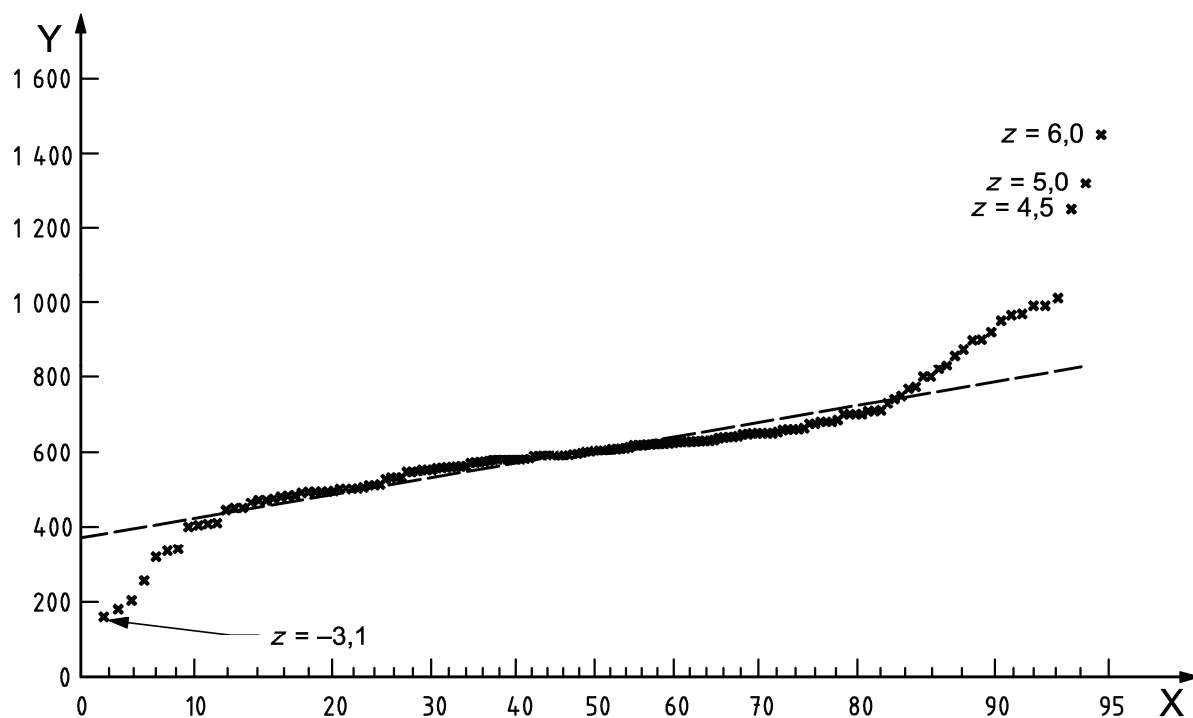
Das Bild zeigt ebenso, dass Ergebnisse, deren z -Werte oberhalb von 3,0 oder unterhalb von -3,0 liegen, deutlich von der gestrichelten Linie abweichen. Dies stützt die Schlussfolgerung, dass diese z -Werte als „Eingriffssignale“ zu werten sind. (Bei einer großen Anzahl von Ergebnissen besteht die Möglichkeit, dass z -Werte oberhalb von 3,0 oder unterhalb von -3,0 auftreten, dass jedoch bei einer Auftragung wie im Bild 6 alle Punkte nahe bei der gestrichelten Linie liegen. In solch einem Fall würde das Bild nicht die Schlussfolgerung stützen, dass sich aus den z -Werten Eingriffssignale ergeben.)

Bild 7 stellt die Ergebnisse ausschließlich für diejenigen Prüfstellen dar, deren z -Werte innerhalb des Bereichs $\pm 3,0$ liegen. Zusätzlich werden die von ihnen angegebenen Unsicherheiten als vertikale Linien dargestellt. (Ergebnisse für zwei Prüfstellen, die sehr große Unsicherheiten angaben, die Prüfstellen 24 und 112, wurden nicht in das Bild übernommen.) Die durchgezogene horizontale Linie in diesem Bild stellt den zugewiesenen Wert dar und die beiden gestrichelten Linien links und rechts davon stellen die erweiterte Unsicherheit des zugewiesenen Werts dar.

Wenn die Prüfstellen gültige Berechnungen der erweiterten Unsicherheiten ihrer Ergebnisse anstellen würden, müssten nahezu alle der vertikalen Linien im Bild 7 durch den Bereich verlaufen, der durch den zugewiesenen Wert abzüglich und zuzüglich seiner erweiterten Unsicherheit definiert wird. Das Bild zeigt jedoch, dass es beiderseits des zugewiesenen Werts viele Prüfstellen gibt, deren vertikale Linien nicht in diesen Bereich fallen. Dies lässt darauf schließen, dass in vielen Prüfstellen keine gültigen Unsicherheitsberechnungen durchgeführt wurden; es ist höchst wahrscheinlich, dass nicht alle wichtigen Unsicherheitsquellen in die Berechnungen eingeflossen sind.

Tabelle 8 — Bestimmung des Bleigehalts von Wasser ($\text{mol/l} \times 10^{10}$) durch 181 Prüfstellen mit Werten für die erweiterten Unsicherheiten (U) der Ergebnisse, wie von den Prüfstellen angegeben

Prüf- stelle	Ergebnis	U	Prüf- stelle	Ergeb- nis	U	Prüf- stelle	Ergeb- nis	U	Prüf- stelle	Ergeb- nis	U
1	−960 000	0	51	545	43	101	618	224	151	740	20
2	−12 100	0	52	545	123	102	618	170	152	748	3
3	−4 800	0	53	550	8	103	620	25	153	767	113
4	−3 860	0	54	550	55	104	620	40	154	772	213
5	−1 500	0	55	550	5	105	621	6	155	800	60
6	−1 010	0	56	555	79	106	622	9	156	800	150
7	−1 000	0	57	556	30	107	622	6	157	821	203
8	−1 000	0	58	557	28	108	623	18	158	830	10
9	−965	0	59	557	28	109	625	15	159	857	27
10	−483	0	60	559	26	110	626	5	160	874	200
11	160	20	61	560	7	111	627	0	161	898	59
12	180	20	62	560	60	112	627	1 010	162	900	100
13	203	0	63	569	116	113	627	15	163	920	140
14	256	13	64	570	86	114	628	3	164	950	110
15	319	0	65	571	16	115	629	26	165	965	0
16	335	18	66	572	40	116	630	40	166	968	0
17	340	180	67	574	35	117	630	580	167	990	0
18	400	20	68	578	0	118	632	50	168	990	80
19	404	36	69	579	52	119	637	96	169	1 010	0
20	407	0	70	579	35	120	639	83	170	1 250	140
21	410	97	71	579	8	121	640	130	171	1 320	410
22	444	58	72	579	10	122	640	77	172	1 450	460
23	450	20	73	579	17	123	642	20	173	1 640	241
24	450	3 400 000	74	579	87	124	647	63	174	1 900	46
25	463	19	75	580	150	125	647	0	175	2 413	20
26	470	10	76	582	122	126	650	160	176	2 460	0
27	470	30	77	589	57	127	650	30	177	2 900	900
28	474	0	78	589	10	128	650	80	178	10 000	0
29	480	100	79	590	0	129	650	48	179	386 000	31 000
30	482	122	80	590	0	130	650	30	180	670 000	60 000
31	483	241	81	590	45	131	653	5	181	630×10 ⁶	60×10 ⁶
32	490	60	82	590	60	132	658	27			
33	492	25	83	590	0	133	660	20			
34	492	1	84	591	112	134	660	120			
35	493	24	85	591	9	135	660	34			
36	493	5	86	594	4	136	663	32			
37	495	0	87	594	119	137	675	280			
38	500	70	88	597	9	138	675	0			
39	500	10	89	600	20	139	680	8			
40	500	10	90	600	300	140	680	50			
41	501	75	91	603	60	141	680	70			
42	504	0	92	603	24	142	685	0			
43	510	130	93	603	13	143	700	0			
44	510	110	94	604	18	144	700	100			
45	512	6	95	608	30	145	700	110			
46	526	26	96	608	8	146	700	300			
47	530	9	97	609	9	147	708	44			
48	530	40	98	610	61	148	709	48			
49	530	60	99	613	22	149	710	100			
50	545	30	100	618	7	150	729	41			

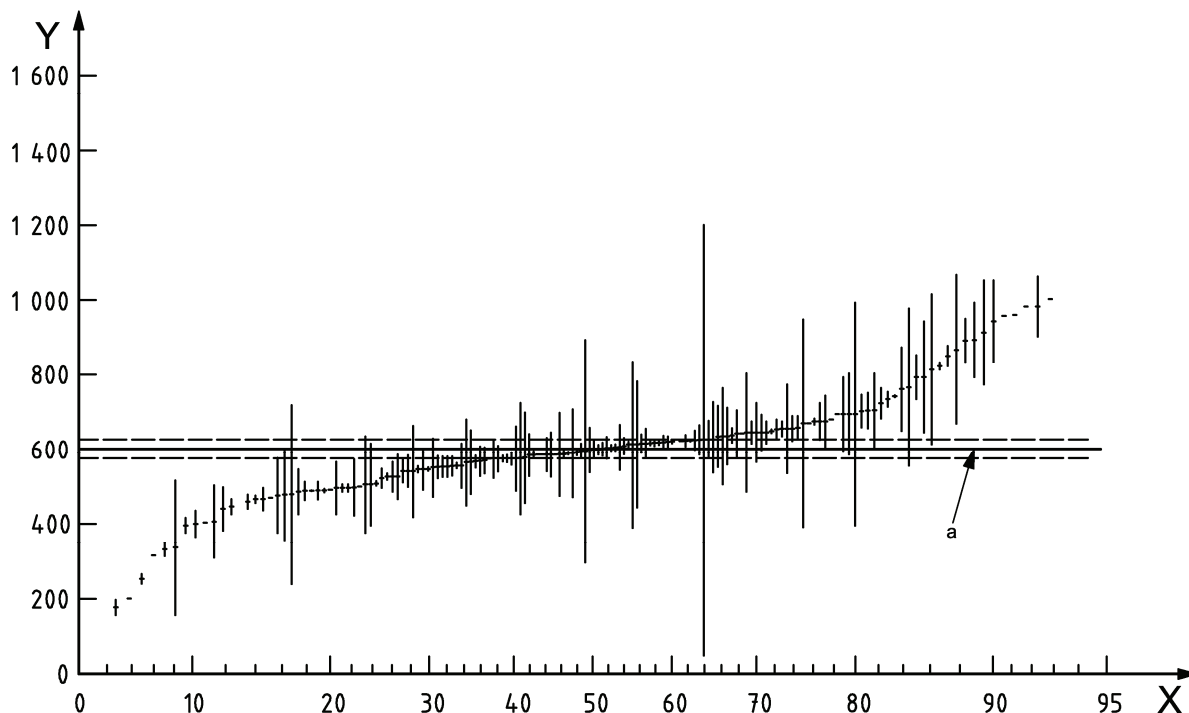
**Legende**

X prozentuale Rangordnung der Prüfstelle, in Prozent (%)

Y Bleigehalt, in $\text{mol/l} \times 10^{10}$

ANMERKUNG Die Ergebnisse von 19 Prüfstellen wurden nicht aufgenommen.

Bild 6 — Normal-Wahrscheinlichkeits-Kurve der Ergebnisse von Bestimmungen des Bleigehalts von Wasser durch 162 Prüfstellen



Legende

- X prozentuale Rangordnung der Prüfstelle, in Prozent (%)
 Y Bleigehalt, in $\text{mol/l} \times 10^{10}$

^a zugewiesener Wert abzüglich und zuzüglich der zugehörigen erweiterten Unsicherheit.

ANMERKUNG Die Ergebnisse von 25 Prüfstellen wurden nicht aufgenommen.

Bild 7 — Normal-Wahrscheinlichkeits-Kurve der erweiterten Unsicherheiten von Bestimmungen des Bleigehalts von Wasser durch 156 Prüfstellen

7.10 Kombiniertes Leistungswert

Es ist üblich, dass innerhalb eines Durchlaufs eines Eignungsprüfungsplans Ergebnisse an mehr als einem Prüfgegenstand oder für mehrere Messgrößen ermittelt werden. In dieser Situation sind die Ergebnisse für jeden Prüfgegenstand und jede Messgröße entsprechend 7.2 bis 7.9 auszuwerten; das heißt, dass die Ergebnisse für jeden Prüfgegenstand und jede Messgröße separat auszuwerten sind. Es gibt Anwendungen, bei denen zwei oder mehr Materialien mit speziell zugeschnittenen Werten in den Plan aufgenommen werden, um andere Aspekte der Leistungsfähigkeit zu messen, zum Beispiel Wiederholpräzision oder Linearität. In solchen Fällen muss der Koordinator den Teilnehmern vollständige Beschreibungen der statistischen Planung und der statistischen Verfahren übermitteln, die zur Anwendung kommen. Wurden in den Plan zwei ähnliche Materialien aufgenommen, die als Youden-Paar behandelt werden sollen, so sind die Sonderverfahren nach 8.5 anzuwenden.

Darüber hinaus wird empfohlen, die grafischen Verfahren nach Abschnitt 8 ebenfalls zu verwenden, wenn Ergebnisse an mehr als einem Prüfgegenstand oder für mehr als eine Messgröße ermittelt werden. Sie kombinieren Werte in einer Weise, die hohe Einzelwerte nicht verbirgt, und sie können zusätzliche Informationen über die Leistungsfähigkeit der Prüfstellen aufzeigen, zum Beispiel Korrelationen zwischen Ergebnissen für verschiedene Messgrößen, die aus den Tabellen der einzelnen Werte nicht erkennbar sind.

Bei Plänen, die eine große Anzahl von Messgrößen beinhalten, kann eine Zählung der Eingriffs- und Warnsignale verwendet werden, um diejenigen Prüfstellen zu erkennen, die ein oder mehrere solche Signale

erhalten. Ihnen kann ein Bericht mit detaillierten Ergebnissen unter Verwendung der Verfahren nach 7.2 bis 7.9 zur Verfügung gestellt werden. Prüfstellen, die kein solches Signal erhalten, kann ein Kurzbericht zur Verfügung gestellt werden.

ANMERKUNG Die Verwendung von zusammengesetzten Werten in Fällen mit mehr als einem Prüfgegenstand für dieselbe Messgröße (mittlerer oder aufsummierter z -Wert oder mittlerer oder aufsummierter Betrag der Differenz beziehungsweise mittlere oder aufsummierte Quadrate der Differenzen) oder die Verwendung von zusammengesetzten Werten in Fällen mit Ergebnissen für mehr als eine Messgröße (mittlerer Betrag des z -Werts oder mittlerer Betrag der Differenz, bezogen auf die Auswertungsgrenzen) ist nicht empfehlenswert. Der mittlere (oder aufsummierte) z -Wert hat den deutlichen Nachteil, dass ein hoher Wert bei einem Prüfgegenstand verborgen bleiben kann, wenn andere Werte niedrig liegen oder wenn ein anderer Wert ebenfalls hoch liegt, aber das entgegengesetzte Vorzeichen hat. Der mittlere (oder aufsummierte) Betrag der Differenz und die aufsummierten Quadrate der Differenzen leiden ebenfalls an dem deutlichen Nachteil, dass ein hoher Wert bei einem Prüfgegenstand verborgen bleiben kann, wenn andere Werte niedrig liegen. Der mittlere Betrag des z -Werts und der mittlere Betrag der Differenz, bezogen auf die Auswertungsgrenzen, haben ebenfalls diesen Nachteil.

8 Grafische Verfahren zur Kombination von Leistungswerten für mehrere Messgrößen in einem Durchlauf einer Eignungsprüfung (siehe ISO/IEC Guide 43-1:1997, A.2.2.1)

8.1 Anwendung

Der Koordinator muss erwägen, die sich bei jedem Durchlauf eines Eignungsprüfungsplans ergebenden Leistungswerte zur Erstellung von grafischen Darstellungen nach 8.2 und 8.3 zu verwenden. Die Verwendung von z -Werten in diesen Darstellungen hat den Vorteil, dass standardisierte Achsen verwendet werden können, wodurch sich Darstellung und Deutung vereinfachen. Die grafischen Darstellungen müssen den Teilnehmern zur Verfügung gestellt werden, so dass jeder Teilnehmer erkennen kann, wie sich seine Ergebnisse in Relation zu denjenigen der anderen Teilnehmer anordnen. Es sind Buchstaben- oder Zahlenschlüssel für die Teilnehmer zu verwenden, so dass jeder Teilnehmer seine eigenen Ergebnisse erkennen kann, jedoch nicht in der Lage ist, zuzuordnen, welcher Teilnehmer welches andere Ergebnis erzielte. Die grafischen Darstellungen sind auch dem Koordinator zur Verfügung zu stellen, damit dieser die Gesamt-Wirksamkeit des Plans beurteilen und erkennen kann, ob das zur Eignungsbeurteilung verwendete Kriterium überprüft werden muss.

8.2 Histogramme von Leistungswerten

8.2.1 Allgemeines

Zur Erstellung eines Histogramms von z -Werten sind die z -Werte für die Messung eines Merkmals aus einem Durchlauf eines Eignungsprüfungsplans wie im Bild 8 dargestellt aufzutragen. In diesem Histogramm ist eine Schrittweite von etwa 0,3 bis 0,5 zu verwenden, damit sich ein guter visueller Eindruck ergibt. Bei $\pm 2,0$ und $\pm 3,0$ sind Linien einzuzeichnen, um das Kriterium zur Eignungsbeurteilung darzustellen. Das Histogramm sollte einen Bereich von ungefähr $\pm 6,0$ überdecken. Der Auswerter darf die Verwendung einer abweichenden Spanne in Betracht ziehen, wenn die Ergebnisse sich nicht über diesen Bereich verteilen oder wenn die zentral liegenden Punkte eng gehäuft liegen.

Für den Fall, dass Histogramme der systematischen Abweichungskomponenten von Prüfstellen oder der prozentualen Differenzen vorgezogen werden, werden nachfolgend die gleichwertigen Histogramm-Schrittweiten und Warn- und Eingriffsgrenzen angegeben. In solchen Fällen kann es einfacher sein, die Eingriffsgrenzen direkt aus den vorgegebenen oder wahrgenommenen Anforderungen abzuleiten, als sie aus der Standardabweichung für die Eignungsbeurteilung zu berechnen.