DIN ISO 13528:2025-07 (D/E)

Statistische Verfahren für Eignungsprüfungen durch Vergleiche zwischen Laboratorien (ISO 13528:2022); Text Deutsch und Englisch

Statistical methods for use in proficiency testing by interlaboratory comparison (ISO 13528:2022); Text in German and English

Inhalt

Seite

Nationales Vorwort 7

Nationaler Anhang NA (informativ) Literaturhinweise 9

Vorwort 10

0 Einleitung 11
0.1 Zweck von Eignungsprüfungen 11
0.2 Begründung für die Bewertung durch Scores in Eignungsprüfungsprogrammen 11
0.3 ISO 13528 und ISO/IEC 17043 11
0.4 Statistischer Sachverstand 12
0.5 Computersoftware 12

1 Anwendungsbereich 14

2 Normative Verweisungen 14

3 Begriffe 14

4 Allgemeine Grundsätze 18
4.1 Allgemeine Anforderungen an statistische Verfahren 18
4.2 Grundmodell 18
4.3 Allgemeine Ansätze für die Leistungsbewertung 19

5 Leitlinien für das statistische Modell von Eignungsprüfungsprogrammen 19
5.1 Einleitung zum statistischen Modell von Eignungsprüfungsprogrammen 19
5.2 Grundlage eines statistischen Modells 20
5.3 Überlegungen zur statistischen Verteilung der Ergebnisse 21
5.4 Überlegungen für kleine Teilnehmeranzahlen 22
5.5 Leitlinien zur Auswahl des Berichtsformats 22
5.5.1 Allgemeine Anforderungen an das Berichtsformat 22
5.5.2 Berichte über Wiederholmessungen 23
5.5.3 Berichte von „kleiner als“ oder „größer als“ ein Grenzwert (zensierte Daten) 23
5.5.4 Anzahl der signifikanten Stellen 24

6 Leitlinien für die anfängliche Überprüfung von Eignungsprüfungsgegenständen und von Ergebnissen 24
6.1 Homogenität und Stabilität von Eignungsprüfungsgegenständen 24
6.2 Betrachtungen zu verschiedenen Messverfahren 26
6.3 Entfernung grober Fehler 26
6.4 Visuelle Überprüfung von Daten 27
6.5 Robuste statistische Verfahren 27
6.6 Ausreißertechniken für Einzelergebnisse 28

7 Bestimmung des zugewiesenen Werts und seiner Standardunsicherheit 29
7.1 Auswahl des Verfahrens zur Bestimmung des zugewiesenen Werts 29
7.2 Bestimmung der Unsicherheit des zugewiesenen Werts 29
7.3 Bekannte Zusammensetzung aus der Probenherstellung 30
7.4 Zertifiziertes Referenzmaterial 31

7.5 Ergebnisse aus einem einzelnen Laboratorium ... 32
7.6 Konsenswert von Expertenlaboratorien ... 33
7.7 Konsenswert aus Teilnehmerergebnissen ... 34
7.8 Vergleich des zugewiesenen Werts mit einem unabhängigen Referenzwert ... 35

8 Bestimmung von Kriterien für die Leistungsbewertung ... 36
8.1 Ansätze für die Bestimmung von Bewertungskriterien ... 36
8.2 Bestimmung anhand von Expertenangaben ... 37
8.3 Bestimmung anhand von Erfahrungen aus früheren Runden eines Eignungsprüfungsprogramms ... 37
8.4 Bestimmung anhand eines allgemeinen Modells ... 38
8.5 Bestimmung anhand der Wiederhol– und Vergleichstandardabweichungen aus einem früheren Vergleich zwischen Laboratorien zur Präzision eines Messverfahrens ... 38
8.6 Werte aus Daten, die in derselben Runde eines Eignungsprüfungsprogramms erzielt wurden ... 39
8.7 Überwachung der Übereinstimmung zwischen Laboratorien ... 40

9 Berechnung von Leistungskenngrößen ... 41
9.1 Allgemeine Betrachtungen zur Leistungsbestimmung ... 41
9.2 Begrenzung der Unsicherheit des zugewiesenen Werts ... 41
9.3 Schätzwerte der Abweichung (Messabweichung) ... 42
9.4 z-Scores ... 43
9.5 z'-Scores ... 44
9.6 Zeta-Scores (ζ) ... 46
9.7 E_n-Scores ... 47
9.8 Bewertung der Teilnehmerunsicherheiten bei der Prüfung ... 48
9.9 Kombinierte Leistungs-Scores ... 49

10 Graphische Verfahren zur Beschreibung von Leistungs-Scores ... 50
10.1 Anwendung graphischer Verfahren ... 50
10.2 Histogramme von Ergebnissen oder Leistungs-Scores ... 50
10.3 Kerndichtediagramme ... 51
10.4 Säulendiagramme von standardisierten Leistungs-Scores ... 52
10.5 Youden-Diagramm ... 53
10.6 Diagramme von Wiederholstandardabweichungen ... 54
10.7 Geteilte Proben ... 55
10.8 Graphische Verfahren zur Kombination von Leistungs-Scores über mehrere Runden eines Eignungsprüfungsprogramms ... 55

11 Planung und Analyse von qualitativen Eignungsprüfungsprogrammen (einschließlich nominaler und ordinaler Eigenschaften) ... 56
11.1 Arten von qualitativen Daten ... 56
11.2 Statistisches Modell ... 57
11.3 Zugewiesene Werte für qualitative Eignungsprüfungsprogramme ... 58
11.4 Leistungsbewertung und Bewertung durch Scores bei qualitativen Eignungsprüfungsprogrammen ... 60

Anhang A (normativ) Symbole und Abkürzungen ... 62

Anhang B (informativ) Homogenität und Stabilität von Eignungsprüfungsgegenständen ... 64
B.1 Allgemeines Verfahren zur Homogenitätsprüfung ... 64
B.2 Beurteilungskriterien für eine Homogenitätsprüfung ... 64
B.3 Formeln für die Homogenitätsprüfung ... 67
B.4 Verfahren zur Stabilitätsprüfung ... 69
B.4.1 Allgemeine Betrachtungen zur Stabilitätsprüfung ... 69
B.4.2 Verfahren zur Stabilitätsprüfung im Verlauf einer Eignungsprüfungsrunde ... 70
B.5 Beurteilungskriterium für eine Stabilitätsprüfung ... 71
B.6 Stabilität unter Transportbedingungen ... 72

Anhang C (informativ) Robuste Analyse ... 73
C.1 Allgemeines ... 73

C.2 Einfache ausreißerresistente Schätzfunktionen für den Mittelwert und die Standardabweichung der Grundgesamtheit ... 73
C.2.1 Median ... 73
C.2.2 Skalierter Median der absoluten Abweichungen MAD_e ... 74
C.2.3 Normierter Interquartilsabstand nIQR ... 74
C.3 Robuste Analyse: Algorithmus A ... 75
C.3.1 Algorithmus A mit iterierter Skala ... 75
C.3.2 Varianten von Algorithmus A ... 76
C.4 Robuste Analyse: Algorithmus S ... 76
C.5 Rechenintensive robuste Schätzfunktionen: Q-Verfahren und Hampel-Schätzer ... 78
C.5.1 Begründung für rechenintensive Schätzfunktionen ... 78
C.5.2 Bestimmung einer robusten Standardabweichung unter Anwendung von Q- und Q_n-Verfahren ... 79
C.5.3 Bestimmung eines robusten Mittelwerts mit dem Hampel-Schätzer ... 82
C.5.4 Q/Hampel-Verfahren ... 84
C.6 Andere robuste Verfahren ... 85

Anhang D (informativ) Zusätzlicher Leitfaden zu statistischen Verfahren ... 86
D.1 Verfahren für kleine Teilnehmeranzahlen ... 86
D.1.1 Allgemeine Überlegungen ... 86
D.1.2 Verfahren zur Ausreißererkennung ... 86
D.1.3 Verfahren für lagebezogene Schätzwerte ... 87
D.1.4 Verfahren für Schätzwerte der Streuung ... 87
D.2 Effizienz und Bruchpunkte für robuste Verfahren ... 88
D.2.1 Verschiedene statistische Schätzer (z. B. robuste Verfahren) können anhand von drei Schlüsselmerkmalen verglichen werden: ... 88
D.2.2 Bruchpunkt ... 88
D.2.3 Relative Effizienz ... 89
D.3 Verwendung von Eignungsprüfdaten zur Bewertung der Vergleich- und Wiederholpräzision eines Messverfahrens ... 90

Anhang E (informativ) Erläuternde Beispiele ... 92
E.1 Wirkung von zensierten Werten (siehe 5.5.3.3) ... 92
E.2 Homogenitäts- und Stabilitätsprüfung — Arsen (As) in Schokolade (siehe 6.1) ... 93
E.3 Umfassendes Beispiel für Atrazin in Trinkwasser ... 95
E.4 Umfassendes Beispiel für Quecksilber in Tierfutter ... 99
E.5 Referenzwert von einem einzelnen Laboratorium: Los-Angeles-Wert von Zuschlagstoffen (siehe 7.5) ... 102
E.6 Beispiel für ein Bootstrap-Verfahren für coliforme Bakterien in Lebensmittelproben (siehe 7.7.6) ... 104
E.7 Vergleich von Referenzwert und Konsens-Mittelwert (siehe 7.8) ... 105
E.8 Bestimmung von Bewertungskriterien durch Erfahrung mit früheren Runden: Toxaphen in Trinkwasser (siehe 8.3) ... 105
E.9 Anhand eines allgemeinen Modells: Horwitz-Gleichung (siehe 8.4) ... 108
E.10 Bestimmung der Leistung aus einem Präzisionsversuch: Bestimmung des Zementgehalts von Festbeton (siehe 8.5) ... 108
E.11 Säulendiagramme von standardisierten systematischen Abweichungskomponenten: Antikörperkonzentrationen (siehe 10.4) ... 108
E.12 Youden-Diagramm — Antikörper-Konzentrationen (siehe 10.5) ... 109
E.13 Diagramm der Wiederholstandardabweichungen: Antikörper-Konzentrationen (siehe 10.6) ... 112
E.14 Graphische Verfahren zur Rückverfolgung der Leistung über die Zeit (siehe 10.8) ... 114
E.15 Qualitative Datenanalyse; Beispiel für eine Ordinalgröße: Hautreaktion auf ein Kosmetikum (siehe Abschnitt 11) ... 116

Anhang F (informativ) Beispiel für einen Computercode zur Darstellung und Resampling-Analyse („Bootstrapping“) von Ergebnissen aus Eignungsprüfungen ... 118

Literaturhinweise ... 119

Bilder

Bild E.1 — Nach Rang geordnete Teilnehmerergebnisse für Atrazin (Daten nach Tabelle E.4) 97

Bild E.2 — Histogramm der Teilnehmerergebnisse 97

Bild E.3 — Kerndichtediagramm für Teilnehmerergebnisse 98

Bild E.4 — Zusammenfassung von robusten Kenngrößen nach Tabelle E.5 99

Bild E.5 — Teilnehmerergebnisse und Unsicherheiten für Ergebnisse in IMEP 111 (Daten nach Tabelle E.6) 101

Bild E.6 — Kerndichtediagramm für Teilnehmerergebnisse 102

Bild E.7 — Kerndichtediagramm für Teilnehmerergebnisse 105

Bild E.8 — Relative Standardabweichung von Teilnehmerergebnissen (%) verglichen mit dem zugewiesenen Wert (µg/l) 107

Bild E.9 — Teilnehmer-Standardabweichung (µg/l) verglichen mit dem zugewiesenen Wert (µg/l) 107

Bild E.10 — Säulendiagramm der *z*-Scores (4,0 bis −4,0) für eine Runde eines Eignungsprüfungsprogramms, während der die Teilnehmer Konzentrationen dreier allergenspezifischer IgE-Antikörper bestimmten 109

Bild E.11 — Youden-Diagramm der *z*-Scores nach Tabelle E.10 110

Bild E.12 — Diagramm von Standardabweichungen über Mittelwerten für 25 Teilnehmer (Daten nach Tabelle E.10) 113

Bild E.13 — Leistungs-Scores für jede Runde des Eignungsprüfungsprogramms (Daten nach Tabelle E.12) 115

Bild E.14 — Leistungs-Scores für verschiedene Niveaus der Messgröße 115

Bild E.15 — Säulendiagramm von prozentualen Reaktionen auf zwei Eignungsprüfungsgegenstände für die Hautreizung — # Modus, @ Median 117

Tabellen

Tabelle B.1 — Faktoren F_1 und F_2 zur Verwendung in der Prüfung auf ausreichende Homogenität 66

Tabelle C.1 — Für die robuste Auswertung notwendige Faktoren: Algorithmus S 77

Tabelle C.2 — Korrekturfaktor b_p für $2 \leq p \leq 12$ 80

Tabelle D.1 — Bruchpunkte für Schätzwerte des Mittelwerts und der Standardabweichung (Anteil von Ausreißern, der zum Versagen des Schätzers führen kann) 89

Tabelle D.2 — Relative Effizienz von robusten Schätzern für Mittelwert und Standardabweichung der Grundgesamtheit für normalverteilte Datensätze mit n = 50 oder 500 Teilnehmer 90

Tabelle E.1 — Stichprobendatensatz mit zensierten (<)-Ergebnissen und drei Wahlmöglichkeiten für entsprechende Ergebnisse 92

Tabelle E.2 — Homogenitätsdaten für Eignungsprüfungsgegenstände für Arsen in Schokolade 93

Tabelle E.3 — Stabilitätsdaten für Eignungsprüfungsgegenstände für Arsen in Schokolade 94

Tabelle E.4 — Berechnung des robusten Mittelwerts und der robusten Standardabweichung für Atrazin in Trinkwasser 95

Tabelle E.5 — Zusammenfassende Kenngrößen für das Atrazin-Beispiel 98

Tabelle E.6 — Eignungsprüfungsergebnisse von 24 Teilnehmern an der IMEP 111-Untersuchung .. 100

Tabelle E.7 — Leistungskenngrößen nach verschiedenen Verfahren 102

Tabelle E.8 — Berechnung der mittleren Differenz zwischen einem ZRM und einem Eignungsprüfungsgegenstand und der Standardunsicherheit dieser Differenz 103

Tabelle E.9 — Eignungsprüfungsrunden für Toxaphen in Trinkwasser und $p \geq 20$ Ergebnisse 106

Tabelle E.10 — Daten und Berechnungen zu Antikörperkonzentrationen für zwei ähnliche Allergene 111

Tabelle E.11 — Konzentrationen bestimmter Antikörper in Serum-Eignungsprüfungsgegenständen (vier Wiederholbestimmungen an einem Eignungsprüfungsgegenstand je Teilnehmer) 112

Tabelle E.12 — P_A-Scores für 5 Runden eines Eignungsprüfungsprogramms, jeweils mit 3 Eignungsprüfungsgegenständen für Kalium in Serum 114

Tabelle E.13 — Ergebnisse für zwei Eignungsprüfungsgegenstände, Hautreizung 116