



TC/52/16

ORIGINAL: englisch

DATUM: 22. Januar 2016

INTERNATIONALER VERBAND ZUM SCHUTZ VON PFLANZENZÜCHTUNGEN

Genf

TECHNISCHER AUSSCHUSS**Zweiundfünfzigste Tagung
Genf, 14. bis 16. März 2016**

ÜBERARBEITUNG VON DOKUMENT TGP/8: TEIL I: DUS-PRÜFUNGSANLAGE UND DATENANALYSE,
NEUER ABSCHNITT: MINIMIERUNG DER VARIATION INFOLGE VERSCHIEDENER ERFASSER IM
SELBEN ANBAUVERSUCH

vom Verbandsbüro erstelltes Dokument

Haftungsausschluß: dieses Dokument gibt nicht die Grundsätze oder eine Anleitung der UPOV wieder

ZUSAMMENFASSUNG

1. Zweck dieses Dokuments ist es, einen Entwurf für einen neuen Abschnitt für Dokument TGP/8 Teil I: DUS-Prüfungsanlage und Datenanalyse über „Minimierung der Variation infolge verschiedener Erfasser“ darzulegen.
2. Der TC wird ersucht:
 - a) den Entwurf einer Anleitung zu „Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch“, wie in der Anlage dieses Dokuments dargelegt, in Verbindung mit den Anmerkungen der TWP auf ihren Tagungen im Jahr 2015 und des TC-EDC auf seiner Tagung im Jahr 2016 als Grundlage für eine künftige Überarbeitung von Dokument TGP/8: „Prüfungsanlage und Verfahren für die Prüfung der Unterscheidbarkeit, der Homogenität und der Beständigkeit“, Teil I, DUS-Prüfungsanlage und Datenanalyse, zu prüfen;
 - b) zu prüfen, ob in dem Dokument ein anderer Abschnitt entwickelt werden soll, um Anleitung zur Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch für PQ-Merkmale, einschließlich der Anwendung nicht parametrischer Verfahren zu erteilen;
 - c) zu prüfen, ob die TWP auf ihren Tagungen im Jahr 2016 ersucht werden sollten, Informationen über Variationen zwischen Erfassern für PQ-Merkmale, wie etwa Häufigkeit von Abweichungen, zu erteilen;
 - d) zur Kenntnis zu nehmen, daß die TWC vereinbart hatte, die Sachverständigen aus Argentinien und Brasilien aufzufordern, auf ihrer vierunddreißigsten Tagung ein Referat über ihre Erfahrungen mit Schulungen über die Minimierung von Variationen zwischen Erfassern bei PQ-Merkmalen zu halten;
 - e) zu prüfen, ob Sachverständige eingeladen werden sollen, um den TWP auf ihren Tagungen im Jahr 2016 ihre Erfahrungen im Hinblick auf die Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch für PQ-Merkmale darzulegen.

3. Der Aufbau dieses Dokuments ist wie folgt:

ZUSAMMENFASSUNG	1
HINTERGRUND	2
ENTWICKLUNGEN IM JAHR 2015	2
TECHNISCHER AUSSCHUß	2
TECHNISCHE ARBEITSGRUPPEN	3
ENTWICKLUNGEN IM JAHR 2016	3
ERWEITERTER REDAKTIONSAUSSCHUß	3
VORSCHLAG	3

ANLAGE: ENTWURF EINER ANLEITUNG FÜR DIE KÜNFTIGE ÜBERARBEITUNG VON DOKUMENT TGP/8 ZUR MINIMIERUNG DER VARIATION INFOLGE VERSCHIEDENER ERFASSER IM SELBEN ANBAUVERSUCH

4. In diesem Dokument werden folgende Abkürzungen verwendet:

CAJ:	Verwaltungs- und Rechtsausschuß
TC:	Technischer Ausschuß
TC-EDC:	Erweiterter Redaktionsausschuß
TWA:	Technische Arbeitsgruppe für landwirtschaftliche Arten
TWV:	Technische Arbeitsgruppe für Gemüsearten
TWC:	Technische Arbeitsgruppe für Automatisierung und Computerprogramme
TWF:	Technische Arbeitsgruppe für Obstarten
TWO:	Technische Arbeitsgruppe für Zierpflanzen und forstliche Baumarten
TWP:	Technische Arbeitsgruppen

HINTERGRUND

5. In Dokument TGP/8/1 Draft 7 TEIL I, Absatz 2.9.1: „Kontrolle der Variation infolge verschiedener Erfasser“, das von den Technischen Arbeitsgruppen auf ihren Tagungen im Jahr 2007 geprüft wurde, heißt es:

„[Ist dieser Abschnitt erforderlich, so sind die TWP aufgefordert, Anleitung zur Kontrolle der Variation infolge verschiedener Erfasser zu erteilen, wenn keine statistische Analyse zur Bestimmung der Unterscheidbarkeit eingesetzt wird, und sie in bezug auf Absatz 2.7.2.9. zu prüfen]“

6. Entwicklungen vor dem Jahr 2015 sind in Dokument TC/51/16 „Überarbeitung von TGP/8: Teil I: DUS-Prüfungsanlage und Datenanalyse, Neuer Abschnitt: Minimierung der Variation infolge verschiedener Erfasser“ dargelegt.

ENTWICKLUNGEN IM JAHR 2015

Technischer Ausschuß

7. Der TC prüfte auf seiner einundfünfzigsten Tagung vom 23. bis 25. März 2015 in Genf Dokument TC/51/16 „Überarbeitung von Dokument TGP/8: Teil I: DUS-Prüfungsanlage und Datenanalyse, Neuer Abschnitt: „Minimierung der Variation infolge verschiedener Erfasser“ und den Entwurf einer Anleitung zur Aufnahme in Dokument TGP/8 über die Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch, wie in der Anlage von Dokument TC/51/16 wiedergegeben, in Verbindung mit den Bemerkungen der TWP auf ihren Tagungen im Jahr 2014 (vergleiche Dokument TC/51/39, Absatz 132).

8. Der TC vereinbarte, den Sachverständigen aus Australien zu ersuchen, mit der Entwicklung des den TWP auf ihren Tagungen im Jahr 2015 vorzustellenden Dokuments fortzufahren und den in der Anlage von Dokument TC/51/16 „Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch“ verwendeten Titel zu ändern (vergleiche Dokument TC/51/39, Absatz 133).

9. Als Antwort auf das Ersuchen des TC stellte der Verfaßer aus Australien (Herr Nik Hulse) einen überarbeiteten Entwurf einer Anleitung für die künftige Überarbeitung von Dokument TGP/8 über die

Minimierung der Variation infolge verschiedener Erfasser zur Prüfung durch die TWP auf ihren Tagungen im Jahr 2015, wie in der Anlage dieses Dokuments dargelegt, bereit.

Technische Arbeitsgruppen

10. Die TWV, TWC, TWA, TWF und die TWO prüften jeweils die Dokumente TWV/49/15, TWC/33/15, TWA/44/15, TWF/46/15 und TWO/48/15.

11. Die TWV und die TWC vereinbarten, daß der Entwurf einer Anleitung in der Anlage der Dokumente TWV/49/15 und TWC/33/15 zur Aufnahme in eine künftige Überarbeitung von Dokument TGP/8 über die Minimierung der Variation infolge verschiedener Erfasser weiter entwickelt werden soll.

12. Die TWV schlug vor, daß eine Anleitung zu PQ-Merkmalen weiter geprüft werden sollte und schlug die Entwicklung eines anderen Abschnitts im Dokument zur Erklärung nicht parametrischer Verfahren vor. Die TWV forderte die anderen TWP auch dazu auf, zu prüfen, ob künftig weiter an PQ-Merkmalen im Entwurf einer Anleitung gearbeitet werden sollte.

13. Die TWC vereinbarte, daß weitere Informationen über die Variation zwischen Verfassern für PQ-Merkmale erteilt werden soll, bevor eine Anleitung zur Verwendung nichtparametrischer Verfahren, wie etwa die Häufigkeit von Abweichungen, aufgesetzt werden könnte.

14. Die TWC vereinbarte, die Sachverständigen aus Argentinien und Brasilien aufzufordern, auf ihrer vierunddreißigsten Tagung ein Referat über ihre Erfahrungen mit Schulungen über die Minimierung von Variationen zwischen Erfassern bei PQ-Merkmalen zu halten.

15. Die TWA stimmte dem Entwurf einer Anleitung in der Anlage von Dokument TWV/44/15 zur Aufnahme in eine künftige Überarbeitung von Dokument TGP/8 über die Minimierung der Variation infolge verschiedener Erfasser zu.

16. Die TWF und die TWO stimmten dem Entwurf einer Anleitung in der Anlage des Dokuments TWV/46/15 zur Aufnahme in eine künftige Überarbeitung von Dokument TGP/8 über die Minimierung der Variation infolge verschiedener Erfasser, vorbehaltlich folgender redaktioneller Änderung, zu:

„Allerdings wurde dieses Verfahren ~~unseres Wissens nach~~ nicht ~~an~~ für PQ-Merkmale ~~angewandtentwickelt~~ und ~~PQ-Merkmale~~ könnte auch Zusatzinformationen zur Kalibrierung erfordern.“

ENTWICKLUNGEN IM JAHR 2016

Erweiterter Redaktionsausschuß

17. Der TC-EDC prüfte auf seiner Tagung am 6. und 7. Januar 2016 in Genf Dokument TC-EDC/Jan16/5 „Überarbeitung von Dokument TGP/8: Teil I: DUS-Prüfungsanlage und Datenanalyse, Neuer Abschnitt: Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch“ und unterbreitete folgende Vorschläge:

Anlage, Abs. 2.2	Sollte lauten: „...Allerdings kann die Art und Weise, in der ein Merkmal erfaßt oder gemessen wird, je nach <u>Jahr</u> , Ort oder Prüfungsbehörde variieren. Von den lokalen Prüfungsbehörden erstellte Kalibrierungshandbücher <u>und Beispielssorten</u> sind sehr nützlich für die Umsetzung der UPOV-Prüfungsrichtlinien vor Ort. ...“
Anlage, Abs. 2.3	Sollte lauten: „Das Glossar der in den UPOV-Dokumenten verwendeten Begriffe (<u>Dokument</u> TGP/14/2/) liefert nützliche Anleitung zur Klarstellung vieler Merkmale, insbesondere von PQ-Merkmalen.“

VORSCHLAG

18. Es wird vorgeschlagen, den Entwurf einer Anleitung zu „Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch“, wie in der Anlage dieses Dokuments dargelegt, in Verbindung mit den Anmerkungen der TWP auf ihren Tagungen im Jahr 2015 und des TC-EDC auf seiner Tagung im Jahr 2016 als Grundlage für eine künftige Überarbeitung von Dokument TGP/8: „Prüfungsanlage

und Verfahren für die Prüfung der Unterscheidbarkeit, der Homogenität und der Beständigkeit“, Teil I: DUS-Prüfungsanlage und Datenanalyse, zu prüfen.

19. *Der TC wird ersucht:*

a) *den Entwurf einer Anleitung zu „Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch“, wie in der Anlage dieses Dokuments dargelegt, in Verbindung mit den Anmerkungen der TWP auf ihren Tagungen im Jahr 2015 und des TC-EDC auf seiner Tagung im Jahr 2016 als Grundlage für eine künftige Überarbeitung von Dokument TGP/8: „Prüfungsanlage und Verfahren für die Prüfung der Unterscheidbarkeit, der Homogenität und der Beständigkeit“, Teil I: DUS-Prüfungsanlage und Datenanalyse, einschließlich insbesondere, ob in dem Dokument ein anderer Abschnitt zu entwickeln ist, um Anleitung zur Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch für PQ-Merkmale, einschließlich der Anwendung nicht parametrischer Verfahren zu erteilen, zu prüfen;*

b) *zu prüfen, ob die TWP auf ihren Tagungen im Jahr 2016 ersucht werden sollten, Informationen über Variationen zwischen Erfassern für PQ-Merkmale, wie etwa Häufigkeit von Abweichungen, zu erteilen;*

c) *zur Kenntnis zu nehmen, daß die TWC vereinbarte, die Sachverständigen aus Argentinien und Brasilien aufzufordern, auf ihrer vierunddreißigsten Tagung ein Referat über ihre Erfahrungen mit Schulungen für die Minimierung von Variationen zwischen Erfassern bei PQ-Merkmalen zu halten;*

d) *zu prüfen, ob Sachverständige eingeladen werden sollen, um den TWP auf ihren Tagungen im Jahr 2016 ihre Erfahrungen im Hinblick auf die Minimierung der Variation infolge verschiedener Erfasser im selben Anbauversuch für PQ-Merkmale darzulegen.*

[Anlage folgt]

TGP/8/1: TEIL I: NEUER ABSCHNITT: MINIMIERUNG DER VARIATION INFOLGE VERSCHIEDENER
ERFASSER IM SELBEN ANBAUVERSUCH

Im Vergleich zu der dem TC auf seiner einundfünfzigsten Tagung dargelegten Fassung sind die vorgeschlagenen Änderungen nachfolgend durch Unterstreichen (Einfügungen) und ~~Durchstreichen~~ (Streichungen) angegeben.

1. Einleitung

In diesem Dokument werden Variationen zwischen Erfassern im selben Anbauversuch auf Ebene der Behörden geprüft. Es wurde im Hinblick auf QN/MG-, QN/MS-, QN/VG- und QN/VS-Merkmale erstellt. Es behandelt nicht ausdrücklich PQ-Merkmale wie Farbe und Form. Das beschriebene Kappa-Verfahren ist an sich weitgehend auf diese Merkmale anwendbar, z. B. wird das Standard-Kappa-Merkmal für nominale Daten entwickelt. Allerdings wurde dieses Verfahren ~~unseres Wissens nach nicht beifür~~ angewandt entwickelt und ~~PQ-Merkmale~~ könnten auch Zusatzinformationen zur Kalibrierung erfordern. Beispielsweise müssen bei der Farbkalibrierung auch die RHS-Farbkarte, die Lichtbedingungen und so weiter berücksichtigt werden. Unterschiede zwischen Erfassern bei PQ-Merkmalen könnten anhand nichtparametrischer Verfahren, wie etwa der Häufigkeit von Abweichungen, geprüft werden. Diese Aspekte werden in diesem Dokument nicht behandelt.

1.1 Variation bei Messungen oder Erfassungen kann durch viele verschiedene Faktoren, wie etwa Pflanzentyp, Merkmalstyp, Jahr, Ort, Prüfungsanlage und -verwaltung, Verfahren und Erfasser, verursacht werden. Besonders bei visuell erfaßten Merkmalen (QN/VG oder QN/VS) können Unterschiede zwischen Erfassern Grund für große Variation und potentielle Abweichung bei den Erfassungen sein. Ein Erfasser könnte weniger gut geschult sein oder die Merkmale anders interpretieren. Erfaßt also Erfasser A die Sorte 1 und Erfasser B die Sorte 2, so könnten die erfaßten Unterschiede auf Unterschiede zwischen Erfasser A und B statt auf Unterschiede zwischen Sorte 1 und 2 zurückzuführen sein. Unser Interesse gilt eindeutig den Unterschieden zwischen den Sorten und nicht den Unterschieden zwischen den Erfassern. Wichtig ist, daß man sich der Tatsache bewußt ist, daß die durch verschiedene Erfasser verursachte Variation nicht ausgeschaltet werden kann, aber daß es Möglichkeiten dafür gibt, sie unter Kontrolle zu halten.

1.2 Es wird empfohlen, nach Möglichkeit einen Erfasser pro Anbauversuch einzusetzen, um die Variation bei den Erfassungen aufgrund verschiedener Erfasser zu minimieren.

2. Schulung und Bedeutung klarer Erklärungen für Merkmale und Erfassungsmethoden

2.1 Die Schulung neuer Erfasser ist grundlegend wichtig für die Stetigkeit und Kontinuität von Erfassungen an Pflanzensorten. Kalibrierungshandbücher, Aufsicht und Anleitung durch erfahrene Erfasser sowie auch die Verwendung von Vergleichssorten, die die Bandbreite an Ausprägungen illustrieren, sind nützliche Wege dafür, dies zu erreichen.

2.2 Mit den UPOV-Prüfungsrichtlinien wird versucht, das Verfahren der Sortenbeschreibung zu harmonisieren und die Merkmale einer Pflanze sowie die Ausprägungsstufen so klar wie möglich zu beschreiben. Das ist der erste Schritt zur Kontrolle von Variation und Abweichung. Allerdings kann die Art und Weise, in der ein Merkmal erfaßt oder gemessen wird, je nach Jahr, Ort oder Prüfungsbehörde variieren. Von den lokalen Prüfungsbehörden erstellte Kalibrierungshandbücher und Beispielsorten sind sehr nützlich für die Umsetzung der UPOV-Prüfungsrichtlinien vor Ort. Bei Bedarf erläutern diese pflanzenspezifischen Handbücher die zu erfassenden Merkmale detaillierter und führen genau aus, wann und wie sie zu erfassen sind. Außerdem enthalten sie möglicherweise Bilder und Zeichnungen zu jedem Merkmal, oft sogar zu jeder Ausprägungsstufe eines Merkmals.

2.43 Das Glossar der in den UPOV-Dokumenten verwendeten Begriffe (Dokument TGP/14/2) liefert nützliche Anleitung zur Klarstellung vieler Merkmale, insbesondere von PQ-Merkmalen.

2.34 Ist ein Erfasser erst einmal geschult, so ist es wichtig, häufige Auffrischkurse und Rekalibrierung zu gewährleisten.

3. Prüfung der Kalibrierung

Nach der Schulung eines Erfassers könnte der nächste Schritt darin bestehen, die Leistung des Erfassers im Rahmen eines Kalibrierungsversuchs zu prüfen. Das ist besonders nützlich für unerfahrene Erfasser, die visuelle Erfassungen vornehmen müssen (QN/VG- und QN/VS-Merkmale). Bei visuellen Erfassungen sollten sie vorzugsweise einer Kalibrierungsprüfung unterzogen werden, bevor sie Erfassungen im Anbauversuch vornehmen. Aber auch für erfahrene Erfasser ist es nützlich, sich selbst regelmäßig zu testen, um zu prüfen, daß sie die Kalibrierungskriterien immer erfüllen.

3.2 Ein Kalibrierungsversuch kann auf verschiedene Art und Weise angelegt und ausgewertet werden. Im Allgemeinen sind mehrere Erfasser daran beteiligt, die denselben Satz an Material erfassen und die Unterschiede zwischen den Erfassern bewerten.

4. Prüfung der Kalibrierung für QN/MG- oder QN/MS- Merkmale

4.1 Bei Erfassungen, die anhand von Messinstrumenten wie etwa Linealen erfolgen (oftmals QN/MS-Merkmale), erfolgt die Messung oft auf einer Intervall- oder Verhältnisskala. In diesem Fall kann nach dem Ansatz von Bland und Altman (1986) vorgegangen werden. Bei diesem Ansatz wird zunächst eine graphische Darstellung der von jedem Erfasserpaar gemessenen Werte in einem Streudiagramm erstellt und mit der Gleichheitslinie (bei der $y=x$ ist) verglichen. Das erleichtert dem Auge, den Grad an Übereinstimmung der Messungen desselben Objekts zu erkennen. Im nächsten Schritt wird anhand des Unterschieds pro Erfassungsobjekt eine graphische Darstellung erstellt, wobei der Unterschied zwischen den Erfassern auf der y-Achse und entweder der Objektverweis oder der Mittelwert des Objekts auf der x-Achse eingetragen werden. Werden die horizontalen Linien $y=0$, $y=\text{Mittel(Differenz)}$ und die zwei Linien $y = \text{Mittel(Differenz)} \pm 2 \times \text{Standardabweichung}$ weitergezogen, so ist die Abweichung zwischen den Erfassern und eventuellen Ausreißern ganz einfach zu erkennen. Auf ähnliche Weise können wir auch die Differenz zwischen der Messung jedes Erfassers und dem Durchschnitt der Messungen aller Erfasser betrachten. Prüfungsmethoden wie der gepaarte t-Test können zur Prüfung einer signifikanten Abweichung des Erfassers von einem anderen Erfasser oder vom Mittelwert der anderen Erfasser angewandt werden.

4.2 Ziehen wir zwei Messungen von jedem Erfasser von jedem Objekt heran, so können wir die Unterschiede zwischen diesen beiden Messungen betrachten. Sind diese Unterschiede im Vergleich zu jenen anderer Erfasser groß, so könnte dieser Erfasser eine geringe Wiederholbarkeit aufweisen. Zählen wir für jeden Erfasser die Zahl der mäßigen und großen Ausreißer (z. B. jeweils 2 und 3 Mal so groß wie die Standardabweichung), so können wir eine Tabelle mit Erfasser versus Anzahl Ausreißer erstellen, die dazu verwendet werden kann, zu entscheiden, ob der Erfasser innerhalb der Qualitätssicherungsgrenzwerte liegt.

4.3 Andere Qualitätsprüfungen können auf Wiederholbarkeits- und Reproduzierbarkeitsprüfungen für Labore basieren, wie in ISO 5725-2 beschrieben. Kostenlose Software ist auf der ISTA-Website verfügbar, um Werte und Diagramme gemäß diesem ISO-Standard zu erhalten.

4.4 Allerdings ist in vielen Fällen von QN/MG oder QN/MS eine gute und klare Anleitung normalerweise ausreichend und die Variation oder Abweichung von Messungen ist zwischen den Erfassern oftmals unerheblich. Besteht Anlaß zu Zweifel, so kann ein Kalibrierungsversuch wie oben beschrieben in dieser Situation Klarheit verschaffen.

4.5 Im Falle von QN/MG-Erfassungen muß eventuell der mögliche Zufall zwischen Parzellen geprüft und berücksichtigt werden.

5. Prüfung der Kalibrierung für QN/VS- oder QN/VG- Merkmale

5.1 Für die Auswertung ordinaler Daten (QN/VS- oder QN/VG-Merkmale) ist die Erstellung von Kontingenztafeln zwischen jedem Erfasserpaar für die unterschiedlichen Ergebnisse sehr aufschlußreich. Eine Prüfung auf strukturelle Differenz (Abweichung) zwischen zwei Erfassern kann unter Verwendung des Wilcoxon-Tests für Paardifferenzen (oft auch Wilcoxon Vorzeichen-Rangtest genannt) durchgeführt werden.

5.2 Zur Messung des Grades an Übereinstimmung wird oftmals Cohens Kappa (κ)-Statistik (Cohen, 1960) verwendet. Die Statistik versucht, zufällige Übereinstimmung zu berücksichtigen: $\kappa = P(\text{Übereinstimmung}) - P(e) / (1 - P(e))$, wobei $P(\text{Übereinstimmung})$ der Bruchteil von Objekten ist, die sich für beide Beurteiler in derselben Klasse befinden (die Mitteldiagonale der Kontingenztafel) und $P(e)$ in Anbetracht der Randverteilungen die Wahrscheinlichkeit zufälliger Übereinstimmung ist (wie in einem Chi-Quadrat-Test). Besteht vollkommene Übereinstimmung zwischen den Erfassern, so ist der Kappa-Wert $\kappa = 1$. Besteht außer der zufällig zu erwartenden Übereinstimmung ($P(e)$) keine Übereinstimmung zwischen den Erfassern, dann ist $\kappa = 0$.

5.3 Die Standard-Kappa-Statistik nach Cohen berücksichtigt nur vollkommene Übereinstimmung versus Nichtübereinstimmung. Soll der Grad an Nichtübereinstimmung ermittelt werden (z.B. im Falle von ordinalen Merkmalen), kann ein lineares oder ein quadratisch gewichtetes Kappa angewandt werden (Cohen, 1968). Soll eine einzige Statistik für alle Erfasser gleichzeitig erstellt werden, kann ein generalisierter Kappa-Koeffizient berechnet werden. Die meisten statistischen Softwarepakete, einschließlich SPSS, Genstat und R (Concord Paket) enthalten Tools zur Berechnung der Kappa-Statistik.

5.4 Wie bereits angemerkt, steht ein niedriger κ -Wert für geringe Übereinstimmung und Werte nahe 1 stehen für hervorragende Übereinstimmung. Oft werden Ergebnisse zwischen 0,6 und 0,8 als grundsätzliche Übereinstimmung betrachtet und Werte über 0,8 gelten praktisch als vollkommene Übereinstimmung. Bei Bedarf stehen auch z-Tests für Kappa zur Verfügung (wenn eine annähernd normale Verteilung angenommen wird). Die Kriterien für erfahrende DUS-Sachverständige könnten stringenter als für unerfahrene Mitarbeiter sein.

6. Prüfungsanlage

6.1 Sind mehrere Erfasser an einer Prüfung beteiligt, so wird am besten so vorgegangen, daß eine Person eine oder mehrere vollständige Wiederholung/en beobachtet. In diesem Fall erklärt die Korrektur aufgrund von Blockeffekten auch die Abweichung zwischen Erfassern. Wird mehr als ein Erfasser pro Wiederholung benötigt, dann sollte zusätzlich auf Kalibrierung und Übereinstimmung geachtet werden. In einigen Fällen kann die Verwendung unvollständiger Block-Designs (wie Alpha-Designs) hilfreich sein und ein Erfasser kann für die untergeordneten Blöcken eingeteilt werden. Auf diese Weise können wir die systematische Abweichung zwischen den Erfassern korrigieren.

7. Beispiel für Cohens Kappa

7.1 In diesem Beispiel gibt es drei Erfasser und 30 Objekte (Parzellen oder Sorten). Das Merkmal wird auf einer Skala von 1 bis 6 erfaßt. Die unaufgearbeiteten Daten und ihre tabellarisierten Werte sind in folgenden Tabellen enthalten:

Sorte	Erfasser		
	1	2	3
V1	1	1	1
V2	2	1	2
V3	2	2	2
V4	2	1	2
V5	2	1	2
V6	2	1	2
V7	2	2	2
V8	2	1	2
V9	2	1	2
V10	3	1	3
V11	3	1	3
V12	3	2	2
V13	4	5	4
V14	2	1	1
V15	2	1	2
V16	2	2	3
V17	5	4	5
V18	2	2	3
V19	1	1	1
V20	2	2	2
V21	2	1	2
V22	1	1	1
V23	6	3	6
V24	5	6	6
V25	2	1	2
V26	6	6	6
V27	2	6	2
V28	5	6	5
V29	6	6	5
V30	4	4	4

Werte für Sorte	1	2	3	4	5	6
V1	3	0	0	0	0	0
V2	1	2	0	0	0	0
V3	0	3	0	0	0	0
V4	1	2	0	0	0	0
V5	1	2	0	0	0	0
V6	1	2	0	0	0	0
V7	0	3	0	0	0	0
V8	1	2	0	0	0	0
V9	1	2	0	0	0	0
V10	1	0	2	0	0	0
V11	1	0	2	0	0	0
V12	0	2	1	0	0	0
V13	0	0	0	2	1	0
V14	2	1	0	0	0	0
V15	1	2	0	0	0	0
V16	0	2	1	0	0	0
V17	0	0	0	1	2	0
V18	0	2	1	0	0	0
V19	3	0	0	0	0	0
V20	0	3	0	0	0	0
V21	1	2	0	0	0	0
V22	3	0	0	0	0	0
V23	0	0	1	0	0	2
V24	0	0	0	0	1	2
V25	1	2	0	0	0	0
V26	0	0	0	0	0	3
V27	0	2	0	0	0	1
V28	0	0	0	0	2	1
V29	0	0	0	0	1	2
V30	0	0	0	3	0	0

Die Kontingenztafel für Erfasser 1 und 2 ist:

O1\O2	1	2	3	4	5	6	Insgesamt
1	3	0	0	0	0	0	3
2	10	5	0	1	0	1	17
3	2	1	0	0	0	0	3
4	0	0	0	1	0	0	1
5	0	0	0	1	0	2	3
6	0	0	1	0	0	2	3
Insgesamt	15	6	1	3	0	5	30

Der Kappa-Koeffizient zwischen Erfasser 1 und 2, $\kappa(O1,O2)$, wird wie folgt berechnet:

- $\kappa(O1,O2) = (P(\text{Übereinstimmung zwischen } O1 \text{ und } O2) - P(e)) / (1 - P(e))$ wobei:
- $P(\text{Übereinstimmung}) = (3+5+0+1+0+2)/30 = 11/30 \approx 0,3667$ (diagonale Elemente)
- $P(e) = (3/30) \cdot (15/30) + (17/30) \cdot (6/30) + (3/30) \cdot (1/30) + (1/30) \cdot (3/30) + (3/30) \cdot (0/30) + (3/30) \cdot (5/30) \approx 0,1867$. (paarweise Randverteilungen)
- So $\kappa(O1,O2) \approx (0,3667 - 0,1867) / (1 - 0,1867) \approx 0,22$

Das ist ein niedriger Wert, der sehr geringe Übereinstimmung zwischen diesen beiden Erfassern zeigt. Es gibt Grund zur Besorgnis und es sollten Maßnahmen ergriffen werden, um die Übereinstimmung zu verbessern. Auf ähnliche Weise können die Werte für die anderen Paare berechnet werden: $\kappa(O1,O3) \approx 0,72$, $\kappa(O2,O3) \approx 0,22$. Erfasser 1 und 3 zeigen gute Übereinstimmung. Erfasser 2 unterscheidet sich deutlich von 1 und 3 und die Gründe für die Abweichung müssen noch weiter untersucht werden (z. B. weiteren Schulungsbedarf prüfen).

8. Literaturhinweise

Cohen, J. (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20: 37-46.

Cohen, J. (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70(4): 213-220.

Bland, J. M. Altman D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement, Lancet: 307–310.

<http://www.seedtest.org/en/stats-tool-box-content---1--1143.html> (auf ISO 5725-2 basierende Software)

[Ende der Anlage und des Dokuments]