



TC/51/16

ORIGINAL: englisch

DATUM: 11. Februar 2015

**INTERNATIONALER VERBAND ZUM SCHUTZ VON PFLANZENZÜCHTUNGEN**  
Genf

**TECHNISCHER AUSSCHUSS**

**Einundfünfzigste Tagung  
Genf, 23. bis 25. März 2015**

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

*vom Verbandsbüro erstelltes Dokument*

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

1. Zweck dieses Dokuments ist es, einen Entwurf für einen neuen Abschnitt für Dokument TGP/8 Teil I: DUS-Prüfungsanlage und Datenanalyse, zur „Minimierung der Variation infolge verschiedener Erfasser“ darzulegen.

2. 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  |

3. Der Aufbau dieses Dokuments ist wie folgt:

|                                 |   |
|---------------------------------|---|
| HINTERGRUND .....               | 1 |
| ENTWICKLUNGEN IM JAHR 2014..... | 2 |
| TECHNISCHER AUSSCHUSS .....     | 2 |
| TECHNISCHE ARBEITSGRUPPEN ..... | 2 |

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

HINTERGRUND

4. 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]“

5. Entwicklungen vor dem Jahr 2014 sind in Dokument TC/50/21 „Überarbeitung von TGP/8: Teil I: DUS-Prüfungsanlage und Datenanalyse, Neuer Abschnitt: Minimierung der Variation infolge verschiedener Erfasser“ dargelegt.

## ENTWICKLUNGEN IM JAHR 2014

### Technischer Ausschuß

6. Der TC prüfte auf seiner fünfzigsten Tagung vom 7. bis 9. April 2014 in Genf Dokument TC/50/21 „Überarbeitung von Dokument TGP/8: Teil I: DUS-Prüfungsanlage und Datenanalyse, Neuer Abschnitt: Minimierung der Variation infolge verschiedener Erfasser“

7. Der TC nahm zur Kenntnis, daß die TWF einen Sachverständigen aus Neuseeland ersucht habe, auf ihrer Tagung im Jahr 2014 über die zuvor erstellte Arbeit über harmonisierte Sortenbezeichnung für Apfel für eine vereinbarte Sortenserie zu berichten (vergleiche Dokument TC/50/36 „Bericht über die Entschlüsseungen“, Absatz 46).

8. Der TC ersuchte den Sachverständigen aus Australien, mit Unterstützung von Sachverständigen aus der Europäischen Union, Frankreich, Deutschland, den Niederlanden und dem Vereinigten Königreich, einen Entwurf einer weiteren Anleitung zur Aufnahme in eine künftige Überarbeitung von Dokument TGP/8 zur Minimierung der Variation infolge verschiedener Erfasser, einschließlich einer Anleitung zu PQ- und QN/MG-Merkmalen zur Prüfung durch die TWP auf ihren Tagungen im Jahr 2014 zu verfassen (vergleiche Dokument TC/50/36 „Bericht über die Entschlüsseungen“, Absatz 47).

9. Als Antwort auf das Ersuchen des TC stellte der Verfaßer aus Australien (Herr Nik Hulse) mit Unterstützung von Sachverständigen aus der Europäischen Union, Frankreich, Deutschland, den Niederlanden und dem Vereinigten Königreich einen Entwurf einer Anleitung für eine künftige Überarbeitung von Dokument TGP/8 über die Minimierung der Variation infolge verschiedener Erfasser, einschließlich einer Anleitung zu PQ- und QN/MG-Merkmalen zur Prüfung durch die TWP auf ihren Tagungen im Jahr 2014, bereit.

10. Der Sachverständige aus Australien schlug vor, daß folgende Punkte von den TWP für die künftige Aufnahme oder Klarstellung in den nächsten Entwurf der Anleitung zur weiteren Prüfung von Dokument TGP/8 über die Minimierung der Variation infolge verschiedener Erfasser geprüft werden sollten:

- Wie bereits im Entwurf angegeben, kann mit QN/MG auf ähnliche Weise wie mit QN/MS verfahren werden. Allerdings ist wichtig anzumerken, daß auch die mögliche Wirkung von zufälliger Variation zwischen Parzellen im Falle von QN/MG geprüft werden sollte.
- Ferner muß geprüft werden, wie Anleitung zu PQ-Merkmalen aufgenommen werden kann. Es sollte erklärt werden, daß TGP/14 ein weiteres nützliches Instrument zur Klarstellung vieler PQ-Merkmale (z. B. Form) ist. Unterschiede zwischen Erfassern könnte möglicherweise anhand der Häufigkeit von Abweichungen geprüft werden.
- Derzeit liegt der Schwerpunkt des Dokuments auf Variation zwischen Erfassern auf Ebene der Behörde. Es könnte geprüft werden, ob die Minimierung der Erfasservariation zwischen Behörden in diesem Dokument oder eventuell in einem getrennten künftigen Dokument angeführt werden sollte. Dabei ist anzumerken, daß solch eine Prüfung eine größere Anzahl an Faktoren, wie etwa die GxE-Variation, Verfahren zur Probenentnahme und Ringprüfungen, einführen würde.

### Technische Arbeitsgruppen

11. Die TWO, TWF, TWC, TWV und TWA prüften die Dokumente TWO/47/15, TWF/45/15, TWC/32/15, TWV/48/15 und TWA/43/15 „Überarbeitung von Dokument TGP/8: Teil I: DUS-Prüfungsanlage und

Datenanalyse, Neuer Abschnitt: Minimierung der Variation infolge verschiedener Erfasser“ (vergleiche Dokument TWO/47/28 „Report“, Absätze 39 und 41, Dokument TWF/45/32 „Report“, Absätze 26 bis 32, Dokument TWC/32/28 „Report“, Absätze 50 und 53, Dokument TWV/48/43 „Report“, Absätze 34 bis 37 und Dokument TWA/43/27 „Report“, Absätze 28 bis 33).

12. Die TWO nahm zur Kenntnis, daß die TWF einen Sachverständigen aus Neuseeland ersucht hatte, auf ihrer Tagung im Jahr 2014 über die zuvor erstellte Arbeit über harmonisierte Sortenbezeichnung für Apfel für eine vereinbarte Sortenserie zu berichten, wie in Dokument TWO/47/15, Absatz 18 dargelegt (vergleiche Dokument TWO/47/28 „Report“, Absatz 40).

13. Die TWO, TWF, TWC, TWV und TWA vereinbarten, daß der auf ihren Tagungen vorgestellte Entwurf einer Anleitung jeweils in der Anlage der Dokumente TWO/47/15, TWF/45/15, TWC/32/15, TWV/48/15 und TWA/43/15 weiterentwickelt werden sollte für die Aufnahme in eine künftige Überarbeitung von Dokument TGP/8 zur Minimierung der Variation infolge verschiedener Erfasser, einschließlich einer Anleitung zu PQ- und QN/MG-Merkmalen in Verbindung mit den vom Sachverständigen aus Australien aufgeworfenen Punkten in den Dokumenten TWO/47/15, TWF/45/15, TWC/32/15, TWV/48/15 und wiedergegeben in Absatz 21 dieses Dokuments (vergleiche Dokumente TWO/47/28 „Report“, Absatz 41, TWF/45/32 „Report“, Absatz 27, TWC/32/28 „Report“, Absatz 52, TWV/48/43 „Report“, Absatz 34 und TWA/43/27 „Report“, Absatz 30).

14. Die TWO und die TWA vereinbarten, daß das Dokument sich auf die Variation zwischen Erfassern auf Ebene einer Behörde und nicht auf die Minimierung der Variation zwischen Erfassern zwischen Behörden konzentrieren sollte (vergleiche Dokumente TWO/47/28 „Report“, Absatz 41 und TWA/43/27 „Report“, Absatz 31).

15. Die TWF hörte ein Referat von den Sachverständigen aus Deutschland und Neuseeland über die zuvor erstellte Arbeit über harmonisierte Sortenbezeichnung für Apfel für eine vereinbarte Sortenserie, wie in Dokument TWF/45/28 wiedergegeben (vergleiche Dokument TWF/45/32 „Report“, Absatz 28).

16. Die TWF war sich darin einig, daß es wichtig sei, die Variation zwischen verschiedenen Erfassern und auch zwischen Behörden zu minimieren und schlug deshalb vor, eine Studie über die Möglichkeit der Aufnahme eines neuen Projekts über harmonisierte Sortenbeschreibungen für eine vereinbarte Sortenserie in Betracht zu ziehen. Der Sachverständige aus Deutschland schlug vor, der TWF auf ihrer sechundvierzigsten Tagung ein Protokoll für das Projekt mit einer zu prüfenden vereinbarten Sortenliste vorzulegen, um zu prüfen, ob es zweckmäßig sein könnte, die Studie weiterzuentwickeln (vergleiche Dokument TWF/45/32 „Report“, Absatz 30).

17. Die TWF und die TWV merkten an, daß die Qualität der Prüfungsrichtlinien im Hinblick auf die Erteilung klarer Anleitung für DUS-Prüfer und die Gewährleistung der Übereinstimmung von Erfassungen zwischen Erfassern innerhalb jeder Behörde sowie auch die kontinuierliche Schulung von Prüfern wichtig seien (vergleiche Dokumente TWF/45/32 „Report“, Absatz 31 und TWV/48/43 „Report“, Absatz 35).

18. Die TWF ersuchte den Sachverständigen aus Australien, auf ihrer sechundvierzigsten Tagung über die Auswirkung von Ort, Erfasser und Jahr auf die Konformität eines Merkmals für eine spezifische Pflanze zu berichten (vergleiche Dokument TWF/45/32 „Report“, Absatz 32).

19. Die TWC nahm zur Kenntnis, daß der Sachverständige aus Neuseeland auf der fünfundvierzigsten Tagung der TWF über die zuvor erstellte Arbeit über harmonisierte Sortenbeschreibungen für Apfel für eine vereinbarte Sortenserie berichtet hatte (vergleiche Dokument TWC/32/28 „Report“, Absatz 51).

20. Die TWC vereinbarte, daß der Entwurf einer Anleitung zunächst einmal mit Variation zwischen Erfassern auf Ebene einer Behörde und in einer künftigen Phase zwischen verschiedenen Behörden beginnen sollte (vergleiche Dokument TWC/32/28 „Report“, Absatz 53).

21. Die TWV schlug die Aufnahme einer Schulungsübung in einen DUS-Anbauversuch als Grundlage für den Austausch von Erfahrungen auf dem Feld und zur Verbesserung der Nutzung der TWV für die Schulung vor (vergleiche Dokument TWV/48/43 „Report“, Absatz 36).

22. Die TWA nahm zur Kenntnis, daß ein Sachverständiger aus Neuseeland auf der fünfundvierzigsten Tagung der TWF über die zuvor erstellte Arbeit über harmonisierte Sortenbeschreibungen für Apfel für eine vereinbarte Sortenserie berichtet habe, wie in TWF/45/28 „Harmonized example varieties for Apple: historical data and possible new developments“ dargelegt (vergleiche Dokument TWA/43/27 „Report“, Absatz 29).

23. Die TWA nahm zur Kenntnis, daß die Qualität der Prüfungsrichtlinien im Hinblick auf die Erteilung klarer Anleitung für DUS-Prüfer und die Gewährleistung der Übereinstimmung von Erfassungen zwischen Erfassern innerhalb jeder Behörde und die kontinuierliche Schulung von Prüfern wichtig seien. Die TWA vereinbarte, eine allgemeine Empfehlung vorzuschlagen, nach der möglichst ein Erfasser pro Anbauversuch eingesetzt werden sollte, um Variation bei den Erfassungen zu vermeiden (vergleiche Dokument TWA/43/27 „Report“, Absatz 32).

24. Die TWA vereinbarte, daß mit QN/MG-Merkmalen auf ähnliche Weise wie mit QN/MS-Merkmalen verfahren werden könnte und daß auch die möglichen Auswirkungen auf zufällige Variation zwischen Parzellen geprüft werden sollte. Die TWA vereinbarte, daß Unterschiede zwischen Erfassern bei PQ Merkmalen anhand nichtparametrischer Verfahren, wie etwa der Häufigkeit von Abweichungen, geprüft werden könnten (vergleiche Dokument TWA/43/27 „Report“, Absatz 33).

25. In Antwort auf die von den TWP auf ihren Tagungen im Jahr 2015 gemachten Bemerkungen, arbeitete Herr Nik Hulse (Australien) einen neuen Entwurf einer Anleitung für die Aufnahme in TGP/8 Teil I: DUS-Prüfung und Prüfungsanlage und Datenanalyse, neuer Abschnitt: Minimierung der Variation infolge verschiedener Erfasser, wie in der Anlage zu diesem Dokument wiedergegeben, zur Prüfung durch die TWP auf ihren Tagungen im Jahr 2015 aus.

26. *Der TC wird ersucht,*

*a) den Entwurf einer Anleitung für die Aufnahme in Dokument TGP/8 über die Minimierung der Variation infolge verschiedener Erfasser, wie in der Anlage dieses Dokuments wiedergegeben, in Verbindung mit den Bemerkungen der TWP auf ihren Tagungen im Jahr 2014 zu prüfen; und*

*b) die TWP zu ersuchen, einen neuen Entwurf, der alle vom TC auf seiner einundfünfzigsten Tagung vereinbarten Änderungen einschließt, auf ihren Tagungen im Jahr 2015 zu prüfen.*

[Anlage folgt]

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

### 1. Einleitung

Dieses Dokument 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 bei PQ-Merkmalen angewandt 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. 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 Ort oder Prüfungsbehörde variieren. Von den lokalen Prüfungsbehörden erstellte Kalibrierungshandbücher 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.

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

2.3 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 die Variation 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 ( $\kappa$ )-Statistik (Cohen, 1960) verwendet. Die Statistik versucht, zufällige Übereinstimmung zu berücksichtigen:  $\kappa = \frac{P(\text{Übereinstimmung}) - P(e)}{1 - P(e)}$ , 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  $\kappa$ -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 | Erfasser 2 | Erfasser 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)
- $P(e) = (3/30).(15/30) + (17/30).(6/30) + (3/30).(1/30) + (1/30).(3/30) + (3/30).(0/30) + (3/30).(5/30) \approx 0,1867$ . (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]