

# Statistische Eigenschaften

## der STATIS-Methode

—

## Propriétés statistiques de la méthode STATIS

### DISSERTATION

zur Erlangung des Grades eines

**DOKTORS DER NATURWISSENSCHAFTEN**

der **Universität Dortmund**

und der **Universität Rennes II – Haute Bretagne**

### THÈSE

pour obtenir le grade de

**DOCTEUR**

de l'**Université de Dortmund**

et de l'**Université Rennes II – Haute Bretagne**

**EN SCIENCES NATURELLES ET EXPERIMENTALES**

Dem Fachbereich Statistik der  
Universität Dortmund  
und der  
Universität Rennes II – Haute Bretagne  
vorgelegt von

Présenté au Département de Statistique  
de l'Université de Dortmund  
et au  
Université Rennes II – Haute Bretagne  
par

**Michael Meyners**

Dortmund

September 2001

Septembre 2001

**Gutachter**

**Rapporteurs**

Prof. Dr. Joachim Kunert

Prof. Dr. El Mostafa Qannari

**Tag der mündlichen Prüfung**

**Date de la soutenance**

21. Februar 2002

21 février 2002

# Inhaltsverzeichnis

Résumé	v
<b>1 Einleitung</b>	<b>1</b>
<b>2 Modellannahmen</b>	<b>5</b>
<b>3 Die Methode STATIS</b>	<b>7</b>
3.1 Originalversion . . . . .	7
3.2 Theoretische Alternativversion . . . . .	9
<b>4 Statistische Eigenschaften von STATIS</b>	<b>11</b>
4.1 Problematik bei der theoretischen Untersuchung der Eigenschaften . .	11
4.2 Konvergenz der STATIS-Gewichte für $n \rightarrow \infty$ . . . . .	12
4.3 Betrachtung von Spezialfällen bei der Konvergenz der Gewichte . . .	17
4.4 Untersuchung des Abstandes des STATIS-Konsens vom wahren Konsens	19
4.5 STATIS-Version mit festen Gewichten für $n \rightarrow \infty$ . . . . .	25
4.6 Konvergenz des Abstandes unter weiteren Annahmen . . . . .	28
4.7 Erwartungswert des STATIS-Konsens bei festen Gewichten . . . . .	31
4.8 Konvergenz des STATIS-Konsens für $p \rightarrow \infty$ . . . . .	35
4.9 Konvergenz, sofern die Fehlervarianzen gegen Null streben . . . . .	36

<b>5</b>	<b>Schätzung der Fehlermatrizen</b>	<b>37</b>
<b>6</b>	<b>Simulationen zum Vergleich der Ergebnisse</b>	<b>39</b>
<b>7</b>	<b>Vergleich mit der Verallgemeinerten Prokrustes Analyse (GPA)</b>	<b>49</b>
<b>8</b>	<b>Grafischer Vergleich der Analysemethoden</b>	<b>53</b>
<b>9</b>	<b>Zusammenfassung und Ausblick</b>	<b>59</b>
	<b>Literaturverzeichnis</b>	<b>63</b>

## Résumé

Pour l'analyse des données sensorielles, plusieurs méthodes sont utilisées, parmi lesquelles on peut citer la méthode Procruste (GPA, Gower, 1975 ; ten Berge, 1977) et la méthode STATIS (Lavit *et al.*, 1994 ; Schlich 1996). Meyners, Kunert et Qanari (2000) comparent ces deux méthodes à l'aide de simulations, lesquelles se sont révélées à l'avantage de GPA. Ils donnent aussi des résultats théoriques montrant que la méthode STATIS fournit un compromis trop complexe, c'est-à-dire avec un nombre trop élevé de dimensions. Même dans les cas où les produits ne présentent pas de différences significatives, STATIS génère une structure à plusieurs dimensions. GPA implique en revanche la mise en œuvre d'un algorithme itératif. La convergence d'un tel algorithme n'est pas attestée. En outre, le nombre de paramètres à estimer est élevé par rapport au nombre d'observations.

C'est la raison pour laquelle nous proposons d'utiliser la méthode STATIS en lui apportant une correction. Dans le modèle de Meyners *et al.* (2000), avec  $p$  juges et  $n$  produits, nous supposons que la moyenne des erreurs est nulle et que les erreurs sont indépendantes entre les juges. En partant de ces hypothèses, STATIS fournit un résultat que l'on considère comme étant le compromis théorique plus une matrice semi-définie positive. Cette matrice est associée aux termes d'erreurs et aux scalaires. Etant donné que cette matrice est semi-définie positive, nous obtenons un compromis présentant un nombre élevé de dimensions. Si nous connaissions les matrices d'erreurs  $E_i$  et les scalaires  $\lambda_i$ , nous pourrions soustraire la matrice  $K$  du

compromis  $W$  de STATIS, sachant que :

$$K := \sum_{i=1}^p u_i \lambda_i^2 E_i E_i^T .$$

Nous étudions plusieurs hypothèses supplémentaires et comparons les estimateurs des deux méthodes. Nous constatons, en théorie, que les résultats de cette variante de STATIS sont plus performants que ceux de la méthode originale, c'est-à-dire par exemple que la distance entre le compromis estimé et le compromis théorique est plus réduite ou que les estimateurs sont plus proches du compromis théorique quand le nombre de juges ou de produits tend vers l'infini.

En pratique, les termes d'erreurs et les scalaires sont inconnus. Nous proposons donc la procédure suivante : dans un premier temps, nous calculons une somme  $W$  pondérée des matrices des produits scalaires associées aux configurations des différents juges. Par la suite, nous calculons les composantes principales afin d'obtenir le compromis de la méthode STATIS, soit  $\hat{C}$ . De manière à déterminer les termes d'erreurs, nous calculons une rotation procrustéenne  $\hat{R}_i$  et un scalaire optimal  $\hat{\lambda}_i$  pour adapter la configuration  $X_i$  de chaque juge au compromis et déterminer les différences entre ces matrices et le compromis. Nous désignons par  $E_i$  la matrice d'erreur correspondant au juge  $i$ , que nous estimons ensuite pour chaque juge selon l'équation :

$$\hat{E}_i = \hat{\lambda}_i X_i \hat{R}_i - \hat{C} .$$

Nous calculons donc la matrice :

$$\hat{K} := \sum_{i=1}^p u_i \hat{\lambda}_i^2 \hat{E}_i \hat{E}_i^T$$

et corrigeons  $W$  par soustraction de  $\hat{K}$ .

Nous comparons cette méthode aux méthodes Procruste et STATIS à l'aide de simulations. De plus, nous examinons deux autres versions de STATIS : l'une utilisant la moyenne arithmétique à la place de la moyenne pondérée, et l'autre

utilisant les poids asymptotiques calculés dans cette thèse. Nous supposons le même modèle que Meyners *et al.* (2000). Les résultats de chaque méthode sont comparés au compromis connu qui est utilisé au cours de la simulation. Il en ressort que les résultats de la nouvelle méthode sont, le plus souvent, sensiblement meilleurs que les compromis estimés par STATIS et GPA.

Enfin, nous étudions les différences entre les résultats par le biais d'une comparaison graphique. Nous calculons les compromis pour différentes bases de données et représentons les deux premières composantes principales. Nous observons que les résultats des méthodes sont très similaires. Nous concluons donc à l'absence de différences notables en termes d'interprétation de l'analyse, et nous proposons que chaque expérimentateur applique la méthode qui lui semble la plus facile en tenant compte des logiciels qui sont à sa disposition.

# 1 Einleitung

Zur Beschreibung und Bewertung der Eigenschaften von Nahrungsmitteln mit dem Ziel der Qualitätskontrolle oder -verbesserung werden neben instrumentellen Daten wie chemischen oder physikalischen Eigenschaften auch sensorische Profildaten erhoben. Dazu werden – teilweise geschulte – Prüfpersonen aufgefordert, die Produkteigenschaften anhand bestimmter Merkmale wie beispielsweise süß, fruchtig, bitter usw. zu bewerten. Diese Bewertung erfolgt in der Regel durch eine Markierung auf einer unstrukturierten Linienskala, deren Enden die Ausprägungen „gar nicht vorhanden“ bzw. „sehr stark vorhanden“ symbolisieren. Bisweilen werden stattdessen auch Bewertungen auf einer Skala mit ganzzahligen Werten etwa zwischen 1 und 10 verwendet.

Im Gegensatz zu instrumentellen Daten, die nur in der Messung mit einem zufälligen Fehler behaftet sind, liegen bei einer sensorischen Profilprüfung subjektive Daten vor, die zusätzliche Variationen aufweisen und entsprechend ausgewertet werden müssen. So wählen die Prüfpersonen häufig unterschiedliche Skalenabschnitte oder eine unterschiedliche Spannweite auf der Skala, sie beschreiben identische Geschmacksempfindungen mit unterschiedlichen Worten oder ordnen umgekehrt den gleichen Bezeichnungen unterschiedliche Geschmackseindrücke zu. Schließlich haben auch nicht alle Prüfpersonen für jede Variable die gleiche Sensibilität. Das führt dazu, dass beispielsweise die eine Prüfperson schon geringe Unterschiede in einem Merkmal stark wahrnimmt und daher auch entsprechend bewertet, während eine andere

Prüfperson diese Unterschiede überhaupt nicht erkennt und daher dieses Merkmal bei allen Produkten in etwa gleich beurteilt.

Will man diese Variationen in der Auswertung der entsprechenden Daten berücksichtigen, müssen dafür geeignete Verfahren verwendet werden. Zwei von diesen sind die Verallgemeinerte Prokrustes Analyse (GPA, Gower, 1975) und die Methode STATIS („Structuration des Tableaux A Trois Indices de la Statistique“, Lavit *et al.*, 1994). Aus den Beobachtungen der Prüfpersonen wird jeweils ein Konsens geschätzt, der die wahre Struktur der Unterschiede zwischen den Produkten möglichst gut widerspiegeln soll. Anhand von Simulationen und ersten theoretischen Überlegungen werden die beiden Verfahren in einem Artikel von Meyners, Kunert und Qannari (2000) miteinander verglichen. Dabei stellt sich die GPA als die überlegene Methode heraus, und zwar in dem Sinne, dass der geschätzte Konsens dem theoretisch zugrunde liegenden möglichst ähnlich ist.

In dieser Arbeit werden nun die statistischen Eigenschaften von STATIS näher untersucht. Darüber hinaus wird eine Verbesserung dieser Methode vorgeschlagen, die häufig auch der GPA überlegen ist. Das zur Untersuchung der theoretischen Eigenschaften erforderliche Modell wird in Kapitel 2 motiviert und vorgestellt. Dieses Modell beschreibt die Beobachtungen einer sensorischen Profilanalyse.

In Kapitel 3 wird dann zunächst die Originalversion STATIS in kurzer Form dargestellt. Die Methode erfordert die Bestimmung eines gewichteten Mittels von Matrizen, die aus den Beobachtungen der einzelnen Prüfpersonen berechnet werden. Dabei gehen wir insbesondere auf die Bestimmung der entsprechenden Gewichte ein. Anschließend wird die Idee der hier vorgeschlagenen theoretischen Alternativversion erläutert. Eine in der Praxis anwendbare Version hiervon wird in Kapitel 5 vorgestellt.

Dazwischen enthält Kapitel 4 die zentralen Ergebnisse der vorliegenden Arbeit. Es werden bestimmte Spezialfälle betrachtet, in denen unter zusätzlichen

Bedingungen die theoretischen Eigenschaften von STATIS sowie der theoretischen Alternativversion untersucht werden.

Nach einer kurzen Einführung in die Problematik in Kapitel 4.1 wird zunächst in 4.2 die Konvergenz der Gewichte zur Bestimmung des gewichteten Mittels bei STATIS für den Fall betrachtet, dass die Anzahl der zu bewertenden Produkte gegen unendlich konvergiert. Der Spezialfall davon, in dem alle Prüfpersonen die gleiche Fehlervarianz aufweisen, wird danach in Kapitel 4.3 untersucht.

Der euklidische Abstand zwischen dem wahren und dem berechneten Konsens wird in 4.4 bestimmt.

In Kapitel 4.5 betrachten wir die Ergebnisse der Methoden, wenn die Anzahl der Produkte gegen unendlich strebt, die Gewichte der Prüfpersonen aber konstant bleiben.

Unter Verwendung des arithmetischen statt eines gewichteten Mittels zeigen wir in 4.6, dass der betrachtete euklidische Abstand bei Verwendung der neu vorgeschlagenen Methode asymptotisch gegen Null konvergiert, falls die Anzahl der Prüfpersonen immer größer wird. Dieses Ergebnis gilt hingegen für die Originalversion nicht.

In Kapitel 4.7 nehmen wir erneut an, dass die Gewichte der Prüfpersonen fest sind. In diesem Fall können wir den Erwartungswert des euklidischen Abstands zwischen dem wahren und dem theoretischen Konsens bestimmen.

Abschließend untersuchen wir die Konvergenz des STATIS-Konsens sowohl für den Fall, in dem unter zusätzlichen Annahmen die Anzahl der Prüfpersonen gegen unendlich konvergiert (4.8), als auch für den Fall, dass die Variabilität der zufälligen Fehler immer kleiner wird (4.9).

Da die theoretischen Untersuchungen nur unter zusätzlichen Bedingungen erfolgen, werden in Kapitel 6 die Ergebnisse einer Simulationsstudie zum Vergleich zwischen den unterschiedlichen Verfahren beschrieben, nämlich zwischen

- STATIS in der Originalversion,
- STATIS in der hier vorgestellten korrigierten Version,
- STATIS unter Verwendung des arithmetischen Mittels und
- STATIS unter Verwendung der asymptotischen Gewichte aus Kapitel 4.2.

In Kapitel 7 erfolgt ein entsprechender Vergleich der hier vorgeschlagenen korrigierten Version von STATIS mit der GPA.

In Kapitel 8 werden schließlich die Verfahren grafisch miteinander verglichen. Dabei gehen wir auf die Frage ein, inwiefern sich die Ergebnisse der verschiedenen Verfahren hinsichtlich der Interpretation der Ergebnisse unterscheiden.

## 2 Modellannahmen

Wir betrachten das folgende Modell für die Beschreibung der Daten einer sensorischen Profilprüfung. Sei zunächst  $n$  die Anzahl der Produkte und  $m$  die Anzahl der Merkmale, anhand derer die Produkte bewertet werden, sowie  $C \sim (n, m)$  die Matrix der fehlerfreien Beschreibung der Produkte anhand dieser Merkmale. Diese Matrix bezeichnen wir als (theoretischen) Konsens. Ohne Beschränkung der Allgemeinheit sei dabei die Spaltensumme von  $C$  jeweils gleich Null, d. h. alle Variablen sind so normiert, dass der Mittelwert über alle Produkte genau im Ursprung liegt.

Weiter sei nun  $p$  die Anzahl der Prüfpersonen im Panel und für  $i = 1, \dots, p$  enthalte die Matrix  $X_i \sim (n, m)$  die Bewertungen der  $i$ -ten Prüfperson. Dann sei  $X_i$  darstellbar in der Form

$$X_i = \lambda_i(C + E_i)R_i + 1_n t_i^T. \quad (1)$$

Dabei ist  $\lambda_i$  ein positiver Skalar,  $R_i$  eine Rotationsmatrix, d. h. eine Matrix, die

$$R_i^T R_i = R_i R_i^T = I_m \quad (2)$$

erfüllt, sowie  $t_i$  ein Vektor der Länge  $m$ . Weiterhin enthalte  $E_i$  die zufälligen Fehler, die bei der Bewertung durch die  $i$ -te Prüfperson auftreten. Diese Fehler seien unabhängig zwischen den Prüfpersonen und zwischen den Produkten. Von der Verteilung sei nur bekannt, dass sie Erwartungswert Null habe und dass weiter gilt

$$\mathbf{E}(E_i^T E_i) = \Sigma_i$$

für alle  $i \in \{1, \dots, p\}$ . Für eine ausführlichere Erläuterung dieses Modells sei auf den Artikel von Meyners *et al.* (2000) verwiesen.

Zu erwähnen ist in diesem Zusammenhang, dass die Datenmatrizen grundsätzlich vor Beginn weitergehender Analysen spaltenweise zentriert werden, was einer Multiplikation von links mit der Matrix

$$\begin{aligned} \omega &= I(n) - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \\ &= \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & -\frac{1}{n} & 1 - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \end{pmatrix} \end{aligned}$$

entspricht. Damit ist aber wegen  $\mathbf{1}_n^T \mathbf{1}_n = n$  offensichtlich

$$\begin{aligned} \omega \mathbf{1}_n t_i^T &= \mathbf{1}_n t_i^T - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \mathbf{1}_n t_i^T \\ &= \mathbf{1}_n t_i^T - \mathbf{1}_n t_i^T \\ &= \mathbf{0}_{n,m} , \end{aligned}$$

so dass der letzte Term im Modell (1) von nun an vernachlässigt werden kann und wir stattdessen das Modell

$$X_i = \lambda_i (C + E_i) R_i \tag{3}$$

betrachten.

## 3 Die Methode STATIS

### 3.1 Originalversion

Die Methode STATIS soll hier nur kurz skizziert werden, Details und Verallgemeinerungen auf beliebige Abstandsmaße findet man etwa bei Lavit *et al.* (1994) und Schlich (1996). Die grundsätzliche Idee dieses Verfahrens besteht darin, dass zunächst die so genannten Assoziationsmatrizen  $W_i$  für alle Prüfpersonen gemäß

$$W_i := X_i X_i^T$$

bestimmt werden. Durch Einsetzen der Modellgleichung (3) wird direkt der wesentliche Vorteil dieser Methode deutlich. Es gilt dann nämlich mit Bedingung (2)

$$W_i = \lambda_i^2 (C + E_i)(C + E_i)^T ,$$

d. h. die Rotationsmatrizen beeinflussen die Werte von  $W_i$  nicht und müssen somit zur Schätzung von  $C$  auch nicht berücksichtigt werden (vgl. im Gegensatz dazu etwa die Verallgemeinerte Prokrustes Analyse (GPA) nach Gower, 1975, und ten Berge, 1977).

Ein weiteres wichtiges Hilfsmittel dieser Methode ist der RV-Koeffizient zwischen zwei Matrizen  $A$  und  $B$ , der durch

$$RV(A, B) := \frac{\text{trace}(AA^T BB^T)}{\sqrt{\text{trace}(AA^T AA^T)} \sqrt{\text{trace}(BB^T BB^T)}} \quad (4)$$

definiert ist. Es ist leicht zu zeigen, dass der RV-Koeffizient nur Werte zwischen Null und Eins annimmt, und weiter gilt offensichtlich allgemein  $RV(A, A) = 1$ . Im hier vorliegenden Fall ergibt sich entsprechend für  $X_i$  und  $X_j$

$$\begin{aligned} RV(i, j) &:= RV(X_i, X_j) \\ &= \frac{\text{trace}(W_i W_j)}{\sqrt{\text{trace}(W_i W_i)} \sqrt{\text{trace}(W_j W_j)}} . \end{aligned}$$

Aus den RV-Koeffizienten für alle Paare von Prüfpersonen erstellen wir nun die symmetrische Matrix

$$RV := \left[ RV(i, j) \right]_{i, j=1, \dots, p} .$$

Bezeichne weiterhin  $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix}$  den Eigenvektor zum größten Eigenwert von  $RV$ , der so normiert ist, dass

$$u^T \mathbf{1}_p = 1$$

gilt, d. h. die Summe der Einträge ist 1. Damit bestimmen wir dann

$$W = \sum_{i=1}^p u_i W_i ,$$

d. h. wir berechnen eine gewichtete Summe der Assoziationsmatrizen. Die Gewichte sind dabei so konstruiert, dass Prüfpersonen, die gut mit den übrigen Probanden übereinstimmen, ein großes Gewicht erhalten, wohingegen denjenigen, die nur wenig mit den anderen übereinstimmen, ein kleines Gewicht zugewiesen wird. Aus der Matrix  $W$  bestimmen wir anschließend mit Hilfe der Hauptkomponentenanalyse einen Schätzer  $\hat{C}$  für  $C$ . Falls  $m < n$ , so verwenden wir für  $\hat{C}$  nur die ersten  $m$  Hauptkomponenten von  $W$ , anderenfalls alle  $n$ .

### 3.2 Theoretische Alternativversion

Betrachten wir die Assoziationsmatrizen  $W_i$ , so erhalten wir für deren Erwartungswert unter der Annahme, dass die Fehlermatrizen Erwartungswert Null haben,

$$\begin{aligned}\mathbf{E}(W_i) &= \mathbf{E}\left(\lambda_i^2(CC^T + CE_i^T + E_iC^T + E_iE_i^T)\right) \\ &= \lambda_i^2CC^T + \lambda_i^2\mathbf{E}(E_iE_i^T) .\end{aligned}$$

Daher ist es nahe liegend zu versuchen, den hinteren Term von der zugehörigen Assoziationsmatrix zu subtrahieren. Äquivalent dazu bestimmen wir erst das gewichtete Mittel  $W$  und subtrahieren danach eine entsprechend gewichtete Summe dieser Terme, also gerade

$$K := \sum_{i=1}^p u_i \lambda_i^2 E_i E_i^T . \quad (5)$$

Allerdings sind natürlich sowohl die Skalare  $\lambda_i$  als auch die Fehlermatrizen  $E_i$  für alle  $i \in \{1, \dots, p\}$  unbekannt, so dass wir in der Praxis auch  $K$  nicht kennen. Für die theoretischen Überlegungen im folgenden Kapitel 4 gehen wir allerdings davon aus, dass  $K$  dennoch bekannt ist. In Kapitel 5 schlagen wir dann eine Methode vor, mit der diese Matrix geschätzt werden kann. Um zu untersuchen, inwieweit diese Schätzung noch zu einer Verbesserung des Schätzers für  $C$  führt, werden dann in Kapitel 6 die Ergebnisse einer Simulationsstudie angegeben, bei der insbesondere diese Alternativversion mit der geschätzten Matrix  $\hat{K}$  untersucht wird.



## 4 Statistische Eigenschaften von STATIS

### 4.1 Problematik bei der theoretischen Untersuchung der Eigenschaften

Bei der Bestimmung des STATIS-Konsens werden sowohl die Assoziationsmatrizen als auch die Gewichte aus denselben Beobachtungen berechnet, d. h. insbesondere sind für eine beliebige Prüfperson  $i$  sowohl  $W_i$  als auch  $u_i$  abhängig von den Fehlern in  $E_i$ . Damit sind aber auch die Gewichte nicht unabhängig von den Assoziationsmatrizen, so dass etwa bei der Bestimmung des Erwartungswertes von  $W$  im Allgemeinen

$$\mathbf{E}(u_i W_i) \neq \mathbf{E}(u_i) \mathbf{E}(W_i)$$

gilt. Da aufgrund der Komplexität des Verfahrens die direkte Bestimmung von  $\mathbf{E}(u_i W_i)$  problematisch ist, werden im Folgenden verschiedene Annahmen ergänzt, um dann aus dem beschränkteren Modell theoretische Ergebnisse herzuleiten.

## 4.2 Konvergenz der STATIS-Gewichte für $n \rightarrow \infty$

Wir untersuchen zunächst den Fall, dass die Anzahl der Produkte  $n$  gegen unendlich konvergiert, d. h. dieselben Prüfpersonen bewerten immer mehr Produkte anhand der vorgegebenen Variablen. Gehen wir nun davon aus, dass die Produkte zufällig aus einer unendlichen Grundgesamtheit ausgewählt werden, in der die Varianz der Bewertungen beschränkt ist (was in der Praxis durch die begrenzte Skala auf jeden Fall erfüllt ist), so gilt offensichtlich für die wahren Urteile  $C_n$

$$\frac{1}{n} C_n^T C_n \longrightarrow Z \quad (n \rightarrow \infty), \quad (6)$$

wobei  $Z$  eine zwar unbekannte, aber feste Matrix ist.

$C_n$  bezeichnet hier das  $n$ -te Element der Folge der theoretischen Konsense, wenn die Anzahl der Produkte gegen unendlich konvergiert. Entsprechende Folgen erhalten wir für die Fehler-, Beobachtungs- und Assoziationsmatrizen. In diesem und dem nachfolgenden Kapitel 4.3 verzichten wir jedoch von nun an zugunsten einer übersichtlicheren Notation auf diesen Index, da aus dem Zusammenhang offensichtlich wird, dass es sich um entsprechende Folgen von Matrizen handelt.

Die Verteilung der Fehlermatrizen  $E_i$  ist unabhängig von den Produkten, daher gilt weiter für  $n \rightarrow \infty$

$$\frac{1}{n} E_i^T C \longrightarrow 0_{m,m} \quad (7)$$

und entsprechend

$$\frac{1}{n} C^T E_i \longrightarrow 0_{m,m} \quad (8)$$

sowie für  $i \neq j$  wegen der Unabhängigkeit zwischen den Prüfpersonen

$$\frac{1}{n} E_i^T E_j \longrightarrow 0_{m,m}. \quad (9)$$

Schließlich gilt wegen der Voraussetzungen aus dem Modell (3)

$$\frac{1}{n}E_i^T E_i \longrightarrow \Sigma_i . \quad (10)$$

Wir untersuchen nun den RV-Koeffizienten zwischen den Prüfpersonen  $i$  und  $j$ . Da aus der Definition des Koeffizienten in (4) ersichtlich ist, dass für beliebige Skalare  $d_1$  und  $d_2$

$$RV(d_1 A, d_2 B) = RV(A, B)$$

gilt, können wir dazu das Modell (3) weiter vereinfachen und betrachten es ohne Skalare und Rotationsmatrizen, also in der Form

$$X_i = (C + E_i) . \quad (11)$$

Es gilt

$$\begin{aligned} RV(i, j) &= \frac{\text{trace}(W_i W_j)}{\sqrt{\text{trace}(W_i W_i)} \sqrt{\text{trace}(W_j W_j)}} \\ &= \frac{\frac{1}{n^2} \text{trace}(W_i W_j)}{\sqrt{\frac{1}{n^2} \text{trace}(W_i W_i)} \sqrt{\frac{1}{n^2} \text{trace}(W_j W_j)}} , \end{aligned}$$

und allgemein erhalten wir dafür mit (11) zunächst

$$\begin{aligned} &\frac{1}{n^2} \text{trace}[W_i W_j] \\ &= \frac{1}{n^2} \text{trace} [C^T C C^T C + C^T C E_i^T C + C^T C C^T E_j + C^T C E_i^T E_j] \\ &\quad + \frac{1}{n^2} \text{trace} [E_j^T C C^T C + E_j^T C E_i^T C + E_j^T C C^T E_j + E_j^T C E_i^T E_j] \\ &\quad + \frac{1}{n^2} \text{trace} [C^T E_i C^T C + C^T E_i E_i^T C + C^T E_i C^T E_j + C^T E_i E_i^T E_j] \\ &\quad + \frac{1}{n^2} \text{trace} [E_j^T E_i C^T C + E_j^T E_i E_i^T C + E_j^T E_i C^T E_j + E_j^T E_i E_i^T E_j] . \end{aligned}$$

Ist also  $i \neq j$ , so gilt mit (6)-(10) für  $n \rightarrow \infty$

$$\text{trace} \left( \frac{1}{n^2} W_i W_j \right) \rightarrow \text{trace}(ZZ),$$

wohingegen für  $i = j$

$$\text{trace} \left( \frac{1}{n^2} W_i W_i \right) \rightarrow \text{trace} (ZZ + 2Z\Sigma_i + \Sigma_i\Sigma_i) .$$

Damit erhalten wir für  $i \neq j$  und  $n \rightarrow \infty$

$$\begin{aligned} RV(i, j) &\rightarrow \frac{\text{trace}(ZZ)}{\sqrt{\text{trace} (ZZ + 2Z\Sigma_i + \Sigma_i\Sigma_i)} \sqrt{\text{trace} (ZZ + 2Z\Sigma_j + \Sigma_j\Sigma_j)}} \\ &=: q_{ij} \end{aligned} \tag{12}$$

sowie wegen  $RV(i, i) = 1$  für alle  $i \in \{1, \dots, p\}$  daher für die Matrix  $RV$  insgesamt

$$RV \rightarrow \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & q_{ij} & \\ & & & \ddots \\ q_{ji} & & & & 1 \end{pmatrix} =: Q \quad (n \rightarrow \infty) .$$

Aufgrund der Stetigkeit von Eigenwerten und Eigenvektoren (vgl. Harville, 1997) folgt daraus unmittelbar, dass der Eigenvektor zum größten Eigenwert von  $RV$  für  $n \rightarrow \infty$  gegen den Eigenvektor zum größten Eigenwert von  $Q$  konvergiert, sofern letzterer ein einfacher Eigenwert ist. Insbesondere konvergiert damit der entsprechende Eigenvektor von  $RV$  gegen einen festen Vektor. Also konvergieren die Gewichte, die den Probanden durch STATIS zugewiesen werden, bei fester Anzahl von Prüfpersonen und Merkmalen gegen feste Werte.

Der Vollständigkeit halber werden diese Gewichte nun noch explizit bestimmt. Dazu ist zunächst zu beachten, dass wegen der Symmetrie des RV-Koeffizienten auch  $Q$  symmetrisch ist. Weiter lässt sich diese Matrix in der Form  $Q = B + ss^T$  darstellen, wobei

$$B := \text{diag} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}$$

mit

$$b_i := \frac{2\text{trace}(Z\Sigma_i) + \text{trace}(\Sigma_i\Sigma_i)}{\text{trace}(ZZ) + 2\text{trace}(Z\Sigma_i) + \text{trace}(\Sigma_i\Sigma_i)}$$

und

$$s := \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_p \end{pmatrix}$$

mit

$$s_i := \frac{\sqrt{\text{trace}(ZZ)}}{\sqrt{\text{trace}(ZZ) + 2\text{trace}(Z\Sigma_i) + \text{trace}(\Sigma_i\Sigma_i)}}$$

ist. Schließlich gilt für  $Z \neq 0_{m,m}$  unter Berücksichtigung der Symmetrie dieser Matrix, dass  $\text{trace}(ZZ) > 0$  und damit  $s_i > 0$  ist für alle  $i = 1, \dots, p$ .

Sei nun  $k \in \{1, \dots, p\}$  so, dass  $b_k = \max\{b_1, \dots, b_p\}$ . Dann ist  $b_k$  der größte Eigenwert der Matrix  $B$  mit zugehörigem Eigenvektor  $e_k$ , also dem  $k$ -ten Einheitsvektor. Damit ist

$$\begin{aligned} e_k^T s s^T e_k &= (s^T e_k)^2 \\ &= s_k^2 \\ &> 0, \end{aligned}$$

da ja  $s_i > 0$  und damit insbesondere  $s_i \neq 0$  für alle  $i = 1, \dots, p$ . Daraus folgt für den größten Eigenwert  $\xi$  von  $B + s s^T$ , dass

$$\begin{aligned} \xi &= \max_{\|x\|=1} \left[ x^T (B + s s^T) x \right] \\ &\geq e_k^T B e_k + \underbrace{e_k^T s s^T e_k}_{>0 \text{ n. Vorbem.}} \\ &> e_k^T B e_k \\ &= b_k, \end{aligned}$$

d. h.  $\xi > \max\{b_1, \dots, b_p\}$ . Damit ist  $\xi I_p - B$  positiv definit und somit  $(\xi I_p - B)^{-1}$  wohldefiniert.

Für den zu  $\xi$  gehörenden Eigenvektor  $u \neq 0_p$  von  $B + ss^T$  muss nach Definition

$$(B + ss^T - \xi I_p)u = 0_p$$

gelten (vgl. Horn und Johnson, 1987). Daraus erhalten wir äquivalent

$$\begin{aligned} (\xi I_p - B - ss^T)u &= 0_p \\ \Leftrightarrow (\xi I_p - B)u &= ss^T u \\ &= s \underbrace{(s^T u)}_{\in \mathbb{R}} \\ \Leftrightarrow u &= (\xi I_p - B)^{-1} s (s^T u) . \end{aligned}$$

Für  $Z \neq 0_{m,m}$  ist  $Q$  eine positive Matrix, d. h. alle Einträge  $q_{ij}$  sind positiv. Nach Minc (1988) ist diese Matrix damit irreduzibel und der erste Eigenvektor  $u$  kann positiv gewählt werden. Folglich sind alle Gewichte  $u_i$  zur Berechnung des STATIS-Konsens  $W$  positiv. Des Weiteren ist nach Definition des Eigenvektors  $u \neq 0_p$  und es gilt wegen  $s_i > 0$  für alle  $i \in \{1, \dots, p\}$ , dass

$$s^T u > 0$$

ist. Normieren wir  $u$  so, dass  $1_p^T u = 1$  gilt, d. h. dass die Summe der Gewichte 1 ergibt, so erhalten wir den normierten Eigenvektor zum Eigenwert  $\xi$  durch

$$\begin{aligned} u^* &= (\xi I_p - B)^{-1} s s^T u \frac{1}{1_p^T (\xi I_p - B)^{-1} s s^T u} \\ &= (\xi I_p - B)^{-1} s \frac{1}{1_p^T (\xi I_p - B)^{-1} s} . \end{aligned}$$

Wir betrachten nun den Fall, dass die Varianz einer Prüfperson  $i$  immer größer wird und damit insbesondere  $trace(\Sigma_i \Sigma_i)$  gegen unendlich konvergiert, während die Varianz bei den übrigen Prüfpersonen unverändert bleibt. Durch Einsetzen der entsprechenden Werte für  $B$  und  $s$  sehen wir unmittelbar, dass das Gewicht der zugehörigen Prüfperson gegen Null konvergiert. Also wird dieser (zunehmend schlechten) Prüfperson asymptotisch das Gewicht Null zugewiesen.

### 4.3 Betrachtung von Spezialfällen bei der Konvergenz der Gewichte

Um das Verhalten der asymptotischen Gewichte zu untersuchen, betrachten wir nun den Spezialfall, bei dem die Korrelationsmatrix der Fehlerterme für alle Prüfpersonen identisch ist, die Varianzen allerdings unterschiedlich groß sein können. In unserem Modell bedeutet dies, dass für alle Prüfpersonen  $i$  für  $i = 1, \dots, p$

$$\Sigma_i = \sigma_i^2 \Sigma$$

gilt. Die Matrix  $\Sigma$ , die ohne Beschränkung der Allgemeinheit  $\text{trace}(\Sigma\Sigma) = 1$  erfülle, ist dabei für alle Prüfpersonen identisch. Des Weiteren nehmen wir an, dass auch  $\text{trace}(ZZ) = 1$  ist.

Verwenden wir dieses nun zusammen mit den Ergebnissen des vorherigen Kapitels, so erhalten wir

$$B = \text{diag} \left( \left[ \frac{2\sigma_i^2 \text{trace}(Z\Sigma) + \sigma_i^4}{1 + 2\sigma_i^2 \text{trace}(Z\Sigma) + \sigma_i^4} \right]_{i=1, \dots, p} \right)$$

und

$$s = \left[ \frac{1}{\sqrt{1 + 2\sigma_i^2 \text{trace}(Z\Sigma) + \sigma_i^4}} \right]_{i=1, \dots, p},$$

und somit ist abgesehen von der Normierung auf Summe 1

$$\begin{aligned} u_i &= \left( \xi - \frac{2\sigma_i^2 \text{trace}(Z\Sigma) + \sigma_i^4}{1 + 2\sigma_i^2 \text{trace}(Z\Sigma) + \sigma_i^4} \right)^{-1} \frac{1}{\sqrt{1 + 2\sigma_i^2 \text{trace}(Z\Sigma) + \sigma_i^4}} \\ &= \frac{\sqrt{1 + 2\sigma_i^2 \text{trace}(Z\Sigma) + \sigma_i^4}}{\xi + (\xi - 1)(2\sigma_i^2 \text{trace}(Z\Sigma) + \sigma_i^4)}. \end{aligned}$$

Damit können verschiedene Spezialfälle betrachtet werden, indem wir Annahmen über die  $\sigma_i^2$  treffen. Zunächst betrachten wir dazu den Fall, in dem alle Prüfpersonen die gleiche Fehlervarianz aufweisen, d. h. den Fall  $\sigma_i^2 = \sigma^2$  für alle  $i = 1, \dots, p$ .

Dann sind offensichtlich alle Einträge von  $u$  identisch und  $u$  ist ein Vielfaches des Einsenvektors. Durch die Normierung dieses Vektors auf Summe 1 erhalten wir dann

$$u^* = \left[ \frac{1}{p} \right]_{i=1, \dots, p} .$$

Also erhält für den Fall, dass alle Prüfpersonen die gleiche Fehlervarianz aufweisen, asymptotisch jede Prüfperson das Gewicht  $\frac{1}{p}$ . Im Grenzfall wird also das arithmetische Mittel berechnet.

Ein weiterer interessierender Fall ergibt sich, wenn sich alle Prüfpersonen vollständig einig sind, abgesehen von einigen wenigen schlechten Prüfpersonen. Die Prüfpersonen können sich natürlich nur dann wirklich einig sein, wenn sie keine Fehler machen und somit die Fehlervarianz  $\sigma^2 = 0$  ist. Die schlechten Prüfpersonen weisen dagegen eine von Null verschiedene Fehlervarianz auf. Damit erhalten wir – abgesehen von der Normierung auf Gesamtsumme 1 – als  $i$ -ten Eintrag des Gewichtsvektors  $\frac{1}{\xi}$  für die Prüfpersonen mit Fehlervarianz 0. Aufgrund der obigen Überlegungen folgt, dass Prüfpersonen mit größerer Fehlervarianz asymptotisch ein Gewicht erhalten, das kleiner ist als das der übrigen Prüfpersonen, und wegen der Normierung des Gewichtsvektors werden diese Personen asymptotisch insbesondere ein Gewicht erhalten, das kleiner als  $\frac{1}{p}$  ist.

Konkretere Konstellationen werden im Rahmen der Simulationsstudie in Kapitel 6 untersucht, wobei wir für  $Z$  eine Matrix verwenden, die aus einer in der Literatur angegebenen beobachteten Matrix bestimmt wird. Darüber hinaus werden zusätzlich verschiedene Annahmen über die Struktur von  $\Sigma$  getroffen.

#### 4.4 Untersuchung des Abstandes des STATIS-Konsens vom wahren Konsens

Wir wollen untersuchen, wie stark sich der geschätzte vom wahren Konsens unterscheidet. Anstatt jedoch den eigentlichen STATIS-Konsens  $\hat{C}$  und den wahren Konsens  $C$  zu vergleichen, betrachten wir hier die quadrierte euklidische Norm der Differenz von  $W$  und  $CC^T$ . Wir erhalten dabei

$$\begin{aligned} \|\hat{C}\hat{C}^T - CC^T\|^2 &= \text{trace} \left[ \left( \hat{C}\hat{C}^T - CC^T \right) \left( \hat{C}\hat{C}^T - CC^T \right)^T \right] \\ &= \text{trace} \left[ \underbrace{\hat{C}\hat{C}^T\hat{C}\hat{C}^T}_{(i)} - \underbrace{\hat{C}\hat{C}^TCC^T}_{(ii)} - \underbrace{CC^T\hat{C}\hat{C}^T}_{(iii)} + \underbrace{CC^TCC^T}_{(iv)} \right]. \end{aligned}$$

Um die Rechnung übersichtlicher zu gestalten, betrachten wir nun nacheinander die Terme (i) - (iv). Wir erhalten zunächst für (i)

$$\begin{aligned} &\text{trace} \left[ \hat{C}\hat{C}^T\hat{C}\hat{C}^T \right] \\ &= \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (CC^T CC^T + CC^T CE_j^T + CC^T E_j C^T + CC^T E_j E_j^T) \right] \\ &+ \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (CE_i^T CC^T + CE_i^T CE_j^T + CE_i^T E_j C^T + CE_i^T E_j E_j^T) \right] \\ &+ \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T CC^T + E_i C^T CE_j^T + E_i C^T E_j C^T + E_i C^T E_j E_j^T) \right] \\ &+ \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i E_i^T CC^T + E_i E_i^T CE_j^T + E_i E_i^T E_j C^T + E_i E_i^T E_j E_j^T) \right]. \end{aligned}$$

Entsprechend lässt sich (ii) darstellen durch

$$\begin{aligned} &\text{trace} \left[ \hat{C}\hat{C}^T CC^T \right] \\ &= \text{trace} \left[ \sum_{i=1}^p u_i \lambda_i^2 (CC^T CC^T + CE_i^T CC^T + E_i C^T CC^T + E_i E_i^T CC^T) \right] \end{aligned}$$

und (iii) durch

$$\begin{aligned} & \text{trace} \left[ CC^T \hat{C} \hat{C}^T \right] \\ &= \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 (CC^T CC^T + CC^T CE_j^T + CC^T E_j C^T + CC^T E_j E_j^T) \right]. \end{aligned}$$

Der Term (iv) bleibt unverändert. Daraus resultiert dann insgesamt

$$\begin{aligned} & \|\hat{C} \hat{C}^T - CC^T\|^2 \\ &= \text{trace} \left[ CC^T CC^T \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \\ &+ \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 CC^T (CE_j^T + E_j C^T + E_j E_j^T) \left( \sum_{i=1}^p u_i \lambda_i^2 - 1 \right) \right] \\ &+ \text{trace} \left[ \sum_{i=1}^p u_i \lambda_i^2 (CE_i^T + E_i C^T + E_i E_i^T) CC^T \left( \sum_{j=1}^p u_j \lambda_j^2 - 1 \right) \right] \\ &+ \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (CE_i^T CE_j^T + CE_i^T E_j C^T + CE_i^T E_j E_j^T) \right] \\ &+ \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T CE_j^T + E_i C^T E_j C^T + E_i C^T E_j E_j^T) \right] \\ &+ \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i E_i^T CE_j^T + E_i E_i^T E_j C^T + E_i E_i^T E_j E_j^T) \right]. \quad (13) \end{aligned}$$

Für den Fall, dass wir von  $W$  die Summe  $K$  aus Gleichung (5) in Kapitel 3.2 subtrahieren, erhalten wir

$$\begin{aligned} \|\hat{C} \hat{C}^T - K - CC^T\|^2 &= \text{trace} \left[ \left( \hat{C} \hat{C}^T - K - CC^T \right) \left( \hat{C} \hat{C}^T - K - CC^T \right)^T \right] \\ &= \text{trace} \left[ \hat{C} \hat{C}^T \hat{C} \hat{C}^T - \hat{C} \hat{C}^T CC^T - CC^T \hat{C} \hat{C}^T + CC^T CC^T \right] \\ &+ \text{trace} \left[ \underbrace{-K \hat{C} \hat{C}^T}_{(v)} - \underbrace{\hat{C} \hat{C}^T K}_{(vi)} + \underbrace{K CC^T}_{(vii)} + \underbrace{CC^T K}_{(viii)} + \underbrace{K K^T}_{(ix)} \right]. \end{aligned}$$

Die ersten vier Terme entsprechen den eben betrachteten, so dass wir nur noch die Terme (v) - (ix) untersuchen müssen. Dabei berechnen wir zunächst für (v)

$$\begin{aligned} & \text{trace} \left[ K \hat{C} \hat{C}^T \right] \\ &= \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i E_i^T C C^T + E_i E_i^T C E_j^T + E_i E_i^T E_j C^T + E_i E_i^T E_j E_j^T) \right] \end{aligned}$$

und analog für (vi)

$$\begin{aligned} & \text{trace} \left[ \hat{C} \hat{C}^T K \right] \\ &= \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C C^T E_j E_j^T + C E_i^T E_j E_j^T + E_i C^T E_j E_j^T + E_i E_i^T E_j E_j^T) \right]. \end{aligned}$$

Weiter ist in (vii)

$$\text{trace} [K C C^T] = \text{trace} \left[ \sum_{i=1}^p u_i \lambda_i^2 E_i E_i^T C C^T \right]$$

und entsprechend in (viii)

$$\text{trace} [C C^T K] = \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 C C^T E_j E_j^T \right]$$

sowie schließlich in (ix)

$$\text{trace} [K K^T] = \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 E_i E_i^T E_j E_j^T \right].$$

Fassen wir die Resultate für (i)-(ix) zusammen, so erhalten wir

$$\begin{aligned}
& \|\hat{C}\hat{C}^T - K - CC^T\|^2 \\
&= \text{trace} \left[ CC^T CC^T \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \\
&\quad + \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 CC^T (CE_j^T + E_j C^T) \left( \sum_{i=1}^p u_i \lambda_i^2 - 1 \right) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p u_i \lambda_i^2 (CE_i^T + E_i C^T) CC^T \left( \sum_{j=1}^p u_j \lambda_j^2 - 1 \right) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (CE_i^T CE_j^T + CE_i^T E_j C^T) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T CE_j^T + E_i C^T E_j C^T) \right]. \tag{14}
\end{aligned}$$

Vergleichen wir das Ergebnis mit Gleichung (13), so stellen wir fest, dass insgesamt sieben Terme durch die Korrektur wegfallen. Diese Terme hängen offensichtlich vom Zufall ab, so dass die Varianz des gesamten Ausdrucks in (13) größer ist als in (14). Obwohl die Terme nicht notwendigerweise positiv sind gehen wir daher dennoch davon aus, dass der Abstand von  $W$  vom entsprechenden Matrixprodukt des wahren Konsens durch die Korrektur im Allgemeinen verringert wird und der zugehörige Schätzer somit bessere Ergebnisse liefert.

Betrachten wir noch den Spezialfall, dass es keine Unterschiede zwischen den Produkten gibt, so hat  $C$  spaltenweise identische Einträge und es ist wegen der Bedingung, dass die Spaltensummen Null ergeben,  $C = 0_{n,m}$ . In diesem Fall sehen wir unmittelbar, dass für die unkorrigierte Version gilt

$$\begin{aligned}
\|\hat{C}\hat{C}^T - CC^T\|^2 &= \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 E_i E_i^T E_j E_j^T \right] \\
&\neq 0,
\end{aligned}$$

während wir mit Korrektur und unter der Voraussetzung, dass wir  $K$  kennen,

$$\|\hat{C}\hat{C}^T - K - CC^T\|^2 = 0$$

erhalten. In diesem Fall führt die Korrektur also offensichtlich zu einer Verbesserung des Ergebnisses.

Die Berücksichtigung der Skalare verkompliziert in diesem Zusammenhang die Rechnung stark. In der Praxis interessiert uns allerdings die Matrix  $CC^T$  nur bis auf ein Vielfaches genau, d. h. eine skalare Multiplikation ist für die Interpretation ohne Bedeutung. Da die Varianz der Fehler ebenfalls nicht genau festgelegt ist, macht es offensichtlich keinen Unterschied, ob wir das Modell (3) in der bisherigen Form betrachten oder stattdessen das Modell

$$\begin{aligned} X_i &= \lambda_i \sqrt{\frac{p}{\sum_{i=1}^p \lambda_i^2}} \left( C \sqrt{\frac{\sum_{i=1}^p \lambda_i^2}{p}} + E_i \sqrt{\frac{\sum_{i=1}^p \lambda_i^2}{p}} \right) R_i \\ &=: \lambda_i^* (C^* + E_i^*) R_i . \end{aligned} \quad (15)$$

In diesem Modell gilt dann  $\sum_{i=1}^p (\lambda_i^*)^2 = p$ . Zur Vereinfachung der Notation nehmen wir an, dass eine solche Normierung implizit im Modell enthalten ist und wir im Folgenden davon ausgehen können, dass  $\sum_{i=1}^p \lambda_i^2 = p$  gilt.

Betrachten wir damit den Fall des arithmetischen Mittels – bei dem also  $u_i = \frac{1}{p}$  für alle  $i = 1, \dots, p$  gilt – so ist

$$\sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 = \left( \frac{1}{p} \sum_{i=1}^p \lambda_i^2 - 1 \right)^2$$

sowie

$$\sum_{i=1}^p u_i \lambda_i^2 - 1 = \frac{1}{p} \sum_{i=1}^p \lambda_i^2 - 1 .$$

Im Modell (15) sind die beiden Ausdrücke jeweils Null und wir erhalten

$$\begin{aligned}
& \|\hat{C}\hat{C}^T - CC^T\|^2 \\
&= \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (CE_i^T CE_j^T + CE_i^T E_j C^T + CE_i^T E_j E_j^T) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T CE_j^T + E_i C^T E_j C^T + E_i C^T E_j E_j^T) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i E_i^T CE_j^T + E_i E_i^T E_j C^T + E_i E_i^T E_j E_j^T) \right].
\end{aligned}$$

Andererseits ist mit Korrektur der Abstand gegeben durch

$$\begin{aligned}
& \|\hat{C}\hat{C}^T - K - CC^T\|^2 \\
&= \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (CE_i^T CE_j^T + CE_i^T E_j C^T) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T CE_j^T + E_i C^T E_j C^T) \right].
\end{aligned}$$

Auch hier erkennen wir, dass im zweiten Fall fünf Terme wegfallen, so dass wir von einer Verbesserung der Schätzung durch die vorgeschlagene Korrektur ausgehen können.

## 4.5 STATIS-Version mit festen Gewichten für $n \rightarrow \infty$

Betrachten wir nun den Spezialfall, dass die Gewichte für die einzelnen Prüferpersonen fest gewählt sind und nicht mehr von den Beobachtungen direkt abhängen, d. h. wir befinden uns bezüglich der Gewichte beispielsweise im Grenzfall aus dem Kapitel 4.2. Somit haben wir also

$$W = \sum_{i=1}^p u_i W_i ,$$

wobei die  $u_i$  zwar unbekannt, aber fest sind mit  $\sum_{i=1}^p u_i = 1$ .

Wie in 4.2 verzichten wir für Folgen von Matrizen wiederum auf den Index  $n$  und betrachten nun für  $n \rightarrow \infty$  den mit  $\frac{1}{n^2}$  normierten quadrierten euklidischen Abstand von  $W = \hat{C}\hat{C}^T$  und  $CC^T$ . Aus Gleichung (13) erhalten wir dann unmittelbar

$$\begin{aligned} & \left\| \frac{1}{n} \hat{C}\hat{C}^T - \frac{1}{n} CC^T \right\|^2 \\ &= \frac{1}{n^2} \text{trace} \left[ CC^T CC^T \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \\ &+ \frac{1}{n^2} \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 CC^T (CE_j^T + E_j C^T + E_j E_j^T) \left( \sum_{i=1}^p u_i \lambda_i^2 - 1 \right) \right] \\ &+ \frac{1}{n^2} \text{trace} \left[ \sum_{i=1}^p u_i \lambda_i^2 (CE_i^T + E_i C^T + E_i E_i^T) CC^T \left( \sum_{j=1}^p u_j \lambda_j^2 - 1 \right) \right] \\ &+ \frac{1}{n^2} \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (CE_i^T CE_j^T + CE_i^T E_j C^T + CE_i^T E_j E_j^T) \right] \\ &+ \frac{1}{n^2} \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T CE_j^T + E_i C^T E_j C^T + E_i C^T E_j E_j^T) \right] \\ &+ \frac{1}{n^2} \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i E_i^T CE_j^T + E_i E_i^T E_j C^T + E_i E_i^T E_j E_j^T) \right] , \end{aligned}$$

was sich durch zyklische Vertauschung unter der Spur schreiben lässt als

$$\begin{aligned}
& \frac{1}{n^2} \text{trace} \left[ C^T C C^T C \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \\
& + \frac{2}{n^2} \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 (C^T C E_j^T C + C^T E_j C^T C + C^T E_j E_j^T C) \left( \sum_{i=1}^p u_i \lambda_i^2 - 1 \right) \right] \\
& + \frac{1}{n^2} \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i^T C E_j^T C + E_i^T E_j C^T C + E_i^T E_j E_j^T C) \right] \\
& + \frac{1}{n^2} \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C^T C E_j^T E_i + C^T E_j C^T E_i + C^T E_j E_j^T E_i) \right] \\
& + \frac{1}{n^2} \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i^T C E_j^T E_i + E_i^T E_j C^T E_i + E_i^T E_j E_j^T E_i) \right].
\end{aligned}$$

Berücksichtigen wir weiterhin, dass  $\frac{1}{n} E_i^T C \rightarrow 0_{m,m}$ ,  $\frac{1}{n} C^T E_i \rightarrow 0_{m,m}$ ,  $\frac{1}{n} E_j^T E_i \rightarrow 0_{m,m}$  für  $i \neq j$  sowie  $\frac{1}{n} E_i^T E_i \rightarrow \Sigma_i$  und insbesondere nach Voraussetzung  $\frac{1}{n} C^T C \rightarrow Z$  (jeweils für  $n \rightarrow \infty$ ), so erhalten wir

$$\begin{aligned}
& \left\| \frac{1}{n} \hat{C} \hat{C}^T - \frac{1}{n} C C^T \right\|^2 \\
& \longrightarrow \text{trace} \left[ Z Z \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \\
& + \text{trace} \left[ \sum_{i=1}^p u_i^2 \lambda_i^4 (2 \Sigma_i Z + \Sigma_i^2) \right] \quad (n \rightarrow \infty).
\end{aligned}$$

Subtrahieren wir nun wiederum  $K$  von  $W$ , so erhalten wir stattdessen analog

$$\begin{aligned}
& \left\| \frac{1}{n} (\hat{C}\hat{C}^T - K) - \frac{1}{n} CC^T \right\|^2 \\
&= \text{trace} \left[ C^T CC^T C \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \\
&\quad + \frac{2}{n^2} \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 (C^T C E_j^T C + C^T E_j C^T C) \left( \sum_{i=1}^p u_i \lambda_i^2 - 1 \right) \right] \\
&\quad + \frac{1}{n^2} \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i^T C E_j^T C + E_i^T E_j C^T C) \right] \\
&\quad + \frac{1}{n^2} \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C^T C E_j^T E_i + C^T E_j C^T E_i) \right].
\end{aligned}$$

Damit folgt für  $n \rightarrow \infty$

$$\begin{aligned}
& \left\| \frac{1}{n} (\hat{C}\hat{C}^T - K) - \frac{1}{n} CC^T \right\|^2 \\
&\longrightarrow \text{trace} \left[ ZZ \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \\
&\quad + 2 \cdot \text{trace} \left[ \sum_{i=1}^p u_i^2 \lambda_i^4 \Sigma_i Z \right].
\end{aligned}$$

Zwar ist auch dieser Ausdruck im Allgemeinen ungleich 0, aber wegen  $u_i > 0$ ,  $\lambda_i^4 \geq 0$  sowie  $\text{trace}[\Sigma_i^2] > 0$  für alle  $i = 1, \dots, p$  gilt

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n} (\hat{C}\hat{C}^T - K) - \frac{1}{n} CC^T \right\|^2 < \lim_{n \rightarrow \infty} \left\| \frac{1}{n} \hat{C}\hat{C}^T - \frac{1}{n} CC^T \right\|^2.$$

Also wird unter den getroffenen Annahmen durch die vorgeschlagene Korrektur die hier betrachtete asymptotische Abweichung zwischen  $W$  und dem zu schätzenden Produkt  $CC^T$  verringert.

## 4.6 Konvergenz des Abstandes unter weiteren Annahmen

Wir betrachten in einer weiteren Überlegung das arithmetische Mittel statt eines gewichteten zur Berechnung des STATIS-Konsens und nehmen zudem an, dass die Prüfpersonen aus einer unendlich großen Grundgesamtheit stammen und für das Experiment zufällig aus dieser Grundgesamtheit ausgewählt werden. Schließlich seien auch noch alle Skalare gleich groß, d. h. wir betrachten hier die ‘pre-scaled’ Variante, bei der die Matrizen  $X_i$  auf einheitlichen euklidischen Abstand vom Ursprung normiert werden. Wegen der Nebenbedingung  $\sum_{i=1}^p \lambda_i^2 = p$  aus Modell (15) gilt damit  $\lambda_i = 1$  für alle  $i = 1, \dots, p$ . Dann gilt für  $p \rightarrow \infty$

$$\begin{aligned} \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C E_i^T C E_j^T) &= \frac{1}{p^2} \sum_{i=1}^p \sum_{j=1}^p C E_i^T C E_j^T \\ &= C \underbrace{\left( \frac{1}{p} \sum_{i=1}^p E_i^T \right)}_{\rightarrow 0_{m,n}} C \underbrace{\left( \frac{1}{p} \sum_{j=1}^p E_j^T \right)}_{\rightarrow 0_{m,n}} \\ &\longrightarrow 0_{n,n} \quad (p \rightarrow \infty) \end{aligned}$$

und analog

$$\begin{aligned} \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C E_i^T E_j C^T) &\longrightarrow 0_{n,n} , \\ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T E_j C^T) &\longrightarrow 0_{n,n} \end{aligned}$$

sowie

$$\sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T C E_j^T) \longrightarrow 0_{n,n} .$$

Des Weiteren gilt wegen der Unabhängigkeit der Fehler zwischen den Produkten

$$\sum_{i=1}^p u_i \lambda_i^2 E_i E_i^T = \frac{1}{p} \sum_{i=1}^p (E_i E_i^T)$$

$$\longrightarrow \begin{pmatrix} \delta_1^2 & & & & \\ & \delta_2^2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ 0 & & & & \delta_n^2 \end{pmatrix},$$

wobei  $\delta_l^2$  die durchschnittliche Fehlervarianz der Prüfpersonen in der Grundgesamtheit bei Produkt  $l$  bezeichne.

Damit folgt dann direkt wie oben, dass für  $p \rightarrow \infty$  gilt

$$\sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C E_i^T E_j E_j^T) \longrightarrow 0_{n,n},$$

$$\sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T E_j E_j^T) \longrightarrow 0_{n,n},$$

$$\sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i E_i^T C E_j^T) \longrightarrow 0_{n,n}$$

und auch

$$\sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i E_i^T E_j C^T) \longrightarrow 0_{n,n}.$$

Schließlich bestimmen wir

$$\sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i E_i^T E_j E_j^T) = \frac{1}{p} \sum_{i=1}^p E_i E_i^T \frac{1}{p} \sum_{j=1}^p E_j E_j^T$$

$$\longrightarrow \begin{pmatrix} \delta_1^4 & & & & \\ & \delta_2^4 & & & \\ & & \ddots & & \\ & & & \ddots & \\ 0 & & & & \delta_n^4 \end{pmatrix}$$

für  $p \rightarrow \infty$ .

Damit erhalten wir unter Verwendung der Resultate aus Kapitel 4.4 für die unkorrigierte Version

$$\|\hat{C}\hat{C}^T - CC^T\|^2 \longrightarrow \sum_{l=1}^n \delta_l^4 > 0,$$

falls  $p \rightarrow \infty$ , wohingegen für den entsprechenden Abstand bei der korrigierten Version

$$\|\hat{C}\hat{C}^T - K - CC^T\|^2 \longrightarrow 0$$

gilt.

Das bedeutet, dass bei einer Erhöhung der Anzahl der Prüfpersonen die Abweichung des STATIS-Konsens vom theoretischen Konsens gegen einen Term konvergiert, der unmittelbar von der Varianz der Fehler abhängt und in jedem nicht deterministischen Fall (der Fall, in dem die Fehler eine Varianz von Null haben und damit wegen der Voraussetzung  $\mathbf{E}(E_i) = 0_{n,m}$  fast sicher gleich Null sind) echt größer ist als Null. Durch die vorgeschlagene Korrektur des berechneten Konsens  $W$  erreichen wir hingegen, dass die Abweichung asymptotisch gegen Null konvergiert. Es sei hier darauf hingewiesen, dass diese Aussage insbesondere auch im Spezialfall  $C = 0_{n,m}$  gilt, d. h. dass dieser Fehler von STATIS selbst für den Fall, dass keine Unterschiede zwischen den Produkten bestehen, nicht durch eine große Anzahl von Prüfpersonen vernachlässigbar wird. Dieses ist eine Verallgemeinerung des Ergebnisses von Meyners *et al.* (2000), die nur den Erwartungswert der Assoziationsmatrix einer einzelnen Prüfperson betrachten.

## 4.7 Erwartungswert des STATIS-Konsens bei festen Gewichten

In diesem Abschnitt nehmen wir an, dass die Gewichte  $u_i$  der einzelnen Prüfpersonen für gegebenes  $p$  fest sind und damit keine Zufallsvariable darstellen, die von den Beobachtungen abhängen. In der einfachsten Form erhalten alle Prüfpersonen beispielsweise das Gewicht  $\frac{1}{p}$ . Wir nutzen wiederum die Ergebnisse aus den Gleichungen (13) und (14) in Kapitel 4.4 und erhalten allgemein als erwartete quadratische Abweichung von  $W$  und  $CC^T$

$$\begin{aligned}
& \mathbf{E} \left( \|\hat{C}\hat{C}^T - CC^T\|^2 \right) \\
&= \mathbf{E} \left( \text{trace} \left[ CC^T CC^T \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \right) \\
&\quad + \mathbf{E} \left( \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 CC^T (CE_j^T + E_j C^T + E_j E_j^T) \left( \sum_{i=1}^p u_i \lambda_i^2 - 1 \right) \right] \right) \\
&\quad + \mathbf{E} \left( \text{trace} \left[ \sum_{i=1}^p u_i \lambda_i^2 (CE_i^T + E_i C^T + E_i E_i^T) CC^T \left( \sum_{j=1}^p u_j \lambda_j^2 - 1 \right) \right] \right) \\
&\quad + \mathbf{E} \left( \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (CE_i^T CE_j^T + CE_i^T E_j C^T + CE_i^T E_j E_j^T) \right] \right) \\
&\quad + \mathbf{E} \left( \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T CE_j^T + E_i C^T E_j C^T + E_i C^T E_j E_j^T) \right] \right) \\
&\quad + \mathbf{E} \left( \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i E_i^T CE_j^T + E_i E_i^T E_j C^T + E_i E_i^T E_j E_j^T) \right] \right)
\end{aligned}$$

$$\begin{aligned}
&= \text{trace} \left[ CC^T CC^T \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \\
&+ \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 CC^T \mathbf{E}(E_j E_j^T) \left( \sum_{i=1}^p u_i \lambda_i^2 - 1 \right) \right] \\
&+ \text{trace} \left[ \sum_{i=1}^p u_i \lambda_i^2 \mathbf{E}(E_i E_i^T) CC^T \left( \sum_{j=1}^p u_j \lambda_j^2 - 1 \right) \right] \\
&+ \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C \mathbf{E}(E_i^T C E_j^T) + C \mathbf{E}(E_i^T E_j) C^T + C \mathbf{E}(E_i^T E_j E_j^T)) \right] \\
&+ \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (\mathbf{E}(E_i C^T C E_j^T) + \mathbf{E}(E_i C^T E_j) C^T + \mathbf{E}(E_i C^T E_j E_j^T)) \right] \\
&+ \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (\mathbf{E}(E_i E_i^T C E_j^T) + \mathbf{E}(E_i E_i^T E_j) C^T + \mathbf{E}(E_i E_i^T E_j E_j^T)) \right].
\end{aligned}$$

Betrachten wir den entsprechenden Erwartungswert für die korrigierte Version, so erhalten wir

$$\begin{aligned}
&\mathbf{E} \left( \|\hat{C}\hat{C}^T - K - CC^T\|^2 \right) \\
&= \mathbf{E} \left( \text{trace} \left[ CC^T CC^T \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \right) \\
&+ \mathbf{E} \left( \text{trace} \left[ \sum_{j=1}^p u_j \lambda_j^2 CC^T (C E_j^T + E_j C^T) \left( \sum_{i=1}^p u_i \lambda_i^2 - 1 \right) \right] \right) \\
&+ \mathbf{E} \left( \text{trace} \left[ \sum_{i=1}^p u_i \lambda_i^2 (C E_i^T + E_i C^T) CC^T \left( \sum_{j=1}^p u_j \lambda_j^2 - 1 \right) \right] \right) \\
&+ \mathbf{E} \left( \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C E_i^T C E_j^T + C E_i^T E_j C^T) \right] \right) \\
&+ \mathbf{E} \left( \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (E_i C^T C E_j^T + E_i C^T E_j C^T) \right] \right)
\end{aligned}$$

$$\begin{aligned}
&= \text{trace} \left[ CC^T CC^T \left( \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 - \sum_{i=1}^p u_i \lambda_i^2 - \sum_{j=1}^p u_j \lambda_j^2 + 1 \right) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C \mathbf{E}(E_i^T C E_j^T) + C \mathbf{E}(E_i^T E_j) C^T) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (\mathbf{E}(E_i C^T C E_j^T) + \mathbf{E}(E_i C^T E_j) C^T) \right].
\end{aligned}$$

Auch hier stellen wir fest, dass durch den Korrekturterm insgesamt weniger Terme unter der Spur übrigbleiben und wir eine Verbesserung des resultierenden Schätzers annehmen können.

Sofern es keine Unterschiede zwischen den Produkten gibt und somit  $C = 0_{n,m}$  gilt, berechnen wir für die unkorrigierte Version

$$\begin{aligned}
\mathbf{E} \left( \|\hat{C} \hat{C}^T - CC^T\|^2 \right) &= \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 \mathbf{E}(E_i E_i^T E_j E_j^T) \right] \\
&\neq 0,
\end{aligned}$$

wohingegen wir für die korrigierte Version

$$\mathbf{E} \left( \|\hat{C} \hat{C}^T - K - CC^T\|^2 \right) = 0$$

erhalten.

Nehmen wir wie in 4.4 an, dass  $\sum_{i=1}^p \lambda_i^2 = p$  und  $u_i = \frac{1}{p}$  für alle  $i \in \{1, \dots, p\}$ , so erhalten wir die Gleichungen

$$\begin{aligned}
&\mathbf{E} \left( \|\hat{C} \hat{C}^T - CC^T\|^2 \right) \\
&= \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (C \mathbf{E}(E_i^T C E_j^T) + C \mathbf{E}(E_i^T E_j) C^T + C \mathbf{E}(E_i^T E_j E_j^T)) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (\mathbf{E}(E_i C^T C E_j^T) + \mathbf{E}(E_i C^T E_j) C^T + \mathbf{E}(E_i C^T E_j E_j^T)) \right] \\
&\quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 (\mathbf{E}(E_i E_i^T C E_j^T) + \mathbf{E}(E_i E_i^T E_j) C^T + \mathbf{E}(E_i E_i^T E_j E_j^T)) \right]
\end{aligned}$$

sowie

$$\begin{aligned} & \mathbf{E} \left( \|\hat{C}\hat{C}^T - K - CC^T\|^2 \right) \\ &= \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 \left( C\mathbf{E}(E_i^T C E_j^T) + C\mathbf{E}(E_i^T E_j)C^T \right) \right] \\ & \quad + \text{trace} \left[ \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 \left( \mathbf{E}(E_i C^T C E_j^T) + \mathbf{E}(E_i C^T E_j)C^T \right) \right], \end{aligned}$$

aus denen der Vorteil durch die Korrektur noch deutlicher wird. Wegen der stochastischen Unabhängigkeit von  $E_i$  und  $E_j$  für  $i \neq j$  ist zudem beispielsweise

$$\mathbf{E}(E_i E_i^T C E_j^T) = \mathbf{E}(E_i E_i^T) C \mathbf{E}(E_j^T) = 0_{n,n}$$

(entsprechend bei den übrigen Termen), und es gilt nach Voraussetzung

$$\mathbf{E}(E_i^T E_i) = \Sigma_i,$$

so dass sich die Ausdrücke wie folgt vereinfachen lassen:

$$\begin{aligned} & \mathbf{E} \left( \|\hat{C}\hat{C}^T - CC^T\|^2 \right) \\ &= \text{trace} \left[ \sum_{i=1}^p u_i^2 \lambda_i^4 \left( 2\mathbf{E}(C E_i^T C E_i^T) + 2C \Sigma_i C^T + 2\mathbf{E}(C E_i^T E_i E_i^T) \right) \right] \\ & \quad + \text{trace} \left[ \sum_{i=1}^p u_i^2 \lambda_i^4 \left( 2\mathbf{E}(E_i E_i^T E_i C^T) \right) + \sum_{i=1}^p \sum_{j=1}^p u_i u_j \lambda_i^2 \lambda_j^2 \left( \mathbf{E}(E_i E_i^T E_i E_i^T) \right) \right] \end{aligned}$$

sowie

$$\mathbf{E} \left( \|\hat{C}\hat{C}^T - K - CC^T\|^2 \right) = \text{trace} \left[ \sum_{i=1}^p u_i^2 \lambda_i^4 \left( 2\mathbf{E}(C E_i^T C E_i^T) + 2C \Sigma_i C^T \right) \right].$$

Diese Gleichungen verdeutlichen erneut, dass durch den Korrekturterm die Abweichung vom theoretischen Wert verringert wird. Insbesondere für  $C = 0_{n,m}$  haben wir hier für die Originalversion von STATIS wiederum eine erwartete Abweichung zwischen berechnetem und theoretischem Konsens, die von Null verschieden ist, wohingegen wir für die korrigierte Version eine Abweichung von Null erwarten.

## 4.8 Konvergenz des STATIS-Konsens für $p \rightarrow \infty$

Betrachten wir hier nochmals wie in Kapitel 4.6 den STATIS-Konsens  $W$  unter der Annahme, dass das arithmetische Mittel der Assoziationsmatrizen bestimmt wird, so erhalten wir für  $p \rightarrow \infty$

$$\begin{aligned} W &= \sum_{i=1}^p u_i \lambda_i^2 X_i X_i^T \\ &= \frac{1}{p} C C^T \sum_{i=1}^p \lambda_i^2 + \frac{1}{p} C \sum_{i=1}^p \lambda_i^2 E_i^T + \frac{1}{p} \left( \sum_{i=1}^p \lambda_i^2 E_i \right) C^T + \frac{1}{p} \sum_{i=1}^p \lambda_i^2 E_i E_i^T \\ &\rightarrow b \cdot C C^T + \Delta, \end{aligned}$$

wobei  $b$  eine unbekannte positive reelle Zahl ist und

$$\Delta = \begin{pmatrix} \delta_1^2 & & & \\ & \delta_2^2 & & 0 \\ & & \ddots & \\ & 0 & & \ddots \\ & & & & \delta_n^2 \end{pmatrix}.$$

Dabei ist  $\delta_l^2$  wie in 4.6 die durchschnittliche Fehlervarianz der Prüfpersonen für das  $l$ -te Produkt.

Weiter gilt (ebenfalls für  $p \rightarrow \infty$ ) offensichtlich für die Korrektur durch Subtraktion der Matrix  $K$

$$\begin{aligned} W - K &= C C^T \sum_{i=1}^p u_i \lambda_i^2 + C \sum_{i=1}^p u_i \lambda_i^2 E_i^T + \left( \sum_{i=1}^p u_i \lambda_i^2 E_i \right) C^T \\ &\quad + \sum_{i=1}^p u_i \lambda_i^2 E_i E_i^T - \sum_{i=1}^p u_i \lambda_i^2 E_i E_i^T \\ &\rightarrow b \cdot C C^T, \end{aligned}$$

d. h. hier erhalten wir asymptotisch die Assoziationsmatrix des wahren Konsens multipliziert mit einem unbekanntem Skalar  $b$ , dessen genauer Wert allerdings für die Interpretation der Ergebnisse ohne Bedeutung ist.

#### 4.9 Konvergenz, sofern die Fehlervarianzen gegen Null streben

Der Vollständigkeit halber wird an dieser Stelle noch kurz der Ansatz von Langron und Collins (1985) betrachtet. Sie untersuchen Eigenschaften der Verallgemeinerten Prokrustes Analyse für den Fall, dass die Varianzen der Zufallsfehler gegen Null konvergieren. Bei STATIS gilt dann zunächst bei der Bestimmung der Gewichte, dass sowohl  $trace(Z\Sigma_i)$  als auch  $trace(\Sigma_i\Sigma_i)$  gegen Null konvergieren. Damit erkennen wir aus Gleichung (12) in Kapitel 4.2 unmittelbar, dass der RV-Koeffizient für beliebige Prüfpersonen  $i$  und  $j$  gegen 1 konvergiert, d. h.

$$RV \longrightarrow 1_p 1_p^T .$$

Damit ist aber  $1_p$  Eigenvektor zum größten Eigenwert dieser Matrix, und nach Normierung wird jeder Prüfperson somit das Gewicht  $\frac{1}{p}$  zugewiesen, d. h. es wird asymptotisch das arithmetische Mittel berechnet.

Weiter erhalten wir für die Assoziationsmatrizen

$$\begin{aligned} W_i &= \lambda_i^2 (C + E_i)(C + E_i)^T \\ &\longrightarrow \lambda_i^2 C C^T , \end{aligned}$$

d. h. wir erhalten ein Vielfaches der Assoziationsmatrix des wahren Konsens und somit auch für den STATIS-Konsens als gewichtetes Mittel über die  $W_i$  ein Vielfaches hiervon. Berechnen wir dann eine Hauptkomponentenanalyse von  $W$ , so erhalten wir ein Vielfaches des theoretischen Konsens  $C$ .

Abschließend sei hierzu bemerkt, dass dieser Fall für die praktische Anwendung nicht von Bedeutung ist, da wir wohl in keinem Fall davon ausgehen können, dass die Varianz der Fehler, die die Prüfpersonen machen, beliebig klein wird.

## 5 Schätzung der Fehlermatrizen

Im vorherigen Kapitel haben wir gezeigt, dass der STATIS-Konsens durch eine geeignete Korrektur dahingehend verbessert werden kann, dass der Abstand des berechneten vom wahren Konsens verringert wird. Allerdings beruht diese Korrektur des STATIS-Konsens darauf, dass die Fehlermatrizen und die Skalare der einzelnen Prüfpersonen bekannt sind. Dies ist aber eine theoretische Annahme, die in praktischen Anwendungen nicht erfüllt ist. Daher stellt sich nun die Frage, wie diese unbekanntes Skalare und Matrizen geeignet geschätzt werden können. Nahe liegend wäre eine Schätzung der Fehler aus  $X_i X_i^T - \hat{C} \hat{C}^T$ . Allerdings tritt dabei das Problem auf, dass sich für diese Differenz auch unter der Annahme  $\hat{C} \hat{C}^T = C C^T$  der Ausdruck  $E_i E_i^T + E_i C^T + C E_i^T$  statt  $E_i E_i^T$  ergibt. Daher verwenden wir einen anderen Schätzer, den wir aus der Prokrustes Analyse herleiten (vgl. Schönemann, 1966). Wir bestimmen zunächst den Schätzer  $\hat{C}$  als Hauptkomponenten von  $\hat{C} \hat{C}^T$  und dann die Differenz zwischen jedem einzelnen  $X_i$  und  $\hat{C}$  nach optimaler Prokrustes Rotation und Skalierung. Konkret bestimmen wir dazu zunächst die optimale Rotation  $\hat{R}_i^T$  von  $X_i$  auf  $\hat{C}$  sowie den zugehörigen Skalar  $\hat{\lambda}_i$ . Anschließend schätzen wir den individuellen Fehler von Prüfperson  $i$  durch

$$\begin{aligned}\hat{E}_i &= \hat{\lambda}_i^{-1} X_i \hat{R}_i^T - \hat{C} \\ &= \hat{\lambda}_i^{-1} \lambda_i (C + E_i) R_i \hat{R}_i^T - \hat{C}\end{aligned}$$

und wir subtrahieren statt  $K$  aus Gleichung (5) dann

$$\hat{K} := \sum_{i=1}^p u_i \hat{\lambda}_i^2 \hat{E}_i \hat{E}_i^T$$

von  $W$ .

Sind die einzelnen Schätzer hinreichend gut, so gilt

$$\begin{aligned}\hat{\lambda}_i^{-1}\lambda_i &\approx 1, \\ R_i\hat{R}_i^T &\approx I_m\end{aligned}$$

sowie

$$\hat{C} \approx C.$$

Damit ist dann aber

$$\begin{aligned}\hat{E}_i &\approx 1 \cdot (C + E_i)I_m - C \\ &= E_i\end{aligned}$$

Wenngleich keine dieser Approximationen in der Praxis exakt gelten wird, können wir doch hoffen, dass diese Schätzer als Korrekturterme insoweit geeignet sind, dass es zu einer Verbesserung der Schätzung von  $C$  kommt. Da es aber bisher keine Erkenntnisse darüber gibt, wie gut die vorstehenden Approximationen sind, wird im Folgenden ein Vergleich anhand von Simulationen durchgeführt.

## 6 Simulationen zum Vergleich der Ergebnisse

Zur Untersuchung der Frage, ob das in Kapitel 5 vorgeschlagene Vorgehen tatsächlich zu besseren Resultaten führt, werden die nachfolgenden Simulationen zum Vergleich der verschiedenen Verfahren durchgeführt. Wir verwenden dazu das hier betrachtete Modell (3) der Form

$$X_i = \lambda_i(C + E_i)R_i ,$$

welches für die Simulationen ohne Rotationsmatrizen in der Form

$$X_i = \lambda_i(C + E_i)$$

verwendet werden kann, vergleiche dazu Meyners *et al.* (2000). Ebenso wie in jenem Artikel werden verschiedene Datenmatrizen aus praktischen Untersuchungen für  $C$  verwendet, um eine möglichst realistische Datenlage zu betrachten. Zudem wird die Varianz der Fehler der Prüfpersonen variiert, so dass gute und weniger gute Prüfpersonen gleichermaßen repräsentiert werden können.

Als Maß für die Güte der Ergebnisse wird dabei der RV-Koeffizient zwischen dem berechneten  $\hat{C}$  und dem wahren  $C$  verwendet. Je größer dieser Koeffizient ist, desto besser ist die Schätzung von  $C$  durch das zugehörige Verfahren.

Verglichen werden die Ergebnisse von STATIS mit und ohne Korrektur um zu untersuchen, ob die vorgeschlagene Korrektur zu besseren Schätzungen führt. Darüber hinaus werden noch zwei weitere Varianten von STATIS betrachtet: Zum einen untersuchen wir eine vereinfachte Variante, bei der wir als Gewichte für alle

Prüfpersonen  $\frac{1}{p}$  verwenden und somit das arithmetische Mittel aus den Assoziationsmatrizen berechnen. Zum anderen verwenden wir die asymptotischen Gewichte, wie sie in Kapitel 4.2 bestimmt wurden. Dieses ist in den Simulationen möglich, weil die Varianzen der Fehler, die die Prüfpersonen machen, bekannt sind.

Wir betrachten verschiedene Matrizen für  $C$  und setzen zur Bestimmung der asymptotischen Gewichte

$$Z := CC^T .$$

Weiterhin verwenden wir verschiedene Annahmen über die Kovarianzmatrizen  $\Sigma_i$  und setzen für diese voraus, dass sie sich in der Form  $\sigma_i^2 \Sigma$  darstellen lassen und somit also eine identische Struktur für alle Prüfpersonen  $i$  haben. Ohne Beschränkung der Allgemeinheit gelte zudem wiederum  $\text{trace}(\Sigma\Sigma) = 1$ .

Ist beispielsweise im einfachsten Fall  $\Sigma$  ein Vielfaches der Einheitsmatrix, so folgt aufgrund der getroffenen Annahmen, dass

$$\Sigma = \frac{1}{\sqrt{m}} I_m .$$

Sind andererseits alle Korrelationen zwischen den Variablen identisch, also etwa gleich  $\rho$ , so ist  $\Sigma$  ein Vielfaches von  $(1 - \rho)I_m + \rho\mathbf{1}_m\mathbf{1}_m^T$ , und wegen der Voraussetzungen folgt

$$\Sigma = \frac{1}{\sqrt{m + m(m-1)\rho^2}} \left( (1 - \rho)I_m + \rho\mathbf{1}_m\mathbf{1}_m^T \right) .$$

Schließlich sei noch der Fall betrachtet, in dem alle Variablen voneinander unabhängig sind mit Ausnahme von zweien, die die Korrelation  $\rho$  besitzen. Dieses könnten beispielsweise zwei sehr ähnliche und für die Gutachter kaum unterscheidbare Merkmale sein. Weist eine Prüfperson einem der Merkmale einen zu großen Wert zu, so wird sie auch dem anderen Merkmal häufig einen zu großen Wert zuweisen. Die zugehörigen Variablen seien ohne Beschränkung der Allgemeinheit die beiden ersten in jeder Matrix, so dass  $\Sigma$  ein Vielfaches der Matrix ist, die aus der

Einheitsmatrix besteht, bei der die Nullen an den Stellen  $(1, 2)$  und  $(2, 1)$  durch den Wert  $\rho$  ersetzt werden. Aus den Voraussetzungen folgt dann

$$\Sigma = \frac{1}{\sqrt{m + 2\rho^2}} \begin{pmatrix} 1 & \rho & 0 & 0 & \dots & 0 \\ \rho & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Wir betrachten nun die gerade eingeführten Strukturen der Kovarianzmatrizen in Verbindung mit verschiedenen Konstellationen für die Varianzen der Prüfpersonen, die durch die Werte der  $\sigma_i^2$  repräsentiert werden. Für unterschiedliche Matrizen  $C$  geben wir im Folgenden das Verhältnis der entsprechenden Werte von  $\sigma_i^2$  für die einzelnen Prüfpersonen und die zugehörigen asymptotischen Gewichte an. Anschließend werden die Ergebnisse der Simulationen zum Vergleich der verschiedenen Versionen von STATIS aufgeführt. Je nach Gesamtzahl der Prüfpersonen werden teilweise zusätzlich bis zu zwei Ausreißer betrachtet. Dazu nehmen wir an, dass zwei Produkte bei der Versuchsdurchführung verwechselt wurden und daher die entsprechenden Zeilen der Datenmatrizen vertauscht sind.

Es gibt natürlich unendlich viele denkbare Konstellationen bezüglich der Anzahl der Prüfpersonen, deren Güte u. a. m. Um die Ergebnisse möglichst repräsentativ zu gestalten, betrachten wir ebenso wie Meyners *et al.* (2000) die drei folgenden Datensätze:

1. die Bewertungen der ersten Prüfperson der sensorischen Profilanalyse von neun Rindfleischsorten bezüglich sieben fest vorgegebenen Merkmalen, die von Gower (1975) zitiert werden;
2. die Bewertungen ( $m = 8$ ) der zweiten Prüfperson eines Free-Choice-Profilings von acht Joghurtsorten, die sich bei Dijksterhuis und Gower (1991) finden;
3. eine Matrix aus gleichverteilten Zufallszahlen auf dem Intervall  $[0, 100]$ , wobei  $n = 11$  und  $m = 5$  angenommen wird.

Da Meyners *et al.* (2000) für die zufälligen Bewertungen der letztgenannten Matrix teilweise andere Ergebnisse erhalten als für die ersten beiden Datensätze, wird dieser Zufallsdatensatz hier ebenfalls in die Simulationsstudie aufgenommen.

Die Datenmatrizen werden ebenso simuliert wie bei Meyners *et al.* (2000), allerdings mit dem Unterschied, dass gute und schlechte Prüfpersonen hier neu definiert werden. Eine durchschnittliche Prüfperson habe eine Fehlervarianz, die der empirischen Varianz im theoretischen Konsens entspricht. Meyners *et al.* (2000) verwenden für schlechte Prüfpersonen eine 25-fache Varianz im Vergleich zu einer durchschnittlichen. Diese Vorgabe stellt sich allerdings als so groß heraus, dass die beobachteten Matrizen im Wesentlichen als reine Zufallszahlen angesehen werden müssen. Deshalb verwenden wir hier nur den Faktor 5 für schlechte Prüfpersonen, und analog wird diese Varianz für gute Prüfpersonen nur durch 5 dividiert statt durch 25 wie bei Meyners *et al.* (2000).

Für jede Situation werden 1000 Simulationen durchgeführt und dabei wird festgehalten, wie oft jedes Verfahren den anderen, die in der Studie betrachtet werden, überlegen ist. Die nachfolgenden Tabellen enthalten die entsprechenden Angaben. Die aufgeführte Zahl gibt jeweils an, wie oft das Verfahren der zugehörigen Spalte besser war als das der zugehörigen Zeile. Gehören Zeile und Spalte zum selben Verfahren, so erhalten wir als Eintrag natürlich eine Null. Darüber hinaus sollten sich die beiden Werte, die zum selben Paar von STATIS-Versionen gehören, zu 1000 aufsummieren. In den Fällen, in denen das nicht zutrifft, war der RV-Koeffizient zwischen dem jeweiligen  $\hat{C}$  der beiden Methoden und  $C$  in einigen Simulationen exakt gleich. Letzteres ist verständlicherweise vor allem der Fall beim Vergleich zwischen dem arithmetischen Mittel und der Version mit den asymptotischen Gewichten, denn falls alle Prüfpersonen die gleiche theoretische Fehlervarianz aufweisen, so sind entsprechend der Ausführungen aus Kapitel 4.3 die asymptotischen Gewichte jeweils gleich  $\frac{1}{p}$  und die berechneten Konsense dieser beiden Methoden stimmen damit exakt überein.

Zunächst betrachten wir den Datensatz von Gower (1975). In einer ersten Simulation gehen wir von nur drei Prüfpersonen aus, die alle eine durchschnittliche Fehlervarianz aufweisen und deren Bewertungen für die verschiedenen Variablen unabhängig voneinander sind. Die zugehörigen Ergebnisse sind in Tabelle 1 angegeben.

Methoden	original	arithmet.	asymptot.	korrigiert
Originalversion	0	440	440	986
arithmetisches Mittel	560	0	0	986
asymptotische Gewichte	560	0	0	986
korrigierte Version	14	14	14	0

**Tabelle 1:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Datensatz von Gower (1975), drei durchschnittliche Prüfpersonen, keine Ausreißer, unkorrelierte Variablen.

Wir erkennen aus Tabelle 1, dass die STATIS-Originalversion in etwas mehr als der Hälfte der Fälle bessere Ergebnisse liefert als die Version, die das arithmetische Mittel verwendet. Letztere Version ist identisch mit der Version mit asymptotischen Gewichten, da ja alle Prüfpersonen die gleiche Fehlervarianz aufweisen und somit alle asymptotischen Gewichte gleich  $\frac{1}{3}$  sind. Darüber hinaus erkennen wir den deutlichen Vorteil, der durch die Korrektur erzielt wird. Diese Methode ist mit wenigen Ausnahmen in allen Wiederholungen besser als die anderen drei Verfahren. Die Tatsache, dass es jeweils 986 Fälle sind, legt die Vermutung nahe, dass die übrigen drei Versionen in identischen Simulationen zu besseren Ergebnissen geführt haben und die Korrektur aus unbekanntem Gründen dort jeweils nicht erfolgreich zum Tragen kam. In anderen Situationen mit nur drei Prüfpersonen und für die übrigen Datensätze ergeben sich vergleichbare Ergebnisse.

An dieser Stelle sei dazu erwähnt, dass eine sensorische Profilprüfung üblicherweise mit mehr als drei Prüfpersonen durchgeführt wird. Da aber Gower (1975) einen solchen Datensatz verwendet, beziehen wir diesen Fall ebenfalls in unsere Untersuchung ein. In allen nachfolgenden Situationen werden wir deutlich mehr Prüfpersonen betrachten.

Untersuchen wir in weiteren Simulationen den gleichen Datensatz mit der geänderten Vorgabe, dass wir neun Prüfpersonen beobachten, die alle eine kleine bzw. alle eine große Fehlervarianz aufweisen und deren Variablen miteinander jeweils mit  $\rho = 0,2$  korreliert sind. Die zugehörigen Ergebnisse sind in den Tabellen 2 und 3 angegeben.

Methode	original	arithmet.	asymptot.	korrigiert
Originalversion	0	387	387	984
arithmetisches Mittel	613	0	0	984
asymptotische Gewichte	613	0	0	984
korrigierte Version	16	16	16	0

**Tabelle 2:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Datensatz von Gower (1975), neun gute Prüfpersonen, keine Ausreißer, gleiche Korrelation  $\rho = 0,2$  zwischen allen Variablen.

Methode	original	arithmet.	asymptot.	korrigiert
Originalversion	0	150	150	1000
arithmetisches Mittel	850	0	0	1000
asymptotische Gewichte	850	0	0	1000
korrigierte Version	0	0	0	0

**Tabelle 3:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Datensatz von Gower (1975), neun schlechte Prüfpersonen, keine Ausreißer, gleiche Korrelation  $\rho = 0,2$  zwischen allen Variablen.

Die Werte in den Tabellen 2 und 3 sind wie im vorherigen Beispiel zu interpretieren. Wir erkennen dabei, dass die korrigierte Version wiederum deutlich bessere Ergebnisse liefert als die übrigen Verfahren. Des Weiteren erkennen wir, dass sich dieser Effekt offensichtlich verstärkt, wenn die Prüfpersonen eine große Fehlervarianz aufweisen. Diese Beobachtung bestätigt, dass die Form der Korrektur sinnvoll ist. Sind nur geringe Zufallsschwankungen vorhanden, so kann das Verfahren nicht wesentlich besser sein als die anderen. Insbesondere führten, dem Ergebnis aus Kapitel 4.9 entsprechend, zwangsläufig alle Verfahren zu identischen Ergebnissen, wenn überhaupt keine Fehler gemacht würden und das Modell (1) mit  $E_i = 0_{n,m}$  für alle

$i \in \{1, \dots, p\}$  gälte. Je größer aber diese Schwankungen werden, desto größer sind auch die Einträge auf der Hauptdiagonalen des Erwartungswertes von  $E_i E_i^T$  und desto wirksamer sollte die vorgeschlagene Korrektur sein.

Wir betrachten nun den Datensatz von Dijksterhuis und Gower (1991), und zwar zunächst wieder für unkorrelierte Variablen. In Tabelle 4 sind die Ergebnisse für neun Prüfpersonen aufgeführt, von denen je drei gut, durchschnittlich bzw. schlecht sind. Dieses führt nun natürlich zu unterschiedlichen asymptotischen Gewichten für die einzelnen Prüfpersonen, und die Verwendung des arithmetischen Mittels ist nicht mehr identisch mit der Verwendung der asymptotischen Gewichte. In diesem speziellen Fall ergibt sich für die guten Prüfpersonen jeweils das Gewicht 0,2026, für die durchschnittlichen 0,0905 und für die schlechten Prüfpersonen 0,0402. Die Tabelle zeigt, dass die Verwendung der asymptotischen Gewichte der Version mit dem arithmetischen Mittel und auch der Originalversion von STATIS weit überlegen ist. Könnten wir diese in der Praxis umsetzen, wären die Ergebnisse vergleichbar gut mit denen der korrigierten Version.

Methode	original	arithmet.	asymptot.	korrigiert
Originalversion	0	0	1000	991
arithmetisches Mittel	1000	0	1000	997
asymptotische Gewichte	0	0	0	452
korrigierte Version	9	3	548	0

**Tabelle 4:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Datensatz von Dijksterhuis und Gower (1991), je drei gute, durchschnittliche bzw. schlechte Prüfpersonen, keine Ausreißer, unkorrelierte Variablen.

Verwenden wir nur je zwei gute und schlechte Prüfpersonen und gehen wir davon aus, dass für eine Prüfperson zudem zwei Produkte vertauscht wurden (Ausreißer), so erhalten wir erneut qualitativ die gleichen Ergebnisse wie vorher. In Tabelle 5 sind die Resultate für den Fall angegeben, dass pro Prüfperson zwei Variablen mit  $\rho = 0,4$  korreliert sind. Die Gewichte der Prüfpersonen sind in diesem Fall 0,216, 0,0965 bzw. 0,0429.

Methoden	original	arithmet.	asymptot.	korrigiert
Originalversion	0	6	931	992
arithmetisches Mittel	994	0	957	997
asymptotische Gewichte	69	43	0	728
korrigierte Version	8	3	272	0

**Tabelle 5:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Datensatz von Dijksterhuis und Gower (1991), neun Prüfpersonen, von denen je zwei gut bzw. schlecht und fünf durchschnittlich sind, ein Ausreißer, zwei mit  $\rho = 0,4$  korrelierte Variablen.

Betrachten wir 15 Prüfpersonen für den gleichen Datensatz, so erhalten wir für beispielsweise acht gute bzw. schlechte und je sieben durchschnittliche Prüfpersonen mit identischer Korrelation  $\rho = 0,2$  zwischen allen Variablen die in den Tabellen 6 und 7 angegebenen Ergebnisse. Auch hier ergibt sich kein qualitativer Unterschied zu den vorher untersuchten Situationen. Die korrigierte Version liefert im Allgemeinen die besten Ergebnisse, bei diesen Konstellationen sogar bessere als die Version unter Verwendung der asymptotischen Gewichte. Die Gewichte sind hierbei im ersten Fall 0,0899 für gute bzw. 0,0401 für durchschnittliche Prüfpersonen und im zweiten Fall entsprechend 0,0947 bzw. 0,0421.

Methoden	original	arithmet.	asymptot.	korrigiert
Originalversion	0	1	997	1000
arithmetisches Mittel	999	0	997	1000
asymptotische Gewichte	3	3	0	871
korrigierte Version	0	0	129	0

**Tabelle 6:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Datensatz von Dijksterhuis und Gower (1991), acht gute und sieben durchschnittliche Prüfpersonen, keine Ausreißer, gleiche Korrelation  $\rho = 0,2$  zwischen allen Variablen.

Methoden	original	arithmet.	asymptot.	korrigiert
Originalversion	0	0	1000	1000
arithmetisches Mittel	1000	0	1000	1000
asymptotische Gewichte	0	0	0	998
korrigierte Version	0	0	2	0

**Tabelle 7:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Datensatz von Dijksterhuis und Gower (1991), acht schlechte und sieben durchschnittliche Prüfpersonen, keine Ausreißer, gleiche Korrelation  $\rho = 0,2$  zwischen allen Variablen.

Verwenden wir hingegen sieben gute und acht schlechte Prüfpersonen, so erhalten die guten Prüfpersonen ein Gewicht von 0,1164, während die schlechten Prüfpersonen nur mit 0,0231 gewichtet werden. Dadurch beeinflussen die guten Probanden, die kaum Fehler aufweisen, den Gesamtkonsens wesentlich stärker. Das führt dazu, dass in dieser Situation die Verwendung der asymptotischen Gewichte zu besseren Ergebnissen führt als die korrigierte Version. Letztere bleibt aber auch in diesem Fall die mit Abstand beste unter den in der Praxis durchführbaren Versionen. Die Ergebnisse hierzu sind in Tabelle 8 aufgeführt.

Methode	original	arithmet.	asymptot.	korrigiert
Originalversion	0	0	1000	1000
arithmetisches Mittel	1000	0	1000	1000
asymptotische Gewichte	0	0	0	318
korrigierte Version	0	0	682	0

**Tabelle 8:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Datensatz von Dijksterhuis und Gower (1991), sieben gute und acht schlechte Prüfpersonen, keine Ausreißer, gleiche Korrelation  $\rho = 0,2$  zwischen allen Variablen.

Betrachten wir schließlich noch die Ergebnisse des Zufallsdatensatzes, so erkennen wir in den entsprechenden Situationen die gleichen Muster wie vorher, d. h. die Ergebnisse sind qualitativ identisch mit denen für die anderen Datensätze. Tabelle 9 zeigt die Ergebnisse für den Fall, dass wir neun Prüfpersonen simulieren, von denen fünf gut (Gewicht 0,1726) und vier schlecht (0,0343) sind und unter denen sich zwei Ausreißer befinden. Die Variablen haben untereinander dieselbe Korrelation.

Methode	original	arithmet.	asymptot.	korrigiert
Originalversion	0	0	991	964
arithmetisches Mittel	1000	0	998	993
asymptotische Gewichte	9	2	0	156
korrigierte Version	36	7	844	0

**Tabelle 9:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Zufallsdatensatz, fünf gute und vier schlechte Prüfpersonen, zwei Ausreißer, gleiche Korrelation  $\rho = 0,2$  zwischen allen Variablen.

Für die Resultate in Tabelle 10 wurden nochmals 15 Prüfpersonen betrachtet, von denen drei gut (0,1313), drei schlecht (0,0261) und die übrigen neun durchschnittlich (0,0587) sind. Hierbei wurden die Variablen unkorreliert erzeugt und es wurden keine Ausreißer simuliert.

Methoden	original	arithmet.	asymptot.	korrigiert
Originalversion	0	0	991	964
arithmetisches Mittel	1000	0	998	993
asymptotische Gewichte	9	2	0	156
korrigierte Version	36	7	844	0

**Tabelle 10:** Vergleich der Ergebnisse der verschiedenen STATIS-Versionen: Zufallsdatensatz, je drei gute und schlechte sowie neun durchschnittliche Prüfpersonen, keine Ausreißer, unkorrelierte Variablen.

Auch die hier nicht explizit dargestellten Ergebnisse für andere Konstellationen zeigen qualitativ völlig identische Resultate. Daher können wir also insgesamt feststellen, dass in allen betrachteten Situationen die Korrektur zu einer deutlichen Verbesserung der Schätzung des Konsens beiträgt. Die Verwendung der asymptotischen Gewichte führt zwar ebenfalls zu guten Ergebnissen, ist aber – wie bereits erwähnt – in der Praxis nicht durchführbar, da diese Gewichte grundsätzlich nicht bekannt sind. Die Version mit dem arithmetischen Mittel über alle Prüfpersonen erzielt erwartungsgemäß die schlechtesten Ergebnisse, da bei der Gewichtung in keiner Weise zwischen Prüfpersonen mit großer und solchen mit geringer Fehlervarianz unterschieden wird.

## 7 Vergleich mit der Verallgemeinerten Prokrustes Analyse (GPA)

Nachdem bisher ausschließlich ein Vergleich zwischen den verschiedenen STATIS-Varianten erfolgt ist, soll nun noch die Frage untersucht werden, ob die vorgeschlagene Korrektur auch zu besseren Ergebnissen führt als die Verallgemeinerte Prokrustes Analyse, für die wir den Algorithmus von ten Berge (1977) verwenden.

Meyners *et al.* (2000) stellen aufgrund vergleichbarer Simulationen wie in Kapitel 6 fest, dass die GPA einen Schätzer liefert, der das theoretische  $C$  besser approximiert als der entsprechende Schätzer von STATIS. Zugleich zeigen sie für den Spezialfall  $C = 0_{n,m}$ , dass STATIS im Gegensatz zur GPA einen Konsens berechnet, der im Erwartungswert von Null verschieden ist, (vgl. dazu auch die Ausführungen in Kapitel 4.6). Eine Veränderung des STATIS-Algorithmus erscheint vor diesem Hintergrund daher nur dann sinnvoll, wenn durch die Korrektur bessere Ergebnisse erzielt werden als mit Hilfe der GPA oder wenn die Ergebnisse mindestens genauso gut sind und der Rechenaufwand gleichzeitig verringert wird. Vorab kann schon festgestellt werden, dass nach diesem Kriterium bei vergleichbar guten Ergebnissen die veränderte Version von STATIS zumindest für wachsendes  $p$  sinnvoll erscheint. Sehen wir von der Berechnung der Gewichte ab, die sowohl für die GPA als auch für STATIS die Bestimmung des größten Eigenvektors einer  $(p, p)$ -Matrix erfordert, so wächst die Anzahl der übrigen Rechenschritte des Algorithmus von STATIS auch in der veränderten Form linear in  $p$ : Zunächst werden die  $p$  Assoziationsmatrizen berechnet, anschließend das gewichtete Mittel gebildet und schließlich

die Fehlermatrizen der Prüfpersonen geschätzt, die dann ebenfalls gewichtet gemittelt werden. Im Gegensatz dazu steigt bei der Verwendung der GPA die Anzahl der Rechenschritte mit wachsendem  $p$  nichtlinear, da der Algorithmus iterativ ist und erst dann abbricht, wenn die Veränderung zwischen zwei Iterationen vernachlässigbar klein ist. Führt man immer die gleiche Anzahl von Iterationen durch, so stiege auch für die GPA die Zahl der Rechenschritte linear in  $p$ . Da es aber mit zunehmender Anzahl von Prüfpersonen immer schwieriger wird, alle Matrizen so zu rotieren, dass die Abweichungen minimal werden, wird eine wachsende Zahl von Iterationen erforderlich. Damit wächst aber die Zahl der Rechenschritte stärker als linear in  $p$ .

Um den Vergleich durchzuführen, werden dieselben Simulationen betrachtet wie in Meyners *et al.* (2000). Ebenso wie dort betrachten wir in diesem Fall als Maß für die Güte der Ergebnisse zum einen den RV-Koeffizienten zwischen dem berechneten  $\hat{C}$  und dem wahren  $C$ , zum anderen den euklidischen Abstand von  $\hat{C}$  und  $C$ , nachdem die Matrizen geeignet skaliert und aufeinander rotiert wurden. Details finden sich auch hierzu bei Meyners *et al.* (2000) sowie bei Meyners (1997). Führen diese beiden Gütemaße zu unterschiedlichen Ergebnissen, so wird die zugehörige Simulation als unentschieden gewertet.

Tabelle 11 zeigt am Beispiel des Datensatzes von Gower (1975) mit  $n = 9$  und  $m = 7$  die Ergebnisse für verschiedene Konstellationen und gibt an, wie oft jedes Verfahren dem anderen überlegen war. In jeder Situation wurden 1000 Wiederholungen durchgeführt; addieren sich die beiden letzten Zahlen nicht zu 1000, so war in den fehlenden Wiederholungen keine eindeutige Entscheidung zugunsten eines der Verfahren möglich.

Insgesamt ist aus den Ergebnissen keine allgemeine Aussage möglich, welches der beiden Verfahren grundsätzlich dem anderen überlegen ist. Allerdings erkennen wir, dass die korrigierte Version von STATIS häufig einen besseren Schätzer liefert als die GPA. Insbesondere gilt das, wenn die Prüfpersonen eine vergleichbare Güte besitzen,

$p$	gut	schlecht	Ausreißer	bessere Ergebnisse bei	
				GPA	korrig. STATIS
9	0	0	0	105	591
9	9	0	0	133	379
9	0	9	0	374	457
9	0	0	1	100	640
9	9	0	1	68	584
9	0	9	1	318	489
9	0	0	2	74	674
9	9	0	2	53	652
9	0	9	2	375	463
9	3	3	0	531	200
9	3	3	1	546	223
9	3	3	2	498	270
9	5	4	0	506	179
9	5	4	1	593	196
9	5	4	2	608	193
15	0	0	0	28	756
15	15	0	0	46	455
15	0	15	0	248	600
15	5	5	0	319	300
15	5	5	1	318	329
15	5	5	2	332	345
15	8	0	0	138	423
15	0	8	0	457	302
15	7	8	0	527	175
15	3	3	0	216	435
15	9	3	0	48	388
15	3	9	0	726	127
15	3	3	1	224	429
15	9	3	1	40	469
15	3	9	1	706	148
15	3	3	2	225	460
15	9	3	2	31	465
15	3	9	2	723	148

**Tabelle 11:** Vergleich der Ergebnisse der korrigierten Version von STATIS und der GPA für den Datensatz von Gower (1975), jeweils 1000 Wiederholungen.

d. h. eine ähnliche Fehlervarianz aufweisen. Je größer die Zahl der Prüfpersonen ist, desto stärker scheint dieser Effekt zu werden. Demgegenüber liefert die GPA bessere Ergebnisse, wenn die Prüfpersonen sehr inhomogen sind, also wenn ein Teil eine kleine und ein anderer – oft größerer Teil – eine sehr große Fehlervarianz aufweist.

Ein Beispiel dafür ist etwa bei insgesamt 15 Prüfpersonen der Fall, in dem drei gute, neun schlechte und somit drei durchschnittliche Prüfpersonen betrachtet werden. Die Ergebnisse für die anderen Datensätze und für weitere Konstellationen sind qualitativ völlig identisch. Sie werden hier daher nicht zusätzlich angegeben.

Aufgrund der Untersuchungen können wir feststellen, dass die GPA dem neu vorgestellten Verfahren keinesfalls mehr überlegen ist. Insbesondere wenn alle Prüfpersonen einen vergleichbaren Kenntnisstand aufweisen (wie etwa in einem geschulten Panel) liefert die korrigierte Version von STATIS im Schnitt die besseren Ergebnisse. Zudem sind diese mit wachsendem  $p$  auch schneller zu berechnen.

## 8 Grafischer Vergleich der Analysemethoden

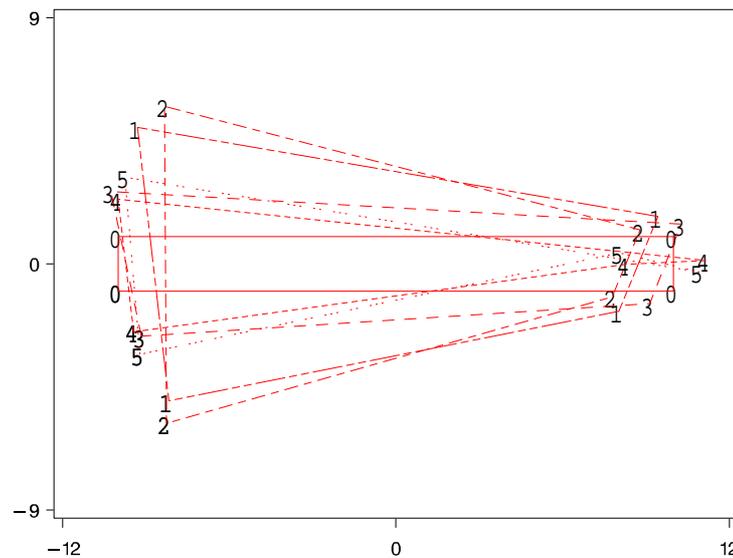
Zum Abschluss sollen in diesem Kapitel die Ergebnisse der unterschiedlichen Verfahren grafisch veranschaulicht werden. Dazu wird zunächst ein einfacher und gut darstellbarer Fall betrachtet: Wir verwenden einen zweidimensionalen theoretischen Konsens, der durch die Matrix

$$\begin{pmatrix} 10 & 1 \\ 10 & -1 \\ -10 & 1 \\ -10 & -1 \end{pmatrix}$$

gegeben ist, d. h. wir haben große Unterschiede in der ersten Dimension und wesentlich kleinere Unterschiede in der zweiten Dimension. Simuliert werden für dieses Beispiel neun Prüfpersonen, von denen jeweils drei eine große, eine mittlere bzw. eine kleine Fehlervarianz aufweisen. Bestimmen wir den jeweiligen Konsens der einzelnen Verfahren, so erhalten wir die in Abbildung 1 gezeigte Darstellung.

Wir erkennen sehr deutlich die vorgegebene rechteckige Struktur des theoretischen Konsens, die von keinem der Verfahren vollständig richtig erkannt wird. Entsprechend den Überlegungen von Meyners *et al.* (2000) sollte die Dimension des Konsens von den Verfahren überschätzt werden, was durch die Grafik bestätigt wird. Ist der theoretische Konsens nahezu eindimensional, so erhalten wir mit allen verwendeten Verfahren einen geschätzten Konsens, der die zweite Dimension zu stark betont.

In unserem Beispiel werden vor allem die Unterschiede zwischen den Produkten 3 und 4 überschätzt, also den Produkten auf der linken Seite in Abbildung 1. Dieser



**Abbildung 1:** Grafische Darstellung der Konsense der verschiedenen Methoden (0 = theoretischer Konsens, 1 = Originalversion von STATIS, 2 = Version mit arithmetischem Mittel, 3 = Version mit asymptotischen Gewichten, 4 = Version mit Korrektur, 5 = GPA).

Effekt ist am stärksten ausgeprägt bei der Version unter Verwendung des arithmetischen Mittels sowie bei der Originalversion von STATIS. In wesentlich geringerem Maße tritt das Problem bei der Verallgemeinerten Prokrustes Analyse auf, bei der die Lage der Produkte (bis auf Produkt 1) der des theoretischen Konsens ähnlicher ist als die von STATIS geschätzte.

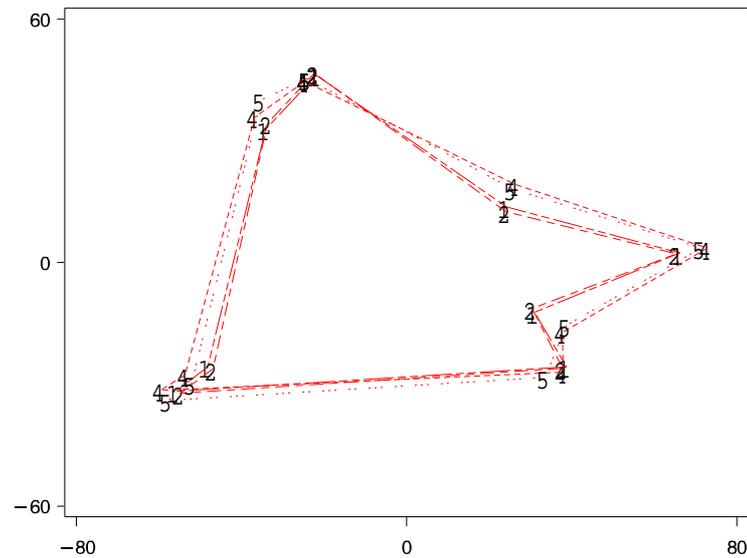
Die Verwendung der asymptotischen Gewichte führt in diesem Beispiel zu einer verhältnismäßig guten Schätzung, und zwar im Vergleich zur korrigierten Version insbesondere auch bei den Produkten 1 und 2, bei denen die korrigierte Version die Unterschiede in der falschen Dimension schätzt. Allerdings ist die korrigierte Version unter den in der Praxis durchführbaren Methoden die geeignetste, da sie zum einen die Lage der Produkte besser schätzt als die anderen Verfahren und zum anderen die zweite Dimension wesentlich weniger stark überschätzt.

Über diesen künstlich erzeugten Fall hinaus wollen wir noch die Ergebnisse der verschiedenen in der Praxis umsetzbaren Verfahren anhand von zwei Beispielen

vergleichen. Sowohl in den theoretischen Überlegungen als auch in den Simulationen wurde deutlich, dass die Verfahren unterschiedliche Schätzer liefern. Allerdings wurde bisher nicht untersucht, ob diese Unterschiede für praktische Anwendungen überhaupt relevant sind. Das Ziel einer Datenauswertung mit Hilfe von GPA oder STATIS ist im Allgemeinen, aus mehreren Prüfurteilen einen Konsens zu bestimmen. Aus diesem berechneten Konsens sollen dann Aussagen darüber getroffen werden, welche Produkte einander ähnlich sind und welche sich voneinander unterscheiden. Mit Hilfe von Biplots (vgl. Gabriel, 1971) kann zudem versucht werden, die Unterschiede inhaltlich zu interpretieren, indem den Richtungen der größten Unterschiede gewisse Variablen zugeordnet werden. Für diese Interpretationen wird üblicherweise eine grafische Darstellung verwendet, die zumeist aus den ersten beiden Hauptkomponenten des berechneten Konsens besteht. Erklären diese Hauptkomponenten einen zu geringen Anteil der Gesamtvarianz, so werden bisweilen auch weitere Dimensionen berücksichtigt.

Um einen Eindruck zu bekommen, wie relevant die Unterschiede zwischen den Schätzern der einzelnen Verfahren sind, betrachten wir erneut die Datensätze von Dijksterhuis und Gower (1991) sowie von Gower (1975), die auch schon für die Simulationen in den Kapiteln 6 und 7 benutzt wurden. Wir verwenden nun jeweils den vollständigen Datensatz und berechnen den Konsens mit den verschiedenen Methoden. Da Rotationen und Skalierungen nicht zu Interpretationsänderungen führen, werden alle Konsense optimal auf einen Konsens (hier ohne Beschränkung der Allgemeinheit auf den der STATIS-Originalversion) rotiert und skaliert, vergleiche dazu Schönemann (1966). Dadurch müssen die Produkte in der Darstellung nicht explizit angegeben werden, da jeweils dicht beieinander liegende Eckpunkte der Polygone immer zum gleichen Produkt gehören. Abbildung 2 zeigt die entsprechenden Konsense für den Datensatz von Dijksterhuis und Gower (1991), Abbildung 3 die für den Datensatz von Gower (1975). Die Nummerierung der Verfahren entspricht derjenigen in Abbildung 1. Da die asymptotischen Gewichte natürlich unbekannt

sind, kann die zugehörige Variante hier nicht berücksichtigt werden. Die entsprechende Ziffer entfällt daher in diesen Darstellungen ebenso wie die des unbekanntem wahren Konsens.

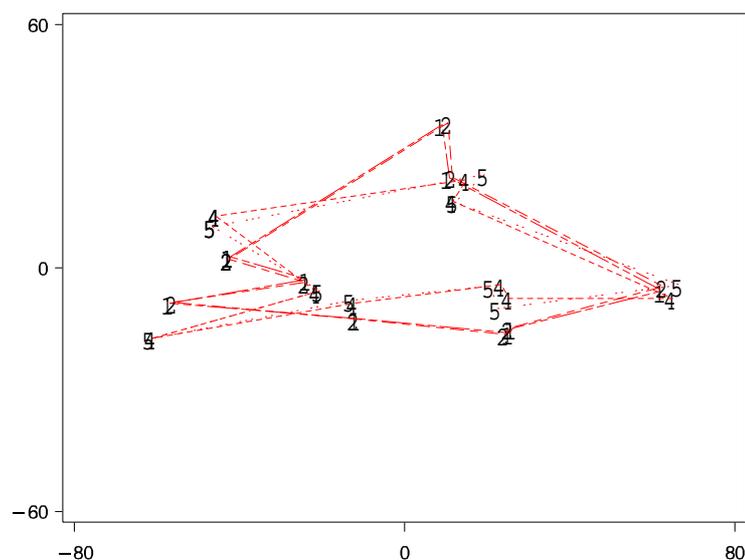


**Abbildung 2:** Grafische Darstellung der Konsense der verschiedenen Methoden für den Datensatz von Dijksterhuis und Gower (1991) (1 = Originalversion von STATIS, 2 = Version mit arithmetischem Mittel, 4 = Version mit Korrektur, 5 = GPA).

In Abbildung 2 erkennen wir zunächst sehr deutlich, dass die Struktur der Daten durch die Verfahren nahezu identisch geschätzt wird. Links unten sowie im oberen Teil sind jeweils zwei Produkte, die sehr ähnlich bewertet wurden und die sich beispielsweise von dem Produkt ganz rechts sehr deutlich unterscheiden. Bei genauerer Betrachtung stellen wir fest, dass die GPA sowie die korrigierte Version im Vergleich zu den anderen Verfahren eine etwas größere Variabilität in der ersten Dimension aufweisen – die Urteile liegen jeweils weiter außen, d. h. die zweite Dimension ist gegenüber der ersten nicht ganz so gewichtig wie bei den Ergebnissen der beiden anderen Verfahren. Dies bestätigt nochmals die Überlegung, dass die Originalversion von STATIS niedere Dimensionen zu stark gewichtet, während dieser Effekt durch die Korrektur zumindest teilweise aufgehoben wird. Allerdings müssen wir

feststellen, dass die Ergebnisse einander sehr ähnlich sind, so dass wir mit jedem Verfahren letztlich zu denselben Interpretationen kommen.

Für den zweiten Datensatz ergibt sich ein ähnliches, wenn auch nicht ganz so eindeutiges Bild (Abbildung 3). Die Struktur scheint in diesem Fall für die verschiedenen Verfahren leicht unterschiedlich zu sein, allerdings liegen auch in dieser Grafik dieselben Produkte jeweils dicht beieinander. Insbesondere sollten wir hierbei berücksichtigen, dass nur die Bewertungen von drei Prüfpersonen vorliegen, d. h. von einer ungewöhnlich kleinen Zahl, bei der Zufallsschwankungen wesentlich stärker ins Gewicht fallen. Festzustellen bleibt aber auch hier, dass wir in allen vier Fällen zu vergleichbaren Interpretationen kommen und die Wahl des Verfahrens die Interpretation gar nicht oder nur wenig beeinflusst.



**Abbildung 3:** Grafische Darstellung der Konsense der verschiedenen Methoden für den Datensatz von Gower (1975) (1 = Originalversion von STATIS, 2 = Version mit arithmetischem Mittel, 4 = Version mit Korrektur, 5 = GPA).

Insgesamt können wir also festhalten, dass die Verfahren – trotz der theoretisch und in den Simulationen gefundenen Unterschiede – so ähnliche Konsense liefern, dass die daraus resultierenden Interpretationen in praktischen Anwendungen identisch sind.



## 9 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurden vor allem die statistischen Eigenschaften der STATIS-Methode untersucht. Aufgrund der Abhängigkeit des einer Prüfperson zugewiesenen Gewichts von den zufälligen Fehlern erfolgte dabei die Untersuchung der Eigenschaften nur unter zusätzlichen einschränkenden Annahmen. In allen Fällen wurde dabei gezeigt, dass dieses Verfahren weder asymptotisch noch im Erwartungswert Resultate liefert, die den wahren, theoretisch zugrunde liegenden Konsens exakt widerspiegeln. Dieses kann als ein wesentlicher Grund dafür angesehen werden, dass STATIS beim Vergleich mit der GPA relativ schlecht abgeschnitten hat, vergleiche Meyners *et al.* (2000).

Um diese Problematik zumindest teilweise zu beheben, wurde eine Korrektur des STATIS-Konsens vorgeschlagen. Diese motiviert sich aus dem Erwartungswert der Assoziationsmatrizen, der nicht gleich dem entsprechenden Matrixprodukt des theoretischen Konsens ist. Unter der Annahme, dass wir die wahren Fehlermatrizen und die Skalare der einzelnen Prüfpersonen kennen, wurde gezeigt, dass solch eine Korrektur asymptotisch bzw. im Erwartungswert zu geringeren Abständen des berechneten vom theoretischen Konsens führt.

Die zur Korrektur notwendigen Skalare und Matrizen sind in der Anwendung unbekannt, daher wurde eine Methode vorgeschlagen, mit der wir diese Werte geeignet schätzen können. Dass diese Korrektur tatsächlich zu einer Verbesserung der Ergebnisse führt, wurde anhand einer Simulationsstudie gezeigt, bei der dieses Verfahren bessere Schätzer liefert als die Originalversion.

Des Weiteren wurde eine vereinfachte Version von STATIS betrachtet, bei der statt eines gewichteten das arithmetische Mittel verwendet wird, was sich aber in den Simulationen als Verschlechterung herausstellte.

Schließlich wurde auch die Version betrachtet, bei der die asymptotischen Gewichte verwendet werden, so wie sie in dieser Arbeit bestimmt wurden. Letzteres führt im Allgemeinen zu guten Ergebnissen, allerdings ist dieses Verfahren in der Praxis nicht anwendbar, da die Gewichte unbekannt sind.

In einer weiteren Simulationsstudie wurde dann gezeigt, dass die korrigierte Version von STATIS mindestens ebenso gute Schätzer liefert wie die GPA. Sie ist gleichzeitig aber leichter zu berechnen, da diese Methode keinen iterativen Algorithmus erfordert.

Abschließend wurden die Ergebnisse aller untersuchten Verfahren grafisch miteinander verglichen. Dabei wurde zunächst an einem fiktiven Beispiel gezeigt, dass es durch die Verwendung von STATIS tatsächlich zu einer Überbewertung niedriger Dimensionen kommt, und zwar stärker als beispielsweise durch die GPA. Dieser Effekt kann jedoch durch die vorgeschlagene Korrektur verringert werden. An zwei Datensätzen aus der Literatur wurde dann deutlich, dass die Unterschiede in den Ergebnissen hinsichtlich der daraus resultierenden Interpretation sehr gering sind. Die zugehörigen grafischen Darstellungen sind sich sehr ähnlich, und insbesondere wären Aussagen darüber, welche Produkte sich stärker voneinander unterscheiden als andere, bei allen Verfahren identisch getroffen worden.

Zu erwähnen ist an dieser Stelle, dass sämtliche Untersuchungen für den Fall durchgeführt wurden, dass die Variablen für die Profilprüfung fest vorgegeben sind bzw. dass alle Prüfpersonen zumindest die gleiche Anzahl an Variablen verwenden. Häufig wird jedoch auch das so genannte Free-Choice-Profiling (FCP) verwendet, welches jeder Prüfperson erlaubt, die Merkmale selbst festzulegen, anhand derer sie die Produkte beurteilen will. Dabei sind insbesondere auch Restriktionen bezüglich

der Anzahl der Variablen unüblich. In diesen Fällen können wir die Beobachtungsmatrizen so lange mit Spalten aus Nullen auffüllen, bis alle Matrizen die gleiche Dimension haben. Allerdings muss das für STATIS im Gegensatz zur GPA nicht explizit durchgeführt werden, da die Assoziationsmatrizen durch diese Spalten nicht beeinflusst werden. Die theoretische Untersuchung der Eigenschaften wird dadurch in der Notation jedoch unübersichtlicher, so dass hier auf diese verallgemeinerte Darstellung verzichtet wurde.

Offen bleibt in diesem Zusammenhang das Verhalten der Verfahren unter anderen Modellannahmen. Eine nahe liegende Erweiterung des von uns verwendeten Modells erhalten wir, indem wir die Annahme fallen lassen, dass jede Prüfperson denselben Skalar für alle Variablen verwendet. Stattdessen könnten wir annehmen, dass unterschiedliche Skalare für die verschiedenen Variablen benutzt werden. Allerdings können diese Faktoren mit keinem der hier betrachteten Verfahren geschätzt werden. Es ist unmöglich zu entscheiden, inwieweit eine große genutzte Spannweite auf einen entsprechend großen Skalar zurückzuführen ist oder ob stattdessen Unterschiede im Geschmack für die spezielle Prüfperson dafür verantwortlich sind. Sicherlich könnten wir als Erweiterung dieser Arbeit unter einer entsprechend veränderten Modellannahme die gleichen Untersuchungen und insbesondere eine Simulationsstudie durchführen, in der die Beobachtungsmatrizen aus einem derartigen Modell erzeugt werden.

Zusammenfassend können wir als Ergebnis dieser Arbeit festhalten, dass es geringe Unterschiede zwischen den einzelnen Verfahren gibt und mit der korrigierten Version von STATIS eine Alternative vorgeschlagen wurde, die zu den besten Ergebnissen unter den untersuchten Verfahren führt. Da die Ergebnisse der Verfahren sich hinsichtlich ihrer Interpretation nicht oder nur sehr wenig voneinander unterscheiden, empfehlen wir, in praktischen Anwendungen jenes Verfahren zu benutzen, das unter den gegebenen Umständen am einfachsten und am schnellsten anzuwenden ist.

Ein Kriterium für diese Entscheidung ist sicherlich das Vorhandensein entsprechender Programme. Falls keine anderen Programme vorliegen, ist die Originalversion von STATIS eine recht einfach zu programmierende Methode, die sich in den Ergebnissen nicht nennenswert von den übrigen Verfahren unterscheidet. Selbst die Verwendung des arithmetischen statt eines gewichteten Mittels führt zu brauchbaren Resultaten und kann daher zur Vereinfachung der Programmierung durchaus in Betracht gezogen werden.

## Literaturverzeichnis

- Dijksterhuis, G. B. und Gower, J. C. (1991)** *The interpretation of generalized procrustes analysis and allied methods*. Food Quality and Preference 3, 67-87.
- Gabriel, K. R. (1971)** *The biplot - graphical display of matrices with application to principal component analysis*. Biometrika 58, 453-467.
- Gower, J. C. (1975)** *Generalized Procrustes Analysis*. Psychometrika 40, 33-51.
- Harville, D. A. (1997)** *Matrix algebra from a statistician's perspective*. Springer, New York.
- Horn, R. A. und Johnson, C. R. (1987)** *Matrix Analysis*. Cambridge University Press, Cambridge.
- Langron, S. P. und Collins, A. J. (1985)** *Perturbation theory for generalized procrustes analysis*. Journal of the Royal Statistical Society Series B 47, 277-284.
- Lavit, C., Escoufier, Y., Sabatier, R. und Traissac, P. (1994)** *The ACT (STATIS method)*. Computational Statistics and Data Analysis 8, 97-119.
- Meyners, M. (1997)** *Vergleich zwischen Verallgemeinerter Prokrustes Analyse und STATIS*. Diplomarbeit, Fachbereich Statistik, Universität Dortmund.
- Meyners, M., Kunert, J. und Qannari, E. M. (2000)** *Comparing Generalized Procrustes Analysis and STATIS*. Food Quality and Preference 11, 77-83.

**Minc, H. (1988)** *Nonnegative Matrices*. Wiley, New York.

**Schlich, P. (1996)** *Defining and validating assessor compromises about product distances and attribute correlations*. In: T. Næs und E. Risvik (Hrsg.), *Multi-variate analysis of data in sensory science*, 259-306.

**Schönemann, P. H. (1966)** *A generalized solution of the orthogonal procrustes problem*. *Psychometrika* 31, 1-10.

**ten Berge, J. M. F. (1977)** *Orthogonal procrustes rotation for two or more matrices*. *Psychometrika* 42, 267-276.