

Simulationsuntersuchungen zur Hypothesenprüfung über feste Effekte bei verschiedenen Freiheitsgradapproximationen in gemischten linearen Modellen bei kleinen und unbalancierten Stichproben

J. Spilke¹, H. P. Piepho², X. Hu¹

1) Martin Luther Universität Halle-Wittenberg, Arbeitsgruppe Biometrie und Agrarinformatik
06108 Halle

2) Universität Gesamthochschule Kassel, Institut für Nutzpflanzenkunde
37213 Witzenhausen

Abstract

Für gemischte lineare Modelle (Blockanlage, Spaltanlage, Streifenanlage) mit unbalancierten Daten werden verschiedene Auswertungsmodelle (Modell I, gemischtes Modell) und innerhalb des gemischten Modells verschieden in SAS (8.1) verfügbare Approximationen der Freiheitsgrade untersucht. Die Nutzung der Vorgehensweise nach Kenward und Roger (DDFM=kenwardroger) führt zur besten Einhaltung des vorgegebenen statistischen Risikos 1. Art.

1. Einführung und Problemstellung

Die überwiegende Anzahl in landwirtschaftlichen Versuchen oder Erhebungen gewonnener Beobachtungsdaten können als Realisationen \underline{y} einer normalverteilten Zufallsvariable \underline{y} angesehen werden, die durch ein gemischtes lineares Modell zu beschreiben sind. Wenngleich überwiegend Effektschätzung und Hypothesenprüfung über die festen Effekte Ziel der Untersuchungen sind, führt eine realistische Modellierung meist zwangsläufig zur zusätzlichen Berücksichtigung zufälliger Effekte im Sinne von Störgrößen.

Für die Modellierung der Realisationen einer solchen Zufallsvariable bedient man sich meist der folgenden Schreibweise:

$$\underline{y} = \mathbf{X}\underline{\beta} + \mathbf{Z}\underline{u} + \underline{e} \quad (1)$$

mit

- $\underline{\beta}$ = $p \times 1$ Vektor der unbekanntesten festen Effekte
- \underline{u} = $q \times 1$ Vektor der unbekanntesten zufälligen Effekte
- \underline{e} = $n \times 1$ Vektor der unbekanntesten zufälligen Resteffekte
- \mathbf{X} = bekannte $n \times p$ Versuchsplanmatrix der festen Effekte
- \mathbf{Z} = bekannte $n \times q$ Versuchsplanmatrix der zufälligen Effekte

Weiter soll gelten:

$$\underline{u} \sim N(0, \mathbf{G}), \underline{e} \sim N(0, \mathbf{R})$$

$$E(\underline{y}) = \mathbf{X}\underline{\beta}, \text{Var}(\underline{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \mathbf{V}$$

$$\text{Var} \begin{pmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{pmatrix} = \begin{bmatrix} \mathbf{V} & \mathbf{Z}\mathbf{G} & \mathbf{R} \\ \mathbf{G}\mathbf{Z}' & \mathbf{G} & 0 \\ \mathbf{R} & 0 & \mathbf{R} \end{bmatrix}$$

(HENDERSON 1963, 1975, 1990 p. 1ff).

Falls \mathbf{G} und \mathbf{R} bekannt sind, lassen sich „Best Linear Unbiased Estimation“ (BLUE) (d.h., lineare erwartungstreue Schätzungen mit kleinster Varianz) für die festen Effekte in (1) aus

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (2)$$

gewinnen.

Ein zentrales Problem besteht aber darin, dass die vorteilhaften Eigenschaften der Schätzungen fester Effekte bzw. Vorhersagen zufälliger Effekte, unverzerzte lineare Schätzungen bzw. Vorhersagen mit kleinster Varianz (BLUE bzw. BLUP) zu sein, an bekannte Matrizen \mathbf{G} und \mathbf{R} gebunden sind (HENDERSON, 1963).

Von großer Bedeutung für die praktische Nutzung ist, dass sich nur bei bekannten \mathbf{G} und \mathbf{R} die Fehlervarianzen der festen Effekte exakt angeben lassen und exakte statistische Tests über die festen Effekte existieren (HENDERSON, 1984, p. 83).

In dieser Arbeit soll der für Anwendungen stets vorliegende Fall unbekannter Varianzkomponenten und die daraus folgenden Konsequenzen für die Prüfung von Hypothesen über feste Effekte bei Beschränkung auf Hypothesen vom Rang=1 mit Hilfe von stochastischen Simulationen untersucht werden. Dabei wird der praktisch ebenfalls bedeutsame Sachverhalt unbalancierter Datenstrukturen und geringer Stichprobenumfänge betrachtet.

2. Konsequenzen unbekannter Matrizen \mathbf{G} und \mathbf{R}

Wie bereits angedeutet, sind im Anwendungsfall die Matrizen \mathbf{G} und \mathbf{R} stets unbekannt. Daraus folgt für die Eigenschaften der Schätzungen fester Effekte:

- Statt BLUE können nur noch eBLUE (empirical BLUE) bereitgestellt werden.
- Die Eigenschaft, „beste“ Schätzung im Sinne der kleinsten Varianz zu sein, geht ebenfalls verloren.
- Falls Varianzkomponenten und Effekte am selben Material geschätzt werden, geht die Eigenschaft der Linearität ebenfalls verloren.
- Die Eigenschaft der Erwartungstreue bleibt jedoch erhalten.

Eine weitere Konsequenz unbekannter Varianzmatrizen ist, dass die Standardfehler der Linearkombination von festen Effekten im allgemeinen eine Funktion mehrerer geschätzter Varianzkomponenten sind und eine Bestimmung der dazugehörigen Freiheitsgrade erforderlich ist. Das findet seinen Ausdruck in der Nutzung mehrerer Konzepte, worauf noch näher eingegangen wird.

Bei der Nutzung gemischter linearer Modelle ist weiterhin zu beachten, dass im Anwendungsfall häufig unbalancierte Daten vorliegen (vgl. Abschnitt 3). Unter diesen Bedingungen verlieren ANOVA-Schätzungen für Varianzkomponenten ihre vorteilhaften Eigenschaften (Erwartungstreue, kleinste Varianz unter allen quadratischen Schätzfunktionen und bei Normalverteilung kleinste Varianz unter allen erwartungstreuen Schätzfunktionen) mit Ausnahme der Erwartungstreue (AHRENS, 1967; SEARLE, 1971). Für unbalancierte gemischte Modelle gewinnen daher seit längerem ausgearbeitete Konzepte [ML bzw. REML (HARTLEY und RAO; 1967; PATTERSON und THOMPSON, 1971); MINQUE bzw. MIVQUE (RAO, 1971; LAMOTTE, 1973)] an Bedeutung. Als Vorteil für REML geben SEARLE et al. (1992, p. 254) die unter Normalverteilung nahezu optimalen Eigenschaften der REML-Schätzer an (asymptotische Konsistenz, Erwartungstreue sowie asymptotische Normalverteilung der Schätzwerte und Möglichkeit der Angabe asymptotischer Fehlervarianzen der Schätzungen). Ein weiterer Grund für die Bevorzugung von REML gegenüber ML ist die Übereinstimmung von ANOVA und REML im balancierten Fall, falls keine negativen ANOVA-Schätzungen auftreten (SEARLE et al., 1992). Ebenso ist es ein

wichtiger Vorteil, für REML-Schätzungen auch für unbalancierte gemischte Modelle den Fehler der geschätzten Varianzkomponenten angeben zu können. Dieser Sachverhalt wird in Approximationen der Freiheitsgrade der Testgröße ausgenutzt (FAI und CORNELIUS, 1996; KENWARD und ROGER, 1997). In dieser Arbeit werden ausschließlich REML-Schätzer für die Varianzkomponenten verwendet, da auf dem ANOVA-Ansatz basierende Verfahren im gemischten unbalancierten Modell zu einer schlechteren Einhaltung des statistischen Risikos 1. Art führen (SPILKE und TUCHSCHERER, 2001).

Für unbalancierte Daten ist weiter wichtig, dass die Schätzfehler fester Effekte bzw. von Differenzen fester Effekte unterschätzt werden (HENDERSON, 1984; KACKAR und HARVILLE, 1984).

Zusätzlich ist für kleine Stichproben von Bedeutung, dass aus wenig Informationen eine vergleichsweise hohe Anzahl von Parametern (feste Effekte, Varianzkomponenten) zu schätzen ist und eine besonders deutliche Unterschätzung des Standardfehlers geschätzter Linearkombinationen fester Effekte zu erwarten ist.

3. Material und Methoden

3.1. Untersuchte Versuchspläne

In die Untersuchungen wurden die folgenden zweifaktoriellen Modelle einbezogen.

Blockanlage:

$$\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \underline{b}_k + \underline{e}_{ijk} \quad (4)$$

Spaltanlage:

$$\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \underline{b}_k + \underline{e}_{ik} + \underline{e}_{ijk} \quad (5)$$

Streifenanlage:

$$\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \underline{b}_k + \underline{e}_{ik} + \underline{e}_{jk} + \underline{e}_{ijk} \quad (6)$$

($i=1, \dots, a$; $j=1, \dots, b$; $k=1, \dots, r$;

$\underline{b}_k \sim N(0, \sigma_{BL}^2)$; $\underline{e}_{ik} \sim N(0, \sigma_{RA}^2)$; $\underline{e}_{jk} \sim N(0, \sigma_{RB}^2)$; $\underline{e}_{jik} \sim N(0, \sigma_{RAB}^2)$)

Dabei bedeuten:

α_i = Effekt der i -ten Stufe des Faktors A (fix),

β_j = Effekt der j -ten Stufe des Faktors B (fix),

$(\alpha\beta)_{ij}$ = Wechselwirkungseffekt zwischen der i -ten Stufe des Faktors A und der j -ten Stufe des Faktors B (fix),

\underline{e}_{ik} ; \underline{e}_{jk} ; \underline{e}_{ijk} = zufällige Effekte.

In Abhängigkeit der unterschiedlichen Randomisationsstruktur ergeben sich für die untersuchten Versuchsanlagen verschiedene zu berücksichtigende zufällige Effekte und entsprechend zu schätzenden Varianzkomponenten.

Die angeführten Versuchspläne sind balancierte Pläne. Dabei wird unter Balanciertheit der Sachverhalt verstanden, dass alle Felder der Klassifikation die selbe Besetzungszahl haben (SEARLE et al., 1992, S.4). Die Werte, die der Index eines Faktors annimmt, sind unabhängig vom Wert der Indizes der übrigen Faktoren. Bezeichnet man diese Form der Balanciertheit als „klassifikationsbalanciert“, so gelten in diesem strengen Sinne auch balancierte unvollständige Blockanlagen (BUB) oder Lateinische Quadrate (LQ) als unbalanciert, obwohl eine Varianzbalanciertheit dahingehend vorliegt, dass jeder paarweise Kontrast von festen Effekten mit derselben Genauigkeit gemessen werden kann (PIEPHO und SPILKE, 1999).

Klassifikationsbalancierte Versuchspläne sind stets auch varianzbalanciert, was umgekehrt nicht gilt.

Im vorliegenden Untersuchungen soll der Fall behandelt werden, dass aus einer balancierten Versuchsanlage (Block-, Spalt- und Streifenanlage) durch Ausfall von Beobachtungen ein Datensatz entsteht, der dann entsprechend obiger Einteilung unbalanciert ist sowohl bezüglich der Klassifikation als auch der Varianz der Differenz fester Effekte.

3.2. Simulationsvarianten

Es wurden zwei Fälle bezüglich der Zahl der Stufen und Wiederholungen untersucht: $a=2, b=3, r=3$ und $a=3, b=4, r=4$ und Wiederholungsanzahl ($r=3$ bzw. $r=4$). Daraus resultieren für den balancierten Fall $N=18$ bzw. $N=48$ Beobachtungen. Die unbalancierten Datenstrukturen werden durch den Ausfall von 1 bzw. 2 zufällig bestimmten Beobachtungen erzeugt.

Weiterhin wurden unterschiedliche Größenordnungen der Varianzkomponenten im Verhältnis zur Restvarianz (100%, 50%, 10%) in die Untersuchungen einbezogen.

Für jede Simulationsvariante wurden 100000 Datensätze erzeugt und ausgewertet. Diese Anzahl ergibt sich aus einer auf dem Konfidenzintervall einer Wahrscheinlichkeit basierenden Versuchsplanung und sichert eine Intervallbreite von 5% des Parameters bei einem Konfidenzniveau von 0.95. Der unterstellte Wahrscheinlichkeitsparameter wurde im Hinblick auf die gewählte Irrtumswahrscheinlichkeit ($\alpha=0.05$) ebenfalls mit 0.05 gewählt.

Zur Simulation und Auswertung (PROC MIXED) wurde das Programmpaket SAS (Ver. 8.1) benutzt.

3.3. Untersuchte Modelle und Approximationen der Freiheitsgrade

Die Auswertung der als gemischtes Modell gemäß (4) bis (6) simulierten Daten erfolgte nach:

- **Modell I**, d.h., alle zufälligen Effekte außer dem Rest werden als zufällig angesehen. Diese Vorgehensweise geht auf einen Vorschlag von Henderson zurück (HENDERSON und HENDERSON, 1979; HENDERSON, 1984).
- **Gemischtes Modell** bei Nutzung verschiedener Approximationen der Freiheitsgrade.

Die Konsequenzen bei Anwendung von Modell I für die Auswertung sind:

- Effektschätzung aus $(\mathbf{W}'\mathbf{W}) \hat{\beta} = \mathbf{W}'\mathbf{y}$
wobei $\mathbf{W}=(\mathbf{X},\mathbf{Z})$ (\mathbf{X} und \mathbf{Z} entsprechen den Matrizen in (1))
- Schätzung ohne Berücksichtigung der Varianzkomponenten
- Ignorierung des Beitrages der zufälligen Effekte außer dem Resteffekt e_{ijk} am Schätzfehler der festen Effekte
- Umgehung der Schätzung von Varianzkomponenten außer für die Restvarianz σ_{RAB}^2
- Freiheitsgrade entsprechen den Restfreiheitsgraden $FG = N - \text{Rang}(\mathbf{W})$

Die Anwendung eines gemischten Modells für die Auswertung führt zur Nutzung von (2) für die Effektschätzung. Entsprechend gehen über $\hat{\mathbf{V}}^{-1}$ die geschätzten Varianzkomponenten in die Schätzung der festen Effekte ein.

Für die Approximation der Freiheitsgrade der Prüfgröße $t = \frac{h'\hat{\beta}}{\sqrt{h'(X'\hat{\mathbf{V}}^{-1}X)^{-1}h}}$ zum Test der

Nullhypothese $h'\beta = 0$ (h' = Zeilenvektor zur Beschreibung der Linearkombination der festen Effekte) werden verschiedene Freiheitsgrade genutzt:

Freiheitsgrade nach der Residual-Methode (DDFM=residual)

- Die Freiheitsgrade aller festen Effekte werden aus dem Stichprobenumfang und dem Rang der Versuchsplanmatrix der festen Effekte bestimmt ($FG=N-\text{Rang}(X)$).

- Die Freiheitsgrade werden damit unabhängig von der formulierten Linearkombination der festen Effekte berechnet.

Freiheitsgrade nach der Containment-Methode (DDFM=contain)

- Die Freiheitsgrade des betrachteten festen Effektes werden aus dem Rangbeitrag der zufälligen Effekte ermittelt, in denen der feste Effekt „enthalten“ ist.
- Für feste Effekte, die in keinem zufälligen Effekt enthalten sind, werden die Freiheitsgrade bei Nutzung des Statement „LSMEANS“ nach $FG=N-Rang(\mathbf{X},\mathbf{Z})$ ermittelt.
- Ein an die SAS-Notation angelehntes Beispiel soll das verdeutlichen:

MODEL y = a b;
 RANDOM c(a) d;

Dann folgt:

$$FG_a = Rang(\mathbf{X},\mathbf{Z}_c,\mathbf{Z}_d) - Rang(\mathbf{X},\mathbf{Z}_d)$$

$$FG_b = N - Rang(\mathbf{X},\mathbf{Z}_c,\mathbf{Z}_d)$$

- Abweichend von dem beschriebenen Containment-Prinzip werden die Freiheitsgrade bei Nutzung des Statement „ESTIMATE“ durchgehend aus $FG=N-Rang(\mathbf{X},\mathbf{Z})$ gebildet!

Freiheitsgrade nach Satterthwaite (DDFM=satterth)

- Die Bestimmung der Freiheitsgrade basiert auf SATTERTHWAITE (1941) und GIESBRECHT und BURNS (1985).
- Die Basisidee ist die Ausnutzung der χ^2 -Verteilung des Quotienten einer Varianzschätzung und ihres Parameters. Bei Nutzung von REML-Schätzern für die Varianzkomponenten gilt diese Verteilungsvoraussetzung nur approximativ. Diese Annahme kann für kleine Stichproben besonders bedeutsam sein.
- Die Berechnung der Freiheitsgrade erfolgt nach

$$FG = \frac{2 \left(\sum_p k_p \cdot \hat{\sigma}_p^2 \right)^2}{\sum_p \sum_q k_p k_q \cdot Cov(\hat{\sigma}_p^2, \hat{\sigma}_q^2)} \tag{7}$$

mit: $k_p = h' (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} Z_p Z_p' \hat{V}^{-1} X (X' \hat{V}^{-1} X)^{-1} h$

(GIESBRECHT und BURNS, 1985; VERBEKE und MOLENBERGHS, 1997)

- Der Ausdruck in den Klammern des Zählers in (7) ist eine Approximation des Erwartungswertes der Varianz der betrachteten Linearkombination fester Effekte. Der Nenner ist eine Approximation der Varianz dieser Varianz. Dabei ist $Cov(\hat{\sigma}_p^2, \hat{\sigma}_q^2)$ die asymptotische (Co)Varianzmatrix der geschätzten Varianzkomponenten.
- Die Bestimmung der Freiheitsgrade gemäß (7) beinhaltet sowohl die Schätzwerte der Varianzkomponenten als auch deren geschätzte (Co)Varianz.

Freiheitsgrade nach Kenward und Roger (DDFM=kenwardroger)

- Diese Approximation basiert grundsätzlich ebenfalls auf dem Ansatz von SATTERTHWAITE (1941). Die Erweiterung besteht in der Korrektur der Verzerrung der geschätzten (Co)Varianzmatrix der festen Effekte $(X' \hat{V}^{-1} X)^{-1}$ (KACKAR und HARVILLE, 1984; KENWARD und ROGER, 1997).
- Die Korrektur wirkt sowohl auf die Fehlervarianz $(h' (X' \hat{V}^{-1} X)^{-1} h)$ der geschätzten Linearkombination der festen Effekte als auch die Approximation der Freiheitsgrade gemäß (7).

4. Ergebnisse

4.1. Wirksamkeit der Korrektur der (Co)Varianzmatrix der festen Effekte

Zunächst soll untersucht werden, wie wirksam die von KACKAR und HARVILLE (1984) vorgeschlagene und bei KENWARD und ROGER (1997) zusammengefasst dargestellte Korrektur der (Co)Varianzmatrix der festen Effekte für die untersuchten Datenstrukturen und Varianzverhältnisse ist. Im weiteren wird diese Vorgehensweise kurz als Kenward-Roger bezeichnet.

Bei Beachtung der im Abschnitt 3.3 formulierten Prüfgröße beziehen sich die weiteren Darstellungen auf den Standardfehler, da dieser in die Testgröße einfließt. Bezugspunkt für die Untersuchung der Wirksamkeit ist der „beobachtete Standardfehler“, berechnet aus den 1000000 Schätzungen des jeweiligen Contrastes je Simulationsvariante. Dieser Standardfehler beruht auf der Nutzung der geschätzten Varianzkomponenten und entspricht damit dem praktisch vorliegenden Sachverhalt. Diesem Bezugspunkt wird der mittlere geschätzte Standardfehler, berechnet aus dem geschätzten Standardfehler je Simulationslauf, ohne und mit Korrektur gegenübergestellt (Abbildung 1-4). Die Darstellungen weisen die Unterschätzung der Standardfehler unterhalb der Bezugslinie aus, die Korrektur dieser Unterschätzung wird aus Übersichtlichkeitsgründen oberhalb der Bezugslinie dargestellt.

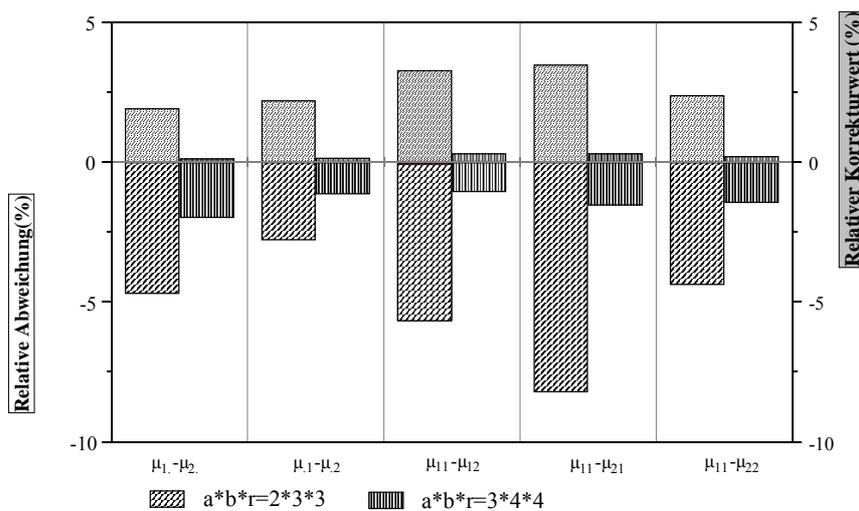


Abbildung 1: Unterschätzung der Standardfehler und Korrektur nach Kenward-Roger für verschiedene Datenstrukturen (Streifenanlage, Varianzverhältnis: 1:1:1:1, zwei fehlende Beobachtungen)

Die in Abbildung 1 dargestellten Ergebnisse zeigen eine deutliche Unterschätzung der Standardfehler bei gleichzeitig deutlicher Abhängigkeit vom betrachteten Kontrast und Datenumfang. Die Korrektur ist wirksam, jedoch nie vollständig. Das gilt auch bei einer Darstellung in Abhängigkeit der Versuchsanlage (Abbildung 2) und Unbalanciertheit (Abbildung 3). Der Vergleich der Versuchsanlagen verdeutlicht, dass für die Blockanlage mit der geringsten Verschätzung zu rechnen ist. Mit zunehmender Unbalanciertheit nimmt die Unterschätzung zu. Die Korrektur wirkt in der Weise, dass nach Korrektur das Ausmaß der Unterschätzung für beide Unbalanciertheiten etwa gleich ist.

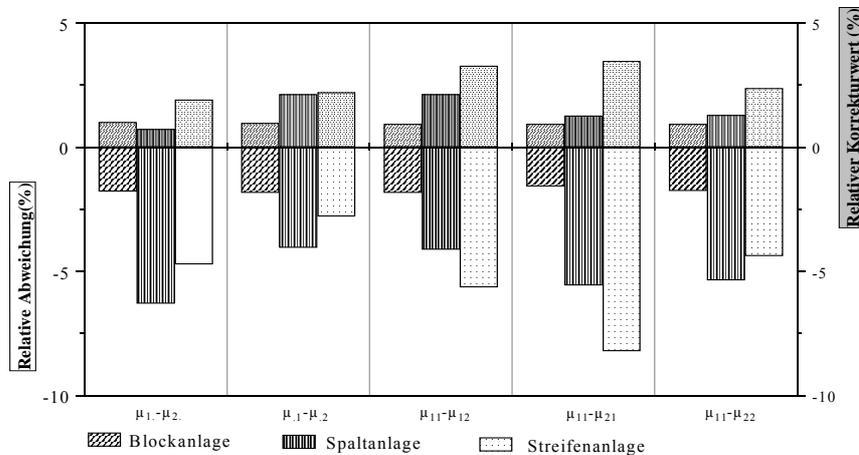


Abbildung 2: Unterschätzung der Standardfehler und Korrektur nach Kenward-Roger für verschiedene Versuchsanlagen (2*3*3, Varianzverhältnis: 1:1:1:1, zwei fehlende Beobachtungen)

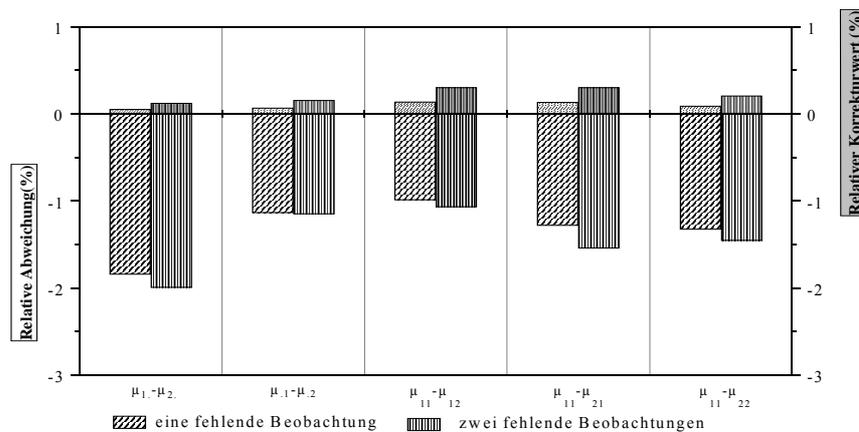


Abbildung 3: Unterschätzung des Standardfehlers und Korrektur nach Kenward-Roger für verschiedene Unbalanciertheiten (Streifenanlage, 3*4*4, Varianzverhältnis: 1:1:1:1)

4.2. Einhaltung des statistischen Risikos 1. Art

Von besonderem Interesse ist, inwiefern bei Nutzung der verschiedenen Modelle für die Auswertung (Modell I, gemischtes Modell) und Approximationen der Freiheitsgrade das vorgegebene statistische Risiko 1. Art eingehalten wird. Für alle durchgeführten Tests wurde ein statistisches Risiko von $\alpha=0.05$ vorgegeben.

Die in Abbildung 4 dargestellten Ergebnisse für einen geringen Stichprobenumfang (N=16) zeigen, dass im Fall der Blockanlage die Nutzung von Modell I (Intra-Block-Analyse) auch dann zur exakten Einhaltung des vorgegebenen statistischen Risikos führt, wenn der Faktor Block entsprechend seiner Randomisationsstruktur als zufällig anzusehen ist. Das beobachtete Risiko weicht nur zufällig vom vorgegebenen Risiko ab. Demgegenüber wirken bei Nutzung der Freiheitsgradoption „Residual“ die entsprechend der Berechnungsvorschrift hohe Anzahl an Freiheitsgraden N-Rang(X). Der Unterschied zwischen der Vorgehensweise

nach Satterthwaite und Kenward-Roger folgt aus der Korrektur der Fehlervarianzen, wobei auch bei Korrektur noch eine geringe Überschätzung zu beobachten ist, die etwa das gleiche Ausmaß aufweist wie für Contain. Bei dieser Option der Freiheitsgrade beziehen sich die Ergebnisse durchgehend auf die bei Nutzung des LSMEANS-Statement erhaltenen Ergebnisse bei Anwendung der in Abschnitt 3.3 beschriebenen Fallunterscheidung.

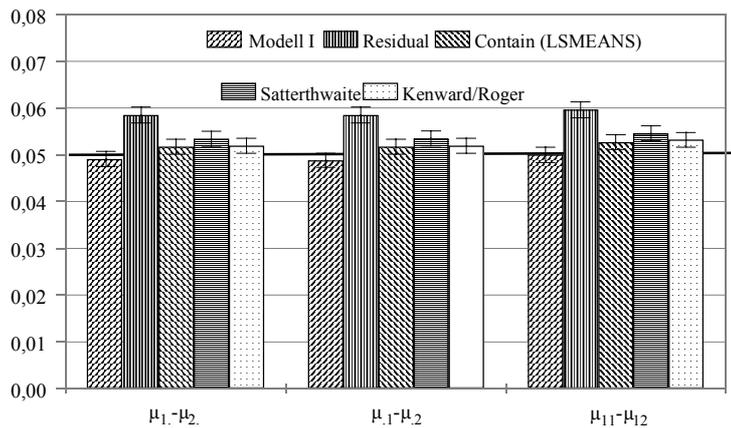


Abbildung 4: Realisierter Fehler 1. Art (vorgegeben $\alpha=0.05$) Blockanlage ($2 \times 3 \times 3$)
Varianzverhältnis 0.5:1, zwei fehlende Beobachtungen

Bedeutend deutlichere Verschätzungen sind für die Spalt- und Streifenanlage zu beobachten. Hier führen sowohl die Nutzung von Modell I als auch der Freiheitsgradapproximation „Residual“ und „Contain“ für die untersuchten Kontraste zu teilweise erheblichen Überschätzungen (Modell I und Residual) als auch Unterschätzungen (Contain). Für die Optionen Satterthwaite und Kenward-Roger werden durchgehend vergleichsweise geringe Abweichungen beobachtet, wobei Kenward-Roger stets etwas geringere Unterschiede zum vorgegebenen statistischen Risiko aufweist.

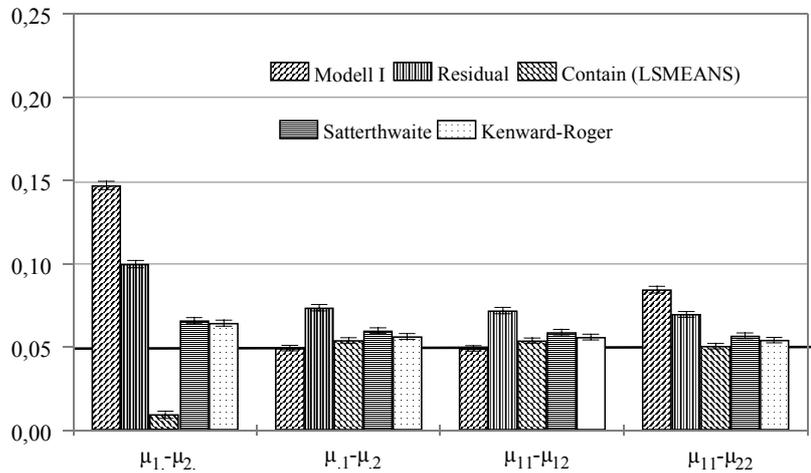


Abbildung 5: Realisierter Fehler 1. Art (vorgegeben $\alpha=0.05$) Spaltanlage ($2 \times 3 \times 3$)
 Varianzverhältnis 0.5:0.5:1, zwei fehlende Beobachtungen

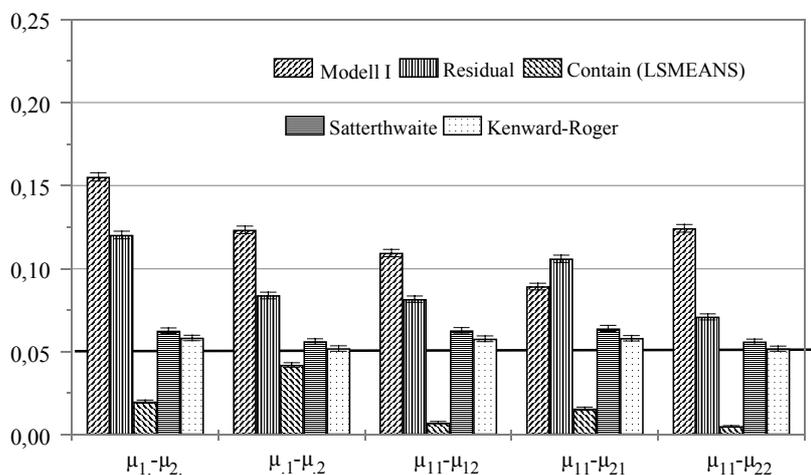


Abbildung 6: Realisierter Fehler 1. Art (vorgegeben $\alpha=0.05$) Streifenanlage ($2 \times 3 \times 3$)
 Varianzverhältnis 0.5:0.5:0.5:1, zwei fehlende Beobachtungen

Die Einhaltung des vorgegebenen statistischen Risikos wird mit der Erhöhung der Anzahl der einbezogenen Stufen der festen Prüffaktoren und Anzahl Wiederholungen (Datenstruktur $3 \times 4 \times 4$) für die Freiheitsgradapproximationen Satterthwaite und Kenward-Roger verbessert. Das gilt für die in Abbildung 7 dargestellte Streifenanlage, aber auch für die nicht dargestellte Block- und Spaltanlage. Demgegenüber sind sowohl für Modell I als auch die Freiheitsgradapproximationen entsprechend „Residual“ und „Contain“ auch bei dem größeren Versuchsumfang teilweise noch erhebliche Abweichungen zu beobachten.

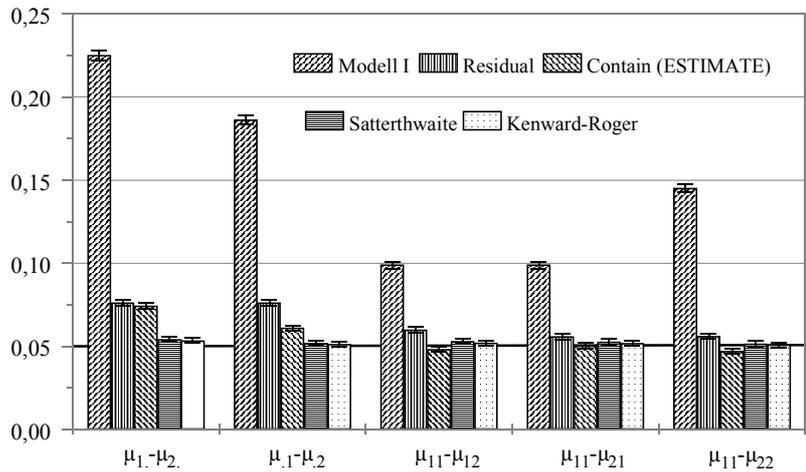


Abbildung 7: Realisierter Fehler 1. Art (vorgegeben $\alpha=0.05$) Streifenanlage ($3 \times 4 \times 4$)
 Varianzverhältnis 0.5:0.5:0.5:1, zwei fehlende Beobachtungen

5. Schlussfolgerungen

Die Hypothesenprüfung über eine Linearkombination fester Effekte im gemischten Modell ist für praktische Anwendungen im allgemeinen mit dem Fall verbunden, dass die durch die zufälligen Effekte hervorgerufenen Varianzkomponenten unbekannt sind.

Falls gleichzeitig ein unbalancierter Datensatz auszuwerten ist, wird bei Nutzung der geschätzten Varianzkomponenten in \hat{V}^{-1} zur Ermittlung der Schätzfehler des betrachteten Kontrasts $[h'(X'\hat{V}^{-1}X)^{-1}h]$ der Schätzfehler des Kontrasts unterschätzt (KACKAR und HARVILLE, 1984). Die Wirksamkeit der Korrektur der Unterschätzung entsprechend der bei KENWARD und ROGER (1997) beschriebenen Vorgehensweise führt zu einer teilweise deutlichen Reduzierung der Unterschätzung. Die Korrektur wirkt um so deutlicher, je stärker die Unterschätzungen ist, ohne dass jedoch die Unterschätzung vollständig ausgeglichen wird.

Der systematische Vergleich zeigt, dass die Unterschätzung abhängig ist von

- Datenstruktur und Stichprobenumfang,
- Ausmaß der Unbalanciertheit,
- Verhältnis der Varianzkomponenten zur Restvarianz,
- Versuchsanlage,
- betrachteten Kontrast.

Die Nutzung der Korrektur in Verbindung mit der Freiheitsgradapproximation nach SATTERTWAITE (1941) führt insgesamt zur besten Einhaltung des statistischen Risikos 1. Art, wobei ebenso eine Abhängigkeit von den bereits genannten Einflussfaktoren besteht.

Für die Datenstruktur $3 \times 4 \times 4$ ist nur noch eine praktisch kaum bedeutsame Unterschätzung der Standardfehler zu beobachten.

Die Nutzung von Modell I führt bei den untersuchten Datenstrukturen zu teilweise erheblichen Verschätzungen. Das gilt auch für die Freiheitsgradapproximation nach den Methoden Residual und Contain.

Wie in Abschnitt 3.3 beschrieben, werden bei Anwendung der Vorgehensweise nach Satterthwaite Approximationen genutzt. Das betrifft

- die χ^2 -Verteilung des Quotienten von REML-Schätzern und ihres Erwartungswertes,
- den Erwartungswert der Varianz des betrachteten Kontrastes (Zähler von (7)),
- die Varianz der Varianz des betrachteten Kontrastes (Nenner von (7)).

Die Vorgehensweise nach Kenward-Roger baut ebenfalls auf diesen Approximationen auf, indem für die Korrektur von \hat{V}^{-1} die bereits im Nenner von (7) enthaltene (Co)Varianzmatrix der geschätzten Varianzkomponenten verwendet wird. Trotz der insbesondere bei geringen Datenumfang nur näherungsweise einzuhaltenen Annahmen kann bei Nutzung dieser

Approximationen für die untersuchten Versuchspläne und Kontraste das statistische Risiko 1. Art vergleichsweise gut eingehalten werden.

Basierend auf unseren Simulationsergebnissen empfehlen wir die Nutzung der Vorgehensweise nach Kenward-Roger.

Literatur

- AHRENS, H., 1967: Die Varianzanalyse. Akademie-Verlag, Berlin
- FAI, A.H.T.; CORNELIUS, P.L., 1996: Approximate F-Tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *J. Statist. Comput. Simul.* **54**: 363-378.
- GIESBRECHT, F.G.; BURNS, J.C., 1985: Two-stage analysis based on a mixed model: large-sample asymptotic theory and small-sample simulation results. *Biometrics* **41**: 477 - 486.
- HARTLEY, H.O.; RAO, C.R., 1967: Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**: 93 - 108.
- HENDERSON, C.R., 1963: Selection index and expected genetic advance. *Statistical Genetics and Plant Breeding*, NAS-NRC Publ. No.982, 141 - 163.
- HENDERSON, C.R., 1975: Best linear estimation and prediction under a selection model. *Biometrics* **31**: 423 - 447.
- HENDERSON, C.R., 1984: Application of linear models in animal breeding. University of Guelph.
- HENDERSON, C.R., 1990: Statistical methods in animal improvement: historical overview. In: *Advances in Statistical Methods for Genetic Improvement of Livestock*. Springer Verlag. New York.
- HENDERSON, C.R. JR, HENDERSON, C.R., 1979: Analysis of covariance in mixed models with unequal subclass numbers. *Communications in Statistics, Series A* **8**: 751-787
- KACKAR, A.N.; HARVILLE, D.A., 1981: Unbiasedness of two-stage estimation and precision procedures for mixed linear models. *Communications in Statistics, Series A* **10**: 1249-1261.
- KENWARD, M.G.; ROGER, J.H., 1997: Small Sample Inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**: 983-997.
- LAMOTTE, L.R., 1973: Quadratic estimation of variance components. *Biometrics* **29**: 311-330.
- PATTERSON, H.D.; THOMPSON, R., 1971: Recovery of inter-block information when block size are unequal. *Biometrika* **58**: 545-554.
- PIEPHO, H.P.; SPILKE, J., 1999: Anmerkungen zur Analyse balancierter gemischter Modelle mit der SAS-Prozedur MIXED. *Zeitschrift für Agrarinformatik* **7**: 39-46
- RAO C.R., 1971: Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis* **1**: 445-456.
- SATTERTHWAITE, F.E., 1941: Synthesis of variance. *Psychometrika* **6**: 309-316.
- SEARLE, S.R.; CASELLA, G.; MCCULLOCH, C.E., 1992: Variance Components. John Wiley and Sons, Inc. New York.
- SEARLE, S.R., 1971.: Topics in variance component estimation. *Biometrics* **27**: 1 - 76.
- SPILKE, J.; TUCHSCHERER, A., 2001: Simulationsuntersuchungen zum Einfluss verschiedener Strategien der Varianzkomponentenschätzung und Hypothesenprüfung auf die statistischen Risiken in gemischten linearen Modellen mit ungleicher Klassenbesetzung. *Zeitschrift für Agrarinformatik* (eingereicht)
- VERBEKE, G.; MOLENBERGHS, G., 1997: Linear mixed models in practice – a SAS-oriented approach. Springer-Verlag, New York.