V. Guiard
Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere
Dummerstorf-Rostock

Erfahrungen mit und Vergleich von GLM und MIXED bei gemischten linearen Modellen

1. Anwendungsbereiche von GLM und MIXED

– Prozedur **GLM** (General Linear Model) liefert Auswertung der Varianzanalyse (auch Kovariablen zugelassen), falls Versuchsplan *nicht balanciert.*

– Bei *balanciertem* Versuchsplan und festen Faktoren (Modell I) sollte **ANOVA** verwendet werden, da *schneller*.

  Hier die üblichen Tests für Haupt- und Wechselwirkungen der Faktoren möglich

– Bei *nicht balancierten* Versuchsplänen sind unterschiedliche Nullhypothesen bezüglich der Faktorwirkungen testbar, d.h. verschiedene F-Tests, bzw. unterschiedliche SQ

– **GLM** vorrangig für Modell I (feste Faktoren), aber auch gemischtes Modell teilweise realisierbar

– Bei Model II (Faktoren zufällig) oder dem gemischten Modell ist **MIXED** günstiger und einfacher als **GLM**, aber nicht in jeder Beziehung.

2. In SAS verwendete Version des gemischten Modells:

Die Versionen des gemischten Modells sollen am folgenden Modellbeispiel erläutert werden:

Modellbeispiel:
Zweifache Varianzanalyse, Kreuzklassifikation
Faktor A mit den a Stufen A_i (i = 1, ..., a) sei *fest*
Faktor B mit den b Stufen B_j (i = 1, ..., b) sei *zufällig*

$$y_{ijk} = \eta_{ijk} + e_{ijk} \quad (\text{Vektordarstellung: } y = \eta + \varepsilon)$$

y_{ijk} = Beobachtungswert des k-ten ($k = 1, ..., n_{ij}$) Versuchsobjektes aus der Klasse (i,j) der mit A_i und B_j behandelten Objekte

n_{ij} = Klassenbesetzung

e_{ijk} = Zufallsfehler dieses Versuchsobjektes
$E(e_{ijk}) = 0$

$\eta_{ijk} = \mu + a_i + b_j + w_{ij}$, (Vektordarstellung: $\eta = X\beta$)
(ist also von Index k unabhängig!)

μ = allgemeines Mittel
a_i = Effekt der Stufe A_i des Faktors A
b_j = Effekt der Stufe B_j des Faktors B (*zufällig*)
w_{ij} = Effekt der Wechselwirkung A∗B zwischen A_i und B_j (*zufällig*)

Für w_{ij} (und analog auch für andere zufällige Wechselwirkungen) gilt:

In Modellversion 1:

Alle zufälligen Wechselwirkungen sind unabhängig!

In Modellversion 2:

Die Summe über den Index eines festen Faktors ergibt Null:

$$\sum_{i=1}^{a} w_{ij} = 0$$

Damit sind aber z.B. w_{ij} und $w_{i'j}$ voneinander abhängig:

$$\mathrm{cov}(w_{ij}, w_{i'j}) = \frac{-\sigma_{ab}^2}{a-1}$$

Entsprechungen zwischen beiden Modellversionen:

Version 1	Version 2
$b_j + \overline{w}_{.j}$	b_j
$w_{ij} - \overline{w}_{.j}$	w_{ij}
$\sigma_b^2 + \frac{1}{a}\sigma_{ab}^2$	σ_b^2
$\frac{a-1}{a}\sigma_{ab}^2$	σ_{ab}^2

- In SAS wird stets Version 1 verwendet.
- Für Schätzung fester Effekte spielt die Modellversion (fast) keine Rolle.
- Die E(MQ) sehen unterschiedlich aus; entsprechen aber einander, wenn man die unterschiedliche Definition der Varianzkomponenten beachtet.

3. Beispiel für weitere Demonstrationen:

Dreifache Kreuzklassifikation mit den Faktoren

A und B (fest; Effekte a_i bzw. b_j ; i=1,...,a=3; j=1,...,b=3)

und

C (zufällig; Effekte c_k ; k=1,...,c=2)

und den Wechselwirkungen

A*B (fest, da A und B fest; Effekte $(ab)_{ij}$)

und

A*C (zufällig, da einer der Faktoren (C) zufällig; Effekte $(ac)_{ik}$)

Klassenbesetzung $n_{ijk} = n = 2$

Gesamtumfang N=abcn=36

Daten:

	B_1		B_2		B_3	
	C_1	C_2	C_1	C_2	C_1	C_2
A_1	8.3 10.0	6.7 12.0	9.1 8.9	9.3 8.6	9.8 8.0	11.3 6.0
A_2	7.6 10.1	7.4 8.8	8.5 11.5	50.0 53.2	9.1 12.6	12.0 16.5
A_3	7.3 10.1	7.8 7.1	8.1 12.9	10.3 15.6	8.8 15.1	12.3 22.2

4. Welche unterschiedlichen F-Tests sind bei *nicht balancierten* Versuchsplänen möglich ?

Diese Frage hängt mit den verschiedenen Typen der SQ (Summen der Abweichungsquadrate) zusammen und wird im Teil II behandelt.
In **GLM** werden 4 SQ-Typen zugelassen.
In **MIXED** wird (bisher) nur der Typ III verwendet.
Wir werden weiterhin hier auch nur diesen Typ III betrachten!
Bei balancierten Versuchsplänen sind alle SQ-Typen identisch.

Im Wesentlichen können die Typ-III-SQ dadurch charakterisiert werden, daß hier zur eindeutigen Definition der Effekte folgende Bedingungen gestellt werden:

Summiert man Effekte über einen Index, so soll die Summe stets Null ergeben.

Beispiele:

$$\sum_{i=1}^{a} a_i = 0, \quad \sum_{j=1}^{b} b_j = 0, \quad \sum_{i=1}^{a} (ab)_{ij} = 0,$$

Einzelheiten, aber auch *Ausnahmen* werden im Teil II erläutert.

In **GLM** werden diese Bedingungen zunächst zur Berechnung der SQ formal auch für zufällige Effekte gefordert, bei der anschließenden Berechnung der E(MQ) jedoch nicht.

In **MIXED** werden diese Bedingungen nur für feste Effekte gefordert.

5. Programmdarstellung in SAS

GLM:

```
proc glm ;
  class a b c ;
  model x = a b c a*c a*b / ss3;
  random c a*c / test;
run;
```

– Zufällige Effekte **auch** in `MODEL`-Zeile angeben.

MIXED:

```
proc mixed;
  class a b c ;
  model x = a b a*b / ddfm=satterth;
  random c  a*c ;
run
```

– Zufällige Effekte **nicht** in `MODEL`-Zeile angeben.
– Option `SATTERTH` gibt an, daß Freiheitsgrade nach Sattethwaite (analog zu Welch-Test) zu berechnen sind, falls Varianzschätzungen sich aus mehreren MQ zusammensetzen.

Allgemein gilt:
Nur durch die Reihenfolge in der `CLASS`-Anweisung wird die Reihenfolge bei allen sortierten Ausgaben festgelegt:
Der Index von C läuft zuerst und der von A zuletzt.

6. Tabelle der E(MQ)

Nur in **GLM** wird durch die `RANDOM`-Zeile die Ausgabe der E(MQ) veranlaßt:

```
General Linear Models Procedure

Source      Type III Expected Mean Square

A           Var(Error) + 6 Var(A*C) + Q(A,A*B)
B           Var(Error) + Q(B,A*B)
C           Var(Error) + 6 Var(A*C) + 18 Var(C)
A*B         Var(Error) + Q(A*B)
A*C         Var(Error) + 6 Var(A*C)
```

- Findet man in der Literatur andere Formeln für E(MQ), so liegt nicht (unbedingt) ein Fehler vor, sondern die andere Modellversion.

- `Q(A,A*B)`, z.B., ist eine quadratische Funktion der nur von dem Index i abhängenden Größen a_i und $\sum_j (ab)_{ij}$. Letztere Summe war aber bei den Typ-III-SQ als Null vorausgesetzt, womit also die in $E(MQ_A)$ stehende quadratische Form `Q(A,A*B)` unter diesen Bedingungen tatsächlich nur von den A-Effekten abhängt.

7. Tests der Haupt- und Wechselwirkungen

GLM:

In **GLM** wird die folgende Varianztabelle ausgegeben.

Die F-Tests dieser Tabelle sind aber zu ignorieren, da diese nur für Modell I (alle Faktoren fest) zutrifft.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	2316.733056	210.612096	3.97	0.0023
Error	24	1271.643333	52.985139		
Corrected Total	35	3588.376389			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	2	433.1872222	216.5936111	4.09	0.0297
B	2	446.9272222	223.4636111	4.22	0.0269
C	1	285.0469444	285.0469444	5.38	0.0292
A*C	2	381.7272222	190.8636111	3.60	0.0428
A*B	4	769.8444444	192.4611111	3.63	0.0189

Es folgt eine zweite Serie von F-Tests, die das gemischte Modell korrekt berücksichtigt:
Hier - *aber auch nur hier* - berechnet **GLM** automatisch den richtigen Fehlerterm (`Error:` ...) und gibt hierfür die entsprechenden Werte für FG (`D-DF`) und MQ (`D-MS`) an.
Gegebenenfalls sind hier - *aber auch nur hier* - Linearkombinationen von verschiedenen MQ möglich.

*Zwei Beispiele: F-Test für A und für A*C:*

```
General Linear Models Procedure
Tests of Hypotheses for Mixed Model Analysis of Variance

Source: A
Error: MS(A*C)
                 Denominator  Denominator
 DF   Type III MS       DF             MS  F Value  Pr > F
  2  216.59361111        2   190.86361111   1.1348  0.4684
 This test assumes one or more other fixed effects are zero.

Source: A*C
Error: MS(Error)
                 Denominator  Denominator
 DF   Type III MS       DF             MS  F Value  Pr > F
  2  190.86361111       24   52.985138889   3.6022  0.0428
```

Der Hinweis „`This test assumes ... fixed effects are zero.`“ kann ignoriert werden, da dieser Test lediglich von Summen weiterer fester Effekte, welche ohnehin als Null vorausgesetzt wurden abhängt.

MIXED:
Die F-Tests für feste Faktoren werden folgendermaßen dargestellt. Das Ergebnis ist - zumindest im balancierten Fall - mit dem von GLM identisch.

```
  Tests of Fixed Effects

Source  NDF   DDF  Type III F  Pr > F

A         2     2        1.13  0.4684
B         2    24        4.22  0.0269
A*B       4    24        3.63  0.0189
```

Die Tests für zufällige Faktoren, d.h., für die Varanzkomponenten werden gemeinsam mit den Varianzkomponentenschätzungen ausgegeben:

```
  Covariance Parameter Estimates (REML)

Cov Parm       Ratio    Estimate   Std Error   Z  Pr > |Z|

C         0.09875236  5.23240741 24.77880162 0.21  0.8328
A*C       0.43370171 22.97974537 31.91258465 0.72  0.4715
Residual  1.00000000 52.98513889 15.29549210 3.46  0.0005
```

Hier wird nicht der F-Test, sondern der Wald-Test für Varianzkomponenten (Prüfgröße Z) durchgeführt. Bei diesem Test wird die Verteilung der Varianzkomponentenschätzung durch eine Normalverteilung approximiert, welche aber auch bei relativ großen Anlagen noch sehr ungenau ist. Außerdem führt SAS diesen Test zweiseitig durch, was bei als positiv vorausgesetzten Varianzkomponenten nicht sinnvoll ist. Der Test von Varianzkomponenten, sofern er interessiert, sollte daher - zumindest bei balancierten Anlagen - mit dem F-Test für gemischte Modelle aus **GLM** durchgeführt werden.

*Vergleich der p-Werte für den Test von A*C:*

MIXED	$p = 0.4715$	$p/2 = 0.2357$
GLM		$p = 0.0428$

Im unbalancierten Fall ist aber auch der F-Test in **GLM**, sofern im Nenner eine Linearkombination mehrerer MQ verwendet wird, approximativ.

8. Mittelwertschätzungen und -Vergleiche im balancierten Fall

8.1 *Theoretisches*

Es gilt z.B. für die Mittelwertschätzungen $\hat{\mu}_{ij}$ der $A_i B_j$-Kombinationen :

$$V(\hat{\mu}_{ij}) = \frac{ab}{N}(\sigma^2 + n\sigma_c^2 + n\sigma_{ac}^2) = \frac{1}{N} E\left[MQ_C + (a-1)MQ_{AC} + a(b-1)MQ_R\right]$$

Diese Varianz hängt also nicht nur von der Restvarianz σ^2 ab, sonder noch von weiteren Varianzen. Die Schätzung dieser Varianz ergibt sich aus einer Linearkombination mehrerer MQ.

Für $i \neq i'$ *gilt*

$$V(\hat{\mu}_{ij} - \hat{\mu}_{i'j'}) = \frac{2ab}{N}(\sigma^2 + n\sigma_{ac}^2) = E\left[\frac{2}{nbc}(MQ_{AC} + (b-1)MQ_R)\right]$$

Für $i = i'$ *gilt*

$$V(\hat{\mu}_{ij} - \hat{\mu}_{ij'}) = \frac{2ab}{N}\sigma^2 = E\left[\frac{2}{nc} MQ_R\right]$$

Hier gibt es also verschiedene Werte für die Varianz der Mittelwertschätzungen $\hat{\mu}_{ij}$ und ihrer verschiedenen Differenzen.

Analog gilt für die Stufen A_i von A:

$$V(\hat{\mu}_{i.}) = \frac{a}{N}(\sigma^2 + bn\sigma_c^2 + bn\sigma_{ac}^2) = E\left[\frac{1}{N}(MQ_c + (a-1)MQ_{AC})\right]$$

und

$$V(\hat{\mu}_{i.} - \hat{\mu}_{i'.}) = \frac{2a}{N}(\sigma^2 + bn\sigma_{ac}^2) = E\left[\frac{2}{nbc} MQ_{AC}\right]$$

Wie berücksichtigt SAS diese unterschiedlichen Varianzen?

8.2 *GLM:*

Hier gibt es zwei mögliche Anweisungen für Mittelwerte:

means:
berechnet für die Stufen (-kombinationen) vorgegebener Faktoren „rohe“ Mittelwerte, ohne dabei die weiteren Faktoren zu berücksichtigen.

Beispiel:
Mittelwerte für A-Stufen:

$$\bar{x}_{i...} = \mu + a_i + \frac{1}{n_{i..}}\left[\sum_j n_{ij.}(b_j + (ab)_{ij}) + \sum_{jk} n_{ijk}(c_k + (ac)_{ik}) + e_{i...}\right]$$

lsmeans:
berechnet für die Stufen (-kombinationen) vorgegebener Faktoren Mittelwerte, welche bezüglich der weiteren Faktoren korrigiert sind.

Beispiel:
Der durch `lsmeans` berechnete Mittelwert für A-Stufen entspricht
(bis auf die Resteffekte) dem folgenden Ausdruck:

$$\mu + a_i + \bar{b}_. + \overline{(ab)}_{i.} + \bar{c}_. + \overline{(ac)}_{i.}$$

Wegen der Summenbedingung für feste Faktoren hängt dieser Ausdruck also außer von $\mu + a_i$ von keinen weiteren festen Effekt ab.

Im hier betrachteten balancierten Fall sind beide Mittelwertarten gleich.

8.2.1 *Verhalten von **means** und **lsmeans** im gemischten Modell:*

In GLM werden bei den Mittelwertschätzungen die Besonderheiten des gemischten Modells nicht so automatisch berücksichtigt, wie bei den F-Tests. Hier wird nämlich die Kenntnis der oben genannten Varianzen von Mittelwerten und -differenzen vorausgesetzt.
D.h., die Anweisungen `means` und `lsmeans` berücksichtigen nicht die Informationen aus der `random`-Anweisung. Diese Informationen sind gesondert anzugeben, werden dann aber nur teilweise berücksichtigt.

Mittelwerte(-vergleiche) für A mit means:

```
means a / t e=a*c;
```

Optionen:

`t` Mittelwertvergleich mit t-Test

`e=a*c` gibt an, daß die Varianzen der Mittelwertdifferenzen unter Verwendung von MQ_{AC} (statt MQ_R) zu berechnen sind.
In dieser Option ist es nicht möglich, Linearkombinationen mehrerer MQ anzugeben!
Aus obiger Theorie wissen wir, daß für $V(\hat{\mu}_{i.} - \hat{\mu}_{i'.})$ MQ_{AC} statt MQ_R zu verwenden ist, daher wurde `e=a*c` gesetzt.
Im unbalancierten Fall ist es im allgemeinen unmöglich, hier mit nur einem MQ auszukommen. Auch die Herleitung der Formel dürfte kompliziert werden.

Ergebnis:

```
              Alpha= 0.05  df= 2  MSE= 190.8636
                  Critical Value of T= 4.30
            Least Significant Difference= 24.267

Means with the same letter are not significantly different.

          T Grouping             Mean      N  A

                   A           17.275     12  2
                   A
                   A           11.467     12  3
                   A
                   A            9.000     12  1
```

Hier sind die Ergebnisse richtig, da das richtige Fehler-MQ angebbar war. Glück gehabt.

8.2.1.2 Mittelwerte(-vergleiche) für A mit lsmeans:

```
lsmeans a / stderr tdiff E=a*c;
```

Optionen:

`stderr` Standardfehler für Mittelwert

`tdiff` t-Test für Differenzen

`E=a*c` „Fehler-MQ" *sowohl für Mittelwert, als auch für Vergleich*

Problem: $V(\hat{\mu}_{i.}) \neq V(\hat{\mu}_{i.} - \hat{\mu}_{i'.})$
Man kann aber das „Fehler-MQ" nur für Varianz **entweder** des Mittelwertes **oder** der Mittelwertdifferenz angeben. Damit sind also stets einige der Ergebnisse falsch!

Wir geben `E=a*c` an, also den Fehler für die Differenzen.

Der Fehler für die Mittelwerte wäre hier ohnehin nicht angebbar, da er sich aus mehreren MQ zusammensetzt.

Ergebnis:

```
                    Least Squares Means

Standard Errors and Probabilities calculated using the Type III
                  MS for A*C as an Error term

 A            X       Std Err     Pr > |T|   LSMEAN
         LSMEAN        LSMEAN   HO:LSMEAN=0   Number
        richtig        falsch        falsch
 1    9.0000000     3.9881450        0.1526     1
 2   17.2750000     3.9881450        0.0494     2
 3   11.4666667     3.9881450        0.1027     3
     richtig wäre:  4.30366

      T for HO: LSMEAN(i)=LSMEAN(j) / Pr > |T|

            i/j       1         2         3
            1     .      -1.46718  -0.43735
                           0.2800    0.7046      richtig
            2  1.467175     .       1.02983
                 0.2800              0.4113
            3  0.437345  -1.02983      .
                 0.7046    0.4113
```

8.2.1.3 Mittelwerte(-vergleiche) für A*B mit means:

```
means a*b ;
```

Für Kombinationen mehrerer Faktoren liefert `means` keinen Mittelwertvergleich. Die Option `e=...` entfällt hier also.

Ergebnis:

Level of A	Level of B	N	------X------ Mean	SD
1	1	4	9.2500000	2.2752289
1	2	4	8.9750000	0.2986079
1	3	4	8.7750000	2.2896506
2	1	4	8.4750000	1.2473305
2	2	4	30.8000000	24.0844348
2	3	4	12.5500000	3.0446675
3	1	4	8.0750000	1.3817260
3	2	4	11.7250000	3.2438403
3	3	4	14.6000000	5.6844818

Hier wurden die Mittelwerte - aber auch die Standardfehler - elementar aus allen Werten der jeweiligen Stufenkombination berechnet, ohne Berücksichtigung der anderen Faktoren. Es gibt daher keinen einheitlichen Standardfehler, obwohl ein balancierter Versuch vorliegt.

Diese Standardfehler lassen auch die Effekte von C unberücksichtigt.

8.2.1.4 Mittelwerte(-vergleiche) für A*B mit lsmeans:

```
lsmeans a*b / stderr tdiff   /* e=? */   ;
```

Hier ist wieder unklar, was man hinter **`e=`** schreiben sollte, da nicht nur die Varianzen der Mittelwerte sich von denen der Differenzen unterscheiden, sondern innerhalb der Differenzen noch zwei Fälle zu unterscheiden sind ($i = i'$ und $i \neq i'$).

Wir entscheiden uns für den Fall $i = i'$, da die anderen Fälle ohnehin nicht durch **nur ein** Fehler-MQ beschreibbar sind.
Im Fall $i = i'$ ist aber das Fehler-MQ gerade MQ_R , somit kann hier also die Angabe **`e=`** entfallen.

Ergebnis:

Least Squares Means

A	B	X LSMEAN	Std Err LSMEAN	Pr > |T| HO:LSMEAN=0	LSMEAN Number
1	1	9.2500000	3.6395446	0.0179	1
1	2	8.9750000	3.6395446	0.0212	2
1	3	8.7750000	3.6395446	0.0239	3
2	1	8.4750000	3.6395446	0.0286	4
2	2	30.8000000	3.6395446	0.0001	5
2	3	12.5500000	3.6395446	0.0021	6
3	1	8.0750000	3.6395446	0.0362	7
3	2	11.7250000	3.6395446	0.0036	8
3	3	14.6000000	3.6395446	0.0005	9
		richtig	*falsch*	*falsch*	

T for HO: LSMEAN(i)=LSMEAN(j) / Pr > |T|

i/j	1	2	3	4	5	6	7
1	.	0.053428	0.092285	0.15057	-4.18683	-0.64114	0.228284
		0.9578	0.9272	0.8816	0.0003	0.5275	0.8214
2	-0.05343	.	0.038857	0.097142	-4.24026	-0.69457	0.174856
	0.9578		0.9693	0.9234	0.0003	0.4940	0.8627
3	-0.09229	-0.03886	.	0.058285	-4.27911	-0.73342	0.135999
	0.9272	0.9693		0.9540	0.0003	0.4704	0.8930
	richtig	*...*	*richtig*	*falsch*		*...*	*falsch*

$i = i'$ | $i \neq i'$

usw.

8.2.2 Ermittlung der theoretischen Varianz mit Hilfe der **`contrast`**-Anweisung

Bei Unsicherheit kann das „Fehler-MQ" für Differenzen auch mit der **`contrast`**-Anweisung bestimmt werden, sofern diese zwischen die **`model`**- und **`random`**-Zeile eingefügt wird.

*Beispiele für A*B-Differenzen:*

Fall $i \neq i'$ *:*

Die Differenz $\overline{\mu}_{2,3} - \overline{\mu}_{1,2}$ setzt sich zusammen aus:

$$\overline{\mu}_{2,3} - \overline{\mu}_{1,2} = (-a_1 + a_2) + (-b_2 + b_3) + (\,(ab)_{2,3} - (ab)_{1,2}\,) + \cdots$$

Daher ist sie durch folgende **`contrast-`**Anweisung darstellbar:

*(Die Koeffizienten müssen nur für A*B und für die in A*B enthaltenen Effekte angegeben werden. Die Koeffizienten der weiteren Effekte werden automatisch gemäß der Typ-III-Bedingungen und der Schätzbarkeitsbedingungen (siehe 9.2) gesetzt.)*

```
contrast "A*B (2,3)-(1,2)"
                        a   -1  1 0
                        b    0 -1 1
                      a*b    0 -1 0   0 0 1   0 0 0   /*  /  e=?  */  ;
```

Nach **`e=`** kann hier nichts sinnvolles angegeben werden.

Fall $i = i'$ *:*

Die Differenz $\overline{\mu}_{1,3} - \overline{\mu}_{1,2}$ setzt sich zusammen aus:

$$\overline{\mu}_{1,3} - \overline{\mu}_{1,2} = (-b_2 + b_3) + (\,(ab)_{1,3} - (ab)_{1,2}\,) + \cdots$$

Daher ist sie durch folgende **`contrast-`**Anweisung darstellbar:

```
contrast "A*B (1,3)-(1,2)"
                        b    0 -1 1
                      a*b    0 -1 1   0 0 0   0 0 0 ;
```

Ergebnis:

```
Contrast          Contrast Expected Mean Square

A*B (2,3)-(1,2) Var(Error) + 2 Var(A*C) + Q(A,B,A*B)
A*B (1,3)-(1,2) Var(Error) + Q(B,A*B)


Contrast          DF  Contrast SS  Mean Square  F Value  Pr > F

A*B (2,3)-(1,2)  1   25.5612500   25.5612500     0.48  0.4940
A*B (1,3)-(1,2)  1    0.0800000    0.0800000     0.00  0.9693
```

Der in den **`Contrast Expected Mean Square`** vor der quadratischen Form **`Q`** stehende Ausdruck entspricht - bis auf den Faktor $\frac{2ab}{N}$ - der Varianz der Differenzschätzung. Bei Wiederholung der Rechnung sollte man also nach **`e=`** ein MQ angeben, dessen Erwartungswert mit diesem Ausdruck identisch ist, Für die Differenz „**`A*B (1,3)-(1,2)`**" (Fall $i = i'$) wäre das MQ_R; für „**`A*B (2,3)-(1,2)`**" existiert kein entsprechendes MQ.

Die **`contrast`**-Anweisung dient nur für den Tests eines Kontrasts oder für den simultanen Test mehrerer Kontraste. Zur Schätzung eines einzelnen Kontrastes ist die **`estimate`**-Anweisung analog anzuwenden; sie liefert aber auch einen Test. Die hier betrachteten $\overline{\mu}_{ij}$ und $\overline{\mu}_{i.}$ sind sogenannte Kombinationseffekte, also die Summe aller in einer vorgegebenen Stufe (-nkombination) auftretenden Effekte. Will man nicht die Differenzen der $\overline{\mu}_{ij}$, sondern die der Effekte $(ab)_{ij}$ schätzen, so werden in der Programmzeile (**`contrast`** bzw. **`estimate`**) nur für **`a*b`** und nicht für **`a`** und **`b`** Koeffizienten angegeben. (siehe auch Rasch u.a., 1996, Verfahren 1/61/0002)

8.3 *MIXED:*

Hier sind keine theoretischen Kenntnisse erforderlich. **MIXED** berechnet automatisch die richtigen Varianzen unter Ausnutzung der Informationen aus der `random`-Zeile.
Folgende Programmzeile ist einzufügen(in MIXED existiert nur die `lsmeans` Anweisung):

```
lsmeans a a*b / diff pdiff;
```

Optionen:
`diff pdiff` Berechnung der Differenzen und ihrer Tests.

Für die `lsmeans`-Ergebnisse soll hier nur ein kleiner Ausschnitt gezeigt werden. Man sieht deutlich die unterschiedlichen Standardfehler (`Std Err.`) und Freiheitsgrade bei den Fällen $i = i'$ und $i \neq i'$.

```
                   Least Squares Means

 Level           LSMEAN     Std Error   DDF       T  Pr > |T|

 A 1         9.00000000    4.30366177  2.88    2.09    0.1312
        ...
 A*B 1 1     9.25000000    5.22994848  6.12    1.77    0.1264
        ...

                    Differences of Least Squares Means

Level 1   Level 2    Difference    Std Error    DDF       T  Pr > |T|

A 1       A 2       -8.27500000   5.64008882      2   -1.47   0.2800
A 1       A 3       -2.46666667   5.64008882      2   -0.44   0.7046
A 2       A 3        5.80833333   5.64008882      2    1.03   0.4113
A*B 1 1   A*B 1 2    0.27500000   5.14709330     24    0.05   0.9578
A*B 1 1   A*B 1 3    0.47500000   5.14709330     24    0.09   0.9272
A*B 1 1   A*B 2 1    0.77500000   7.03365586   4.72    0.11   0.9168
A*B 1 1   A*B 2 2  -21.55000000   7.03365586   4.72   -3.06   0.0302
```

Besonderheit in MIXED:

MIXED benutzt die theoretischen Darstellungen von Varianzen, wie z.B.

$$V(\hat{\bar{\mu}}_{ij} - \hat{\bar{\mu}}_{i'j'}) = \frac{2ab}{N}(\sigma^2 + n\sigma^2_{ac}) \qquad (i \neq i')$$

$$V(\hat{\bar{\mu}}_{ij} - \hat{\bar{\mu}}_{ij'}) = \frac{2ab}{N}\sigma^2 \qquad (i = i'),$$

und setzt für die Varianzkomponenten die REML-Schätzwerte ein. Bei den hier betrachteten balancierten Versuchsplänen ist REML=ANOVA, sofern die ANOVA-Schätzwerte positiv sind. Ist jedoch, z.B. der ANOVA-Schätzwert $\hat{\sigma}^2_{ac}$ negativ, so geht MIXED - gemäß dem REML-Prinzip - folgendermaßen vor:

Eliminiere den Effekt A*C, aus dem Modell, d.h., setze $\hat{\sigma}^2_{ac} = 0$, und wiederhole die Auswertung.

Die neue Varianztabelle ergibt sich durch Assimilation der fehlenden Wirkung A*C zur umfassenden Wirkung R. In unserem Beispiel erhält man

$$FG_{R(neu)} = FG_{AC} + FG_R$$

und

$$MQ_{R(neu)} = (SQ_{AC} + SQ_R) / FG_{R(neu)}.$$

Nun ist aber $V(\hat{\bar{\mu}}_{ij} - \hat{\bar{\mu}}_{i'j'})$ in beiden Fällen $i = i'$ und $i \neq i'$ gleich.
Dieses Vorgehen ist sinnvoll, da damit die theoretische Ungleichung

$$V(\hat{\bar{\mu}}_{ij} - \hat{\bar{\mu}}_{i'j'}) \geq V(\hat{\bar{\mu}}_{ij} - \hat{\bar{\mu}}_{ij'})$$

berücksichtigt wird, was bei dem traditionellen Vorgehen nicht der Fall wäre.

9. Unterschiede zwischen GLM und MIXED im unbalancierten Fall

Bisher wurden die Unterschiede aber auch Gemeinsamkeiten zwischen GLM und MIXED nur im balancierten Fall demonstriert. Im unbalancierten Fall gibt es noch weitere Unterschiede in der prinzipiellen Vorgehensweise.

Allgemeines Matrizenmodell:

$$y = X\beta + Z\gamma + \varepsilon$$

y = Beobachtungsvektor
ß = feste Effekte
γ = zufällige Effekte
ε = Zufallsfehler

9.1 Grundsätzliches Vorgehen

GLM:
Berechnet zunächst die Schätzungen wie im Modell I und berücksichtigt anschließend - mehr oder weniger korrekt - die Verteilung dieser Größen im gemischten Modell.
MIXED:

- Darstellung des Modells durch
  $$y = X\beta + \tilde{\varepsilon}$$
- mit $\tilde{\varepsilon} = Z\gamma + \varepsilon$ und $V(\tilde{\varepsilon}) = V = ZV(\gamma)Z' + V(\varepsilon)$
- Schätztung der Parameter von V, also der Varianzkomponenten, und Einsetzen dieser in V $\Rightarrow$ $\hat{V}$
- Multiplikation des Modells mit $\hat{V}^{-1/2}$:
  $$y^* = X^*\beta + \varepsilon^* \quad = \hat{V}^{-1/2}y^* = \hat{V}^{-1/2}X\beta + \hat{V}^{-1/2}\tilde{\varepsilon}$$
- Weitere Auswertung bezüglich ß wie im Modell I

9.2 Schätzbare Funktionen

Eine Funktion $L'\beta$ des Parametervektors β kann nur durch eine Linearkombination $K'y$ der y-Komponenten geschätzt werden, d.h., es muß
$$E(K'y) = K'X\beta = L'\beta$$
gelten. Die Funktion $L'\beta$ ist also nur dann schätzbar, wenn L' sich als Linearkombination $L' = K'X$ der Zeilen von X darstellen läßt.

Mit dieser Schätzbarkeitsbedingung und den Typ-III-Summenbedingungen werden in den Anweisungen **`estimate`** und **`contrast`** die nur für die interessierenden Effekte angegebenen Koeffizienten automatisch
  für alle festen Effekte (in MIXED)
bzw.
  für alle festen und analog auch zufälligen Effekte (in GLM)
komplettiert.
Eine Funktion, welche in MIXED schätzbar ist, könnte - in Abhängigkeit von Fehlstellen - eventuell (überflüssigerweise) in GLM nicht schätzbar sein.