

Hans-Peter Piepho

# Vertrauensintervalle für multiple Vergleiche bei ungleichen Standardfehlern der Differenzen

Die vorliegende Arbeit stellt ein Verfahren vor, mit welchem Vertrauensintervalle für Behandlungsmittelwerte berechnet werden können. Diese Intervalle eignen sich zum multiplen Vergleich von Mittelwerten. Sie können auch bei unbalancierten Daten angewendet werden und sind von besonderem Vorteil, wenn die Zahl der Behandlungen so groß ist, dass eine paarweise Betrachtung aller Differenzen unpraktisch ist.

## 1 Einleitung

Multiple Vergleiche sind ein sehr häufig angewendetes Verfahren zur Auswertung von Versuchen und Erhebungen. Hierbei werden  $I$  verschiedene Behandlungen miteinander verglichen. Angenommen,  $\mathbf{q}$  sei ein zu schätzender Parameter der  $i$ -ten Behandlung ( $i = 1, \dots, I$ ), beispielsweise der Erwartungswert oder ein additiver Effekt, und es sollen alle paarweisen Vergleiche durchgeführt werden. Hierzu können simultane Konfidenzintervalle für die Differenzen aufgestellt werden nach

$$\mathbf{q}_i - \mathbf{q}_{i'} \in [\hat{\mathbf{q}}_i - \hat{\mathbf{q}}_{i'} \pm d_{ii'}] \quad (1)$$

wobei  $\hat{\mathbf{q}}_i$  die Schätzung von  $\mathbf{q}$  ist und  $d_{ii'}$  die halbe Intervallbreite der Differenz darstellt. Enthält ein Intervall den Wert 0, so sind die betreffenden beiden Behandlungen nicht signifikant voneinander verschieden. Die halbe Intervallbreite  $d_{ii'}$  berechnet sich im allgemeinen nach

$$d_{ii'} = q_{ii'} \sqrt{\text{geschätzte Varianz von } (\hat{\mathbf{q}}_i - \hat{\mathbf{q}}_{i'})} \quad (2)$$

wobei  $q_{ii'}$  das Quantil einer geeigneten Prüfverteilung ist. Die Wahl von  $q_{ii'}$  hängt u.a. davon ab, ob der vergleichsbezogene oder der versuchsbezogene Fehler 1. Art eingehalten werden soll (Hochberg und Tamhane, 1987; Hsu, 1996). Das Ziel der vorliegenden Arbeit ist die graphische Darstellung der durch die Vertrauensintervalle in (1) gegebenen Information. Hierbei soll um jeden Schätzwert ein Intervall der Form

$$[\hat{\mathbf{q}}_i \pm w_i] \quad (3)$$

gelegt werden, so dass folgende Aussage möglich ist: Überlappen sich die Intervalle zweier Behandlungen, so sind diese Behandlungen nicht signifikant voneinander verschieden (vgl. Abb.1). Eine solche Darstellung ist den paarweisen Vertrauensintervallen für alle Differenzen (1) vorzuziehen, da anstelle von  $I(I-1)/2$  Intervallen lediglich  $I$  Intervalle zu betrachten sind. Es sollte an dieser Stelle ausdrücklich darauf hingewiesen werden, dass die in dieser Arbeit diskutierten Intervalle der Form in (3) ausschliesslich zur Beurteilung von

Mittelwertdifferenzen, nicht aber der Mittelwerte  $\mathbf{q}$  selbst herangezogen werden können.

## 2 Methode

Das in der Einleitung beschriebene Vorgehen zur Berechnung von Vertrauensintervallen der Differenzen (1) gilt ganz allgemein, insbesondere für das lineare Modell (LM), das lineare gemischte Modell (LMM) und das generalisierte lineare Modell (GLM). Zur Einhaltung des vergleichsbezogenen Fehlers wird für das LM ein  $t$ -Quantil für  $q_{ii'}$  gewählt. Für das LMM ist ein approximatives Intervall basierend auf der  $t$ -Verteilung (Giesbrecht und Burns, 1985) oder der Standardnormalverteilung möglich. Für das GLM ist  $q_{ii'}$  das Quantil der Standardnormalverteilung (McCullagh und Nelder, 1989). Bei Überdispersion ist das Quantil einer  $t$ -Verteilung vorzuziehen (McCullagh, 1983). Tests im GLM sind meist nur approximativ gültig. Die Einhaltung des versuchsbezogenen Fehlers 1. Art für LM, LMM und GLM ist im allgemeinen mit Hilfe der Ungleichung von Bonferroni möglich. Im LM können alternativ die Ungleichung von Sidak (Hsu, 1996) oder das Simulationsverfahren von Edwards und Berry (1987) verwendet werden. In einigen Fällen sind im LM spezielle Verfahren verfügbar wie z.B. die Tukey-Kramer Prozedur für die unbalancierten einfaktoriellen Varianzanalyse.

Nun zur Berechnung der Vertrauensintervalle (3). Im varianzbalancierten Fall, also wenn die geschätzte Varianz von  $\hat{\mathbf{q}}_i - \hat{\mathbf{q}}_{i'}$  konstant für alle Vergleiche ist und somit  $d_{ii'} = d$ , so wählen wir einfach  $w_i = 0.5d$ . Hiermit werden alle Vergleiche exakt abgebildet. Im unbalancierten Fall ist dagegen lediglich eine konservative Lösung möglich (Hochberg et al., 1982). Wir wählen  $w_i$  unter den Nebenbedingungen

$$w_i + w_{i'} \geq d_{ii'} \quad \text{für alle } i \neq i' \quad \text{und} \quad (4)$$

$$w_i \geq 0 \quad (5)$$

Die Nebenbedingung (4) stellt sicher, dass die Intervalle die vorgegebene Irrtumswahrscheinlichkeit konservativ einhalten. Die  $w_i$  werden nun so gewählt, dass die Intervalle möglichst klein sind, dabei aber (4) und (5) eingehalten werden. Zur Wahl von  $w_i$  gibt es zwei verschiedene, relativ

ähnliche Vorschläge (Hochberg et al., 1982; Hochberg und Tamhane, 1987):

$$\text{Minimiere } \sum_{i=1}^{I-1} \sum_{i'=i+1}^I (w_i + w_{i'}) \quad \text{und} \quad (6)$$

$$\text{Minimiere } \sum_{i=1}^{I-1} \sum_{i'=i+1}^I (w_i + w_{i'}) / d_{ii'}, \quad (7)$$

jeweils unter den Nebenbedingungen (4) und (5). Beide Anweisungen sind Spezialfälle der Anweisung

$$\text{Minimiere } \sum_{i=1}^{I-1} \sum_{i'=i+1}^I (w_i + w_{i'}) / (d_{ii'})^f \quad (f \geq 0), \quad (8)$$

welche im folgenden verwendet wird. Hierbei ist  $f$  ein Parameter, dessen Wert durch Probieren so zu wählen ist, dass alle Signifikanzen abgebildet werden können. Es gibt in der Regel viele Werte für  $f$ , die diese Bedingung erfüllen. Allerdings kann auch der Fall eintreten, dass kein Wert für  $f$  eine Darstellung aller Signifikanz erlaubt. In anderen Worten, es ist möglich, dass die Intervalle für zwei Behandlungen überlappen, obwohl diese Behandlungen signifikant verschieden sind. Dies tritt vor allem in solchen Fällen auf, in denen der betreffende Vergleich an der Signifikanzgrenze ist. Für einen gegebenen Wert von  $f$  stellt die Anweisung (8) ein klassisches Problem der linearen Optimierung dar, für das als Standardlösung der Simplex-Algorithmus zur Verfügung steht (Bronstein, 1989). Wir verwenden hier die SAS/IML Subroutine LINPROG (SAS Institute, 1989, S. 138ff.) zur Lösung von (8).

Man beachte, dass wir durch die Vertrauensintervalle (3) implizit die halbe Intervallbreite approximieren durch

$$d_{ii'} \approx (w_i + w_{i'}) \quad (9)$$

Wegen der Nebenbedingung (4) sind die Intervalle in der Regel konservativ, also breiter als zur Einhaltung der Irrtumswahrscheinlichkeit notwendig. Im varianzbalancierten Fall gilt  $d_{ii'} = w_i + w_{i'}$ , so dass kein Approximationsfehler vorliegt und die Irrtumswahrscheinlichkeit exakt eingehalten wird.

### 3 Beispiel

Zur Erläuterung der Methode verwenden wir eine unbalancierte Serie von Versuchen mit Triticale (vgl. auch Piepho, 2000). Die Einzelversuche waren als Blockanlage ausgelegt. Von den insgesamt  $13 \times 5 \times 6 \times 4 = 1560$  möglichen Kombinationen Sorte  $\times$  Jahr  $\times$  Ort  $\times$  Block wurden nur 196 beobachtet, so dass die Daten stark unbalanciert sind. Zur Auswertung wurde das folgende gemischte Modell mit der SAS Prozedur MIXED angepasst (zufällige Terme unterstrichen):

$$\text{Ertrag} = \text{Sorte} + \text{Jahr} + \text{Ort} + \text{Sorte} * \text{Ort} + \text{Sorte} * \text{Jahr} + \text{Jahr} * \text{Ort} + \text{Sorte} * \text{Ort} * \text{Jahr} + \text{Block}(\text{Ort} * \text{Jahr}) + \text{Fehler} \quad (10)$$

Es wurde ein multipler  $t$ -Test zum vergleichsbezogenen Niveau  $\alpha = 5\%$  durchgeführt. Die Freiheitsgrade wurden

nach Giesbrecht und Burns (1985) approximiert. Die generalisierten Kleinstquadrat-Mittelwerte sind in Tab.1 aufgeführt. Es stellte sich heraus, dass bei Anwendung von (8) mit  $f = 0$  und  $f = 1$  die Signifikanz der Differenz von C52 und C78 verloren geht, nicht aber mit  $f = 0,5$ . Daher wird hier  $f = 0,5$  verwendet. Die paarweisen Vergleiche finden sich in Tab.2. Die approximativen Intervalle (3) sind in Abb.1 wiedergegeben.

Die Summe der approximierten halben Intervallbreiten ist  $\sum_{i=1}^{I-1} \sum_{i'=i+1}^I (w_i + w_{i'}) = 1220$ , während die Summe der tatsächlichen Intervallbreiten gegeben ist durch

$$\sum_{i=1}^{I-1} \sum_{i'=i+1}^I d_{ii'} = 1175. \text{ Somit ist die Approximation als relativ}$$

gut zu bezeichnen. Im Einzelfall kann die Diskrepanz aber recht deutlich ausfallen. Die größte Abweichung zwischen  $d_{ii'}$  und  $w_i + w_{i'}$  ergibt sich für den Vergleich C7 vs. C10 ( $d_{ii'} = 18,99$ ;  $w_i + w_{i'} = 24,44$ ). Der "reale  $\alpha$ -Wert", d.h. derjenige  $\alpha$ -Wert, für den der berechnete Wert für  $d_{ii'}$  der approximativen halben Intervallbreite  $w_i + w_{i'}$  für das nominale  $\alpha$  entsprechen würde, beträgt hier 0,014 statt des nominalen Niveaus von 0,050 (Tab.2), so dass das Intervall relativ konservativ ist. Für die anderen Vergleiche ist das reale  $\alpha$  zum größten Teil deutlich näher am nominalen Wert, so dass der Verlust an Trennschärfe geringer ausfällt, und für viele Vergleiche sind reales und nominales  $\alpha$  sogar identisch. Trotz des mit der Approximation einhergehenden Informationsverlustes können alle Signifikanzen abgebildet werden. Abb.1 ist m.E. einer paarweisen Betrachtung aller  $13 \times 12/2 = 78$  (!) Intervalle oder Tests für Differenzen (Tab.2) vorzuziehen.

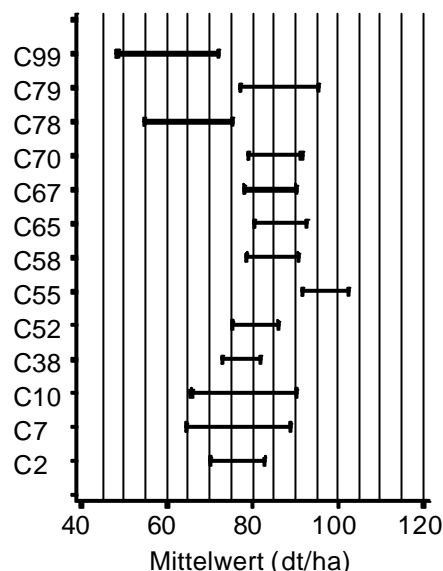


Abb.1: Vertrauensintervalle um Mittelwerte (dt/ha) der Sorten des Triticale-Datensatzes zur Beurteilung von Mittelwertdifferenzen [Behandlungen, deren Intervalle sich überlappen, sind nicht signifikant voneinander verschieden ( $\alpha = 5\%$ , vergleichsbezogen). Es geht keine signifikante Differenz bei dieser Darstellung verloren.  $f = 0,5$ .]

Tab.1: Generalisierte Kleinstquadratschätzungen der Sortenmittelwerte (dt/ha) der Triticale-Daten.

Sorte	Mittel	Standardfehler	Gruppierung*
C55	97,1	6,89	a
C65	86,6	7,27	a b
C79	86,2	8,13	a b
C70	85,2	7,27	b
C58	84,7	7,15	b
C67	84,2	7,15	b
C52	80,6	6,96	b
C10	78,0	9,84	b c d
C38	77,3	6,67	b c
C7	76,8	9,84	b c d
C2	76,6	7,30	b c d
C78	65,1	8,73	c d
C99	60,3	9,70	d

\*: Behandlungen, die mit demselben Buchstaben versehen sind, sind nicht signifikant voneinander verschieden. Vergleiche mit *t*-Test ( $\alpha = 5\%$  vergleichsbezogen) und Freiheitsgraden nach Giesbrecht und Burns (1985). Buchstabendarstellung nach Piepho (2000).

#### 4 Ein SAS Macro

Ich habe ein SAS Macro geschrieben, welches das hier beschriebene Verfahren für LM und LMM mit Hilfe der des Modules IML umsetzt. Das Macro wird unter [www.wiz.uni-kassel.de/fts](http://www.wiz.uni-kassel.de/fts) verfügbar gemacht. Es verwendet Output der SAS Prozedur MIXED (Version 6.12) und ist wie folgt definiert:

```
%macro mcagraph(diff=, lsmean=, error='comp', alpha=0.05, phi=0);
```

Die Datei `diff` enthält die paarweisen Vergleiche wie sie mit Hilfe der `Diff's` Tabelle von MIXED (MAKE Anweisung) erhalten werden kann. Die Datei `lsmean` enthält die Behandlungsmittelwerte und wird über die `'LsMeans'` Tabelle von MIXED erzeugt (MAKE Anweisung). Die Variable `error` gibt vor, ob der vergleichsbezogene Fehler 1. Art (`error='comp'`; Voreinstellung) oder der versuchsbezogene Fehler 1. Art (`error='exp'`) mit Hilfe der Bonferroni-Ungleichung kontrolliert werden soll. Die Variable `alpha` gibt den  $\alpha$ -Fehler vor (Voreinstellung  $\alpha = 0.05$ ). Der Parameter `f` des zu minimierenden Kriteriums (8) wird durch die Variable `phi` festgelegt (Voreinstellung  $f = 0$ ). Für das Triticale-Beispiel benötigen wir die Daten in einem Datensatz mit den Variablen `sorte`, `ort`, `jahr`, `block` und `ertrag` mit einer offensichtlichen Belegung dieser Bezeich-

nungen. Zunächst wird das Macro verfügbar gemacht (Macro in den Editor laden und 'abschicken' oder die %INCLUDE Anweisung verwenden). Dann werden die Daten mit der Prozedur MIXED durch untenstehende Anweisungen ausgewertet:

```
data temp;
input sorte ort jahr block ertrag;
cards;
<Daten>
;
proc mixed data=temp;
class sorte ort jahr block;
model ertrag=sorte/ddfm=satterth;
random jahr ort jahr*ort sorte*jahr
sorte*ort sorte*ort*jahr block(ort*jahr);
lsmeans sorte/pdiff;
make 'diffs' out=diff;
make 'lsmeans' out=lsmean;
run;
```

Abschliessend wird das Macro wie folgt aufgerufen:

```
%mcagraph(diff=diff, lsmean=lsmean, error=
'comp', alpha=0.05, phi=0.5);
```

Der Output enthält zum einen die Mittelwerte (`_LSMEAN_`) und Vertrauensintervalle (`_LOWER_`, `_UPPER_`):

lsmeans with confidence limits			
SORTE	_LSMEAN_	_LOWER_	_UPPER_
2	76.61633187	70.2112	83.021
7	76.77573861	64.5571	88.994
10	77.97573861	65.7571	90.194
38	77.32127618	72.9246	81.718
52	80.61553394	75.2582	85.973
55	97.04345859	91.7249	102.362
58	84.70027677	78.6477	90.753
65	86.62154485	80.3964	92.847
67	84.16514025	78.1126	90.218
70	85.24045830	79.0153	91.466
78	65.06038171	54.8740	75.247
79	86.22061468	77.0556	95.386
99	60.30723704	48.4630	72.151

Zum anderen werden die 'verlorenen' signifikanten Differenzen (`_DIFF_`) mit Überschreitungswahrscheinlichkeit des *t*-Tests (`_PT_`) ausgegeben. Mit  $f = 0,5$  ergibt sich keine verlorene Signifikanz. Mit  $f = 0$  dagegen erhalten wir eine verlorene Signifikanz:

lost significances			
SORTE	_SORTE	_DIFF_	_PT_
52	78	15.55515223	0.0389

Da das Macro Output der SAS Prozedur MIXED verwendet (SAS Institute, 1997), ist es für LM und LMM geeignet. Zur Anwendung bei GLM muss zuvor ein SAS-Datensatz erstellt werden, welcher die Information über die paarweisen Vergleiche im selben Format wie die `Diff's` Tabelle von MIXED enthält. Auch die Sortierung der Vergleiche muss sich an diese Tabelle anlehnen. Die paarweisen Vergleiche können z.B. mit der CONTRAST Anweisung in der Prozedur GENMOD (SAS Institute, 1997) durchgeführt werden.

Tab. 2: Geschätzte Behandlungsdifferenzen ( $dt/ha$ ) und kritische Differenzen für Triticale-Daten [S: Freiheitsgrade nach Giesbrecht und Burns (1985). \*: Signifikant nach t-Test bei  $\alpha = 0,05$  (vergleichsbezogen). S: Halbe Intervallbreite basierend auf t-Test. #: Standardfehler einer Differenz. +:  $\phi = 0,5$ . &: Beschreibung siehe Text. Bei einem realen  $\alpha$ -Wert kleiner dem nominalen Wert von 0,05 liegt ein Trennschrägenverlust vor].

Vergleich	Differenz	SED <sup>#</sup>	FG <sup>S</sup>	$(d_{0,5})^S$	$(w_1 + w_2)^+$	Reales $\alpha$ &	Vergleich	Differenz	SED <sup>#</sup>	FG <sup>S</sup>	$(d_{0,5})^S$	$(w_1 + w_2)^+$	Reales $\alpha$ &
C2							C7	-0,16	8,06	23,05	16,66	18,62	0,030
C2	C10	8,06	23,05	16,66	18,62	0,030	C7	-8,90	6,68	14,51	14,28	14,58	0,046
C2	C38	8,06	23,05	16,66	18,62	0,030	C79	17,01	6,08	10,17	13,52	13,56	0,049
C2	C38	4,58	15,52	9,73	10,80	0,032	C99	-16,43	7,91	26,67	16,24	16,24	0,050
C2	C52	-4,00	13,74	10,62	11,76	0,032	C55	4,63	4,63	11,72	10,11	10,68	0,040
C2	C55	-20,43	11,07	14,72	11,07	0,039	C52	-4,08	5,16	10,65	11,40	11,41	0,050
C2	C58	-8,08	5,76	13,28	12,42	0,049	C65	-6,01	5,29	11,54	11,58	11,58	0,050
C2	C65	-10,01	5,90	14,27	12,63	0,050	C67	-3,55	5,16	10,65	11,40	11,41	0,050
C2	C67	-7,55	5,76	13,28	12,42	0,049	C70	-4,62	5,29	11,54	11,58	11,58	0,050
C2	C70	-8,62	5,90	14,27	12,63	0,050	C78	15,56	6,82	13,81	14,64	15,54	0,039
C2	C78	11,56	6,95	14,10	14,90	0,032	C79	-5,61	6,55	10,39	14,52	14,52	0,050
C2	C79	-9,60	6,96	12,05	15,16	0,045	C99	20,31	8,03	25,08	16,53	17,20	0,042
C2	C99	16,31	8,15	25,01	16,78	0,034	C58	12,34	4,85	9,80	10,85	11,37	0,042
C7	C10	-1,20	9,15	21,89	18,99	0,044	C55	10,42	5,03	10,99	11,06	11,54	0,042
C7	C38	-0,55	7,92	23,84	16,35	0,047	C55	12,88	4,85	9,80	10,85	11,37	0,042
C7	C52	-3,84	8,39	21,38	17,43	0,048	C55	11,80	5,03	10,99	11,06	11,54	0,042
C7	C55	-20,27	8,44	21,49	17,54	0,050	C55	31,98	7,19	14,19	15,41	15,50	0,049
C7	C58	-7,92	8,76	19,89	18,27	0,050	C79	10,82	6,19	9,68	13,85	14,48	0,042
C7	C65	-9,85	8,86	20,49	18,44	0,050	C55	36,74	8,33	24,61	17,16	17,16	0,050
C7	C67	-7,39	8,76	19,89	18,27	0,050	C58	-1,92	5,27	10,14	11,72	12,28	0,042
C7	C70	-8,46	8,86	20,49	18,44	0,050	C58	0,54	5,10	9,10	11,53	12,11	0,042
C7	C78	11,72	9,77	19,09	20,45	0,033	C70	-0,54	5,27	10,14	11,72	12,28	0,042
C7	C79	-9,44	9,57	17,40	20,16	0,039	C58	19,64	7,54	13,38	16,24	16,24	0,050
C7	C99	16,47	10,74	26,41	22,06	0,034	C58	-1,52	6,37	9,21	14,36	15,22	0,040
C10	C38	0,65	7,92	23,84	16,36	0,047	C58	24,39	8,64	22,54	17,89	17,90	0,050
C10	C52	-2,64	8,39	21,38	17,43	0,048	C65	2,46	5,27	10,14	11,72	12,28	0,042
C10	C55	-19,07	8,44	21,49	17,54	0,050	C65	1,38	5,38	10,85	11,86	12,45	0,041
C10	C58	-6,72	8,76	19,89	18,27	0,050	C65	21,56	7,64	13,97	16,40	16,41	0,050
C10	C65	-8,65	8,86	20,49	18,44	0,050	C65	0,40	6,54	10,16	14,55	15,39	0,040
C10	C67	-6,19	8,76	19,89	18,27	0,050	C65	26,31	8,72	23,10	18,03	18,07	0,050
C10	C70	-7,26	8,86	20,49	18,44	0,050	C70	-1,08	5,27	10,14	11,72	12,28	0,042
C10	C78	12,92	9,77	19,09	20,45	0,033	C67	19,10	7,54	13,38	16,24	16,24	0,050
C10	C79	-8,24	9,57	17,40	20,16	0,039	C67	-2,06	6,37	9,21	14,36	15,22	0,040
C10	C99	17,67	10,74	26,41	22,06	0,034	C67	23,86	8,64	22,54	17,89	17,90	0,050
C38	C52	-3,29	4,21	12,18	9,16	0,039	C70	20,18	7,64	13,97	16,40	16,41	0,050
C38	C55	-19,72	4,19	11,95	9,14	0,039	C70	-0,98	6,54	10,16	14,55	15,39	0,040
C38	C58	-7,38	4,70	10,64	10,39	0,049	C99	24,93	8,72	23,10	18,03	18,07	0,050
C38	C65	-9,30	4,87	11,97	10,62	0,050	C78	-21,16	8,50	12,45	18,45	19,35	0,041
C38	C67	-6,84	4,70	10,64	10,39	0,049	C99	4,75	9,30	23,83	19,21	22,03	0,026
C38	C70	-7,92	4,87	11,97	10,62	0,050	C79	25,91	9,48	19,17	19,84	21,01	0,039

## 5 Abschließende Bemerkung

Die graphische Darstellung von Konfidenzintervallen für Mittelwertdifferenzen hat den Vorteil der Übersichtlichkeit und der Möglichkeit, neben den Signifikanzen auch die Genauigkeit der Schätzungen von Mittelwertdifferenzen darstellen zu können. Ein Nachteil besteht in der Einbusse von Trennschärfe bei unbalancierten Daten. Als Alternative für eine kompakte graphische Darstellung kommt die Buchstabendarstellung in Frage, welche auch im unbalancierten Fall immer eine Abbildung aller Signifikanzen erlaubt (Piepho, 2000). Allerdings kann diese Darstellung bei einer großen Zahl von Mittelwerten unübersichtlich werden. Außerdem liefert die Buchstabendarstellung im Gegensatz zu den hier vorgestellten Intervallen keine Aussage über die Genauigkeit der Mittelwertdifferenzen.

## 6 Danksagung

Ich danke Frau PD Dr. M. Karpenstein-Machan (Institut für Nutzpflanzenkunde, Universität-Gesamthochschule Kassel) sowie den beteiligten Landesanstalten und Landwirtschaftskammern für die Triticale Daten. Außerdem bedanke ich mich bei Herrn Dipl. Ing. agr. V. Michel (Landesforschungsanstalt für Landwirtschaft und Fischerei Mecklenburg-Vorpommern, Gülzow) sowie den Gutachtern für hilfreiche Kommentare.

## 7 Literatur

- BRONSTEIN, I.N., Semendajew, K.A. (1989): Taschenbuch der Mathematik. Thun, Frankfurt/Main.
- EDWARDS, D.G., Berry, J.J. (1987): The efficiency of simulation-based multiple comparisons. *Biometrics* 43, 913-928.
- 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.
- Hochberg, Y., Tamhane, A. (1987): Multiple comparison procedures. Wiley, New York.
- HOCHBERG, G.Y., Weiss, G., Hart, S. (1982): On graphical procedures for multiple comparisons. *Journal of the American Statistical Association* 77, 767-772.
- HSU, J.C. (1996): Multiple comparisons. Theory and methods. Chapman and Hall, London.
- MCCULLAGH, P. (1983): Quasi likelihood functions. *Annals of Statistics* 11, 59-67.
- MCCULLAGH, P., Nelder, J. (1989): Generalized linear models. 2nd edition. Chapman and Hall, London.
- PIEPHO, H.-P. (2000): Multiple treatment comparisons in linear models when the standard error of a difference is not constant. *Biometrical Journal* (im Druck).
- SAS Institute, Inc. (1989): SAS/IML software: usage and reference, version 6, first edition. SAS Institute, Cary.
- SAS Institute, Inc. (1997): SAS/STAT software: changes and enhancements through release 6.12. SAS Institute, Cary.

*PD Dr. sc. agr. Hans-Peter Piepho ist Heisenberg-Stipendiat der DFG und arbeitet zur Zeit als Biometriker am Institut für Nutzpflanzenkunde (INK), Universität Gesamthochschule Kassel (Steinstrasse 19, 37213 Witzenhausen, email: piepho@wiz.uni-kassel.de).*

## Vertrauensintervalle für multiple Vergleiche bei ungleichen Standardfehlern der Differenzen (H.-P. Piepho)

### Zusammenfassung

*Paarweise Vergleiche von Behandlungen im linearen Modell können mit Hilfe simultaner Vertrauensintervalle für alle Differenzen durchgeführt werden. Bei einer sehr großen Zahl von Behandlungen ist die Zahl der zu betrachtenden Intervalle so groß, dass eine Interpretation schwierig bis unmöglich wird. Daher bietet sich als Alternative die Berechnung von Vertrauensintervallen um die Behandlungsmittelwerte an, welche zum paarweisen Vergleich herangezogen werden können. Bei Konstanz der Standardfehler der Mittelwertdifferenzen ist eine solche Darstellung ohne Informationsverlust möglich. Bei unbalancierten Standardfehlern/Varianzen ist dagegen eine Approximation notwendig, und es ist ein gewisser Informationsverlust unvermeidbar. Dennoch sind die erhaltenen graphischen Darstellungen in der Regel einer paarweisen Betrachtung aller Vergleiche vorzuziehen.*

**Stichworte:** *Vertrauensintervall, multiple Vergleiche, t-Test, unbalancierte Daten, lineare Optimierung, Simplex Algorithmus.*

## Confidence intervals for multiple comparisons in case of non-constant standard errors of a difference (H.-P. Piepho)

### Summary

*Pairwise treatment comparisons in the linear model can be performed using simultaneous confidence intervals for all differences. This approach becomes infeasible with a large number of treatments. An alternative is to compute confidence limits around the treatment means amenable to pairwise comparisons. When the standard error of a difference is constant, such intervals can be computed without loss of information. In case of variance imbalance an approximation is necessary, and some loss of information is unavoidable. Nevertheless, the resulting graphical displays are usually preferable to a line-by-line display of all pairwise comparisons.*

**Key words:** *Confidence interval, multiple comparisons, t-test, unbalanced data, linear programming, simplex algorithm.*