

Unterschiedliche Signifikanzentscheidungen bei signifikanter Wechselwirkung

Jörg Sellmann
 Julius Kühn-Institut
 Stahnsdorfer Damm 81
 14532 Kleinmachnow
 joerg.sellmann@julius-kuehn.de

Zusammenfassung

In unseren Getreide-Feldversuchen werden über mehrere Jahre verschiedene Strategien des Anbaus mit Hinsicht auf den Ertrag verglichen. Es gilt herauszufinden, welche Strategie den höchsten Ertrag liefert. Kommen signifikante Wechselwirkungen, z. B. mit der Zeit (Jahre) hinzu, ist es außerdem interessant herauszufinden, welche Strategien sich in welchen Jahren unterscheiden. Mit anderen Worten, wir untersuchen die Slices.

SAS stellt zur Analyse linearer Modelle verschiedenste Prozeduren mit ebenso unterschiedlichen Slice-Statements zur Verfügung: GLM und MIXED mit slice, GENMOD und PLM mit sliceby, GLIMMIX mit slicediff. Lässt man alle diese Prozeduren gleichartig formal durchlaufen, gibt es einen reproduzierbaren Unterschied in der Anzahl der Signifikanzentscheidungen. Gibt es dafür eine Erklärung? Gibt es eine beste oder richtige oder zumindest optimale Prozedur für diese Fragestellung? Hierzu wird versucht, eine Antwort geben zu können.

Schlüsselwörter: PROC MIXED, PROC GLM, PROC PLM, PROC GLIMMIX, PROC GENMOD, slice, sliceby, slicediff

1 Daten

Wie bereits in [1] praktiziert, dient der langjährige reale Feldversuch als Ideengeber für die statistische Auswertung. Es werden aber wiederum die zur Analyse verwendeten Daten simuliert. Um die Reproduzierbarkeit der Ergebnisse zu gewährleisten, wird der Zufallszahlengenerator definiert gestartet (`call streaminit(&z.)`). Die zur Diskussion stehende Wechselwirkung wird mittels `(Jahr-2011)*(nStrategie-9)` in die Daten integriert.

```
/* Konstanten */
%LET n=5;
%LET m=5;
%LET s=5;
%LET z=12;

/* Wiederholungen */
/* Mittelwert */
/* Standardabweichung */
/* Streaminit */

data Strategien;
call streaminit(&z.);
```

```
do nStrategie = 5, 10, 12;  
  do Jahr=2008 to 2014;  
    do i=1 to &n.;  
      Ertrag= 60 + nStrategie  
              + (Jahr-2011)*(nStrategie-9)  
              + rand("Normal",&m.,&s.);  
    output;  
  end;  
end;  
end;  
run;
```

2 Grundmodell

Eine erste Ansicht der Daten (Abb. 1) zeigt deutlich die unterschiedliche, jedoch klar strukturierte Entwicklung der Erträge der 3 Anbau-Strategien. Strategie A liefert kontinuierlich fallende, Strategie B nahezu gleich bleibende und Strategie C stark wachsende Erträge. Demzufolge ist (grafisch) eine Wechselwirkung zwischen Zeit (Jahr) und Anbau-Strategie vorhanden.

Zusätzlich sind die 95 % Konfidenzintervalle der einzelnen Mittelwerte eingezeichnet. Daraus lässt sich schlussfolgern, dass signifikante Unterschiede der Erträge zwischen den Anbau-Strategien zweifelsohne jahresunterschiedlich vorhanden sind.

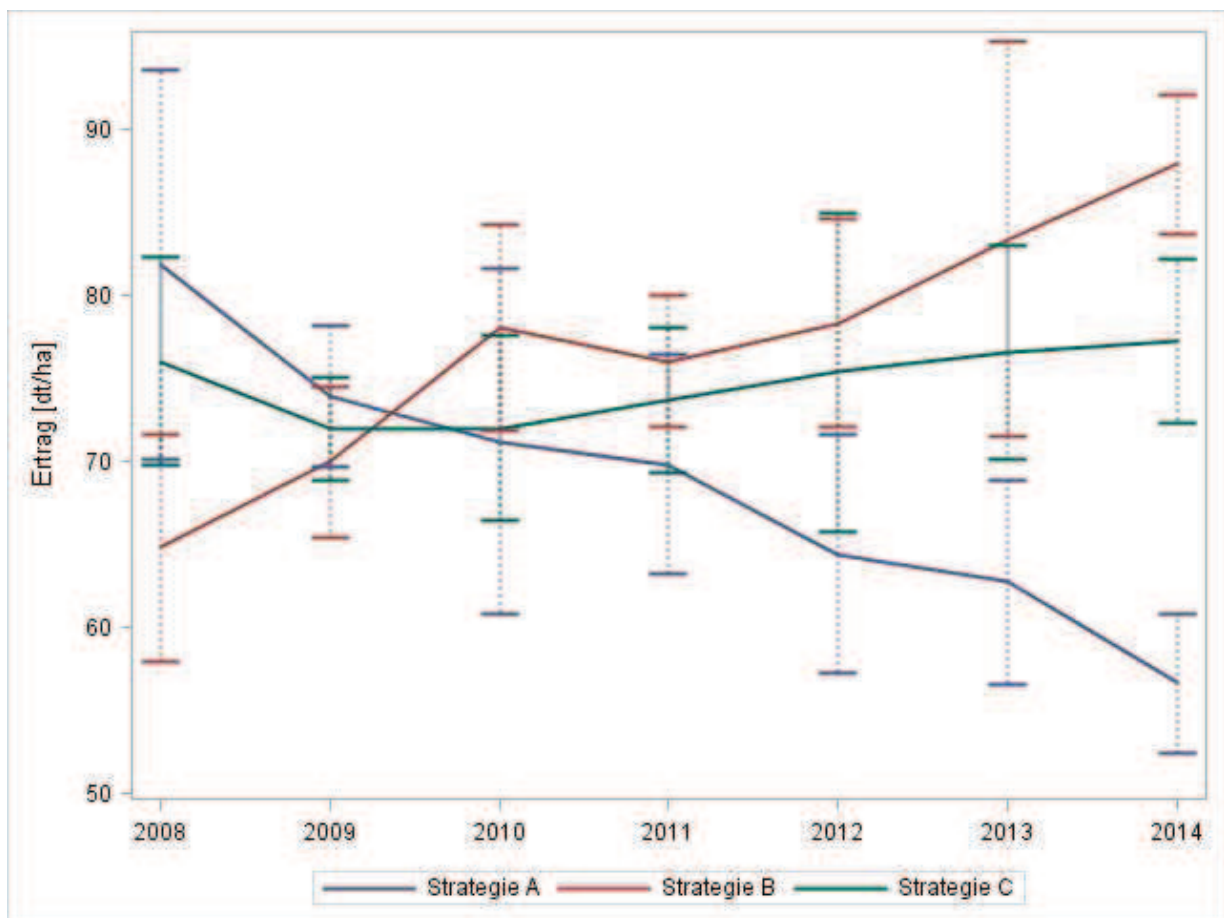


Abbildung 1: (Simulierte) Erträge der Anbau-Strategien

Folgende Fragen sind zu klären:

- Gibt es ein adäquates lineares Modell? (1)
- Gibt es Haupteffekte? (2)
- Gibt es Wechselwirkungen? (3)
- Im Fall von Wechselwirkungen: wo gibt es signifikante Unterschiede auf den einzelnen Stufen? (4)

Die Modellfrage lässt sich einfach lösen. Da bei der Generierung der Daten auf Kovarianzen und weitere `repeated` oder `subject` Effekte verzichtet wurde, reicht ein einfaches lineares Modell aus:

```
proc mixed data=Strategien;
class Jahr Strategie;
model Ertrag=Jahr|Strategie;
run;
```

Als ein Ergebnis erhalten wir den Test auf die fixen Effekte (Tab. 1), der, wie zu erwarten war, einen signifikanten Einfluss des Faktors `Strategie` (Haupteffekt) und der Wechselwirkung `Jahr*Strategie` zeigt. Das `Jahr` für sich allein genommen hat keinen signifikanten Einfluss auf die Erträge. Die Fragen (1) – (3) sind damit beantwortet.

Tabelle 1: Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Jahr	6	84	0.36	0.9043
Strategie	2	84	20.87	<.0001
Jahr*Strategie	12	84	10.66	<.0001

3 Slicing

Zur Beantwortung der Frage (4) erweitern wir das Statement um die `lsmeans` Anweisung mit der `slice` Option:

```
proc mixed data=Strategien;
class Jahr Strategie;
model Ertrag=Jahr|Strategie / ddfm=kr;
lsmeans Jahr*Strategie / slice=jahr;
run;
```

Auch dieses Ergebnis war zu erwarten (Tab. 2). Es gibt offensichtlich Unterschiede zwischen den Anbau-Strategien in den Jahren 2008 und 2012-2014.

Tabelle 2: Tests of Effect Slices

Effect	Jahr	Num DF	Den DF	F Value	Pr > F
Jahr*Strategie	2008	2	84	12.30	<.0001
Jahr*Strategie	2009	2	84	0.63	0.5368
Jahr*Strategie	2010	2	84	2.29	0.1071
Jahr*Strategie	2011	2	84	1.60	0.2072
Jahr*Strategie	2012	2	84	8.74	0.0004
Jahr*Strategie	2013	2	84	18.06	<.0001
Jahr*Strategie	2014	2	84	41.22	<.0001

Bis an diese Stelle funktionieren alle im Folgenden betrachteten Ansätze gleich. In der Frage, wo gibt es signifikante Unterschiede auf den einzelnen Stufen, zeigen sich jedoch Unterschiede in der Syntax und im Ergebnis.

3.1 PROC GLM und PROC MIXED mit slice und pdiff

Da wir hier nur das einfache lineare Modell ohne zufällige Faktoren betrachten, verhalten sich PROC GLM und PROC MIXED vollkommen identisch. Als Adjustierung zur Einhaltung des globalen Signifikanzniveaus wird immer TUKEY genutzt.

```
%LET a=tukey;                               /* Adjustierung */

proc mixed data=Strategien;
class Jahr Strategie;
model Ertrag=Jahr|Strategie / ddfm=kr;
lsmeans Jahr*Strategie / slice=jahr pdiff adjust=&a.;
store mixed;
run;
```

Der Autor war dann an dieser Stelle der Meinung, mittels pdiff im Nachgang zur slice-Anweisung die interessierenden Vergleiche auf gleicher Stufe zu erhalten. Man erhält jedoch eine endlos lange Tabelle (Tab. 3) mit den Vergleichen jeder mit jedem, also jede Kombination von Jahr und Strategie im Vergleich mit jeder beliebig anderen Kombination (210 Vergleiche).

Tabelle 3: Differences of Least Squares Means

Effect	Strategie	Jahr	Strategie	Jahr	Estimate	Standard Error	D F	t Value	Pr > t	Adjustment	Adj P
Jahr*Strat.	A	2008	B	2008	17.1078	3.5048	84	4.88	<.0001	Tukey	0.0009
Jahr*Strat.	A	2008	C	2008	5.8867	3.5048	84	1.68	0.0968	Tukey	0.9819
Jahr*Strat.	A	2008	A	2009	8.0016	3.5048	84	2.28	0.0250	Tukey	0.7610
Jahr*Strat.	A	2008	B	2009	11.9250	3.5048	84	3.40	0.0010	Tukey	0.1083
Jahr*Strat.	A	2008	C	2009	9.9058	3.5048	84	2.83	0.0059	Tukey	0.3773
Jahr*Strat.	A	2008	A	2010	10.6909	3.5048	84	3.05	0.0031	Tukey	0.2456
...											
Jahr*Strat.	B	2014	C	2014	10.6446	3.5048	84	3.04	0.0032	Tukey	0.2525

Der erhaltene Tipp war, aus der Tabelle nur die herauszusuchen, in denen das Jahr gleich war. Als Resultat erhält man 6 signifikante Differenzen (Tab. 4).

Tabelle 4: Differences of Least Squares Means (auf gleicher Jahresstufe)

Effect	Jahr	Strategie	Strategie	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
Jahr* Strat.	2008	A	B	17.1078	3.5048	84	4.88	<.0001	Tukey	0.0009
Jahr* Strat.	2012	A	B	-13.9121	3.5048	84	-3.97	0.0002	Tukey	0.0213
Jahr* Strat.	2013	A	B	-20.6743	3.5048	84	-5.90	<.0001	Tukey	<.0001
Jahr* Strat.	2013	A	C	-13.8168	3.5048	84	-3.94	0.0002	Tukey	0.0232
Jahr* Strat.	2014	A	B	-31.2931	3.5048	84	-8.93	<.0001	Tukey	<.0001
Jahr* Strat.	2014	A	C	-20.6486	3.5048	84	-5.89	<.0001	Tukey	<.0001

Der Irrtum, wie sich später herausstellte, war zu glauben, dass sich die `pdiff`-Anweisung auf die `slice`-Anweisung bezieht. Dieses ist jedoch keineswegs der Fall. Es werden demnach wirklich alle Vergleiche vorgenommen und man verschenkt sehr viel Alpha.

Im Falle von `PROC MIXED` existiert jedoch eine Alternative: analog zu `PROC PLM` gibt es eine separate `slice`-Anweisung, die syntaktisch gleich ist und vollständig identische Resultate liefert. Deshalb sei an dieser Stelle auf Abschnitt 3.3 verwiesen. Für `PROC GLM` gibt es eine solche `slice`-Anweisung nicht. Hier ist `PROC PLM` die einzig vorhandene Alternative.

3.2 PROC GLM und PROC MIXED mit by

Eine andere Möglichkeit ist, die Analyse jahresweise mit Hilfe der `by`-Anweisung vorzunehmen. Der Autor spricht jedoch bewusst nicht von Alternative.

```
proc sort data=Strategien; by jahr; run;
proc mixed data=Strategien;
  by jahr;
  class Strategie;
  model Ertrag=Strategie /ddfm=kr;
  lsmeans Strategie / pdiff adjust=&a.;
run;
```

Hierzu wird in [2] angemerkt: „Slicing ist wesentlich effektiver als ein „BY“-Statement, da beim Slicen der Restfehler nach wie vor aus allen Beobachtungen geschätzt wird, während man durch ein „BY“ den Versuch in Teilversuche zerteilen würde!“

In unserem Fall führen wir 7 Analysen zu einem vergleichswisen Risiko von $\alpha=0.05$ durch. Es treffen zwei entgegengesetzte Situationen aufeinander: Erhöhung des versuchsweisen Risikos aufgrund der 7 Einzelvergleiche auf der einen Seite und Schätzung des Restfehlers mittels weniger, evtl. homogenerer Daten auf gleicher Jahresstufe auf der anderen.

In verschiedenen Simulationsläufen kamen dabei unterschiedliche Signifikanzen heraus, mal mehr, aber auch mal weniger als in allen anderen hier diskutierten Varianten. Im konkreten Fall sind es 8 (Tab. 5).

Tabelle 5: Differences of Least Squares Means (by jahr)

Effect	Jahr	Strategie	Strategie	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
Jahr* Strat.	2008	A	B	17.1078	4.4055	12	3.88	0.0022	Tukey	0.0057
Jahr* Strat.	2012	A	B	-13.9121	3.9753	12	-3.50	0.0044	Tukey	0.0113*
Jahr* Strat.	2012	A	C	-10.9379	3.9753	12	-2.75	0.0176	Tukey	0.0431*
Jahr* Strat.	2013	A	B	-20.6743	4.3797	12	-4.72	0.0005	Tukey	0.0013
Jahr* Strat.	2013	A	C	-13.8168	4.3797	12	-3.15	0.0083	Tukey	0.0210*
Jahr* Strat.	2014	A	B	-31.2931	2.2920	12	13.65	<.0001	Tukey	<.0001
Jahr* Strat.	2014	A	C	-20.6486	2.2920	12	-9.01	<.0001	Tukey	<.0001
Jahr* Strat.	2014	B	C	10.6446	2.2920	12	4.64	0.0006	Tukey	0.0015

Legen wir als Grenze jedoch $\alpha=0.05/7$ (Jahre) an, fallen die in Tab. 5 mit * markierten Zeilen 2, 3 und 5 auch noch heraus. Dem Autor ist an dieser Stelle keine andere Möglichkeit bekannt, α statistisch korrekt zu adjustieren.

3.3 PROC PLM mit sliceby

Wollen wir statistisch korrekter arbeiten, müssen wir einen anderen Weg einschlagen. Wer in 3.1 das Statement richtig angeschaut hat, dem ist die Anweisung

```
store mixed;
```

aufgefallen. Mit dieser Anweisung wird das berechnete Modell für eine weiterführende Verarbeitung gespeichert. Diese Weiterverarbeitung geschieht mittels `restore=mixed` in PROC PLM. Der kleine syntaktische Unterschied zu PROC MIXED lautet hier `sliceby=jahr` statt `slice=jahr`. Beide Ansätze nutzen aber das Wort `pdiff`.

```
proc plm restore=mixed;
slice Jahr*Strategie / sliceby=Jahr pdiff adjust=&a.;
run;
```

Tabelle 6: Differences of Least Squares Means (jahr=2014)

Effect	Strategie	Strategie	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
Strat.	A	B	-31.2931	2.2920	12	-13.65	<.0001	Tukey	<.0001
Strat.	A	C	-20.6486	2.2920	12	-9.01	<.0001	Tukey	<.0001
Strat.	B	C	10.6446	2.2920	12	4.64	0.0006	Tukey	0.0015

Tabelle 7: Simple Differences of Jahr*Strategie Least Squares Means Adjustment for Multiple Comparisons: Tukey

Slice	Strategie	Strategie	Estimate	Standard Error	DF	t Value	Pr > t	Adj P
Jahr 2014	A	B	-31.2931	3.5048	84	-8.93	<.0001	<.0001
Jahr 2014	A	C	-20.6486	3.5048	84	-5.89	<.0001	<.0001
Jahr 2014	B	C	10.6446	3.5048	84	3.04	0.0032	0.0088

Während Tab. 6 das Ergebnis eines Jahres mittels PROC MIXED und by darstellt, ist Tab. 7 das Resultat der sliceby-Anweisung von PROC PLM. Auf den ersten Blick sehen beide Tabellen gleich aus, bei näherem Hinsehen fällt jedoch die unterschiedliche Zahl der Freiheitsgrade DF auf. In PROC PLM wird zur Schätzung des Restfehlers der komplette Datensatz herangezogen.

Zur weiteren Diskussion wird auf 3.4 verwiesen, da PROC GLIMMIX mit slicediff vollständig identische Ergebnisse zu PROC PLM geliefert hat. Das lässt den Schluss zu, dass intern die gleichen Methoden verwendet werden.

3.4 PROC GLIMMIX mit slicediff

Unter Verwendung von PROC GLIMMIX lautet das Statement jetzt

```
proc glimmix data=Strategien;
class Jahr Strategie;
model Ertrag=Jahr|Strategie /ddfm=kr;
lsmeans Jahr*Strategie /slicediff=jahr adjust=&a.;
run;
```

Als Ergebnis der slicediff-Anweisung erhält man im Gegensatz zu PROC PLM eine geschlossene Tabelle (Tab. 8), allerdings mit vollständig identischen Inhalten.

Tabelle 8: Simple Effect Comparisons of Jahr*Strategie Least Squares Means By Jahr Adjustment for Multiple Comparisons: Tukey

Simple Effect Level	Strategie	Strategie	Estimate	Standard Error	DF	t Value	Pr > t	Adj P
...								
Jahr 2014	A	B	-31.2931	3.5048	84	-8.93	<.0001	<.0001
Jahr 2014	A	B	-20.6486	3.5048	84	-5.89	<.0001	<.0001
Jahr 2014	A	C	10.6446	3.5048	84	3.04	0.0032	0.0088

Das Wort `slicediff` scheint demnach nur die Verkürzung von `slice` und `pdiff` zu sein.

Bestärkt wird diese Vermutung durch den Blick ins Handbuch [3]: „When the ADJUST= option is specified, the GLIMMIX procedure also adjusts the tests for multiplicity. The adjustment is based on the number of comparisons within each level of the SLICEDIFF= effect;“

Offensichtlich wird auch hier, analog zum Vorgehen mittels `by`, das α nur jahresweise adjustiert und nicht, wie vom Autor erhofft, global über alle (21) Vergleiche.

3.5 PROC GENMOD mit sliceby

Eine wiederum leicht veränderte Syntax liefert PROC GENMOD mit `sliceby`:

```
proc genmod data=strategien;
class Jahr Strategie;
model Ertrag=Jahr|Strategie / type3;
slice Jahr*Strategie / sliceby=jahr pdiff adjust=&a.;
effectplot interaction(sliceby=Strategie) / clm;
run;
```

Bei eingeschaltetem ODS GRAPHICS erhält man mittels `effectplot...` sehr einfach eine anschauliche Grafik über mögliche Signifikanzen (Abb. 2). Dieser `effectplot` ist nur möglich innerhalb der Statistik-Prozeduren GENMOD, LOGISTIC, ORTHOREG und PLM.

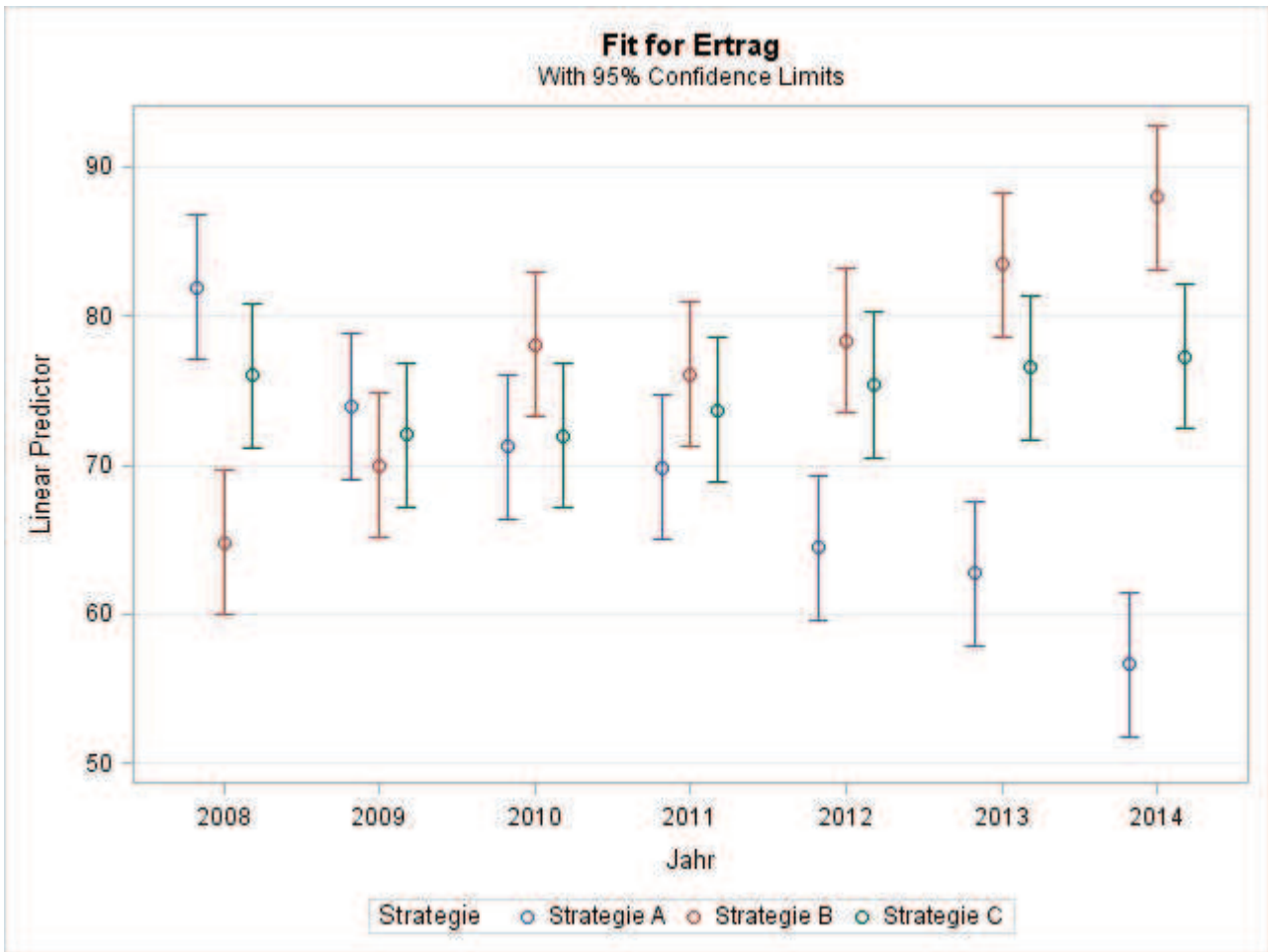


Abbildung 2: Effectplot der Prozeduren GENMOD und PLM

Im Hinblick auf unsere Slice-Fragestellung erhalten wir für jedes Jahr eine Tabelle analog zu Tab. 9 mit den paarweisen Differenzen der 3 Strategien.

Tabelle 9: Simple Differences of Jahr*Strategie Least Squares Means
Adjustment for Multiple Comparisons: Tukey

Slice	Strategie	Strategie	Estimate	Standard Error	z Value	Pr > z	Adj P
...							
Jahr 2014	A	B	-31.2931	3.1348	-9.98	<.0001	<.0001
Jahr 2014	A	C	-20.6486	3.1348	-6.59	<.0001	<.0001
Jahr 2014	B	C	10.6446	3.1348	3.40	0.0032	0.0020

Formal sieht das Ergebnis identisch aus mit den Ergebnissen aus PLM, jedoch werden hier keine F-Tests durchgeführt, sondern Tests auf Basis der Normalverteilung. Vermutet wurde als Ursache die Modellbildung, wo keinerlei Verteilungsannahmen getätigt wurden (Tab. 10). Aber auch bei der probeweisen Verwendung anderer Verteilungsannahmen wird die z-Verteilung für die Tests genutzt.

Die p-Adjustierung erfolgt jedoch wiederum auf Basis von Tukey. Daraus ergeben sich andere p-Werte und zum Teil andere Signifikanzentscheidungen.

Tabelle 10: Model Information (aus GENMOD)

Data Set	WORK.STRATEGIEN
Distribution	Normal
Link Function	Identity
Dependent Variable	Ertrag

4 Schlussfolgerungen und -fragen

Fasst man die Signifikanzentscheidungen aller betrachteten Varianten in einer Tabelle zusammen (Tab. 11), so ergibt sich das folgende Bild:

Tabelle 11: Signifikanzentscheidungen der einzelnen Prozeduren

Jahr	Strategie	_Strategie	GLM	Mixed	MixedBy	PLM	Glimmix	Genmod
2008	Strategie A	Strategie B	0.0009	0.0009	0.0057	<.0001	<.0001	<.0001
2008	Strategie B	Strategie C	-	-	-	0.0054	0.0054	0.0010
2012	Strategie A	Strategie B	0.0213	0.0213	0.0113	0.0004	0.0004	<.0001
2012	Strategie A	Strategie C	-	-	0.0431	0.0069	0.0069	0.0014
2013	Strategie A	Strategie B	<.0001	<.0001	0.0013	<.0001	<.0001	<.0001
2013	Strategie A	Strategie C	0.0232	0.0232	0.0210	0.0005	0.0005	<.0001
2014	Strategie A	Strategie B	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2014	Strategie A	Strategie C	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2014	Strategie B	Strategie C	-	-	0.0015	0.0088	0.0088	0.0020

Dividiert man formal durch 7 (Jahre), so fallen die grau unterlegten Signifikanzen auch noch heraus.

Schlussfolgernd kann der Autor lediglich 2 Empfehlungen geben:

- MIXED und GLM sollten immer zusammen mit der store-Anweisung verwendet werden, um eine Weiterverarbeitung zu ermöglichen.
- Die Analyse mittels by sollte unterlassen werden, da der gesamte Versuch in formal unabhängige Teilversuche zerlegt würde, was u. a. auch der Versuchsplanung widersprechen würde.

Dennoch bleiben (leider) einige Fragen offen:

- Ist das Vorgehen von PLM bzw. GLIMMIX etwas anderes als ein by-Statement, nur mit anderer Restfehlerschätzung? Welches Verfahren ist konservativer?
- Bringen die hier nicht untersuchten Methoden contrast, estimate und lsmeans-estimate etwas Besseres?

- Kann man von einem globalen $\alpha=0.05$ ausgehen? Die Vermutung liegt eher bei Nein. Wie muss dann das α optimal adjustiert werden?
- Wie wird bei GENMOD vorgegangen, da das Handbuch [3] bei sliceby auch nur auf die allgemeinen Erläuterungen zu slice verweist, aber die z-Verteilung verwendet wird anstelle der F-Tests in den anderen Prozeduren?

Vielleicht hat der eine oder andere Leser zu einer oder allen Fragen eine Antwort parat. Der Autor würde sich über eine Resonanz sehr freuen.

5 Programme

Das hier verwendete Programm *compare_slicing.sas*, wie auch die Programme des Autors aus den letzten Jahren, finden Sie unter:

<http://sf.julius-kuehn.de/sas>

Besonders sei hier auf den Abschnitt `/* Slice with GLM */` verwiesen, der die Matrix-Ausgabe von PROC GLM so umformt, so dass sie nahezu dem Tabellen-Layout der anderen hier verwendeten Prozeduren entspricht.

Literatur

- [1] J. Sellmann (2017): Signifikanzen grafisch veranschaulichen. In: M. Weyermann, R. Minkenber, R. Muche (Hrsg.): KSFE 2017 - Proceedings der 21. Konferenz der SAS-Anwender in Forschung und Entwicklung (KSFE); Shaker Verlag, Aachen (2017), S. 311 - 325.
- [2] A. Büchse (2006): Auswertung landwirtschaftlicher Experimente mit SAS. Eine Einführung...
https://biostatistik.uni-hohenheim.de/fileadmin/einrichtungen/biostatistik/SAS-Praktikum_HerrBuechse/Skripte/SAS_Skript_Juli_2006.pdf (15.01.2018)
- [3] SAS Institute Inc. (2017): SAS/STAT® 14.3 User's Guide. Cary, NC: SAS Institute Inc.