

Ausgangspunkt ist eine recht häufige Beobachtung: Je umfangreicher die Versuchsserien, je größer die Orthogonalität der Einzelversuche, d.h. je mehr die angelegten Serien den theoretischen Modellvorstellungen entsprechen, desto kleiner sind die statistischen Maßzahlen, die die Interaktionen PO, PJ, POJ kennzeichnen. Es liegt also die Vermutung nahe, daß in umfangreichen Serien echte prüfgliedtypische Reaktionen auf spezifische Umwelten durch die Masse der übrigen Streuungsbeiträge - also der Prüfglieder und der Umwelten, deren Reaktionen weitgehend parallel mit dem Durchschnitt gehen - überdeckt werden und in dem globalen Interaktionsterm nicht zum Ausdruck kommen können. - Daraus ergibt sich fast zwangsläufig, daß es in solchen Fällen sinnvoll ist, - vorher oder nachher - mit Aufgliederungen des Datenmaterials die gleichen Verrechnungen vorzunehmen. Mit anderen Worten: Die Aufgliederung des Datenmaterials (nach genau definierbaren Gesichtspunkten) muß hier als wichtiges Hilfsmittel der explorativen (ausforschenden) Datenanalyse angesehen werden.

Hierfür ein Beispiel: In einer Modellserie von 8 Orten und einem Sortiment von 16 Prüfgliedern wurde bei zusammenfassender Verrechnung über alle Orte ein F-Wert für die Wechselwirkung PO von nur 1,4 gefunden. Verrechnet man aber jeden Ort mit allen anderen einzeln und stellt dann die Werte zusammen, so ergibt sich eindeutig (Abbildung 1 auf Seite 218), daß der Ort 5 und der Ort 8 bei den Kombinationen mit den anderen Orten in der Regel höhere, fast immer signifikante ($p = 5\%$) F-Werte aufweisen. Diese Orte weichen also von der Gesamttendenz der Serie ab. Bei einer weiteren Analyse zeigte sich, daß bei diesen beiden Orten die meisten und die höchsten Abweichungen einzelner Prüfglieder von ihrem durchschnittlichen Leistungsverhalten gefunden werden; Abweichungen, die in Einzelfällen hoch signifikant sind. Hier brachte also die Aufgliederung der Serie die erwartete zusätzliche Information.

Dieses Verfahren ist gar nicht so neu. Bereits 1970 zeigte KOSSWIG, daß durch paarweise Vergleiche eine klare Unterscheidung der Standorte möglich ist. HAUFE (1976) erwähnte ebenfalls die Differenzierung durch eine solche Aufgliederung (s. auch Abbildung 1 auf Seite 218). Beide Veröffentlichungen erfolgten zu einem Zeitpunkt, zu dem der Begriff 'explorative Datenanalyse' im deutschen Sprachgebrauch noch weitgehend unbekannt war und 'Data-Snooping' bei den Puristen unter den Biometrikern als suspekt angesehen wurde.

In geradezu klassischer Weise demonstrierte dann BLEIHOLDER (1980) die Zweckmäßigkeit dieses Verfahrens bei Serien mehrfaktorieller Versuche. Unter seiner Mitwirkung wurde auch in der Landwirtschaftlichen Versuchsstation Limburgerhof der BASF ein Software-Paket entwickelt, mit dem generell sämtliche Versuchsserien analysiert werden (s.a. BLEIHOLDER, BEHRENDT, HOCHADEL und SARWAR, 1978). Bei dieser Auswertung kann man in der Regel drei Stufen unterscheiden:

1. Auswertung über alle Orte
2. paarweise Verrechnung der beteiligten Orte
3. aufgrund der Befunde unter 1. und 2. Vornahme von Gruppierungen und erneute Verrechnung.

Für die zuletzt genannte Stufe als Beispiel die Zusammenstellung der Werte $\beta(i)$, s und $-x$ vom Ertrag verschiedener Pflanzenschutzmaßnahmen für alle

Zusammenfassung alle 8 Orte

F-Wert Interaktion PO = 1,4

Zusammenstellung der F-Werte PO
aller 2er-Kombinationen der Orte

		Ort							
		2	3	4	5	6	7	8	
Ort	1	0,2	1,1	0,7	3,9	1,6	1,8	2,2	
	2		0,5	0,3	2,8	0,9	1,6	2,3	
	3			0,2	2,4	0,6	1,9	3,0	
	4				0,9	0,2	0,8	1,5	
	5					1,9	2,8	3,9	
	6						1,2	2,3	
	7							0,7	
	8								

Abbildung 1: Zusammenfassung alle 8 Orte F-Wert Interaktion PO = 1,4

Standorte sowie für die Gruppierungen Ertrag N1>Ertrag N2 und Ertrag N1<Ertrag N2 (Abbildung 2 auf Seite 219) (BLEIHOLDER und HAUFE, 1980).

Die Programme hierfür sind in FULL-FORTRAN IV geschrieben, der maximale Kernspeicherbedarf beträgt 342 K; die Jobs laufen auf einer Honeywell Bull 66/80 DPS mit 512 K Kernspeichergröße.

Das Prinzip der Aufgliederung und Neugruppierung im Rahmen der Gesamtanalyse hat sich auch bei Problemen bewährt, bei denen nicht die globalen Einflußgrößen (z.B. mm Niederschlag pro Jahr), sondern deren Aufsplittungen (Zeitabschnitte = Dekade, Doppeldekade, Monat etc.) die Zusammenhänge aufdecken können - und die optimale Wirkung der Länge der Zeitabschnitte in der Regel nicht bekannt ist. WEBER, BAROCKA, HAUFE und OLTSMANN (1966) beschreiben in ihrer Arbeit 'Die Beziehungen zwischen Witterungsfaktoren und Ertragsmerkmalen bei der Zuckerrübe' aufgedgliederte Regressionsrechnungen mit verschiedenen Zeitabschnitten der Witterungsfaktoren als Einflußgrößen und den Proberodungsergebnissen verschiedener Erntetermine als Zielgrößen. Interpretiert werden dann die durchschnittlichen Einfachkorrelationen und die durchschnittlichen Anteile an der multiplen Bestimmtheit für bestimmte Zeitabschnitte. Als Beispiel hierfür die Zusammenstellung der Ergebnisse für den Rübenertrag aus dieser Arbeit (Abbildung 3 auf Seite 220).

Die verwendeten ad-hoc-Programme wurden von WEBER (heute KFZ Heidelberg) in ALGOL geschrieben und liefen damals auf einer X1 in Kiel. Nach einer Mitteilung von WEBER (1981) stehen diese Programme auch heute noch in Kiel zur Verfügung und laufen dort auf einer X7.

Nicht nur für die beiden abgehandelten Problemkreise, sondern bei allen hierarchisch gegliederten Daten kann das Bedürfnis der Aufgliederung der Analyse bestehen. Das hat im wesentlichen zwei Gründe:

- o einmal die schon erwähnte Möglichkeit der Überdeckung prüfglied-typischer Reaktionen,
- o zum anderen können gefundene Reaktionen oft auf einzelne oder wenige fehlerhafte Werte zurückgeführt werden; wobei als Prüfkriterium gilt: Bei kontinuierlichen Abstufungen kann es keine signifikanten Kontinuitätssprünge geben.

Rüben-ertrag		Mittelwerte statistischer Maßzahlen aus Regressionsrechnungen, die mit dem aufgegliederten Datenmaterial durchgeführt wurden.																						
Monat / Dekade	A. Einfachkorrelationen							B.			C. Anteil an mult. B innerhalb der Witterungsfaktoren					D. Zusammensetzung der Schwerpunkte A und C								
	Nieder-schlag	rel. Luft-feuchte	Klima-Wasser-bilanz	Temperatur-mittel	Tag	min.	max.	Sonnen-schein-dauer	mult. B innerhalb der Dekade	Nieder-schlag	rel. Luft-feuchte	Klima-Wasser-bilanz	Tag	min.	max.	Sonnen-schein-dauer	Nieder-schlag	rel. Luft-feuchte	Klima-Wasser-bilanz	Tag	min.	max.	Sonnen-schein-dauer	
März	1	-15	-12	-24	.11	.05	.21	.32	.37	.10	.16	.16	.06	.07	.06	.02								
	2	.02	-16	-07	.12	.05	.17	.21	.45	.05	.09	.09	.10	.09	.04	.01								
	3	.07	-17	.06	.01	.09	.13	.08	.46	.02	.04	.04	.08	.07	.02	.02								
April	1	-17	-25	-25	.22	.03	.31	.40	.53	.03	.06	.05	.06	.17	.09	.14								
	2	-25	-34	-36	.23	.04	.26	.28	.49	.04	.04	.05	.11	.23	.10	.14								
	3	-02	-18	-13	.13	-01	.04	.02	.45	.05	.01	.04	.16	.16	.07	.06								
Mai	1	.09	-11	.02	-04	-10	-03	-14	.32	.04	.03	.03	.11	.17	.06	.11								
	2	.13	-05	.10	.12	.08	.10	-20	.26	.02	.05	.02	.03	.14	.03	.06								
	3	-01	-10	-03	.15	.06	.04	-07	.24	.04	.08	.06	.03	.09	.02	.03								
Juni	1	-38	-20	-39	.16	.01	.15	.20	.34	.10	.05	.10	.03	.15	.01	.05								
	2	-33	-16	-36	.27	.14	.30	.27	.37	.09	.12	.07	.18	.23	.11	.07								
	3	-15	-21	-27	.43	.25	.43	.36	.42	.03	.11	.04	.36	.24	.11	.06								
Juli	1	-33	-40	-40	.13	.09	.30	.31	.47	.13	.17	.16	.16	.17	.06	.01								
	2	-48	-43	-56	-08	-01	.18	.34	.49	.15	.19	.16	.07	.15	.06	.02								
	3	-33	-36	-37	-17	-15	.07	.21	.51	.04	.03	.04	.18	.15	.03	.04								
Aug.	1	.18	-31	.02	-02	-09	.08	.02	.55	.04	.01	.03	.17	.02	.05	.03								
	2	.10	-27	.22	.14	.19	.19	-08	.53	.18	.05	.15	.06	.20	.04	.02								
	3	.35	-21	.22	-05	.12	-04	-17	.54	.16	.05	.17	.02	.24	.03	.06								
Sept.	1	-05	-26	-11	-28	-20	-17	.14	.67	.05	.05	.04	.00	.07	.03	.06								
	2	-10	-14	-10	-13	.04	-06	.05	.51	.06	.10	.02	.09	.10	.04	.04								
	3	.07	-11	.08	-12	.08	-12	.01	.45	.02	.05	.00	.17	.15	.04	.08								
B. mult. B innerhalb eines Witterungsfaktors										.73	.69	.75	.64	.64	.47	.56								

Abbildung 3: Rüben-ertrag

Auf Anregung von GEIDEL, Hohenheim wurden bei der KWS Programme entwickelt, mit denen hierarchisch gegliederte Daten verrechnet werden konnten. Die Verrechnung wird bei jedem Gruppenwechsel getrennt durchgeführt, und bei jedem Übergang zu einem höheren Faktor wird eine Zusammenfassung vorgenommen. Beispielsweise für 3 Faktoren (A = 1, 2; B = 1, 2; C = 1, 2):

A	B	C
1	1	1
1	1	2
1	1	.
1	2	1
1	2	2
1	2	.
1	.	.
2	1	1
2	1	2
2	1	.
2	2	1
2	2	2
2	2	.
2	.	.
.	.	.

Auf dieser Grundlage wurden für verschiedene Probleme Programme entwickelt:

Programm L08L00 = Mittelwerte, Streuung, Matrix der Einfachkorrelationen.

Max. 5 Faktoren (A, B, C, D, E);

Faktor A mit max. 999, übrige Faktoren mit max. 9 Stufen. Das Programm besteht aus einer ROOT-Phase und 5 Nachladephasen, ist in Full-FORTRAN IV geschrieben und hat einen max. Kernspeicherbedarf von 110 K.

Programm L07L00 = Verteilungsmatrix für 2 Variable mit Randsummen, dazu in den Schnittpunkten Mittelwerte einer 3. Variablen (Abbildung 4 auf Seite 222).

Max. 3 Faktoren (A, B, C); Faktor A = max. 999, Faktor B und C = max. 9 Stufen.

ROOT-Phase + 3 Nachladephasen, FULL-FORTRAN IV, max. 55 K Kernspeicherbedarf.

Programm L22L00 = Varianzanalysen - Einfachklassifikation und Blockversuche -: Varianztabelle, Prüfgliedmittelwerte sowie s und v für die Prüfglieder - jeweils für 1 Merkmal.

Max. 5 Faktoren (A, B, C, D, E); Faktor A = max. 999, Faktor B = max. 200, Faktor C, D, E = max. 9 Stufen.

ROOT-Phase + 5 Nachladephasen, FULL-Fortran IV, max. 100 K Kernspeicherbedarf.

Nach meinen Informationen gibt es Programmsysteme nach dem gleichen Prinzip für gleiche oder ähnliche Probleme im RZ Hohenheim und im RZ der FAL Braunschweig-Völkenrode.

Bei den bisher beschriebenen Verfahren wurden im wesentlichen für die Aufgliederungen und Neugruppierungen die gleichen Verrechnungsmethoden verwendet wie für die Gesamtanalyse. Man kann aber auch in bestimmten Fällen eine hohe Transparenz schon bei der Gesamtanalyse erreichen, wenn man entsprechende Erweiterungen vornimmt:

Bei der routinemäßigen Varianzanalyse einfach gegliederter Versuchsserien ist bei der KWS eine sogenannte 'weitere Auswertung' (=Programm J15J00) enthalten (die per Option ganz oder teilweise abgehängt werden kann).

1 = Merkmal 3
 J = Merkmal 4
 Mittelwerte von Merkmal 2

Zusammenfassung über Faktor C und Faktor B
 Faktor A = 4

	1	2	3	4	5	6	7	8	9	10	
1	103 2.21	60 5.80	15 17.31	0	0	0	0	0	0	0	178 4.69
2	3 10.75	33 16.50	26 19.42	5 41.03	3 87.74	0	0	1 25.14	0	0	71 22.18
3	0	3 18.62	12 9.30	6 44.86	6 57.59	5 67.57	1 79.50	0	0	0	33 43.62
4	0	1 31.40	5 41.70	4 49.25	5 65.09	4 58.22	0	1 68.31	0	0	20 53.18
5	0	0	1 39.28	1 92.51	6 70.77	3 65.45	2 97.48	0	0	0	13 72.90
6	0	0	0	2 64.29	1 35.24	3 113.95	2 90.98	0	0	0	8 85.95
7	0	0	0	1 37.95	2 56.17	4 83.84	0	0	0	0	7 69.38
8	0	0	0	0	1 72.88	1 83.05	0	0	0	105,60	3 87.18
9	0	0	0	0	0	1 91.04	0	0	0	0	1 91.04
10	0	0	0	0	0	105.70	0	0	0	0	1 105.70
	106 2.45	97 10.10	59 23.12	19 48.97	24 65.80	22 78.37	5 91.28	2 46.73	0 0.0	1 105,60	335 22.37

Abbildung 4: Zusammenfassung über Faktor C und Faktor B

Diese 'weitere Auswertung' enthält:

- o 3 Zwei-Wegetafeln Prüfglieder/Orte mit
- o Absolutwerten
- o Relativwerten
- o Rangordnungsziffern
- o ein Tableau der Differenzen; dazu $-/x$ und 2 Stabilitätsparameter für die Prüfglieder; dafür als Beispiel ein Auszug aus einer entsprechenden Auf-tabellierung (Abbildung 5 auf Seite 224):

Kapazität: max. 169 Prüfglieder, max. 20 Orte; es können max. 19 Merkmale nacheinander verarbeitet werden. Das Programm besteht aus ROOT-Phase + 15 Nachladephasen, FULL-FORTRAN IV, max. 85 K Kernspeicherbedarf.

Für mehrfach gegliederte Serien (Orte/Jahre) wurde das jeder zusammenfassenden Verrechnung vorangehende Einlese-Programm in der Weise erweitert, daß nicht nur die gewünschten Prüfglieder der gewünschten Versuche aus der Archivdatei selektiert und auf eine Hilfsdatei ausgegeben werden, sondern es werden auch sämtliche Parameter, die für eine Aufgliederung der Analyse erforderlich sind, automatisch aufgrund der für die Gesamtanalyse notwendigen Angaben erstellt. Beispiel für die Aufgliederung einer Serie über die Orte A, B und C in den Jahren 1979 und 1980:

- o Ort A über 1979 und 1980
- o Ort B über 1979 und 1980
- o Ort C über 1979 und 1980
- o Jahr 1979 über die Orte A, B, C
- o Jahr 1980 über die Orte A, B, C

Erst nach der Verrechnung dieser Aufgliederungen erfolgt die gemeinsame Analyse über alle Orte und alle Jahre. Anschließend wird die doppelt gegliederte Serie als einfach gegliederte Serie betrachtet und wieder das Programm 'weitere Auswertung' eingesetzt. Musterbeispiel einer solchen Recherche bringt dieser Auszug von Kartoffelversuchen aus der Türkei (Abbildung 6 auf Seite 225) über das Verhalten der mitgeprüften Sorte 2. Die stark unterschiedlichen Reaktionen bei den Verrechnungen für die Orte und für die Jahre kann man in diesem Falle sogar sachlogisch interpretieren.

Die beschriebenen Aktivitäten für die Orts-/Jahres-Serien werden durch 6 Programme (Mehrphasenprogramme, FULL-FORTRAN IV) im Rahmen des geschlossenen Verrechnungssystems (ZGES) abgedeckt.

Alle vorgestellten Programme der KWS werden z.Z. auf einer IBM 370-125 mit 512 K gefahren.

Ausgangswert und Regressionsgerade(β_1) des Ertrages (dt/ha) der Kartoffelsorte 2 der in der Türkei durchgeführten Versuche
 Die Daten wurden

- a) in Jahre und
- b) in Standorte aufgliedert.

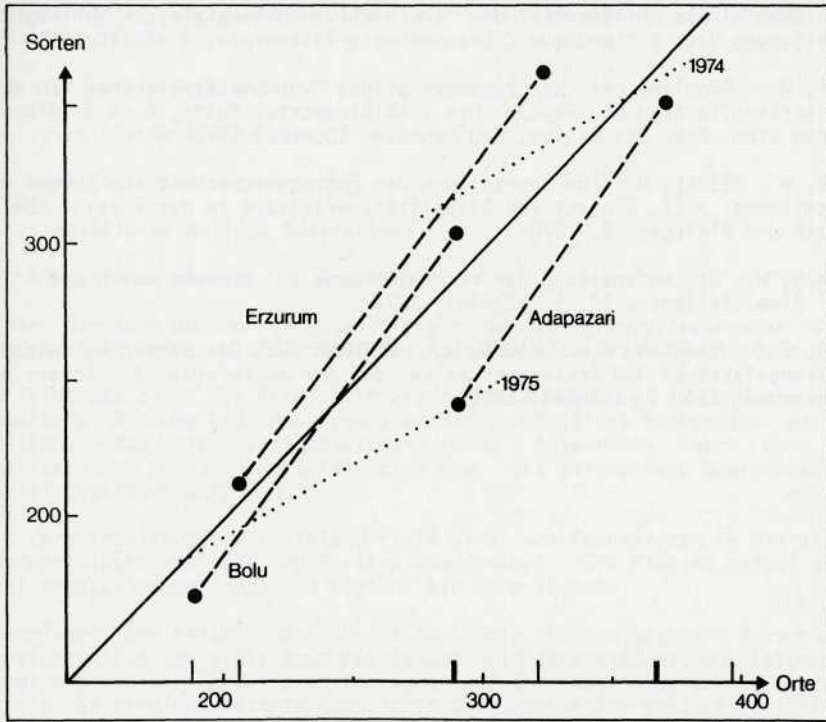


Abbildung 6: Ausgangswert und Regressionsgerade des Ertrages (dt/ha) der Kartoffelsorte 2 der in der Türkei durchgeführten Versuche
 Die Daten wurden in Jahre und Standorte aufgliedert.

Literatur

1. BLEIHOLDER, H., BEHRENDT, S., HOCHADEL, H., SARWAR, E.: Organisation eines Datenbanksystems zur Speicherung und Bewertung von Versuchsergebnissen aus Feldversuchen mit Pflanzenbehandlungsmitteln; 1978; Z. Pflanzenkrankheiten und Pflanzenschutz, 86 (7), S. 385-440, 1979 sowie Vortrag 3. Intern. Congr. of Plant Pathologie, München, 16.-23. Aug. 1978.
2. BLEIHOLDER, H., HAUFE, W.: Zusammenfassende Verrechnung mit aufgegliedertem Datenmaterial als Hilfsmittel der explorativen Datenanalyse. Vortrag 24. Jahrestagung Ges. f. Landbauw., Braunschweig-Völkenrode, 9.-10.10.1980
3. HAUFE, W.: Probleme bei der Zusammenfassung von Einzelversuchen zu Versuchsserien; Vortrag Dt. Phytom. Ges., AK Biometrie, Fulda, 4.-5.3.1976 und Vortrag Biom. Ges. Dt. Region, Bad Nauheim, 10.-12.3.1976
4. HAUFE, W., GEIDEL, H.: Zur Beurteilung der Ertragssicherheit von Sorten und Zuchtstämmen. - II. Einsatz von Stabilitätsparametern in der Praxis. EDV in Medizin und Biologie, 2, 1978
5. KOSSWIG, W.: Die Aufspaltung der Wechselwirkung 1. Ordnung aus $n \times k$ - Tafeln. Biom. Zeitspr., 12, S. 137-151, 1970
6. WEBER, E.E., BAROCKA, K.H., HAUFE, W., OLTSMANN, W.: Die Beziehung zwischen Witterungsfaktoren und Ertragsmerkmalen bei der Zuckerrübe. Z. Acker- und Pflanzenbau, 124, S. 134-164, 1966