

Die Ermittlung von Korrelationskoeffizienten zwischen aufeinanderfolgenden Messungen in pflanzenzüchterischen Versuchen

W.E. Weber

Institut für Angewandte Genetik der Technischen Universität Hannover (BRD)

The Question of Correlation Coefficients between Successive Measurements in Plant Breeding Research

Summary. The analysis of variance normally is conducted under the assumption of no correlations between residuals. Sometimes this assumption is violated in plant breeding experiments. If it is ignored, the estimates of mean squares and variance components are biased. For some restricted models the correlation is estimated from experiments with Gerbera, tulips and potatoes. Often a positive correlation was found, but never a negative one. The consequences of such correlations are discussed.

Zusammenfassung. Die Annahme nicht korrelierter Restabweichungen, von der bei der Varianzanalyse in der Regel ausgegangen wird, trifft in pflanzenzüchterischen Versuchen häufig nicht zu. Aus Versuchen mit Gerbera, Tulpen und Kartoffeln wurde die Korrelation unter der Annahme verschiedener Modelle geschätzt. In vielen Fällen trat eine positive Korrelation auf, aber niemals eine negative. Die Auswirkung solcher Korrelationen auf die Interpretation von Versuchsergebnissen wird aufgezeigt.

1. Einleitung

Eine wichtige Aufgabe der Pflanzenzüchtung ist es, quantitative Eigenschaften der Kulturarten zu verbessern. Dabei stellt die Heritabilität eine häufig benutzte Maßzahl zur Abschätzung des Selektionserfolges dar. Neben Eltern-Nachkommenschafts-Regressionsanalysen können vor allem Varianzkomponenten aus Versuchen nach hierarchischen, faktoriellen und diallelen Kreuzungsplänen für die Ermittlung des Erblichkeitsanteiles benutzt werden. Die Schätzung von Varianzkomponenten ist zwar verhältnismäßig einfach durchzuführen, die Schätzwerte sind aber i. a. mit einem großen Fehler behaftet. Abweichungen von den Voraussetzungen für die Varianzanalyse wirken sich stark auf die Abschätzung von Varianzkomponenten aus. Eine der Voraussetzungen für die Varianzanalyse ist, daß die Restabweichungen - in dieser Arbeit kurz Reste genannt - nicht korreliert sind. Diese Voraussetzung ist jedoch in pflanzenzüchterischen und pflanzenbaulichen Versuchen nicht immer erfüllt. Werden solche Korrelationen nicht beachtet, führt das zu falschen Schätzwerten für die mittleren Abweichungsquadrate und damit auch zu verzerrten Schätzungen für die Varianzkomponenten, während die Mittelwerte nach wie vor erwartungstreu bleiben.

In dieser Arbeit soll an pflanzenzüchterischen Versuchen bei Gerbera, Tulpen und Kartoffeln erläutert werden, welchen Einfluß solche Korrelationen auf die Größe von MQ-Werten und Varianzkomponenten haben und wie diese korrigiert werden können.

2. Modelle und Schätzmethodik

Korrelationen zwischen Restabweichungen können auftreten, wenn wiederholte Messungen an derselben Pflanze vorgenommen werden. Dieser Fall tritt z.B. ein, wenn an einer Zierpflanze Meßwerte an mehreren aufeinanderfolgenden Blüten ermittelt werden (Gerbera), oder auch, wenn Klone jedes Jahr geprüft und zwischendurch vegetativ weitervermehrt werden (Tulpen und Kartoffeln). Solche Versuche können durch das folgende lineare Modell für eine Zweiwegklassifikation beschrieben werden (Modell 1):

$$y_{ij} = \mu + a_i + m_j + e_{ij} \quad (1)$$

mit $e_{ij} \sim N(0, \sigma^2)$;

$$E(e_{ij}^2) = \sigma^2; E(e_{ij}e_{ik}) = \rho_{jk}\sigma^2;$$

übrige Kovarianzen Null.

Dabei ist y_{ij} = Beobachtungswert
 μ = allgemeiner Mittelwert
 a_i = Effekt des i -ten Prüfgliedes, $i = 1, p$
 m_j = Effekt der j -ten Messung, $j = 1, J$
 e_{ij} = Restabweichung = Rest.

Tab. 1. Varianztabelle zum Modell $y_{ij} = \mu + a_i + m_j + e_{ij}$
 $E(e_{ij}^2) = \sigma^2$; $E(e_{ij}e_{ik}) = \rho_{jk}\sigma^2$; übrige Kovarianzen Null

Varianz- ursache	FG	E(MQ)
Prüfglieder (A)	p-1	$\sigma^2 \left(1 + \frac{2}{J} \sum\right) + J \sigma_A^2$
Messungen (M)	J-1	$\sigma^2 \left(1 - \frac{2}{J(J-1)} \sum\right) + p \sigma_M^2$
Rest	(p-1)(J-1)	$\sigma^2 \left(1 - \frac{2}{J(J-1)} \sum\right)$

$$\sum = \sum_{j=1}^{J-1} \sum_{k=j+1}^J \rho_{jk}$$

Tabelle 1 gibt für dieses Modell die Erwartungswerte für MQ an, wenn die Varianzanalyse wie üblich durchgeführt wird. Sind die ρ_{jk} nicht Null, dann ergeben sich Fehler bei der Schätzung von MQ-Werten und damit der Varianzkomponenten für Prüfglieder, Messungen und den Rest. Ist $\sum_{j=1}^{J-1} \sum_{k=j+1}^J \rho_{jk}$ positiv, wird die Restvarianz unterschätzt und die Prüfgliedvarianz überschätzt und umgekehrt. Es ist nun nicht möglich, die ρ_{jk} und damit die Restvarianz σ^2 in Modell 1 zu schätzen, obwohl aus der Matrix $\left\{ \sum_{i=1}^p \hat{e}_{ij} \hat{e}_{ik} \right\}$ ersichtlich ist, ob die Annahme nicht korrelierter Reste zutrifft. Wenn nämlich die ρ_{jk} alle Null sind, dann sind die Erwartungswerte $E(\hat{e}_{ij}^2) = \frac{J-1}{J} \sigma^2$ und $E(\hat{e}_{ij} \hat{e}_{ik}) = -\frac{1}{J} \sigma^2$ (siehe Anscombe und Tukey 1963). Weber (1976) hat einige Modelle beschrieben, bei denen die Matrix $R = \{\rho_{jk}\}$ mit J Zeilen und Spalten gewissen Einschränkungen unterliegt, und Methoden für die Schätzung der Korrelationskoeffizienten und der Restvarianz vorgeschlagen. Folgende Modelle werden bei den hier zu betrachtenden Versuchen benutzt und deren Schätzmethoden kurz skizziert (Einzelheiten bei Weber 1976): Modell 2: $\rho_{jk} = \prod_{s=j}^{k-1} \rho_{s, s+1}$, d.h., nur die unmittelbar vorhergehenden Messungen wirken direkt, weiter zurückliegende nur indirekt auf den Rest

ein. Neben der Restvarianz sind somit $(J-1)$ Korrelationskoeffizienten zu schätzen. Für $J = 3$ ergibt sich die folgende Lösung:

$$\begin{aligned} \sigma^2 &= (2b - a - d)(a + 4b + 4d)/(36a + 36b - 72d) \\ \rho_{12} &= (3a + 6b)/(a + 4b + 4d) \\ \rho_{23} &= (3a - 3d)/(a - 2b + d) \\ \rho_{13} &= \rho_{12} \cdot \rho_{23} \end{aligned} \quad (2)$$

mit

$$a = cE\left(\sum_i \hat{e}_{i1}^2\right); b = cE\left(\sum_i \hat{e}_{i1} \hat{e}_{i2}\right); d = cE\left(\sum_i \hat{e}_{i2}^2\right)$$

und $c = J^2/(p-1)$.

Werden bei a , b und d statt der Erwartungswerte die Schätzwerte eingesetzt, so erhält man Schätzungen für σ^2 , ρ_{12} und ρ_{23} , die allerdings nicht erwartungstreu sind, da i.a. $E(xy) \neq E(x)E(y)$ und $E(x/y) \neq E(x)/E(y)$ ist. Bei $J > 3$ sind mehr nicht-lineare Gleichungen zu lösen als Parameter zu schätzen.

Modell 3: $\rho_{jk} = \rho_{|j-k|}$, d.h., die Größe des Korrelationskoeffizienten hängt nur vom Abstand der Messungen ab. Auch in diesem Falle gibt es neben der Restvarianz $(J-1)$ verschiedene Korrelationskoeffizienten, es können aber maximal nur $(J-1)$ Parameter geschätzt werden, da bei diesem Modell aus den geschätzten Resten nur $(J-1)$ linear unabhängige Gleichungen gebildet werden können. Sicherlich werden aber bei genügend großem J die Korrelationskoeffizienten ρ_{J-1} , ρ_{J-2} u.s.w. klein und damit vernachlässigbar sein. Mit Hilfe eines Regressionsansatzes können dann σ^2 , $\rho_1 \sigma^2$, $\rho_2 \sigma^2$ u.s.w. geschätzt werden: Der Erwartungswert für $\sum_i \hat{e}_{ij} \hat{e}_{ik}$ ist eine lineare Funktion der Parameter σ^2 , $\rho_1 \sigma^2$, $\rho_2 \sigma^2$ u.s.w., wobei die Koeffizienten der Parameter bekannt sind und vom Versuchsumfang abhängen. Werden die Koeffizienten der Parameter als x -Werte und die $\sum_i \hat{e}_{ij} \hat{e}_{ik}$ als y -Werte betrachtet und geht die Regressionsfunktion durch den Anfangspunkt, dann sind die Regressionskoeffizienten b_0 , b_1 , b_2 u.s.w. bis b_{J-2} Schätzwerte für σ^2 , $\rho_1 \sigma^2$, $\rho_2 \sigma^2$ u.s.w. bis $\rho_{J-2} \sigma^2$. Die üblichen Regressionsprogramme liefern einen Test, ob die Hinzunahme einer weiteren x -Variablen lohnt. Wird der Reihe nach zu-

nächst nur b_0 geschätzt, dann neben b_0 auch noch b_1 , dann zusätzlich b_2 u.s.w., so kann man ermitteln, bis zu welchem Grade die Korrelation zwischen Resten bedeutend ist. Für die Korrelationskoeffizienten gilt dann $\rho_i = E(b_i)/E(b_0)$. Auch in diesem Fall ist die Schätzung b_i/b_0 nicht erwartungstreu.

Modell 4: $\rho_{jk} = \rho^{|j-k|}$, d.h., es gelten die Einschränkungen von Modell 2 und Modell 3. In diesem Fall ist neben der Restvarianz nur ein Korrelationskoeffizient zu bestimmen. Schätzt man, wie bei Modell 3 ausgeführt, b_0 und b_1 , so ist

$$E(b_1)/E(b_0) = \frac{\sum_{j=1}^{J-2} c_j \rho^j}{\sum_{j=0}^{J-2} d_j \rho^j} \quad (3)$$

wobei die Koeffizienten c_j und d_j nur von der Zahl der Messungen abhängen und bekannt sind. Wird $E(b_1)/E(b_0)$ durch b_1/b_0 ersetzt, ergibt sich für ρ ein Polynom $(J-2)$ -ten Grades. Die Lösung von ρ ist nicht erwartungstreu, da $E(b_1)/E(b_0) \neq E(b_1/b_0)$. Weber (1976) fand, daß für eine genügend große Zahl von Prüfgliedern (mindestens 50 Prüfglieder bei nur 3 Messungen, bei mehr Messungen weniger Prüfglieder erforderlich) die Differenz zwischen $E(b_1)/E(b_0)$ und $E(b_1/b_0)$ vernachlässigt werden kann. Dieses Modell sowie das folgende werden auch von Larson (1969) betrachtet.

Modell 5: Für $j \neq k$ gilt: $\rho_{jk} = \rho$ für $|j-k| = 1$, ρ_{jk} sonst Null. Auch bei diesem Modell ist neben der Restvarianz nur ein Korrelationskoeffizient zu schätzen. Wie bei Modell 3 ausgeführt, sind b_0 und b_1 Schätzungen für σ^2 und $\rho\sigma^2$. Der Quotient b_1/b_0 ist unter den gleichen Einschränkungen wie bei Modell 4 eine brauchbare Schätzung für ρ . Bei diesem Modell wird angenommen, daß eine Messung nicht mit der übernächsten korreliert ist, auch wenn sie mit der nächsten sehr eng korreliert ist, eine Annahme, die in vielen pflanzenzüchterischen Versuchen wohl nicht zutrifft.

Modell 6: Zusätzlich zu den bei Weber (1976) betrachteten Modellen soll hier noch ein Modell für eine einfache Varianzanalyse herangezogen werden, das u.a. von Durbin (1960), Zellner und Tiao (1964) sowie von Tiao und Tan (1966) untersucht wurde. Es lautet:

$$y_{ij} = \mu + a_i + e_{ij} \quad (4)$$

mit $e_{ij} = \rho \cdot e_{i,j-1} + \varepsilon_{ij}$, alle $\varepsilon_{ij} \sim N(0, \sigma^2)$ und nicht korreliert. Damit auch Werte für y_{i1} existieren, muß man Werte für e_{i0} annehmen, etwa $e_{i0} = 0$. Tiao und Tan geben zwei Schätzmethoden für ρ an, einmal mittels eines Bayes-Ansatzes, zum anderen nach der Maximum Likelihood-Methode. Letztere soll auch hier herangezogen werden.

Die unter der Annahme eines der Modell geschätzten Korrelationskoeffizienten können nun zu einer Korrektur der MQ-Werte unter Berücksichtigung der in Tab.1 angegebenen Erwartungswerte benutzt werden.

3. Experimentelle Ergebnisse

3.1. Gerbera

3.1.1. Beschreibung der Versuche

Die drei Versuche mit Gerbera, die hier besprochen werden, sind bei Horn u.a. (1974) ausführlich beschrieben. Sie waren zur Untersuchung der Vererbung quantitativer Eigenschaften durchgeführt worden und sollen hier im Hinblick auf die Korrelation zwischen Messungen kurz skizziert werden. Es wurden nur Pflanzen mit mindestens 8 Blütenständen (bei Versuch II mit 7 Blütenständen) betrachtet. An den Blütenständen - im folgenden kurz "Blüten" genannt - wurden folgende Eigenschaften untersucht: Stengellänge (in cm), Durchmesser des Blütenstandes (in mm), Blütenscheibendurchmesser (in mm) und Zungenblütenzahl (180°).

Der Versuch I enthielt 64 Klone, die in zwei Wiederholungen (Blöcken) über zwei Vegetationsperioden untersucht wurden. 17 Klone brachten in jeder Wiederholung und Vegetationsperiode eine Pflanze mit mindestens 8 Blüten, nur die ersten 8 Blüten werden hier ausgewertet. Bei den Untersuchungen soll nicht zwischen Klonen, Wiederholungen und Vegetationsperioden unterschieden werden. Es standen also insgesamt $17 \times 2 \times 2 = 68$ Pflanzen für eine Vegetationsperiode zur Verfügung, d.h., $p = 68$ und $J = 8$.

Der Versuch II umfaßte 28 F_1 -Nachkommenschaften einer diallelen Kreuzung aus 8 Klonen. 11 der 28 Nachkommenschaften produzierten in jeder der 4 Wiederholungen (Blöcke) eine Pflanze mit mindestens 7 Blüten. Der Versuch lief über eine Vegetationsperiode. Somit ist $p = 44$ und $J = 7$.

Der Versuch IIa enthielt dasselbe Material. Bei ihm wurden jedoch von den Einzelwerten vor Beginn der Analyse die Wochendurchschnittswerte (Gerbera blühen in mehr oder weniger unregelmäßigen Abständen) abgezogen. Dadurch war es möglich, die jahreszeitlichen Einflüsse auszuschalten (für Einzelheiten siehe Horn u.a., 1974).

Im Versuch III wurden 80 Kreuzungen zwischen I_0 - und I_2 -Linien in 4 Wiederholungen (Blöcken) mit je 5 Sämlingen über eine Vegetationsperiode geprüft. Zahlreiche Sämlinge fielen aus oder blühten wenig, so daß nur bei 20 Kreuzungen in jeder Wiederholung eine Pflanze 8 Blüten ausbildete. Zur Auswertung gelangte wieder eine Vegetationsperiode. Bei diesem Versuch ist daher $p = 80$ und $J = 8$.

Der Versuch IIIa enthielt dieselben Prüfglieder wie Versuch III, wobei wie bei Versuch IIa die jahreszeitlichen Einflüsse durch Abzug der Wochendurchschnittswerte ausgeschaltet wurden.

3.1.2. Ergebnisse

In einer früheren Arbeit hatten Weber und Wricke (1973) die Kovarianzen zwischen den geschätzten Resten $\hat{\epsilon}_{ij}$ und $\hat{\epsilon}_{ik}$ bestimmt, wobei die Schätzung wie üblich nach der Methode der kleinsten Quadrate unter der Annahme nicht korrelierter Reste erfolgte.

Bildet man $r_{jk} = \frac{\sum_i (\hat{\epsilon}_{ij} \hat{\epsilon}_{ik})}{(\text{SQ Rest}/J)}$, so ist

$E(r_{jk}) = -\frac{1}{J-1}$, wenn die Reste nicht korreliert sind.

In Tab.2 sind die Werte von r_{jk} bei Versuch I für die Eigenschaften Stengellänge und Zungenblütenzahl angegeben. Während bei der Stengellänge $r_{j,j+1}$ stets und im Schnitt auch $r_{j,j+2}$ größer als der Erwartungswert ist, schwanken bei der Zungenblütenzahl die Einzelwerte mehr oder weniger zufällig um den Erwartungswert. Weber und Wricke schlossen daraus, daß bei der Stengellänge die Reste zwischen aufeinanderfolgenden Messungen positiv korreliert sein müssen. Da auch $r_{j,j+2}$ im Schnitt größer als der Erwartungswert war und die übrigen r_{jk} dementsprechend kleiner, scheint Modell 4 die Situation am besten zu beschreiben.

Tabelle 3 enthält für Versuch I den geschätzten Korrelationskoeffizienten nach Modell 4, die MQ-Werte sowie für die Variationsursache Pflanzen die Varianzkomponente und den F-Wert ohne und mit Berücksichtigung der Korrelation. Die bei der Stengellänge ermittelte Korrelation führt zu einer beträchtlichen Veränderung der Werte, der Rest wird unterschätzt, der Einfluß der

Pflanzen dagegen überschätzt. Beim Blütenscheibendurchmesser ist die Korrelation schwächer, bei den anderen Eigenschaften nicht feststellbar. Werden die bei Weber (1976) angegebenen Vertrauensbereiche für den Wert b_1/b_0 (Tab.4) benutzt, so ergibt sich, daß die Nullhypothese, d.h., es liegt keine Korrelation

Tab.2. Normierte Kovarianzmatrix r_{jk} des Versuches I bei Gerbera

$\bar{r} = -1/7 = -0,143$

a) Stengellänge

j	k						
	2	3	4	5	6	7	8
1	0,25	0,14	-0,10	-0,53	-0,36	-0,31	-0,24
2		0,24	0,06	-0,48	-0,32	-0,31	-0,44
3			-0,12	-0,27	-0,23	-0,21	-0,22
4				0,12	-0,17	-0,30	-0,18
5					0,34	-0,03	-0,27
6						-0,04	-0,16
7							0,16

b) Zungenblütenzahl

j	k						
	2	3	4	5	6	7	8
1	-0,05	-0,00	-0,12	-0,11	-0,36	-0,18	-0,18
2		-0,15	-0,07	-0,33	-0,16	-0,03	-0,25
3			-0,27	-0,24	-0,02	-0,02	-0,25
4				0,05	-0,18	-0,28	-0,19
5					-0,30	-0,21	-0,01
6						0,18	-0,11
7							-0,16

Tab.3. Schätzwerte aus Versuch I bei Gerbera (unkorrigiert und nach Modell 4 korrigiert)

	Stengellänge (cm)		Durchmesser des Blütenstandes (mm)	
	unkorrigiert	korrigiert	unkorrigiert	korrigiert
b_1/b_0		0,32		0,05
r		0,35		0,05
MQ Pflanzen	523,55	479,59	1369,35	1363,83
MQ Blüten	273,51	279,79	510,70	511,49
MQ Rest	43,42	49,70	54,49	55,28
s^2 Pflanzen	60,02	53,74	164,36	163,57
F-Wert (Pflanzen)	12,06	9,65	25,13	24,67
	Blütenscheibendurchmesser (mm)		Zungenblütenzahl (180°)	
	unkorrigiert	korrigiert	unkorrigiert	korrigiert
b_1/b_0		0,17		0,04
r		0,18		0,04
MQ Pflanzen	52,34	50,87	113,38	112,67
MQ Blüten	27,70	27,91	14,53	14,63
MQ Rest	3,68	3,89	9,64	9,74
s^2 Pflanzen	6,08	5,87	12,97	12,87
F-Wert (Pflanzen)	14,23	13,08	11,76	11,57

zwischen den Resten vor, bei der Stengellänge und dem Blütenscheibendurchmesser abgelehnt werden muß.

Tabelle 5 gibt die geschätzten Korrelationskoeffizienten bei Annahme der verschiedenen beschriebenen Modelle für alle Gerbera-Versuche an. Modell 6 gilt nur für die Einwegklassifikation ohne Blüteneffekte. Von den Einzelwerten wurden daher vor der Analyse die mittleren Blüteneffekte abgezogen. Daß diese Effekte aus dem Versuch selbst geschätzt wurden, hat möglicherweise einen Einfluß auf die Schätzwerte nach Modell 6. Für die von Wocheneinflüssen bereinigten Daten, bei denen ohnehin keine signifikanten Blüteneffekte mehr vorhanden waren (siehe Horn u.a., 1974), führte die Rechnung mit und ohne Abzug der Blüteneffekte zu sehr ähnlichen Ergebnissen.

Bedingt durch das Schätzverfahren müssen die Werte für Modell 4 und 5 bei geringer Korrelation sehr ähnlich sein. Modell 6 führte besonders bei der

Stengellänge zu etwas höheren Werten. Allen Versuchen gemeinsam ist die positive Korrelation bei der Stengellänge und das Fehlen der Korrelation bei der Zungenblütenzahl. Bei den beiden anderen Eigenschaften wurde z.T. keine, z.T. eine schwach positive Korrelation gefunden. Die unter Modell 5 aufgeführten Werte können zur Überprüfung der Nullhypothese mit den Vertrauensbereichen, die in Tab.4 angegeben sind, verglichen werden. Der Tabelle ist zu entnehmen, daß die Grenze bei Versuch I und III etwa bei 0,10 liegt und bei Versuch II bei 0,15.

3.2. Tulpen

Die Tulpe ist eine Pflanze mit einer sehr langen Generationsdauer. Bis zur ersten Blüte werden in der Regel sechs bis acht Vegetationsperioden benötigt. Der Zuchtgang erstreckt sich über viele Jahre, und der Züchter hat ein großes Interesse daran, über wirt-

Tab.4. 95%-Bereiche von b_1/b_0 für $H_0: \rho = 0$ [aus: Weber (1976)]

J	p = 10		p = 20		p = 50		p = 100	
	von	bis	von	bis	von	bis	von	bis
3	-		-		-0,63	0,32	-0,40	0,25
4	-0,87	0,47	-0,52	0,34	-0,29	0,22	-0,20	0,17
6	-0,42	0,33	-0,28	0,23	-0,16	0,15	-0,12	0,11
8	-0,31	0,27	-0,21	0,19	-0,13	0,12	-0,09	0,09
10	-0,26	0,23	-0,17	0,16	-0,11	0,10	-0,07	0,07

Tab.5. Korrelationskoeffizienten zwischen aufeinanderfolgenden Resten bei Gerbera für verschiedene Modelle

Modell	Stengellänge			Durchmesser des Blütenstandes		
	4	5	6	4	5	6
Versuch I	0,35	0,32	0,44	0,05	0,05	0,08
Versuch II	0,22	0,20	0,38	0,19	0,18	0,20
Versuch II a	0,14	0,13	0,26(0,28) ⁺	0,08	0,07	0,10(0,12) ⁺
Versuch III	0,25	0,23	0,30	0,12	0,12	0,14
Versuch III a	0,24	0,22	0,28(0,29) ⁺	0,12	0,11	0,13(0,13) ⁺
	Blütenscheibendurchmesser			Zungenblütenzahl		
Versuch I	0,18	0,17	0,24	0,04	0,04	0,07
Versuch II	0,20	0,19	0,19	0,00	0,00	0,06
Versuch II a	0,16	0,15	0,15(0,14) ⁺	0,01	0,01	0,06(0,04) ⁺
Versuch III	-0,01	-0,01	0,02	0,01	0,01	0,01
Versuch III a	-0,03	-0,03	-0,01(0,00) ⁺	-0,04	-0,04	-0,04(-0,05) ⁺

⁺ Werte in Klammern: Blüteneffekte nicht abgezogen

Tab.6. Varianz der Werte für Zwiebelanzahl und Zwiebelgewicht bei Tulpen nach logarithmischer Transformation

a) Zwiebelanzahl

	Jahr					
	1	2	3	4	5	6
Versuch 1	0,116	0,141	0,160	-	-	-
Versuch 2	0,094	0,181	0,216	0,224	0,363	0,341
Versuch 3	0,232	0,123	0,111	0,115	-	-

b) Zwiebelgewicht

	Jahr					
	1	2	3	4	5	6
Versuch 1	0,360	0,342	0,165	-	-	-
Versuch 2	0,240	0,110	0,125	0,073	0,168	0,238
Versuch 3	0,381	0,103	0,098	0,095	-	-

schaftlich wichtige Eigenschaften, wie z.B. die Vermehrungsrate und das Zwiebelgewicht, möglichst früh Aussagen zu erhalten. In drei über mehrere Jahre laufenden Versuchen⁴ sind für jede Parzelle Zwiebelzahl und Zwiebelgewicht ermittelt und der Zwiebelertrag als Ganzes weitervermehrt worden. Bei Versuch 1 wurden über 3 Jahre 24 Nachkommenschaften in 3 Wiederholungen angebaut, das ergab in jedem Jahr 72 Parzellen. Bei Versuch 2 wurden entsprechend $36 \times 3 = 108$ Parzellen über 6 Jahre geprüft und bei Versuch 3 in gleicher Weise $26 \times 4 = 104$ Parzellen über 4 Jahre.

Die Ertragswerte steigen von Jahr zu Jahr stark an, da zwischen den Werten verschiedener Jahre ein multiplikativer Zusammenhang vorliegt: Gewicht im Jahre $z = \text{Gewicht im Jahre } (z-1) \times \text{Vermehrungsrate}$. Durch eine logarithmische Transformation ist es möglich, diese Verknüpfung in eine additive Beziehung zu überführen.

Tabelle 6 zeigt, daß auch nach der logarithmischen Transformation immer noch beachtliche Unterschiede zwischen den Jahren vorhanden sind. Jedoch ist zu vermuten, daß diese Unterschiede auf spezifische Effekte einzelner Jahre und weniger auf eine unzureichende Transformation zurückzuführen sind.

Soll eine Aussage über die Unterschiede zwischen den Klonen bezüglich der Zwiebelzahl und des Zwie-

belgewichtes gemacht werden, so ist zu berücksichtigen, daß die Werte einer Parzelle in einem Jahr von den Ergebnissen der Vorjahre abhängen. Die Jahre entsprechen somit den Messungen bei Modell 1, die Parzellen der Prüfgliedzahl.

Obwohl die Bedingung der Homogenität der Varianzen für die einzelnen Jahre auch nach der logarithmischen Transformation nicht voll erfüllt war, wurde eine Varianzanalyse mit Parzellen und Jahren durchgeführt. Sie ergibt bei allen Versuchen und Eigenschaften eine überraschende Einheitlichkeit: Die Restvarianz ist für das erste und das letzte Jahr wesentlich größer als für die mittleren Jahre. Ein solcher Sachverhalt ist bei starker positiver Korrelation zwischen Resten zu erwarten. Die Anwendung des Modelles 4 auf Versuch 1 mit 3 Jahren ergibt:

$$E(\hat{\epsilon}_{i1}^2) = E(\hat{\epsilon}_{i3}^2) = c(6 - 2\rho - 4\rho^2) \text{ und} \\ E(\hat{\epsilon}_{i2}^2) = c(6 - 8\rho + 2\rho^2) \quad (5)$$

$$\text{mit } c = \sigma^2/9.$$

Für $\rho = 0,5$ z.B. erhält man daraus $E(\hat{\epsilon}_{i1}^2) = E(\hat{\epsilon}_{i3}^2) = c \cdot 4$ und $E(\hat{\epsilon}_{i2}^2) = c \cdot 2,5$.

Die Ableitung der Gleichung (5) findet sich bei Weber (1976). In den Tulpenversuchen sind die Unterschiede zwischen den Restvarianzen in den einzelnen Jahren so groß, daß sich bei der Schätzung von Korrelationskoeffizienten zum Teil unsinnige Werte ergeben (Tab.7). Dabei ist zu beachten, daß bei Modell 5

⁴ Die Versuchsdaten wurden von Prof. Horn für diese Untersuchungen freundlicherweise zur Verfügung gestellt.

Tab.7. Korrelationskoeffizienten zwischen aufeinanderfolgenden Resten bei Tulpen

Versuch 1

	Modell				
	2		4	5	6
Zwiebelanzahl	r ₁₂ = 1,31; r ₂₃ = 1,18		1,23	0,55	0,69
Zwiebelgewicht	r ₁₂ = 0,86; r ₂₃ = 0,89		0,88	0,47	0,62

Versuche 2 und 3

a) Zwiebelanzahl

	Modell					
	3 ₃ ⁺	3 ₄	3 ₅	4	5	6
Versuch 2, Jahr 1-6	0,69	0,74	0,77	0,98	0,62	0,91
Versuch 2, Jahr 1-5	0,63	0,72	-	0,83	0,54	0,96
Versuch 3	0,62	-	-	0,73	0,46	0,33

b) Zwiebelgewicht

	Modell					
	3 ₃	3 ₄	3 ₅	4	5	6
Versuch 2, Jahr 1-6	0,72	0,78	0,80	1,00	0,63	0,74
Versuch 2, Jahr 1-5	0,64	0,75	-	0,77	0,51	0,70
Versuch 3	0,70	-	-	1,14	0,59	0,38

⁺ Der Index bei Modell 3 bezieht sich auf die Zahl der in die Regression einbezogenen Parameter.

$|\rho| < \frac{1}{2\cos[\pi/(J+1)]}$ (Box, 1954) sein muß. Das ergibt folgende Werte:

J	$ \rho $ muß kleiner sein als
3	0,7071
4	0,6180
5	0,5774
6	0,5550
8	0,5412
⋮	⋮
∞	0,5000

Alle Schätzungen weisen somit darauf hin, daß die Ergebnisse der Vorjahre einen sehr starken Einfluß ausüben und die Restvarianz bei der Analyse über mehrere Jahre erheblich unterschätzt wird. Für den Züchter ist das ein sehr unerfreuliches Ergebnis, besagt es doch, daß weitere Jahre nur wenig neue Information liefern und statt dessen die Zahl der Wiederholungen in einem Jahr erhöht werden muß. Eine solche

Erhöhung kann jedoch an der begrenzten Zahl der vorhandenen Zwiebeln scheitern.

Der Versuch 2 wurde auch ohne das letzte Jahr gerechnet, da die Nachkommenschaften in diesem Jahr blühten, was sich möglicherweise unterschiedlich auf den Zwiebelertrag auswirkt. Am Ergebnis änderte sich jedoch wenig.

3.3. Kartoffeln

Die Kartoffelzüchtung würde sich wesentlich verbilligen, wenn man mit der Selektion schon beim Sämling beginnen könnte. Bei den von Krug u.a. (1974) beschriebenen Versuchen über mehrere Jahre (Sämlingsgeneration und 2 oder 3 Klongenerationen) zeigte sich allerdings, daß der Erfolg einer Selektion bei Sämlingen infolge von Nachwirkungen überschätzt wurde. Bevor auch bei diesem Versuch die Korrelation zwischen Resten aufeinanderfolgender Jahre geschätzt werden konnte, mußte bei den betrachteten Eigenschaften "End-

Tab.8. Korrelationskoeffizienten zwischen Resten aufeinanderfolgender Vegetationsperioden bei Kartoffeln

Eigenschaft	Anzahl ¹	Art ²	Zahl der Veg.per.	Korrelationskoeffizienten bei	
				Modell 4	Modell 5
Endertrag	23	K	3	0,246 ± 0,119	0,189 ± 0,083
Endertrag	12	K	4	0,222 ± 0,136	0,182 ± 0,097
Endertrag	11	Z	3	0,223 ± 0,113	0,176 ± 0,076
Frühertrag	23	K	3	0,325 ± 0,143	0,236 ± 0,091

¹ Anzahl der Nachkommenschaften

² K = Sämlingsanzucht unter kontrollierten Bedingungen

Z = Sämlingsanzucht unter Bedingungen des praktischen Zuchtbetriebes

ertrag" und "Frühertrag, 60 Tage nach Pflanzung" die unterschiedliche Größenordnung der Meßwerte (Sämlinge ergaben einige Gramm, Klone z.T. dagegen mehrere kg) für jedes Jahr getrennt durch Transformation in standardisierte Werte ausgeschaltet werden. Beim Sämling kann naturgemäß nur eine Ernte erfolgen, das dabei erzielte Knollengewicht wurde für beide Eigenschaften herangezogen. Tabelle 8 zeigt die aus der Analyse geschätzten Korrelationskoeffizienten nach Modell 4 und 5. Es sind Mittelwerte aus mehreren Nachkommenschaften angegeben. Jede Nachkommenschaft ging aus einer Kreuzung hervor, als Kreuzungspartner wurden Sorten und Wildformen benutzt. Die Mittelwerte zeigen innerhalb der Modelle eine gute Übereinstimmung, die Standardabweichungen sind jedoch beträchtlich, was entweder auf Unterschiede zwischen den Nachkommenschaften oder auf die Unsicherheit der Schätzmethode zurückzuführen ist. In den Versuchen mit Sämlingsanzucht unter kontrollierten Bedingungen wurden 4 Kreuzungen jeweils unter 6 Kulturbedingungen herangezogen. Unterschiede zwischen den Nachkommenschaften dürften nicht die Ursache für die großen Standardabweichungen sein, da die Schwankungen innerhalb einer Kreuzung oder Anzuchtbedingung auch nicht geringer waren.

4. Diskussion

Wie die vorstehenden Beispiele aus pflanzenzüchterischen Versuchen an den drei Arten Gerbera, Tulpen und Kartoffeln zeigen, trifft die für die Varianzanalyse normalerweise zugrunde gelegte Annahme unkorrelierter Reste häufig nicht zu. Für die Bestimmung von Varianzkomponenten, einer in der Quantitativen Genetik häufig benutzten Methode, ist daher eine Schätzung der Korrelationsmatrix R der Reste unumgänglich.

Wie Tiao and Tan (1966) ausführen, kann das Auftreten negativer Varianzkomponenten in manchen Fällen auf eine negative Reihenkorrelation zurückgeführt werden. Bei den in dieser Arbeit beschriebenen Versuchen scheint der umgekehrte Fall vorzuliegen: Bei der üblichen Schätzmethode wird die Restvarianz unterschätzt und die Bedeutung von Prüfgliedeffekten überschätzt, da die Reste aufeinanderfolgender Meßwerte positiv korreliert sind.

In der allgemeinen Form, wie sie in Modell 1 angegeben ist, sind die Korrelationskoeffizienten der Matrix R nicht schätzbar. Bei den Versuchen an Gerbera, Tulpen und Kartoffeln wurden die Korrelationskoeffizienten daher für eingeschränkte Modelle (Modell 2 bis 5), die bei Weber (1976) aufgeführt sind, und für das von Tiao und Tan (1966) sowie anderen Autoren beschriebene Modell, als Modell 6 bezeichnet, bestimmt.

Bei den Versuchen an Gerbera mit der sich in Grenzen haltenden positiven Korrelation wird die korrigierte Restvarianz der wahren Größe wahrscheinlich nahekommen. Dafür spricht auch die gute Übereinstimmung der Schätzungen mit dem Regressionsansatz nach Modell 4 bzw. 5 und der Maximum Likelihood-Methode nach Modell 6. Besonders beim Merkmal Stengellänge, bei dem in allen Versuchen eine deutlich positive Korrelation festgestellt wurde, lieferte die Maximum Likelihood-Methode höhere Schätzwerte.

Die Kartoffelversuche ergaben für Ertrag und Frühertrag eine deutlich positive Korrelation im Mittel zahlreicher Nachkommenschaften. Die Ähnlichkeit der Mittelwerte aus verschiedenen Gruppen von Nachkommenschaften war mit einer großen Schwankungsbreite innerhalb der Gruppen verbunden. Nach Möglichkeit sollte daher statt des aus einem Ver-

sich geschätzten Korrelationskoeffizienten der Mittelwert aus mehreren ähnlichen Versuchen zur Berechnung einer korrigierten Restvarianz herangezogen werden.

In den Versuchen mit Tulpen waren die Schätzwerte für die Korrelation in allen Versuchen bei den beiden Eigenschaften Zwiebelzahl und Zwiebelgewicht sehr hoch. Es wurde untersucht, ob diese hohen Werte durch die Ungleichheit der Varianzen in den einzelnen Jahren bedingt sein könnten. Zu diesem Zweck wurden bei Versuch 2 die Werte für jedes Jahr einzeln standardisiert und dann der Rechenvorgang wiederholt. Das Ergebnis änderte sich jedoch nur unwesentlich. Somit können die Varianzunterschiede nicht für die hohen Schätzwerte verantwortlich gemacht werden. Für den Tulpenzüchter bedeutet die hohe Korrelation, daß bei dem hier beschriebenen Prüfverfahren weitere Jahre bezüglich der Vermehrungsrate nur wenig zusätzliche Information liefern.

Allen Versuchen war gemeinsam, daß niemals negative, sondern immer nur positive Korrelationen gefunden wurden. Die Auswirkung einer positiven Korrelation ist jedoch schwerwiegender. Wird sie nicht berücksichtigt, dann wird die Restvarianz unterschätzt und beim F-Test der Fehler 1. Art erhöht (Box, 1954). Der Versuch täuscht somit eine Präzision vor, die nicht vorhanden ist. Weber (1976) weist darauf hin, daß die prozentuale Verkleinerung der Restvarianz bei positiver Korrelation erheblich stärker ist als die Erhöhung durch eine gleich große negative Korrelation. Es empfiehlt sich also, bei Versuchen, in denen eine positive Korrelation zwischen den Restabweichungen vermutet werden kann, die hier beschriebenen

Methoden zur Abschätzung der Korrelationskoeffizienten anzuwenden, um eine Überschätzung der Prüfgliedeinflüsse zu verhindern und damit bessere Voraussagen für die Züchtung zu gewinnen.

Literatur

- Anscombe, F.J.; Tukey, J.W.: The examination and analysis of residuals. *Technometrics* 5, 141-160 (1963)
- Box, G.E.P.: Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Ann. Math. Stat.* 25, 484-498 (1954)
- Durbin, J.: Estimation of parameters in time-series regression models. *J. Royal Stat. Soc. Ser. B*, 22, 139-153 (1960)
- Horn, W.; Wricke, G.; Weber, W.E.: Erbliche und umweltabhängige Merkmalsausprägung bei *Gerbera jamesonii*. *Gartenbauwissenschaft* 39(21), 289-300 (1974)
- Krug, H.; Wriedt, G.; Weber, W.E.: Untersuchungen zur Frühselektion in der Kartoffelzüchtung. I. Untersuchungen an der Sämlingsgeneration. *Z. Pflanzenzüchtung* 73, 141-162 (1974)
- Larson, W.A.: The analysis of variance for the two-way classification fixed-effects model with observations within a row serially correlated. *Biometrika* 56, 509-514 (1969)
- Tiao, G.C.; Tan, W.Y.: Bayesian analysis of random-effect models in the analysis of variance. *Biometrika* 53, 477-495 (1966)
- Weber, W.E.: Eine Methode zur Bestimmung von Korrelationskoeffizienten zwischen Resten in der Varianzanalyse (Zweiwegklassifikation). *Biometrische Zeitschrift*, im Druck (1976)
- Weber, W.E.; Wricke, G.: Implications of serial correlations in plant breeding. *Proc. 1st Meeting of the Section "Biometrics in Plant Breeding"*, EUCARPIA, Hannover, 75-82 (1973)
- Zellner, A.; Tiao, G.C.: Bayesian analysis of the regression model with autocorrelated errors. *Americ. Stat. Assoc. J.* 59, 763-778 (1964)

Eingegangen am 23. Dezember 1975
Angenommen durch W. Seyffert

Dr. W. Eberhard Weber
Institut für Angewandte Genetik der
Technischen Universität Hannover
Herrenhäuser Str. 2
D-3000 Hannover 21 (Germany/BRD)