

# Korrigierte Regressionsanalyse in Feldversuchen mit SAS

KRISTINA WARNSTORFF, HALLE  
HANNELORE DÖRFEL, HALLE  
JOACHIM SPILKE, HALLE

## Abstract

*In multiple regression analysis a function will be estimated by least square method. When the data are surveyed from field experiments then the estimated variance explained  $R^2$  and the standard error of estimate must be corrected. The estimation of a regression function with the SAS procedures GLM, REG and MIXED is discussed.*

## 1. Einleitung

In einem landwirtschaftlichen Feldversuch wird die Wirkung eines quantitativen Prüffaktors (z. B. N-Düngung) auf ein Prüfmerkmal (Ertrag) bei Winterweizen in einer Blockanlage geprüft. Die Abhängigkeit des Ertrages von der N-Düngung wird mit Hilfe der Regressionsanalyse geschätzt. In Versuchen wirkt auf den Ertrag außer der Einflußgröße N-Düngung die durch Blockbildung erfaßte Störgröße Boden. Bestimmtheitsmaß und Reststandardabweichung als Maße für die Güte der geschätzten Funktion werden in Versuchen unter- bzw. überschätzt (Bätz u.a., 1987). Die Bereinigung dieser Größen von systematischen Störeinflüssen (z.B. Boden) wird diskutiert. Wird für den Blockeinfluß im Modell der Regression fix (Modell I) oder zufällig (Modell II) angenommen, ergeben sich Unterschiede für den Test der Regressionskoeffizienten und Konfidenzintervallschätzungen

## 2. Regressionsanalyse

Mit dem Verfahren der Regressionsanalyse wird die Wirkung ein oder mehrerer quantitativer Prüffaktoren auf ein Prüfmerkmal durch Schätzung einer geeigneten Funktion beschrieben. Maße für die Güte der Anpassung der gewählten Funktion an die Punktwolke sind dabei das Bestimmtheitsmaß B und die Reststandardabweichung  $s_R$ . Das Bestimmtheitsmaß gibt den Anteil der durch die geschätzte Funktion erklärten Variabilität an der Gesamtvariabilität an. Die Reststandardabweichung ist ein Maß für die durchschnittliche Abweichung der Einzelwerte vom Funktionswert  $\bar{y}$ .

Der Test der Regressionskoeffizienten erfolgt durch den t-Test.

Zur Beurteilung der Regressionskoeffizienten und der Regressionsfunktion werden Konfidenzintervalle geschätzt. Vorhersageintervalle geben an, in welchem Bereich mit einer Wahrscheinlichkeit P bei zukünftigen Untersuchungen Einzelwerte zu erwarten sind.

## 3. Regressionsanalyse in Feldversuchen

Wurde das Datenmaterial aus Versuchen in einer Blockanlage gewonnen, so ist in der Regressionsanalyse folgendes zu beachten:

Als Beispiel sei die Wirkung einer gestaffelten N-Düngung auf den Ertrag von Winterweizen in einem Blockversuch geprüft worden. Wird der Ertrag in Abhängigkeit von den N-Gaben grafisch dargestellt, so gehören zu jedem Wert der Einflußgröße x (N-Düngung) r Werte der Zielgröße y (Ertrag), wobei r die Anzahl der Blocks angibt.

Eine solche Abhängigkeit kann durch eine polynomiale Funktion  $\bar{y} = a + b_1x + b_{11}x^2$  mit p=3 Parametern geschätzt werden. Im vorliegenden Beispiel sei der Prüffaktor N-Düngung fix (Modell I) mit a Stufen gegeben, die Schätzung der Funktion soll auf Basis der Einzelwerte des Versuches erfolgen.

## • Bestimmtheitsmaß B

Die durch die geschätzte Funktion erklärbare Variabilität SQ(Regression) kann in Versuchen höchstens den Wert der durch den Prüffaktor verursachten Variabilität SQ(N-Düngung) erreichen. In Versuchen wird also das Bestimmtheitsmaß unterschätzt.

Es gilt in der Varianzanalyse für eine einfaktorielles Blockanlage die Zerlegung:

$$SQ(\text{Gesamt}) = SQ(N-Dg) + SQ(\text{Block}) + SQ(R/VA)$$

$$FG: ar-1 = a-1 + r-1 + (a-1)(r-1)$$

mit

$$SQ(N-Dg) = SQ(\text{Regression}) + SQ(\text{Anpassungsmangel})$$

$$FG: a-1 = p-1 + a-p$$

Die Korrektur des Bestimmtheitsmaßes lautet deshalb:  $B_{\text{Prüfglied}} = \frac{SQ(\text{Regression})}{SQ(N - Dg)}$ .

## • Reststandardabweichung $s_R$

In Versuchen wird die Reststandardabweichung überschätzt, d.h. der Rest der Regression enthält den systematischen Störeinfluß „Block“.

Wegen

$$SQ(\text{Gesamt}) = SQ(\text{Regression}) + SQ(\text{Anpass.mangel}) + SQ(\text{Block}) + SQ(R/VA)$$

$$FG: ar-1 = p-1 + a-p + r-1 + (a-1)(r-1)$$

gilt:

$$SQ(R/RA) = SQ(\text{Anpass.mangel}) + SQ(\text{Block}) + SQ(R/VA)$$

$$FG: ar-p = a-p + r-1 + (a-1)(r-1)$$

Durch Elimination der systematischen Störgröße SQ(Block) aus dem Rest der Regression ergibt sich:

$$SQ(R^*/RA) = SQ(R/RA) - SQ(\text{Block})$$

$$FG: r(a-1)-(p-1) = ar-p - (r-1)$$

Die Reststandardabweichung  $s_R$  wird zu  $s_R^*$  korrigiert.

## • Test des Bestimmtheitsmaßes und der Regressionskoeffizienten

Mit der bereinigten Reststandardabweichung  $s_R^*$  kann der Test des Bestimmtheitsmaßes mit

$$F = \frac{MQ(\text{Regression})}{MQ(R^*/RA)} \text{ und } F(1-\alpha, p-1, FG(R^*/RA))$$

und der Test der Regressionskoeffizienten mit

$$t = \frac{b}{s_b} \text{ und } t(1-\frac{\alpha}{2}, FG(R^*/RA)) \text{ „verbessert“ werden.}$$

Bei der Schätzung von  $s_b^*$  muß beachtet werden, ob im Modell der Regression Block fix (SAS Proc GLM und Proc REG) oder zufällig (SAS Proc Mixed) angenommen wird. Die Blockvarianzkomponente  $\sigma_{\text{Block}}^2$  beeinflusst bei Block zufällig die Schätzung von  $s_b^2(a)$ , nicht aber die von  $s_b^2(b_1)$ . Die Standardabweichung für das durchschnittliche Anfangs-

niveau a vergrößert sich mit steigendem Verhältnis von  $\frac{\sigma_{\text{Block}}^2}{\sigma_R^2}$ .

## • Konfidenzintervalle

Die Beurteilung der Regressionskoeffizienten durch ein Konfidenzintervall sollte in ihrer Umwelt (Block) erfolgen, die Konfidenzgrenzen werden daher bei Block fix im Modell der Regressionsanalyse nicht korrigiert, ebenso wird das Konfidenz- und Vorhersageintervall für die geschätzte Funktion nicht bereinigt.

Bei Block zufällig sind die Konfidenzgrenzen für das durchschnittliche Anfangsniveau  $\alpha$  von  $\sigma_{Block}^2$  beeinflusst.

Die mit Proc MIXED geschätzten Konfidenzgrenzen für die geschätzte Funktion sind von der Varianzkomponente  $\sigma_{Block}^2$  beeinflusst.

### 5. Beispiel

Der Ertrag (dt ha<sup>-1</sup>) von Winterweizen in Abhängigkeit von einer gestaffelten Stickstoffdüngung (0 30 60 90 120 kg N ha<sup>-1</sup>) wurde in einer Blockanlage mit 4 Blocks geprüft. Die vom Störeinfluß Block bereinigte Regressionsanalyse wird mit den Prozeduren GLM, REG und MIXED durchgeführt.

Stark verkürzter SAS-output:

#### Procedure GLM - Varianzanalyse:

Dependent Variable: ERTRAG					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1369.6170	342.40425	7.28	0.0018
Error	15	705.8125	47.05416		
Corrected Total	19	2075.4295			

R-Square	C.V.	Root MSE	ERTRAG Mean
0.65992	14.74	6.8596	46.545

$$SQ(N-Düng.)=1369.617 \quad SQ(R/VA)=705.8125$$

#### Procedure GLM - Regressionsanalyse mit Störgröße Block (fix)

Dependent Variable: ERTRAG

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	5	1702.7265	340.5453	12.79	0.0001
Error	14	372.7030	26.6216		
Corrected Total	19	2075.4295			

$$SQ(R^*/RA)=372.703 \quad FG(R^*/RA)=14$$

	R-Square	C.V.	Root MSE	ERTRAG Mean
	0.8204	11.0852	5.1596	46.545

  

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Block	3	437.9455	145.9818	5.48	0.0105
X1	1	1147.0410	1147.0410	43.09	0.0001
X1*X1	1	117.7400	117.7400	4.42	0.0540

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Block	3	437.9455	145.9818	5.48	0.0105
X1	1	400.4773	400.4773	15.04	0.0017
X1*X1	1	117.7400	117.7400	4.42	0.054

#### Procedure REG - multiple Regressionsanalyse

Model: MODEL1

Dependent Variable: ERTRAG

Analysis of Variance

Source	DF	Squares	Sum of Mean Square	F Value	Prob>F
Model	2	1264.7810	632.3905	13.262	0.0003
Error	17	810.6485	47.6852		
C Total	19	2075.4295			

Root MSE	6.9055	R-square	0.6094
Dep Mean	46.5450	Adj R-sq	0.5635
C.V.	14.8361		

$$SQ(\text{Regression})=1264.781 \quad FGR=17 \quad MQ(R/RA)=47.685 \quad s_R=6.91$$

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	32.93500	3.249441	10.136	0.0001
X1	1	0.37183	0.128307	2.898	0.0100
X11	1	-0.00161	0.001025	-1.571	0.1345

#### korrigierte Regressionsanalyse:

$$B_{\text{Prüfglied}} = \frac{SQ(\text{Regression})}{SQ(N - D\ddot{u}ng.)} = \frac{1264.781}{1369.617} = 0.9234$$

von Störeinfluß bereinigte Restvarianz				
OBS	FGR	MQRKORR	SRKORR	
1	14	26.6216	5.15962	

bereinigter Test der Regressionskoeffizienten				
OBS	B	SB	T	PROB
1	32.9350	2.42792	13.5651	0.00000
2	0.3718	0.09587	3.8786	0.00084
3	-0.0016	0.00077	-2.1030	0.97299

#### Procedure MIXED - Regressionsanalyse mit Störgröße Block (zufällig)

Covariance Parameter Estimates (REML)

Cov Parm	Subject	Estimate
Block	Block	23.87203
Residual		26.62164

Solution for Fixed Effects

Effect	Estimate	Std Error	DF	t	Pr >  t	Alpha	Lower	Upper
INTERCEPT	32.9350	3.44424	3	9.56	0.0024	0.05	21.9739	43.8961
X1	0.37183	0.09587	14	3.88	0.0017	0.05	0.1662	0.5775
X1*X1	-0.00161	0.00077	14	-2.10	0.0540	0.05	-0.0033	-0.0000

### 6. Literatur

G. BÄTZ u.a. Einführung in die Methodik des Feldversuches, Deutscher Landwirtschaftsverlag, Berlin (1987)

SAS/STAT Changes and Enhancements, Release 6.11 (1996)