BU-664-M

January 1979

STATISTICAL METHODS BY SAS

Annotated Output from SAS for Selected Examples from
Statistical Methods, Snedecor and Cochran

by

Valerie Arneson

Patricia A. Firey

Charles McCulloch

Table of Contents

Note: The version of SAS used was version 76.6B. The output has been edited to conserve space.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

p. 384, S+C.

```
1          DATA IOWAPH;  INPUT  X1 X2 Y;
2          X0=1;
3          CARDS;
```

Set a column of 1's only for printing the X matrix. SAS GLM supplies its own.

```
NOTE: DATA SET WORK.IOWAPH HAS 18 OBSERVATIONS AND 4 VARIABLES. 361 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.22 SECONDS AND 102K.

22         PROC PRINT;
23    (A)  VARIABLES X0 X1 X2;
```

Print the X matrix.

```
NOTE: THE PROCEDURE PRINT USED 0.18 SECONDS AND 102K.

24         PROC GLM;
25    (B)  MODEL Y=X2 X1/ SOLUTION P;
```

Regress Y on X2 and X1.

```
NOTE: THE PROCEDURE GLM USED 0.44 SECONDS AND 152K.

26         PROC GLM;
27         MODEL Y=X1 X2/SOLUTION P;
28    (C)  OUTPUT  OUT=NEW(DROP=X0 X1 X2 Y)
29            RESIDUAL=RESID
30             PREDICTED=YHAT;
```

Regress Y on X1 and X2.

The OUTPUT statement must be used in order to plot residuals.

```
NOTE: DATA SET WORK.NEW HAS 18 OBSERVATIONS AND 2 VARIABLES. 651 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.47 SECONDS AND 172K.

31         PROC PLOT;
32    (D)  PLOT RESID*YHAT;
```

Plot residuals as Y, predicted as X.

```
NOTE: THE PROCEDURE PLOT USED 0.33 SECONDS AND 118K.


NOTE: BARR, GOODNIGHT, SALL AND HELWIG
      SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605
```

S T A T I S T I C A L A N A L Y S I S S Y S T E M

OBS	XO	X1	X2
1	1	0.4	53
2	1	0.4	23
3	1	3.1	19
4	1	0.6	34
5	1	4.7	24
6	1	1.7	65
7	1	9.4	44
8	1	10.1	31
9	1	11.6	29
10	1	12.6	58
11	1	10.9	37
12	1	23.1	46
13	1	23.1	50
14	1	21.6	44
15	1	23.1	56
16	1	1.9	36
17	1	26.8	58
18	1	29.9	51

(A)

The X matrix for fitting X1 first.

Exchange X1 and X2 to fit X2 first.

The ABDO's would look like this:

X'X / X'Y

XO	X1	X2	Y	
18	215	758	1463	
	4321.0209	10139.5011	20706.2021	
		35076	63825	
18	215	758	1463	
1	11.9444	42.1111	81.2778	$= \bar{Y} = b0$
	1752.9649	1085.6113	3231.4782	
	1	.6193	1.8434	$= b1$
		2483.4588	215.2117	
		1	1.7898	$= b2$

$Y = \bar{Y} + b_1 (X_1 - \hat{X}_1(0)) + b_2 (X_2 - \hat{X}_2(0,1))$

X'X / X'Y

XO	X2	X1	Y	
18	758	215	1463	
	35076	10139.5011	63825	
		4321.0209	20706.2021	
18	758	215	1463	
1	42.1111	11.9444	81.2778	$= \bar{Y} = b0$
	3155.7778	1085.6113	2216.4444	
	1	.3444	.7023	$= b1$
		1479.5064	2469.0046	
		1	.0867	$= b2$

Sequential b's.

$Y = \bar{Y} + b_2 (X_2 - \hat{X}_2(0)) + b_1 (X_1 - \hat{X}_1(0,2))$

Fitting X1 alone, we get

18	215	1463
	4321.0209	20706.2021
18	215	1463
1	11.9444	81.2778 $= \bar{Y} = b_0$
	1752.9649	3231.4782
	1	1.8434 $= b_1$ (sequential b.)

		SS	DF	MS	F	b	SE(b)	t
CFM	X0	118909.3889	1	118909.3889	295.7674	59.2589	7.420	7.986
X1 alone	X1	5957.0225	1	5957.0225	14.8171	1.8434 $= b_1$	.479	3.849
	RESID	6432.5886	16	402.0368		Intercept		
Uncorrected	TOTAL	131299.	18					

Fitting X2 alone, we get

18	758	1463
	35076	63825
18	758	1463
1	42.1111	81.2778 $= \bar{Y} = b_0$
	3155.7778	2216.4444
	1	.7023 $= b_1$ (sequential b.)

		SS	DF	MS	F	b	SE(b)	t
CFM	X0	118909.3889	1	118909.3889	175.6270	51.7013	20.4469	2.529
X2 alone	X2	1556.7085	1	1556.7085	2.2992	.7023 $= b_2$	.4632	1.516
	RESID	10832.9026	16	677.0564		Intercept		
	TOTAL	131299.	18					

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

Regress Y on X2 and X1 with intercept.

(B)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL (Regression)	2	5975.66853237	2987.83426619	6.99	0.0072	0.482313	25.4416
ERROR (Deviations)	15	6413.94257874	427.59617192		STD DEV		Y MEAN
CORRECTED TOTAL	17	12389.61111111			20.67839868	$\bar{Y}$ =	81.27777778

SOURCE		DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X2	alone (S+C Table 13.3.2)	1	1556.70846341	3.64	0.0757	1	X2 after X1 18.64603750	0.04	0.8374
X1	after X2	1	4418.96006896	10.33	0.0058	1	X1 after X2 4418.96006896	10.33	0.0058

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	56.25102409 = a	3.45	0.0036	16.31073734
X2	0.08664925 = b_2	0.21	0.8374	0.41494299
X1	1.78977412 = b_1	3.21	0.0058	0.55674341

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	64.00000000	61.55934398	2.44065602
2	60.00000000	58.95986648	1.04013352
3	71.00000000	63.44565960	7.55434040
4	61.00000000	60.27096305	0.72903695
5	54.00000000	66.74254443	-12.74254443
6	77.00000000	64.92584133	12.07415867
7	81.00000000	76.88746778	4.11253222
8	93.00000000	77.01386941	15.98613059
9	93.00000000	79.52523208	13.47476792
10	51.00000000	83.82783445	-32.82783445
11	76.00000000	78.96558420	-2.96558420
12	96.00000000	101.58067167	-5.58067167
13	77.00000000	101.92726867	-24.92726867
14	93.00000000	98.72271199	-5.72271199
15	95.00000000	102.44716417	-7.44716417
16	54.00000000	62.77096791	-8.77096791
17	168.00000000	109.24262690	58.75737310
18	99.00000000	114.18438191	-15.18438191

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	6413.94257874
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.18471654
DURBIN-WATSON D	2.32568745

© STATISTICAL ANALYSIS SYSTEM

GENERAL LINEAR MODELS PROCEDURE

Regress Y on X1 and X2, with intercept.

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	5975.66853237	2987.83426619	6.99	0.0072	0.482313	25.4416
ERROR	15	6413.94257874	427.59617192		STD DEV		Y MEAN
CORRECTED TOTAL	17	12389.61111111			20.67839868	$\bar{Y}$ =	81.27777778

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1 alone	1	5957.02249487	13.93	0.0020	1	X1 after X2 4418.96006896	10.33	0.0058
X2 after X1	1	18.64603750	0.04	0.8374	1	X2 after X1 18.64603750	0.04	0.8374

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	56.25102409 = a	3.45	0.0036	16.31073734
X1	1.78977412 = b_1	3.21	0.0058	0.55674341
X2	0.08664925 = b_2	0.21	0.8374	0.41494299

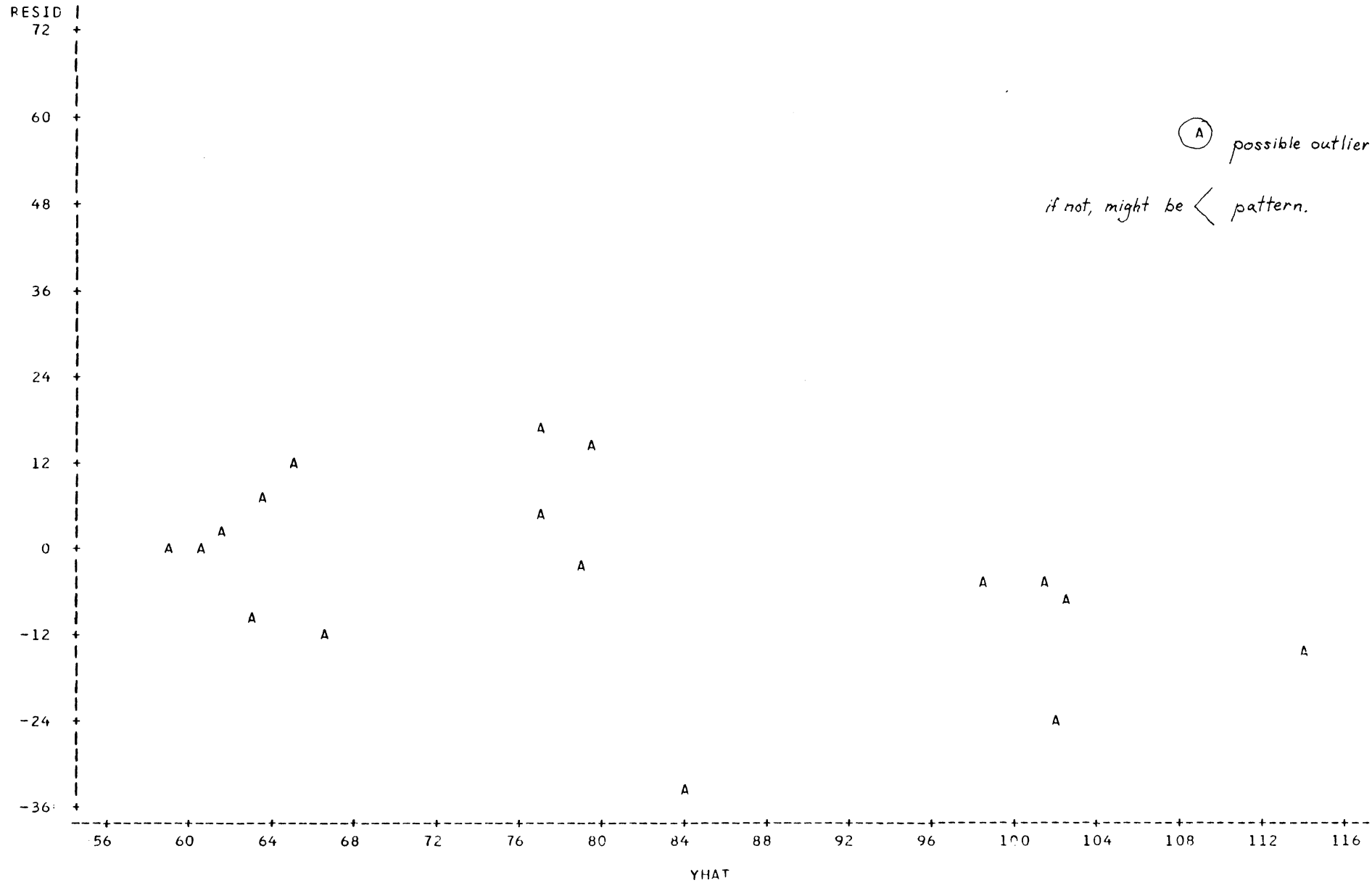
OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	64.00000000	61.55934398	2.44065602
2	60.00000000	58.95986648	1.04013352
3	71.00000000	63.44565960	7.55434040
4	61.00000000	60.27096305	0.72903695
5	54.00000000	66.74254443	-12.74254443
6	77.00000000	64.92584133	12.07415867
7	81.00000000	76.88746778	4.11253222
8	93.00000000	77.01386941	15.98613059
9	93.00000000	79.52523208	13.47476792
10	51.00000000	83.82783445	-32.82783445
11	76.00000000	78.96558420	-2.96558420
12	96.00000000	101.58067167	-5.58067167
13	77.00000000	101.92726867	-24.92726867
14	93.00000000	98.72271199	-5.72271199
15	95.00000000	102.44716417	-7.44716417
16	54.00000000	62.77096791	-8.77096791
17	168.00000000	109.24262690	58.75737310
18	99.00000000	114.18438191	-15.18438191

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	6413.94257874
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.18471654
DURBIN-WATSON D	2.32568745

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(D)

PLOT OF RESID*YHAT LEGEND: A = 1 OBS, B = 2 OBS, ETC.



(A) possible outlier?

if not, might be < pattern.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

p. 259 S•C.

```
1          DATA DONUT; INPUT FAT ABS;
2          X0=1;
3          X1=0;
4          IF FAT=1 THEN X1=1;
5          X2=0;
6          IF FAT=2 THEN X2=1;
7          X3=0;
8          IF FAT=3 THEN X3=1;
9          X4=0;
10         IF FAT=4 THEN X4=1;
11         CARDS;
```

Set up X matrix by defining columns of 0's and 1's.

```
NOTE: DATA SET WORK.DONUT HAS 24 OBSERVATIONS AND 7 VARIABLES. 217 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.24 SECONDS AND 102K.

36 (A)     PROC PRINT;
37         VARIABLES X1 X2 X3 X4;
```

Print X matrix, no intercept.

```
NOTE: THE PROCEDURE PRINT USED 0.20 SECONDS AND 102K.

38 (B)     PROC GLM;
39         MODEL ABS=X1 X2 X3 X4/NOINT SOLUTION P;

NOTE: THE PROCEDURE GLM USED 0.51 SECONDS AND 152K.

40 (C)     PROC PRINT;
41         VARIABLES X0 X1 X2 X3 X4;
```

Print X matrix, with column of 1's for intercept.

```
NOTE: THE PROCEDURE PRINT USED 0.20 SECONDS AND 102K.

42         PROC GLM;
43 (D)     MODEL ABS=X1 X2 X3 X4/ SOLUTION P;
44         OUTPUT  OUT=NEW(DROP=X1 X2 X3 X4 FAT ABS)
45            RESIDUAL=RESID
46             PREDICTED=YHAT;
```

Output residuals and predicted in order to plot them.

```
NOTE: DATA SET WORK.NEW HAS 24 OBSERVATIONS AND 3 VARIABLES. 465 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.50 SECONDS AND 172K.

47 (E)     PROC PLOT;
48         PLOT RESID*YHAT;
```

Plot residuals as Y, predicted as X.

```
NOTE: THE PROCEDURE PLOT USED 0.30 SECONDS AND 118K.


NOTE: BARR, GOODNIGHT, SALL AND HELWIG
      SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605
```

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(A)

OBS	X1	X2	X3	X4
1	1	0	0	0
2	1	0	0	0
3	1	0	0	0
4	1	0	0	0
5	1	0	0	0
6	1	0	0	0
7	0	1	0	0
8	0	1	0	0
9	0	1	0	0
10	0	1	0	0
11	0	1	0	0
12	0	1	0	0
13	0	0	1	0
14	0	0	1	0
15	0	0	1	0
16	0	0	1	0
17	0	0	1	0
18	0	0	1	0
19	0	0	0	1
20	0	0	0	1
21	0	0	0	1
22	0	0	0	1
23	0	0	0	1
24	0	0	0	1

X matrix (no intercept.)

ABDO looks like the following:

X'X | X'Y

6	0	0	0	432	
	6	0	0	510	Trt totals
		6	0	456	
			6	372	
6	0	0	0	432	
1	0	0	0	72	$= \overline{X1}$
	6	0	0	510	
	1	0	0	85	$= \overline{X2}$
		6	0	456	
		1	0	76	$= \overline{X3}$
			6	372	
			1	62	$= \overline{X4}$

With no intercept, X'X is diagonal and sequential b's are the same as the nonsequential b's. In this case they are equal to the trt means.

STATISTICAL ANALYSIS SYSTEM

GENERAL LINEAR MODELS PROCEDURE

(B) No intercept.

DEPENDENT VARIABLE: ABS

Since there is no intercept, the CFM is not removed from model SS.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL = Between + CFM	4	132174.00000000	33043.50000000	327.49	0.0001	0.984962	13.6202
ERROR Within fats	20	2018.00000000	100.90000000		STD DEV		ABS MEAN
UNCORRECTED TOTAL	24	134192.00000000			10.04489920		73.75000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	1	31104.00000000	308.27	0.0001	1	31104.00000000	308.27	0.0001
X2	1	43350.00000000	429.63	0.0001	1	43350.00000000	429.63	0.0001
X3	1	34656.00000000	343.47	0.0001	1	34656.00000000	343.47	0.0001
X4	1	23064.00000000	228.58	0.0001	1	23064.00000000	228.58	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X1	72.00000000 = $\bar{X}_1$	17.56	0.0001	4.10081293
X2	85.00000000 = $\bar{X}_2$	20.73	0.0001	4.10081293
X3	76.00000000 = $\bar{X}_3$	18.53	0.0001	4.10081293
X4	62.00000000 = $\bar{X}_4$	15.12	0.0001	4.10081293

Note: with no intercept model, the estimated b's are the trt means.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	64.00000000	72.00000000	-8.00000000
2	72.00000000	72.00000000	0.00000000
3	68.00000000	72.00000000	-4.00000000
4	77.00000000	72.00000000	5.00000000
5	56.00000000	72.00000000	-16.00000000
6	95.00000000	72.00000000	23.00000000
7	78.00000000	85.00000000	-7.00000000
8	91.00000000	85.00000000	6.00000000
9	97.00000000	85.00000000	12.00000000
10	82.00000000	85.00000000	-3.00000000
11	85.00000000	85.00000000	0.00000000
12	77.00000000	85.00000000	-8.00000000
13	75.00000000	76.00000000	-1.00000000
14	93.00000000	76.00000000	17.00000000
15	78.00000000	76.00000000	2.00000000
16	71.00000000	76.00000000	-5.00000000
17	63.00000000	76.00000000	-13.00000000
18	76.00000000	76.00000000	0.00000000
19	55.00000000	62.00000000	-7.00000000
20	66.00000000	62.00000000	4.00000000
21	49.00000000	62.00000000	-13.00000000
22	64.00000000	62.00000000	2.00000000

Note that the predicted values are just the $\bar{X}_i$'s.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: ABS

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
23	70.00000000	62.00000000	8.00000000
24	68.00000000	62.00000000	6.00000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	2018.00000000
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	-0.30336133
DURBIN-WATSON D	2.54212091

(C)

OBS	X0	X1	X2	X3	X4
1	1	1	0	0	0
2	1	1	0	0	0
3	1	1	0	0	0
4	1	1	0	0	0
5	1	1	0	0	0
6	1	1	0	0	0
7	1	0	1	0	0
8	1	0	1	0	0
9	1	0	1	0	0
10	1	0	1	0	0
11	1	0	1	0	0
12	1	0	1	0	0
13	1	0	0	1	0
14	1	0	0	1	0
15	1	0	0	1	0
16	1	0	0	1	0
17	1	0	0	1	0
18	1	0	0	1	0
19	1	0	0	0	1
20	1	0	0	0	1
21	1	0	0	0	1
22	1	0	0	0	1
23	1	0	0	0	1
24	1	0	0	0	1

X matrix with column of 1's for intercept (mean)

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

(D)

DEPENDENT VARIABLE: ABS

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL = *between fats*	3	1636.50000000	545.50000000	5.41	0.0069	0.447804	13.6202
ERROR = *within fats*	20	2018.00000000	100.90000000		STD DEV		ABS MEAN
CORRECTED TOTAL	23	3654.50000000			10.04489920		73.75000000

SOURCE		DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	*between fats*	1	24.50000000	0.24	0.6275	0	0.00000000	.	.
X2		1	1024.00000000	10.15	0.0046	0	0.00000000	.	.
X3		1	588.00000000	5.83	0.0255	0	0.00000000	.	.
X4		0	0.00000000	.	.	0	0.00000000	.	.

PARAMETER	ESTIMATE		T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	62.00000000	B= $\bar{X4}$	15.12	0.0001	4.10081293
X1	10.00000000	B= $\bar{X1}-\bar{X4}$	1.72	0.1001	5.79942526
X2	23.00000000	B= $\bar{X2}-\bar{X4}$	3.97	0.0008	5.79942526
X3	14.00000000	B= $\bar{X3}-\bar{X4}$	2.41	0.0255	5.79942526
X4	0.00000000	B	.	.	.

Note: the intercept model results in estimated β's which are certain linear combinations of the trt. means.

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	64.00000000	72.00000000	-8.00000000
2	72.00000000	72.00000000	0.00000000
3	68.00000000	72.00000000	-4.00000000
4	77.00000000	72.00000000	5.00000000
5	56.00000000	72.00000000	-16.00000000
6	95.00000000	72.00000000	23.00000000
7	78.00000000	85.00000000	-7.00000000
8	91.00000000	85.00000000	6.00000000
9	97.00000000	85.00000000	12.00000000
10	82.00000000	85.00000000	-3.00000000
11	85.00000000	85.00000000	0.00000000
12	77.00000000	85.00000000	-8.00000000
13	75.00000000	76.00000000	-1.00000000

The predicted values stay the same as in B.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: ABS

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
14	93.00000000	76.00000000	17.00000000
15	78.00000000	76.00000000	2.00000000
16	71.00000000	76.00000000	-5.00000000
17	63.00000000	76.00000000	-13.00000000
18	76.00000000	76.00000000	0.00000000
19	55.00000000	62.00000000	-7.00000000
20	66.00000000	62.00000000	4.00000000
21	49.00000000	62.00000000	-13.00000000
22	64.00000000	62.00000000	2.00000000
23	70.00000000	62.00000000	8.00000000
24	68.00000000	62.00000000	6.00000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	2018.00000000
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.30336133
DURBIN-WATSON D	2.54212091

ABDO for the above looks like this:

$X'X \mid X'Y$

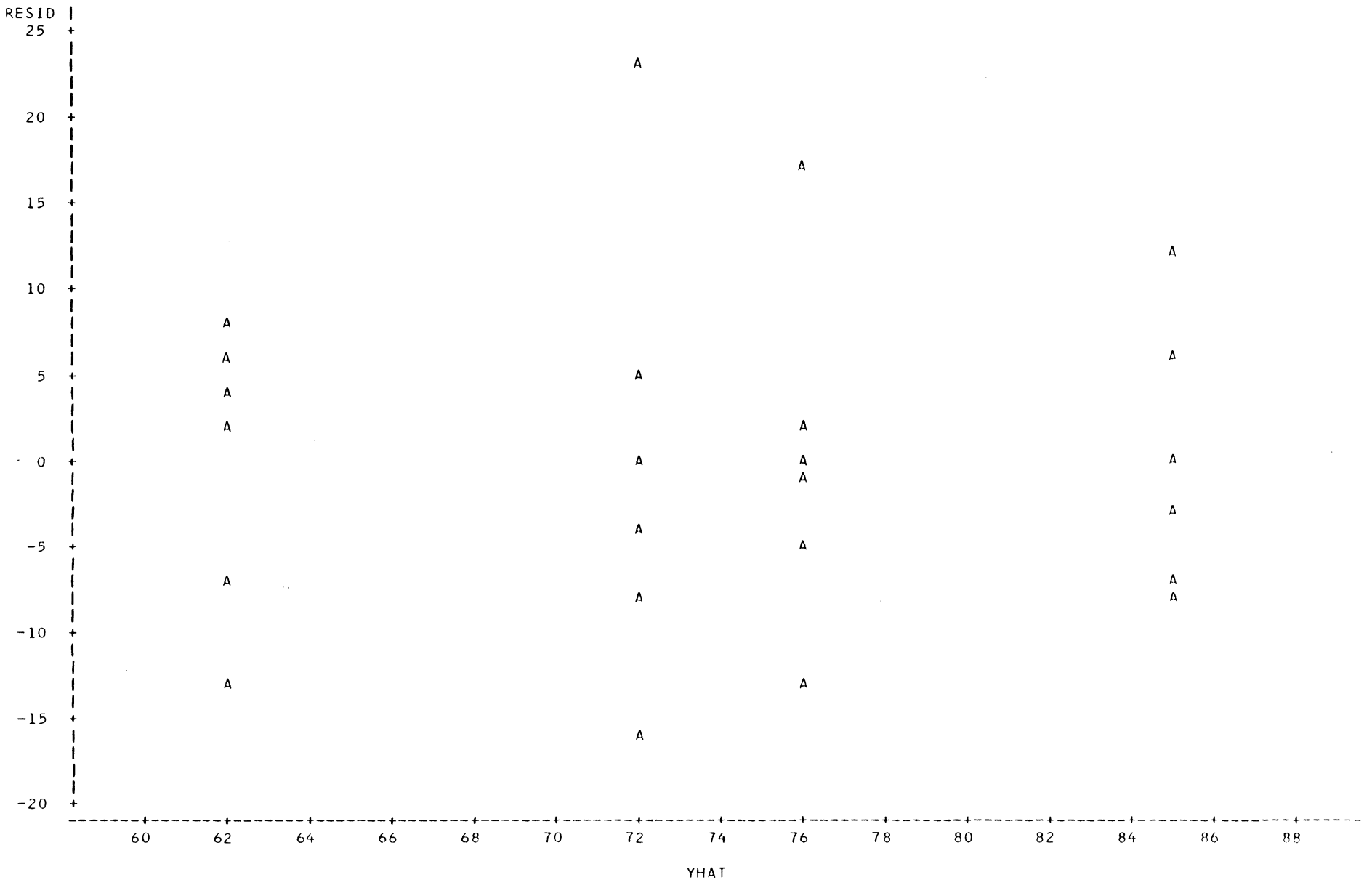
24	6	6	6	6	1770	
	6	0	0	0	432	
		6	0	0	510	
			6	0	456	
				6	372	
24	6	6	6	6	1770	
1	.25	.25	.25	.25	73.75	$= \bar{X}..$
	4.5	-1.5	-1.5	-1.5	-10.5	
	1	-.33	-.33	-.33	-2.33	$= \bar{X}1 - (\bar{X}2 + \bar{X}3 + \bar{X}4)/3$
		4	-2	-2	64	
		1	-.5	-.5	16	$= \bar{X}2 - (\bar{X}3 + \bar{X}4)/2$
			3	-3	42	
			1	-1	14	$= \bar{X}3 - \bar{X}4$
				0	0	
				0	0	

S T A T I S T I C A L A N A L Y S I S S Y S T E M

Note that for this type of design the plot has only 4 discrete values of $\hat{Y}$.

(E)

PLOT OF RESID*YHAT LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Page 287 S-C.

```
1          DATA TURNIP; INPUT PLANT LEAF CALC;
2          X0=1;
3          X1=(PLANT=1);
4          X2=(PLANT=2);
5          X3=(PLANT=3);
6          X4=(PLANT=4);
7          X5=(PLANT=1 AND LEAF=1);
8          X6=(PLANT=1 AND LEAF=2);
9          X7=(PLANT=1 AND LEAF=3);
10         X8=(PLANT=2 AND LEAF=1);
11         X9=(PLANT=2 AND LEAF=2);
12         X10=(PLANT=2 AND LEAF=3);
13         X11=(PLANT=3 AND LEAF=1);
14         X12=(PLANT=3 AND LEAF=2);
15         X13=(PLANT=3 AND LEAF=3);
16         X14=(PLANT=4 AND LEAF=1);
17         X15=(PLANT=4 AND LEAF=2);
18         X16=(PLANT=4 AND LEAF=3);
19         CARDS;

NOTE: DATA SET WORK.TURNIP HAS 24 OBSERVATIONS AND 20 VARIABLES. 79 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.30 SECONDS AND 102K.

44         PROC PRINT;
45    (A)  VARIABLES X0-X16;

NOTE: THE PROCEDURE PRINT USED 0.31 SECONDS AND 116K.

46         PROC GLM;
47    (B)  MODEL CALC= X1-X16;

NOTE: THE PROCEDURE GLM USED 0.74 SECONDS AND 160K.

48         PROC GLM;
49    (C)  CLASSES PLANT LEAF;
50         MODEL CALC= PLANT LEAF(PLANT);

NOTE: THE PROCEDURE GLM USED 0.54 SECONDS AND 158K.

51         PROC GLM;
52         CLASSES PLANT LEAF;
53    (D)  MODEL CALC= PLANT LEAF(PLANT);
54         TEST H=PLANT E=LEAF(PLANT)/HTYPE=1 ETYPE=1;

NOTE: THE PROCEDURE GLM USED 0.54 SECONDS AND 160K.

NOTE: SAS USED 160K MEMORY.

NOTE: BARR, GOODNIGHT, SALL AND HELWIG
      SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605
```

Note the logic for setting up a nested design matrix. Using a parenthesis gives a value of 1 if the statement inside the parenthesis is true and a 0 otherwise. The statement

X1 = (PLANT = 1)

is equivalent to the two statements

X1 = 0
IF PLANT = 1 THEN X1 = 1

(A) Printing the X matrix

(B) Using GLM on the X matrix. X0 is not included since an intercept is automatically put in.

(C) A nested analysis using the CLASSES statement. LEAF(PLANT) indicates leaves are nested within plants.

(D) Specifying the correct F-test

(A)

S T A T I S T I C A L A N A L Y S I S S Y S T E M

A printing of the X matrix, including an intercept.

OBS	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
3	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
6	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
9	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
10	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
11	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
12	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
13	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
14	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
15	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
16	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
17	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
18	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
19	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
20	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
21	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
22	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
23	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
24	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

(B)

DEPENDENT VARIABLE: CALC

This corresponds to the analysis on p. 286.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	10.19054583	0.92641326	139.22	0.0001	0.992225	2.7082
ERROR	12	0.07985000	0.00665417		STD DEV		CALC MEAN
CORRECTED TOTAL	23	10.27039583			0.08157308		3.01208333

10.1905 = 7.5603 + 2.6302
= SS (plants) + SS (leaves in plants)

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	1	0.21233472	31.91	0.0001	0	0.00000000	.	.
X2	1	5.46780278	821.71	0.0001	0	0.00000000	.	.
X3	1	1.88020833	282.56	0.0001	0	0.00000000	.	.
X4	0	0.00000000	.	.	0	0.00000000	.	.
X5	1	0.00030000	0.05	0.8354	0	0.00000000	.	.
X6	1	0.43560000	65.46	0.0001	0	0.00000000	.	.
X7	0	0.00000000	.	.	0	0.00000000	.	.
X8	1	0.22140833	33.27	0.0001	0	0.00000000	.	.
X9	1	0.08702500	13.08	0.0035	0	0.00000000	.	.
X10	0	0.00000000	.	.	0	0.00000000	.	.
X11	1	0.16803333	25.25	0.0003	0	0.00000000	.	.
X12	1	1.08160000	162.54	0.0001	0	0.00000000	.	.
X13	0	0.00000000	.	.	0	0.00000000	.	.
X14	1	0.02000833	3.01	0.1085	0	0.00000000	.	.
X15	1	0.61622500	92.61	0.0001	0	0.00000000	.	.
X16	0	0.00000000	.	.	0	0.00000000	.	.

SS for plants

SS for leaves in plants

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	3.31000000 B	57.38	0.0001	0.05768087
X1	-0.47000000 B	-5.76	0.0001	0.08157308
X2	-1.12000000 B	-13.73	0.0001	0.08157308
X3	-0.76000000 B	-9.32	0.0001	0.08157308
X4	0.00000000 B	.	.	.
X5	0.34500000 B	4.23	0.0012	0.08157308
X6	0.66000000 B	8.09	0.0001	0.08157308
X7	0.00000000 B	.	.	.
X8	0.26000000 B	3.19	0.0078	0.08157308
X9	-0.29500000 B	-3.62	0.0035	0.08157308
X10	0.00000000 B	.	.	.
X11	0.16500000 B	2.02	0.0660	0.08157308
X12	1.04000000 B	12.75	0.0001	0.08157308
X13	0.00000000 B	.	.	.
X14	0.51500000 B	6.31	0.0001	0.08157308
X15	0.78500000 B	9.62	0.0001	0.08157308
X16	0.00000000 B	.	.	.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

(C)

DEPENDENT VARIABLE: CALC

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	10.19054583	0.92641326	139.22	0.0001	0.992225	2.7082
ERROR	12	0.07985000	0.00665417		STD DEV		CALC MEAN
CORRECTED TOTAL	23	10.27039583			0.08157308		3.01208333

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
PLANT	3	7.56034583	378.73	0.0001	3	7.56034583	378.73	0.0001
LEAF(PLANT)	8	2.63020000	49.41	0.0001	8	2.63020000	49.41	0.0001

This is the correct test for leaves in plants:

$$F = \frac{MS(\text{leaves in plants})}{MS(\text{error})}$$

This is the incorrect test for plants since we have a random model. SAS gave us

$$F = \frac{MS(\text{plants})}{MS(\text{error})}$$

when we actually wanted

$$F = \frac{MS(\text{plants})}{MS(\text{leaves in plants})}$$

(D)

DEPENDENT VARIABLE: CALC

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	10.19054583	0.92641326	139.22	0.0001	0.992225	2.7082
ERROR	12	0.07985000	0.00665417		STD DEV		CALC MEAN
CORRECTED TOTAL	23	10.27039583			0.08157308		3.01208333

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
PLANT	3	7.56034583	378.73	0.0001	3	7.56034583	378.73	0.0001
LEAF(PLANT)	8	2.63020000	49.41	0.0001	8	2.63020000	49.41	0.0001

TESTS OF HYPOTHESES USING THE TYPE I MS FOR LEAF(PLANT) AS AN ERROR TERM

SOURCE	DF	TYPE I SS	F VALUE	PR > F
PLANT	3	7.56034583	7.67	0.0097

Here we have specified the correct F test to be made. This agrees with the text, p. 287.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

p.p. 301 and 309, S+C.

```
1          DATA CITRUS;  INPUT FRUIT $ SUN $ LEAF;
2          X0=1;
3          X1=(FRUIT='ORANGE');
4          X2=(FRUIT='GRAPEFRU');
5          X3=(FRUIT='MADARIN');
6          X4=(SUN='SUN');
7          X5=(SUN='HALF_SHA');
8          X6=(SUN='SHADE');
9          Y1=0;
10         Y2=0;
11         IF SUN='SUN' THEN Y1=1;
12         IF SUN='SHADE' THEN Y1=-1;
13         IF SUN='SUN' THEN Y2=1;
14         IF SUN='HALF_SHA' THEN Y2=-2;
15         IF SUN='SHADE' THEN Y2=1;
16         CARDS;
```

$ — Indicates a character variable.

Setting up the X matrix

Setting up contrasts

```
NOTE: DATA SET WORK.CITRUS HAS 9 OBSERVATIONS AND 12 VARIABLES. 130 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.30 SECONDS AND 102K.

26    (A)  PROC PRINT;
27         VARIABLES X0-X6;

NOTE: THE PROCEDURE PRINT USED 0.20 SECONDS AND 102K.

28    (B)  PROC GLM;
29         MODEL LEAF= X1-X6;

NOTE: THE PROCEDURE GLM USED 0.41 SECONDS AND 156K.

30    (C)  PROC GLM;
31         CLASSES FRUIT SUN;
32         MODEL LEAF= FRUIT SUN;

NOTE: THE PROCEDURE GLM USED 0.39 SECONDS AND 156K.

33         PROC GLM;
34    (D)  CLASSES FRUIT SUN;
35         MODEL LEAF= FRUIT SUN FRUIT*SUN;

NOTE: THE PROCEDURE GLM USED 0.48 SECONDS AND 158K.

36         PROC GLM;
37    (E)  CLASSES FRUIT SUN;
38         MODEL LEAF= SUN FRUIT/ XPX;

NOTE: THE PROCEDURE GLM USED 0.46 SECONDS AND 156K.

39    (F)  PROC PRINT;
40         VARIABLES X0-X3 Y1 Y2;

NOTE: THE PROCEDURE PRINT USED 0.31 SECONDS AND 102K.
```

(A) Printing the X matrix.

(B) Running GLM using the X matrix.

(C) Running using the CLASSES statement.

(D) Running using the CLASSES statement and including an interaction term.
Interactions are denoted by asterisks between variables.

(E) reversing the order of variables entered.
XPX; — asking for the X'X matrix to be printed.

(F) Printing the matrix and using the contrasts.

STATISTICAL ANALYSIS SYSTEM

```
41      PROC GLM;
42  (G) CLASS FRUIT;
43      MODEL LEAF= Y1 Y2 FRUIT/ SOLUTION;
```

Running the model specified by the above matrix, but using the CLASSES statement for FRUIT.

(A)

OBS	X0	X1	X2	X3	X4	X5	X6
1	1	1	0	0	1	0	0
2	1	1	0	0	0	1	0
3	1	1	0	0	0	0	1
4	1	0	1	0	1	0	0
5	1	0	1	0	0	1	0
6	1	0	1	0	0	0	1
7	1	0	0	1	1	0	0
8	1	0	0	1	0	1	0
9	1	0	0	1	0	0	1

The X matrix.

(B)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: LEAF

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	2735.11111111	683.77777778	31.40	0.0028	0.969134	5.2764
ERROR	4	87.11111111	21.77777778 — agrees with MS(error) in text.		STD DEV		LEAF MEAN
CORRECTED TOTAL	8	2822.22222222			4.66666667		88.44444444

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	1	80.22222222	3.68	0.1274	0	0.00000000	.	.
X2	1	770.66666667	fruit 35.39	0.0040	0	0.00000000	.	.
X3	0	0.00000000	.	.	0	0.00000000	.	.
X4	1	1780.05555556	81.74	0.0008	0	0.00000000	.	.
X5	1	104.16666667	shading 4.78	0.0940	0	0.00000000	.	.
X6	0	0.00000000	.	.	0	0.00000000	.	.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	83.55555556 B	24.02	0.0001	3.47832796
X1	-5.00000000 B	-1.31	0.2597	3.81031738
X2	-22.66666667 B	-5.95	0.0040	3.81031738
X3	0.00000000 B	.	.	.
X4	34.00000000 B	8.92	0.0009	3.81031738
X5	8.33333333 B	2.19	0.0940	3.81031738
X6	0.00000000 B	.	.	.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(C)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: LEAF

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	2735.11111111	683.77777778	31.40	0.0028	0.969134	5.2764
ERROR	4	87.11111111	21.77777778		STD DEV		LEAF MEAN
CORRECTED TOTAL	8	2822.22222222			4.66666667		88.44444444

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
FRUIT	2	850.88888889	19.54	0.0086	2	850.88888889	19.54	0.0086
SUN	2	1884.22222222	43.26	0.0020	2	1884.22222222	43.26	0.0020

2 MS (sun) = 1884.222/2 = 942.111 which agrees with the mean square for shading in the text, p. 301.

GENERAL LINEAR MODELS PROCEDURE

(D)

DEPENDENT VARIABLE: LEAF

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	8	2822.22222222	352.77777778	99999.99	0.0000	1.000000	0.0000
ERROR	0	0.00000000	0.00000000		STD DEV		LEAF MEAN
CORRECTED TOTAL	8	2822.22222222			0.00000000		88.44444444

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
FRUIT	2	850.88888889	.	.	2	850.88888889	.	.
SUN	2	1884.22222222	.	.	2	1884.22222222	.	.
FRUIT*SUN	4	87.11111111	.	.	4	87.11111111	.	.

When there is only one observation per cell we cannot include an interaction term and still get an estimate of error. Before we were assuming there was no interaction term and using SS (fruit * sun) (= 87.111) as the error term.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

(E)

THE X'X MATRIX

DEPENDENT VARIABLE: LEAF

	INTERCEPT	SUN HALF_SHA	SUN SHADE	SUN SUN	grapefruit DUMMY001	mandarin DUMMY002	orange DUMMY003
INTERCEPT	9	3	3	3	3	3	3
SUN HALF_SHA	3	3	0	0	1	1	1
SUN SHADE	3	0	3	0	1	1	1
SUN SUN	3	0	0	3	1	1	1
DUMMY001	3	1	1	1	3	0	0
DUMMY002	3	1	1	1	0	3	0
DUMMY003	3	1	1	1	0	0	3

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: LEAF

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	2735.11111111	683.77777778	31.40	0.0028	0.969134	5.2764
ERROR	4	87.11111111	21.77777778		STD DEV		LEAF MEAN
CORRECTED TOTAL	8	2822.22222222			4.66666667		88.44444444

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
SUN	2	1884.22222222	43.26	0.0020	2	1884.22222222	43.26	0.0020
FRUIT	2	850.88888889	19.54	0.0086	2	850.88888889	19.54	0.0086

Note that changing the order of variables does not affect the SS. This is also reflected in the Type IV SS column.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(F)

OBS	X0	X1	X2	X3	Y1	Y2
1	1	1	0	0	1	1
2	1	1	0	0	0	-2
3	1	1	0	0	-1	1
4	1	0	1	0	1	1
5	1	0	1	0	0	-2
6	1	0	1	0	-1	1
7	1	0	0	1	1	1
8	1	0	0	1	0	-2
9	1	0	0	1	-1	1

Here is the X matrix set up using contrasts. The two contrasts Y1 and Y2 are equivalent to the columns X4, X5, and X6 and have the same total SS. We are separating the SS (sun) into two single degree of freedom contrast SS. Y1 is a contrast of sun versus shade and Y2 is a contrast of half-shade versus the average of the other two.

(G)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: LEAF

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	2735.11111111	683.77777778	31.40	0.0028	0.969134	5.2764
ERROR	4	87.11111111	21.77777778		STD DEV		LEAF MEAN
CORRECTED TOTAL	8	2822.22222222			4.66666667		88.44444444

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
Y1	1	1734.00000000	79.62	0.0009	1	1734.00000000	79.62	0.0009
Y2	1	150.22222222	6.90	0.0584	1	150.22222222	6.90	0.0584
FRUIT	2	850.88888889	19.54	0.0086	2	850.88888889	19.54	0.0086

SS for sun/shade contrast (p. 309)

SS for half-shade vs. rest (p. 309)

PARAMETER		ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		92.66666667 B	34.39	0.0001	2.69430126
Y1		17.00000000	8.92	0.0009	1.90515869
Y2		2.88888889	2.63	0.0584	1.09994388
FRUIT	GRAPEFRU	-17.66666667 B	-4.64	0.0098	3.81031738
	MADARIN	5.00000000 B	1.31	0.2597	3.81031738
	ORANGE	0.00000000 B	.	.	.

= $L_1/6$ (p. 309)

= $L_2/18$

Note that the total SS (sun) is the same: 1734 + 150.22 = 1884.22.

STATISTICAL ANALYSIS SYSTEM

p. 343 S+C.

```
1          DATA COLLARD;  INPUT REP SSIZE PERM RIBO;
2          X0=1;
3          R1=0;
4          IF REP=1 THEN R1=1;
5          R2=0;
6          IF REP=2 THEN R2=1;
7          R3=0;
8          IF REP=3 THEN R3=1;
9          X1=0;
10         IF SSIZE=0.25 AND PERM=0 THEN X1=1;
11         X2=0;
12         IF SSIZE=0.25 AND PERM=1 THEN X2=1;
13         X3=0;
14         IF SSIZE=1.00 AND PERM=0 THEN X3=1;
15         X4=0;
16         IF SSIZE=1.CO AND PERM=1 THEN X4=1;
17         S0=0;
18         IF SSIZE=0.25 THEN S0=1;
19         S1=0;
20         IF SSIZE=1.00 THEN S1=1;
21         P0=0;
22         IF PERM=0 THEN P0=1;
23         P1=0;
24         IF PERM=1 THEN P1=1;
25         S0XP0=0;
26         IF SSIZE=0.25 AND PERM=0 THEN S0XP0=1;
27         S0XP1=0;
28         IF SSIZE=0.25 AND PERM=1 THEN S0XP1=1;
29         S1XP0=0;
30         IF SSIZE=1.00 AND PERM=0 THEN S1XP0=1;
31         S1XP1=0;
32         IF SSIZE=1.00 AND PERM=1 THEN S1XP1=1;
33         SAMP=1;
34         IF SSIZE=0.25 THEN SAMP=-1;
35         PMAN=1;
36         IF PERM=0 THEN PMAN=-1;
37         SXP=SAMP*PMAN;
38         CARDS;
```

Setting up X matrix for A, B.

X matrix for C, D.

X matrix for E, F.

```
NOTE: DATA SET WORK.COLLARD HAS 12 OBSERVATIONS AND 23 VARIABLES. 69 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.39 SECONDS AND 102K.

51         PROC PRINT;
52 (A)     VARIABLES X0 R1 R2 R3 X1 X2 X3 X4;
```

Print X matrix which treats each combination of factors as a treatment.

```
NOTE: THE PROCEDURE PRINT USED 0.22 SECONDS AND 106K.

53 (B)     PROC GLM;
54         MODEL RIBO=R1 R2 R3 X1 X2 X3 X4;

NOTE: THE PROCEDURE GLM USED 0.51 SECONDS AND 158K.

55 (C)     PROC PRINT;
```

S T A T I S T I C A L A N A L Y S I S S Y S T E M

```
56          VARIABLES X0 R1 R2 R3 S0 S1 P0 P1 S0XP0 S0XP1 S1XP0 S1XP1;

NOTE: THE PROCEDURE PRINT USED 0.21 SECONDS AND 110K.

57          PROC GLM;
58 (D)      MODEL  RIBO=R1 R2 R3 S0 S1 P0 P1 S0XP0 S0XP1 S1XP0 S1XP1;

NOTE: THE PROCEDURE GLM USED 0.49 SECONDS AND 158K.

59          PROC PRINT;
60 (E)      VARIABLES X0 R1 R2 R3 SAMP PMAN SXP;

NOTE: THE PROCEDURE PRINT USED 0.19 SECONDS AND 104K.

61          PROC GLM;
62 (F)      MODEL RIBO=R1 R2 R3 SAMP PMAN SXP/SOLUTION ;

NOTE: THE PROCEDURE GLM USED 0.47 SECONDS AND 158K.

63          PROC GLM;
64 (G)      CLASSES REP SSIZE PERM;
65          MODEL RIBO=REP SSIZE PERM SSIZE*PERM;

NOTE: THE PROCEDURE GLM USED 0.47 SECONDS AND 158K.

NOTE: SAS USED 158K MEMORY.

NOTE: BARR, GOODNIGHT, SALL AND HELWIG
      SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605
```

This uses 0-1 indicator variables to do an analysis like that on p. 343.

Use CLASSES statement to get same X matrix as E or F.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(A)

X matrix

OBS	X0	R1	R2	R3	X1	X2	X3	X4
1	1	1	0	0	1	0	0	0
2	1	1	0	0	0	0	1	0
3	1	1	0	0	0	1	0	0
4	1	1	0	0	0	0	0	1
5	1	0	1	0	1	0	0	0
6	1	0	1	0	0	0	1	0
7	1	0	1	0	0	1	0	0
8	1	0	1	0	0	0	0	1
9	1	0	0	1	1	0	0	0
10	1	0	0	1	0	0	1	0
11	1	0	0	1	0	1	0	0
12	1	0	0	1	0	0	0	1

X0 Mean
R1-R3 correspond to Rep 1 - Rep 3

"Trt 1"	X1	corresponds to	.25	sample size and	without Perm.
"Trt 2"	X2	"	1.00	" "	without "
"Trt 3"	X3	"	.25	" "	with "
"Trt 4"	X4	"	1.00	" "	with "

Treats each combination of factors as a treatment.

(B)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: RIBO

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL (Reps + Trts)	5	769.29083333	153.85816667	18.81	0.0013	0.940029	8.9120
ERROR	6	49.07833333	8.17972222		STD DEV		RIBO MEAN
COPRECTED TOTAL	11	818.36916667			2.86002137		32.09166667

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
R1	1	0.88166667	0.11	0.7538	0	0.00000000	.	.
P2 } Reps	2 1	3.76 2.88000000	0.35	0.5746	0	0.00000000	.	.
R3	0	0.00000000	.	.	0	0.00000000	.	.
X1	1	441.70027778	54.00	0.0003	0	0.00000000	.	.
X2 } Trts	3 1	765.5 55.82722222	6.83	0.0400	0	0.00000000	.	.
X3	1	268.00166667	32.76	0.0012	0	0.00000000	.	.
X4	0	0.00000000	.	.	0	0.00000000	.	.

STATISTICAL ANALYSIS SYSTEM

(C)

X matrix

mean | reps (R1, R2, R3) | .25 (S0), 1.00 (S1) | without (P0), with (P1) | SOXP0 | SOXP1 | S1XP0 | S1XP1

OBS	XO	R1	R2	R3	S0	S1	P0	P1	SOXP0	SOXP1	S1XP0	S1XP1
1	1	1	0	0	1	0	1	0	1	0	0	0
2	1	1	0	0	0	1	1	0	0	0	1	0
3	1	1	0	0	1	0	0	1	0	1	0	0
4	1	1	0	0	0	1	0	1	0	0	0	1
5	1	0	1	0	1	0	1	0	1	0	0	0
6	1	0	1	0	0	1	1	0	0	0	1	0
7	1	0	1	0	1	0	0	1	0	1	0	0
8	1	0	1	0	0	1	0	1	0	0	0	1
9	1	0	0	1	1	0	1	0	1	0	0	0
10	1	0	0	1	0	1	1	0	0	0	1	0
11	1	0	0	1	1	0	0	1	0	1	0	0
12	1	0	0	1	0	1	0	1	0	0	0	1

sample size | permanganate | interactions

(D)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: RIBO — This is the same as on p. 343. If one uses ORTHO on the above matrix, the matrix on the next page results.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL (Rep + Trts.)	5	769.29083333	153.85816667	18.81	0.0013	0.940029	8.9120
ERROR	6	49.07833333	8.17972222		STD DEV		RIBO MEAN
CORRECTED TOTAL	11	818.36916667			2.86002137		32.09166667

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
R1 } Reps	1	0.88166667	0.11	0.7538	0	0.00000000	.	.
R2 } Reps	1	2.88000000	0.35	0.5746	0	0.00000000	.	.
R3 } Reps	0	0.00000000	.	.	0	0.00000000	.	.
S0 } Sample size	1	36.40083333	4.45	0.0794	0	0.00000000	.	.
S1 } Sample size	0	0.00000000	.	.	0	0.00000000	.	.
P0 } Permanganate	1	716.10750000	87.55	0.0001	0	0.00000000	.	.
P1 } Permanganate	0	0.00000000	.	.	0	0.00000000	.	.
SOXP0 } Interaction	1	13.02083333	1.59	0.2539	0	0.00000000	.	.
SOXP1 } Interaction	0	0.00000000	.	.	0	0.00000000	.	.
S1XP0 } Interaction	0	0.00000000	.	.	0	0.00000000	.	.
S1XP1 } Interaction	0	0.00000000	.	.	0	0.00000000	.	.

ts

(E)

S T A T I S T I C A L A N A L Y S I S S Y S T E M

X matrix

OBS	X0	R1	R2	R3	SAMP	PMAN	SXP
1	1	1	0	0	-1	-1	1
2	1	1	0	0	1	-1	-1
3	1	1	0	0	-1	1	-1
4	1	1	0	0	1	1	1
5	1	0	1	0	-1	-1	1
6	1	0	1	0	1	-1	-1
7	1	0	1	0	-1	1	-1
8	1	0	1	0	1	1	1
9	1	0	0	1	-1	-1	1
10	1	0	0	1	1	-1	-1
11	1	0	0	1	-1	1	-1
12	1	0	0	1	1	1	1

Note: since there are only two levels of sample size and of permanganate, we can combine the two levels of each into one column, where

-1 represents the lower level
+1 represents the higher level.

Variables	Correspond to
X0	Mean
R1-R3	Reps 1-3
Samp	Sample size
Pman	Permanganate
SxP	Interaction

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: RIBO

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	769.29083333	153.85816667	18.81	0.0013	0.940029	8.9120
ERROR	6	49.07833333	8.17972222		STD DEV		RIBO MEAN
CORRECTED TOTAL	11	818.36916667			2.86002137		32.09166667

	SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
Reps	R1	1	0.88166667	0.11	0.7538	0	0.00000000	.	.
	R2	1	2.88000000	0.35	0.5746	0	0.00000000	.	.
	R3	0	0.00000000	.	.	0	0.00000000	.	.
Trt	SAMP	1	36.40083333	4.45	0.0794	1	36.40083333	4.45	0.0794
	PMAN	1	716.10750000	87.55	0.0001	1	716.10750000	87.55	0.0001
	SXP Interaction	1	13.02083333	1.59	0.2539	1	13.02083333	1.59	0.2539

3.76

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	31.30000000 B	21.89	0.0001	1.43001068
R1	1.17500000 B	0.58	0.5824	2.02234050
R2	1.20000000 B	0.59	0.5746	2.02234050
R3	0.00000000 B	.	.	.
SAMP	-1.74166667	-2.11	0.0794	0.82561705
PMAN	-7.72500000	-9.36	0.0001	0.82561705
SXP	1.04166667	1.26	0.2539	0.82561705

$\frac{1}{2}$ × Factorial Effect Means.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(G)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: RIBO

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	769.29083333	153.85816667	18.81	0.0013	0.940029	8.9120
ERROR	6	49.07833333	8.17972222		STD DEV		RIBO MEAN
CORRECTED TOTAL	11	818.36916667			2.86002137		32.09166667

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
REP	2	3.76166667	0.23	0.8013	2	3.76166667	0.23	0.8013
Trts SSIZE	1	36.40083333	4.45	0.0794	1	36.40083333	4.45	0.0794
PERM	1	716.10750000	87.55	0.0001	1	716.10750000	87.55	0.0001
SSIZE*PERM	1	13.02083333	1.59	0.2539	1	13.02083333	1.59	0.2539

p. 352, S+C.

```
1          DATA COWPEA;
2          INPUT VARIETY SPACING BLOCK YIELD;
3          VART1=0;
4          VART2=0;
5          VART3=0;
6          IF VARIETY=1 THEN VART1=1;
7          IF VARIETY=2 THEN VART2=1;
8          IF VARIETY=3 THEN VART3=1;
9          SPACE1=0;
10         SPACE2=0;
11         SPACE3=0;
12         IF SPACING=4 THEN SPACE1=1;
13         IF SPACING=8 THEN SPACE2=1;
14         IF SPACING=12 THEN SPACE3=1;
15         BLOCK1=0;
16         BLOCK2=0;
17         BLOCK3=0;
18         BLOCK4=0;
19         IF BLOCK=1 THEN BLOCK1=1;
20         IF BLOCK=2 THEN BLOCK2=1;
21         IF BLOCK=3 THEN BLOCK3=1;
22         IF BLOCK=4 THEN BLOCK4=1;
23         INT11=VART1*SPACE1;
24         INT12=VART1*SPACE2;
25         INT13=VART1*SPACE3;
26         INT21=VART2*SPACE1;
27         INT22=VART2*SPACE2;
28         INT23=VART2*SPACE3;
29         INT31=VART3*SPACE1;
30         INT32=VART3*SPACE2;
31         INT33=VART3*SPACE3;
32         INTRCPT=1;
33         LSPACE=0;
34         QSPACE=1;
35         IF SPACING=4 THEN LSPACE=-1;
36         IF SPACING=12 THEN LSPACE=1;
37         IF SPACING=8 THEN QSPACE=-2;
38         LSPACE1=LSPACE*VART1;
39         LSPACE2=LSPACE*VART2;
40         LSPACE3=LSPACE*VART3;
41         QSPACE1=QSPACE*VART1;
42         QSPACE2=QSPACE*VART2;
43         QSPACE3=QSPACE*VART3;
44         CARDS;
```

These set up three indicator variables for variety...

...these do the same for spacing...

... and these for blocks.

These are the interaction variables for VARIETY by SPACING. Please note that they are formed by multiplication.

These set up the orthogonal polynomials for SPACING, while ...

...these set up the six degrees of freedom for blocks.

```
NOTE: DATA SET WORK.COWPEA HAS 36 OBSERVATIONS AND 32 VARIABLES. 50 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.45 SECONDS AND 102K.

81         PROC PRINT DATA=COWPEA;
82    (A)  VARIABLES INTRCPT BLOCK1 BLOCK2 BLOCK3 BLOCK4 VART1 VART2 VART3 SPACE1 SPACE2
83         SPACE3 INT11 INT12 INT13 INT21 INT22 INT23 INT31 INT32 INT33 YIELD;

NOTE: THE PROCEDURE PRINT USED 0.44 SECONDS AND 122K.
```

This is the X matrix for the first analysis.

```
84          PROC GLM DATA=COWPEA;
85      (B) MODEL YIELD=BLOCK1 BLOCK2 BLOCK3 BLOCK4 VART1 VART2 VART3 SPACE1 SPACE2 SPACE3
86          INT11 INT12 INT13 INT21 INT22 INT23 INT31 INT32 INT33;
```

This is the first analysis, as on p. 352.

```
NOTE: THE PROCEDURE GLM USED 0.79 SECONDS AND 162K.

87          PROC PRINT DATA=COWPEA;
88      (C) VARIABLES INTRCPT BLOCK1 BLOCK2 BLOCK3 VART1 VART2 LSPACE1 LSPACE2 LSPACE3
89          QSPACE1 QSPACE2 QSPACE3;
```

This is the X matrix for the more complicated analysis,

```
NOTE: THE PROCEDURE PRINT USED 0.36 SECONDS AND 112K.

90          PROC GLM DATA=COWPEA;
91      (D) MODEL YIELD=BLOCK1 BLOCK2 BLOCK3 VART1 VART2 LSPACE1 LSPACE2 LSPACE3
92          QSPACE1 QSPACE2 QSPACE3;
```

which is done here.

```
NOTE: THE PROCEDURE GLM USED 0.50 SECONDS AND 158K.

93          PROC GLM DATA=COWPEA;
94      (E) CLASSES SPACING BLOCK VARIETY;
95          MODEL YIELD=BLOCK VARIETY SPACING VARIETY*SPACING;
```

Using CLASSES statements, we can do B much more easily.

```
NOTE: THE PROCEDURE GLM USED 0.50 SECONDS.

96          PROC GLM DATA=COWPEA;
97      (F) CLASSES BLOCK VARIETY;
98          MODEL YIELD=BLOCK VARIETY SPACING*VARIETY SPACING*SPACING*VARIETY;
```

similarly, this is how one does D with a CLASSES statement; please note that SPACING is not included in the CLASSES statement.

```
NOTE: THE PROCEDURE GLM USED 0.61 SECONDS AND 158K.

NOTE: SAS USED 162K MEMORY.

NOTE: BARR, GOODNIGHT, SALL AND HELWIG
      SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605
```

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(A)

OBS	INTRCPT	BLOCK1	BLOCK2	BLOCK3	BLOCK4	VART1	VART2	VART3	SPACE1	SPACE2	SPACE3	INT11	INT12	INT13	INT21	INT22	INT23	INT31	INT32	INT33	YIELD
1	1	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	56
2	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	45
3	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	43
4	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	46
5	1	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	60
6	1	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	50
7	1	0	0	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	45
8	1	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	48
9	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	66
10	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	57
11	1	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	50
12	1	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	50
13	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	65
14	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	61
15	1	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	60
16	1	0	0	0	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	63
17	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	60
18	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	58
19	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	56
20	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	60
21	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	53
22	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	53
23	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	48
24	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	55
25	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	60
26	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	61
27	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	50
28	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	53
29	1	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	62
30	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	68
31	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	67
32	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	60
33	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	73
34	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	77
35	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	77
36	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	65
		Blocks				Varieties			Spacing			Variety × Spacing									Y

GENERAL LINEAR MODELS PROCEDURE

(B) This is the same analysis as on the bottom of p. 352.

DEPENDENT VARIABLE: YIELD

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	2203.52777778	200.32070707	11.34	0.0001	0.838596	7.2722
ERROR	24	424.11111111 -SS(error)	17.67129630 -MS (error)		STD DEV		YIELD MEAN
CORRECTED TOTAL	35	2627.63888889			4.20372410		57.80555556

These, added together, give 255.64, SS (blocks).

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
BLOCK1	1	178.89814815	10.12	0.0040	0	0.00000000	.	.
BLOCK2	1	75.85185185	4.29	0.0492	0	0.00000000	.	.
BLOCK3	1	0.88888889	0.05	0.8244	0	0.00000000	.	.
BLOCK4	(0)	0.00000000	.	.	0	0.00000000	.	.
VART1	1	754.01388889	42.67	0.0001	0	0.00000000	.	.
VART2	1	273.37500000	15.47	0.0006	0	0.00000000	.	.
VART3	(0)	0.00000000	.	.	0	0.00000000	.	.
SPACE1	1	117.55555556	6.65	0.0165	0	0.00000000	.	.
SPACE2	1	37.50000000	2.12	0.1581	0	0.00000000	.	.
SPACE3	(0)	0.00000000	.	.	0	0.00000000	.	.
INT11	1	14.69444444	0.83	0.3709	0	0.00000000	.	.
INT12	1	18.75000000	1.06	0.3132	0	0.00000000	.	.
INT13	(0)	0.00000000	.	.	0	0.00000000	.	.
INT21	1	507.00000000	28.69	0.0001	0	0.00000000	.	.
INT22	1	225.00000000	12.73	0.0016	0	0.00000000	.	.
INT23	(0)	0.00000000	.	.	0	0.00000000	.	.
INT31	(0)	0.00000000	.	.	0	0.00000000	.	.
INT32	(0)	0.00000000	.	.	0	0.00000000	.	.
INT33	(0)	0.00000000	.	.	0	0.00000000	.	.

These 0's denote the redundant variables, or columns in the X matrix. They may be omitted without any effect.

give 1027.39, SS (var.)

give 155.06, SS (spacings)

give 765.444, SS (interactions)

These are all zero because the redundant columns are included.

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	70.75000000 B	29.15	0.0001	2.42702124
BLOCK1	6.11111111 B	3.08	0.0051	1.98165454
BLOCK2	3.33333333 B	1.68	0.1055	1.98165454
BLOCK3	-0.44444444 B	-0.22	0.8244	1.98165454
BLOCK4	0.00000000 B	.	.	.
VART1	-17.25000000 B	-5.80	0.0001	2.97248182
VART2	-20.75000000 B	-6.98	0.0001	2.97248182
VART3	0.00000000 B	.	.	.
SPACE1	-17.00000000 B	-5.72	0.0001	2.97248182
SPACE2	-8.75000000 B	-2.94	0.0071	2.97248182
SPACE3	0.00000000 B	.	.	.
INT11	8.75000000 B	2.08	0.0482	4.20372410
INT12	3.75000000 B	0.89	0.3812	4.20372410
INT13	0.00000000 B	.	.	.
INT21	27.00000000 B	6.42	0.0001	4.20372410
INT22	15.00000000 B	3.57	0.0016	4.20372410

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YIELD

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INT23	0.00000000 B	.	.	.
INT31	0.00000000 B	.	.	.
INT32	0.00000000 B	.	.	.
INT33	0.00000000 B	.	.	.

This is the X matrix for the second part of the analysis. Y is as before.

OBS	INTRCPT	BLOCK1	BLOCK2	BLOCK3	VART1	VART2	LSPACE1	LSPACE2	LSPACE3	QSPACE1	QSPACE2	QSPACE3
(C)												
1	1	1	0	0	1	0	-1	0	0	1	0	0
2	1	0	1	0	1	0	-1	0	0	1	0	0
3	1	0	0	1	1	0	-1	0	0	1	0	0
4	1	0	0	0	1	0	-1	0	0	1	0	0
5	1	1	0	0	1	0	0	0	0	-2	0	0
6	1	0	1	0	1	0	0	0	0	-2	0	0
7	1	0	0	1	1	0	0	0	0	-2	0	0
8	1	0	0	0	1	0	0	0	0	-2	0	0
9	1	1	0	0	1	0	1	0	0	1	0	0
10	1	0	1	0	1	0	1	0	0	1	0	0
11	1	0	0	1	1	0	1	0	0	1	0	0
12	1	0	0	0	1	0	1	0	0	1	0	0
13	1	1	0	0	0	1	0	-1	0	0	1	0
14	1	0	1	0	0	1	0	-1	0	0	1	0
15	1	0	0	1	0	1	0	-1	0	0	1	0
16	1	0	0	0	0	1	0	-1	0	0	1	0
17	1	1	0	0	0	1	0	0	0	0	-2	0
18	1	0	1	0	0	1	0	0	0	0	-2	0
19	1	0	0	1	0	1	0	0	0	0	-2	0
20	1	0	0	0	0	1	0	0	0	0	-2	0
21	1	1	0	0	0	1	0	1	0	0	1	0
22	1	0	1	0	0	1	0	1	0	0	1	0
23	1	0	0	1	0	1	0	1	0	0	1	0
24	1	0	0	0	0	1	0	1	0	0	1	0
25	1	1	0	0	0	0	0	0	-1	0	0	1
26	1	0	1	0	0	0	0	0	-1	0	0	1
27	1	0	0	1	0	0	0	0	-1	0	0	1
28	1	0	0	0	0	0	0	0	-1	0	0	1
29	1	1	0	0	0	0	0	0	0	0	0	-2
30	1	0	1	0	0	0	0	0	0	0	0	-2
31	1	0	0	1	0	0	0	0	0	0	0	-2
32	1	0	0	0	0	0	0	0	0	0	0	-2
33	1	1	0	0	0	0	0	0	1	0	0	1
34	1	0	1	0	0	0	0	0	1	0	0	1
35	1	0	0	1	0	0	0	0	1	0	0	1
36	1	0	0	0	0	0	0	0	1	0	0	1

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(D)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YIELD

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	2203.52777778	200.32070707	11.34	0.0001	0.838596	7.2722
ERROR	24	424.11111111	17.67129630		STD DEV		YIELD MEAN
CORRECTED TOTAL	35	2627.63888889			4.20372410		57.80555556

Error is the same as for B.

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
BLOCK1	1	178.89814815	10.12	0.0040	1	168.05555556	9.51	0.0051
BLOCK2	1	75.85185185	4.29	0.0492	1	50.00000000	2.83	0.1055
BLOCK3	1	0.88888889	0.05	0.8244	1	0.88888889	0.05	0.8244
VART1	1	754.01388889	42.67	0.0001	1	1027.04166667	58.12	0.0001
VART2	1	273.37500000	15.47	0.0006	1	273.37500000	15.47	0.0006
LSPACE1	1	136.12500000	7.70	0.0105	1	136.12500000	7.70	0.0105
LSPACE2	1	200.00000000	11.32	0.0026	1	200.00000000	11.32	0.0026
LSPACE3	1	578.00000000	32.71	0.0001	1	578.00000000	32.71	0.0001
QSPACE1	1	2.04166667	0.12	0.7369	1	2.04166667	0.12	0.7369
QSPACE2	1	4.16666667	0.24	0.6317	1	4.16666667	0.24	0.6317
QSPACE3	1	0.16666667	0.01	0.9234	1	0.16666667	0.01	0.9234

SS (blocks) same

SS (var.) same

The six terms as on bottom of p. 353

These are the same as Type I SS, as the polynomials are already orthogonalized to the intercept.

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	62.16666667	36.22	0.0001	1.71616318
BLOCK1	6.11111111	3.08	0.0051	1.98165454
BLOCK2	3.33333333	1.68	0.1055	1.98165454
BLOCK3	-0.44444444	-0.22	0.8244	1.98165454
VART1	-13.08333333	-7.62	0.0001	1.71616318
VART2	-6.75000000	-3.93	0.0006	1.71616318
LSPACE1	4.12500000	2.78	0.0105	1.48624091
LSPACE2	-5.00000000	-3.36	0.0026	1.48624091
LSPACE3	8.50000000	5.72	0.0001	1.48624091
QSPACE1	0.29166667	0.34	0.7369	0.85808159
QSPACE2	-0.41666667	-0.49	0.6317	0.85808159
QSPACE3	0.08333333	0.10	0.9234	0.85808159

the three b's for the linear part

the three b's for the quadratic part.

Note that the coefficients for the second variety differ greatly from those for the first and third.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

(E)

DEPENDENT VARIABLE: YIELD

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	2203.52777778	200.32070707	11.34	0.0001	0.838596	7.2722
ERROR	24	424.11111111	17.67129630		STD DEV		YIELD MEAN
CORRECTED TOTAL	35	2627.63888889			4.20372410		57.80555556

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
BLOCK	3	255.63888889	4.82	0.0091	3	255.63888889	4.82	0.0091
VARIETY	2	1027.38888889	29.07	0.0001	2	1027.38888889	29.07	0.0001
SPACING	2	155.05555556	4.39	0.0238	2	155.05555556	4.39	0.0238
SPACING*VARIETY	4	765.44444444	10.83	0.0001	4	765.44444444	10.83	0.0001

This is the same analysis as B. Note that Type IV is the same as Type I - i.e. it doesn't matter if we run BLOCK, VARIETY, or SPACING first, second, or third. This is because there are no missing data points.

GENERAL LINEAR MODELS PROCEDURE

(F)

DEPENDENT VARIABLE: YIELD

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	2203.52777778	200.32070707	11.34	0.0001	0.838596	7.2722
ERROR	24	424.11111111	17.67129630		STD DEV		YIELD MEAN
CORRECTED TOTAL	35	2627.63888889			4.20372410		57.80555556

equals 136.125 + 200 + 578.

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
BLOCK	3	255.63888889	4.82	0.0091	3	255.63888889	4.82	0.0091
VARIETY	2	1027.38888889	29.07	0.0001	2	38.16666667	1.08	0.3556
SPACING*VARIETY	3	914.12500000	17.24	0.0001	3	9.24744898	0.17	0.9126
SPACIN*SPACIN*VARIET	3	6.37500000	0.12	0.9473	3	6.37500000	0.12	0.9473

equals 2.0416 + 4.166 + .166

This is the hardest to understand. First, remember that the orthogonal polynomials may be calculated using ORTHO; one starts with $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$ and orthogonalizes each successive column. Regression has ORTHO "built in" so to speak, so instead of using the polynomials -1, 0, 1 and 1, -2, 1 one may just use 1, 2, 3 and 1, 4, 9, or 4, 8, 12 and 16, 64, 144, looking only at Type I sums of squares.

p. 433, S+C.

```
1          DATA SERUM; INPUT STATE $ AGE CHOL;
2          X0=1;
3          X4=0;
4          X5=0;
5          X1=(STATE='IOWA');
6          X2=(STATE='NEBRASKA');
7          X3=AGE;
8          IF STATE='IOWA' THEN X4=AGE;
9          IF STATE='NEBRASKA' THEN X5=AGE;
10         IF STATE='IOWA' THEN X8='I';
11         IF STATE='NEBRASKA' THEN X8='N';
12         CARDS;
```

Setting up the X matrix.

A special variable for fancy plots later.

```
NOTE: DATA SET WORK.SERUM HAS 30 OBSERVATIONS AND 10 VARIABLES. 169 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.25 SECONDS AND 102K.

43 (A)     PROC PRINT;
44         VARIABLES X0-X5;

NOTE: THE PROCEDURE PRINT USED 0.28 SECONDS AND 102K.

45 (B)     PROC GLM;
46         MODEL CHOL= X1-X5;

NOTE: THE PROCEDURE GLM USED 0.48 SECONDS AND 154K.

47 (C)     PROC GLM;
48         MODEL CHOL=X1-X3/ NOINT SOLUTION;

NOTE: THE PROCEDURE GLM USED 0.43 SECONDS AND 152K.

49 (D)     PROC GLM;
50         CLASS STATE;
51         MODEL CHOL= STATE AGE STATE*AGE;

NOTE: THE PROCEDURE GLM USED 0.45 SECONDS AND 152K.

52 (E)     PROC GLM;
53         MODEL CHOL= X0-X3/ NOINT SOLUTION;
54         OUTPUT OUT=NEW1 PREDICTED=YHAT1 RESIDUAL=RESID1;

NOTE: DATA SET WORK.NEW1 HAS 30 OBSERVATIONS AND 12 VARIABLES. 140 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.52 SECONDS AND 174K.

55 (F)     PROC PLOT;
56         PLOT YHAT1*RESID1=X8;

NOTE: THE PROCEDURE PLOT USED 0.37 SECONDS AND 120K.

57 (G)     PROC GLM;
58         MODEL CHOL= X3 X0 X1 X2/ NOINT SOLUTION;

NOTE: THE PROCEDURE GLM USED 0.38 SECONDS AND 154K.
```

Printing the X matrix.

Getting an F test of different slopes by the successive models approach. When the X4 variable enters the equation (Type I SS) we get the SS explained by two different slopes.

Get estimates of the pooled slope and intercepts.

A fancy way to test whether the slopes are the same. The interaction term will give the required F test.

Fitting a model with just a single slope but two intercepts. The last line is so that we can plot residuals.

This plots predicted versus residuals. The "=X8" just means we plot variable X8 on the graph (N for Nebraska, I for Iowa)

Getting a test for different intercepts by sucessive models. When X1 enters we will get the required SS.

STATISTICAL ANALYSIS SYSTEM

```
59  (H)   PROC GLM;
60        MODEL CHOL= X3;
61        OUTPUT OUT=NEW2 PREDICTED=YHAT2 RESIDUAL=RESID2;
```

Running a one intercept, one slope model so we can look at the residuals.

```
NOTE: DATA SET WORK.NEW2 HAS 30 OBSERVATIONS AND 14 VARIABLES. 119 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.45 SECONDS AND 170K.

62  (I)   PROC PLOT;
63        PLOT YHAT2*RESID2=X8;
```

Our residual plot.

```
NOTE: THE PROCEDURE PLOT USED 0.35 SECONDS AND 120K.


NOTE: BARR, GOODNIGHT, SALL AND HELWIG
      SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605
```

S T A T I S T I C A L A N A L Y S I S S Y S T E M

OBS	X0	X1	X2	X3	X4	X5
1	1	1	0	46	46	0
2	1	1	0	52	52	0
3	1	1	0	39	39	0
4	1	1	0	65	65	0
5	1	1	0	54	54	0
6	1	1	0	33	33	0
7	1	1	0	49	49	0
8	1	1	0	76	76	0
9	1	1	0	71	71	0
10	1	1	0	41	41	0
11	1	1	0	58	58	0
12	1	0	1	18	0	18
13	1	0	1	30	0	30
14	1	0	1	44	0	44
15	1	0	1	47	0	47
16	1	0	1	33	0	33
17	1	0	1	58	0	58
18	1	0	1	78	0	78
19	1	0	1	70	0	70
20	1	0	1	51	0	51
21	1	0	1	67	0	67
22	1	0	1	43	0	43
23	1	0	1	31	0	31
24	1	0	1	44	0	44
25	1	0	1	21	0	21
26	1	0	1	58	0	58
27	1	0	1	56	0	56
28	1	0	1	63	0	63
29	1	0	1	19	0	19
30	1	0	1	42	0	42

(A) printing the X matrix

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(B)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: CHOL

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	55141.80269439	18380.60089813	9.87	0.0002	0.532582	20.1919
ERROR	26	48394.86397228	1861.34092201		STD DEV		CHOL MEAN
CORRECTED TOTAL	29	103536.66666667			43.14326045		213.66666667

F test for differences in slopes. We conclude no difference.

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	1	612.69537480	0.33	0.5711	0	0.00000000	.	.
X2	0	0.00000000	.	.	0	0.00000000	.	.
X3	1	53820.05813042	28.91	0.0001	0	0.00000000	.	.
X4	1	709.04918917	0.38	0.5425	0	0.00000000	.	.
X5	0	0.00000000	.	.	0	0.00000000	.	.

This is the "708" in S+C, Table 14.6.2, line 5.

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	101.29776609 B	3.57	0.0014	28.35661703
X1	-65.48655225 B	-1.06	0.3005	61.98336764
X2	0.00000000 B	.	.	.
X3	2.52043808 B	4.36	0.0002	0.57833894
X4	0.71770686 B	0.62	0.5425	1.16284514
X5	0.00000000 B	.	.	.

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS HO: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

STATISTICAL ANALYSIS SYSTEM

GENERAL LINEAR MODELS PROCEDURE

(C)

DEPENDENT VARIABLE: CHOL

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	1424036.08683855	474678.69561285	261.00	0.0001	0.966667	19.9590
ERROR	27	49103.91316145	1818.66345042		STD DEV		CHOL MEAN
UNCORRECTED TOTAL	30	1473140.00000000			42.64579054		213.66666667

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	1	474656.81818182	260.99	0.0001	1	8808.98883833	4.84	0.0365
X2	1	895559.21052632	492.43	0.0001	1	25654.10555201	14.11	0.0008
X3	1	53820.05813042	29.59	0.0001	1	53820.05813042	29.59	0.0001

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X1	64.48977167	2.20	0.0365	29.30245306
X2	93.14079645	3.76	0.0008	24.79920040
X3	2.69796663	5.44	0.0001	0.49595316

estimates of, respectively, α_I, α_N, b (pooled slope)

$\hat{\alpha}_I = \bar{Y}_I - \hat{b}\bar{X}_I = 207.7 - 2.70(53.1) = 64.33$; $\hat{\alpha}_N = \bar{Y}_N - \hat{b}\bar{X}_N = 217.1 - 2.70(45.9) = 93.17$

see p. 435.

GENERAL LINEAR MODELS PROCEDURE

(D)

DEPENDENT VARIABLE: CHOL

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	55141.80269439	18380.60089813	9.87	0.0002	0.532582	20.1919
ERROR	26	48394.86397228	1861.34092201		STD DEV		CHOL MEAN
CORRECTED TOTAL	29	103536.66666667			43.14326045		213.66666667

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
STATE	1	612.69537480	0.33	0.5711	1	2077.68565493	1.12	0.3005
AGE	1	53820.05813042	28.91	0.0001	1	45647.12507786	24.52	0.0001
AGE*STATE	1	709.04918917	0.38	0.5425	1	709.04918917	0.38	0.5425

F test for differences in slope.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(E)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: CHOL

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	1424036.08683855	474678.69561285	261.00	0.0001	0.966667	19.9590
ERROR	27	49103.91316145	1818.66345042		STD DEV		CHOL MEAN
UNCORRECTED TOTAL	30	1473140.00000000			42.64579054		213.66666667

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X0	1	1369603.33333333	753.08	0.0001	0	0.00000000	.	.
X1	1	612.69537480	0.34	0.5664	0	0.00000000	.	.
X2	0	0.00000000	.	.	0	0.00000000	.	.
X3	1	53820.05813042	29.59	0.0001	1	53820.05813042	29.59	0.0001

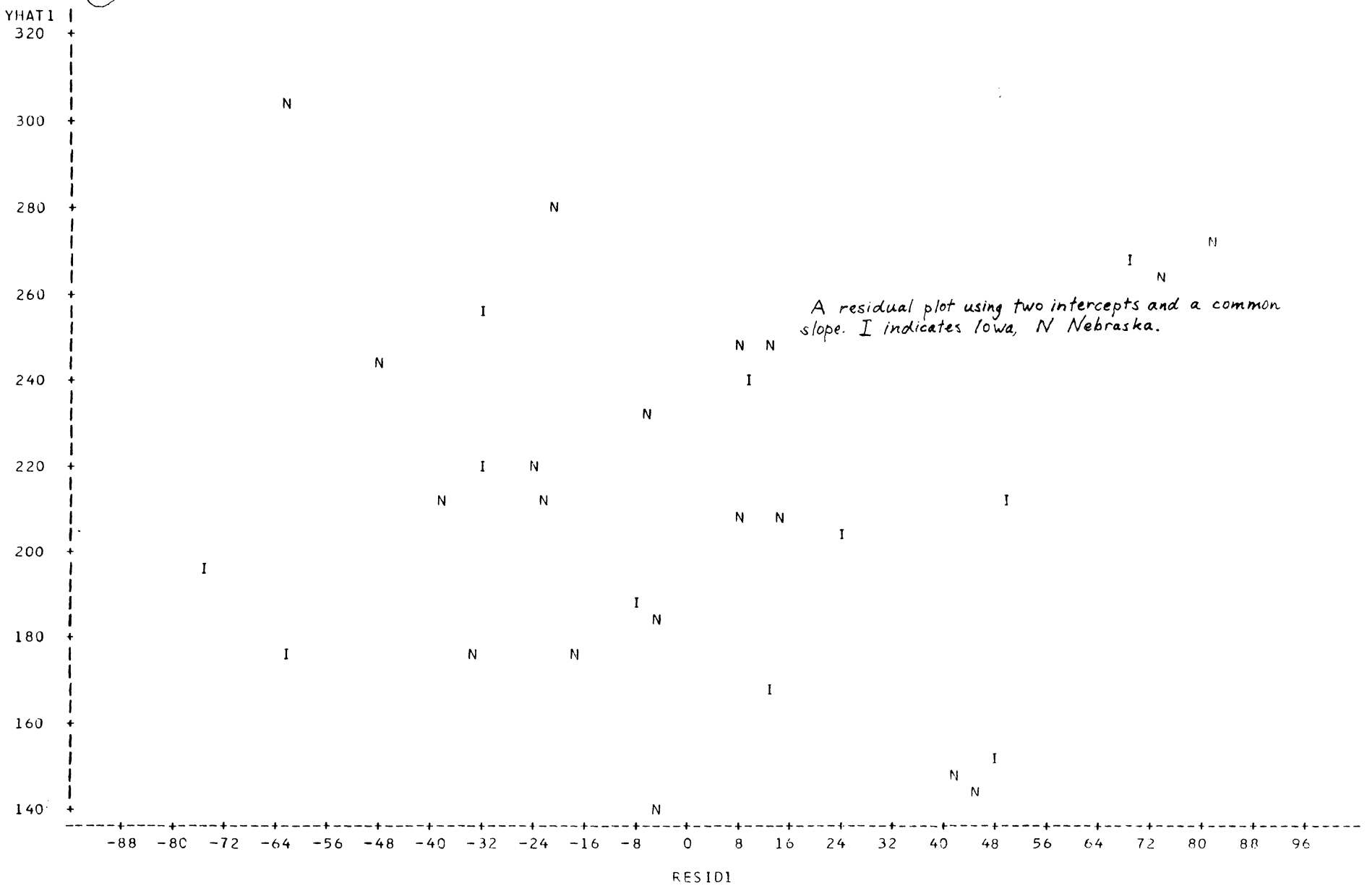
PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X0	93.14079645 B	3.76	0.0008	24.79920040
X1	-28.65102477 B	-1.73	0.0947	16.54098056
X2	0.00000000 B	.	.	.
X3	2.69796663	5.44	0.0001	0.49595316

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS HO: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(F)

PLOT OF YHAT1*RESID1 SYMBOL IS VALUE OF X8



A residual plot using two intercepts and a common slope. I indicates Iowa, N Nebraska.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

GENERAL LINEAR MODELS PROCEDURE

(G)

DEPENDENT VARIABLE: CHOL

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	1424036.08683855	474678.69561285	261.00	0.0001	0.966667	19.9590
ERROR	27	49103.91316145	1818.66345042		STD DEV		CHOL MEAN
UNCORRECTED TOTAL	30	1473140.00000000			42.64579054		213.66666667

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X3	1	1393749.54383462	766.36	0.0001	1	53820.05813042	29.59	0.0001
X0	1	24830.09329701	13.65	0.0010	0	0.00000000	.	.
X1	1	5456.44970691	3.00	0.0947	0	0.00000000	.	.
X2	0	0.00000000	.	.	0	0.00000000	.	.

This is the "5450" in S+C Table 14.6.2, line 8, to test for different intercepts.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X3	2.69796663	5.44	0.0001	0.49595316
X0	93.14079645 B	3.76	0.0008	24.79920040
X1	-28.65102477 B	-1.73	0.0947	16.54098056
X2	0.00000000 B	.	.	.

(H)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: CHOL

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	48976.30379830	48976.30379830	25.13	0.0001	0.473033	20.6596
ERROR	28	54560.36286836	1948.58438816		STD DEV		CHOL MEAN
CORRECTED TOTAL	29	103536.66666667			44.14277277		213.66666667

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X3	1	48976.30379830	25.13	0.0001	1	48976.30379830	25.13	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	91.57157789	3.57	0.0013	25.65258477
X3	2.51396888	5.01	0.0001	0.50144869

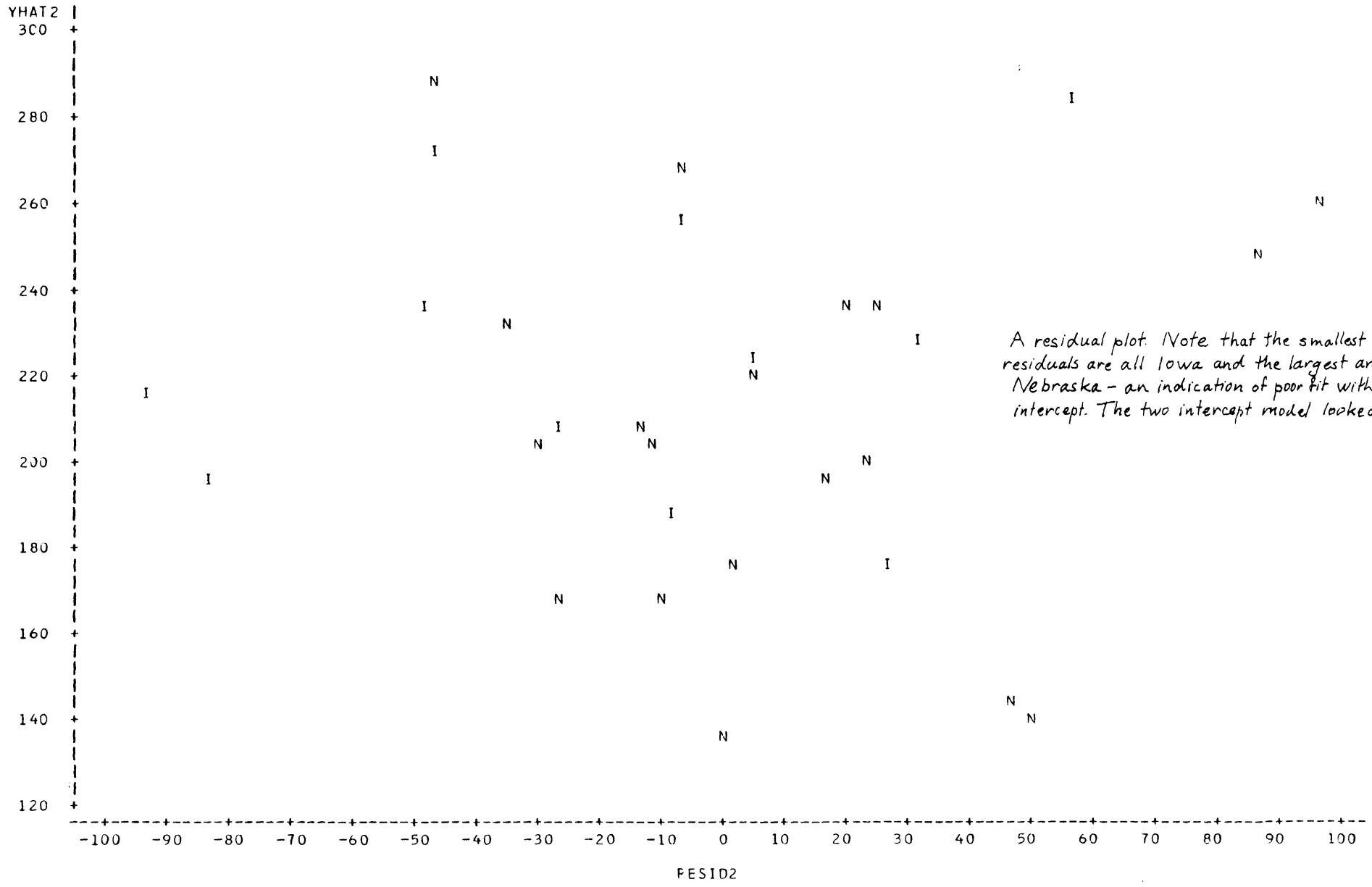
pooled intercepts

pooled slope

(I)

S T A T I S T I C A L A N A L Y S I S S Y S T E M

PLOT OF YHAT2*RESID2 SYMBOL IS VALUE OF X8



A residual plot. Note that the smallest residuals are all Iowa and the largest are all Nebraska - an indication of poor fit with just one intercept. The two intercept model looked better.

STATISTICAL ANALYSIS SYSTEM

p. 422, S+C

```
1          DATA LEPROSY;
2          INPUT DRUG $ X Y;
3          DRUG1=0;DRUG2=0;DRUG3=0;
4          IF DRUG='A' THEN DRUG1=1;
5          IF DRUG='D' THEN DRUG2=1;
6          IF DRUG='F' THEN DRUG3=1;
7          INTRCPT=1;
8          CARDS;
```

This $ signifies a character value of a variable.

These statements create three 0-1 indicator variables.

NOTE: DATA SET WORK.LEPROSY HAS 30 OBSERVATIONS AND 7 VARIABLES. 217 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.26 SECONDS AND 102K.

```
39   (A)   PROC PRINT DATA=LEPROSY;
40         VARIABLES INTRCPT DRUG1 DRUG2 DRUG3 X Y;
```

This is to print out the X matrix, which is used in the next GLM.

NOTE: THE PROCEDURE PRINT USED 0.23 SECONDS AND 102K.

```
41   (B)   PROC GLM DATA=LEPROSY;
42         MODEL Y=DRUG1 DRUG2 DRUG3 X ;
```

A first analysis, with all three dummy variables.

NOTE: THE PROCEDURE GLM USED 0.76 SECONDS AND 152K.

```
43   (C)   PROC GLM DATA=LEPROSY;
44         MODEL Y=X DRUG1 DRUG2 DRUG3/ SOLUTION NOINT;
```

This is an analysis to calculate treatment means. Please note this.

NOTE: THE PROCEDURE GLM USED 0.88 SECONDS AND 152K.

```
45         PROC GLM DATA=LEPROSY;
46   (D)   CLASSES DRUG;
47         MODEL Y=DRUG X;
```

Using a CLASSES statement, we can with much less effort produce the same results as B.

NOTE: THE PROCEDURE GLM USED 0.47 SECONDS AND 152K.

```
48         PROC GLM DATA=LEPROSY;
49   (E)   CLASSES DRUG;
50         MODEL Y=X DRUG;
51         OUTPUT OUT=RS1 RESIDUAL=RES1 PREDICTED=PRE1;
```

The X is run first here. Also residuals are requested; if residuals had been requested from D, the same numbers would have resulted. Calculation of residuals depends only on what variables are in the model, not in what order they enter.

NOTE: DATA SET WORK.RS1 HAS 30 OBSERVATIONS AND 9 VARIABLES. 171 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.41 SECONDS AND 172K.

```
52   (F)   PROC PLOT DATA=RS1;
53         PLOT RES1*PRE1;
```

A residual plot.

NOTE: THE PROCEDURE PLOT USED 0.30 SECONDS AND 118K.

```
54   (G)   PROC GLM DATA=LEPROSY;
55         CLASSES DRUG;
56         MODEL Y=DRUG X X*DRUG;
```

Classical covariance assumes equal slopes for all treatments. Here we test that assumption, using the quickest approach.

NOTE: THE PROCEDURE GLM USED 0.50 SECONDS AND 154K.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(A)

These three together.

OBS	INTRCPT	DRUG1	DRUG2	DRUG3	X	Y
1	1	1	0	0	11	6
2	1	1	0	0	8	0
3	1	1	0	0	5	2
4	1	1	0	0	14	8
5	1	1	0	0	19	11
6	1	1	0	0	6	4
7	1	1	0	0	10	13
8	1	1	0	0	6	1
9	1	1	0	0	11	8
10	1	1	0	0	3	0
11	1	0	1	0	6	0
12	1	0	1	0	6	2
13	1	0	1	0	7	3
14	1	0	1	0	8	1
15	1	0	1	0	18	18
16	1	0	1	0	8	4
17	1	0	1	0	19	14
18	1	0	1	0	8	9
19	1	0	1	0	5	1
20	1	0	1	0	15	9
21	1	0	0	1	16	13
22	1	0	0	1	13	10
23	1	0	0	1	11	18
24	1	0	0	1	9	5
25	1	0	0	1	21	23
26	1	0	0	1	16	12
27	1	0	0	1	12	5
28	1	0	0	1	12	16
29	1	0	0	1	7	1
30	1	0	0	1	12	20

GENERAL LINEAR MODELS PROCEDURE

(B)

DEPENDENT VARIABLE: Y

To test regression significance, fit it last.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	871.49740304	290.49913435	18.10	0.0001	0.676261	50.7060
ERROR	26	417.20259696	16.04625373 — *Deviation mean square.*		STD DEV		Y MEAN
CORRECTED TOTAL	29	1288.70000000	*Deviations from regression.*		4.00577754		7.90000000

Since we included all three, they cancel out.

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
DRUG1	1	101.40000000	6.32	0.0185	0	0.00000000	.	.
DRUG2	1	192.20000000	11.98	0.0019	0	0.00000000	.	.
DRUG3	0	0.00000000	.	.	0	0.00000000	.	.
X	1	577.89740304	36.01	0.0001	1	577.89740304	36.01	0.0001

Between drugs has two degrees of freedom, so the last fitted is ignored.

This is the reduction due to regression.

F test for regression

One tests the regression fitting it after the treatments.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	-0.43467116 B	-0.18	0.8617	2.47135356
DRUG1	-3.44613828 B	-1.83	0.0793	1.88678065
DRUG2	-3.33716695 B	-1.80	0.0835	1.85386642
DRUG3	0.00000000 B	.	.	.
X	0.98718381	6.00	0.0001	0.16449757

b for regression — $6^2 = 36$, the F test for regression.

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

(c)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

This is a special run to estimate adjusted means.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	2743.79740304	685.94935076	42.75	0.0001	0.868016	50.7060
ERROR	26	417.20259696	16.04625373		STD DEV		Y MEAN
UNCORRECTED TOTAL	30	3161.00000000			4.00577754		7.90000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X	1	2602.04391072	162.16	0.0001	1	577.89740304	36.01	0.0001
DRUG1	1	55.22783408	3.44	0.0749	1	61.25923855	3.82	0.0616
DRUG2	1	86.02926534	5.36	0.0287	1	52.95963781	3.30	0.0838
DRUG3	1	0.49639289	0.03	0.8617	1	0.49639289	0.03	0.8617

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X	0.98718381	6.00	0.0001	0.16449757
DRUG1	-3.88080944	-1.95	0.0616	1.98620170
DRUG2	-3.77183811	-1.82	0.0838	2.07619131
DRUG3	-0.43467116	-0.18	0.8617	2.47135356

These are adjusted means - $b\bar{X}_{..}$. To get the adjusted means, add $b\bar{X}_{..}$, or $.998 \cdot 10.73$. (See ABDO for verification.)

The $\underset{\sim}{X}'\underset{\sim}{X}|\underset{\sim}{X}'\underset{\sim}{Y}$ matrix and ABDO are:

10	0	0	ΣX_1	ΣY_1
	10	0	ΣX_2	ΣY_2
		10	ΣX_3	ΣY_3
			ΣX	ΣY
10	0	0	ΣX_1	ΣY_1
1	0	0	$\bar{X}_1$	$\bar{Y}_1$
	10	0	ΣX_2	ΣY_2
	1	0	$\bar{X}_2$	$\bar{Y}_2$
		10	ΣX_3	ΣY_3
		1	$\bar{X}_3$	$\bar{Y}_3$
			$\Sigma(X-\bar{X})^2$	$\Sigma(X-\bar{X})(Y-\bar{Y})$
			1	b

Solving up, $b = (X-\bar{X})(Y-\bar{Y})/(X-\bar{X})^2$, $\mu_3 = \bar{Y}_3 - b\bar{X}_3$, $\mu_2 = \bar{Y}_2 - b\bar{X}_2$, $\mu_1 = \bar{Y}_1 - b\bar{X}_1$. Therefore, we must add in a $b\bar{X}$, in order to get adjusted means. This raises an interesting possibility. Can we put in columns instead of 0's and 1's that would produce estimates (and exact standard errors) for $\mu_A(adj) - \mu_D(adj)$, etc.?

The answer is yes. First we note that the $b\bar{X}$ term will cancel, so that problem is eliminated. We can break up the treatment SS into the two contrasts (1, -1, 0) and (1, 1, -2) . This will add up to the treatment SS, as they are orthogonal, and for the same reason the two estimates are not in any way cancelled by each other (i.e., the TYPE I SS will be the same independent of the order of the model). So, instead of putting in 0's and 1's, we put in the above contrasts in the IF-THEN section. The b for the (1, -1, 0) contrast will then be the estimated difference between the adjusted means and the standard error will be exact.

S T A T I S T I C A L A N A L Y S I S S Y S T E M

(D) This is the same as B.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	871.49740304	290.49913435	18.10	0.0001	0.676261	50.7060
ERROR	26	417.20259696	16.04625373		STD DEV		Y MEAN
CORRECTED TOTAL	29	1288.70000000			4.00577754		7.90000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
DRUG	2	293.60000000	9.15	0.0010	2	68.55371060	2.14	0.1384
X	1	577.89740304	36.01	0.0001	1	577.89740304	36.01	0.0001

This is equal to 101.4+192.2, from B.

Test for regression.

(E)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

Here, this is what happens, using the convenience of the CLASSES statement, for MODEL Y= X DRUG1 DRUG2. This order is used to calculate the between drugs SS (adjusted for covariance.)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	871.49740304	290.49913435	18.10	0.0001	0.676261	50.7060
ERROR	26	417.20259696	16.04625373		STD DEV		Y MEAN
CORRECTED TOTAL	29	1288.70000000			4.00577754		7.90000000

Note that one really need not run both D+E; Type IV SS give the same result.

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X	1	802.94369243	50.04	0.0001	1	577.89740304	36.01	0.0001
DRUG	2	68.55371060	2.14	0.1384	2	68.55371060	2.14	0.1384

This is the between drugs SS, adjusted for the covariate. It is used to test an overall difference between drugs.

(G)

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

This is a special run, to test the assumption $b_1 = b_2 = b_3 = b$.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C
MODEL	5	891.14204818	178.22840964	10.76	0.0001	0.691505	51.5
ERROR	24	397.55795182	16.56491466		STD DEV		Y M
CORRECTED TOTAL	29	1288.70000000			4.07000180		7.90000

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR
DRUG	2	293.60000000	8.86	0.0013	2	8.50258450	0.26	0.7
X	1	577.89740304	34.89	0.0001	1	564.56752833	34.08	0.0
X*DRUG	2	19.64464514	0.59	0.5606	2	19.64464514	0.59	0.5

Not significant.

This is the correct SS for the test; it tests how much fitting different slopes contributes after a common slope and intercept have been fitted.

(F)

STATISTICAL ANALYSIS SYSTEM

However, another assumption has been violated. Here residuals equal to Y_{ij} - trt mean - $b(X_{ij} - \bar{X}_{i.})$ are plotted versus the predicted $\hat{Y}_{ij}$ = trt mean + $b(X_{ij} - \bar{X}_{i.})$.

(See p. 48 for verification.)

There is some indication of a cone shape.

PLOT OF RES1*PRE1 LEGEND: A = 1 OBS, B = 2 OBS, ETC.