

## Week 10: Autocorrelated errors

This week, I have done one possible analysis and provided lots of output for you to consider.

### Case study: predicting body fat

Body fat is an important health measure, but accurately measuring body fat is not easy. The best method requires weighing someone underwater. A quicker, easier method, based on physical measurements, would be desirable.

Bodyfat.txt on the class web site includes data on bodyfat and physical measurements for 252 men. A multiple regression using all variables was fit. The regression output, diagnostic tables and some plots are included on subsequent pages. Some of the numerical output has been condensed.

Questions for discussion:

1. Is this a good model?
2. How well does this model predict body fat?
3. What, if anything, concerns you?
4. What, if anything, would you do next?

## The SAS System

1

## The REG Procedure

Model: MODEL1

Dependent Variable: fat

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	13168	1012.88783	54.65	<.0001
Error	238	4411.44804	18.53550		
Corrected Total	251	17579			

Root MSE	4.30529	R-Square	0.7490
Dependent Mean	19.15079	Adj R-Sq	0.7353
Coeff Var	22.48098		

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Variance Inflation
Intercept	1	-18.18849	17.34857	-1.05	0.2955	0
age	1	0.06208	0.03235	1.92	0.0562	2.25045
weight	1	-0.08844	0.05353	-1.65	0.0998	33.50932
height	1	-0.06959	0.09601	-0.72	0.4693	1.67459
neck	1	-0.47060	0.23247	-2.02	0.0440	4.32446
chest	1	-0.02386	0.09915	-0.24	0.8100	9.46088
abdomen	1	0.95477	0.08645	11.04	<.0001	11.76707
hip	1	-0.20754	0.14591	-1.42	0.1562	14.79652
thigh	1	0.23610	0.14436	1.64	0.1033	7.77786
knee	1	0.01528	0.24198	0.06	0.9497	4.61215

ankle	1	0.17400	0.22147	0.79	0.4329	1.90796
biceps	1	0.18160	0.17113	1.06	0.2897	3.61974
forearm	1	0.45202	0.19913	2.27	0.0241	2.19249
wrist	1	-1.62064	0.53495	-3.03	0.0027	3.37751

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: fat

Output Statistics

Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITs	-----DFBETAS----- Intercept	age
(output condensed)							
30	-2.8265	-0.6703	0.0429	1.0792	-0.1418	0.0249	0.0180
31	-2.7030	-0.7546	0.3090	1.4844	-0.5046	0.1079	0.0263
32	-5.3312	-1.2775	0.0580	1.0228	-0.3168	0.1600	0.1313
33	5.6649	1.3645	0.0668	1.0187	0.3649	0.0838	-0.0503
34	-2.5185	-0.6005	0.0534	1.0970	-0.1427	0.0154	-0.0134
35	0.0297	0.007005	0.0312	1.0948	0.0013	-0.0001	0.0000
36	2.5204	0.6538	0.2001	1.2930	0.3270	-0.1603	-0.0191
37	0.0800	0.0188	0.0234	1.0862	0.0029	0.0003	-0.0007
38	6.4966	1.5538	0.0513	0.9701	0.3613	-0.1469	0.0347
39	-8.8349	-2.6280	0.3751	1.1354	-2.0362	-0.3904	-0.2800
40	0.0747	0.0178	0.0580	1.1260	0.0044	0.0003	-0.0011
41	-2.1759	-0.5693	0.2140	1.3239	-0.2970	0.0650	0.0125
42	0.3652	0.1660	0.7400	4.0735	0.2801	0.1325	-0.0015
43	-2.6349	-0.6337	0.0697	1.1134	-0.1734	0.0404	-0.0344
44	5.6247	1.3291	0.0306	0.9862	0.2361	0.0621	-0.0256
45	-3.2041	-0.7627	0.0495	1.0784	-0.1741	-0.0834	0.0255
46	3.9435	0.9329	0.0366	1.0460	0.1819	-0.0525	0.0515
47	3.0287	0.7147	0.0331	1.0644	0.1322	0.0522	0.0163
48	-3.9409	-0.9280	0.0276	1.0368	-0.1563	-0.0206	0.0553

(output condensed)

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: fat

## Output Statistics

Obs	weight	height	neck	chest	abdomen	hip	thigh
-----	--------	--------	------	-------	---------	-----	-------

(output condensed)

30	0.0325	-0.0168	-0.0124	-0.0609	0.0519	-0.0562	-0.0343
31	0.1024	-0.0529	-0.0938	-0.0372	-0.0300	-0.0532	0.0684
32	0.1740	-0.0664	-0.1092	-0.0391	0.0113	-0.1812	0.0613
33	0.1243	-0.0798	0.0241	-0.0542	-0.0627	-0.0743	0.0524
34	0.0005	0.0145	0.0142	-0.0421	0.0485	0.0121	0.0007
35	0.0001	0.0002	-0.0001	-0.0002	0.0002	-0.0001	0.0002
36	-0.1892	0.0725	0.0361	0.1533	0.0047	0.2276	-0.0643
37	0.0003	-0.0004	-0.0015	-0.0003	0.0005	0.0001	-0.0002
38	-0.1171	-0.0443	0.2067	0.0224	-0.0331	-0.0001	0.1453
39	-0.7244	0.4026	-0.4632	0.6348	0.2821	-0.3777	0.2553
40	-0.0001	-0.0011	0.0005	0.0010	0.0014	-0.0017	-0.0008
41	0.0328	-0.0008	0.0502	-0.0531	0.0064	-0.1183	-0.0226
42	0.0965	-0.2603	-0.0287	-0.0477	-0.0499	-0.0159	-0.0286
43	0.0420	-0.0442	0.0914	-0.0283	-0.0143	-0.0705	-0.0132
44	0.0722	-0.0220	0.0602	-0.1279	0.0875	-0.0734	0.0028
45	-0.0427	0.0136	0.0676	0.0229	-0.0237	0.0227	0.0302
46	-0.0231	0.0641	-0.0571	0.0878	-0.1036	0.0359	0.0742
47	0.0380	-0.0279	-0.0538	0.0114	-0.0369	-0.0344	0.0449
48	-0.0137	0.0075	0.0513	0.0281	-0.0141	-0.0035	0.0647

(output condensed)

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: fat

## Output Statistics

Obs	knee	ankle	biceps	forearm	wrist
(output condensed)					
30	0.0732	-0.0123	0.0223	0.0336	-0.0193
31	0.0531	-0.4833	-0.0466	0.0559	0.1020
32	-0.1241	0.0925	-0.0031	0.0961	-0.0691
33	-0.1512	0.0667	-0.1941	0.1002	0.1052
34	-0.0870	-0.0049	-0.0339	-0.0098	0.0457
35	-0.0002	0.0003	-0.0002	0.0001	-0.0000
36	-0.0106	0.0156	-0.0225	0.0723	-0.0800
37	-0.0003	-0.0015	0.0005	0.0006	0.0014
38	0.0872	0.0513	-0.1288	0.0323	0.0376
39	0.4034	-0.1187	-0.0871	0.8192	0.2869
40	0.0016	0.0011	-0.0003	0.0009	-0.0013
41	0.1832	-0.0325	0.0795	-0.0401	-0.1296
42	0.0364	-0.0153	-0.0201	0.0049	-0.0037
43	0.0382	-0.0393	-0.0495	-0.0124	0.0155
44	-0.0266	-0.0655	0.0056	0.0719	-0.0334
45	-0.0048	0.0061	-0.0082	0.0711	0.0146
46	-0.0121	0.0068	-0.0536	-0.0198	0.0436
47	-0.0542	0.0242	-0.0238	-0.0120	0.0268
48	-0.0562	0.0443	0.0013	0.0193	-0.0619

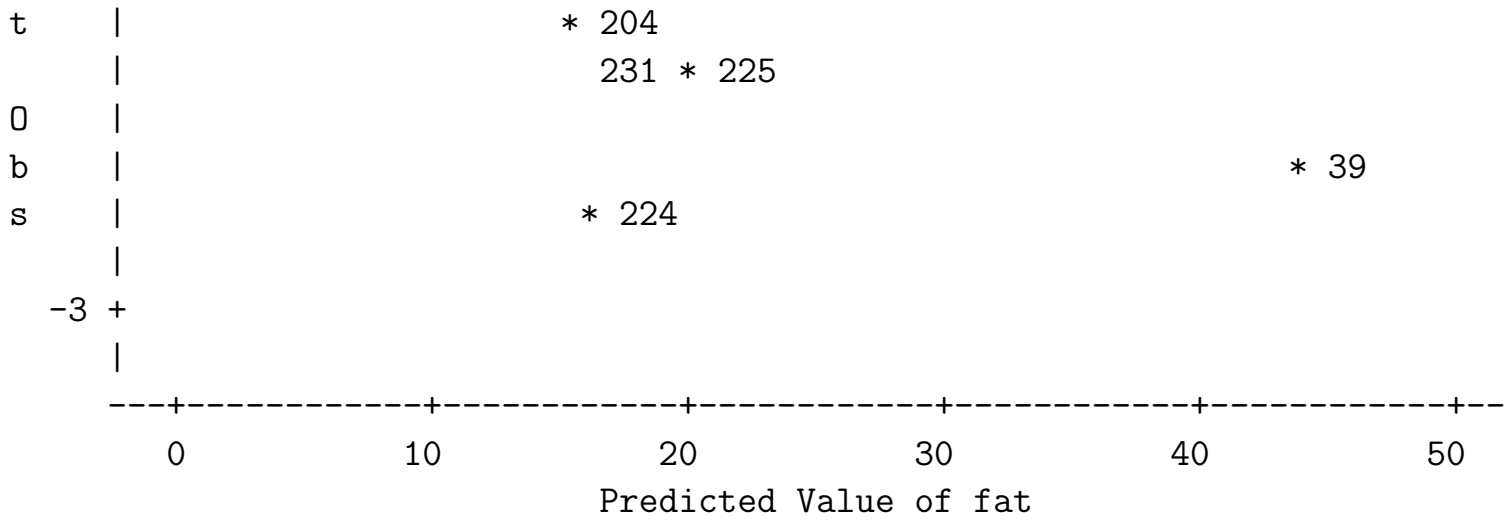
(output condensed)

Sum of Residuals	0
Sum of Squared Residuals	4411.44804
Predicted Residual SS (PRESS)	5018.17684

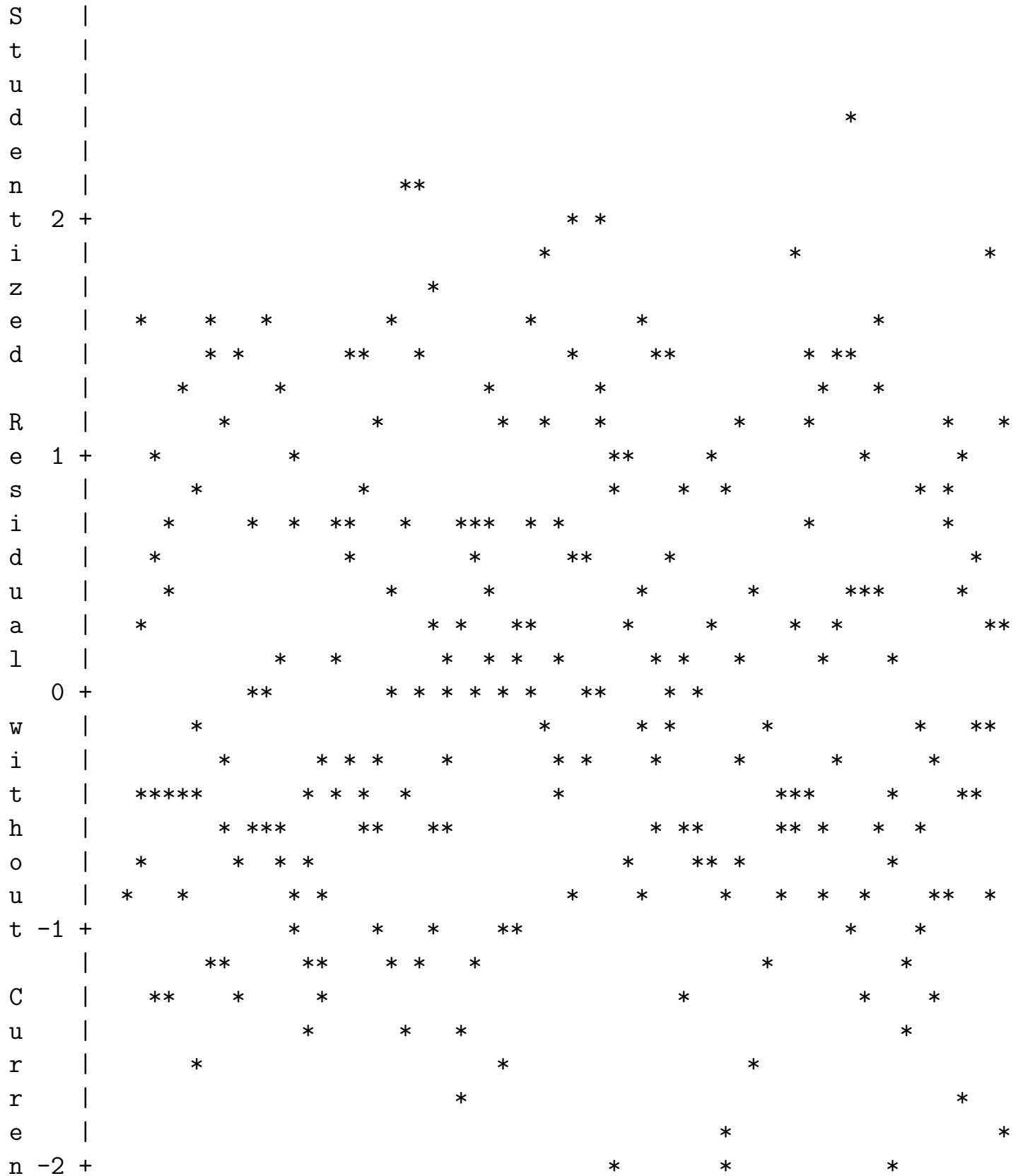
Residual vs. predicted value

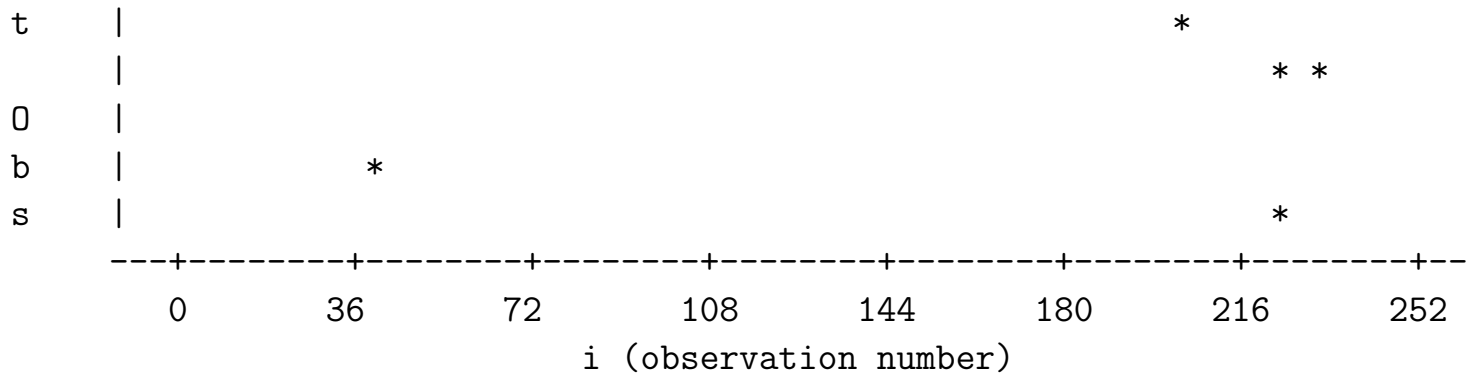
```

S      |
t      |
u      |
d      |                * 207
e      |
n      |                135 * 82 * 81
t 2 +  |            * 128      * 121
i      |                119 * *   249 *   * 192
z      |                76      3 115*386
e      |            33 24 * * 206195* * * * 148                * 216
d      | 153 * * ** * 156 ***2084 * * 62* 208
      |            25 23 2151*7 **9717127* * 44
R      |            109 ** 120** *8**134 17 * 252
e 1 +  |            * 144 * 46 1*1**6*39214138* 240
s      |                21311*10*91*0*196* 66
i      |            * 47      1037**2*77***2363 ** 59                * 36
d      |            2171420612*1* *012218157**2*365
u      |            10 * **74**241691*425* *05* 178
a      |            191**246 *17*9*6***** 88*5 ** 203
l      | 199 *14* 1*39*131** ***113* * 6*858 * 42
0 +    |            151 * 7715*****991121* 37* 1340 * 35
w      | 149 * 2717611*7*23**131 *913315016* *45* 244
i      |            218* 4*22*68**18515*79**56* 64 247 * 205
t      | 50 * 11 *69*8 ***16*219* ***16**60                * 242
h      | 29 * 2*86**7161*5**28* **1*81* 194 43 * * 41
o      |            52** **930*2*0* 156*214934190                * 169
u      |            55 * * 45893***1721**1*023* * 187
t -1 + |            48 54 ***7201* *923* 108 14 * 112
      | 182 * 26 * *1209* **982*7 * 83                * 222
C      |            9 ** *51*1875 22* 232 * 57
u      |            3211 **158*      180 * 80
r      |            53 95 223 20 ** * 107
r      |                * 97                * 238
e      |            * 172                * 250
n -2 + |            * 171                * 221 * 140
    
```



Residual vs. observation number



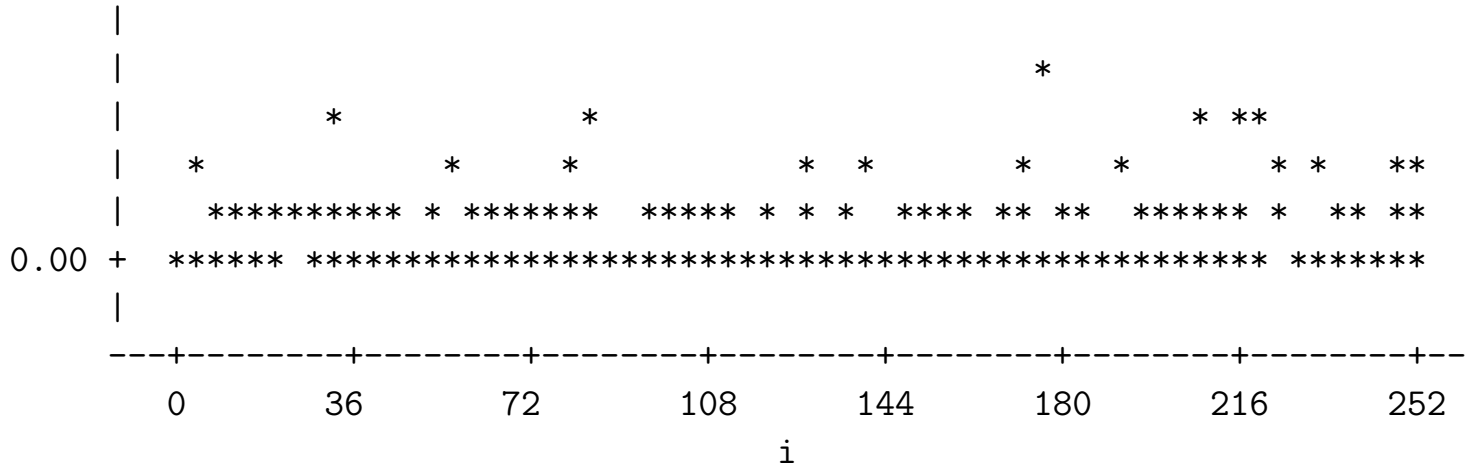


Cook's distance vs. observation number

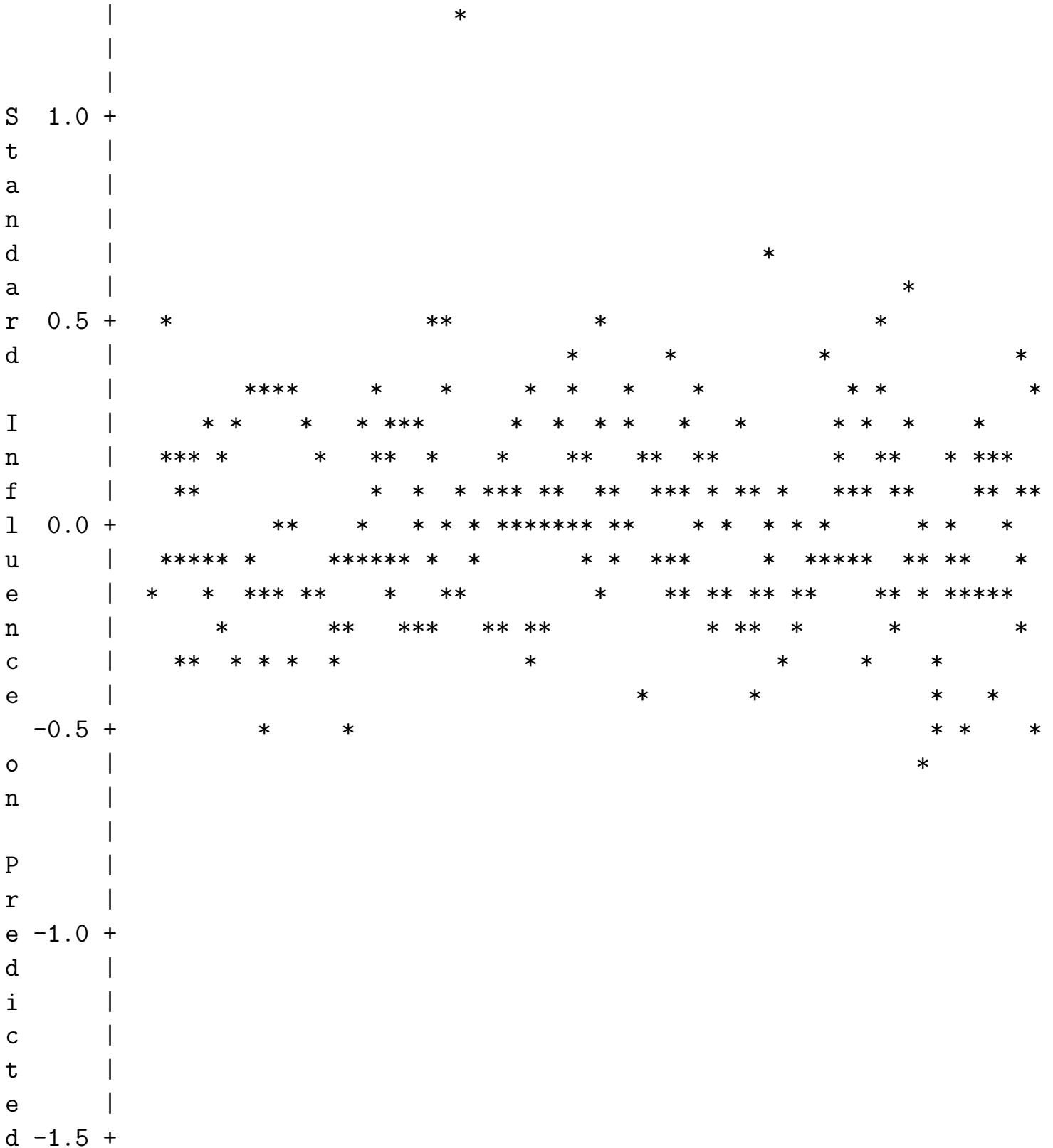
C 0.25 +  
o |  
o |  
k |  
' |  
s |  
D 0.20 +  
I |  
n |  
f |  
l |  
u |  
e 0.15 +  
n |  
c |  
e |  
S |  
t |  
a 0.10 +  
t |  
i |  
s |  
t |  
i |  
c |  
0.05 +  
|

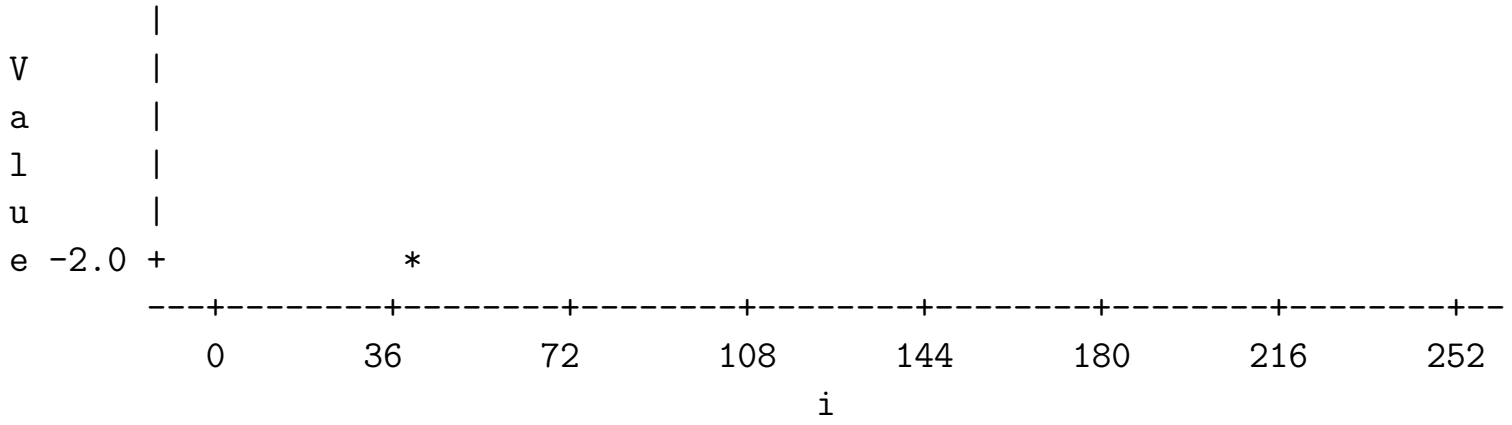
\*

\*

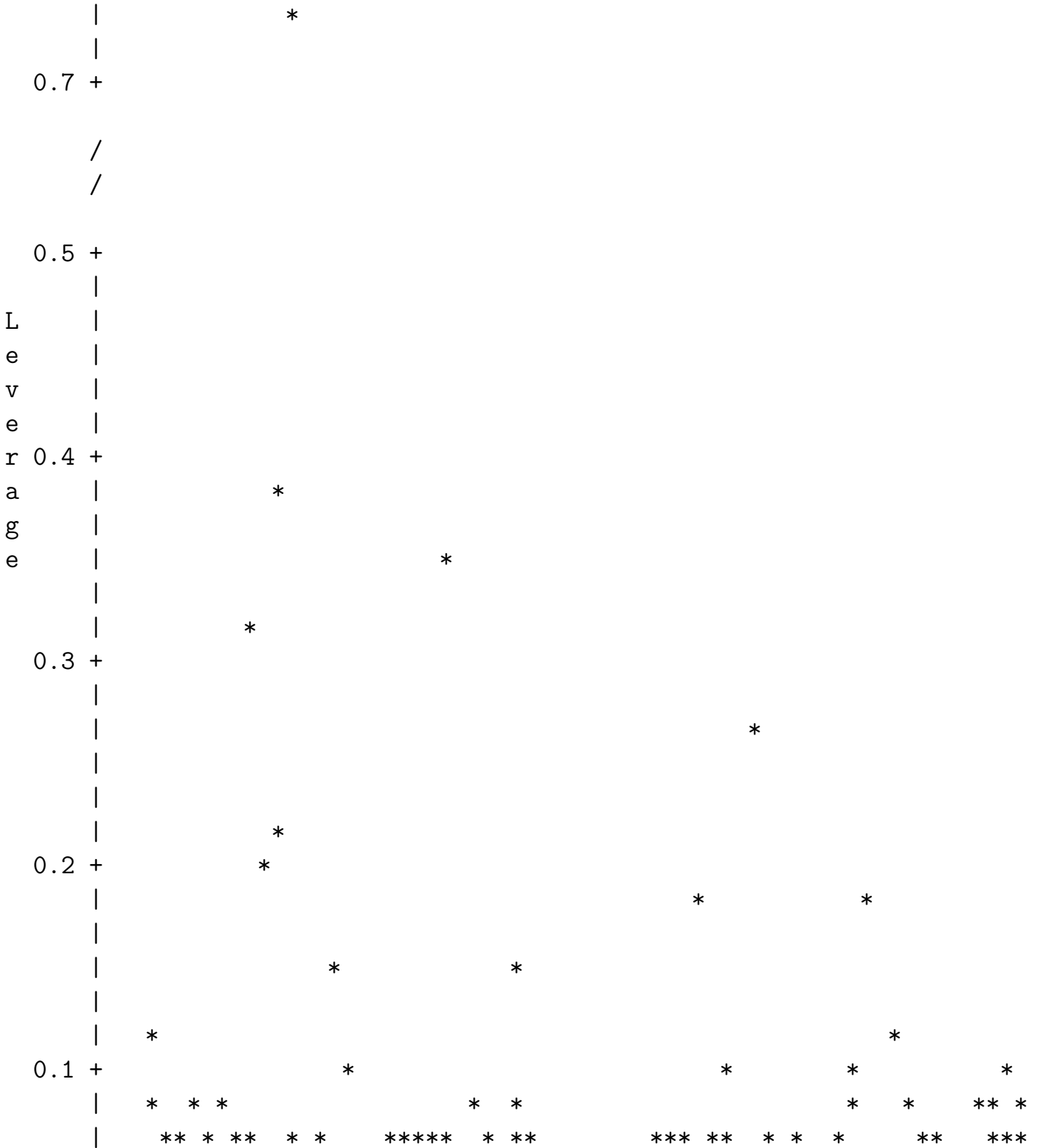


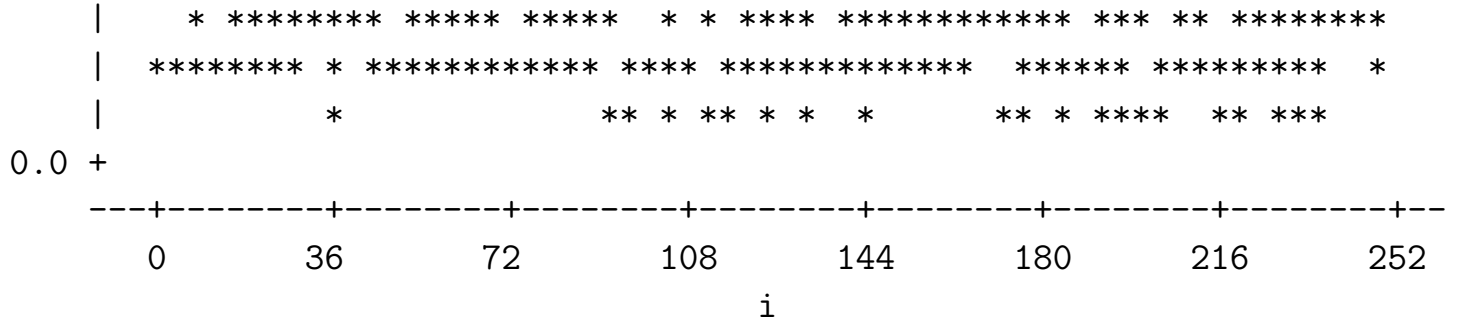
DFfit vs observation number





influence (Hii) vs. observation number





**births3.sas**

```
data births;
  infile 'birth.txt';
  input year denmark netherlands canada usa;

  y = netherlands;
  x = year - 1949;

/* calculate Durbin-Watson statistic in proc reg */
/* by adding the option dw to the model statement */

proc reg;
  model y = x /dw dwprob;
  title 'regression with Durbin-Watson statistic';

/* fit a regression model using an AR(1) error structure */
/* using proc mixed */
/* there are other procedures that do this (e.g. autoreg */
/* in the Econometrics / Time Series collection of procs) */
/* you are welcome to use autoreg if you are familiar with it */
/* we will use mixed for other models in the next few weeks */
/* so this is an introduction to it */

proc mixed;
  model y = x /solution;
  repeated /subject = intercept type = ar(1);
  title 'regression with ar(1) error ';
run;
```

## births3.lst

regression with Durbin-Watson statistic

1

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: y

Number of Observations Read	46
Number of Observations Used	45
Number of Observations with Missing Values	1

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00004961	0.00004961	32.61	<.0001
Error	43	0.00006542	0.00000152		
Corrected Total	44	0.00011502			

&lt;some output omitted&gt;

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	0.51483	0.00037395	1376.74	<.0001
x	1	-0.00008084	0.00001416	-5.71	<.0001

Durbin-Watson D	1.998
Number of Observations	45
1st Order Autocorrelation	-0.031

< dwprob option provides a p-value for the test of  $H_0: \rho_1 = 0$  >

regression with ar(1) error

4

The Mixed Procedure

Model Information

Data Set	WORK.BIRTHS
Dependent Variable	y
Covariance Structure	Autoregressive
Subject Effect	Intercept
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Prasad-Rao-Jeske- Kackar-Harville
Degrees of Freedom Method	Kenward-Roger

<some output omitted>

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	-441.25449304	
1	2	-441.26340421	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
----------	---------	----------

AR(1)	Intercept	0.01509
Residual		1.523E-6

## Fit Statistics

-2 Res Log Likelihood	-441.3
AIC (smaller is better)	-437.3
AICC (smaller is better)	-437.0
BIC (smaller is better)	-433.7

## The Mixed Procedure

## Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	0.01	0.9248

## Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	0.5148	0.000371	13.4	1388.99	<.0001
x	-0.00008	0.000014	13.6	-5.75	<.0001

## Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
x	1	13.6	33.11	<.0001