

```

name: <unnamed>
log: \fall_2011_381_logs\class11.log
log type: text
opened on: 1 Nov 2011, 13:42:24

. use "fifty_state_dataset.dta", clear

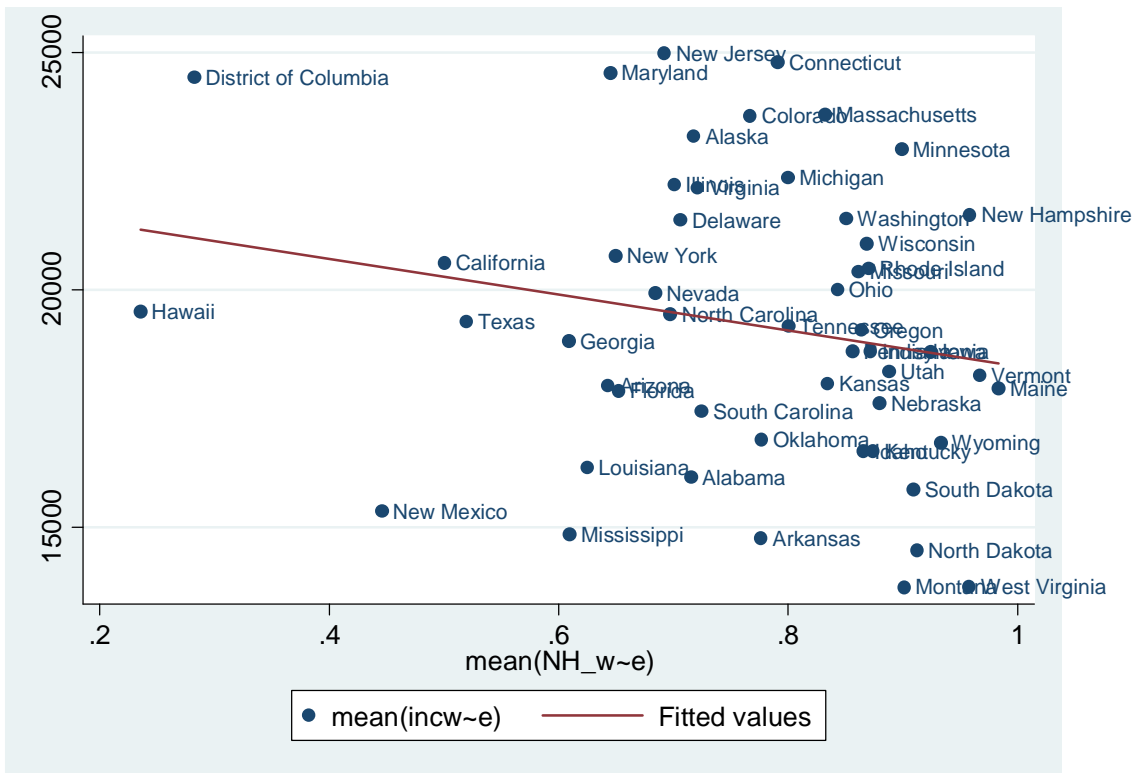
```

```

. twoway (scatter incwage NH_White_proportion, mlabel(statefip)) (lfit incwage
NH_White_proportion)

```

\*I am reproducing this log as a pdf file because you need to see the figure, embedded here, to make sense of the dfbeta and residual lists below...



\* According to our notes below, the states with the biggest residuals (in absolute value) are Connecticut, New Jersey, and New Mexico. These are simply the points furthest from the line. But the states with the biggest DFbetas, that is the biggest influence on the slope of the line, are DC, Hawaii, and New Mexico. The reason these 3 states have the biggest influence on the slope is that they are outliers in X. By being further away from the cluster of points between .7 and 1 in X, they exert more leverage on the slope.

```

. regress incwage NH_White_proportion

```

Source	SS	df	MS	Number of obs =	51
Model	18878316.5	1	18878316.5	F( 1, 49) =	2.14
Residual	432407199	49	8824636.71	Prob > F =	0.1500
Total	451285515	50	9025710.3	R-squared =	0.0418
				Adj R-squared =	0.0223
				Root MSE =	2970.6

incwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
NH_White_p~n	-3761.36	2571.649	-1.46	0.150	-8929.282 1406.562
_cons	22161.23	2004.928	11.05	0.000	18132.18 26190.29

-----

\*And note that in the regression calculation (which is the same line Stata plots as "lfit", above) the slope is not significantly different from zero here. This is because we have used only fifty data points and we have not used the weights here in any way, so the fifty points represent a very small (unweighted) sample size, and furthermore the plot is not very linear, i.e. the line does not fit very well. The 95% CI for the slope of the line includes zero.

\* How one generates the dfbetas is a little different in Stata 10, see the online help.

```
. dfbeta
      _dfbeta_1: dfbeta(NH_White_proportion)
```

```
. predict M1_predicted
(option xb assumed; fitted values)
```

```
. gen residual= incwage- M1_predicted
```

\* First generate predicted values, then use those to generate the residuals, then we will want the absolute value of residuals and the absolute value of dfbetas, so that we can list the ones that are furthest from zero (in either direction).

```
. gen abs_residual=abs(residual)
```

```
. gen abs_dfbeta=abs( _dfbeta_1)
```

```
. gsort -abs_residual
```

\* gsort allows sorting from largest to smallest, which is what the - does.

```
. list statefip abs_residual residual abs_dfbeta _dfbeta_1
```

	statefip	abs_re~1	residual	abs_df~a	_dfbeta_1
1.	Connecticut	5617.314	5617.314	.0487289	.0487289
2.	New Jersey	5432.736	5432.736	.1172426	-.1172426
3.	New Mexico	5139.932	-5139.932	.5375599	.5375599
4.	Montana	5024.921	-5024.921	.2145686	-.2145686
5.	Mississippi	5009.293	-5009.293	.2366829	.2366829
6.	Maryland	4840.275	4840.275	.1743481	-.1743481
7.	West Virginia	4798.199	-4798.199	.2921322	-.2921322
8.	Massachusetts	4666.598	4666.598	.098265	.098265
9.	Arkansas	4460.403	-4460.403	.0185422	-.0185422
10.	Colorado	4390.833	4390.833	.0056333	.0056333
11.	North Dakota	4204.203	-4204.203	.192666	-.192666
12.	Minnesota	4187.81	4187.81	.1740437	.1740437
13.	Alaska	3780.312	3780.312	.0506594	-.0506594
14.	Louisiana	3555.554	-3555.554	.1484154	.1484154
15.	Alabama	3416.563	-3416.563	.0482474	.0482474
16.	District of Columbia	3381.999	3381.999	.5903278	-.5903278
17.	Michigan	3208.876	3208.876	.0356351	.0356351
18.	New Hampshire	3021.199	3021.199	.1813799	.1813799
19.	South Dakota	2937.162	-2937.162	.1303558	-.1303558
20.	Virginia	2706.413	2706.413	.0332524	-.0332524
21.	Illinois	2687.422	2687.422	.04929	-.04929
22.	Washington	2536.313	2536.313	.0665131	.0665131
23.	Oklahoma	2383.868	-2383.868	.0102179	-.0102179
24.	Idaho	2289.952	-2289.952	.070259	-.070259
25.	Kentucky	2257.391	-2257.391	.075205	-.075205
26.	Wisconsin	2079.558	2079.558	.0656909	.0656909
27.	South Carolina	1986.75	-1986.75	.022419	.022419
28.	Delaware	1970.206	1970.206	.0329438	-.0329438
29.	Wyoming	1860.133	-1860.133	.0957266	-.0957266

30.	Florida	1833.188	-1833.188	.0603009	.0603009
31.	Arizona	1755.73	-1755.73	.062599	.062599
32.	Hawaii	1728.106	-1728.106	.3419163	.3419163
33.	Rhode Island	1562.148	1562.148	.0499289	.0499289
34.	Missouri	1460.693	1460.693	.0429018	.0429018
35.	Nebraska	1231.4	-1231.4	.0429309	-.0429309
36.	Ohio	1017.316	1017.316	.024228	.024228
37.	Kansas	1001.302	-1001.302	.021283	-.021283
38.	New York	998.9494	998.9494	.0335865	-.0335865
39.	Georgia	939.2884	-939.2884	.0433774	.0433774
40.	Texas	873.475	-873.475	.0654734	.0654734
41.	Utah	544.7642	-544.7642	.0203697	-.0203697
42.	Maine	536.482	-536.482	.0362305	-.0362305
43.	Nevada	347.9611	347.9611	.0080656	-.0080656
44.	Vermont	328.0741	-328.0741	.0203959	-.0203959
45.	California	294.7873	294.7873	.0240138	-.0240138
46.	Oregon	241.7805	241.7805	.0072873	.0072873
47.	Pennsylvania	230.294	-230.294	.0063625	-.0063625
48.	Indiana	165.3883	-165.3883	.0053708	-.0053708
49.	Tennessee	89.31062	89.31062	.0009869	.0009869
50.	North Carolina	46.82497	-46.82497	.0009033	.0009033
51.	Iowa	17.83441	17.83441	.0008664	.0008664

. gsort - abs\_dfbeta

. list statefip abs\_dfbeta \_dfbeta\_1 abs\_residual M1\_predicted

	statefip	abs_df~a	_dfbeta_1	abs_re~l	M1_pre~d
1.	District of Columbia	.5903278	-.5903278	3381.999	21099.99
2.	New Mexico	.5375599	.5375599	5139.932	20484.66
3.	Hawaii	.3419163	.3419163	1728.106	21275.74
4.	West Virginia	.2921322	-.2921322	4798.199	18558.29
5.	Mississippi	.2366829	.2366829	5009.293	19867.71
6.	Montana	.2145686	-.2145686	5024.921	18771.17
7.	North Dakota	.192666	-.192666	4204.203	18729.07
8.	New Hampshire	.1813799	.1813799	3021.199	18555.83
9.	Maryland	.1743481	-.1743481	4840.275	19735.32
10.	Minnesota	.1740437	.1740437	4187.81	18779.89
11.	Louisiana	.1484154	.1484154	3555.554	19810.51
12.	South Dakota	.1303558	-.1303558	2937.162	18740.31
13.	New Jersey	.1172426	-.1172426	5432.736	19557.67
14.	Massachusetts	.098265	.098265	4666.598	19031.36
15.	Wyoming	.0957266	-.0957266	1860.133	18651.91
16.	Kentucky	.075205	-.075205	2257.391	18873.26
17.	Idaho	.070259	-.070259	2289.952	18905.92
18.	Washington	.0665131	.0665131	2536.313	18961.83
19.	Wisconsin	.0656909	.0656909	2079.558	18894.24
20.	Texas	.0654734	.0654734	873.475	20206.79
21.	Arizona	.062599	.062599	1755.73	19741.79
22.	Florida	.0603009	.0603009	1833.188	19707.55
23.	Alaska	.0506594	-.0506594	3780.312	19460.74
24.	Rhode Island	.0499289	.0499289	1562.148	18888.72
25.	Illinois	.04929	-.04929	2687.422	19524.39
26.	Connecticut	.0487289	.0487289	5617.314	19185.95

27.	Alabama	.0482474	.0482474	3416.563	19470.32
28.	Georgia	.0433774	.0433774	939.2884	19871.7
29.	Nebraska	.0429309	-.0429309	1231.4	18852.31
30.	Missouri	.0429018	.0429018	1460.693	18920.82
-----					
31.	Maine	.0362305	-.0362305	536.482	18461.66
32.	Michigan	.0356351	.0356351	3208.876	19152.48
33.	New York	.0335865	-.0335865	998.9494	19717.74
34.	Virginia	.0332524	-.0332524	2706.413	19448.1
35.	Delaware	.0329438	-.0329438	1970.206	19504.86
-----					
36.	Ohio	.024228	.024228	1017.316	18989.97
37.	California	.0240138	-.0240138	294.7873	20279.2
38.	South Carolina	.022419	.022419	1986.75	19436.02
39.	Kansas	.021283	-.021283	1001.302	19022.22
40.	Vermont	.0203959	-.0203959	328.0741	18523.26
-----					
41.	Utah	.0203697	-.0203697	544.7642	18820.45
42.	Arkansas	.0185422	-.0185422	4460.403	19240.7
43.	Oklahoma	.0102179	-.0102179	2383.868	19238.16
44.	Nevada	.0080656	-.0080656	347.9611	19587.5
45.	Oregon	.0072873	.0072873	241.7805	18910.3
-----					
46.	Pennsylvania	.0063625	-.0063625	230.294	18941.73
47.	Colorado	.0056333	.0056333	4390.833	19276.55
48.	Indiana	.0053708	-.0053708	165.3883	18881.21
49.	Tennessee	.0009869	.0009869	89.31062	19151.45
50.	North Carolina	.0009033	.0009033	46.82497	19538.41
-----					
51.	Iowa	.0008664	.0008664	17.83441	18683.9

```

. log close
  name: <unnamed>
  log: C:\Documents and Settings\Michael Rosenfeld\My Documents\newer web
> pages\soc_meth_proj3\fall_2011_381_logs\class11.log
  log type: text
  closed on: 1 Nov 2011, 15:34:07
-----

```