

Data Set Page Size	8192
Number of Data Set Pages	50
First Data Page	1
Max Obs per Page	119
Obs in First Data Page	73
Number of Data Set Repairs	0
Filename	C:\Users\nxnguy01\Desktop\star.sas7bdat
Release Created	9.0202M0
Host Created	XP_PRO

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
5	absent	Num	3	days absent
19	aide	Num	3	regular class with aide
11	black	Num	3	black student
9	boy	Num	3	male student
14	freelunch	Num	3	free lunch provided
1	id	Num	4	student id
7	mathscore	Num	4	math score
6	readscore	Num	4	reading score
18	regular	Num	3	regular class
2	schid	Num	6	school id
16	schrural	Num	3	school rural
15	schurban	Num	3	school urban or inner city
17	small	Num	3	small class
4	tchexper	Num	3	teacher years of experience
3	tchid	Num	6	teacher id
13	tchmasters	Num	3	teacher with masters degree
12	tchwhite	Num	3	white teacher
8	totalscore	Num	4	combined math and reading score
10	white_asian	Num	3	white or asian student

```
* construct new variables;
```

```
data stardata;  
set tmp1.star;  
if aide = 0;  
run;
```

```
* sort data;
```

```
proc sort data=work.stardata;  
by small;  
run;
```

```
* obtain descriptive statistics;
```

```
*small=1 is small class without a teachers aid
```

```
*totalscore should be the only thing showing a difference
```

```
*1 if in small classroom, 0 if regular classroom
```

```
*all of the means should be about the same. We should see if the average totalscore  
higher in the small class. They look different, but are they statistically significant?
```

```
*is there a statistically significant difference in the means? In a paired t-test is the  
931.94 different than the 918.04? you would do a regression where the explanatory  
variable "small" and see if that variable is statistically significant.
```

```
*if you randomly assign to small and regular classrooms, all of the other variables  
should be about the same value.
```

```
options nolabel;
```

```
proc means data=work.stardata;
```

```
var totalscore small tchexper boy freelunch white_asian tchwhite tchmasters schurban  
schrural;
```

```
by small;
```

```
run;
```

The SAS System 17:07 Tuesday, February 5, 2019 3

----- small=0 -----

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
totalscore	2005	918.0428928	73.1379917	635.0000000	1229.00
small	2005	0	0	0	0
tchexper	2005	9.0683292	5.7244459	0	24.0000000
boy	2005	0.5132170	0.4999500	0	1.0000000
freelunch	2005	0.4738155	0.4994385	0	1.0000000
white_asian	2005	0.6812968	0.4660899	0	1.0000000
tchwhite	2005	0.7980050	0.4015887	0	1.0000000
tchmasters	2005	0.3650873	0.4815747	0	1.0000000
schurban	2005	0.3012469	0.4589142	0	1.0000000
schrural	2005	0.4997506	0.5001247	0	1.0000000

----- small=1 -----

Variable	N	Mean	Std Dev	Minimum	Maximum
totalscore	1738	931.9418872	76.3586330	747.0000000	1253.00
small	1738	1.0000000	0	1.0000000	1.0000000
tchexper	1738	8.9953970	5.7315679	0	27.0000000
boy	1738	0.5149597	0.4999200	0	1.0000000
freelunch	1738	0.4718067	0.4993482	0	1.0000000
white_asian	1738	0.6846951	0.4647709	0	1.0000000
tchwhite	1738	0.8624856	0.3444887	0	1.0000000
tchmasters	1738	0.3176064	0.4656795	0	1.0000000

schurban	1738	0.3060990	0.4610040	0	1.0000000
schrural	1738	0.4626007	0.4987428	0	1.0000000

```

* estimate model using proc reg;
*intercept=small
*then small and add in the control variable "teacher experience"
*the 13.89. see the difference 918.89....is the treatment effect statistically significant?
Yes because p-value is less than .05
*small classrooms have an impact on student achievements
*1.5551 is the value of the control variable. The treatment effect is 13.98 higher for
smaller classrooms than regular classrooms
options nolabel;
proc reg data=work.stardata;
model totalscore = small;
model totalscore = small tchexper;
run;
quit;

```

The SAS System 17:07 Tuesday, February 5, 2019 4

The REG Procedure
Model: MODEL1
Dependent Variable: totalscore

Number of Observations Read	3743
Number of Observations Used	3743

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	179850	179850	32.27	<.0001
Error	3741	20847551	5572.72158		
Corrected Total	3742	21027402			

Root MSE	74.65066	R-Square	0.0086
Dependent Mean	924.49666	Adj R-Sq	0.0083
Coeff Var	8.07474		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	918.04289	1.66716	550.66	<.0001
small	1	13.89899	2.44659	5.68	<.0001

The REG Procedure
 Model: MODEL2
 Dependent Variable: totalscore

Number of Observations Read 3743
 Number of Observations Used 3743

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	343722	171861	31.08	<.0001
Error	3740	20683680	5530.39563		
Corrected Total	3742	21027402			

Root MSE 74.36663 R-Square 0.0163
 Dependent Mean 924.49666 Adj R-Sq 0.0158
 Coeff Var 8.04401

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	907.56434	2.54241	356.97	<.0001
small	1	13.98327	2.43733	5.74	<.0001
tchexper	1	1.15551	0.21228	5.44	<.0001

```
* estimate model using proc autoreg;
*here you control for "school id"(schid)
*see the log, the warning in GREEN means that the class statement is experimental in this
release
*79 levels because there are 79 different schools
*you are controlling for the impact of different schools in different systems
*smaller classrooms with more individual attention means improved performance
options nolabel;
proc autoreg data=work.stardata;
class schid;
model totalscore = small schid / noint;
model totalscore = small tchexper schid / noint;
run;
quit;
```

The AUTOREG Procedure

Dependent Variable totalscore

Class Level Information

Class	Levels	Values
schid	79	112038 123056 128068 128076 128079 130085 159171 161176 161183 162184 164198 165199
		166203 168211 168214 169219 169229 169231 169280 170295 173312 176329 180344 189378
		189382 189396 191411 193422 193423 201449 203452 203457 205488 205489 205490 205491
		205492 208501 208503 209510 212522 215533 216536 218562 221571 221574 225585 228606
		230612 231616 234628 244697 244708 244723 244727 244728 244736 244745 244746 244755

244764 244774 244776 244780 244796 244799 244801 244806 244818 244831 244839 252885
 253888 257899 257905 259915 261927 262937 264945

Ordinary Least Squares Estimates

SSE	16028908.4	DFE	3663
MSE	4376	Root MSE	66.15056
SBC	42580.3309	AIC	42082.1195
MAE	50.4721972	AICC	42085.6585
MAPE	5.43927562	HQC	42259.3194
Durbin-Watson	1.9975	Total R-Square	0.9950

NOTE: No intercept term is used. R-squares are redefined.

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
small	1	15.9978	2.2228	7.20	<.0001
schid 112038	1	838.7586	11.5596	72.56	<.0001
schid 123056	1	894.2657	11.3870	78.53	<.0001
schid 128068	1	887.0332	11.9175	74.43	<.0001
schid 128076	1	886.1151	11.2119	79.03	<.0001
schid 128079	1	881.0535	10.7669	81.83	<.0001
schid 130085	1	907.1034	9.5412	95.07	<.0001
schid 159171	1	975.7788	7.8583	124.17	<.0001
schid 161176	1	890.2975	9.0863	97.98	<.0001
schid 161183	1	955.8947	8.1826	116.82	<.0001
schid 162184	1	914.6291	10.1639	89.99	<.0001
schid 164198	1	931.3645	11.5486	80.65	<.0001
schid 165199	1	993.7342	12.1101	82.06	<.0001
schid 166203	1	887.4906	9.7332	91.18	<.0001
schid 168211	1	910.4503	9.5351	95.48	<.0001
schid 168214	1	959.6572	11.7287	81.82	<.0001
schid 169219	1	969.7593	12.3032	78.82	<.0001
schid 169229	1	944.3419	7.0370	134.20	<.0001
schid 169231	1	852.9127	11.3870	74.90	<.0001

The AUTOREG Procedure

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
schid 169280	1	922.7282	11.5539	79.86	<.0001
schid 170295	1	936.4431	10.1639	92.13	<.0001
schid 173312	1	973.5303	11.3816	85.54	<.0001
schid 176329	1	969.3421	10.0460	96.49	<.0001
schid 180344	1	912.8812	9.4376	96.73	<.0001
schid 189378	1	885.0218	9.5476	92.70	<.0001
schid 189382	1	906.0499	10.3877	87.22	<.0001
schid 189396	1	890.7575	10.9031	81.70	<.0001
schid 191411	1	921.6306	12.7690	72.18	<.0001
schid 193422	1	940.9746	10.7718	87.36	<.0001
schid 193423	1	923.8836	9.3430	98.89	<.0001
schid 201449	1	948.4009	7.9600	119.15	<.0001
schid 203452	1	918.9716	8.0817	113.71	<.0001
schid 203457	1	953.5425	13.5286	70.48	<.0001
schid 205488	1	926.1151	11.2119	82.60	<.0001
schid 205489	1	926.4455	11.0749	83.65	<.0001
schid 205490	1	833.4333	10.9076	76.41	<.0001
schid 205491	1	922.6603	10.0400	91.90	<.0001
schid 205492	1	962.6167	9.2626	103.92	<.0001
schid 208501	1	926.5965	10.2860	90.08	<.0001
schid 208503	1	883.7035	10.9076	81.02	<.0001
schid 209510	1	889.1441	9.5351	93.25	<.0001
schid 212522	1	934.5362	10.1703	91.89	<.0001
schid 215533	1	948.3523	7.7425	122.49	<.0001
schid 216536	1	907.6243	7.5985	119.45	<.0001
schid 218562	1	944.0013	10.4126	90.66	<.0001
schid 221571	1	851.6225	7.7425	109.99	<.0001
schid 221574	1	881.4537	10.2860	85.69	<.0001
schid 225585	1	891.6791	8.8548	100.70	<.0001
schid 228606	1	934.6875	9.3488	99.98	<.0001
schid 230612	1	969.2377	10.7669	90.02	<.0001
schid 231616	1	931.2068	11.3816	81.82	<.0001
schid 234628	1	925.1189	9.3488	98.96	<.0001
schid 244697	1	841.9792	9.9577	84.56	<.0001
schid 244708	1	852.3134	8.3218	102.42	<.0001
schid 244723	1	861.6680	8.8559	97.30	<.0001
schid 244727	1	915.2713	10.9229	83.79	<.0001
schid 244728	1	858.4073	11.7402	73.12	<.0001
schid 244736	1	933.4072	11.7287	79.58	<.0001
schid 244745	1	946.0303	11.3765	83.16	<.0001
schid 244746	1	916.8010	11.2276	81.66	<.0001
schid 244755	1	933.2620	7.0822	131.78	<.0001
schid 244764	1	899.6945	18.4810	48.68	<.0001
schid 244774	1	880.7335	8.9363	98.56	<.0001
schid 244776	1	893.2247	6.9502	128.52	<.0001
schid 244780	1	970.4453	11.0589	87.75	<.0001
schid 244796	1	876.8513	10.5371	83.22	<.0001

The AUTOREG Procedure

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
schid 244799	1	895.3057	9.8402	90.98	<.0001
schid 244801	1	897.2413	9.4497	94.95	<.0001
schid 244806	1	944.7963	6.8834	137.26	<.0001
schid 244818	1	879.5093	8.7765	100.21	<.0001
schid 244831	1	889.8582	11.2335	79.21	<.0001
schid 244839	1	972.2676	8.5942	113.13	<.0001
schid 252885	1	938.5325	9.6686	97.07	<.0001
schid 253888	1	923.3054	13.8343	66.74	<.0001
schid 257899	1	898.6085	7.4968	119.87	<.0001
schid 257905	1	939.3175	7.5189	124.93	<.0001
schid 259915	1	899.9126	11.3765	79.10	<.0001
schid 261927	1	932.1973	8.4898	109.80	<.0001
schid 262937	1	982.8649	10.0522	97.78	<.0001
schid 264945	1	959.1413	9.4497	101.50	<.0001

The AUTOREG Procedure

Dependent Variable totalscore

Class Level Information

Class	Levels	Values
schid	79	112038 123056 128068 128076 128079 130085 159171 161176 161183 162184 164198 165199
		166203 168211 168214 169219 169229 169231 169280 170295 173312 176329 180344 189378
		189382 189396 191411 193422 193423 201449 203452 203457 205488 205489 205490 205491
		205492 208501 208503 209510 212522 215533 216536 218562 221571 221574 225585 228606
		230612 231616 234628 244697 244708 244723 244727 244728 244736 244745 244746 244755
		244764 244774 244776 244780 244796 244799 244801 244806 244818 244831 244839 252885
		253888 257899 257905 259915 261927 262937 264945

Ordinary Least Squares Estimates

SSE	15957533.5	DFE	3662
MSE	4358	Root MSE	66.01212
SBC	42571.8542	AIC	42067.4151
MAE	50.429299	AICC	42071.0436
MAPE	5.43416488	HQC	42246.83
Durbin-Watson	1.9965	Total R-Square	0.9950

NOTE: No intercept term is used. R-squares are redefined.

*teacher experience is almost 1 point higher for every year of teacher experience

*

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
small	1	16.0656	2.2183	7.24	<.0001
tchexper	1	0.9132	0.2256	4.05	<.0001
schid 112038	1	830.7579	11.7036	70.98	<.0001
schid 123056	1	883.6532	11.6618	75.77	<.0001
schid 128068	1	881.8790	11.9606	73.73	<.0001
schid 128076	1	881.4978	11.2465	78.38	<.0001
schid 128079	1	873.7210	10.8960	80.19	<.0001
schid 130085	1	893.8123	10.0717	88.74	<.0001
schid 159171	1	966.2360	8.1886	118.00	<.0001
schid 161176	1	881.7027	9.3126	94.68	<.0001
schid 161183	1	946.4889	8.4898	111.49	<.0001
schid 162184	1	910.9384	10.1835	89.45	<.0001
schid 164198	1	924.8069	11.6378	79.47	<.0001
schid 165199	1	987.6799	12.1770	81.11	<.0001
schid 166203	1	879.3493	9.9189	88.65	<.0001
schid 168211	1	900.6271	9.8198	91.72	<.0001
schid 168214	1	948.7281	12.0116	78.98	<.0001
schid 169219	1	958.2129	12.6046	76.02	<.0001
schid 169229	1	932.4737	7.6100	122.53	<.0001

The AUTOREG Procedure

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
schid 169231	1	847.7258	11.4353	74.13	<.0001
schid 169280	1	911.2151	11.8755	76.73	<.0001
schid 170295	1	927.9952	10.3552	89.62	<.0001
schid 173312	1	963.4570	11.6273	82.86	<.0001
schid 176329	1	961.7088	10.2009	94.28	<.0001
schid 180344	1	907.2544	9.5199	95.30	<.0001
schid 189378	1	875.0467	9.8412	88.92	<.0001
schid 189382	1	893.6550	10.8090	82.68	<.0001
schid 189396	1	886.7600	10.9250	81.17	<.0001
schid 191411	1	911.7581	12.9736	70.28	<.0001
schid 193422	1	932.0542	10.9729	84.94	<.0001
schid 193423	1	913.8726	9.6460	94.74	<.0001
schid 201449	1	942.0716	8.0958	116.37	<.0001
schid 203452	1	903.6184	8.9124	101.39	<.0001
schid 203457	1	941.5311	13.8227	68.11	<.0001
schid 205488	1	917.8449	11.3735	80.70	<.0001
schid 205489	1	916.3681	11.3288	80.89	<.0001
schid 205490	1	825.5589	11.0573	74.66	<.0001
schid 205491	1	917.8927	10.0880	90.99	<.0001
schid 205492	1	951.7243	9.6271	98.86	<.0001
schid 208501	1	917.3386	10.5163	87.23	<.0001
schid 208503	1	876.3722	11.0344	79.42	<.0001
schid 209510	1	879.4514	9.8119	89.63	<.0001
schid 212522	1	922.9223	10.5469	87.51	<.0001
schid 215533	1	942.4506	7.8627	119.86	<.0001
schid 216536	1	901.0249	7.7560	116.17	<.0001
schid 218562	1	937.2795	10.5227	89.07	<.0001
schid 221571	1	842.4629	8.0509	104.64	<.0001
schid 221574	1	878.2839	10.2943	85.32	<.0001
schid 225585	1	879.3676	9.3452	94.10	<.0001
schid 228606	1	922.7593	9.7837	94.32	<.0001
schid 230612	1	964.4526	10.8092	89.23	<.0001
schid 231616	1	917.0509	11.8842	77.17	<.0001
schid 234628	1	916.9868	9.5431	96.09	<.0001
schid 244697	1	835.8692	10.0509	83.16	<.0001
schid 244708	1	847.6331	8.3845	101.09	<.0001
schid 244723	1	857.7515	8.8902	96.48	<.0001
schid 244727	1	904.4543	11.2230	80.59	<.0001
schid 244728	1	850.6703	11.8706	71.66	<.0001
schid 244736	1	931.1536	11.7174	79.47	<.0001
schid 244745	1	940.5519	11.4331	82.27	<.0001
schid 244746	1	912.2039	11.2615	81.00	<.0001
schid 244755	1	923.7259	7.4498	123.99	<.0001
schid 244764	1	887.7549	18.6768	47.53	<.0001
schid 244774	1	876.3231	8.9839	97.54	<.0001
schid 244776	1	887.6565	7.0708	125.54	<.0001
schid 244780	1	964.1279	11.1456	86.50	<.0001

The AUTOREG Procedure

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
schid 244796	1	864.1871	10.9708	78.77	<.0001
schid 244799	1	888.0991	9.9798	88.99	<.0001
schid 244801	1	890.3515	9.5824	92.92	<.0001
schid 244806	1	939.1424	7.0097	133.98	<.0001
schid 244818	1	870.7302	9.0228	96.50	<.0001
schid 244831	1	876.0488	11.7178	74.76	<.0001
schid 244839	1	965.0391	8.7602	110.16	<.0001
schid 252885	1	929.8285	9.8852	94.06	<.0001
schid 253888	1	921.0892	13.8162	66.67	<.0001
schid 257899	1	887.8421	7.9400	111.82	<.0001
schid 257905	1	931.2394	7.7641	119.94	<.0001
schid 259915	1	893.2256	11.4723	77.86	<.0001
schid 261927	1	923.3318	8.7506	105.52	<.0001
schid 262937	1	970.2697	10.5029	92.38	<.0001
schid 264945	1	954.4432	9.5011	100.46	<.0001

```

* sort data;

proc sort data=work.stardata;
by schid;
run;

* estimate model using proc glm;
*use "absorb" to control for the school level fixed effects
*it suppresses the 79 fixed effects, it ran those, it just didn't print them out
options nolabel;
proc glm data=work.stardata;
absorb schid;
model totalscore = small;
run;
quit;

```

The SAS System 17:07 Tuesday, February 5, 2019 12

The GLM Procedure

Number of Observations Read	3743
Number of Observations Used	3743

The SAS System 17:07 Tuesday, February 5, 2019 13

The GLM Procedure

Dependent Variable: totalscore

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	79	4998493.34	63272.07	14.46	<.0001
Error	3663	16028908.37	4375.90		
Corrected Total	3742	21027401.71			

R-Square	Coeff Var	Root MSE	totalscore Mean
0.237713	7.155305	66.15056	924.4967

Source	DF	Type I SS	Mean Square	F Value	Pr > F
schid	78	4771837.250	61177.401	13.98	<.0001
small	1	226656.090	226656.090	51.80	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
small	1	226656.0900	226656.0900	51.80	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
small	15.99777661	2.22284636	7.20	<.0001

```

* estimate model using proc glm;
*this include teacher experience control variable
*16 points higher for small classroom
*fixed effect for the 79 schools is not printed, but it did run
options nolabel;
proc glm data=work.stardata;
absorb schid;
model totalscore = small tchexper;
run;
quit;

```

The SAS System 17:07 Tuesday, February 5, 2019 14

The GLM Procedure

Number of Observations Read	3743
Number of Observations Used	3743

The SAS System 17:07 Tuesday, February 5, 2019 15

The GLM Procedure

Dependent Variable: totalscore

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	80	5069868.17	63373.35	14.54	<.0001
Error	3662	15957533.54	4357.60		
Corrected Total	3742	21027401.71			

R-Square	Coeff Var	Root MSE	totalscore Mean
0.241108	7.140331	66.01212	924.4967

Source	DF	Type I SS	Mean Square	F Value	Pr > F
schid	78	4771837.250	61177.401	14.04	<.0001
small	1	226656.090	226656.090	52.01	<.0001
tchexper	1	71374.825	71374.825	16.38	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
small	1	228569.3106	228569.3106	52.45	<.0001
tchexper	1	71374.8251	71374.8251	16.38	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
small	16.06561257	2.21825793	7.24	<.0001
tchexper	0.91321395	0.22564397	4.05	<.0001

```

* estimate probit using proc reg;
*
options nolabel;
proc reg data=work.stardata;
model small = boy white_asian tchexper freelunch;
run;
quit;

```

The SAS System 17:07 Tuesday, February 5, 2019 16

The REG Procedure
Model: MODEL1
Dependent Variable: small

Number of Observations Read	3743
Number of Observations Used	3743

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	0.05880	0.01470	0.06	0.9936
Error	3738	930.92972	0.24904		
Corrected Total	3742	930.98851			

Root MSE	0.49904	R-Square	0.0001
Dependent Mean	0.46433	Adj R-Sq	-0.0010
Coeff Var	107.47535		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.46646	0.02516	18.54	<.0001
boy	1	0.00141	0.01634	0.09	0.9312
white_asian	1	0.00441	0.01960	0.22	0.8221
tchexper	1	-0.00060255	0.00144	-0.42	0.6754
freelunch	1	-0.00088588	0.01819	-0.05	0.9612