

QUESTION 0. Load the following dataset and run the necessary diagnostic routines in SAS to become comfortable with your data, which comprise high school student results for final exams. After looking over your data, what is one limitation of your data? [10 Points]*/

```
proc print data=grades;
run;
```

Obs	id	gender	race	ses	sctype	read	write	math	science	socst	sex
1	70	0	4	1	pub	57	52	41	47	57	Male
2	121	1	4	2	pub	68	59	53	63	61	Female
3	88	0	4	3	pub	44	33	54	58	31	Male
4	141	0	4	3	pub	63	44	47	53	56	Male
5	172	0	4	2	pub	47	52	57	53	61	Male
6	113	0	4	2	pub	44	52	51	63	61	Male
7	50	0	3	2	pub	50	59	42	53	61	Male
8	11	0	1	2	pub	34	46	45	39	36	Male
9	84	0	4	2	pub	63	57	54	.	51	Male
10	48	0	3	2	pub	57	55	52	50	51	Male
11	75	0	4	2	pub	60	46	51	53	61	Male
12	60	0	4	2	pub	57	65	51	63	61	Male
13	95	0	4	3	pub	73	60	71	61	71	Male
14	104	0	4	3	pub	54	63	57	55	46	Male
15	38	0	3	1	pub	45	57	50	31	56	Male
16	115	0	4	1	pub	42	49	43	50	56	Male
17	76	0	4	3	pub	47	52	51	50	56	Male
18	195	0	4	2	pri	57	57	60	.	56	Male
19	114	0	4	3	pub	68	65	62	55	61	Male
20	85	0	4	2	pub	55	39	57	53	46	Male

```
proc means mean std data=newgrades;
class sex;
var read write math science socst;
run;
```

The MEANS Procedure

sex	N Obs	Variable	Mean	Std Dev
Female	109	read	51.7339450	10.0578348
		write	54.9908257	8.1337152
		math	52.3944954	9.1510153
		science	50.6972477	9.0385026
		socst	52.9174312	10.2344086
Male	91	read	52.8241758	10.5087105
		write	50.1208791	10.3051607
		math	52.9450549	9.6647845
		science	52.8837209	10.7543902
		socst	51.7912088	11.3338397

Mean grads are very close among male and females.

The MEANS Procedure

sctype	race	N Obs	Variable	Mean	Std Dev
pri	1	2	read	66.5000000	9.1923882
			write	57.5000000	4.9497475
			math	53.5000000	4.9497475
			science	52.5000000	3.5355339
			socst	58.5000000	10.6066017
	2	1	read	55.0000000	.
			write	59.0000000	.
			math	52.0000000	.
			science	42.0000000	.
			socst	56.0000000	.
	3	2	read	49.5000000	3.5355339
			write	51.5000000	3.5355339
			math	47.5000000	2.1213203
			science	47.0000000	4.2426407
			socst	58.5000000	3.5355339
	4	27	read	53.6666667	9.1357287
			write	55.5555556	7.6174563
			math	55.4814815	9.3782497
			science	53.6400000	8.2959830
			socst	54.1851852	10.0996460
pub	1	22	read	44.8636364	8.3625893
			write	45.4545455	7.8177288
			math	46.8636364	6.9576642
			science	44.7142857	8.4566120
			socst	46.8181818	8.7430863
	2	10	read	51.6000000	8.0027773
			write	57.9000000	8.3193216
			math	57.8000000	10.5071404
			science	52.4000000	9.4422220
			socst	50.5000000	10.1242284
	3	18	read	46.5000000	7.4142153
			write	47.8333333	9.7452854
			math	46.6666667	6.8341702
			science	42.3333333	9.8159534
			socst	48.4444444	10.9610123
	4	118	read	53.9830508	10.5554968
			write	53.7118644	9.4878255
			math	53.6271186	9.3897531
			science	54.0258621	9.2599758
			socst	53.5677966	11.0077508

Race group 1 has highest grades in private school but in public school it is very low.

Limitation: Sample size is not large enough. We do not know which school area or how many schools have been considering for this study. The scope of sample is small.

Question 1. A fellow data scientist wants to do analysis only on public school students. Create a separate dataset for her that is called 'only public' [10 points]*/

```
proc sql;  
    create table only_public as  
    select *  
    from grades  
    where schtype='pub';  
quit;  
run;
```

```
proc print data=only_public;  
run;
```

Obs	id	gender	race	ses	schtype	read	write	math	science	socst	sex
1	70	0	4	1	pub	57	52	41	47	57	Male
2	121	1	4	2	pub	68	59	53	63	61	Female
3	86	0	4	3	pub	44	33	54	58	31	Male
4	141	0	4	3	pub	63	44	47	53	56	Male
5	172	0	4	2	pub	47	52	57	53	61	Male
6	113	0	4	2	pub	44	52	51	63	61	Male
7	50	0	3	2	pub	50	59	42	53	61	Male
8	11	0	1	2	pub	34	46	45	39	36	Male
9	84	0	4	2	pub	63	57	54	.	51	Male
10	48	0	3	2	pub	57	55	52	50	51	Male
11	75	0	4	2	pub	60	46	51	53	61	Male
12	60	0	4	2	pub	57	65	51	63	61	Male
13	95	0	4	3	pub	73	60	71	61	71	Male
14	104	0	4	3	pub	54	63	57	55	46	Male
15	38	0	3	1	pub	45	57	50	31	56	Male
16	115	0	4	1	pub	42	49	43	50	56	Male
17	76	0	4	3	pub	47	52	51	50	56	Male
18	114	0	4	3	pub	68	65	62	55	61	Male
19	85	0	4	2	pub	55	39	57	53	46	Male
20	167	0	4	2	pub	63	49	35	66	41	Male

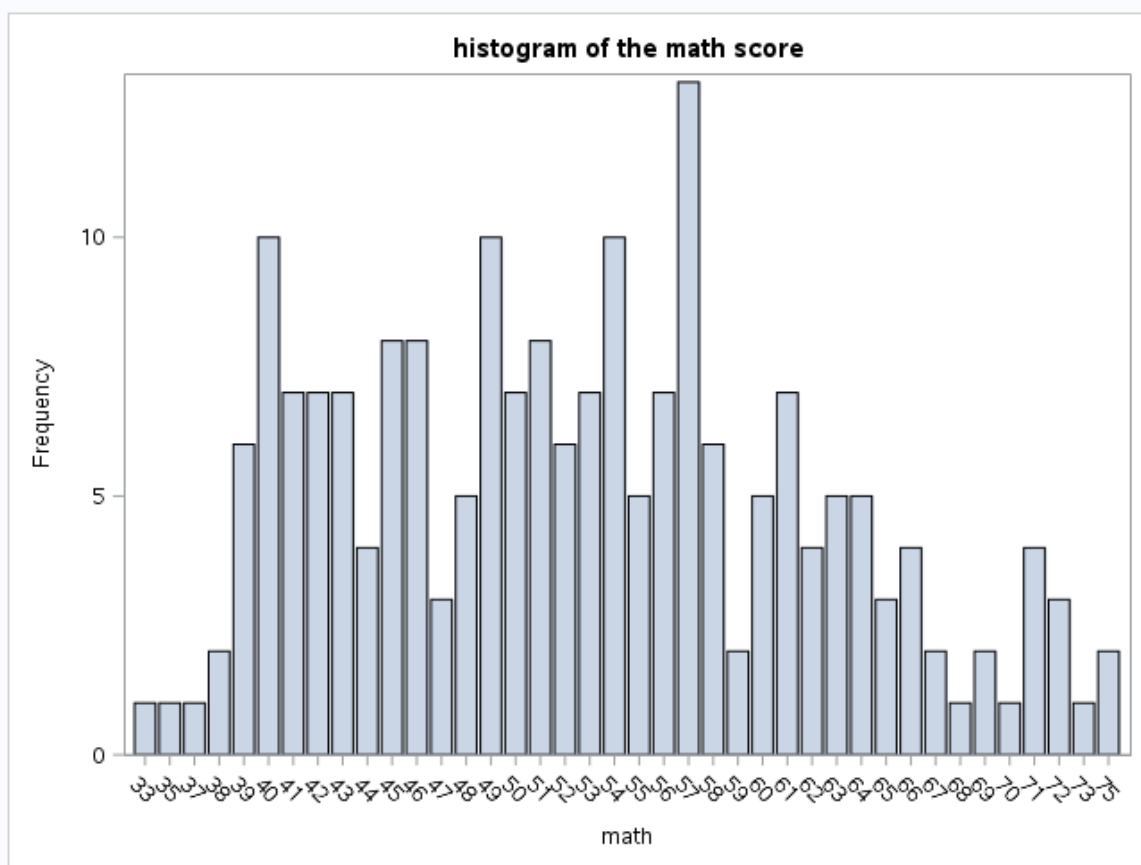
QUESTION 2. Produce a histogram of the math score [5 points]. After looking over the exam results, you decide to grade on a curve. Everybody with a math test score greater than 1 standard deviation above the average will receive an A. Between the mean and 1 standard deviation will receive a B. Between 1 standard deviation and the average a C. Between 2 standard deviations and 1 standard deviation below the average will receive a D. And below 2 standard deviations below average will receive an F. All lower bounds may be interpreted as greater than or equal to the lower bound. You may round all numbers to the hundreths place. Using PROC SGPLOT, produce a frequency table of the post-curve grade distribution. [10 Points] What percent of students will receive a B after the curve has been imposed? [10 Points] [25 Points Total]*/

```
PROC SGPLOT DATA =grades;
```

```
  VBAR math ;
```

```
  TITLE 'histogram of the math score ';
```

```
RUN;
```



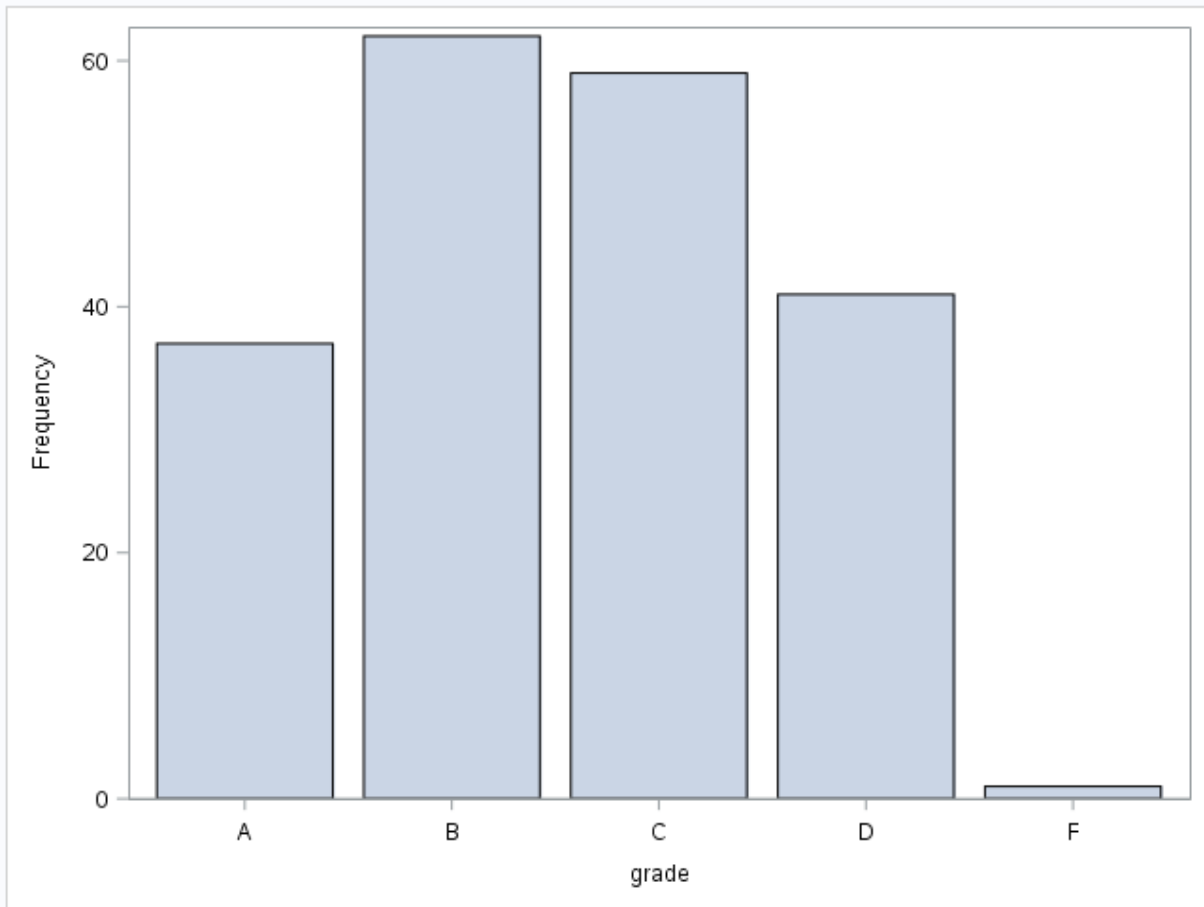
```
proc means mean std data=grades;  
var math;  
run;
```

The MEANS Procedure

Analysis Variable : math	
Mean	Std Dev
52.6450000	9.3684478

```
data freq;  
  set grades;  
    if math >=(53+9) then grade = "A";  
    if 62>math >=53 then grade = "B";  
    if 53>math >=44 then grade = "C";  
    if 44>math >=35 then grade = "D";  
    if math <35 then grade = "F";  
  
run;
```

```
proc SGPLOT data=frec;  
vbar grade;  
run;
```



```
proc print data=frec;
run;
```

Obs	id	gender	race	ses	sctype	read	write	math	science	socst	sex	grade
1	70	0	4	1	pub	57	52	41	47	57	Male	D
2	121	1	4	2	pub	68	59	53	63	61	Female	B
3	88	0	4	3	pub	44	33	54	58	31	Male	B
4	141	0	4	3	pub	63	44	47	53	56	Male	C
5	172	0	4	2	pub	47	52	57	53	61	Male	B
6	113	0	4	2	pub	44	52	51	63	61	Male	C
7	50	0	3	2	pub	50	59	42	53	61	Male	D
8	11	0	1	2	pub	34	46	45	39	36	Male	C
9	84	0	4	2	pub	63	57	54	.	51	Male	B
10	48	0	3	2	pub	57	55	52	50	51	Male	C
11	75	0	4	2	pub	60	46	51	53	61	Male	C
12	60	0	4	2	pub	57	65	51	63	61	Male	C
13	95	0	4	3	pub	73	60	71	61	71	Male	A
14	104	0	4	3	pub	54	63	57	55	46	Male	B
15	38	0	3	1	pub	45	57	50	31	56	Male	C
16	115	0	4	1	pub	42	49	43	50	56	Male	D
17	76	0	4	3	pub	47	52	51	50	56	Male	C
18	195	0	4	2	pri	57	57	60	.	56	Male	B
19	114	0	4	3	pub	68	65	62	55	61	Male	A
20	85	0	4	2	pub	55	39	57	53	46	Male	B
21	167	0	4	2	pub	63	49	35	66	41	Male	D
22	143	0	4	2	pub	63	63	75	72	66	Male	A
23	41	0	3	2	pub	50	40	45	55	56	Male	C
24	20	0	1	3	pub	60	52	57	61	61	Male	B
25	12	0	1	2	pub	37	44	45	39	46	Male	C
26	53	0	3	2	pub	34	37	46	39	31	Male	C
27	154	0	4	3	pub	65	65	66	61	66	Male	A
28	178	0	4	2	pri	47	57	57	58	46	Male	B
29	196	0	4	3	pri	44	38	49	39	46	Male	C
30	29	0	2	1	pub	52	44	49	55	41	Male	C
31	126	0	4	2	pub	42	31	57	47	51	Male	B

What percent of students will receive a B after the curve has been imposed?

```
proc freq data=frec;
  table grade ;
run;
```

31.0 percent of students will receive a B after the curve has been imposed

The FREQ Procedure

grade	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	37	18.50	37	18.50
B	62	31.00	99	49.50
C	59	29.50	158	79.00
D	41	20.50	199	99.50
F	1	0.50	200	100.00

QUESTION 3. The race variable is coded from 1-4. A value of 1 corresponds to the student being black. A value of 2 equates to Asian. A value of 3 corresponds to Hispanic. 4 is White. Use two different SAS commands to produce the same summary table depicting the mean and standard deviation of the reading test score by race and gender. You do not need to show subtotals. Each SAS command should show the same results. Make certain that the summary table is clearly labeled (e.g. 4 is listed as 'White', and not '4'). [20 Points Total]*/

```
data newgrades;
    set grades;
        if race=1 then newrace = "black";
        if race=2 then newrace = "Asian";
        if race=3 then newrace = "Hispanic";
        if race=4 then newrace = "White";

run;

proc print data=newgrades;
run;
```


Obs	id	gender	race	ses	schtype	read	write	math	science	socst	sex	newrace
1	70	0	4	1	pub	57	52	41	47	57	Male	White
2	121	1	4	2	pub	68	59	53	63	61	Female	White
3	86	0	4	3	pub	44	33	54	58	31	Male	White
4	141	0	4	3	pub	63	44	47	53	56	Male	White
5	172	0	4	2	pub	47	52	57	53	61	Male	White
6	113	0	4	2	pub	44	52	51	63	61	Male	White
7	50	0	3	2	pub	50	59	42	53	61	Male	Hispa
8	11	0	1	2	pub	34	46	45	39	36	Male	black
9	84	0	4	2	pub	63	57	54	.	51	Male	White
10	48	0	3	2	pub	57	55	52	50	51	Male	Hispa
11	75	0	4	2	pub	60	46	51	53	61	Male	White
12	60	0	4	2	pub	57	65	51	63	61	Male	White
13	95	0	4	3	pub	73	60	71	61	71	Male	White
14	104	0	4	3	pub	54	63	57	55	46	Male	White
15	38	0	3	1	pub	45	57	50	31	56	Male	Hispa
16	115	0	4	1	pub	42	49	43	50	56	Male	White
17	76	0	4	3	pub	47	52	51	50	56	Male	White
18	195	0	4	2	pri	57	57	60	.	56	Male	White
19	114	0	4	3	pub	68	65	62	55	61	Male	White
20	85	0	4	2	pub	55	39	57	53	46	Male	White

```

proc means mean std data=newgrades;
class newrace sex;
var read;
run;

```

The MEANS Procedure

Analysis Variable : read				
newrace	sex	N Obs	Mean	Std Dev
Asian	Female	8	51.7500000	9.1456469
	Male	3	52.3333333	0.5773503
Hispa	Female	13	46.7692308	7.2933057
	Male	7	46.8571429	7.3581830
White	Female	77	53.4025974	9.7969648
	Male	68	54.5147059	10.8373217
black	Female	11	45.9090909	12.2429943
	Male	13	47.3076923	8.6639542

```
proc sql;
select newrace,sex, mean(read) as mean, std(read) as std

from newgrades
group by sex, newrace;
quit;
run;
```

newrace	sex	mean	std
Asian	Female	51.75	9.145647
Hispa	Female	46.76923	7.293306
White	Female	53.4026	9.796965
black	Female	45.90909	12.24299
Asian	Male	52.33333	0.57735
Hispa	Male	46.85714	7.358183
White	Male	54.51471	10.83732
black	Male	47.30769	8.663954

QUESTION 4. You noted that some students didn't take all of their exams. A colleague tells you that this isn't a problem and that she can impute how the students would do on their missing exams based on the performance of their other exams. Does your data support this hypothesis? Why or why not? [10 Points]*/

```
proc corr data=grades_m2;
var read write math science socst mean_G;
run;
```

The CORR Procedure

6 Variables: read write math science socst mean_G

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
read	200	52.23000	10.25294	10446	28.00000	76.00000
write	200	52.77500	9.47859	10555	31.00000	67.00000
math	200	52.64500	9.36845	10529	33.00000	75.00000
science	195	51.66154	9.86803	10074	26.00000	74.00000
socst	200	52.40500	10.73579	10481	26.00000	71.00000
mean_G	200	52.37325	8.17543	10475	33.00000	68.50000

Pearson Correlation Coefficients						
Prob > r under H0: Rho=0						
Number of Observations						
	read	write	math	science	socst	mean_G
read	1.00000 200	0.59678 <.0001 200	0.66228 <.0001 200	0.61706 <.0001 195	0.62148 <.0001 200	0.85671 <.0001 200
write	0.59678 <.0001 200	1.00000 200	0.61745 <.0001 200	0.56713 <.0001 195	0.60479 <.0001 200	0.82068 <.0001 200
math	0.66228 <.0001 200	0.61745 <.0001 200	1.00000 200	0.61683 <.0001 195	0.54448 <.0001 200	0.83456 <.0001 200
science	0.61706 <.0001 195	0.56713 <.0001 195	0.61683 <.0001 195	1.00000 195	0.45292 <.0001 195	0.79302 <.0001 195
socst	0.62148 <.0001 200	0.60479 <.0001 200	0.54448 <.0001 200	0.45292 <.0001 195	1.00000 200	0.79725 <.0001 200
mean_G	0.85671 <.0001 200	0.82068 <.0001 200	0.83456 <.0001 200	0.79302 <.0001 195	0.79725 <.0001 200	1.00000 200

No

If we check the mean values for all subject, we can see that almost all the mean values very close to each other's. But if we look at the correlation table that is not true. All the subject grades have strong correlation with mean grades. That's why we can see that all the subjects have very closer means. But if we check the correlation in between each other some of them have weak correlation.

On other hand for exam grades there are other reason to consider. May be missed exam because they do not know anything about that subject. Those students may use the extra time to study for subject that they know that they can pass by missing some exams intentionally.

QUESTION 5. A colleague is interested in whether or not there is a difference in average test score between students who attend public or private schools. Produce a report that compares the mean student test score across all five of the different tests by school type. [5 Points] You note to your colleague that school type may not have a causal relationship with student grades, but instead may just be a proxy for other factors determining a student's performance, like socio-economic status. With that, you notice that socio-economic status is a variable on the data (SES) but that you do not know the coding definitions for this variable. What assumptions can you make on socio-economic status when exploring the impact of school type on grades? How did you arrive at this conclusion? After controlling for socioeconomic status, does school type explain as much of a student's average test grade? Why or why not? [20 points] [25 Points Total]*/

```
proc means data=grades;
class schtype ;
var read write math science socst;
run;
```

The MEANS Procedure

schtype	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
pri	32	read	32	54.2500000	9.1967736	38.0000000	73.0000000
		write	32	55.5312500	7.1796503	38.0000000	67.0000000
		math	32	54.7500000	8.8827488	41.0000000	75.0000000
		science	30	52.7333333	8.0598337	36.0000000	66.0000000
		socst	32	54.7812500	9.5773064	31.0000000	71.0000000
pub	168	read	168	51.8452381	10.4227918	28.0000000	76.0000000
		write	168	52.2500000	9.7855754	31.0000000	67.0000000
		math	168	52.2440476	9.4301701	33.0000000	75.0000000
		science	165	51.4666667	10.1690188	26.0000000	74.0000000
		socst	168	51.9523810	10.9099798	26.0000000	71.0000000

```
proc means data=grades;
class schtype ses ;
var read write math science socst;
run;
```

The MEANS Procedure

schtype	ses	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
pri	1	2	read	2	53.5000000	9.1923882	47.0000000	60.0000000
			write	2	54.0000000	0	54.0000000	54.0000000
			math	2	48.0000000	2.8284271	46.0000000	50.0000000
			science	2	50.0000000	0	50.0000000	50.0000000
			socst	2	53.5000000	3.5355339	51.0000000	56.0000000
	2	19	read	19	53.5789474	7.6761955	44.0000000	68.0000000
			write	19	54.8421053	6.1847766	41.0000000	65.0000000
			math	19	55.6315789	9.0012995	41.0000000	75.0000000
			science	17	52.3529412	7.5493475	39.0000000	66.0000000
			socst	19	53.4210526	7.8691789	41.0000000	71.0000000
	3	11	read	11	55.5454545	12.0363087	36.0000000	73.0000000
			write	11	57.0000000	9.3380940	38.0000000	67.0000000
			math	11	54.4545455	9.3419873	42.0000000	69.0000000
			science	11	53.8181818	9.7037949	36.0000000	66.0000000
			socst	11	57.3636364	12.6670654	31.0000000	71.0000000
pub	1	45	read	45	48.0444444	9.3831463	28.0000000	68.0000000
			write	45	50.4666667	9.6756489	31.0000000	65.0000000
			math	45	49.2222222	9.0626328	39.0000000	72.0000000
			science	44	47.6590909	10.7507559	29.0000000	69.0000000
			socst	45	47.0444444	10.9854357	26.0000000	71.0000000
	2	76	read	76	51.0789474	9.7949145	31.0000000	73.0000000
			write	76	51.1973684	9.5944703	31.0000000	67.0000000
			math	76	51.3552632	9.3033384	33.0000000	75.0000000
			science	74	50.9594595	9.0257458	34.0000000	74.0000000
			socst	76	51.6842105	10.4488730	26.0000000	71.0000000
	3	47	read	47	56.7234043	10.6927487	34.0000000	76.0000000
			write	47	55.6595745	9.5491329	33.0000000	67.0000000
			math	47	56.5744681	8.5890342	38.0000000	71.0000000
			science	47	55.8297872	9.8804135	26.0000000	69.0000000
			socst	47	57.0851064	9.3850709	31.0000000	71.0000000

For private school and public school with socioeconomic status 3 have higher grades than other two. So ses =3 should be the best factor. And then ses =2 less better than ses= 3 but it is better than ses=1. Least good one is ses=1

And Ses =1 should represent the lower income families. Because ses=1 most of the students go to public schools. Ses =2 should be represented middle class families. Ses=3 rich families.

Yes

School type explain as much of a student's average test grade when we consider ses factor with the school type. when Ses=1,2 private school have higher grades than public schools. But when ses=3 public school have higher grades.

```

proc sql;
  create table grades_m as
  select *,mean(read, write, math, science, socst) as mean_G
  from grades;

quit;
run;

data grades_m2;
  set grades_m;
  if schtype='pub' then schcode =0 ;
  if schtype='pri' then schcode =1 ;

run;

```

Obs	id	gender	race	ses	schtype	read	write	math	science	socst	sex	mean_G	schcode
1	70	0	4	1	pub	57	52	41	47	57	Male	50.80	0
2	121	1	4	2	pub	68	59	53	63	61	Female	60.80	0
3	88	0	4	3	pub	44	33	54	58	31	Male	44.00	0
4	141	0	4	3	pub	63	44	47	53	56	Male	52.60	0
5	172	0	4	2	pub	47	52	57	53	61	Male	54.00	0
6	113	0	4	2	pub	44	52	51	63	61	Male	54.20	0
7	50	0	3	2	pub	50	59	42	53	61	Male	53.00	0
8	11	0	1	2	pub	34	46	45	39	36	Male	40.00	0
9	84	0	4	2	pub	63	57	54	.	51	Male	56.25	0
10	48	0	3	2	pub	57	55	52	50	51	Male	53.00	0
11	75	0	4	2	pub	60	46	51	53	61	Male	54.20	0
12	60	0	4	2	pub	57	65	51	63	61	Male	59.40	0
13	95	0	4	3	pub	73	60	71	61	71	Male	67.20	0
14	104	0	4	3	pub	54	63	57	55	46	Male	55.00	0
15	38	0	3	1	pub	45	57	50	31	56	Male	47.80	0
16	115	0	4	1	pub	42	49	43	50	56	Male	48.00	0
17	76	0	4	3	pub	47	52	51	50	56	Male	51.20	0
18	195	0	4	2	pri	57	57	60	.	56	Male	57.50	1
19	114	0	4	3	pub	68	65	62	55	61	Male	62.20	0
20	85	0	4	2	pub	55	39	57	53	46	Male	50.00	0
21	167	0	4	2	pub	63	49	35	66	41	Male	50.80	0
22	143	0	4	2	pub	63	63	75	72	66	Male	67.80	0
23	41	0	3	2	pub	50	40	45	55	56	Male	49.20	0
24	20	0	1	3	pub	60	52	57	61	61	Male	58.20	0
25	12	0	1	2	pub	37	44	45	39	46	Male	42.20	0
26	53	0	3	2	pub	34	37	46	39	31	Male	37.40	0
27	154	0	4	3	pub	65	65	66	61	66	Male	64.60	0
28	178	0	4	2	pri	47	57	57	58	46	Male	53.00	1
29	196	0	4	3	pri	44	38	49	39	46	Male	43.20	1
30	29	0	2	1	pub	52	44	49	55	41	Male	48.20	0
31	126	0	4	2	pub	42	31	57	47	51	Male	45.60	0

Regression model with only school type

```
proc reg data=grades_m2;  
model mean_G= schcode;  
run;
```

Number of Observations Read	200
Number of Observations Used	200

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	175.61438	175.61438	2.65	0.1052
Error	198	13125	66.28833		
Corrected Total	199	13301			

Root MSE	8.14176	R-Square	0.0132
Dependent Mean	52.37325	Adj R-Sq	0.0082
Coeff Var	15.54566		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	51.96429	0.62815	82.73	<.0001
schcode	1	2.55603	1.57038	1.63	0.1052

Regression model with school type and ses

```
proc reg data=grades_m2;  
model mean_G= schcode ses;  
run;
```

Dependent Variable: mean_G

Number of Observations Read	200
Number of Observations Used	200

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1607.03679	803.51839	13.54	<.0001
Error	197	11694	59.35872		
Corrected Total	199	13301			

Root MSE	7.70446	R-Square	0.1208
Dependent Mean	52.37325	Adj R-Sq	0.1119
Coeff Var	14.71068		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	44.44380	1.64277	27.05	<.0001
schcode	1	1.54922	1.50011	1.03	0.3030
ses	1	3.73799	0.76120	4.91	<.0001

R^2 improved a lot after add the ses. Which mean after controlling for socioeconomic status, school type does explain as much of a student's average test grade

/*Bonus Question */

/*After controlling for a student's gender, school type, and socioeconomic status, do female students who attend private schools perform better than those who attend public schools? How do you test this? [10 Points]*/

```
proc sql;
  create table grades_m as
  select *,mean(read, write, math, science, socst) as mean_G
  from grades;
```

```
quit;
run;
```

```
data grades_m2;
  set grades_m;
  if schtype='pub' then schcode =0 ;
  if schtype='pri' then schcode =1 ;
```



```
run;
```

```
proc print data=grades_m2;  
run;
```

Obs	id	gender	race	ses	schtype	read	write	math	science	socst	sex	mean_G	schcode
1	70	0	4	1	pub	57	52	41	47	57	Male	50.80	0
2	121	1	4	2	pub	68	59	53	63	61	Female	60.80	0
3	86	0	4	3	pub	44	33	54	58	31	Male	44.00	0
4	141	0	4	3	pub	63	44	47	53	56	Male	52.60	0
5	172	0	4	2	pub	47	52	57	53	61	Male	54.00	0
6	113	0	4	2	pub	44	52	51	63	61	Male	54.20	0
7	50	0	3	2	pub	50	59	42	53	61	Male	53.00	0
8	11	0	1	2	pub	34	46	45	39	36	Male	40.00	0
9	84	0	4	2	pub	63	57	54	.	51	Male	56.25	0
10	48	0	3	2	pub	57	55	52	50	51	Male	53.00	0
11	75	0	4	2	pub	60	46	51	53	61	Male	54.20	0
12	60	0	4	2	pub	57	65	51	63	61	Male	59.40	0
13	95	0	4	3	pub	73	60	71	61	71	Male	67.20	0
14	104	0	4	3	pub	54	63	57	55	46	Male	55.00	0
15	38	0	3	1	pub	45	57	50	31	56	Male	47.80	0
16	115	0	4	1	pub	42	49	43	50	56	Male	48.00	0
17	76	0	4	3	pub	47	52	51	50	56	Male	51.20	0
18	195	0	4	2	pri	57	57	60	.	56	Male	57.50	1
19	114	0	4	3	pub	68	65	62	55	61	Male	62.20	0
20	85	0	4	2	pub	55	39	57	53	46	Male	50.00	0
21	167	0	4	2	pub	63	49	35	66	41	Male	50.80	0
22	143	0	4	2	pub	63	63	75	72	66	Male	67.80	0
23	41	0	3	2	pub	50	40	45	55	56	Male	49.20	0
24	20	0	1	3	pub	60	52	57	61	61	Male	58.20	0
25	12	0	1	2	pub	37	44	45	39	46	Male	42.20	0
26	53	0	3	2	pub	34	37	46	39	31	Male	37.40	0
27	154	0	4	3	pub	65	65	66	61	66	Male	64.60	0
28	178	0	4	2	pri	47	57	57	58	46	Male	53.00	1
29	196	0	4	3	pri	44	38	49	39	46	Male	43.20	1
30	29	0	2	1	pub	52	44	49	55	41	Male	48.20	0
31	126	0	4	2	pub	42	31	57	47	51	Male	45.60	0

```
proc reg data=grades_m2;  
model mean_G= schcode ses gender;  
run;
```

The REG Procedure
 Model: MODEL1
 Dependent Variable: mean_G

Number of Observations Read	200
Number of Observations Used	200

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1661.78585	553.92862	9.33	<.0001
Error	196	11639	59.38224		
Corrected Total	199	13301			

Root MSE	7.70599	R-Square	0.1249
Dependent Mean	52.37325	Adj R-Sq	0.1115
Coeff Var	14.71359		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	43.67956	1.82572	23.92	<.0001
schcode	1	1.50166	1.50122	1.00	0.3184
ses	1	3.83259	0.76769	4.99	<.0001
gender	1	1.05956	1.10348	0.96	0.3381

'pub'=0
 'pri'=1

Parameter of school type is positive. So privet school code is 1.
 Privet school female grades are better than public schools.