11. The following equations were estimated using the data in ECONMATH, with standard errors reported under coefficients. The average class score, measured as a percentage, is about 72.2; exactly 50% of the students are male; and the average of *colgpa* (grade point average at the start of the term) is about 2.81.

$$\begin{aligned}\widehat{score} &= \underset{(2.00)}{32.31} + \underset{(0.70)}{14.32}\, colgpa \\ n &= 856,\ R^2 = .329,\ \overline{R}^2 = .328.\end{aligned}$$

$$\begin{aligned}\widehat{score} &= \underset{(2.04)}{29.66} + \underset{(0.74)}{3.83}\, male + \underset{(0.69)}{14.57}\, colgpa \\ n &= 856,\ R^2 = .349,\ \overline{R}^2 = .348.\end{aligned}$$

$$\begin{aligned}\widehat{score} &= \underset{(2.86)}{30.36} + \underset{(3.96)}{2.47}\, male + \underset{(0.98)}{14.33}\, colgpa + \underset{(1.383)}{0.479}\, male \cdot colgpa \\ n &= 856,\ R^2 = .349,\ \overline{R}^2 = .347.\end{aligned}$$

$$\begin{aligned}\widehat{score} &= \underset{(2.86)}{30.36} + \underset{(0.74)}{3.82}\, male + \underset{(0.98)}{14.33}\, colgpa + \underset{(1.383)}{0.479}\, male \cdot (colgpa - 2.81) \\ n &= 856,\ R^2 = .349,\ \overline{R}^2 = .347.\end{aligned}$$

i. Interpret the coefficient on *male* in the second equation and construct a 95% confidence interval for β_{male}. Does the confidence interval exclude zero?

ii. In the second equation, how come the estimate on *male* is so imprecise? Should we now conclude that there are no gender differences in *score* after controlling for *colgpa*? [*Hint*: You might want to compute an *F* statistic for the null hypothesis that there is no gender difference in the model with the interaction.]

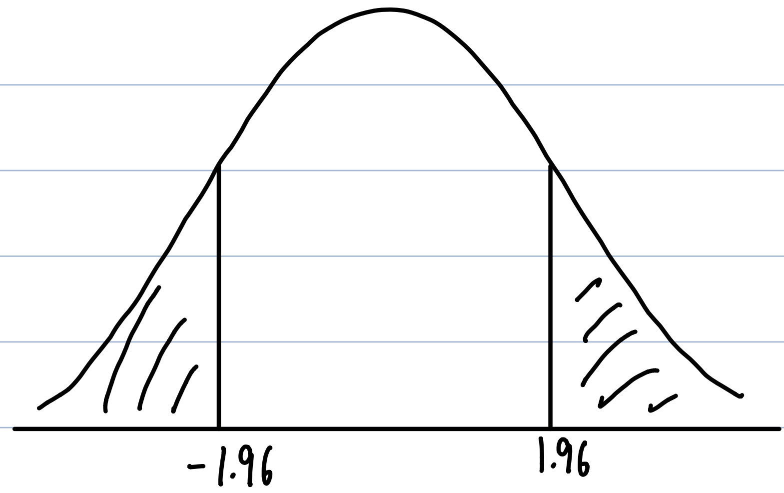
iii. Compared with the third equation, how come the coefficient on *male* in the last equation is so much closer to that in the second equation and just as precisely estimated?

i)

i. Interpret the coefficient on *male* in the second equation and construct a 95% confidence interval for β_{male}. Does the confidence interval exclude zero?

$$\begin{aligned}\widehat{score} &= \underset{(2.04)}{29.66} + \underset{(0.74)}{3.83}\, male + \underset{(0.69)}{14.57}\, colgpa \\ n &= 856,\ R^2 = .349,\ \overline{R}^2 = .348.\end{aligned}$$

An additional male would increase total score by 3.83 unit.



The 95% confidence interval for β_{male}

$$[\hat{\beta}_j - 1.96 s.e(\hat{\beta}_j), \hat{\beta}_j + 1.96 s.e(\hat{\beta}_j)]$$

$$[3.83 - 1.96(0.79), 3.83 + 1.96(0.79)]$$

$$[2.3796, 5.2804]$$

ii. In the second equation, how come the estimate on *male* is so imprecise? Should we now conclude that there are no gender differences in *score* after controlling for *colgpa*? [*Hint*: You might want to compute an *F* statistic for the null hypothesis that there is no gender difference in the model with the interaction.]

$H_0 : \hat{\beta}_{male} = 0$, $H_a : \beta_{male} \times Colgpa = 0$

$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)}$$

$$= \frac{0.348 - 0.324}{2} \times \frac{856-3-1}{9 - 0.348}$$

$$= 13.09$$

$\rightarrow 13.09 > 3.00$ So, we rejected H_0 that there is no gender differences in model at 5% significant level.

C4. Use the data in GPA2 for this exercise.

i. Consider the equation

$$colgpa = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 hsperc + \beta_4 sat$$
$$+\beta_5 female + \beta_6 athlete + u,$$

where *colgpa* is cumulative college grade point average; *hsize* is size of high school graduating class, in hundreds; *hsperc* is academic percentile in graduating class; *sat* is combined SAT score; *female* is a binary gender variable; and *athlete* is a binary variable, which is one for student-athletes. What are your expectations for the coefficients in this equation? Which ones are you unsure about?

ii. Estimate the equation in part (i) and report the results in the usual form. What is the estimated GPA differential between athletes and nonathletes? Is it statistically significant?

iii. Drop *sat* from the model and reestimate the equation. Now, what is the estimated effect of being an athlete? Discuss why the estimate is different than that obtained in part (ii).

iv. In the model from part (i), allow the effect of being an athlete to differ by gender and test the null hypothesis that there is no ceteris paribus difference between women athletes and women nonathletes.

v. Does the effect of *sat* on *colgpa* differ by gender? Justify your answer.

ii. Estimate the equation in part (i) and report the results in the usual form. What is the estimated GPA differential between athletes and nonathletes? Is it statistically significant?

```
. reg colgpa hsize hsizesq hsperc sat female athlete
```

Source	SS	df	MS			
				Number of obs	=	4,137
				F(6, 4130)	=	284.59
Model	524.819305	6	87.4698842	Prob > F	=	0.0000
Residual	1269.37637	4,130	.307355053	R-squared	=	0.2925
				Adj R-squared	=	0.2915
Total	1794.19567	4,136	.433799728	Root MSE	=	.5544

colgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsize	β_1 -.0568543	.0163513	-3.48	0.001	-.0889117	-.0247968
hsizesq	β_2 .0046754	.0022494	2.08	0.038	.0002654	.0090854
hsperc	β_3 -.0132126	.0005728	-23.07	0.000	-.0143355	-.0120896
sat	β_4 .0016464	.0000668	24.64	0.000	.0015154	.0017774
female	β_5 .1548814	.0180047	8.60	0.000	.1195826	.1901802
athlete	β_6 .1693064	.0423492	4.00	0.000	.0862791	.2523336
_cons	β_0 1.241365	.0794923	15.62	0.000	1.085517	1.397212

$$Colgpa = 1.241 - 0.0569\,hsize + 0.00468\,hsize^2 - 0.0132\,hsperc + 0.00165\,sat + 0.156\,female + 0.169\,athlete$$

$N = 1437$, $R^2 = 0.293$

The GPA athlete is approximately 0.169 higher than nonathlete.

$$t_{athlete} = \frac{\hat{\beta}_{athlete} - \beta_{athlete}}{s.e.\,\hat{\beta}_{athlete}}$$

$$= \frac{0.169 - 0}{0.042} = 4.02$$

iii. Drop *sat* from the model and reestimate the equation. Now, what is the estimated effect of being an athlete? Discuss why the estimate is different than that obtained in part (ii).

```
. reg colgpa hsize hsizesq hsperc female athlete
```

Source	SS	df	MS
Model	338.217123	5	67.6434247
Residual	1455.97855	4,131	.35245184
Total	1794.19567	4,136	.433799728

Number of obs	=	4,137
F(5, 4131)	=	191.92
Prob > F	=	0.0000
R-squared	=	0.1885
Adj R-squared	=	0.1875
Root MSE	=	.59368

colgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsize	-.0534038	.0175092	-3.05	0.002	-.0877313	-.0190763
hsizesq	.0053228	.0024086	2.21	0.027	.0006007	.010045
hsperc	-.0171365	.0005892	-29.09	0.000	-.0182916	-.0159814
female	.0581231	.0188162	3.09	0.002	.0212333	.095013
athlete	.0054487	.0447871	0.12	0.903	-.0823582	.0932556
_cons	3.047698	.0329148	92.59	0.000	2.983167	3.112229

Coefficient (athlete) : 0.0054

S.e athlete : 0.0448

$$t_{atlete} = \frac{\hat{\beta}_{athlete} - \beta_{athlete}}{s.e.\ \hat{\beta}_{athlete}}$$

$$= \frac{0.0054 - 0}{0.0448} = 0.1205$$

→ Not statiscally significant since there are no SAT score to control (Athletes score is lower than nonathletes.

iv. In the model from part (i), allow the effect of being an athlete to differ by gender and test the null hypothesis that there is no ceteris paribus difference between women athletes and women nonathletes.

```
. gen femath= female==1& athlete==1

. gen male=female==0

. gen maleath=male==1& athlete==1

. gen malenonath=male==1& athlete==0

. reg colgpa hsize hsizesq hsperc sat femath maleath malenonath
```

Source	SS	df	MS
Model	524.821272	7	74.9744674
Residual	1269.3744	4,129	.307429015
Total	1794.19567	4,136	.433799728

Number of obs	=	4,137
F(7, 4129)	=	243.88
Prob > F	=	0.0000
R-squared	=	0.2925
Adj R-squared	=	0.2913
Root MSE	=	.55446

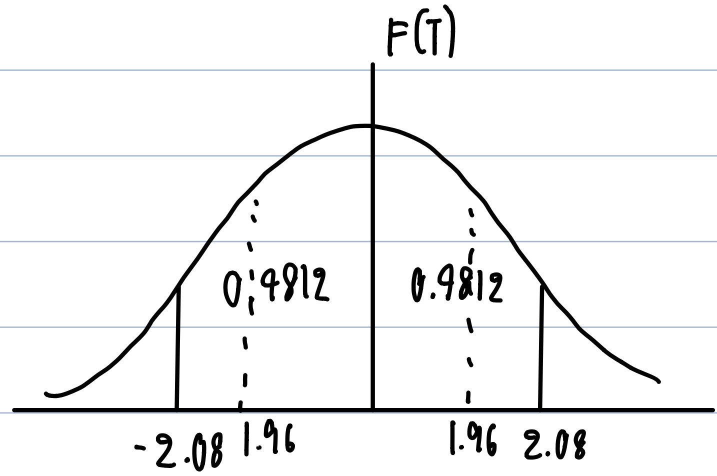
colgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsize	-.0568006	.0163671	-3.47	0.001	-.0888889	-.0247124
hsizesq	.0046699	.0022507	2.07	0.038	.0002573	.0090825
hsperc	-.0132114	.000573	-23.06	0.000	-.0143349	-.012088
sat	.0016462	.0000669	24.62	0.000	.0015151	.0017773
femath	.1751106	.0840258	2.08	0.037	.0103748	.3398464
maleath	.0128034	.0487395	0.26	0.793	-.0827523	.1083591
malenonath	-.1546151	.0183122	-8.44	0.000	-.1905168	-.1187133
_cons	1.39619	.0755581	18.48	0.000	1.248055	1.544324

$$\widehat{Colgpa} : 1.396 - 0.0568\, hsize + 0.00467\, hsize^2 - 0.0132\, hsperc + 0.00165\, Sat + 0.175\, female + 0.0128\, maleath - 0.155\, male\ nonath$$

$N = 4137 \quad R^2 = 0.293.$

$$t_{female} = \frac{\hat{\beta}_{female} - \beta_{female}}{s.e.\ \hat{\beta}_{female}}$$

$$= \frac{0.175 - 0}{0.084} = 2.08$$



→ one side → $0.05 \div 2 = 0.025$

$0.5 - 0.025 = 0.475 \rightarrow 1.96$

$t = 2.08$ is significantly at 5% level of significant.

v. Does the effect of *sat* on *colgpa* differ by gender? Justify your answer.

No, since female sat = female × SAT

$$t_{femalesat} = \frac{\hat{\beta}_{femalesat} - \beta_{femalesat}}{s.e\ \beta_{femalesat}}$$

$$= \frac{0.000051 - 0}{0.0001291}$$

$$= 0.41$$

→ too small amount of evidence to say that effect of sat on colgpa differ on gender.

```
. gen femalesat = female * sat

. reg colgpa hsize hsizesq hsperc sat female femalesat athlete
```

Source	SS	df	MS
Model	524.867644	7	74.981092
Residual	1269.32803	4,129	.307417784
Total	1794.19567	4,136	.433799728

Number of obs	=	4,137
F(7, 4129)	=	243.91
Prob > F	=	0.0000
R-squared	=	0.2925
Adj R-squared	=	0.2913
Root MSE	=	.55445

colgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsize	-.0569121	.0163537	-3.48	0.001	-.0889741	-.0248501
hsizesq	.0046864	.0022498	2.08	0.037	.0002757	.0090972
hsperc	-.013225	.0005737	-23.05	0.000	-.0143497	-.0121003
sat	.0016255	.0000852	19.09	0.000	.0014585	.0017924
female	.1023066	.1338023	0.76	0.445	-.1600179	.3646311
femalesat	.0000512	.0001291	0.40	0.692	-.000202	.0003044
athlete	.1677568	.0425334	3.94	0.000	.0843684	.2511452
_cons	1.263743	.0974952	12.96	0.000	1.0726	1.454887