Chapter 6

3) $\widehat{rdintens} = 2.613 + 0.003\, sales - 0.000000007\, sales^2$
$\quad\quad (0.429) \quad (0.00014) \quad (0.0000000037)$

$n = 32$, $R^2 = 0.1484$

i) what point does the marginal effect of sales on rdintens become negative?

$$\frac{d\widehat{rdintens}}{dsales} = 0.0003 - 0.000000014 sales$$

for the marginal effect to be negative

$$\frac{d\widehat{rdintens}}{d\,sales} < 0$$

it implies, $0.0003 - 0.000000014 sales < 0$

$0.000000014 sales > 0.0003$

$sales > \frac{0.0003}{0.000000014}$

$sales > 21428.5714$

Hence, sales = 21428.5714 is the marginal effect of sales on rdintens become negative.

ii) would you keep the quandratic in the model?

$$t = \frac{\hat{\beta}sales^2 - \beta sales^2}{std.\ error}$$
$$= \frac{-0.000000007 - 0}{0.0000000037}$$
$$= -1.89$$

So, $H_0: \beta sales^2 < 0$

iii) Define salesbil as sales measured in billions of dollars;
sales = 1000 salesbils → $salesbil = \frac{sales}{1000}$. Rewrite the estimated equation with salesbil and salesbil² as the independent variables. $\left[salesbil^2 = \frac{sales^2}{1000^2} \right]$

Ans $\widehat{rdintens} = 2.613 + 0.0003 sales - 0.000000007 sales^2$
$= 2.613 + 0.0003 (1000 \times salebil) - 0.000000007 (1000 \times salesbil)^2$
$= 2.613 + 0.3\, salesbil - 0.007\, salesbil^2$
$\quad (0.429) \quad (0.14) \quad (0.0037)$

$n = 32$, $R^2 = 0.1484$

iv) for the purpose of reporting the results, which equation do you prefer?

<u>Ans</u> I prefer (iii) equation since it consists of few zeros and easier to understant rather than the original equation providing many zeros

Chapter 7

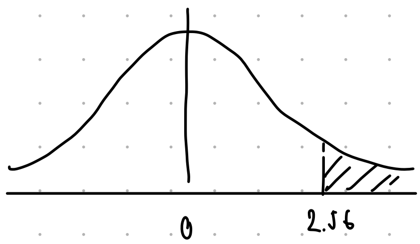
1) $\widehat{sleep} = \underset{(235.11)}{3840.83} - \underset{(0.018)}{0.163}\, totwrk - \underset{(5.86)}{11.71}\, educ - \underset{(11.21)}{8.70}\, age + \underset{(0.134)}{0.128}\, age^2 + \underset{(34.33)}{87.75}\, male$

$n = 706$, $R^2 = 0.123$, $\bar{R}^2 = 0.117$

i) All other factors being equal, is there evidence that men sleep more than women? How strong is the evidence?

<u>Ans</u> coefficient on male is 87.75

$$t_{male} = \frac{\hat{\beta}_{male} - \beta_{male}}{se.(\hat{\beta}_j)} = \frac{87.75 - 0}{34.33} = 2.56$$



p-value : $0.5 - 0.4948 = 0.0052$

1% level of significant : one-side $= 0.005$

So, p-value > significant level is $0.0052 > 0.005$

Hence, the significant at 1% level

ii) Is there a statistically significant trade-off b/w working and sleeping? What is the estimated trade-off?

<u>Ans</u> The coefficient on totwrk is negative and 1% significant level

$$t_{totwrk} = \frac{\hat{\beta}_{totwrk} - \beta_{totwrk}}{se.\hat{\beta}_{totwrk}} = \frac{-0.163 - 0}{0.018} = -9.06$$ → statistically significant

1 more hour of work : $0.163(60) = 9.8$ minutes less sleep

iii) what other regression do you need to run to test the null hypothesis that, holding other factors fixed, age has no effect on sleeping?

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

8) Suppose you collect data from a survey on wages, education, experience, and gender. In addition, you ask for information about marijuana usage. The original question is: "On how many separate occasions last month did you smoke marijuana?"

i. Write an equation that would allow you to estimate the effects of marijuana usage on wage, while controlling for other factors. You should be able to make statements such as, "Smoking marijuana five more times per month is estimated to change wage by *x*%."

$$\log(wage) = \beta_0 + \beta_1 usage + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + u$$

So, $100\beta_1$ is the percentage change in wage when marijuana usage rise by 1 time per month

ii. Write a model that would allow you to test whether drug usage has different effects on wages for men and women. How would you test that there are no differences in the effects of drug usage for men and women?

$$\log(wage) = \beta_0 + \beta_1 usage + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + \beta_6 female \cdot usage + u$$

$H_0 : \beta_6 = 0$
$H_a : \beta_6 \neq 0$

iii. Suppose you think it is better to measure marijuana usage by putting people into one of four categories: nonuser, light user (1 to 5 times per month), moderate user (6 to 10 times per month), and heavy user (more than 10 times per month). Now, write a model that allows you to estimate the effects of marijuana usage on wage.

3 groups user : 1 light user : lghtuser
2 mode user : moduser
3 heavy user : hvyuser

$$\log(wage) = \beta_0 + \delta_1 lghtuser + \delta_2 moduser + \delta_3 hvyuser + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + u$$

iv. Using the model in part (iii), explain in detail how to test the null hypothesis that marijuana usage has no effect on wage. Be very specific and include a careful listing of degrees of freedom.

$H_0 : \delta_1 = \delta_2 = \delta_3 = 0$

total of q = 3 restrictions
n = sample size
df = unrestricted models

Thus, we obtain $F_{q, n-k}$ distribution (the critical value come from)

v. What are some potential problems with drawing causal inference using the survey data that you collected?

the potential problems are factors especially family background that are directly affect wages and correlated with marijuana usage. We are keen on the impacts of an individual's drug usage on their wage, so we might want to hold other confounding factors fixed. We could attempt to gather information on relevant background information

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.

1)
$$\widehat{score} = \underset{(2.00)}{32.31} + \underset{(0.70)}{14.32}\, colgpa$$
$$n = 856, R^2 = .329, \bar{R}^2 = .328.$$

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

3)
$$\widehat{score} = \underset{(2.86)}{30.36} + \underset{(3.96)}{2.47}\, male + \underset{(0.98)}{14.33}\, colgpa + \underset{(1.383)}{0.479}\, male \cdot colgpa$$
$$n = 856, R^2 = .349, \bar{R}^2 = .347.$$

4)
$$\widehat{score} = \underset{(2.86)}{30.36} + \underset{(0.74)}{3.82}\, male + \underset{(0.98)}{14.33}\, colgpa + \underset{(1.383)}{0.479}\, male \cdot (colgpa - 2.81)$$
$$n = 856, R^2 = .349, \bar{R}^2 = .347.$$

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?

coeff male : 3.83

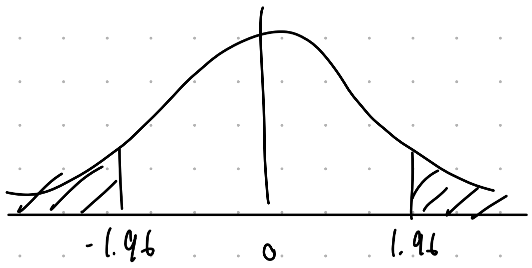
The 95% confidence interval for β male

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

$: [3.83 - 1.96(0.74), 3.83 + 1.96(0.74)]$

$:$ 2.3796, 5.2804

The confidence interval exclude 0.



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.]

$H_0 : \beta_{male} = 0, \quad H_0 : \beta_{male \cdot colgpa} = 0$

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

$$= \frac{0.348 - 0.329}{2} \times \frac{856-3-1}{1-0.348}$$

$$= 13.09$$

$F_{2,854} = 3.00$, since $13.09 > 3.00$ so I rejected Ho that there is no gender difference in model at 5% significant level

Computer Questions

(4)

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?

There are some controlling of hsperc and sat on any gender & athlete.

- The student athletes would be β_5.
- The coefficient in the equation can tell me about what factor is better or worse. For example in part (ii) athlete do better on grading comparing with other variables
- unsure : β_1 and β_2 since it's size of graduating whether males and female with different GPAs. No clear information about gender that the regression in part (ii) didn't include the specific gender on hsize.

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
Model	524.819305	6	87.4698842
Residual	1269.37637	4,130	.307355053
Total	1794.19567	4,136	.433799728

Number of obs	=	4,137
F(6, 4130)	=	284.59
Prob > F	=	0.0000
R-squared	=	0.2925
Adj R-squared	=	0.2915
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_7 1.241365	.0794923	15.62	0.000	1.085517	1.397212

$$\widehat{Colgpa} = \underset{0.079}{1.241} - \underset{0.0164}{0.0569}\,hsize + \underset{0.00225}{0.00468}\,hsize^2 - \underset{0.0006}{0.0132}\,hsperc + \underset{0.00007}{0.00165}\,sat$$

$$+ \underset{0.018}{0.155}\,female + \underset{0.042}{0.169}\,athlete$$

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

The GPA of athlete is around 0.169 point higher than nonathlete.

$$t_{athlete} = \frac{\hat{\beta}_{athlete} - \beta_{athlete}}{s.e.\,\hat{\beta}_{athlete}} = \frac{0.169 - 0}{0.042} = 4.02$$

It's statistically significant at 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

Coefficent (athlete) : 0.0054
s.e. athlete : 0.0448

$$t_{athlete} = \frac{\hat{\beta}_{athlete} - \beta_{athelete}}{s.e\ \hat{\beta}_{athlete}} = \frac{0.0054 - 0}{0.0448} = 0.1205$$

It's not statistically significant since there are no SAT score to control. Athletes score is lower than nonathletes. Part (i) has shown that when we include SAT score, athlete will do better than nonathlete.

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.

no difference b/w women athletes and women nonathletes

I chose female nonathlete

femath : female athlete
maleath : male athlete
malenonath : male nonathlete

```
. 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

$$\text{Colgpa} : \underset{0.076}{1.396} - \underset{0.0164}{0.0568}\,\text{hsize} + \underset{0.00225}{0.00467}\,\text{hsize}^2 - \underset{0.0006}{0.0132}\,\text{hsperc}$$

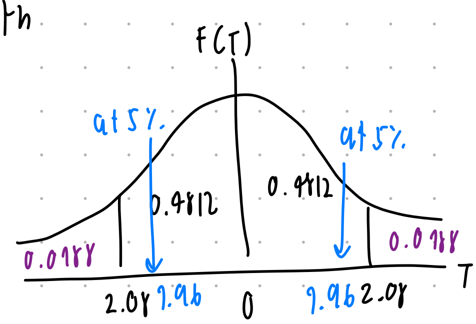
$$+ \underset{0.00007}{0.00165}\,\text{sat} + \underset{0.084}{0.175}\,\text{female} + \underset{0.049}{0.0128}\,\text{maleath}$$

$$- \underset{0.018}{0.155}\,\text{malenonath}$$

$N = 4137$, $R^2 = 0.293$

The colgpa predicted that female athlete is 0.175 that is higher than female nonathlete

$$t_{femath} = \frac{\hat{\beta}_{femath} - \beta_{femath}}{s.e.\ \hat{\beta}_{femath}} = \frac{0.175 - 0}{0.084} = 2.08$$



if 5% significant level

only one side : 0.05

two side : 0.025 → 0.5 − 0.025 = 0.475 → 1.96

t = 2.08 which is statistically significant at 5% level of significant against 2 sided alternative

iv) Does the effect of sat on colgpa differ by gender?

No, only athlete because whether I added femalesat to the equation in part (ii) or (iv), the result will be the same. So, when I generate femalesat = female • sat

```
. 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

$$t_{femalesat} = \frac{\hat{\beta}_{femalesat} - \beta_{femalesat}}{s.e\ \hat{\beta}_{femalesat}} = \frac{0.000051 - 0}{0.0001291} = 0.4$$

There is little evidence that the effect of sat on colgpa differ by gender