Athicha K.
6104640062

3. Using the data in RDCHEM, the following equation was obtained by OLS:

$$\widehat{rdintens} = \underset{(.429)}{2.613} + \underset{(.00014)}{.00030}\, sales - \underset{(.0000000037)}{.0000000070}\, sales^2$$
$$n = 32, R^2 = .1484.$$

i. At what point does the marginal effect of *sales* on *rdintens* become negative?

$$\frac{drdintens}{dsales} = 0.00030 - 2(0.0000000070\, sales)$$

In order to have the marginal effect of sales on rdintens to become negative.

$$\frac{drdintens}{dsales} < 0 \longrightarrow 0.00030 - 0.000000014 sales < 0$$

$$0.00030 < 0.000000014 sales$$

$$21,428.5714 < sales$$

∴ at sales of 21,428.14, the marginal effect of sales on rdintens becomes negative.

ii. Would you keep the quadratic term in the model? Explain.

$$t = \frac{-0.0000000070}{0.0000000037} = -1.8919 > -1.96$$

As $t > -1.96$, the quadratic term is statistically significant so, we will kept this variable at 5% significant level.

iii. Define *salesbil* as sales measured in billions of dollars: $salesbil = sales/1{,}000$. Rewrite the estimated equation with *salesbil* and $salesbil^2$ as the independent variables. Be sure to report standard errors and the *R*-squared. [*Hint*: Note that $salesbil^2 = sales^2/(1{,}000)^2$.]

Given that $salesbil = sales/1000$, $sales = 1000 salesbil$
and $salesbil^2 = sales^2/1000^2$, $sale^2 = (1000 salesbil)^2$

$$\widehat{rdintens} = 2.613 + 0.0003(1000 salesbil) - 0.0000000070(1000 salesbil)^2$$

$$\widehat{rdintens} = \underset{(.429)}{2.613} + \underset{(.14)}{0.3} salesbil - \underset{(.0037)}{0.007} salesbil^2$$

iv. For the purpose of reporting the results, which equation do you prefer?

Although the two equations are identical, the second one on (iii) is more comfortable to read as the equation is shorter.

1. Using the data in SLEEP75 (see also Problem 3 in Chapter 3), we obtain the estimated equation

$$\begin{aligned}\widehat{sleep} &= 3{,}840.83 - .163\ totwrk - 11.71\ educ - 8.70\ age \\ &\quad\ (235.11)\quad (.018) \qquad\qquad (5.86) \qquad\ (11.21) \\ &\quad + .128\ age^2 + 87.75\ male \\ &\quad\ (.134) \qquad\ (34.33) \\ n &= 706,\ R^2 = .123,\ \overline{R}^2 = .117.\end{aligned}$$

The variable *sleep* is total minutes per week spent sleeping at night, *totwrk* is total weekly minutes spent working, *educ* and *age* are measured in years, and *male* is a gender dummy.

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

yes there is, as the coefficient of male is 87.75. This can imply that male sleeps 87.75 minutes/week more than female.

ii. Is there a statistically significant tradeoff between working and sleeping? What is the estimated tradeoff?

one minute spent more with work will result in 0.163 minutes spent less with sleep. However, this is not a significant trade off between working and sleeping.

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?

we need to run a regression with restricted model excluding age, we have

$H0: \beta 3 = \beta 4 = 0$

Ha: otherwise

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?"

usage = marijuana usage.

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*%."

$$wage = \beta_0 + \beta_1 usage + \beta_2 educ + \beta_3 exper + \beta_4 female + u$$

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?

$$wage = \beta_0 + \beta_1 useage + \beta_2 educ + \beta_3 exper + \beta_4 female + \beta_5 useage \cdot female + u$$

Test $H_0: \beta_5 = 0$

$H_a: \beta_5 \neq 0$

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.

$$wage = \beta_0 + \beta_1 light + \beta_2 moderate + \beta_3 heavy + \beta_4 educ + \beta_5 exper + \beta_6 female + u$$

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.

รอเขียน

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

รอเขียน

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.

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

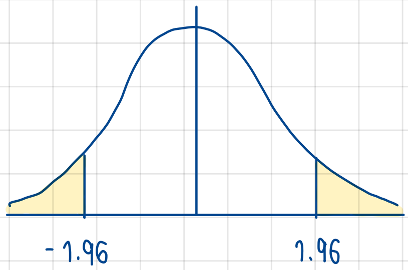
$$\widehat{score} = \underset{(2.04)}{29.66} + \underset{(0.74)}{3.83}\, male + \underset{(0.69)}{14.57}\, colgpa$$
$$n = 856, R^2 = .349, \bar{R}^2 = .348.$$

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

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

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?

The coefficient of 3.83 means male is scored more than woman by 3.83



$= [3.83 - 1.96(0.74),\ 3.83 + 1.96(0.74)]$

$= [2.3796,\ 5.2704]$

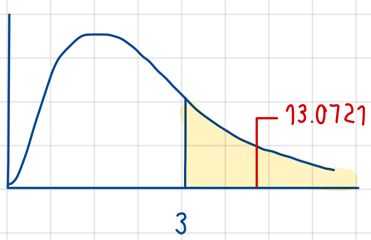
The confidence interval excludes zero

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.]

$$F = \frac{SSR_{ur} - SSR_r / q}{1 - SSR_{ur} / n-k-1} = \frac{(0.349 - 0.329)/2}{1 - 0.349 / 836-7-1}$$

$$= 13.0721$$

$R^2_{ur} - R^2_r / q$



we reject $H_0: \beta_{male} = \beta_{male \cdot colgpa} = 0$. So, there is a gender difference at 5% significant level.
The estimate on male is implicise as we exclude many significant variables.

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?

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?

. regress colgpa hsize hsize2 hsperc sat female athlete

Source	SS	df	MS			
Model	524.819305	6	87.4698842	Number of obs	=	4,137
Residual	1269.37637	4,130	.307355053	F(6, 4130)	=	284.59
				Prob > F	=	0.0000
				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	-.0568543	.0163513	-3.48	0.001	-.0889117	-.0247968
hsize2	.0046754	.0022494	2.08	0.038	.0002654	.0090854
hsperc	-.0132126	.0005728	-23.07	0.000	-.0143355	-.0120896
sat	.0016464	.0000668	24.64	0.000	.0015154	.0017774
female	.1548814	.0180047	8.60	0.000	.1195826	.1901802
athlete	.1693064	.0423492	4.00	0.000	.0862791	.2523336
_cons	1.241365	.0794923	15.62	0.000	1.085517	1.397212

β_1 is expected to have a negative effect.

β_3 is expected to have a negative effect.

β_4 is expected to have a negative effect.

β_6 is expected to have a negative effect.

We are unsure about β_2 and β_5 as β_2 is in a quadratic form and β_5 is a dummy variable.

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?

$$\widehat{colgpa} = 1.24 - 0.569 hsize + 0.00468 hsize^2 - 0.0132 hsperc + 0.00165 sat + 0.155 female + 0.169 athlete + u$$

Athlete is expected to score better than non-athlete by 0.169 at 1% significant level.

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).

. regress colgpa hsize hsize2 hsperc female athlete

Source	SS	df	MS			
Model	338.217123	5	67.6434247	Number of obs	=	4,137
Residual	1455.97855	4,131	.35245184	F(5, 4131)	=	191.92
				Prob > F	=	0.0000
				R-squared	=	0.1885
				Adj R-squared	=	0.1875
Total	1794.19567	4,136	.433799728	Root MSE	=	.59368

colgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsize	-.0534038	.0175092	-3.05	0.002	-.0877313	-.0190763
hsize2	.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

The estimated effect is lower by a significant level to 0.00545 as we don't control SAT any more which result in the fact that athlete will still do better than non-athlete

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.

. regress 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

The coefficient of 0.175 of female athletes means female athletes is scored better than female non-athletes by 0.175

As 2.08 > 1.96, there is a difference in colgpa between female athletes and female non-athletes

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

. regress colgpa hsize hsperc femsat female athlete

Source	SS	df	MS
Model	411.075682	5	82.2151365
Residual	1383.11999	4,131	.334814813
Total	1794.19567	4,136	.433799728

Number of obs	=	4,137
F(5, 4131)	=	245.55
Prob > F	=	0.0000
R-squared	=	0.2291
Adj R-squared	=	0.2282
Root MSE	=	.57863

colgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsize	-.0190793	.0051959	-3.67	0.000	-.0292661	-.0088925
hsperc	-.0156907	.0005655	-27.74	0.000	-.0167995	-.014582
femsat	.0015785	.0001058	14.92	0.000	.0013712	.0017859
female	-1.522999	.1076139	-14.15	0.000	-1.73398	-1.312018
athlete	.0226718	.0436648	0.52	0.604	-.0629347	.1082783
_cons	2.977225	.0226209	131.61	0.000	2.932876	3.021574

yes as 14.92 > 1.96, so, we reject Ho: $\beta_{female\ sat} = 0$.