Pitchapovn Inthisorn
6104641219

Chapter Review

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?

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

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 $salesbil^2 = sales^2/(1{,}000)^2$.]

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

i) $\frac{d\,rdintens}{d\,sales} = 0.0030 - 2(0.0000000070\ sales)$

to have maginal effect of sales on rdintens to be negative.

$\frac{d\,rdintens}{d\,sales} < 0 \rightarrow 0.00030 - 0.000000014\ sales < 0$

$21{,}428.5714 < sales.$

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

$t > -1.96$ means than this variable is significant at 5% significant level.

iii) $\widehat{rdintens} = 2.613 + 0.3\ sales - 0.000007\ sale^2$

$(.429) \quad (0.14) \quad (0.0037)$

iv) I prefer the second one because it's easier to understand eventhough they both have the same implications.

PAPER AND PENCIL QUESTIONS

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

min working

$$\underset{\text{min sleep/wk}}{\widehat{sleep}} = \underset{(235.11)}{\overset{\beta_1}{3{,}840.83}} - \underset{(.018)}{\overset{\beta_2}{.163}}\, \underline{totwrk} - \underset{(5.86)}{\overset{\beta_3}{11.71}}\, \underline{educ} - \underset{(11.21)}{\overset{\beta_4}{8.70}}\, \underline{age}$$

$$+ \underset{(.134)}{\overset{\beta_5}{.128}}\, \underline{age^2} + \underset{(34.33)}{\overset{\beta_6}{87.75}}\, \underline{male}$$

$$n = 706,\ R^2 = .123,\ \bar{R}^2 = .117.$$

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?

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

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?

i) yes, there is an evidence that men sleep more than woman by 87.75 min / week.

ii) yes, we need to trade 1 minute weekly spend working to gain 0.163 min of sleeping more.

iii) $H_0 = \beta_4 + \beta_5 = 0$

$H_a = \beta_4 + \beta_5 \neq 0$

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*%."

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?

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

Print Preview

than 10 times per month). Now, write a model that allows you to estimate the effects of marijuana usage on wage.

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.

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

i) $\widehat{\log wage} = \beta_0 - 5\,mrjn + \beta_1 education + \beta_2 experience + \delta_1 female$

ii) ♀ : $E(wage \mid female = 1, educ) = \beta_0 + \delta_0(1) - 5\,mrjn + \beta_1 education + \beta_2 experience$

♂ : $E(wage \mid male = 0, educ) = \beta_0 + \delta_0(0) - 5\,mrjn + \beta_1 education + \beta_2 experience$

$H_0: \delta_0 = 0$

$H_a: \delta_0 \neq 0$

iii) $\log(wage) = \beta_0 + \delta_1 nonuser + \delta_2 light + \delta_3 moderate + \delta_4 heavy + \beta_2 educ + \beta_3 exper + \beta_4 female + u$

iv) $H_0: \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0$ test by f-test $k = 7$

D.f. $n - k - 1 = n - 7 - 1 = n - 8$; $F_{3, n-8}$

iv) law, self-selection bias.

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?

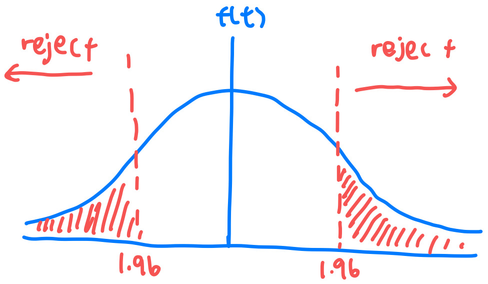
i) From second equation $\dfrac{\partial \widehat{score}}{\partial male} = 3.83$

It implies that male participants will have higher score by 3.83%.

Ho : $\beta male = 0$

Ha : $\beta male \neq 0$

at 95% confidence interval



$t_{male} = \dfrac{3.83}{0.74} = 5.18 > 1.96$

we reject Ho at 95% confidence interval

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.]

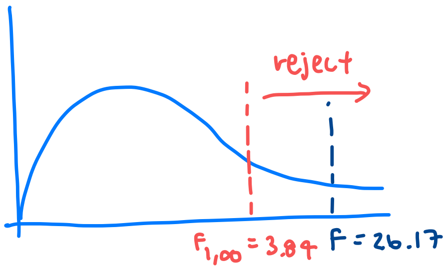
iii. Compared with the third equation, how come the coefficient on *male* in the last equation is so much closer to that in the second equation and just as precisely estimated?

ii) Conduct F-test

$H_0 : \beta male = 0$

$H_a : \beta male \neq 0$

$$F = \frac{(0.348 - 0.328)/1}{(1 - 0.348)/853} = \frac{0.02}{0.00076} = 26.17$$



$F = 26.17 > 3.84$, we reject H_0 at 5% significant level

iii) Since, averange colgpa is 2.81, this will mak colgpa - 2.81 equal to 0 and 0.479 male (0) = 0, so the value of the coefficient on male are similar.

COMPUTER EXERCISE

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.

i) my expectation are $\beta_3 < 0$, the less percentile, the higher GPA

$\beta_4 > 0$, the more SAT score the higher GPA and $\beta_6 < 0$

because athlete need to spend time on practicing

I am not sure about whether the size of the school and gender will affect GPA or not.

ii)

. regress colgpa hsize hsizesq 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
hsizesq	.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

$$\widehat{colgpa} = 1.241 - 0.057\, hsize + 0.005\, hsizesq - 0.013\, hsperc + 0.002\, sat + 0.155\, female + 0.169\, athlete + u$$

An athlet is predicted to have higher GPA by .169

$t = 4.00$

iii)

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
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

Dropped SAT out of the regression, the coefficient of athlete become .005 with st. error of .045 which is not significant.

This predicted that athlete have lower SAT score than non athlete.

iv)

Source	SS	df	MS		
Model	524.821272	7	74.9744674	Number of obs =	4,137
Residual	1269.3744	4,129	.307429015	F(7, 4129) =	243.88
				Prob > F =	0.0000
Total	1794.19567	4,136	.433799728	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
femaleath	.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

$\widehat{colgpa} = 1.396 - 0.057\, hsize + 0.005\, hsizesq - 0.013\, hsperc + 0.002\, sat + 0.175\, femaleath + 0.013\, maleath - 0.155\, malenonath.$

the coefficient on femaleath is .175 which is higher than female nonathlete.

$H_0: \beta_{femaleath} = 0$

H_a: otherwise

v.)

. regress colgpa hsize hsizesq hsperc sat female athlete femalesat

Source	SS	df	MS		
Model	524.867644	7	74.981092	Number of obs =	4,137
Residual	1269.32803	4,129	.307417784	F(7, 4129) =	243.91
				Prob > F =	0.0000
Total	1794.19567	4,136	.433799728	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
athlete	.1677568	.0425334	3.94	0.000	.0843684	.2511452
femalesat	.0000512	.0001291	0.40	0.692	-.000202	.0003044
_cons	1.263743	.0974952	12.96	0.000	1.0726	1.454887

$\widehat{colgpa} = 1.264 - .0569\, hsize + .00469\, hsizesq - .0132\, hsperc + .001626\, sat + .1023\, female + .16776\, athlete + .0000512\, femalesat$

Coefficient of femalesat is only 0.00005 and .4 t-stat, Gender rarly have impact on SAT.