Daniel Carpenter

2/24/2019

PS3

Exercise 4.1

Which of the following can cause the usual OLS t statistics to be invalid (that is, not to have t distributions under)?

1. Both Heteroskedasticity and high correlation between the two independent variables can cause the t-statistics to be invalid.

Exercise 4.5

Consider the estimated equation from Example 4.3, which can be used to study the effects of skipping class on college GPA:

1. **Using the standard normal approximation, find the 95% confidence interval for .**
   a. c = -1.96, 1.96

2. **Can you reject the hypothesis =.4 against the two-sided alternative at the 5% level?**
   a. $t_{140} = \frac{0.412}{0.094} = 4.382$
   b. 4.382 > 1.96, so reject H_0

3. **Can you reject the hypothesis =1 against the two-sided alternative at the 5% level?**
   a. $t_{140} = \frac{1}{0.094} = 10.638$
   b. 10.638 > 1.96, so reject H_0

Exercise 8.4

Using the data in GPA3, the following equation was estimated for the fall and second semester students:

$$\begin{aligned}
\widehat{trmgpa} = {} & -2.12 + .900\ crsgpa + .193\ cumgpa + .0014\ tothrs \\
& (.55) \quad (.175) \qquad (.064) \qquad (.0012) \\
& [.55] \quad [.166] \qquad [.074] \qquad [.0012] \\
& + .0018\ sat - .0039\ hsperc + .351\ female - .157\ season \\
& (.0002) \qquad (.0018) \qquad (.085) \qquad (.098) \\
& [.0002] \qquad [.0019] \qquad [.079] \qquad [.080] \\
n = 269, R^2 = .465. &
\end{aligned}$$

Here, trmgpa is term GPA, crsgpa is a weighted average of overall GPA in courses taken, cumgpa is GPA prior to the current semester, tothrs is total credit hours prior to the semester, sat is SAT score, hsperc is graduating percentile in high school class, female is a gender dummy, and season is a dummy variable equal to unity if the student's sport is in season during the fall. The usual and heteroskedasticity-robust standard errors are reported in parentheses and brackets, respectively.

1. **Do the variables crsgpa, cumgpa, and tothrs have the expected estimated effects? Which of these variables are statistically significant at the 5% level? Does it matter which standard errors are used?**
    a. Expected estimates:
        i. crsgpa: This seems high compared to other measures but seems reasonable.
        ii. cumgpa: Yes, this estimate makes sense.
        iii. tohrs: I would have assumed higher effect on term GPA.
    b. Statistically Significant at 5% significance:
        i. Standard Errors:
            1. crsgpa: Yes: 0.9/0.175 = 5.14
            2. cumgpa: Yes: 0.193/.064 = 3.02
            3. tohrs: No: 0.0014/0.0012 = 1.17
        ii. Robust-Standard Errors
            1. crsgpa: Yes: 0.9/0.166 = 5.42
            2. cumgpa: Yes: 0.193/ .074 = 2.61
            3. tohrs: No: 0.0014/0.0012 = 1.17
    c. Yes, it does matter which standard errors are used; however, when testing both ways, their appears no difference in statistical significance for the stated variables.

2. **Why does the hypothesis H_0 crsgpa = 1 make sense? Test this hypothesis against the two-sided alternative at the 5% level, using both standard errors. Describe your conclusions.**

a. It makes sense because you would assume that your GPA from all courses should have a positive correlation to your current GPA.
b. H_0: crsgpa = 1
c. H_1: crsgpa <> 1
    i. SE: reject at 5%: 0.9/0.175 = 5.14
    ii. RSE: reject at 5%: 0.9/0.166 = 5.42
d. The t statistics fall within the rejection area, which are above the critical value of 1.96.

3. **Test whether there is an in-season effect on term GPA, using both standard errors. Does the significance level at which the null can be rejected depend on the standard error used?**
    a. Using the prior model above, there appears not to be a difference. I am not sure what an "in season effect" means, so this answer may be meaningless.

Computer Exercise 4.C12

Use the data in ECONMATH to answer the following questions.

1. **Estimate a model explaining colgpa to hsgpa, actmth, and acteng. Report the results in the usual form. Are all explanatory variables statistically significant?**

	term	estimate	std.error	statistic	p.value
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	(Intercept)	0.0282	0.168	0.168	8.66e- 1
2	hsgpa	0.659	0.0530	12.4	1.47e-32
3	actmth	0.0130	0.00515	2.53	1.17e- 2
4	acteng	0.0122	0.00504	2.42	1.56e- 2

   a. At 95% level of significance, all variables are significant. P values very small, t statistics above 2.

2. **Consider an increase in hsgpa of one standard deviation, about .343. By how much does Formula increase, holding actmth and acteng fixed. About how many standard deviations would the actmth have to increase to change Formula by the same amount as a one standard deviation in hsgpa? Comment.**
   a. A 0.343-point increase in high school GPA leads to a (1.3714-0.938437) = 0.432963 increase in GPA.
   b. Both scores need to increase by over 9.9 points.

3. **Test the null hypothesis that actmth and acteng have the same effect (in the population) against a two-sided alternative. Report the p-value and describe your conclusions.**
   a. P value = .9253, which is significant. So reject the null hypothesis.

4. **Suppose the college admissions officer wants you to use the data on the variables in part (i) to create an equation that explains at least 50 percent of the variation in colgpa. What would you tell the officer?**
   a. Using the equation above, we can see that high school GPA influences college GPA the most.
   b. However, when implanting other combinations for factors leading to your total GPA, it appears impossible to create an equation that accounts for the entirety of your GPA. Advanced methods may have to be used in order to understand this relationship.

Computer Exercise 8.C4

Use VOTE1 for this exercise.

1. **Estimate a model with voteA as the dependent variable and prtystrA, democA, log(expendA), and log(expendB) as independent variables. Obtain the OLS residuals, u_i, and regress these on all of the independent variables. Explain why you obtain R^2.**

	term	estimate	std.error	statistic	p.value
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	(Intercept)	37.7	4.74	7.95	2.56e-13
2	prtystrA	0.252	0.0713	3.53	5.30e- 4
3	democA	3.79	1.41	2.70	7.72e- 3
4	lexpendA	5.78	0.392	14.7	4.03e-32
5	lexpendB	-6.24	0.397	-15.7	9.34e-35

   a. R^2 is 0 because the SSR's are minimized in the error term.

2. **Now, compute the Breusch-Pagan test for heteroskedasticity. Use the F statistic version and report the p-value.**
   a. BP = 9.0934, df = 4, <u>p-value = 0.05881</u>
   b. Reject at 5% significance level; accept at a 10% significance level.

3. **Compute the special case of the White test for heteroskedasticity, again using the F statistic form. How strong is the evidence for heteroskedasticity now?**
   a. BP = 5.49, df = 2, <u>p-value = 0.06425</u>
   b. Reject at 5% significance level; accept at a 10% significance level.

Computer Exercise 8.C7

Use the data in LOANAPP for this exercise.

1. **Estimate the equation in part (iii) of Computer Exercise C8 in Chapter 7, computing the heteroskedasticity-robust standard errors. Compare the 95% confidence interval on B_{white} with the nonrobust confidence interval.**

Term	(Intercept)	white	hrat	obrat	loanprc	unem	male	married	dep	sch	cosign	chist	pubrec	mortlat1	mortlat2	vr
Estimate	0.93700	0.12900	0.00183	-0.00543	-0.14700	-0.00730	-0.00414	0.04580	-0.00683	0.00175	0.00977	0.13300	-0.24200	-0.05730	-0.11400	-0.03140
SE	0.05270	0.01970	0.00126	0.00110	0.03750	0.00320	0.01890	0.01630	0.00670	0.01660	0.04110	0.01930	0.02820	0.05000	0.06700	0.01400
Robust-SE	0.05967	0.02593	0.00147	0.00134	0.03813	0.00373	0.01933	0.01727	0.00693	0.01718	0.03995	0.02469	0.04298	0.06723	0.09329	0.01450
Variance	**0.00697**	**0.00623**	**0.00021**	**0.00024**	**0.00063**	**0.00053**	**0.00043**	**0.00097**	**0.00023**	**0.00058**	**-0.00115**	**0.00539**	**0.01478**	**0.01723**	**0.02629**	**0.00050**

2. **Obtain the fitted values from the regression in part (i). Are any of them less than zero? Are any of them greater than one? What does this mean about applying weighted least squares?**
   a. Cosign is the only fitted value less than zero. There are none greater than one.
   b. This means that the using the weighted least squares could mean that some representations of variables will be excluded from the regression.

Exercise 8.1

Which of the following are consequences of heteroskedasticity?

1. The usual F statistic no longer has an F distribution.
2. The OLS estimators are no longer BLUE.