Assignment-3

1. Estimate one linear model with slope coefficients that have the following interpretation:

a. "Upper division courses on average are evaluated X units higher than lower division courses"

Table 1: Ratings by Division

	Dependent variable:
	Rating
Upper Division	-0.135^{**}
	(0.054)
Observations	463
R^2	0.013
Adjusted R^2	0.011
Residual Std. Error	0.552 (df = 461)
F Statistic	6.179^{**} (df = 1; 461)
Note:	*p<0.1; **p<0.05; ***p<0.01

b. "An additional student enrolled in the course increases the course's evaluation by X units"

Table 2: Ratings by Enrolment

	Dependent variable:
	Rating
Enrolment	-0.00001
	(0.0003)
Observations	463
R^2	0.00000
Adjusted R^2	-0.002
Residual Std. Error	0.555 (df = 461)
F Statistic	0.001 (df = 1; 461)
Note:	*p<0.1; **p<0.05; ***p<0.01

2. Explain why we might be skeptical that we are measuring the causal effect of these variables.

It may be – and is in fact probably the case – that the effects on ratings of enrolment and/or division are dependent on the instructor. There are instructor specific fixed effects. We may try to fix these issues through several means, many of which are listed in question 4. Including them as a dummy variable will remove the confounding problem, but it may not be enough. De-mean evaluation is probably more effective.

3. Now suppose that we add instructor-specific fixed effects to our model. What kind of endogeneity problems are we solving by doing this?

See (2.) Instructor probably causes both division and enrolment, and also ratings

Justification: “instructor” causes “ratings” because it should be that anything we missed the inclusion of in the data set – such as personality, agreeableness, etc. . . – that had to do with the instructor, is part of the instructor term. Thus, it causes ratings by whatever mechanisms we missed putting in the dataset.

4. Estimate this model with instructor-specific fixed effects three ways (these should all produce the same results if you cluster your standard errors at the instructor level):

a. Using “lm”, include dummy variables for each instructor (i.e. as.factor(prof))

See tables 3 and 4.

b. Using “lm”, de-mean eval, division, and allstudents before including them in the regression.

See tables 5 and 6.

I’ll need to talk with you about this one.

c. Using the “plm” library, estimate the model with the “plm” function.

See tables 7 and 8.

5. Explain why you cannot include a variable like “beauty” in these models. Is this a problem if all you want to do is comment on the coefficients on eval and allstudents?

It’s to do with how the data was collected. Who collected the data? What was their subjective view of attractiveness? Were they a representative sample of students as a whole? How does beauty interact with the other variables? Does age factor in? Could it be that their view of the teacher’s beauty is in part caused by their view of how well they are doing as a teacher? So many questions, so few answers. I’d never include this in a model.

As for the second part, it really depends on how the data was collected. If the class was larger, maybe the participants of the data collection couldn’t get a good look at the professor. Who knows? I don’t, so I wouldn’t include it.

6. What kind of endogeneity problems have we not solved by including instructor-specific fixed effects?

For the case of the two models covered, instructor-specific fixed effects do not account for any endogeneity which does not have to do with the instructor. It could be that there is endogeneity between course evaluations, and division or enrollment. This could not be accounted for by instructor fixed-specific effects.

Table 3: Ratings by Division

	Dependent variable:
	Rating
Upper Division	0.160^{**}
	(0.071)
Observations	463
R^2	0.568
Adjusted R^2	0.458
Residual Std. Error	0.409 (df = 368)
F Statistic	5.145^{***} (df = 94; 368)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 4: Ratings by Enrolment

	Dependent variable:
	Rating
Enrolment	-0.003^{***}
	(0.001)
Observations	463
R^2	0.599
Adjusted R^2	0.496
Residual Std. Error	0.394 (df = 368)
F Statistic	5.837^{***} (df = 94; 368)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5: Ratings by Division

	Dependent variable:
	Rating
Upper Division	-0.135^{**}
	(0.054)
Observations	463
R^2	0.013
Adjusted R^2	0.011
Residual Std. Error	0.552 (df = 461)
F Statistic	6.179^{**} (df = 1; 461)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: Ratings by Enrolment

	Dependent variable:
	Rating
Enrolment	-0.00001
	(0.0003)
Observations	463
R^2	0.00000
Adjusted R^2	-0.002
Residual Std. Error	0.555 (df = 461)
F Statistic	0.001 (df = 1; 461)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Ratings by Division

	Dependent variable:
	Rating
Upper Division	0.160^{**}
	(0.071)
Observations	463
R^2	0.014
Adjusted R^2	-0.238
F Statistic	5.058^{**} (df = 1; 368)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Ratings by Enrolment

	Dependent variable:
	Rating
Enrolment	-0.003^{***}
	(0.001)
Observations	463
R^2	0.084
Adjusted R^2	-0.151
F Statistic	33.550^{***} (df = 1; 368)
Note:	*p<0.1; **p<0.05; ***p<0.01