

dummies and interactions

Adam Okulicz-Kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Tuesday 13th February, 2018 16:48

outline

midterm: bimodal distribution

- ◇ lets very briefly go over posted ans
- ◇ for more detailed explanations lets meet outside of the class
- ◇ especially those of you who did less than fabulous!
 - you should rework it and make sure you arrive at right ans
 - and do ask questions, stop by my office etc
 - it is important that you dont fall behind on this
 - again this is qualifying exam material!

this class is about research, and so is PhD!

- ◇ you do phd to do research and research only!
 - otherwise it doesnt make sense!
- ◇ if you want to *do* (carry out) public administration, community development etc, just get BA/MA
- ◇ if you want to research them and other things do PhD
- ◇ typically, successful PhDs become academics or researchers
- ◇ obtaining PhD and not doing research is failure and waste of time
- ◇ if you want to be manager, leader etc, just get MBA!!

intuition

- ◇ dummies and interactions are fun !
- ◇ this is one of the most interesting things in regression
- ◇ you can test some interesting hypotheses
 - and you can contribute to the literature

what is it?

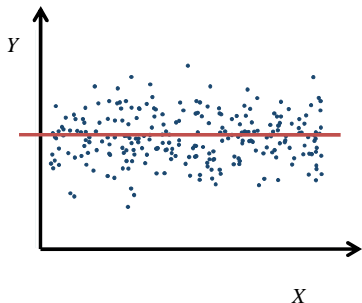
- ◇ dummies identify nominal or ordinal characteristics, such as gender, race, region, religion, or education (measured as highest degree attained)
- ◇ dummies are binary indicators of a specific attribute
 - you either have the attribute or you do not
 - 1 if the condition is true and 0 otherwise
 - say male dummy=1 if a guy; 0 if a girl
- ◇ never name a dummy like 'gender', which is impossible to figure out what it means!

what is it?

- ◇ can use dummies to create separate intercepts and/or slopes for subgroups within one regression
- ◇ dummies must always be interpreted relative to
 - “base case,” “omitted category”, “reference group”

regression on constant only

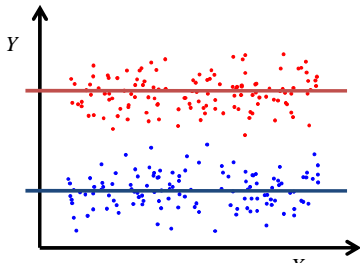
- ◇ $\hat{\beta}_2 = 0$
- ◇ $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = \bar{Y}$



- ◇
- ◇ remember sums of squares discussions?
 - our best bet before regression
 - our best prediction of y is mean of y

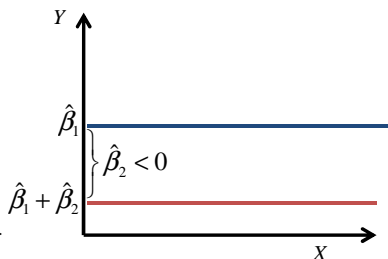
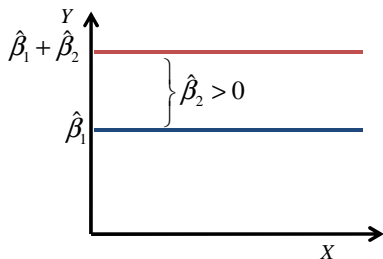
now add a dummy

- ◇ $Y_i = \beta_1 + \beta_2 \text{female}_i + u_i$
- ◇ if $\text{female}_i = 1$ $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2(1) = \hat{\beta}_1 + \hat{\beta}_2$
 - $E[Y | \text{female} = 1] = \hat{\beta}_1 + \hat{\beta}_2$
- ◇ if $\text{female}_i = 0$ $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2(0) = \hat{\beta}_1$
 - $E[Y | \text{female} = 0] = \beta_1$
- ◇ hence, β_2 is the difference between \bar{Y} for males and females



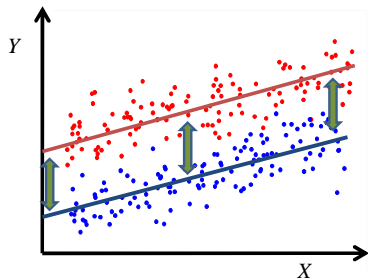
schematic

- ◇ $Y_i = \beta_1 + \beta_2 \text{female}_i + u_i$
- ◇ $\hat{\beta}_2 = \bar{Y}_{\text{female}} - \bar{Y}_{\text{male}}$
- ◇ this is like a t-test!



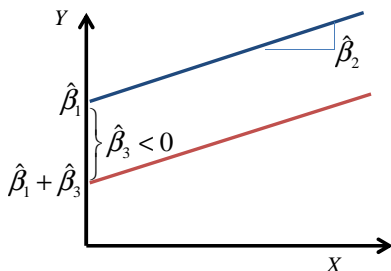
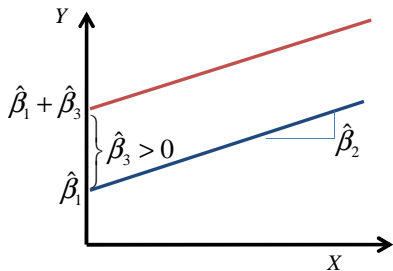
and add a continuous var

- ◇ $Y_i = \beta_1 + \beta_2 X_i + \beta_3 \text{female}_i + u_i$
- ◇ if $\text{female}_i = 1$ $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3(1) = (\hat{\beta}_1 + \hat{\beta}_3) + \hat{\beta}_2 X_i$
- ◇ if $\text{female}_i = 0$ $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3(0) = (\hat{\beta}_1) + \hat{\beta}_2 X_i$

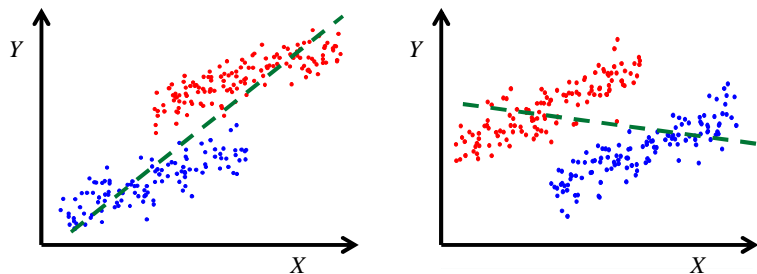


schematic

◇ $Y_i = \beta_1 + \beta_2 X_i + \beta_3 \text{female}_i + u_i$



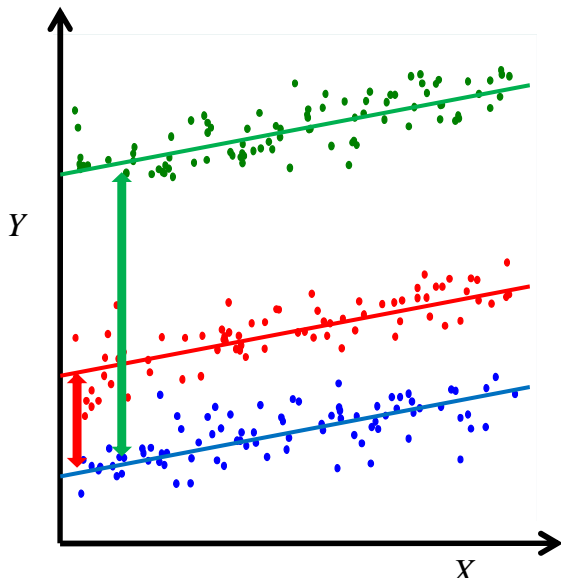
bias from omitting a dummy...



- ▶ difficult to know if there should be a dummy or not
- ▶ can experiment; and again use theory! it will tell you!

ordinal vars: blackboard: asst assoc full prof

◇ omit one category (ie pick base case)

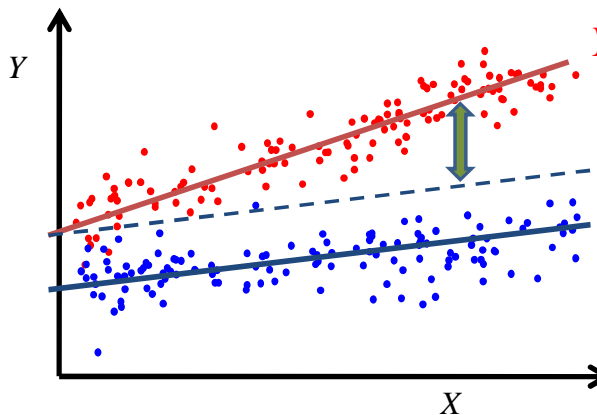


choosing the base case

- ◇ be meaningful, eg pick typical or default situation
 - eg in my paper on wrk hrs I picked 40,
 - and dummies are then relative to the typical case
- ◇ think about what hypotheses you are most interested in
- ◇ remember that a different base case can change which coefficients are significant!
- ◇ make your choice(s) clear in your tables **and** text

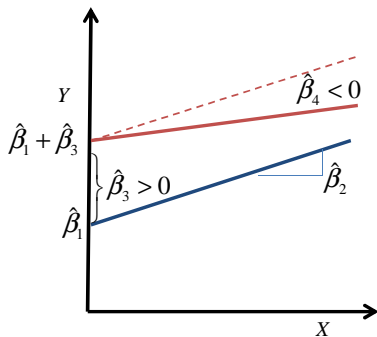
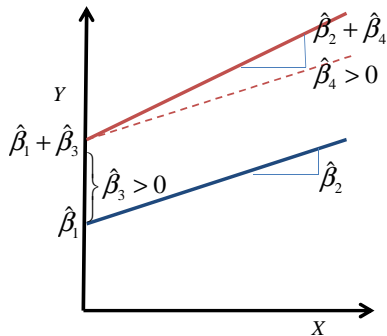
continuous/dummy interactions

◇ $Y_i = \beta_1 + \beta_2 X_i + \beta_3 \text{female}_i + \beta_4 \text{female}_i * X_i + u_i$



schematic

◇ $Y_i = \beta_1 + \beta_2 X_i + \beta_3 \text{female}_i + \beta_4 \text{female}_i * X_i + u_i$



interaction of dummies

- ◇ if there is an interaction effect between two variables, the effect of one variable depends on the level of the other
- ◇ eg the effect of marriage on wage depends on gender.
- ◇ interactions go both ways:
 - the effect of gender depends on marital status, too

interaction of dummies

◇ $Y_i = \beta_1 + \beta_2 \text{female} + \beta_3 \text{married} + \beta_4 \text{female} * \text{married} + u_i$

	Male	Female	Gender Difference
Unmarried	$\hat{\beta}_1$	$\hat{\beta}_1 + \hat{\beta}_2$	$\hat{\beta}_2$
Married	$\hat{\beta}_1 + \hat{\beta}_3$	$\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4$	$\hat{\beta}_2 + \hat{\beta}_4$
Effect of Marriage	$\hat{\beta}_3$	$\hat{\beta}_3 + \hat{\beta}_4$	$\hat{\beta}_4$

◇

example [let's calc tab from reg]

```
. table married female, c(mean wage) row col f(%7.2f)
```

Married	Gender		Total
	male	female	
no	8.35	8.26	8.31
yes	10.88	7.68	9.40
Total	9.99	7.88	9.02

```
. gen femxmar = female*married  
. reg wage female married femxmar
```

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
$\hat{\beta}_2$ female	-.0951892	.7350367	-0.13	0.897	-1.539132	1.348754
$\hat{\beta}_3$ married	2.521222	.6120814	4.12	0.000	1.318819	3.723626
$\hat{\beta}_4$ femxmar	-3.09704	.9072785	-3.41	0.001	-4.879344	-1.314737
$\hat{\beta}_1$ _cons	8.354677	.4936728	16.92	0.000	7.384882	9.324473

dummy practice

- ◇ in addition to the dofile, see the links on the website for the code
- ◇ let's especially focus on the dummy variables
- ◇ we'll do it in the class if we have time...

interactions of continuous variables

- ◇ $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 (X_{2i} X_{3i}) + u_i$
- ◇ $\frac{\Delta Y_i}{\Delta X_{2i}} = \beta_2 + \beta_4 X_{3i}$
- ◇ $\frac{\Delta Y_i}{\Delta X_{3i}} = \beta_3 + \beta_4 X_{2i}$

interactions links

- ◇ again, interactions are a great way to contribute

- ◇ see sections 3.7 and 3.8

- `http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter3/statareg3.htm`

- ◇ `http://www.stata.com/support/faqs/stat/anoregcoef.html`

- ◇ `http://nd.edu/~rwilliam/stats2/l51.pdf`

- ◇ `http://www.stata.com/manuals13/rmarginsplot.pdf`: scroll down to examples

- it is terrific command!