dummies and interactions

`adam.okulicz.kozaryn@gmail.com`

this version: Tuesday 13th February, 2024   19:04

outline

fa24

$\diamond$ note i beefed up ps3 with dummies and interactions

intuition

◇ fun !

◇ test new hypotheses and contribute to literature

what is it?

$\diamond$ dummies identify nominal or ordinal characteristics, eg:

$\circ$ gender, race, region, rel, edu (highest degree attained)

$\diamond$ dummies are binary indicators

$\circ$ 1 if the condition is true and 0 otherwise

$\circ$ say 'male' dummy=1 if a guy; 0 if a girl

$\diamond$ not 'gender'!: impossible to figure out what it means!

$\diamond$ dummies create separate lines (intercepts and slopes) for subgroups within one regression

$\diamond$ dummies must always be interpreted relative to "base case," "omitted category", "reference group"
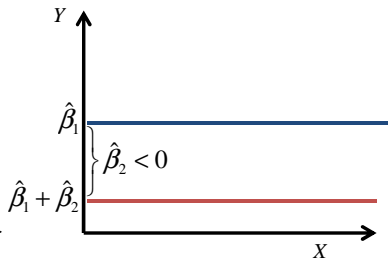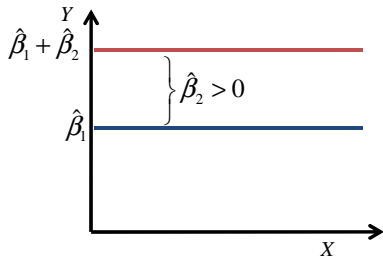
dummy

◇ $Y_i = \beta_1 + \beta_2 female_i + u_i$

◇ if $female_i = 1$ $\quad \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2(1) = \hat{\beta}_1 + \hat{\beta}_2$

○ $E[Y|female = 1] = \hat{\beta}_1 + \hat{\beta}_2$

◇ if $female_i = 0$ $\quad \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2(0] = \hat{\beta}_1$

○ $E[Y|female = 0] = \beta_1$

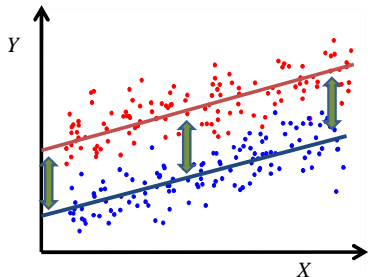◇ β_2: how much more or less female make relative to male

◇
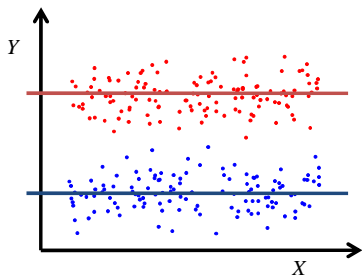
schematic

$\diamond\ Y_i = \beta_1 + \beta_2 female_i + u_i$

$\diamond\ \hat{\beta}_2 = \bar{Y}_{female} - \bar{Y}_{male}$
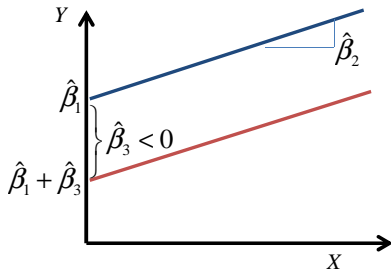
$\diamond$ this is like a t-test!

$\diamond$

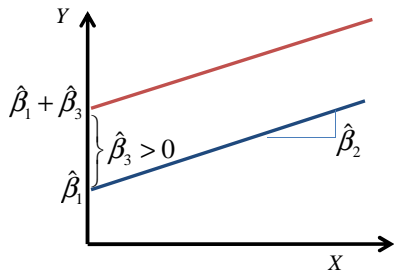## and add a continuous var

◇ $Y_i = \beta_1 + \beta_2 X_i + \beta_3 female_i + u_i$

◇ if $female_i = 1$ $\quad \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3(1) = (\hat{\beta}_1 + \hat{\beta}_3) + \hat{\beta}_2 X_i$

◇ if $female_i = 0$ $\quad \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3(0) = (\hat{\beta}_1) + \hat{\beta}_2 X_i$
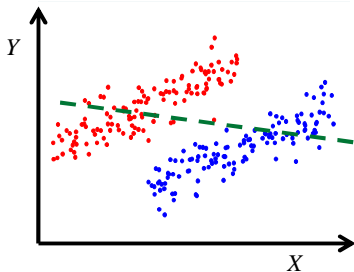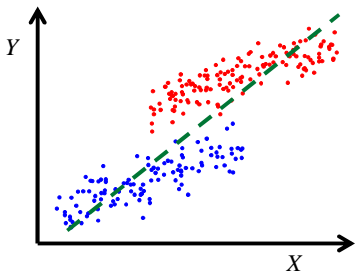

◇

schematic

$\diamond$ $Y_i = \beta_1 + \beta_2 X_i + \beta_3 female_i + u_i$
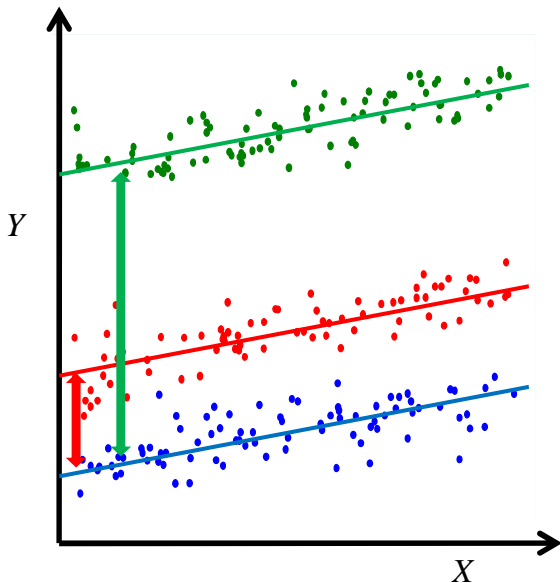

$\diamond$

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

$\diamond$ be meaningful, eg pick typical or default situation

$\circ$ eg in my paper on wrk hrs I picked 40,

$\circ$ and dummies are then relative to the typical case

$\diamond$ think about what hypotheses you are most interested in

$\diamond$ base case affects coefficients' significance

$\circ$ eg 40-50 hrs may not be significant relative to 40, but
would be significant relative to ¡20

$\diamond$ make your choice(s) clear in your tables **and** text

continuous/dummy interactions

◇ $Y_i = \beta_1 + \beta_2 X_i + \beta_3 female_i + \beta_4 female_i * X_i + u_i$

◇
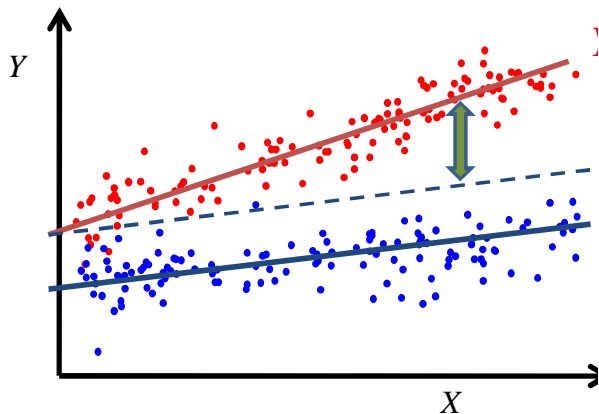
schematic

$\diamond$ $Y_i = \beta_1 + \beta_2 X_i + \beta_3 female_i + \beta_4 female_i * X_i + u_i$

$\diamond$
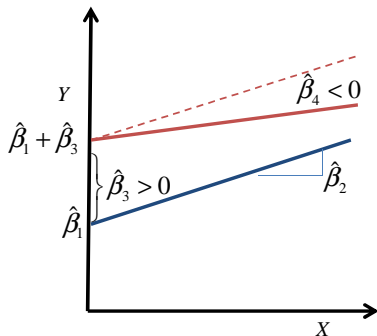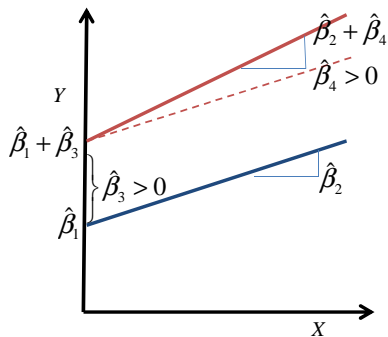
interaction of dummies

$\diamond$ if there is an interaction effect between two variables, the effect of one variable depends on the level of the other

$\diamond$ eg the effect of marriage on wage depends on gender.

$\diamond$ interactions go both ways:

$\circ$ the effect of gender depends on marital status, too

interaction of dummies

$\diamond$ $Y_i = \beta_1 + \beta_2 female + \beta_3 married + \beta_4 female * married + u_i$

	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	$\hat{\beta}_3$	$\hat{\beta}_3 + \hat{\beta}_4$	$\hat{\beta}_4$

$\diamond$

example [let's calc tab from reg]

```
. table married female, c(mean wage) row col f(%7.2f)
----------------------------------
          |        Gender
  Married |   male  female   Total
----------+-----------------------
       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`

```
-------------------------------------------------------------------------------
```

interactions of continuous variables

$\diamond$ $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 (X_{2i} X_{3i}) + u_i$

$\diamond$ $\frac{\Delta Y_i}{\Delta X_{2i}} = \beta_2 + \beta_4 X_{3i}$

$\diamond$ $\frac{\Delta Y_i}{\Delta X_{3i}} = \beta_3 + \beta_4 X_{2i}$

interactions links

◇ again, interactions are a great way to contribute

◇ see sections 3.7 and 3.8 `https://stats.idre.ucla.edu/stata/webbooks/reg/`

   `chapter3/regression-with-statachapter-3-regression-with-categorical-predictors/`

◇ `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

○ marginsplot is a terrific command!