Class 3, Bivariate Regression

Adam Okulicz-Kozaryn

adam.okulicz.kozaryn@gmail.com

(note: for official utd business use ajo021000@utdallas.edu)

this version: Monday 9th March, 2015 17:40

outline

misc

hypothesis testing

measurement

outline

misc

hypothesis testing

measurement

today and looking ahead

- ◇ much of today's class is the repetition of the last class
- ◇ and we are adding hypothesis testing
- ◇ next week we will start multiple regression
  - · class material will get more applied
  - · we will have more examples
  - · and more programming

paper?

- how is the paper going ?
- again, start early
- if you do not know where to start – email me...

basic calculations again

- let's do some basic calculations again
- here we will cover some of the ps1
- and later demonstrate hypothesis testing on ps1 data

basic calculations [blackboard;dofile:ps1]

Y	X	y	y2	x	x2	xy
1	17					
3	13					
5	8					
7	10					
9	2					
Sum:						
25	50					

$\bar{Y} = 5 \quad \bar{X} = 10$

predicted values and residuals

[blackboard:dofile:ps1]

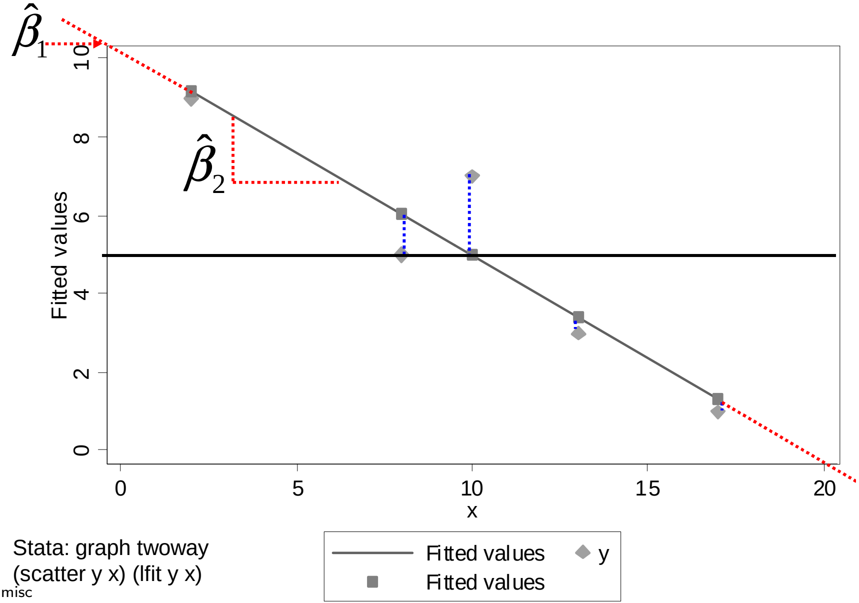
Y	X	Y_hat	e	e^2
1	17			
3	13			
5	8			
7	10			
9	2			

- $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$
- for obs 1:
- $\hat{Y}_1 = 10.24 + (-0.524)(17) = 1.332$
- $e_1 = 1 - 1.33 = -0.33$
- let's calculate TSS and Rsq and p-value! (this is your ps)

the coefficients–interpretation [dofile:ps1]

- Beta hat two is the slope coefficient. Thus, a one unit change in X leads to a 0.524 decrease in Y. Beta hat one is the intercept term. It is the predicted value for Y when X is equal to zero.

regression plot again



Stata: graph twoway
(scatter y x) (lfit y x)

sum of squared residuals

- $\sum e_i = 0$
- $\sum e_i^2 = 5.42$
- $s = \sqrt{\frac{\sum e_i^2}{n-2}} =$
- The sum of the squared residuals is the quantity that was minimized. s is the estimate of sigma, the standard deviation of the disturbance terms under the assumption of homoskedastic (constant variance) disturbance terms.

standard error of the slope

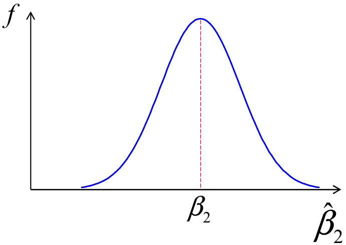
- $s_{\hat{\beta}_2} = \frac{s}{\sqrt{\sum x_i^2}}$
- This is the standard error of the slope coefficient, and gives us information about the reliability (like sd or se) of our estimate of the slope. We can use it to construct confidence intervals or conduct hypothesis tests.

coefficient distribution

- we know the mean and variance of the estimator
- but what about the shape of the sampling distribution of the slope coefficient?
- If the disturbance term is normally distributed, then the coefficients also have a normal distribution. That is: $u_i \sim N(0, \sigma^2) \rightarrow$ If the distribution of the disturbance term is not normal or unknown, the distribution of the OLS estimators is still normal in large samples - "asymptotically".
- so you want to use big samples!

sampling distribution of the slope

The probability distribution $\hat{\beta}_2$ is centered on the true value of the parameter (i.e. unbiased) and is normally distributed with variance estimated by:



- $s^2_{\hat{\beta}_2} = \frac{s^2}{\sum x_i^2}$
- $s_{\hat{\beta}_2} = \sqrt{\frac{s^2}{\sum x_i^2}} = \frac{s}{\sqrt{\sum x_i^2}}$
- $t = \frac{\hat{\beta}_2 - \beta_2}{s_{\hat{\beta}_2}} = \frac{\hat{\beta}_2}{s_{\hat{\beta}_2}}$
- CI: $\hat{\beta}_2 \pm (t_{n-2,\frac{\alpha}{2}})(s_{\hat{\beta}_2})$
- $H_0 : \beta_2 = 0 \; H_A : \beta_2 \neq 0$

outline

misc

hypothesis testing

measurement

confidence intervals

- In general, a confidence interval is the point estimator plus or minus a margin of error, which consists of a distribution parameter (z or t) times the standard error of the estimator. In this case (small sample, σ unknown, we use the t distribution.
- $PE \pm (t_{\frac{\alpha}{2},DOF})(SE) = \hat{\beta}_2 \pm t_{0.025,3} s_{\hat{\beta}_2}$

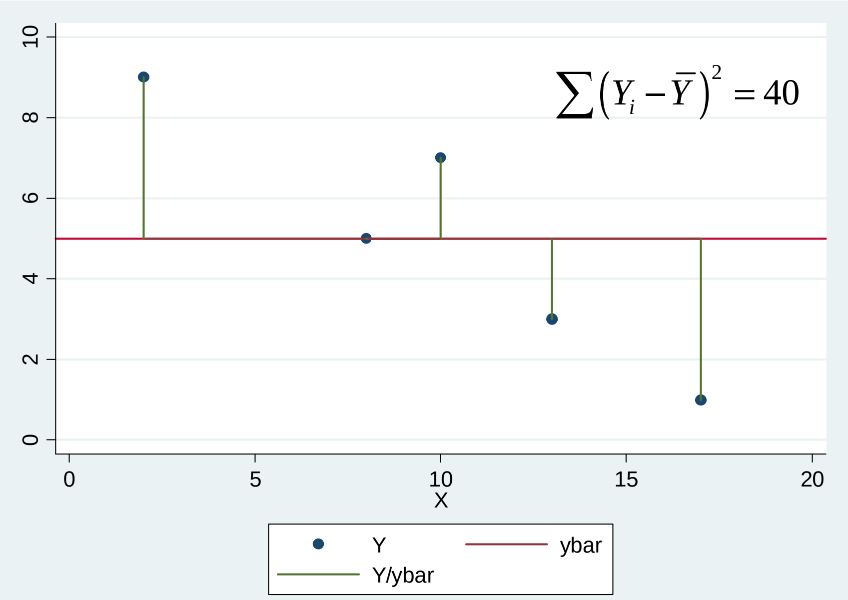
hypothesis test

- the null is that slope ("the onobserved true parameter")is zero (i.e. no effect)
- $H_0 : \beta_2 = 0$
- $H_A : \beta_2 \neq 0$
- $t = \frac{\hat{\beta}_2 - \beta_2}{s_{\hat{\beta}_2}}$
- What if our null hypothesis was that the slope was -1? Can you think of an example where a null hypothesis other than 0 would be appropriate?

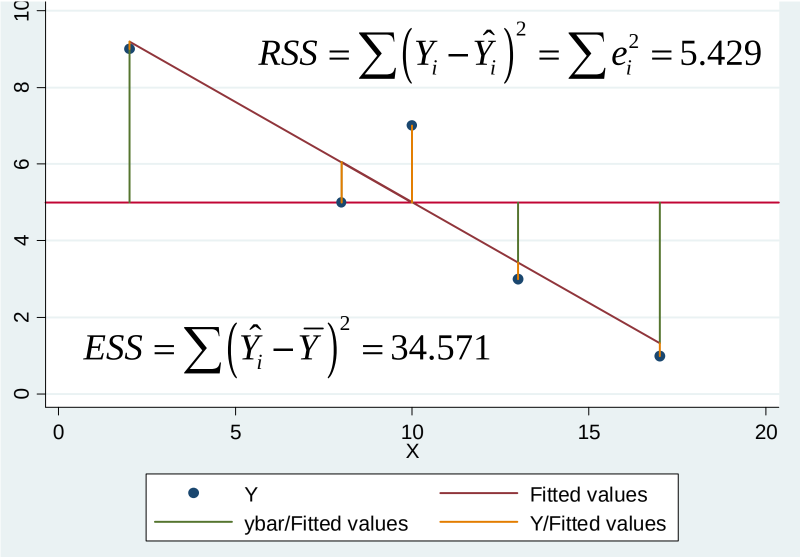
partitioning the variance in Y

- ◇ before regression $E[Y_i] = \bar{Y}$
  - $TSS = \sum(Y_i - \bar{Y})^2 = \sum y_i^2 = 40$
- ◇ after regression $E[Y_i|X_i] = \hat{Y}_i$
  - $RSS = \sum(Y_i - \hat{Y}_i)^2 = \sum e_i^2 = 5.43$
  - $ESS = TSS - RSS = 40 - 5.4 = 34.57$

yet another look at tss



yet another look at rss



R^2 again

- $R^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2}$
- The R2 is the proportion of the total variance in the Y variable that is explained by the model. However, do not overestimate the importance of this statistic. We will see (when we move on to multiple regression) that it can be misleading. Notes:
- $0 \leq R^2 \leq 1$
- $R^2 = (r_{xy})^2$ for bivariate case only

exercise 1

- you regressed car's price on its weight

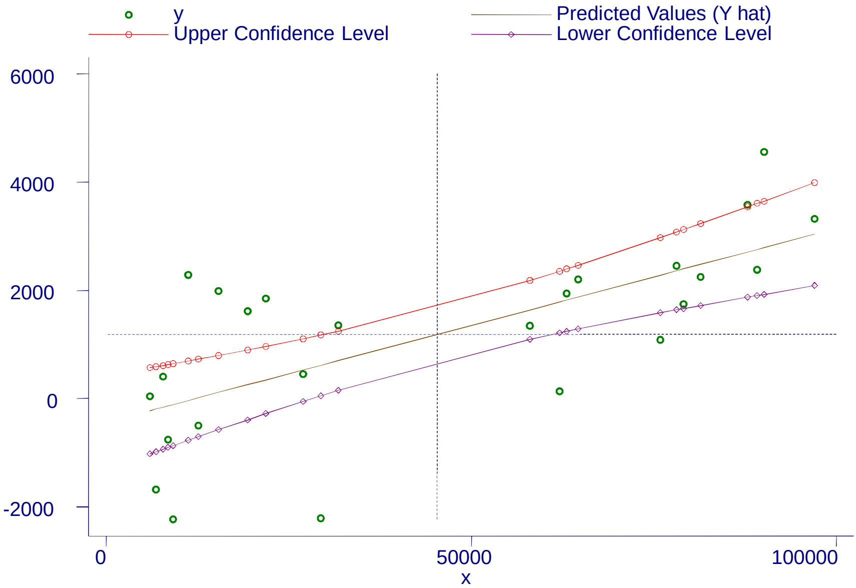
```
------------------------------------------
       price |      Coef.   Std. Err.
-------------+----------------------------
      weight |   2.044063    .3768341
```

- interpret the coefficient
- is it significant ?
- calculate 95% CI
- (check with answer in the dofile)

reliability of predict. val. (se of of $E(Y|X)$)

- We have discussed the fact that parameter estimates are random variables, and so they have standard errors. Predicted values are also random variables because they are linear combinations of the coefficients.
- The further from the mean of X, the wider the confidence interval around the predicted value.
- $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ why no residual in this equation?
- $var(\hat{Y}_0|X_0) = \sigma^2(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(X_i - \bar{X})^2})$

another illustration...



outline

misc

hypothesis testing

measurement

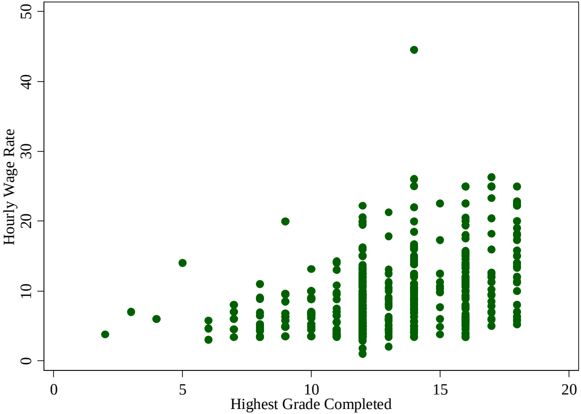
intuition

- what happens to the regression estimates if we change the measurement of the variables
- but first let's look at the relationship between correlation coefficient and regression coefficient

correlation vs regression

- $r = \frac{\sum y_i x_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}}$ $\hat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2}$
- bivariate slope equals corr coef scaled by std dev of Y and X:
  $\hat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2} = r\left(\frac{s_Y}{s_X}\right)$

education and wages [dofile:wages]




```
. corr wage educ
(obs=534)

             |     wage     educ
-------------+------------------
        wage |   1.0000
        educ |   0.3819   1.0000
```

```
. sum wage educ
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       534    9.023939    5.138876          1       44.5
        educ |       534    13.01873    2.615373          2         18
```

education and wages [dofile:wages]

```
. regress wage educ

  Source |       SS       df       MS              Number of obs =     534
---------+------------------------------           F(  1,   532) =   90.86
   Model |  2053.22494     1  2053.22494           Prob > F      =  0.0000
Residual |  12022.2635   532  22.5982396           R-squared     =  0.1459
---------+------------------------------           Adj R-squared =  0.1443
   Total |  14075.4884   533  26.4080458           Root MSE      =  4.7538

------------------------------------------------------------------------------
    wage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .7504488     .07873      9.532   0.000       .5957891    .9051086
   _cons |   -.745949   1.045404     -0.714   0.476      -2.799576    1.307678
------------------------------------------------------------------------------
```

The estimated regression line:

$$\widehat{wage}_i = \hat{\beta}_1 + \hat{\beta}_2 educ_i = -0.75 + 0.75 educ_i$$

Interpret the coefficients.

anatomy of stata output

```
. regress DV IV
```

Source	SS	df	MS
Model	$ESS = \sum\left(\hat{Y}_i - \bar{Y}\right)^2$	1	....
Residual	$RSS = \sum e_i^2$	$n-2$	$s^2 = \dfrac{RSS}{n-2}$
Total	$TSS = \sum\left(Y_i - \bar{Y}\right)^2$	$n-1$	$s_Y^2 = \dfrac{TSS}{n-1}$

Number of obs	=	n
$\mathrm{F}(1, n-2)$	=	....
Prob > F	=	....
R-squared	=	r^2
Adj R-Squared	=	....
Root MSE	=	s

DV	Coef.	Std.Err.	t	P>|t|	[95% Conf.	Interval]
IV	$\hat{\beta}_2$	$s_{\hat{\beta}_2}$	$\left(\dfrac{\hat{\beta}_2}{s_{\hat{\beta}_2}}\right)$	p val. for H_0 that $\beta_2 = 0$	$\hat{\beta}_2 - t_{0.025} s_{\hat{\beta}_2}$	$\hat{\beta}_2 + t_{0.025} s_{\hat{\beta}_2}$
Intercept	$\hat{\beta}_1$	$s_{\hat{\beta}_1}$	$\left(\dfrac{\hat{\beta}_1}{s_{\hat{\beta}_1}}\right)$	p val. for H_0 that $\beta_1 = 0$	$\hat{\beta}_1 - t_{0.025} s_{\hat{\beta}_1}$	$\hat{\beta}_1 + t_{0.025} s_{\hat{\beta}_1}$

some examples

- $\hat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2} = r\left(\frac{s_Y}{s_X}\right)$
- if x is packs instead of cigarettes per capita, how does s_X change? how does the slope change?
- if y is deaths per 1,000 instead of deaths per million, how does s_Y change? how does the slope change?
- if both x and y have a variance of 1 (e.g. z scores), what is the slope coefficient?

shifting X by a constant c

- $Y_i = \beta_1 + \beta_2 X_i^{'} + u_i \quad X_i^{'} = X_i + c$
- The mean of X changes, but the deviations from the mean do not.
- $\bar{X}^{'} = \frac{\sum(X_i + c)}{n} = \bar{X} + c$
- $x_i^{'} = (X_i^{'} - \bar{X}^{'}) = [(X_i + c) - (\bar{X} + c)] = x_i$
- Thus, the slope is unaffected. The intercept shifts to reflect in the opposite direction of the constant.
- $\hat{\beta}_2 = \frac{\sum y_i x_i^{'}}{\sum x_i^{'2}} = \frac{\sum y_i x_i}{\sum x_i^2}$
- $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}^{'} = \bar{Y} - \hat{\beta}_2(\bar{X} + c) = \bar{Y} - \hat{\beta}_2 \bar{X} - \hat{\beta}_2 c$
- If X is shifted by subtracting $\bar{X}$ $(e.g. \bar{X} = c)$: $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} - \hat{\beta}_2(\bar{X}) = \bar{Y}$

shifting Y by a constant d

- $Y_i^{'} = \beta_1 + \beta_2 X_i + u_i \quad Y_i^{'} = Y_i + d$
- The mean of Y changes, but the deviations from the mean do not.
- $\bar{Y}^{'} = \frac{\sum(Y_i + d)}{n} = \bar{Y} + d$
- $y_i^{'} = (Y_i^{'} - \bar{Y}^{'}) = [(Y_i + d) - (\bar{Y} + d)] = y_i$
- Again, the slope is unaffected. The intercept shifts in the same direction and amount as the constant.
- $\hat{\beta}_2 = \frac{\sum y_i^{'} x_i}{\sum x_i^2} = \frac{\sum y_i x_i}{\sum x_i^2}$
- $\hat{\beta}_1 = \bar{Y}^{'} - \hat{\beta}_2 \bar{X} = (\bar{Y} + d) - \hat{\beta}_2 \bar{X} = \bar{Y} - \hat{\beta}_2 \bar{X} + d$

regression on Z scores

- $z_{Yi} = \beta_1 + \beta_2 z_{Xi} + u_i$
  $z_{Xi} = \frac{X_i - \bar{X}}{s_X} = \frac{x_i}{s_x}$
  $z_{Yi} = \frac{Y_i - \bar{Y}}{s_U} = \frac{y_i}{s_Y}$
- z scores always have a mean of 0 and a variance (and standard deviation) of 1
- $\hat{\beta}_2 = r_{Z_Y Z_X} \frac{s_{Z_Y}}{s_{Z_X}} = r_{YX}$
  $\hat{\beta}_1 = \bar{z}_Y - \hat{\beta}_2 \bar{z}_X = 0 - r(0) = 0$
- Thus, a regression of the z scores of Y on the z scores of X produces a slope equal to the correlation coefficient of X and Y and a zero intercept.

exercise 2 [do it at home; see dofile for ans]

- confirm the above by using data from ps1 in stata
- run regression of Y on X
- add 3 to X (gen new var) and repeat regression
- load original data, substruct 3 from Y, and repeat regression
- load original data and run regression on Z scores

interpretation: transforming variables

- Linear: One unit change in X leads to a β_2 unit change in Y.
- Log-Lin: One unit change in X leads to a $100 * \beta_2$ % change in Y. (guj ed4:p180 fig6.4; ed5:p163 ex6.4)
- Lin-Log: One percent change in X leads to a $\beta_2/100$ unit change in Y. (guj: ed4:p182 fig6.5; ed5:p165-6 ex6.5)
- Log-Log (aka log-linear or "linear in logs"): One percent change in X leads to a β_2 % change in Y (elasticity).
- These interpretations are valid for small changes (coefficients less than 0.10, i.e 10%).
- for interpretation and examples see

`http://www.ats.ucla.edu/stat/mult_pkg/faq/general/log_`

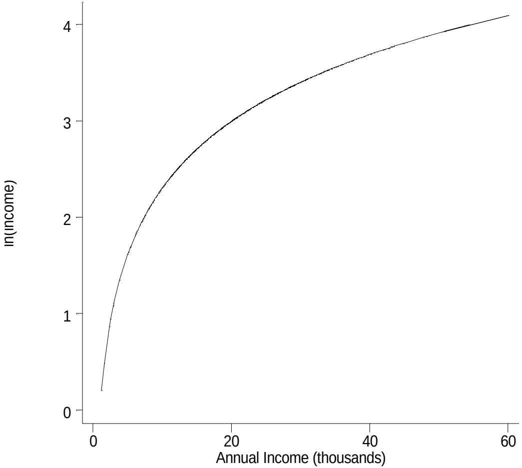
review of logarithms

- $y = log_b(x) \rightarrow x = b^y$
  - e.g. $log_2(8) = 3$ because $2^3 = 8$
  - $log_{10}100 = 2$ because $10^2 = 100$
- $ln\ x = log_e x; \quad e = 2.72...$
  - e.g. $e^{0.5} = 1.65; \quad ln\ 1.65 = 0.5$
- $ln(xy) = ln(x) + ln(y)$
- $ln(\frac{x}{y}) = ln(x) - ln(y)$
- $ln(x^n) = n\ ln(x)$
- $ln(e^x) = x\ ln(e) = x$

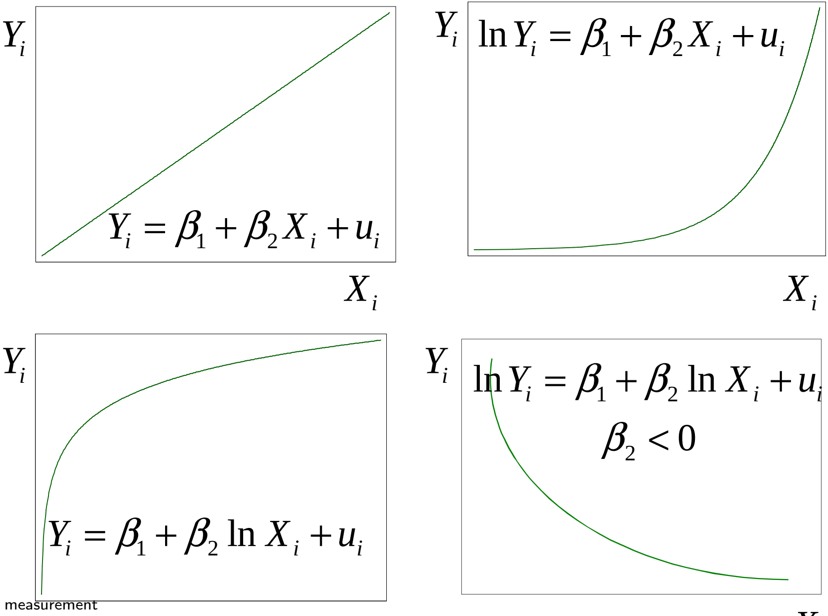
logarithms and relative change

For small changes, the change in ln(x) is the percentage change in x. E.g. a 0.05 increase in ln(x) is a 5 percent increase in x. Example:

x	ln(x)	chg
100	4.6052	
105	4.6540	0.0488
12,345	9.4210	
12,962	9.4698	0.0488



it makes a difference



lin-lin [dofile: measurment]

- e.g. people with more education earn higher wages...
- $Y_i = \beta_1 + \beta_2 X_i + u_i \;\; \beta_2 = \frac{\Delta Y_i}{\Delta X_i}$
- This model specifies that the change is constant regardless of the level of X (because β is constant)
- $wage_i = \beta_1 + \beta_2 educ_i + u_i$
  - $\widehat{wage}_i = -0.75 + 0.75 educ_i$
  - $\widehat{wage}_{10} = \$6.75 \quad \widehat{wage}_{11} = \$7.50 \quad \Delta\widehat{wage} = \$0.75$
    The change is the same for any 1 year change in educ.

relative change: log-lin

- We estimate this model by taking the natural log of Y first, then regression lnY on X. The regression treats lnY the same as any other variable.
- $lnY_i = \beta_1 + \beta_2 X_i + u_i$
- $\beta_2 = \frac{\Delta(lnY_i)}{\Delta X_i} = \frac{\Delta Y_i / Y_i}{\Delta X_i} = \frac{\text{relative change in} Y_i}{\text{unit change in} X_i}$

  ...for small changes. The percent change in Y per unit change in X is 100*β_2 times the unit change in X. This is still a linear regression, but with a new dependent variable (lnY). It is not linear in terms of Y.

e.g. log-lin [dofile: measurment]

- $ln(wage_i) = \beta_1 + \beta_2 educ_i + u_i$
- $\widehat{ln(wage)}_i = 1.06 + 0.08 educ_i$
- $\widehat{ln(wage)}_{10} = 1.06 + 0.08(10) = 1.86$
- This is the predicted ln(wage). But what about the predicted wage?
- $\widehat{wage}_{10} = e^{1.86} = \$6.42$
- $\widehat{wage}_{11} = e^{1.94} = \$6.96$
- $\%\Delta\widehat{wage}_{10\rightarrow 11} = \frac{\$6.96-\$6.42}{\$6.42} = 0.08 = 8\%$

the change varies in dollar terms

- But let's examine the change in wage for an additional year of graduate school, e.g. master's degree years.
- $\widehat{ln(wage)}_i = 1.06 + 0.08 educ_i$
- $\widehat{ln(wage)}_{17} = 1.06 + 0.08(17) = 2.42 \quad \widehat{wage}_{17} = \$11.25$
- $\widehat{ln(wage)}_{17} = 1.06 + 0.08(18) = 2.50 \quad \widehat{wage}_{18} = \$12.18$
- $\%\Delta\widehat{wage}_{17\to 18} = 0.08 = 8\%$
- The change in relative (percentage) terms is constant at 0.08 (8 percent), but the dollar change is larger.

lin-log [dofile: measurment]

- $Y_i = \beta_1 + \beta_2 lnX_i + u_i$
- In this model, we generate the natural log of education and regress dollar wage on the log of education
- $\beta_2 = \frac{\Delta Y_i}{\Delta lnX_i} = \frac{\Delta Y_i}{\Delta X_i / X_i} = \frac{\text{unit change in} Y_i}{\text{relative change in} X_i}$

relative change in education

educ	%change	educ	%change
1		11	10%
2	100%	12	9%
3	50%	13	8%
4	33%	14	8%
5	25%	15	7%
6	20%	16	7%
7	17%	17	6%
8	14%	18	6%
9	13%	19	6%
10	11%	20	5%

The relative change in education per year is declining because the base is getting larger. So the lin-log model will predict a smaller impact on wage each year (see graph few slides back)

e.g.: lin-log

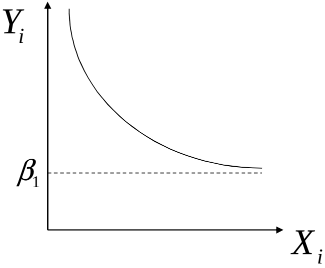
- Wage as a function of relative change in education
- $wage_i = \beta_1 + \beta_2 ln(educ_i) + u_i$
- $\widehat{wage}_i = -10.15 + 7.54 ln(educ_i)$
- $\widehat{wage}_{10} = -10.15 + 7.54 ln(10) = 7.21$
- $\widehat{wage}_{11} = -10.15 + 7.54 ln(11) = 7.93$
- $\%\Delta\widehat{wage}_{10\rightarrow 11} = \$0.72$
- $\widehat{wage}_{17} = -10.15 + 7.54 ln(17) = 11.21$
- $\widehat{wage}_{18} = -10.15 + 7.54 ln(18) = 11.64$
- $\%\Delta\widehat{wage}_{17\rightarrow 18} = \$0.43$
- $\frac{1}{10} = 0.1 \quad \frac{1}{17} = 0.06 \quad 0.06 * 7.54 = .44$
- For a 1% (0.01) change in X, the change in Y is $\beta_2/100$, in this case 0.0754.

log-log (elasticity model)

- $lnY_i = \beta_1 + \beta_2 lnX_i + u_i$
- $\beta_2 = \frac{\Delta lnY_i}{\Delta lnX_i} == \frac{\text{relative change in} Y_i}{\text{relative change in} X_i}$
- Thus, for small changes, the relative (percentage) change in Y is β_2 time the relative (percentage) change in X.

reciprocal models

- In the case of diminishing marginal returns, more of X produces less and less effect on Y. Eventually Y reaches a max/min value and more of X has no effect.
  - Example: effect of GDP on infant mortality rate. β_1 is the value that Y approaches as X $\rightarrow \infty$



  - $Y_i = \beta_1 + \beta_2(\frac{1}{X_i}) + u_i$

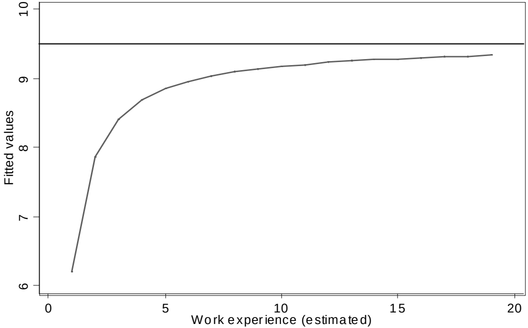
e.g.: wage and experience

- After a point, you are experienced enough, and more experience has less and less effect on wage. In this model, X can not be equal to 0, because 1/0 is undefined. Thus, you must either drop zeros or add a positive constant to all X.

- ```
  . reg wage exp if exp>0
  ------------------------------------------------------------------------------
          wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
           exp |   .0262689   .0183825     1.43   0.154     -.009844    .0623817
         _cons |   8.640484   .4029475    21.44   0.000     7.848883    9.432086
  ------------------------------------------------------------------------------

  . gen recipexp = 1/exp
  . reg wage recipexp
  ------------------------------------------------------------------------------
          wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
      recipexp |  -3.297158    1.33971    -2.46   0.014    -5.929055   -.6652601
  ```
  ```

interpreting reciprocal model [dofile: measurment]

- 

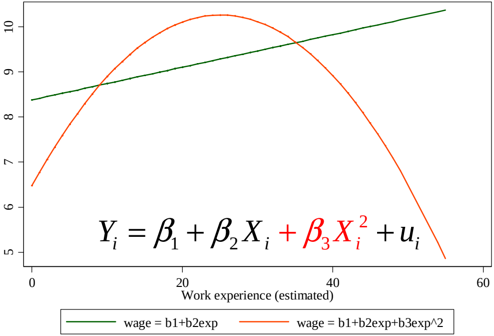
- $Y_i = \beta_1 + \beta_2(\frac{1}{X_i}) + u_i$
- changes in Y are smaller and smaller for larger Xs; yhats:
- di 9.5 -3.3*1
- di 9.5 -3.3*1/2
- di 9.5 -3.3*1/3

and now quadratic regression

- not bivariate regression anymore, but trivariate–see next week
- we introduce it here because the third var is just a sq of 2nd var
- and it does what logs or reciprocals do–fits a curve as opposed to a line
- and i think it is more intuitive than logs or reciprocals !
- the idea is that quadratic coef is smaller than linear, and opposite sign
- but as X gets bigger, its square get huge, and so quadratic coef with opposite sign overpowers fiorst term and curve flips to positive or negative

quadratic model

If a *non-linear relationship* between X and Y is suspected, a *polynomial function of* X can be used to model it.



Source: see quadratic_graph.do

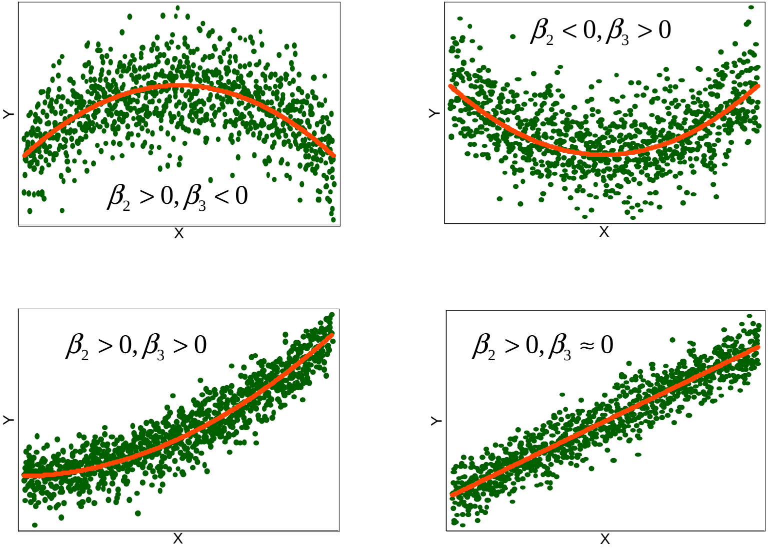
$$\frac{\Delta Y_i}{\Delta X_i} = \underbrace{\beta_2 + 2\beta_3 X_i}_{\text{The slope is no longer a constant}}$$

$$\beta_2 + 2\beta_3 X_i^* = 0$$

$$X_i^* = -\frac{\beta_2}{2\beta_3}$$

This curve reaches a maximum wage at the point were the marginal effect of experience is zero.

quadratic model

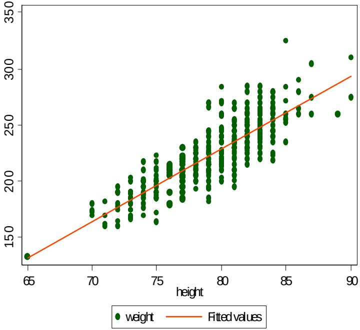


quadratic: interpretation

- the slope changes with X, it is not constant
- the best way to show the quadratic relationship is to graph it
- there is always a tipping point, but it may be outside the range of the data; in fact, the estimated line may be approximately linear for the observed data range even if the quadratic term is significant !
- the t test on squared term has a null hypothesis of linearity
  - if it is not significant, only linear term is left
- more practice http://www.ats.ucla.edu/stat/mult_pkg/faq/general/curves.htm
- dofile: quadratic

regression through the origin (RTO)

- often the intercept has to be zero, for reasons of logic or theory.
- rto seems logical ? be careful !
- rto is virtually always a bad idea
- just don't use it

- 

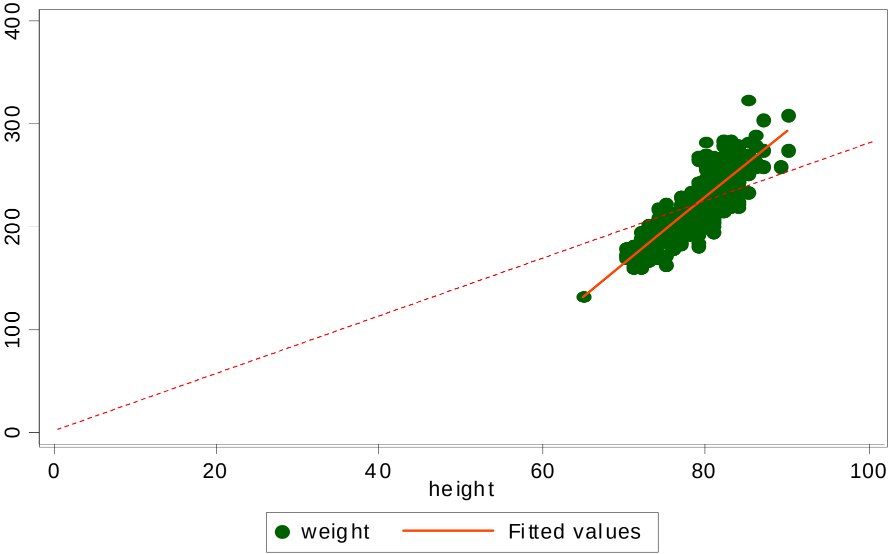
- i talk about it here to show you that you should not use it...

very different answers... why?

```
. reg weight height
      Source |       SS       df       MS              Number of obs =     425
-------------+------------------------------           F(  1,   423) = 1025.84
       Model |  253964.038     1  253964.038           Prob > F      =  0.0000
    Residual |   104720.96   423  247.567281           R-squared     =  0.7080
-------------+------------------------------           Adj R-squared =  0.7074
       Total |  358684.998   424  845.955183           Root MSE      =  15.734
------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   6.453962   .2015055    32.03   0.000     6.057885    6.850039
       _cons |  -287.2806   15.98884   -17.97   0.000    -318.7081   -255.8531
------------------------------------------------------------------------------
```

```
. reg weight height, nocons
      Source |       SS       df       MS              Number of obs =     425
-------------+------------------------------           F(  1,   424) =49471.98
       Model |    21544113     1    21544113           Prob > F      =  0.0000
    Residual |  184643.978   424  435.481079           R-squared     =  0.9915
-------------+------------------------------           Adj R-squared =  0.9915
       Total |    21728757   425  51126.4871           Root MSE      =  20.868
------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   2.837525   .0127573   222.42   0.000      2.81245    2.862601
```

oops !



bonus

- http://www.ats.ucla.edu/stat/stata/webbooks/reg/default.htm
- dofile:bonus