Review of key concepts

Rajeev Dehejia

About the course: mechanics

- Office hours Tuesday 3.45-4.45
- Default location is my Puck office or via Zoom
- By appointment.
  - Online signup: http://goo.gl/TnJEM
  - Schedule appointment with Erilia (as needed)
- Assignments are due via NYU Brightspace: Assignments

About the course: fine print

- Late assignments: point-reductions as noted on syllabus (for missing the deadline, then up to a one-week delay, zero thereafter).
- E-mail blackout 24 hours prior to submission of assignments.

About the course: Stata

- We make extensive use of Stata in this course.
- First learn where to access it:
  - Buy.
  - DDS workstation.
  - Virtual Computer Lab.
- Second, learn how to use it!
  - The hard way is the best way: teach yourself.

About the course: Stata

- Getting help:
  - DDS: http://nyu.libguides.com/dataservicestudio/
  - Princeton guide: http://data.princeton.edu/stata/
  - UCLA guide: http://www.ats.ucla.edu/stat/stata/
  - Built-in help; i.e. stata "help" commands
  - Visit Stata forums; e.g. statalist.org
  - Ask each other.
  - Ask me.
  - Special OH and classes.
  - Ask the TA!

TA

- We’re luck to have a fantastic TA
- Lots of course knowledge, lots of applied knowledge
- TA will run recitations and OH

About the course: goal

- Not just talk – let's learn to use these methods, when and how they're best applied, and how they get us closer to inferring causality
- The only way to learn is to **use real data**, **try**, and **figure it out**.
- In terms of interpretation methods, we will stick to concepts where possible, but will need a bit of math from time to time.

Statistics: just need the basics

- Measures of dispersion.
- Population vs. sample.
- Standard deviation vs. standard error.
- Expectations.
- Iterated expectations.
- Analogy principle.

Econometrics is a means to an end, not an end in itself

- Two types of ends (What do we want to know?)
  - Causal Effects: **Did** X cause Y?
  - Forecasting: How **will** X impact Y?
- Causal effects are answers to ‘what if’ questions:
  - What would happen to smoking if cigarette taxes were raised?
  - ‘What if questions’ = considering counterfactuals
- Forecasting – just want best currently available predictors – don’t (necessarily) worry about causality

Purpose, scope, and examples

Goal in policy-oriented econometrics is to assess the causal effect of public policy interventions

- Focus here is causal rather than forecasting aims.
- Includes part of "program evaluation" as used in other course (i.e., impact evaluation).

Examples include effects of:

- Job training programs on earnings and employment
- Class size on test scores
- Minimum wage on employment
- Military service on earnings and employment
- Tax-deferred saving programs on savings accumulation

➢ I.e., not just explicit / formal programs

Emphasis on means to an end…

- Recommended texts:
  - ✓ *Wooldridge Intermediate Econometrics* (not very technical)
  - ✓ *Cross-Sectional Econometrics* (more advanced)
  - ✓ *Mastering Metrics* – Angrist & Pischke (good review of regression and key ideas)
  - ✓ *Mostly Harmless Econometrics*- Angrist & Pischke (opinionated, informative – more advanced)

Today

- A review of preliminaries.
- But first…

Standard deviation vs standard error

- Standard deviation is a measure of central tendency – it is descriptive and can be applied to any data.
- Standard error is the variance (or standard deviation) of the sampling distribution of a statistic of interest.
    - If the statistic of interest is the mean from a random sample, then as you take a sequence of random samples, the sample mean will bounce around. That distribution is the sampling distribution of the sample mean, and the standard error is the variance of that distribution.

Standard deviation vs standard error

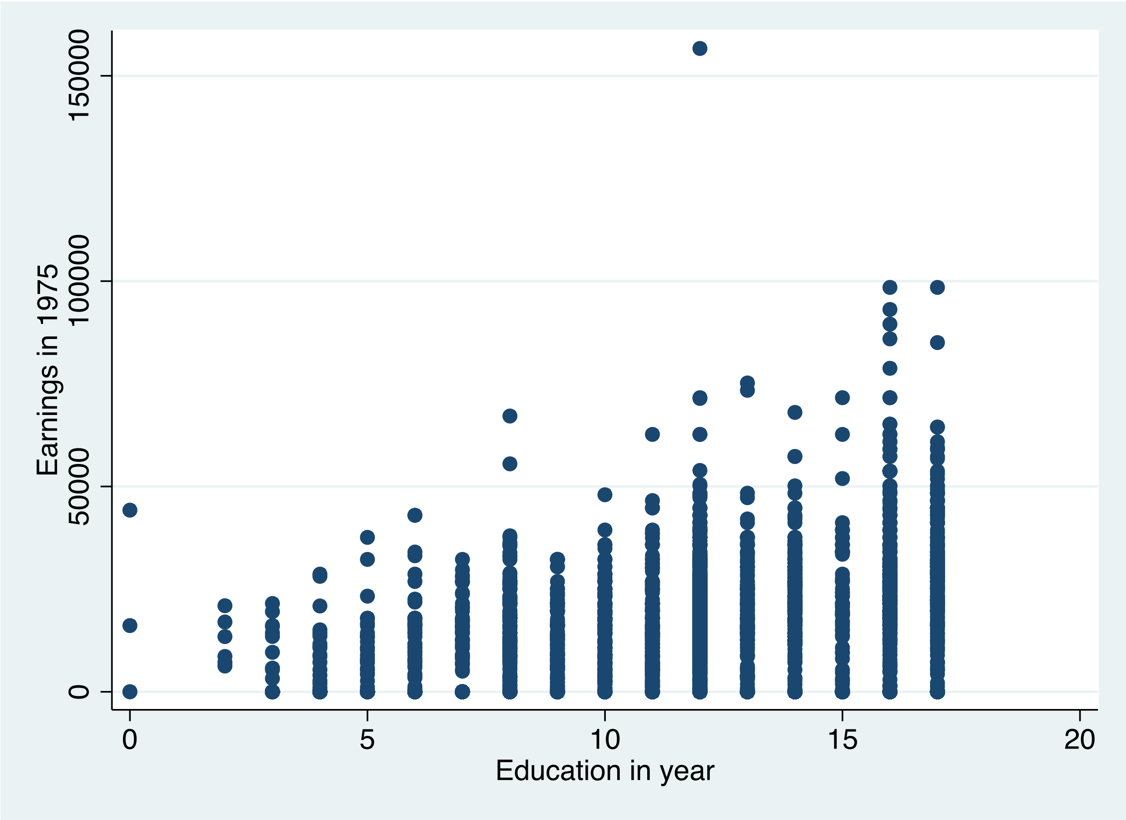
- So the standard error of the sample mean is s^2/n, where s is the standard deviation.

(Linear) (Mean) Regressions

- Now that we're at the mean, let's go back to the population mean.
- You might notice or think that the mean of one variable (earnings, Y) might depend on another variable (education X).
- Would be great if we could figure out how the mean Y depended on X.

Regressions

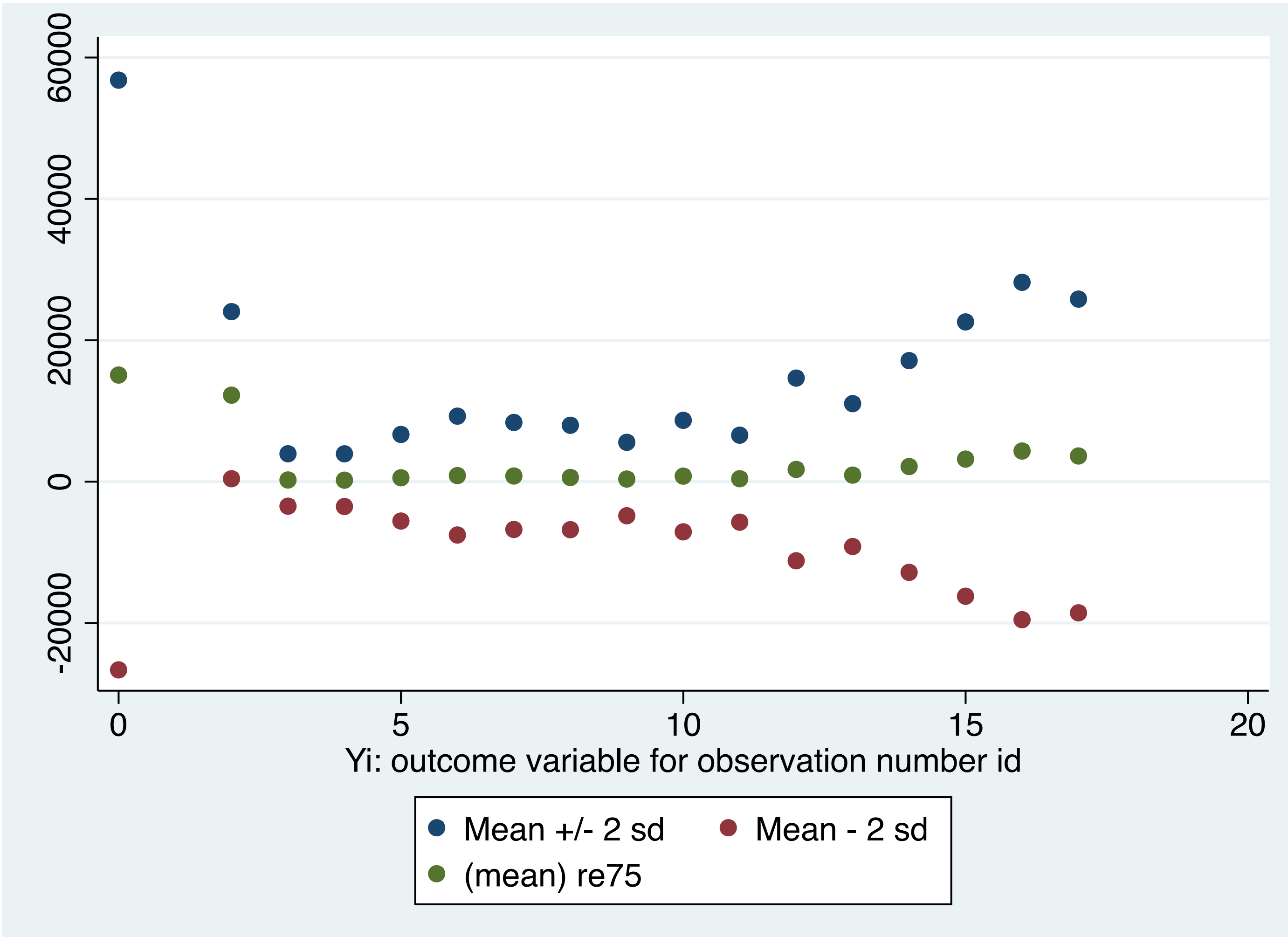
- One way to do this is just to plot the data.
- An example from NSW data (you'll use later).



(Linear) (Mean) Regressions

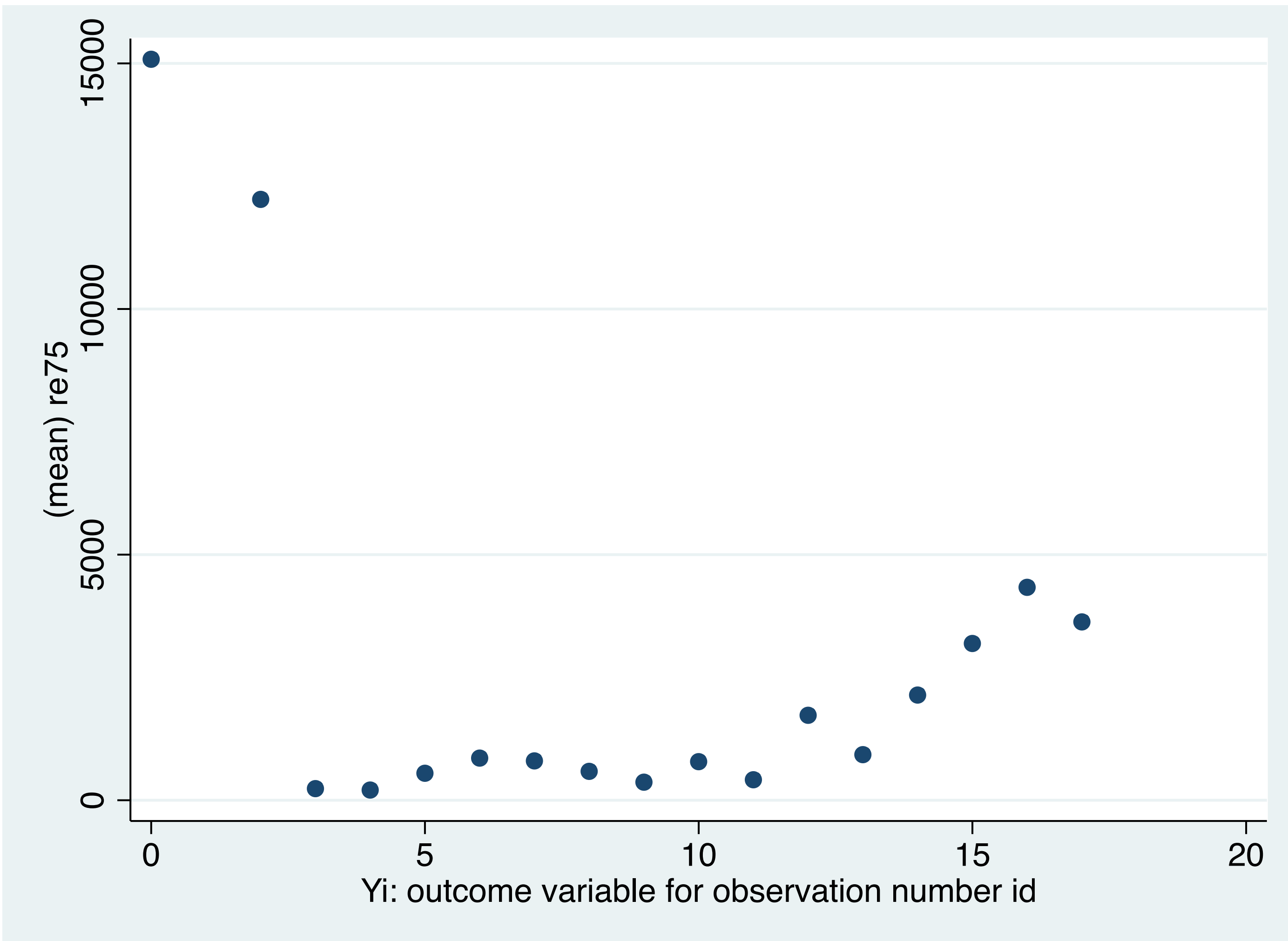
- That's a bit messy.
- Another simple idea. Supposed education takes on only integer values (e.g., years of education).
- The compute $E(Y)$ for each subsample defined on X, or $E(Y|X=1)$, $E(Y|X=2)$, etc.

From NSW



60000
40000
20000
0
-20000
0
5
10
15
20
Yi: outcome variable for observation number id
Mean +/- 2 sd
Mean - 2 sd
(mean) re75

From NSW



15000
10000
5000
0
(mean) re75
0
5
10
15
20
Yi: outcome variable for observation number id

(Linear) (Mean) Regressions

- That's interesting.
- Now suppose we want to think of this as a functional relationship, to keep it simple linear.
- I.e., mean of $y = f(x)$
- Or $Y=a+bX+e$
- Linear regression *is* mean regression.
- $E(Y|X)$=“the mean of y, given x” $=a+bX$

NSW example again

```
      Source |       SS       df       MS              Number of obs =   38062
-------------+------------------------------           F(  1, 38060) =  759.08
       Model |  2.5571e+10     1  2.5571e+10           Prob > F      =  0.0000
    Residual |  1.2821e+12 38060  33686421.3           R-squared     =  0.0196
-------------+------------------------------           Adj R-squared =  0.0195
       Total |  1.3077e+12 38061  34357374.4           Root MSE      =    5804

------------------------------------------------------------------------------
        re75 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   education |   275.2402   9.990032    27.55   0.000     255.6595     294.821
       _cons |  -1735.699    112.593   -15.42   0.000    -1956.385   -1515.014
------------------------------------------------------------------------------
```

- But what does 275 mean?

NSW (cont'd)

- For that you need to know this:
- . sum re75 education

Variable	Obs	Mean	Std. Dev.	Min	Max
re75	38062	1256.162	5861.516	0	156653.2
education	38062	10.87	2.977971	0	17

NSW (cont'd)

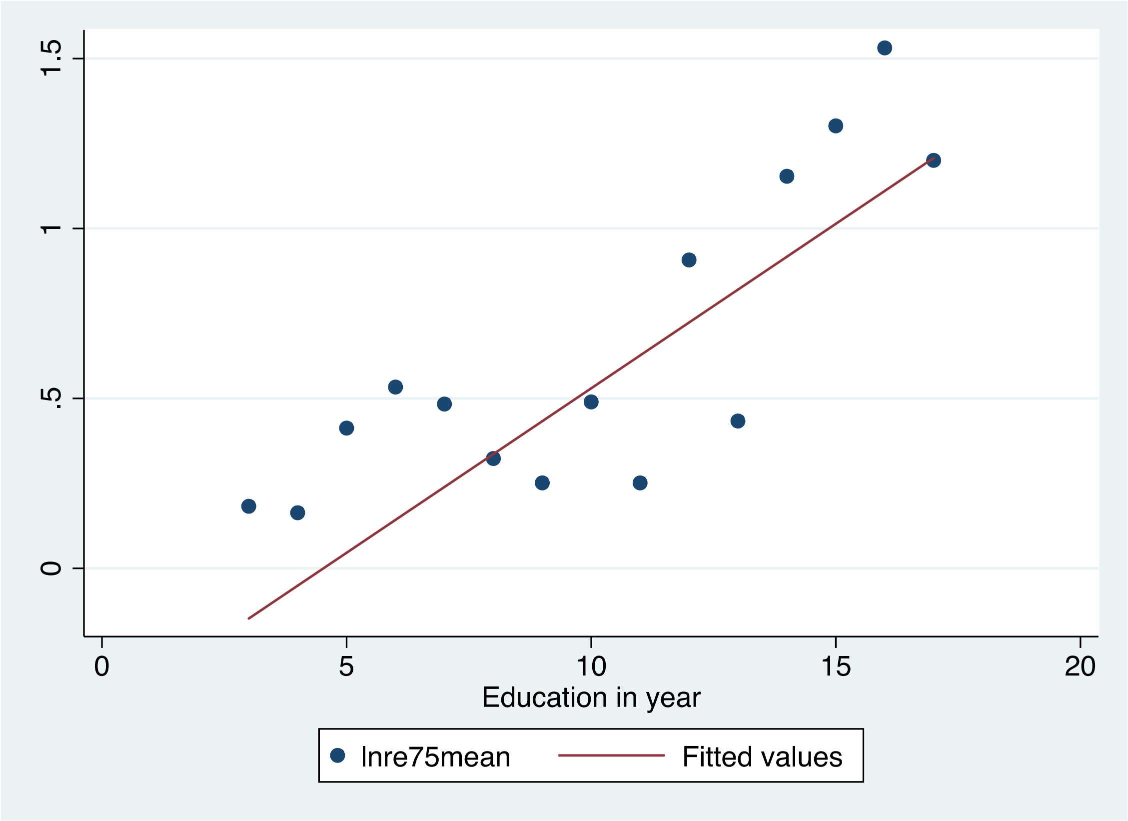

15000
10000
5000
0
0
5
10
15
20
Yi: outcome variable for observation number id
Mean earnings in 1975
Linear regression fit

NSW (cont'd)

- Before we leave this example, can we improve the functional form?

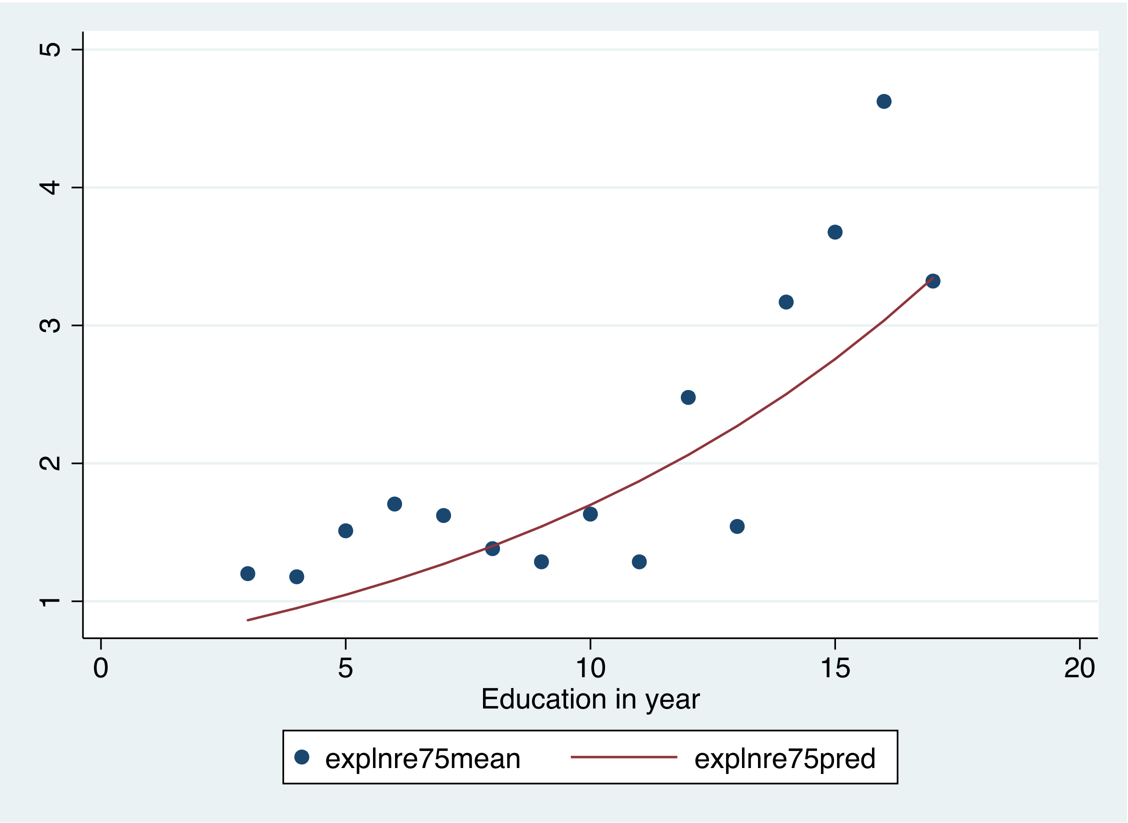
NSW (cont'd)

- Dropping the weird outliers in education, in a quasi log scale



NSW (cont'd)

- The same in levels

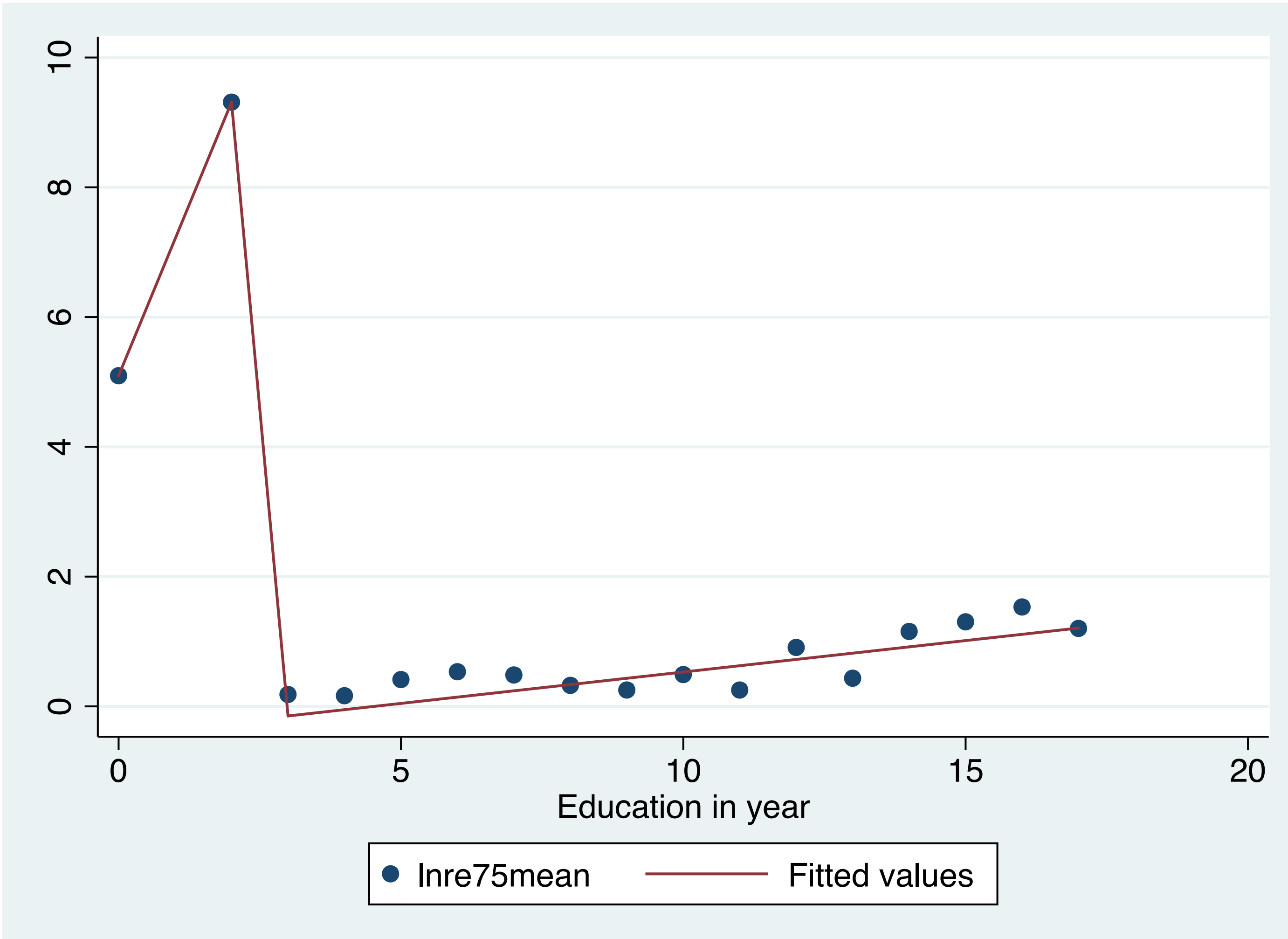


Regressions (con't)

- As we've seen a linear regression can be very non-linear (linear=linear in parameters).
- Dummy variables are another trick of the trade to deal with jumps. Recall those outliers at low education levels.
- Try this instead:

$$Y=a+b\,1(X=0) + c\,1(X=2) + dX+e$$

NSW (con’t)



0
2
4
6
8
10
0
5
10
15
20
Education in year
lnre75mean
Fitted values

Regressions (con't)

- Use of dummies can be extended to interactions of dummies as well.
- Recall the one dummy model:

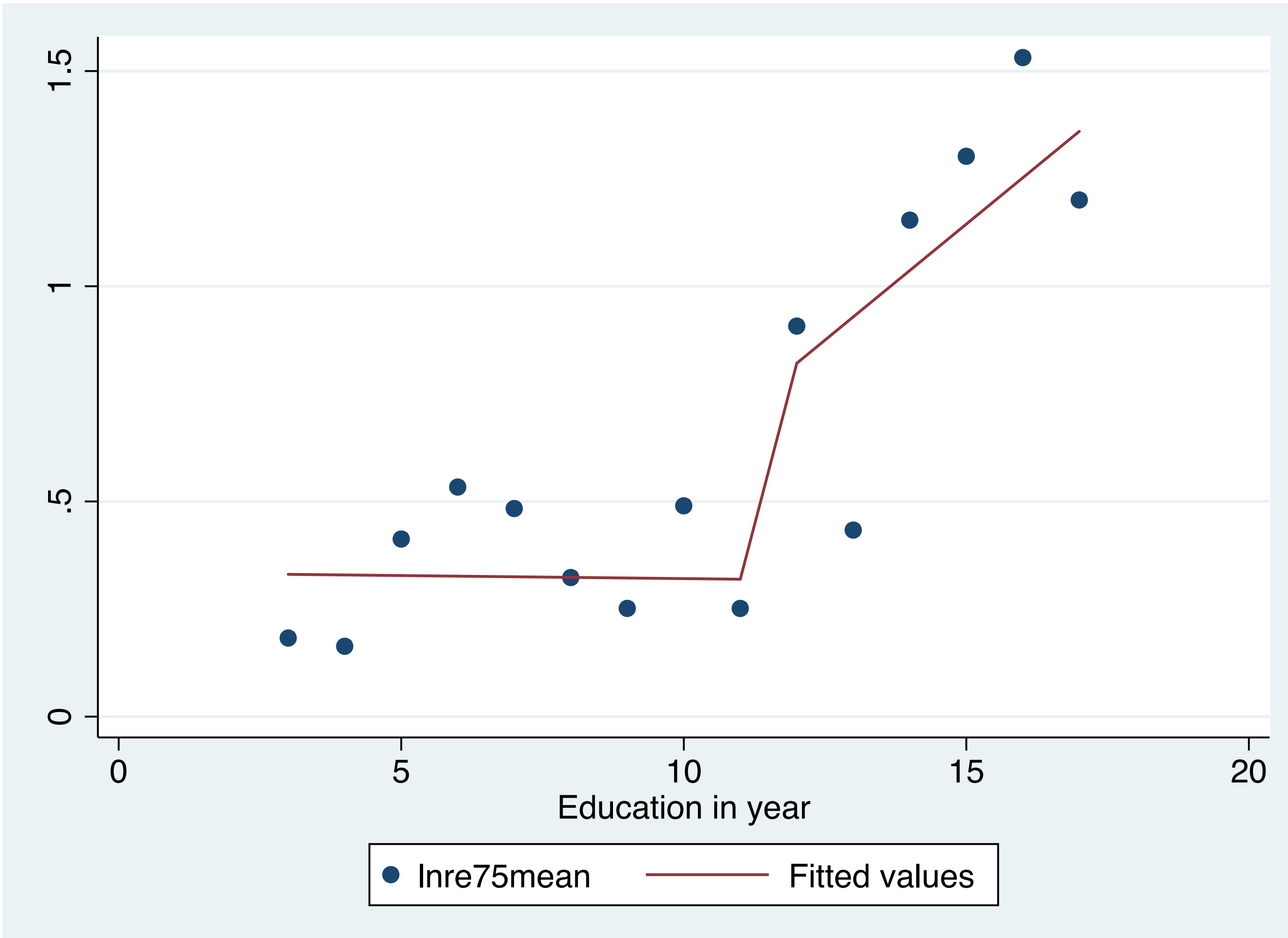
$$Y = a + bT + e$$

- Here *b* is the difference in means in Y of T=1 vs T=0.
- Recall the two-dummy interaction model:

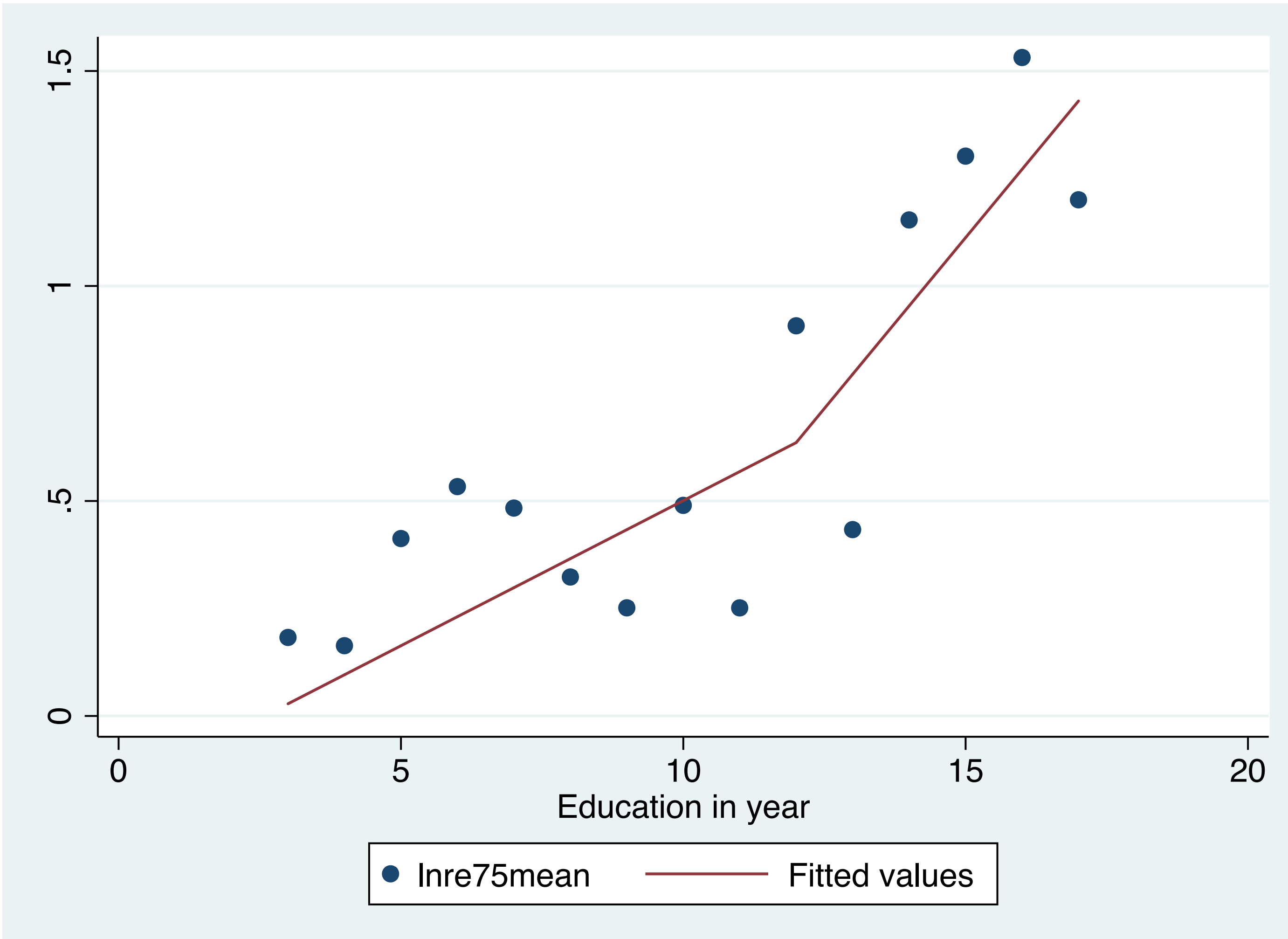
$$Y = a + bT + cTZ + dZ + e$$

- Here through the combination of 4 dummies, we can estimate the 4 means of the (0,1) x (0,1) combinations.
- This is a fully interacted ("fully flexible") model – there is no linearity assumption of any kind.

Example: NSW discontinuity



Example: NSW continuous slope shift



Quantile regressions

- Why you ask should all our machinery be devoted to a single summary statistic, namely the mean? Why not a median regression? Why not a 99th percentile regression?
- Why not?
- A median regression would tell you how the 50th percentile of a variable varied whatever explanatory variables you had.
- Practical issue: computationally much harder.

Expectations

- Suppose we have a population with a random variable taking on values y_i, for i=1,2,…,N with probability p_i. The population mean is

$$\mu_N = \sum_{i=1}^{N} p_i y_i \equiv E(y_i)$$

Expectations with a continuous random variable

- Suppose we have a population with a random variable taking on continuous values $y_i = g(i)$, for i = [0,1] with probability density dy.
- The probability of any range of values is

$$\Pr(i \le y \le j) = \int_i^j dy$$

- The population mean is

$$\mu_y = \int_i^j g(y)\,dy \equiv E(y_i)$$

Example 1

- Recall from Flash Quiz 1 that we found that $E(X) =(?)\ E(X|T=1)Pr(T=1)+ E(X|T=0)Pr(T=0)$.
- It isn't luck.
- Even though *T* doesn't appear in E(X), we can still "iterate the expectation".

$$E(X)=E_T(E(X|T))$$

- In words, on the inside take the expectation of X given T, and then take the expectation over the distribution of T.

Example 1

- But in $E(X)=E_T(E(X|T))$ *T* happens to be binary.
- Step back, what is the expectation (i.e., mean) of a binary variable?
- $E(T)=Pr(T=1)1+Pr(T=0)0$
- Instead of the expected value we are taking the expected value of $E(X|T)$ at $T=1$ and $T=0$
- Or $Pr(T=1)E(X|T=1) + Pr(T=0)E(X|T=0)$

Example 2: joint expectations

		Price		
		1	2	3
Yield	5	0.16 (1/6)	0.03 (1/30)	0.083 (1/12)
	10	0.14 (2/15)	0.14 (2/15)	0.16 (1/6)
	15	0.03 (1/30)	0.16 (1/6)	0.083 (1/12)

$$E(\mathrm{Re}venue) = E(py) = \sum_{y=5,10,15} \sum_{p=1,2,3} p \cdot y \cdot d(py)$$

$$E_p\left(E(y|p)p\right) = \sum_{p=1,2,3} \sum_{y=5,10,15} p \cdot y \cdot d(py)$$

$$= \sum_{p=1,2,3} \left(\sum_{y=5,10,15} y \cdot d(y|p) \right) p d(y)$$

$$= (1)(5)(1/6) + (1)(10)(2/15) + (1)(15)(1/30) +$$

$$(2)(5)(1/30) + (2)(10)(2/15) + (2)(15)(1/6) +$$

$$(3)(5)(1/12) + (3)(10)(1/6) + (3)(15)(1/12)$$

$$= \left[(1)(5)(0.5) + (1)(10)(0.4) + (1)(15)(0.1)\right]\left[\tfrac{1}{3}\right] +$$

$$\left[(2)(5)(0.1) + (2)(10)(0.4) + (2)(15)(0.5)\right]\left[\tfrac{1}{3}\right] +$$

$$\left[(3)(5)(0.25) + (3)(10)(0.5) + (3)(15)(0.25)\right]\left[\tfrac{1}{3}\right] +$$

Law of iterated expectations

- $E(x) = E_z(E(x \mid z))$
- Taking the expectation of a random variable, e.g. *x*, you can condition on some other random variable, *z*, as long as you then take the expectation over *z* as well.
- Easiest to see using summations rather than expectations.

Look at E(Y_i)

Imagine a situation with an outcome interest Y, where some people are in a treatment group (T_i=1), and others in a control group (T_i=0)

$$
\begin{aligned}
E(Y_i) &= \frac{1}{N}\sum_{i=1}^{N} Y_i \\
&= \frac{1}{N}\left(\sum_{i=1}^{N_1}(Y_i) + \sum_{i=1}^{N_0}(Y_i)\right), \text{ where first sum is for treated, second for untreated,} \\
&\quad \text{N}_1 = \#(\text{T}_\text{i} = 1 \text{ obs}), \text{N}_0 = \#(\text{T}_\text{i} = 0 \text{ obs}) \\
&= \frac{N_1}{N}\left(\frac{1}{N_1}\sum_{i=1}^{N_1}(Y_i)\right) + \frac{N_0}{N}\left(\frac{1}{N_0}\sum_{i=1}^{N_0}(Y_i)\right) \\
&= \Pr(T_i = 1)E(Y_i \mid T_i = 1) + \Pr(T_i = 0)E(Y_i \mid T_i = 0) \\
&= E_T(E(Y_i \mid T_i))
\end{aligned}
$$

Why the last step? Recall that if x is a binary variable taking on values x_1 and x_2 then E(x) = Pr(x=x_1)x_1 + Pr(x=x_2)x_2

Analogy principle

- If we observed the entire population, we could compute μ_Y , σ_Y^2 , etc.
- Usually, we observe a sample, $Y_1, . . . , Y_N$, of N observations, but we want to estimate a population parameter, like μ_Y or σ_Y^2 .
- Notation:
  - T : Estimand. Parameter in the population (e.g., μ_Y)
  - $T_N(Y_1,...,Y_N)$: Estimator. Function of the sample (e.g., the sample mean)
  - The values that our estimators take for particular samples are called estimates.

Analogy principle

- The analogy principle tells us to estimate a characteristic in the population using the same characteristic in the sample.
- For example, following the analogy principle, we estimate the population mean μ_Y using the sample mean:

$$\frac{1}{N}\sum_{i=1}^{N} x_i \rightarrow \frac{1}{n}\sum_{i=1}^{n} x_i,$$

where n observations are sample from the population of N,

- Similarly, we estimate the population variance σ_Y^2 sample variance.

Analogy principle

Estimators are functions of random variables. Therefore, they are random variables themselves.

Let

$$Y = \begin{cases} 1 & \text{with prob. } 0.5, \\ 0 & \text{with prob. } 0.5. \end{cases}$$

Therefore

$$E[Y] = 1 \times 0.5 + 0 \times 0.5 = 0.5$$

Suppose that $N = 4$.

Then, $\bar{Y}$ has the distribution shown in the table.

Y_1	Y_2	Y_3	Y_4	$\bar{Y}$
1	1	1	1	1.00
1	1	1	0	0.75
1	1	0	1	0.75
1	0	1	1	0.75
0	1	1	1	0.75
1	1	0	0	0.50
1	0	1	0	0.50
1	0	0	1	0.50
0	1	1	0	0.50
0	1	0	1	0.50
0	0	1	1	0.50
1	0	0	0	0.25
0	1	0	0	0.25
0	0	1	0	0.25
0	0	0	1	0.25
0	0	0	0	0.00

Analogy principle

The distribution of $\bar{Y}$ in this example can be summarized as:

$$\bar{Y} = \begin{cases} 0.00 & \text{with prob. } 1/16, \\ 0.25 & \text{with prob. } 4/16, \\ 0.50 & \text{with prob. } 6/16, \\ 0.75 & \text{with prob. } 4/16, \\ 1.00 & \text{with prob. } 1/16. \end{cases}$$

Notice that $E[\bar{Y}] = E[Y] = 0.5$.

That is, $\bar{Y}$ is **unbiased**.