Lecture 1: Introduction to Regression

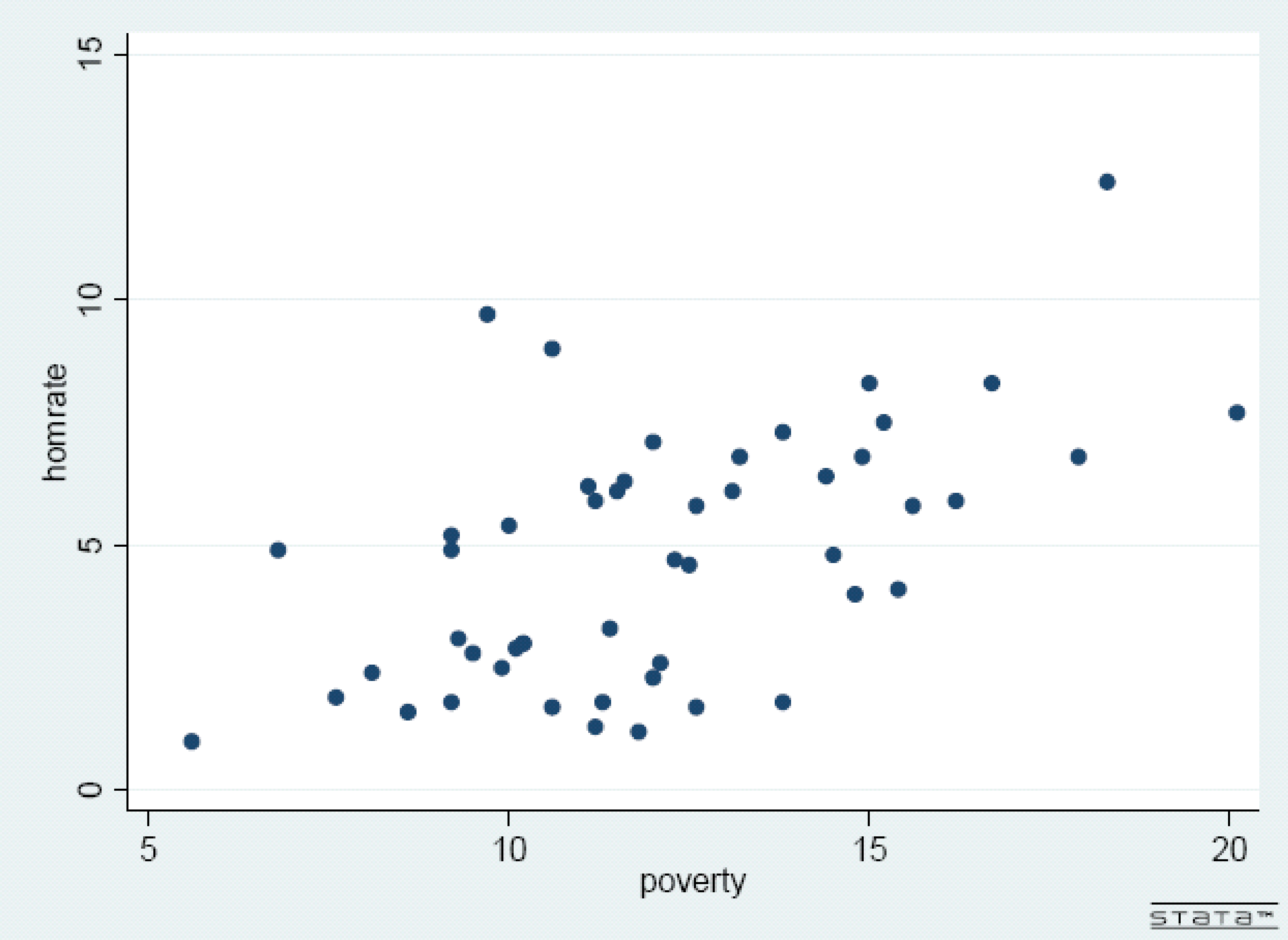
An Example: Explaining State Homicide Rates

- What kinds of variables might we use to explain/predict state homicide rates?

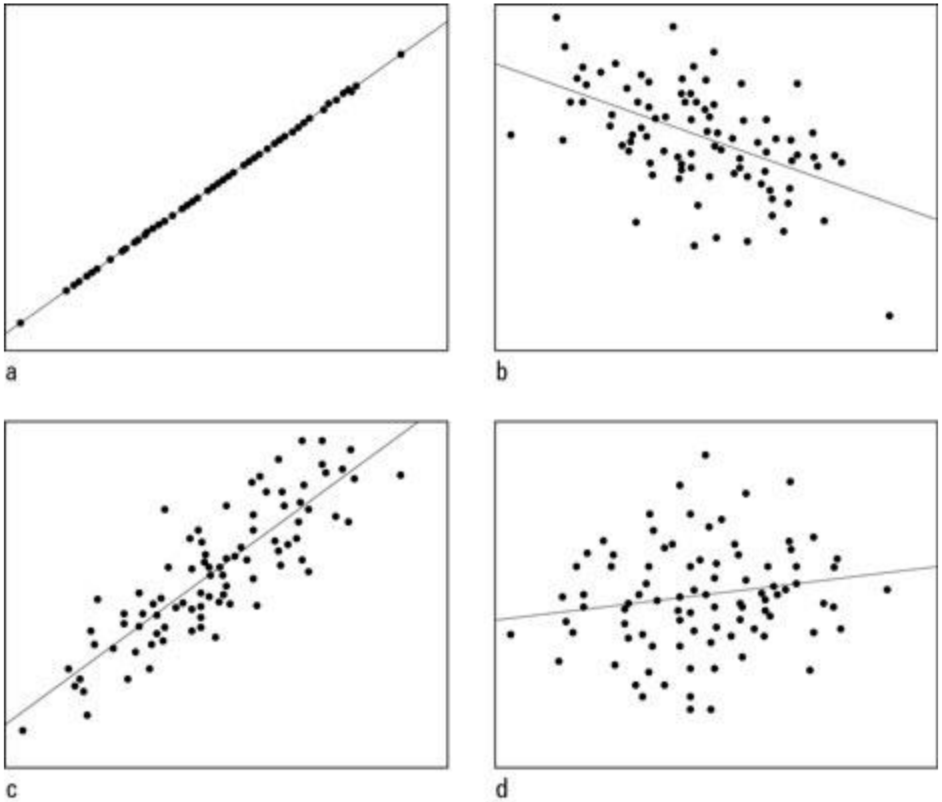
- Let's consider just one predictor for now: poverty
  - Ignore omitted variables, measurement error
  - How might this be related to homicide rates?

Poverty and Homicide

- These data are located here:
  - http://www.public.asu.edu/~gasweete/crj604/data/hom_pov.dta
- Download these data and create a scatterplot in Stata.
- Does there appear to be a relationship between poverty and homicide? What is the correlation?


15
10
5
0
homrate
5
10
15
20
poverty
stata

Scatterplots and correlations



Scatterplots with correlations of a) +1.00; b) –0.50; c) +0.85; and d) +0.15.

Poverty and Homicide

- There appears to be some relationship between poverty and homicide rates, but it’s not perfect.
- But there is a lot of “noise” which we will attribute to unobserved factors and random error.

Poverty and Homicide, cont.

- There is some nonzero value of expected homicides in the absence of poverty. (β_0)
- We expect homicide rates to increase as poverty rates increase. (β_1)
- Thus, $Y = \beta_0 + \beta_1 X$
- This is the **Population Regression Function**

Poverty and Homicide, Sample Regression Function

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + u_i$$

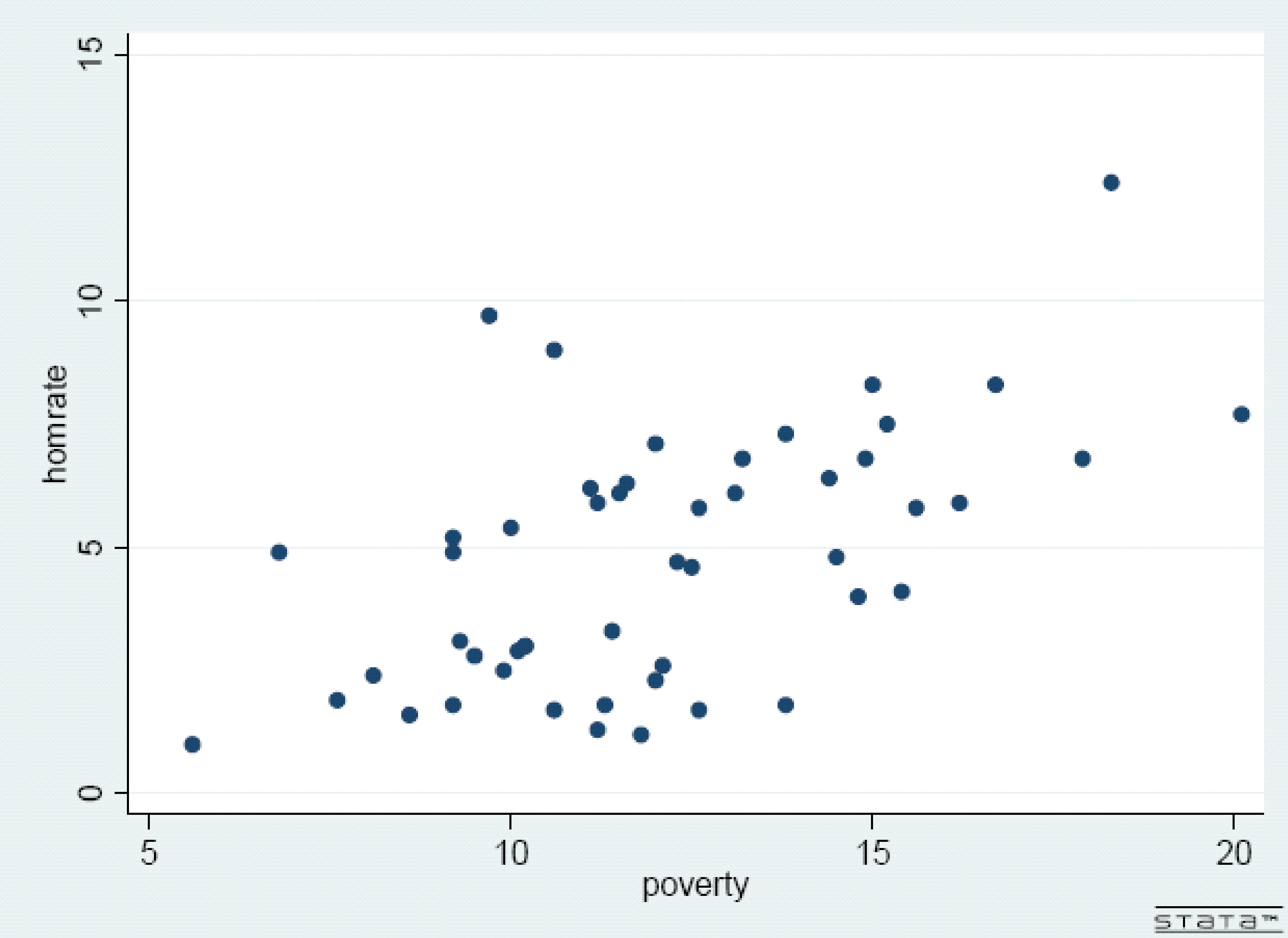
- y_i is the dependent variable, homicide rate, which we are trying to explain.
- $\hat{\beta}_0$ represents our *estimate* of what the homicide rate would be in the absence of poverty*
- $\hat{\beta}_1$ is our *estimate* of the “effect” of a higher poverty rate on homicide
- u_i is a “noise” term reflecting other things that influence homicide rates

*This is extrapolation outside the range of data. Not recommended.

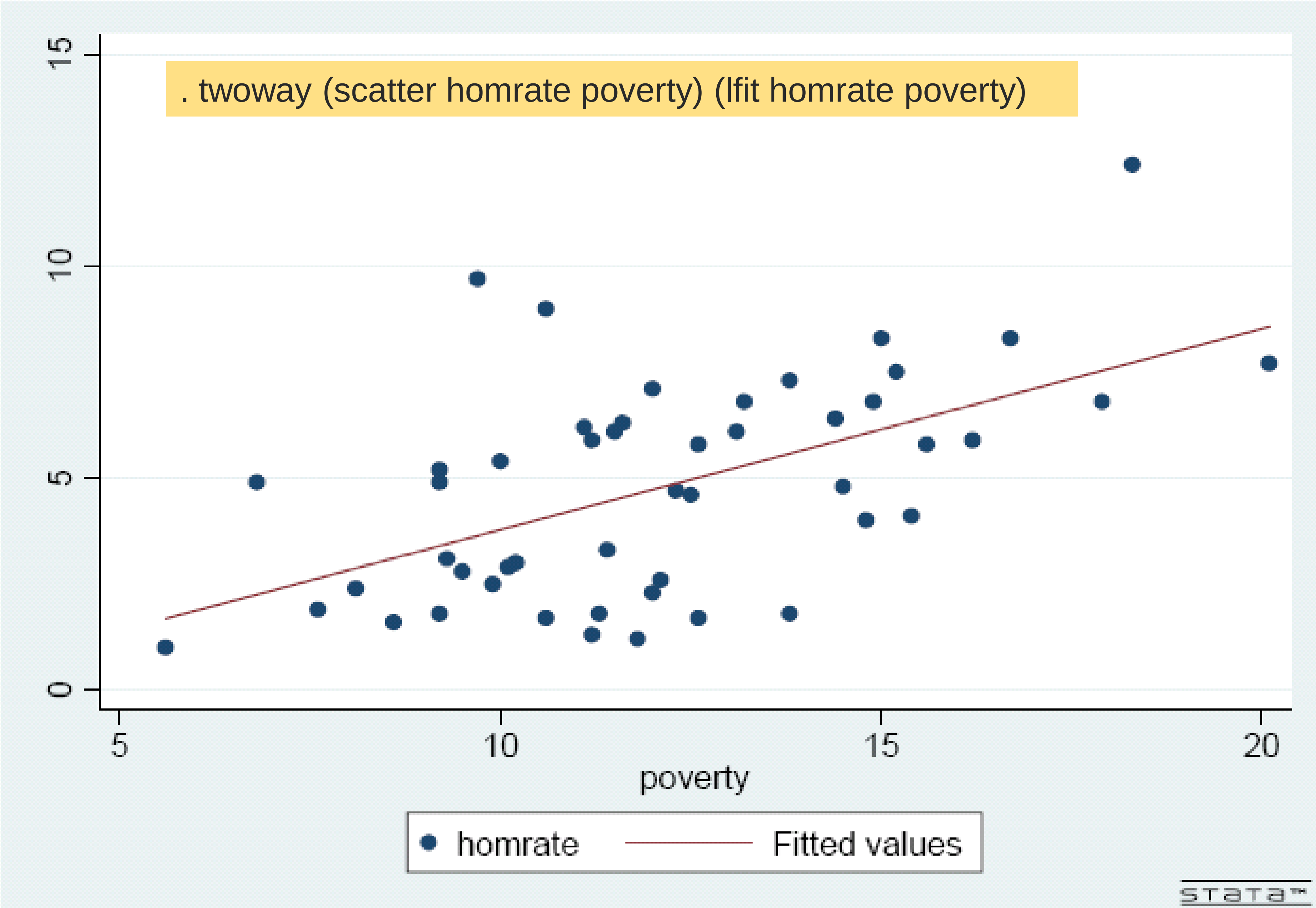
Poverty and Homicide, cont.

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + u_i$$

- Only y_i and x_i are directly observable in the equation above. The task of a regression analysis is to provide estimates of the slope and intercept terms.
- The relationship is assumed to be linear. An increase in *x* is associated with an increase in *y*.
  - Same expected change in homicide going from 6 to 7% poverty as from 15 to 16%



homrate
0
5
10
15
poverty
5
10
15
20
stata™



$$\beta_0 = -.973 \qquad \beta_1 = 0.475$$

Ordinary Least Squares

$$y_i = -.973 + .475x_i + u_i$$

- Substantively, what do these estimates mean?
- -.973 is the expected homicide rate if poverty rates were zero. This is never the case, except perhaps in the case of a zombie apocalypse, so it's not a meaningful estimate.
- .475 is the effect of a 1 unit increase in the poverty rate on the homicide rate. You need to know how you are measuring poverty. In this case, 1 unit increase is an increase of 1 percentage point.
- So a 1 percentage *point* increase (not "percent increase") in the poverty rate is associated with an increase of .475 homicides per 100,000 people in the state.
    - In AZ, this would be ~31 homicides.

Ordinary Least Squares

$$y_i = -.973 + .475x_i + u_i$$

- How did we arrive at this estimate? Why did we draw the line exactly where we did?
  - Minimize the sum of the "squared error", aka Ordinary Least Squares (OLS) estimation

$$\min \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

- Why squared error?
- Why vertical error? (Not perpendicular).

Ordinary Least Squares Estimates

$$\min \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

- Solving for the minimum requires calculus (set derivative with respect to β to 0 and solve)
- The book shows how we can go from some basic assumptions to estimates for β_0 and β_1 without using calculus.
- I will go through two different ways to obtain these estimates: Wooldridge's and Khan's (khanacademy.org)

Ordinary Least Squares: Estimating the intercept (Wooldridge's method)

$$E(u) = 0$$

$$u = y - \beta_0 - \beta_1 x$$

$$E(y - \beta_0 - \beta_1 x) = 0$$

$$\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- Assuming that the average value of the error term is zero, it is a trivial matter to calculate β_0 once we know β_1.

Ordinary Least Squares: Estimating the intercept (Wooldridge)

- Incidentally, these last sets of equations also imply that the regression line passes through the point that corresponds to the mean of x and the mean of y: $\left(\bar{x}, \bar{y}\right)$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Ordinary Least Squares: Estimating the slope (Wooldridge)

$$E(u) = 0$$

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + u_i$$

$$u_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$n^{-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

- First, we use the fact that the expected value of the error term is zero, to create generate a new equation equal to zero.
- We saw this before, but here I use the exact formula used in the book.

Ordinary Least Squares: Estimating the slope (Wooldridge)

$$Cov(x,u) = E(xu) = 0$$

$$n^{-1}\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$n^{-1}\sum_{i=1}^{n} x_i(y_i - (\bar{y} - \hat{\beta}_1\bar{x}) - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i) = 0$$

- We can multiply this last equation by x_i since the covariance between *x* and *u* is assumed to be zero and the terms in the parentheses are equal to *u*.
- Next, we plug in our formula for the intercept and simplify

Ordinary Least Squares: Estimating the slope (Wooldridge)

$$\sum_{i=1}^{n} x_i(y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i) = 0$$

- Re-arranging . . .

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) + \sum_{i=1}^{n} x_i(\hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) + \hat{\beta}_1\sum_{i=1}^{n} x_i(\bar{x} - x_i) = 0$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) = \hat{\beta}_1\sum_{i=1}^{n} x_i(x_i - \bar{x})$$

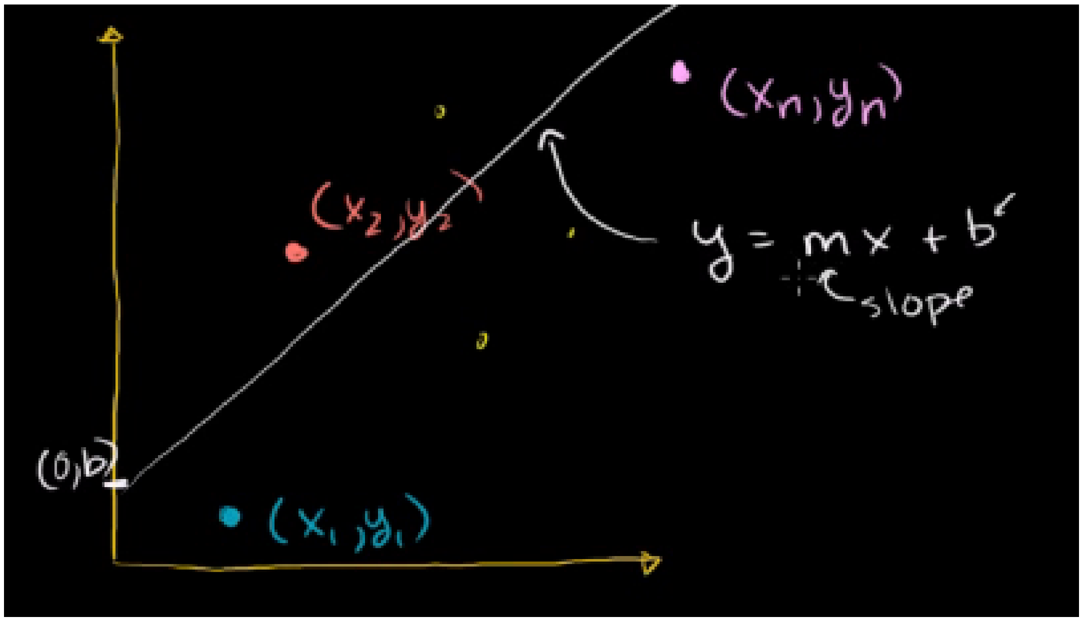
Ordinary Least Squares: Estimating the slope (Wooldridge)

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)=\hat{\beta}_1\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)}$$

- Re-arranging . . .
- Interestingly, the final result leads us to the relationship between covariance of x and y and variance of x.

Ordinary Least Squares: Estimates (Khan's method)



- Khan starts with the actual points, and elaborates how these points are related to the squared error, the square of the distance between each point (x_n,y_n) and the line $y=mx+b=\beta_1 x+\beta_0$

Ordinary Least Squares: Estimates (Khan's method)

- The vertical distance between any point (x_n,y_n), and the regression line y= $\beta_1 x+\beta_0$ is simply $y_n-(\beta_1 x_n+\beta_0)$

$$Total\ Error = (y_1 - (\beta_1 x_1 + \beta_0)) + (y_2 - (\beta_1 x_2 + \beta_0)) + \ldots + (y_n - (\beta_1 x_n + \beta_0))$$

- It would be trivial to minimize the total error. We could set β_1 (the slope) equal to zero, and β_0 equal to the mean of y, and then the total error would be zero.
- Another approach is to minimize the absolute difference , but this actually creates thornier math problems than squaring the differences, and results in situations where there is not a unique solution.
- In short, what we want is the sum of the squared error (SE), which means we have to square every term in that equation.

Ordinary Least Squares: Estimates (Khan's method)

$$SE = (y_1 - (\beta_1 x_1 + \beta_0))^2 + (y_2 - (\beta_1 x_2 + \beta_0))^2 + \ldots + (y_n - (\beta_1 x_n + \beta_0))^2$$

- We need to find the β_1 and β_0 that minimize the SE. Let's expand this out.
- To be clear, the subscripts for the β estimates just refer to our two regression line estimates, whereas the subscripts for our x's and y's refer to the first observation, second observation and so on.

$$SE = (y_1^2 - 2y_1(\beta_1 x_1 + \beta_0) + (\beta_1 x_1 + \beta_0)^2) + \ldots + (y_n^2 - 2y_n(\beta_1 x_n + \beta_0) + (\beta_1 x_n + \beta_0)^2)$$

$$= y_1^2 - 2y_1\beta_1 x_1 - 2y_1\beta_0 + \beta_1^2 x_1^2 + 2\beta_1 x_1 \beta_0 + \beta_0^2 + \ldots$$

$$+ y_n^2 - 2y_n\beta_1 x_n - 2y_n\beta_0 + \beta_1^2 x_n^2 + 2\beta_1 x_n \beta_0 + \beta_0^2$$

Ordinary Least Squares: Estimates (Khan's method)

- Summing these columns . . .

$$SE = \sum_{i=1}^{n} y_i^2 - 2\beta_1 \sum_{i=1}^{n} y_i x_i - 2\beta_0 \sum_{i=1}^{n} y_i + \beta_1^2 \sum_{i=1}^{n} x_i^2 + 2\beta_0\beta_1 \sum_{i=1}^{n} x_i + n\beta_0^2$$

$$= n * mean(y^2) - 2n\beta_1 * mean(xy) - 2n\beta_0 * mean(y) +$$

$$n\beta_1^2 * mean(x^2) + 2n\beta_0\beta_1 * mean(x) + n\beta_0^2$$

- Everything but the regression line coefficients are known entities here.
- This equation represents a 3D surface, where different values of β_1 and β_0 correspond to different values of the squared error. We just need to pick the values of β_1 and β_0 that minimize the SE.

Ordinary Least Squares: Estimates (Khan's method)

- Those familiar with calculus will know that the minimum of the squared error surface occurs where the partial derivative (slope) with respect to β_1 is equal to zero and the partial derivative with respect to β_0 is equal to zero.

$$\frac{\partial SE}{\partial \beta_0} = -2n * mean(y) + 2n\beta_1 * mean(x) + 2n\beta_0 = 0$$

$$-\bar{y} + \beta_1 \bar{x} + \beta_0 = 0$$

$$\bar{y} = \beta_1 \bar{x} + \beta_0$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

- We've seen that before. How about the other derivative?

Ordinary Least Squares: Estimates (Khan's method)

$$\frac{\partial SE}{\partial \beta_1} = -2n * mean(xy) + 2n\beta_1 * mean(x^2) + 2n\beta_0 * mean(x) = 0$$

$$-mean(xy) + \beta_1 * mean(x^2) + \beta_0 * \bar{x} = 0$$

- Replacing β_0 . . .

$$-mean(xy) + \beta_1 * mean(x^2) + \bar{y} * \bar{x} - \beta_1 \bar{x} * \bar{x} = 0$$

$$\beta_1 (mean(x^2) - \bar{x} * \bar{x}) = mean(xy) - \bar{y} * \bar{x}$$

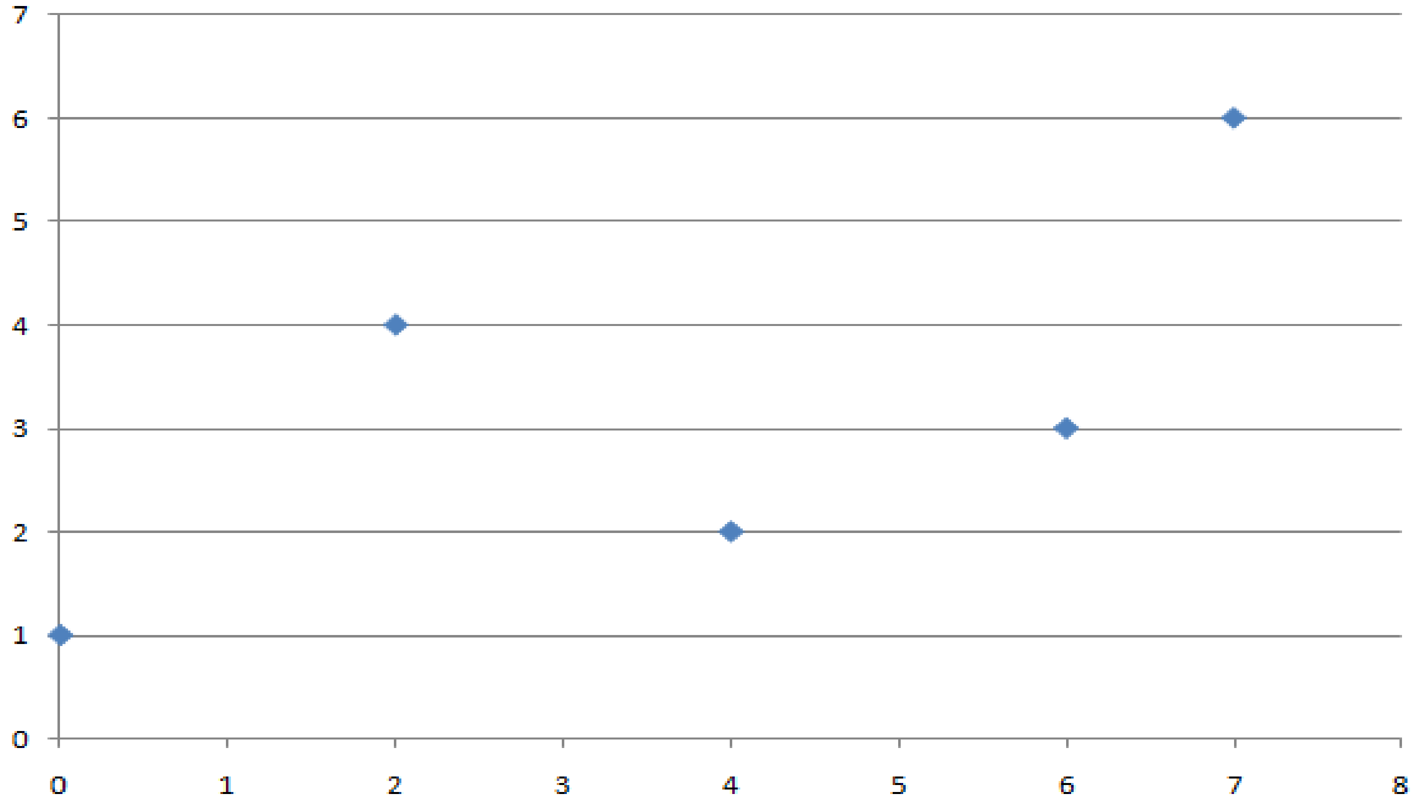
$$\beta_1 = \frac{mean(xy) - \bar{y} * \bar{x}}{mean(x^2) - \bar{x} * \bar{x}} = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)}$$

Ordinary Least Squares Estimates

- Hopefully it is reassuring to know that we can obtain the same answers from two very different methods.
- These formulas allow us, in a bivariate regression, to calculate the regression line “by hand” without using fancy statistical packages. All we need to do is find the mean of x, the mean of y, the mean of the products of x and y, and the mean of the squares of x, and then we can plug this into the formulas and crank out our solutions.

OLS by hand, example

- Let's look at a set of 5 points, and see how to calculate a regression line "by hand".
- Here are our five points: (4,2) (7,6) (0,1) (6,3) (2,4)



OLS by hand, example

- We can generally guess that the slope will be positive, but we can find the slope exactly if we calculate four things: the mean of x, the mean of y, the mean of the products of x and y, and the mean of the squares of x
- The x’s are 4,7,0,6, and 2. Their mean is 19/5=3.8
- The y’s are 2,6,1,3, and 4. Their mean is 16/5=3.2
- The products are 8,42,0,18 and 8. Their mean is 76/5=15.2.
- The squared x’s are 16,49,0,36, and 4. Their mean is 105/5=21.

OLS by hand, example

- Recall the formula for the slope:

$$\beta_1 = \frac{mean(xy) - \bar{y} * \bar{x}}{mean(x^2) - \bar{x} * \bar{x}} = \frac{15.2 - 3.2 * 3.8}{21 - 3.8 * 3.8} = \frac{3.04}{6.56} \cong .463$$

- Once we have the slope, the intercept is trivial:

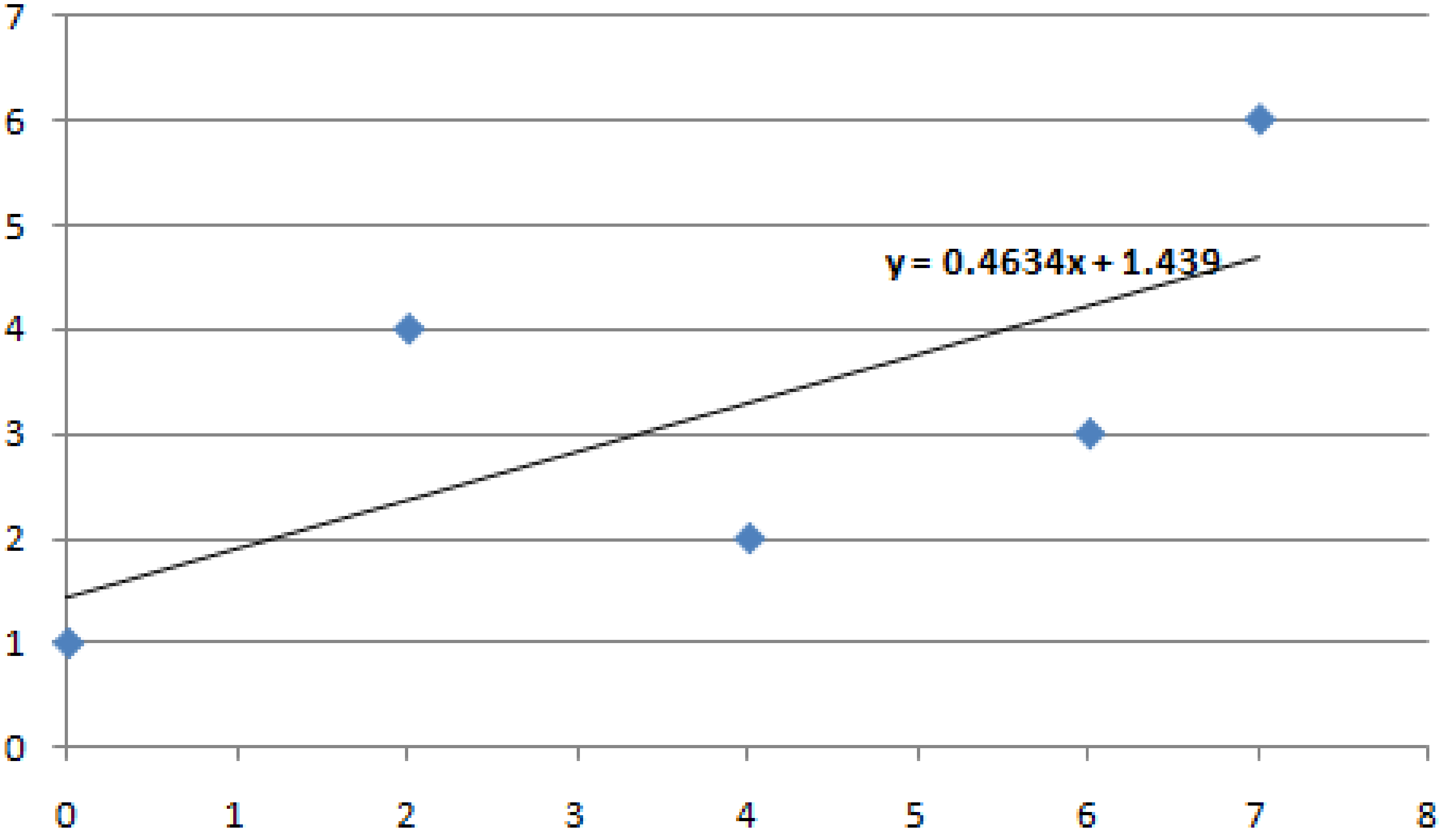
$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 3.2 - .463 * 3.8 = 1.44$$

- And our regression line that minimizes the sum of squared differences:

$$y_i = 1.44 + .463 x_i + u_i$$

OLS by hand, example

- Checking our work . . .



Analysis of Variance

- Once we have our regression line, we can define a “fitted value” as follows:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- This is our estimated value for y given our slope and intercept estimates and the value of x. It’s also sometimes called a “predicted value.”
- All of the “y-hats” fall on the regression line. For purposes of evaluating our regression, it makes sense to compare the y-hats to the actual values of y.

Analysis of Variance

- **The total variation in Y** is partitioned into two parts:

$$y_i - \bar{y} = y_i - \bar{y} - \hat{y}_i + \hat{y}_i = \underbrace{\left(y_i - \hat{y}_i\right)}_{\text{Residuals (variation not explained by the model)}} + \underbrace{\left(\hat{y}_i - \bar{y}\right)}_{\text{Variation explained by the model}}$$

- Of course, in order to assess variance, we square all of these terms:

$$\sum\left(y_i - \bar{y}\right)^2 = \sum\left(y_i - \hat{y}_i\right)^2 + \sum\left(\hat{y}_i - \bar{y}\right)^2$$

SST SSR SSE

- Where SST is the total sum of squares, SSE is the explained sum of squares, and SSR is the residual sum of squares.

R^2 “R-squared”

- R^2 represents the portion of the variance in *y* that is “explained” by the model.

$$R^2 = \frac{SSE}{SST} = \frac{\sum\left(\hat{y}_i - \bar{y}\right)^2}{\sum\left(y_i - \bar{y}\right)^2}$$

- Typically, in social science applications, our standards for R^2 are pretty low. Individual-level regressions rarely exceed .3

Ordinary Least Squares Estimates by hand

- See Excel file: “bivariate regression by hand.xls”
- http://www.public.asu.edu/~gasweete/crj604/misc/

state	hom	poverty	xi-xbar	yi-ybar	x*y	xi-xbar2
Alabama	8.3	16.7	4.61	3.53	16.27	21.3
Alaska	5.4	10	-2.09	0.63	-1.32	4.37
Arizona	7.5	15.2	3.11	2.73	8.49	9.67
Arkansas	7.3	13.8	1.71	2.53	4.326	2.92
California	6.8	13.2	1.11	2.03	2.253	1.23

Ordinary Least Squares Estimates by hand, cont.

- We can also get β_1 from the covariance (“`. corr hom pov, c`”) matrix in Stata, which shows that the covariance of homicide and poverty is 4.304 and the variance of poverty is 9.06.
- β_1=4.304/9.06=.475
- The mean of homicide rates is 4.77, and the mean of poverty rates is 12.09.
- β_0=4.77-12.09*.475=-.973
- Or, in Stata “`. reg hom pov`”

Stata output

- β_1=4.304/9.06=.475
- β_0=4.77-12.09*.475=-.973

```
. reg hom pov

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  1,    48) =   21.36
       Model |  100.175656     1  100.175656           Prob > F      =  0.0000
    Residual |  225.109343    48  4.68977798           R-squared     =  0.3080
-------------+------------------------------           Adj R-squared =  0.2935
       Total |  325.284999    49  6.63846936           Root MSE      =  2.1656

------------------------------------------------------------------------------
     homrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     poverty |    .475025   .1027807     4.62   0.000     .2683706    .6816795
       _cons |  -.9730529   1.279803    -0.76   0.451     -3.54627    1.600164
------------------------------------------------------------------------------
```

Assumptions of the Classical Linear Regression Model

1) X & Y are linearly related in the population.
2) We have a random sample of size *n* from the population.
3) The values of x_1 through x_n are not all the same.
4) The error has an expected value of zero for all values of *x*: $E(u_i|x) = 0$ (**zero conditional mean)**
5) The error term has a constant variance for all values of *x*: $\mathrm{Var}(u|x) = \sigma^2$ (**homoscedasticity)**

1) Linearity

- If X and Y are not linearly related, the estimates will be incorrect. Look at your data!
- Example, how do these data compare?:

```
. summ

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          x1 |        11           9    3.316625          4         14
          x2 |        11           9    3.316625          4         14
          x3 |        11           9    3.316625          4         14
          x4 |        11           9    3.316625          8         19
          y1 |        11    7.500909    2.031568       4.26      10.84
-------------+--------------------------------------------------------
          y2 |        11    7.500909    2.031657        3.1       9.26
          y3 |        11         7.5    2.030424       5.39      12.74
          y4 |        11    7.500909    2.030579       5.25       12.5
```

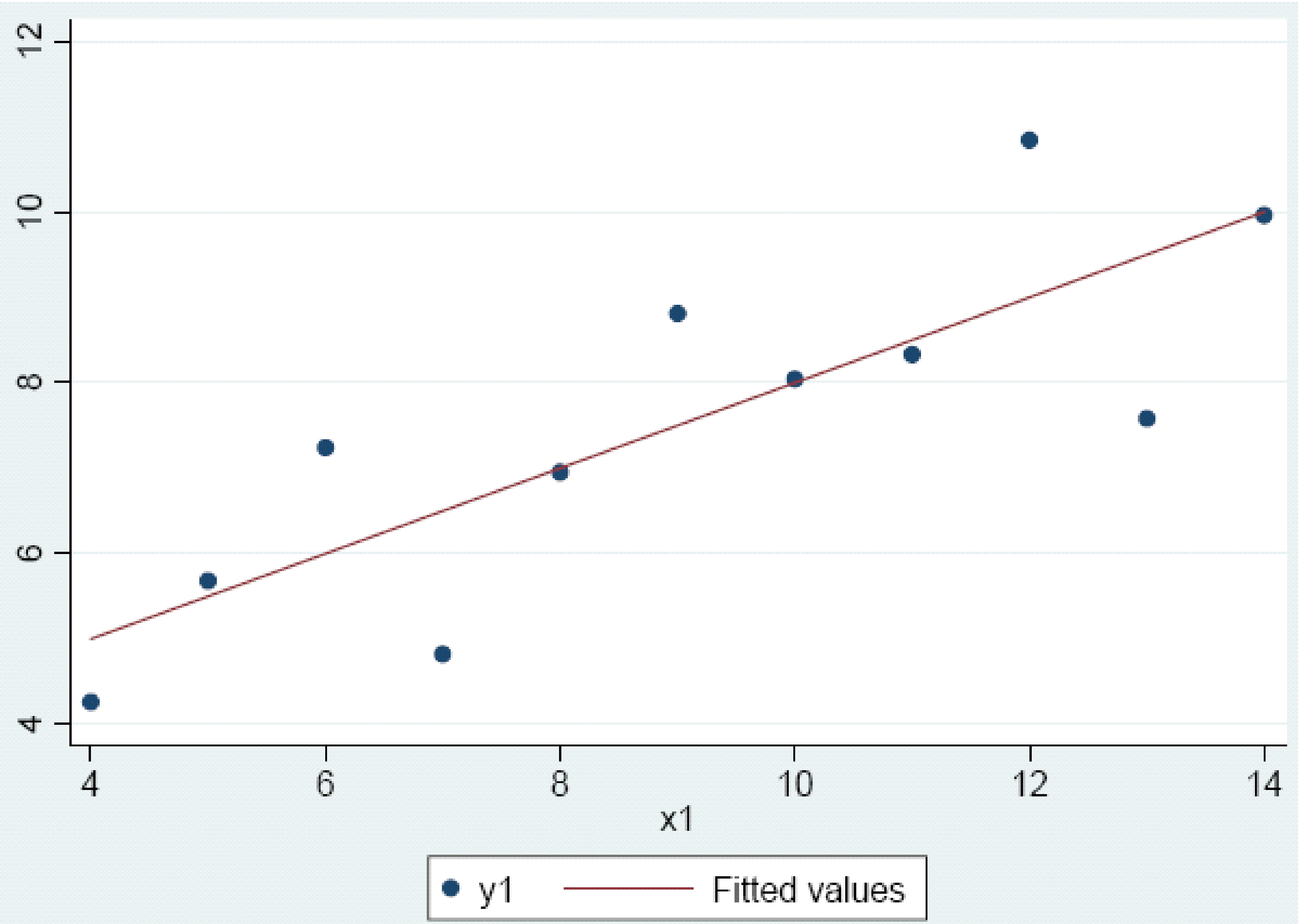
```
. reg y1 x1

      Source |       SS       df       MS              Number of obs =      11
-------------+------------------------------           F(  1,     9) =   17.99
       Model |  27.5100011     1  27.5100011           Prob > F      =  0.0022
    Residual |  13.7626904     9  1.52918783           R-squared     =  0.6665
-------------+------------------------------           Adj R-squared =  0.6295
       Total |  41.2726916    10  4.12726916           Root MSE      =  1.2366

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5000909   .1179055     4.24   0.002     .2333701    .7668117
       _cons |   3.000091   1.124747     2.67   0.026     .4557369    5.544445
------------------------------------------------------------------------------

. reg y2 x2

      Source |       SS       df       MS              Number of obs =      11
-------------+------------------------------           F(  1,     9) =   17.97
       Model |  27.5000024     1  27.5000024           Prob > F      =  0.0022
    Residual |   13.776294     9  1.53069933           R-squared     =  0.6662
-------------+------------------------------           Adj R-squared =  0.6292
       Total |  41.2762964    10  4.12762964           Root MSE      =  1.2372

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |         .5   .1179638     4.24   0.002     .2331475    .7668526
       _cons |   3.000909   1.125303     2.67   0.026     .4552978     5.54652
------------------------------------------------------------------------------
```

```
. reg y3 x3

      Source |       SS       df       MS              Number of obs =      11
-------------+------------------------------           F(  1,     9) =   17.97
       Model |  27.4700075     1  27.4700075           Prob > F      =  0.0022
    Residual |  13.7561905     9  1.52846561           R-squared     =  0.6663
-------------+------------------------------           Adj R-squared =  0.6292
       Total |  41.2261979    10  4.12261979           Root MSE      =  1.2363

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x3 |   .4997273   .1178777     4.24   0.002     .2330695    .7663851
        cons |   3.002455   1.124481     2.67   0.026     .4587014    5.546208
------------------------------------------------------------------------------

. reg y4 x4

      Source |       SS       df       MS              Number of obs =      11
-------------+------------------------------           F(  1,     9) =   18.00
       Model |  27.4900007     1  27.4900007           Prob > F      =  0.0022
    Residual |  13.7424908     9  1.52694342           R-squared     =  0.6667
-------------+------------------------------           Adj R-squared =  0.6297
       Total |  41.2324915    10  4.12324915           Root MSE      =  1.2357

------------------------------------------------------------------------------
          y4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x4 |   .4999091   .1178189     4.24   0.002     .2333841    .7664341
        cons |   3.001727   1.123921     2.67   0.026     .4592411    5.544213
------------------------------------------------------------------------------
```

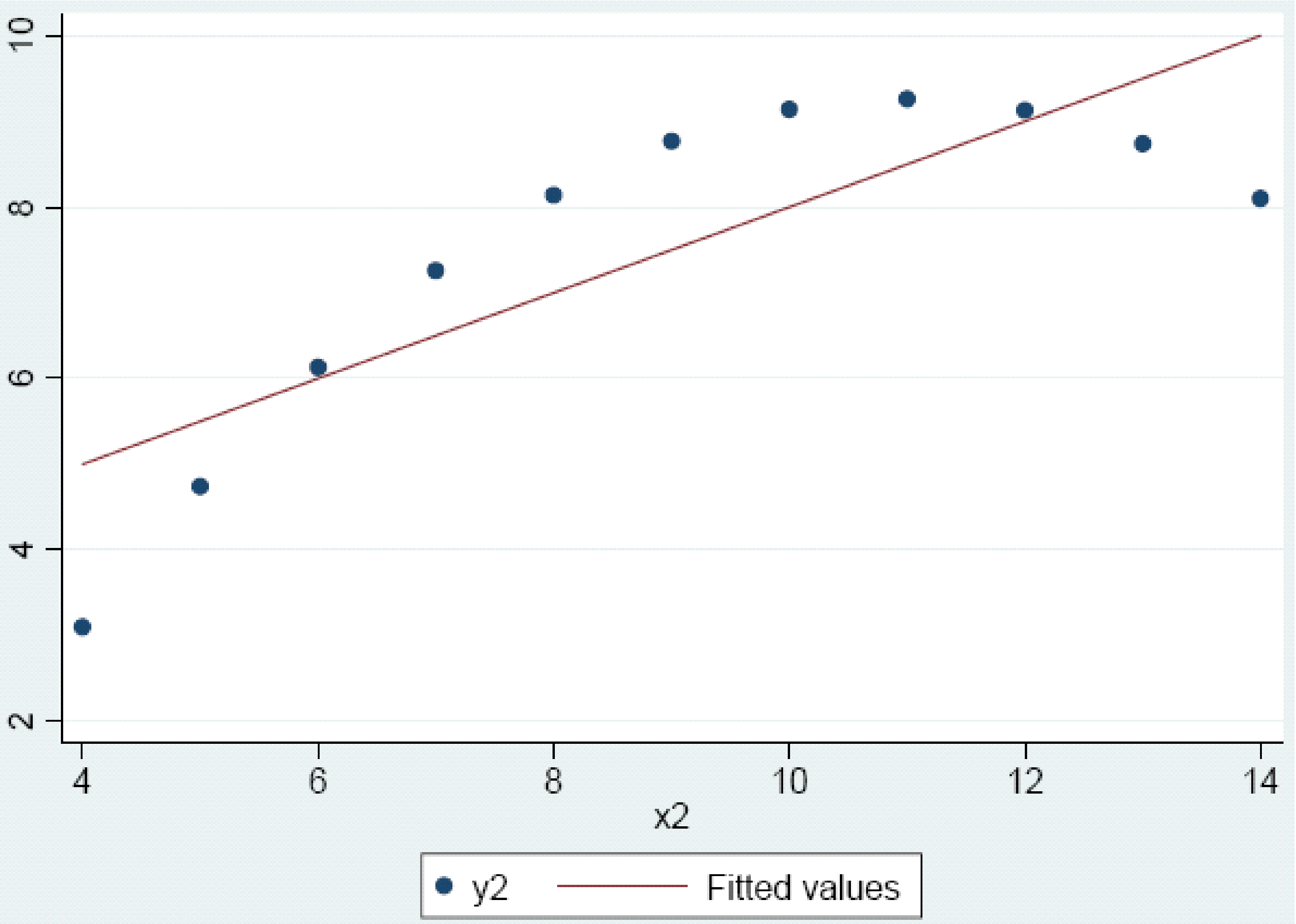
1) Linearity, cont.

- How do these models compare?
- β_0=3
- β_1=.5
- Let's look at each of them separately

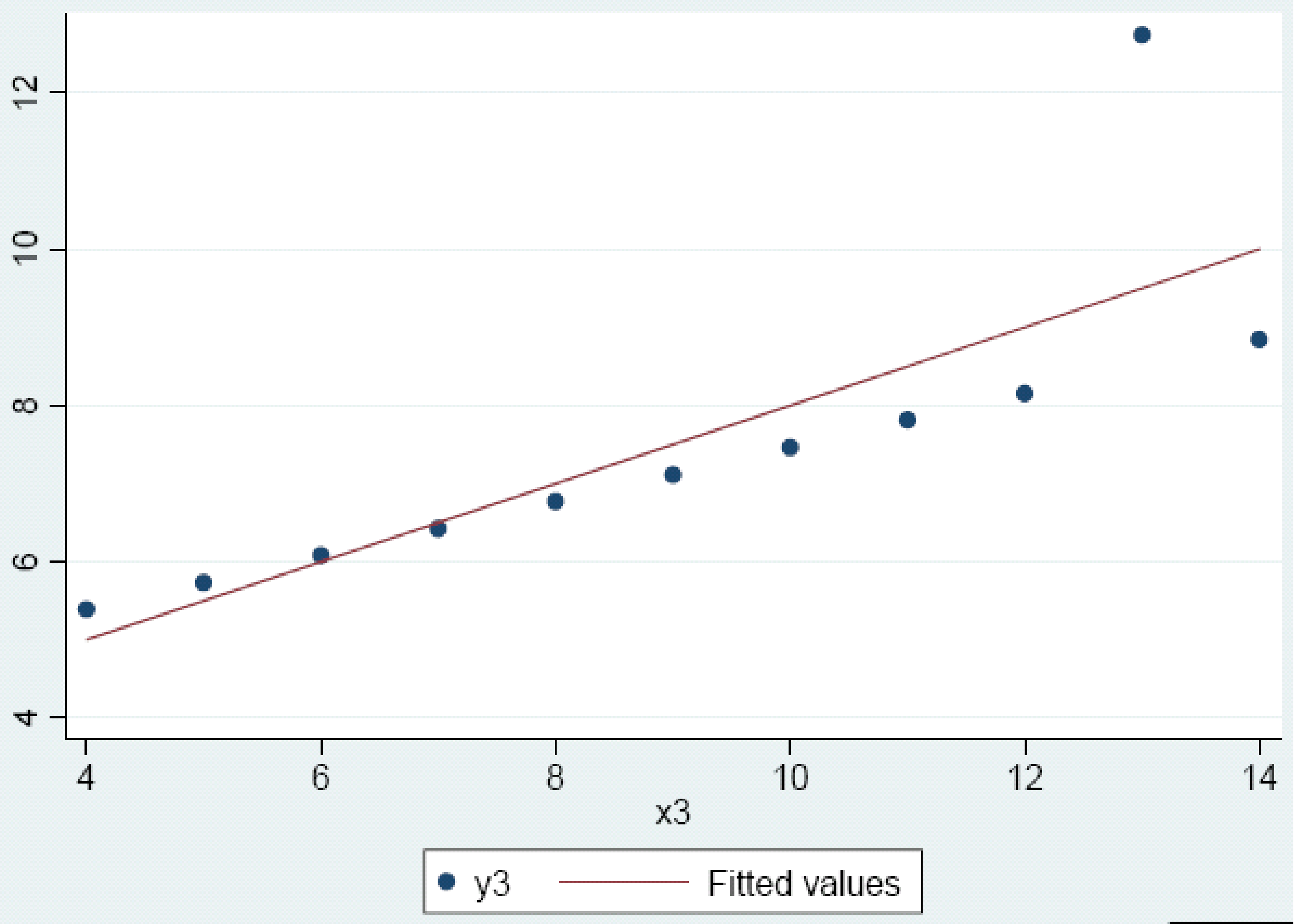
1) Linearity, cont., Regression 1



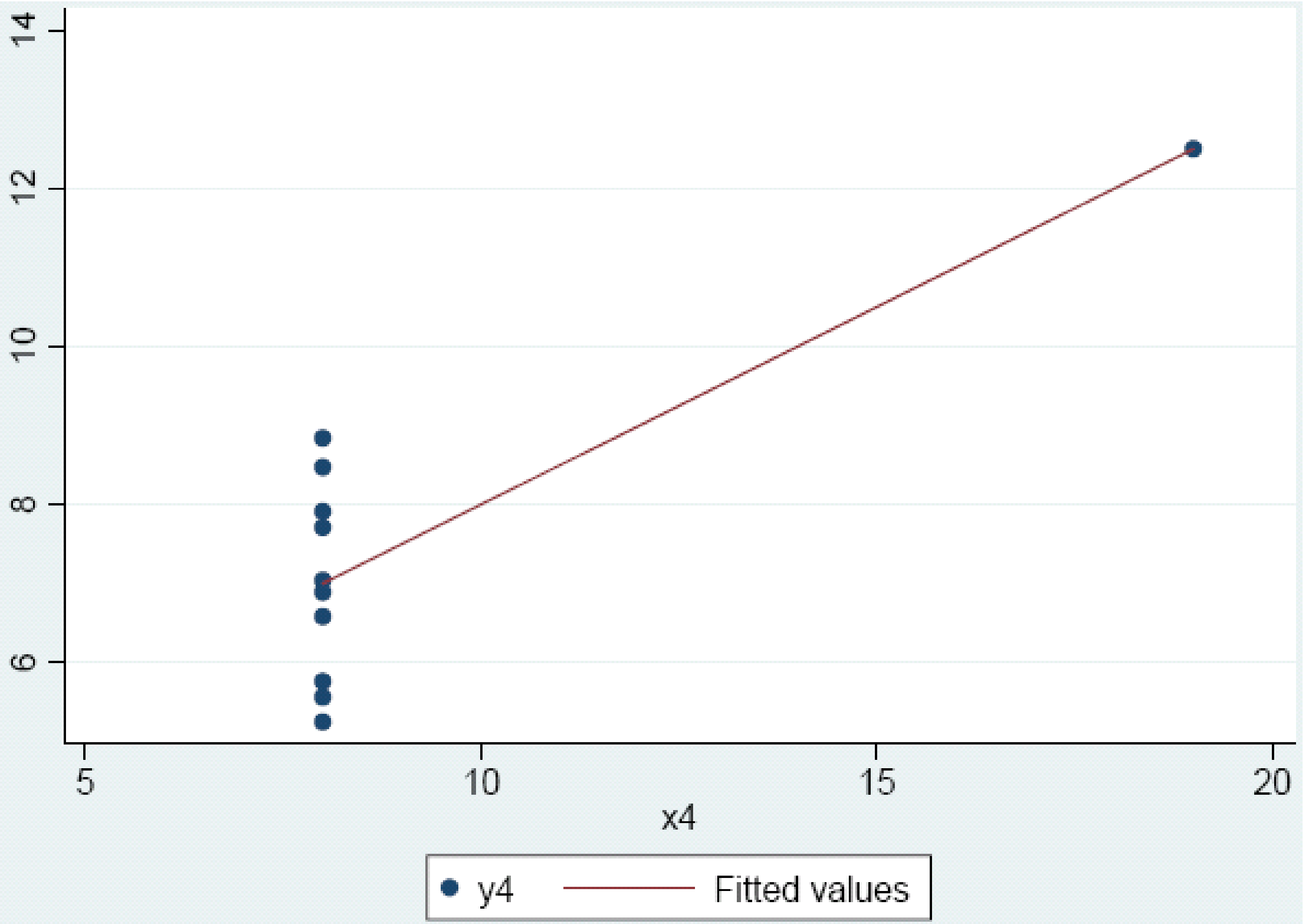
1) Linearity, cont., Regression 2



1) Linearity, cont., Regression 3



1) Linearity, cont., Regression 4



3) Sample variation

- If there is no variation in the values of *x*, it is not possible to estimate a regression line. The line of best fit would point straight up and pass through every point.
- Minimal variation in *x* is sometimes problematic as well, as it makes regression estimates very unstable.
- This assumption is easy to check by looking at summary statistics.

4) Zero conditional mean $E(u_i|x) = 0$

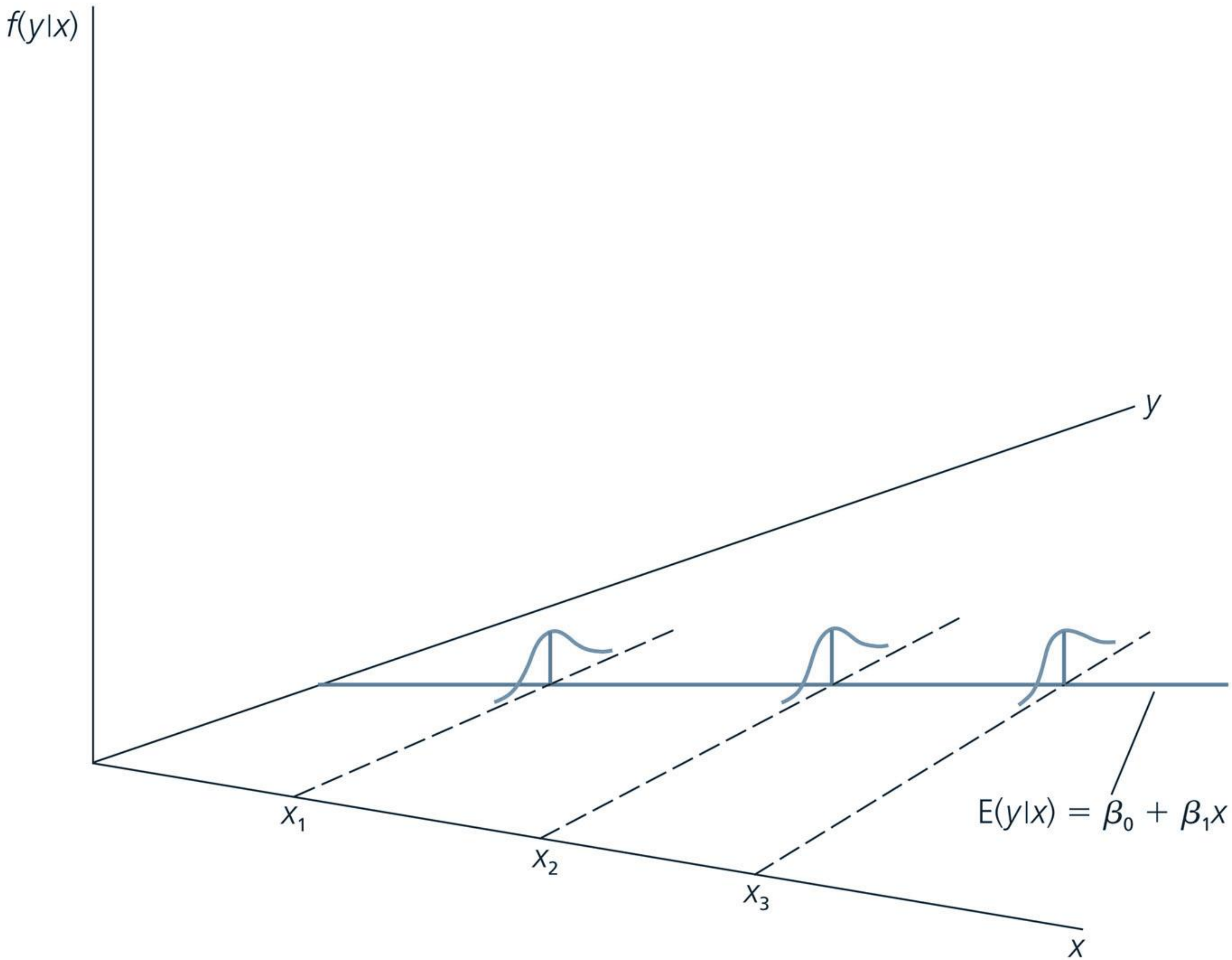
- In practical terms, this means that the sum of the unobserved variables is not related to x.
- Also, it means that variation in our estimates of the intercept and slope are all due to variations in the error terms.
- Should this assumption hold true, our estimates of the slope and intercept are unbiased, meaning that on average we're going to get the right answer.

5) Var($u|x$) = σ^2 (homoscedasticity)

- In practical terms, this means that the variance of the error term is unrelated to the independent variables.

FIGURE 2.8

The simple regression model under homoskedasticity.



Root Mean Squared Error (RMSE)

- Root mean squared error gives us an indication of how well the regression line fits the data.

$$RMSE = \sqrt{\frac{SSR}{n-k}}$$

- This is the square root of the residual sum of squares divided by the sample size minus the number of parameters being estimated (k=2 in simple bivariate regression).

Root Mean Squared Error, cont.

- Provided the error term is distributed normally, the RMSE tells us:
- 68.3% of the observations fall within the band that is ±1*RMSE of the regression line
- 95.4% of the observations fall within the band that is ±2*RMSE of the regression line
- 99.7% of the observations fall within the band that is ±3*RMSE of the regression line
- RMSE is also an element in calculating the standard errors of β_0 and β_1

Regression estimates, standard errors

$$SE(\beta_1) = \frac{RMSE}{\sqrt{\sum\left(x_i - \overline{x}\right)^2}}$$

$$SE(\beta_0) = RMSE \cdot \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum\left(x_i - \overline{x}\right)^2}}$$

Regression estimates, standard errors, cont.

- While these two standard error formulas may not appear very intuitive, we can glean some important information from them:
  1. As uncertainty about the regression line increases (RMSE increases), the standard errors of both β_0 and β_1 increase.
  2. As the variability of x increases, the standard errors of both β_0 and β_1 decrease.

Formal test of model fit, F-test

$$F_{k-1,N-k} = \frac{SSE/k-1}{SSR/n-k}$$

- Where k = the number of parameters in the model, and n is the sample size
- This is a general test of model fit. If the F-test is statistically significant, it means that the model explains some of the variance in Y.

Next time:

Homework: Problems 2.4i, 2.4ii, C2.4i, C2.4ii

Read: Wooldridge Chapters 19 & Appendix C.6, and Bushway, Sweeten & Wilson (2006) article