Regression Forced March

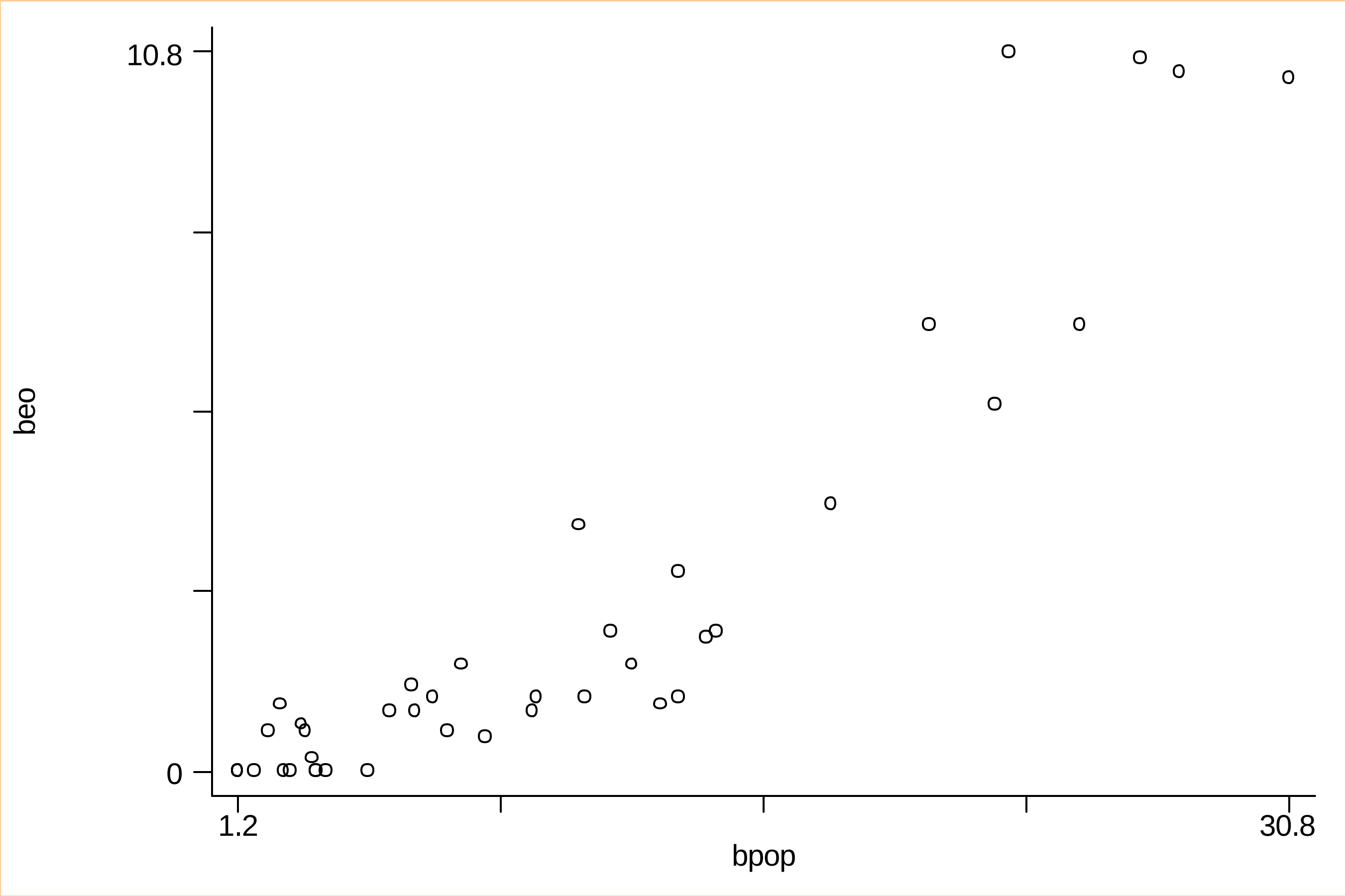
17.871

Spring 2002

2/28/02

Regression quantifies how one variable can be described in terms of another

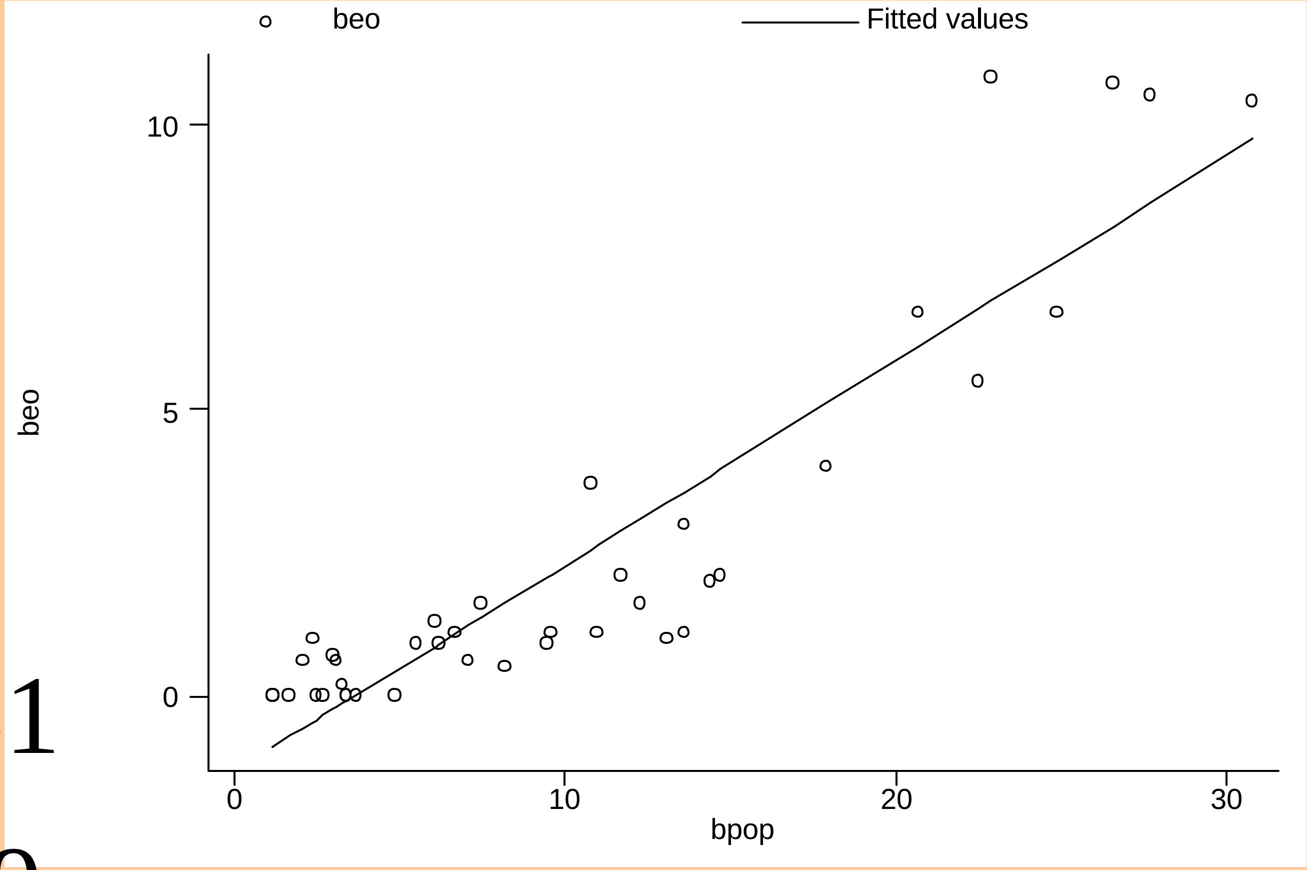
Black Elected Officials Example I



The Linear Relationship between Two Variables

$$Y_i = b_0 + b_1 X_i + e_i$$

The Linear Relationship between African American Population & Black Legislators

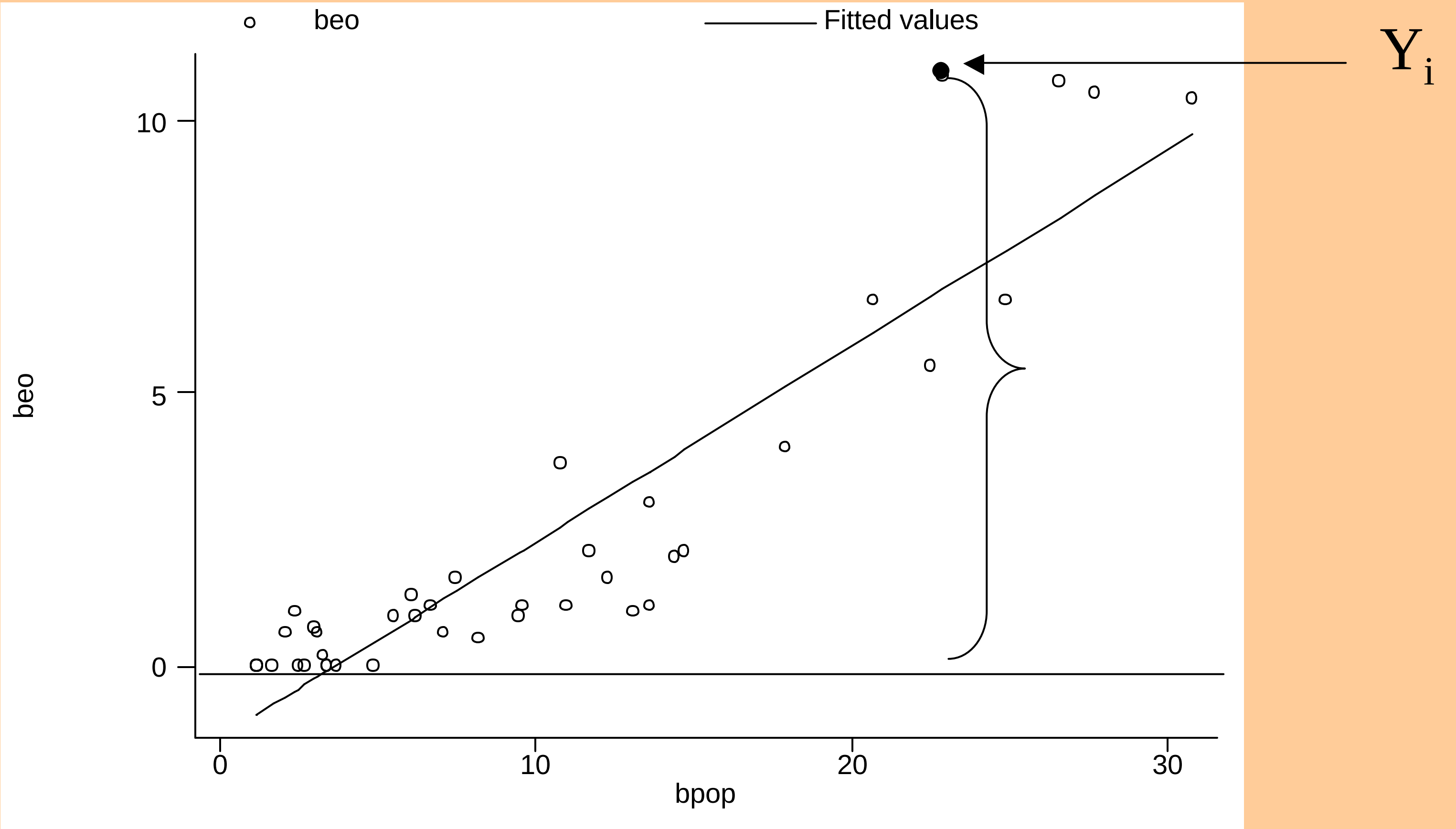


$$b_0 = -1.31$$

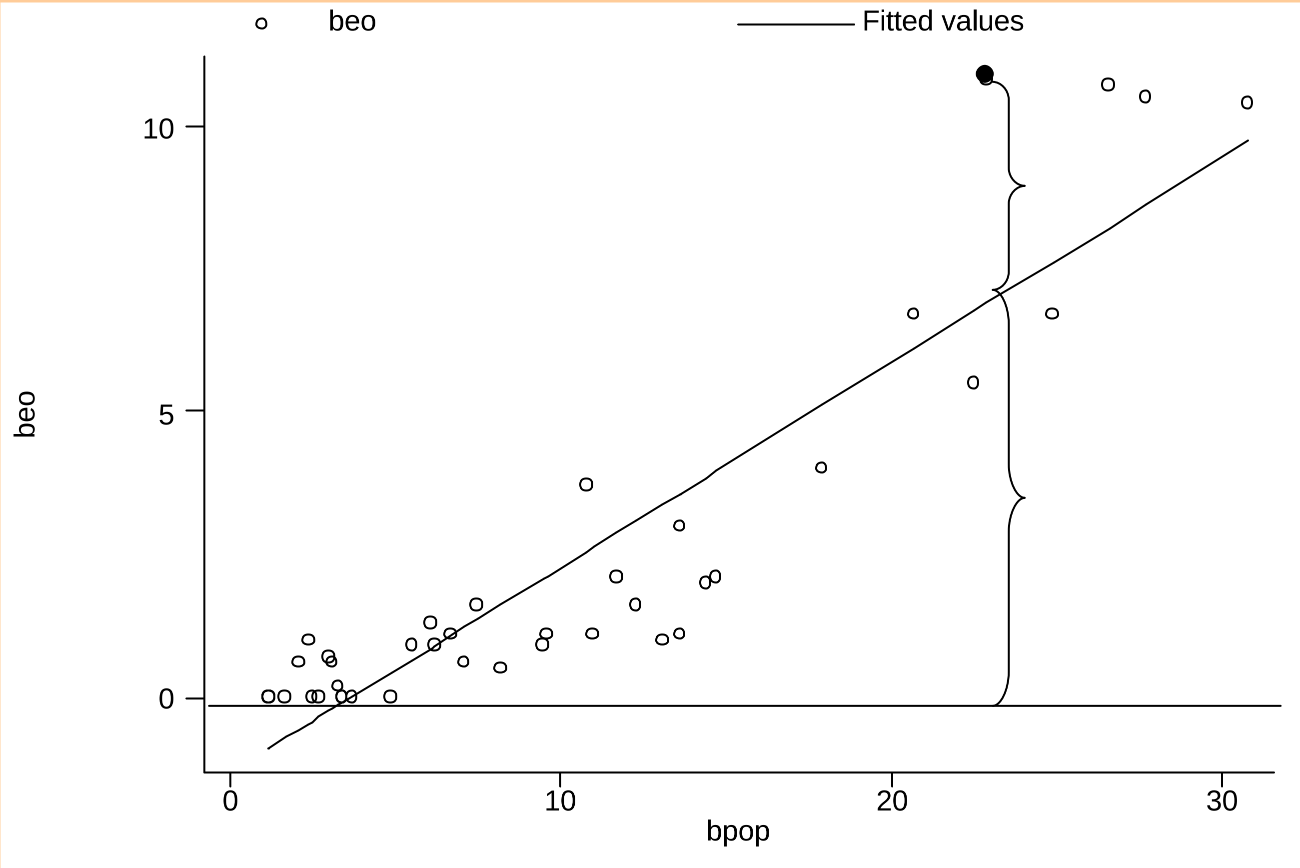
$$b_1 = 0.359$$

How did we get that line?

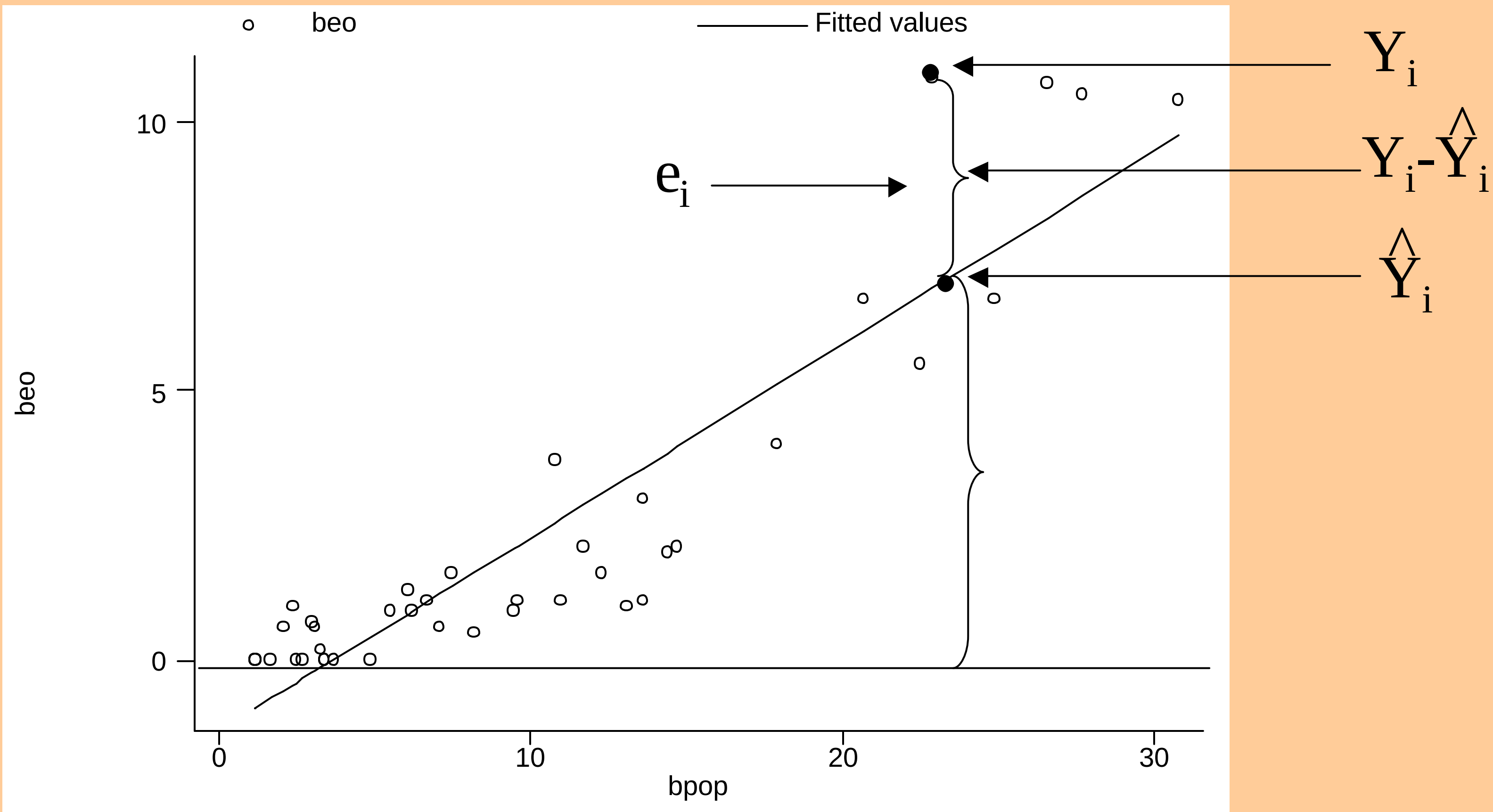
1. Pick a representative value of Y_i



How did we get that line?
2. Decompose Y_i into two parts



How did we get that line?
3. Label the points



The Method of Least Squares

Pick b_0 and b_1 to minimize

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

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2$$

Solve for $\dfrac{\partial \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2}{\partial b_1} = 0$

$$b_1 = \frac{\sum_{i=1}^{n} (\bar{Y} - Y_i)(\bar{X} - X_i)}{\sum_{i=1}^{n} (\bar{X} - X_i)^2} \text{ or}$$

$$\frac{\operatorname{cov}(X, Y)}{\operatorname{var}(X)}$$

Solve for $$\frac{\partial \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2}{\partial b_0} = 0$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

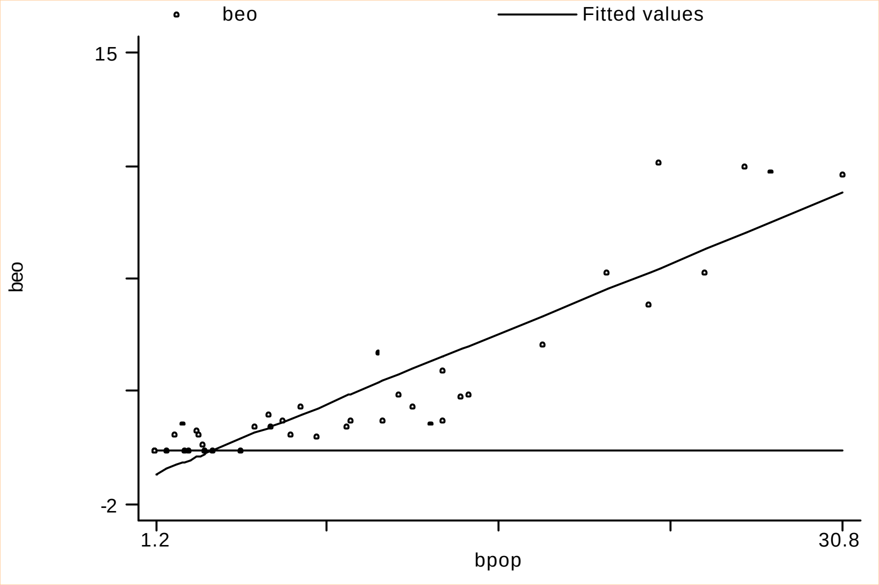
Note that if you rearrange.....

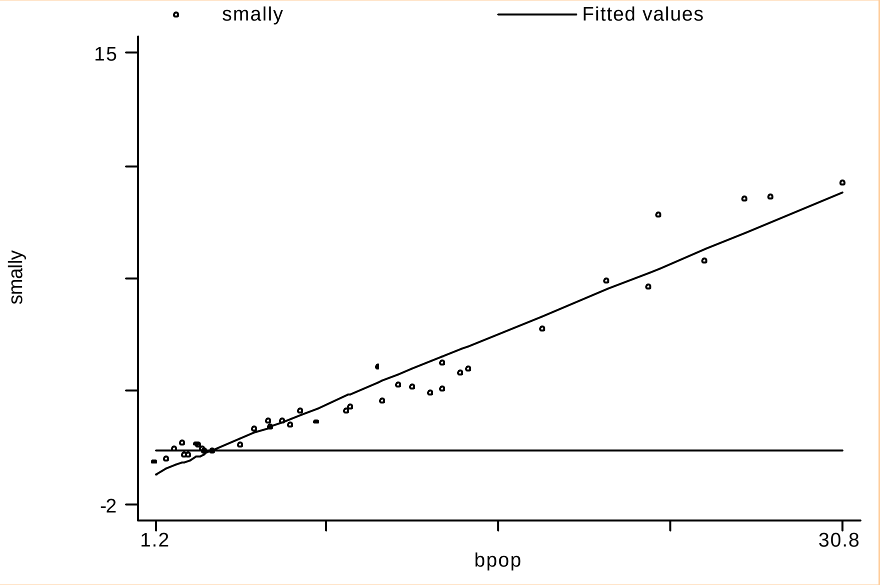
$$\bar{Y} = b_0 + b_1 \bar{X}$$

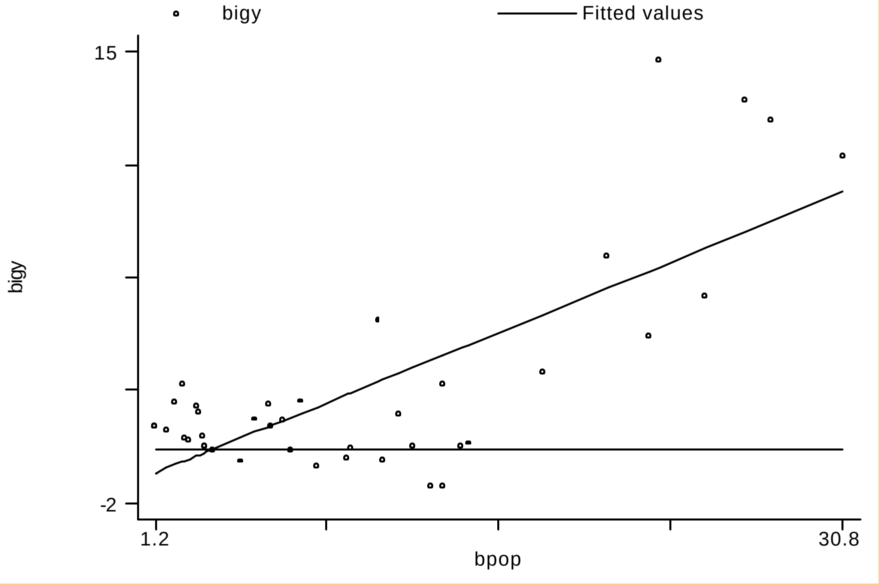
How to Think About Regression Results

- Deterministically
- Expectations

How “good” is the fitted line?







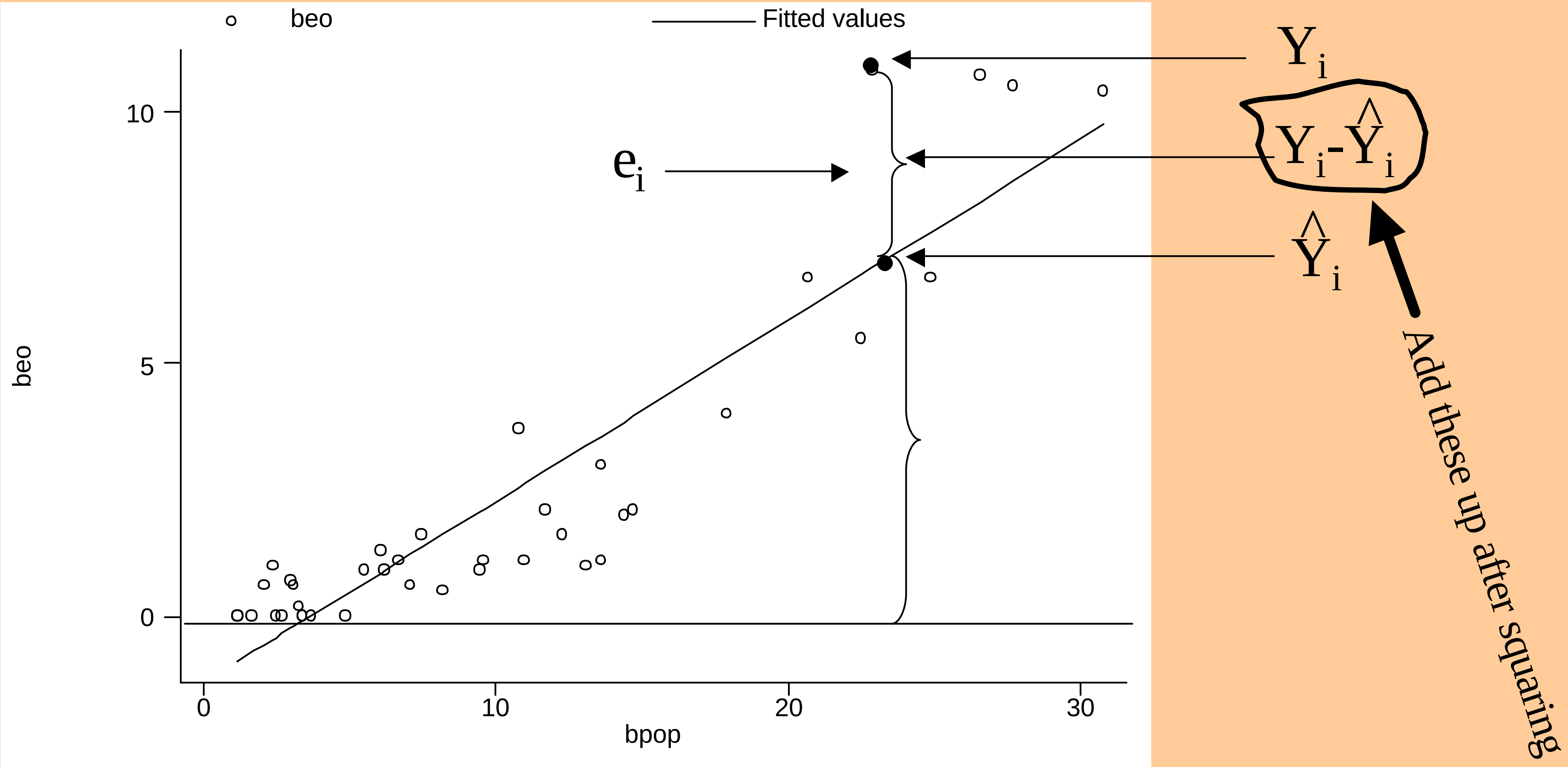
Judging results

- Substantive interpretation of coefficients
- Technical judgment of regression
  - Judgment of coefficients
  - Judgment of overall fit

Determining Goodness of Fit I

- Coefficients
  - Standard error of a coefficient
  - *t*-statistic: *coeff./s.e.*

Standard error of the regression picture

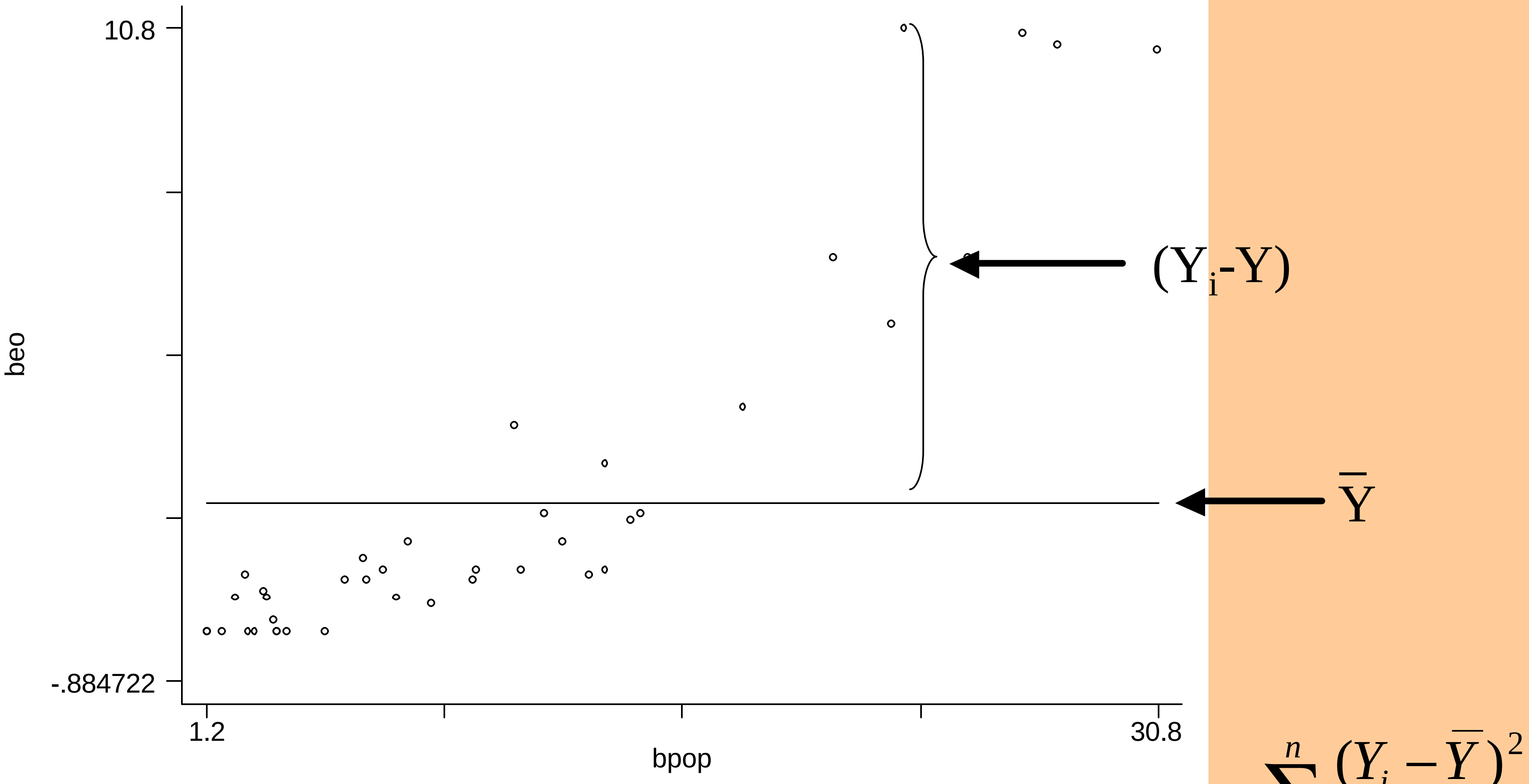


Determining Goodness of Fit

- Standard error of the regression (Root mean square error in STATA and Freedman)

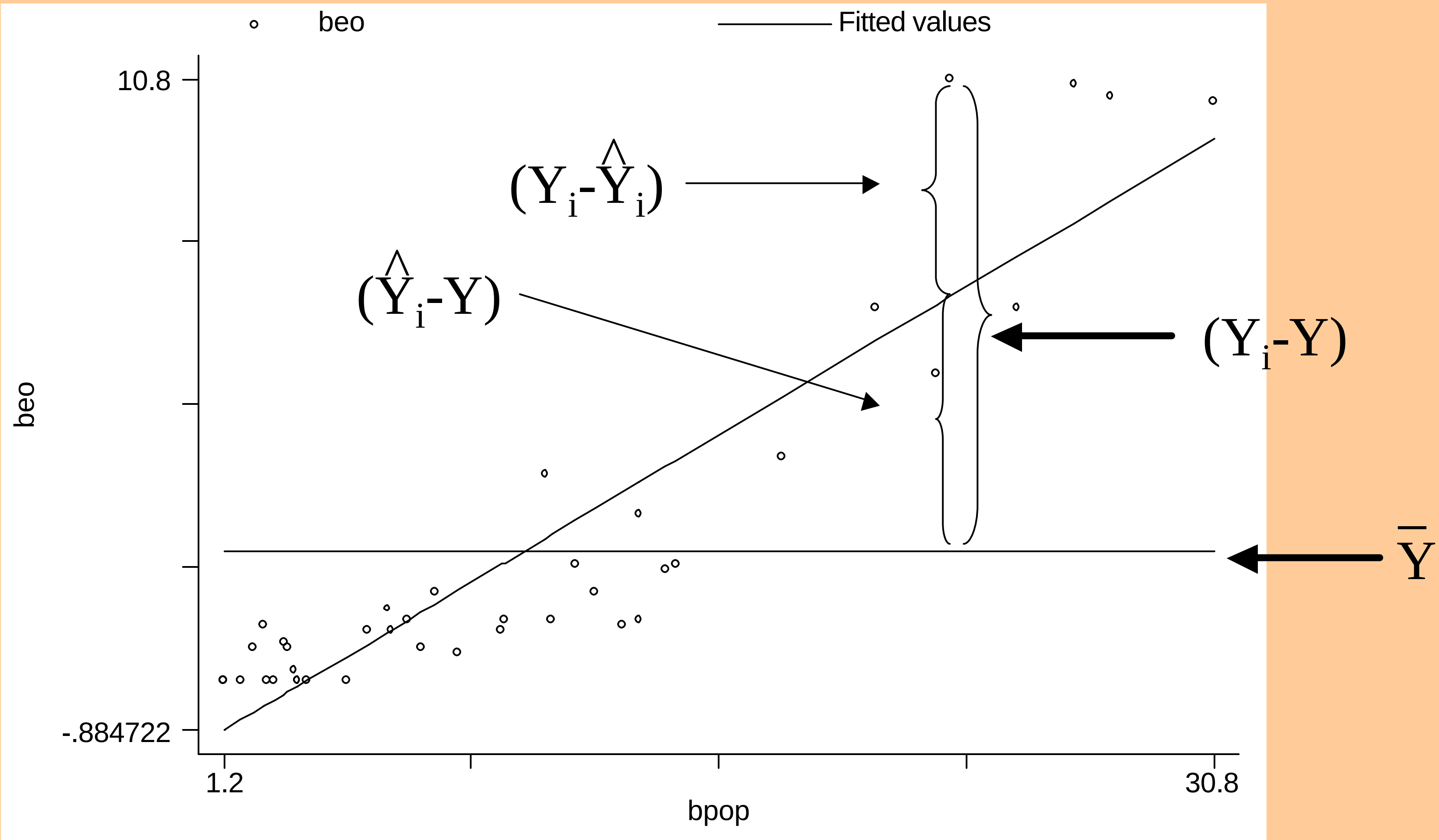
$$s.e.e. = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{d.f.}}$$

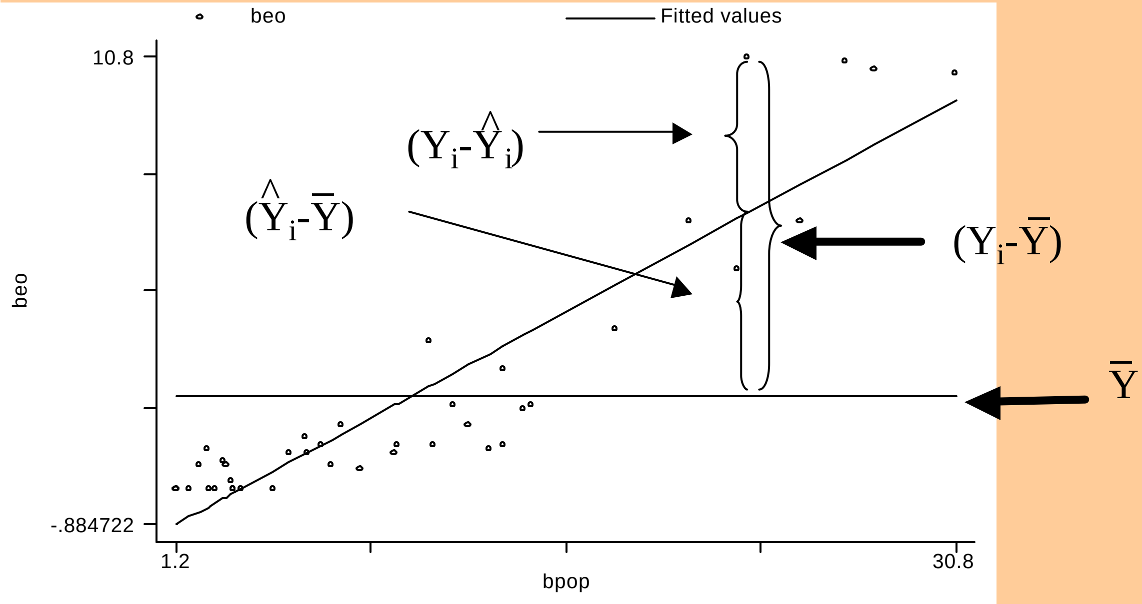
R^2 picture I



$$\sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{n}$$

R^2 picture II





$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \text{"}\textit{total sum of squares}\text{"}$$

$$\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = \text{"}\textit{regression sum of squares}\text{"}$$

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \text{"}\textit{residual sum of squares}\text{"}$$

Determining Goodness of Fit

- R-squared

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \quad \text{or}$$

$$\text{percent variance "explained"}$$

Return to Black Elected Officials Example

```
. reg beo bpop

      Source |       SS       df       MS              Number of obs =      41
-------------+------------------------------           F(  1,    39) =  202.56
       Model |   351.26542     1   351.26542           Prob > F      =  0.0000
    Residual |  67.6326195    39  1.73416973           R-squared     =  0.8385
-------------+------------------------------           Adj R-squared =  0.8344
       Total |  418.898039    40   10.472451           Root MSE      =  1.3169

------------------------------------------------------------------------------
         beo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bpop |   .3584751   .0251876    14.23   0.000     .3075284    .4094219
       _cons |  -1.314892   .3277508    -4.01   0.000    -1.977831   -.6519535
------------------------------------------------------------------------------
```

Looking at Residuals