Images created with STATA® software.

La marcha forzada de la regresión

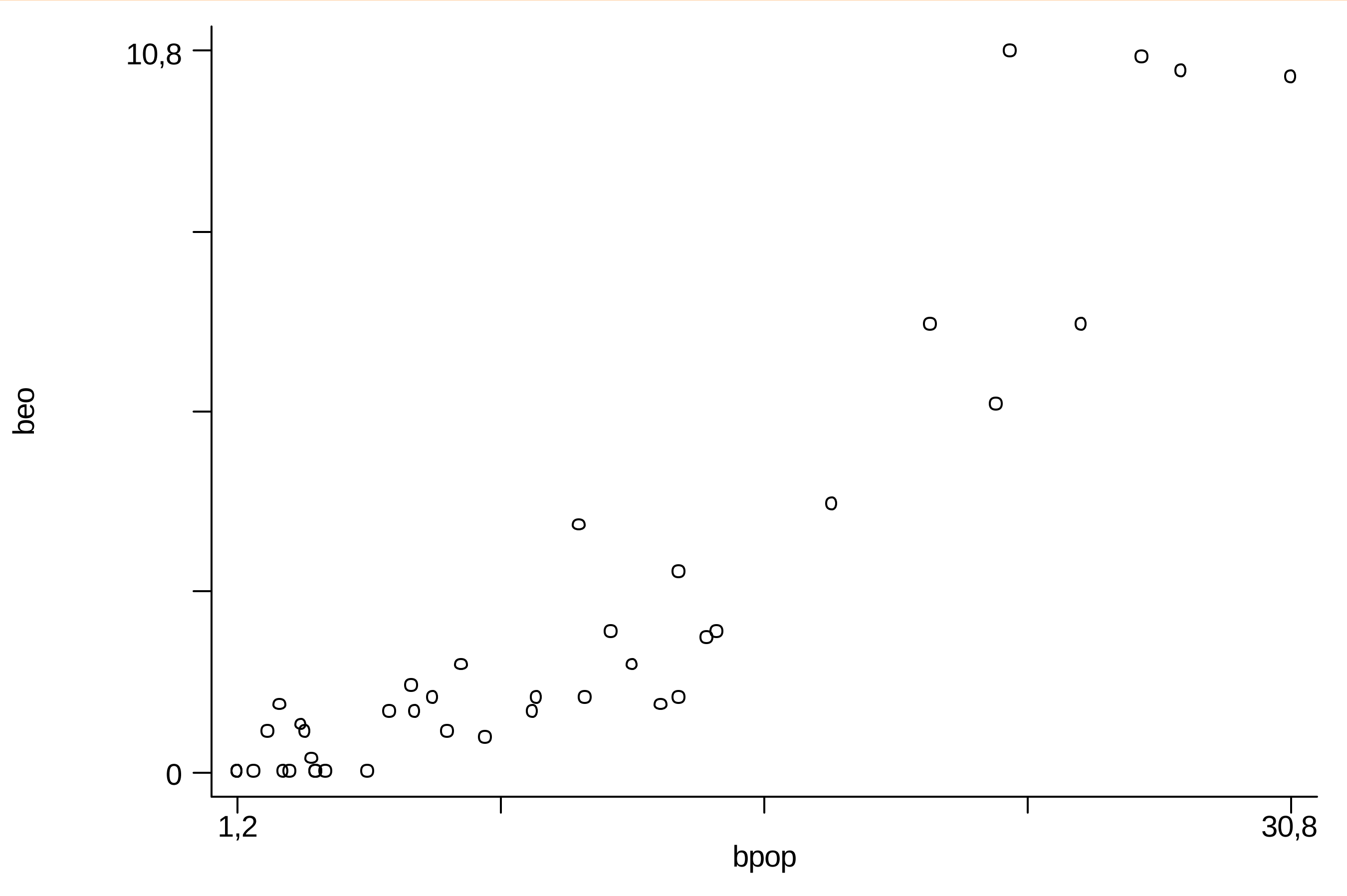
17.871

Primavera de 2002

28/02/02

La regresión cuantifica cómo se puede describir una variable respecto de otra

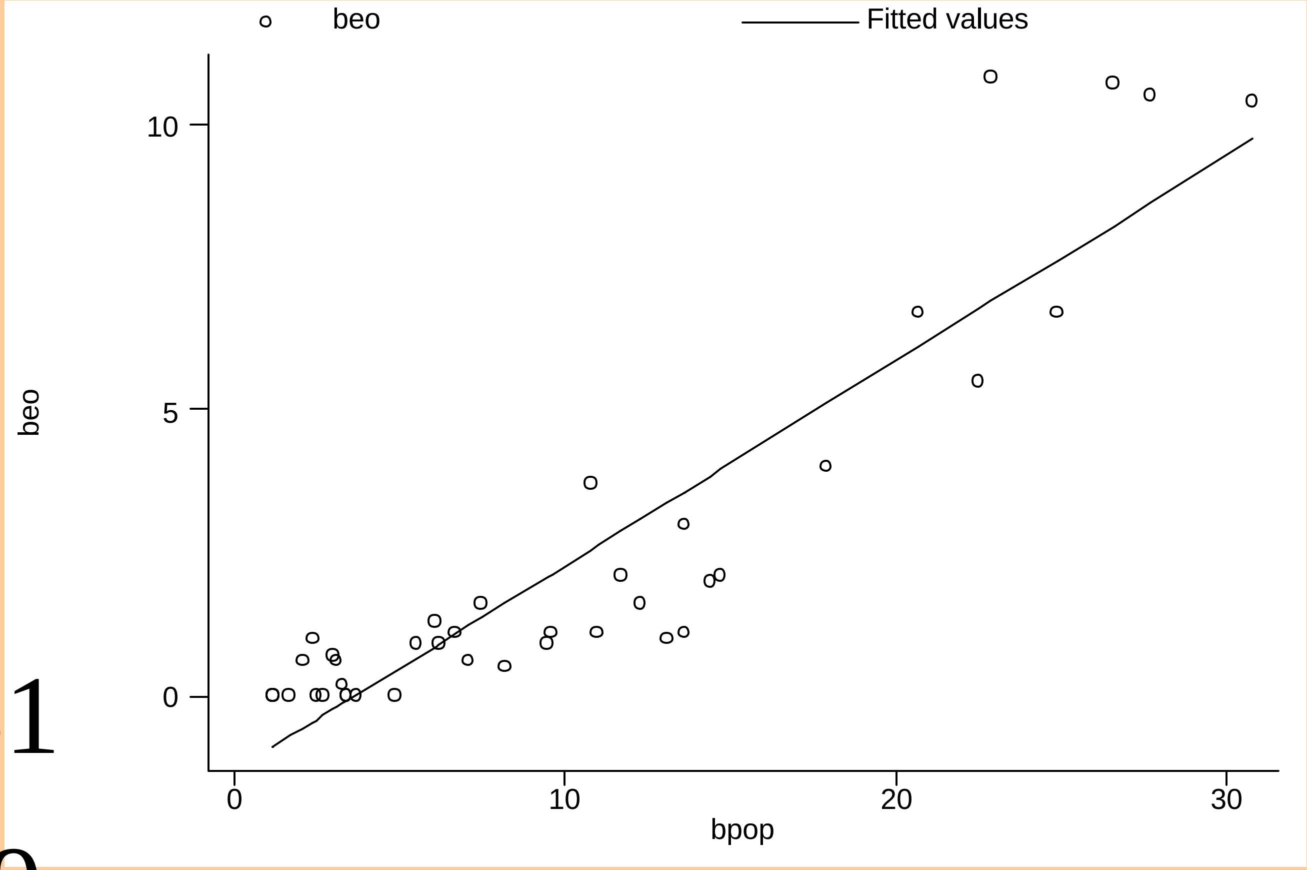
Cargos públicos negros elegidos, ejemplo I



La relación lineal entre dos variables

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

La relación lineal entre la población afroamericana y el número de legisladores negros

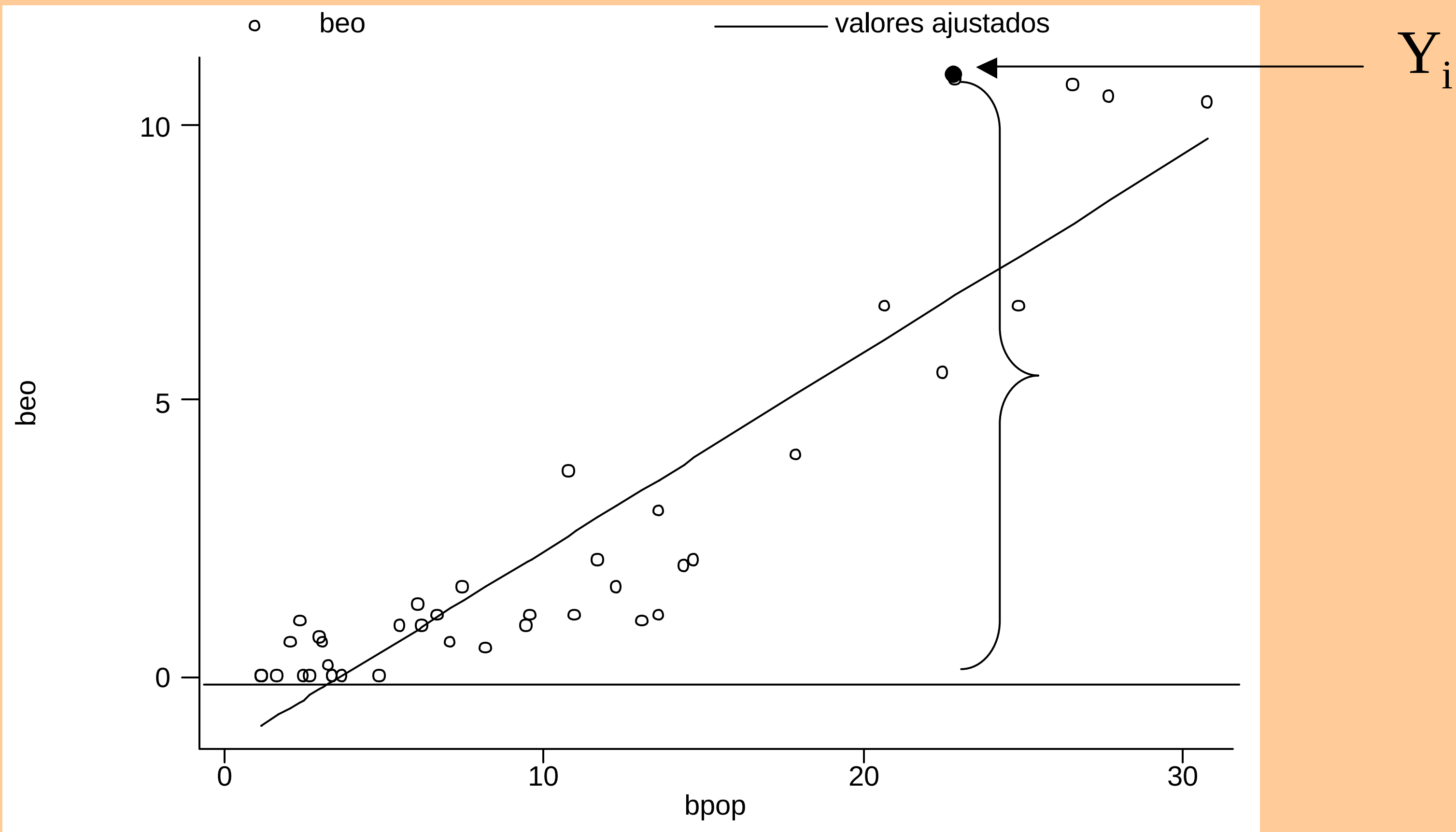


$$b_0 = -1.31$$

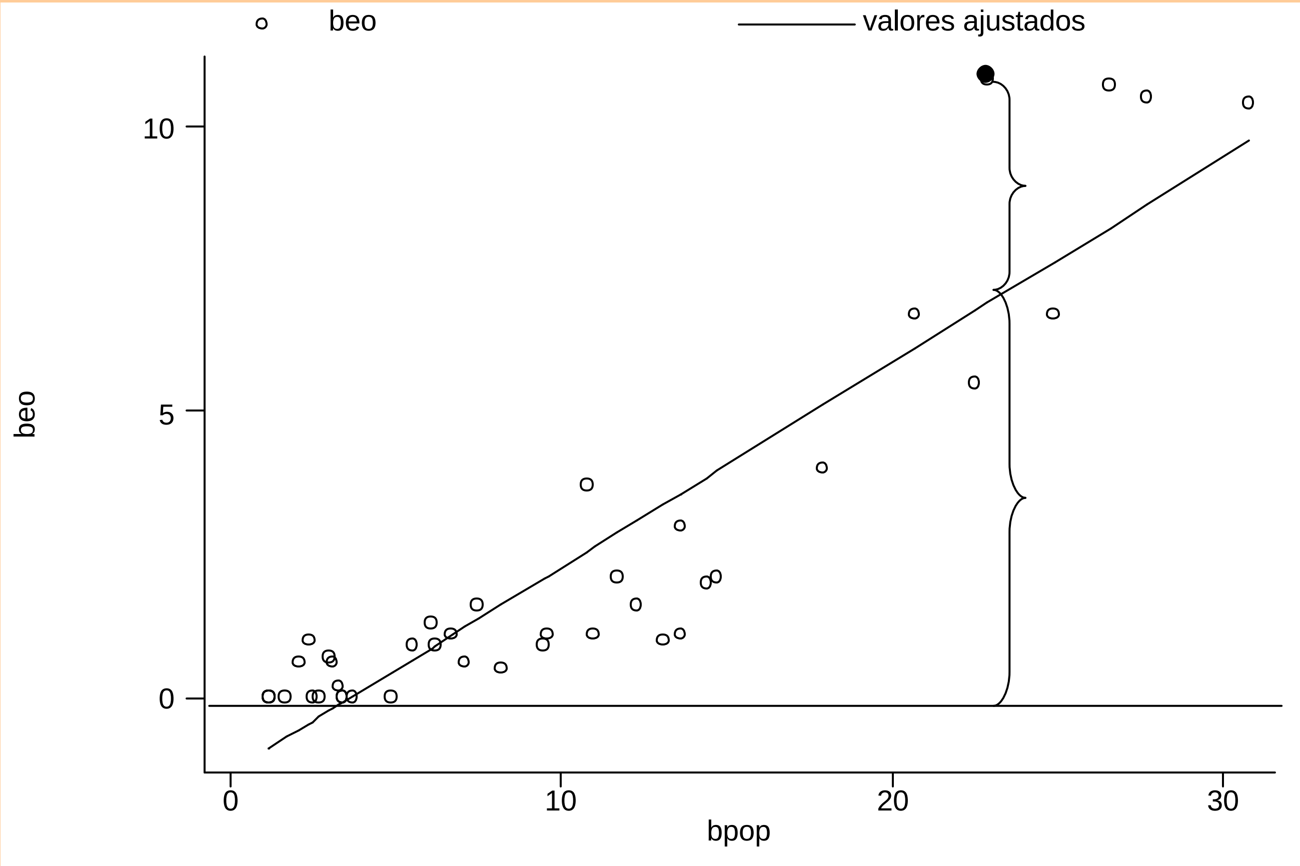
$$b_1 = 0.359$$

¿Cómo obtuvimos esa línea?

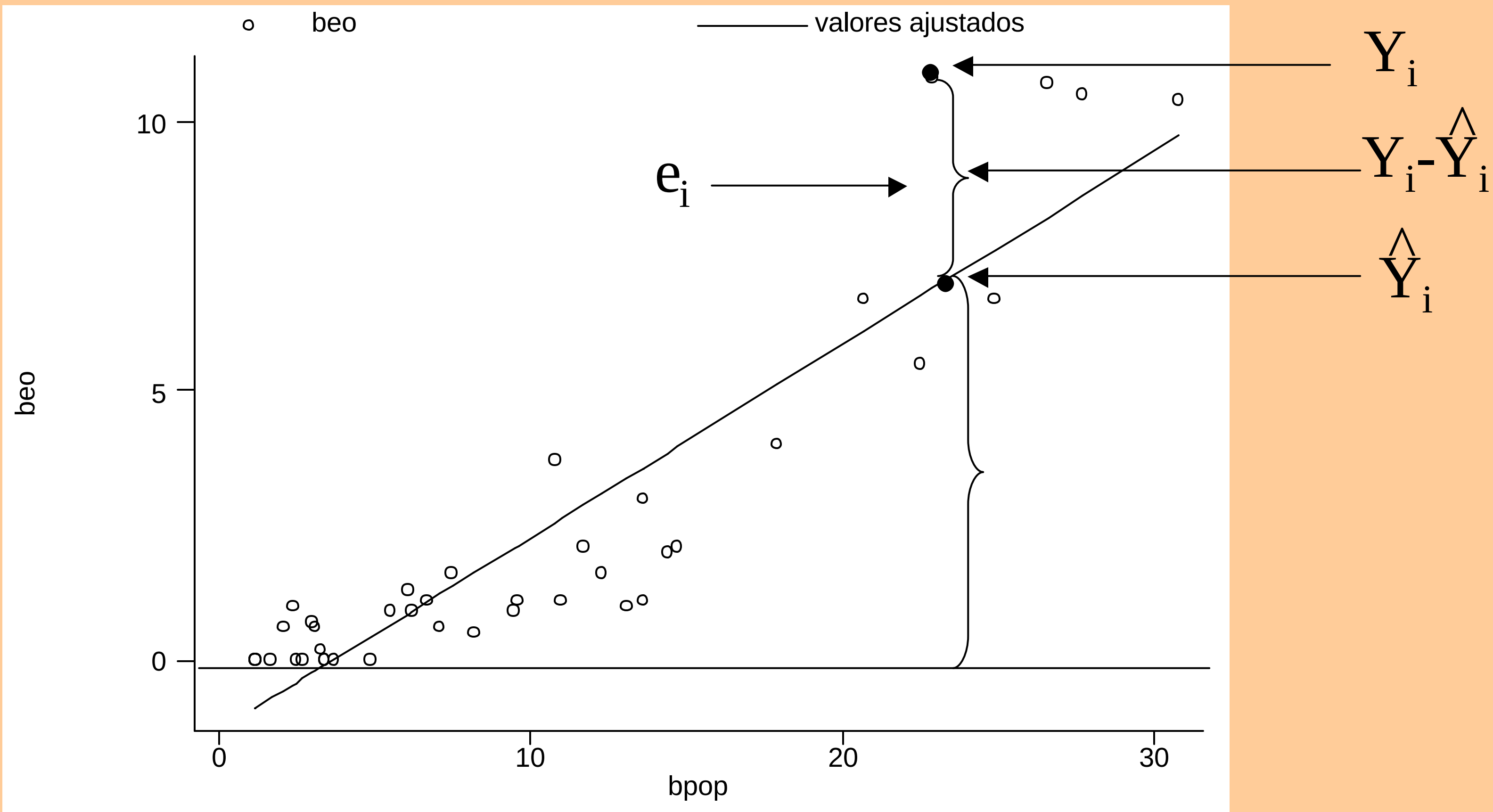
1. Escoja un valor representativo de Y_i



¿Cómo obtuvimos esa línea?
2. Descomponga Y_i en dos partes



¿Cómo obtuvimos esa línea?
3. Etiquete los puntos



El método de mínimos cuadrados

Escoja b_0 y b_1 para minimizar

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

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

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

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

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

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

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

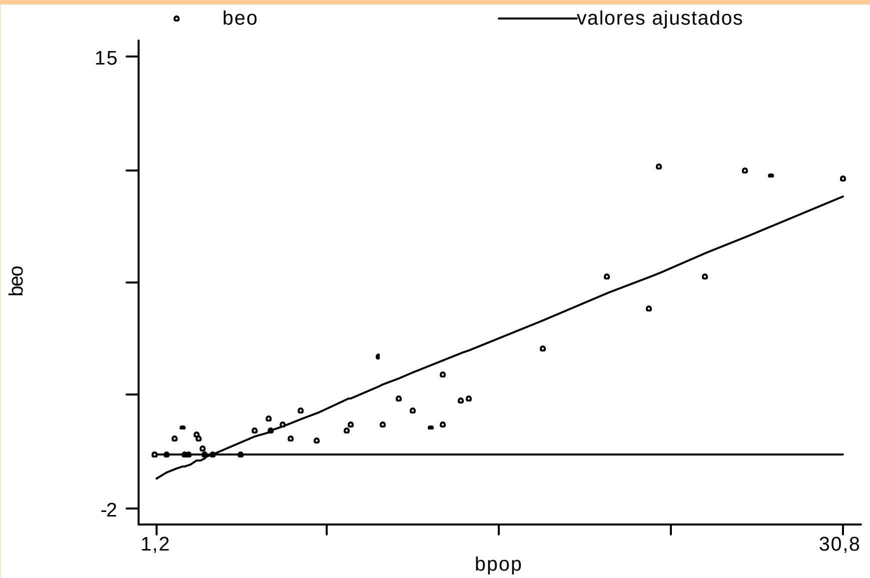
Observe que si reordena.....

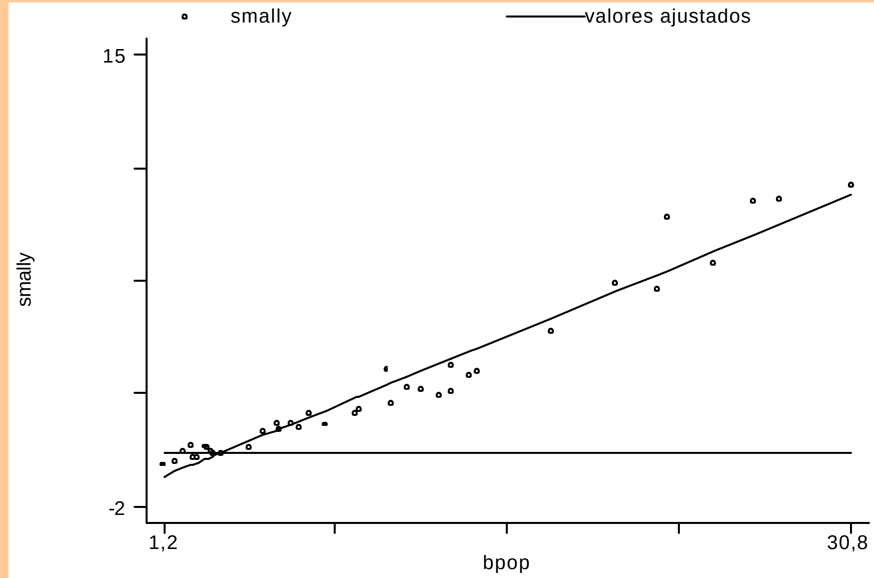
$$\overline{Y} = b_0 + b_1 \overline{X}$$

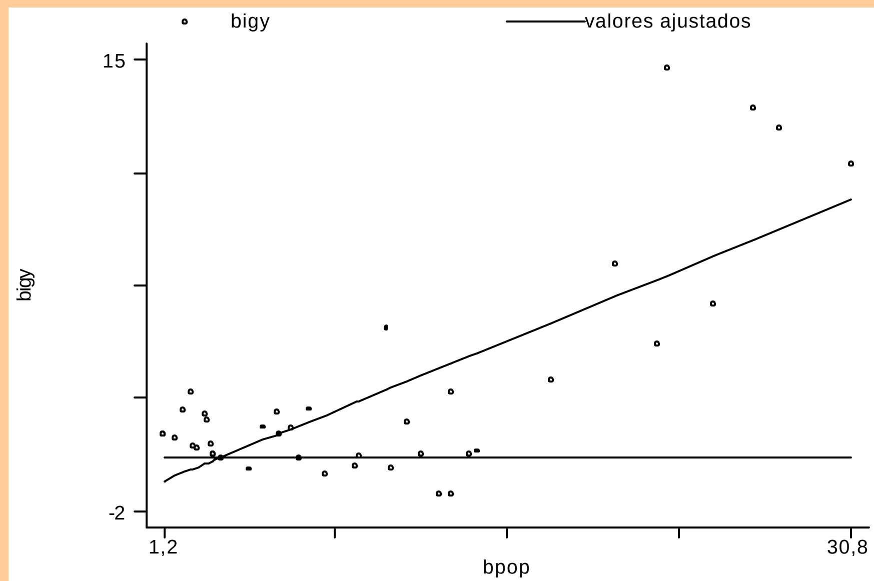
Cómo pensar en resultados de regresión

- Determinísticamente
- Expectativas

Bondad de la línea ajustada







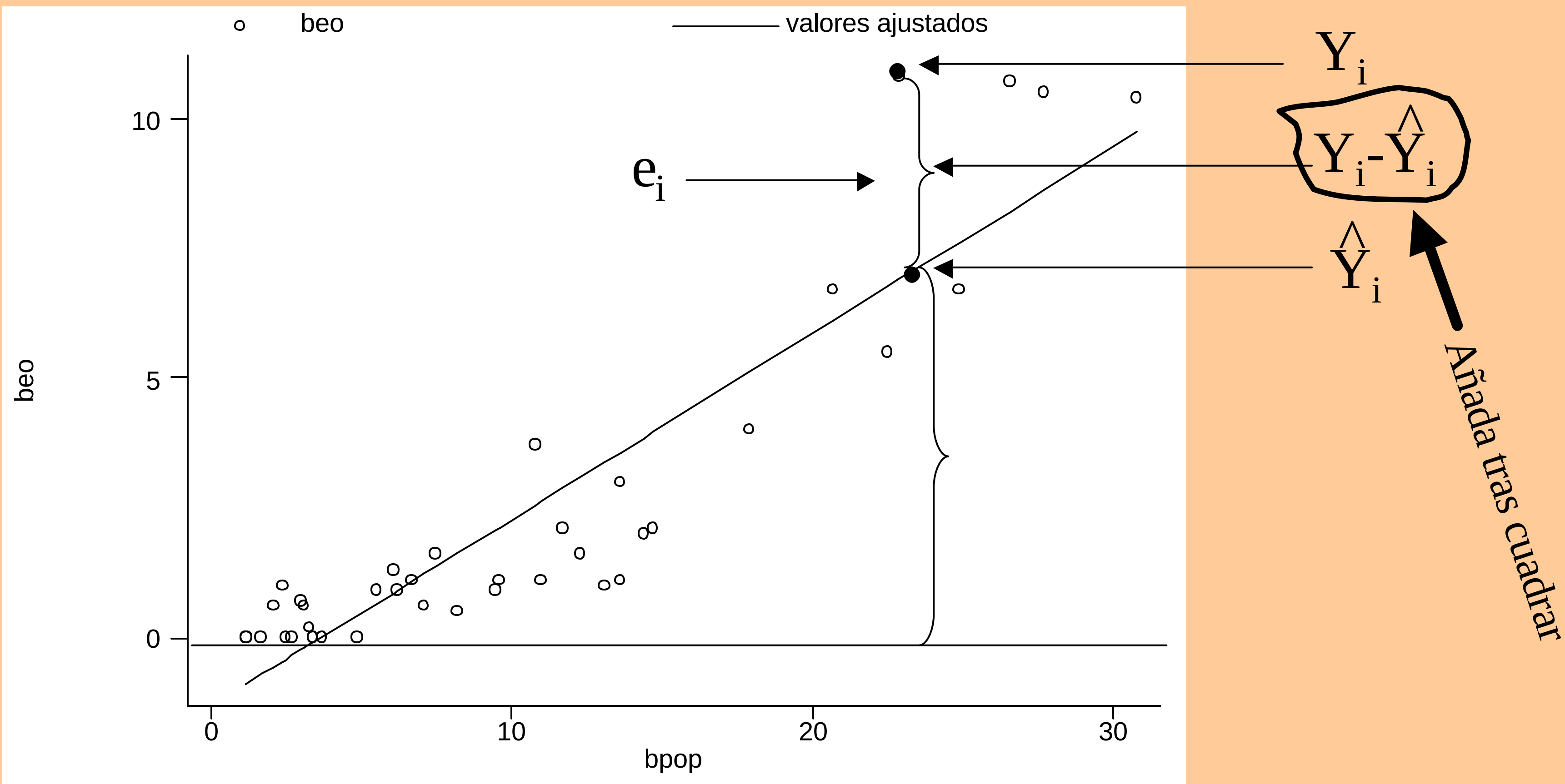
Evaluación de los resultados

- Interpretación sustantiva de los coeficientes
- Evaluación técnica de la regresión
  - Estimación de los coeficientes
  - Estimación del ajuste general

Determinación de la bondad del ajuste I

- Coeficientes
  - Error estándar de un coeficiente
  - *t*-estadística: *coeff./s.e.*

Error estándar del cuadro de regresión

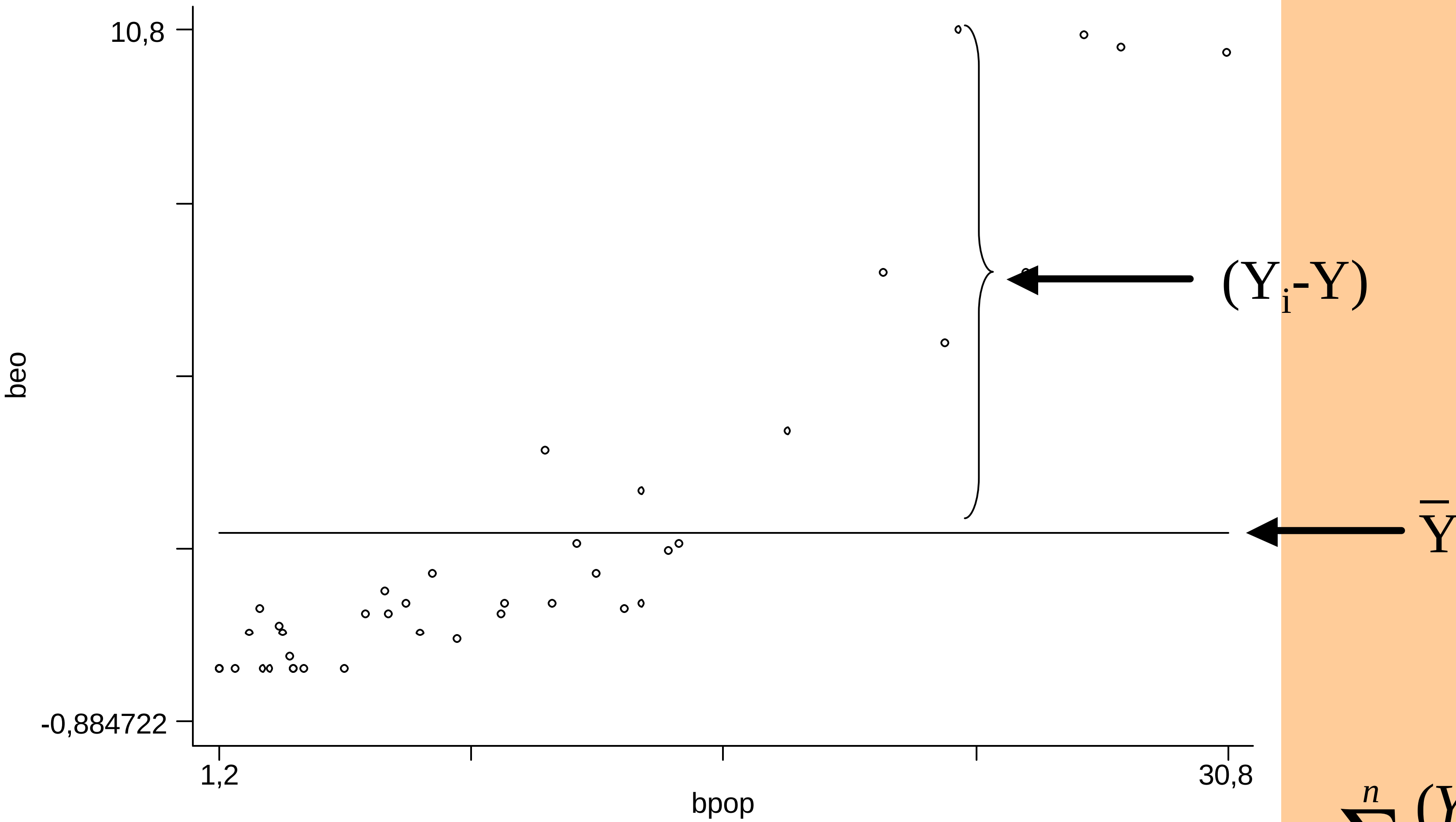


Determinación de la bondad del ajuste

- Error estándar de la regresión (error *root mean square* en STATA y Freedman)

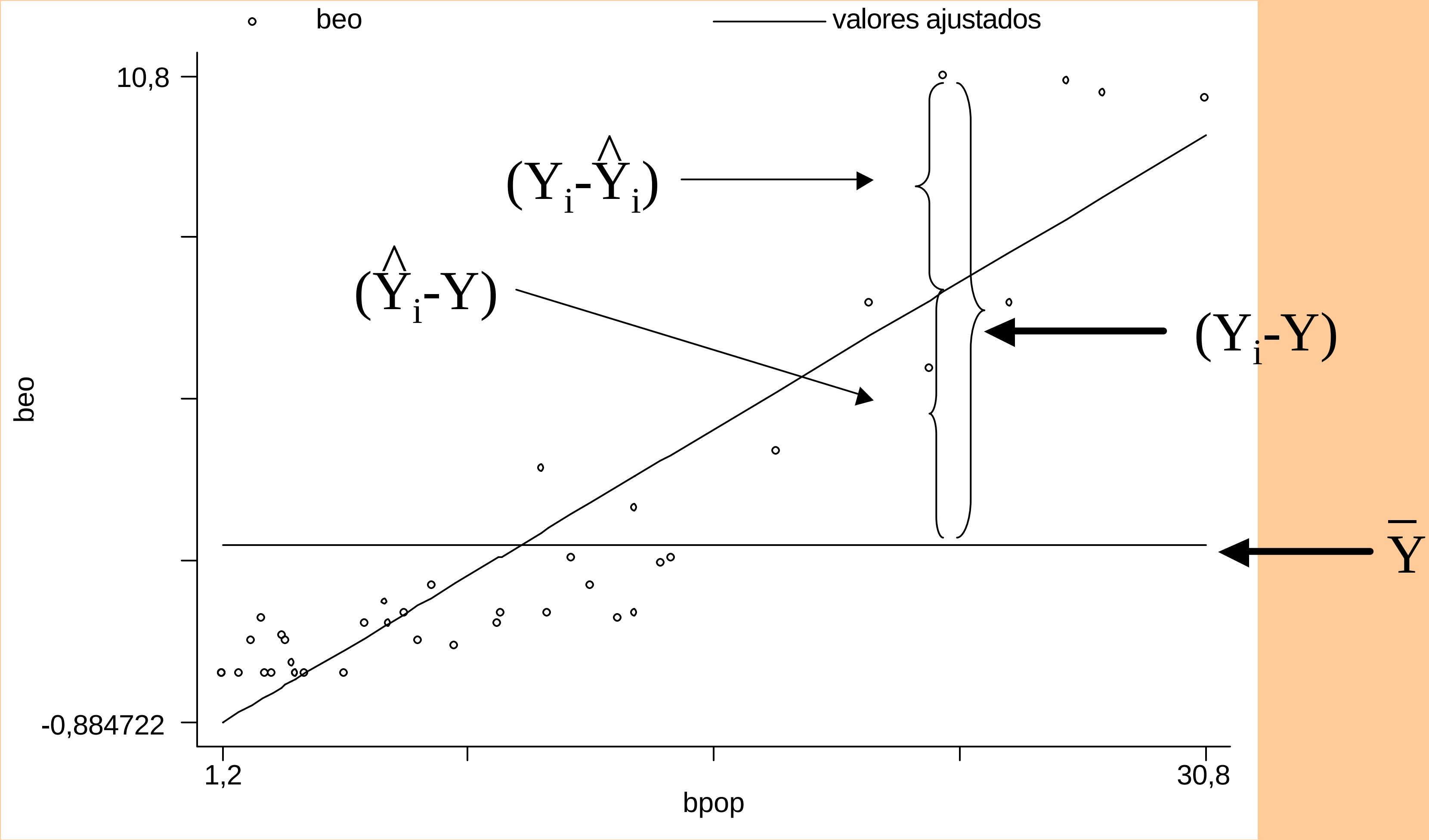
$$s.e.e. = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{d.f.}}$$

R^2 cuadro I

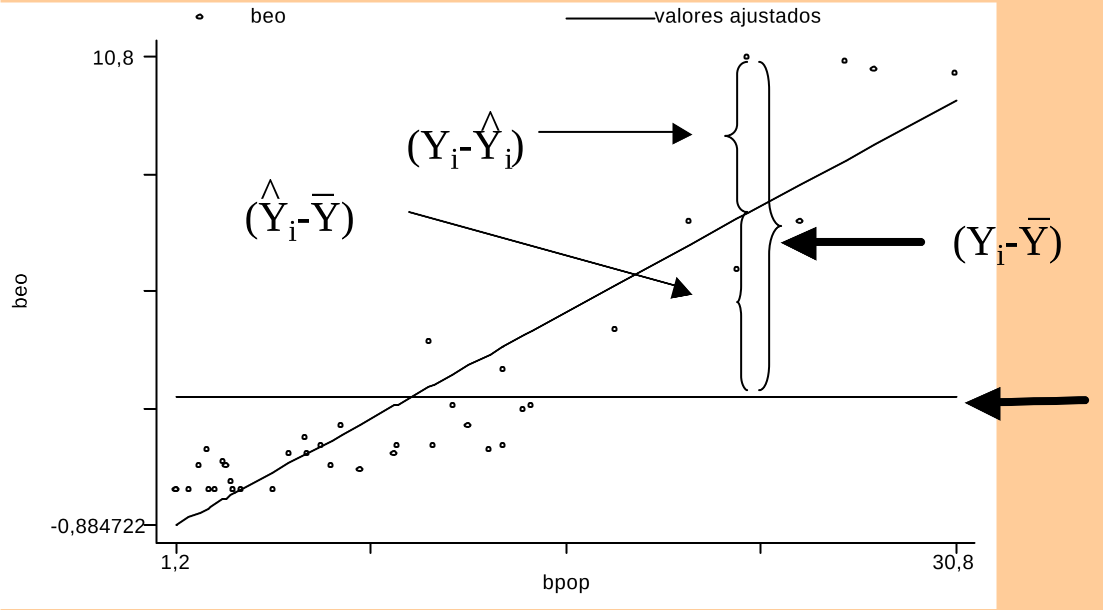


$$\sum_{i=1}^{n} \frac{(Y_i - \overline{Y})^2}{n}$$

R^2 cuadro II



beo
valores ajustados
10,8
-0,884722
1,2
30,8
bpop
beo
$(Y_i - \hat{Y}_i)$
$(\hat{Y}_i - \bar{Y})$
$(Y_i - \bar{Y})$
$\bar{Y}$



$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \text{"}\textit{suma total de cuadrados}\text{"}$$

$$\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 = \text{"}\textit{suma de regresión de cuadrados}\text{"}$$

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \text{"}\textit{suma residual de cuadrados}\text{"}$$

Determinación de la bondad del ajuste

- R-cuadrado

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

varianza de % "explicada"

De nuevo el ejemplo de los cargos públicos negros elegidos

```
. 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
------------------------------------------------------------------------------
```

Estudio de los residuales