UST 803 QUANTITATIVE Research Methods I

COMPUTER LAB#1: Regression

Data from Plain Dealer Sunday August 17,1997, Section A, pg1 and 23, *Prisoners not doing the same time for the same crime* by Mark Tatge. The article gives in a Table the percentage of Ohio prison inmates released annualy after review by the Ohio Adult Parole Authority.

$$\text{release} := \begin{pmatrix} 62 \\ 60 \\ 55 \\ 57 \\ 47 \\ 39 \\ 43 \\ 39 \\ 37 \\ 39 \\ 43 \\ 34 \\ 35 \\ 26 \\ 24 \\ 21 \\ 20 \end{pmatrix} \qquad \text{year} := \begin{pmatrix} 1980 \\ 1981 \\ 1982 \\ 1983 \\ 1984 \\ 1985 \\ 1986 \\ 1987 \\ 1988 \\ 1989 \\ 1990 \\ 1991 \\ 1992 \\ 1993 \\ 1994 \\ 1995 \\ 1996 \end{pmatrix} \qquad i := 0..16$$

Regress Release on Year, and compute

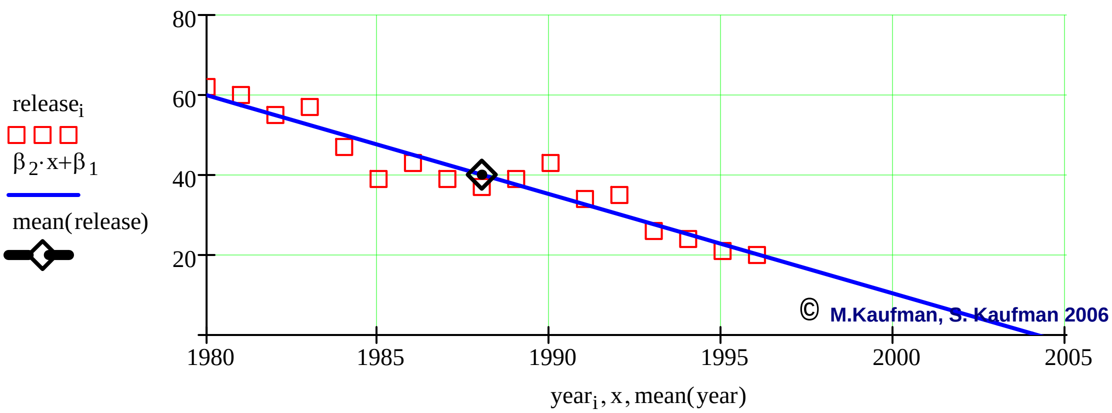
- **b estimates for intercept and slope (do you expect it to be positive or negative?)**
- **correlation coefficient between release and year (do you expect it to be positive or negative?)**

$\beta_2 := \text{slope(year, release)}$ $\quad$ $\beta_1 := \text{intercept(year, release)}$ $\quad$ $r := \text{corr(year, release)}$

$\beta_2 = -2.483$ $\quad$ $\beta_1 = 4.976 \times 10^3$ $\quad$ $r = -0.956$

Graph the scatter and the regression line between years 1980 and 2010 (beyond the sample)

$x := 1980 .. 2010$



The linear regression predicts that in the year 2004 the percentage of paroled inmates will be zero. (What does this mean? is it likely?)

We define the residuals e and check two OLS properties:

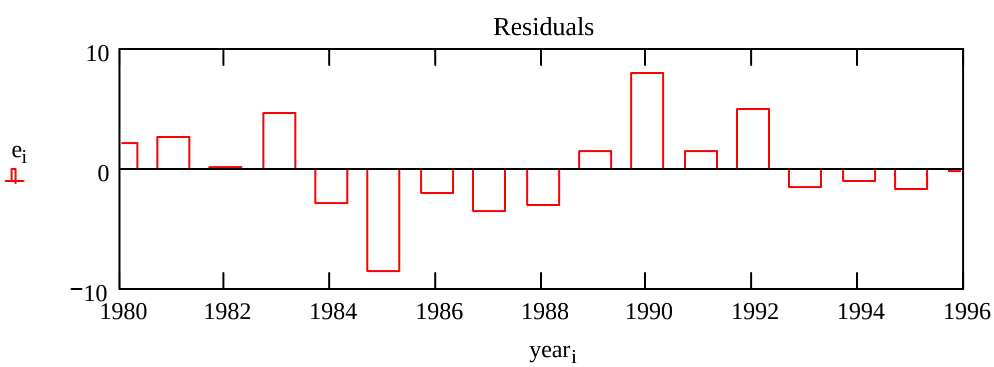
- sum of e is zero;
- sum of e times X is zero.

Note that due to the finite precision of the computer, the numerical values are very small, though not exactly zero.

$$e_i := release_i - \beta_1 - \beta_2 \cdot year_i$$

$$\sum_{j=0}^{16} e_j = 9.095 \times 10^{-12}$$

$$\sum^{16} e_j \cdot year_j = 1.808 \times 10^{-8}$$



Next we compute the standard error of estimate: σhat.

$$\sigma hat := \sqrt{\frac{1}{17-2}\cdot\sum_{j=0}^{16}\left(e_j\right)^2}$$

$\sigma hat = 3.965$ **Is this large? small? How can you tell?**

$\sigma hat^2 = 15.721$

We compute the variances, the standard errors and the covariance of the β parameters.

$$var\beta_2 := \frac{\sigma hat^2}{17\cdot var(year)} \qquad se\beta_2 := \sqrt{var\beta_2}$$

$var\beta_2 = 0.039$ $se\beta_2 = 0.196$

$$var\beta_1 := \frac{\sigma hat^2}{17\cdot var(year)}\cdot mean\left(\overrightarrow{year^2}\right) \qquad se\beta_1 := \sqrt{var\beta_1}$$

$var\beta_1 = 1.523\times 10^5$ $se\beta_1 = 390.241$

$$cov\beta1\beta2 := -\frac{\sigma hat^2}{17\cdot var(year)}\cdot mean(year)$$

$cov\beta1\beta2 = -76.603$

Why is the covariance of parameter estimates negative?