Influential Observations

Looking for trouble with the model.

Let $H = [h_{ij}]$, diagonal elements are h_{ii}. Moral of the story is small h_{ii} are good & big h_{ii} are bad.

(a) Average h_{ii} is small

$$tr(H) = tr(X(X'X)^{-1}X') = k+1$$

and $\frac{1}{n}\sum_{i=1}^{n} h_{ii} = \frac{k+1}{n} \longrightarrow 0$

(b) $0 \le h_{ii} \le 1$

First H is non-negative definite

$$v'Hv = v'H'Hv = (Hv)'Hv = z'z$$
$$= \sum_{j=1}^{n} z_j^2 \ge 0$$

Setting $v = 0$ except for a 1 in position i

$$v'Hv = h_{ii} \ge 0$$

Next.

$$\text{cov}(e) = \text{cov}((I-H)\underset{\sim}{y})$$
$$= (I-H)\sigma^2(I-H)'$$
$$= \sigma^2(I-H)$$

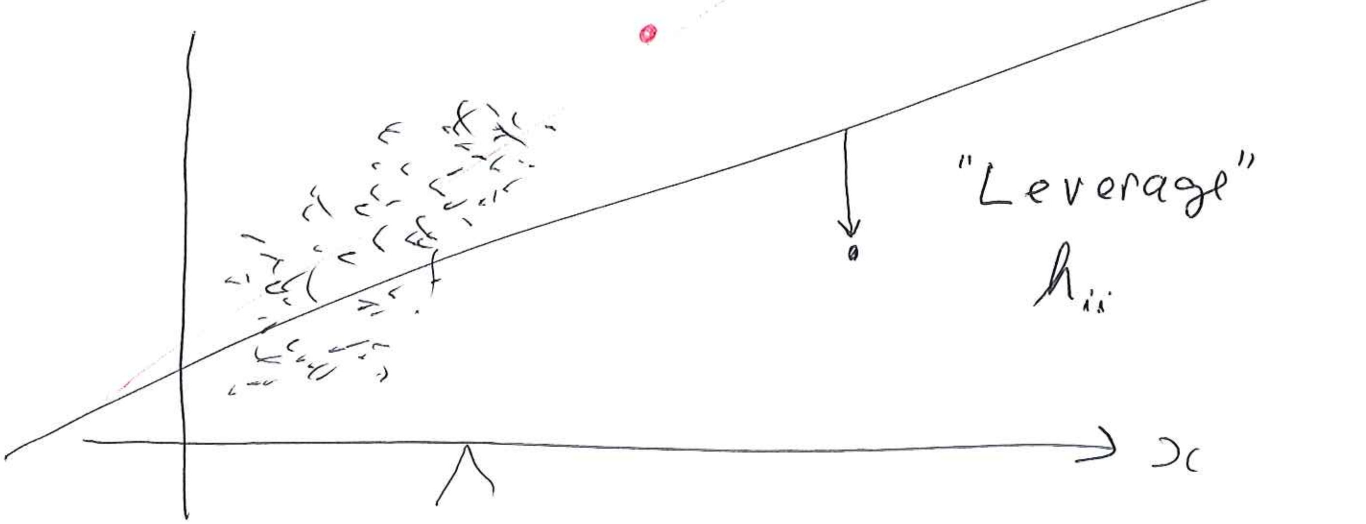
so $\text{var}(e_i) = \sigma^2(1-h_{ii}) \geq 0$

so $h_{ii} \leq 1$ & have $0 \leq h_{ii} \leq 1$

(c) h_{ii} <u>indirectly</u> reflects how far $\underset{\sim}{x}_i$ (row i of X) is from $\bar{\underset{\sim}{x}}$, vector of sample IV means.

"centroid"

Stannard section 8.2.1 p. 156



"Leverage" h_{ii}

(1) (a) Residuals e_i reflect ε_i better when h_{ii} are small.

$$y = X\beta + \varepsilon$$
$$y = Xb + e$$

Denote cols (rows) of H by $H = (h_1 \,\vdots\, h_2 \cdots \vdots\, h_n)$

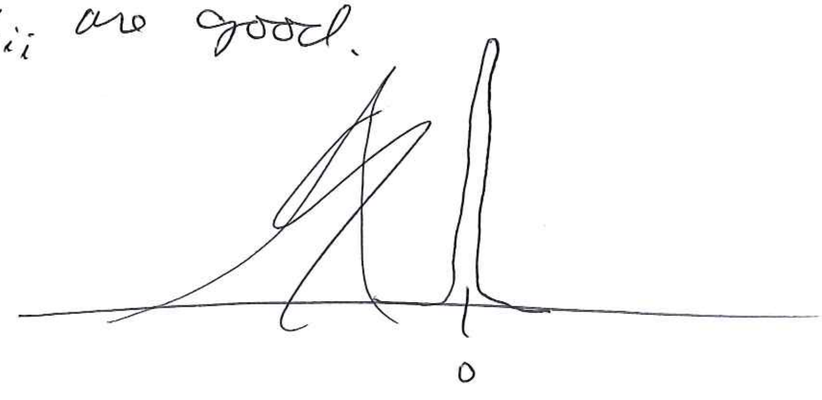
$$H = HH = h_i' h_i = h_{ii}$$

using $e = M\varepsilon = (I - H)\varepsilon = \varepsilon - H\varepsilon$

so
$$e_i = \varepsilon_i - \underbrace{h_i'\varepsilon}_{\text{difference term}}$$

$$h_i'\varepsilon \sim N(0, \sigma^2 h_i' h_i) = N(0, \sigma^2 h_{ii})$$

And small h_{ii} are good.



2) DFBETA $= b - b_{(i)}$

Transpose of row i

$$= \frac{(X'X)^{-1} x_i e_i}{1 - h_{ii}}$$

(8.8), p. 158

DFBETAS: e_i^* instead of e_i

3) DFFIT $= \hat{y}_i - \hat{y}_{(i)} = \frac{h_{ii} e_i}{1 - h_{ii}}$

p. 157

DFFITS

use e_i^*

4) Thm 5.1, p. 106

~~Tests are n.~~

Normality does not matter for tests & CIs provided $\max(h_{ii}) \to 0$ as $n \to \infty$

Rule of thumb

$\text{Max}(h_{ii}) < 0.2$