Multivariate

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \quad + \beta_{p-1} x_{ip-1} + \varepsilon_i$

$\rightarrow$ if 5 basic assumptions hold $\Rightarrow$ LS fit OK

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p-1} \\ 1 & x_{21} & x_{22} & & x_{2,p-1} \\ & & \vdots & & \\ 1 & x_{n1} & x_{n2} & & x_{np-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\underset{n \times 1}{\underset{\sim}{y}} = \underset{n \times p}{\mathbb{X}} \quad \underset{p \times 1}{\underset{\sim}{\beta}} + \underset{n \times 1}{\underset{\sim}{\varepsilon}}$$

Minimize LS $\quad Q = \sum (y_i - \underset{\sim}{x}_i' \underset{\sim}{\beta})^2$ where $\underset{\sim}{x}_i' = (1, x_{i1}, \ldots \quad x_{ip-1})$

$= \sum e_i^2$

$= \underset{\sim}{e}'\underset{\sim}{e} = \underbrace{(y - \mathbb{X}\beta)'(y - \mathbb{X}\beta)}_{\text{Scaler } 1 \times 1}$

$\Rightarrow \frac{\partial Q}{\partial \beta} = 0 \quad \Rightarrow$ $\boxed{\text{Normal Equations } \quad (\mathbb{X}'\mathbb{X})\underset{\sim}{\beta} = \mathbb{X}'y}$

(A) In simple linear regression $\rightarrow$ solution $\hat{\beta}_1 = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_j - \bar{x})^2} = \dfrac{\mathrm{Cov}(x, y)}{\mathrm{Cov}(x, x)} \sim \mathrm{Corr}(x, y)$

(B) Multiple case $\rightarrow$ $(\mathbb{X}'\mathbb{X}) \sim \mathrm{Cov}(\mathbb{X})$ where $V(x_1), V(x_2) \ldots \quad V(x_{p-1})$ on the diagonal, and cross-covariance $\mathrm{Cov}(x_1, x_2)$, $\mathrm{Cov}(x_k, x_l)$ off-diagonal

$$\rightarrow (X'y) = \begin{pmatrix} \mathrm{Cov}(x_1, y) \\ \mathrm{Cov}(x_2, y) \\ \vdots \\ \mathrm{Cov}(x_{p-1}, y) \end{pmatrix}$$

- If $(\underline{X}'\underline{X}) = \Lambda$ (diagonal), all x's are uncorrelated.

→ then $\hat{\beta} = (\underline{X}'\underline{X})^{-1}\underline{X}'y \sim \begin{pmatrix} \mathrm{corr}(x_1, y) \\ \vdots \\ \mathrm{corr}(x_{p-1}, y) \end{pmatrix}$

i.e. each coefficient has a <u>direct</u> interpretation as the dependency between an individual x-variable and y.

- If $(\underline{X}'\underline{X})$ not diagonal → no direct interpretation

→ can't separate influence of one x on y from another (correlated) x.

→ That is, $\hat{\beta}_k$ involves both $\mathrm{corr}(x_k, y)$ and

$\mathrm{corr}(x_\ell, y)$, $\ell \neq k$.
$\mathrm{corr}(x_\ell, x_k)$

- Moreover → if x's are highly correlated, $(\underline{X}'\underline{X})^{-1}$ may ~~not~~ exist! $(\det(\underline{X}'\underline{X}) \to 0)$

or inverse $(\underline{X}'\underline{X})^{-1}$ numerically unstable

→ When this happens, magnitude & even sign of $\hat{\beta}$'s may become meaningless. [Many models fit the data equally well]

Ex i) $X_2 = a + b X_1$

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

} infinite set of β_0 & β_1 fit these data equally well

The Hat-matrix

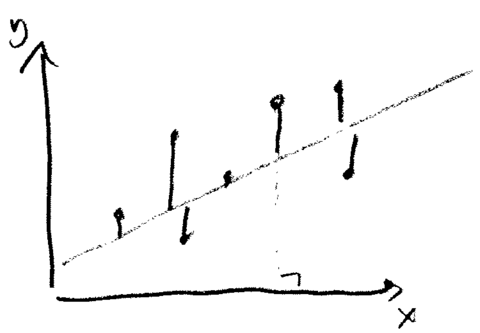
Assume we can solve the normal equations & obtain

$$\hat{\beta} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'y$$

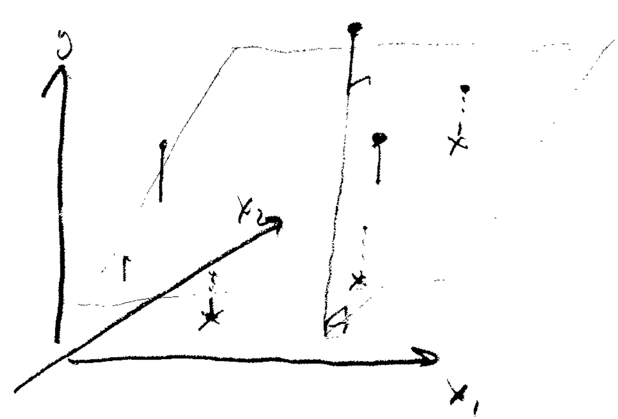
- Fitted values $\underset{\sim}{\hat{y}} = \underset{\sim}{X}\hat{\beta} = \underbrace{\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'}\,\underset{\sim}{y}$

  Hat-matrix $H = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$

- H is a <u>projection</u> matrix — projects the data $\underset{\sim}{y}$ onto the plane spanned by $\underset{\sim}{X}$



Simple case



projection $\perp$ to both x_1 & x_2 in multiple case.

Facts

- $HH = H$ { (H is an idempotent matrix)
  ($H' = H$) (No point in redoing regression)
  (Symmetric)

- $e = \hat{\varepsilon} = y - \hat{y} = y - Hy = (I - H)y$

- $X'e = X'(I - X(X'X)^{-1}X')y = 0$

  (orthogonal projection)

- $\hat{y}'e = y'H(I-H)y = 0$

  (fitted values $\perp$ to residuals)

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & & & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

ith row of X

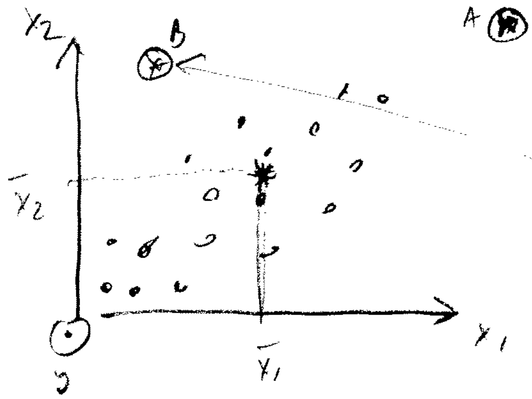
leverage $h_{ii} = \underset{\sim}{x}_i (X'X)^{-1} \underset{\sim}{x}_i'$

Note, h_{ii} is large when $\underset{\sim}{x}_i$ is extreme compared with $\bar{\underset{\sim}{x}}$

vector, average X-values.

vector of all x-variable values for observation i

Extreme how? wrt mass of data



Extreme in both x_1 & x_2

Extreme cmp. w. mass of data

(extreme in x_2 only.)

Both ⊙ observations have high leverage.

Properties

- As before, $E(\hat{\beta}) = \beta$ (unbiased)

- $V(\hat{\beta}) = V(\underbrace{(X'X)^{-1}X'}_{\text{constant}}y) = (X'X)^{-1}X' \cdot \underbrace{V(y)}_{=\sigma^2 I} \cdot X(X'X)^{-1} = \sigma^2(X'X)^{-1}$

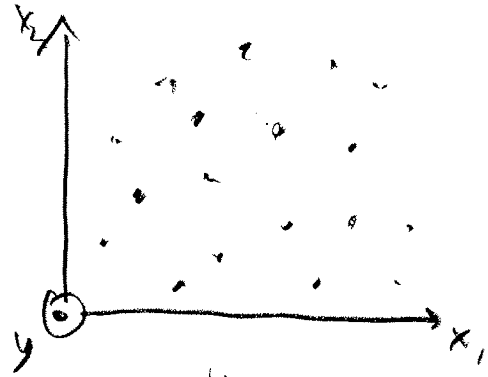
Meaning? $\rightarrow$ More spread in any x ($var(x_k)$ large)

$\rightarrow$ better estimate of $\hat{\beta}$.

$\rightarrow$ $\sigma^2 \uparrow$ $\rightarrow$ variance $\uparrow$

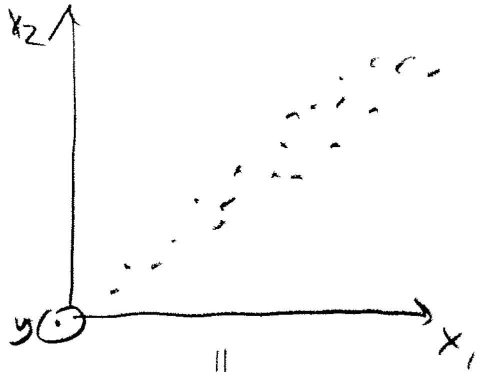
$\rightarrow$ Unless $(X'X)$ is diagonal, $\hat{\beta}$'s are correlated $\Rightarrow$ Think about impact on inference!

$\rightarrow$ The more dependent x's are, the larger the $V(\hat{\beta})$.



$\Downarrow$

$\hat{\beta}_1$ & $\hat{\beta}_2$ nearly uncorrelated



$\Downarrow$

$\hat{\beta}_1$ & $\hat{\beta}_2$ heavily dependent

$\rightarrow$ large variance associated with estimates

$\rightarrow$ Difficult to make direct inferences.

Fitted values $\quad \hat{y} = Hy \quad \Rightarrow \quad E(\hat{y}) = E(y)$

$$V(\hat{y}) = H\,V(y)\,H = \sigma^2 H$$

(i.e. large variance @ locations of high leverage)

Residuals $\quad e = (I - H)y \quad \Rightarrow \quad E(e) = 0$

$$V(e) = \sigma^2 (I - H)$$

$\rightarrow$ smaller variance @ locations w. high leverage

$\downarrow$

e's are correlated!