Multiple Linear Regression.

$$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad \underset{\sim}{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad X = \begin{array}{c} \begin{array}{ccccc} \beta_0 & \beta_1 & \beta_2 & \cdots & \beta_k \end{array} \\ \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & & x_{22} \\ & & & & \\ 1 & x_{n1} & x_{n2} & & x_{nk} \end{pmatrix} \end{array} \quad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$\underset{n\times 1}{\underset{\sim}{Y}} = \underset{n\times(k+1)}{X}\,\underset{(k+1)\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{\epsilon}}.$$

$$= [\cancel{c_1}\ c_2 \cdots c_k] = [1\ c_1\ c_2 \cdots c_k] \qquad k < n$$

$$Y \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n) \quad \text{as} \quad \underset{\sim}{\epsilon} \sim N(0, I_n\sigma^2).$$

Assume.

$|X^T X| \neq 0$

$$\underset{\sim}{\hat{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{Y} \sim N\left(\underset{\sim}{\beta}, (X^T X)^{-1}\sigma^2\right)$$

$$\hat{Y} = X\hat{\beta} = [X(X^T X)^{-1} X^T]\underset{\sim}{Y} = P_x \underset{\sim}{Y} \sim N(X\beta, \sigma^2 P_x)$$

$$\begin{cases} Z \sim N(\underset{\sim}{\mu}, \Sigma) \\ AZ \sim N(A\underset{\sim}{\mu}, A\Sigma A^T). \end{cases}$$

$$\underset{\sim}{e} = (Y - \hat{Y}) = (I_n - P_x)\underset{\sim}{Y} \sim N\left(0, \sigma^2(I_n - P_x)\right)$$

$$\begin{cases} P_x = X(X^T X)^{-1} X^T \\ P_x = P_x^T \\ P_x = P_x^2 \\ \text{rank}(P_x) = k+1 \end{cases}$$

$$\frac{\underset{\sim}{e}^T \underset{\sim}{e}}{\sigma^2} = \left(\frac{\underset{\sim}{e}}{\sigma}\right)^T \left(\frac{\underset{\sim}{e}}{\sigma}\right) = \left(\frac{\underset{\sim}{Y}}{\sigma}\right)^T (I_n - P_x)^T (I_n - P_x)\left(\frac{\underset{\sim}{Y}}{\sigma}\right).$$

$$= \left(\frac{\underset{\sim}{Y}}{\sigma}\right)^T (I_n - P_x)\left(\frac{\underset{\sim}{Y}}{\sigma}\right) \sim \chi^2_{\substack{df = n-k-1 \\ ncp = 0}}$$

②.

$$\underset{\sim}{e}^T \underset{\sim}{e} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \underline{\text{SS Error}} = \text{SS Residual}$$

$\text{SS Error} \to \text{SSE}$, $\quad \text{SS Residual} \to \text{SS Res}$

$\text{SSReg} \neq \text{SSRes}$
$\text{SSReg} = \text{SS Model}$, $\quad \text{SSRes} = \text{SSE}$
$\underline{\text{SS Model}} \neq \underline{\text{SSE}}$

$$\frac{\underset{\sim}{e}^T \underset{\sim}{e}}{n-k-1} = \text{MS Error} = \text{MS Residual}$$

$$E\left(\frac{\underset{\sim}{e}^T \underset{\sim}{e}}{n-k-1}\right) = \frac{(n-k-1)\sigma^2}{(n-k-1)} = \sigma^2$$

LS: So MS Error = MS Res is an unbiased estimator of σ^2.
$\hat{\beta}$ is an unbiased estimator of $\underset{\sim}{\beta}$.

MLE. $\underset{\sim}{\epsilon} \sim N(0, \sigma^2 I_n) \Rightarrow \underset{\sim}{y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n)$.

$$f(\underset{\sim}{y}) = \frac{e^{-\frac{1}{2}\left[(\underset{\sim}{y} - X\underset{\sim}{\beta})^T [\sigma^2 I_n]^{-1} (\underset{\sim}{y} - X\underset{\sim}{\beta})\right]}}{(\sqrt{2\pi})^n \sigma^n}$$

$$= \frac{e^{-\frac{1}{2}\frac{1}{\sigma^2}\left[\sum_{i=1}^{n} (y_i - x_i^T \underset{\sim}{\beta})^2\right]}}{(\sqrt{2\pi})^n \sigma^n}$$

$$\hat{\underset{\sim}{\beta}}_{MLE} = \hat{\beta}_{LS}$$

$$\hat{\sigma}^2_{MLE} = \frac{\underset{\sim}{e}^T \underset{\sim}{e}}{n} \neq \frac{\underset{\sim}{e}^T \underset{\sim}{e}}{n-k-1} = \hat{\sigma}^2_{LS}.$$

biased estimator.

ANOVA (Analysis of Variance)

H_0: $\beta_1 = \beta_2 = \cdots\cdots = \beta_k = 0$ (multiple hypothesis).

H_1: H_0 is not true. (at least one of $\beta_1, \beta_2, \ldots \beta_k$ are nonzero).

If H_0 is true $\Leftrightarrow$ $Y_i \overset{iid}{\sim} N(\beta_0, \sigma^2)$ $\Leftrightarrow$ Model building is not required.

$H_1 \Leftrightarrow$ model building is needed.

$$\text{SS Total} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \underset{\sim}{Y}^T \left(I_n - \tfrac{1}{n} 1 1^T\right) \underset{\sim}{Y}.$$

$$\text{SS Error} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \underset{\sim}{Y}^T (I_n - P_x) \underset{\sim}{Y}.$$

$$\text{SS Model} = \text{SS Regression} = \underset{\sim}{Y}^T \left(P_x - \tfrac{1}{n} 1 1^T\right) \underset{\sim}{Y}.$$

$$SST = \underset{\sim}{Y}\left(I_n - \tfrac{1}{n} 1 1^T\right)\underset{\sim}{Y} = \underset{\sim}{Y}\left\{[I_n - P_x] + \left[P_x - \tfrac{1}{n} 1 1^T\right]\right\}\underset{\sim}{Y}.$$

We can apply cochran's theorem.

$$SST = \text{SS Error} + \text{SS Model}.$$

$$\underbrace{\left(I_n - \tfrac{1}{n} 1 1^T\right)}_{(n-1)} = \underbrace{(I_n - P_x)}_{(n-k-1)} + \underbrace{\left(P_x - \tfrac{1}{n} 1 1^T\right)}_{k}.$$

square, symmetric, idempotent (HW).

$$\frac{SSE}{\sigma^2} = \frac{1}{\sigma^2} \underset{\sim}{Y}^T (I_n - P_x) \underset{\sim}{Y} \sim \chi^2_{n-k-1}$$

$$\frac{SSModel}{\sigma^2} = \frac{1}{\sigma^2} \underset{\sim}{Y}^T \left(P_x - \frac{1}{n} \underset{=}{1}\underset{=}{1}^T\right) \underset{\sim}{Y} \sim \chi^2_{k,\, ncp = \lambda} \qquad \text{since } \lambda > 0$$

~~SSModel~~

$$\frac{SS\ Total}{\sigma^2} = \frac{1}{\sigma^2} \underset{\sim}{Y}^T \left(I_n - \frac{1}{n} \underset{=}{1}\underset{=}{1}^T\right) \underset{\sim}{Y} \sim \chi^2_{n-1,\, ncp = \lambda}$$

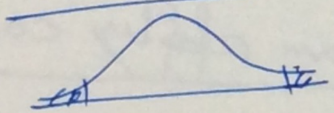
To show λ is a function of $(\beta_1, \beta_2, \cdots, \beta_k)$ only. and free from β_0.

$\underset{\sim}{x} \sim N(\mu, I)$ $\qquad E(x^T x) = n + \mu^T \mu$

$\Rightarrow x_i \overset{iid}{\sim} N(\mu, 1)$ $\qquad \Rightarrow E\left(\sum_{i=1}^{n} x_i^2\right) = n + n\mu^2$

$P(\bar{x} \neq 0) = 1$ $\qquad \Rightarrow E\left(\frac{\sum x_i^2}{n}\right) = 1 + \mu^2$

$\sqrt{n}\ \bar{x} \sim N(0,1)$ under H_0.



$E(ax + by)$

$= a E(x) + b E(y)$

$n\ \bar{x}^2 \sim \chi^2_{1}$ under H_0

$\frac{\chi^2_{1,\, ncp\, \mu^2}}{1 + \mu^2}$ under H_1.

(5)

$$\underset{\sim}{Y} \sim N\left(X\underset{\sim}{\beta}, \sigma^2 I_n\right) \Rightarrow \frac{\underset{\sim}{Y}}{\sigma} \sim N\left(\frac{X\underset{\sim}{\beta}}{\sigma}, I_n\right)$$

$$\underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \hdashline \underset{\sim}{\beta}_R \end{pmatrix} \qquad \underset{n \times (k+1)}{X} = \left(\underset{\sim}{1} \,\vdots\, X_R\right)$$

$$NCP\left(\frac{SSModel}{\sigma^2}\right) = \left(\frac{X\underset{\sim}{\beta}}{\sigma}\right)^T\left(P_x - \frac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T\right)\left(\frac{X\underset{\sim}{\beta}}{\sigma}\right)$$

$$= \frac{1}{\sigma^2}\underset{\sim}{\beta}^T\left[X^T\left(P_x - \frac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T\right)X\right]\underset{\sim}{\beta}.$$

$$= \frac{1}{\sigma^2}\underset{\sim}{\beta}^T\left[X^TX - \frac{1}{n}X^T\underset{\sim}{1}\underset{\sim}{1}^TX\right]\underset{\sim}{\beta}.$$

$$= \frac{1}{\sigma^2}\underset{\sim}{\beta}^T\left[\begin{pmatrix} n & \underline{1}^TX_R \\ X_R^T\underline{1} & X_R^TX_R \end{pmatrix} - \frac{1}{n}\begin{pmatrix} \underline{1}^T \\ X_R^T \end{pmatrix}\underline{1}\underline{1}^T\left(\underline{1}\ X_R\right)\right]\underset{\sim}{\beta}.$$

$$= \frac{1}{\sigma^2}\underset{\sim}{\beta}^T\left[\begin{pmatrix} n & \underline{1}^TX_R \\ X_R^T\underline{1} & X_R^TX_R \end{pmatrix} - \frac{1}{n}\begin{pmatrix} n \\ X_R^T\underline{1} \end{pmatrix}\left(n\ \ \underline{1}^TX_R\right)\right]\underset{\sim}{\beta}.$$

$$= \frac{1}{\sigma^2}\underset{\sim}{\beta}^T\left[\begin{pmatrix} n & \underline{1}^TX_R \\ X_R^T\underline{1} & X_R^TX_R \end{pmatrix} - \begin{pmatrix} n & \underline{1}^TX_R \\ X_R^T\underline{1} & \frac{1}{n}X_R^T\underline{1}\underline{1}^TX_R \end{pmatrix}\right]\underset{\sim}{\beta}.$$

$$= \frac{1}{\sigma^2}\begin{pmatrix} \beta_0 \\ \hdashline \beta_R \end{pmatrix}^T\begin{bmatrix} 0 & \underset{\sim}{0}^T \\ \underset{\sim}{0} & X_R^TX_R - \frac{1}{n}X_R^T\underset{\sim}{1}\underset{\sim}{1}^TX_R \end{bmatrix}\begin{pmatrix} \beta_0 \\ \hdashline \beta_R \end{pmatrix}$$

$$= \frac{1}{\sigma^2}\underset{\sim}{\beta}_R^T\left(X_R^TX_R - \frac{1}{n}X_R^T\underset{\sim}{1}\underset{\sim}{1}^TX_R\right)\underset{\sim}{\beta}_R$$

Note.

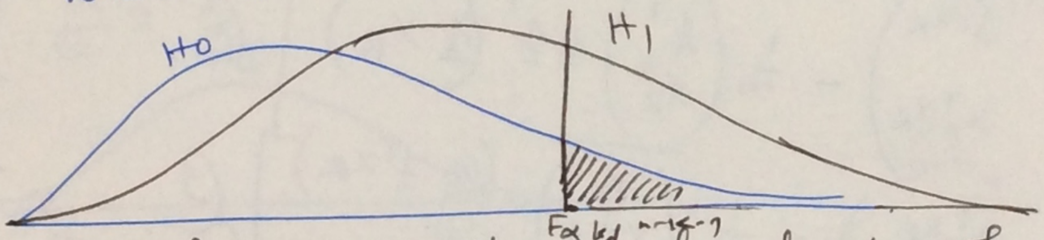
→ No contribution of β_0

→ it is a quadratic form of β_R only

when $\beta_R = \underline{0}$ then
NCP $\left(\frac{SS\,Model}{\sigma^2}\right) = 0$ ~~then~~

$$F = \frac{\left(\frac{SS\,Model}{\sigma^2}\right) / k}{\left(\frac{SS\,Error}{\sigma^2}\right) / (n-k-1)} \sim F_{k, n-k-1, \lambda}$$

$\lambda = 0$ when H_0 is true.



We reject H_0 in favour of H_1 at level α if

$$F(\text{observed}) > F_{\alpha, k, n-k-1}$$

HW 1. $\left[(X_R^T X_R) - \frac{1}{n} X_R^T \underline{1}\,\underline{1}^T X_R\right]$ can be expressed as ~~centered~~ $X_c^T X_c$ when X_c represents the centered regressors. (appendix Montgomery).

HW 2. Consider simple linear regression and the line is passing through origin. Perform ANOVA for the model.

Ex. 1 $\hat{\beta} \sim N\left(\underset{\sim}{\beta},\ (X^TX)^{-1}\sigma^2\right)$

H_0: $\beta_j = b_j$ vs $\beta_j \neq b_j$

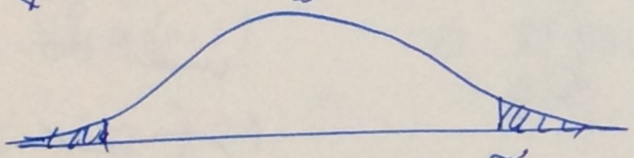
j is a fixed number between $0, 1, 2, \cdots k$.

$\underset{\sim}{l}^T \underset{\sim}{\hat{\beta}} = \hat{\beta}_j \sim N\left(\beta_j,\ c_{jj}\sigma^2\right)$

when $\begin{array}{c} 0\ 1\ 2 \cdots \quad j \quad \cdots k \\ l = (0 \cdots 0\ 1\ 0 \cdots 0)^T \end{array}$

$$T = \frac{\hat{\beta}_j - b_j}{\sqrt{\hat{\sigma}^2\, c_{jj}}} \overset{\text{under } H_0}{\sim} t_{n-k-1,\ ncp=0}$$

$C = (X^TX)^{-1}$ $(0, 1, 2, \cdots k)$



$-\tau_{\alpha/2}$ $\tau_{\alpha/2,\ n-k-1}$

$$\hat{\sigma}^2 = \frac{SS\ Error}{n-k-1}$$

We reject H_0 in favour of H_1 when $|T(obs)| > \tau_{\alpha/2,\ n-k-1}$ at level α.

Ex2.

$H_0: \beta_i - 2\beta_j = b_0$ vs $H_1 \; \beta_i - 2\beta_j \neq b_0$

$$\underset{\sim}{l} = (\overset{0}{0} \cdots\cdots \overset{i}{1}0 \cdots\cdots \overset{j}{-2} \cdots\cdots \overset{k}{0})^T \qquad \hat{\beta} \sim N(\beta, \sigma^2 C)$$

$$\hat{\beta}_i - 2\hat{\beta}_j = \underset{\sim}{l}^T \underset{\sim}{\hat{\beta}} \sim N\left(\beta_i - 2\beta_j, \sigma^2(c_{ii} + 4c_{jj} - 4c_{ij})\right)$$

$$V(ax+by) = a^2 V(x) + b^2 V(y) + 2\,\mathrm{cov}(x,y) \quad \Bigg| \quad \hat{\sigma}^2 = \frac{SSError}{n-k-1}$$

$$T = \frac{(\hat{\beta}_i - 2\hat{\beta}_j) - b_0}{\sqrt{\hat{\sigma}^2 (c_{ii} + 4c_{jj} - 4c_{ij})}} \overset{\text{under } H_0}{\sim} \underline{t_{n-k-1,\, ncp=0}}$$

We reject H_0 in favour of H_1 at level α if.

$$\underline{|T(obs)| > t_{\alpha/2,\, n-k-1}}.$$

Ex3. $H_0: \beta_i = \beta_j = 0$ VS H_1: H_0 is not true.

$$H_0 \iff \begin{cases} \beta_i = 0 \\ \beta_i - \beta_j = 0 \end{cases} \iff \underbrace{\begin{array}{c} \begin{array}{cccccccc} & & & \beta_i & & & & \end{array} \\ \begin{pmatrix} 0 & 0 & \cdots & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & 1 & 0 & \cdots & -1 & \cdots 0 \end{pmatrix} \\ \begin{array}{cccccccc} & & & & & & \beta_j & \end{array} \end{array}}_{A} \underset{\sim}{\beta} = \underbrace{\underset{\sim}{0}}_{\underset{\sim}{b}} \iff A\underset{\sim}{\beta} = \underset{\sim}{b}$$

H_1: $A\underset{\sim}{\beta} \neq \underset{\sim}{b}$.

$$Z \sim N(\mu, \Sigma)$$
$$(Z-\mu) \sim N(0, \Sigma)$$
$$W = \Sigma^{-1/2}(Z-\mu) \sim N(0, I)$$
$$\underset{\sim}{W}^T \underset{\sim}{W} = (Z-\mu)^T \Sigma^{-1} (Z-\mu)$$

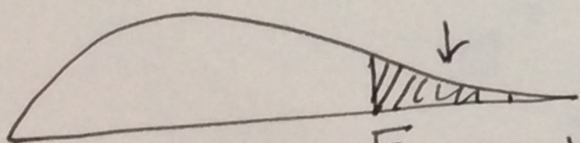
$$\hat{\beta} \sim N(\underset{\sim}{\beta}, \sigma^2 C)$$
$$A\hat{\underset{\sim}{\beta}} \sim N(A\underset{\sim}{\beta}, \sigma^2 ACA^T)$$
$$\sim N(0, \sigma^2 ACA^T) \text{ under } H_0.$$
$$\Rightarrow (A\hat{\underset{\sim}{\beta}} - \underset{\sim}{b}) \sim N(0, \sigma^2 ACA^T)$$
$$\Rightarrow (A\hat{\underset{\sim}{\beta}} - \underset{\sim}{b})^T \frac{(ACA^T)^{-1}}{\sigma^2} (A\underset{\sim}{\beta} - \underset{\sim}{b}) \sim \chi^2_{\text{rank}(A),\, ncp=0} \quad \underline{\text{under } H_0}.$$

$$I = \Sigma^{-1/2} \Sigma \Sigma^{-1/2}$$

let $\text{rank}(A) = r$ under H_0

$$F = \frac{(A\hat{\beta} - \underset{\sim}{b})^T (ACA^T)^{-1} (A\underset{\sim}{\beta} - b)/r}{(\underset{\sim}{Y}^T (I_n - P_x) \underset{\sim}{Y})/(n-k-1)} \sim F_{r,\, n-k-1}$$

upper tail test.



$F_{\alpha,\, r,\, n-k-1}$

$$Z \sim N(\mu, \Sigma)$$
$$AZ \sim N(A\mu, \underline{A\Sigma A^T}).$$

$$\Sigma = PDP^T \quad \Sigma^{1/2} = P D^{1/2} P^T$$

Result 1. The iff condition of a LPF $\underset{\sim}{p}^T \underset{\sim}{\beta}$ to be estimable is $\underset{\sim}{p} \in \mathcal{C}(X^T)$.

$$E(\underset{\sim}{l}^T \underset{\sim}{y}) = \underset{\sim}{p}^T \underset{\sim}{\beta} \quad \forall \underset{\sim}{\beta} \in \mathbb{R}^{k+1}$$

$$\Leftrightarrow \quad \underset{\sim}{l}^T X \underset{\sim}{\beta} = \underset{\sim}{p}^T \underset{\sim}{\beta} \quad \forall \underset{\sim}{\beta} \in \mathbb{R}^{k+1}$$

$$\Leftrightarrow \quad \underset{\sim}{l}^T \underset{\sim}{X} = \underset{\sim}{p}^T$$

$$\Leftrightarrow \quad X^T \underset{\sim}{l} = \underset{\sim}{p}$$

$$\Leftrightarrow \quad p \in \mathcal{C}(X^T).$$

Result 2. Show that $\underset{\sim}{l}^T \underset{\sim}{y}$ is a LUE of $\underset{\sim}{p}^T \underset{\sim}{\beta}$ iff $X^T l = p$.

Result 3 Show that $\underset{\sim}{l}^T \underset{\sim}{y}$ is a LZF iff $\underset{\sim}{X}^T \underset{\sim}{l} = \underset{\sim}{0}$.

$$E(\ell^T Y) = 0 \quad \forall \underset{\sim}{\beta} \in \mathbb{R}^{k+1}.$$

$$\Leftrightarrow \quad \ell^T X \underset{\sim}{\beta} = 0 \quad \forall \underset{\sim}{\beta} \in \mathbb{R}^{k+1}.$$

LZF

$$\Leftrightarrow \quad \underset{\sim}{\ell}^T \underset{\sim}{X} = \underset{\sim}{0}^T$$

$$\Leftrightarrow \quad X^T \underset{\sim}{\ell} = \underset{\sim}{0} \quad \Leftrightarrow \quad \underset{\sim}{\ell} \in \mathcal{C}(X)^{\perp}$$

$$\Leftrightarrow \quad \ell \in \mathcal{C}(I - P_x).$$

$\Rightarrow$ There exists a vector $\underset{\sim}{m}$ such that. $\underset{\sim}{\ell} = (I_n - P_x)\underset{\sim}{m}$.

$$\Rightarrow \quad \ell^T X = \underset{\sim}{m}^T (I_n - P_x) X = m^T (X - X) = \underset{\sim}{0}^T.$$

Every LZF has the representation. $\underline{\underset{\sim}{m}^T (I_n - P_x) \underset{\sim}{Y}}$.

$$E\left(\underset{\sim}{m}^T (I_n - P_x) \underset{\sim}{Y}\right)$$

$$= \underset{\sim}{m}^T \underline{(I_n - P_x) X} \underset{\sim}{\beta}$$

$$= \underset{\sim}{m}^T \underbrace{\underset{\sim}{0}}_{\text{matrix}} \underset{\sim}{\beta} = \underbrace{0}_{\text{scalar}}.$$

Th. $\underset{\sim}{l}^T \underset{\sim}{Y}$ is said to be the BLUE of $E(\underset{\sim}{l}^T Y)$ iff it is uncorrelated to all linear zero functions.

proof. D. Sengupta & Jammalamadaka.

Application. If we know that $\underset{\sim}{l}^T \underset{\sim}{Y}$ is a LUE of $\underset{\sim}{p}^T \underset{\sim}{\beta} = E(\underset{\sim}{l}^T Y)$ then immidiately we can have the BLUE of $\underset{\sim}{p}^T \beta$ as $\underset{\sim}{l}^T P_x \underset{\sim}{Y}$.

$$\underset{\sim}{l}^T \underset{\sim}{Y} = \underset{\sim}{l}^T I_n \underset{\sim}{Y} = \underset{\sim}{l}^T (P_x + (I_n - P_x)) \underset{\sim}{Y} = \underset{\sim}{l}^T P_x \underset{\sim}{Y} + \underset{\sim}{l}^T (I_n - P_x) \underset{\sim}{Y}$$

① $E(l^T P_x y) = E(l^T y) = \underset{\sim}{p}^T \underset{\sim}{\beta}$. (unbiased)

② $E(l^T (I_n - P_x) y) = 0$ hence (LZF).

③ $\operatorname{cov}(\underset{\sim}{l}^T P_x \underset{\sim}{Y},\ l^T (I_n - P_x) \underset{\sim}{Y}) = 0 \Rightarrow \underset{\sim}{l}^T P_x \underset{\sim}{Y}$ is the BLUE of $\underset{\sim}{p}^T \underset{\sim}{\beta}$.

HW Show that BLUE is unique.

BLUE of a linear parametric function is unique.

Pf. Let if possible there exists two BLUEs $\underset{\sim}{l}_1^T \underset{\sim}{Y}$ and $\underset{\sim}{l}_2^T \underset{\sim}{Y}$ for $\underset{\sim}{p}^T \underset{\sim}{\beta}$

$$\underset{\sim}{l}_1^T \underset{\sim}{Y} = \underset{\sim}{l}_2^T \underset{\sim}{Y} + \underset{\sim}{l}_1^T \underset{\sim}{Y} - \underset{\sim}{l}_2^T \underset{\sim}{Y}$$

$$= \underset{\sim}{l}_2^T \underset{\sim}{Y} + (\underset{\sim}{l}_1^T - \underset{\sim}{l}_2^T) \underset{\sim}{Y}$$

Note $(\underset{\sim}{l}_1^T - \underset{\sim}{l}_2^T)\underset{\sim}{Y}$ is LZF.

$\updownarrow$

as $\underset{\sim}{l}_2^T \underset{\sim}{Y}$ is blue.

$$Var(\underset{\sim}{l}_1^T \underset{\sim}{Y}) = Var(\underset{\sim}{l}_2^T \underset{\sim}{Y}) + Var((\underset{\sim}{l}_1^T - \underset{\sim}{l}_2^T)\underset{\sim}{Y}) + 0$$

Case 1 $\quad Var(\underset{\sim}{l}_1^T \underset{\sim}{Y}) > Var(\underset{\sim}{l}_2^T \underset{\sim}{Y}) \Rightarrow$ contradiction.

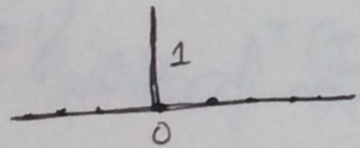
Case 2 $\quad V(\underset{\sim}{l}_1^T \underset{\sim}{Y} - \underset{\sim}{l}_2^T \underset{\sim}{Y}) = 0.$

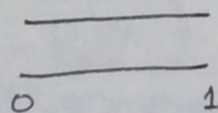
also we know $E(\underset{\sim}{l}_1^T \underset{\sim}{Y} - \underset{\sim}{l}_2^T \underset{\sim}{Y}) = 0.$

$$\Rightarrow P((\underset{\sim}{l}_1^T \underset{\sim}{Y} = \underset{\sim}{l}_2^T \underset{\sim}{Y})) = 1.$$

unique mean

$E(X) = 0$
$V(X) = 0$





$P(\text{rational number}) = 0$

H, palm length, feet length.

Polynomial Regression

$(Y, x) \longrightarrow Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \underline{\underline{\epsilon}}$

curve

or $(Y, x_1, x_2 \cdots) \longrightarrow Y = \beta_0 + (\beta_1 x_1 + \beta_2 x_2) + (\beta_3 x_1^2 + \beta_4 x_1 x_2 + \beta_5 x_2^2) + \epsilon$

surface.

In general we can write.

$$\underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{\epsilon}.$$

Case 1.

$$X\beta = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ & & & \vdots & \\ 1 & x_n & x_n^2 & \cdots & x_n^k \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

Case 2

$$X\underset{\sim}{\beta} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & (x_{11}x_{12}) & x_{12}^2 \\ 1 & x_{21} & x_{22} & x_{21}^2 & (x_{21}x_{22}) & x_{22}^2 \\ \vdots & & & & & \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & (x_{n1}x_{n2}) & x_{n2}^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{Y}$$

$$\hat{Y} = X \hat{\beta} = X (X^T X)^{-1} X^T \underline{Y}$$

Linear regression problem

$\mathbb{D} = \{ (y_i, x_i) \quad i = 1, 2, \cdots n. \}$

Q1 How may degrees can we use in polynomial regression.?

let $k = n-1$

Interpolation

$$\underset{\sim}{Y} = \underset{n\times n}{X}\, \underset{\sim}{\beta} \Rightarrow \hat{\underset{\sim}{\beta}} = X^{-1}\underset{\sim}{Y}. \quad \text{if } x_i \neq x_j \ \forall\, i \neq j$$

$$|X| \neq 0$$

$$\Rightarrow \hat{\underset{\sim}{Y}} = X\hat{\beta} = X X^{-1} \underset{\sim}{Y} = \underset{\sim}{Y}$$

$$\Rightarrow \underset{\sim}{e} = \underset{\sim}{Y} - \hat{\underset{\sim}{Y}} = \underset{\sim}{0}$$

we cannot estimate σ^2

* curve must pass through all the data points.

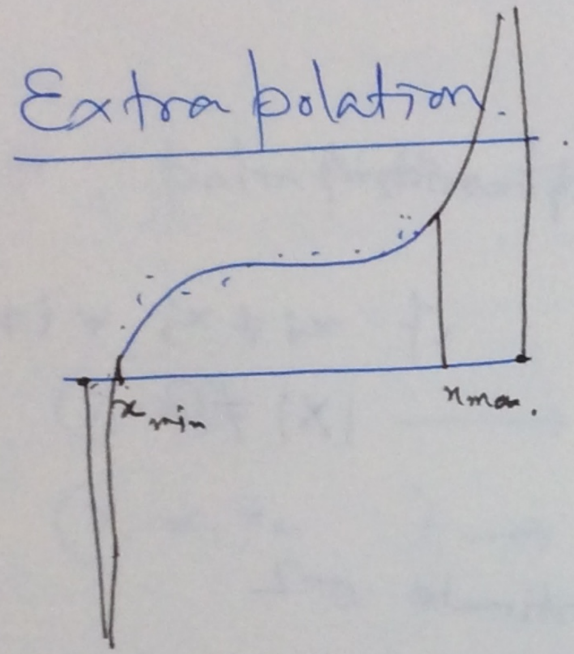
we need to take $k \leq n-2$

* Regression problem does not need that the fitted curve must pass through all the data points.

Issues with polynomial regression.

1. order of the model: ① Find the max degree. ② Having good predictibility.
2. Model building process. (i) Forward selection. (ii) Backward selection.

(III) Extrapolation.



$x_{new} < x_{imin}$

$x_{new} > x_{max}$.

then the prediction error mabe large with ~~large~~ increasing degree.

(IV) Ill-conditioning: $\left|(X^T X)\right| \approx 0$ for exampe $0 < x_i < 1$

unstable $\hat{\underset{\sim}{\beta}}$.

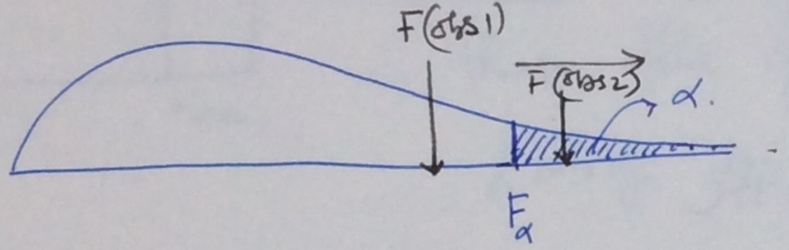
(V) Hieraarchy: $Y = \beta_0 + \underline{\beta_1} x + \underline{\beta_2} x^2 + \underline{\beta_3} x^3 + \beta_4 \cdot x^4 + \epsilon$.

when

There might be case some of β_i whre $i <$ max degree are significatly close to zero.

P value.

Assume it is an upper tail test F, or χ^2.



Test conditon.

$$F(obs) > F_\alpha$$

we reject Ho in favour of H_1

$$P\left(F(obs\ 1) < F\right) > \alpha \Leftrightarrow \text{we donot reject Ho.}$$

$$P\left(F(obs\ 2) < F\right) < \alpha \Leftrightarrow \text{we reject Ho.}$$

Orthogonal polynomial.

$\sum_{i=1}^{n} u_i v_i = \underset{\sim}{u}^T \underset{\sim}{v} = 0$ then we say $\underset{\sim}{u} \perp \underset{\sim}{v}$. orthogonal vectors.

$\langle u, v \rangle$ inner product.

$\theta \in [0,1]$ $u, v, : [0,1] \to \mathbb{R}$

$\langle u, v \rangle = \int_0^1 u(\theta)\, v(\theta)\, d\theta$ ~~$\neq 1$~~

$\langle u, v \rangle = 0 \iff u \perp v$ orthogonal functions.

$u(\theta) = \sin(2\pi\theta)$
$v(\theta) = \cos(2\pi\theta)$

Orthogonal Polynomials.

$(x_i, y_i,)$ $i = 1, 2, \ldots n$. Data given.

$\boxed{P_j(x_i) = \sum_{m=0}^{j} \theta_m (x_i)^m \quad j\text{th degree orthogonal polynomial.}}$

$$\left\{ \begin{array}{l} P_0(x_i) = 1. \quad \forall x_i \\ \sum_{i=1}^{n} P_j(x_i)\, P_k(x_i) = \underset{\sim}{P_j}^T \underset{\sim}{P_k} = 0 \quad \forall j \neq k. \text{ for given data } (x_i, y_i,) \end{array} \right.$$

To fit k-degree ~~poly~~ orthogonal polynomial to the given data set. (12)

$$y_i = \sum_{j=0}^{k} \alpha_j P_j(x_i) + \epsilon_i \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

Model parameters are. $(\alpha_0\, \alpha_1 \cdots \alpha_k)$ and σ^2. are unknown.

$$X_0 \alpha = \begin{pmatrix} P_0(x_1) & P_1(x_1) & P_2(x_1) & \cdots & P_k(x_1) \\ P_0(x_2) & P_1(x_2) & P_2(x_2) & & P_k(x_2) \\ \vdots & & & & \\ P_0(x_n) & P_1(x_n) & P_2(x_n) & & P_k(x_n) \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}$$

$$\underset{\sim}{Y} = X_0 \underset{\sim}{\alpha} + \underset{\sim}{\epsilon}.$$

$k < n-1$

X_0 is known as data $(\underset{\sim}{x})$ is known and all $P_j(\cdot)$ are known.

S estimate.

$$\hat{\underset{\sim}{\alpha}} = (X_0^T X_0)^{-1} X_0^T \underset{\sim}{Y} = \begin{pmatrix} \sum_{i=1}^{n} P_0^2(x_i) & & & 0 \\ & \sum_{i=1}^{n} P_1^2(x_i) & & \\ & & \ddots & \\ 0 & & & \sum_{i=1}^{n} P_k^2(x_i) \end{pmatrix}^{-1} X_0^T \underset{\sim}{Y}.$$

$$\hat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(x_i)\, y_i}{\sum P_j^2(x_i)} = \frac{\underset{\sim}{P_j}^T \underset{\sim}{Y}}{\underset{\sim}{P_j}^T \underset{\sim}{P_j}}$$

$$\hat{\alpha}_j = \frac{\underset{\sim}{P}_j^T \underset{\sim}{Y}}{\underset{\sim}{P}_j^T \underset{\sim}{P}_j} = \frac{\sum_{i=1}^{n} P_j(x_i)\, y_i}{\sum_{i=1}^{n} P_j^2(x_i)} \Rightarrow \hat{\alpha}_0 = \bar{y}$$

$$\hat{\alpha}_j \sim N\left(\frac{\underline{P}_j^T X_0 \underset{\sim}{\alpha}}{\underline{P}_j^T \underline{P}_j},\ \sigma^2 \frac{(\underline{P}_j^T I_n \underline{P}_j)}{(\underline{P}_j^T \underline{P}_j)^2} \right) \qquad \underset{\sim}{Y} \sim N\left(X_0 \underset{\sim}{\alpha},\ I_n \sigma^2\right)$$

$$\hat{\alpha}_j \sim N\left(\alpha_j,\ \frac{\sigma^2}{\sum_{i=1}^{n} P_j^2(x_i)} \right) \qquad \hat{\alpha}_j = \underset{\sim}{l}_j^T \underset{\sim}{Y}$$

$$l_j^T = \frac{\underline{P}_j^T}{\underline{P}_j^T \underline{P}_j}$$

Now we can perform test $H_0: \alpha_j = a_j$ vs $H_1: \alpha_j \neq a_j$ $\quad \underline{\forall\, j}$

$$\hat{\sigma}^2 = \frac{SS\ Error}{n-k-1}$$

SS Error.

$$SSE = \underset{\sim}{y}^T (I_n - P_{X_0}) \underset{\sim}{y} \qquad P_{X_0} y = X_0 (X_0^T X_0)^{-1} X_0^T . \underset{\sim}{y}$$

$$= X_0 \hat{\underset{\sim}{\alpha}}$$

$$= \underset{\sim}{y}^T \underset{\sim}{y} - \underset{\sim}{y}^T P_{X_0} \underset{\sim}{y}.$$

$$\sum_{i=1}^{n} P_0(x_i) y_j = n \bar{y}$$

$$= \sum_{i=1}^{n} y_i^2 - \sum_{j=0}^{K} \hat{\alpha}_j \left(\sum_{i=1}^{n} P_j(x_i) y_i \right)$$

$$= \sum_{i=1}^{n} y_i^2 - \hat{\alpha}_0 \sum_{i=1}^{n} P_0(x_i) y_i - \sum_{j=1}^{K} \hat{\alpha}_j \left(\sum_{i=1}^{n} P_j(x_i) y_i \right)$$

$$= \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 - \sum_{j=1}^{K} \hat{\alpha}_j \left(\sum_{i=1}^{n} P_j(x_i) y_i \right)$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{j=1}^{K} \hat{\alpha}_j \left(\sum_{i=1}^{n} P_j(x_i) y_i \right)$$

$$= SS\,Total. - \underline{SS\,Model}.$$

$$SST = SS\,Model + SSE.$$

HW.

$H_0: \alpha_j = 0$.

$H_1: \alpha_j \neq 0$.

$$F = \frac{\left(\hat{\alpha}_j \sum_{i=1}^{n} P_j(x_i) y_i\right)/1}{\hat{\sigma}^2} \sim F_{1, n-k-1}$$

under Ho.

perform the test.

do the proof.