Error Analysis.

→ After filling the model if we want to cross check that the model fitting is justifiable or not. then error analysis will helps us.

→ Out lier analysis.

Model

$$Y \sim N(X\beta, \sigma^2 I_n)$$

Estimated error.

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

$$\hat{\sigma}^2 = \frac{\underset{\sim}{e}^T \underset{\sim}{e}}{n-k-1} \text{ a.s. } \frac{\underset{\sim}{e}^T \underset{\sim}{e}}{\sigma^2} \sim \chi^2_{n-k-1}. \qquad P_x = H = \underset{(k+1)\times(k+1)}{((h_{ij}))}$$

$$\text{cov}(e_i, e_j) = \begin{cases} \sigma^2(1 - h_{ii}) & \text{if } i = j \\ \sigma^2(-h_{ij}) & \text{if } i \neq j \end{cases}$$

$$\frac{e_i}{\sqrt{\sigma^2(1-h_{ii})}} \sim N(0,1) \quad \text{Marginally.}$$

$$\Rightarrow r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1-h_{ij})}} \sim t_{n-k-1}.$$

~~If n is large.~~

$$\frac{e_i}{\sqrt{\hat{\sigma}^2 (1-h_{ii})}} \overset{a}{\sim} N(0,1) \quad \frac{n \gg k}{\text{because.}}$$

$\xrightarrow{d}$ converges in distribut.

$$T = \frac{Z}{\sqrt{Y/n}} \sim t_n \quad \Big| \quad \begin{matrix} Z \sim N(0,1) \\ Y \sim \chi^2_n \end{matrix} > \text{ind.}$$

$$t_1 \equiv C(0,1)$$

$$Y/n \xrightarrow{P} 1 \text{ WLLN as } n \to \infty.$$

w.p 1.

$$Y \sim \chi^2_n.$$

$$Y = \sum_{i=1}^{n} z_i^2 \qquad z_i \overset{iid}{\sim} N(0,1)$$

$$E(z_i) = 1.$$

$$\frac{Y}{n} = \frac{1}{n} \sum_{i=1}^{n} z_i^2$$

$$\begin{cases} X_n \xrightarrow{d} X \\ Y_n \xrightarrow{P} c \\ Y_n X_n \xrightarrow{d} cX. \end{cases}$$

- Slutsky's theorem.

??

Standardized residual.

$$d_i = \frac{e_i}{\sqrt{MSError}}$$

$$MSError = \frac{SSError}{n-k-1}$$

$$e_i = y_i - \hat{y}_i$$

Studentized residual.

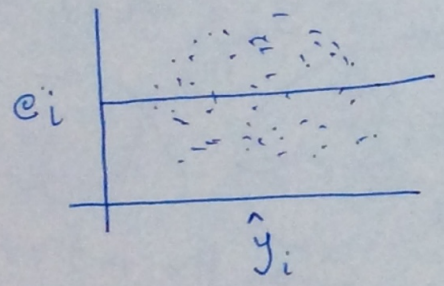
$$r_i = \frac{e_i}{\sqrt{MSError\,(1-h_{ii})}}$$

For simple linear regression.

$$r_i = \frac{e_i}{\sqrt{MSError\left[1-\left(\frac{1}{n}+\frac{(x_i-\bar{x})^2}{S_{xx}}\right)\right]}}$$

① residuals are marginally normally distributed $\left(k \ll n.\right)$

② residuals are uncorrelated to the predicted values.



test for correlation

{ Correlation zero should be reflected.

$r_i \overset{a}{\sim} N(0,1)$

→ diagram based. ✓ (q-q-plot.)

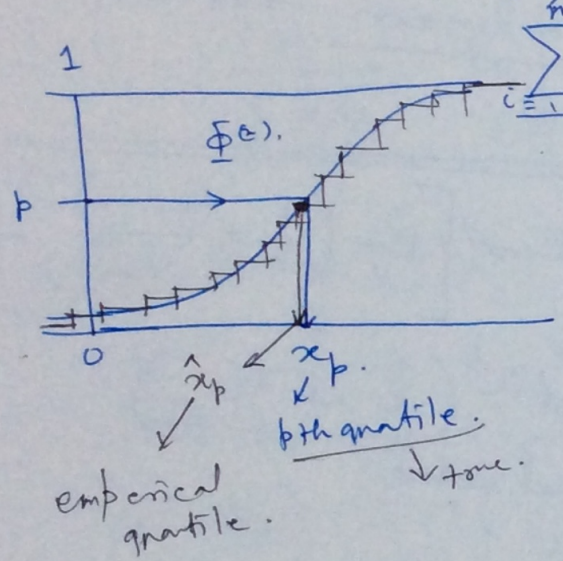
→ pdf based. ✓

→ cdf based. ✓ } testing.

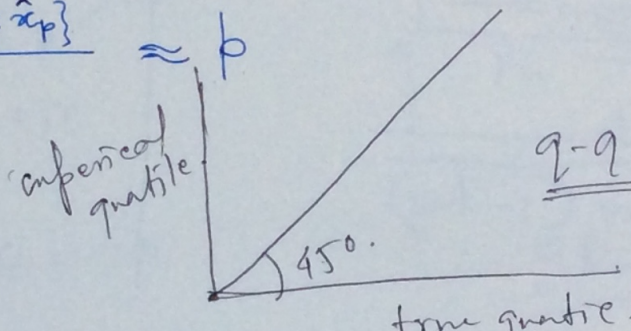
① p-th quantile x_p if $p = \Phi(x_p) = P(X \le x_p)$ $\quad X \sim N(0,1)$.

Let r_i's be sorted in increasing order.

$\min\{r_i\} = r_{(1)} < r_{(2)} < r_{(3)} \cdots \cdots r_{(n)} = \max\{r_i\}$



$$\sum_{i=1}^{n} \frac{I_{\{r_i \le \hat{x}_p\}}}{n} \approx p$$

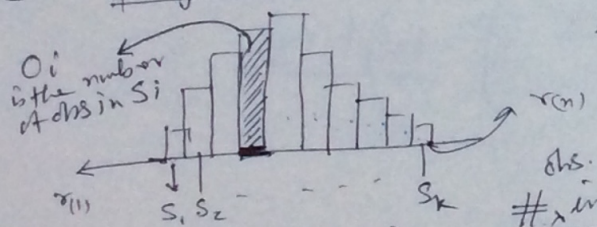


$$I_A = \begin{cases} 1 & \text{if } A \text{ is satisfied} \\ 0 & \text{o/w.} \end{cases}$$

q-q plot.

diagrammatic approach to check whether errors are ~~for~~ following normal distribution or not. (qqplot).

② bdf based.



construct the histogram.

$\rightarrow P(\text{data} \in S_i) = \frac{\#_s \text{ obs. in } S_i}{n} = \frac{O_i}{n}$

$\rightarrow P(x \in S_i) =$ calculate from true cdf $= p_i$ say.

expected frequency is $S_i = n \cdot p_i = E_i$

$$\sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{K-1}$$

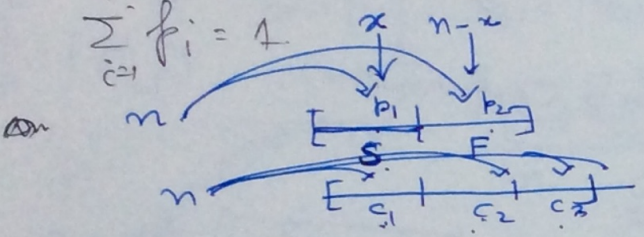
When H_0 is true.

$(S_1, S_2, \ldots, S_k)$

n obs.

multinomial $(n, p_1, p_2, \ldots, p_k)$

$\sum_{i=1}^{n} p_i = 1$



and $n \uparrow \infty$.

Multinomial ⟶ ~~X~~ Multivariate normal with k conformation. but dim- will be $(k-1)$ hence.

χ^2 - statstic will follow χ^2_{k-1} dof.

When expected frequencies are completely known as the true cdf is completely specified

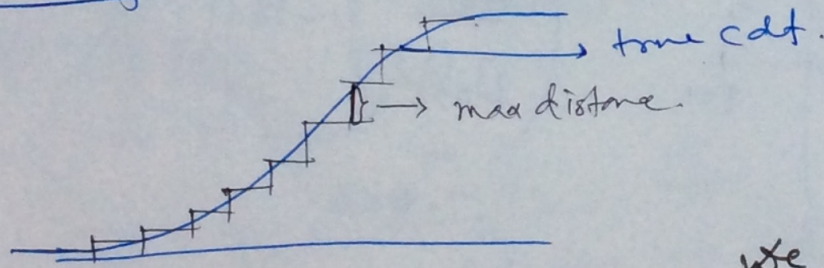
when ~~the~~ r parameter of the null distribution are estimated from data. then χ^2 statistic follows.

$$\sum_{i=1}^{k} \frac{(O_i - \hat{E}_i)^2}{\hat{E}_i} = \chi^2_{k-1-r}$$

$\hat{E}_i = n\hat{p}_i$

$= n\hat{p}_i(\hat{\mu}, \hat{\sigma}^2)$

③ Cdf based test. [only for continuous r.v.s]



$F_n(x) = \frac{\sum_{i=1}^{n} I_{\{x_i \le x\}}}{n}$ empirical cdf.

$F(x)$ is true cdf.

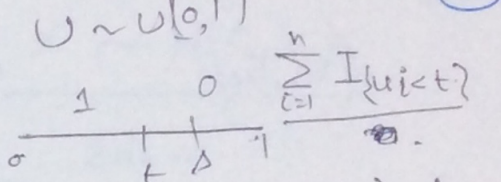
use ~~Kolmogorov~~ Kolmogorov

We study the distribution of

$$\max_x \sqrt{n}\left|F_n(x) - F(x)\right|$$

↙ Supremum.

$U \sim U[0,1]$



$$N(0, t(1-t)) \xrightarrow{d} \sqrt{n}\left(\frac{\sum I_{\{u_i \le t\}}}{n} - t\right)$$

$$G_n(y) = \frac{\sum_{i=1}^{n} I_{\{y_i \le y\}}}{n}.$$

If we know $X \sim F_\cdot(\cdot)$ which is continuous.

$F(X) \sim U[0,1]$?? How

transfer the data. $F(x_i) = y_i$

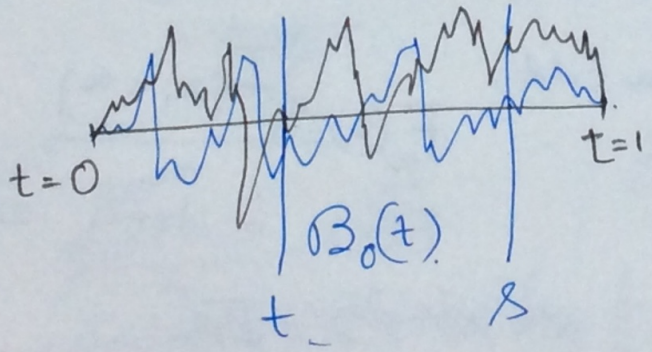
$$Y = F(X)$$

Kolmogorov-Smirnov test
KS-test.

$$\sup_{0 \le y \le 1} \sqrt{n}\left|G_n(y_i) - y\right| \xrightarrow{d} \sup_{0 \le t \le} \left|B_0(t)\right|$$

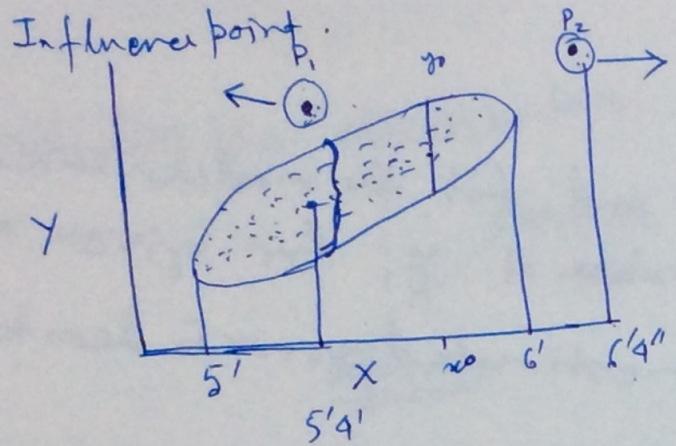
Kolmogorov distribution.

$B_0(t)$ is known as the standard Brownian bridge on $[0,1]$.



$$\begin{pmatrix} B_0(t) \\ B_0(s) \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} t(1-t) & t(1-s) \\ t(1-s) & s(1-s) \end{pmatrix}\right]$$

$t < s$



$X = h$

$Y = W$

all data.

✓ $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_0$ ⟶

✓ $\hat{y} = \underset{(-1)}{\hat{\beta}_0} + \underset{(-1)}{\hat{\beta}_1} x_0$ ⟶

P_1 removed.

① If the number of data point is less then we may not discard the outlier.

② If the data is logically incorrect then we discard it.

$P_1 = (y_1, x_1)$.

case 1 $|x_0 - x_1| < \delta$ ⟶ prefer to include P_1

case 2 $|x_0 - x_1| >> \delta$ ⟶ prefer not to include P_1.

③

Validation.	Model fitting.
Test.	train.
k	n − k

$\binom{n}{k}$ $\binom{100}{10}$ lags.

Leave-one-out / Jack-knife.

x_1 x_2 ... x_n.

x_1 x_2 x_3 ... x_n.

method is repeated for each data point.

using (n−1) data point for model fitting.

validation with one data point.

$(n-2)\hat{\sigma}^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$?!

without fitting the model.

$D = \{(x_i, y_i) \mid i = 1, 2, \cdots n.\}$ data set.

If we remove (x_i, y_i) from the data set · and fit the model based on $(n-1)$ data points. and then predict the value of y_i for given x_i then we denote it as $\hat{y}_{(i)}$ and ~~xxx~~ · estimated error is denoted as $e_{(i)} = y_i - \hat{y}_{(i)}$

where $\hat{y}_{(i)}$ is the predicted value with model based on $(n-1)$ data pts. not using (x_i, y_i)

$\sum_{i=1}^{n} e_{(i)}^2$ = Predicted residual sum of square ≡ PRESS.

$= \sum_{i=1}^{n} (y_i - \hat{y}_{(i)})^2$

$\underset{\sim}{Y} \sim N(X\underset{\sim}{\beta}, I_n \underset{=}{\sigma}^2)$

~~xxx~~ $e_i = (y_i - \hat{y}_i) \sim N(0, \sigma^2(1 - h_{ii}))$.

It can be shown $e_{(i)} = \frac{e_i}{1 - h_{ii}} \sim N\left(0, \frac{\sigma^2}{1 - h_{ii}}\right)$

Montgomery C.7 appendix.

$$\frac{e_{(i)}}{\sqrt{V(e_{(i)})}} = \frac{e_i}{\sqrt{\sigma^2(1-h_{ii})}} = \frac{e_i}{\sqrt{V(e_i)}}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1}$$

σ^2 unknown.

$\hat{\sigma}^2 \stackrel{?}{=}$ MSError. not perferable tobe for jack-knife method.

$$\begin{cases} \sigma^2 \stackrel{\wedge}{=} S^2_{(i)} = \dfrac{(n-k-1)\,MSE_{rrorr} - \dfrac{e_i^2}{1-h_{ii}}}{n-k-2} \\ \text{when } (x_i, y_i) \text{ not used.} \end{cases}$$

$$T = \frac{e_i}{\sqrt{S^2_{(i)}(1-h_{ii})}} \sim t_{n-k-2}$$

To test.

H_0: ith obs is not outlier.

H_1: ith obs is an outlier.

if $n \gg k$.

then approximate it by Normal dist.

C-8

$$\left(A + \underset{\sim}{u}\underset{\sim}{v}^T\right)^{-1} = A^{-1} - \frac{A^{-1}\underset{\sim}{u}\underset{\sim}{v}^T A^{-1}}{1 + \underset{\sim}{v}^T A^{-1}\underset{\sim}{u}}$$

$$X_{(i)}^T X_{(i)} = \left(X^T X - \underset{\sim}{x}_i \underset{\sim}{x}_i^T\right)$$

Leverage point in a data set is observed because of $\underset{\sim}{x}$ values.
it is identified with the quantity - $h_{ii} = \underset{\sim}{x}_i^T (X^T X)^{-1} \underset{\sim}{x}_i$.

$$tr(P_x) = tr(H) = \sum_{i=1}^{n} h_{ii} > \frac{(k+1)}{(n/2)}$$ we may suspect that there can be atlast one leverage point(s) in the dataset.

Larg magnitude. h_{ii} or/and $|e_i|$ canbe considered as an influencial point which has a dominant impact in estimation of $\underset{\sim}{\beta}$ or $\underset{\sim}{e}_i$.

VIF = Variance inflation factor. $= \frac{1}{1 - h_{ii}}$

$\hat{\underset{\sim}{\beta}}_{j(i)}$ is estimated value of $\underset{\sim}{\beta}$ when $(\underset{\sim}{x}_i, y_i)$ is not used in estimation. i.e. $(n-1)$ observations are used

Impact of influence point on y_i

Cook's distance: [1977, 79]

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T X^T X (\hat{\beta}_{(i)} - \hat{\beta}) / (k+1)}{\sum_{i=1}^{n} (y_i - \bar{y})^2 / (n-k-1)} = \frac{(\hat{Y}_{(i)} - \hat{Y})^T (\hat{Y}_{(i)} - Y) / (k+1)}{Y(I - P_X)Y / (n-k-1)}$$

NOT an F statistic

$\hat{Y}$ is the predicted value of Y for give X when all data are used to estimate β.

$\hat{Y}_{(i)}$ is the predicted value of Y for given X when ith obs is not used to estimate β

$$\hat{Y}_{(i)} = X \hat{\beta}_{(i)}$$

$$\hat{Y} = X \hat{\beta}$$

$$\hat{\beta}_{(i)} - \hat{\beta} = \frac{-(X^T X)^{-1} x_i e_i}{1 - h_{ii}}$$

ch 6 Montgom.

$$D_i = \frac{e_i^2 h_{ii} / (k+1)}{(1-h_{ii})^2 \, MS\,Error} = \left(\frac{e_i}{\sqrt{(1-h_{ii}) \, MS\,Error}} \right)^2 \frac{h_{ii}}{(1-h_{ii})(k+1)}$$

In prediction how much impact is present of the data. (x_i, y_i) is measured by cook's distance. It can be expressed as a fraction elements which are available from complete data estimation.

For $n \gg k$ D_i is compared with 1. (If $D_i > 1$ we may consider it has an impact in prediction

DFBETA Belsley, Kuh, Welsch (1980).

$\hat{\beta}_j - \hat{\beta}_{j(i)} \sim ?$

$\hat{\beta} \sim N(\beta, C\sigma^2)$.

Denote:

$R = (X^T X)^{-1} X^T$.

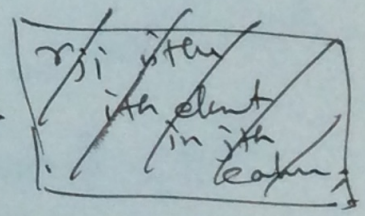
$R^T R = (X^T X)^{-1} = (C_{ij})$.

r_j^T is the jth row of R.

$$\hat{\beta}_j - \hat{\beta}_{j(i)} = \frac{r_{ji}\, e_i}{1 - h_{ii}}$$

everything is known from old estimation process with all data points.

$$\frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{r_j^T r_j\, s_{(i)}^2}} = \frac{r_{ji}}{\sqrt{r_j^T r_j}} \cdot \left(\frac{e_i}{\sqrt{s_{(i)}^2 (1-h_{ii})}}\right) \frac{1}{\sqrt{1-h_{ii}}}$$

$\overset{?}{=}$ 

$$t_{(i)} = \frac{e_i}{\sqrt{s_{(i)}^2 (1-h_{ii})}} \sim t_{n-k-2}.$$

t test to check. H_0: $E(\hat{\beta}_j - \hat{\beta}_{j(i)}) = 0$.

H_1: $E(\hat{\beta}_j - \hat{\beta}_{j(i)}) \neq 0$.

Cov Ratio

$$\text{COVRatio} = \frac{\left|\hat{V}\left(\hat{\beta}_{(i)}\right)\right|}{\left|\hat{V}\left(\hat{\beta}\right)\right|} = \frac{\left|\hat{\sigma}^2_{(i)}\left(X_{(i)}^T X_{(i)}\right)^{-1}\right|}{\left|\hat{\sigma}^2\left(X^T X\right)^{-1}\right|}$$

$$= \frac{\left|s^2_{(i)}\left(X_{(i)}^T X_{(i)}\right)^{-1}\right|}{\left|MSE_{rror}\left(X^T X\right)^{-1}\right|} = \left(\frac{s^2_{(i)}}{MS\,Error}\right)^{k+1} \frac{\left|\left(X_{(i)}^T X_{(i)}\right)^{-1}\right|}{\left|\left(X^T X\right)^{-1}\right|}$$

$$= \left(\frac{s^2_{(i)}}{MS\,Error}\right)^{k+1} \frac{1}{1-h_{ii}}$$

$$\hat{\beta}_{(i)} \sim N\left(\beta,\ \sigma^2\left(X_{(i)}^T X_{(i)}\right)^{-1}\right)$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \underset{n \times k}{\left(\quad X \quad\right)} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \underset{\sim}{\epsilon}.$$

- Cov Ratio. $\gtrless 1 \pm 3\left(\frac{k+1}{n}\right)$. → ith obs has impact.
- $|DFBETA_{j(i)}| > 2/\sqrt{n}$. ⇒ ith obs has impact.
- $DFFIT_i = \dfrac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{s_{(i)}^2 h_{ii}}} = \left(\dfrac{h_{ii}}{1-h_{ii}}\right)^{1/2} t_{(i)}$.
- $|DFFIT_i| > \dfrac{2(k+1)}{n}$ then ith obs has impact.

$x_i^T (X^T X)^{-1} x_i$ is large enough.

- or. $\text{Tr}(H) > \dfrac{2(k+1)}{n}$. then there may exists leverage points.

Variable selection & Model building.

Assume the true model $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$ $\quad \underset{\sim}{\beta} \in \mathbb{R}^{k+1}$, $\underset{\sim}{y} \in \mathbb{R}^n$, $\underset{\sim}{\epsilon} \in \mathbb{R}^n$

But when we fit the model we use $p < k+1$ variables instead $(k+1)$.

$$\underset{\sim}{Y} = \begin{bmatrix} X_p & \vdots & X_r \end{bmatrix} \begin{pmatrix} \underset{\sim}{\beta}_p \\ \underset{\sim}{\beta}_r \end{pmatrix} + \underset{\sim}{\epsilon}$$

$$\Rightarrow \underset{\sim}{Y} = X_p \underset{\sim}{\beta}_p + X_r \underset{\sim}{\beta}_r + \underset{\sim}{\epsilon} \text{. where. } p+r = k+1.$$

$$\underset{\sim}{\beta}_p = (\beta_0 \beta_1 \cdots \beta_{k-r})^T \qquad \underset{\sim}{\beta}_r = (\beta_{k-r+1} \cdots \beta_k)^T$$

Fitted model.

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

$$\hat{\underset{\sim}{\beta}}_p = (X_p^T X_p)^{-1} X_p^T \underset{\sim}{Y} \longrightarrow \text{Reduced model.}$$

$$\begin{pmatrix} \underset{\sim}{\beta}_p \\ \underset{\sim}{\beta}_r \end{pmatrix} \triangleq \hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{Y} \longrightarrow \text{complete model.}$$

Q1 whether $\hat{\beta}_p$ is an unbiased estimator of β_p under reduced model?

$\hat{\beta}_p = (X_p^T X_p)^{-1} X_p^T Y$ under reduced fitted model.

True model. $Y \sim N(X\beta, \sigma^2 I_n)$.

$$E(\hat{\beta}_p) = (X_p^T X_p)^{-1} X_p^T (X\beta)$$

$$= (X_p^T X_p)^{-1} X_p^T [X_p \beta_p + X_r \beta_r]$$

$$= (X_p^T X_p)^{-1} (X_p^T X_p) \beta_p + (X_p^T X_p)^{-1} (X_p^T X_r) \beta_r$$

Assume $(X_p^T X_p)$ is invertable.

$$= \beta_p + (X_p^T X_p)^{-1} (X_p^T X_r) \beta_r$$

$\Rightarrow$ In general $E(\hat{\beta}_p) \neq \beta_p$ but if $X_p^T X_r = 0$ ie. the colums of X_p are orthogonal to the columns of X_r then only $E(\hat{\beta}_p) = \beta_p$

→ orthogonal polynomial.

→ Principal component regression.

Q2 Is $\hat{\sigma}_p^2$, under reduced fitted model, an unbiased estimator of σ^2?

$$\hat{\sigma}_p^2 = \frac{\underset{\sim}{Y}^T \left(I_n - X_p (X_p^T X_p)^{-1} X_p^T \right) \underset{\sim}{Y}}{n-p}.$$

Theorem.
$Y \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n)$

$$E\left(\hat{\sigma}_p^2\right) = \frac{\sigma^2 (n-p) + \underset{\sim}{\beta}^T X^T \left(I_n - X_p (X_p^T X_p)^{-1} X_p \right) X\underset{\sim}{\beta}}{(n-p)}$$

$$= \sigma^2 + \frac{\underset{\sim}{\beta}^T X^T \left(I_n - X_p (X_p^T X_p)^{-1} X_p^T \right) X\underset{\sim}{\beta}}{(n-p)}.$$

$$= \sigma^2 + \frac{\underset{\sim}{\beta}_r^T X_r^T \left(I_n - X_p (X_p^T X_p)^{-1} X_p^T \right) X_r \underset{\sim}{\beta}_r}{n-p}.$$

$\geq \sigma^2$ as $\left(I_n - X_p (X_p^T X_p)^{-1} X_p^T \right)$ is a p.s.d matrix.

$\underset{\sim}{Z} \sim N(\underset{\sim}{\mu}, \sigma^2 I_n)$

$E(\underset{\sim}{Z}^T A \underset{\sim}{Z})$
$= tr(A I_n \sigma^2) + \underset{\sim}{\mu}^T A \underset{\sim}{\mu}$

A is square, symmetric & idempotent.

$X\underset{\sim}{\beta} = X_p \underset{\sim}{\beta}_p + X_r \underset{\sim}{\beta}_r$

If $\mathcal{C}(X_p) = \mathcal{C}(X_r)^{\perp}$

then. $\left(I_n - X_p (X_p^T X_p)^{-1} X_p^T \right)$ is the projection matrix of $\mathcal{C}(X_r)$

Q3 When $\hat{\beta}_p$ (reduced model) has less variance compared the estimator of β_p under complete model.?

$$D\left(\hat{\beta}_p\right) = \sigma^2 \left(X_p^T X_p\right)^{-1}$$

$$D\left(\hat{\beta}\right) = \sigma^2 \left(X^T X\right)^{-1} = \sigma^2 \left(\left[X_p : X_r\right]^T \left[X_p : X_r\right]\right)^{-1}$$

$$= \sigma^2 \begin{pmatrix} X_p^T X_p & ! & X_p^T X_r \\ - & - & - \\ X_r^T X_p & ! & X_r^T X_r \end{pmatrix}^{-1}$$

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}.$$

$$A^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & \left(-\left[A_{11} - A_{12}A_{22}^{-1}A_{21}\right]^{-1} A_{12}A_{22}^{-1}\right) \\ \left(-A_{21}A_{11}^{-1}\left[A_{22} - A_{21}A_{11}^{-1}A_{12}\right]\right) & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$$