NAME: Vandana Chudasama

STD: MSC DIV: Previous ROLL NO.:

YOUVA

SUBJECT: Linear Model

INDEX

SR. NO.	DATE	TITLE	PAGE NO.	TEACHER'S SIGN

* Linear Estimator

The estimator which are linear functions of the observations are known as linear estimator

Concept of unbiased linear estimator of linear function of parameter

→ Sample mean $\bar{Y} = \frac{1}{n} \Sigma Y_i$ is an unbiased estimator of population mean μ, $\hat{\mu} = \bar{Y}$

→ $S^2 = \frac{1}{n-1} \Sigma (Y_i - \bar{Y})^2$ is an unbiased estimator of population variance σ^2
But S^2 is not a linear estimator S^2 is a quadratic estimator of σ^2

If we have more than one unbiased linear estimator to select the best we use BLUE.

The principle of BLUE is to choose that linear estimator from among all unbiased linear estimators which has smallest variance

The sampling distribution of that estimator which have maximum concentration amount the unknown through parameteric function Such an estimator is known as minimum variance unbiased linear estimator.

Note:- This minimum variance unbiased linear estimator have variance and covariance which again lend themself to unbiased estimators

The minimum variance approach to the estimation of the parameters in the linear model is given by markov (1990)

The least square method which is a practical method of estimating unknown parameter in a linear model was published independently by Gauss (1809) and legendre (1806) in books on astronomical problems the combine result is Gauss Markov theorem.

* ## Gauss - Markov Linear Model

Consider a set of 'n' indep r.v's $Y_1, \ldots Y_n$ with a common variance σ^2 whoose expectations are linear function with known coeff (a_{ij}'s) of p unknown parameters $\beta_1, \beta_2, \ldots \beta_p$ with ($p < n$)

Thus
$$\left.\begin{aligned} E(Y_i) &= a_{i1}\beta_1 + a_{i2}\beta_2 + \ldots + a_{ip}\beta_p \\ V(Y_i) &= \sigma^2 \quad \text{for } i = 1, 2, \ldots n \\ Cov(Y_i, Y_j) &= 0 \quad \forall\ i \neq j \end{aligned}\right\} \quad (1)$$

Eq (1) is called gauss Markov Linear Model

We represent (1) in matrix form
let
$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \qquad \underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

Matrix of coefficient
$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1p} \\ a_{21} & a_{22} & \ldots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{np} \end{bmatrix}$$

where $D(\underline{Y})$ stands for Dispension - matrix and I is identity of order n

Eq (1) can be written as

$$\left.\begin{aligned} E(\underline{Y}) &= A\underline{\beta} \\ D(\underline{Y}) &= \sigma^2 I \end{aligned}\right\} \quad (2)$$

An Alternative way of representing eq (1) using column vector $\underline{e}$ of indp errors.

$$\left.\begin{aligned} &\underline{Y} = A\underline{\beta} + \underline{e} \\ &E(\underline{e}) = \underline{0} \text{ and } D(\underline{e}) = \sigma^2 I \\ &\text{where } \underline{0} \text{ is a null vector} \end{aligned}\right\} \quad (3)$$

The unknown parameter β_j's in model are called effects.

Fixed Effect Model

In linear Estimations the effects are called fixed quantities and such a model where all effects are unknown parameters is called a fixed effect model or model I.

Sometimes one of the β_j's is a constant with $a_{ij} = 1$ for that j and call $i = 1, 2, \ldots n$ such an effect is called general effect or an edit constant

Random Effect Model

If effects are Random in a model then it is called random effect model.

Mixed Effect Model

If some effects are fixed and some are random then it is called Mixed effect Model

Gauss - Markov linear Model may be classified into three Broad categories depending on the nature of the values taken by coeff. a_{ij}'s.

→ Model in which a_{ij}'s are indicator variables taking values 1 or 0 then such a model is called analysis of variance model

→ A model in which a_{ij}'s are values taken by independent variables also called concomitant or regressor variables then we have a Regression Model

→ A model in which both types of a_{ij}'s are present i.e indicator as well as independent variables then the model is called ~~ANCOV~~ ANCOVA Model

→ Definition : Parametric function
A linear function of the parameter $\beta_1, \beta_2, \ldots \beta_p$ is called a parametric function
The constant of a function to be known

Thus $l_1\beta_1 + l_2\beta_2 + \cdots + l_p\beta_p = l'\beta$ is a parametric function.

If there exists a vector $c = (c_1, c_2, \ldots c_n)'$ of constant such that $E(\underline{c}'\underline{Y}) = \underline{l}'\underline{\beta}$ indp of parameter then the parameteric function $\underline{l}'\underline{\beta}$ is said to be linearly estimable or to be estimable function.

Thm:- (1) A Neccessary and sufficient condition for the linear function $\underline{l}'\underline{\beta}$ of the parameter to be linearly estimate is

$$\text{rank}(A) = \text{rank}\begin{pmatrix} A \\ l' \end{pmatrix}$$

where $\begin{pmatrix} A \\ l' \end{pmatrix}$ is the matrix obtained from A by adjoining the row vector l'

Proof We see that $l'\beta$ is estimable iff there exists a vector of constant $\underline{c}$ such that

$$E(\underline{c}'\underline{Y}) = \underline{l}'\underline{\beta} \quad \forall \beta$$

But $E(\underline{c}'\underline{Y}) = \underline{c}'E(\underline{Y}) = \underline{c}'A\underline{\beta}$ $(\because E(Y) = A\underline{\beta})$
So $l'\beta$ is estimable iff

$$\underline{c}'A\underline{\beta} = \underline{l}'\underline{\beta} \text{ identically in } \beta$$
$$\text{iff } \quad \underline{c}'A = \underline{l}' \quad -(1)$$

To find an unbiased linear estimator for $\underline{l}'\underline{\beta}$ we have to solve (1) for the unknown vector $\underline{c}$ and the condition for existence of a solution to (1) is

$$\text{rank}(A) = \text{rank}\begin{pmatrix} A \\ l' \end{pmatrix}$$

* **Corollary 1:** If rank (A) = P then every linear function of the parameter is linearly estimate

Proof: whatever may be l giving parametric function $\underline{l'\beta}$

$$\text{rank}\begin{pmatrix} A \\ l' \end{pmatrix} \geq \text{rank}(A)$$

But rank $\begin{pmatrix} A \\ l' \end{pmatrix}$ cannot exceeds p since $\begin{pmatrix} A \\ l' \end{pmatrix}$ has only p columns. Hence if

$$\text{rank}(A) = P$$

then $$\text{rank}\begin{pmatrix} A \\ l' \end{pmatrix} = \text{rank}(A)$$

So by thm (1) $\underline{l'\beta}$ is estimable.

* **Corollary (2)** Every linearly estimable parameter function is of the form $b'A\beta$

OR

$b_1 E(Y_1) + b_2 E(Y_2) + \ldots + b_n E(Y_n)$, where $\underline{b'} = (b_1, b_2, \ldots b_n)$

Proof: If $l'\beta$ is estimable then by (1) l' is linearly independent in the rows of A. So that $\underline{l'}$ is of the form $\underline{b'A}$

Hence $\underline{l'\beta} = \underline{b'}A\underline{\beta}$

* **Definition:-** A linear function of $Y_1, Y_2, \ldots Y_n$ is said to belong to error if its expectations vanishes independently of parameters. If $d' = (d_1, d_2, \ldots d_n)$ then $\underline{d'Y}$ belongs to error if and only if

$$E(\underline{d}'\underline{Y}) = \underline{d}'E(\underline{Y}) = \underline{d}'A\underline{\beta} = 0 \ \forall \beta$$

i.e iff $\underline{d}'A = \underline{0}$ or $A'\underline{d} = 0$

Thm A Necessary and Sufficient condition for the linear function $\underline{d}'\underline{Y}$ to belong to Error is that $\underline{d}'$ is orthogonal to the vector space $V(A')$ generated by the row vector of A

Error Space

Suppose $p < n$ then the rank of vector space $V(A')$ cannot exceed p. Let rank of $V(A') = n_0$ $n_0 \leq p$. Then the rank of vector space orthogonal to $V'(A')$ is let us say $n_e = n - n_0$.

Let the row vector of the matrix B of order $n_e \times n$ generate a basis of the vector space orthogonal to $V(A')$ the $BA = 0$

From above result the coeff vector d' of any linear function belonging to error lies in vector Space B.

Hence the vector Space $V(B)$ which is orthogonal to vector Space $V(A')$ is called error space and it has rank $n_e = n - n_0$ and the space $V(A')$ which contains the coefficient vector of the best estimators of estimable functions is known as estimable space and it has rank n_0.

rm If $\underline{l}'\underline{\beta}$ is any estimable linear function of the parameter $\beta_1, \beta_2, \dots \beta_p$ then. Prove that

(1) There exists a unique linear function $\underline{C}'Y$ of the random var $Y_1, Y_2, \dots Y_n$ $\exists$ $c \in V(A')$ and $E(\underline{c}'\underline{Y}) = \underline{l}'\underline{\beta}$

(2) Var $(c'Y)$ is less than the variance of any other linear unbiased estimator of $\underline{l}'\underline{\beta}$

roof (1) Since $\underline{l}'\underline{\beta}$ is estimable, $\exists$ a linear function $\underline{b}'\underline{Y}$ of the r.v's $\ni E(\underline{b}'\underline{Y}) = \underline{l}'\underline{\beta}$

Now can uniquely resolve $\underline{b}'$ into $\underline{c}'$ & $\underline{d}'$ such that $\underline{c}' \in V(A')$ and $\underline{d}' \in V(\beta)$ which is orthogonal to $V(A')$

Hence $\underline{b}'\underline{Y} = \underline{C}'\underline{Y} + d'Y$
where
$d'Y$ belongs to error and also

$E(\underline{C}'\underline{Y}) = E(\underline{b}'\underline{Y}) = \underline{l}'\beta$
$E(\underline{d}'\underline{Y}) = 0$
$E(\underline{b}'\underline{Y}) = E(\underline{C}'\underline{Y}) + E(\underline{d}'\underline{Y})$
$E(\underline{b}'\underline{Y}) = E(\underline{C}'\underline{Y})$

Thus $\exists$ a linear function $c'Y$ with $c' \in V(A')$ such that $\underline{E}(\underline{C}'\underline{Y}) = \underline{l}'\underline{\beta}$

Now to show this is unique.
If possible let there be another row vector $\underline{C}_0' \in V(A')$
such that $E(\underline{C_0}'\underline{Y}) = \underline{l}'\underline{\beta}$

Define : the row vector $\underline{c}_1' = \underline{c}' - c_0'$

Then $E(\underline{c}_1'\underline{Y}) = E(\underline{c}'Y) - E(\underline{c_0}'\underline{Y})$
$= 0 - 0 = 0$

Thus c_1' belongs to Error Space and being a linear combination of $\underline{c}'$ and $\underline{c_0}'$ also belongs to estimation space.

But this is impossible unless c_1' is null vector because same non-null vector cannot lie in two orthogonal spaces.

Thus, so $\underline{c_1}'$ is null vector $\Rightarrow \underline{c}' = c_0'$
Thus, the vector $\underline{c}'$ which lies in vector space $V(A')$ and for what
$E(\underline{c}'Y) = \underline{l}'\beta$ is unique.

(ii) Next to show that $\underline{l}'\underline{Y}$ has minimum variance among all linear unbiased estimators

let $b'Y$ be any arbitary unbiased estimator of $\underline{l}'\beta$

$$
\begin{aligned}
var(\underline{b}'\underline{Y}) &= \underline{b}'\underline{b}\ var(\underline{Y}) \\
&= \underline{b}'\underline{b}\ \sigma^2 \\
&= (\underline{c}'+\underline{d}')(\underline{c}+\underline{d})\sigma^2 \quad (\because \underline{b}' = \underline{c}'+\underline{d}') \\
&= (\underline{c}'\underline{c} + \underline{c}'\underline{d} + \underline{d}'\underline{c} + \underline{d}'\underline{d})\sigma^2 \\
&= \underline{c}'\underline{c}\sigma^2 + \underline{d}'\underline{d}\ \sigma^2 \quad (\because \underline{c}'\underline{d} = 0) \\
var(\underline{b}'\underline{Y}) &= var(\underline{c}'\underline{Y}) + var(\underline{d}'\underline{Y})
\end{aligned}
$$

Hence $var(\underline{b}'\underline{Y}) \geq var(\underline{c}'\underline{Y})$

and equality hold iff $var(\underline{d}'Y) = 0$
i.e iff $d'\underline{d} = 0$ or d' is a null vector

In that case $\underline{b'Y}$ coincides with $\underline{c'Y}$ and 0
Hence the proof

(1) Corrollary (1) : The best estimator of any estimable function $\underline{l}'\beta$ must be of the form $q'A'\underline{Y}$ where $q' = (q_1, q_2, \ldots q_p)$ is a row vector and satisfies the eqn $q'A'A = \underline{l}$

Proof The coefficient vector c' of the best estimator lies in $V(A')$ Hence $\underline{c}' = q'A$ for a suitable q' and the best estimator is of the form

$$\underline{c'Y} = q'A\underline{Y}$$

Since $E(q'A\underline{Y}) = \underline{l}'\beta$ we have $q'AA = \underline{l}'$

Corollary 2 Let $\underline{l}_i'\beta$ $i = 1, 2, \ldots k$ be k estimable function and let $T_i = \underline{C_i'Y}$ for $i = 1, 2, \ldots k$ be there best estimator. Then the best estimator of

$$\sum_i b_i \underline{l}_i'\beta \text{ is } T = \sum b_i T_i$$

Proof Clearly $E(T) = E(\Sigma b_i T_i) = \Sigma b_i E(T_i)$
$= \Sigma b_i E(C_i'\underline{Y})$
$= \Sigma b_i \underline{l}_i'\beta$

Hence T is an unbiased estimator of $\Sigma b_i \underline{l}_i \beta$

Since, $T = (b_1 C_1^2 + b_2 C_2^2 + \ldots + b_k C_k^2) \underline{Y}$
the coeff. vector of T lies in $V(A')$ which shows that T must be the least estimator of $\Sigma b_i \underline{l_i}' \beta$

⋆ least Square estimators and normal equations

A set of measurable function $\underline{Y}$ say
$\hat{\beta}_1 = \hat{\beta}_1(\underline{Y}),\ \hat{\beta}_2 = \hat{\beta}_2(\underline{Y}), \ldots \hat{\beta}_p = \hat{\beta}_p(\underline{Y})$

Such that the values $\hat{\beta}_1, \hat{\beta}_2, \ldots \hat{\beta}_p$ minimize the sum of square of the deviations $Y_1, Y_2, \ldots Y_n$ from expectations i.e

$$S = (\underline{Y} - A\underline{\beta})'(\underline{Y} - A\underline{\beta})$$

is called a set of LS estimator of the unknown parameter $\beta_1, \beta_2, \ldots \beta_p$ of linear model (2).

To show that the min value of S is attained when $\hat{\underline{\beta}}$ is a solution of a set of eq^n which are called normal equations.

we have $\quad \underline{\beta}'A'\underline{Y} = \underline{Y}'A\underline{\beta}$

$$= \sum_{\alpha j} a_{\alpha j} \beta_j Y_\alpha \quad \text{where } i = 1, 2, \ldots p, \quad \alpha = 1, 2, \ldots n$$

Hence

$$\frac{d}{d\beta_j}(\underline{\beta}'A'\underline{Y}) = \frac{d}{d\beta_j}(\underline{Y}'A\underline{\beta}) = \sum_{\alpha j} a_{\alpha j} Y_\alpha = d_j'\underline{Y}$$

where $d_1, d_2, \ldots d_p$ are column vectors of A
Let $A'A = C = (c_{ij})$

Clearly c is a symmetric matrix of order p.

Now

$\underline{\beta}' A' A \underline{\beta} = \underline{\beta}' C \underline{\beta} = \sum_{i,j} C_{ij} \beta_i \beta_j$ with $i,j = 1, 2, \dots p$

Hence

$$\frac{d}{d\beta_j} (\underline{\beta}' A' A \underline{\beta}) = \frac{d}{d\beta_j} (\underline{\beta}' C \underline{\beta})$$

$$= 2 \sum_j C_{ij} \beta_i = \underline{C}_j' \beta$$

where $\underline{c}_j' = (C_{1j}, C_{2j}, \dots C_{pj})$

Differentiating $S = (\underline{Y} - A\underline{\beta})'(\underline{Y} - A\underline{\beta})$ w.r.t $\beta_1, \beta_2, \dots \beta_p$ and equating derivative to zero

$$S = (\underline{Y} - A\underline{\beta})'(\underline{Y} - A\underline{\beta})$$
$$= (\underline{Y}'\underline{Y} - \underline{Y}'A\underline{\beta} - \underline{\beta}'A'\underline{Y} + \underline{\beta}'A'A\underline{\beta})$$
$$= (\underline{Y}'Y - 2\underline{\beta}'A'\underline{Y} - \beta'A'A\underline{\beta})$$

$$\frac{dS}{d\beta_j} = -2A'\underline{Y} + 2A'A\underline{\beta}$$
$$= -2d_j'\underline{Y} + 2\underline{C}_j'\underline{\beta} = 0 \quad , j = 1, 2, \dots p$$

$$\Rightarrow -\underline{d}_j'\underline{Y} + \underline{C}_j'\underline{\beta} = 0 \quad \text{--- } (*)$$

Equation (*) are normal Eqns and are equivalent to

$$\left.\begin{array}{l} A'A\underline{\beta} = A'\underline{Y} \\ \text{or } C\underline{\beta} = A'\underline{Y} \end{array}\right\} (**) \text{ where } C > A'A$$

The normal Equations always admit a solution Since, $A'Y$ lies in the vector space generated by the columns of C

Let $\hat{\underline{\beta}}$ be a solution of these equations

$$\hat{\underline{\beta}} = (A'A)^{-1}A'Y$$

Note :- Every solution of the normal equations is a set of list square estimators and every set of least square estimators satisfies the normal eq.

Result : The solutions of normal eq^{ns} $\underline{\beta} = \hat{\underline{\beta}}$ gives an extreme value of S and this extreme value is min value of S.

i.e $\underline{\beta} = \hat{\underline{\beta}}$ minimies S

proof Consider $(\underline{Y} - A\underline{\beta})'(\underline{Y} - A\underline{\beta})$

$$= (\underline{Y} - A\hat{\underline{\beta}} + A\hat{\underline{\beta}} - A\underline{\beta})'(\underline{Y} - A\hat{\underline{\beta}} + A\hat{\underline{\beta}} - A\hat{\underline{\beta}})$$

$$= (\underline{Y} - A\hat{\underline{\beta}})'(\underline{Y} - A\hat{\underline{\beta}}) + (\hat{\underline{\beta}} - \underline{\beta})'A'A(\hat{\underline{\beta}} - \underline{\beta})$$

$$\geq (\underline{Y} - A\hat{\underline{\beta}})'(\underline{Y} - A\hat{\underline{\beta}}) \quad — (***)$$

Since the quddratic form.

$[A(\hat{\underline{\beta}} - \underline{\beta})]'[A(\hat{\underline{\beta}} - \underline{\beta})]$ cannot be negative the equality holds only when $\underline{\beta} = \hat{\underline{\beta}}$

Thus $\underline{\beta} = \hat{\underline{\beta}}$ minimize S

Further if $\hat{\underline{\beta}}$ and $\hat{\hat{\underline{\beta}}}$ are any two solutions of (*) then.

$$(\underline{Y} - A\hat{\underline{\beta}})'(\underline{Y} - A\hat{\underline{\beta}}) = (\underline{Y} - A\hat{\hat{\underline{\beta}}})'(\underline{Y} - A\hat{\hat{\underline{\beta}}})$$

Thus valong with (***) show that every solution of the normal eq is a set of ~~linear~~ least square estimating

Theorem **Gauss Markov theorem**

The best estimator of the estimable linear function $l'\beta$ of the parameters is $l\hat{\beta}$ where $\hat{\beta}_1, \hat{\beta}_2, \ldots \hat{\beta}_p$ are a set of least square estimators $\beta_1, \beta_2, \ldots \beta_p$. In other words the least square estimators $l'\beta$ is identical with the minimum variance linear unbiased estimator $l'\hat{\beta}$

proof from corollary (1) we have

BLUE of $\underline{l}'\underline{\beta} = q'A'Y = q'A'A\hat{\underline{\beta}} = \underline{l}'\hat{\underline{\beta}}$

Since from the normal equations we have since q' satisfies $q'A'A = l'$

$$A'A\hat{\underline{\beta}} = A'\underline{Y}$$

Since $\underline{l}'\beta = q'A\underline{Y}$

$\underline{l}'\hat{\beta}$ is the same for all LS estimator $\hat{\underline{\beta}}'$ of $\underline{\beta}$ obtained by solutiong normal Eq^n

Result To show that $\hat{\beta}_j$ is a unique linear unbiased estimator of β_j for $j = 1, 2, \ldots p$ with $var(\hat{\beta}_j) = \sigma^2 C^{ij}$

$$cov(\hat{\beta}_j, \hat{\beta}_j) = \sigma^2 C^{ij}$$

proof If rank (A) = p the no. of unknown parameter β_j, Since usually $p < n$ this is known as the cause of full rank

Since rank(C) = rank (A) $[\because C > A'A)$

So C is non-singular.

Thus the inverse matrix $C^{-1} = C^{ij}$ exists and unique solution of β is given by

$$\hat{\beta} = C^{-1}A^{-1}Y \qquad (C^{-1} = (A^{-1}A)^{-1})$$

To show $\hat{\beta}$ is unbiased

$$\hat{\beta} = C^{-1}A'Y = (A'A)^{-1}A'Y$$

$$\begin{aligned}
E(\hat{\beta}) &= E[C^{-1}A'Y] \\
&= E[C^{-1}A'(A\beta + \varepsilon)] \\
&= E[C^{-1}A'A\beta] + E[C^{-1}A'\varepsilon] \\
&= E[C^{-1}A'A\beta] + 0 \\
&= E[C^{-1}C\beta] = \beta
\end{aligned}$$

Dispersion Matrix of $\hat{\beta}$ is $D(\hat{\beta})$

$$\begin{aligned}
D(\hat{\beta}) &= E[\{\hat{\beta} - E(\hat{\beta})\}\{\hat{\beta} - E(\hat{\beta})\}'] \\
&= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\
&= E[(C^{-1}A'\varepsilon)(C^{-1}A'\varepsilon)'] \\
&= C^{-1}A'AC^{-1}E(\varepsilon\varepsilon') \\
&= C^{-1}\sigma^2 \\
&= \sigma^2 \begin{bmatrix} c^{11} & c^{12} & \cdots & c^{1p} \\ c^{21} & c^{22} & & c^{2p} \\ \vdots & \vdots & & \vdots \\ c^{p1} & c^{p2} & \cdots & c^{pp} \end{bmatrix}_{p\times p}
\end{aligned}$$

$$V(\varepsilon) = E(\varepsilon\varepsilon') - (E(\varepsilon))^2$$

$$\sigma^2 = E(\varepsilon\varepsilon') - 0$$

$$\begin{aligned}
&\hat{\beta} - \beta \\
&= C^{-1}A'Y - \beta \\
&= C^{-1}A'(A\beta + \varepsilon) - \beta \\
&= C^{-1}A'A\beta + C^{-1}A\varepsilon - \beta \\
&= (A^{-1}A)^{-1}A'A\beta + C^{-1}A'\varepsilon - \beta \\
&= C^{-1}A'\varepsilon
\end{aligned}$$

$\hat{\beta}_j$ is a unique linear unbiased estimator of β_j for $j = 1, 2, \ldots p$ with

$var(\hat{\beta}_j) = \sigma^2 C^{ii}$
$cov(\hat{\beta}_j, \hat{\beta}_j) = \sigma^2 C^{ii'}$

G-Inverse of a matrix

Let A be $m \times n$ matrix then a g-inverse of A denoted by A^- is $n \times m$ matrix. Such that for any arbitary vector Y for which the Equations $Ax = \underline{Y}$ are consistant $x = A^-\underline{Y}$ is a Solutions of equations

A^- is not unique but A^- exists and the identity $AA^- = A$ holds and when A is non-singular theyn the Soln is uniquely determined.

Model 2

Model with linear restrictions on parameter when the parameter are Subject to a Set of consistant linear restrictions β

$$P'\underline{\beta} = \underline{L}$$

rank $(p) = P - S$, the appropriate linear model is

$$\left.\begin{array}{l} E(\underline{Y}) = A\underline{\beta} \\ D(\underline{Y}) = \sigma^2 \underline{I} \\ P'\beta = \underline{L} \end{array}\right\}$$

In this case there are two methods first we eliminate some of the parameters in the observational equations with the help of equations in the linear restrictions and obtain a different set of observational equation with fewer parameter and having no restrictions.

(B)

On this fewer parameter this will be similar to model

$$\left.\begin{aligned} E(\underline{Y}) &= A\underline{\beta} \\ D(\underline{Y}) &= \sigma^2 I \end{aligned}\right\}$$

In this approach $\underline{\beta}_0 + F\underline{\beta}^*$ be general soln of $\underline{P}'\underline{\beta} = \underline{L}$ and $\underline{\beta}_0$ is a particular soln and $P'F = 0$ & β^* being the arbitary vector of fewer parameter (s in no.) with

$$\underline{Z} = \underline{Y} - A\underline{\beta}_0$$

$$E(\underline{Z}) = AF\underline{\beta}^*$$

$$D(\underline{Z}) = \sigma^2 I$$

Thus with $\underline{Z}$ the model reduces to the original model less parameters $\beta_1^*, \beta_2^*, \ldots \beta_s^*$ and no. restrictions on β^*'s.

To estimate $\underline{l}'\underline{\beta} = \underline{l}'(\underline{\beta}_0 + F\underline{\beta}^*)$

$$= \underline{l}'\underline{\beta}_0 + \underline{l}'F\underline{\beta}^*$$

and we need to consider only $\underline{l}'F\underline{\beta}^*$

Second Approach

Second approach is to minimize
$(\underline{Y}-A\underline{\beta})'(\underline{Y}-A\underline{\beta})$ subject to the conditions
$\underline{P}'\underline{\beta} = L$

Using a lagrangan multiplies λ the minimize variance unbiased estimator

$\underline{l}'\underline{\beta}$ is $\underline{l}\hat{\beta}$ where $\hat{\beta}$ is a solution of normal equation obtained by minimizing

$(\underline{Y}-A\underline{\beta})'(\underline{Y}-A\underline{\beta})$ such the restriction
i.e

$$S^2 = (\underline{Y}-A\underline{\beta})'(\underline{Y}-A\underline{\beta}) + \lambda(\underline{P}'\underline{\beta}-\underline{L})$$

The normal Equation are

$$A'A\underline{\beta} + \underline{P}\underline{\lambda} = A'\underline{Y}$$

$$\underline{P}'\underline{\beta} = \underline{L}$$

$$\begin{bmatrix} A'A & \underline{P} \\ \underline{P}' & \underline{0} \end{bmatrix} \begin{bmatrix} \underline{\beta} \\ \underline{\lambda} \end{bmatrix} = \begin{bmatrix} A'Y \\ \underline{L} \end{bmatrix}$$

$$\begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} A'A & \underline{P} \\ \underline{P}' & 0 \end{pmatrix}^{-1} \begin{pmatrix} A'\underline{Y} \\ \underline{L} \end{pmatrix}$$

$$\hat{\underline{\beta}} = (A'A)^{-1}A'\underline{Y} + P^{-1}\underline{L}$$
$$\hat{\underline{\lambda}} = \underline{P}'A'Y$$

Model 3

linear estimator with correlated variables
we consider the model which is more general that gauss markov model

$$E(\underline{Y}) = A\underline{\beta}$$
$$D(\underline{Y}) = \sigma^2 B \quad \text{—(1)}$$

where σ^2 is unknown positive constant & B is a known matrix. B is symmetrical and we assume that $|B| \neq 0$

Since B is assumed to be a non-singular matrix of order n. Their exists a non-singular matrix H of order n such that
$$H'BH = I$$

Consider the transformation
$$\underline{Y}^* = H'\underline{Y}$$
$$E(\underline{Y}^*) = E(H'\underline{Y})$$
$$= H'E(\underline{Y})$$
$$= H'A\underline{\beta}$$
$$= A^*\beta \qquad \text{where } A^* = H'A$$

Dispersion
$$\left.\begin{aligned} D(\underline{Y}^*) &= D(H'\underline{Y}) \\ &= H' D(\underline{Y}) H \\ &= H' \sigma^2 BH \\ &= \sigma^2 I \end{aligned}\right\} \text{(2)}$$

So, $\exists$ a linear transformation which reduces the case of correlated variables to uncorrelated variables with matrix of coefficient

$$A^* = H'A$$

$$\text{rank}(A^*) = \text{rank}(H'A)$$
$$= \text{rank}(A).$$

So all the result of model 1 will also be true for this model.

Note:- In applications it is simple to use actual observations $(Y_1, Y_2, \ldots Y_n)$ instead of transformed observations $(Y_1^*, Y_2^*, \ldots Y_n^*)$

The least square estimators may be found by minimizing following sum of square in terms of transformed variables.

$$(\underline{Y}^* - A\underline{\beta}^*)'(\underline{Y}^* - A\beta^*)$$

Since $(\underline{Y}^* - A\beta^*) = (H'\underline{Y} - H'A\underline{\beta})$
$= H'(\underline{Y} - A\underline{\beta})$ & $H'H = B^{-1}$

we have $(Y' - A\underline{\beta})' HH' (\underline{Y} - A\underline{\beta})$
$= (\underline{Y}' - A\beta)' \beta^{-1} (\underline{Y} - A\underline{\beta})$

where $\beta^{-1} = \beta^{ij}$

the above model can be explained as.

$$\sum_i \sum_j b^{ij} (Y_i - a_{i1}\beta_1 - a_{i2}\beta_2 \ldots a_{ip}\beta_p)(Y_j - a_{j1}\beta_1 - a_{j2}\beta_2 \ldots a_{jp}\beta_p)$$

and is called weighted sum squared least square estimators of β's are found by minimizing weighted sum of squares in this case of correlated variance

Remark

(1) If the variables are uncorrelated but do not have constant variance σ^2 then matrix B will be a diagonal matrix with $b_1, b_2, \ldots b_n$ as diagonal elements.

Here b_i will be proportional to the variance of Y_i the weighted sum of squared which is to be minimized in this case is to ~~be minimized~~ ~~in this~~ ~~great~~ get the least square estimator of β's is

$$\Sigma (Y_i - a_{i1}\beta_1 - a_{i2}\beta_2 - \cdots a_{ip}\beta_p{}^2) / b_i$$

Model 1 is a special case with weights $1/b_i$ all equal i.e ~~also~~ case of constant variance

(2) In case of correlated variables if the parameters are also subject to linear restrictions then $(\underline{Y} - A\underline{\beta}')\beta^{-1}(\underline{Y} - A\underline{\beta})$ has to be minimized with linear restrictions

$\underline{P'\beta} = \underline{L}$ in order to get the normal eq^n

The normal eq^{n}s are

$$\left.\begin{array}{c} A'\beta^{-1}A\underline{\beta} + \underline{P\lambda} = A'B^{-1}\underline{Y} \\ P' = \underline{L} \end{array}\right\}$$

Simultaneous Estimation of parametric funcs

Consider the general setup $(\underline{Y}, X\beta, \Sigma)$ with or without restriction on the parameter B, let $\underline{P}_1'\hat{\underline{\beta}}, \underline{P}_2'\hat{\underline{\beta}}, \dots \underline{P}_k'\hat{\underline{\beta}}$

Individual least square estimator of the parametric function $\underline{P}_1'\underline{\beta}, \underline{P}_2'\underline{\beta}, \dots \underline{P}_k'\underline{\beta}$

then that A be the dispersion matrix of the estimators $\underline{P}_1'\hat{\underline{\beta}}, \underline{P}_2'\hat{\underline{\beta}}_2, \dots \underline{P}_k'\hat{\underline{\beta}}$

then we have the following optimum property of LS estimator.

(1) let $\underline{L}_1'Y, \underline{L}_2'Y, \dots \underline{L}_k'Y$ be any unbiased estimator of $\underline{P}_1'\beta_1, \underline{P}_2'\beta_2, \dots \underline{P}_k'\beta_k$ and let dispersion matrix of the estimator be B then B − A is non-negative definite implying

(i) trace $B \geq$ trace A
(ii) $|B| \geq |A|$
(iii) trace $QB \geq$ trace QA where Q is non-negative definite matrix
(iv) max latent root of $B \geq$ max latent root of A

Result Take the model as

$$E(Y_i) = \delta + \beta(x_i - \bar{x}), \quad i = 1, 2, \dots n$$
$$D(Y) = \sigma^2 I$$

obtain LS estimator of δ and β and show that $\operatorname{cov}(\hat{\delta}, \hat{\beta}) = 0$

$$E(\underline{Y}) = A\underline{\beta}$$
$$A\underline{\beta} = \delta + \beta(x_i - \bar{x})$$
$$S^2 = (\underline{Y} - A\underline{\beta})'(\underline{Y} - A\underline{\beta})$$
$$= \Sigma(Y_i - A\underline{\beta})^2$$

$$S^2 = \Sigma\left(y_i - \delta - \beta(x_i - \bar{x})\right)^2$$

$$\frac{\partial S^2}{\partial \delta} = 0 \Rightarrow \frac{\partial S^2}{\partial \delta} = 2\Sigma\left[y_i - \delta - \beta(x_i - \bar{x})\right](-1) = 0$$

$$\Sigma y_i - \beta\Sigma(x_i - \bar{x}) = n\delta$$

$$n\delta = \Sigma y_i$$

$$\hat{\delta} = \frac{1}{n}\Sigma y_i = \bar{y}$$

$$\frac{\partial S^2}{\partial \beta} = 0 \Rightarrow 2\Sigma\left[y_i - \delta - \beta(x_i - \bar{x})\right]\left(-(x_i - \bar{x})\right) = 0$$

$$\Sigma y_i - n\delta\ (x_i - \bar{x}) = \beta\Sigma(x_i - \bar{x})^2$$

$$\hat{\beta} = \frac{\Sigma y_i(x_i - \bar{x}) - \delta\,\Sigma(x_i - \bar{x})}{\Sigma(x_i - \bar{x})^2}$$

$$\hat{\beta} = \frac{\Sigma y_i(x_i - \bar{x})}{\Sigma(x_i - \bar{x})^2}$$

$$\operatorname{cov}(\hat{\delta}, \hat{\beta}) = \operatorname{cov}\left(\bar{y}, \frac{\Sigma y_i(x_i - \bar{x})}{\Sigma(x_i - \bar{x})^2}\right)$$

$$= \operatorname{cov}\left(\frac{\Sigma y_i}{n}, \frac{\Sigma y_i(x_i - \bar{x})}{\Sigma(x_i - \bar{x})^2}\right)$$

$$= \frac{1}{n\Sigma(x_i - \bar{x})^2}\operatorname{cov}\left(\Sigma y_i, \Sigma y_i(x_i - \bar{x})\right)$$

$$= \frac{1}{n\Sigma(x_i - \bar{x})^2}\Sigma\operatorname{cov}\left(y_i, y_i(x_i - \bar{x})\right)$$

$$= \frac{\sigma^2\,\Sigma(x_i - \bar{x})}{n\Sigma(x_i - \bar{x})^2}$$

$$= 0$$

① Let $Y_1, Y_2, \ldots Y_n$ be 'n' independent obs. from a population with mean μ and variance 'σ^2'. Obtain the BLUE of μ and an unbiased quadratic estimator for σ^2

$$E\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \mu = \underline{a}\mu \text{ (say)}$$

Normal eqn is $\quad \underline{a}'\underline{a}\,\mu = \underline{a}'\underline{Y}$

$$\hat{\underline{\mu}} = (\underline{a}'\underline{a})^{-1}\underline{a}'\underline{Y}$$

$$\hat{\underline{\mu}} = \frac{(1)}{n}(1\ 1 \ldots 1)\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$\hat{\underline{\mu}} = \frac{\Sigma Y_i}{n} = \bar{Y}$$

Blue of population mean is the Sample mean $\bar{Y}$ rank of estimation Space is 1, So the rank of Error Space is $(n-1)$

The S.S due to all BLUE's is

$$S_0^2 = \frac{(\underline{a}'\underline{Y})^2}{(\underline{a}'\underline{a})} = \frac{(\Sigma Y_i)^2}{n}$$

S.S due to error is

$$SSE = \underline{Y}'\underline{Y} - S_0^2$$

$$= \Sigma Y_i^2 - \frac{(\Sigma Y_i)^2}{n}$$

and has degree of freedom $(n-1)$

Thus $\dfrac{SSE}{(n-1)} = \dfrac{\Sigma(Y_i - \bar{Y})^2}{n-1}$ is an unbiased quadratic estimator of σ^2

The variance of the best estimator $\hat{\mu}$ is

$$var(\hat{\mu}) = var(\bar{Y}) = \frac{\sigma^2}{n}$$

Ex (2) Suppose a firm producing gas fielded electric bulbs is trying p methods of filling the bulbs with gas. n_i bulbs are made using the i^{th} method of filling $(i = 1, \dots p)$ then the appropriate linear model is this case is

$$E(Y_{ij}) = \mu_i$$
$$Var(Y_{ij}) = \sigma^2, \quad i = 1, 2, \dots p$$
$$j = 1, 2, \dots n_i$$

Obtain the BLUE of parametric function $\Sigma l_i \mu_i$ the S.S due to error and the dispersion matrix of $\hat{\mu}$

$Y_{ij} \rightarrow$ denotes the life in hrs of the j^{th} bulbs prepared by using i^{th} ~~used~~ method

$\mu_i \rightarrow$ population mean life in hrs for bulbs prepared by i^{th} method $Y's$ are assumed to be independent

Observation vector $\underline{Y}$ is

$$\underline{Y}' = (Y_{11}, Y_{12}, \dots Y_{1n}, Y_{21}, Y_{22}, \dots Y_{2n_2}, \dots Y_{p1}, Y_{p2} \dots Y_{pn_p})$$

Matrix of known coefficients is

$$A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{pmatrix} \quad A_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & \end{bmatrix}_{n_1 \times p} \quad A_2 = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 1 & \cdots & 0 \end{bmatrix}_{n_2 \times p}$$

$$A_p = \begin{bmatrix} 0 & \cdots & 1 \\ 0 & \cdots & 1 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{bmatrix}_{n_p \times p}$$

~~Ra~~ Since rank of A, rank (A) = p
Hence every linear function of parameters is linearly estimable
Unique solution of Normal eq^{n}s

$$\hat{\mu} = (A'A)^{-1} A'Y$$

Here $A'A = \text{diag}(n_1, n_2, \ldots n_p)$

$$(A'A)^{-1} = \text{dig}\left(\frac{1}{n_1}, \frac{1}{n_2}, \ldots \frac{1}{n_p}\right)$$

$$\hat{\mu}_i = \sum_{j=1}^{n_i} \frac{y_{ij}}{n_i} = y_{io}$$

= sample mean life (in hrs)
of bulbs produced by the ith method
$(i = 1, 2, \ldots p)$

BLUE of any parameter function $\sum l_i \mu_i$ is $\sum_i l_i y_{io}$

The rank of estimation space is p and the rank of error space is $(n-p)$

$$SSE = \sum_i \sum_j (y_{ij} - y_{io})^2 \text{ and has } (n-p) \text{ d.f}$$

Then an unbiased estimator of σ^2 is $\sum_i \sum_j \frac{(y_{ij} - y_{io})^2}{n-p}$

Dispersion metric of $\hat{\mu}$

$$D(\hat{\mu}) = (a'a)^{-1} \sigma^2$$

Thus $var(\hat{\mu}_i) = \frac{\sigma^2}{n}$, $cov(\hat{\mu}_i, \hat{\mu}_{j'}) = 0 \ \forall i \neq j'$

The above mode is one-way classification model

Ex(3) In an educational program 'p' different methods of teaching arithmetic to beginners where under study for this purpose 'q' experienced teachers wh were taken and each teacher tried out each method on 1 beginner

There where were pq beginners, so that each teacher tried each method on a separate individual.
We shall assume that we are concerned with the results of this 'q' teachers only then at the end of the program a test is given to pq learners and there scores are noted.
Let Y_{ij} → denote the score of the individual on whom j^{th} teacher tried i^{th} method
Y_{ij}'s are independent

Appropriate model in this case is

$$E(Y_{ij}) = \mu + \alpha_i + \beta_j$$
$$V(Y_{ij}) = \sigma^2 \quad , \quad i = 1, \ldots p, \quad j = 1, \ldots q$$

Here parameters $\mu, \alpha_1, \alpha_2, \ldots \alpha_p, \beta_1, \beta_2, \ldots \beta_q$
And because it is fixed effect $\sum_{i=1}^{p} \alpha_i = 0$, $\sum_{j=1}^{q} \beta_j = 0$

So we have case of linear estimation with linear restriction

$$\left.\begin{aligned} E(\underline{Y}) &= A\underline{\beta} \\ D(\underline{Y}) &= \sigma^2 I \\ P'\underline{\beta} &= \underline{L} \end{aligned}\right\}$$

Note :- The above model is two-way classified data with one obs. per cell

Ex 4 Consider three independent Random Variables Y_1, Y_2, Y_3 having common variance σ^2 and Expectations, $E(Y_1) = \mu_1 + \mu_3$

$E(Y_2) = \mu_1 + \mu_2$

$E(Y_3) = \mu_1 + \mu_3$

Determine the condition of estimability of the parametric function $\underline{l}'\underline{\mu} = l_1\mu_1 + l_2\mu_2 + l_3\mu_3$

$$E(\underline{Y}) = A\underline{\mu} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

The condition of estimability of $\underline{l}'\underline{\mu}$ is

$$\text{rank}(A) = \text{rank}\begin{pmatrix} A \\ \underline{l}' \end{pmatrix}$$

Since rank(A) = 2

with (1 0 1) & (1 1 0) as two indp row vectors of A,

Hence l' should depend on linearly on them.

$$\begin{pmatrix} A \\ l' \end{pmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ l_1' & l_2' & l_3' \end{bmatrix}$$

$$(l_1^* \; l_2^* \; l_3^*) = a(1\;0\;1) + b(1\;1\;0)$$

$l_1^* = a+b$

$l_2^* = b$ $\qquad \boxed{l_1 = l_2 + l_3}$

$l_3^* = a$

$\hat{y} = (A'A)^{-1} A'Y$

So $\hat{y}_1 = 0$

$\hat{y}_2 = Y_2$

$\hat{y}_3 = \frac{Y_1 + Y_3}{3}$

Some Concepts of Matrix

1 - norm

$||A||$ ← notation of norm of matrix A

Example ① $A = \begin{bmatrix} 1 & -7 \\ -2 & -3 \end{bmatrix}$

Absolute colum sums of A are

$1 + |-2| = 3$ and $|-7| + |-3| = 10$

$||A|| = 10$ (Norm of A is 10)

② $B = \begin{bmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{bmatrix}$

Absolute colum Sum of A are

$5 + |-1| + |-2| = 8$, $|-4| + |2| + |1| = 7$, $|2| + |3| + 0 = 5$

$||B|| = 8$ (Norm of A is 8)

③ Infinity Norm

$A = \begin{bmatrix} 1 & -7 \\ -2 & -3 \end{bmatrix}$, $||A||_\infty = 8$

$||B||_\infty = 11$

Euclidean Norm

$$||A||_E = \sqrt{\Sigma\Sigma(a_{ij})^2}$$

$$||A|| = \sqrt{(1)^2+(7)^2+(2)^2+(3)^2} = \sqrt{1+49+4+9} = \sqrt{63}$$
$$= 7.9$$

Projection

A projection is a linear transformation P from a vector space to itself such that $P^2 = P$ whenever P is applied twice to any value it gives the same result as if it were applied once

Projection Operator

let S_1 and S_2 be two distinct subspaces Such that

$$V = S_1 \oplus S_2 \qquad \otimes \rightarrow \text{kronecher product}$$

kronecher sum.

If A is $n\times n$ and B is $m \times m$ and I_k denotes $K\times K$ identity matrix

- kronecher sum is given by

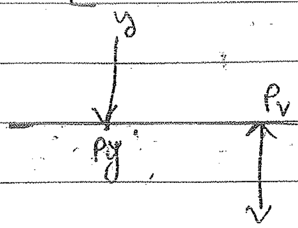
$$A \oplus B = A \otimes I_n + I_n \otimes B$$

Ex

$$A = \begin{bmatrix} 1 & -2 \\ -1 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 4 & -3 \\ 2 & 3 \end{bmatrix}$$

$$A \otimes B = \begin{bmatrix} 1\times 4 & 1\times -3 & -2\times 4 & -2\times -3 \\ 1\times 2 & 1\times 3 & -2\times 2 & -2\times 3 \\ -1\times 4 & -1\times -3 & 0\times 4 & 0\times -3 \\ -1\times 2 & -1\times 3 & 0\times 2 & 0\times 3 \end{bmatrix}$$

$$A \otimes B = \begin{bmatrix} 1 & -2 \\ -1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 2 & 3 \end{bmatrix}$$



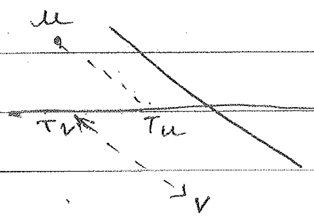
Orthogonal Projection

The function which maps the point (x, y, z) in three dimensional space R^3 to point $(x, y, 0)$ is a projection on $x \cdot y$ plan.
The function is represented by the matrix P

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow P\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

To see that P is indeed a projection it should satisfy satisfy $P^2 = P$
we compute

$$P^2\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = P\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$



Oblique

$$P = \begin{bmatrix} 0 & 0 \\ \alpha & 1 \end{bmatrix}$$

The projection P is orthogonal if and only if $\alpha = 0$

$$P^2 = \begin{bmatrix} 0 & 0 \\ \alpha & 1 \end{bmatrix} = 0$$

UNIT-2

Theue fundamental theorem of least Square theory

- Consider the independent variables

$Y_i \sim N(x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{im}\beta_m, \sigma^2)$, $i = 1, 2 .. n$ —(1)

where x_{ij}'s are known coeff & β_i are unknown para.

In matrix notation $\underline{Y}$ stands for colu^n vector of the variables y_i, $\underline{\beta}$ for the parameters β_i and $x = (x_{ij})$ for the matrix coefficient

$$\Sigma(y_i - x_i\beta_1 - x_{i2}\beta_2 - \dots - x_{im}\beta_m)^2$$
$$= (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$$

Hence pdf of $Y_1, Y_2, .. Y_n$ can be written as

$$C \cdot e^{-(\underline{Y} - X\underline{\beta})'(\underline{Y} - X\beta)/2\sigma^2} \quad -(2)$$

Thm (1) Let $R_0^2 = \min_{\beta} (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$

Then $R_0^2 \sim \chi^2_{(n-r)}$ where r is the rank of x

proof $(\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$ is minimum when $X\underline{\beta}$ is the projection of $\underline{Y}$ on $M(x)$ on manifold of x

(linear manifold is a vector space V is any subset of vector M closed under addition and scalar multiplication)

But projection of any vector of $M(x)$ is obtained through an operator a matrix P which is symmetric idempotent of rank$(x) = r$

The projection $\underline{\hat{Y}}$ on $M(x)$ equals $P\underline{Y}$ and therefore $(\underline{Y} - \underline{Y}P)$ is perpendicular

Hence $R_0^2 = (\underline{Y} - P\underline{Y})'(\underline{Y} - P\underline{Y})$

$= \underline{Y}'(I-P)(I-P)\underline{Y}$

$= \underline{Y}'(I-P)\underline{Y}$

Since $(I-P)$ is idempotent

$\therefore$ rank $(I-P)$ = trace $(I-P)$

= trace I − trace P

$= n - r$

The distn of quadratic form

$\underline{Y}'(I-P)\underline{Y} \sim \sigma^2 \chi^2_{(n-r, \lambda)}$ — (*)

where $\lambda = \dfrac{E[\underline{Y}'(I-P)\underline{Y}]}{\sigma^2}$

$\therefore \lambda\sigma^2 = E[\underline{Y}'(I-P)\underline{Y}]$

$= E(\underline{Y}')(I-P)E(\underline{Y})$

$= (X\underline{\beta})'(I-P)(X\underline{\beta})$

$= \underline{\beta}'X'(I-P)(X\underline{\beta})$

$= \underline{\beta}'(X'X - X'PX)\underline{\beta}$

Since $PX = X$ ($\because$ vectors of $M(X)$ ~~remains~~ remains unchanged by projection operator P)

$\lambda\sigma^2 = \underline{\beta}'(X'X - X'X)\underline{\beta} = 0$

Hence distn of (*) is central

Thm (2) Let H be a matrix of order $(m \times k)$ and rank (k) such that $M(H) \subset M(X')$ and

$R_1^2 = \min\limits_{\underline{\beta}} (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$

such that condition $H'\underline{\beta} = \xi$ (given) Then

Then Prove that.

(a) R_0^2 & $R_1^2 - R_0^2$ are independently distributed

(b) $R_0^2 \sim \sigma^2 \chi^2_{(n-r)}$ and $R_1^2 - R_0^2$ as a non-central χ^2 on K d.f

(c) If $H'\beta = \xi$ is true theory

$$R_1^2 - R_0^2 \sim \sigma^2 \chi^2_{(k)} \text{ and } \frac{R_1^2 - R_0^2}{K} \div \frac{R_0^2}{n-r} \sim F(k, n-r)$$

proof If $H'\beta = \xi$
then $\beta = \beta_0 + \gamma$ where β_0 is a particular soln. of $H'\beta = \xi$ and γ is a general soln of $H'\beta = 0$

Hence

$$\min_{H'\beta = \xi} (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$$

$$= \min_{H'\gamma = 0} (\underline{Y} - X\underline{\beta}_0 - X\gamma)'(\underline{Y} - X\beta_0 - X\gamma)$$

But $X\gamma$ with restriction $H'\gamma = 0$ is a Subspace $\mathcal{S} \subset M(X)$ with

$$d[\mathcal{S}] = \text{rank}(X' \vdots H) - \text{rank}(H) = S \text{ (say)}$$

irrespective of H satisfying $M(H) \subset M(X')$

Let P be the operator for projecting a vector on manifold of X and u be the operator for projecting on $\mathcal{S}$

Then $\text{rank}(P) = r$ and $\text{rank}(u) = S$

Now

$$R_1^2 = (\underline{Y} - X\beta_0)'(I - u)(\underline{Y} - X\underline{\beta}_0)$$

$$R_0^2 = \underline{Y}'(I - P)\underline{Y}$$

$$= (\underline{Y} - X\beta_0)'(I - P)(\underline{Y} - X\beta_0)$$

Introduction of factor vector $X\beta_0$ in the expression of R_0^2 does not change its value. Since $(I-U)$ is idempotent and rank $(I-U) = n-S$

$$\text{So } R_1^2 \sim \sigma^2 \cdot \chi^2(n-s, \lambda)$$

where

$$\sigma^2\lambda = E[(Y - X\beta_0)'](I-U)[E(Y - X\beta_0)]$$
$$= (X\beta - X\beta_0)'(I-U)(X\beta - X\beta_0)$$

Now $R_0^2 \sim \sigma^2 \cdot \chi^2(n-r)$ by thm (1) and

$$R_1^2 - R_0^2 \geq 0$$

Hence we have $R_1^2 - R_0^2 \sim \sigma^2 \chi^2_{(r-s,\lambda)}$ indep of R_0^2

If $H'\beta = \xi$ is true $\beta = \beta_0 + \gamma$ where $H'\gamma = 0$

Hence
$$\sigma^2\lambda = (X\beta - X\beta_0)'(I-U)(X\beta - X\beta_0)$$
$$= (X\beta_0 + X\gamma - X\beta_0)'(I-U)(X\beta_0 + X\gamma - X\beta_0)$$
$$= (X\gamma)'(I-U)(X\gamma)$$
$$= (X\gamma)' - UX\gamma$$
$$= 0 \qquad (\text{since } XU\gamma = X\gamma)$$

Thus $R_1^2 - R_0^2 \sim \sigma^2\chi^2(r-s)$ a central χ^2

In the special case

$M(H) \subset M(X')$ and H is of rank k

$$S = \text{rank}(X' \vdots H) - \text{rank}(H)$$
$$r-k \Rightarrow r-s = k$$

Hence $R_1^2 - R_0^2 \sim \sigma^2 \chi^2_{(k)}$

Hence $\dfrac{R_1^2 - R_0^2}{k} \div \dfrac{R_0^2}{n-r} \sim F(k, n-r)$

Thm 3 Let $\underline{Y}' = (\underline{Y}_1' \vdots \underline{Y}_2')$ with the corresponding partition of the expectation vector. Then

$\beta' X' = \beta'(X_1' \vdots X_2') = (\underline{\beta}' X_1' \vdots \underline{\beta}' X_2')$

the statistic $u = \dfrac{\min_{\beta} (\underline{Y}_1 - X_1 \underline{\beta}')(\underline{Y}_1 - X_1 \underline{\beta})}{\min_{\underline{\beta}} (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})}$

$$\sim \beta\left(\frac{q-r}{2}, \frac{n-q-r+r_1}{2}\right)$$

where q is the column in X_1

$r_1 = \text{rank}(X_1)$ and n & r are as defined in (1)

Proof Using the same logic as in (1) & (2)

$$u = \frac{\chi^2(q-r_1)}{\chi^2(q-r_1) + \chi^2(r_1 - r - q + r_1)}$$

and both are indep and ratio of 2 χ^2 is beta distribution

Testing of hypothesis & Interval estimation

The least square estimate of a parameter funtion is only a point estimate and no exact statement of probability of its deviation from the true value of parametric function can be made without a specific $dist^n$ for the variable Y being considred.

Let $Y \sim N(X\underline{\beta}, \sigma^2 I)$. Let H' be a matrix of order $k \times m$ of rank k such that its rows depend on rows of X. This implies that k parametric functions $H'\beta$ are individually estimable under the setup $(\underline{Y}, X\underline{\beta}, \sigma^2 I)$ in which was the least square estimable are

$$H'\underline{B} = Z.$$

In terms of X & Y $\quad H'\underline{\beta} = H'(X'X)^{-1}X'\underline{Y}$

and $R_0^2 = \text{error } SS = \underline{Y}'(I - X(X'X)^{-1}X') = 0$

Showing that $H'\hat{\beta}$ and R_0^2 are independent distributed. The $dist^n$ of Z is k-variate model between $H'\underline{\beta}$ and dispersion matrix $\sigma^2 D$ (Say) and that of R_0^2 is $\sigma^2 \cdot \chi^2(n-r)$

Single Parametric function

Given an estimable parametric function
$\theta = \underline{P'\beta}$
Let y be its L.S.E with variance $P^2\sigma^2$
Then $y \sim N(\theta, P^2, \sigma^2)$ and $R_0^2 \sim \sigma^2 - \chi^2(n-r)$ are indep so that

$$S^2 = \frac{R_0^2}{n-r} \quad \text{then} \quad t = \frac{\frac{u-\theta}{P\sigma}}{S/\sigma} = \frac{u-\theta}{PS} \sim S_{(n-r)} \quad \text{—(i)}$$

which is student's dist^n on $(n-r)$ d.f

If t_α is α-probability point of $|t|$

i.e $P(|t| > t_\alpha) = \alpha$ then

$$P\left[\frac{|u-\theta|}{PS} \leq t_\alpha\right] = 1-\alpha \quad \text{—(ii)}$$

$$P\left[-t_\alpha \leq \frac{|u-\theta|}{PS} \leq t_\alpha\right] = 1-\alpha$$

$$P[-pst_\alpha \leq u-\theta \leq pst_\alpha] = 1-\alpha$$

$$P[u - pst_\alpha < \theta \leq u + pst_\alpha] = 1-\alpha \quad \text{—(iii)}$$

<u>Testing of hypothesis</u>

let the null hypothesis be $H_0 : P'\beta = \theta_0$ (an assigned value)

If the null hypothesis is true then using (ii) we have

$$P\left\{\frac{|u-\theta_0|}{PS} \leq t_\alpha\right\} = 1-\alpha$$

The null hypothesis is thus rejected at α level of significance if for an observed u we have

$$\boxed{\frac{|u-\theta_0|}{Ps} > t_\alpha} \longrightarrow \text{(iv)}$$

Power of test is given by
$$P[|t| > t_{\alpha/2}, n-r \mid H_1]$$

→ t has non-central distⁿ with λ as non-central parameter the power of t is monotonically increasing function of λ

n = parameter

$$\lambda = \frac{\text{value of } P'\beta \text{ under } H_1 - \text{value of } P'\beta \text{ under } H_0}{\sqrt{\text{var of best estimator of } \underline{P}'\beta}}$$

Interval Estimation

→ Eq (3) implies that for an observed u the ~~the~~ interval $(u - ps t_{\alpha_1}, u + ps t_{\alpha_1})$

$\beta(1-\alpha)$ confidence interval of θ

Note than one param. function

→ Consider k indept estimable linear parametric funtion
$\theta_1 = \underline{H}_1'\beta$; $\theta_2 = \underline{H}_2'\underline{\beta}$, ..., $\theta_k = \underline{H}_k'\underline{\beta}$

In matrix notation it maybe written as $\underline{\theta} = H'\underline{\beta}$ where H' is $k \times m$ matrix hand θ is a column vector of $\theta_1, \theta_2, .. \theta_k$

→ The LSE of $\theta_1, \theta_2, ..., \theta_k$ are represented by $(z_1, z_2, .. z_k) = z'$ and its dispersion matrix is $\sigma^2 D$

Since $E(z) = \theta$, $D(z) = \sigma^2 D$ we have
$$(z-\theta)' D (z-\theta) \sim \sigma^2 \chi^2_{(k)}$$

$R_0^2 \sim \sigma^2 . \chi^2_{(n-r)}$ and are indept

Hence $F = \dfrac{\dfrac{(z-\theta)' D^{-1} (z-\theta)}{K}}{\dfrac{R_0^2}{n-r}} \sim F(k, n-r) \quad -(v)$

* Tests of Multiple Hypothesis

Let us test the null hypothesis that K parametric f^n have assigned the values

$$H_0 = H_1' \beta = \theta_{10}, \; H_2' \beta = \theta_{20}, \ldots H_k' \beta = \theta_{k0}$$

which maybe written in matrix notation as

$$H' \beta = \theta_0$$

$(z-\theta_0)$ the vector of deviations between least square estimators and assigned values

If infact θ_0 is not true the deviations $(z-\theta_0)$ are likely to be large.

Then let us consider the compound deviation a single measure of deviation

$$(z-\theta_0)' D^{-1} (z-\theta_0)$$

Now $\dfrac{E(z-\theta_0)' D^{-1} (z-\theta_0)}{K}$

| $E(Y'AY)$ $= \text{trace } AV + E(Y')AE(Y)$ where $V =$ var-cov matrix of Y

$$= \frac{\sigma^2}{K} \text{trace}(D^{-1} D) + \frac{1}{K} E(z-\theta_0)' D^{-1} E(z-\theta_0)$$

$$= \sigma^2 + \frac{1}{K} \left[(H'\beta - \theta_0)' D^{-1} (H'\beta - \theta_0) \right]$$

$= \sigma^2$ if null hypothesis is true i.e $H'\beta = \theta_0$

$> \sigma^2$ if null hypothesis is not true

Since $(H'\underline{\beta} - \underline{\theta}_0)' D^{-1} (H'\underline{\beta} - \underline{\theta}_0)$ is positive definite

Also in the ratio $F = \frac{(\underline{z} - \underline{\theta}_0)' D^{-1} (\underline{z} - \underline{\theta}_0)}{K} \div \frac{R_0^2}{n-r}$

The numerator is on an average larger in magnitude than denominator if null hypothesis is not true

Thus the larger value of F indicates departure from null hypothesis choosing a value F_α such that

$$P(F > F_\alpha) = \alpha \quad \text{where } F \sim F(k, n-r)$$

Reject Ho at α-significance level if

$$F = \frac{(\underline{z} - \underline{\theta}_0)' D^{-1} (\underline{z} - \underline{\theta}_0)}{KS^2} > F_\alpha$$

ANOVA Table

Source	d.f	S.S	MSS
Deviation from hypo	k	$R_1^2 - R_0^2$	$\frac{(R_1^2 - R_0^2)}{K}$
Residual	$n-r$	$\min_{\underline{\beta}} (Y - X\underline{\beta})'(Y - X\underline{\beta}) = R_0^2$	$\frac{R_0^2}{n-r}$
Total	$n-r+k$	$\min_{H'\beta = \theta_0} (Y - X\underline{\beta})'(Y - X\underline{\beta}) = R_1^2$	

Simultaneous Confidence Intervals

~~If~~ It follows that $P\left\{\frac{(z-\theta)' D^{-1}(z-\theta)}{ks^2} \le F_\alpha\right\} = 1-\alpha$ — (*)

where

$\theta = H'\beta$ stands for indep estimable parametric function of β. The above expression (*) is called $(1-\alpha)$ confidence region of $\theta = H'\beta$ let it be represented by 'C' then simultaneous ~~$\theta = H'\beta$ stands for indept estimable parametric function of β. The above expression (*)~~

Confidence interval for functions $g_i(\underline{\theta})$, $i = 1, 2, \ldots k$ with confidence coefficient possibly greater than $(1-\alpha)$ are given by

$$\left[\underset{\theta \in C}{\text{Min}}\; g_i(\underline{\theta}),\ \underset{\underline{\theta}' \in \underline{C}}{\text{Max}}\; g_i(\underline{\theta})\right] \quad p = 1, 2, \ldots k$$

for any ~~parametric~~ particular $g(\underline{\theta})$

$$P\left[\underset{\theta \in C}{\text{Min}}\; g(\underline{\theta}),\ \underset{\theta \in \underline{C}}{\text{Max}}\; g_i(\underline{\theta})\right]$$

So that confidence co-efficient for any particular f^n is greater than $1-\alpha$. $F(\alpha)$ is upper (α) prob value of F-statistic.

Second Method (using Cauchy-Schwartz inequality

If Y and $\underline{A}$ are column vector and B is positive definite matrix than

$$A'B^{-1}A = \frac{\underset{y}{\text{Max}}\,(\underline{y}'\underline{A})^2}{\underline{y}'B\underline{y}}$$

let $A = \underline{z} - \underline{\theta}$ $B = D$ then we have

$$\frac{(\underline{z}-\underline{\theta})' D^{-1} (\underline{z}-\underline{\theta})}{KS} = \frac{1}{KS^2} \underset{\underline{y}}{\operatorname{Max}} \left[\frac{(\underline{y}'(\underline{z}-\underline{\theta}))^2}{\underline{y}' D \underline{y}} \right]$$

then using * we have

$$P\left\{ \underset{\underline{y}}{\operatorname{Max}} \frac{|\underline{y}'(\underline{z}-\underline{\theta})|}{\sqrt{\underline{y}' D \underline{y}}} \leq S\sqrt{KF_\alpha} \right\} = 1-\alpha$$

i.e $P\left[|\underline{y}'(\underline{z}-\underline{\theta})| \leq S\sqrt{KF_\alpha \underline{y}' D \underline{y}} \right]$

$= 1-\alpha \quad \forall \underline{y}$

$$P\left[\underline{y}'\underline{\theta} \in \underline{y}'\underline{z} \pm S\sqrt{KF_\alpha \underline{y}' D \underline{y}} \right] = 1-\alpha \quad \forall \underline{y}$$

The above eq^n provides simultaneously conf intervals for all f^n of the form $\underline{y}'\underline{\theta} = \underline{y}' H \underline{\beta}$ ~~for~~ all linear combination of the given parametric f^n $H'\underline{\beta}$

PWR → R-Package → for sample size

Power of F-Test :-

OC curve is a plot of type - II error proba. of a statistical test for a particular sample size vs a parametric that reflects the ~~exact~~ extent to which the null hypothesis is false. This curves can be used to guide the experimenter in selecting the no. of replicates so that the design will be sensitive to important potential difference in the treatments.

we consider the prob. of type II error of fixed effect model for the case of equal sample sizes per treatment

$$\beta = 1 - P\{\text{Reject } H_0 \mid H_0 \text{ is false}\}$$
$$= 1 - P\{F_0 > F_{\alpha, a-1, N-a} \mid H_0 \text{ is false}\} \quad — (*)$$

a = no. of treatment
N = no. of obs.

To evaluate the prob we need to know the distn of test statistic F_0 if null hypothesis is false
If H_0 is false the statistic

$$F_0 = \frac{MSS_{treat}}{MSS_E} \sim \text{Non-central } F \text{ with } (a-1, N-a) \text{ df}$$

and noncentrality parameter δ. If $\delta = 0$ then n.c.f becomes central f-distribution

OC curves are used to prob statements in (*) this curves plot the prob (type-II error β) against a parameter

$$\Phi = \frac{n \sum_{i=1}^{a} \tau_i^2}{a\sigma^2}$$

Φ^2 is related to non-centrality parameter δ

Curves are available for $\alpha = 0.05$ and $\alpha = 0.01$ & a range of degree of freedom for numerator and denominator.
One way to determine Φ to choose the actual values of the treatment means for which we would like to reject null hypo with high probability. Thus, if $\mu_1, \mu_2 \ldots \mu_x$ are specified treatment means we find τ_i as

$$\tau_i = \mu_i - \bar{\mu} \quad \text{where } \bar{\mu} = \frac{1}{a}\sum_{i=1}^{a} \mu_i$$

we also required an estimate of σ^2.

sometimes this is available from prior experience or a preliminary test. when we are uncertain about the value of σ^2. sample sizes is could been determine form a range of likely values of σ^2. to study the effect of this parameter of on the required sample size before the final choice is made. An alternate approach is to select a sample size such that if the difference between any two treatments means exceeds a specified value the null hypothesis should be rejected if the difference between any two treatment means is as large as D. The minimum value of

$$\Phi^2 = \frac{nD^2}{2a\sigma^2}$$ because this is a minimum of Φ^2 the corresponding sample size obtained from OC curve is a conservative value i-e it provides a power atleast as great as specified by the experiment

Post hoc Test

(1) Turkey's test

Suppose that following an analysis of variance in which we have rejected the null hypotheses of equal treatment mean we wish to test all pairwise mean comparisons.

$H_0: \mu_i = \mu_j$

$H_1: \mu_i \neq \mu_j \quad \forall\, i \neq j$

Tukey proposed a procedure for testing hypothesis for which overall significance level is exactly α when sample sizes are equal and at the at most α when sample size are unequal. This procedure can also be used to construct confi. intervals on the differences in all pairs of means, for this intervals the simultaneous confidence level is $100(1-\alpha)\%$ when the sample sizes are equal and at most $100(1-\alpha)\%$ when sample sizes are unequal

Tukey procedure make use of studentized range & statistics

$$q = \frac{\bar{Y}_{max} - \bar{Y}_{min}}{\sqrt{MSSE/n}}$$

where $\bar{Y}_{max}$ and $\bar{Y}_{min}$ are the largest and the smallest sample means respectively. Out of a group of 'p' sample means $q_\alpha(p,f)$ is the upper α percentage points of q and f is no. of degrees of freedom associated with MSSE.
For equal sample sizes tukey's test declares two means significantly different if the absolute value of the sample differences exceeds T_α i.e $\bar{Y}_{i.} - \bar{Y}_{j.} > T_\alpha$

and

$$T_\alpha = q_\alpha(p,f)\sqrt{\frac{MSSE}{n}}$$

we could construct $100(1-\alpha)$ percent confi. int. for all pairs of means as follows.

$$\bar{y}_{i.} - \bar{y}_{j.} - q_\alpha(p,f)\sqrt{\frac{MSSE}{n}} \leq \mu_i - \mu_j \leq \bar{y}_{i.} - \bar{y}_{j.} + q_\alpha(p,f)\sqrt{\frac{MSSE}{n}}$$

$i \neq j$

when sample sizes are not equal then

$$T_\alpha = \frac{q_\alpha(p,f)}{\sqrt{2}}\sqrt{MSSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

and confidence interval is given by.

$$\bar{y}_{i.} - \bar{y}_{j.} - \frac{q_\alpha(p,f)}{\sqrt{2}}\sqrt{MSSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \leq \mu_i - \mu_j \leq$$
$$\bar{y}_{i.} - \bar{y}_{j.} + \frac{q_\alpha(p,f)}{\sqrt{2}}\sqrt{MSSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

The unequal sample size version is some times called Tukey-kramer

★ Scheffe's Method for comparing all contrasts

~~If~~ In many situation experimentor may not know in advance which contrast they wish to compare or they may be interested in more than (a-1) possible comparisions. In many exploratory experiments the comparision of interest are discovered only after preliminary examination of the data. ~~She~~ Scheffe has proposed a method for comparing any and all possible contrast between treatment means.

In scheffe's method type I error is atmost α for any of the possible comparisons. Suppose that 'm' contrast in the treatment means

$$\gamma_u = C_{1u}\mu_1 + C_{2u}\mu_2 + \dots + C_{au}\mu_a \quad , u=1,2..m$$

of interest have been determine. the corresponding contrast in the treatment averages $\bar{Y}_{i.}$ is

$$C_u = C_{1u}\bar{y}_{1.} + C_{2u}\bar{y}_{2.} + \dots + C_{au}\bar{y}_{a.} , \quad u=1,2..m$$

and the S.E of this contrast is

$$S.E_u = \sqrt{MSSE \sum_{i=1}^{a} \left(\frac{C_{iu}^2}{n_i}\right)}$$

where n_i is the no of obs in ith treatment the critical value against which C_u can be compared is

$$S_{\alpha,u} = S_{cu}\sqrt{(a-1)F_{\alpha,a-1,N-a}}$$

If $|C_u| > S_{\alpha,u}$
the hypo that the contrast γ_u equals zero is rejected.
The scheffe procedure can also be used to contrast among treatment means. The resulting interval

$$C_u - S_{\alpha,u} \leq \Gamma_u \leq C_u + S_{\alpha,u}$$

are simultaneous confidence interval. In that probability is atmost α that all of them simultaneously true

Ex A development engineer is interested in determining if the cotton weight percentage in a synthetic fibre affects the tensile strength for this a CRD with 5 levels of cotton weight percentage and 5 replicates is run for

Weight of % of cotton	Tensile Strength (lb/in^2) 1	2	3	4	5
15	7	7	15	11	9
20	12	17	12	8	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

Tukey's test

$\alpha = 5\%$ $\quad \bar{y}_{1.} = 9.8 \quad \bar{y}_{3.} = 17.6 \quad \bar{y}_{5.} = 10.8$

$\bar{y}_{2.} = 15.4 \quad \bar{y}_{4.} = 21.6$

$\bar{y}_{1.} - \bar{y}_{2.} = -5.6 \qquad \bar{y}_{2.} - \bar{y}_{4.} = -6.2$

$\bar{y}_{1.} - \bar{y}_{3.} = -7.8 \qquad \bar{y}_{2.} - \bar{y}_{5.} = 4.6$

$\bar{y}_{1.} - \bar{y}_{4.} = -11.8 \qquad \bar{y}_{3.} - \bar{y}_{4.} = -4$

$\bar{y}_{1.} - \bar{y}_{5.} = -1 \qquad \bar{y}_{3.} - \bar{y}_{5.} = 6.8$

$\bar{y}_{2.} - \bar{y}_{3.} = -2.2 \qquad \bar{y}_{4.} - \bar{y}_{5.} = 10.8$

$\cancel{T_{tab}} = T_{tab} = q_{(.05)(5,20)}\sqrt{\frac{MSSE}{n}}$

$= 4.23\sqrt{\frac{8.06}{5}} = 5.37$

Hypothesis :- $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Rejection criteria :- $|\bar{y}_{1.} - \bar{y}_{2.}| > T_{tab}$ Reject H_0

suppose that contrast of interest is

$\gamma_1 = \mu_1 + \mu_3 - \mu_4 + \mu_5$

$C_1 = \bar{y}_{1.} + \bar{y}_{3.} - \bar{y}_{4.} - \bar{y}_{5.} = -5$

$S_{C_1} = \sqrt{MSSE \sum_{i=1}^{5}\left(\frac{C_i^2}{n_i}\right)} = \sqrt{8.06\left(\frac{1+1+1+1}{5}\right)}$

$= 2.54$

critical values are

$S_{0.05} = S_{C_1}\sqrt{(a-1)F_{(0.05, a-1, N-a)}} = 2.54\sqrt{(4)F_{(0.05,4,20)}}$
$= 8.606$

$H_0: \mu_1 + \mu_3 - \mu_4 - \mu_5 = 0$
$H_1: \mu_1 + \mu_3 - \mu_4 - \mu_5 \neq 0$

Here $|C_1| = 5 < 8.606$ Hence we donot reject H_0

Conclusion: Joint effect of μ_1 & μ_3 are same as joint effect of μ_4 and μ_5

Note: the sampling distribution for 't' assumes only one 't' test from any given sample. hence there is a substantial increase in α if multiple test are performed the estimated value of α is $1-(1-\alpha)^c$ where c is total no. of comparision or contrast or test when α is 0.05 then $1-(1-0.05)^{10}$ is 0.4

Random Effect Model (3rd Unit)

If the effect in the linear model are all random variable ~~are~~ except for additive constant which is a fixed quantity is called a random effect Model, effects become random due the sampling of the levels of the factors included.

It is also called variance component model thus we have variance component model for one-way layout, two-way layout or multiway layout. The complete p-way classification or layout is called balanced if the number of observations in different cells are equal.

the one-way classification is balanced if the number of observations under the categories are same

General random effect model in the balanced case

let an observable random variable $y_{ijk\ldots m}$ for a balanced case be such that

$$y_{ijk\ldots m} = \mu + a_i + b_{ij} + c_{ijk} + \ldots + e_{ijk\ldots m} \quad -(1)$$

where μ is a constant, the r.v's $a_i, b_{ij}, c_{ijk} \ldots e_{ijk\ldots m}$ are completely independent

$a_i \sim N(0, \sigma_a^2)$

$b_{ij} \sim N(0, \sigma_b^2)$

$c_{ijk} \sim N(0, \sigma_c^2)$ —— $e_{ijk\ldots m} \sim N(0, \sigma_e^2)$

let the dispersion matrix of y_{ijk-m} be +ve definite when the elements of eq (1) satisfy the condition that follow it the relationship (1) is called variance component model or random effect model.

We observe that

$E(y_{ijk-m}) = \mu$ and y_{ijk-m} are not indep.

The model is thus different from corresponding fixed effect model where y_{ijk-m} do not have same expectations and y_{ijk-m} are independent

Also under (i)

$V(y_{ijk-m}) = \sigma_a^2 + \sigma_b^2 + \ldots + \sigma_e^2$

So, the diagonal elements of the dispersion matrix of Y are all same. $\sigma_a^2, \sigma_b^2, \ldots \sigma_e^2$ are the components of the variance of the obs so they are called variance component and Hence the model is called variance component model

ANOVA (table 1)

Source	d.f	SS	MSS	E(MSS)
1	f_1	S_1	s_1^2	σ_1^2
2	f_2	S_2	s_2^2	σ_2^2
3	f_3	S_3	s_3^2	σ_3^2
⋮	⋮	⋮	⋮	⋮
k	f_k	S_k	s_k^2	σ_k^2

generally, each σ_i^2 will be a linear function of the variance components $\sigma_a^2, \sigma_b^2, \ldots, \sigma_e^2$

Thm (1) Let the model be given by eq (1) and it is balanced an appropriate ANOVA table is given by table (1) then $\frac{S_1}{\sigma_1^2}, \frac{S_2}{\sigma_2^2}, \ldots, \frac{S_k}{\sigma_k^2}$ are mutually independently distributed as central χ^2 - with $f_1, f_2, \ldots f_k$ d.f resp. Since σ_i^2 are linear funⁿs of the variance components we can obtain unbiased estimators of the variance component by equating of MSS to the corresponding MSS of the table (1). This estimators have following properties.

Thm (2) Let the model be (1) and be balanced then the estimators of the variance components obtain by equating each E(MSS) to the corresponding MSS of table (1) are minimum variance unbiased. It may so happen that the estimate of variance component obtained is negative. in such a case the estimate is taken to be zero

Thm (3) let the model be (1) and balanced with the corresponding ANOVA given in table (1) then the test of $H_0: \sigma_i^2 = \sigma_j^2 \quad (i \neq j)$

$H_1: \sigma_i^2 > \sigma_j^2$

is given by s_i^2/s_j^2 which is distributed as $\frac{\sigma_i^2}{\sigma_j^2}F$ and where F has central F distribution with (f_i, f_j) d.f

Power

Power of test Ho depends on $H_1: \frac{\sigma_i^2}{\sigma_j^2} = \lambda < 1$ and which is given by

$$\beta(\lambda) = P\left[\frac{s_i^2}{s_j^2} > F(\alpha, f_i, f_j) \middle| H_1\right]$$

$$= P\left[\frac{s_i^2 \lambda}{s_j^2} > \lambda F(\alpha, f_i, f_j) \middle| H_1\right]$$

$$= \int_{\lambda F(\alpha, f_i, f_j)}^{\infty} f(u)\, du$$

where f is pd.f of central F with d.f (f_i, f_j). Thus in the variance component model both test and power are given by central F distⁿ where as in fixed effect model power is given by non-central F distⁿ

⇒ One-way classification random effect model
Consider the following balance one-way classification random effect model

$$Y_{ij} = \mu + a_i + e_{ij} \qquad i = 1, 2, \ldots p$$
$$a_i \sim N(0, \sigma_a^2) \qquad j = 1, 2, \ldots n$$
$$e_{ij} \sim N(0, \sigma_e^2)$$

Here a_i is the random effect due to ith classification and e_{ij} are errors.

To test $H_0: \sigma_a^2 = 0$ ag $H_1: \sigma_a^2 > 0$, which means testing the homogenity of the effects of the classification.

We have $\sum_i \sum_j (y_{ij} - y_{00})^2 = \sum\sum (y_{ij} - y_{i0} + y_{i0} - y_{00})^2$

$$= \sum_i \sum_j (y_{ij} - y_{i0})^2 + \sum_i \sum_j (y_{i0} - y_{00})^2$$

$$= n \sum_i (y_{i0} - y_{00})^2 + \sum_i \sum_j (y_{ij} - y_{i0})^2$$

$$= S.S\ (\text{bet}^n \text{ classes}) + S.S\ (\text{with clas})$$

$$= SSB + SSE$$

Also

$$y_{i0} - y_{00} = \mu + a_i + e_{i0} - \mu - a_0 - e_{00}$$
$$= a_i + e_{i0} - a_0 - e_{00}$$

where $a_0 = \frac{\sum a_i}{p}$, $e_{i0} = \frac{\sum_j e_{ij}}{n}$, $e_{00} = \frac{\sum e_{i0}}{p}$

Thus $SSB = n \sum (a_i + e_{i0} - a_0 + e_{00})^2$

where

$a_i + e_{i0}$ are independently distributed as

$$N\left(0, \sigma_a^2 + \frac{\sigma_e^2}{n}\right)$$

$$\boxed{\begin{aligned} var(a_i + e_{i0}) &= var(a_i) + var(\sum e_{ij}/n) \\ &= \sigma_a^2 + \frac{1}{n^2} \sum \sigma_e^2 \\ &= \sigma_a^2 + \frac{\sigma_e^2}{n} \end{aligned}}$$

$$\frac{n \sum_{i=1}^{p} (a_i + b_{i0} - a_0 - b_{00})^2}{\sigma_e^2 + n\sigma_a^2} \sim \chi^2_{(p-1)} \text{ df}$$

Hence SSB is $(\sigma_e^2 + n\sigma_a^2) \cdot \chi^2_{(p-1)}$

$$E(MSSB) = E\left[\frac{SSB}{p-1}\right] = \sigma_e^2 + n\sigma_a^2$$

Similarly $SSE = \sum_i \sum_j (y_{ij} - y_{io})^2$

$$= \sum_i \sum_j (e_{ij} - e_{io})^2$$

and e_{ij} are independently distributed as $N(0, \sigma_e^2)$

$\frac{\Sigma(e_{ij} - e_{io})^2}{\sigma_e^2}$ are independently distributed as $\chi^2(n-1)$

Hence $\frac{SSE}{\sigma_e^2} = \frac{\sum_i \sum_j (e_{ij} - e_{io})^2}{\sigma_e^2} \sim \chi^2 p(n-1)$

To show that SSB & SSE are independently distributed we first show that

$U_i = a_i + e_{io} - a_o - e_{oo}$ and
$V_{ij} = e_{ij}' - e_{io}'$ have zero co-variance

$$Cov(U_i, V_{i'j}) = E(U_i, V_{i'j}) - E(U_i)E(V_{ij})$$

$E(U_i) = a_i - a_o$
$E(V_{i'j}) = 0$

Hence $cov(U_i, V_{i'j}) = E(U_i, V_{i'j})$

$$= E\{(a_i - a_o) + (e_{io} - e_{oo})\}\{e_{ij} - e_{i'o}\}$$

$$= E\{(a_i - a_o)(e_{ij} - e_{i'o})\} + E\{(e_{io} - e_{oo})(e_{i'j} - e_{i'o})\}$$
$$= 0 + E\{(e_{io} - e_{oo})(e_{i'j} - e_{i'o}\}\} \quad (\because a_i \,\&\, e_{ij} \text{ are indp})$$
$$= E(e_{io}e_{ij}) - E(e_{io}e_{i'o}) - E(e_{oo}e_{ij}) + E(e_{oo}e_{i'o})$$

$$= \delta\delta_{ii'} \frac{\sigma_e^2}{n} - \delta_{ii'} \frac{\sigma_e^2}{n} - \frac{\sigma_e^2}{np} + \frac{\sigma_e^2}{np}$$

$var(e_{ij})$
$= E(e_{ij}^2) - [E(e_{ij})]^2$

$= 0$

This implies U_i & V_{ij} are independent & they are normally distributed with zero mean;

Thus SSB & SSE being functions of independent quantities are also independent

Thus $\frac{SSE}{\sigma_e^2} \sim \chi^2_{p(n-1)}$ & $\frac{SSB}{(\sigma_e^2 + n\sigma_a^2)} \sim \chi^2_{(p-1)}$

and independently distributed

ANOVA

Source	df	S.S	MSS	E(MSS)
Bet classes	$p-1$	SSB	MSSB	$\sigma_e^2 + n\sigma_a^2$
within classes	$p(n-1)$	SSE	MSSE	σ_e^2
Total	$np-1$	Total SS	-	

The test of $H_0: \sigma_a^2 = 0$ is given by

$F_{cal} = \frac{MSSB}{MSSE}$ $\quad F_{tab} = F_{(p-1,\, p(n-1))}$

we reject H_0 at level α if $F_{cal} > F_{tab}$

The power of test against

$H_1: \lambda = \frac{\sigma_e^2}{(\sigma_e^2 + n\sigma_a^2)} < 1$ is given by

$$\beta(\lambda) = \int_{\lambda F(\alpha, p-1, p(n-1))}^{\infty} f(v)\, dv$$

where f is p.d.f of central F distⁿ above ANOVA table we have

$$\hat{\sigma}_e^2 = MSSE \quad , \quad \hat{\sigma}_\alpha = \frac{MSSB}{n} \cdot MSSE$$

Minimum norm quadratic unbiased estimator (MINQUE)

We consider the general linear model $Y = X\beta + e$

$$E = U_1 \xi_1 + U_2 \xi_2 + \dots + U_k \xi_k \quad -(1)$$

ξ_i being a n_i-vector of hypothetical (unobservable) variables such that

$$E(\xi_i) = 0 \quad D(\xi_i) = \sigma_i^2 I \quad i = 1, 2, \dots k$$
$$E(\xi_i, \xi_j) = 0 \quad i \neq j \quad -(2)$$

So that $D(E) = \sigma_1^2 U_1 U_1' + \sigma_2^2 U_2 U_2' + \dots + \sigma_k^2 U_k U_k'$
$= \sigma_1^2 V_1 + \sigma_2^2 V_2 + \dots, \sigma_k^2 V_k \quad -(3)$

The design matrices X, U_i (or V_i) and the obs
Obs vector Y are known β and $\sigma = (\sigma_1^2, \dots, \sigma_k)$ are unknown parameters under estimator the variance $\sigma_1^2, \sigma_2^2, \dots \sigma_k^2$ of the hypothestical variables in $\xi_1, \xi_2, \dots \xi_k$ are called variance component. Let $\alpha_1, \alpha_2, \dots, \alpha_k$ be approx, or approximate values of $\sigma_1, \sigma_2, \dots, \sigma_k$. If no aprior information is available then each α_i may be taken unity. We can rewrite eqn (1) and (3) as

$$e = w_1 \eta_1 + w_2 \eta_2 + \dots + w_k \eta_k$$

$w_i = \alpha_i u_i$
$\xi_i = \alpha_i \eta_i$

$$D(E) = \gamma_1^2 T_1 + \gamma_2^2 T_2 + \dots \gamma_k^2 T_k^2$$

$$T_i = \omega_i \omega_i' = \alpha_i^2 V_i$$

where $\gamma_i^2 = \sigma_i^2 / \alpha_i^2$ the scaled variance component to be estimated

<u>MINQUE Theory</u>

Let us consider the estimators of a linear func
$$\Sigma p_i \sigma_i^2 = \Sigma q_i V_i^2 \quad \text{—(i)}$$

of the variance component of a quadratic f^n $\underline{Y'AY}$ st the following conditions.

(a) <u>Invariance & Unbiasedness</u>

If instead of $\underline{\beta}$ we considered;

$\beta d = \underline{\beta} - \beta_0$ then the model becomes
$$\underline{Y}_d = X \underline{\beta}_d + \varepsilon$$
$$= X(\underline{\beta} - \beta_0) + \varepsilon$$
$$= X\underline{\beta} - X\beta_0 + \varepsilon$$

$$\Rightarrow Y_d = \underline{Y} - X\beta_0$$

In such a case the estimator is $\underline{Y}_d' A \underline{Y}_d$ which is same as $\underline{Y'AY}$ for all β_0 iff $\underline{AX = 0}$

$$(Y_d' A Y_d) = (\underline{Y} - X\beta_0)' A (\underline{Y} - X\beta_0)$$
$$= Y'A\underline{Y} - \underline{Y}'AX\beta_0 - \beta_0'X'A\underline{Y} + \beta_0'X'AX\beta_0$$
$$= \underline{Y'AY} \quad \text{if } (AX = 0)$$

This is called condition of invariance for translation in vector (β) parameter

Not for such an Estimator

$$E(\underline{Y}'A\underline{Y}) = Tr\,A\,D(\varepsilon)$$
$$= \sum \gamma_i^2 \quad Tr\,A\,T_i$$
$$= \sum \gamma_i^2 q_i$$

$$\Rightarrow Tr\,A\,T_i = q_i \quad i = 1, 2, \ldots \quad -(ii)$$

Which are the conditions for unbiasedness under invariance

(b) Minimum Norm :-

If the hypothetical variables η_i were known then the natural estimator of γ_i^2 and $\eta_i'\eta_i/n_i$ and Hence that of $\sum q_i \gamma_i^2$ is

$$\left(\frac{q_1}{n_1}\right)(\eta_1'\eta_1) + \cdots + \left(\frac{q_k}{n_k}\right)(\eta_k'\eta_k) = \eta' A \eta \quad (iii)$$

where A is a suitably defined diagonal matrix and $\eta' = (\eta_1', \ldots \eta_k')$ but the proposed estimator is $\underline{Y}'A\underline{Y} = (X\underline{\beta} + \varepsilon)'A(X\underline{\beta} + \varepsilon)$

$$= \underline{\varepsilon}'A\underline{\varepsilon} \quad (\because AX = 0)$$

$$= \eta' \omega' A \omega \eta \quad -(iv)$$

where $\omega = (\omega_1 \,\vdots \ldots \vdots\, \omega_k)$ The difference between (iii) & (iv) is $\eta'(\omega' A \omega - A)\eta$ which can be made small in some sence by minimizing $\| \omega' A \omega - A \| \quad -(v)$

where norm is suitably choosen that estimator $Y'AY$ where A is such that

$\|W'AW - A\|$ is minimum subject to $AX=0$

and $Tr\, A\, T_i = q_i \quad i = 1, 2, 3, \ldots$ is called MINQUE

→ RESIDUAL ANALYSIS :-

2019

The major assumption that are made in the study of regression analysis are as follows

(1) The relationship between independent and dependent variables is linear atleast approximately

(2) The error term ε has zero mean

(3) the error term ε has constant variance (iii) σ^2

(4) The errors are uncorrelated

(5) Errors are normally distributed (why normal → Because to do parametric test like F-test)

Assumption (4) & (5) imply that errors are indept R.V.

Assumption (5) is required for hypothesis testing and interval estimation

Dreams violation of the assumption may give an unstable model in the sense that a different sample will give a totally different model with opposite conclusions

We usually cannot detect departure from the underlined assumptions by examination of the standard summary statistics such as t or F or R^2 these are global model properties and as such they do not ensure mode adequacy.

Residuals

We define residual as $e_i = y_i - \hat{y}_i$; $i = 1, 2, \ldots n$
where y_i is an observed value
$\hat{y}_i$ is fitted value

Since a residual is a deviation between the data and the fitted value it is also a measure of variability in the response variable not explained by the regression model. Thus, any departure from the assumption on the errors should show up in the residuals

2019 Properties of Residual

(i) They have zero means $\bar{e} = \frac{\Sigma e_i}{n} = \frac{0}{n} = 0$

(ii) Approximate average variance is estimated by $\frac{\Sigma (e_i - \bar{e})^2}{n-p} = \frac{\Sigma e_i^2}{n-p} = \frac{SSE}{n-p} = MSSE$

(iii) Residuals are not independent however as the n residuals have only $n-p$ degree of freedom associate with them. The non-independence of Residuals has little effect on their views for model adequety checking as long as n is not small relative to no. of parameters p.

2019 **Methods of Scaling Residuals**

Sometimes it is useful to work with scaled residuals, these scale residuals are helpfull in finding observations that are outliers or extreme values

(i) **Standardized Residuals**

It is given by $d_i = \frac{e_i}{\sqrt{MSSE}}$, $i = 1, 2, \dots n$

$\bar{d_i} = 0$ and $var(d_i) \simeq 1$

— If $d_i > 3$ it indicates outlier

(ii) **Studentied Residuals**

It is given by $r_i = \frac{e_i}{\sqrt{MSSE\ (1-h_{ii})}}$, $i = 1, 2, — n$

$h_{ii} \leftarrow i^{th}$ diagonal element of the hat matrix H

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

$Var(r_i) = 1$ and $\bar{r_i} = 0$

$Var(r_i) = 1$, regardless of location x_i when the form of the model is correct in many situations variance of the residuals stabilizes particularly for large datasets in this cases their may be little difference between standardized and Studentied residuals. Thus, standardized & studented residuals often convey equivalent informations.

(iii) PRESS Residuals

Another approach for making residuals useful in finding outliers is ~~the~~ to examine the quantity that is computed from $y_i - \hat{y}_{(i)}$

$\hat{y}_{(i)} \leftarrow$ is the fitted value of the ith response based on all obs except the ith one

The logic behind is that if the ith observation y_i is really unusual the regression model based on all obs. will be overly influenced by this obs. this could produce a fitted value $\hat{y}_i$ that is very similar to observed value y_i & hence, the ordinary residual e_i will be small

Therefore it will be hard to detect the outlier however if we delete ith obs. then $\hat{y}_{(i)}$ cannot be influenced by that obs. so the resulting residual should be likely to indicate the presence of the outlier if we delete the ith obs. fit the regression model to the remaining $n-1$ obs. and calculate the predicted value of y_i corresponding to the ~~devitale~~ deleted obs the corresponding prediction error is $e_{(i)} = y_i - \hat{y}_{(i)}$

This prediction error calculation is repeated for each obs 1 to n. This prediction errors are called PRESS residuals. Calculating PRESS residual requires fitting n different regression

ith PRESS residual is

$$e_{(i)} = \frac{e_i}{1-h_{ii}}, \quad i = 1, 2, \ldots n$$

$$var[e_{(i)}] = var\left[\frac{e_i}{1-h_{ii}}\right] = \frac{1}{(1-h_{ii})^2} var(e_i)$$

$$= \frac{\sigma^2(1-h_{ii})}{(1-h_{ii})^2} = \frac{\sigma^2}{1-h_{ii}}$$

So that a standardized PRESS residual is

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

If σ^2 is unknown than $\hat{\sigma}^2 = MSSE$

(iv) R- Student

In computing r_i it is customary to use MSSE as an estimate of σ^2 this is reffered as internal scaling of residual because MSSE is an internally generated estimate of σ^2 obtained from fitting the model to all 'n' obs

Another approach would be use, to use an estimate of σ^2 based on data set with ith obs removed Let $\hat{\sigma}^2 = S^2_{(i)}$

where

$$S^2_{(i)} = \frac{(n-p)MSSE - e_i^2/(1-h_{ii})}{n-p-1}.$$

The above estimate of σ^2 is used instead of MSSE to produce an externally studentized residual called R- student given by

$$t_i = \frac{e_i}{\sqrt{S^2_{(i)}(1-h_{ii})}} \qquad i = 1,2,\ldots n$$

$$t_i \sim t_{(n-p-1)}$$

If $|t_i| > t_{(\frac{\alpha}{2n}, n-p-1)}$ it indicates an outlier

* Residual Plots

Graphical analysis of residuals is a very effective way to investigate the adequacy ~~of~~ of the fit of a regression model and to check underlying assumption

(i) Normal Prob plots

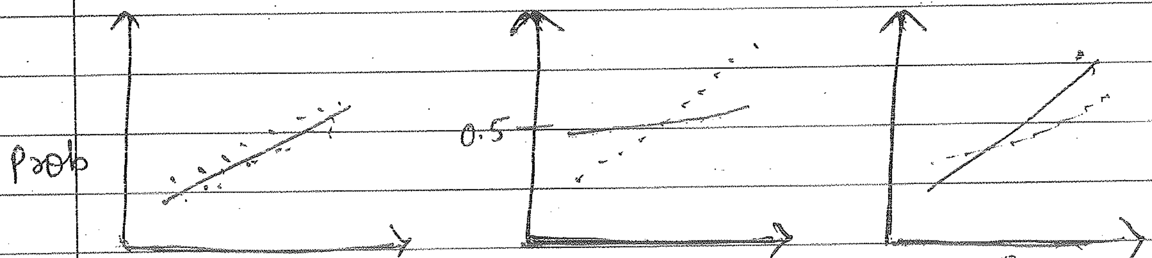
Small departures from the normality assumption doesn't affect the model greatly but gross non-normality is potentially more serious as t or F statistic and confidence and prediction interval depends on normality assumption

A very simple method of checking normallity assumption is to construct a normal prob plots of the residuals. this graph is so designed that cummulative normal $dist^n$ will plot as a straight line.

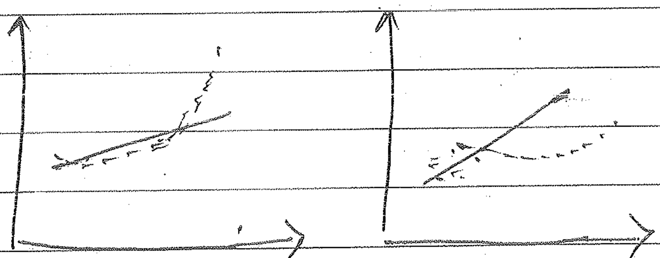
Let $e_{[1]} < e_{[2]} < \text{---} < e_{[n]}$ be the residual ranked in increasing order. If we plot $e_{[i]}$ against cumulative prob. $P_i = (i - 1/2)/n$, $i = 1, 2 \ldots n$ on the normal prob point the resulting points should lie approximately on a straight line This straight line is usually determined visually with amphasis on central values (eg 0.33 & 0.67 cummulative prob. points) rather than the extrems.

Substantial departures from a straight line indicate that $dist^n$ is non-normal sometimes normal prob plots are constructed by plotting the ranked residual $e_{[i]}$ against the expected normal value.

$\Phi^{-1}[(i-1/2)/n]$ where Φ denotes the Standard normal cummulative



$e_{(i)}$ (a) $e_{(i)}$ (b) $e_{(i)}$ (c)



$e_{(i)}$ (d) $e_{(i)}$ (e)

→ figure (a), This is idealized normal prob plot

→ figure (b), it shows a sharp ~~app~~ upward and downward curve at both extreme indicating that the tails of this $dist^n$ are too heavy for it to be considered normal

→

→ figure (c), shows flattening at the extreme which is a pattern typical of samples from a distribution with thinner tails than the normal.

→ Figure (d) indicates positive skewness

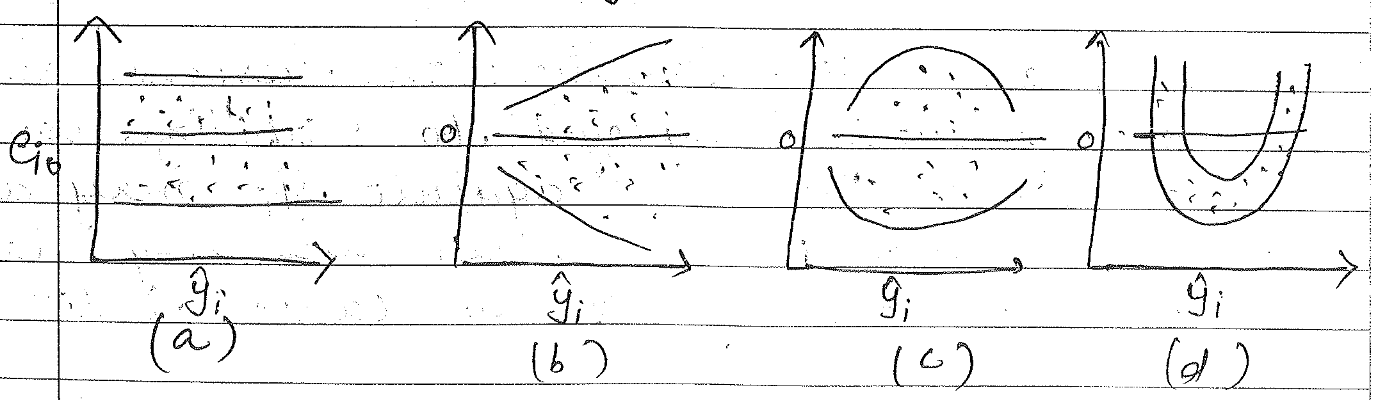
→ Figure (e) indicate -ve skewness

Small sample size ($n \leq 16$) often produce normal prob plots that deviate substantially from linearity for larger sample size ($n \geq 32$) the plots are much better ~~deviate~~ behaved. Usually above 20 obs are required to produced normal prob plots that are stable in a to be easily interpreted.

A common defect that shows up on the normal prob plot is the accurance of 1 or 2 large residual sometimes it is an indication that corresponding observations are outliers.

* Plot of residuals against the fitted values $\hat{y}_i$

A plot of residual e_i (or scaled residuals r_i, d_i or t_i) versus the corresponding fitted values $\hat{y}_i$ it is useful for detecting several types of model inadequacy.



→ Figure (a) Indicates that the residuals can be contained in a horizontal band and there is no model defect.

→ Figure (b) & (c) indicates that variance of the errors is not constant

→ figure (b) implies that variance is an increasing function of y and inversed opening funnel is also possible indicating variance increases as y decreases

→ Figure (c) occurs when y is a proportion b/w 0 and 1, the variance of a binomial proportion is greater than 0.5. ~~figure~~

→ Figure (d) indicates non-linearity. This means that other regressor variables are needed in the model for eg. a squared term can be added etc.

The usual approach for dealing with inequality of variance is to apply a suitable transfor-mation to either regressor variables or response variables. or to use the method of weighted ~~ls~~ least square.

In practice; transformations on response variable are generally employed to stabilize variance a plot of residuals against $\hat{y}_i$ may also ~~rea~~ reveal one or more unusually large residuals these points are called outliers

Remark :- large residuals that occurs at the extreme $\hat{Y}_i$ values could also indicate that either the variance is not constant or true relationship between y and x is not linear.

* Plot of residuals against regressors

Plotting the residuals against the corresponding values of each regressor variable can also be helpful here on the horizontal scale x_{ij} is plotted for the j^{th} regressor instead of $\hat{Y}_i$ and the conclusion remains same

* Detection and treatment of outliers

5M 2019

An outlier is an extreme observations residuals that are considerably larger in absolute values than the others say three or four standard deviation from the mean indicate potential y-space outlier.

Outliers are data points that are not typical of the rest of the data depending on their location in x-space outliers can have moderate to sevior ~~eff~~ seviors effects on regression effect against $\hat{Y}_i$ and normal probability plots are helpful in identifying outliers. examining scaled residuals such as studentized and r-student Residual is an excellent way to identify potential outliers. Outliers should be carefully investigated to see if a reason for their unusual behaviours can be found.

Sometimes, outliers are bad values occuring as a result of unusual but explainable events for eg. faulty measurement or analysis, incorrect recording of data and failure of a measuring instrument. If this is the case then the outlier should be corrected if possible or deleted from the data set clearly discarding bad values is desirable because least square pull the fitted eqn^s toward the outlier as it minimizes error sum of squares.

However, there should be strong non-Statistical evidence that outlier is a bad value before it is discarded. Ocassionally we find that the outlier is more important then the rest of the data because it may control many key model ~~modern~~ properties.

Outliers may also point out inadequacies in the model such as failure to fitt the data when in a certain region of x-space. If the outlier is a point of particularly desirable response. knowledge of regressor values when that response was observed may be extremly valueable. the effect of outliers on the regression model can be easily checked by dropping this points and refitting the regression equation

* Test of an outliers

(i) Suppose the ith case is suspected to be an outlier, delete the ith case from the data so $(n-1)$ cases remain in the data

(ii) Using the reduced data set estimate β and σ^2. Call this estimates as $\hat{\beta}_{(i)}$ and $\hat{\sigma}^2_{(i)}$ to remind that case (i) (obs i) was not used in the estimation. $\sigma^2_{(i)}$ has $(n-p-1)$ degree of ~~per~~ freedom

(iii) For the deleted case compute $\hat{y}_{i(i)} = x_i' \hat{\beta}_{(i)}$ Since ith case was not used in the estimation. $\hat{y}_i$ & $y_{i(i)}$ are indep.

(iv) Now under

$H_0: E(y_i - \hat{y}_{i(i)}) = \delta = 0$ i.e case i is not an outlier. $H_1: \delta \neq 0$

Assuming normal errors the test statistic for ith case is given by

$$t_i = r_i \left(\frac{n-p-1}{n-p-r_i^2} \right)^{1/2} \quad \text{where } r_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$$

We can use a Benferroni type approach & compare all 'n' values of $|t_i|$ to $t_{(\alpha/2,\, n-p-1)}$

We reject Ho if $|t_i| > (t_{(\alpha/2,\, n-p-1)})$

UNIT-4

Introduction to non-linear models

There are many situations where a linear regression model may not be appropriate for example the engineer or scientists may have direct knowledge of the form of relationship between the response variable and the regressors perhaps from the theory underline the phenomena

The true relationship between the response and the regressors may be a differential equation or the solution to a differential equation often lead to a model of non-linear form any model that is not linear in unknown parameters is a non-linear regression model

$$y = \theta_1 e^{\theta_2 x} + \varepsilon$$

is not linear in unknown parameter θ_1 & θ_2

let non linear regression model be represented as

$$y = f(x, \underline{\theta}) + \underline{\varepsilon}$$

$\underline{\theta}$ is a $p \times 1$ vector of unknown parameter
$\underline{\varepsilon}$ is an uncorrelated random error term with $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2 I$

$$E(\underline{y}) = f(x, \underline{\theta})$$

$f(x, \theta)$ is expectation func for the nonlinear regression model

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

$$S^2 = (Y - \beta_0 - \beta_1 x - \cancel{\varepsilon})^2$$

$$\frac{\partial S^2}{\partial \beta_0} = 2(Y - \beta_0 - \beta_1 x)$$

In a non-linear regression model atleast one of the derivatives of the expectation function with respect to parameters depends on atleast one of the parameters in linear regression these derivatives are not functions of unknown parameters

$$\left.\begin{aligned} Y &= \theta_1 e^{\theta_2 x} + \varepsilon \\ E(Y) &= \theta_1 e^{\theta_2 x} \\ \frac{\partial E(Y)}{\partial \theta_1} &= e^{\theta_2 x} \\ \frac{\partial E(Y)}{d\theta_2} &= \theta_1 \times e^{\theta_2 x} \end{aligned}\right\} \text{Non linear}$$

eg Consider the model $Y = \theta_1 e^{\theta_2 x} + \varepsilon$
The least Square normal eqns are

$$S^2 = \Sigma (y_i - \theta_1 e^{\theta_2 x_i})^2$$

$$\frac{\partial S^2}{\partial \theta_1} = 2\Sigma (y_i - \theta_1 e^{\theta_2 x_i})(e^{\theta_2 x_i}) = 0$$

$$\Rightarrow \Sigma y_i e^{\theta_2 x_i} - \theta_1 \Sigma (e^{\theta_2 x_i}) = 0$$

$$\frac{dS^2}{\partial 2} 2\Sigma (y_i - \theta_1 e^{\theta_2 x_i})(-\theta_1 x_i e^{\theta_2 x_i}) = 0$$

$$\Rightarrow \Sigma x_i y_i \theta_1 e^{\theta_2 x_i} - \theta_1^2 \Sigma x_i (e^{\theta_2 x_i})^2 = 0$$

This eqns are not linear in θ_1 and θ_2 no simple closed form solution exists

In general iterative procedure must be used to estimate θ_1 and θ_2.

It is sometimes useful to consider a transformation that induces linearity in the model expectation function

$$y = f(x, \theta) + \varepsilon$$

$$y = \theta_1 e^{\theta_2 x} + \varepsilon \quad -(1) \rightarrow \text{Non linear for}$$

$$E(y) = \theta_1 e^{\theta_2 x}$$

taking log transformation

$$\log E(y) = \log(\theta_1 e^{\theta_2 x})$$

$$y' = \log \theta_1 + \theta_2 x$$

$$y' = \theta_1' + \theta_2 x \quad -(ii) \rightarrow \text{linear form}$$

Use OLS method to estimate the parameters

However least square estimate of parameters in equation (2) will not in general be equivalent to non-linear parameter estimate in original model Y.

A non-linear model which can be transformed to an equivalent linear form is said to be intrinsically linear.

Egs of non-linear regression models

① Logistic growth model is

$$y = \frac{\theta_1}{1 + \theta_2 \exp(-\theta_3 x)}$$

② Gompertz model

$$y = \theta_1 \exp(-\theta_2 e^{-\theta_3 x}) + \varepsilon$$

③ Weibull growth model

$$y = \theta_1 - \theta_2 \exp(-\theta_3 x^{\theta_4}) + \varepsilon$$

Regressor - Independent variable
- used for prediction

* Multicolinearity

If there is no linear relationship between the regressor they are said to be orthogonal. when the regressors are orthogonal inferences such as

1 → Identifying the relative effects of the regressor variable.
2 → Prediction or estimation.
3 → Selection of appropriate set of variables for the model can be made relatively easy

unfortunately in most application of regression the regressor are not orthogonal sometimes the lack of orthogonality is not serious however in some situations the regressors are nearly perfectly linearly related and in such cases the inference based on regression model can be misleading
when there are near linear dependencies among the regressors the problem of multicolinearity is said to exists.

We write multiple regression model as

$$\underline{Y} = X\beta + \varepsilon$$

X: $n \times p$ matrix of the regression variables
Y: $n \times 1$ vector of responses
β: $p \times 1$ vector of unknown constants
ε: $n \times 1$ vector of random errors

$$\varepsilon_i \sim N(0, \sigma^2)$$

It is convinient to assume that the regressor variables and response has been centred and scaled to unit length

$X'X$: It is a $p \times p$ matrix of correlation betn the regressors.

$X'Y$: Is a $p \times 1$ vector of correlation betn the regressors and response.

Let the j^{th} column of x matrix be denoted by X_j so that $X = [X_1, X_2 \ldots X_j \ldots X_p]$ thus X_j contains the n levels of the j^{th} regressor variable. we formly define multicolinearity in terms of linear dependence of the columns of X the vector $X_1, X_2, \ldots X_p$ are linearly dependent if there is a set of constants $t_1, t_2, \ldots t_p$ not all zero such that $\sum_{j=1}^{p} t_j X_j = 0$ — (*)

If equation (*) holds exactly for a subset of the colums of X then the rank of $X'X$ matrix is $< p$ (less than p) thus $(X'X)^{-1}$ does not exists. However suppose eqn (*) is approximately true for some subset of columns of X then there will be near linear dependency in $X'X$ and the problem of multicolinearity is said to be exists in a form of ~~in~~ illconditioning in $X'X$ matrix

Sources of Multicolinearity

1. the data collection method employed
2. Constraints on the model or in the pop^n
3. Model specification
4. An overdefined model

(1) the data collection method can lead to multi-colinearity problem when the analyst sample only subspace of the reason of the regressor defined by the (x).

(2) Constraints on the model or in the population bel being sample can cause multicolinearity constraints often occurs in problems involving production or chemical process where the regressors are the components of a product and these components add o to a constant

(3) Multicolinearity may also be induced by choice of model for example adding polynomial terms to a regression model causes illconditioning in $X'X$. furthermore if range of X is small adding an X^2 term can result insignificant multicolinearity. We often encounter situations such as these where 2 ~~to~~ or more regressor are nearly linearly dependent and retaining all these regressor may contribute to multicolinearity.

(4) An overdefined model has more regressor variables than observation. these model are sometimes encountered in medical and

behavioural researche where there will be a small number of subjects. An information is collected for a large number of regressor for each subjects. The usual approach to deal with multicolinearity in this context is to elliminati some off the regressor variables from consideration.

* There are three specific recomendation made for the above case

(1) Redifine the model in terms of smaller sets of regressor

(2) Perform preliminary studies using only subsets of the original regressors.

(3) Use principle component type regression method to decide which regressors to remove from the model

* Effects of Multicolinearity

1st Effect

(1) Suppose there are only ~~large~~ two regressor variable x_1 & x_2 and if there is strong multicolinearity with the x_1 & x_2.
⇒ Means correlation coefficient r_{12} will be large which will result in large variances and co-variances for the least square estimator for of the regression coefficient. This implies that different samples taken at same x levels would lead to same x levels. widely different

estimate of the model parameter

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

$$(X'X)\hat{\underline{\beta}} = (X'\underline{Y})$$

$$\begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

$r_{12} \rightarrow$ Sample correlation betn x_1 & x_2

$r_{jy} \rightarrow$ Sample correlation bet x_j & y $j = 1, 2$

$$C = (X'X)^{-1} = \begin{bmatrix} \frac{1}{(1-r_{12}^2)} & \frac{-r_{12}}{(1-r_{12}^2)} \\ \frac{-r_{12}}{(1-r_{12}^2)} & \frac{1}{(1-r_{12}^2)} \end{bmatrix}$$

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12} r_{2y}}{(1-r_{12}^2)} \qquad \hat{\beta}_2 = \frac{r_{2y} - r_{12} r_{1y}}{(1-r_{12}^2)}$$

$$var(\hat{\beta}_j) = C\sigma^2$$

If there is strong multicolinearity between x_1 and x_2 then correlation coefficient r_{12} will be large so $|r_{12}| \rightarrow 1$. Hence variance of $var(\hat{\beta}_j) = C_{jj}\sigma^2 \rightarrow \infty$

$cov(\hat{\beta}_1, \hat{\beta}_2) = C_{12}\sigma^2 \rightarrow \pm\infty$ depending on $r_{12} \rightarrow \pm 1$.

Therefore strong multicolinearity between x_1 and x_2 results in large variances and covariance for the least square estimator there are more than two regressors multicolinearity produces similar effects the diagonal elements of C matrix $C = (X'X)^{-1}$ and $C_{jj} = \frac{1}{1-R_j^2}$, $j = 1, 2, \ldots, p$

$R_j^2 \leftarrow$ coefficient of Multiple determination from regression of x_j on remaining $(p-1)$ regressor variable.

If there is a strong multicolinearity b/w x_j and any subset of other $(p-1)$ regressors then $R_j^2 \to 1$ Hence $var(\hat{\beta}_j) \to \infty$

— X —

(continue on page no. 94)

* Choosing of K

Horel and kennard proved that there exist a non-zero value of k for which MSE of $\hat{\beta}_R$ is less than the variance of least square estimator $\hat{\beta}$ provided $\hat{\beta}'\beta$ is Bounded of least square estimator $\hat{\beta}$ provided $\beta'\beta$ is bounded.

Residual sum square is

$$SS_{Res} = (\underline{Y} - X\hat{\beta}_R)'(\underline{Y} - X\hat{\beta}_R)$$

$$= (\underline{Y} - X\hat{\beta}_R - X\hat{\underline{\beta}} + X\hat{\underline{\beta}})(\underline{Y} - X\hat{\beta}_R - X\hat{\underline{\beta}} + X\hat{\underline{\beta}})$$

$$= (\underline{Y} - X\hat{\underline{\beta}})'(\underline{Y} - X\hat{\underline{\beta}}) + (\hat{\beta}_R - \hat{\underline{\beta}})'(\hat{\beta}_R - \hat{\underline{\beta}}) \quad -(i)$$

First term is error sum of square for least square estimating. as K increases $\hat{\beta}_R$ decreases hence error sum of square increases Consequently, because TSS is fixed, R^2 decreases therefore the Ridge estimate will not necesserily provide the best fit to the data this should not concern us, since we are more interested in obtaining a stable set of parameter estimate

- Methods ~~of~~ to select the value of k

(i) Horel and kennard had suggested an appropriate value of k can be determined by inspection of Ridge trace. The ridge trace is a plot of $\hat{\beta}_R$ vs k for values of k usually in 0 to 1 interval.

If multicolinearity is serious, the instability in the regression coefficient will be seen from the ridge trace, as k is increased some of the ridge estimates will change at some value of k the ridge estimate $\hat{\beta}_R$ will stabilized.

The objective is to select a reasonably a small value of k at which ridge estimators $\hat{\beta}_R$ are stable.

(2) Hoerel, kennard and Baldwin had suggested an appropriate choice for k is

$$k = \frac{p\hat{\sigma}^2}{\hat{\underline{\beta}}'\hat{\underline{\beta}}} \quad \text{---} (*) \qquad p - \text{no. of parameters}$$

where $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are found from OLS they have shown that resulting ridge estimator has significant improvement in MSE over least square

(3) Horel and Kennard proposed an iterative estimation procedure based on (*) term. Specifically they suggested the following sequence of estimates of $\hat{\beta}$ and k

$$\hat{\underline{\beta}} : k_0 = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

$$\hat{\beta}_R(k_0) : k_1 = \frac{p\hat{\sigma}^2}{\hat{\beta}'_R(k_0)\,\hat{\beta}_R(k_0)}$$

$$\hat{\beta}_R(k_1) : k_2 = \frac{p\hat{\sigma}^2}{\hat{\beta}_R'(k_1)\,\hat{\beta}_R(k_1)}$$

Thus the relative change in k_j is used to terminate the procedure

If

$$\frac{k_{j+1} - k_j}{k_j} > 20T^{-1.3}$$

$$T = \frac{\text{Trace}\,(X'X)^{-1}}{p}$$

the algorithm should continue otherwise terminate

(4) Mc Donald and Galarneau suggest choosing k so that $\hat{\beta}_R'\hat{\beta}_R = \hat{\underline{\beta}}'\hat{\underline{\beta}} - \hat{\sigma}^2\left(\sum_{j=1}^{p} \frac{1}{\lambda_j}\right)$ — (**)

If RHS of (**) was negative they took $k=0$ or $k=\infty$

Note: None of the methods discussed above are best it is left to the experimentor to use appropriate method fined the value of k

★ **Some properties of Ridge estimate**

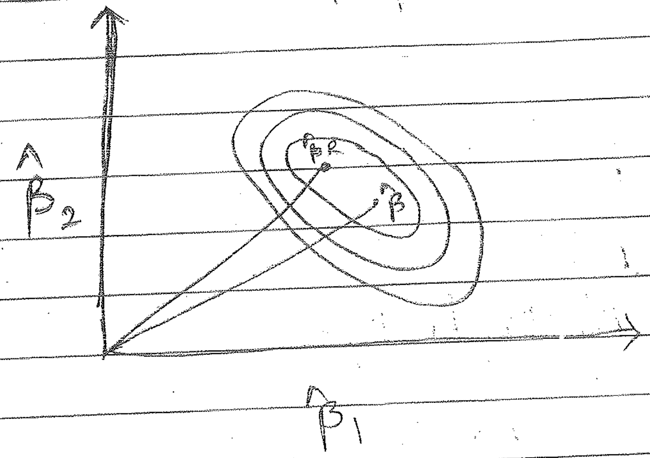


Figure indicated the geometry of ridge regression for a two regressor problem $\hat{\beta}$ at the centre of ellipse correspondse to least square solution where ESS takes minimum value. The small ellipse indicates locus of points in $\hat{\beta}_1$ $\hat{\beta}_2$ plane where ESS is constant at some value greater than the minimum.

The ridge estimate $\hat{\beta}_R$ is the shortest vector from the origin that produces ESS equal to the value represented by small ellipse. That is the ridge estimate $\hat{\beta}_R$ produces the vector of regression coefficient with smallest norm consistent with a specified increase in Residual sum of squares.

Ridge estimator shrinks the least square estimator towards the origin. Consequently ridge estimators are some times called shrinked estimators

Hoeking has observed that the ridge estimators shrinks the least square estimator with respect to contours of $X'X$ i.e $\hat{\beta}_R$ is the solution to minimizing

$$\min (\underline{\beta}-\hat{\underline{\beta}})' X' X (\hat{\underline{\beta}}-\underline{\beta})$$
$$\text{st } \underline{\beta}'\underline{\beta} \le d^2$$

where d is radius which depends on k

* Principle component Regression

2019 (5M)

Biased estimators of regression coefficients can also be obtained by using a procedure known as principle component regression consider canonical form of the model

$$\underline{Y} = Z\underline{\alpha} + \underline{\varepsilon} \quad \text{where} \quad Z = XT$$
$$\underline{\alpha} = T'\underline{\beta}$$
$$T'X'XT = Z'Z = \lambda$$

$\lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots \lambda_p)$ is a $p\times p$ diagonal matrix of the eigen values $X'X$ and T is $p\times p$ orthogonal matrix whose columns are eigen vectors associated with $\lambda_1, \lambda_2, \ldots \lambda_p$

The columns of Z which define a new set of orthogonal regressors $Z = [Z_1, Z_2, — Z_p]$ are referred as principle components. So the least square estimator of $\hat{\underline{\alpha}} = (Z'Z)^{-1} Z'\underline{Y}$
$$= \Lambda^{-1} Z'\underline{Y}.$$

and covariance matrix of $\hat{\underline{\alpha}}$ is

$$\operatorname{var}(\hat{\underline{\alpha}}) = \sigma^2 (Z'Z)^{-1}$$
$$= \sigma^2 \Lambda^{-1}$$

Thus a small eigen value of $x'x$ means that the variance of corresspording orthogonal regression coefficient will be large since

$$z'z = \sum_{i=1}^{p}\sum_{j=1}^{p} z_i z_j' = \Lambda$$

We often refer eigen value of λ_j has the variance of j^{th} principle component if all λ_j are equal to 1 the original regressors are orthogonal while if λ_j exactly equal to 0 this implies a perfect linear relationship between the original regressors

One or more of λ_j near zero implies that multicolinearity is present.

$$Var(\hat{\beta}) = var(T\hat{\alpha}) = T\Lambda^{-1}T'\sigma^2$$

$$\Rightarrow Var(\hat{\beta}_j) = \sigma^2 \sum_{j=1}^{p} t_{ji}^2/\lambda_i$$

Therefore variance of $\hat{B}_j$ is a linear combination of the reciprocals of eigen values

So this demonstrate how one or more small eigen values can destroy the precision of the least square estimate $\hat{\beta}_j$

Since $Z = XT$

we have $z_i = \sum_{j=1}^{p} t_{ji} x_j$ —(i)

where $\underline{x}_j$ is the j^{th} column of x matrix and t_{ji} are the elements of the i^{th} column of T. If the variance of i^{th} principle component λ_i is small this implies the Z_i is nearly constant and eq (1) indicates that there is a linear combination of original regressors that is ~~A~~ nearly constant

This is the def^n of Multicolinearity, therefore eq (1) explains ~~eig~~ by the ~~eig~~ elements of eigen vector associated with a small eigen value of $x'x$ identify the regressors involved in multicolinearity

The principle component regression approach deals with multicolinearity by using a subset of regressors which are called principle components. To obtain the principle component estimator assume that regressors are arranged in order of decreasing eigen value ($\lambda_1 \geq \lambda_2 \geq \dots \geq 0$

Suppose that last s of this eigen values are approximately to 0 (zero) in principle component regression the principle component corresponding to neer zero eigen values are removed from analyses and least square method is applied to remaining components

i.e $\hat{\alpha}_{pc} = B\hat{\alpha}$

where $b_1 = b_2 = \dots b_{p-s} = 1$

and $b_{p-s+1} = b_{p-s+2}$ —— $= b_p = 0$

Thus p.c estimator

$$\hat{\alpha}_{pc} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_{ps} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} p-s \text{ component} \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} s-\text{component} \end{matrix}$$

In terms of standardised regressors

$$\hat{\beta}_{pc} = T\hat{\alpha}_{pc}$$

$$= \sum_{j=1}^{p-s} \lambda_j^{-1} t_j' X' \underline{y} \, t_j$$

(*) Variable selection & model Building

In most practical problem the analyst have a pool of candidate regressor that should include all the influencial factors but the actual subset of regressors that should be used in the model need to be determined

Finding an appropriate subset of regressor for the model is called variable selection problem. Building a regression model that includes only a subset of available regressors involved ~~to~~ 2 conflicting objectives

(1) We would like to include as many regressors as possible so that the information content in this factors can influence the predicted value of Y

(2) We want to include few regressors in the model because the variance of $\hat{y}$ increases as the no. of regressors increasing. Also if there are more regressors in the

model cost of data collection and cost of maintaining the model increasing
the process of finding a model that is compromise between these two objectives is called selecting the best regression equation

* Criteria for evaluating subset regression model

(i) Coeff of Multiple determination
(ii) Residual Mean Square
(iii) Mellow Cp Statistic

Mellow Cp Statistics

Mellow has proposed a criterian that is related to mean Square of fitted value

$$E[\hat{y}_i - E(y_i)]^2 = \underbrace{E[y_i - E(\hat{y}_i)]^2}_{\text{squared bias}} + \underbrace{var(\hat{y}_i)}_{\text{variance component}} \quad \text{--- (1)}$$

$E(y_i)$ = expected response from the true regression equation
$E(\hat{y}_i)$ = Expected response from the ~~true~~ p-term subset model

Thus $E(y_i) - E(\hat{y}_i)$ is the bias at the ith data point.

let the total squared bias for p-term is

$$SS_B(p) = \sum_{i=1}^{n} [E(y_i) - E(\hat{y}_i)]^2$$

define the standardized total mean square error

$$\Gamma_p = \frac{1}{\sigma^2}\left\{\sum_{i=1}^{n}\left[E(y_i)-E(\hat{y}_i)\right]^2+\sum_{i=1}^{n} Var(\hat{y}_i)\right\}$$

$$\Gamma_p = \frac{SS_B(p)}{\sigma^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n} var(\hat{y}_i)$$

Now $\Sigma var(\hat{y}_i) = p\sigma^2$

So ~~expectation of~~

$$E(SS_{Res}(p)) = SS_B(p)+(n-p)\sigma^2$$

Substituting for $SS_B(p)$ and $\Sigma\, var(\hat{y}_i)$ Gives

$$\Gamma_p = \frac{1}{\sigma^2}\left\{E(SS_{Res}(p)) - (n-p)\sigma^2 + p\sigma^2\right\}$$

$$= \frac{E[SS_{Res}(p)]}{\sigma^2} - n + 2p$$

If σ^2 is unknown we replace it by $\hat{\sigma}^2$ also replacing ~~expects~~ $E[SS_{Res}(p)]$ by observed value $SS_{Res}(p)$, iff produced estimate Γ_p say C_p

$$C_p = \frac{SS_{Res}(p)}{\hat{\sigma}^2} - n + 2p$$

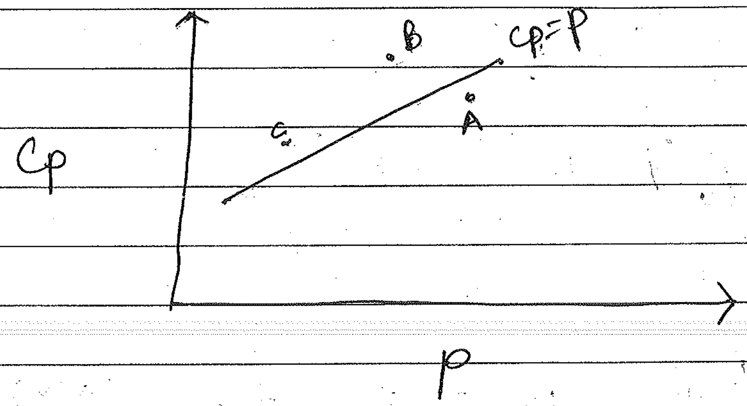
If p-term model has negligible bias then

$SS_B(P) = 0$

Consequently
So

$$E(SS_{Res}(P)) = (n-p)\sigma^2$$

$$E[C_p \mid Bias = 0] = \frac{(n-p)\cdot\sigma^2}{\sigma^2} - n + 2p = p$$



So regression equations with small Bias will have values of Cp that fall near the line while those with substantial bias will fall above the line.

Generally, small values of Cp are desirable

2nd Effect

MCT also tends to produce least square estimate $\hat{\beta}_j$ that are too large in absolute values.

For this consider the squared distance from β_j to the true parameter vector β

$$L_1^2 = (\hat{\underline{\beta}} - \underline{\beta})'(\hat{\underline{\beta}} - \underline{\beta})$$

$$E(L_1^2) = E\left[(\hat{\underline{\beta}} - \underline{\beta})'(\hat{\underline{\beta}} - \underline{\beta})\right]$$

$$= \sum_{j=1}^{p} E(\hat{\beta}_j - \beta_j)^2$$

$$= \sum_{j=1}^{p} Var(\beta_j) = \sigma^2 = trace\,(X'X)^{-1}$$

where trace of a matrix is the sum of main diagonal elements

When multicollinearity is present some of the eigen values of $X'X$ will be small since trace of a matrix is equal to sum of its eigen values then, above Eqn becomes

$$E(L_1^2) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}$$

where $\lambda_j > 0$, $j = 1, 2, \ldots p$ are upon values of $X'X$. Thus if $X'X$ matrix is ill-conditioned because of multicollinearity at least one of the λ_j will be small and the above equation implies that the distance from the least square Estimate $\hat{\beta}$ to the true parameter $\hat{\beta}$ may be large.

Detection of Multicolinearity (2019-5M)

① Examination of correlation matrix

(the no. of variates)

A very simple measure of MC is the inspection of the off diagonal elements r_{ij} in $X'X$. if the regressors X_i & X_j are nearly linearly dependent then $|r_{ij}|$ will be near 1.

Examining r_{ij} b/w the regressors is useful in detecting near linear dependence between pair of regressors only.

Unfortunately when more than 2 regresors are involved in a near linear dependence there is no assurence that any of the pairwise correlation r_{ij} will be large

Generally inspection of r_{ij} is not sufficient for detecting anything more complex than pairwise multicollinearity.

② Variance Inflation Method (VIF)

It is defined as.

$$VIF_j = C_{jj} = \frac{1}{1-R_j^2}$$

(Best model)

In VIF for each term in the model measures the combined effect of dependences among the regressors on the variances of that term.

One or more large VIF indicates multicollinearity. Practical Experience indicates that if any of VIF exceed 5 or 10 it is an indication that the associated regression coefficient are poorly estimated because of multicollinearity.

(3) Eigen System Analysis of $X'X$ $\{k_j\}$

The characteristic roots / eigen values of $X'X$ (say $\lambda_1, \lambda_2, \dots \lambda_p$) can be used to measure the extent of MCT in the data.

If there are one or more near linear dependencies in the data, then one or more characteristic roots will be small.

The condition indices of the $X'X$ matrix are $k_j = \frac{\lambda_{max}}{\lambda_j}$ $j = 1, 2, \dots p$

if $\boxed{k_j < 100}$ ⇒ no series problem of MCT
if $\boxed{k_j > 1000}$ ⇒ there is severe MCT
if in between moderate MCT

(4) The determinant of $X'X$ can be used as an index of Multicollinearity

Since $X'X$ matrix is a correlation matrix, the possible range of values of the determinant is $0 \le |X'X| \le 1$

If $|X'X| = 0 \Rightarrow$ there is multicollinearity (linear dependency among variables)

If $|X'X| = 1 \Rightarrow$ there is no multicollinearity (not linear dependency)

While this measure is easy to apply it does not provide any information among which variables are dependent

(5) The Sign and Magnitude of the Regression Coefficient will sometimes provide an indication that Multicollinearity is Present

In particular if adding / removing a regressor produces large changes in the estimates of regression coefficient multicollinearity is indicated.

If deletion / addition of one or more datapoints results in large changes in the Regression Coefficient then there may be MC? present

Finally if sign or magnitude of the Regression Coefficient in regression model are contrary to prior expectation there is possibility of multicollinearity

* Methods for dealing with Multicollinearity

1. Collecting Additional Information } not statistical Methods
2. Model Re-specification
3. Ridge Regression
4. Principal Component Regression

* Ridge Regression

When the method of least square is applied to non-independent orthogonal data very poor estimates of regression coefficients are obtained. The problem with method of least square is the requirement that $\hat{\beta}$ is an unbiased estimator of β

The gauss Markov Property assumes that least square estimator has min variance in the class of unbiased linear estimators but there is no guarantee that this variances will be small.

One way to solve this problem is to drop the requirement that the estimator of β be unbiased suppose that we can find a biased estimator of β, say $\hat{\beta}^*$

that has smaller variance than the unbiased estimator of the MSE of the estimator β^* is define as

$$MSE(\hat{\beta}^*) = E[\hat{\beta}^* - \beta]^2 + var\beta^* + [E(\beta^*) - \beta^*]^2$$

$$MSE(\hat{\beta}^*) = var(\hat{\beta}^*) + (Bias\ in\ \hat{\beta}^*)^2$$

By allowing a small amount of bias in $\hat{\beta}^*$ the variance of $\hat{\beta}^*$ can be made small such that

$MSE(\hat{\beta}^*) <$ the variance of $\hat{\beta}$ i.e unbiased estimator.

The small variance for the Biased Estimator also implis that $\hat{\beta}^*$ is a more stable estimator of β than unbiased estimator $\hat{\beta}$.

A no. of procedures have been developed for obtaining biased estimators of regression coffient One of them is Ridge Regression.

Ridge Estimator is $\hat{\beta}_R$ is the solution to

$$((x'x) + kI)\beta_R = x'y$$

OR

$$\hat{\beta}_R = (x'x + kI)^{-1} x'y$$

$k > 0$ is a constant selected by the analyst when $k=0$ Ridge Estimator is OLS estimator

~~where~~
since $(X'X)\hat{\beta} = X'y$ is OLS, we have

$$\hat{\beta}_R = (X'X + kI)^{-1} \cdot (X'X)\hat{\beta}$$
$$= Z_k \hat{\beta}$$

$$E(\hat{\beta}_R) = E[Z_k \hat{\beta}]$$
$$= Z_k [E(\hat{\beta})]$$
$$= Z_k \beta \qquad (\because E(\hat{\beta}) = \beta)$$

$\hat{\beta}_R$ is a biased estimator of β k is called biasing parameter.

$$V(\hat{\beta}_R) = var[(X'X + kI)^{-1} (X'X) \hat{\beta}]$$

$$= (X'X + kI)^{-1} (X'X)\, var(\beta)\, (X'X) (X'X + kI)^{-1}$$

$$= \sigma^2 ((X'X + kI))^{-1} (X'X) (X'X + kI)^{-1}$$

$$MSE(\hat{\beta}_R) = var(\hat{\beta}_R) + (bias\ in\ \hat{\beta}_R)^2$$

$$= \sigma^2 tr[(X'X + kI)^{-1} (X'X) (X'X + kI)^{-1}] + k^2 \beta' (X'X + kI)^{-2} \beta$$

$$= \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \beta' (X'X + kI)^{-2} \beta \qquad (*)$$

where $\lambda_1, \lambda_2, \ldots \lambda_p$ are the eigen values of $X'X$
If $k > 0$ the bias in $\hat{\beta}_R$ increases with ~~How~~ However variance decreases as k ~~invariance~~ increases

In using Ridge Regression, we would like to choose a value K such that the Reduction in variance term is greater than the increase in squared Bias

If this can be done, MSE of Ridge Estimator will be less than, the variance of least square estimator $\hat{\beta}$.