27/6/25

Unit - 1

PAGE NO.:
DATE: / /

* Linear Estimation :-

Def^n :- Linear Estimator

- The Estimator which are linear funct^{ns} of the obs^{ns} are known as Linear Estimator.

* Concept of Unbiased Linear Estimators of Linear funct^{ns} of parameters :-

- Sample Mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is an unbiased estimator of pop^n mean μ
  i.e. $\boxed{\hat{\mu} = \bar{y}}$
  which is a linear unbiased estimator.

Linear Models

- $s^2 = \frac{1}{n-1}\sum(y_i - \bar{y})^2$ is an unbiased estimator of pop^n variance σ^2 i.e. $\hat{\sigma}^2 = s^2$
  But s^2 is not a linear estimator but a quadratic estimator of pop^n var. σ^2.

- If we have more than one unbiased linear estimator then to select the best estimator we use the concept of BLUE (Best Linear Unbiased Estimator)

- The principle which is used to make this selection is to choose that linear estimator from among all unbiased linear estimators which has the smallest variance.

- The sampling dist^n of that estimator will have the maximum concentration around the unknown true parametric function. Such an estimator is known as Minimum Variance Unbiased Linear Estimator or BLUE.

- This Minimum Variance Unbiased Linear Estimators have variances and covariances which again lead lend themselves to unbiased estimation.

- The Minimum Variance approach to the estimation of the parameters in the linear model was given by Markov (1900).

- 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 book on astronomical problems. The combined results is the famous Gauss - Markov Theorem.

* Gauss - Markov Linear Model :-

- Consider a set of 'n' independent r.v's $y_1, y_2, \ldots, y_n$ with a common variance σ^2 whose expectation are linear funct^{ns} with known coefficients (a_{ij}'s) of p unknown parameters $\beta_1, \beta_2, \ldots, \beta_p$ $(p<n)$

Thus,

$$\left.\begin{aligned} E(y_i) &= a_{i1}\beta_1 + a_{i2}\beta_2 + \cdots + a_{ip}\beta_p \\ Var(y_i) &= \sigma^2 \quad \forall\, i = 1, 2, \ldots, n \\ Cov(y_i, y_j) &= 0, \quad i \neq j \end{aligned}\right\} \quad \text{--- (1)}$$

Eq^n (1) is called the Gauss - Markov Linear Model.
Eq^n (1) can be represented in matrix form as

Let $\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$; $\underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$

and matrix of the known Coefficient as

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}$$

Eq^n (1) can be written as

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

where $D(\underline{Y})$ stands for Dispersion Matrix and I is the identity matrix of order n.

- An alternative way of representing eq^n (2) is using column vector $\underline{e}$ of independent errors $e_1, e_2, \ldots, e_n$ as

$$\left.\begin{aligned} &\underline{Y} = A\underline{\beta} + \underline{e} \\ &E(\underline{e}) = \underline{0}, \quad D(\underline{e}) = \sigma^2 I \end{aligned}\right\} \text{———} (3)$$

where $\underline{0}$ is a null vector

- The unknown parameters β_j's in the Model are called effects.

* **Fixed Effect Model :-**

- In linear estimation, the effects are all fixed quantities (parameters) and such a model where all effects are unknown parameters is called Fixed

Effect Model or Model I. Sometimes, one of the β_j's is a constant with $a_{ij} = 1$ for that j and all $i = 1, 2, \ldots, n$. Such an effect is called General Effect or an additive 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 effects are random then it is called Mixed Effect Model.

- Gauss Markov Linear Model may be classified into three broad areas depending on the nature of the values taken by coefficient a_{ij}'s.

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

Design

→ A Model in which a_{ij}'s takes the values as indep. variables then we have a regression model.

Regression

→ A Model in which a_{ij}'s take indicator as well as indep. variable then the model is Analysis of Covariance Model.

5/7/25

* Concept of Estimable funct^ns is due to R.C. Bose

Def^n: A linear funct^ns of the parameters $\beta_1, \beta_2, \ldots, \beta_p$ is called a Parametric funct^n.
Thus $l_1\beta_1 + l_2\beta_2 + \cdots + l_p\beta_p = \underline{l}'\underline{\beta}$ is a parametric funct^n where $\underline{l}' = (l_1, l_2, \ldots, l_p)$ is known.

Def^n: An estimable funct^n is a parametric funct^n $\underline{l}'\underline{\beta}$ for which an unbiased linear estimator exists.
Thus if $\exists$ a vector $\underline{c} = (c_1, c_2, \ldots, c_n)'$ of constants $\ni$ $E(\underline{c}'\underline{y}) = \underline{l}'\underline{\beta}$ indep. of the parameters then the parametric funct^n $\underline{l}'\underline{\beta}$ is such to be linearly estimable or to be an estimable funct^n

Theorem (1): A necessary and sufficient condition for a parametric funct^n or a linear parametric funct^n $\underline{l}'\underline{\beta}$ of the parameters to be linearly estimable is
$\text{rank}(A) = \text{rank}\begin{pmatrix} A \\ \underline{l}' \end{pmatrix}$ where $\begin{pmatrix} A \\ \underline{l}' \end{pmatrix}$ is the matrix obtained from A by adjoining the row vector $\underline{l}'$.

Proof: From the def^n, we see that $\underline{l}'\underline{\beta}$ is estimable iff $\exists$ a vector of constants $\underline{c}$ $\ni$

$$E[\underline{c}'\underline{y}] = \underline{l}'\underline{\beta} \quad \forall \ \underline{\beta}$$

But $$E[\underline{c}'\underline{y}] = \underline{c}'E(\underline{y}) = \underline{c}'A\underline{\beta}$$

So $\underline{l}'\underline{\beta}$ is estimable iff $\underline{c}'A\underline{\beta} = \underline{l}'\underline{\beta}$ identically in $\underline{\beta}$ iff $\underline{c}'A = \underline{l}'$ —— (i) —— (*)

To find an unbiased estimator for $\underline{l}'\underline{\beta}$ we have to solve (i) for an unknown vector $\underline{c}$

And the condition for existence of eq^n (i) is

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

Corollary-(1): If rank $(A) = P$ every linear function of the parameters is linearly estimable.

Proof :- Whatever may be $\underline{l}$ giving parametric funct^n $\underline{l}'\underline{\beta}$

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

But $\text{rank}\begin{pmatrix} A \\ \underline{l}' \end{pmatrix}$ cannot exceed P,

Since $\begin{pmatrix} A \\ \underline{l}' \end{pmatrix}$ has only 'P' columns.

Hence if rank $(A) = P$, then

$\text{rank}\begin{pmatrix} A \\ \underline{l}' \end{pmatrix} = \text{rank}(A)$ so by thm (1)

every $\underline{l}'\underline{\beta}$ is estimable.

Corollary-(2) :- Every linearly estimable parametric funct^n is of the form $\underline{b}'A\underline{\beta}$ or

$b_1 E(y_1) + b_2 E(y_2) + \cdots + b_n E(y_n)$ where

$\underline{b}' = (b_1, b_2 \ldots b_n)$

Proof :- If $\underline{l}'\underline{\beta}$ is estimable then by (*) $\underline{l}'$ is linearly dependent on the rows of A, so that $\underline{l}'$ is of the form $\underline{b}'A$.

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

PAGE NO. :
DATE : / /

(r.v)

Def^n :- A linear funct^n of $y_1, y_2, \dots, y_n$ is said to belong to error if the expectation vanishes (becomes zero) independently of the parameters.

- If $\underline{d}' = (d_1, d_2, \dots, d_n)$, then $\underline{d}'\underline{y}$ belong to error iff

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

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

Theorem (2) :- A necessary and sufficient condition for the linear funct^n $\underline{d}'\underline{y}$ to belong to error is that $\underline{d}'$ is orthogonal to the vector space $V(A')$ generated by the row vectors of A'.

- The vector space $V(B)$ which is orthogonal to the space $V(A')$ is called the Error space

and

- The vector space $V(A')$ which contains the coefficient vectors of the best estimators of estimable funct^n is known as Estimation Space.

Theorem (3) :- If $\underline{l}'\underline{\beta}$ is any estimable linear funct^n of the parameters $\beta_1, \beta_2, \dots, \beta_p$ then

(i) $\exists$ a unique linear funct^n $\underline{c}'\underline{y}$ of the r.v's $y_1, y_2, \dots, y_n \ni \underline{c} \in V(A')$ and $E(\underline{c}'\underline{y}) = \underline{l}'\underline{\beta}$.

(ii) $Var(\underline{c}'\underline{y})$ is less than the var of any other linear unbiased estimator of $\underline{l}'\underline{\beta}$.

Proof :- (i) Since $\underline{l}'\underline{\beta}$ is estimable, $\exists$ a linear funct^n $\underline{b}'\underline{y}$ of r.v.'s $\ni E(\underline{b}'\underline{y}) = \underline{l}'\underline{\beta}$

PAGE NO.:
DATE: / /

Now we can uniquely resolve $\underline{b}'$ into $\underline{c}'$ and $\underline{d}'$ $\ni$ $\underline{c}' \in V(A')$ and $\underline{d}' \in V(B)$ which is orthogonal to $V(A')$

Hence $\underline{b}'\underline{y} = \underline{c}'\underline{y} + \underline{d}'\underline{y}$

where $\underline{d}'\underline{y}$ belong to error

Also,

$$E(\underline{b}'\underline{y}) = E(\underline{c}'\underline{y}) + \underbrace{E(\underline{d}'\underline{y})}_{\text{zero } (0)}$$

$$\Rightarrow E[\underline{c}'\underline{y}) = E(\underline{b}'\underline{y}) = \underline{l}'\underline{\beta}$$

$$(\because E(\underline{d}'\underline{y}) = 0)$$

Thus $\exists$ a linear funct^n $\underline{c}'\underline{y}$ with $\underline{c}' \in V(A')$ $\ni$ $E(\underline{c}'\underline{y}) = \underline{l}'\underline{\beta}$

Now to show that this is unique.

If possible let there be another row vector $\underline{c}_0' \in V(A') \ni$

$$E(\underline{c}_0'\underline{y}) = \underline{l}'\underline{\beta}$$

Define the row vector $\underline{c}_1' = \underline{c}' - \underline{c}_0'$

Then $E(\underline{c}_1'\underline{y}) = E(\underline{c}'\underline{y}) - E(\underline{c}_0'\underline{y})$

Thus $\underline{c}_1'$ belongs to error space and being a linear combination $\underline{c}'$ and $\underline{c}_0'$ also belongs to estimation space but this is impossible unless $\underline{c}_1'$ is a null vector has non-null vector cannot lie into two orthogonal space.

$\underline{c}_1'$ is a null vector implying $\underline{c}' = \underline{c}_0'$. Thus $\underline{c}'$ which lies in $V(A')$ for which $E(\underline{c}'\underline{y}) = \underline{l}'\underline{\beta}$

(ii) Let $\underline{b}'\underline{y}$ be any arbitrary unbiased linear estimator of $\underline{l}'\underline{\beta}$.

then $Var(\underline{b}'\underline{y}) = \underline{b}'\, Var(\underline{y})\, \underline{b}$

$$= \underline{b}'\sigma^2\underline{b}$$

Corollary

Proof :

Corollar

Proof :-

PAGE NO.:
DATE: / /

$$= (\underline{c}' + \underline{d}')(\underline{c} + \underline{d})\sigma^2$$
$$= \underline{c}'\underline{c}\sigma^2 + \underline{d}'\underline{d}\sigma^2$$
$$= var(\underline{c}'\underline{y}) + Var(\underline{d}'\underline{y})$$

Hence,
$$Var(\underline{b}'\underline{y}) \geq Var(\underline{c}'\underline{y})$$
and equality holds iff $Var(\underline{d}'\underline{y}) = 0$
ie. $\underline{d}'\underline{d} = 0$ or $\underline{d}'$ is a null vector.

$\Rightarrow$ $\underline{c}'\underline{y}$ has minimum variance.

Corollary :- The best estimator of any estimable funct^n $\underline{l}'\beta$ must be of the form $\underline{q}'A'\underline{y}$ where $\underline{q}' = (q_1, q_2, \ldots, q_p)$ is a row vector and satisfies the eq^n $\underline{q}'A'A = \underline{l}'$

Proof :- The coefficient vector $\underline{c}'$ of the best estimator lies in $V(A')$
Hence $\underline{c}' = \underline{q}'A'$ for a suitable $\underline{q}'$ and the best estimator is of the form $\underline{c}'\underline{y} = \underline{q}'A'\underline{y}$
Since $E(\underline{q}'A'\underline{y})$ must be $\underline{l}'\underline{\beta}$
$$\Rightarrow \underline{q}'A'A = \underline{l}'$$

Corollary :- Let $\underline{l}_i'\underline{\beta}$ for $i = 1, 2, \ldots, k$ be 'k' estimable parametric funct^n and let $T_i = \underline{c}_i'\underline{y}_i$ for $i = 1, 2, \ldots k$ be their estimators. Then the best estimator of $\Sigma b_i \underline{l}_i'\underline{\beta}$ is $T = \Sigma b_i T_i$

Proof :-
$$E(T) = E(\Sigma b_i T_i)$$
$$= \Sigma b_i E(T_i)$$
$$= \Sigma b_i E(\underline{c}_i'\underline{y}_i)$$
$$= \Sigma b_i \underline{l}_i'\underline{\beta}$$
T_i is an unbiased estimator of $\Sigma b_i \underline{l}_i'\underline{\beta}$.

PAGE NO. :
DATE : / /

* Least Square Estimators :

- Let $b_1, b_2, \ldots, b_p$ denotes any set of p known quantities which can be used as estimates of $\beta_1, \beta_2, \ldots, \beta_p$.

$\hat{\beta}_1 = b_1, \ldots, \hat{\beta}_p = b_p$.

Def^n :- A set of measurable funct^n $\underline{y}$, 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 squares of the deviation $y_1, y_2, \ldots, y_n$ from their expectation

i.e. $S = (\underline{y} - A\underline{\beta})'(\underline{y} - A\underline{\beta})$ is called a set of least square estimators of the known parameters $\beta_1, \beta_2, \ldots, \beta_p$ of the linear model.

11/7/25

* Least Square Estimators and Normal eq^{ns} :-

$$S = [\underline{y} - A\underline{\beta}]'[\underline{y} - A\underline{\beta}]$$

To show that the minimum value of S is attained when $\underline{\hat{\beta}}$ is a sol^n of a set of eq^{ns} which are called Normal eq^{ns}.

We have $\underline{\beta}'A'\underline{y} = \underline{y}'A\underline{\beta} = \sum_{\alpha, j} a_{\alpha j}\beta_j y_\alpha$; $j = 1, 2, \ldots, p$
$\alpha = 1, 2, \ldots, n$

Hence,

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

$$= \sum_{\alpha, j} a_{\alpha j} y_\alpha = \boxed{\underline{d}_j'\underline{y}}$$

PAGE NO. :
DATE : / /

where $d_1, d_2, \ldots, d_p$ are the column vector of A.

Let $A'A = C = (C_{ij})$
where C is a symmetric matrix of order p.
Now,

$$\underline{\beta}' A' A \underline{\beta} = \underline{\beta}' C \underline{\beta} = \Sigma C_{ij} \beta_i \beta_j \quad ; \; i,j = 1,2,\ldots,p$$

Hence,

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

$$= 2 \sum_i C_{ij} \beta_i$$

$$= 2 \underline{C}_j' \underline{\beta}$$

where

$$\underline{C}_j' = (C_{1j}, C_{2j}, \ldots, C_{pj})$$

$$S = [\underline{y} - A\underline{\beta}]' [\underline{y} - A\underline{\beta}]$$

$$\frac{\partial S}{\partial \beta} = 0$$

$$S = [\underline{y}'\underline{y} - \underline{y}' A \underline{\beta} - A'\underline{\beta}'\underline{y} - A'\underline{\beta}' A \underline{\beta}]$$

$$\frac{\partial S}{\partial \beta} = 0 \Rightarrow -2 \frac{\partial}{\partial \beta_j}(\underline{\beta}' A' \underline{y}) + \frac{\partial}{\partial \beta_j}(\underline{\beta}' A' A \underline{\beta}) = 0$$

$$\Rightarrow -2 \underline{d}_j' \underline{y} + 2 \underline{C}_j' \underline{\beta} = 0 \quad ; \; j = 1,2,\ldots,p$$

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

Eq^n $(*)$ are normal eq^{ns} and are equivalent to

$$\left. \begin{array}{c} A' A \underline{\beta} = A' \underline{y} \\ \text{or} \\ C \underline{\beta} = A' \underline{y} \end{array} \right\} \text{——— } (**)$$

where $C = A'A$

PAGE NO. :
DATE : / /

The normal eq^{ns} always admit a solution since $\underline{A'y}$ lies in the vector space generator by the columns of C.

Let $\hat{\underline{\beta}}$ be a sol^{ns} of these eq^{ns}

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

NOTE :- Every solution of the normal eq^{ns} is a set of least square Estimators and every set of least square estimators satisfies the normal equations.

Result :- The sol^{n} of normal eq^{ns} $\underline{\beta} = \hat{\underline{\beta}}$ gives an extreme value of S and thus extreme value is the minimum value of S.

ie $\underline{\beta} = \hat{\underline{\beta}}$ minimizes S.

Proof :- Consider $(\underline{y} - A\underline{\beta})'(\underline{y} - A\underline{\beta})$

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

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

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

($\because$ Since the quadratic form $[A(\hat{\underline{\beta}} - \underline{\beta})'][A(\hat{\underline{\beta}} - \underline{\beta})]$ cannot be negative)

The equality holds only when $\underline{\beta} = \hat{\underline{\beta}}$

$\Rightarrow \underline{\beta} = \hat{\underline{\beta}}$ minimizes S.

Further if $\underline{\beta}$ and $\hat{\underline{\beta}}$ are any two sol^{ns} of (*) then $(\underline{y} - A\hat{\underline{\beta}})'(\underline{y} - A\hat{\underline{\beta}}) = (\underline{y} - A\hat{\hat{\underline{\beta}}})'(\underline{y} - A\hat{\hat{\underline{\beta}}})$
this along with (***) show that every sol^{ns} of the normal eq^{ns} is a set of LS estimators.

* Gauss - Markov Theorem :-

The best estimator of the estimable linear funct^{ns} $\underline{l}'\underline{\beta}$ of the parameters is $\underline{l}'\hat{\underline{\beta}}$ where $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$ are a set of LSE of $\beta_1, \beta_2, \ldots, \beta_p$.

In other words, LSE of $\underline{l}'\underline{\beta}$ is identical with min. variance linear unbiased estimator of $\underline{l}'\underline{\beta}$.

Proof :- From Corollary (1) we have BLUE of $\underline{l}'\underline{\beta}$.

$$\underline{l}'\underline{\beta} = \underline{q}'A'\underline{y} = \underline{q}'A'A\hat{\underline{\beta}} = \underline{l}'\hat{\underline{\beta}}$$

Since from the normal eq^{ns} we have

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

Since $\underline{q}'$ satisfies $\underline{q}'A'A = \underline{l}'$ and $\underline{l}'\hat{\underline{\beta}} = \underline{q}'A'\underline{y}$

$\underline{l}'\hat{\underline{\beta}}$ is the same for all LSE ~~of~~ $\hat{\underline{\beta}}'$ of $\underline{\beta}$ obtained by solving Normal eq^{n}.

* Variance and Covariance of LSE :-

The normal eq^{ns} are given by

$$A'A\underline{\beta} = A'\underline{y}$$

Let $A'A = C$

$\Rightarrow \quad C\underline{\beta} = \underline{Q} \quad$ where $\underline{Q} = A'\underline{y}$ (say)

$E(\underline{y}) = A\underline{\beta}$

$$E(A'\underline{y}) = E(\underline{Q}) = A'E(\underline{y}) = A'A\underline{\beta}$$
$$= E(\underline{Q}) = E(\underline{Q})$$

PAGE NO.:
DATE: / /

Remark

$\Rightarrow \quad E(\underline{q}) = C\underline{\beta}$

$E(\underline{q}) = \underline{q}$

Let $\bar{C}$ be a g-inverse of C then a solⁿ of Normal eqⁿˢ is given by

$$\underline{\hat{\beta}} = \bar{C}\underline{q}$$

$$D(\underline{\hat{\beta}}) = D(\bar{C}\underline{q})$$

$$D(\underline{q}) = D(C\underline{\beta}) = D(A'\underline{y}) = A'D(\underline{y})A = A'A\sigma^2 = C\sigma^2$$

$$D(\underline{\hat{\beta}}) = D(\bar{C}\underline{q}) = (\bar{C})D(\underline{q})(\bar{C})' = (\bar{C}\cdot C)\sigma^2\bar{C} = \sigma^2\bar{C} = \boxed{\sigma^2(A'A)^-}$$

Theorem:- If $\underline{l}_1'\underline{\hat{\beta}}$ and $\underline{l}_2'\underline{\hat{\beta}}$ LSE of two estimable functⁿ $\underline{l}_1'\underline{\beta}$ and $\underline{l}_2'\underline{\beta}$ respectively then

$$Var(\underline{l}_1'\underline{\hat{\beta}}) = \sigma^2\underline{l}_1'\bar{C}\underline{l}_1$$

$$Cov(\underline{l}_1'\underline{\hat{\beta}}, \underline{l}_2'\underline{\hat{\beta}}) = \sigma^2\underline{l}_1'\bar{C}\underline{l}_2$$

where $\bar{C}$ is a g-inverse of $C = A'A$

Proof:- $Var(\underline{l}_1'\underline{\hat{\beta}}) = \underline{l}_1' Var(\underline{\hat{\beta}})\underline{l}_1 = \boxed{\sigma^2\underline{l}'\bar{C}\underline{l}}$

$$Cov(\underline{l}_1'\underline{\hat{\beta}}, \underline{l}_2'\underline{\hat{\beta}}) = \underline{l}_1' Cov(\underline{\hat{\beta}}, \underline{\hat{\beta}})\underline{l}_2 = \underline{l}_1' Var(\underline{\hat{\beta}})\underline{l}_2$$

$$= \boxed{\sigma^2\underline{l}_1'\bar{C}\underline{l}_2}$$

Remark: If $(A'A) = C$ is a non-singular matrix then $\text{rank}(A) = p$

$\Rightarrow C^{-1}$ exists and a unique sol^n of $\underline{\beta} = \hat{\underline{\beta}}$ is obtained $(A'A)\underline{\beta} = A'\underline{y}$

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

$$\because E(\hat{\underline{\beta}}) = E[(A'A)^{-1}A'\underline{y}]$$
$$= (A'A)^{-1}A'E(\underline{y})$$
$$= (A'A)^{-1}(A'A)\underline{\beta}$$
$$\boxed{E(\hat{\underline{\beta}}) = \underline{\beta}}$$

$$D(\hat{\underline{\beta}}) = D[(A'A)^{-1}A'\underline{y}]$$
$$= [(A'A)^{-1}A']' D(\underline{y}) [(A'A)^{-1}A']$$
$$= ((A'A)^{-1}A')' \sigma^2 (A'A)^{-1}A'$$
$$= (A'A)^{-1}\sigma^2 \quad (A'A \text{ is symmetric})$$

* **Linear Estimation with linear restrictions on parameters** :-

When the parameters are subject to a set of consistent linear restrictions $\underline{P}'\underline{\beta} = \underline{L}$ with $\text{rank}(P) = P\_s$, the appropriate linear model is

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

In this case there are two method to estimate the parameters. First we eliminate some of parameters in the obs^n eq^n with the help of restrictions and obtained a different eq^n in linear set of observational eq^{ns} with less parameters and having non restrictions on these fewer parameters.

This case is then similar to model

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

and in this approach, let $\beta_0 + F\beta^*$ be a general sol^n of $P'\beta = L$ where β_0 is a particular a sol^n and $P'F' = 0$.

- β^* being the arbitrary vector of fewer parameters (s in no.) with $z = y - A\beta_0$

$$E(z) = AF\beta^*$$

$$D(z) = \sigma^2 I$$

Thus with z the model reduces to the original model with fewer parameters $\beta_1^*, \beta_2^*, \ldots \beta_s^*$ and no restrictions β^*'s.

To estimate,

$$l'\beta = l'(\beta_0 + F\beta^*) = l'\beta_0 + l'F\beta^*$$

we need to consider only $l'F\beta^*$.

12/7/25

- Second approach is to minimize the S.S. $(y - A\beta)'(y - A\beta)$ such that the conditions $P'\beta = L$ using a lagrangian multiplier λ.

$$S^2 = (y - A\beta)'(y - A\beta) + \lambda(P'\beta - L)$$

The normal eq^n are

$$S^2 = (y'y - y'A\beta - \beta'A'y + \beta'A'A\beta + \lambda(P'\beta - L)$$

$$\frac{\partial S^2}{\partial \beta} = 0 \rightarrow 2\underline{\beta}' A'A + \underline{P}'\underline{\lambda} - 2A'\underline{y} = 0$$

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

$$P\underline{\lambda}$$

$$\frac{\partial S^2}{\partial \lambda} = 0 \Rightarrow \underline{P}'\underline{\beta} = \underline{L}$$

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

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

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

$$\hat{\lambda} = (\underline{P}')^{-1} A'\underline{y}.$$

* Linear Estimation with correlated variables :-

- We consider a model, it is more general then the Gauss - Markov Linear Model

$$\left.\begin{aligned} E(\underline{y}) &= A\underline{\beta} \\ D(\underline{y}) &= \sigma^2 \underline{B} \end{aligned}\right\} \text{——— } (1)$$

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

- The above model can be reduced to the original Gauss - Markov model as follows :-
Since 'B' is assumed to be non - singular matrix of order 'n' $\exists$ a n.s. matrix H of order 'n' $\ni$ $H'BH = I$

PAGE NO.:
DATE: / /

Then we consider the transformation

$$\underline{y}^* = H'\underline{y}$$

$$\left.\begin{aligned} E(\underline{y}^*) &= E(H'\underline{y}) = H'E(\underline{y}) = H'A\underline{\beta} = A^*\underline{\beta} \quad (\text{say}) \\ D(\underline{y}^*) &= D(H'\underline{y}) = H'D(\underline{y})H \\ &= H'\sigma^2 BH \\ &= \sigma^2 I \end{aligned}\right\} (2)$$

There exist a linear transformation which ~~requires~~ reduces the case of correlated variable to the earlier model with the matrix of coefficient

$$A^* = H'A$$

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

So all the results for model (2) (the original model) will also be true above model.

<u>Remark</u>:- However it is simple to used the actual obs^n $(y_1, y_2, \ldots, y_n)$ instead of transformed obs^n $(y_1^*, y_2^*, \ldots, y_n^*)$

LSE is found by minimising the following S.S in terms of transformed variables.

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

Since $(\underline{y}^* - A\underline{\beta}^*) = H'\underline{y} - H'A\underline{\beta}$

$= H'(\underline{y} - A\underline{\beta})$ and $HH' = B^{-1}$

we have,

$$(\underline{y} - A\underline{\beta})'HH'(\underline{y} - A\underline{\beta})$$

$$= (\underline{y} - A\underline{\beta})'B^{-1}(\underline{y} - A\underline{\beta}) \quad \text{with } B^{-1} = b^{ij}$$

The above model can be expressed as

$$\sum_i \sum_j b^{ij} (y_i - a_{i_1}\beta_1 - a_{i_2}\beta_2 - \cdots a_{ip}\beta_p)$$
$$\times (y_j - a_{j_1}\beta_1 - a_{j_2}\beta_2 - \cdots a_{jp}\beta_p)$$

is called the Weighted Sum of Squares.

- LSE of β^* are then found by minimizing the weighted S.S. in the case of the correlated variables.

Remarks :- (1) If the variables are uncorrelated but do not have a constant variance σ^2 then B will be a diagonal matrix with say $b_1, b_2, \ldots, b_n$ as diagonal elements. Here b_i will be proportional to the $Var(y_i)$.

The weighted S.S. in this case, in order to get LSE of β is

$$\sum_i (y_i - a_{i_1}\beta_1 - a_{i_2}\beta_2 - \cdots a_{ip}\beta_p)^2 / b_i$$

if all $\frac{1}{b_i}$ are equal i.e. case of constant variance, the model reduces original Gauss-Markov Model.

18/7/25

(1) Let y_1, y_2, y_3 be an independent r.v. with a common variance σ^2 and $E(y_1) = \theta_1 + \theta_3$, $E(y_2) = \theta_1 + 2\theta_2 + \theta_3$, $E(y_3) = \theta_1 + \theta_2$, obtain LSE of $2\theta_1 - \theta_2 - \theta_3$

→ $E(\underline{y}) = A\underline{\theta}$ (∵ Normal eq^n

$D(\underline{y}) = \sigma^2 I$ $(A'A)\hat{\underline{\beta}} = A'\underline{y}$

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

$\sum_i (y_i - a_{i1}\beta_1 - a_{i2}\beta_2 \ldots - a_{ip}\beta_p)^2 / b_i$

if all $\frac{1}{b_i}$ are equal i.e case of constant var. the model reduces original Gauss-Markov model.

$\boxed{18/7}$

① Let y_1, y_2, y_3 be a independent r.v. with a common variance σ^2 and $E(y_1) = \theta_1 + \theta_3$, $E(y_2) = \theta_1 + 2\theta_2 + \theta_3$, $E(y_3) = \theta_1 + \theta_2$, obtain LSE of $2\theta_1 - \theta_2 - \theta_3$

$\Rightarrow E(\underline{y}) = A\theta$

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

$\boxed{\begin{array}{c}(\because \text{normal eq}^n)\\ (A'A)\beta = A'\underline{y}\\ \hat{\theta} = (A'A)^{-1}A'y\end{array}}$

$$\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \quad \underline{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \quad A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix}_{3\times 3}$$

normal eq^n

$(A'A)\hat{\beta} = A'y$

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

$$= \left[\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}\right]^{-1}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$(A'A)^{-1}A'Y$

Date ____
Page ____

$$\Rightarrow \frac{1}{4}\begin{bmatrix} 6 & -2 & -4 \\ -2 & 2 & 0 \\ -4 & 0 & 6 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & -1 & 0 \end{bmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$\Rightarrow \begin{bmatrix} 3/2 & -1/2 & -1 \\ -1/2 & 1/2 & 0 \\ -1 & 0 & 3/2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$\Rightarrow \begin{bmatrix} 1/2 & -1/2 & 1 \\ -1/2 & 1/2 & 0 \\ 1/2 & 1/2 & -1 \end{bmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}$$

$$\hat{\theta}_1 = \frac{1}{2}y_1 - \frac{1}{2}y_2 + y_3$$

$$\hat{\theta}_2 = \frac{-1}{2}y_1 + \frac{1}{2}y_2$$

$$\hat{\theta}_3 = \frac{1}{2}y_1 + \frac{1}{2}y_2 - y_3$$

LSE of $2\theta_1 - \theta_2 - \theta_3$ is $2\hat{\theta}_1 - \hat{\theta}_2 - \hat{\theta}_3$

$$= 2\left(\frac{1}{2}y_1 - \frac{1}{2}y_2 + y_3\right) - \left(\frac{-1}{2}y_1 + \frac{1}{2}y_2\right) - \left(\frac{1}{2}y_1 + \frac{1}{2}y_2 - y_3\right)$$

$$= y_1 - y_2 + 2y_3 + \frac{1}{2}y_1 - \frac{1}{2}y_2 - \frac{y_1}{2} - \frac{y_2}{2} + y_3$$

$$\Rightarrow \boxed{y_1 - 2y_2 + 3y_3}$$

(2) Let $(Y_i; i = 1, 2 \ldots 6)$ be independent

H.W r.v's with common var σ^2

$E(Y_1) = E(Y_2) = \theta_1$, $E(Y_3) = E(Y_4) = \theta_2$

$E(Y_5) = E(Y_6) = \theta_1 + \theta_2$ obtain LSE of $\underline{\theta}$

$E(\underline{Y}) = A\underline{\theta}$

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix}_{6\times 1} \qquad \underline{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{bmatrix}_{6\times 1}$$

$\cancel{A'} = \begin{bmatrix} 1 & \cancel{0} & \cancel{0} & 0 & \cancel{0} & \cancel{0} \\ 0 & \cancel{1} \end{bmatrix}$

$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cc} \theta_1 & \theta_2 \end{array} \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}_{6\times 2} \end{array}$$

normal eqn

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

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

$$(A'A)^{-1} = \left[\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}\right]^{-1}$$

$$= \frac{1}{8}\begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/4 \\ -1/4 & 1/2 \end{bmatrix}$$

$$(A'A)^{-1}A'\underline{Y} = \begin{bmatrix} 1/2 & -1/4 \\ -1/4 & 1/2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix}$$

$$\hat{\theta} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{bmatrix} 1/2 & -1/4 \\ -1/4 & 1/2 \end{bmatrix}\begin{bmatrix} y_1+y_2+y_5+y_6 \\ y_3+y_4+y_5+y_6 \end{bmatrix}$$

$$\hat{\theta}_1 = \frac{1}{2}\left[(y_1+y_2+y_5+y_6) - \frac{1}{2}\left[y_3+y_4+y_5+y_6\right]\right]$$

$$\hat{\theta}_2 = \frac{1}{2}\left[-(y_1+y_2+y_5+y_6) + \frac{1}{2}(y_3+y_4+y_5+y_6)\right]$$

③ Y_1, Y_2, Y_3 are uncorrelated random variables with common variance σ^2 and $E(Y_1) = \theta_1 + \theta_2$ $E(Y_2) = \theta_1 + \theta_3$ $E(Y_3) = \theta_1 + \theta_2$ Show that $l_1\theta_1 + l_2\theta_2 + l_3\theta_3$ is estimable iff $l_1 = l_2 + l_3$

$\rightarrow$ $\underline{l} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix}$ $\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$ $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

(According to thm) $\underline{l}'\underline{\theta}$ is estimable iff $\text{rank}(A') = \text{rank}(A', \underline{l})$

or

$\text{rank}(A'A) = \text{rank}(A'A, \underline{l})$

$$A'A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$(A'A, \underline{l}) = \begin{pmatrix} 3 & 2 & 1 & l_1 \\ 2 & 2 & 0 & l_2 \\ 1 & 0 & 1 & l_3 \end{pmatrix}$$

Rank

$$3R_3 - R_1 = \begin{bmatrix} 0 & -2 & 2 & 3l_3 - l_1 \\ 2 & 2 & 0 & l_2 \\ 1 & 0 & 1 & l_3 \end{bmatrix}$$

$2R_2$

$\theta_1 + 2\theta_2 + \theta_3$
$V(y_1 + 2y_2 + 3y_3)$
$V(y_1) + 4V(y_2) + 9V(y_3)$
$\sigma^2 + 4\sigma^2 + 9\sigma^2 = 14\sigma^2$

$$2R_3 - R_2 = \begin{bmatrix} 0 & -2 & 2 & 3l_3 - l_1 \\ 0 & -2 & 2 & 2l_3 - l_2 \\ 1 & 0 & 1 & l_3 \end{bmatrix}$$

$$R_2 - R_1 = \begin{bmatrix} 0 & 0 & 0 & 2l_1 - l_2 - l_3 \\ 0 & -2 & 2 & 2l_1 - 2l_3 \\ 1 & 0 & 1 & l_3 \end{bmatrix}$$

$\text{rank}(A'A, l) = \boxed{2}$

* Simultaneous estimation of parametric functions

· Consider the general setup $(Y, X\beta, \Sigma)$, with or without restrictions on the parameter β

Let $P_1'\hat{\beta}, P_2'\hat{\beta} — P_k'\hat{\beta}$ be the individual lse of the parametric functions. $P_1'\beta, P_2'\beta — P_k'\beta$ Let A be the dispersion matrix of the estimators. $P_1'\hat{\beta}, P_2'\hat{\beta}, — P_k'\hat{\beta}$

Then we have the following optimum property of LSE

i) Let $L_1'Y, L_2'Y, --- L_k'Y$ be any unbiased estimators of $P_1'\beta, P_2'\beta — P_k'\beta$ and Let the dispersion matrix of estimators be B, ~~them~~ then B-A is non-neg

$\theta_1 + 2\theta_2 + \theta_3$

$V(y_1 + 2y_2 + 3y_3)$

$V(y_1) + 4(y_2) + 9(y_3)$

$\sigma^2 + 4\sigma^2 + 9\sigma^2 = 14\sigma^2$

$$2R_3 - R_2 = \begin{bmatrix} 0 & -2 & 2 & 3l_3 - l_1 \\ 0 & -2 & 2 & 2l_3 - l_2 \\ 1 & 0 & 1 & l_3 \end{bmatrix}$$

$$R_2 - R_1 = \begin{bmatrix} 0 & 0 & 0 & l_1 - l_2 - l_3 \\ 0 & -2 & 2 & l_1 - 2l_3 \\ 1 & 0 & 1 & l_3 \end{bmatrix}$$

$\text{rank}(A'A, l) = \boxed{2}$

* Simultaneous estimation of parametric functions

Consider the general step $(\underline{Y}, X\underline{\beta}, \Sigma)$, with or without restrictions on the parameter $\underline{\beta}$

Let $P_1'\hat{\underline{\beta}}, P_2'\hat{\underline{\beta}} \text{ --- } P_k'\hat{\underline{\beta}}$ be the individual lse of the parametric functions. $P_1'\underline{\beta}, P_2'\underline{\beta} \text{ --- } P_k'\underline{\beta}$ Let A be the dispersion matrix of the estimators $P_1'\hat{\underline{\beta}}, P_2'\hat{\underline{\beta}}, \text{ --- } P_k'\hat{\underline{\beta}}$

Then we have the following optimum property of LSE

i) Let $L_1'\underline{Y}, L_2'\underline{Y}, \text{---} L_k'\underline{Y}$ be any unbiased estimators of $P_1'\beta, P_2'\underline{\beta} \text{ --- } P_k'\beta$ and Let the dispersion matrix of estimators be B. ~~them~~ then B-A is non-negative definite implying

a) trace $B \geq$ trace A

b) $|B| \geq |A|$

c) trace $QB \geq$ trace QA where Q is non-negative definite matrix

d) Maximum latent root of $B \geq$ max latent root of A

Consider the linear parametric functions

$a_1 \underline{P_1}'\underline{\beta} + a_2 \underline{P_2}'\underline{\beta} + \_\_\_ + a_k \underline{P_k}'\underline{\beta}$ where MSE is

$a_1 \underline{P_1}'\hat{\underline{\beta}} + a_2 \underline{P_2}'\hat{\underline{\beta}} \_\_\_ a_k \underline{P_k}'\underline{\beta}$

$$V(\underline{a}'\underline{P}'\hat{\underline{\beta}}) = \underline{a}' var(\underline{P}'\hat{\underline{\beta}})\underline{a}$$

$$= \underline{a}'A\underline{a}$$

where $\underline{a}' = (a_1, a_2 \_\_\_ a_k)$

$$V(a_1 \underline{L_1}'\underline{y} + a_2 \underline{L_2}'\underline{y} + \_\_\_ + a_k \underline{L_k}'\underline{y})$$

$$= \underline{a}'B\underline{a}$$

Then for all $\underline{a}$ we have,

$$\underline{a}'B\underline{a} \geq \underline{a}'A\underline{a}$$

Unit - 2

Prerequisite for

- Consider the indep ~~ver~~ variables $y_i \sim N(x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{im}\beta_m, \sigma^2)$, $i = 1, 2, \ldots n$ —— (1)

where x_{ij} are known coefficient and β_i are unknown parameters

~~In~~

- In matrix notation $\underline{y}$ stands for col^n vector of the variables y_i, $\underline{\beta}$ for the parameters β_i and $X = (x_{ij})$ for matrix of coefficient then

$$\Sigma(y_i - x_{i1}\beta_1 - \ldots - x_{im}\beta_m)^2$$
$$= (\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})$$

Hence the pdf of $y_1, y_2 \ldots y_n$ can be written as

$$c e^{-(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})/2\sigma^2} \quad \text{—— (2)}$$

19/7

* First Fundamental thm :-

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

then $R_0^2 \sim \sigma^2 \chi^2_{(n-r)}$ 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 $M(X)$ (many fold of X)

[→ A linear subspace or a linear manifold in a vector space V is any subset of vectors M closed under addition and scalar multiplication.

if $\underline{x} \in M$, $\underline{y} \in M$ then $(c\underline{x}+d\underline{y}) \in M$ for any pair of scalars c and d. Any such subset M is itself a vector space.

→ All linear combination of a given fixed set S of vectors $\alpha_1, \alpha_2 \text{---} \alpha_k$ is a subspace called the linear manifold M(S) spanned by S.

This is the smallest subspace containing S.]

A projection is a linear transformation P from a V.S.P (vector space) to itself ∋ $P^2 = P$

i.e whenever P is applied twice to any value it gives the same result as if it were applied once.

The mapping $P: \underline{x} \mapsto \underline{x}_1$ ($P\underline{x} = \underline{x}_1$) is called the projection of $\underline{x}$ on M and P is called the projection operator.

thm Proof: But Projection of any vector on $M(X)$ is obtained through an operator a matrix P which symmetric, idempotent $(P^2 = P)$ and of rank = rank of $(X) = r$

Thus the Projection $\underline{y}$ on $M(X)$ equal $P\underline{y}$. Therefore $\underline{y} - P\underline{y}$ is perpendicular

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

chi-square dist.

$$= (\underline{y} - P\underline{y})'(\underline{y} - P\underline{y})$$

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

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

(standard normal dist)

The matrix $(I-P)$ is also idempotent and $\therefore$ rank $(I-P)$ = trace I − trace P $= n - r$

The dist^n of quadratic from

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

where, λ is non-centrality parameter

$$\lambda = E\left[\frac{\underline{y}'(I-P)\underline{y}}{\sigma^2}\right]$$

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

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

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

Projection operator

$$= \underline{\beta}'[X'X - X'PX]\underline{\beta} \quad [\because PX = X]$$
$$= 0$$

$\Rightarrow$ The dist^n of

$$\frac{Y'(I-P)Y}{\sigma^2} \sim \text{central } \chi^2 \text{ with } (n-r) \text{ d.f.}$$

Thm: ② : Second fundamental thm :

- Let H be a matrix of order $(m \times k)$ and rank $k \ni M(H) \subset M(X')$ and
$$R_1^2 = \min_{\underline{\beta}} (Y - X\underline{\beta})'(Y - X\underline{\beta})$$

s.t

$$H\underline{\beta} = \xi \text{ (given)}$$

a) R_0^2 and $R_1^2 - R_0^2$ are indepently distributed

~~b) $R_0^2 \sim \sigma^2$.~~

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

c) If $H'\underline{\beta} = \xi$ is true then

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

$$\boxed{\frac{R_1^2 - R_0^2}{k} \div \frac{R_0^2}{n-r} \sim F_{(k, n-r)}}$$

$(R_1^2 = (Y - X\beta_0)'(I - M)(Y - X\beta_0)$

Proof: If $H'\beta = \xi$ then
$\beta = \beta_0 + r$ where β_0 is a particular sol^n of $H'\beta = \xi$ and r is a general sol^n of $H'\beta = 0$

Hence,

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

$$= \min_{H'r = 0} (Y - X\beta_0 - Xr)'(Y - X\beta_0 - Xr)$$

But Xr with the restriction $H'r = 0$ is a subspace $S \subset M(X)$ with

$$\text{rank}(S) = \text{rank}[X' : H] - \text{rank}(H) = s \text{ (say)}$$

whether or not H satisfies the condition $M(H) \subset M(X')$

Let P be a operator for projection projecting $M(X)$ and U be the operator for projection projecting on S rank(P) = rank$(X) = r$

$$\text{rank}(U) = s$$

$$R_1 = (Y - X\beta_0)'(I - U)(Y - X\beta_0)$$

$$R_0^2 = y'(I-P)y$$
$$= (y - X\beta_0)'(I-P)(y - X\beta_0)$$

Introduction of the factor $X\beta_0$ in the experssion of R_0^2 does not alter its value.

— Since in R_1^2 is $(I-U)$ is idempotent.
rank $(I-U) = n-s$

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

$$\sigma^2\lambda = E[(y - X\beta_0)'(I-U)(y - X\beta_0)]$$
$$= (X\beta - X\beta_0)'(I-U)(X\beta - X\beta_0)$$
$$\Rightarrow (X\beta_0 + x\gamma - X\beta_0)'(I-U)(X\beta_0 + x\gamma - x\beta_0)$$
($\because$ If $H'\beta = \xi$ is true then $\beta = \beta_0 + \gamma$)

$$= (X\gamma)'(I-U)(X\gamma)$$
$$= \boxed{\gamma'X'X\gamma - \gamma'X'UX\gamma} \quad (\text{since } UX\gamma = X\gamma)$$
$$\Rightarrow \gamma'X'X\gamma - \gamma'X'X\gamma = 0$$

Since $R_0^2 \sim \sigma^2 \chi^2_{(n-r)}$ by 1st thm and $R_1^2 - R_0^2 \geq 0$ also $R_1^2 \sim \sigma^2 \chi^2_{(n-s)}$

$u = \chi^2_{(q-r_1)}$ $E(\underline{y}) = X'\underline{\beta}$

Date
Page

This follows $\chi^2_{(q-r)} + \chi^2_{(n-r-q+r_1)}$ (∵ treatment contrast)

$R_1^2 - R_0^2 \sim \sigma^2 \cdot \chi^2_{(r-s)}$

In the special case $M(H) \subset M(X')$ and H is of rank K.

$s = \text{rank}(X' : H) - \text{rank}(H)$
$= K$

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

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

<u>Thm</u> 3.) Let $\underline{y}' = (\underline{y}_1' : \underline{y}_2')$ with the corresponding partition of the expectation vector $\underline{\beta}' X' = \underline{\beta}'(\underline{X}_1' : \underline{X}_2') = (\underline{\beta}\underline{X}_1' : \underline{\beta}\underline{X}_2')$ then the statistic

$$u = \frac{\min\limits_{\underline{\beta}} (\underline{y}_1 - \underline{X}_1\underline{\beta})'(\underline{y}_1 - \underline{X}_1\underline{\beta})}{\min\limits_{\underline{\beta}} (\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})}$$

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

where, $\text{rank}(X_1) = r_1$ and $\text{rank}(X) = r$ and q-col^n ~~matrix~~ in X_1

25/7 * Test of hyp and interval estimation

- The LSE of a parametric funct^n is only a point estimate and ~~to~~ no

Scanned with OKEN Scanner

$H'\hat{\beta}_1 ; H'\hat{\beta}_2 \ldots H'\hat{\beta}_k$ $H_0 : \mu = 0$

$(\hat{\beta}) = \beta$ $H'\underline{\beta} = H'\hat{\beta}$ large $n > 0$ σ is known

$(H'\beta) = H' Var(\beta) H$

exact statement of probability of its deviation from the true value of the parametric funt^n can be made without a specific dist^n for the variable $\underline{y}$ being considered we shall assume that

$$\underline{y} \sim N(X\beta, \sigma^2 I)$$

Let H' be matrix of order $k \times m$ of row k 'k' such that its rows dependes in the rows of X

$\Rightarrow$ K parametric funcs $H'\underline{\beta}$ are individually estimable under the set up $(\underline{y}, X\underline{\beta}, \sigma^2 I)$ in which case the ~~Ros~~ LSE are $H'\hat{\underline{\beta}} = Z$

In terms of X, Y we have

$$\boxed{H'\hat{\underline{\beta}} = H'(X'X)^{-1}X'\underline{y}}$$

$$R_0^2 = \text{error S.S} = \underline{y}'(I - (X'X)^{-1}X')\underline{y}$$

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

$$= (\underline{y} - X(X'X)^{-1}X'\underline{y})'(\underline{y} - X(X'X)^{-1}X'\underline{y})$$

$$= \underline{y}'(I - X(X'X)^{-1}X')$$

$$\underline{y}(I - X(X'X)^{-1}X'\underline{y}$$

$V(H'\underline{\beta}) = H' var(\underline{\beta}) H$ $\quad E(\hat{\beta}) = \beta$ $\quad (\because E(H'\underline{\hat{\beta}}) = H'E(\underline{\hat{\beta}}) = H'\underline{\beta})$

Idempotent ~~matrix ($A'A = A$)~~

$= \underline{y}'(I - X(X'X)^{-1}X')\underline{y}$

then $H'(X'X)^{-1}X'(I - X(X'X)^{-1}X')$

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

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

$= 0$

$\Rightarrow H'\hat{\underline{\beta}}$ and R_0^2 are independently distributed

The distn of z is a K-variate normal with mean $H'\underline{\beta}$ and dispersion matrix $\sigma^2 D$ (say) and $R_0^2 \sim \sigma^2 \chi^2_{(n-r)}$

$\therefore$ i.e $H'\hat{\underline{\beta}} \sim N(H'\underline{\beta}, \sigma^2 D)$

$(\because E(H'\hat{\underline{\beta}}) = H'E(\hat{\beta}) = H'\underline{\beta}$

$Var(H'\hat{\underline{\beta}}) = H'Var(\hat{\underline{\beta}})H = \sigma^2 D$ $\quad$ as $H'H = D)$

* Single Parametric function

- Given an estimable parametric func $\theta = P\underline{\beta}$ Let u be its LSE with var $P^2\sigma^2$ then $u \sim N(\theta, P^2\sigma^2)$ and $R_0^2 \sim \sigma^2\chi^2_{(n-r)}$ and are independent so that $s^2 = \frac{R_0^2}{(n-r)}$

then,

$$t = \frac{u-\theta}{P.\sigma} \div \frac{s}{\sigma} = \frac{u-\theta}{Ps} \sim S_{(n-r)} \quad \text{--- (i)}$$

which is student's $dist^n$ on $(n-r)$ df

If t_α is the α prob Point of $|t|$

i.e $P\left[|t| > t_\alpha\right] = \alpha$ then

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

i.e

$$P\left[-t_\alpha \le \frac{(u-\theta)}{Ps} \le t_\alpha\right] = 1-\alpha$$

$$\Rightarrow P\left[u - Pst_\alpha \le \theta \le u + Pst_\alpha\right] = 1-\alpha \quad \text{--- (iii)}$$

Let the Null hyp. be $H_0 : P\beta = \theta_0$ (an assigned value)

Null hyp is Rejected at α level of significance if for an observed u we have

$$\left|\frac{u-\theta_0}{Ps}\right| > t_\alpha \quad \text{--- (iv)}$$

The Power of the test is given by $(1-\beta)$ i.e. $P\left[|t| > t_{\alpha/2}, n-r / H_1\right]$ Here, effect size is based on Non-centrality Parameter

$\theta = \underline{P}'\underline{\beta}$
$H_0 : \mu = 1$
$H_1 : \mu = 2$
$\mu \neq 1$

t has non-central t-dist^n with λ as non-central parameter. The power of t is monotonically increasing funct^n of $|\lambda|$.

~~λ = value of $\underline{P}'\underline{\beta}$ under H_1~~ (1)

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

26/7

If $Z \sim N(0,1)$ ($\because$ Z and V are indep)
$V \sim \chi^2_{(\nu)}$
then non-central 't' is given by

$$\boxed{t_{\nu,\delta} = \frac{Z+\delta}{\sqrt{\frac{V}{\nu}}}}$$ δ is n.c Parameter

e.g. Suppose we want to $H_0 : \mu = 70$
$H_1 : \mu > 70$
for $n = 25$, $\alpha = 0.05$
find power of 't' when $\sigma = 10$

$$\delta = \frac{75-70}{\sigma/\sqrt{n}} = \frac{5}{10/5} = \boxed{2.5}$$

effect size d is given by
$\delta = d \cdot \sqrt{n}$ $(\delta/\sqrt{n})$

Statistics kingdom
Power T2 calculator

R → PWR
Pakage

$Y = X\beta + \epsilon$

$(Y - X\beta)'(Y - X\beta)$
$= \epsilon'\epsilon$

$d = \frac{2.5}{5} = 0.5$

$\Rightarrow \boxed{d = \frac{\mu - \mu_0}{\sigma}}$

(1) Let $y_1, y_2 \ldots y_n$ be indep obs from $N(\mu, \sigma^2)$ test the hyp $H_0 : \mu = \mu_0$ (given)

In this case

$X\beta$ = is the μ

$$SSE = \min_{\mu} \Sigma (y_i - \mu)^2$$

$$\neq \sum_{i=1}^{n} \cancel{(y_i - \bar{y})^2} = \cancel{\sum_{i=1}^{n} (y_i - \bar{y})}$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \bar{y})^2 \text{ has } (n-1) \text{df}$$

The normal eq^n is

$$\Sigma y_i = n\mu$$

$$\Rightarrow \hat{\mu} = \frac{\Sigma y_i}{n} = \boxed{\bar{y}}$$

Also $V(\bar{y}) = V\left(\frac{\Sigma y_i}{n}\right) = \frac{1}{n^2} \Sigma V(y_i)$

$$= \frac{\Sigma \sigma^2}{n^2} = \frac{n\sigma^2}{n^2}$$

$$= \boxed{\frac{\sigma^2}{n}}$$

Thus appropriate test statistic for testing H_0 is

$$t = \frac{(\bar{y} - \mu_0) \cdot \sqrt{n}}{\sqrt{MSSE}}$$ which is t

Scanned with OKEN Scanner

with $(n-1)$ df

Reject H_0 if
$|t|_{cal} > t_{tab}$ where
$t_{tab} = t_{(\alpha/2, n-1)}$

The Power of the t-test w.r.t
$H_1 : \mu = \mu_1$ is

$$P\left[\frac{\sqrt{n}\,|\bar{y} - \mu_0|}{\sqrt{MSSE}} > t_{(\alpha/2, n-1)} \,\middle|\, H_1 \right]$$

where t has non central t-dist^n with n-c parameter

$$\boxed{\lambda = \frac{|\mu_1 - \mu_0|\sqrt{n}}{\sigma}}$$

eg ~~The~~ Let $y_1, y_2 \text{---} y_{n_1}$ be a r.s from $N(\mu, \sigma^2)$ and $y_1', y_2' \text{---} y_{n_2}'$ be a r.s from $N(\mu_2, \sigma^2)$ the two samples being mutually indep Test the hyp
$H_0 : \mu_1 = \mu_2$ $\quad (\because H_0 : P'\beta = 0,\ \mu_1 - \mu_2 = 0)$

$E(y_i) = \mu_1 \quad i = 1, 2 \text{---} n$
$E(y_i') = \mu_2 \quad i = 1, 2 \text{---} n$

$$\boxed{SSE = \lim_{\mu_1, \mu_2} \left[\sum_{i=1}^{n_1} (y_i - \mu_1)^2 + \sum_{i=1}^{n_2} (y_i' - \mu_2)^2 \right]}$$

The normal eq^n are

Date
Page

$$n_1 \mu_1 = \Sigma y_i = n_1 \bar{y}$$
$$n_2 \mu_2 = \Sigma y_i' = n_2 \bar{y}'$$

Hence

$$SSE = \sum_{i=1}^{n_1} (y_i - \bar{y})^2 + \sum_{i=1}^{n_2} (y_i' - \bar{y}')^2$$

with (n_1+n_2-2) d.f

Variance for use of parametric $(\mu_1-\mu_2)$ func is

$$V(\bar{y}-\bar{y}') = var(\bar{y}) + var(\bar{y}')$$
$$= \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}$$
$$= \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Hence the appropriate 't' statistic is

$$\cancel{t = \frac{\bar{y}-\bar{y}'}{\sqrt{\left(\frac{1}{n_1}\right)+\frac{1}{(n_2}}}}$$

$$t = \frac{\bar{y}-\bar{y}'}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right) MSSE}}$$

with d.f (n_1+n_2-2)

(∴ Fisher's t)

1/8

* More than one Parametric func

- Let us consider 'K' indep estimable linear parametric func θ_1 $\theta_i = H_i'\beta$, $\theta_2 = H_2'\beta$, ..., $\theta_K = H_K'\beta$
In matrix notation it may be written as

$$\underline{\theta} = H'\underline{\beta}$$

where H' is a $K \times m$ matrix and $\underline{\theta}$ is the col^n vector $\theta_1, \theta_2 — \theta_K$

- The LSE of $(\theta_1, \theta_2 — \theta_K)$ are represented by $(z_1, z_2 — z_K) = z'$ and its dispersion matrix by $\sigma^2 D$ then

$$E(\underline{z}) = \underline{\theta}, \quad D(\underline{z}) = \sigma^2 D$$

We have

$$(\underline{z} - \underline{\theta})' D^{-1} (\underline{z} - \underline{\theta}) \sim \sigma^2 \chi^2_{(K)} \text{ d.f.}$$

and

$$R_0^2 \sim \sigma^2 \chi^2_{(n-r)} \text{ are indep}$$

Hence

$$\boxed{F = \frac{(\underline{z} - \underline{\theta})' D^{-1} (\underline{z} - \underline{\theta})}{K} \div \frac{R_0^2}{(n-r)} \sim F_{(K, n-r)}}$$

* Test of multiple hypothesis *

(ANOVA)

— Let it be required to test the null hyp. that 'k' parametric func have the assigned values as

$$H_0 : \underline{H}_1'\beta = \theta_{10}, \underline{H}_2'\beta = \theta_{20} \ldots H_k'\beta = \theta_k$$

which may be written in matrix notation as

$$\boxed{H'\beta = \theta_0}$$

Let $Z_1, Z_2 \ldots Z_k$ be LSE of $\underline{H}_1'\beta, \underline{H}_2'\beta, \ldots H_k'\beta$.

$(Z - \theta_0)$ is the vector of deviations Defⁿ l.s.e and assigned values If infact θ_0 is not true the deviations $(Z - \theta_0)$ are likely to be large Let us consider the compound deviation (a single measure of deviations)

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

Now

$$\frac{E[(Z - \theta_0)' D^{-1} (Z - \theta_0)]}{K}$$

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

$\boxed{D \text{ matrix} \geq 0}$

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

$$\left(\because E(\underline{z}) = \underline{\theta} = \underline{H}'\underline{\beta} \right.$$
$$\left. D(\underline{z}) = \sigma^2 D \right)$$

$= \sigma^2$ if the null hyp is true i.e $\underline{H}'\underline{\beta} = \theta_0$

$> \sigma^2$ if the null hyp is not true

(Since $(\underline{H}'\underline{\beta} - \theta_0)' D^{-1} (\underline{H}'\underline{\beta} - \theta_0)$ is Positive definite)

Also $E(S^2) = \frac{R_0^2}{n-r} = \sigma^2$

Hence the test statistics becomes

$$\boxed{F_{cal} = \frac{(\underline{z}-\theta_0)' D^{-1} (\underline{z}-\theta_0)}{K} \div \frac{R_0^2}{(n-r)}}$$

Reject H_0 if

$F_{cal} > F_{tab}$

where $F_{tab} = F_{(K, n-r)}$

* Simultaneous C.I

It follows that

$$P\left\{\frac{(\underline{z}-\underline{\theta})'D^{-1}(\underline{z}'-\underline{\theta})}{ks^2} < F_\alpha\right\} = 1-\alpha \quad (1)$$

where $\underline{\theta} = H'\beta$ stands for 'K' indep estimable parametric functions of β

Expression (1) is called $1-\alpha$ confidence region of $\underline{\theta} = H'\beta$

Let it be represented by C. Then simultaneous C.I for funcs $g_i(\underline{\theta})$ $i=1,2 \ldots K$ with confidence coefficients possible greater than $1-\alpha$ are given by

$$\left[\min_{\underline{\theta}\in C} g_i(\underline{\theta}),\ \max_{\underline{\theta}\in C} g_i(\underline{\theta})\right], \quad i=1,2 \ldots K$$

For any particular $g(\underline{\theta})$

$$P\left[\min_{\underline{\theta}\in C} g(\underline{\theta}),\ \max_{\underline{\theta}\in C} g(\underline{\theta})\right] \geq 1-\alpha$$

* Second Method

(Using Cauchy - Schwartz inequality

- If u and A are Col^n vectors and

B is a positive definite matrix then

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

Let $\underline{A} = (\underline{z} - \underline{\theta})$; $B = D$

then we have

$$\frac{(\underline{z} - \underline{\theta})' D^{-1} (\underline{z} - \underline{\theta})}{KS^2} = \frac{1}{KS^2} \underset{\underline{u}}{Max} \frac{[\underline{u}'(\underline{z} - \underline{\theta})]^2}{\underline{u}' D \underline{u}} \quad \text{—(I)}$$

then using (I) we have

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

$$\therefore P\left\{ \sqrt{\frac{(\underline{z} - \underline{\theta})' D^{-1} (\underline{z} - \underline{\theta})}{KS^2}} \leq \sqrt{F_\alpha} \right\} = 1 - \alpha$$

$$P\left\{ Max \frac{|\underline{u}(\underline{z} - \underline{\theta})|}{\sqrt{KS} \cdot \sqrt{\underline{u}' D \underline{u}'}} \leq \sqrt{F_\alpha} \right\}$$

$\Rightarrow$ ~~$P\{|\underline{u}'(\underline{z}$~~

$$\Rightarrow P\{ |\underline{u}'(\underline{z} - \underline{\theta})| \leq S\sqrt{KF_\alpha \underline{u}' D \underline{u}} \text{ for all } \underline{u} \} = 1 - \alpha$$

LSD ⟶ Part of Post Hoc test

$$\Rightarrow P\{-S\sqrt{KF_\alpha \underline{u}'D\underline{u}} \le \underline{u}'(\underline{z}-\underline{\theta}) \le S\sqrt{KF_\alpha \underline{u}'D\underline{u}}\} = 1-\alpha$$

$$\Rightarrow P\{-S\sqrt{KF_\alpha \underline{u}'D\underline{u}} - \underline{u}'\underline{z} \le \underline{u}'\underline{\theta} \le S\sqrt{KF_\alpha \underline{u}'D\underline{u}} - \underline{u}'\underline{z}\} = 1-\alpha$$

$$\Rightarrow P\{\underline{u}'\underline{z} + S\sqrt{KF_\alpha \underline{u}'D\underline{u}} \ge \underline{u}'\underline{\theta} \ge \underline{u}'\underline{z} - S\sqrt{KF_\alpha \underline{u}'D\underline{u}}\} = 1-\alpha$$

$$\Rightarrow P\{\underline{u}'\underline{z} - S\sqrt{KF_\alpha \underline{u}'D\underline{u}} \le \underline{u}'\underline{\theta} \le \underline{u}'\underline{z} + S\sqrt{KF_\alpha \underline{u}'D\underline{u}}\} = 1-\alpha$$

($100(1-\alpha)\%$ C.I for

$$\boxed{\underline{u}'\underline{\theta} = \underline{u}'H'\underline{\beta}}$$

* Tukey's test :-

- Suppose that following and analysis of variance in which we have rejected the null hyp. of equal treatment means (For one way ANOVA) we wish ~~to~~ test all pairwise mean comparison

$H_0 : \mu_i = \mu_j$

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

- Tukey propose a procedure for testing hyp. for which overall significance level is exactly α when sample sizes are equal and at the most α when the sample sizes are unequal.

- The procedure can also be used to construct confidence intervals on the differences in all pairs of means for this intervals the simultaneously confidence level $100(1-\alpha)\%$ when the sample sizes are equal and atleast $100(1-\alpha)\%$ when the sample sizes are not equal.

- Tukey procedure makes use of the distribution of students ties their range statistic

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

When $\bar{y}_{max} \leftarrow$ largest sample mean
$\bar{y}_{min} \leftarrow$ smallest sample mean
out of a group of p samples.

$q_\alpha(P,f)$ is the upper α percentage points of P and f is the no. of d.f associated with MSSE

- For equal sample sizes, Tukey's test declares two mean significantly different if the absolute value of their sample difference exceeds T_α.

$$|\bar{y}_{i} - \bar{y}_j| > T_\alpha$$

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

$100(1-\alpha)\%$ C.I for all pairs of means are given by

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

$$i \neq j$$

When sample sizes are not equal then,

$T_\alpha = q_\alpha$

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

and $100(1-\alpha)\%$ C-I is given by

$$(\bar{y}_{i.} - \bar{y}_{j.}) - \frac{q_\alpha(P,f)}{\sqrt{2}}\left(\sqrt{MESS\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}\right)$$

$$\leq \mu_i - \mu_j$$

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

$i \neq j$

unequal sample size version is sometimed called Tukey-Kramer procedure.

28

* Scheffe's test (for any type of treat contrast)

- In many situations, 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 may exploratory experiment

Date
Page

the comparision of interest are discovered only after preliminary examination of the data.

- Scheffe's has proposed a method for comparing any and/or possible contrast between treatment means.
- In the scheffe's method, Type-I error is atmost α for any of the possible comparision.

- ~~Suppose that $\forall$ a set~~

- Suppose a set of 'm' contrast in the treatment means

$$\Gamma_u = C_{1u}\mu_1 + C_{2u}\mu_2 + \text{---} + C_{au}\mu_a \quad , u = 1, 2 \text{---} m$$

of interest have been determined

- The corresponding contrast in the treatment average $\bar{y}_{i.}$ is

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

and the S.E of the contrast is

$$S_{C_u} = \left[\sqrt{MSSE \sum_{i=1}^{a} \left(c_{iu}^2 / n_i\right)}\right]$$

$$S_{Cu} = \sqrt{MSE \sum_{i=1}^{a} \left(c_{iu}^2 / n_i \right)}$$

Where n_i is the no. of obs in the i^{th} 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 hyp that the contrast $\gamma_u = 0$ is rejected.

- The scheffe's Procedure can also be used to from C.I for all possible contrast among treatment means the resulting interval.

$$C_u - S_{\alpha,u} \le \Gamma_u \le C_u + S_{\alpha,u}$$ are simultoneous C.I.

- Although for pair-wise treatment comparison scheffe's method can be applied but it is not the most sensitive procedure for such comparision.

e.g

% weight of cotton	Observed tensile strength (lb/in^2) 1	2	3	4	5	Total $y_{i.}$	Avg $\bar{y}_{i.}$
15	7	7	15	11	9	49	9.8
20	12	17	12	18	18	77	15.4
25	14	18	18	19	19	88	17.6
30	19	25	22	19	23	108	21.6
35	7	10	11	15	11	54	10.8

$$SS_{Total} = \sum\sum y_{ij}^2 - \frac{y_{..}^2}{N} = 6292 - \frac{141376}{25}$$

$$= 6292 - 5655.04$$

$$= \boxed{636.96}$$

$$SS_{treat} = \frac{1}{n}\sum y_{i.}^2 - \frac{y_{..}^2}{N} = 6130.8 - 5655.04$$

$$= \boxed{475.76}$$

$$SS_E = SS_{total} - SS_{treat}$$

$$= \boxed{161.2}$$

hypothesis

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$

$H_1 :$ Atleast one $\mu_i \neq 0$

* ANOVA

Source	d.f	S.S	MESS	Fcal	Ftab
treat	$a-1=4$	475.76	118.94	14.76	$F_{(4,20,0.05)} = 2.87$
error	20	161.2	8.06		
total	$N-1=24$	636.96			

conclusion :- Fcal > Ftab $(14.76 > 2.87)$ we reject our H_0 at 5% level of significant.

Date
Page

Hypothesis : $H_0 : \mu_i = \mu_j$
$H_1 : \mu_i \neq \mu_j$

$$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{\frac{MSSE}{n}}} = \frac{11.8}{1.270} = \boxed{9.291}$$

$$T_\alpha = q_\alpha(p, f)\sqrt{\frac{MSSE}{n}}$$
$$= q_{0.05}(5, 20)\sqrt{\frac{8.06}{5}}$$
$$= 4.23 \times 1.270 = \boxed{5.3721}$$

(*significantly different)

$\bar{y}_{1.} - \bar{y}_{2.} = +5.6$ *	$\bar{y}_{2.} - \bar{y}_{4.} = +6.2$ *
$\bar{y}_{1.} - \bar{y}_{3.} = +7.8$ *	$\bar{y}_{2.} - \bar{y}_{5.} = 4.6$
$\bar{y}_{1.} - \bar{y}_{4.} = +11.8$ *	$\bar{y}_{3.} - \bar{y}_{4.} = +4$
$\bar{y}_{1.} - \bar{y}_{5.} = +1$	$\bar{y}_{3.} - \bar{y}_{5.} = 6.8$ *
$\bar{y}_{2.} - \bar{y}_{3.} = +2.2$	$\bar{y}_{4.} - \bar{y}_{5.} = 10.8$ *

Test the contrast

$$\gamma = \mu_1 + \mu_3 - \mu_4 - \mu_5$$
$$\Rightarrow \mu_1 + \mu_3 = \mu_4 + \mu_5$$

$H_0 : \gamma = 0$
$H_1 : \gamma \neq 0$

$$C_u = C_{1u}\bar{y}_{1.} + C_{2u}\bar{y}_{2.} + \dots + C_{au}\bar{y}_{a.}$$
$$C_1 = \bar{y}_{1.} + \bar{y}_{3.} - \bar{y}_{4.} - \bar{y}_{5.}$$
$$= -5$$

Date
Page

$$S_{c_1} = \sqrt{MSSE \sum_{i=1}^{a} \left[(C_{iu})^2 / n \right]}$$

$$= \sqrt{8.06 \times \left(\frac{1+1+1+1}{5} \right)}$$

$$= \sqrt{6.448} = \boxed{2.54}$$

$$S_{\alpha;u} = S_{c_1} \sqrt{(a-1) F_{\alpha, a-1, N-a}}$$

$$= 2.5 \sqrt{4 \times 2.87} = 2.5 \times 3.39$$

$$= \boxed{8.47}$$

Test criteria: $|C_1| > 8.47$

<u>Conclusion</u>: since $|C_1| \not> S_{\alpha;u}$ $(5 \not> 8.47)$ we do not reject H_0 at 5% Level of significiance Hence we conclude that $\gamma \neq 0$

$(\mu_1 + \mu_3 = \mu_4 + \mu_5)$

*** <u>Note</u> :-**

- If multiple test are performed on a given sample the value of α changes since sampling distribution for it assumes only one t-test from any given sample.

- The true α level given multiple test or comparision can be estimate as $1-(1-\alpha)^c$ where,

c = total no. of comparision, contrast or test performed

(e.g $\alpha = 0.05$, $c = 10$
$1-(1-0.05)^{10} = \boxed{0.4}$)

* Power of F-test :-

- OC-curve is a plot of Type-II error probability of a statistical test for a particular sample size versus a parameter that ~~reflects~~ reflects the extend to which the null hyp. is False.

- This curve can be used to ~~go~~ guide the experimenta in selecting the no. of replicates so that the design will be sensitive to imp. potential differences in the treatment.

- We consider the prob. of Type-II error of the fixed effect model for the case of equal sample sizes per treatment.

$P_i = 1 - P\{\text{Reject } H_0 / H_0 \text{ is false}\}$

$= 1 - P\{F_{cal} \geq F_{\alpha, a-1, N-a} / H_0 \text{ is false}\}$ ——(*)

To evaluate the prob we need to know the $dist^n$ of the test statistics F_{cal} if null hyp. is false.

— If H_0 is false the statistic

$$\boxed{F_{cal} = \frac{MSS_{treat}}{MSS_E} \sim \text{non-central } F}$$

with $(a-1, N-a)$ d.f and non-centrality parameter δ, if $\delta = 0$ then non-central F becomes usual (central) F $dist^n$.

— OC curve are used to evaluate the probability statement in the (*)

— This curves plot the Prob. of Type-II error (β) against a parameter Φ where,

$$\boxed{\Phi = \frac{n\sum_{i=1}^{a} \tau_i^2}{a\sigma^2}}$$ (∵ a = no. of treatment)

The quantity Φ is related to non-centrality parameter ~~to~~ δ.

curves are available for $\alpha = 0.5$ and $\alpha = 0.01$

- One way to determine Φ is to choose the actual values of the treatment means for which we would like to reject the null hyp. with high Prob.

- Thus if $\mu_1, \mu_2 \ldots \mu_a$ are the specified treatment means we find τ_i as

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

- We also required and estimate of σ^2 sometimes this is available from prior experience or preliminary test or a judgement ~~the exp.~~ estimate an alternate approach is to select a sample size such that the difference between any two treatments means if it exceeds a specified value the null hyp. should be rejected.

- If difference between any two treatment means is as large as

1) then $\boxed{\Phi = \frac{nD^2}{2a\sigma^2}}$

22/8

Unit - 3

If no. of obs^n in each cell are same → Balanced layout.

Random effect Model :-

- If the effects in the linear model are all random variables except for the additive constant if any which is a fixed quantity is called a random effect model.

- The effects become random give to the sampling of the levels of the factors included. It is called Variance-component Model.

- Thus we have variance-component model for one-way layout, two-way layout or multi-way layout. The complete p-way classification or layout is called Balanced if the number of obs^n in the different cells are equal.

- The one-way classification is balanced if no. of obs^n under the categories are same.

* General Random effect Model in the balanced case.

- Let an observable r.v. $Y_{ijk \ldots m}$ for a balanced case be such that

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

where μ is a constant the r.v's $a_i, b_{ij}, c_{ijk} \ldots e_{ijk\ldots m}$ are completely indep and $a_i \sim N(0, \sigma_a^2)$

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

$c_{ijk} \sim N(0, \sigma_c^2)$

$\vdots$

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

$E(y_{ijk\ldots m}) = \mu$

$D(y_{ijk\ldots m}) = var(\mu + a_i + b_{ij} \ldots e_{ijk\ldots m})$

$= \sigma_a^2 + \sigma_b^2 + \ldots \sigma_e^2$

- Diagonal elements of the dispersion matrix of y's are all same then $\sigma_a^2, \sigma_b^2 \ldots \sigma_e^2$ are the component of the variance of the obs^n. So there are ~~class~~ ~~ctck~~ called variance - component and hence the model is called variance - component model.

- The dispersion matrix $(y_{ijk\ldots m})$ is positive definite if it satisfies the condition of the model (1)

Source
1
2
$\vdots$
K

Thm

$a_1, a_2, a_3 \sim N(0, \sigma_a^2)$

$b_1, b_2, b_3 \sim N(0, \sigma_b^2)$

ANOVA (table 1)

Source	d.f	S.S	M.S.S	E(MSS)
1	f_1	S_1	s_1^2	σ_1^2
2	f_2	S_2	s_2^2	σ_2^2
⋮	⋮	⋮	⋮	⋮
k	f_k	S_k	s_k^2	σ_k^2
	$\sum_{i=1}^{k} f_i$	$\sum S_i$		

Generally each σ_i^2 will be a linear function a variance-component $\sigma_a^2, \sigma_b^2, \sigma_e^2$

Thm (1): Let the model be (1) and be balanced and the appropriate ANOVA tables be given as table(1) then,

$$\frac{S_1}{\sigma_1^2}, \frac{S_2}{\sigma_2^2} \text{ --- } \frac{S_k}{\sigma_k^2} \text{ are mutually}$$

independently distributed as central χ^2 with f_1, f_2 --- f_k df resp. Since σ_i^2 are linear funcs of the variance components we can obtain unbiased estimators of the variance components by equating each E(MSS) to the corresponding M.S.S of the table(1)

* These estimators have the following properties

Thm(2): Let the model be (1) and be balanced then the estimators of the variance components obtained by equating each E(MSS) to the corresponding MSS of table (1) are minimum variance unbiased.

Thm(3): Let the model be (1) and be balanced with the corresponding ANOVA given in table (1) then the test of

$H_0 : \sigma_i^2 = \sigma_j^2 \ (i \neq j)$

$H_1 : \sigma_i^2 > \sigma_j^2$ is given by

$$F = \frac{s_i^2}{s_j^2} \sim F_{central} \text{ with } (f_i, f_j) \text{ df}$$

* Power

- Power of the test H_0 depends on $H_1 : \frac{\sigma_i^2}{\sigma_j^2} = \lambda < 1$ and is given by

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

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

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

where f is the pdf of central F dist^n with (f_i, f_j) df.

- Thus in the variance-component model, both the test and the power are given by central F-dist^n whereas in fixed effect model power is given by non-central F-dist^n.

* One-way classification : random effect Model

- Consider the following balanced one way classification random effect Model

$$y_{ij} = \mu + a_i + e_{ij} \,, \quad i = 1, 2 \ldots p$$
$$j = 1, 2 \ldots n$$

where a_i and e_{ij} are completely indep and $a_i \sim N(0, \sigma_a^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$, μ is additive constant

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

$var(a_i + e_{io})$
$= var(a_i) + var(e_{io})$
$= var(a_i) + var\left(\sum \frac{e_{ij}}{n}\right)$
$= \sigma_a^2 + \frac{1}{n^2} \sum var(e_{ij})$
$= \sigma_a^2 + \frac{n}{n^2}\sigma_e^2$
$= \sigma_a^2 + \frac{1}{n}\sigma_e^2$
$= \sigma_a^2 + \sigma_e^2/n$

To test

$H_0 : \sigma_a^2 = 0$ ag $H_1 : \sigma_a^2 > 0$

which means testing the homogeneity of the effects of the classification

we have,

$$\sum_i \sum_j (y_{ij} - y_{oo})^2 = \sum_i \sum_j (y_{ij} - y_{io} + y_{io} - y_{oo})^2$$

$$= \sum_i \sum_j (y_{ij} - y_{io})^2 + \sum_i \sum_j (y_{io} - y_{oo})^2$$

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

$$= SS(\text{bet}^n \text{ classes}) + SS(\text{within classes})$$

$$= SSB + SSE$$

Let $y_{io} - y_{oo} = \mu + a_i + e_{io} - \mu - a_o - e_{oo}$

$$= a_i + e_{io} - a_o - e_{oo}$$

where $a_o = \frac{\sum a_i}{p}$, $e_{io} = \sum_j \frac{e_{ij}}{n}$

$$e_{oo} = \sum_i \frac{e_{io}}{p}$$

Thus $SSB = n \sum_i (a_i + e_{io} - a_o - e_{oo})^2$

where $a_i + e_{io} \sim N(0, \sigma_a^2 + \sigma_e^2/n)$

So $\dfrac{n \sum_{i=1}^{p} (a_i + e_{io} - a_o - e_{oo})^2}{(n\sigma_a^2 + \sigma_e^2)} \sim \chi^2_{(p+1)} \text{ df.}$

i.e 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

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

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

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

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

So for different i

$\sum_{j=1}^{n} \frac{(e_{ij} - e_{io})^2}{\sigma_e^2}$ are independently distributed $\chi^2_{(n-1)}$

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

* <u>To show SSB and SSE are indep.</u>

We first show that

$u_i = a_i + e_{io} - a_o - e_{oo}$ and

$v_{ij}' = e_{ij} - e_{io}$

$e_{ij} \sim N(0, \sigma_e^2)$

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

$\Rightarrow V(e_{ij}) = \sigma_e^2 = E[e_{ij}^2]$

have zero cov.

$$cov(u_i, v_{ij}') = E[u_i, v_{ij}'] - E[u_i]E[v_{ij}']$$

$\cancel{= E\{u}$

$$\left(\begin{array}{l} \because E[u_i] = a_i - a_0 \\ E[v_{ij}'] = 0 \end{array}\right)$$

$$cov(u_i, v_{ij}') = E[u_i, v_{ij}']$$

$$E[u_i v_{ij}'] = E[\{(a_i - a_0) + (e_{i0} - e_{00})\}\{(e_{ij} - e_{i0})\}]$$

$$= E[\{(a_i - a_0)(e_{ij}' - e_{i0}')\} + \{(e_{i0} - e_{00})(e_{ij}' - e_{i0}')\}]$$

(Since a_i and e_{ij} are indep

$$E[(a_i - a_0)(e_{ij}' - e_{i0}')] = 0 \,)$$

$$= E[(e_{i0} - e_{00})(e_{ij}' - e_{i0}')]$$

$$= E[e_{i0}e_{ij}' - e_{i0}e_{i0}' - e_{00}e_{ij}' + e_{00}e_{i0}']$$

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

$$\left(\because \text{where } \begin{array}{l} \delta_{ii'} = 1 \text{ if } i = i' \\ \delta_{ii'} = 0 \text{ if } i \neq i' \end{array}\right)$$

$$cov = \boxed{0}$$

Scanned with OKEN Scanner

$_{00})\}\{(e_{ij} - e_{i0})\}]$

$+\{(e_{i0} - e_{00})(e_{ij}' - e_{i0}')\}]$

indep

$= 0$

$_{i'0})]$

$' + e_{00}e_{i0}']$

in

source

Bet^n classes

within classe

Total

The

$\Rightarrow u_i$ and v_{ij}' are indep.

Thus SSB and SSE being funcs of indep quantities are also indep

Hence

$$\frac{SSE}{\sigma_e^2} \sim \chi^2_{p(n-1)} \text{ and}$$

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

independently distributed.

ANOVA

Source	d.f	S.S	MSS	E(MSS)
Bet^n 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 S.S		

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

$$F_{cal} = \frac{MSSB}{MSSE}$$

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

The power of the test against

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

$$\beta(\lambda) = \int_{\lambda \cdot F(\alpha, (p-1), n(p-1))} f(u)\,du$$

where f is p.d.f of central F-dist^n.

16/9

* Residual Analysis :-

- The nature assumptions that we have made in study of regression analysis are as follows:

(i) The relationship between the response y and the regressors is linear atleast approximate.

ii) The error term ϵ has zero mean

iii) The error term ϵ has constant var σ^2

iv) The error are uncorrelated

v) The error are normally distributed

Assumption (iv) and (v) imply that the errors are independent r.v's

Assumption (v) is required for hypothesis testing and interval estimation.

- Gross violations of the assumption may yield an unstable model. In the sense that a different sample could lead to a totally different model with oppositive conclusions.

- We usually cannot detect departure form the underline assumption by examination of the standard summary statistic such as t- or F or r^2. This are global model properties and as such they do not ensure model addequacy.

* **Def^n of residuals:-**

- $e_i = y_i - \hat{y}_i \quad , \quad i = 1, 2 \ldots n$

  $y_i \leftarrow$ observed value

  $\hat{y}_i \leftarrow$ fitted value

- Since a residual may be viewed as deviation bet^n the observed and the fitted value It is also a measure of the variability in the response variable not explain by the regression model. Thus, any departure from the assumption on the errors should show up in the residuals.

* Properties of Residuals

i) They have zero mean

$$\bar{e} = \frac{\sum e_i}{n} = \frac{0}{n} = 0$$

ii) Approximate average variance is estimated by

$$\frac{\sum (e_i - \bar{e})^2}{n-p} = \frac{\sum e_i^2}{n-p} = \frac{SSE}{n-p} = MSE$$

iii) Residual are not independent. However, as the n residuals have only n-p degrees of freedom associated with them. This non-independence of residuals has little effect on their use for model adequacy checking as long as n is ~~not~~ small relative to the number of parameters p.

* Methods of scaling residuals:

- Sometimes, it is useful to work with scaled residuals. This scaled residuals are helpful ~~in finding~~ in finding obs^{ns} that are outliers or extreme values.

$$\hat{\beta} = (X'X)^{-1}X'\underline{y} = H\underline{y}$$

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

* **standardized residuals**

- It is given by

$$d_i = \frac{e_i}{\sqrt{MSSE}}, \quad i = 1, 2 -- n$$

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

If $d_i > 3$ it indicates outliers

* **Studentized residuals :-**

- It is given by

$$r_i = \frac{e_i}{\sqrt{MSSE(1-h_{ii})}}, \quad i = 1, 2 - n$$

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

h_{ii} is the measure of the location of i^{th} point in X-space.

The studentized residual have constant variance.

$$var(r_i) = 1$$

regardless of the location X_i when the form of the model is correct in many situations the variance of the residuals stabilizes particularly for large data sets. In this cases there may be little difference bet^n standardized and studentized residuals.

Thus standardized and studentized residuals often convey equivalent information.

22/9 * PRESS Residuals :-

Another approach to making residuals useful in finding outliers is to examine the quantity that is computed from $y_i - \hat{y}_i$ where $\hat{y}_{(i)}$ is the fitted value of the i^{th} response based on all observation accept the i^{th} one.

- The logic behind this if the obs^n y_i is really unusual the regression model based on all the observation may be overly influenced by this obs^n this would produce a fitted value $\hat{y}_i$ i.e very similar to the observed value y_i and hence the ordinary residual e_i will be small. therefore it will be hard to defect the outlier however if we we delete the i^{th} obs^n than $\hat{y}_i$ cannot be influenced by that ~~obs^n, so,~~ Obs^n So, the resulting residual should be likely to indicate the presence of the outlier. If we delete the i^{th} obs^n, fit the remaining $(n-1)$ obs^n and calculate the predicted value of y_i corresponding to the ~~deted obs~~ deleted obs^n, the corresponding prediction error is

$e_{(i)} = y_i - \hat{y}_{(i)}$

- This prediction error calculation is repeated for each obs^n. This prediction errors are called <u>PRESS residuals</u>. Calculating press residuals requires fitting n different regression however it can be calculated from the results of a single least-square fit to all n obs^{ns}.

- i^{th} PRESS residual is

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

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

$$= \frac{1}{(1-h_{ii})^2}\,\text{var}(e_i)$$

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

So that a standardize PRESS residuals is,

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

If σ^2 is unknown then

$$\hat{\sigma}^2 = MSE$$

* R-Student :-

- In computing r_i, it is customary to use MSSE as an estimate of σ^2. This is referred as an internal scaling of residual because MSSE is an internally generated estimate of σ^2 obtained from fitting the model to all n obs^ns.

- Another approach would be use an estimate of σ^2 based on a dataset with the i^{th} obs^n remove.

- 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

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

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

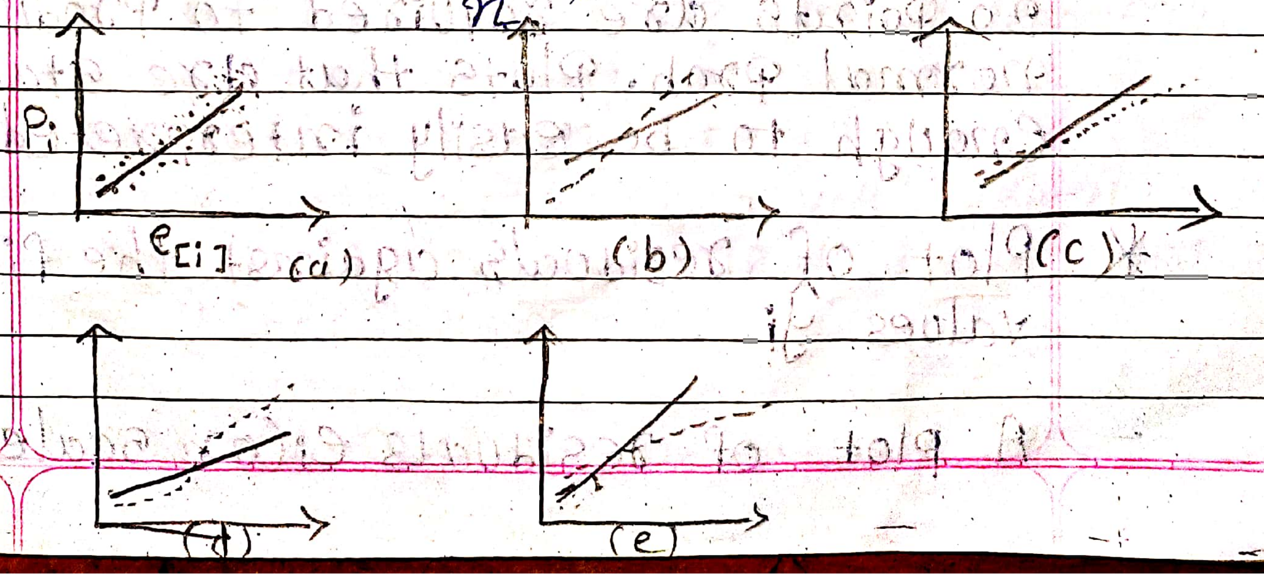
- One who compare all n values of $|t_i|$ to $t_{(\alpha/2n,\, n-p-1)}$ to provide guidance regarding outliers.

* Normal probability Plots :-

- Small departures from the normality assumption do not affect the model greatly but gross non-normality is potentially more serious as t or F statistic depend on the normality assumptions.

- A very simple method of checking normality assumption is to construct a normal prob. plot of the residuals.

- Let $e_{[1]} < e_{[2]} < \cdots e_{[n]}$ be the residuals ranked in increasing order. Then use plot $e_{[i]}$ against cumulative prob $P_i = \left(\frac{i - 1/2}{n}\right)$



- Figure (a) is an ideal Normal prob. plot. Figure (b) shows a sharp upward and downward curve at both extremes indicating that the tails of this dist^{ns} are too heavy for it to be considered normal.
Figure (c) shows flatening at the extremes which is a pattern typical of samples from a dist^{n} with thiner tails than the normal.
Figure (d) indicates positive skewness.
Figure (e) indicates negative skewness.

- Study of this plots is helpful in acquiring a feel for how much deviation from the straight line is acceptable. Some sample sizes $n < 16$ often produced normal prob. plots that deviates substantially from linearity. For larger sample sizes $n \geq 32$, the plots are much better behave. Usually 20 points are required to produce normal prob. plots that are stable enough to be easily interpreted.

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

- A plot of residuals e_i (or scaled

residuals) versus the corresponding fitted values $\hat{y}_i$ is useful for detecting several common types of model inadequacy

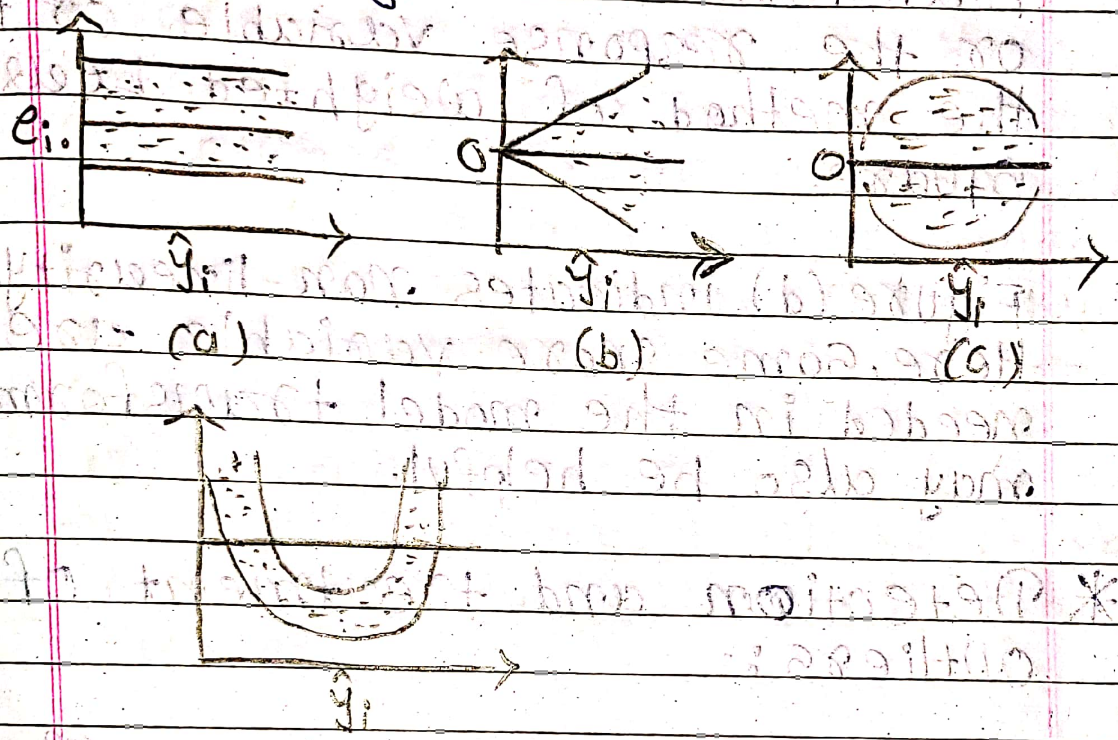


Figure (a) indicates that there is no model defect.

Figure (b) and (c) indicate that variance of the errors is not constant.

The outward opening funnel pattern in (b) implies that the variance is an increasing funct^n of y.

An inward opening funnel is also possible indicating variance increases as y decreases

Figure (c) occurs when y is a

Proportional betⁿ 0 and 1. The usual approach for dealing with inequality of variance is to apply a suitable transformation to either regressor or the response variable or to use the method of weighted Least Square.

Figure (d) indicates non-linearity. Here, some more variable are needed in the model transformation may also be helpful.

* Detection and treatment of outliers:-

- An outlier is an extreme obs^{ns} residual that are considerably larger in absolute value than the others say 3 or 4 standard deviation from the mean indicate potentially space outliers.

- Outliers are data points that are not typical of the rest of the data. Outliers can be detected by

(i) Residual plots against $\hat{y}_i$ and also normal prob plots.

ii) Examining scale residuals such as Studentized and R-student:

- Outlier should be carefully investigated to see if a reason for their unusual behaviour can be found. Sometimes outliers are bad values occuring as a result of unusual but explanable events. For example: Faulty measurement, incorrect, recording of data, failure of instrument, etc. If this is the case then the outlier should be corrected if possible or deleted from the dataset.

- Ocassionally, we find that outliers is more important than the rest of the data. The effect of outliers on regression model may be easily check by droping this points and we fitting the regression eq^n.

* <u>Test of an outliers</u>:-

- Suppose that i^{th} case is ~~sus~~ suspected to be an outlier, we proceed as follows:-

3/10

i) Delete the i^{th} case from the data so $(n-1)$ cases remain in the data set.

ii) Using the reduce dataset, estimate β and σ^2 call these estimates $\hat{\beta}_{(i)}$ and $\hat{\sigma}^2_{(i)}$ To indicate that case (i) (obs^n i) was not used in the estimation.

iii) For the deleted case, compute $\hat{y}_{i(i)} = x_i'\hat{\beta}_{(i)}$ since i^{th} case was not used in the estimation.

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 i^{th} case is given by

$$t_i = r_i\left(\frac{n-p-1}{n-p-r_i^2}\right)^{1/2}$$

where $r_i = \dfrac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$

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

we reject H_0 if $|t_i| \geq t_{(\alpha/2, n-p-1)}$

3/10

Unit – 4

* Introduction to non linear model

- These are many situations where a linear Regression model may not be appropriate for $gx \rightarrow$ The engineer or scientist may have direct knowledge of the form of relationship between the response variable and Regressors perhaps from the theory underlying the Phenomenon

- The true relationship b/w the response and the regressons may be a differential eq^n or sol^n toward differential eq^n often this leave to a model of Non-linear form any model i.e not linear in unknown parameters is non-linear Regression model

$y = \theta_1 e^{\theta_2 x} + \varepsilon$ is not linear in the unknown parameters θ_1 and θ_2

Let non-linear Regression model represented as

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

θ is $p \times 1$ vector unknowns parameters

ε is an uncorrelated random error term with

$E(\varepsilon) = 0, \ var(\varepsilon) = \sigma^2$

$E(\underline{y}) = E[f(\underline{x}, \underline{\theta})]$ $\quad (\because E(\varepsilon) = 0)$
$= f(\underline{x}, \underline{\theta})$

In non-linear Regression model, atleast one of the derivative of expected f^n w.r.t parameters depends on Atleast one of the parameters. In linear Regression, these derivatives are not functions of unknown parameters.

- for eg for linear model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$\frac{\partial y}{\partial \beta_j} = x_j \ \forall \ j$$

* Non linear least squares

Consider the non-linear regression model as

$$y_i = f(x_i, \theta) + \varepsilon_i \quad i = 1, 2, \ldots n$$

The error s.s is given by

$$S(\theta) = \sum [(y_i - f(x_i, \theta)]^2 \quad (1)$$

The normal eq^n are

$$\sum [y_i - f(x_i, \theta)] \left[\frac{\partial f(x_i, \theta)}{\partial \theta_j} \right] = 0$$

In Non-linear regression model, the derivaties in 2^{nd} bracket will be f^n of unknown parameters

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

$$\sum_{i=1}^{n} (y_i - \theta_1 e^{\theta_2 x_i})^2$$

The least square normal eq^{ns} are

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

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

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

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

These eq^n are not linear in θ_1 and θ_2 and hence no explicit sol^n exist. In general, interative method must be used to find the sol^n of θ_1 and θ_2

* Transformation to a linear model

- It is sometime useful to consider a transformation that induces linearity the model.

e.g. $y = \theta_1 e^{\theta_2 x} \epsilon$

$\log y = \log \theta_1 + \theta_2 x + \log \epsilon$
$= \beta_0 + \beta_1 + \epsilon'$

A Non-linear model that can be transformed to an equivalent form is said to be Intrinsically linear.

* Multicollinearity

- If there is ~~kno~~ no linear Relationship among the regressors, they are said to be orthogonal. when regressors are orthogonal inferences such as

(1) Identifing relative effects of regressor variables.

(2) Prediction or estimation.

(3) Selection of appropriate set of variables for the model can be make

relatively eassily. unfortunation in most regression the regressor are ~~unfortun: more application~~ not orthogonal.

Sometime Lack of orthogonality is not serious, However in some situations the regressors are nearly, perfectly Linearly related and in such cases the inferences based on regression model can be misleading, when there are near linear dependency among the regressor the problem of multicollinearity is said to be exist.

Multiple regression model is given by

$$\underline{y} = X\underline{\beta} + \underline{\epsilon}$$

$\underline{y}$: $n \times 1$ vector of responses
$\underline{X}$: $n \times p$ Matrix of regressor variables
$\underline{\beta}$: $p \times 1$ vector of unknown constant
ε : $n \times 1$ vector of random error
$\varepsilon \sim N(0, \sigma^2)$

We assume that, the regressor

variable and Response variables have been centered and scaled to Unit length.

$X'X$: is a $p \times p$ matrix of correlations between the regressors

$X'Y$: is a $p \times 1$ vector of correlations between the regressors and Response

Let the j^{th} column of the X matrix be denoted by $\underline{X}_j$ so that

$$X = [\underline{X}_1, \underline{X}_2 \ldots \underline{X}_p]$$

Thus $\underline{X}_j$ contains 'n' levels of the j^{th} Regressor variables.

we may formally defined multicollinearity in terms of linear dependence of columns of 'X'. The vector $\underline{X}_1, \underline{X}_2, \ldots \underline{X}_p$ are linearly dependent if there is a set of constant $t_1, t_2 \ldots t_p$ not all zero $\ni$ $\sum_{j=1}^{p} t_j \underline{X}_j = 0$ ——— $(*)$

If eq $(*)$ holds exactly for a subsets of columns of 'X' then rank of $X'X$ matrix is less than p. And $(X'X)^{-1}$ does not exist.

However, suppose eq^n (*) is approximately true for some subset of columns of 'X' then there will be near linear dependencies in 'x' then problem of multicollinearity is said to be exist multicollinearity inconditioning in $X'X$ matrix

* Sources of Multicollinearity

i) The data collection method employed Here when the analyst samples only a subspace of regions of the regressors defined by eq^n (*) then multicollinearity enters.

ii) Constraints on the model or in the Pop^n being sampled can cause multicollinearity. Contraints often occur in ~~for~~ problem involving production or chemical processes where regressors are components of product and these components add to a ~~con~~ constant.

iii) Model specification:

- Multicollinearity may also be in used by choice of model. for eg - Adding

Polynomial forms to a Regression model causes ill-conditioning in $X'X$. furthermore Range of X is small adding x^2 term can result insignificant multicollinearity. we often encounter situations such as these where two or more regressors are nearly linearly dependent and retaining all this regressors may contribute to multicollinearity

iv) Over defined Model

- An overdefined model has more regressor or variables than observation. These model are sometimes encountered in medical and behaviourial research where there may be only a small no. of subjects availables find information is collected for large no. of regressors on each subject. The usual approach to deal with multicollinearity in this context is to eliminate same of regressor variable from consideration.

* Effect of Multicollinearity

Suppose there are only two

$$Y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$
$$(X'X)\hat{\beta} = X'Y$$

Regressor x_1 and x_2 and if there is strong multicollinearity between x_1 and x_2. Then correlation coefficient r_{12} will be large which will result in large variances and covariances for the least square estimator of the regression coefficient.

- This imply that, different sample taken at some x-levels would lead to widely different estimates of model parameters.

This model:

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

$$(X'X)\hat{\beta} = X'y$$

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

$r_{12} \leftarrow$ sample correlation betn x_1 and x_2

$r_{jy} \leftarrow$ sample correlation betn x_j and y

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

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

Scanned with OKEN Scanner

$$\begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{pmatrix} = \begin{bmatrix} \frac{1}{(1-r_{12}^2)} & \frac{-r_{12}}{(1-r_{12}^2)} \\ \frac{-r_{12}}{(1-r_{12}^2)} & \frac{1}{(1-r_{12}^2)} \end{bmatrix} \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

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

- If there is strong multicollinearity between x_1 and x_2 then correlation coefficient will be large hence from above $|r_{12}| \to 1$.

  So, $var(\hat{\beta}_j) = C_{jj}\sigma^2 \to \infty$

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

- Therefore strong multicollinearity between x_1 and x_2 result in large variance and covariance for the LSE of the Regression coefficient. When there are more than two regressors multicollinearity produces similar effect. The diagonal elements of $C = (X'X)^{-1}$ matrix are c_{ij}

$$c_{ij} = \frac{1}{1-R_j^2}, \quad j = 1, 2 \ldots p$$

where, $R_j^2 \leftarrow$ coeff of multiple determination from the regression of x_j on the remaining $(p-1)$

regressor variables.

If there is strong multicollinearity between x_j and any subset of other $(p-1)$ regressors then $R_j^2 \to 1$ and hence $V(\hat{\beta}_j) \to \infty$.

→ Multicollinearity also tends to produces LSE $\hat{\beta}_j$ that are ~~too~~ to large in absolute values for this considered squared distance from $\hat{\beta}$ to the true parameter vector β.

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

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

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

$$= \sum_{j=1}^{p} var(\hat{\beta}_j)$$

$$= \sigma^2 \, trace(X'X)^{-1}$$

where, trace is sum of the main diagonal elements.

when there is multicollinearity presents ~~some~~ sum of the eigen values of $X'X$ will be small and since trace of matrix is equal to the sum of the eigen value then above eq^n 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 eigen value of $(X'X)$

The above eq^n implies that the distance from the LSE $\hat{\beta}$ to the true parameter value β may be large.

4/10

* Detection of Multicollinearity

→ Examination of correlation matrix

A very simple measure of multicollinearity is the inspection of the off diagonal elements r_{ij} in $X'X$ if the regressors x_i and x_j are nearly, linearly dependent then $|r_{ij}|$ will be near unity. examining r_{ij} between the regressor is useful in detecting near linear dependents bet^n. pairs of regressors only unfortunately when more than two regressors are involved in a near linear dependence there is no assurance that any of the pairwise correlation r_{ij} will be large.

Generally inspection of r_{ij} is not sufficient for detecting anything more complex then pairwise multicollinearity

→ Variance inflation factor (VIF)

It is defiened as

$$VIF_j = C_{ij} = \frac{1}{1-R_j^2}$$

The VIF for each term in the model measures the combine effect of the dependencies among the regressors on the variance of that term one or more large VIF indicate multicollinearity ~~pact~~ Paractical practical experience indicates that if of VIF exceeds 5 or 10. It is an indication that the associated regression coefficient are poorly estimated because of multicollinearity.

→ Eigen system analysis of $X'X$

- The characteristic roots or eigen value of $X'X$ say $\lambda_1, \lambda_2, \ldots, \lambda_p$ can be used to measure the exceeds of multicollinearity in the data. ~~If~~
- If there are one or more near linear independencies in the data than one or more characteristic roots will be small.
- The condition indices of $X'X$ matrix are,

$$K_j = \frac{\lambda_{max}}{\lambda_j}, \quad j = 1, 2 \ldots p$$

If $K_j < 100$ there is no serious problem with multicollinearity and

If $K_j > 1000$ there is serious multicollinearity

→ The determinate of $X'X$ can be used as a ~~indie~~ index of multicollinearity since $X'X$ matrix is a correlation matrix the possible rang of the values of the determinate is

$$0 \leq |X'X| \leq 1$$

Scanned with OKEN Scanner

– If $|X'X| = 1$ the regressors are orthogonal.
If $|X'X| = 0$ there is a exact linear dependence among the regressors.

The degree of multicollinearity becomes more severe as determinant of $|X'X|$ approaches zero.

→ The sign and magnitude of the regression coefficient will sometimes provide an indication that multicollinearity is present.

* Methods for dealing with multicollinearity

① Collecting additional Data
② Model respecification (montgomery)
③ Ridge regression
④ Principal component regression.

* Ridge Regression

– When the method of Least square is applied to non orthogonal data very poor estimate of the regression coefficient can be obtain the problem with method of ~~Leo~~ Least square is the requirement that $\hat{\beta}$ be an unbiased estimator of β.

- The ~~Gauss~~ Gauss Markov property assures that least square estimates has minimum variance in the class of unbiased linear estimators but there is not guareenty that this variance 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 then the unbiased estimator $\hat{\beta}$ the mean square error of the estimator $\hat{\beta}^*$ is defined as

$$MSE(\hat{\beta}^*) = E[\hat{\beta}^* - \beta]^2$$

$$= var(\hat{\beta}^*) + [E(\hat{\beta}^*) - \beta]^2$$

$$= var(\hat{\beta}^*) + (\text{Bias in } \hat{\beta}^*)^2$$

MSE is the expected squared distance

($\hat{\beta}$ = unbiased
$\hat{\beta}^*$ = bias)

form $\hat{\beta}^*$ to β by allowing a small amount of Bias in $\hat{\beta}^*$ the var($\hat{\beta}^*$) can be made small such that MSE($\hat{\beta}^*$) is less than the ~~var~~ variance of the unbiased estimator $\hat{\beta}$.

- The small variance for the bias estimator also implies that $\hat{\beta}^*$ is more stable estimator of β than unbiased estimator $\hat{\beta}$

- A number of procedure have been developed for obtaining biased estimators of regression coefficient one of this procedure is Ridge Regression.

- So ~~Re~~ Ridge estimator $\hat{\beta}_R$ is the solution of

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

or

$$\hat{\beta}_R = (X'X + KI)^{-1}X'\underline{Y}, \quad K \geq 0 \text{ is a constant selected by the analyst}$$

when $K = 0$ the ridge estimator is LSE

$$E(\hat{\beta}_R) = E[(X'X + KI)^{-1}X'\underline{Y}]$$

$$= (X'X + KI)^{-1} X' E(Y)$$

$$= (X'X + KI)^{-1} X' E(X\hat{\beta}) \quad \left(\because Y = X\beta + \varepsilon \quad E(Y) = E(X\hat{\beta})\right)$$

$$\boxed{E(\hat{\beta}_R) = (X'X + KI)^{-1} (X'X) \beta}$$

$\hat{\beta}_R$ is a biased estimator of β

$$var(\hat{\beta}_R) = var((X'X + KI)^{-1} X' Y)$$

$$= var[(X'X + KI)^{-1} (X'X) \hat{\beta}]$$

$$= (X'X + KI)^{-1} (X'X) \, var(\hat{\beta}) (X'X) (X'X + KI)$$

$$= \sigma^2 (X'X + KI)^{-1} (X'X) (X'X)^{-1} (X'X) (X'X + KI)^{-1}$$

$$\boxed{var(\hat{\beta}_R) = \sigma^2 (X'X + KI)^{-1} (X'X) (X'X + KI)^{-1}} \quad \left(\because var(\hat{\beta}) = \sigma^2 (X'X)^{-1}\right)$$

MSE of the ridge estimator

$$MSE(\hat{\beta}_R) = var(\hat{\beta}_R) + (\text{Bias in } \hat{\beta}_R)^2$$

$$= \sigma^2 Tr[(X'X + KI)^{-1} (X'X) (X'X + KI)^{-1}] + K^2 \beta' (X'X + KI)^{-2} \beta$$

$$= \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + K)^2} + K^2 \beta' (X'X + KI)^{-2} \beta \quad \text{(*)}$$

where $\lambda_1, \lambda_2 - \lambda_p$ are the eigen values of $X'X$

- The first terms on RHS of (*) is sum of variances of the parameters in $\hat{\beta}_R$ and

Second terms is squared on the Bias

- If $K>0$ the bias in $\hat{\beta}_R$ increases with K however variance decreases as K 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 the squared biase.

* Methods for choosing K

- Choosing K by inspection of the Ridge trace, is a subjective procedure requiring judgement on the part of the analysis

- Ridge trace is a plot of the elements of $\hat{\beta}_R$ vs K for values of K in the interval 0 to 1.

→ Hoerl, Kennard and Baldwin have suggested that an appropriate no. of variable

choice for $K = \frac{p\hat{\sigma}^2}{\hat{\underline{\beta}}'\hat{\underline{\beta}}}$ ----- $(*)$

where $\hat{\underline{\beta}}$ and $\hat{\sigma}$ are found from the Least squared sol^n they have shown that the resulting Ridge estimator held significant improvement in MSE over least squares.

→ Hoerl and kennard proposed an iterative estimation procedure based on $(*)$ specifically they ~~sugge~~ suggested they following squence of estimate of $\underline{\beta}$ and K

$$\hat{\underline{\beta}} : K_0 = \frac{p\hat{\sigma}^2}{\hat{\underline{\beta}}'\hat{\underline{\beta}}}$$

$$\hat{\beta}_R(K_0) : K_1 = \frac{p\hat{\sigma}^2}{\hat{\underline{\beta}}'(K_0)\,\hat{\beta}_R(K_0)}$$

$$\hat{\beta}_{R(K_1)} : K_2 = \frac{p\hat{\sigma}^2}{\hat{\beta}'_{R(K_1)}\,\hat{\beta}_{R(K_1)}}$$

The relative change in K_i is use to terminate the procedure if

$$\frac{K_{j+1} - K_j}{K_j} > 20T^{-1.3}$$

$$T = \frac{tr(X'X)^{-1}}{p}$$

the algorithm should continue. o.w terminate

→ MC Donald and Galarneau suggested choosing K so that

$$\hat{\beta}_R'\hat{\beta}_R = \hat{\beta}'\hat{\beta} - \hat{\sigma}^2 \sum_{j=1}^{p}\left(\frac{1}{\lambda_j}\right) \qquad (**)$$

If RHS of (**) was -ve they took $K=0$ or $K=\infty$

$\hat{\beta}_{R(k_0)}$: $k_1 = \dfrac{p\hat{\sigma}^2}{\hat{\beta}_R'^{(k_0)}\hat{\beta}_R^{(k_0)}}$

$\hat{\beta}_{R(k_1)}$: $k_2 = \dfrac{p\hat{\sigma}^2}{\hat{\beta}'_{R(k_1)}\hat{\beta}_{R(k_1)}}$

The relative change in k_j is used to terminate the procedure if $\dfrac{k_{j+1}-k_j}{k_j} > 20T^{-1.3}$

$T = \dfrac{tr(X'X)^{-1}}{p}$

the algorithm should continue otherwise terminate.

→ McDonald & Galarneau : Suggested choosing k so that $\hat{\beta}_R'\hat{\beta}_R = \hat{\beta}'\hat{\beta} - \hat{\sigma}^2\sum_{j=1}\left(\frac{1}{\lambda_j}\right)$ (**)

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

* **Principal components Regression**

→ Biased estimators of regr^n coeff^n can also be obtain by using a procedure known as principal component regr^n. Consider the canonical form of the model $Y = Z\alpha + \varepsilon$ where $Z = XT$, $\alpha = T'\beta$, $T'X'XT = Z'Z = \Lambda$

Recall $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$ is a $p\times p$ diagonal matrix of the eigen values of $X'X$ & T is a $p\times p$ orthogonal matrix whose columns are the eigen vectors associated with $\lambda_1, \lambda_2, \ldots, \lambda_p$.

The columns of Z which defines a new set of orthogonal regressors such as $Z = (Z_1, Z_2, \cdots Z_p)$ are reffered as principal components.

The least squares estimators of $\hat{\underline{\alpha}}$ is:
$\hat{\underline{\alpha}} = (Z'Z)^{-1} Z'\underline{Y} = \Lambda^{-1} Z'\underline{Y}$ and var-cov matrix of $\hat{\underline{\alpha}}$ is $var(\hat{\underline{\alpha}}) = \sigma^2 (Z'Z)^{-1} = \sigma^2 \Lambda^{-1}$
Thus a small eigen value of $X'X$ means that the variance of the corresponding orthogonal regression coefficient will be large since $Z'Z = \sum_{i=1}^{p} \sum_{j=1}^{p} Z_i Z_j = \Lambda$

We often reffer to the eigen value, λ_j as the variance of j^{th} principal component. If all the λ_j are equal to unity, the original regressors are orthogonal while if a λ_j is exactly equal to $0 \Rightarrow$ A perfect linear relationship b/w the original regressors. One or more of λ_j near 0 is multicollinearity is present.

$Var(\hat{\underline{\beta}}) = Var(T\hat{\underline{\alpha}}) = T\Lambda^{-1}T'\sigma^2$
$Var(\hat{\beta}_j) = \sigma^2 \sum_{j=1}^{p} t_{jj}^2 / \lambda_i$
$\therefore Var(\hat{\beta}_j)$ is a linear combination of the reciprocals of the eigen values.
This demonstrates how one or more small eigen values can destroy the precision of the least square estimate $\hat{\beta}_j$. Since $Z = XT$

Date

P.No.

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

where x_j is the j^{th} column of the X matrix & t_{ji} are the elements of the i^{th} colⁿ of T

If the var of the i^{th} principal component (λ_i) is small $\Rightarrow$ Z_i is nearly constant and eqⁿ (1) indicates that there is a linear combination of the original regressors that is nearly constant. This is the defination of multicollinearity. $\therefore$ eq (i) explains why the elements of the eigen vector associated with a small eigen value of $X'X$ identify the regressors involved in the multicollinearity.

The ~~principal~~ PCR approach deals with multicollinearity by using less regressor variables which are principal component in the model.

To obtain the PC estimator assume that the regressors are arranged in the order of decreasing eigen values. $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_p > 0$

Suppose that last s of this eigen values are approximatly $= 0$
In PCR, the PCs corresponding to near 0 eigen values are removed from the analyses.
Then and least square method is applied to the remaining components.

$\hat{\alpha}_{pc} = B\hat{\alpha}$

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

$$b_{p-s+1} = b_{p-s+2} \cdots = b_p = 0$$

Thus pc estimation is

$$\hat{\alpha}_{pc} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_{p-s} \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \\ a \end{matrix}}\right\} (p-s) \text{ components} \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} s \text{ components} \end{matrix}$$

or in forms of standardized regressors

$$\hat{\beta}_{pc} = T\hat{\alpha}_{pc} = \sum_{j=1}^{p-s} \lambda_j^{-1} t_j' X' y \, t_j$$

residual plots were obtained (not shown here). All of these plots indicate that each of the predictor variables is linearly associated with Y', with X_3 and X_4 showing the highest degrees of association and X_1 the lowest. The scatter plot matrix and the correlation matrix further show intercorrelations among the potential predictor variables. In particular, X_4 has moderately high pairwise correlations with X_1, X_2, and X_3.

On the basis of these analyses, the investigator concluded to use, at this stage of the model-building process, $Y' = \ln Y$ as the response variable, to represent the predictor variables in linear terms, and not to include any interaction terms. The next stage in the model-building process is to examine whether all of the potential predictor variables are needed or whether a subset of them is adequate. A number of useful measures have been developed to assess the adequacy of the various subsets. We now turn to a discussion of these measures.

9.3 Criteria for Model Selection

From any set of $p-1$ predictors, 2^{p-1} alternative models can be constructed. This calculation is based on the fact that each predictor can be either included or excluded from the model. For example, the $2^4 = 16$ different possible subset models that can be formed from the pool of four X variables in the surgical unit example are listed in Table 9.2. First, there is the regression model with no X variables, i.e., the model $Y_i = \beta_0 + \varepsilon_i$. Then there are the regression models with one X variable (X_1, X_2, X_3, X_4), with two X variables (X_1 and X_2, X_1 and X_3, X_1 and X_4, X_2 and X_3, X_2 and X_4, X_3 and X_4), and so on.

TABLE 9.2 SSE_p, R_p^2, $R_{a,p}^2$, C_p, AIC_p, SBC_p, and $PRESS_p$ Values for All Possible Regression Models—Surgical Unit Example.

X Variables in Model	(1) p	(2) SSE_p	(3) R_p^2	(4) $R_{a,p}^2$	(5) C_p	(6) AIC_p	(7) SBC_p	(8) $PRESS_p$
None	1	12.808	0.000	0.000	151.498	−75.703	−73.714	13.296
X_1	2	12.031	0.061	0.043	141.164	−77.079	−73.101	13.512
X_2	2	9.979	0.221	0.206	108.556	−87.178	−83.200	10.744
X_3	2	7.332	0.428	0.417	66.489	−103.827	−99.849	8.327
X_4	2	7.409	0.422	0.410	67.715	−103.262	−99.284	8.025
X_1, X_2	3	9.443	0.263	0.234	102.031	−88.162	−82.195	11.062
X_1, X_3	3	5.781	0.549	0.531	43.852	−114.658	−108.691	6.988
X_1, X_4	3	7.299	0.430	0.408	67.972	−102.067	−96.100	8.472
X_2, X_3	3	4.312	0.663	0.650	20.520	−130.483	−124.516	5.065
X_2, X_4	3	6.622	0.483	0.463	57.215	−107.324	−101.357	7.476
X_3, X_4	3	5.130	0.599	0.584	33.504	−121.113	−115.146	6.121
X_1, X_2, X_3	4	3.109	0.757	0.743	3.391	−146.161	−138.205	3.914
X_1, X_2, X_4	4	6.570	0.487	0.456	58.392	−105.748	−97.792	7.903
X_1, X_3, X_4	4	4.968	0.612	0.589	32.932	−120.844	−112.888	6.207
X_2, X_3, X_4	4	3.614	0.718	0.701	11.424	−138.023	−130.067	4.597
X_1, X_2, X_3, X_4	5	3.084	0.759	0.740	5.000	−144.590	−134.645	4.069

In most circumstances, it will be impossible for an analyst to make a detailed examination of all possible regression models. For instance, when there are 10 potential X variables in the pool, there would be $2^{10} = 1{,}024$ possible regression models. With the availability of high-speed computers and efficient algorithms, running all possible regression models for 10 potential X variables is not time consuming. Still, the sheer volume of 1,024 alternative models to examine carefully would be an overwhelming task for a data analyst.

Model selection procedures, also known as subset selection or variables selection procedures, have been developed to identify a small group of regression models that are "good" according to a specified criterion. A detailed examination can then be made of a limited number of the more promising or "candidate" models, leading to the selection of the final regression model to be employed. This limited number might consist of three to six "good" subsets according to the criteria specified, so the investigator can then carefully study these regression models for choosing the final model.

While many criteria for comparing the regression models have been developed, we will focus on six: R_p^2, $R_{a,p}^2$, C_p, AIC_p, SBC_p, and $PRESS_p$. Before doing so, we will need to develop some notation. We shall denote the number of potential X variables in the pool by $P-1$. We assume throughout this chapter that all regression models contain an intercept term β_0. Hence, the regression function containing all potential X variables contains P parameters, and the function with no X variables contains one parameter (β_0).

The number of X variables in a subset will be denoted by $p-1$, as always, so that there are p parameters in the regression function for this subset of X variables. Thus, we have:

$$1 \leq p \leq P \tag{9.1}$$

We will assume that the number of observations exceeds the maximum number of potential parameters:

$$n > P \tag{9.2}$$

and, indeed, it is highly desirable that n be substantially larger than P, as we noted earlier, so that sound results can be obtained.

R_p^2 or SSE_p Criterion

The R_p^2 criterion calls for the use of the coefficient of multiple determination R^2, defined in (6.40), in order to identify several "good" subsets of X variables—in other words, subsets for which R^2 is high. We show the number of parameters in the regression model as a subscript of R^2. Thus R_p^2 indicates that there are p parameters, or $p-1$ X variables, in the regression function on which R_p^2 is based.

The R_p^2 criterion is equivalent to using the error sum of squares SSE_p as the criterion (we again show the number of parameters in the regression model as a subscript). With the SSE_p criterion, subsets for which SSE_p is small are considered "good." The equivalence of the R_p^2 and SSE_p criteria follows from (6.40):

$$R_p^2 = 1 - \frac{SSE_p}{SSTO} \tag{9.3}$$

Since the denominator $SSTO$ is constant for all possible regression models, R_p^2 varies inversely with SSE_p.

The R_p^2 criterion is not intended to identify the subsets that maximize this criterion. We know that R_p^2 can never decrease as additional X variables are included in the model. Hence, R_p^2 will be a maximum when all $P-1$ potential X variables are included in the regression model. The intent in using the R_p^2 criterion is to find the point where adding more X variables is not worthwhile because it leads to a very small increase in R_p^2. Often, this point is reached when only a limited number of X variables is included in the regression model. Clearly, the determination of where diminishing returns set in is a judgmental one.

Example

Table 9.2 for the surgical unit example shows in columns 1 and 2 the number of parameters in the regression function and the error sum of squares for each possible regression model. In column 3 are given the R_p^2 values. The results were obtained from a series of computer runs. For instance, when X_4 is the only X variable in the regression model, we obtain:

$$R_2^2 = 1 - \frac{SSE(X_4)}{SSTO} = 1 - \frac{7.409}{12.808} = .422$$

Note that $SSTO = SSE_1 = 12.808$.

Figure 9.4a contains a plot of the R_p^2 values against p, the number of parameters in the regression model. The maximum R_p^2 value for the possible subsets each consisting of $p-1$ predictor variables, denoted by $\max(R_p^2)$, appears at the top of the graph for each p. These points are connected by solid lines to show the impact of adding additional X variables. Figure 9.4a makes it clear that little increase in $\max(R_p^2)$ takes place after three X variables are included in the model. Hence, consideration of the subsets (X_1, X_2, X_3) for which $R_4^2 = .757$ (as shown in column 3 of Table 9.2) and (X_2, X_3, X_4) for which $R_4^2 = .718$ appears to be reasonable according to the R_p^2 criterion.

Note that variables X_3 and X_4, correlate most highly with the response variable, yet this pair does not appear together in the $\max(R_p^2)$ model for $p = 4$. This suggests that X_1, X_2, and X_3 contain much of the information presented by X_4. Note also that the coefficient of multiple determination associated with subset (X_2, X_3, X_4), $R_4^2 = .718$, is somewhat smaller than $R_4^2 = .757$ for subset (X_1, X_2, X_3).

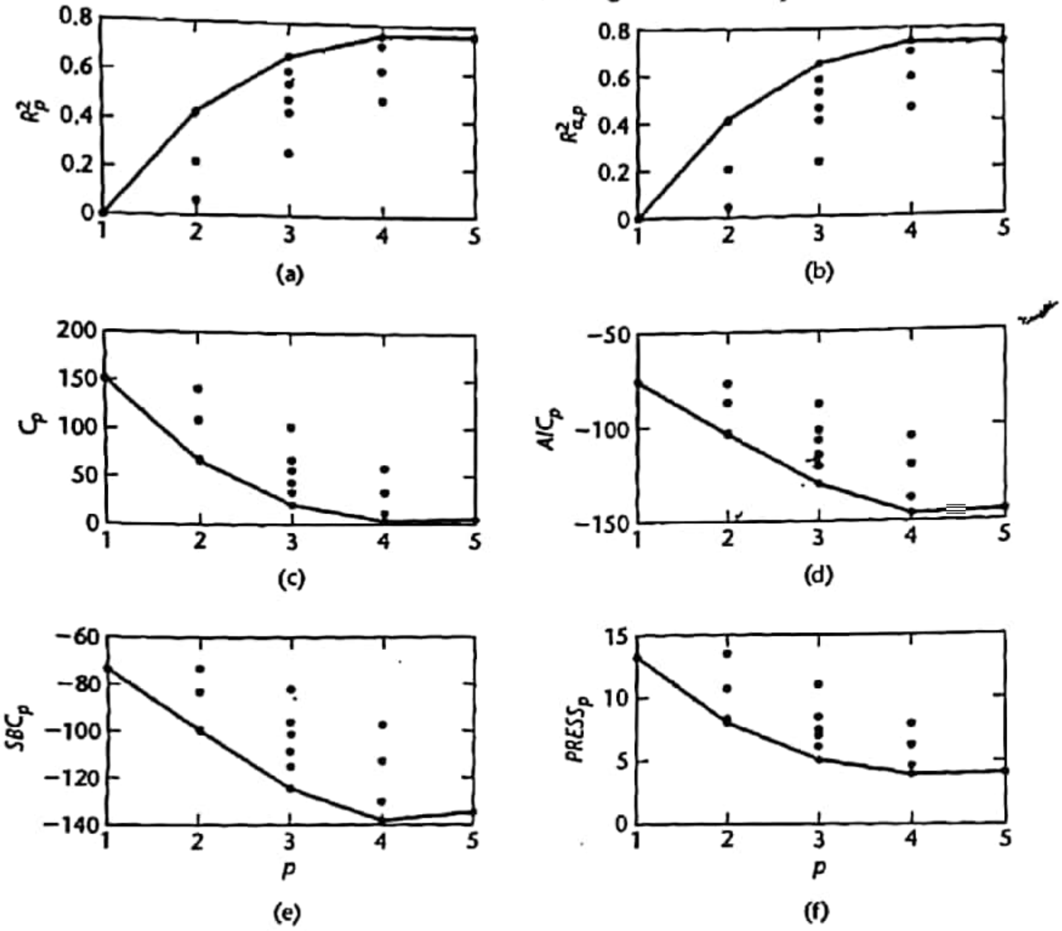
$R_{a,p}^2$ or MSE_p Criterion

Since R_p^2 does not take account of the number of parameters in the regression model and since $\max(R_p^2)$ can never decrease as p increases, the adjusted coefficient of multiple determination $R_{a,p}^2$ in (6.42) has been suggested as an alternative criterion:

$$R_{a,p}^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE_p}{SSTO} = 1 - \frac{MSE_p}{\frac{SSTO}{n-1}} \tag{9.4}$$

This coefficient takes the number of parameters in the regression model into account through the degrees of freedom. It can be seen from (9.4) that $R_{a,p}^2$ increases if and only if MSE_p decreases since $SSTO/(n-1)$ is fixed for the given Y observations. Hence, $R_{a,p}^2$ and MSE_p provide equivalent information. We shall consider here the criterion $R_{a,p}^2$, again showing the number of parameters in the regression model as a subscript of the criterion. The largest $R_{a,p}^2$ for a given number of parameters in the model, $\max(R_{a,p}^2)$, can, indeed, decrease as p increases. This occurs when the increase in $\max(R_p^2)$ becomes so small that it is not

FIGURE 9.4 Plot of Variables Selection Criteria—Surgical Unit Example.



sufficient to offset the loss of an additional degree of freedom. Users of the $R_{a,p}^2$ criterion seek to find a few subsets for which $R_{a,p}^2$ is at the maximum or so close to the maximum that adding more variables is not worthwhile.

Example

The $R_{a,p}^2$ values for all possible regression models for the surgical unit example are shown in Table 9.2, column 4. For instance, we have for the regression model containing only X_4:

$$R_{a,2}^2 = 1 - \left(\frac{n-1}{n-2}\right)\frac{SSE(X_4)}{SSTO} = 1 - \left(\frac{53}{52}\right)\frac{7.409}{12.808} = .410$$

Figure 9.4b contains the $R_{a,p}^2$ plot for the surgical unit example. We have again connected the $\max(R_{a,p}^2)$ values by solid lines. The story told by the $R_{a,p}^2$ plot in Figure 9.4b is very similar to that told by the R_p^2 plot in Figure 9.4a. Consideration of the subsets (X_1, X_2, X_3) and (X_2, X_3, X_4) appears to be reasonable according to the $R_{a,p}^2$ criterion. Notice that $R_{a,4}^2 = .743$ is maximized for subset (X_1, X_2, X_3), and that adding X_4 to this subset—thus using all four predictors—decreases the criterion slightly: $R_{a,5}^2 = .740$.

Mallows' C_p Criterion

This criterion is concerned with the *total mean squared error* of the n fitted values for each subset regression model. The mean squared error concept involves the total error in each fitted value:

$$\hat{Y}_i - \mu_i \tag{9.5}$$

where μ_i is the true mean response when the levels of the predictor variables X_k are those for the ith case. This total error is made up of a bias component and a random error component:

1. The bias component for the ith fitted value $\hat{Y}_i$, also called the model error component, is:

$$E\{\hat{Y}_i\} - \mu_i \tag{9.5a}$$

where $E\{\hat{Y}_i\}$ is the expectation of the ith fitted value for the given regression model. If the fitted model is not correct, $E\{\hat{Y}_i\}$ will differ from the true mean response μ_i and the difference represents the bias of the fitted model.

2. The random error component for $\hat{Y}_i$ is:

$$\hat{Y}_i - E\{\hat{Y}_i\} \tag{9.5b}$$

This component represents the deviation of the fitted value $\hat{Y}_i$ for the given sample from the expected value when the ith fitted value is obtained by fitting the same regression model to all possible samples.

The mean squared error for $\hat{Y}_i$ is defined as the expected value of the square of the total error in (9.5)—in other words, the expected value of:

$$(\hat{Y}_i - \mu_i)^2 = [(E\{\hat{Y}_i\} - \mu_i) + (\hat{Y}_i - E\{\hat{Y}_i\})]^2$$

It can be shown that this expected value is:

$$E\{\hat{Y}_i - \mu_i\}^2 = (E\{\hat{Y}_i\} - \mu_i)^2 + \sigma^2\{\hat{Y}_i\} \tag{9.6}$$

where $\sigma^2\{\hat{Y}_i\}$ is the variance of the fitted value $\hat{Y}_i$. We see from (9.6) that the mean squared error for the fitted value $\hat{Y}_i$ is the sum of the squared bias and the variance of $\hat{Y}_i$.

The total mean squared error for all n fitted values $\hat{Y}_i$ is the sum of the n individual mean squared errors in (9.6):

$$\sum_{i=1}^{n}[(E\{\hat{Y}_i\} - \mu_i)^2 + \sigma^2\{\hat{Y}_i\}] = \sum_{i=1}^{n}(E\{\hat{Y}_i\} - \mu_i)^2 + \sum_{i=1}^{n}\sigma^2\{\hat{Y}_i\} \tag{9.7}$$

The criterion measure, denoted by Γ_p, is simply the total mean squared error in (9.7) divided by σ^2, the true error variance:

$$\Gamma_p = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(E\{\hat{Y}_i\} - \mu_i)^2 + \sum_{i=1}^{n}\sigma^2\{\hat{Y}_i\}\right] \tag{9.8}$$

The model which includes all $P-1$ potential X variables is assumed to have been carefully chosen so that $MSE(X_1, \dots, X_{P-1})$ is an unbiased estimator of σ^2. It can then be shown that an estimator of Γ_p is C_p:

$$C_p = \frac{SSE_p}{MSE(X_1, \dots, X_{P-1})} - (n - 2p) \tag{9.9}$$

where SSE_p is the error sum of squares for the fitted subset regression model with p parameters (i.e., with $p-1$ X variables).

When there is no bias in the regression model with $p-1$ X variables so that $E\{\hat{Y}_i\} \equiv \mu_i$, the expected value of C_p is approximately p:

$$E\{C_p\} \approx p \qquad \text{when } E\{\hat{Y}_i\} \equiv \mu_i \tag{9.10}$$

Thus, when the C_p values for all possible regression models are plotted against p, those models with little bias will tend to fall near the line $C_p = p$. Models with substantial bias will tend to fall considerably above this line. C_p values below the line $C_p = p$ are interpreted as showing no bias, being below the line due to sampling error. The C_p value for the regression model containing all $P-1$ X variables is, by definition, P. The C_p measure assumes that $MSE(X_1, \dots, X_{P-1})$ is an unbiased estimator of σ^2, which is equivalent to assuming that this model contains no bias.

In using the C_p criterion, we seek to identify subsets of X variables for which (1) the C_p value is small and (2) the C_p value is near p. Subsets with small C_p values have a small total mean squared error, and when the C_p value is also near p, the bias of the regression model is small. It may sometimes occur that the regression model based on a subset of X variables with a small C_p value involves substantial bias. In that case, one may at times prefer a regression model based on a somewhat larger subset of X variables for which the C_p value is only slightly larger but which does not involve a substantial bias component. Reference 9.1 contains extended discussions of applications of the C_p criterion.

Example

Table 9.2, column 5, contains the C_p values for all possible regression models for the surgical unit example. For instance, when X_4 is the only X variable in the regression model, the C_p value is:

$$C_2 = \frac{SSE(X_4)}{\dfrac{SSE(X_1, X_2, X_3, X_4)}{n-5}} - [n - 2(2)]$$

$$= \frac{7.409}{\dfrac{3.084}{49}} - [54 - 2(2)] = 67.715$$

The C_p values for all possible regression models are plotted in Figure 9.4c. We find that C_p is minimized for subset (X_1, X_2, X_3). Notice that $C_p = 3.391 < p = 4$ for this model, indicating little or no bias in the regression model.