Mathematical Derivation of least Squares.

@¶▼

▸@@▸@●▸▼@ É●@H●@▸ @©ñ▸

Matrix Form.

$$Y = XB + e$$

$[N \times 1]$ $\quad N \times (R+1) \quad (R+1) \times 1 \quad (N+1)$

Sum of Square of errors.

$$SSE = e'e$$

Now $\quad e = Y - XB$

Thus $\quad SSE = (Y - XB)'(Y - XB)$

Thus $SSE = (Y' - B'X')(Y - XB)$

or $SSE = Y'Y - Y'XB - B'X'Y + B'X'XB$

Thus we have $SSE = Y'Y - 2Y'XB + B'X'XB$

Note. $Y'XB = B'X'Y$ in Scalar Sense

$$\frac{\partial SSE}{\partial B} = \frac{\partial}{\partial B}\left(Y'Y - 2Y'XB + B'X'XB\right) = 0$$

$$\cancel{X'X} = X'XB.$$

Thus $$\frac{\partial SSE}{\partial B} = -2X'Y + 2X'XB$$

$$\Rightarrow X'XB = X'Y$$

or $$\boxed{B = (X'X)^{-1}X'Y}$$

Basics

LINEAR ALGEBRA REVISIONS.

Equality of Matrices : $A = B$ if $a_{ij} = b_{ij}$ for all i, j.

Transpose and symmetric matrices:

$(A')' = A$ Transpose of a matrix = original matrix

Symmetric matrix.

Vector of 1's. $j = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

Vector of Zeros $0 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

Square matrix of 1's $J =$ e.g. $J_{(3\times3)} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

Matrix of Zeros $\underset{3\times3}{O} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

Operations.

Summation. $\sum_{i=1}^{n} a_i = a_1 + a_2 + \dots a_n.$

$$\sum_{i=1}^{2}\sum_{j=1}^{3} a_{ij} = a_{11} + a_{12} + a_{13} + a_{21} + a_{22} + a_{23}.$$

Diagonal Matrix

Say $A = \begin{bmatrix} 5 & -2 & 4 \\ 7 & 9 & 3 \\ -6 & 8 & 1 \end{bmatrix}$ $D = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix}$ All Non diagonal terms are Zero.

$\text{diag}(A) = D_A = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Identity 3×3 $\underset{3\times3}{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Upper Δ. $T = \begin{bmatrix} 8 & 3 & 4 & 7 \\ 0 & 0 & -2 & 3 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 6 \end{bmatrix} \rightarrow$ All elements below diag. are Zeros.

Such a matrix is Upper Δ

Multiplication of Matrices and vectors.

$C = AB \Rightarrow$

$$c_{ij} = \sum_{k} a_{ik} b_{kj}$$

product of Sequence of Numbers.

$$\prod_{i=1}^{n} a_i = (a_1)(a_2)\cdots(a_n)$$

Addition

$A+B = B+A.$] Addition is commutative.

Other Matrix Algebra.

$$(A+B)' = A'+B'$$
$$(A-B)' = A'-B'$$
$$(x+y)' = x'+y'$$
$$(x-y)' = x'-y'$$

Triple product

$$ABC = A(BC) = (AB)C.$$

$$ABC + ADC = A(B+D)C.$$

Consider the following:

Let $\underset{n\times p}{X}$, $\underset{n\times n}{A}$.

Then we have

$$X'X - X'AX = X'(X - AX) = X'(I-A)X$$

$$A(B+C) = AB + AC.$$
$$A(B-C) = AB - AC.$$
$$(A+B)C = AC + BC.$$
$$(A-B)C = AC - BC.$$

] Multiplication is Distributive. Over addition or subtraction.

Transpose. $(AB)' = B'A'$

Getting Row sums and col sums.

Consider $j = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}_{3\times 1}$, Consider $\underset{3\times 3}{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$.

Row sum $\underline{Aj}$.

$[A]\{j\}$

$[3\times 3][3\times 1] = [3\times 1]$

$$Aj = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_i a_{1i} \\ \sum_i a_{2i} \\ \sum a_{3i} \end{pmatrix}$$

Colsum.

$\{j'\}[A]$ $\quad j'A = (1\ 1\ 1) \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = (\Sigma_{i1}\ \Sigma_{i2}\ \Sigma$

$[1\times 3][3\times 3] = [1\times 3]$

*

important Relation where product is scalar

$a = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$ $\quad b = \begin{pmatrix} 2 \\ 3 \\ 9 \end{pmatrix}$. $\quad a'b$ is a scalar.

Then $(a'b)' = b'a = a'b$

Thus $(a'b)^2 = (a'b)(a'b) = (a'b)(b'a)$ — ①

Using ① we can rewrite

Middle term as per ①.

$\underline{(x-y)'(x-y)}$ As $x'x - 2x'y + yy'$

↳ This kind of product is encountered in evaluating Least Squares. for multiple regression.

Dot Product & Matrix Products.

@♥
Ü

@◆► '@◄☺@☺☻ ☻ ☻ ÇBÇ☻☺Bê►@↔☺?☺`α?ÇJÇ⌂' ε ΩÇ⌂Ç:Ç?└§ °¿☺ ♣☺ ▓╥ ►▼@►►☻@☺☻►►@♣Ç$é

Consider $\underset{n\times 1}{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix}$ $a'a = (a_1 \ a_2 \cdots a_n)\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1^2 + a_2^2 + \cdots a_n^2$

$=$ Sum of Squares.

also we have $aa' = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}(a_1 \ a_2 \cdots a_n) = \begin{pmatrix} a_1^2 & a_1 a_2 \cdots & a_1 \\ & \ddots & \\ & & a_n \end{pmatrix}$

Thus $a'a$ = dot product

aa' = matrix product.

Length of $a = \sqrt{a'a} = \sqrt{\sum_{i=1}^{n} a_i^2}$

$j'j = n$ $\quad jj' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & & & 1 \\ \vdots & & & \vdots \\ & & & 1 \end{pmatrix} = J$

Rank

Linear dependence:

$\vec{a_1}, \vec{a_2} \cdots \vec{a_n}$ said to be linearly dependent

if constants $c_1, c_2 \cdots c_n$. Not all zero can be found such that: $c_1 a_1 + c_2 a_2 + \cdots c_n a_n = 0$

Linear Independent: : If no constants can be found (Non zero) then they are independent.

RF=6

@®▶◀@☺▶Ç▶♣@ ▶

Rank (A) =

of a square matrix

= Nos of linearly independent rows of A.

= Nos of linearly independent columns of A.

if A is $(n \times p)$ the maximum possible Rank is the smaller of n and p.

Inverse:

if A is square and full rank. A is said to be non singular, A has unique inverse A^{-1}

$$AA^{-1} = A^{-1}A = I$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

Positive Definite Matrices:

If $x'Ax > 0$ for all vectors x then A is said to be positive definite.

obtaining a +ve definite Matrix.

Consider $A = B'B$, where B is $n \times p$, of Rank $p < n$.

Then $B'B$ is +ve definite.

$$x'Ax = ~~x'B'Ax~~ \quad x'B'Bx = (Bx)'Bx = zz'$$

$$= z_i^2$$

Thus $x'Ax > 0$,

Positive Definite

[RF7]

@@▸♠@♣D@Æd ☻▸ @♥●ç▸ ▲a▸

$A = T'T$ T is a non singular upper

Cholesky Decomposition.

Let $A = a_{ij}$

$T = t_{ij}$ be $n \times n$

$$t_{11} = \sqrt{a_{11}}, \quad t_{1j} = \frac{a_{1j}}{t_{11}} \quad 2 \le j \le n.$$

$$t_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} t_{ki}^2} \quad 2 \le i \le n.$$

$$t_{ij} = \frac{a_{ij} - \sum_{k=1}^{i-1} t_{ki} t_{kj}}{t_{ii}} \quad 2 \le i < j < n.$$

$$t_{ij} = 0 \quad 1 \le j < i \le n.$$

Determinants:

2x2 Matrix $|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$

For a diagonal Matrix the following holds.

$$|D| = \prod_{i=1}^{n} d_i$$

$$|AB| = |A||B|.$$

$$|A'| = |A|$$

$$|A^{-1}| = \frac{1}{|A|} = |A|^{-1}$$

@♦@@ü▶

if B is non singular. And c is a vector

then. $|B+cc'| = |B|(1 + c'B^{-1}c)$

TRACE

Sum of diagonal elements for a $n \times n$ (Square) Matrix

$\therefore \quad tr(A) = \sum_{i=1}^{n} a_{ii}$

$tr(A+B) = tr(A) + tr(B)$

$tr(AB) = tr(BA)$

$tr(A'A) = tr(AA')$

Orthogonal Vectors and Matrices.

2 vectors of same size a, b

a, b. are said to be orthogonal

if $a'b = a_1b_1 + a_2b_2 + \dots a_nb_n = 0$

if $[a'a = 1]$ Vector $\underline{a}$ is said to be normalized.

$c = \dfrac{a}{\sqrt{a'a}}$ Thus $c'c = 1$

Consider 'C' matrix whose columns are mutually orthogonal and are normalized.

We have $c'c = I$.

also $cc' = I$.

Thus for an orthogonal Matrix. $c^{-1} = c'$

Rotation Effect

If C is orthogonal, Consider $z = Cx$

$\$\blacktriangleright\blacklozenge$ $\heartsuit\spadesuit$ $\$\blacktriangleright$ Then $z'z = (Cx)'(Cx) = x'C'Cx$

$= x'Ix$

$= x'x$

Thus distance from origin to z is Same as distance from origin to x

Thus $z = Cx$, leads to a rotation effect.

Eigen Values & Eigen Vectors:

For every Sq Matrix A.

we can find a scalar λ, vector x, such that

$Ax = \lambda x \Rightarrow (A - \lambda I)x = 0$.

if $|A - \lambda I| \neq 0$ then $x = 0$ is the only solution.

Thus for a non zero solution. $|A - \lambda I| = 0$.

$|A - \lambda I| = 0$] → Char. eqn.

$I + A$, $I - A$

$Ax = \lambda x$

$x + Ax = x + \lambda x$

$(I + A)x = (1 + \lambda)x$.

trace and eigen values.

$tr(A) = \sum_{i=1}^{n} \lambda_i$

$|A| = \prod_{i=1}^{n} \lambda_i$

Spectral Decomposition.

Consider $n \times n$ matrix A which is symmetric.

Consider a matrix 'C'

$C = (x_1, x_2 \dots x_n)$ → The columns of C are eigen vectors of A.

If the n eigen vectors are normalized, then C' is orthogonal.

Thus we have $I = CC'$

Then $A = ACC'$

Now subs. $C = x_1, x_2 \dots x_n$.

$\therefore A = A(x_1, x_2 \dots x_n)C'$

$= (Ax_1, Ax_2, Ax_n)C'$

$= (\lambda_1 x_1, \lambda_2 x_2 \dots \lambda_n x_n)C'$

$A = CDC'$ → $\therefore D = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots \\ 0 & \lambda_2 & \dots & \\ & & \ddots & \\ \dots & & & \lambda_n \end{pmatrix}$

Thus $C'AC = D$

Thus a symmetric Matrix A.

Can be diagonalized by an orthogonal Matrix containing normalized eigen vectors.

$\boxed{A = CDC'}$ → Spectral Decomposition of A.

Square root Matrix

$$A^{1/2} = CD^{1/2}C' \quad \Big] \; D^{1/2}$$

also $A^2 = CD^2C'$

$$A^{-1} = CD^{-1}C'$$

OLS Theory

$$y = X\beta + e.$$

Let $\hat{\beta}$ be an estimate of β. Then we can predict y_i with $\hat{y}_i$ as follows.

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

also we have $\hat{e}_i = y_i - \hat{y}_i$

Sum of Squared residuals $\hat{e}'\hat{e}$.

in matrix notation we get.

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

$$\frac{\partial s}{\partial \hat{\beta}} = 0 \Rightarrow \hat{\beta} = (X'X)^{-1}X'y$$

Now $\hat{y} = X\hat{\beta}$] Thus. $\hat{y} = X(X'X)^{-1}X'y$

or $\hat{y} = Hy$

hat matrix.

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

Properties of the hat matrix.

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

Hat matrix is symmetric. $H = H'$

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

then $H' =$

$$\text{transpose } [\underbrace{X(X'X)^{-1}}_{A}\underbrace{X'}_{B}]$$

$$= X \cdot \text{transpose } [X(X'X)^{-1}]$$

$$= X \cdot \text{transpose } [(X'X)^{-1}] \cdot X'$$

Now we know $X'X$ is symmetric.

$$\therefore \text{transpose } [(X'X)^{-1}] = (X'X)^{-1}$$

$\therefore H' =$

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

$(I-H)$ is idempotent: i.e. $(I-H) = (I-H)(I-H)$

Check. hat matrix is idempotent:

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

$$= X\underbrace{(X'X)^{-1}(X'X)}_{I}(X'X)^{-1}X'$$

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

$$= H.$$

$$(I-H)(I-H)$$

$$= I(I-H) - H(I-H)$$

$$= I - H - H + H \cdot H$$

$$= I - H \cancel{- H} \cancel{+ H}$$

$$= I - H$$

$\hat{y} = Hy$

Var operator:

$y - \hat{y} = (1 - H)y$

$Var(\hat{e}) = Var[(I - H)y]$

Now in general if X is a matrix, v is a column

$Var(\hat{e}) = Var(X \cdot v) = X \cdot Var(v) \cdot X'$

$\therefore Var(\hat{e}) = (1 - H) Var(y) (1 - H)'$

Digression:

$$Var(y) = \begin{bmatrix} Var\, y_1 & Cov(y_1, y_2) & & \\ & Var\, y_2 & & \\ & & \ddots & Var\, y_N \end{bmatrix}$$

Now $Var(ax + by) = a^2 Var(x) + b^2 Var(y) + 2ab\, Cov(x, y)$.

Now $y_i = X_i \beta + e_i$

$Var(y_i) = Var(X_i \beta) + Var(e_i) + 2\, Cov(X_i \beta, e_i)$

Now $Var(\beta) = 0$, also X is uncorr with e.

So $Var(y_i) = Var(e_i) = \sigma^2_{e_i}$

Now → Homoskedasticity Assumption.

$$Var\, y = \begin{bmatrix} \sigma^2_e & & \\ & \sigma^2_e & \\ & & \sigma^2_e \end{bmatrix} = \sigma^2_e [I]$$

Thus $Var(\hat{e}) = \sigma^2_e (1 - H)(1 - H')$

$= \sigma^2_e (I - H)$

Thus

$Var(\hat{e}_{10}) = \hat{\sigma}^2_e (1 - h_{10,10})$

Regression Diagnostics with hat matrix.

Standardized Residual. $= \dfrac{\hat{e_i}}{\sqrt{\hat{\sigma}_e^2}} = \dfrac{\hat{e_i}}{\hat{\sigma}_e}$

Studentized residual $r_i = \dfrac{\hat{e_i}}{\hat{\sigma}_e \sqrt{1-h_{ii}}}$

R-student residual.
external estimate. use $\hat{\sigma}_e^2(-i)$ $= \dfrac{\hat{e_i}}{\hat{\sigma}_{e(-i)} \sqrt{1-h_{ii}}}$

$$\hat{\sigma}^2_{e(-i)} = \frac{(N-p)\hat{\sigma}_e^2 - \frac{e_i^2}{(1-h_{ii})}}{N-p-1}$$

Press Residual.

- drop ith observation from data.
- Recalculate Regression Model.
- Use X_i to Predict y_i.
- $\hat{e}_{(i)}$ is the $\text{PRESS} = \sum_{i=1}^{N} \hat{e}^2_{(i)}$

$$R^2_{\text{prediction}} = 1 - \frac{\text{PRESS}}{\text{TOT Sum of Squares}}$$

Inspecting H For Leverage points:

h_{ii}, the diagonal elements help us understand the effect of ith observation on ith Predicted value.

DFFITs

Compute change in predicted value of the jth observation due to deletion of jth observation from dataset.

$$DFFIT_j = \hat{y}_j - \hat{y}(-j)$$

$$\hat{y}_i = [X][\beta]$$

$$y$$

(a standardised approach.

$$DFFITS_j = \frac{\hat{y}_j - \hat{y}(-j)}{\hat{\sigma}_e(-j)\sqrt{h_{jj}}}$$

$$y = X\beta$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_n \end{bmatrix}_{n \times 1} = [\ ]_{n \times (k+1)} [\beta]$$

if $DFFITS_j$ is Large, jth observation is influential

DFBETA

Let $\hat{\beta}(-j)$ represent vector of estimates based on dataset with ~~day~~ observation 'j' omitted. $\hat{y}_2$ $\beta(-j)$ $(k+1)$

$$d_j = \hat{\beta} - \hat{\beta}(-j) \quad \hat{y} \quad [\beta]$$

$$\text{jth} \quad \hat{y}_2 = [\ ]$$

$$\hat{y} \quad \beta$$

Standardise:] Consider estimate of parameter 'i' when jth observation is omitted.

$$d[i]_j^* = \frac{d[i]_j}{\sqrt{Var(\hat{\beta}_i(-j))}}$$

$$d[i]_j = \frac{\hat{e}(X'X)^{-1}x_j}{1 - h_{ii}}$$

$$[X][\beta_{(-j)}]$$

Cook's distance:

Q] is the vector of estimates obtained with observation 'j' omitted, meaningfully different from the vector obtained when all observations are used.

Cook's distance in essence measures – overall distance between

$$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2 \dots \hat{\beta}_p)$$

and $$\hat{\beta}_{(-j)} = (\hat{\beta}_{1(-j)}, \hat{\beta}_{2(-j)} \dots)$$

in the

Square of total distance =

$$(\hat{\beta}(-j) - \hat{\beta})'(\hat{\beta}(-j) - \hat{\beta})$$

↳ we have

$$D_j = \frac{X[(\beta_{-j}) - \hat{\beta}]' [X \cdot (\hat{\beta}_{(-j)} - \hat{\beta})]}{p \cdot \hat{\sigma}^2_\epsilon}$$

Myers, Montgomery, Vining,

$$D_j = \frac{r_j^2 \; h_{jj}}{p\,(1-h_{jj})}$$