Parameter estimation in SLR (MLR the same way).

$$Y_i = b_0 + b_1 x_i + e_i$$

LSE: minimize error

MLE: maximize likelihood (probability to happen)

LSE: minimize e_i → many ways ⇒ one is $\sum e_i^2$ (Gates).

$$\sum e_i^2 = \sum (Y_i - b_0 - b_1 x_i)^2$$

if $\sum e_i^2$ minimum then partial derivatives in respect with b_0 & b_1 are 0.

(1) $$\frac{\partial \sum e_i^2}{\partial b_0} = \frac{\partial \sum (Y_i - b_0 - b_1 x_i)^2}{\partial b_0} = 2 \cdot \sum (Y_i - b_0 - b_1 x_i) \cdot (-1).$$

(2) $$\frac{\partial \sum e_i^2}{\partial b_0} = \frac{\partial \sum (Y_i - b_0 - b_1 X_i)^2}{\partial b_1} = 2 \sum x_i (Y_i - b_0 - b_1 x_i) \cdot (-1)$$

From (1) ⇒ $\sum (Y_i - b_0 - b_1 x_i) = 0$ ⇒ $\sum Y_i = n b_0 + b_1 \sum x_i$

From (2) ⇒ $\sum (x_i Y_i - b_0 x_i - b_1 x_i^2) = 0$. $\sum x_i Y_i = b_0 \sum x_i + b_1 \sum x_i^2$

Solve for b_1: $n \sum x_i Y_i - \sum x_i \sum Y_i = b_1 \left(n \sum x_i^2 - (\sum x_i)^2\right)$

$$b_1 = \frac{\sum x_i Y_i - \frac{\sum x_i \sum Y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

$$b_0 = \frac{\sum Y_i}{n} - b_1 \frac{\sum x_i}{n} = \bar{Y} - b_1 \bar{x}$$

$$b_0 = \frac{\sum y_i}{n} - b_1 \cdot \frac{\sum x_i}{n} = \frac{\sum y_i}{n} - \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} \cdot \frac{\sum x_i}{n} =$$

$$= \frac{\sum y_i\left(n\sum x_i^2 - (\sum x_i)^2\right) - \sum(x_i)\left(n\sum x_i y_i - \sum x_i \sum y_i\right)}{n\left(n\sum x_i^2 - (\sum x_i)^2\right)} =$$

$$= \frac{\cancel{n}\sum y_i \sum x_i^2 - \cancel{\sum y_i (\sum x_i)^2} - \cancel{n}(\sum x_i)\sum x_i y_i + \cancel{\sum y_i (\sum x_i)^2}}{\cancel{n}\left(n\sum x_i^2 - (\sum x_i)^2\right)} =$$

$$= \frac{\sum y_i \cdot \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

---

properties:

① $E[e_i] = E[Y_i - b_0 - b_1 x_i] = E[Y_i] - \beta_0 - \beta_1 E[x_i] =$

$= \bar{Y} - \beta_0 - \beta_1 \bar{X} = 0$

therefore LSE has $E[e_i] = 0$ → unbiased

② $E[Y_i] = E[b_0 + b_1 X + \varepsilon_i] = \beta_0 + \beta_1 E[x_i] + E[\varepsilon_i] \Rightarrow$

$E[Y_i] = \beta_0 + \beta_1 x_i$ → mean of Y for a given X

③ If variance is constant across X then:

$var[Y_i] = var[b_0 + b_1 x + \varepsilon_i] = var[\varepsilon_i] = \sigma^2$

if $var(\varepsilon_i) = \sigma^2$

this is true if the errors are uncorrelated.

④ $\sum Y_i = \sum \hat{Y}_i$

$$\sum \hat{Y}_i = \sum (b_0 + b_1 x_i) = \sum (\bar{Y} - b_1 \bar{x} + b_1 x_i) =$$

$$= \sum (\bar{Y} + b_1 (x_i - \bar{x})) = \sum \bar{Y} + b_1 \sum (x_i - \bar{x}) =$$

$$= \sum \bar{Y} + b_1 \left(\sum x_i - \sum \frac{\sum x_i}{n} \right) = \sum \bar{Y} + b_1 \left(\sum x_i - \sum x_i \right) = \sum \bar{Y}$$

⑤ Regression line goes thru $(\bar{x}, \bar{y})$

or $(\bar{x}, \bar{y})$ is on the regression line

$$\hat{y} = b_0 + b_1 x \Rightarrow \bar{y} = b_0 + b_1 x \Rightarrow$$

$$\Rightarrow x = \frac{\bar{y} - b_0}{b_1} = \frac{\bar{y} - (\bar{y} - b_1 \bar{x})}{b_1} = \bar{x}$$

Variance of parameters.

$$\sigma^2(b_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

$$\sigma^2(b_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right]$$

$$\frac{b_1 - \beta_1}{s(b_1)} \sim t(n-2) \quad \& \quad \frac{b_0 - \beta_0}{s(b_0)} \sim t(n-2)$$

Variance of e

$$MSE = \frac{SSE}{n-2} = \frac{\sum (Y_i - \hat{Y})^2}{n-2} = \frac{\sum e_i^2}{n-2}$$ is minimal

$$E[MSE] = \sigma^2 \quad \text{and} \quad E[\sqrt{MSE}] = \sigma$$

Note: if $e \sim N(0, \sigma^2)$ then uncorrelation leads to independence

Interval estimation $\Rightarrow$ multiple records for same X

$\hat{Y}_h \to$ distribution of Y given X_h.

$$E[\hat{Y}_h] = E[Y_h] = E[b_0 + b_1 X_h] = \beta_0 + \beta_1 X_h.$$

$$\sigma^2[\hat{Y}_h] = \sigma^2 \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right].$$

$\hat{Y}_h$ is normal because $\hat{Y}_h$ linear combination of $Y_i \sim N$.

Notes: b_1 & $\bar{Y}$ are uncorrelated.

$$\sigma(\bar{Y}, b_1) = 0.$$

$$\frac{\hat{Y}_h - E[\hat{Y}_h]}{s[\hat{Y}_h]} \sim t(n-2).$$

Prediction of a new observation

$$E[Y_{h,new}] = E[Y_h] = b_0 + b_1 X_h.$$

$$\sigma^2(\text{pred}) = \sigma^2(Y_{h,new} - \hat{Y}_h) = \sigma^2(Y_{h\,new}) + \sigma^2(\hat{Y}_h) =$$

$$= \sigma^2 + \sigma^2(\hat{Y}_h)$$

$$E[s^2[\text{pred}]] = \sigma^2(\text{pred}) = MSE + s^2(\hat{Y}_h).$$

$$\frac{Y_{h,new} - \hat{Y}_h}{s(\text{pred})} \sim t(n-2).$$

Confidence band for Regression line.

$E\{Y\} = \beta_0 + B_1 X$

$\hat{Y}_h \pm W\, s(\hat{Y}_h)$

wher W – Working-Hotelling value

$W^2 = 2F_{2, n-2\,;\,1-\alpha}$

Diagnostic and Remedial measures.

① Graphical

② Analytical

① Graphical → Plot
→ Box-Plot

② Analytical → Residuals.

Important statistic: Semi-studentized Residual

$$e_i^* = \frac{e_i - \bar{e}}{\sqrt{MSE}} = \frac{e_i}{\sqrt{MSE}}$$

Semistudentized because $\sqrt{MSE} \neq \sigma(e_i)$ it is just an approximation

Used in assessment of 6 assumptions:

① → departure from linear.

② → non-constant variance → heteroskedasticity.

③ → error terms are not independent.

(4) presence of outliers
(5) error is not normally distributed
(6) predictor variables.

Visual diagnostic of residuals
a) residual vs. $\hat{Y}$: e_i vs. $\hat{Y}_i$
b) $|e_i|$ & e_i^2 vs $\hat{Y}$
c) Box Plot
d) Normal probability plot

Tests involving residuals.
1) → Randomness ⇒ Durbin - Watson
2) → Homoskedasticity = constant variance
- Levene & Breush-Pagan
3) → Outliers → multiple test (
4) → Normality ⇒ Shapiro - Wilk
→ Kolmogorov - Smirnov.

To fulfill assumptions most of the time we apply transformations of the variables.
→ transformation of Y ⇒ most of the time solve the problems encountered by linear relationship
→ biased results

→ transformation of X → X becomes f(x) then $f(x) = x'$ and linear model is applied.

Matrix Approach to SLR.

$A = [a_{ij}]$

$$\begin{array}{c} \\ A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}. \\ \uparrow \\ \text{columns.} \end{array} \quad \leftarrow \text{rows}$$

vector a matrix with one column.

transpose $(a_{ij})' = (a_{ji})$

equality of matrices $A = B \Rightarrow a_{ij} = b_{ij} \ \forall i,j$.

$A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$

$A - B = (a_{ij}) - (b_{ij}) = (a_{ij} - b_{ij})$.

$\lambda A = A \cdot \lambda = (\lambda a_{ij})$

$A \cdot B = (a_{ij})(b_{jk}) = (\sum_j a_{ij} b_{jk}) = (c_{ik})$

Inverse of a matrix: $A^{-1} \cdot A = A \cdot A^{-1} = I$ (A square)

Determinant of a matrix $\det(A) = \sum_{p \in P} (\operatorname{sign} p \prod a_{i,pi})$

where p_i is a permutation of the # lines (rows).

P all permutations with # columns (rows)

if $\det(A) = 0 \Rightarrow$ matrix is singular $\Rightarrow$ no inverse.

Special matrices:

① Symmetric matrices: $a_{ij} = a_{ji} \Rightarrow A = A^t$.

②. diagonal matrix $a_{ij} = 0$ if $i \neq j$

identity matrix : $\begin{cases} a_{ij} = 0 & i \neq j \\ a_{ii} = 1. \end{cases}$

symbolized as I.

scalar matrix : $\lambda \cdot I$

③ All elements unity : $a_{ij} = 1 \Rightarrow$ symbolized as $\mathbb{1}$ or J

Linear dependencies in a matrix

Let $A = (a_{ij})$ and C_j is a column in A.
If there are λ_j such that not λ_j are 0 and

$\sum \lambda_j c_j = 0$ then columnvectors are linear dependent.

Otherwise vectors are <u>linear independent</u>

The same result holds for rows.
Rank of a matrix is the maximum number of linear independent rows (columns).
Rank is unique for a matrix

$\det A \neq 0$ if $\operatorname{rank}(A) = n$
remember A is square for an inverse to exist.

Regression in matrix form.

$y_i = b_0 + b_1 x_i + \varepsilon_i$

$Y = (y_i)$

$X_i = (x_i)$

$\varepsilon_i = (\varepsilon_i)$

$b = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$

$Y = Xb + \varepsilon$

$E\{Y\} = (E\{Y_{ij}\})$

$E[\varepsilon] = (0)$

$$\sigma^2[Y] = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma^2(Y_1, Y_2) & \cdots \\ \sigma^2(Y_2, Y_1) & \sigma^2(Y_2) & \\ \vdots & & \end{bmatrix}$$

$\sigma^2(Y_1, Y_2) = \sigma^2(Y_2, Y_1)$ $\Rightarrow \sigma^2\{Y\} \Rightarrow$ symmetric matrix

$$\sigma^2\{Y\} = E\left[[Y - E\{Y\}] \cdot [Y - E\{Y\}]^t\right] =$$

$$= E\left[\begin{bmatrix} Y_1 - E\{Y_1\} \\ Y_2 - E\{Y_2\} \\ \vdots \end{bmatrix} \begin{bmatrix} Y_1 - E\{Y_1\} & Y_2 - E\{Y_2\} & \cdots \end{bmatrix}\right]$$

in normal regression model:

$\sigma^2[\varepsilon] = \sigma^2 \cdot I$

<u>OLS estimation</u> in matrix form.

$Y = Xb$

$X^t Y = X^t X b$. multiply with X^t to have a square matrix $\Rightarrow$ inverse

$(X^t X)^{-1} X^t Y = (X^t X)^{-1} (X^t X) \cdot b = b$

$b = (X^t X)^{-1} (X^t Y)$

Fitted values: $\hat{Y} = Xb$

Hat matrix $\Rightarrow$ plays a significant role in assessment.

$\hat{Y} = Xb = \underbrace{X \cdot (X^t X)^{-1} (X^t}) Y) \Rightarrow$

$H = X (X^t X)^{-1} X^t$ $\Rightarrow$ symmetric

$\Rightarrow$ idempotent: $H \cdot H = H$

Residuals

$$e = Y_i - \hat{Y}_i$$

$$e = Y - \hat{Y} = Y - Xb = Y - HY = (I-H)Y$$

$$\sigma^2\{e\} = (I-H)\,\sigma^2\,(I-H)^t$$

But

$\sigma^2\{Y\} = \sigma^2\{e\} = \sigma^2 I$ → according to normality and indep.

$(I-H)^t = (I-H)$ as both are symetric

$$\sigma^2\{e\} = \sigma^2 (I-H)\, I\, (I-H) = \sigma^2 (I-H)(I-H)$$

$$s^2\{e\} = MSE(I-H)$$

Abstract algebra or real numbers vs. matrices

Ring.

Let K be a set on which two operations are defined; called for convenience addition and multiplication:

if $x, y \in K \Rightarrow$ $x+y$ is addition and
$x \cdot y$ is multiplication

If the following properties of the $(K, +, \cdot)$ occurs. then $(K, +, \cdot)$ is called a ring.

The properties are:

a) K is a commutative group under addition
b). multiplication is associative
c) the distributive laws holds

a) commutative group under +

1) + is associative: $(x+y)+z = x+(y+z)$

2) neutral element for +: $0+x = x+0 = x$

3) $\exists$ inverse for +: $x+(-x) = (-x)+x = 0$

where $-x$ is the inverse of x under +

0 is the neutral element.

4). + is commutative: $x+y = y+x$

b). associative under $*$

5) $\cdot$ is associative: $(x \cdot y) \cdot z = x \cdot (y \cdot z)$

c) distributive of $\cdot$ with $\cdot$

6) $x \cdot (y+z) = xy + xz$

7) $(x+y) \cdot z = xz + yz$

A ring that satisfies the conditions 1-7 and also the following three conditions is a FIELD

8) $\exists$ neutral element for $\cdot$: $1 \cdot x = x \cdot 1 = x$

9) $\exists$ inverse for $\cdot$: $x \cdot x^{-1} = x^{-1} \cdot x = 1$

where x^{-1} is the inverse under $\cdot$ and

1 is the neutral element under $\cdot$

10). $1 \neq 0$ the neutral elements are differents for the two operations.

Observation: Fields $\subset$ Rings:

any field is a ring.

Notes 1) real numbers $\to R$ form a field ~~with~~ + & •

2) matrices form a ring ~~with~~ + & •

3) set square matrices of order n with determinant non zero form a field.

> set of diagonal matrices of order n with elements from R of $\mathbb{C}$ is a field.

4) differences: Real field and matrix field:

1) non-commutative

5) similarities: a square matrix = a real (complex) number as a set.