B A B I I

T E O R I   P E N U N J A N G

2.1 *ESTIMASI - M*

<u>Definisi 2.1 :</u>

Estimator $T_n(x_1,\ldots,x_n)$ adalah estimator M untuk suatu fungsi ρ dan sampel $x_1,x_2,\ldots,x_n$, jika terdapat harga t yang meminimalkan fungsi obyektif $\sum_{i=1}^{n}\rho(x_i;t)$ Jika diketahui ψ adalah derivatif dari ρ ke t, maka harga T dapat dihitung dengan jalan meminimalkan harga t yang memenuhi :

$$\sum_{i=1}^{n}\psi\,(x_i;t) = 0 \qquad \ldots\ldots.(2.1)$$

Estimator M merupakan peminimalan fungsi-fungsi deviasi dari observasi yang lebih umum dari pada estimasi jumlah kwadrat deviasi atau jumlah harga mutlak deviasi.

Pada estimasi M, mean dan median merupakan suatu kasus khusus. Untuk lebih jelasnya akan diuraikan sebagai berikut :

Pada estimasi kwadrat terkecil, fungsi ρ adalah kwadrat residu :

$$\rho\,(x\;;\;t)= (x - t)^2 \qquad \ldots.(2.2)$$

Untuk meminimalkan $\sum_{i=1}^{n}\rho(x_i,t)$ maka harus diminimalkan $\sum_{i=1}^{n}(x_i-t)^2$ dengan jalan mendifferensialkannya ke t :

$$\sum_{i=1}^{n} \rho(x_i - t) = \sum_{i=1}^{n} (x_i - t)^2$$

$$\sum_{i=1}^{n} \rho'(x_i;t) = 2 \sum_{i=1}^{n} (x_i - t) = 0$$

$$\sum_{i=1}^{n} (x_i - t) = 0 \qquad ....(2.3)$$

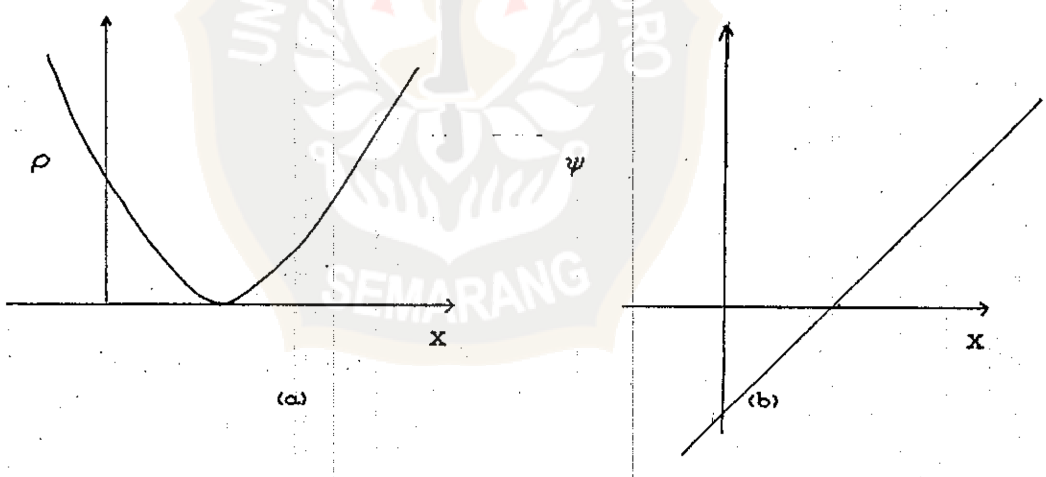
Harga t yang memenuhi persamaan (2.3) adalah

$$t = \frac{\Sigma x_i}{n} .$$

Dengan demikian T_n sesungguhnya adalah mean sampel.

Gambar berikut akan memperlihatkan fungsi ρ dan ψ untuk mean sampel.

<u>GAMBAR 2.1:</u> *Estimasi dengan fungsi obyektif kwadrat residu*



<u>Keterangan Gambar:</u>

Gambar (a) Fungsi obyektif : $\rho(x\ ;\ t) = (x - t)^2$

Gambar (b) Fungsi ψ , dengan $\psi(x\ ;\ t) = x - t$

Sedangkan apabila fungsi obyektif diambil sebagai harga mutlak residu, maka

$$\rho\ (x;t) = |\ x - t\ | \qquad .....(2.4)$$

dan fungsi ψ yang merupakan derivatif ρ ke t adalah

$$\psi(x;t) = \operatorname{sgn}(x - t) \qquad \text{.........(2.5)}$$

dengan

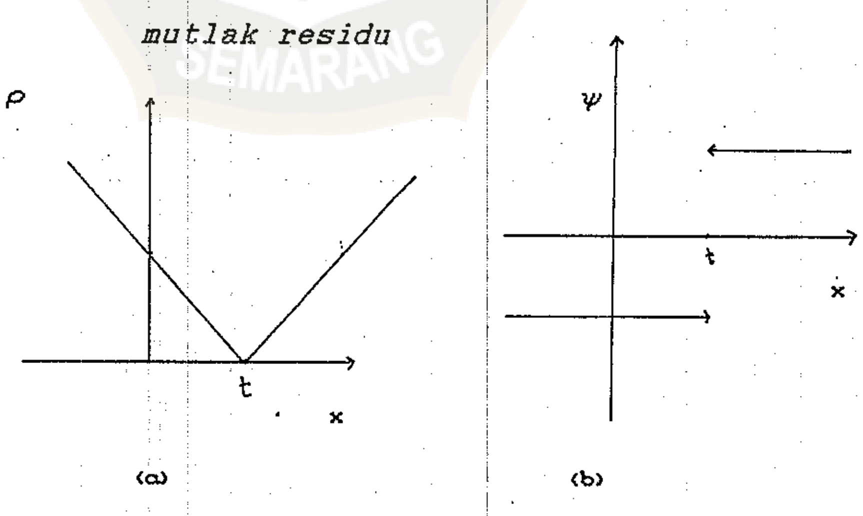
$$\operatorname{sgn}(u) = \begin{cases} 1 & \text{jika } u > 0 \\ 0 & \text{jika } u = 0 \\ -1 & \text{jika } u < 0 \end{cases}$$

Sehingga didapatkan persamaan :

$$\sum_{i=1}^{n} \psi(x_i;t) = \sum_{i=1}^{n} \operatorname{sgn}(x_i - t) \qquad \text{......(2.6)}$$

Persamaan (2.6) memasangkan setiap x yang lebih besar dari t pada $\psi(x) = 1$ dan setiap x yang lebih kecil dari t pada $\psi(x) = -1$, sehingga dapat dilihat bahwa t adalah median sampel.

GAMBAR 2.2 : *Estimasi dengan fungsi obyektif harga mutlak residu*



Keterangan gambar :

Gambar (a) Fungsi obyektif $\rho(x;t) = | x - t |$

Gambar (b) Fungsi ψ, dengan $\psi(x;t) = \operatorname{sgn}(x - t)$

2.2 *UKURAN LOKASI DAN UKURAN SKALA*

Yang termasuk dalam ukuran lokasi atau ukuran tendensi sentral antara lain mean, median, modus. Sedangkan yang termasuk dalam ukuran skala, misalnya jangkauan, varian dan standar deviasi.

Misalkan $x_1, x_2, \ldots, x_n$ adalah variabel acak diskrit dengan nilai kemungkinan masing-masing $p_1, p_2, \ldots, p_n$ dan $\sum_{i=1}^{n} p_i = 1$, maka expektasi matematik dari mean adalah

$$E(x) = \mu = \sum_{i=1}^{n} x_i \cdot p_i \qquad \ldots.(2.7)$$

Misalkan $x_1, x_2, \ldots, x_n$ adalah variabel acak dengan mean μ maka variansi x adalah :

$$Var(x) = \sigma_x^2 = E(x - \mu)^2$$

sedangkan standar deviasi adalah :

$$\sigma_x = \sqrt{Var(x)} = \sqrt{E(x - \mu)^2}$$

Berdasarkan uraian di atas, maka dapat diturunkan sifat-sifat berikut :

Jika a dan b adalah konstanta, maka berlaku :

$$\mu_{ax+b} = a\,\mu_x + b = a\mu + b \qquad ..(2.8)$$

Hal ini dapat dibuktikan demikian :

$$\mu_{ax+b} = E(ax + b) = \sum_{i=1}^{n}(ax_i + b) \cdot p_i$$
$$= (ax_1 + b)p_1 + \ldots + (ax_n + b)p_n$$
$$= a(x_1p_1 + x_2p_2 + \ldots + x_np_n) + b(p_1 + p_2 + \ldots + p_n)$$
$$= a \sum_{i=1}^{n} x_ip_i + b\sum_{i=1}^{n} p_i$$

Karena $\sum_{i=1}^{n} x_i\, p_i = \mu$ dan $\sum_{i=1}^{n} p_i = 1$ maka diperoleh

$$\mu_{ax+b} = a\,\mu + b$$

Estimator yang mempunyai sifat seperti ini disebut estimator *ekuivarian lokasi dan skala.*

Jika b adalah konstanta maka berlaku :

$$\sigma^2_{x+b} = \sigma^2_x = \sigma^2 \qquad ....(2.9)$$

Hal ini dapat dibuktikan demikian :

$$\sigma^2_{x+b} = E\{(x + b) - \mu_{x+b})^2\}$$
$$= E\{(x + b - \mu - b)^2\}$$
$$= E\{(x - \mu)^2\}$$
$$= \sigma^2$$

Estimator yang memiliki sifat demikian disebut estimator *invarian lokasi.*

2.3 *METODE KWADRAT TERKECIL*

Model regresi linier biasa dapat ditulis dalam bentuk :

$$Y = \beta_o + \beta_1 x + \varepsilon \qquad \text{....(2.10)}$$

Dimana β_o merupakan jarak titik asal O dengan perpotongan antara sumbu Y dan garis fungsi liniernya. β_1 merupakan koefisien arah, koefisien regresi, besarnya pengaruh X terhadap Y, sedang ε merupakan komponen error random yang diasumsikan memiliki mean nol dan varian konstan.

Diberikan pasangan data $(y_1,x_1),,...,(y_n,x_n)$ yang didapatkan dari suatu eksperimen. Dari data yang tersedia dapat ditemukan harga-harga $\hat{\beta}_o$ dan $\hat{\beta}_1$ yang merupakan harga estimasi β_o dan β_1. Sehingga persamaan (2.13) berubah menjadi :

$$\hat{y}_i = \hat{\beta}_o + \hat{\beta}_1 x_i \qquad \text{...(2.11)}$$

Untuk mendapatkan β_o dan β_1 digunakan metode kwadrat terkecil yaitu :

$$S(\beta_o,\beta_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$S(\beta_o,\beta_1) = \sum_{i=1}^{n} (y_i - \hat{\beta}_o + \hat{\beta}_1 x_i)^2 \qquad \text{...(2.12)}$$

$$\frac{\partial S}{\partial \beta_o} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_o + \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial S}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i)\ x_i = 0$$

Didapatkan persamaan normal sebagai berikut :

$$n\hat{\beta}_0 + \hat{\beta}_1 \Sigma x_i = \Sigma y_i$$
$$\hat{\beta}_0 \Sigma x_i + \hat{\beta}_1 \Sigma x_i^2 = \Sigma y_i x_i \qquad \text{...(2.13)}$$

Solusi dari persamaan normal :

$$n\ \hat{\beta}_0 = \Sigma y_i - \hat{\beta}_1 \Sigma x_i$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \text{dengan } \bar{y} = \frac{\Sigma y_i}{n},\ \bar{x} = \frac{\Sigma x_i}{n}$$

dan $$\hat{\beta}_1 = \frac{n\ \Sigma y_i x_i - \Sigma y_i\ \Sigma x_i}{n\ \Sigma x_i^2 - (\Sigma x_i)^2} \quad \text{atau}$$

$$\hat{\beta}_1 = \frac{Sxy}{Sxx}$$

Sehingga model pendekatan regresi linier sederhana menjadi : $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

dengan $\hat{y}$ merupakan harga perkiraan dari y.

Selisih antara harga y dengan $\hat{y}$ disebut residu.

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

Jumlah kwadrat error : $SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Rata-rata kwadrat residu : $MSE = \frac{SSE}{n-2}$

dimana (n-2) adalah derajad kebebasan dari error. MSE adalah merupakan estimasi dari varians (σ^2).

2.4 *MODEL REGRESI LINIER BERGANDA*

Bentuk persamaan regresi linier berganda dengan k regresor adalah

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \qquad \ldots(2.14)$$

Untuk mengestimasi koefisien-koefisien regresinya digunakan metode kwadrat terkecil, dengan asumsi bahwa $E(\varepsilon) = 0$ dan $Var(\varepsilon) = \sigma^2$. Dimisalkan tersedia n observasi dengan $n > k$, sebagai berikut :

Tabel 2.1 : *Data untuk regresi linier berganda*

Observasi i	y	x_1	x_2	...	x_n
1	y_1	x_{11}	x_{12}	...	x_{1k}
2	y_2	x_{21}	x_{22}	...	x_{2k}
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
n	y_n	x_{n1}	x_{n2}	...	x_{nk}

Sesuai dengan model regresi linier pada persamaan (2.14), maka dari tabel dapat dibentuk persamaan :

$$y_i = \beta_o + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i \qquad \ldots(2.15)$$

$$= \beta_o + \sum_{i=1}^{n} \beta_j x_{ij} + \varepsilon_i \quad , i = 1,2 \ldots n$$

Untuk mendapatkan estimasi parameter-parameter β digunakan metode kwadrat terkecil :

$$S(\beta_o, \beta_1, \ldots, \beta_k) = \sum_{i=1}^{n} \varepsilon_i^2 \qquad \ldots \quad (2.16)$$

$$= \sum_{i=1}^{n} (y_i - \beta_o - \sum_{j=1}^{n} \beta_j x_{ij})^2$$

Kemudian fungsi S diturunkan ke β_o, β_1, $\beta_2, \ldots,$ β_k sebagai berikut :

$$\frac{\delta S}{\delta \beta_o} = -2 \sum_{i=1}^{n} \left[y_i - \hat{\beta}_o - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right] = 0 \quad ..(2.17)$$

$$\frac{\delta S}{\delta \beta_1} = -2 \sum_{i=1}^{n} \left[y_i - \hat{\beta}_o - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right] x_{ij} = 0 \quad (2.18)$$

dan seterusnya, dengan $i = 1,2,\ldots,n$, $j = 1,2,\ldots,k$

Selanjutnya didapatkan :

$$\sum_{i=1}^{n} y_i - \hat{\beta}_o - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} = 0$$

$$\sum_{i=1}^{n} y_i - n\,\hat{\beta}_o - [\,\hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \ldots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}\,] = 0$$

$$\sum_{i=1}^{n} y_i = n\hat{\beta}_o + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \ldots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}$$

Sedangkan dari persamaan (2.18) diperoleh :

$$\sum_{i=1}^{n} y_i - \hat{\beta}_o - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \quad x_{ij} = 0$$

$$\Sigma x_{i1} y_i = \hat{\beta}_o \Sigma x_{i1} + \hat{\beta}_1 \Sigma x_{i1}^2 + \ldots + \hat{\beta}_k \Sigma x_{i1} x_{ik}$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\Sigma x_{ik} y_i = \hat{\beta}_o \Sigma x_{ik} + \hat{\beta}_1 \Sigma x_{ik} x_{i1} + \ldots + \hat{\beta}_k \Sigma x_{ik}^2 \quad \ldots\ldots(2.19)$$

Solusi dari persamaan (2.19) merupakan estimator kwadrat terkecil parameter-parameter β -nya. Penyelesaian persamaan tersebut akan menjadi lebih jelas dan cepat jika disajikan dalam bentuk matrik.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_{11} & x_{12} \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \ldots & x_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{dan} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

<u>Keterangan :</u>

y adalah vektor observasi (n x 1)

X adalah matriks (n x p) merupakan matriks dari regresor-regresornya dengan p = k + 1

β adalah vektor (p x 1) merupakan koefisien-koefisien regresi yang harus dicari , dan ε adalah vektor (n x 1) dari random error

Sehingga persamaan (2.19) bila disajikan dalam bentuk matriks menjadi :

$$y = X\beta + \varepsilon$$

Untuk mencari estimator β akan diminimalkan fungsi residu :

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$

$$= y'y - \beta'X'y - y'X\beta + \beta'X'X\beta$$

$$= y'y - \beta'X'y - \beta'X'y + \beta'X'X\beta$$

$$= y'y - 2\beta'X'y + \beta'X'X\beta \qquad ....(2.20)$$

Dan turunan pertama dari persamaan (2.20) adalah :

$$-2\,X'y + 2\,X'X\beta = 0$$

$$X'X\beta = X'y \qquad ....(2.21)$$

Selanjutnya persamaan (2.21) dikalikan invers dari $X'X$ diperoleh :

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

Sehingga diperoleh hasil estimasi β yaitu :

$$\hat{\beta} = (X'X)^{-1} X'y \quad ....(2.22)$$

2.5 *ANALISA RESIDUAL*

<u>Definisi 2.2 :</u>

Didefinisikan residu adalah :

$$e_i = y_i - \hat{y}_i \quad ......(2.23)$$

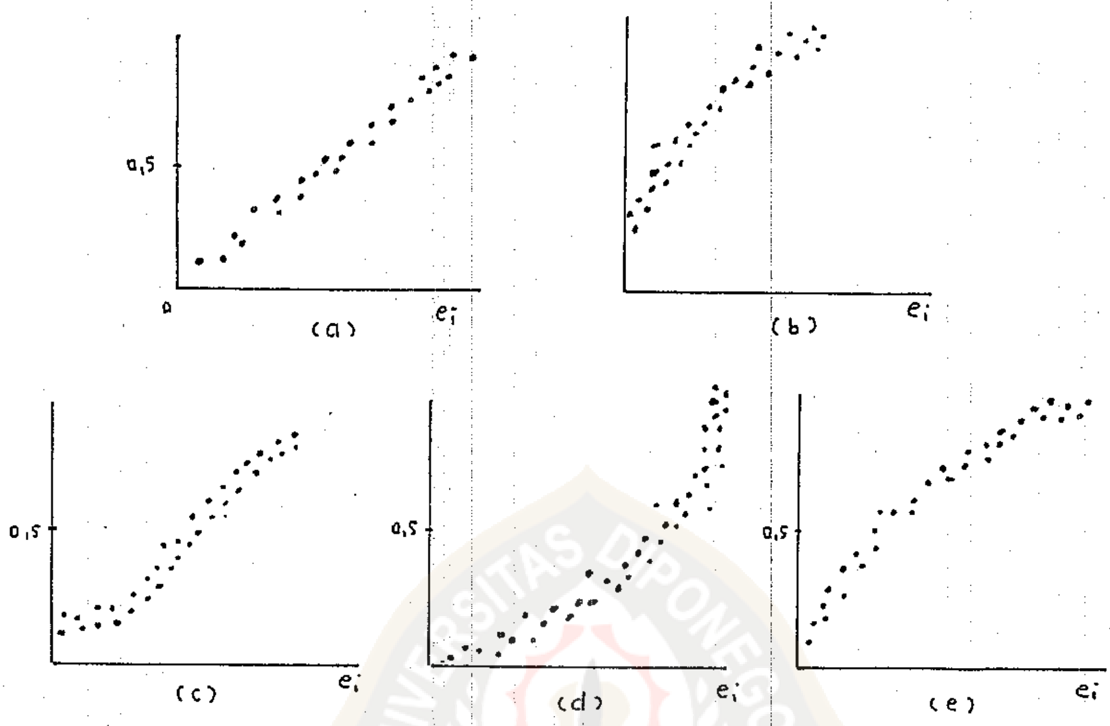
Dimana y_i adalah y observasi dan $\hat{y}_i$ adalah nilai y teoritis berdasarkan persamaan regresi.

Asumsi-asumsi dalam analisa residual :

1. Residu memiliki mean nol
2. Residu berdistribusi normal dan memiliki varians σ^2.

Penggunaan rumus kwadrat terkecil adalah untuk error yang berdistribusi normal. Apabila ternyata error memiliki distribusi non normal maka akan membentuk adanya outlier yang akan mempengaruhi besarnya estimasi $\hat{\beta}_0$ dan $\hat{\beta}_1$. Cara sederhana untuk mengetahui bentuk distribusi adalah dengan membandingkan residu-residu pada plot probabilitas normal. Gambar berikut menunjukkan jenis-jenis plot probabilitas normal.

Gambar 2.4 : *Bentuk plot probabilitas normal*



Keterangan Gambar :

Sumbu horisontal merupakan harga e_i yang sudah terurut dengan $e_1 < e_2 \ldots < e_n$

Sumbu vertikal merupakan expected normal value :

$\varnothing^{-1}$ [(i - 1/2)/n] dimana $\varnothing$ adalah distribusi kumulatip normal dengan i = 1,...,n (diperoleh dari tabel distribusi kumulatip normal).

Gambar (a) :Plot probabilitas normal ideal

Gambar (b) :Plot probabilitas normal untuk heavy-tailed distribution

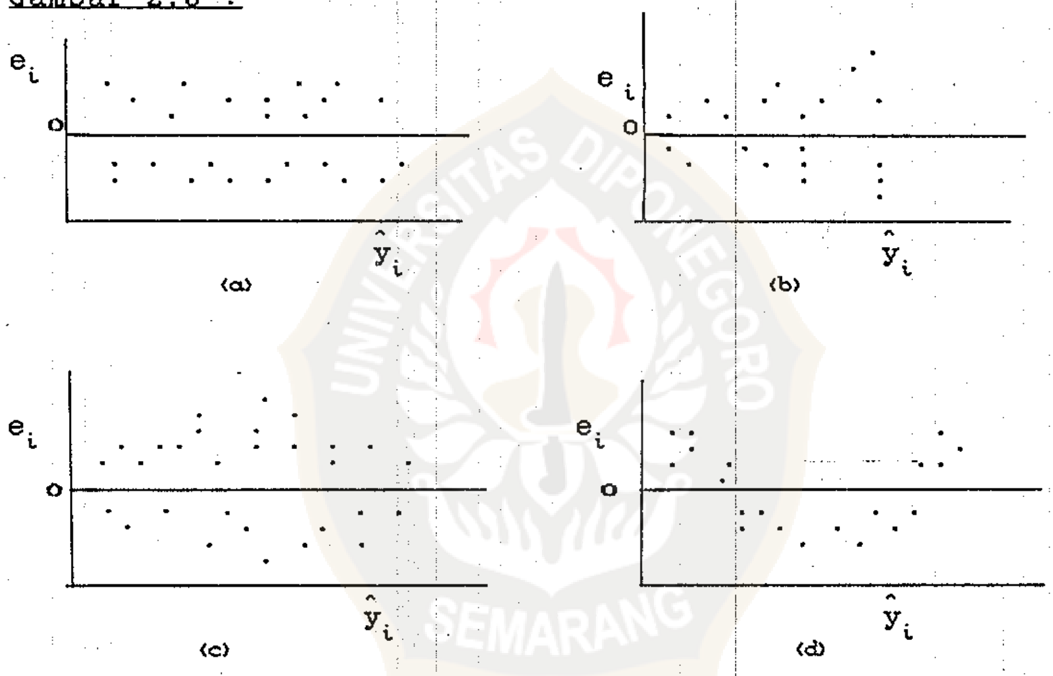
Gambar (c) :Plot probabilitas normal untuk light-tailed distribution

Gambar (d) :Plot probabilitas normal positif skew

Gambar (e) : Plot probabilitas normal negatif skew

Selanjutnya akan dibahas plot yang lain yaitu plot residu melawan $\hat{y}_i$. Suatu plot residu melawan $\hat{y}_i$ dipergunakan untuk mendeteksi beberapa tipe dari suatu model yang tidak sempurna, misalnya tidak memenuhi asumsi bahwa error memiliki varian konstan. Berikut ini disajikan beberapa tipe plot residu melawan $\hat{y}_i$.

Gambar 2.5 :



Keterangan Gambar :

Gambar (a) menunjukkan bahwa residu termuat dalam sebuah pita horinsontal, bentuk ini berarti bahwa error memiliki varian konstan.

Gambar (b) merupakan bentuk *Funel* yang menunjukkan bahwa residu membesar sejalan dengan membesarnya $\hat{y}_i$ sehingga mengakibatkan varian tidak konstan.

Gambar (c) disebut bentuk *double bow*, biasanya terjadi pada harga y yang merupakan proporsi antara nol dan satu. Bentuk ini juga tidak memiliki varian konstan. Gambar (d) menunjukkan *non linier*.

Selanjutnya akan dibicarakan mengenai outlier.

<u>Definisi 2.3 :</u>

Outlier adalah residu yang nilai mutlaknya jauh lebih besar dari pada residu-residu lainnya, bisa sampai tiga atau empat kali simpangan baku.

Untuk mengidentifikasi adanya outlier, dapat dengan menggunakan plot probabilitas normal. Kadang-kadang outlier juga disebut *bad values*, sehingga dapat dikoreksi atau dihilangkan dari perhitungan. Hal ini hanya dapat dilakukan apabila ada alasan yang secara rasional dapat diterima, misalnya karena adanya kesalahan dalam pengukuran. Namun ada kalanya outlier tidak dapat dihilangkan begitu saja, apabila kejadiannya memang demikian, contohnya data tersebut adalah nilai terendah, pendapatan tertinggi dan sebagainya. Sehingga sebelum menerapkan suatu perlakuan pada outlier diperlukan data-data non statistik yang mendukung.

Munculnya outlier akan berpengaruh pada ketepatan garis regresi, sehingga pendeteksian adanya outlier sangat penting. Identifikasi outlier dengan plot probabilitas normal merupakan cara yang mudah.

Pendekatan kwadrat terkecil memberikan harga standard residu yang relatif kecil, sehingga tidak dapat untuk mendeteksi adanya outlier. Sedangkan metode robust akan memberikan harga yang cukup besar, sehingga ada atau tidaknya outlier akan segera terdeteksi.

Contoh 2.1 :

Dalam suatu penelitian didapatkan data mengenai kekuatan roket dan umur bahan bakar (radio aktif) yang digunakan. Diasumsikan model regresi linier :

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

Dari 20 observasi diperoleh data kekuatan roket dalam satuan *psi* dan umur bahan bakar dalam satuan minggu.

Tabel 2.2 : *Data 20 observasi Contoh 2.1*

Obs i	Kekuatan roket y_i	Umur bahan bakar x_i
1	2198,70	15,50
2	1678,15	23,75
3	2316,00	8,00
4	2061,30	17,00
5	2270,50	5,50
6	1708,30	19,00
7	1784,70	24,00
8	2575,00	2,50
9	2357,90	7,50
10	2256,70	11,00
11	2165,20	13,00
12	2399,55	3,75
13	1779,80	25,00
14	2336,75	9,75
15	1765,30	22,00
16	2053,50	18,00
17	2414,40	6,00
18	2200,50	12,50
19	2654,20	2,00
20	1753,70	21.50

Sumber : Introduction to Linear Regression Analysis

oleh: DC. Montgomery dan Elisabeth A. Peck

Untuk mengestimasi parameter-parameter model, pertama-tama dihitung :

$$S_{xx} = n \Sigma x_i^2 - (\Sigma x_i)^2 = 1106,56$$

$$S_{xy} = n \Sigma x_i y_i - \Sigma x_i \Sigma y_i = -41112,65$$

$$\bar{x} = 13,3525 \quad \text{dan} \quad \bar{y} = 2131,3575$$

Sehingga didapatkan :

$$\beta_1 = \frac{-41112,65}{1106,56} = -37,15$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 2627,82$$

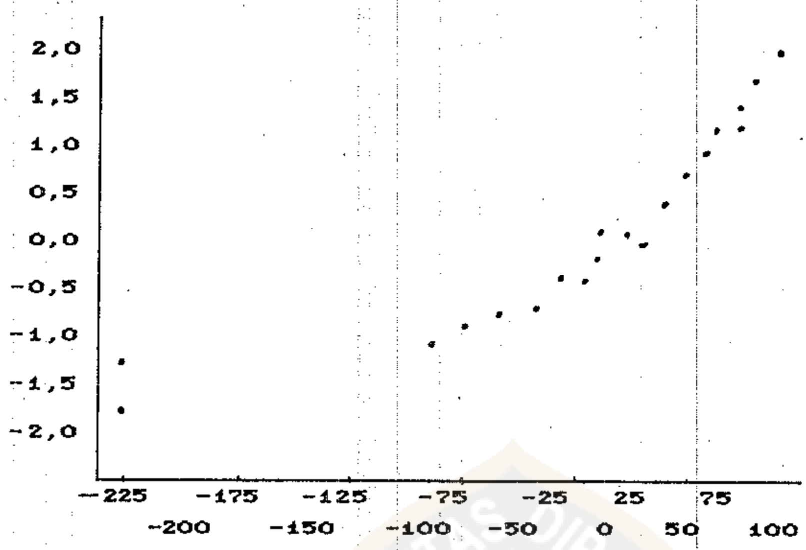
Jadi persamaan regresinya : $\hat{y} = 2627,82 - 37,15\,x$

Tabel 2.2 berikut merupakan perhitungan residu dari masing-masing observasi.

Tabel 2.3: *Hasil Perhitungan Residu Contoh* 2.1

Harga y_i	Harga $\hat{y}_i$	Residu (e_i)
2158,70	2051,94	106,76
1678,15	1745,42	-67,27
2316,00	2330,59	-14,59
2061,30	1996,21	65,09
2207,50	2423,48	-215,98
1708,30	1921,90	-213,60
1784,70	1736,14	48,56
2575,00	2534,94	40,56
2357,90	2349,17	8,73
2256,70	2219,13	37,57
2165,20	2144,83	20,37
2399,55	2488,50	-88,95
1779,80	1698,98	80,82
2336,75	2265,58	71,17
1765,30	1810,44	-45,14
2053,50	1959,06	94,44
7414,40	2040,90	9,50
2200,50	2163,40	37,10
2654,20	2553,52	100,68
1753,70	1829,02	-75,32
42627,15	42627,15	00,00

Gambar 2.6: *Plot probabilitas normal*



Sumber : Introduction to Linear Regression Analysis
oleh DC Montgomery dan Elisabeth A. Peck

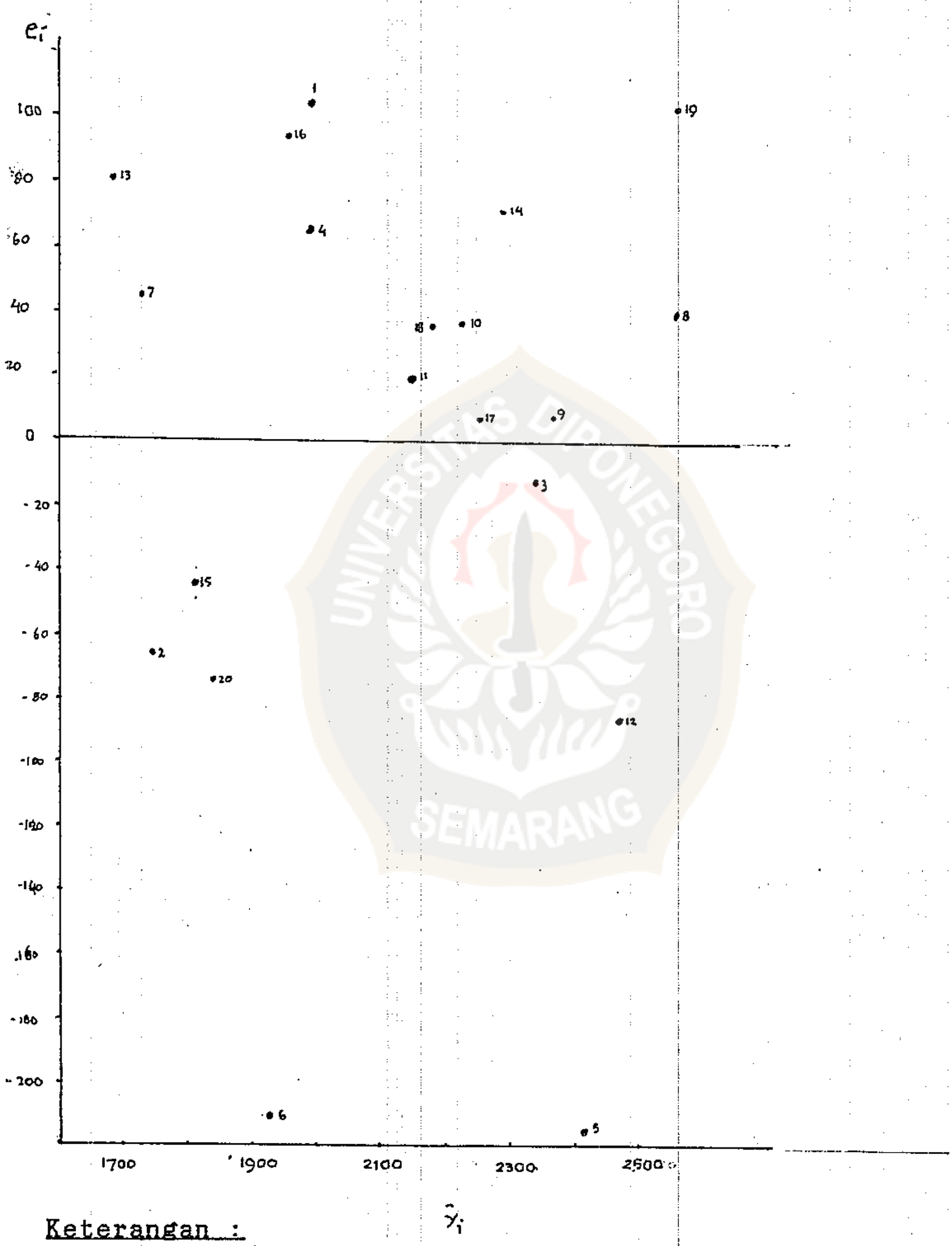
Keterangan :

Sumbu horisontal adalah residu ke-i dan sumbu vertikal adalah expected normal value : $\varnothing^{-1}[(i - 1/2)/n]$ yang diperoleh dari distribusi kumulatip normal.

Selanjutnya dari tabel 2.2 yang memuat hasil perhitungan residu dan $\hat{y}_i$ yang diperoleh dari perhitungan dengan kwadrat terkecil akan dibuat plot residu melawan $\hat{y}_i$ untuk mengetahui apakah residu sudah memenuhi asumsi memiliki varian konstan.

Hasilnya akan disajikan dalam Gambar 2.7 berikut ini :

Gambar 2.7 : *Plot residu malawan* $\hat{y}_i$



Keterangan :

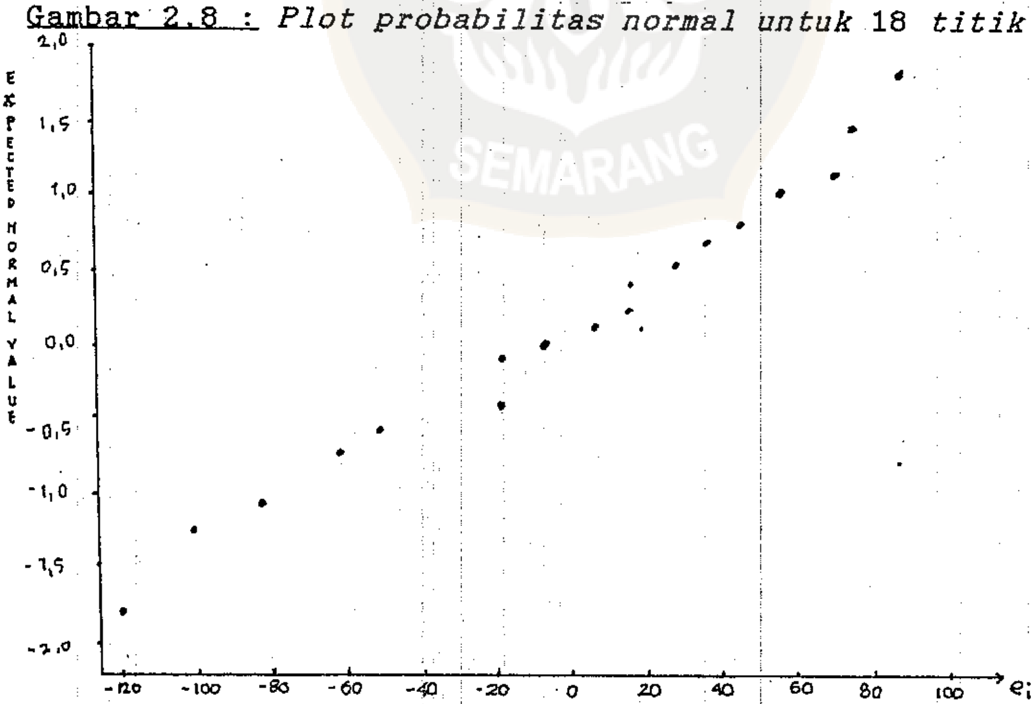
Indeks 1,2,...,20 menunjukkan observasi ke-i

Dari plot probabilitas normal gambar 2.6 tampak bahwa observasi 5 dan 6 terletak paling jauh diantara titik-titik yang lain. Demikian pula pada plot residu melawan $\hat{y}_i$, kedua titik tersebut terletak jauh dari garis nol. Sehingga kedua titik tersebut berpotensi sebagai outlier. Untuk meyelidikinya, dicoba untuk menghitung kembali $\hat{\beta}_0$ dan $\hat{\beta}_1$dengan kwadrat terkecil, diperoleh hasil :

$$\hat{\beta}_0 = 2658,97 \quad \text{dan} \quad \hat{\beta}_1 = -37,69$$

Ternyata menghilangkan titik 5 dan 6 memberikan pengaruh terhadap harga $\hat{\beta}_0$dan $\hat{\beta}_1$. Dan selajutnya akan disajikan plot probilitas normal dan plot residu melawan $\hat{y}_i$.

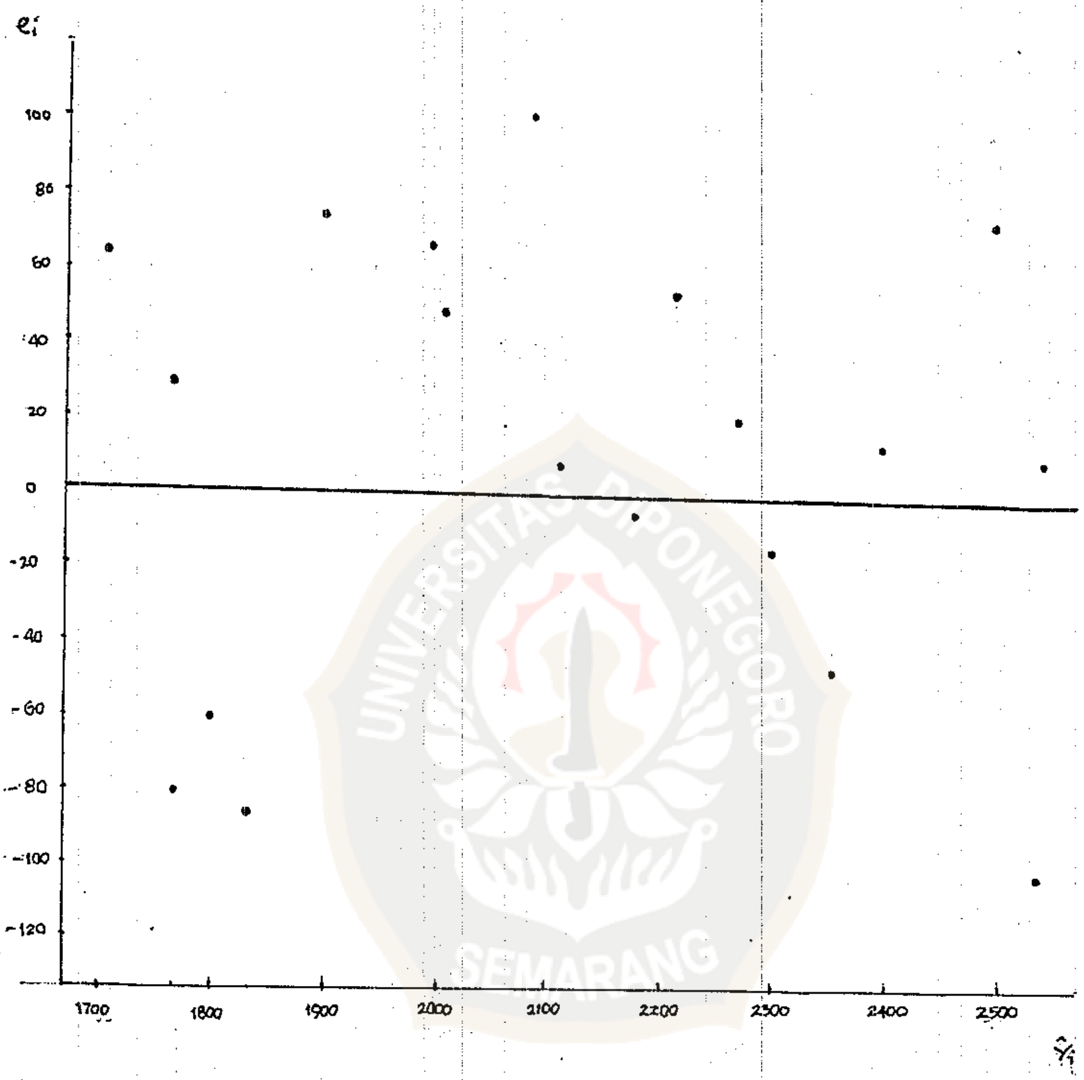
<u>Gambar 2.8 :</u> *Plot probabilitas normal untuk* 18 *titik*



Sumber :Introduction to Linear Regression Analysis

oleh :DC.Montgomery dan Elisabeth A.Peck

Gambar 2.9 : *Plot residu melawan* $\hat{y}_i$ *untuk* 18 *titik*



Ternyata plot probabilitas normal dan plot residu melawan $\hat{y}_i$ memberikan hasil yang lebih baik, karena tidak ada lagi titik-titik yang merupakan outlier. Dengan demikian terdapat persamaan regresi baru yang hanya melalui 18 titik yaitu :

$$\hat{y} = 2658,97 - 37,69\ x$$