REGRESI NON LINIER

A. Macam-macam regresi non linier

1. Parabola kuadratik: $\hat{Y} = a + bX + cX^2$
   Dengan menggunakan metode kuadrat terkecil, maka a, b, dan c dapat dihitung dari sistem persamaan:
$$\sum Y_i = na + b\sum X_i + c\sum X_i^2$$
$$\sum X_iY_i = a\sum X_i + b\sum X_i^2 + c\sum X_i^3$$
$$\sum X_i^2Y_i = a\sum X_i^2 + b\sum X_i^3 + c\sum X_i^4$$

2. Parabolik kubik: $\hat{Y} = a + bX + cX^2 + dX^3$
   Untuk menentukan nilai a, b, c, dan d gunakan sistem persamaan berikut:
$$\sum Y_i = na + b\sum X_i + c\sum X_i^2 + d\sum X_i^3$$
$$\sum X_iY_i = a\sum X_i + b\sum X_i^2 + c\sum X_i^3 + d\sum X_i^4$$
$$\sum X_i^2Y_i = a\sum X_i^2 + b\sum X_i^3 + c\sum X_i^4 + d\sum X_i^5$$
$$\sum X_i^3Y_i = a\sum X_i^3 + b\sum X_i^4 + c\sum X_i^5 + d\sum X_i^6$$

3. Eksponen : $\hat{Y} = ab^X$
   Besar nilai a dan b ditentukan menggunakan persamaan:
$$\log a = \frac{\sum \log Y_i}{n} - (\log b)\left(\frac{\sum X_i}{n}\right)$$
$$\log b = \frac{n(\sum X_i \log Y_i) - (\sum X_i)(\sum \log Y_i)}{n\sum X_i^2 - (\sum X_i)^2}$$

4. Geometrik: $\hat{Y} = aX^b$
   Besar nilai a dan b ditentukan menggunakan persamaan:
$$\log a = \frac{\sum \log Y_i}{n} - b\frac{\sum \log X_i}{n}$$
$$b = \frac{n(\sum \log X_i \log Y_i) - (\sum \log X_i)(\sum \log Y_i)}{n\sum log^2 X_i - (\sum \log X_i)^2}$$

5. Gompertz: $\hat{Y} = pq^{bX}$
6. Logistik: $\hat{Y} = \frac{1}{ab^X + c}$
$$\log a = \frac{\sum \log\left(\frac{1}{Y_i}\right)}{n} - (\log b)\left(\frac{\sum X_i}{n}\right)$$
$$\log b = \frac{n\left(\sum X_i \log\left(\frac{1}{Y_i}\right)\right) - (\sum X_i)\left(\sum \log\left(\frac{1}{Y_i}\right)\right)}{n\sum X_i^2 - (\sum X_i)^2}$$

7. Hiperbola: $\hat{Y} = \frac{1}{a + bX}$
   Jika $\hat{Y}$ tidak ada yang bernilai nol, maka a dan b adalah

$$a = \frac{\left(\sum \frac{1}{Y_i}\right)\left(\sum X_i^2\right) - \left(\sum X_i\right)\left(\sum X_i \frac{1}{Y_i}\right)}{n \sum X_i^2 - \left(\sum X_i\right)^2}$$

$$b = \frac{n \sum X_i \frac{1}{Y_i} - \left(\sum X_i\right)\left(\sum \frac{1}{Y_i}\right)}{n \sum X_i^2 - \left(\sum X_i\right)^2}$$

B. Regresi Linier Ganda

Banyak data pengamatan yang terjadi sebagai akibat lebih dari dua variabel. Misalnya, rata-rata pertambahan berat daging sapi (Y) bergantung pada berat permulaan (X_1), umur sapi ketika pengamatan dimulai dilakukan (X_2), berat makanan yang diberikan setiap hari (X_3) dan mungkin masih ada faktor lain.

Akan ditentukan hubungan antara Y dan $X_1, X_2, \dots, X_k$ sehingga didapat regresi Y atas $X_1, X_2, \dots, X_k$. Yang akan ditinjau di sini hanyalah garis regresi sederhana ialah yang dikenal dengan *regresi linier ganda*. Model tersebut ditaksir oleh:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + \cdots + +a_k X_k$$

Koefisien-koefisien $a_0, a_1, \dots, a_k$ ditentukan dengan menggunakan metode kuadrat terkecil.

Untuk regresi linier ganda dua variabel bebas:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2$$

maka untuk mengetahui koefisien-koefisiennya harus menyelesaikan persamaan-persamaan berikut:

$$\boxed{\begin{aligned} \sum Y_i &= a_0 n + a_1 \sum X_{1_i} + a_2 \sum X_{2_i} \\ \sum Y_i X_{1_i} &= a_0 \sum X_{1_i} + a_1 \sum X_{1_i}^2 + a_2 \sum X_{1_i} X_{2_i} \\ \sum Y_i X_{2_i} &= a_0 \sum X_{2_i} + a_1 \sum X_{1_i} X_{2_i} + a_2 \sum X_{2_i}^2 \end{aligned}} \quad \text{persamaan (*)}$$

Untuk regresi linier ganda tiga variabel bebas:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3$$

Maka untuk mengetahui koefisien-koefisiennya harus menyelesaikan persamaan-persamaan berikut:

$$\boxed{\begin{aligned} \sum Y_i &= a_0 n + a_1 \sum X_{1_i} + a_2 \sum X_{2_i} + a_3 \sum X_{3_i} \\ \sum Y_i X_{1_i} &= a_0 \sum X_{1_i} + a_1 \sum X_{1_i}^2 + a_2 \sum X_{1_i} X_{2_i} + a_3 \sum X_{1_i} X_{3_i} \\ \sum Y_i X_{2_i} &= a_0 \sum X_{2_i} + a_1 \sum X_{1_i} X_{2_i} + a_2 \sum X_{2_i}^2 + a_3 \sum X_{2_i} X_{3_i} \\ \sum Y_i X_{3_i} &= a_0 \sum X_{3_i} + a_1 \sum X_{1_i} X_{3_i} + a_2 \sum X_{2_i} X_{3_i} + a_3 \sum X_{3_i}^2 \end{aligned}} \quad \text{aan(**)}$$

Sistem persamaan (*) dan (**) dapat disederhanakan sedikit, apabila diambil $x_1 = X_1 - \overline{X_1}$, $x_2 = X_2 - \overline{X_2}$, $x_3 = X_3 - \overline{X_3}$ dan $y = Y - \bar{Y}$ maka persamaan regresi linier ganda dua variabel bebas menjadi:

$$y = a_1 x_1 + a_2 x_2$$

Koefisien a_1 dan a_2 dapat dihitung dengan persamaan:

$$\begin{aligned} \sum y_i x_{1_i} &= a_1 \sum x_{1_i}^2 + a_2 \sum x_{1_i} x_{2_i} \\ \sum y_i x_{2_i} &= a_1 \sum x_{1_i} x_{2_i} + a_2 \sum x_{2_i}^2 \end{aligned}$$

sedangkan persamaan regresi linier ganda tiga variabel bebas, menjadi:

$$y = a_1 x_1 + a_2 x_2 + a_2 x_2$$

dan koefisien-koefisiennya dihitung menggunakan persamaan:

$$\sum y_i x_{1_i} = a_1 \sum x_{1\,i}^2 + a_2 \sum x_{1_i} x_{2_i} + a_3 \sum x_{1_i} x_{3_i}$$
$$\sum y_i x_{2_i} = a_1 \sum x_{1_i} x_{2_i} + a_2 \sum x_{2\,i}^2 + a_3 \sum x_{2_i} x_{3_i}$$
$$\sum y_i x_{3_i} = a_1 \sum x_{1_i} x_{3_i} + a_2 \sum x_{2_i} x_{3_i} + a_3 \sum x_{3\,i}^2$$

Selain menggunakan persamaan di atas untuk mencari koefisien-koefisien regresi linier dua variabel bebas dapat juga digunakan:

$$a_0 = \bar{Y} - a_1 \overline{X_1} - a_2 \overline{X_2}$$
$$a_1 = \frac{\left(\sum x_{1\,i}^2\right)\left(\sum x_{1_i} y_i\right) - \left(\sum x_{1_i} x_{2_i}\right)\left(\sum x_{2_i} y_i\right)}{\left(\sum x_{1\,i}^2\right)\left(\sum x_{2\,i}^2\right) - \left(\sum x_{1_i} x_{2_i}\right)^2}$$
$$a_2 = \frac{\left(\sum x_{1\,i}^2\right)\left(\sum x_{2_i} y_i\right) - \left(\sum x_{1_i} x_{2_i}\right)\left(\sum x_{1_i} y_i\right)}{\left(\sum x_{1\,i}^2\right)\left(\sum x_{2\,i}^2\right) - \left(\sum x_{1_i} x_{2_i}\right)^2}$$

Koefisien a_1 merupakan perubahan rata-rata Y unutuk setiap perubahan satuan dalam variabel X_1 apabila $X_2, X_3, \ldots, X_k$ semua dianggap tetap, begitu juga a_2 merupakan perubahan rata-rata Y unutuk setiap perubahan satuan dalam variabel X_2 apabila $X_1, X_3, \ldots, X_k$ semua dianggap tetap dan begitu seterusnya. Jelas bahwa di sini setiap koefisien hanya memberikan gambaran parsial apa yang terjadi pada Y untuk perubahan X yang berhubungan dengan koefisien dimaksud. Karenanya, koefisien-koevisien $a_1, a_2, \ldots, a_k$ disebut pula *koefisien regresi parsil.*

Untuk regresi linier ganda variabel, maka ukuran kekelirua yang digunakan adalah:

$$s_{y.1,2,\ldots,k}^2 = \frac{\sum\left(Y_i - \widehat{Y}_i\right)^2}{n-k-1}$$

dimana Y_i = nilai data hasil pengamatan dan $\widehat{Y}_i$ = nilai harapan yang didapat dari persamaan regresi.

C. Indeks Determinasi

Akan ditinjau seberapa kuat hubungan antara variabel-variabel itu terjadi. Secara umum, untuk pengamatan yang terdiri atas dua variabel X dan Y, kita tinjau hal berikut.

Misalkan persamaan regresi Y atas X, tidak harus linier, yang dihitung dari sampel, berbentuk: $\hat{Y} = f(x)$. Jika regresinya linier, maka $f(X) = a + bX$ dan jika parabola, maka $f(X) = a + bX + cX^2$ dan seterusnya. Apabila $\bar{Y}$ menyatakan rata-rata untuk data variabel Y, maka kita dapat membentuk jumlah kuadrat total, $JK_{tot} = \sum(Y_i - \bar{Y})^2$ dan jumlah kuadrat residu, $JK_{res} = \sum\left(Y_i - \widehat{Y}_i\right)^2$ dengan menggunakan harga-harga $\widehat{Y}_i$ yang didapat dari regresi $\hat{Y} = f(X)$.

Besaran yang ditentukan oleh rumus

$$I = \frac{\sum(Y_i - \bar{Y})^2 - \sum\left(Y_i - \widehat{Y}_i\right)^2}{\sum(Y_i - \bar{Y})^2}$$

Atau

$$I = \frac{JK_{tot} - JK_{res}}{JK_{tot}}$$

dinamakan *indeks determinasi* yang mengukur hubungan antara variabel X dan Y. Jika titik pencar semakin dekat dengan garis regresi, maka harga I makin dekat ke 1. Sebaliknya jika titik-titik pencar semakin jauh dari garis regresi maka harga I makin dekat ke nol, $0 \leq I \leq 1$.

D. Korelasi dalam Regresi Linier

Apabila garis regresi yang terbaik sekumpulan data berbenyuk linier, maka derajat hubungannya akan dinyatakan dengan r dan biasa dinamakan *koefisien korelasi*. Untuk regresi linier Y dan X dalam hal ini I diganti oleh r^2 dan diperoleh:

$$r^2 = \frac{\sum(Y_i - \bar{Y})^2 - \sum\left(Y_i - \widehat{Y}_i\right)^2}{\sum(Y_i - \bar{Y})^2}$$

r^2 dinamakan *koefisien determinasi* atau *koefisien penentu.* Dinamakan demikian karena 100 r^2% dari variasi yang terjadi dalam variabel takbebas Y dapat dijelaskan oleh variabel bebas X dengan adanya regresi linier Y atas X. Harga $\sqrt{1-r^2}$ dinamakan *koefisien aliensi* atau *koefisien perenggangan*. Harga $1-r^2$ sendiri dinamakan *koefisien non determinasi*.

Koefisien korelasi r tentu saja didapat dengan jalan mengambil akar dari r^2.

Pada indeks determinasi dapat dilihat $0 \le r^2 \le 1$ sehingga untuk koefisien korelasi didapat hubungan $-1 \le r \le +1$. Harga r menyatakan

$r = -1$	**Hubungan linier sempurna tak langsung**. Titik-titik yang ditentukan oleh (X_i, Y_i) seluruhnya terletak pada garis regresi dan harga X besar berpasangan dengan Y kecil dan X kecil berpasangan dengan Y besar
$r = +1$	**Hubungan linier sempurna langsung.** Titik-titik yang ditentukan oleh (X_i, Y_i) seluruhnya terletak pada garis regresi dan harga X besar berpasangan dengan Y besar dan X kecil berpasangan dengan Y kecil
$r < 1$	Korelasi tak langsung atau korelasi negatif
$r > 1$	Korelasi langsung atau korelasi positif
$r = 0$	Tidak terdapat hubungan linier antara variabel X dan Y

Untuk keperluan perhitungan koefisien korelasi r berdasarkan sekumpulan data (X_i, Y_i) berukuran n dapat digunakan rumus:

$$r = \frac{n\sum X_iY_i - (\sum X_i)(\sum Y_i)}{\sqrt{\left\{n\sum X_i^2 - (\sum X_i)^2\right\}\left\{n\sum Y_i^2 - (\sum Y_i)^2\right\}}}$$

Bentuk lain dapat pula digunakan:

$$r = \sqrt{1 - \frac{s_{y.x}^2}{s_y^2}}$$

dengan $s_{y.x}$ = kekeliruan baku taksiran dan s_y = simpangan baku untuk variabel Y.

Jika persamaan regresi linier Y atas X telah ditentukan dan sudah didapat koefisien arah,b, maka koefisien determinasi, r^2, dapat ditentukan oleh rumus:

$$r^2 = \frac{b\{n\sum X_iY_i - (\sum X_i)(\sum Y_i)\}}{n\sum Y_i^2 - (\sum Y_i)^2}$$

atau dapat juga menggunakan formula:

$$r = \frac{bs_x}{s_y}$$

dengan s_x simpangan baku untuk variabel X dan s_y simpangan baku untuk variabel Y.

Jika b_1 adalah koefisien arah regresi Y atas X dan b_2 adalah koefisien arah regresi X atas Y untuk data yang sama, maka

$$r^2 = b_1b_2$$

Rumus ini menyatakan bahwa koefisien korelasi r adalah rata-rata ukur daripada koefisien-koefisien arah b_1 dan b_2.