MODUL 4

Analisis Regresi III

Prof. Dr. Zanzawi Soejoeti

PENDAHULUAN

Pembahasan pada Modul 4 mata kuliah *Metode Statistika 2* ini juga merupakan kelanjutan dari Modul 2 dan Modul 3 yang membahas tentang analisis regresi. Kalau pada Modul 2 dan Modul 3 Anda telah mempelajari bagaimana mencari dan mengestimasi suatu model regresi linear sederhana dan aplikasinya untuk melakukan penaksiran peubah tak bebas bila diketahui nilai peubah bebasnya maka pada Modul 4 ini juga Anda akan mempelajari penggunaan dari analisis regresi sederhana dengan lebih luas lagi. Anda akan mempelajari bagaimana membandingkan dua persamaan regresi, bagaimana membandingkan lebih dari dua persamaan regresi, dan bagaimana mentransformasikan variabel yang tidak linear menjadi model regresi linear sederhana. Pembahasan regresi linear pada Modul 4 ini akan ditutup dengan mempelajari bagaimana menerapkan prinsip membangun model regresi linear bila terdapat lebih dari satu variabel bebas, dan ini sering dinamakan sebagai model regresi linear berganda. Setelah mempelajari modul 4 ini, secara khusus Anda diharapkan dapat memperoleh gambaran untuk dapat:

1. membandingkan kesamaan dua model regresi linear sederhana;
2. membandingkan kesamaan lebih dari dua model regresi linear sederhana;
3. memilih transformasi yang cocok dan tepat untuk model regresi yang tak linear menjadi model regresi linear;
4. melakukan analisis regresi linear berganda.

KEGIATAN BELAJAR 1

Membandingkan Beberapa Persamaan Regresi Garis Lurus

A. MEMBANDINGKAN DUA PERSAMAAN REGRESI

Misalkan, seorang pembuat percobaan ingin menentukan bagaimana respons y dipengaruhi oleh dosis x dari masing-masing dua perlakuan yang dapat dibandingkan. Perlakuan 1 digunakan untuk n_1 subjek dalam dosis yang berbeda, dan pengukuran responsnya dicatat. Demikian juga, perlakuan 2 digunakan untuk grup independen yang lain terdiri dari n_2 subjek, dan respons-nya dicatat. Maka, struktur datanya berbentuk, seperti berikut.

Tabel 4.1

Dosis x_1 perlakuan 1	$x_{11} x_{12} \ldots x_{1i} \ldots x_{1n_1}$
Respons y_1	$y_{11} y_{12} \cdots y_{1i} \cdots y_{1n_1}$

Dosis x_2 perlakuan 2	$x_{21} x_{22} \ldots x_{2i} \ldots x_{2n_2}$
Respons y_2	$y_{21} y_{22} \cdots y_{2i} \cdots y_{2n_2}$

Dengan anggapan bahwa hubungan linear sesuai untuk masing-masing perlakuan maka:

Perlakuan 1: $y_{1i} = \alpha_1 + \beta_1 x_{1i} + e_{1i}$; $i = 1,2,\ldots, n_1$

Perlakuan 2: $y_{2i} = \alpha_2 + \beta_2 x_{2i} + e_{2i}$; $i = 1,2,\ldots, n_2$

Kecuali anggapan standar untuk model regresi linear, di sini kita perlu menambahkan anggapan bahwa variansi sesatan kedua model itu sama, jadi

$$\text{var}(e_{1i}) = \text{var}(e_{2i}) = \sigma^2$$

Sering kali kita ingin menguji hipotesis nol bahwa kedua garis regresi itu mempunyai *lerengan yang sama*, yakni, $H_0 : \beta_1 = \beta_2$. Secara grafis, ini ekuivalen dengan hipotesis bahwa kedua garis itu *sejajar*. Langkah pertama analisis ini adalah secara terpisah menaksir persamaan regresi untuk kedua himpunan data itu, kemudian menghitung jumlah kuadrat sesatan masing-masing. Misalkan, $\hat{\beta}_1$ dan $\hat{\beta}_2$ masing-masing menunjukkan penaksir kuadrat terkecil untuk $\hat{\beta}_1$ dan $\hat{\beta}_2$, misalkan JKS (1) dan JKS (2) masing-masing menunjukkan jumlah kuadrat sesatan yang berkaitan. Selanjutnya kita tulis:

$$S_{xx_1} = \sum_{i=1}^{n_1} \left(x_{1i} - \overline{x}_1 \right)^2 \quad , \quad S_{xx_2} = \sum_{i=1}^{n_2} \left(x_{2i} - \overline{x}_2 \right)^2$$

Statistik $\left(\hat{\beta}_1 - \hat{\beta}_2 \right)$ berdistribusi normal dengan mean $\left(\beta_1 - \beta_2 \right)$ dan variansi

$$\sigma^2 \left(\frac{1}{S_{xx_1}} + \frac{1}{S_{xx_2}} \right)$$

Penaksir gabungan untuk σ^2 diberikan dengan

$$\hat{\sigma}^2 = s_{\text{gab}}^2 = \frac{JKS(1) + JKS(2)}{n_1 + n_2 - 4} \; ; \;\; \text{db} = n_1 + n_2 - 4$$

dan

$$\frac{\left(\hat{\beta}_1 - \hat{\beta}_2 \right) - \left(\beta_1 - \beta_2 \right)}{s_{\text{gab}} \sqrt{\frac{1}{S_{xx_1}} + \frac{1}{S_{xx_2}}}}$$

berdistribusi Student's t dengan db = $n_1 + n_2 - 4$.

Uji hipotesis $H_0 : \beta_1 = \beta_2$ didasarkan atas statistik:

$$t = \frac{\left(\hat{\beta}_1 - \hat{\beta}_2 \right)}{s_{\text{gab}} \sqrt{\frac{1}{S_{xx_1}} + \frac{1}{S_{xx_2}}}}$$

dengan db = $n_1 + n_2 - 4$.

Interval kepercayaan $100(1-\alpha)\%$ untuk $\left(\beta_1 - \beta_2 \right)$ adalah

$$\left(\hat{\beta}_1 - \hat{\beta}_2 \right) \pm t_{\frac{\alpha}{2}} s_{\text{gab}} \sqrt{\frac{1}{S_{xx_1}} + \frac{1}{S_{xx_2}}}$$

Contoh 4.1

Dua obat insulin A dan insulin B dipelajari untuk menentukan pengaruhnya dalam menurunkan kadar gula darah pada tikus. Tiga belas tikus yang sejenis dibagi secara random menjadi dua kelompok, yaitu 7 ekor tikus dan 6 ekor tikus. Subjek dalam masing-masing kelompok disuntik dengan insulin A dan insulin B dalam berbagai dosis yang berbeda. Pengurangan dalam kadar gula darahnya adalah sebagai berikut.

Tabel 4.2

Dosis A (x_1)	0,20	0,25	0,25	0,30	0,40	0,50	0,50
Pengurangan kadar gula darah (y_1)	30	26	40	35	54	56	65

Dosis B (x_2)	0,20	0,25	0,30	0,40	0,40	0,50
Pengurangan kadar gula darah (y_2)	23	24	42	49	55	70

Untuk obat insulin A kita hitung:

$$\sum x_1 = 2,40\,; \qquad \sum y_1 = 306\,; \qquad \sum x_1^2 = 0,915\,;$$

$$\sum y_1^2 = 14678$$

$$\sum x_1 y_1 = 115,1$$

$$S_{xx_1} = 0,0921\,; \qquad S_{yy_1} = 1301,4286\,; \qquad S_{xy_1} = 10,1857$$

$$\hat{\beta}_1 = \frac{10,1857}{0,0921} = 110,594; \quad \hat{\alpha}_1 = 5,796$$

$$JKS(1) = 1301,4286 - \frac{(10,1857)^2}{0,0921} = 174,9521$$

Jadi, persamaan regresi taksiran

$$\hat{y}_1 = 5,796 + 110,594 x_1$$

Untuk obat insulin B kita hitung

$$\sum x_2 = 2,05 \ ; \qquad \sum y_2 = 263 \ ; \qquad \sum x_2^2 = 0,7625 \, ;$$

$$\sum y_2^2 = 13195 \, ; \qquad \sum x_2 y_2 = 99,8$$

$$S_{xx_2} = 0,0621 \ ; \qquad S_{yy_2} = 1666,8333 \ ; \qquad S_{xy_2} = 9,9417$$

$$\hat{\beta}_2 = \frac{9,9417}{0,0621} = 160,0918 \ ; \qquad \hat{\alpha}_2 = -10,8647$$

$$JKS(2) = 1666,8333 - \frac{(9,9417)^2}{0,0621} = 75,2488$$

Jadi, persamaan regresi taksiran adalah:

$$\hat{y}_2 = -10,8647 + 160,0918 x_2$$

Selanjutnya untuk menguji bahwa kedua garis regresi itu sejajar, yakni $H_0 : \beta_1 = \beta_2$, kita hitung

$$s_{gab}^2 = \frac{JKS(1) + JKS(2)}{n_1 + n_2 - 4} = \frac{174,9521 + 75,2488}{7 + 6 - 4} = 27,8001$$

Sesatan standar taksiran $\left(\hat{\beta}_1 - \hat{\beta}_2\right)$ adalah ….

$$\sqrt{27,8001\left(\frac{1}{0,0921} + \frac{1}{0,0621}\right)} = 27,3772$$

Maka,

$$t = \frac{110,594 - 160,0918}{27,3772} = -1,81$$

Dengan tingkat signifikansi 5%:

H_0 ditolak jika $t \geq 2,262$ atau $t \leq -2,262$

H_0 tidak ditolak jika $-2,262 < t < 2,262$

Oleh karena t = -1,81 maka H_0 tidak ditolak. Ini berarti kita terima bahwa kedua garis itu sejajar.

Contoh 4.2

Panjang sayap pada berbagai waktu setelah menetas untuk jenis burung A dan burung B tertuang dalam tabel berikut.

Tabel 4.3 (a) Jenis burung A

Umur/hari (x_1)	3	4	5	6	8	9	10	11	12	14	15	16	17
Panjang sayap (y_1) (cm)	1,4	1,5	2,2	2,4	3,1	3,2	3,2	3,9	4,1	4,7	4,5	5,2	5,0

Tabel 4.3 (b) Jenis burung B

Umur/hari (x_2)	3	4	5	7	7	8	9	10	12	14
Panjang sayap (y_2) (cm)	2,1	2,5	3,1	3,0	3,8	3,2	4,3	3,9	4,4	4,8

Untuk jenis burung A kita hitung:

$$\sum x_1 = 130; \qquad \sum x_1^2 = 1562; \qquad \sum y_1 = 44,4;$$

$$\sum y_1^2 = 171,30$$

$$\sum x_1 y_1 = 514,80$$

$$S_{xx_1} = 262,00; \qquad S_{yy_1} = 19,6569; \qquad S_{xy_1} = 70,80$$

$$\hat{\beta}_1 = \frac{70,80}{262,00} = 0,270; \quad \hat{\alpha}_1 = 3,42 - (0,270)(10,0) = 0,72$$

$$JKS(1) = 19,6569 - \frac{(70,80)^2}{262,00} = 0,5247$$

Maka, persamaan regresi taksiran adalah $\hat{y}_1 = 0,72 + 0,270x_1$

Untuk jenis burung B kita hitung:

$\sum x_2 = 79$; $\sum x_2^2 = 733$; $\sum y_2 = 35,1$;

$\sum y_2^2 = 130,05$

$\sum x_2 y_2 = 302,7$

$S_{xx_2} = 108,9$; $S_{yy_2} = 6,85$; $S_{xy_2} = 25,41$

$$\hat{\beta}_1 = \frac{25,41}{108,9} = 0,233 \;;\quad \hat{\alpha}_2 = 1,67$$

$$JKS(2) = 6,85 - \frac{25,41^2}{108,9} = 0,921$$

Maka, persamaan regresi taksiran adalah:

$\hat{y}_2 = 1,67 + 0,233x_2$

Selanjutnya untuk menguji bahwa kedua garis regresi itu mempunyai lerengan yang sama, $H_0 : \beta_1 = \beta_2$, kita hitung:

$$s_{gab}^2 = \frac{JKS(1) + JKS(2)}{n_1 + n_2 - 4} = \frac{0,5247 + 0,921}{21} = 0,0688$$

Sesatan standar taksiran $\left(\hat{\beta}_1 - \hat{\beta}_2\right)$ adalah

$$\sqrt{(0,0688)\left(\frac{1}{262,00} + \frac{1}{108,9}\right)} = 0,0299$$

Maka,

$$t = \frac{0,270 - 0,233}{0,0299} = 1,24$$

Dengan tingkat signifikansi 10%:

H_0 ditolak jika $t \geq 1,721$ atau $t \leq -1,721$

H_0 tidak ditolak jika $-1,721 < t < 1,721$.

Oleh karena t = 1,24 terletak dalam daerah penerimaan maka H_0 diterima (tidak ditolak). Ini berarti kita terima bahwa kedua garis itu mempunyai lerengan yang sama (sejajar).

Dapat juga kita menghitung interval kepercayaan 95% untuk $\left(\hat{\beta}_1 - \hat{\beta}_2\right)$, kita peroleh

$$(0,270 - 0,233) \pm 2,08(0,0299) = 0,037 \pm 0,062 \quad \text{atau} \quad (-0,025; 0,099)$$

Lihat kembali contoh 4.1 jika $H_0 : \beta_1 = \beta_2$ tidak ditolak maka taksiran β berdasarkan $\hat{\beta}_1$ dan $\hat{\beta}_2$ dinamakan *koefisien regresi bersama*, dan dihitung sebagai berikut:

$$\hat{\beta}_{ber} = \frac{S_{xy_1} + S_{xy_2}}{S_{xx_1} + S_{xx_2}}$$

Untuk Contoh 4.1:

$$\hat{\beta}_{ber} = \frac{10,1857 + 9,9417}{0,0921 + 0,0621} = 130,528$$

Untuk contoh 4.2

$$\hat{\beta}_{ber} = \frac{70,80 + 25,41}{262,00 + 108,9} = 0,259$$

Membandingkan Dua Elevasi (Titik Potong dengan Sumbu y)

Jika $H_0 : \beta_1 = \beta_2$ ditolak maka kita dapat menyimpulkan bahwa kedua sampel yang kita punyai diambil dari dua populasi yang berbeda. Tetapi, 2 garis regresi populasi itu disimpulkan sejajar maka mungkin kita ingin menentukan apakah kedua garis regresi populasi itu *berimpit* (yakni, mempunyai elevasi yang sama atau $\alpha_1 = \alpha_2$). Prosedur uji t yang dapat digunakan untuk menguji hipotesis $H_0 : \alpha_1 = \alpha_2$ memanfaatkan koefisien regresi bersama $\hat{\beta}_{ber}$ seperti yang telah kita pelajari sebelumnya. Statistik penguji yang kita gunakan adalah:

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - \hat{\beta}_{ber}(\bar{x}_1 - \bar{x}_2)}{s_{gab}\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1 - \bar{x}_2)}{S_{xx_1} + S_{xx_2}}}}$$

= yang berdistribusi t dengan db $= n_1 + n_2 - 3$

Contoh 4.3

Kembali ke Contoh 4.1 kita peroleh:

$$\overline{y}_1 - \overline{y}_2 = \frac{306}{7} - \frac{263}{6} = 43,71 - 43,83 = -0,12$$

$$\hat{\beta}_{ber}\left(\overline{x}_1 - \overline{x}_2\right) = 130,528\left(\frac{2,40}{7} - \frac{2,05}{6}\right) = 0,155$$

$$S_{gab}\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{\left(\overline{x}_1 - \overline{x}_2\right)^2}{S_{xx_1} + S_{xx_2}}} = 5,2726\sqrt{\frac{1}{7} + \frac{1}{6} + \frac{\left(0,343 - 0,342\right)^2}{0,092 - 0,064}} = 2,9334$$

Jadi,

$$t = \frac{-0,12 - 0,155}{2,9334} = -0,094$$

dengan tingkat signifikansi 5%:
$H_0 : \alpha_1 = \alpha_2$ ditolak jika $t \geq 2,228$ atau $t \leq -2,228$. Oleh karena t = -0,094 maka H_0 tidak ditolak (diterima). Dengan demikian, dapat kita simpulkan bahwa garis regresi populasinya berimpit.

Contoh 4.4

Kembali ke contoh 4.2, kita peroleh $\overline{y}_1 - \overline{y}_2 = 4,44 - 3,51 = 0,93$.

$$\hat{\beta}_{ber}\left(\overline{x}_1 - \overline{x}_2\right) = \left(0,259\right)\left(13,0 - 7,9\right) = 1,3209$$

$$s_{gab}\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{\left(\overline{x}_1 - \overline{x}_2\right)^2}{S_{xx_1} + S_{xx_2}}} = 0,2623\sqrt{\frac{1}{13} + \frac{1}{10} + \frac{\left(13 - 7,9\right)^2}{262,00 + 108,9}}$$
$$= 0,1304$$

Jadi,

$$t = \frac{0,93 - 1,3209}{0,1304} = -2,998$$

Dengan tingkat signifikansi 5%.
$H_0 : \alpha_1 = \alpha_2$ ditolak jika $t \geq 2,086$ atau $t \leq -2,086$. Oleh karena t = -2,998 ada di dalam daerah penolakan maka H_0 ditolak. Kita simpulkan bahwa garis regresi populasinya sejajar, tetapi tidak berimpit.

B. MEMBANDINGKAN LEBIH DARI 2 PERSAMAAN REGRESI

Di sini kita akan mempelajari bagaimana menguji $k(>2)$ garis regresi populasi sejajar, yakni $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k$, dengan alternatif: tidak semua sejajar. Di sini kita juga harus menganggap bahwa variansi sesatan semua model regresi sama, yakin:

$$\operatorname{var}(e_1) = \operatorname{var}(e_2) = \ldots = \operatorname{var}(e_k) = \sigma^2$$

Hitungan-hitungan dasar yang diperlukan untuk membandingkan k garis regresi ini adalah S_{xx}, S_{yy}, S_{xy}, dan JKS serta derajat bebasnya.

Nilai-nilai $S_{xx_1,\ldots,S_{xx_k}}$; juga $S_{yy_1,\ldots,S_{yy_k}}$, dan $S_{xy_1,\ldots,S_{xy_k}}$ dapat kita jumlahkan, dan dari jumlah ini dapat kita hitung jumlah kuadrat sesatan bersama atau JKS_{ber}.

Jadi,

$$JKS_{gab} = JKS(1) + \ldots + JKS_{(k)}$$

$$db_{gab} = (n_1 - 2) + \ldots + (n_k - 2) = \sum_{i=1}^{k} n_i - 2k$$

Nilai-nilai $S_{xx_1,\ldots,S_{xx_k}}$; juga $S_{yy_1,\ldots,S_{yy_k}}$, dan $S_{xy_1,\ldots,S_{xy_k}}$ dapat kita jumlahkan, dan dari jumlah ini dapat kita hitung jumlah kuadrat sesatan bersama atau JKS_{ber}.

Jadi,

$$JKS_{ber} = \left(\sum_{i=1}^{k} S_{yy_i}\right) - \frac{\left(\sum_{i=1}^{k} S_{xy_i}\right)^2}{\left(\sum_{i=1}^{k} S_{xx_i}\right)}$$

$$db_{ber} = \sum_{i=1}^{k} n_i - k - 1$$

Untuk menguji $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k$ kita gunakan statistik penguji.

$$F = \frac{\left(JKS_{ber} - JKS_{gab}\right) / (k-1)}{JKS_{gab} \Big/ \left(\sum_{i=1}^{k} n_i - 2k\right)}$$

yang berdistribusi F dengan db pembilang = $(k - 1)$ dan db penyebut = $\left(\sum_{i=1}^{k} n_i - 2k\right)$

Jika dalam uji ini kita tidak menolak H_0 maka dapat kita hitung lerengan bersama sebagai berikut:

$$\hat{\beta}_{ber} = \frac{S_{xy_1} + \ldots + S_{xyk}}{Sxx_1 + \ldots + S_{xx_k}}$$

Jika dalam uji ini kita menolak H_0 maka kita dapat melakukan uji pembandingan ganda untuk menentukan mana saja β yang sama. Uji pembandingan ganda baru akan kita pelajari dalam modul 6 mendatang.

Membandingkan Lebih dari Dua Elevasi

Jika kita telah menyimpulkan bahwa semua k garis regresi populasinya sejajar maka kita dapat melanjutkan dengan pertanyaan apakah garis-garis itu berimpit? Untuk menjawab pertanyaan ini dapat kita lakukan uji hipotesis, $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k$. Uji hipotesis dapat kita lakukan dengan menghitung $S_{xx,\ldots,S_{yy}}$, dan S_{xy} untuk semua $\left(n_1 + n_2 + \ldots + n_k\right)$ pasang nilai (x, y). Selanjutnya kita hitung *JKS*, yang dinamakan jumlah kuadrat total, ditulis JKS_t, dengan derajat bebas $\left(\sum_{i=1}^{k} n_i - 2\right)$.

Hipotesis itu diuji dengan statistik penguji:

$$F = \frac{\left(JKS_t - JKS_{ber}\right) / (k-1)}{JKS_{ber} \Big/ \left(\sum_{i}^{k} n_i - k - 1\right)}$$

yang berdistribusi F dengan derajat bebas pembilang = $(k - 1)$ dan derajat bebas penyebut = $\left(\sum_{i}^{k} n_i - k - 1 \right)$.

Jika hipotesis nol itu ditolak maka kita dapat melakukan uji pembandingan ganda untuk menentukan mana saja α yang sama.

Contoh 4.5

Dipunyai data tentang panjang sayap (x) dan panjang ekor 3 jenis burung (A, B, dan C) sebagai berikut.

Jenis A:		Jenis B:		Jenis C:	
x_1	y_1	x_2	y_2	x_3	y_3
10,4	7,4	10,7	7,9	11,2	8,1
10,8	7,6	10,5	8,1	11,7	9,2
11,1	7,9	10,9	8,5	10,5	9,3
10,2	7,2	11,7	8,9	10,9	9,8
10,6	8,1	11,2	8,8	10,3	8,9
11,3	8,5	11,6	9,1	11,5	9,5
11,6	8,7	11,9	9,4	11,8	9,7
11,4	8,3	12,1	9,8	11,4	9,3
10,7	7,5	12,4	9,9	12,1	9,9
10,9	7,4	12,5	10,1	12,7	10,3
11,3	7,9	11,4	10,5	12,5	10,2
11,4	8,9			12,3	10,5
11,5	8,8				
11,7	9,7				

Untuk jenis A kita hitung:

$\sum x_1 = 154,9$; $\sum y_1 = 113,9$; $\sum x_1^2 = 1716,71$;

$\sum y_1^2 = 933,37$

$\sum x_1 y_1 = 1263,87$

$S_{xx_1} = 2,8521$; $S_{yy_1} = 6,7121$; $S_{xy_1} = 3,6479$

$JKS(1) = 2,0464$; $db_1 = 12$

Untuk jenis B kita hitung

$$\sum x_2 = 126{,}9 \ ; \qquad \sum y_2 = 101{,}0 \ ; \qquad \sum x_2^2 = 1468{,}43;$$
$$\sum y_2^2 = 934{,}40;$$
$$\sum x_2 y_2 = 1169{,}63$$
$$S_{xx_2} = 4{,}4655 \ ; \quad S_{yy_2} = 7{,}0364 \ ; \quad S_{xy_2} = 4{,}4573$$
$$JKS(2) = 2{,}5873; \ \ db_2 = 9$$

Untuk jenis C kita hitung

$$\sum x_3 = 138{,}9 \ ; \quad \sum y_3 = 114{,}7 \ ; \quad \sum x_3^2 = 1614{,}17;$$
$$\sum y_3^2 = 1101{,}21;$$
$$\sum x_3 y_3 = 1331{,}46$$
$$S_{xx_3} = 6{,}4025 \ ; \quad S_{yy_3} = 4{,}8692 \ ; \quad S_{xy_3} = 3{,}8075$$
$$JKS(3) = 2{,}6049; \ \ \mathrm{db}_3 = 10$$

Selanjutnya kita hitung

$$\begin{aligned} JKS_{gab} &= JKS(1) + JKS(2) + JKS(3) \\ &= 2{,}0464 + 2{,}5873 + 2{,}6049 = 7{,}2386 \\ db_{gab} &= 12 + 9 + 10 = 31 \end{aligned}$$

$$\begin{aligned} JKS_{ber} &= (6{,}7121 + 7{,}0364 + 4{,}8692) - \frac{(3{,}6479 + 4{,}4573 + 3{,}8075)^2}{(2{,}8521 + 4{,}4655 + 6{,}4025)} \\ &= 18{,}6177 - \frac{11{,}9127^2}{13{,}7201} = 8{,}2743 \\ db_{ber} &= 14 + 11 + 12 - 3 - 1 = 37 - 4 = 33 \end{aligned}$$

Untuk menghitung JKS_t, pertama-tama kita hitung

$$\sum x_t = 154{,}9 + 126{,}9 + 138{,}9 = 420{,}7$$
$$\sum x_t^2 = 1716{,}17 + 1468{,}43 + 1614{,}17 = 4798{,}77$$
$$\sum y_t = 113{,}9 + 101{,}0 + 114{,}7 = 329{,}6$$
$$\sum y_t^2 = 933{,}37 + 934{,}40 + 1101{,}21 = 2968{,}98$$
$$\sum (xy)_t = 1263{,}87 + 1169{,}63 + 1331{,}46 = 3764{,}96$$

$$
\begin{aligned}
S_{xx_t} &= 15,2973 \\
S_{yy_t} &= 32,8676 \\
S_{xy_t} &= 17,3189 \\
JKS(t) &= 13,2599 \\
\text{db}_t &= 14+11+12-2=35
\end{aligned}
$$

Untuk menguji $H_0 : \beta_1 = \beta_2 = \beta_3$ versus

H_1 : tidak semua sama

Kita gunakan statistik penguji, kita peroleh

$$F = \frac{(8,2743-7,2386)/2}{7,2386/31} = 2,22$$

Dengan tingkat signifikansi 5%.

H_0 ditolak jika $F \geq 2,93$

H_0 tidak ditolak jika $F < 2,93$

Oleh karena F = 2,22 < 2,93 maka H_0 tidak ditolak. Maka, dapat kita simpulkan bahwa 3 garis regresi populasinya sejajar.

Dapat kita hitung taksiran untuk β bersama:

$$\hat{\beta}_{ber} = \frac{3,6479+4,4573+3,8075}{2,8521+4,4655+6,4025} = 0,87$$

Selanjutnya dapat kita uji apakah garis-garis regresi populasi-populasi itu berimpit, yakni $H_0 : \alpha_1 = \alpha_2$. Untuk ini kita hitung statistik penguji.

$$
\begin{aligned}
F &= \frac{(JKS_t - JKS_{ber})/2}{JKS_{ber}/33} = \frac{(13,2599-9,2743)/2}{9,2743/33} \\
&= 7,09
\end{aligned}
$$

Dengan tingkat signifikansi 5%.

H_0 ditolak jika $F \geq 3,34$

H_0 tidak ditolak jika $F < 3,34$

Karena F = 7,09 > 3,34 maka H_0 ditolak. Maka, kita simpulkan bahwa ketiga garis regresi itu tidak semuanya sejajar.

Uji Sekali Langkah bahwa k garis regresi berimpit

Kita dapat melakukan uji hipotesis nol bahwa k garis regresi populasi berimpit, yakni H_0: semua β sama dan semua α sama. Untuk ini kita gunakan statistik penguji:

$$F = \frac{\left(JKS_t - JKS_{gab}\right)/(2k-2)}{JKS_{gab}/\left(\sum n_i - 2k\right)}$$

yang berdistribusi F dengan derajat bebas pembilang = (2k-2) dan derajat bebas penyebut = $\left(\sum n_i - 2k\right)$.

Beberapa orang senang menggunakan prosedur uji ini daripada uji dua langkah yang kita bicarakan sebelumnya. Jika hipotesis nol ditolak, kita masih perlu menggunakan prosedur di atas jika kita ingin menentukan apakah perbedaan di antara garis regresi itu karena berbeda lerengan atau berbeda elevasinya.

Contoh 4.6

Kembali ke data contoh, sebelumnya (Contoh 4.5), dan kita ingin menguji $H_0: \beta_1 = \beta_2 = \beta_3$ dan $H_0: \alpha_1 = \alpha_2 = \beta_3$ (ketiga garis regresi populasi berimpit).

Versus H_1 : Tidak semua berimpit.

Untuk uji ini kita hitung statistik penguji:

$$F = \frac{(13,2599 - 7,2386)/4}{7,2386/31} = 6,45$$

Dengan tingkat signifikansi 5%.

H_0 ditolak jika $F \geq 2,70$

H_0 tidak ditolak jika $F < 2,70$

Oleh karena F = 6,45 > 2,70 maka H_0 ditolak. Jadi tidak semua ketiganya garis regresi populasinya berimpit. Tentu saja hasil ini konsisten dengan hasil uji sebelumnya.

LATIHAN

Untuk memperdalam pemahaman Anda mengenai materi di atas, kerjakanlah latihan berikut!

Dari suatu studi untuk menentukan hubungan antara x dan y beberapa populasi yang berbeda, namun diperkirakan serupa, telah diambil observasi (x, y) yang tidak disajikan di sini, namun telah dihitung berbagai statistik yang dapat digunakan sebagai landasan analisis selanjutnya.

1) Yakni:

Sampel 1: $n_1 = 18$; $\bar{x}_1 = 14,7$; $\bar{y}_1 = 32,0$;
$S_{xx_1} = 142,35$
$S_{yy_1} = 108,77$; $S_{xy_1} = 69,47$

Sampel 2: $n_2 = 20$; $\bar{x}_2 = 15,8$; $\bar{y}_2 = 27,4$;
$S_{xx_2} = 181,32$
$S_{yy_2} = 153,59$; $S_{xy_2} = 97,40$

(a) Ujilah $H_0 : \beta_1 = \beta_2$ versus $H_1 : \beta_1 \neq \beta_2$

(b) Jika H_0 dalam (a) tidak ditolak, ujilah H_0: Kedua garis regresi populasi itu berimpit, versus H_1: Kedua garis regresi populasi tidak berimpit.

2) Yakni:

Sampel 1: $n_1 = 29$; $S_{xx_1} = 49,76$; $S_{yy_1} = 2304,45$; $S_{xy_1} = 331,40$
Sampel 2: $n_2 = 37$; $S_{xx_2} = 31,34$; $S_{yy_2} = 1524,81$; $S_{xy_2} = 210,71$
Sampel 3: $n_3 = 33$; $S_{xx_3} = 66,38$; $S_{yy_3} = 1875,98$; $S_{xy_3} = 341,27$
Sampel 4: $n_4 = 42$; $S_{xx_4} = 42,12$; $S_{yy_4} = 1130,15$; $S_{xy_4} = 204,18$

Untuk total semua 4 sampel:
$n_t = 141$; $S_{xx_t} = 274,31$; $S_{yy_t} = 9216,55$; $S_{xy_t} = 1493,20$

(a) Ujilah $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$ versus H_1: tidak semua sama
(b) Jika H_0 dalam (a) tidak ditolak, ujilah H_0: Keempat garis regresi populasi berimpit, versus H_1: Keempat garis itu tidak semuanya berimpit.
(c) Lakukan uji satu langkah H_0: keempat garis regresi populasi itu berimpit, versus H_1: Keempat garis itu tidak semuanya berimpit.

RANGKUMAN

Sering kali kita ingin menguji bahwa beberapa garis regresi populasi berimpit. Untuk ini kita pelajari 2 cara sebagai berikut.

1. Pertama-tama diuji bahwa beberapa garis regresi itu sejajar. Jika ini tidak ditolak (diterima), selanjutnya kita uji bahwa garis regresi itu berimpit.
2. Uji dalam satu langkah, yakin kita uji bahwa beberapa garis regresi populasi itu berimpit.

Rumus-rumus:

Untuk menguji $H_0: \beta_1 = \beta_2$ versus $H_1: \beta_1 \neq \beta_2$ kita gunakan statistik penguji:

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{s_{gab}\sqrt{\frac{1}{S_{xx_1}} + \frac{1}{S_{xx_2}}}}$$

yang berdistribusi t dengan derajat bebas $n_1 + n_2 - 4$.

Atau kita dapat menghitung interval kepercayaan $100(1-\alpha)\%$ untuk

$$(\beta_1 - \beta_2): \left(\hat{\beta}_1 - \hat{\beta}_2\right) \pm t_{\frac{\alpha}{2}} s_{gab}\sqrt{\frac{1}{S_{xx_1}} + \frac{1}{S_{xx_2}}}$$

Jika H_0 di atas tidak ditolak, dapat kita hitung taksiran β bersama dengan rumus:

$$\hat{\beta}_{ber} = \frac{S_{xy_1} + S_{xy_2}}{S_{xx_1} + S_{xx_2}}$$

Selanjutnya dapat diuji $H_0 : \alpha_1 = \alpha_2$ dengan statistik penguji

$$t = \frac{\left(\bar{y}_1 - y_2\right) - \hat{\beta}_{ber}\left(\bar{x}_1 - \bar{x}_2\right)}{s_{gab}\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{\left(\bar{x}_1 - \bar{x}_2\right)^2}{S_{xx_1} + S_{xx_2}}}}$$

yang berdistribusi t dengan db = $n_1 + n_2 - 3$

Untuk menguji $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k$ kita gunakan statistik penguji $F = \dfrac{\left(JKS_{ber} - JKS_{gab}\right) / (k-1)}{JKS_{gab} / \left(\sum n_i - 2k\right)}$

yang berdistribusi F dengan db pembilang (k – 1) dan db penyebut $\left(\sum n_i - 2k\right)$.

Nilai taksiran untuk β bersama adalah:

$$\hat{\beta}_{ber} = \frac{S_{xy_1} + S_{xy_2} + \ldots + S_{xy_k}}{S_{xx_1} + S_{xx_2} + \ldots + S_{xx_k}}$$

Jika H_0 tidak ditolak kita uji $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$ dengan statistika penguji:

$$F = \frac{\left(JKS_t - JKS_{ber}\right) / (k-1)}{JKS_{ber} / \left(\sum n_i - k - 1\right)}$$

yang berdistribusi F dengan db pembilang $= (k-1)$ dan db penyebut $= \left(\sum n_i - k - 1\right)$.

Untuk uji sekali langkah bahwa semua garis regresi populasi berimpit kita gunakan statistik penguji:

$$F = \frac{\left(JKS_t - JKS_{gab}\right) / (2k-2)}{JKS_{gab} / \left(\sum n_i - 2k\right)}$$

yang berdistribusi F dengan db pembilang = (2k – 2) dan db penyebut = $\left(\sum n_i - 2k\right)$.

TES FORMATIF 1

Pilihlah satu jawaban yang paling tepat!

1) Dipunyai data sebagai berikut.

Sampel 1: $n_1 = 23$; $S_{xx_1} = 494,47$; $S_{yy_1} = 3229,25$; $S_{xy_1} = 1231,28$;

$\overline{x}_1 = 41,7$; $\overline{y}_1 = 68,6$

Sampel 2: $n_2 = 25$; $S_{xx_1} = 399,21$; $S_{yy_2} = 1368,39$; $S_{xy_2} = 696,71$;

$\overline{x}_2 = 53,2$; $\overline{y}_2 = 64,9$

a. Kita hitung $JKS(1)$ dan derajat bebasnya, db, sama dengan ….
   A. 134,167
   B. 151,635
   C. 163,239
   D. 191,192

b. Kita hitung $JKS(2)$ dan derajat bebasnya, db_2, sama dengan ….
   A. 110,635
   B. 121,334
   C. 141,112
   D. 152,477

c. Untuk menguji $H_0 : \beta_1 = \beta_2$ kita hitung statistik penguji *t*, kita peroleh ….
   A. 3,929
   B. 5,431
   C. 6,937
   D. 7,846

d. Kita hitung $\hat{\beta}_{gab}$ sama dengan ….
   A. 0,11
   B. 0,29
   C. 0,46
   D. 0,58

e. Untuk menguji $H_0 : \alpha_1 = \alpha_2$ jika $\beta_1 = \beta_2$ kita hitung statistik penguji t sama dengan
   A. 98,56
   B. 107,63
   C. 114,39
   D. 124,68

2) Dipunyai data sebagai berikut
   Sampel 1: $n_1 = 23$; $S_{xx_1} = 744,32$; $S_{yy_1} = 7498,91$; $S_{xy_1} = 2341,37$
   Sampel 2: $n_2 = 24$; $S_{xx_2} = 973,14$; $S_{yy_2} = 10366,97$; $S_{xy_2} = 3147,68$
   Sampel 3: $n_3 = 19$; $S_{xx_3} = 664,42$; $S_{yy_3} = 6503,32$; $S_{xy_3} = 2047,73$
   Total semua 3 sampel
   $n = 66$; $S_{xx_t} = 3146,72$; $S_{yy_t} = 20599,33$; $S_{xy_t} = 7938,25$

   a. Kita hitung JKS_{gab}, sama dengan …..
      A. 423,11
      B. 471,75
      C. 511,64
      D. 551,12
   b. Kita hitung JKS_{ber}, sama dengan
      A. 501,14
      B. 521,21
      C. 597,31
      D. 607,77
   c. Kita hitung db_{gab} dan db_{ber}, sama dengan ….
      A. $\text{db}_{gab} = 60$
         $\text{db}_{ber} = 62$
      B. $\text{db}_{gab} = 65$
         $\text{db}_{ber} = 60$
      C. $\text{db}_{gab} = 63$
         $\text{db}_{gab} = 61$
      D. $\text{db}_{gab} = 62$
         $\text{db}_{ber} = 60$

d. Untuk uji $H_0 : \beta_1 = \beta_2 = \beta_3$ kita hitung statistik penguji F, sama dengan ….
   A. 0,56
   B. 0,66
   C. 0,72
   D. 0,78
e. Untuk uji $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$ (jika semua garis regresi sejajar) kita hitung statistik F, sama dengan ….
   A. 1,96
   B. 2,23
   C. 3,62
   D. 4,91
f. Untuk uji satu langkah bahwa ketiga garis regresi itu berimpit kita hitung statistik penguji F, sama dengan ….
   A. 1,55
   B. 1,93
   C. 2,47
   D. 3,01

Cocokkanlah jawaban Anda dengan Kunci Jawaban Tes Formatif 1 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar. Kemudian, gunakan rumus berikut untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 1.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{\text{Jumlah Soal}} \times 100\%$$

Arti tingkat penguasaan: 90 - 100% = baik sekali
80 - 89% = baik
70 - 79% = cukup
< 70% = kurang

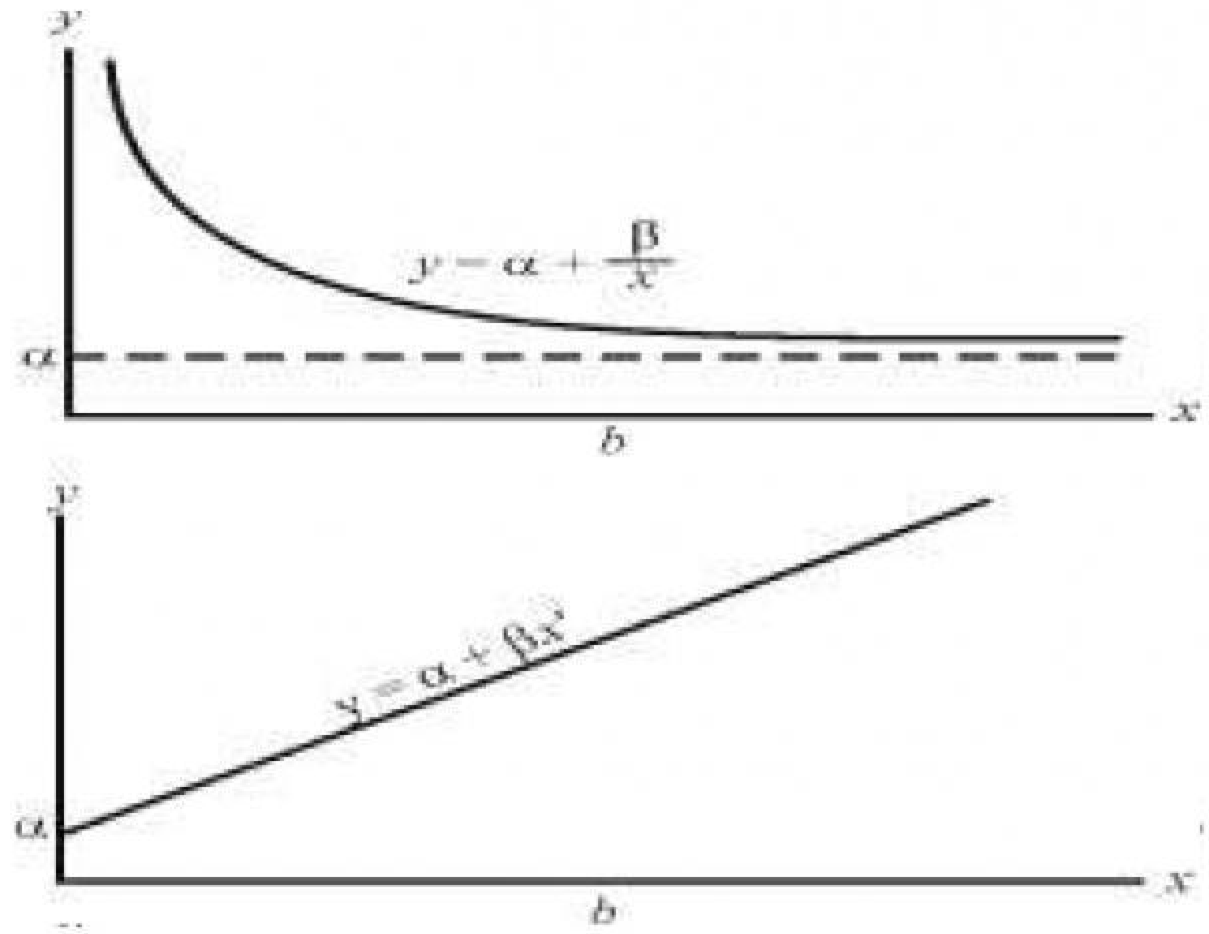
Apabila mencapai tingkat penguasaan 80% atau lebih, Anda dapat meneruskan dengan Kegiatan Belajar 2. **Bagus!** Jika masih di bawah 80%, Anda harus mengulangi materi Kegiatan Belajar 1, terutama bagian yang belum dikuasai.

KEGIATAN BELAJAR 2

Analisis Regresi Berganda

A. HUBUNGAN TAK LINEAR DAN TRANSFORMASI LINEARNYA

Telah kita pelajari analisis hubungan antara variabel dependen y dan variabel independen x yang dapat dirumuskan sebagai model regresi linear. Model garis lurus sangat menarik karena sederhana, dan tersedianya prosedur inferensi statistik yang baik. Tetapi, pembicaraan kita tentang prosedur-prosedur ini harus tidak memberi kesan bahwa model regresi garis lurus dapat digunakan untuk kebanyakan himpunan data yang ada dalam kehidupan ini. Dalam banyak keadaan, gambar data dalam diagram titik menunjukkan adanya hubungan yang tidak linear. Ini dapat dikuatkan dengan analisis statistik yang ditunjukkan oleh kecilnya nilai r^2 jika ditaksir dengan garis lurus atau dengan melakukan uji ketidaksesuaian, seperti yang pernah kita pelajari. Prosedur statistik untuk menangani hubungan tidak linear lebih rumit daripada prosedur yang untuk menangani hubungan linear, dengan perkecualian model jenis tertentu yang dinamakan *model regresi polinomial,* yang akan kita bicarakan nanti. Tetapi, dalam banyak hal mungkin dapat kita lakukan untuk mentransformasi variabel x dan/atau y sedemikian hingga hubungan baru sangat dekat dengan linear. Maka, model regresi linear dapat dirumuskan untuk variabel-variabel yang ditransformasi, dan analisis yang sesuai dapat dilakukan untuk data ini. Analisis ini harus meliputi pemeriksaan residu bagi model ditransformasi karena anggapan sesatan berdistribusi normal dan independen serta variansi konstan sekarang berlaku untuk model transformasi ini.



Gambar 4.1
Transformasi $x' = 1/x$ melinearkan $y = \alpha + \beta/x$

Untuk melukiskan garis penalaran ini, kita pandang masalah menentukan hubungan antara pertumbuhan dalam keterampilan y dan lamanya pelatihan *x* yang pernah kita sebutkan dalam modul terdahulu. Khususnya, misalkan *x* banyak jam pelatihan yang diterima oleh seorang mekanik dan *y* adalah waktu yang diperlukan oleh mekanik itu untuk merakit satu bagian mesin yang rumit. Waktu untuk menyelesaikan pekerjaan itu *y* diharapkan menurun dengan bertambahnya waktu pelatihan *x*, tetapi pada suatu titik diharapkan telah dicapai suatu tingkat yang nilai *y* tidak lagi dapat diturunkan dengan menambah waktu pelatihan. Mungkin kita dapat memandang hubungan dalam bentuk:

$$y = \alpha + \frac{\beta}{x}$$

yang grafiknya ditunjukkan dalam Gambar 4.1 (a). Hubungan ini sangat tidak linear.

Sekarang kita pandang variabel transformasi $x^1 = 1/x$ sehingga model kita menjadi $y = \alpha + \beta x^1$ yang merupakan garis lurus seperti ditunjukkan dalam Gambar 4.1 (b). Maka, kita dapat mulai melakukan analisis data pada pertumbuhan keterampilan dengan melakukan analisis regresi linear sederhana variabel y dan variabel x yang baru, yang merupakan kebalikan dari banyak jam pelatihan.

Beberapa model tak linear yang sering kita jumpai dan transformasi linearnya yang sesuai kita sajikan dalam Tabel 4.4.

Dalam banyak hal hubungan linear tertentu dengan jelas ditunjukkan oleh datanya atau oleh pandangan teoretis. Meskipun informasi awal tentang bentuk hubungan sangat kurang, studi tentang grafik titik sering menunjukkan transformasi linearnya yang sesuai. Akan sangat membantu untuk menggambarkan titik-titik $(x_i\ ,\ y_i)$ pada berbagai jenis kertas grafik, seperti kertas semilog atau double log, untuk melihat apakah hubungan transformasinya mendekati linear. Misalnya, hubungan (a) dalam Tabel 4.4 grafiknya dalam kertas grafik semilog akan berbentuk garis lurus. Kadang-kadang jika diagram titik memperlihatkan hubungan pada suatu kurva yang menunjukkan nilai y bertambah dengan cepat dibandingkan dengan nilai x maka gambar $\sqrt{y}$ atau y berpangkat pecahan yang lain dapat membantu melinearkan hubungannya. Keadaan ini dilukiskan dalam Contoh 4.7. Beberapa petunjuk analitis guna menentukan pangkat yang sesuai dalam transformasi jenis ini telah tersedia, namun di luar jangkauan pembicaraan kita di sini.

Tabel 4.4
Beberapa Model Tak Linear dan Transformasi Linearnya

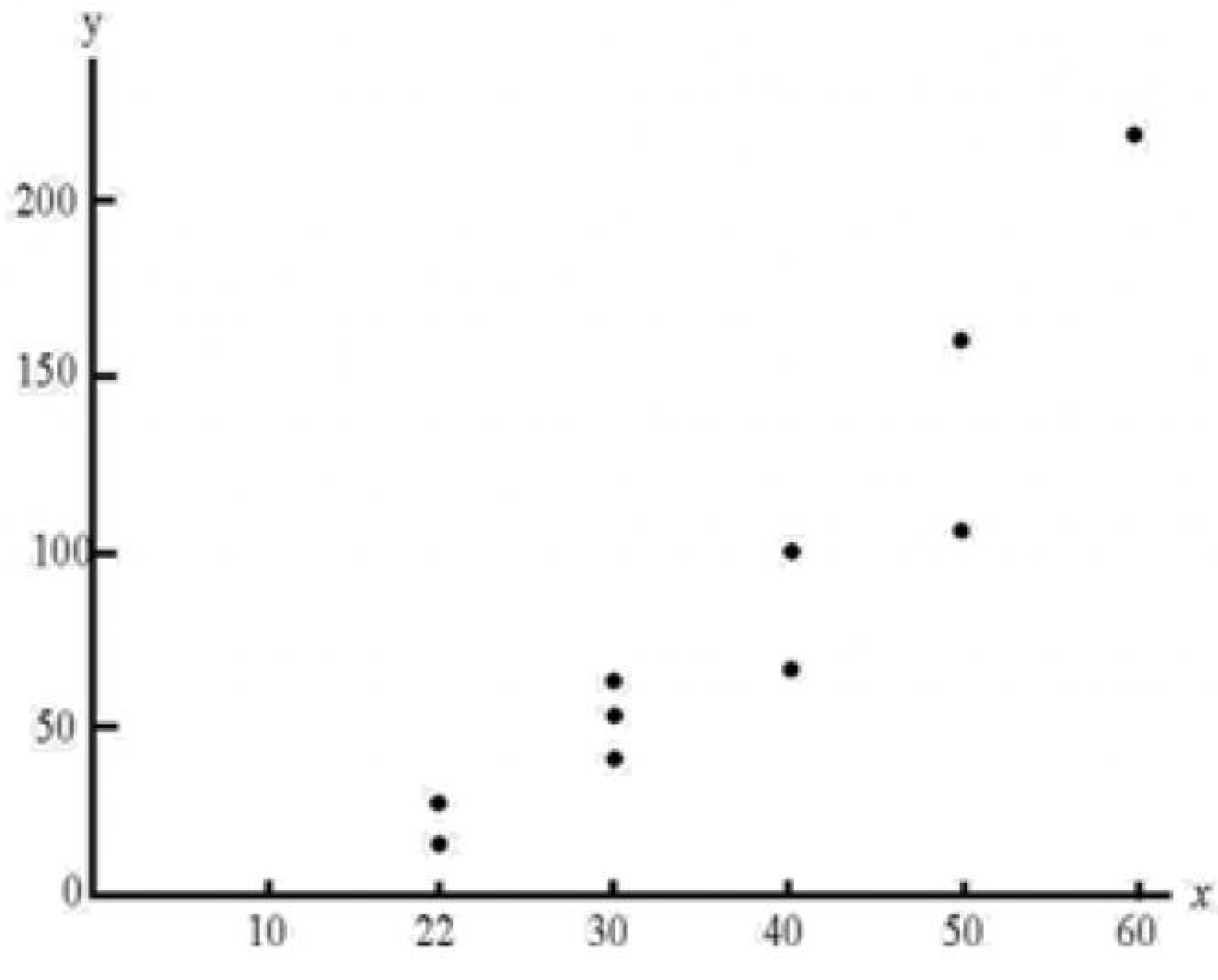
Model Tak Linear	Transformasi	Model Transformasi
(a) $y = ae^{bx}$	$y' = {}^{e}\log y$; $x' = x$	$y' = \alpha + \beta x'$; $\alpha = {}^{e}\log a$; $\beta = b$
(b) $y = ax^{b}$	$y' = \log y$; $x' \log x$	$y' = \alpha + \beta x'$; $\alpha = \log a$; $\beta = b$
(c) $y = \dfrac{1}{a+bx}$	$y' = \dfrac{1}{y}$; $x' = x$	$y' = \alpha + \beta x'$; $\alpha = a$; $\beta = b$
(d) $y = \dfrac{1}{(a+bx)^2}$	$y' = \dfrac{1}{\sqrt{y}}$; $x' = x$	$y' = \alpha + \beta x'$; $\alpha = a$; $\beta = b$
(e) $\dfrac{1}{y} = a + \dfrac{b}{1+x}$	$y' = \dfrac{1}{y}$; $x' = \dfrac{1}{1+x}$	$y' = \alpha + \beta x'$; $\alpha = a$; $\beta = b$
(f) $y = a + b\sqrt{x}$	$x' = y$; $x' = \sqrt{x}$	$y' = \alpha + \beta x'$; $\alpha = a$; $\beta = b$

Contoh 4.7

Untuk menentukan kemampuan berhenti yang maksimum mobil merek Gagah jika direm secara penuh, 10 mobil merek itu dikendarai masing-masing dengan kecepatan tertentu, dan jarak yang diperlukan sampai berhenti penuh diukur. Berbagai kecepatan dipilih untuk masing-masing 10 mobil itu dan jarak sampai berhenti dicatat dan dituangkan dalam Tabel 4.5. Diagram titik data itu tampak dalam Gambar 4.2.

Tabel 4.5
Data kecepatan dan jarak berhenti

Kecepatan (x)	20	20	30	30	30	40	40	50	50	60
Jarak berhenti (y)	16,3	26,7	39,2	63,5	51,3	98,4	65,7	104,1	155,6	217,2



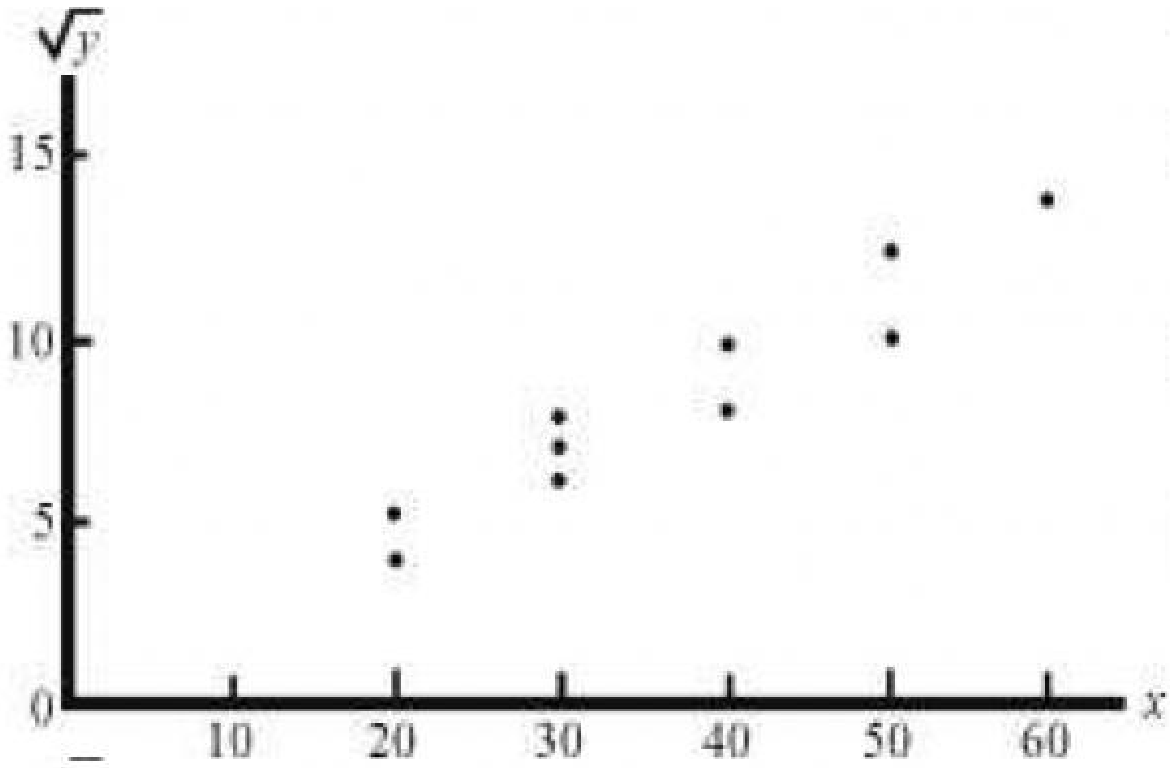
Gambar 4.2
Diagram titik untuk data dalam Tabel 4.5

Terlihat hubungan yang tidak linear, khususnya nilai y naik jauh lebih cepat pada nilai-nilai x yang besar daripada nilai-nilai x yang kecil. Ini menunjukkan kepada kita untuk mencoba melinearkan hubungan dengan menggambarkan √y atau y berpangkat pecahan yang lain dengan x. Gambar

$\sqrt{y}$ menghasilkan data transformasi, seperti dalam Tabel 4.6 dan diagram titik data ini, yang mengesankan hubungan yang kira-kira linear, tampak dalam Gambar 4.3.

Tabel 4.6
Data Kecepatan dan Akar Jarak Berhenti

x	20	20	30	30	30	40	40	50	50	60
$y' = \sqrt{y}$	4,037	5,167	6,261	7,969	7,162	9,920	8,106	10,203	12,474	14,738



Gambar 4.3
Diagram titik data dalam Tabel 4.6

Dengan program komputer untuk analisis regresi dapat kita peroleh hasil hitungan berikut.

$\overline{x} = 37$ $\overline{y} = 8,604$

$S_{xx} = 1610$ $S_{yy} = 97,773$ $S_{xy} = 381,621$

$\hat{\alpha} = -0,166$ $\hat{\beta} = 0,237$

Jadi, persamaan regresi taksiran:

$$\hat{y}' = -0,166 + 0,237x$$

Bagian variasi y' yang dijelaskan oleh model garis lurus adalah:

$$r^2 = \frac{(381,621)^2}{(1610)(97,773)} = 0,92$$

Kita ingatkan kembali bahwa semua inferensi tentang model transformasi didasarkan pada anggapan hubungan linear dan sesatan berdistribusi normal independen serta variansi konstan. Sebelum kita mempercayai inferensi ini, model transformasi ini harus diperiksa untuk menentukan apakah ada pelanggaran yang serius terhadap anggapan-anggapan ini.

> Pola yang diungkapkan oleh diagram titik sering menunjukkan adanya hubungan yang tidak linear. Kadang-kadang dapat diperoleh transformasi data aslinya sehingga hubungan variabel-variabel baru itu kira-kira linear. Jika ini mungkin maka kita dapat melakukan analisis regresi garis lurus seperti biasa dan menarik inferensi dari model transformasi. Anggapan mengenai struktur sesatan variabel transformasi harus diperiksa dengan cara yang biasa.

B. REGRESI LINEAR BERGANDA

Setelah melakukan analisis regresi linear biasa antara y dan x, mungkin kita mendapatkan nilai r^2 yang kecil, tetapi pemeriksaan diagram titik atau uji kekurangcocokkan tidak dapat mendiskreditkan hubungan linear karena variansi sesatan yang besar. Lagi pula, pemeriksaan lebih jauh tentang metode percobaan dan proses pengumpulan data dapat mengungkap adanya variabel penyebab selain x yang mempengaruhi variabel respons y, tetapi kita lupakan dalam analisis regresi linear sederhana itu. Jika variasi dalam variabel-variabel berpengaruh ini tidak terkendali selama percobaan, hal ini dapat mengaburkan hubungan yang sebenarnya antara y dan x dengan memperbesar variansi sesatan σ^2. Dalam contoh pengurangan nitrogen oksida yang telah kita bicarakan dalam modul yang lalu jika faktor yang mempengaruhi, seperti kecepatan rata-rata, lama waktu berhenti, dan temperatur sekeliling diperbolehkan berubah-ubah selama percobaan, diagram titik y terhadap x mungkin menunjukkan fluktuasi yang lebar dan JKS-nya dapat sangat besar. Setiap informasi yang ada tentang faktor-faktor tambahan ini dapat digunakan untuk meningkatkan perkiraan.

Dengan demikian, guna memperoleh model perkiraan yang bermanfaat dan juga penaksiran parameter yang efisien dan tak bias, kita harus mencatat data yang relevan dengan semua variabel yang diketahui berpengaruh terhadap variabel respons *y* dan menggabungkannya secara eksplisit dalam analisis regresi. Untuk lebih spesifik, misalkan variabel respons *y* dalam suatu percobaan diharapkan akan dipengaruhi oleh tiga variabel penyebab x_1, x_2, dan x_3, dan data yang relevan dengan ketiga variabel penyebab ini dicatat bersama dengan pengukuran *y*. Dengan analogi model regresi linear sederhana, kita sementara dapat merumuskan model hubungan yang sederhana antara *y* dengan x_1, x_2, dan x_3.

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i, \quad i = 1, 2, \ldots, n$$

dengan x_{i1}, x_{i2}, dan x_{i3} adalah nilai-nilai tertentu tiga variabel independen dalam percobaan ke-i dan y_i adalah respons yang berkaitan. Suku sesatan e_i dianggap merupakan variabel normal independen dengan mean = 0 dan variansi = σ^2. Parameter α, β_1, β_2, dan β_3 kuantitas tertentu yang tidak diketahui.

Model ini menegaskan bahwa di samping suku sesatan, respons berubah-ubah secara linear dengan tiap variabel independen jika dua yang lain tetap. Karena itu fungsi respons ini suatu *bidang datar* dengan dua prediktor dan *bidang datar dimensi tinggi (hyperplane)* dengan lebih dari dua prediktor. Karena adanya lebih dari satu variabel prediktor, model ini dinamakan *model regresi berganda.*

Meskipun diagram titik tidak dapat digambarkan, azas kuadrat terkecil berguna dalam menaksir parameter regresi. Untuk model ini, kita perlu memandang α, β_1, β_2, dan β_3 secara bersama berubah-ubah untuk meminumkan jumlah kuadrat penyimpangan:

$$\sum_{i=1}^{n} (y_i - \alpha - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}$$

Taksiran kuadrat terkecil $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \text{dan } \hat{\beta}_3$ dapat diperiksa sebagai penyelesaian *persamaan normal* berikut yang merupakan perluasan dari persamaan-persamaan untuk penyelesaian kuadrat terkecil bagi model garis lurus yang pernah kita pelajari.

$$\hat{\beta}_1 Sx_1x_1 + \hat{\beta}_2 Sx_1x_2 + \hat{\beta}_3 Sx_1x_3 = Sx_i y$$
$$\hat{\beta}_1 Sx_1x_2 + \hat{\beta}_2 Sx_2x_2 + \hat{\beta}_3 Sx_2x_3 = Sx_2 y$$
$$\hat{\beta}_1 Sx_1x_3 + \hat{\beta}_2 Sx_2x_3 + \hat{\beta}_3 Sx_3x_3 = Sx_3 y$$
$$\hat{\alpha} = \overline{y} - \hat{\beta}_1\overline{x}_1 - \hat{\beta}_2\overline{x}_2 - \hat{\beta}_3\overline{x}_3$$

dengan $S_{x_1x_1}$, dan $S_{x_1x_1}$, dan seterusnya adalah jumlah kuadrat dan hasil kali silang variabel-variabel dalam indeks itu dan dapat dihitung, seperti dalam model regresi garis lurus, paling banyak hanya dua variabel yang dapat masuk dalam setiap S. Tersedia metode untuk penaksiran interval, uji hipotesis, uji kekurangcocokkan model dan untuk menentukan variabel x mana yang cukup penting untuk masuk dalam model dan mana yang dapat ditinggalkan. Dalam prinsipnya, metode-metode ini serupa dengan yang digunakan dalam model regresi sederhana, tetapi rumus-rumus aljabar yang diperlukan menjadi bertambah ruwet jika banyak variabel x yang terlibat bertambah. Pembicaraan lengkap tentang analisis statistik model ini di luar jangkauan pembicaraan kita dalam kuliah ini. Tetapi, pemeriksaan residu untuk model regresi berganda tepat sama, seperti untuk model regresi garis lurus, dan merupakan cara yang penting untuk menilai model yang ditaksir dengan metode kuadrat terkecil.

Selanjutnya, kita bicarakan satu contoh yang melukiskan metode hitungan guna menaksir *model linear* yang mempunyai dua variabel prediktor.

Contoh 4.8

Kita ingin mempelajari tekanan darah sistolik *y* dalam hubungannya dengan berat x_1, dan umur x_2 kelompok orang laki-laki yang kira-kira tingginya sama. Dari 13 subjek yang dipilih menurut berat dan umurnya, kita peroleh data yang tertuang dalam 3 kolom pertama Tabel 4.7. Taksirlah model regresi berganda $y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$ untuk data itu.

Tabel 4.7
Data dan hitungan-hitungan untuk Analisis regresi berganda

x_1	x_2	y	x_1^2	x_2^2	x_1x_2	x_1y	x_2y	$\hat{y}$
152	50	120	23104	2500	7600	18240	6000	11489
183	20	141	33489	400	3660	25803	2820	14052
171	20	124	29241	400	3420	21204	2480	12759
165	30	126	27225	900	4950	20790	3780	12538
158	30	117	24964	900	4740	18486	3510	11784
161	50	129	25921	2500	8050	20769	6450	12958
149	60	123	22201	3600	8940	18327	7380	12091
158	50	125	24964	2500	7900	19750	6250	12635
170	40	132	28900	1600	6800	22440	5280	13502
153	55	123	23409	3025	8415	18819	6765	12309
164	40	132	26896	1600	6560	21648	5280	12856
190	40	155	36100	1600	7600	29450	6200	15657
185	20	147	34225	400	3700	27195	2940	14267

$\sum x_1 = 2159$; $\sum x_1^2 = 360639$; $\sum x_1x_2 = 82335$;
$\sum x_2y = 65135$; $\sum x_2 = 505$; $\sum x_2^2 = 21925$;
$\sum x_1y = 282921$; $\sum y = 1694$

Analisis regresi berganda biasanya dilakukan dengan bantuan program komputer yang memberikan taksiran kuadrat terkecil dan residu. Tetapi, untuk melukiskan prosedur ini, dan jenis-jenis penghitungan yang dilakukan disajikan dalam Tabel 4.7. Semua hasil-hasil penghitungan sesudah kolom ketiga dilakukan dengan komputer.

Dari jumlah-jumlah kolom dalam Tabel 4.7 dapat kita hitung $S_{x_1x_2} = 82355 - (2159)(505)/13 = -1533{,}8$ dan seterusnya maka kita peroleh persamaan normal sebagai berikut:

$$2078{,}9\hat{\beta}_1 - 1533{,}8\hat{\beta}_2 = 1586{,}7$$

$$-1533{,}8\hat{\beta}_1 + 2307{,}7\hat{\beta}_2 = -670{,}38$$

$$\hat{\alpha} = 130{,}308 - 166{,}077\hat{\beta}_1 - 38{,}846\hat{\beta}_2$$

Mengalikan persamaan pertama dengan 1533,8 dan persamaan kedua dengan 2078,9 kita hilangkan $\hat{\beta}_1$ dengan menjumlah kedua persamaan yang telah dikalikan itu. Selanjutnya dapat kita hitung nilai $\hat{\beta}_2$, kita peroleh $\hat{\beta}_2 = 0,425$; selanjutnya dari persamaan pertama atau kedua dapat kita hitung $\hat{\beta}_1 = 1,077$. Akhirnya, persamaan ketiga menghasilkan $\hat{\alpha}_1 = -65,10$. Maka, kita peroleh persamaan regresi taksiran

$$\hat{y} = -65,10 + 1,077\, x_1 + 0,425\, x_2$$

Kolom terakhir Tabel 4.7 memberikan nilai perkiraan $\hat{y}$ yang kita hitung dari persamaan regresi taksiran itu.

Ini berarti mean tekanan darah akan naik dengan 1,077 jika berat x_1 satu unit dan umur x_2 tetap konstan. Demikian juga, kenaikan satu tahun dalam umur dengan berat dibuat tetap hanya akan menaikkan mean tekanan darah 0,425. Tentu akan sangat baik bagi kita untuk menghitung residu guna melihat apakah dapat ditentukan adanya pola sistematik.

Dalam Contoh 4.8 melukiskan analisis model regresi berganda dengan 2 variabel prediktor x_1 dan x_2. Analisis model regresi garis lurus dengan satu variabel prediktor x telah dibicarakan secara terperinci dalam modul yang lalu, dan beberapa transformasi yang melinearkan model telah kita bicarakan juga di muka. Meskipun untuk kasus 1 prediktor, diagram titiknya mungkin memperlihatkan hubungan kurva yang transformasi, yang melinearkannya tidak dapat dibentuk. Satu metode alternatif untuk menangani hubungan linear semacam itu adalah memasukkan suku dengan x berpangkat lebih tinggi dalam model $y = \alpha + \beta x + e$. Misalnya, dengan memasukkan x pangkat dua kita peroleh model:

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + e_i \ ; \ i = 1,2,\ldots, n$$

yang menyatakan bahwa di samping suku sesatan e_1, respons y adalah *fungsi kuadrat* (atau *polinomial derajat dua)* variabel independen x. Model seperti itu dinamakan *model regresi polinomial* y dengan x, dan pangkat tertinggi x yang terjadi dalam model itu dinamakan *derajat* atau *orde* regresi polinomial itu. Menarik untuk dicatat bahwa analisis model regresi polinomial tidak memerlukan teknik khusus selain yang digunakan dalam analisis regresi berganda. Dengan identifikasi x dan x^2 sebagai dua variabel independen x_1 dan x_2, model regresi polinomial derajat dua ini berubah menjadi bentuk model regresi berganda

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i \ ; \ i = 1,2,\ldots, n$$

dengan $x_{i1} = x_i$ dan $x_{i2} = x_i^2$. Kedua jenis model ini dan banyak lagi jenis yang lain adalah kasus khusus dari kelas umum yang dinamakan *model linear*, yang respons y dinyatakan sebagai

$$y_i = \ell_{i0}\alpha + \ell_{i1}\beta_1 + ... + \ell_{ik}\beta_k + e_i$$

Hubungan ini adalah linear dalam parameter (α, β_1, …, β_k), dan koefisien ℓ parameter-parameter ini diketahui dari nilai-nilai variabel prediktor yang dipilih untuk percobaan itu. Koefisien-koefisien ini dapat merupakan nilai variabel-variabel penyebab yang berbeda, seperti halnya dalam regresi berganda atau nilai-nilai dari pangkat-pangkat yang berbeda variabel penyebab yang sama, seperti halnya dalam regresi polinomial. Bahkan mungkin kedua macam variabel itu dalam satu model. Oleh karena semua model ini mempunyai sifat penting sebagai linear di dalam parameternya, bentuk analisisnya tetap sama seperti yang telah kita bicarakan di muka.

Aspek lain dari regresi berganda

Oleh karena luas penggunaannya, model regresi linear berganda memegang peranan penting dalam pekerjaan seorang statistisi. Meskipun analisis yang lengkap tidak dapat diberikan di sini, beberapa aspek tertentu regresi berganda pantas mendapat perhatian lebih lanjut. Contoh-contoh yang kita bicarakan di sini menjelaskan pokok-pokok berikut.

a. Jika suatu variabel penting tidak dimasukkan dalam model, penaksir tidak akan tak bias.
b. Koefisien regresi individual βi harus diinterpretasikan dengan hati-hati.
c. Jika rancangan percobaan begitu jelek sehingga satu variabel input dapat diperoleh dari menambah kelipatan konstan variabel-variabel yang lain maka timbul ketidaktentuan.

Contoh 4.9

Penaksir itu bias jika garis lurus ditaksirkan untuk respons yang sebenarnya juga bergantung pada suatu variabel lain.

Misalkan, mean sebenarnya adalah $4 + 3x_1 + 2x_2$, dengan

x_1	1	-3	2
x_2	2	3	4

Oleh karena sesuatu alasan peneliti hanya mencatat x_1 dan menaksir model $y = \alpha + \beta x_1 + e$. Maka penaksir kuadrat terkecil adalah

$$\hat{\alpha} = \hat{y} - \hat{\beta}\overline{x} = \overline{y}, \quad \text{karena } \overline{x} = 0$$

$$\hat{\beta} = \frac{s_{x_1 y}}{s_{x_1 x_1}} = \frac{1}{14}(y_1 - 3y_2 + 2y_3)$$

dengan $E(y_1) = 4 + 3(1) + 2(2) = 11$; $E(y_2) = 4 + 3(-3) + 2(3) = 1$, dan $E(y_3) = 18$. Dengan demikian:

$$E(\hat{\alpha}) = \frac{1}{3}\left[E(y_1) + E(y_2) + E(y_3)\right] = \frac{11 + 1 + 18}{3} = 10$$

$$E(\hat{\beta}) = \frac{1}{14}\left[E(y_1) - 3E(y_2) + 2E(y_3)\right] = \frac{1}{14}(11 - 3 + 36) = \frac{22}{7}$$

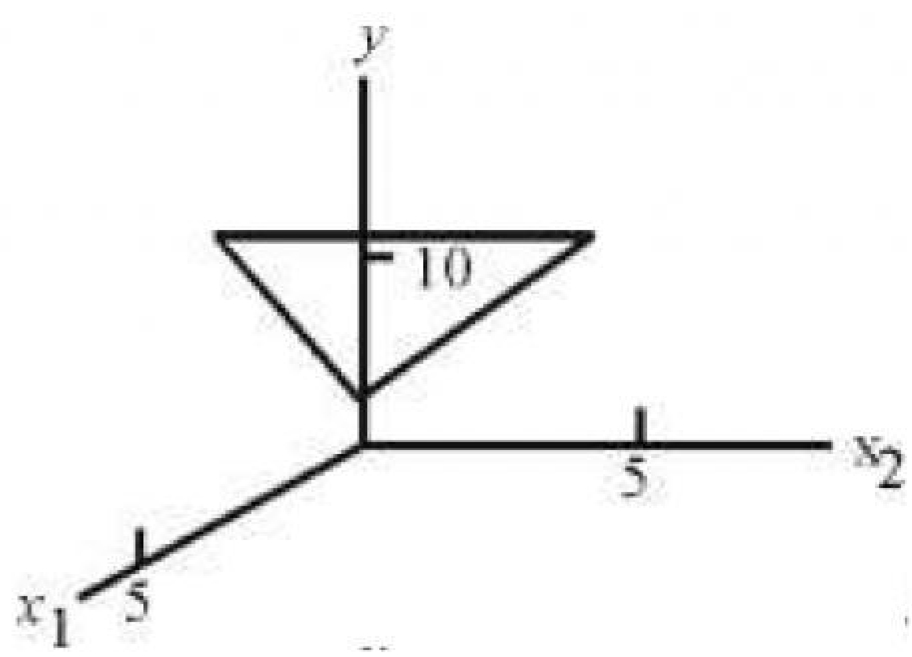
Respons harapan pada x_1^*, sebenarnya pada $\left(x_1^*; x_2^*\right)$, adalah:

$$E\left(\hat{\alpha} + \hat{\beta}x_1^*\right) = 10 + \frac{22}{7}x_1^*$$

bukannya $4 + 3x_1^* + 2x_2^*$ yang merupakan nilai yang sebenarnya. Model garis lurus sederhana ini tidak cukup sehingga diperlukan model yang lebih rumit.

Contoh 4.10

Interpretasi koefisien regresi.



Gambar 4.4
Bidang respons harapan $4 + 3x_1 + 2x_2$

Dalam model regresi berganda $\alpha + \beta_1 x_1 + \beta_2 x_2 = 4 + 3x_1 + 2x_2$, nilai $\beta_1 = 3$ menggunakan pengaruh x_1, jika x_2 *tetap*. Respons ini dilukiskan dalam Gambar 4.4. Jika x_2 tetap maka hubungan garis lurus ada antara E(y) dan x_1.

Untuk memandang regresi taksiran, misalkan kita peroleh $3,5+3,6x_1+2,8x_2$ berdasarkan 20 observasi. Meskipun 3,6 menggambarkan pengaruh taksiran x_1 dengan x_2 dibuat tetap, kita tidak dapat menyimpulkan bahwa x_1 lebih penting daripada x_2. Sebenarnya (x_1, x_2) biasanya terjadi dengan komponen kedua lebih dari dua kali yang pertama, kebalikannya mungkin yang benar. Kita dapat juga menggunakan R^2, yang akan kita pelajari dalam soal latihan selanjutnya, untuk menentukan relatif pentingnya x_1 dan x_2.

Contoh 4.11

Kasus rancangan yang jelek dengan koefisien tidak tunggal.

Misalkan, respons harapan $E(y) = 4 + 3x_1 + 2x_2$ dan bahwa $x_2 = 2x_1$ untuk semua percobaan. Nilai-nilai x_1, x_2 adalah:

x_1	1	3	-2	2
x_2	2	6	-4	4
$4 + 3x_1 + 2x_2$	11	25	-10	18
$4 + 5x_1 + x_2$	11	25	-10	18

Di sini $4 + 3x_1 + 2x_2 = 4 + 3x_1 + 2x_1 + x_2 = 4 + 5x_1 + x_2$ sehingga kita tidak dapat membedakan $\beta_1 = 3$, $\beta_2 = 2$, dan $\beta_1 = 5$, $\beta_2 = 1$ atau dengan sebarang bilangan lain yang dipilih. Tidak mengherankan, ketidaktentuan ini juga berlaku untuk nilai-nilai taksiran, dan kedua suku itu tidak boleh ditahan di dalam model itu.

Telah kita sebutkan bahwa kebanyakan analisis kuadrat terkecil model regresi linear berganda dilakukan dengan bantuan komputer. Program untuk melakukan analisis itu menuntut peneliti untuk menyediakan nilai-nilai respons y_i, dan ρ variabel input $x_{i1}, \ldots, x_{ip}$ untuk setiap i = 1, 2, …, n. Model itu adalah:

$$\underbrace{y_i}_{\text{Observasi}} = 1.\alpha + \underbrace{x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ip}\beta_p}_{\text{Variabel input}} + \underbrace{e_i}_{\text{Sesatan}}$$

Kuantitas dasar itu dapat disusun dalam bentuk susunan sebagai berikut.

$$\begin{array}{cccc} & \text{Observasi} & & \text{Variabel input:} \\ \underset{\sim}{y} = & \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_1 \\ \vdots \\ y_n \end{pmatrix} & ; \quad \underset{\sim}{x} = & \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{i1} & \cdots & x_{ip} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & & x_{np} \end{pmatrix} \text{ nilai percobaan ke-i} \end{array}$$

Hanya susunan $\underset{\sim}{y}$ dan $\underset{\sim}{x}$ yang diperlukan untuk menghitung taksiran kuadrat terkecil bagi α, β_1, β_2, …, β_p yang meminimumkan:

$$\sum_{i=1}^{n}(y_1 - \alpha - x_{i1}\beta_1 - ... - x_{i\rho}\beta_\rho)^2$$

Susunan input $\underset{\sim}{x}$ dinamakan *matriks rancangan.* Dengan cara serupa, kita tulis:

$$\underset{\sim}{e} = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix} \qquad \text{dan} \quad \underset{\sim}{\beta} = \begin{bmatrix} \alpha \\ \beta_1 \\ . \\ . \\ . \\ \beta_p \end{bmatrix}$$

sehingga model itu dapat ditulis dalam bentuk:

$$\begin{array}{ccccccc} \text{Observasi} & & \text{Matriks rancangan} & & \text{Parameter} & & \text{Sesatan} \\ \underset{\sim}{y} & = & \underset{\sim}{x} & & \underset{\sim}{\beta} & + & \underset{\sim}{e} \end{array}$$

yang merupakan representasi matriks model linear.

LATIHAN

Untuk memperdalam pemahaman Anda mengenai materi di atas, kerjakanlah latihan berikut!

1) Dipunyai beberapa pasang nilai (x, y) sebagai berikut:

x	0,5	1	2	4	5	6	7
y	4,6	3,8	1,8	1,3	0,9	0,7	0,8

   a. Gambarkan diagram titik data di atas.
   b. Carilah taksiran garis lurus yang paling cocok untuk data itu dan gambarkan garis itu melalui diagram titik itu.
   c. Berapa persen variabilitas y dijelaskan oleh garis taksiran itu.
   d. Pandang transformasi $y' = 1/y$, dan gambarkan diagram titik y' versus x.
   e. Taksirlah regresi garis lurus pada data transformasi ini.
   f. Hitunglah r^2 dan berikan komentar pada kecukupan kesesuaian taksiran ini.

2) Seorang peneliti mencari informasi tentang dimensi dasar pohon di hutan, memperoleh pengukuran diameter pohon satu meter di atas tanah dan tinggi pohon untuk 12 pohon tertentu. Diperoleh data sebagai berikut

Diameter x (dalam meter)	0,9	1,2	2,9	3,1	3,3	3,9	4,3	6,2	9,6	12,6	16,1	25,8
Tinggi y (dalam meter)	18	26	32	36	44,5	35,6	40,5	57,5	67,3	84	67	87,5

   a. Gambarkan diagram titik dan tentukan apakah hubungan garis lurus memadai!
   b. Tentukan transformasi yang membuat linear. Khususnya coba $x' = \log x$; $y' = \log y$!
   c. Taksirlah regresi garis lurus untuk data transformasi itu!
   d. Berapa proporsi variabilitas dijelaskan oleh model taksiran itu!

3) Model regresi $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + e$ ditaksir untuk himpunan data yang diperoleh dari 20 kali percobaan dengan dua prediktor x_1 dan x_2 diamati bersama dengan respons y. Taksiran kuadrat terkecil adalah:

   $\hat{\alpha} = 4,21$; $\hat{\beta}_1 = 11,37$; $\hat{\beta}_2 = -0,513$

   Perkirakan respons untuk

   a. $x_1 = 8$; $x_2 = 30$
   b. $x_1 = 8$; $x_2 = 50$
   c. $x_1 = 3$; $x_2 = 50$

4) Pandang kembali soal nomor 3 di atas.

   Misalkan, jumlah kuadrat sesatan = 46,25 dan jumlah kuadrat karena regresi = 236,70.

   a) Taksirlah deviasi standar sesatan σ. Sebutkan derajat bebasnya!
   b) Hitunglah R^2 dan interpretasikan hasil itu!

5) Pandang kembali soal nomor 3 di atas.

   Misalkan, sesatan standar taksiran $\hat{\alpha}$, $\hat{\beta}_1$, dan $\hat{\beta}_2$ adalah masing-masing 2,26 ; 1,08 ; dan 0,098.

   a) Hitunglah interval kepercayaan 95% untuk α dan β_2!
   b) Ujilah $H_0: \beta_1 = 10$ versus $H_1: \beta_1 > 10$ dengan tingkat signifikansi 5%.

RANGKUMAN

Jika diagram titik menunjukkan hubungan garis lengkung (kurva) maka mungkin bagi kita memilih transformasi untuk satu atau kedua variabelnya sehingga data transformasi menunjukkan hubungan linear (garis lurus). Maka, analisis regresi linear sederhana dapat digunakan untuk data transformasi itu.

Analisis regresi berganda adalah teknik aneka guna untuk membina model perkiraan dengan beberapa variabel *input* (masukan). Sebagai tambahan apa yang kita peroleh dari taksiran kuadrat terkecil, kita dapat menghitung interval kepercayaan dan uji hipotesis tentang pengaruh masing-masing variabel masukan.

Model regresi polinomial adalah kasus khusus regresi berganda dengan pangkat satu prediktor x, yakni x, x^2, x^3, dan seterusnya, memainkan peranan individu prediktor.

Ukuran R^2, dinamakan kuadrat koefisien korelasi berganda, menunjukkan proporsi variabilitas y yang dijelaskan oleh model regresi berganda taksiran.

TES FORMATIF 2

Pilihlah satu jawaban yang paling tepat!

1) Dipunyai nilai-nilai berpasangan:

x	100	150	250	250	400	650	1000	1600
y	22	21	26	17	18	10	12	8

Transformasikan nilai x menjadi $x' = {}^{10}\log x$

a. Kita taksir persamaan regresi garis lurus untuk data transformasi, kita peroleh ….
   A. $y = 41,22 + 11,71 \log x$
   B. $y = 45,32 - 15,11 \log x$
   C. $y = 42,76 + 18,41 \log x$
   D. $y = 50,37 - 13,06 \log x$

b. Kita hitung interval kepercayaan 90% untuk lerengan garis regresinya, kita peroleh
   A. (−19,37 ; −6,75)
   B. (−11,21 ; 4,23)
   C. (−6,12 ; 11,93)
   D. (2,17 ; 12,11)

c. Kita taksir nilai y harapan yang berkaitan dengan x = 300, sama dengan ….
   A. 15,34
   B. 18,02
   C. 21,19
   D. 27,89

d. Kita hitung interval kepercayaan 95% untuk nilai y harapan yang berkaitan dengan x = 300, kita peroleh …
   A. (11,78 ; 17,95)
   B. (13,65 ; 18,31)
   C. (14,84 ; 21,17)
   D. (15,78 ; 28,34)

2) Transformasi yang melinearkan $y = \dfrac{1}{(1+ae^{bx})^2}$ adalah …

A. $\left(\dfrac{1}{\sqrt{y}} - 1\right) = a + b\,\ell n\,x$

B. $\ell n\left(\dfrac{1}{\sqrt{y}} + 1\right) = a + x\,\ell n\,b$

C. $\ell n\left(\dfrac{1}{y^2} + 1\right) = \ell na + x\,\ell n\,b$

D. $\ell n\left(\dfrac{1}{\sqrt{y}} - 1\right) = \ell na + bx$

3) Suatu regresi linear berganda ditaksir untuk himpunan data dari 27 kali percobaan dengan empat prediktor x_1, x_2, x_3, dan x_4 diamati bersama dengan respons y. Diperoleh hasil sebagai berikut:

$\hat{\alpha} = -8{,}51$; $\hat{\beta}_1 = 2{,}37$; $\hat{\beta}_2 = 20{,}2$; $\hat{\beta}_3 = -0{,}828$; $\hat{\beta}_4 = 5{,}91$

Jumlah kuadrat karena regresi = 925,50

Jumlah kuadrat sesatan = 82,86

a. Nilai respons y taksiran untuk $x_1 = 16$; $x_2 = 0{,}5$; $x_3 = 5$; dan $x_4 = 4{,}6$ adalah ….
   A. 51,44
   B. 54,11
   C. 58,93
   D. 62,56

b. Deviasi standar sesatan taksiran $\hat{\sigma}$ adalah ….
   A. 0,83
   B. 1,94
   C. 5,19
   D. 7,43

c. Proporsi variabilitas y yang dijelaskan oleh regresi taksiran adalah ….
   A. 0,52
   B. 0,79

C. 0,92
D. 0,99

4) Pandang kembali soal 3 di atas. Sesatan standar taksiran untuk $\hat{\beta}_1$ dan $\hat{\beta}_2$ adalah masing-masing 0,062 dan 2,51. Maka,
   a. Interval kepercayaan 90% untuk $\hat{\beta}_1$ adalah ….
      A. (2,071 ; 2,963)
      B. (2,264 ; 2,476)
      C. (2,363 ; 2,831)
      D. (2,444 ; 3,010)
   b. Untuk menguji $H_0 : \beta_2 = 25$ kita hitung statistik penguji t, sama dengan ….
      A. $(-8,37)$
      B. $(-6,84)$
      C. $(-3,52)$
      D. $(-1,91)$

Cocokkanlah jawaban Anda dengan Kunci Jawaban Tes Formatif 2 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar. Kemudian, gunakan rumus berikut untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 2.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{\text{Jumlah Soal}} \times 100\%$$

Arti tingkat penguasaan: 90 - 100% = baik sekali
80 - 89% = baik
70 - 79% = cukup
< 70% = kurang

Apabila mencapai tingkat penguasaan 80% atau lebih, Anda dapat meneruskan dengan Kegiatan Belajar 3. **Bagus!** Jika masih di bawah 80%, Anda harus mengulangi materi Kegiatan Belajar 2, terutama bagian yang belum dikuasai.

Kunci Jawaban Tes Formatif

Tes Formatif 1

1) a. C b. D c. A d. C
   e. A
2) a. C b. B c. A d. A
   e. C f. A

Tes Formatif 2

1) a. D b. A c. B d. C
2) D
3) a. D b. B c. C
4) a. B b. D

Daftar Pustaka

Battacharyya, G.K. and R.A. Johnson. (1977). *Statistics Concepts and Methods.* John Willey. New York.

Freud. J. (1979). *Modern Elementary Statistics.* Prentice Hall.

Kooros, A. (1965). *Elements of Mathematical Economics.* Houghton Miffing Company. Boston.

Pfeffenberger, R.C. dan J.H. Peterson. (1977). *Statistical Methods for Business and Economics.* Richard D. Irwin. Illinois.

Robbins, H. dan J.V. Ryzin. (1975). *Introduction to Statistics.* Science Research Associates. Inc.

Siegel, S. (1956). Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill. New York.