MODUL 3

Penggunaan Persamaan Regresi, Pemeriksaan Model Regresi

Prof. Dr. Zanzawi Soejoeti

PENDAHULUAN

Pembahasan pada Modul 3 mata kuliah *Metode Statistika 2* ini merupakan kelanjutan dari modul kedua yang membahas tentang analisis regresi. Kalau pada modul kedua Anda telah mempelajari bagaimana mencari suatu persamaan regresi linier sederhana maka pada modul 3 ini lebih dititik-beratkan kepada penggunaan persamaan regresi sederhana. Persamaan regresi taksiran digunakan untuk menaksir atau memperkirakan mean y untuk nilai x tertentu, memperkirakan sesatan standar taksiran, dan menaksir mean y dengan interval kepercayaan. Dibahas juga memperkirakan satu respon untuk x tertentu, sesatan standar taksiran, dan interval kepercayaan. Pada penutup kegiatan belajar pertama dibahas juga mengenai kekuatan hubungan linier dengan menggunakan koefisien korelasi dan koefisien determinasi. Sedangkan pada kegiatan belajar kedua dibahas tentang pemeriksaan terhadap model regresi melalui residual atau analisis sisaan. Setelah mempelajari Modul 3 ini, secara khusus Anda diharapkan dapat memperoleh gambaran untuk:

a. melakukan penaksiran titik dan interval kepercayaan mean y untuk x tertentu;
b. melakukan penaksiran titik dan interval kepercayaan nilai tunggal y untuk x tertentu;
c. melihat dan menguji kecocokan model regresi linier sederhana.

KEGIATAN BELAJAR 1

Penggunaan Persamaan Regresi Taksiran

A. PERKIRAAN MEAN Y UNTUK X TERTENTU

Tujuan yang paling penting dalam analisis regresi adalah menggunakan model regresi taksiran untuk menaksir respons harapan yang berkaitan dengan tingkat variabel terkendali tertentu, misalnya mungkin kita ingin menaksir pengurangan yang diharapkan dalam nitrogen oksida untuk banyak bahan tambahan x^* tertentu, dalam semua mobil merek A, dengan menggunakan model hubungan linear. Menurut model linear yang kita pelajari dalam modul yang lalu, respons harapan pada nilai x^* variabel terkendali diberikan oleh $E(Y|x^*) = \alpha + \beta x^*$. Penaksir tak bias untuk ini adalah $\hat{\alpha} + \hat{\beta}x^*$ karena $\hat{\alpha}$ dan $\hat{\beta}$ adalah masing-masing penaksir tak bias untuk α dan β. Dengan perkataan lain, titik pada garis regresi taksiran yang berkaitan dengan nilai x^* memberikan taksiran tak bias untuk respons harapan. Sifat penaksir itu adalah berikut ini.

Untuk menaksir $E(Y|x^*) = \alpha + \beta x^*$, kita gunakan penaksir $\hat{\alpha} + \hat{\beta}x^*$.

Sesatan standar taksiran $s\sqrt{\dfrac{1}{n} + \dfrac{(x^* - \overline{x})^2}{S_{xx}}}$.

Distribusi: $\dfrac{(\hat{\alpha} + \hat{\beta}x^*) - (\alpha + \beta x^*)}{s\sqrt{\dfrac{1}{n} + \dfrac{(x^* - \overline{x})^2}{S_{xx}}}}$ adalah distribusi t dengan db = $n - 2$

Distribusi t ini dapat digunakan untuk menghitung interval kepercayaan dan melakukan uji hipotesis seperti biasa.

Interval kepercayaan 95% untuk respons harapan $E(Y|x^*) = \alpha + \beta x^*$ adalah:

$$\hat{\alpha} + \hat{\beta}x^* \pm t_{0,025}\, s\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$$

Untuk menguji hipotesis bahwa $E\left(Y \middle| x^*\right) = \mu_0$ kita gunakan statistik penguji

$$t = \frac{\hat{\alpha} + \hat{\beta} x^* - \mu_0}{s\sqrt{\frac{1}{n} + \frac{\left(x^* - \overline{x}\right)}{S_{xx}}}} \quad ; \quad \text{db} = n - 2$$

Contoh 3.1

Kita pandang kembali data pengurangan nitrogen oksida yang disajikan dalam Tabel 2.2 dan hitungan-hitungan untuk analisis regresi yang dituangkan dalam Tabel 2.4 Modul 2. Kita peroleh garis regresi taksiran.

$$\hat{y} = 2 + 0,387x$$

Pengurangan harapan yang berkaitan dengan nilai x^* = 4 bahan tambahan ditaksir dengan:

$$\hat{\alpha} + \hat{\beta} x^* = 2 + \left(0,387\right)4 = 3,548$$

Sesatan standar taksiran:

$$s\sqrt{\frac{1}{10} + \frac{\left(4 - 3,9\right)^2}{40,9}} = \left(0,304\right)\left(0,317\right) = 0,096$$

Interval kepercayaan 95% untuk mean pengurangan dalam nitrogen oksida pada x* = 4 adalah

$$\begin{aligned} 3,548 \pm t_{0,025}.0,096 &= 3,548 \pm \left(2,306\right)\left(0,096\right) \\ &= 3,548 \pm 0,22 \\ \text{atau } &\left(3,33; 3,77\right) \end{aligned}$$

Maka, kita 95% yakin akan mencapai suatu pengurangan dalam nitrogen oksida antara 3,33 dan 3,77.

Misalkan, kita juga ingin menaksir mean pengurangan pada x* = 7,5. Kita ikuti langkah-langkah yang sama, seperti di atas dan taksiran titiknya adalah:

$$\hat{\alpha} + \hat{\beta} x^* = 2 + \left(0,387\right)\left(7,5\right) = 4,90$$

Sesatan standar taksiran

$$(0,304)\sqrt{\frac{1}{10}+\frac{(7,5-3,9)^2}{40,9}}=(0,304)(0,646)=0,196$$

Internal kepercayaan 95% adalah

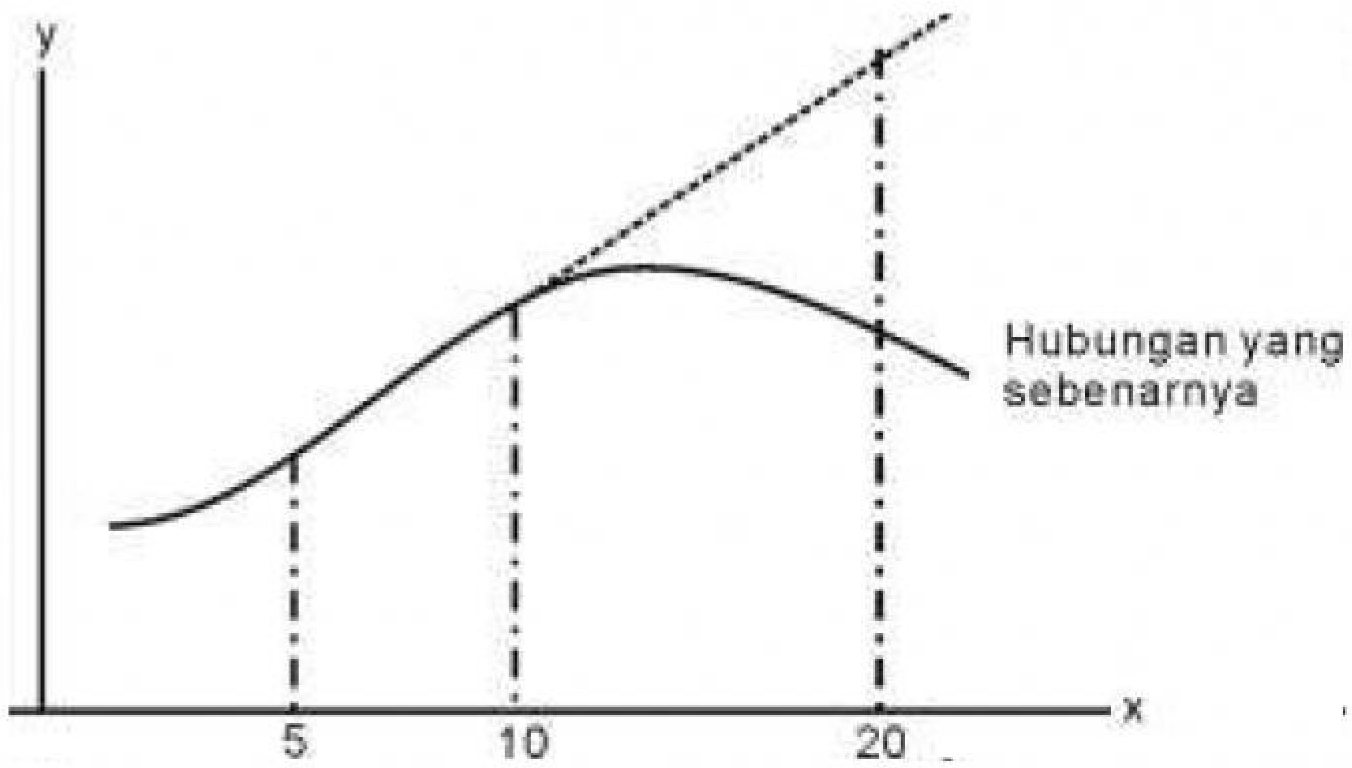
$$4,90\pm(2,304)(0,196)=4,90\pm0,45 \text{ atau } (4,45;\ 5,35)$$

Rumus untuk sesatan standar perkiraan menunjukkan jika x^* dekat dengan $\overline{x}$, sesatan standar itu lebih kecil dari sesatan standar jika x^* jauh dari $\overline{x}$. Hal ini ditegaskan oleh Contoh 3.1, yang menunjukkan bahwa sesatan standar untuk perkiraan pada x^* = 7,5 lebih dari dua kali sebesar sesatan standar pada nilai x^* = 4. Akibatnya, interval kepercayaan untuk x^* = 7,5 juga lebih lebar. Jadi, kita dapat menyimpulkan bahwa pada umumnya, perkiraan akan lebih tepat untuk x^* dekat dengan $\overline{x}$ daripada untuk nilai-nilai x^* yang terletak jauh dari mean $\overline{x}$.

Lagi pula, rumus untuk sesatan standar perkiraan mencerminkan kekuatan teknik regresi karena untuk x^* pada pusat nilai-nilai x, yakni $\overline{x}$, sesatan standar sama dengan $\frac{\sigma}{\sqrt{n}}$. Dengan perkataan lain, metode regresi tidak hanya memungkinkan kita untuk menentukan bagaimana x menjelaskan y, tetapi taksiran untuk mean $\alpha+\beta\overline{x}$ mempunyai variansi sama dengan jika sekiranya semua observasi dilakukan menggunakan satu kuantitas bahan tambahan $\overline{x}$ untuk semua percobaan.

Catatan:

Peringatan keras harus diperhatikan dalam melebarkan garis regresi taksiran untuk membuat perkiraan rentang panjang jauh dari rentang nilai-nilai x yang digunakan dalam percobaan.



Gambar 3.1
Bahaya dalam Perkiraan Rentang Panjang

Tidak hanya karena interval kepercayaan menjadi begitu lebar sehingga perkiraan berdasarkan persamaan regresi menjadi sangat tidak terpercaya, tetapi bahkan terjadi bahaya yang lebih besar. Jika pola hubungan antara variabel-variabel itu berubah drastis pada nilai x yang jauh maka data tidak memberi informasi apa pun yang dapat digunakan untuk menyidik perubahan seperti itu. Gambar 3.1 melukiskan keadaan seperti ini. Kita akan mengamati hubungan linear yang baik jika kita melakukan percobaan dengan nilai-nilai x dalam rentang 5 – 10, tetapi jika garis regresi taksiran diperluas untuk menaksir respons pada $x^* = 20$ maka taksiran kita akan berubah drastis dari yang kita duga sebelumnya.

B. PERKIRAAN SATU RESPON UNTUK X TERTENTU

Misalkan, kita berikan sejumlah tertentu x^* bahan tambahan pada satu mobil merek A dan kita ingin memperkirakan pengurangan dalam nitrogen oksida. Masalah ini berbeda dengan masalah yang telah kita pelajari di atas, di mana kita ingin menaksir mean pengurangan untuk populasi semua mobil merek A untuk sejumlah x^* bahan tambahan. Perkiraan itu masih ditentukan dari garis regresi taksiran, yakni nilai perkiraan respons adalah $\hat{\alpha} + \hat{\beta}x^*$, seperti dalam kasus di atas, tetapi sesatan standar nilai perkiraan di sini mengembang menjadi besar karena satu observasi lebih tidak pasti daripada mean beberapa observasi. Sekarang kita berikan rumus sesatan standar taksiran untuk kasus ini.

Sesatan standar taksiran untuk perkiraan satu observasi y pada nilai x^* tertentu adalah:

$$s\sqrt{1+\frac{1}{n}+\frac{\left(x^*-\bar{x}\right)}{S_{xx}}}$$

Rumus-rumus untuk menghitung interval kepercayaan dan statistik penguji harus disesuaikan seperlunya.

Contoh 3.2

Sekali lagi, kita pandang data pengurangan dalam nitrogen oksida yang tertuang dalam Tabel 2.2. Satu percobaan baru dilakukan dengan satu mobil merek A dan dengan x* = 4,5. Perkiraan pengurangan dalam nitrogen oksida adalah

$$\hat{\alpha}+\hat{\beta}x^* = 2+\left(0,387\right)\left(4,5\right) = 3,74$$

dan interval kepercayaan 95% untuk pengurangan dalam nitrogen oksida dan mobil ini adalah:

$$3,74 \pm \left(2,306\right)\left(0,304\right)\sqrt{1+\frac{1}{10}+\frac{\left(4,5-3,9\right)^2}{40,9}}$$
$$= 3,74 \pm 0,74 \quad \text{atau} \quad \left(3,00;\ 4,48\right)$$

Ini berarti kita 95% yakin bahwa mobil baru ini akan mempunyai pengurangan dalam nitrogen oksida antara 3,00 dan 4,48 jika digunakan 4,5 unit bahan tambahan. Ini benar karena 95% dari interval-interval yang dihitung dengan cara ini dari sampel yang diulang-ulang akan memuat pengukuran baru itu.

Dalam pembicaraan di atas, kita telah menggunakan data yang disajikan dalam Tabel 2.2 dalam Modul 2 untuk melukiskan berbagai inferensi yang berkaitan dengan model regresi garis lurus. Berikut ini akan kita sajikan Contoh 3.3 yang memberikan penerapan analisis regresi garis lurus untuk himpunan data yang berbeda.

Contoh 3.3

Dalam suatu studi untuk menentukan bagaimana keterampilan dalam mengerjakan suatu pekerjaan perakitan yang rumit dipengaruhi oleh banyak pelatihan. 15 orang karyawan baru diberi berbagai kuantitas pelatihan yang berkisar antara 3 sampai 12 jam. Setelah pelatihan, waktu mereka melakukan pekerjaan itu dicatat. Kita tulis x = lama pelatihan (dalam jam) dan y = waktu mengerjakan pekerjaan (dalam menit), diperoleh statistik sebagai berikut.

$\overline{x} = 7,2$; $S_{xx} = 33,6$; $S_{xy} = -57,2$; $S_{yy} = 160,2$; $\overline{y} = 45,6$

Hitunglah persamaan garis regresi taksiran!

1. Apakah data mendukung pernyataan bahwa waktu menyelesaikan pekerjaan menurun dengan bertambahnya jam pelatihan?
2. Taksirlah waktu menyelesaikan pekerjaan untuk 9 jam pelatihan, dan hitunglah interval kepercayaan 95%!
3. Hitunglah perkiraan y untuk x = 35 jam, dan berilah komentar tentang hasilnya!
4. Hitunglah interval kepercayaan 95% untuk satu observasi y yang berkaitan dengan 9 jam pelatihan!

Penyelesaian:

1. Taksiran kuadrat terkecil adalah:

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{-57,2}{33,6} = -1,702$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x} = 45,6 - (-1,702)(7,2) = 57,85$$

Maka, persamaan garis regresi taksiran adalah:

$$\hat{y} = 57,85 - 1,702x$$

2. Untuk menjawab pertanyaan itu, kita harus menguji $H_0 : \beta = 0$ terhadap $H_1 : \beta < 0$. Taksiran β adalah $\hat{\beta} = -1,702$. Untuk memperoleh sesatan standar taksiran kita hitung:

$$JKS = S_{yy} - \frac{S_{xy}^2}{S_{xx}} = 160,2 - \frac{(-57,2)^2}{33,6} = 62,824$$

$$s = \sqrt{\frac{JKS}{n-2}} = \sqrt{\frac{62,824}{13}} = 2,198$$

Sesatan standar taksiran $\left(\hat{\beta}\right) = \frac{s}{\sqrt{S_{xx}}} = \frac{2,198}{\sqrt{33,6}} = 0,379$

Statistik penguji t mempunyai nilai

$$t = \frac{-1,702}{0,379} = -4,49 \;\; ; \quad \text{db} = 13$$

Dengan db = 13 dan luas 0,01 di bawah dibatasi oleh titik $t_{0,01} = -2,650$. Karena nilai t hitungan = -4,49 lebih kecil dari –2,650 maka H_0 ditolak pada tingkat signifikansi 0,01. Kita simpulkan bahwa kenaikan lamanya pelatihan signifikan menurunkan mean waktu menyelesaikan pekerjaan pada rentang yang digunakan dalam percobaan.

3. Waktu menyelesaikan pekerjaan harapan yang berkaitan dengan x^* = 9 ditaksir sebagai

$$\hat{\alpha} + \hat{\beta}x^* = 57,85 + (-1,702)9 = 42,53 \text{ menit}$$

dan

$$\text{Sesatan standar taksiran} = s\sqrt{\frac{1}{15} + \frac{(9-7,2)^2}{33,6}} = 0,888$$

Oleh karena $t_{0,025} = 2,160$ dengan db = 13 maka interval kepercayaan yang dicari adalah:

$$42,5 \pm (2,160)(0,888) = 42,53 \pm 1,92$$
$$\text{atau} \quad (40,6;\ 44,5)$$

Oleh karena x = 35 jam jauh di luar rentang percobaan 3 sampai 12 jam maka tidak masuk akal untuk memperkirakan y pada x = 35 menggunakan garis regresi taksiran. Di sini, hitungannya memberikan:
Perkiraan waktu menyelesaikan pekerjaan

$$= 37,85 - 1,702(35)$$
$$= -21,72 \text{ menit}$$

suatu hasil yang tidak masuk akal.

4. Untuk menghitung interval kepercayaan, pertama-tama kita hitung nilai perkiraan:

$$\hat{\alpha} + \hat{\beta}x^* = 57,85 - 1,702(9) = 42,53 \text{ menit}$$

dan

$$\text{Sesatan standar taksiran} = s\sqrt{1 + \frac{1}{15} + \frac{(9-7,2)^2}{33,6}} = 2,37$$

Oleh karena $t_{0,025} = 2,160$ dengan db = 13 maka interval kepercayaan 95% yang dicari adalah:

$$42,53 \pm (2,160)(2,37) = 42,53 \pm 5,1192$$

atau $(37,41;\ 47,65)$

C. KEKUATAN HUBUNGAN GARIS LURUS

Untuk sampai pada ukuran kecukupan model garis lurus kita pelajari berapa banyak variasi dalam variabel respons dijelaskan oleh garis regresi taksiran. Untuk ini, kita pandang y_i observasi sebagai terdiri dari dua komponen ini.

$$\underbrace{y_i}_{\substack{\text{Nilai y}\\ \text{observasi}}} = \underbrace{\left(\hat{\alpha} + \hat{\beta}x_i\right)}_{\substack{\text{Dijelaskan oleh}\\ \text{hubungan linear}}} + \underbrace{\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)}_{\substack{\text{Residu atau penyimpangan}\\ \text{dari hubungan linear}}}$$

Dalam keadaan yang ideal, di mana semua titik terletak tepat pada garis itu, residu itu semuanya nol, dan nilai *y* selengkapnya dijelaskan oleh ketergantungan linearnya dengan *x*.

Kita dapat memandang jumlah kuadrat residu

$$JKS = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} \left(y_i - \alpha - \beta x_i\right)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

yang JKS membentuk satu bagian. Selisih

$$S_{yy} - JKS = S_{yy} - \left(S_{yy} - \frac{S^2xy}{Sxx}\right)$$

$$= \frac{S^2xy}{Sxx}$$

membentuk bagian yang lain. Dimotivasi oleh pemisahan observasi y di atas sekarang kita dapat memandang pemisahan variabilitas nilai *y* =

Syy	$=$	$\frac{S^2xy}{S_{xx}}$	$+$	JKS
Variabilitas total y		Variabilitas dijelaskan oleh hubungan linear		Residu atau variabilitas tak dijelaskan

Suku pertama ruas kanan persamaan ini dinamakan jumlah kuadrat (*JK*) karena regresi linear. Demikian juga, variabilitas total S_{yy} juga dinamakan JK total *y*. Supaya model garis lurus dapat dipandang memberikan kecocokan yang baik pada data, *JK* karena regresi linear harus merupakan bagian utama dari S_{yy}. Dalam keadaan yang ideal di mana semua titik terletak pada garis itu, *JKS* sama dengan nol sehingga S_{yy} dijelaskan secara lengkap oleh kenyataan bahwa nilai-nilai *x* berubah-ubah dalam percobaan itu, yakni hubungan linear antara *y* dan *x* semata-mata tanggung jawab terhadap variabilitas dalam nilai-nilai *y*.

Sebagai petunjuk seberapa baik model garis lurus cocok maka cukup beralasan untuk memandang bagian variabilitas *y* yang dijelaskan oleh hubungan linear.

$$\frac{\text{JK karena regresi linear}}{\text{JK total y}} = \frac{S_{xy}^2 / S_{xx}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

Ingat dari Modul 1 bahwa kuantitas:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

Dinamakan koefisien korelasi sampel. Jadi, kuadrat dari koefisien korelasi sampel merupakan bagian dari variabilitas *y* yang dijelaskan oleh hubungan linear.

Kekuatan hubungan linear diukur dengan

$$r^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

yang merupakan kuadrat dari koefisien korelasi sampel *r*.

Contoh 3.4

Dipunyai data hasil percobaan obat alergi yang dilakukan untuk mempelajari hubungan antara dosis obat dan lamanya alergi disembuhkan. Sepuluh orang pasien diikutkan dalam percobaan tersebut. Dalam Tabel 3.1 berikut disajikan data dosis (*x*) obat dan lamanya penyembuhan (*y*) untuk 10 orang pasien beserta hitungan-hitungan yang diperlukan untuk analisis selanjutnya.

Tabel 3.1
Dosis (x) dan Banyak Hari Penyembuhan (y) Sepuluh Orang Pasien dan Hitungan-hitungan

Dosis: *x*	**Lama Sembuh *y***	x^2	y^2	***xy***	$\hat{\alpha}+\hat{\beta}x$	**Residu** $\overline{e}$
3	9	9	81	27	7,15	1,85
3	5	9	25	15	7,15	-2,15
4	12	16	144	48	9,89	2,11
5	9	25	81	45	12,63	-3,63
6	14	36	196	84	15,37	-1,37
6	16	36	256	96	15,37	0,63
7	22	49	484	154	18,11	3,89
8	18	64	324	144	20,85	-2,85
8	24	64	576	192	20,85	3,15
9	22	81	484	198	23,59	-1,59
Jumlah = 59	151	389	2651	1003		0,04

$$\overline{x} = 5,9\ ; \quad \overline{y} = 15,1$$

$$S_{xx} = 389 - \frac{59^2}{10} = 40,9$$

$$S_{yy} = 2651 - \frac{151^2}{10} = 370,9$$

$$S_{xy} = 1003 - \frac{(59)(151)}{10} = 112,1$$

$$\hat{\beta} = \frac{112,1}{40,9} = 2,74$$

$$\hat{\alpha} = 15,1 - 2,74(5,9) = -1,07$$

$$JKS = 370,9 - \frac{112,1^2}{40,9} = 63,6528$$

Jadi, persamaan regresi taksiran: $\hat{y} = -1,07 + 2,74x$. Berapa besar variabilitas dalam y dijelaskan oleh model regresi garis lurus?

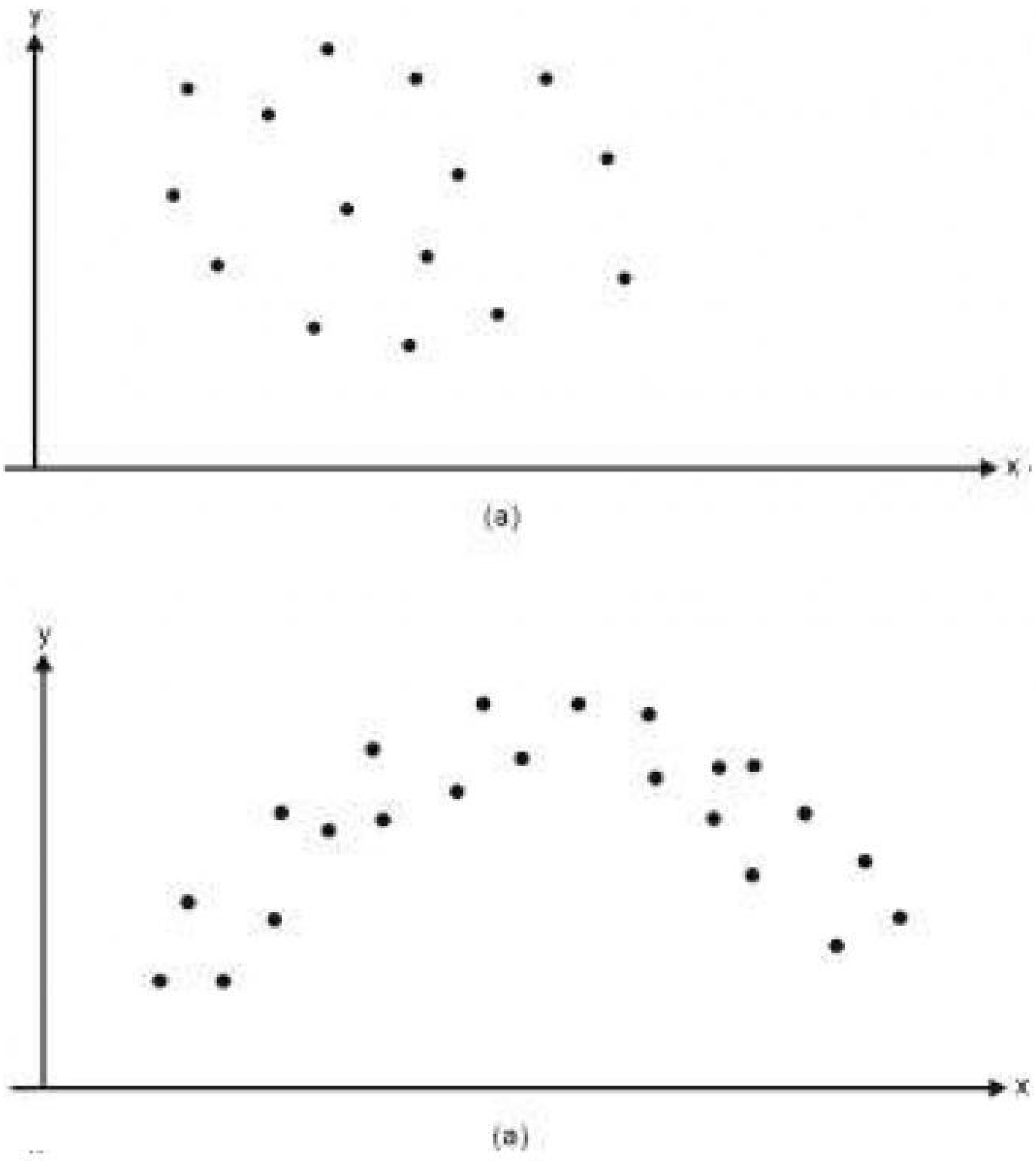
Untuk menjawab pertanyaan ini kita hitung

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{112,1^2}{(40,9)(370,9)} = 0,83$$

Ini berarti bahwa 83% variabilitas dalam y dijelaskan oleh model regresi linear, dan model linear itu kelihatannya memuaskan dalam hal ini.

Jika nilai r^2 kecil, kita hanya dapat menyimpulkan bahwa hubungan garis lurus tidak cukup cocok untuk data yang kita punyai. Hal semacam itu dapat timbul karena alasan-alasan sebagai berikut.

1. Hanya ada sedikit hubungan antara variabel-variabel itu, dalam arti bahwa diagram titik gagal untuk menunjukkan sesuatu pola, sebagaimana dilukiskan dalam Gambar 3.2 (a). Dalam hal ini, penggunaan model regresi yang lain tidak mungkin mengurangi JKT atau menjelaskan bagian yang besar dalam S_{yy} .
2. Ada hubungan yang menonjol, tetapi bersifat tidak linear, yakni diagram titik bergerombol di sekitar kurva bukan di sekitar garis lurus. Bagian S_{yy} yang dijelaskan oleh regresi garis lurus kecil karena model itu tidak sesuai. Suatu hubungan yang lain mungkin meningkatkan cukup banyak kesesuaian. Gambar 3.2 (b) melukiskan hal seperti itu, di mana JKT dapat dikurangi dengan menaksir kurva yang sesuai pada datanya.



Gambar 3.2
Pola Diagram Titik (a) Tidak Ada Hubungan (b) Hubungan Tidak Linear

LATIHAN

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

1) Untuk mempelajari kekerasan air pada rasa, diperoleh data berikut dari contoh air minum dari delapan tempat.
   x = banyak magnesium (miligram per liter)
   y = nilai rasa

x	8,7	9	11	8,5	9,2	12	12	18
y	25	25	26	48	65	87	90	100

   a. Dengan memandang model regresi linear, hitunglah interval kepercayaan 90% untuk lerengan garis regresinya!
   b. Taksirlah nilai rasa yang diharapkan jika $x = 10$, dan hitung pula interval kepercayaan 95%!
   c. Untuk satu contoh yang mempunyai $x = 10$, perkirakan nilai rasanya dan hitung pula interval kepercayaan 95%!
   d. Berapa besar variabilitas dalam y dijelaskan oleh model regresi linear itu?
   e. Hitunglah koefisien korelasi sampel!

2) Dalam suatu percobaan yang dirancang guna menentukan hubungan antara dosis pupuk kompos x dan hasil tanaman y, dicatat beberapa nilai statistik sebagai berikut.

   $n = 15$ $\quad\quad$ $\overline{x} = 10,8$ $\quad\quad$ $\overline{y} = 122,7$

   $S_{xx} = 70,6$ $\quad\quad$ $S_{yy} = 98,5$ $\quad\quad$ $S_{xy} = 68,3$

   Anggap hubungan linear
   a. Hitunglah persamaan garis regresi kuadrat terkecil!
   b. Hitunglah *JKS* dan taksirlah σ^2 !
   c. Apakah data mendukung dugaan pembuat percobaan bahwa rata-rata kenaikan dalam hasil tanaman per kenaikan satu unit dalam dosis paling sedikit 0,6? (Untuk rentang x dalam percobaan)?
   d. Hitunglah interval kepercayaan 95% untuk hasil tanaman harapan yang berkaitan dengan $x = 12$!

e. Hitunglah interval kepercayaan 95% untuk satu observasi y yang berkaitan dengan $x = 12$!
f. Hitunglah interval kepercayaan 90% untuk hasil tanaman harapan yang berkaitan dengan $x = 15$!
g. Hitunglah interval kepercayaan 90% untuk satu observasi y yang berkaitan dengan $x = 15$!
h. Hitunglah koefisien korelasi sampel!
i. Berapa besar variabilitas dalam y dijelaskan oleh model regresi linear?

3)

a. Tunjukkan bahwa koefisien korelasi sampel r dan lerengan $\hat{\beta}$ garis regresi taksiran mempunyai hubungan sebagai:

$$r = \beta \frac{\sqrt{S_{xx}}}{\sqrt{S_{yy}}}$$

b. Tunjukkan bahwa $JKS = \left(1 - r^2\right) S_{yy}$.

c. Tunjukkan bahwa jumlah kuadrat karena regresi S_{xy}^2 / S_{xx} dapat juga ditulis sebagai $\hat{\beta}^2 S_{xx}$

RANGKUMAN

Kecocokan garis lurus taksiran pada data diukur dengan r^2, yang menunjukkan bagian variabilitas y yang dijelaskan oleh hubungan linear antara y dan x. Nilai r^2 yang rendah hanya menunjukkan bahwa hubungan linear tidak cocok, tetapi mungkin masih ada hubungan kurva antara dua variabel itu.

Rumus-rumus:

Pada nilai $x = x^*$ tertentu, respons harapan adalah $(\alpha + \beta x^*)$. Inferensi tentang respons harapan didasarkan atas:

Penaksir: $\hat{\alpha} + \hat{\beta} x^*$

Sesatan standar taksiran: $s\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$

Interval kepercayaan $100(1-\alpha)\%$ untuk respons harapan pada x^* adalah

$$\hat{\alpha}+\hat{\beta}x^* \pm t_{\frac{\alpha}{2}} s\sqrt{\frac{1}{n}+\frac{(x^*-\overline{x})^2}{S_{xx}}}$$

Satu respons pada nilai tertentu x = x^* diperkirakan dengan $\hat{\alpha}+\hat{\beta}x^*$, mempunyai:

Sesatan standar taksiran: $s\sqrt{1+\frac{1}{n}+\frac{(x^*-\overline{x})^2}{S_{xx}}}$

Interval kepercayaan $100(1-\alpha)\%$ untuk perkiraan satu respons adalah:

$$\hat{\alpha}+\hat{\beta}x^* \pm t_{\frac{\alpha}{2}} s\sqrt{1+\frac{1}{n}+\frac{(x^*-\overline{x})^2}{S_{xx}}}$$

Pemisahan variabilitas:

Variabilitas yang dijelaskan oleh hubungan linear = $\frac{S_{xy}^2}{S_{xx}}=\hat{\beta}^2 S_{xx}$

Residu atau variabilitas tak terjelaskan = *JKS*.
Jumlah variabilitas $y = S_{yy}$.

Bagian variabilitas y yang dijelaskan oleh regresi linear:

$$r^2=\frac{S_{xy}^2}{S_{xx}S_{yy}}$$

Koefisien korelasi sampel: $r=\frac{S_{xy}}{\sqrt{S_{xx}.S_{yy}}}$

TES FORMATIF 1

Pilihlah satu jawaban yang paling tepat!

I. Dipunyai data berpasangan (x, y) yang tidak kita sajikan di sini, tetapi telah kita hitung beberapa statistik sebagai berikut.
$n = 15$; $\overline{x} = 8,3$; $\overline{y} = 54,8$; $S_{xx} = 5,6$; $S_{xy} = -12,4$;
$S_{yy} = 38,7$

1) Kita hitung persamaan regresi taksiran $\hat{y} = \hat{\alpha} + \hat{\beta}x$, kita peroleh ....
   A. $\hat{\beta} = 1,83$
   B. $\hat{\beta} = 3,12$
   C. $\hat{\beta} = -2,21$
   D. $\hat{\beta} = -4,72$

2) Dengan soal nomor 1 kita peroleh ....
   A. $\hat{\alpha} = 73,18$
   B. $\hat{\alpha} = 62,12$
   C. $\hat{\alpha} = 56,74$
   D. $\hat{\alpha} = 47,77$

3) Variansi sesatan σ^2 kita taksir dengan s^2, kita peroleh ....
   A. 0,865
   B. 0,999
   C. 1,873
   D. 1,998

4) Untuk menguji $H_0 : \beta = -2$, kita hitung statistik penguji t, kita peroleh ....
   A. 0,37
   B. 0,53
   C. 0,83
   D. 1,36

5) Kita hitung interval kepercayaan 95% untuk nilai y harapan yang berkaitan dengan x = 10, kita peroleh ....
   A. (35,11 ; 53,17)
   B. (49,52 ; 52,61)
   C. (51,63 ; 59,37)
   D. (52,22 ; 59,88)

6) Kita hitung interval kepercayaan 90% untuk satu nilai y yang berkaitan dengan x = 10, kita peroleh ....
   A. (48,85 ; 53,31)
   B. (49,11 ; 54,47)
   C. (50,06 ; 59,31)
   D. (52,67 ; 61,17)

II. Suatu percobaan dilakukan untuk menentukan hubungan antara variabel independen x dan variabel respons y, diperoleh data yang tidak kita sajikan di sini, tetapi telah kita hitung beberapa statistik sebagai berikut.
$n = 20$; $\sum x = 160$; $\sum y = 240$; $\sum x^2 = 1536$;
$\sum xy = 1832$; $\sum y^2 = 2965$

7) Kita hitung S_{xx} sama dengan ....
   A. 172
   B. 188
   C. 202
   D. 256

8) Kita hitung S_{xy} , sama dengan ....
   A. 47
   B. 69
   C. 85
   D. 98

9) Kita hitung S_{yy} sama dengan ....
   A. –88
   B. –62
   C. 53
   D. 78

10) Kita hitung koefisien korelasi sampel sama dengan ....
    A. -0,975
    B. –0,597
    C. 0,527
    D. 0,834

11) Kita hitung lerengan garis regresi taksiran maka kita peroleh ....
    A. –0,122
    B. –0,344
    C. 0,433
    D. 0,544

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.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{11} \times 100\%$$

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

Pemeriksaan Model Regresi

Sangat penting bagi kita untuk menyadari kenyataan bahwa analisis regresi belum lengkap dengan hanya menaksir model dengan metode kuadrat terkecil (MKT) menghitung interval kepercayaan, dan menguji berbagai hipotesis. Langkah-langkah ini hanya mengatakan separuh dari ceritanya, yaitu inferensi statistik yang dapat dibuat jika model yang dipostulasikan memadai. Dalam banyak studi dalam bidang ilmu sosial dan kealaman, hubungan antara variabel-variabel adalah empiris dalam bentuknya sehingga kita tidak pernah dapat yakin bahwa suatu model itu benar. Oleh karena itu, kita harus memakai strategi berikut.

1. Sementara diterima suatu model.
2. Hitunglah taksiran kuadrat terkecil dan hitung nilai-nilai residu.
3. Tinjaulah model itu dengan memeriksa nilai-nilai residu.

Dalam banyak penelitian langkah (3) mengusulkan cara-cara mengubah model yang layak. Kembali ke langkah (1), model berubah diterima, dan iterasi ini berlanjut sampai suatu model diperoleh dari data itu tidak lagi kelihatan bertentangan dengan anggapan-anggapan yang dibuat tentang model itu.

A. MEMERIKSA RESIDU

Sekarang kita membicarakan beberapa teknik untuk meninjau model itu yang tidak hanya berlaku untuk model garis lurus tetapi juga untuk semua perluasan yang kita pandang kemudian. Sekali model telah ditaksir dengan metode kuadrat terkecil, semua informasi tentang variasi yang tidak dapat dijelaskan oleh model itu termuat dalam residu.

$$e_i = y_i - \hat{y}_i \;\; ; \;\; i = 1, 2, ..., n$$

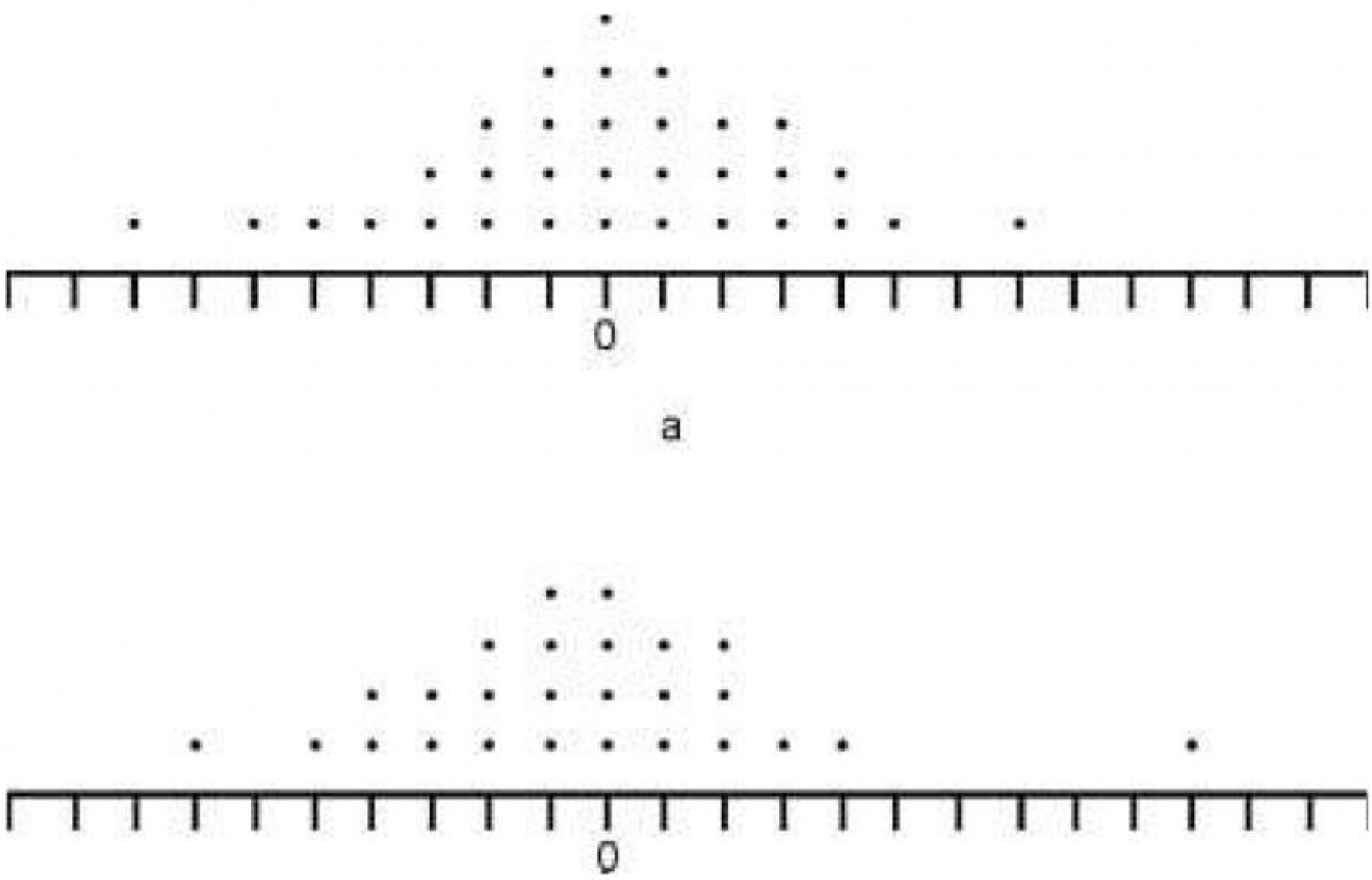
dengan y_i adalah nilai observasi dan $\hat{y}_i$ menunjukkan nilai yang berkaitan yang diperkirakan dengan model taksiran kuadrat terkecil, misalnya dalam hal model regresi linear sederhana, $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$.

Kita ingat bahwa dari model regresi garis lurus yang dipelajari di atas, kita telah membuat anggapan independensi, variansi konstan, dan distribusi normal untuk komponen sesatan e_i. Prosedur inferensi didasarkan atas anggapan-anggapan ini. Jika model itu benar, residu dapat dipandang sebagai taksiran untuk sesatan e_i yang berdistribusi $N\left(0;\sigma^2\right)$.

Untuk menentukan kepantasan model yang sementara diterima, kita dapat memeriksa residu dengan menggambarkannya dalam kertas grafik. Maka, kita mengenali adanya pola sistematik yang dibentuk oleh gambar residu itu kita dapat mencurigai bahwa beberapa anggapan tentang model itu tidak berlaku. Ada banyak cara menggambarkan residu, bergantung pada aspek apa yang ingin diperiksa. Kita pelajari beberapa di sini untuk melukiskan teknik-teknik itu.

1. Histogram atau Diagram Titik

Untuk menggambarkan keseluruhan tingkah laku residu kita dapat menggambarkan suatu histogram dengan interval kelas yang sesuai diagram titik pada suatu garis untuk observasi yang lebih sedikit, misalnya dalam diagram titik, seperti yang tertuang dalam Gambar 3.3(a) pola data kelihatan serupa dengan suatu sampel dari populasi normal dan di sana tidak tampak adanya observasi yang "liar".

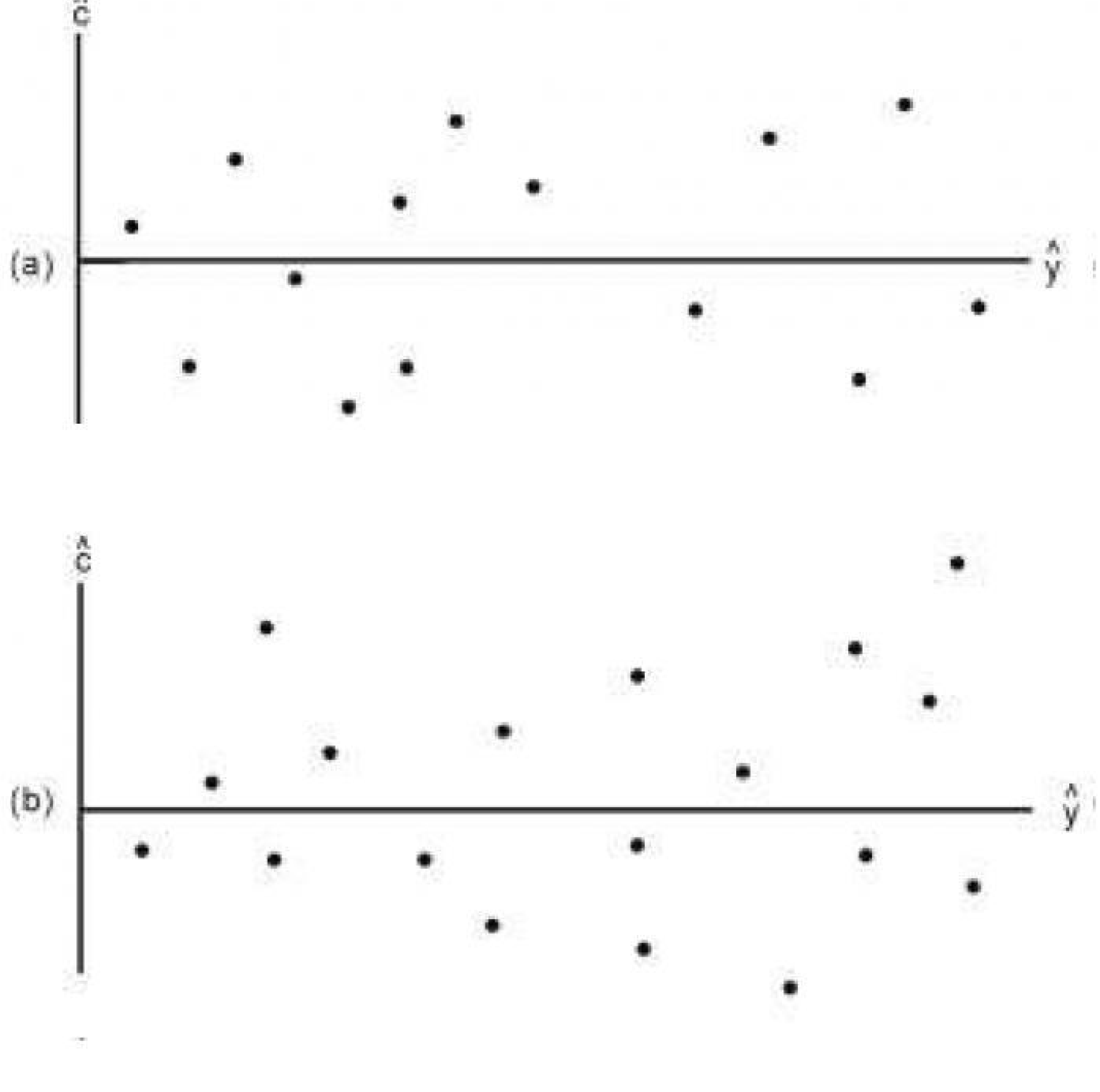


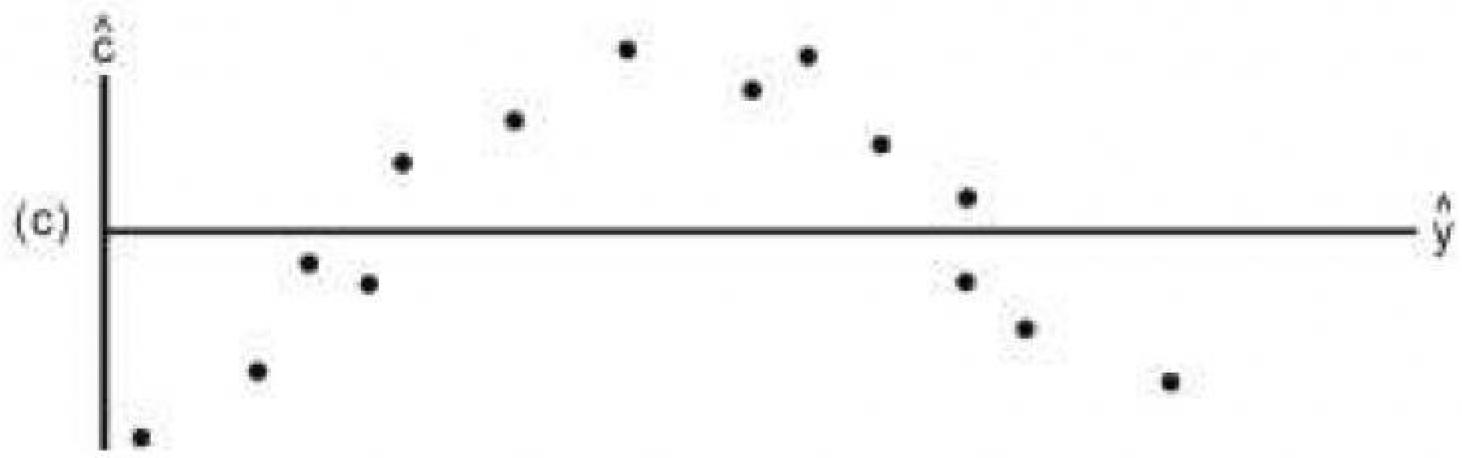
Gambar 3.3
Diagram Titik Residu

Berbeda dengan itu, Gambar 3.3 (b) melukiskan keadaan di mana distribusi itu tampak sangat normal, kecuali satu residu yang terletak jauh di kanan. Keadaan yang menghasilkan observasi yang bersangkutan menuntut perhatian serius untuk meyakinkan bahwa observasi ini bukan pengaruh yang tidak semestinya dari penaksiran model.

2. Residu versus Nilai Perkiraan

Gambar residu $\hat{e}_i$ versus nilai perkiraan $\hat{y}_i$ sering kali membantu menyidik ketidaksesuaian suatu hubungan yang dianggapkan atau pelanggaran anggapan variansi sesatan konstan. Gambar 3.4 melukiskan beberapa fenomena yang sering kita jumpai. Jika titik-titik membentuk (memberi kesan) gerombolan mendatar di sekitar nol, seperti dalam Gambar 3.4 (a) maka tidak ada hal-hal yang abnormal yang ditunjukkan. Jika lebar gerombolan titik-titik itu bertambah dengan jelas jika nilai-nilai $\hat{y}$ bertambah besar, seperti ditunjukkan Gambar 3.4 (b) ini menunjukkan bahwa variansi sesatan σ^2 cenderung naik dengan naiknya tingkat respons. Maka kita akan mencurigai keabsahan anggapan variansi konstan dalam model itu. Ada banyak cara mentransformasi data guna menstabilkan variansi, tetapi kita tidak akan membicarakannya di sini.



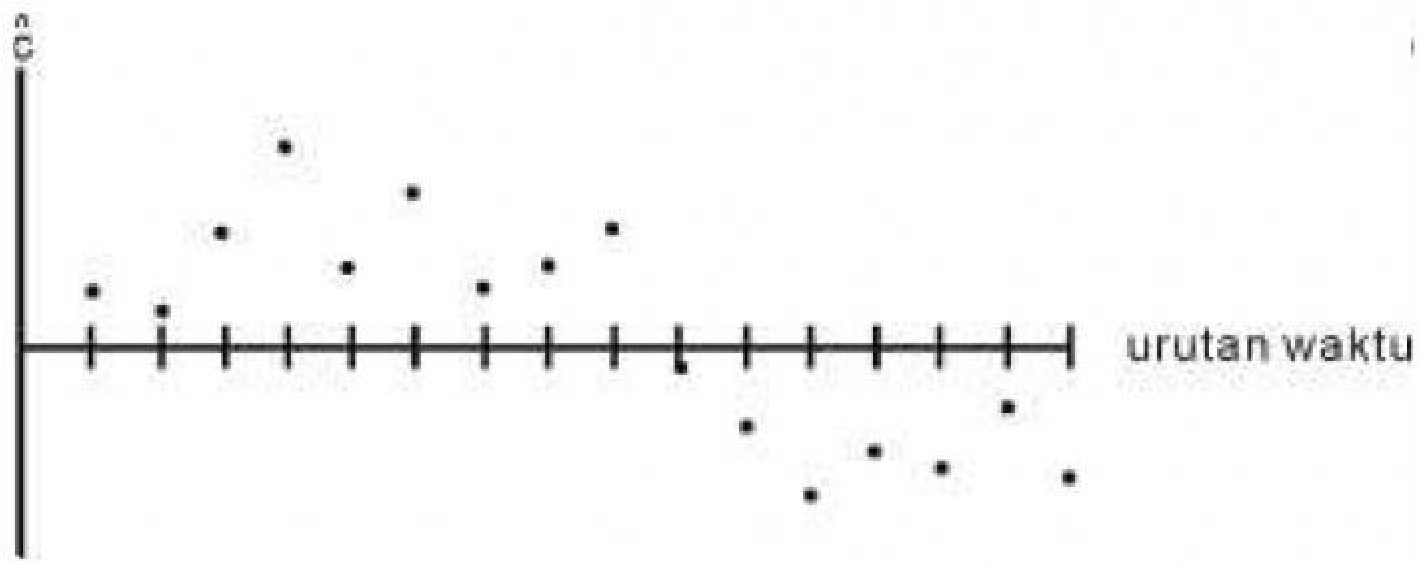


Gambar 3.4
Gambar Grafik Residu versus Nilai Perkiraan

Gambar 3.4 (c) menunjukkan bahwa residu membentuk suatu pola sistematik. Titik-titik tidak terdistribusi secara random di sekeliling sumbu $\hat{y}$, tetapi berkecenderungan pertama-tama naik terus, kemudian menurun. Ini akan membuat kita mencurigai bahwa model itu tidak memadai dan bahwa suatu suku kuadrat atau suatu model tidak linear yang lain harus dicoba.

3. Residu versus Urutan Waktu

Telah kita sebutkan sebelumnya bahwa pelanggaran anggapan yang paling serius terjadi jika sesatan e_i tidak independen. Tidak independensi ini kerap kali terjadi dalam penerapan di bidang bisnis dan ekonomi, di mana observasi dikumpulkan dalam urutan waktu dengan maksud menggunakan teknik regresi untuk memperkirakan *trend* di masa mendatang. Dalam banyak percobaan yang lain, percobaan dilakukan dalam waktu berturut-turut. Dalam peristiwa yang mana pun, bilamana waktu berbeda cukup besar selama percobaan, suatu gambar grafik residu versus urutan waktu sering kali menyidik pelanggaran terhadap anggapan independensi, misalnya grafik dalam Gambar 3.5 menunjukkan pola sistematik dalam hal adanya deretan nilai-nilai yang tinggi diikuti oleh deretan nilai-nilai yang rendah. Ini menunjukkan bahwa residu yang berurutan berkorelasi positif dan kita mencurigai adanya pelanggaran terhadap anggapan independen. Anggapan independen juga dapat diperiksa dengan menggambarkan pasangan berurutan $(\hat{e}_i, \hat{e}_{i-1})$, dengan $\hat{e}_1$ menunjukkan residu dari nilai y yang pertama diamati, $\hat{e}_2$ menunjukkan yang kedua, dan seterusnya. Independensi dikuatkan jika diagram titik merupakan kelompok titik tanpa pola, sedangkan titik-titik yang berkelompok sepanjang garis mengesankan kurang kuatnya independensi di antara observasi yang berdekatan.



Gambar 3.5
Grafik Residu vs Urutan Waktu

Penting untuk diingat bahwa keyakinan kita dalam prosedur inferensi statistik berkaitan dengan keabsahan anggapan dalam modelnya. Inferensi yang dilakukan secara mekanis dapat menyesatkan jika beberapa anggapan dalam modelnya nyata sekali dilanggar. Pemeriksaan residu adalah suatu bagian yang penting dalam analisis regresi karena ini membantu untuk menyidik setiap ketidakkonsistenannya antara data dan model yang dipostulasikan. Jika tidak ada keabnormalan tampak dalam proses ini maka model dapat kita pandang memadai dan selanjutnya kita ikuti dengan inferensi yang bersangkutan. Jika tidak demikian, kita harus mencari model yang lebih sesuai.

Gambaran teknik-teknik standar yang digunakan dalam menangani data jika anggapan tertentu terlihat dilanggar akan membawa kita jauh di luar jangkauan kuliah ini. Karena itu kita di sini hanya menekankan pentingnya memeriksa anggapan-anggapan dan menyajikan metode deskriptif untuk digunakan dalam mempelajari residu.

B. UJI KEKURANGCOCOKAN BERDASARKAN PERCOBAAN BERREPLI-KASI

Kita pelajari metode untuk melakukan uji ketidakcocokkan model garis lurus taksiran berdasarkan data replikasi suatu percobaan. Konsep ini dapat diperluas untuk setiap model taksiran, tetapi untuk pertama kali sebaiknya kita bekerja untuk model yang telah kita kenal dengan baik, misalkan k tingkat variabel independen $x_1, x_2, \ldots, x_k$ yang berbeda kita ikutkan dalam percobaan, dan observasi y masing-masing diulang $n_1, n_2, \ldots, n_k$ kali pada tingkat-tingkat x ini. Tabel 3.2 menunjukkan keadaan data hasilnya. Dalam setiap baris tabel itu, nilai x tetap sehingga variabilitas nilai-nilai y semata-mata karena komponen sesatan yang variansinya σ^2.

Tabel 3.2
Pola Data dengan Replikasi

Nilai x berbeda	Nilai y replikasi	Mean JK	db
x_1 x_2 $\vdots$ x_k	$y_{11}\ y_{12} \cdots y_{1n_1}$ $x_2\ y_{22} \cdots y_{2n_2}$ $\vdots$ $x_k\ y_{k_2} \cdots y_{kn_k}$	$\overline{y}_1$ S_1 $\overline{y}_2$ S_2 $\vdots$ $\vdots$ $\overline{y}_k$ S_k	$n_1 - 1$ $n_2 - 1$ $\vdots$ $n_k - 1$
		Jumlah JK_{sm}	n – k dengan $n = \sum_{i=1}^{k} n_i$

Dari setiap baris kita dapat menghitung jumlah kuadrat penyimpangan (deviasi) dari *mean*-nya, yang memberikan nilai taksiran untuk σ^2 jika dibagi dengan derajat bebasnya. Misalnya:

$$S_1 = \sum_{i=1}^{n_1} \left(y_{1i} - \overline{y}_1 \right)^2 \quad ; \text{db} = n_1 - 1$$

dengan $\overline{y}_1$ mean nilai-nilai y dalam baris pertama dalam Tabel 3.2.

Jumlah kuadrat penyimpangan dan derajat bebasnya disajikan dalam dua kolom terakhir tabel itu. Jumlah semua jumlah kuadrat ini.

$$JK_{sm} = \sum_{i=1}^{k} S_i$$

dinamakan jumlah kuadrat sesatan murni, dan derajat bebas yang berkaitan adalah $\sum_{1}^{k}(n_i - 1) = (n-k)$, dengan n banyak semua observasi y. Kuadrat rata-ratanya KR_{sm} didefinisikan sebagai

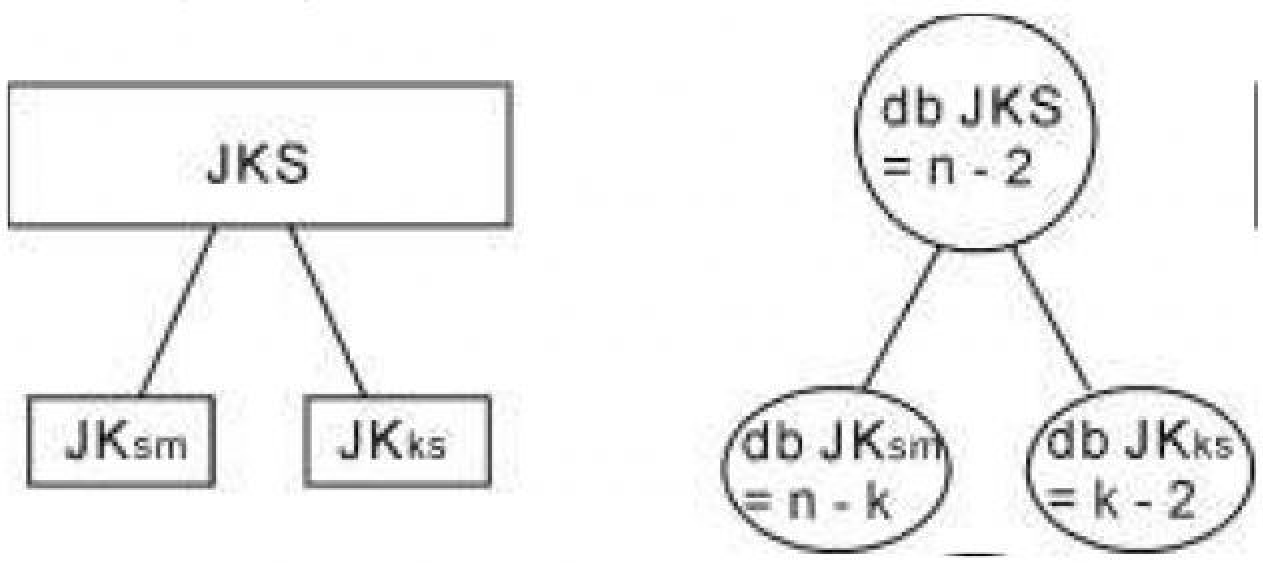
$$KR_{sm} = \frac{JK_{sm}}{n-k} = \text{kuadrat rata-rata sesatan murni.}$$

Memberikan taksiran tak bias untuk variansi sesatan σ^2, apa pun model hubungan y dengan x.

Dengan mengambil n titik-titik data, kita dapat menaksir garis regresi dan menghitung jumlah kuadrat sesatan JKS, seperti biasa.

$$JKS = S_{yy} - \hat{\beta}^2 S_{xx} \;;\quad \text{db} = n - 2$$

Gambar 3.6 menunjukkan pemisahan JKS ini menjadi dua bagian: pertama JK_{sm}, seperti yang kita jelaskan di atas dan *kedua* sisanya $JK_{ks} = JKS - JK_{sm}$ yang dinamakan jumlah kuadrat kurang sesuai. Derajat bebas yang berkaitan dipecah menjadi dua dengan cara yang sama, dan memberikan $db = (k-2)$ untuk jumlah kuadrat kurang sesuai



Gambar 3.6
Pemisahan JKS dan Derajat Bebas

Sekarang kita ingat:

$$KR_{ks} = \frac{JK_{ks}}{k-2} = \text{kuadrat rata-rata kurang sesuai.}$$

Jika model linear itu benar, KR_{ks} juga menaksir variansi sesatan σ^2, seperti hanya KR_{sm}. Sebaiknya jika model tak linear yang sesuai maka KR_{ks} terlalu besar sebagai taksiran untuk σ^2 dan mungkin nyata-nyata (signifikan) lebih besar dari KR_{sm}. Jadi, uji kekurangsesuaian untuk model linear dapat dilakukan dengan menghitung nilai F sebagai berikut.

$$F = \frac{KR_{ks}}{KR_{sm}} = \frac{\left(JKS - JK_{sm}\right)/\left(k-2\right)}{JK_{sm}/\left(n-k\right)}$$

dengan db = (k – 2 ; n – k)

Jika uji F ini menunjukkan signifikansi, artinya F hitungan lebih besar dari tabel dengan *db* yang sesuai pada tingkat signifikansi yang dipilih maka kita simpulkan bahwa kekurangsesuaian untuk model garis lurus signifikan sehingga kita harus mencari model alternatif untuk hubungan itu. Jika F tidak signifikan, sedikit manfaat yang dapat kita peroleh dengan menggunakan model yang lebih rumit. Kenyataan ini tidak dengan sendirinya meyakinkan bahwa model garis lurus cukup memadai. Memadainya model linear ditentukan oleh besarnya r^2, seperti yang telah kita pelajari di muka.

Contoh 3.5

Lima tingkat variabel independen x diikutkan dalam suatu percobaan, beberapa di antaranya dengan observasi replikasi dalam y sehingga memberikan 11 titik-titik data seperti tertuang dalam Tabel 3.3

Taksirlah persamaan regresi garis lurus dan lakukan uji ketidaksesuaian. Apakah kesimpulan yang dapat ditarik?

Tabel 3.3
Data dengan Replikasi

x	2	2	2	3	3	4	5	5	6	6	6
y	4	3	8	18	22	24	24	18	13	10	16

Pertama-tama kita susun data itu, seperti ditunjukkan dalam dua kolom pertama Tabel 3.4.

Tabel 3.4
Hitungan-hitungan untuk Uji Kekurangsesuaian

x	y	$\overline{y}$	JK	db
2	4 ; 3 ; 8	5	14	2
3	18 ; 22	20	8	1
4	24	24	0	0
5	24 ; 18	21	18	1
6	13 ; 10 ; 16	13	18	2
		Jumlah	58	6

Untuk $x = 2$, mean y adalah 5 dan jumlah kuadratnya dihitung:

$$(4-5)^2+(3-5)^2+(8-5)^2=14$$

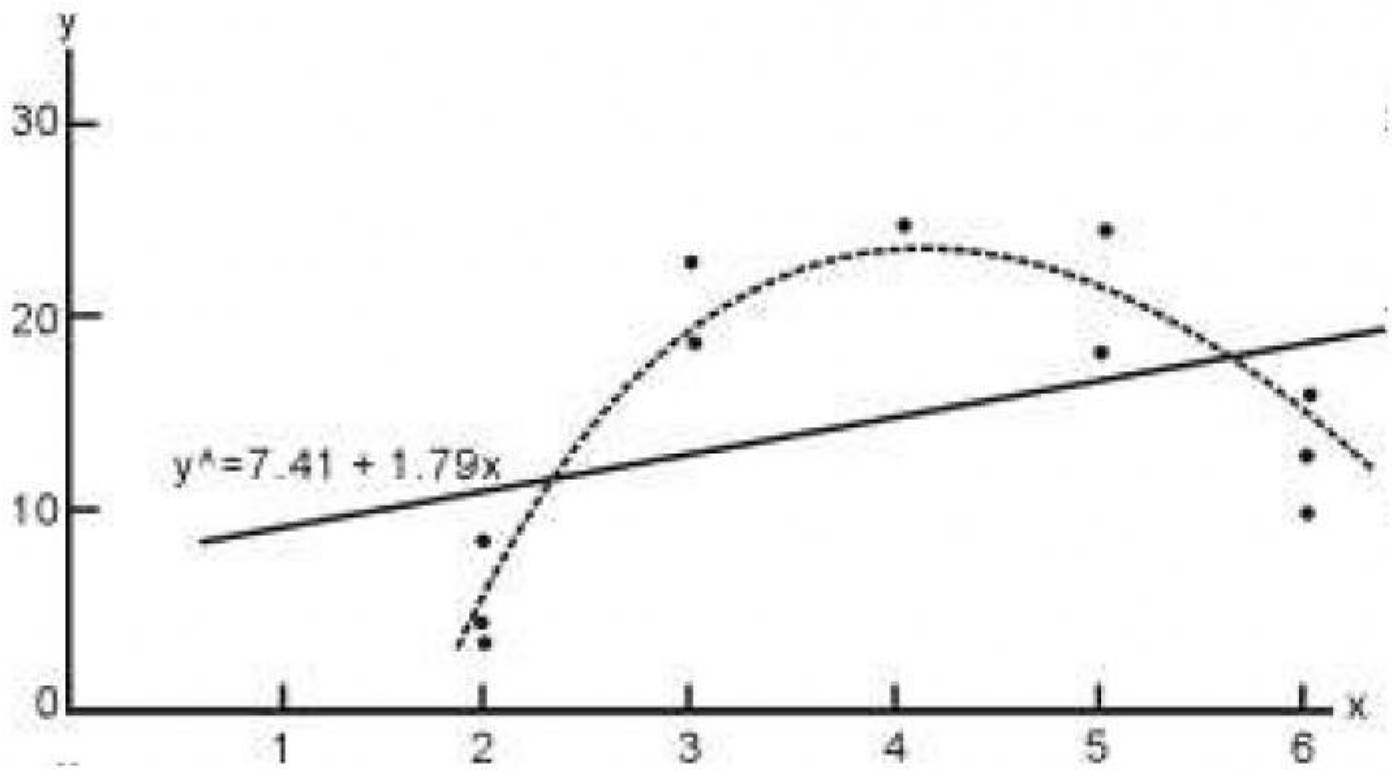
Hitungan-hitungan yang sama dilakukan untuk tiap baris dalam Tabel 3.4 untuk mendapatkan:

$$JK_{sm}=14+8+18+18=58$$
$$db=2+1+1+2=6$$

Menggunakan rumus-rumus yang telah kita pelajari untuk menaksir persamaan regresi garis lurus dari 11 pasang nilai (x, y), kita peroleh: $\hat{\beta}=1,786$; $\hat{\alpha}=7,41$; $S_{yy}=571$; dan JKS = 481 dengan db = 11 – 2 = 9. Dengan demikian, $JK_{ks}=481-58=423$ dengan db = 9 – 6 = 3. Jadi, ukuran kita untuk kekurangsesuaian adalah:

$$F=\frac{\frac{423}{3}}{\frac{58}{6}}=14,6 \quad ; \quad \text{db} = (3;\ 6)$$

yang sangat signifikan karena nilai tabel 5% untuk F dengan db = (3 ; 6) adalah 4,76. Ini menunjukkan adanya kekurangcocokkan yang signifikan untuk model regresi garis lurus. Diagram titik data yang ada dalam Tabel 3.3 digambarkan dalam Gambar 3.7. Kesimpulan yang dapat kita tarik dari uji ketidakcocokkan adalah dengan jelas dibenarkan oleh pandangan visual diagram titik ini, yang menunjukkan adanya hubungan kurva linear. Dalam keadaan ini akan baik sekali digambarkan nilai-nilai residu untuk melihat bagaimana titik-titik itu memperlihatkan adanya kekurangcocokkan.



Gambar 3.7
Diagram Titik Data Tabel 3.3 dan Garis Taksiran Hubungannya Lebih Baik Digambarkan oleh Garis Patah-Patah

Dalam modul mendatang akan kita pelajari kemungkinan mentransformasi model hubungan tertentu sehingga menjadi linear.

LATIHAN

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

1) Taksiran garis regresi kuadrat terkecil menghasilkan residu sebagai berikut.

$\hat{y}$	11,3	14,8	18,4	22,0	25,5	27,0	31,2	32,7
Residu	-0,3	-0,2	-5,4	2,0	0,5	5,0	-9,2	6,3

34,1	36,2	39,8	43,4	45,5	46,2	46,9	46,9
12,9	-4,2	-14,8	7,6	-1,5	16,8	1,1	-16,9

Gambarlah residu ini terhadap nilai perkiraan $\hat{y}$ dan tentukan adanya anggapan yang tampak dilanggar.

2) Seorang mahasiswa menggunakan metode kuadrat terkecil untuk mendapatkan persamaan garis taksiran $\hat{y} = 264,3 + 18,77x$ bagi produk nasional bruto y dalam \$45. Hasil untuk 26 tahun terakhir, x = 1, 2, ..., 26, seperti tampak berikut ini. Anggapan yang mana untuk model regresi linear tampak dilanggar secara serius oleh data? (*Catatan*: metode regresi biasanya tidak sesuai untuk data jenis ini).

Tahun	1	2	3	4	5	6	7	8	9	10
y	309,9	323,7	324,1	355,3	383,4	395,1	412,8	407	438	446,1
$\hat{y}$	283,1	301,9	320,6	339,4	358,2	376,9	395,9	414,5	433,2	452,0
Residu	26,8	21,8	3,5	15,9	25,2	18,2	17,1	-7,5	4,8	-5,9

11	12	13	14	15	16	17	18	19	20	21
452,5	447,3	475,9	487,7	497,2	529,8	551	581,1	617,8	658,1	675,2
470,8	489,5	508,3	527,1	545,8	564,6	583,4	602,1	620,9	639,7	658,4
-18,3	-42,2	-32,4	-39,4	-48,6	-34,8	-32,4	-21	-3,1	18,4	16,8

22	23	24	25	26
706,6	725,6	722,5	745,4	790,7
677,2	696,0	714,7	733,5	752,3
29,4	29,6	7,8	11,9	38,4

3) Dipunyai data berpasangan (x, y) sebagai berikut.

x	4,5	4,5	4,5	4,0	4,0	4,0	5,0	5,0	5,5	5,0
y	619	1049	1033	495	723	681	890	1522	987	1194

0,5	0,5	6,0	6,0	1,0	1,0	1,0
163	182	764	1373	978	466	549

a. Apakah model hubungan garis lurus memadai?
b. Dapatkah dicari model yang lebih baik?

4) Dipunyai observasi (x, y) sebagai berikut.

No. Obs	y	x	No. Obs	y	x	No. Obs	y	X
1	2,3	1,3	9	1,7	3,7	17	3,5	5,3
2	1,8	1,3	10	2,8	4,0	18	2,8	5,3
3	2,8	2,0	11	2,8	4,0	19	2,1	5,3
4	1,5	2,0	12	2,2	4,0	20	3,4	5,7
5	2,2	2,7	13	5,4	4,7	21	3,2	6,0
6	3,8	3,3	14	3,2	4,7	22	3,0	6,0
7	1,8	3,3	15	1,9	4,7	23	3,0	6,3
8	3,7	3,7	16	1,8	5,0	24	5,9	6,7

a. Taksirlah persamaan regresi garis lurus
b. Lakukan uji kekurangsesuaian

RANGKUMAN

Untuk menjaga terhadap penggunaan analisis regresi yang salah, kita harus memeriksa dengan cermat kesesuaian data dengan anggapan-anggapan model. Pemeriksaan residu, khususnya dengan gambar grafik sangat penting untuk menyidik pelanggaran terhadap anggapan yang mungkin terjadi dan juga dalam mengidentifikasi perubahan-perubahan yang sesuai dengan model awalnya.

Konsep uji ketidakcocokkan model regresi garis lurus hanya dapat dilakukan jika untuk beberapa tingkat variabel independen x terdapat (berkaitan) observasi replikasi dalam y. Konsep ini dapat diperluas untuk setiap model taksiran.

Teknik uji ketidakcocokkan ini dilakukan dengan pemisahan JKS menjadi dua komponen, pertama JK_{sm} dan kedua, JK_{ks} :

$$\text{JKS} = JK_{sm} + JK_{ks}$$

$$\text{db} = db_{sm} + db_{ks}$$

Jika komponen JK_{ks} cukup besar ini berarti kemungkinan adanya kekurangcocokkan model dengan data.

Uji kekurangcocokkan dilakukan dengan statistik penguji

$$F = \frac{JK_{ks}/db_{ks}}{JK_{sm}/db_{sm}} = \frac{KR_{ks}}{KR_{sm}}$$

TES FORMATIF 2

Pilihlah satu jawaban yang paling tepat!

I. Kelembaban adonan basah suatu produk dipandang mempunyai pengaruh terhadap kepadatan hasil akhir produk itu. Kelembaban adonan dikendalikan dan kepadatan produk akhir diukur.
Diperoleh data berikut.

x	y
4,7	3
5,0	3
5,2	4
5,2	5
5,9	10
4,7	2
5,9	9
5,2	3
5,3	7
5,9	6
5,6	6
5,0	4

1) Model regresi taksiran $\hat{y} = \hat{\alpha} + \hat{\beta}x$ kita hitung maka kita peroleh $\hat{\beta}$ adalah ....
   A. 3,2
   B. 5
   C. 6,61
   D. 7

2) Seperti soal nomor 1 juga kita peroleh $\hat{\alpha}$ adalah ....
   A. 14,45
   B. 19,12
   C. -18,17
   D. -21,33

3) Kita hitung JKS, sama dengan ...
   A. 16,16
   B. 17,17
   C. 18,18
   D. 19,19

4) Kita hitung interval kepercayaan 95% untuk β ....
   A. $4,1 < \beta < 5,9$
   B. $4,7 < \beta < 6,3$
   C. $4,9 < \beta < 6,6$
   D. $4,9 < \beta < 6,9$

5) Kita hitung JK_{sm} sama dengan ....
   A. 8,67
   B. 10,17
   C. 12,87
   D. 14,47

6) Kita hitung JK_{ks} sama dengan ....
   A. 6,50
   B. 7,50
   C. 8,50
   D. 9,50

7) Untuk uji ketidakcocokkan kita hitung statistik penguji F, kita peroleh ....
   A. 0,47
   B. 1,47
   C. 2,47
   D. 3,47

II. Dalam suatu analisis regresi garis lurus telah diperoleh persamaan regresi taksiran $\hat{y} = 1,436 + 0,338x$. Untuk menguji adanya ketidakcocokkan model telah dihitung JKS = 21,192 dengan db = 22, dan juga JK_{sm} = 12,470 dengan db = 11.

8) Maka, JK_{ks} dan *db*-nya sama dengan ....
   A. $JK_{ks} = 8{,}722$
      db = 11
   B. $JK_{ks} = 9{,}124$
      db = 10
   C. $JK_{ks} = 9{,}867$
      db = 9
   D. $JK_{ks} = 9{,}999$
      db = 8

9) KR_{sm} sama dengan ....
   A. 0,927
   B. 1,134
   C. 2,116
   D. 2,678

10) KR_{ks} sama dengan ....
   A. 0,444
   B. 0,576
   C. 0,793
   D. 0,897

11) Untuk menguji tidak ada ketidakcocokkan model kita hitung statistik penguji *F*, sama dengan ....
   A. 0,699
   B. 0,811
   C. 0,899
   D. 0,911

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.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{11} \times 100\%$$

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

I.

1) C
2) A
3) A
4) B
5) B
6) A

II.

7) D
8) C
9) A
10) B
11) B

Tes Formatif 2

I.

1) B
2) D
3) B
4) A
5) A
6) C
7) B

II.

8) A
9) B
10) C
11) A

Daftar Pustaka

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

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

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

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

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

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