

Tujuan Pembelajaran

Seselah proses pembelajaran, peserta didik diharapkan dapat:

- menggambar diagram pencar, kemudian menggunakan untuk menyelidiki dan menjelaskan hubungan antara dua variabel numerik;
- menggambar, menentukan, dan menginterpretasikan persamaan garis regresi linear;
- menerapkan interpolasi dan ekstrapolasi data berdasarkan persamaan garis regresi linear;
- menghitung koefisien korelasi Pearson dan koefisien determinasi, kemudian menginterpretasikannya dalam analisis regresi linear.

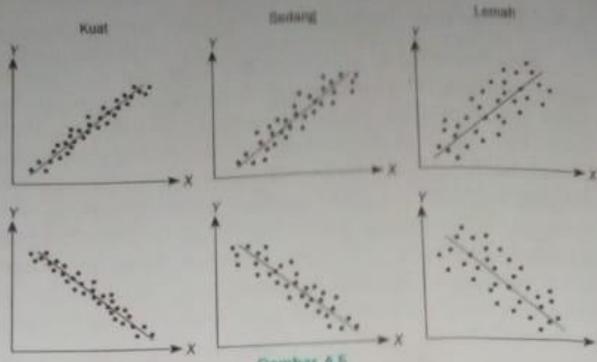
PROFIL PELAJAR PANCASILA

Kreatif, Bernalar Kritis,
Mandiri

Kata Kunci:
Diagram Pencar,
Ekstrapolasi, Koefisien
Determinasi, Interpolasi,
Koefisien Korelasi Pearson,
Rgresi Linear

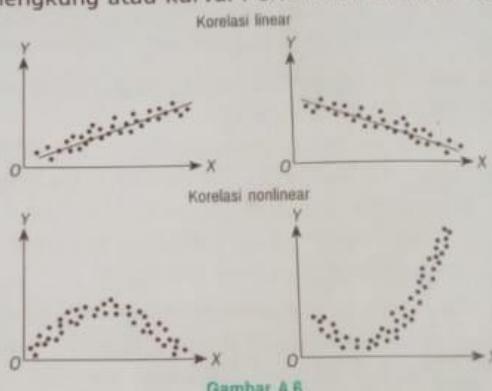
Bab 4

Statistika Regresi



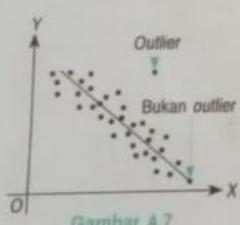
Gambar 4.5

- Perhatikan pola penyebaran titik-titik pada diagram pencar untuk menentukan hubungan linear atau nonlinear. Hubungan linear terjadi jika titik-titik pada diagram pencar membentuk pola yang menyerupai garis lurus. Hubungan nonlinear terjadi jika titik-titik pada diagram membentuk garis lengkung atau kurva. Perhatikan Gambar 4.6.



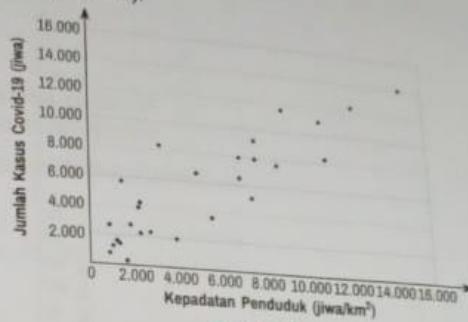
Gambar 4.6

- Carilah beberapa titik pada diagram pencar yang terletak jauh dari pola penyebaran. Titik-titik tersebut menunjukkan adanya **outlier**. **Outlier** atau penciran adalah data yang menyimpang terlalu jauh dari data lain dalam sekumpulan data. **Outlier** dapat muncul karena adanya kesalahan prosedur dalam memasukkan data. Perhatikan Gambar 4.7.



variabel yang memiliki korelasi, kita tidak dapat langsung menyimpulkan bahwa ada hubungan sebab-akibat keduanya. Ada variabel lain yang perlu dipertimbangkan sebelum menyimpulkan adanya hubungan sebab-akibat. Sebagai contoh, perhatikan hubungan antara tingkat kepadatan penduduk dan jumlah kasus Covid-19 di suatu daerah pada diagram pencar berikut (Gambar 4.8).

Kepadatan penduduk dan jumlah kasus Covid-19 memiliki **korelasi positif** yang artinya semakin padat penduduk, semakin banyak jumlah kasus Covid-19. Apakah ini berarti kepadatan penduduk adalah penyebab tingginya jumlah kasus Covid-19? Masih ada variabel lain yang perlu dipertimbangkan untuk memastikan hal tersebut, misalnya mobilitas penduduk.



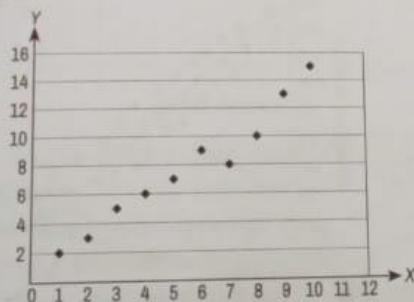
Gambar 4.8

Contoh Soal dan Pembahasan

1. Perhatikan himpunan pasangan data (x, y) berikut.
a. $\{(1, 2), (2, 3), (3, 5), (4, 6), (5, 7), (6, 9), (7, 8), (8, 10), (9, 13), (10, 15)\}$
b. $\{(1, 11), (2, 4), (3, 10), (4, 3), (5, 12), (6, 2), (7, 1), (8, 15), (9, 2), (10, 15)\}$
Gambarkan diagram pencar dari himpunan tersebut. Kemudian, selidiki hubungan antara x dan y .

Jawab:

- a. Diagram pencar:



Perhatikan sebaran titik-titik pada diagram.

- Oleh karena peningkatan nilai x diikuti dengan peningkatan nilai y , maka terdapat korelasi positif antara x dan y .
- Oleh karena titik-titik tersebut membentuk pola menyerupai garis lurus, maka terdapat korelasi linear antara variabel x dan y .

Jadi, x dan y memiliki korelasi linear positif.



Menentukan Faktor-Faktor yang Memengaruhi Penyebaran Covid-19 dengan Statistika Regresi

Sumber: www.shutterstock.com

Sebelum virus Covid-19 terdeteksi di Indonesia, Prof. Marc Lipsitch, seorang pakar epidemiologi dari Universitas Harvard, menyatakan bahwa virus Covid-19 sudah masuk ke Indonesia. Ia juga menyatakan bahwa setidaknya sudah ada 5 kasus Covid-19 di Indonesia tapi belum terdeteksi. Pernyataan tersebut berdasarkan penelitian yang ia lakukan. Ia menyelidiki hubungan statistik antara jumlah pengunjung ke suatu negara dengan jumlah kasus Covid-19 yang terdeteksi. Apa metode yang dapat digunakan untuk menyelidiki hubungan tersebut?

Setelah virus Covid-19 terdeteksi di Indonesia, penyebarannya terjadi sangat cepat. Hal tersebut dipengaruhi oleh banyak faktor. Selain karakteristik Covid-19 yang mudah menular, faktor lain yang memengaruhi penyebaran Covid-19 seperti kepadatan penduduk, mobilitas penduduk, interaksi sosial, kondisi ekonomi masyarakat, tingkat kemiskinan, dan masih banyak lagi. Bagaimana kita dapat memastikan faktor-faktor tersebut benar memengaruhi penyebaran Covid-19?

Untuk menjawab pertanyaan-pertanyaan di atas, Anda perlu mempelajari materi tentang diagram pencar, regresi linear, dan analisis regresi pada bab ini.

4.1 Diagram Pencar

Diagram pencar atau scatter plot dapat digunakan untuk menunjukkan hubungan antara dua variabel yang digambarkan dalam bentuk titik-titik (points) pada bidang koordinat Cartesius. Variabel pertama adalah variabel **variabel independen** (bebas) yang digambarkan pada sumbu X, menyebabkan berubahnya variabel yang memengaruhi atau merupakan variabel dependen (terikat) yang digambarkan pada sumbu Y. **Variabel dependen** adalah variabel yang lainnya dipengaruhi oleh variabel independen. Perhatikan Gambar 4.1.

Berdasarkan diagram pencar, terdapat tren hubungan linear, yakni makin besar nilai X, maka makin besar pula nilai Y. Jika titik-titik pada diagram pencar mendekati garis lurus, maka dapat dikatakan hubungan linear antara variabel X dan Y makin kuat. Untuk menganalisis kuat-lemahnya hubungan linear antara X dan Y digunakan analisis korelasi. Dalam melihat korelasi antara dua variabel, kita harus mengikuti langkah-langkah berikut.

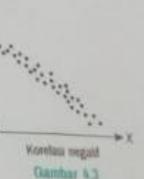
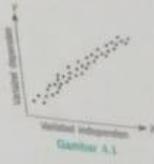
1. Perhatikan tren data pada diagram pencar untuk menentukan jenis korelasi.

a. Jika tren data menunjukkan kenaikan, maka terdapat hubungan positif atau **korelasi positif** antara dua variabel. Jika nilai variabel independen (X) meningkat, nilai variabel dependen (Y) juga meningkat. Perhatikan Gambar 4.2.

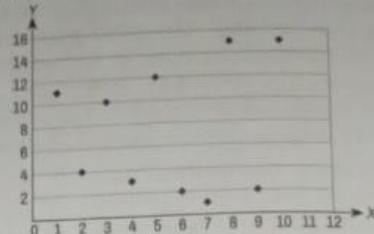
b. Jika tren data menunjukkan penurunan, maka terdapat hubungan negatif atau **korelasi negatif** antara dua variabel. Jika nilai variabel independen (X) meningkat, nilai variabel dependen (Y) menurun. Perhatikan Gambar 4.3.

c. Jika titik-titik pada diagram tampak menyebar secara acak (tidak terlihat tren), maka tidak ada korelasi antara dua variabel. Perhatikan Gambar 4.4.

2. Perhatikan penyebaran titik-titik pada diagram pencar untuk menentukan **kekuatan korelasi** antara dua variabel. Jika titik-titik makin rapat, maka hubungannya semakin kuat, sebaliknya jika titik-titik semakin menyebar, maka hubungannya semakin lemah. Perhatikan Gambar 4.5.



b. Diagram pencar:



Perhatikan sebaran titik-titik pada diagram. Titik-titik tersebut tersebar secara acak. Jadi, tidak ada korelasi antara x dan y .

2. Perhatikan diagram pencar berikut.

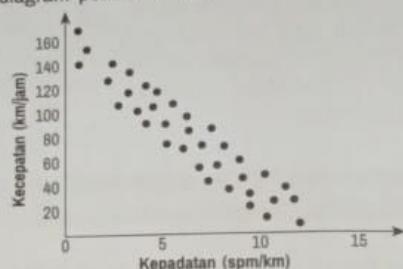
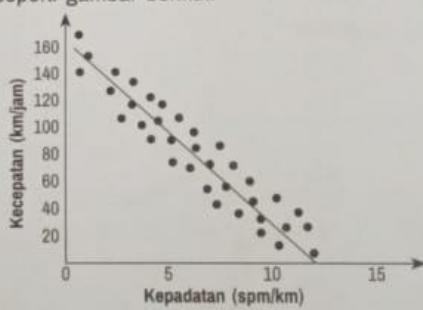


Diagram tersebut menunjukkan hubungan antara kecepatan kendaraan dan kepadatan lalu lintas. Jelaskan hubungan tersebut.

Jawab:

Titik-titik pada diagram pencar di atas membentuk pola menyerupai garis lurus seperti gambar berikut.



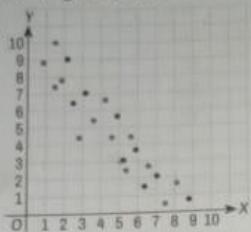
Oleh karena peningkatan kepadatan kendaraan diikuti dengan penurunan kecepatan kendaraan, maka diagram tersebut menunjukkan korelasi negatif antara kedua variabel tersebut. Jadi, terdapat korelasi linear negatif antara kepadatan kendaraan dan kecepatan kendaraan.

3. Seorang guru ingin mengetahui apakah durasi belajar siswa di luar jam sekolah mempengaruhi nilai rata-rata ujian sekolah. Ia melakukan survei kepada 20 siswa dan mencatat hasilnya dalam bentuk tabel seperti berikut.

Uji Pemahaman

A. Pilihlah satu jawaban yang benar.

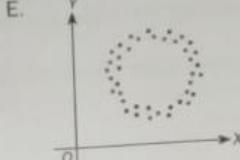
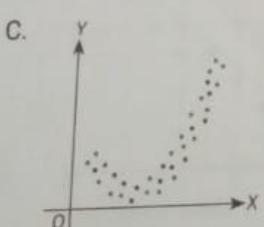
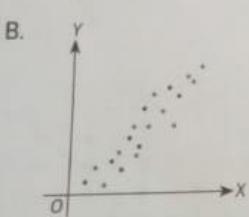
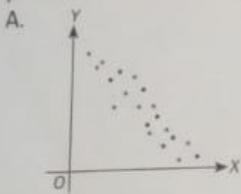
1. Perhatikan diagram pencar berikut.



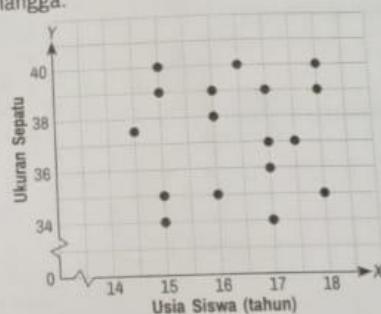
Hubungan antara nilai x dan y pada diagram tersebut adalah

- A. hubungan linear positif
- B. hubungan linear negatif
- C. hubungan nonlinear
- D. hubungan kuadratik
- E. tidak ada hubungan

2. Diagram pencar yang menunjukkan korelasi positif antara x dan y adalah



3. Berikut adalah diagram pencar yang menunjukkan hubungan antara usia dan ukuran sepatu beberapa siswa SMA Erlangga.



Pernyataan yang tepat sesuai dengan diagram tersebut adalah

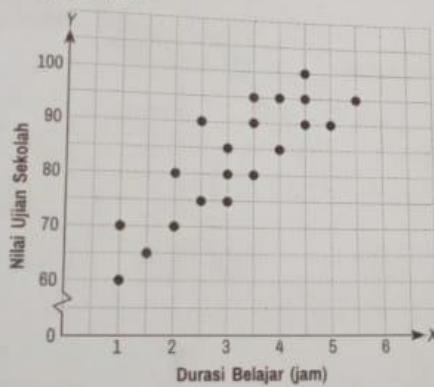
- A. Diagram tersebut menunjukkan korelasi positif.
- B. Diagram tersebut menunjukkan korelasi negatif.
- C. Diagram tersebut menunjukkan tidak ada korelasi.
- D. Diagram tersebut menunjukkan korelasi kuadratik.
- E. Diagram tersebut menunjukkan korelasi yang kuat.

Siswa	Durasi Belajar (jam)	Nilai Ujian Sekolah	Siswa	Durasi Belajar (jam)	Nilai Ujian Sekolah
1	1 jam	60	11	3,5 jam	80
2	1 jam	70	12	3,5 jam	90
3	1,5 jam	65	13	3,5 jam	95
4	2 jam	70	14	3,5 jam	85
5	2 jam	80	15	4 jam	95
6	2,5 jam	75	16	4 jam	90
7	2,5 jam	90	17	4,5 jam	95
8	3 jam	75	18	4,5 jam	90
9	3 jam	80	19	4,5 jam	95
10	3 jam	85	20	5 jam	100

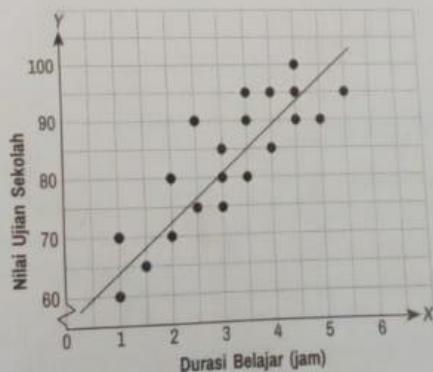
- Gambarkan diagram pencar yang menunjukkan hasil survei guru tersebut.
- Berdasarkan diagram pencar tersebut, jelaskan hubungan antara durasi belajar siswa di luar jam sekolah dengan nilai rata-rata ujian sekolah.

Jawab:

- Diagram pencar sesuai dengan data hasil survei pada tabel adalah sebagai berikut.



- Perhatikan sebaran titik-titik pada diagram pencar di samping. Titik-titik pada diagram tersebut membentuk pola menyerupai garis lurus. Oleh karena peningkatan durasi belajar diikuti dengan peningkatan nilai ujian sekolah, maka diagram tersebut menunjukkan korelasi positif antara kedua variabel tersebut.
Jadi, terdapat korelasi linear positif antara durasi belajar siswa di luar jam sekolah dan nilai ujian sekolah.



3. Perhatikan tabel data absensi dan nilai rata-rata rapor semester ganjil dari 20 siswa berikut ini.

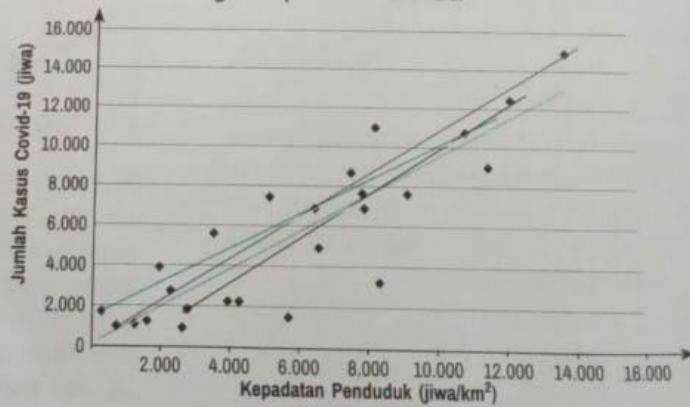
Siswa	Absensi (hari)	Nilai	Siswa	Absensi (hari)	Nilai
1	1	93	11	7	70
2	2	95	12	8	65
3	6	70	13	4	86
4	4	79	14	7	68
5	10	65	15	1	100
6	3	87	16	6	75
7	5	85	17	5	80
8	2	85	18	2	88
9	8	70	19	5	79
10	9	62	20	10	60

- a. Gambarlah diagram pencar berdasarkan data tersebut.
- b. Berdasarkan diagram pencar yang Anda gambar, jelaskan hubungan antara dua kelompok data tersebut.
4. **(NOTA)** Tika ingin menyelidiki hubungan antara indeks massa tubuh (IMS) dengan tekanan darah sistolik. Ia mengumpulkan data berat badan, tinggi badan, dan tekanan darah sistolik beberapa orang dalam bentuk tabel seperti berikut.
- | Responden | Berat Badan (kg) | Tinggi Badan (cm) | Tekanan Darah Sistolik (mmHg) |
|-----------|------------------|-------------------|-------------------------------|
| 1 | 72 | 168 | 110 |
| 2 | 57 | 160 | 100 |
| 3 | 80 | 170 | 110 |
| 4 | 105 | 165 | 130 |
| 5 | 88 | 165 | 110 |
| 6 | 73 | 152 | 110 |
| 7 | 40 | 151 | 85 |
| 8 | 59 | 161 | 110 |
| 9 | 59 | 173 | 100 |
| 10 | 58 | 166 | 110 |
| 11 | 73 | 167 | 110 |
| 12 | 102 | 168 | 140 |
| 13 | 82 | 157 | 120 |
| 14 | 62 | 157 | 130 |
| 15 | 86 | 165 | 120 |
- a. Hitunglah IMS setiap responden sesuai data pada tabel.
- b. Gambarlah diagram pencar yang menunjukkan hubungan antara IMS dan tekanan darah sistolik.
- c. Berdasarkan diagram pencar yang Anda gambar, bagaimana hubungan antara IMS dan tekanan darah sistolik? Jelaskan jawaban Anda.

4.2 Regresi Linear

4.2.1 Garis Regresi atau Garis Best-Fit

Perhatikan diagram pencar berikut.

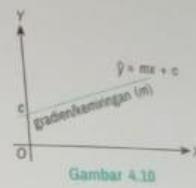


Gambar 4.9

Titik-titik pada diagram pencar tersebut membentuk pola yang menyerupai garis lurus. Artinya, ada korelasi antara kepadatan penduduk dan jumlah kasus Covid-19. Jika diagram pencar menunjukkan adanya korelasi antara dua variabel, kita dapat membuat garis-garis yang ditarik dari titik-titik pada diagram pencar tersebut. Di antara semua garis tersebut, hanya ada satu garis yang paling tepat untuk menggambarkan hubungan antara dua variabel. Garis tersebut disebut **garis regresi** atau **garis best-fit**. Persamaan garis regresi sederhana sering dituliskan dalam bentuk umum berikut ini.

$$\hat{y} = mx + c$$

di mana: \hat{y} = nilai variabel dependen yang diprediksi
 x = nilai variabel independen
 m = koefisien regresi (gradien garis regresi)
 c = konstanta (nilai \hat{y} jika $x = 0$)



Gambar 4.10

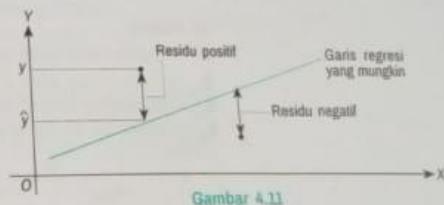
4.2.2 Metode Kuadrat Terkecil

Salah satu metode yang digunakan untuk menentukan garis regresi adalah metode kuadrat terkecil. Prinsip pemilihan garis regresi dengan metode kuadrat terkecil adalah meminimalkan jumlah kuadrat residu. **Residu** adalah jarak vertikal antara titik-titik data dan garis regresi. Perhatikan Gambar 4.11.

Berdasarkan gambar tersebut, nilai dari residu adalah sebagai berikut.

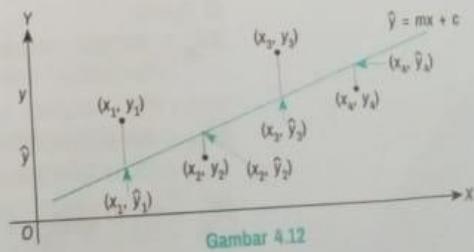
$$\text{Residu } (\varepsilon) = y - \hat{y}$$

di mana: y = nilai variabel dependen yang diamati
 \hat{y} = nilai variabel dependen yang diprediksi



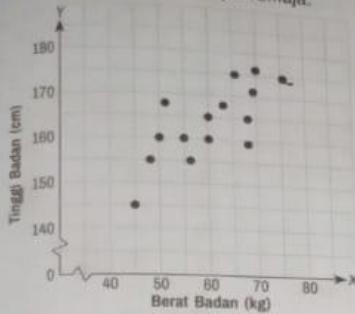
Gambar 4.11

Untuk mendapatkan garis regresi yang baik, kita harus meminimalkan jumlah dari residu. Oleh karena residu dapat bernilai positif dan negatif, jumlah dari residu mungkin saja bernilai nol. Padahal, kenyataannya ada residu yang jaraknya jauh dari garis regresi. Hal tersebut dapat dihindari dengan cara mengkuadratkan setiap residu.



Gambar 4.12

4. Berikut adalah diagram pencar yang menunjukkan hubungan antara berat badan dan tinggi badan beberapa remaja.



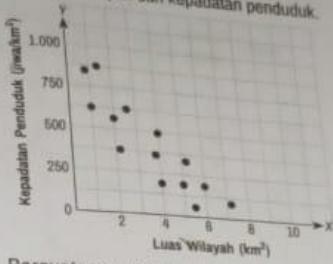
Pernyataan yang tepat sesuai dengan diagram tersebut adalah . . .

- A. Berat badan cenderung bertambah sesuai dengan pertambahan tinggi badan.
- B. Berat badan cenderung bertambah sesuai dengan pengurangan tinggi badan.
- C. Berat badan cenderung berkurang sesuai dengan pertambahan tinggi badan.
- D. Berat badan dan tinggi badan bervariasi sesuai dengan usia remaja.
- E. Hubungan antara berat badan dan tinggi badan tidak pasti.

B. Jawablah dengan jelas dan benar.

1. Gambarlah diagram pencar untuk himpunan data (x, y) berikut ini. Kemudian, selidikilah hubungan antara x dan y .
 - a. $\{(1, 8), (1, 9), (1, 10), (2, 7), (3, 7), (3, 6), (4, 5), (5, 4), (7, 1), (8, 0)\}$
 - b. $\{(1, 2), (2, 5), (3, 10), (4, 17), (5, 26), (6, 37), (7, 50), (8, 65)\}$
2. Andi sedang mencari informasi tentang dampak pemanasan global terhadap permukaan laut. Ia mendapatkan data perubahan permukaan laut dalam bentuk diagram pencar berikut ini.

5. Diagram pencar berikut menunjukkan hubungan antara luas wilayah dan kepadatan penduduk.



Pernyataan yang tepat sesuai dengan diagram tersebut adalah . . .

- A. Luas wilayah tidak mempengaruhi kepadatan penduduk.
- B. Kepadatan penduduk selalu mengalami peningkatan.
- C. Semakin luas suatu wilayah, kepadatan penduduk cenderung mengalami peningkatan.
- D. Semakin luas suatu wilayah, kepadatan penduduk cenderung mengalami penurunan.
- E. Semakin sempit suatu wilayah, kepadatan penduduk cenderung mengalami penurunan.

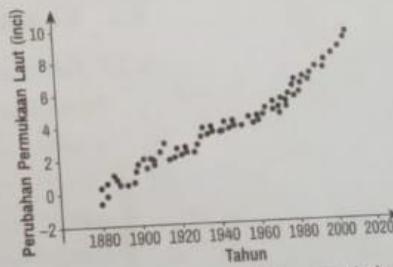


Diagram tersebut menunjukkan perubahan permukaan laut setiap tahun sejak tahun 1880. Bagaimana seharusnya Andi menafsirkan diagram pencar tersebut? Apakah diagram tersebut menunjukkan hubungan antara tahun dan perubahan permukaan laut? Jelaskan jawaban Anda.

Titik-titik pada diagram pencar tersebut membentuk pola yang menyerupai garis lurus. Artinya, ada korelasi antara kepadatan penduduk dan jumlah kasus Covid-19. Jika variabel, kita dapat membuat garis-garis yang ditarik dari titik-titik pada diagram pencar tersebut. Di antara semua garis tersebut, hanya ada satu garis yang paling tepat untuk menggambarkan hubungan antara dua variabel. Garis tersebut disebut **garis regresi** atau **garis best-fit**. Persamaan garis regresi sederhana sering dituliskan dalam bentuk umum berikut ini.

$$\hat{y} = mx + c$$

di mana: \hat{y} = nilai variabel dependen yang diprediksikan
 x = nilai variabel independen
 m = koefisien regresi (gradien garis regresi)
 c = konstanta (nilai \hat{y} jika $x = 0$)

4.2.2 Metode Kuadrat Terkecil

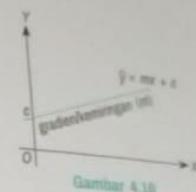
Salah satu metode yang digunakan untuk menentukan garis regresi adalah metode kuadrat terkecil. Prinsip pemilihan garis regresi dengan metode kuadrat terkecil adalah meminimalkan jumlah kuadrat residu. **Residu** adalah jarak vertikal antara titik-titik data dan garis regresi. Perhatikan Gambar 4.11.

Berdasarkan gambar tersebut, nilai dari residu adalah sebagai berikut.

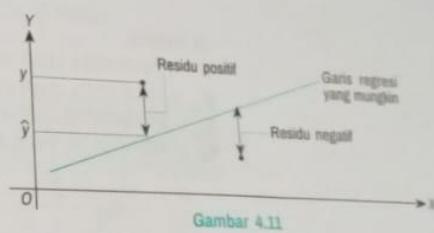
$$\text{Residu } (\varepsilon) = y - \hat{y}$$

di mana: y = nilai variabel dependen yang diamati
 \hat{y} = nilai variabel dependen yang diprediksikan

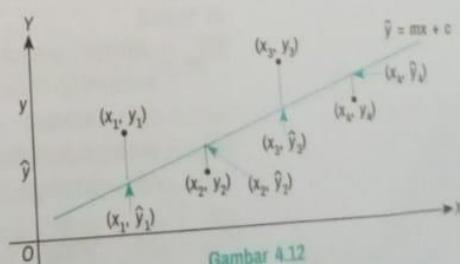
Untuk mendapatkan garis regresi yang baik, kita harus meminimalkan jumlah dari residu. Oleh karena residu dapat bernilai positif dan negatif, jumlah dari residu mungkin saja bernilai nol. Padahal, kenyataannya ada residu yang jaraknya jauh dari garis regresi. Hal tersebut dapat dihindari dengan cara mengkuadratkan setiap residu.



Gambar 4.10



Gambar 4.11



Gambar 4.12

3. Perhatikan tabel data absensi dan nilai rata-rata rapor semester ganjil dari 20 siswa berikut ini.

Siswa	Absensi (hari)	Nilai	Siswa	Absensi (hari)	Nilai
1	1	93	11	7	70
2	2	95	12	8	65
3	6	70	13	4	86
4	4	79	14	7	68
5	10	65	15	1	100
6	3	87	16	6	75
7	5	85	17	5	80
8	2	85	18	2	88
9	8	70	19	5	79
10	9	62	20	10	60

- a. Gambarlah diagram pencar berdasarkan data tersebut.
 b. Berdasarkan diagram pencar yang Anda gambar, jelaskan hubungan antara dua kelompok data tersebut.
4. **(NOTE)** Tika ingin menyelidiki hubungan antara indeks massa tubuh (IMS) dengan tekanan darah sistolik. Ia mengumpulkan data berat badan, tinggi badan, dan tekanan darah sistolik beberapa orang dalam bentuk tabel seperti berikut.

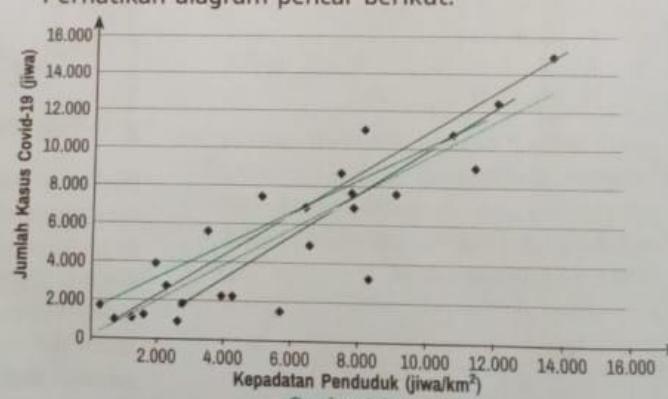
Responden	Berat Badan (kg)	Tinggi Badan (cm)	Tekanan Darah Sistolik (mmHg)
1	72	168	110
2	57	160	100
3	80	170	110
4	105	165	130
5	88	165	110
6	73	152	110
7	40	151	85
8	59	161	110
9	59	173	100
10	58	166	110
11	73	167	110
12	102	168	140
13	82	157	120
14	62	157	130
15	86	185	120

- a. Hitunglah IMS setiap responden sesuai data pada tabel.
 b. Gambarlah diagram pencar yang menunjukkan hubungan antara IMS dan tekanan darah sistolik.
 c. Berdasarkan diagram pencar yang Anda gambar, bagaimana hubungan antara IMS dan tekanan darah sistolik? Jelaskan jawaban Anda.

4.2 Regresi Linear

4.2.1 Garis Regresi atau Garis Best-Fit

Perhatikan diagram pencar berikut.



Gambar 4.9

Contoh Soal dan Pembahasan

1. Dua orang siswa menyelidiki hubungan antara tinggi badan (variabel independen) dan berat badan (variabel dependen) siswa di kelasnya. Data yang diperoleh disajikan dalam tabel berikut.

Siswa	1	2	3	4	5	6	7	8	9	10
Berat Badan (kg)	40	45	36	49	50	34	42	54	64	68
Tinggi Badan (cm)	154	150	142	160	159	150	145	160	159	165

Kedua siswa tersebut mendapatkan persamaan garis regresi yang berbeda berikut ini.

$$\text{Siswa 1: } \hat{y}_1 = 1,2x - 143$$

$$\text{Siswa 2: } \hat{y}_2 = 1,2x - 144$$

Siapa yang mendapatkan persamaan garis regresi yang lebih tepat untuk menunjukkan hubungan tinggi badan dan berat badan tersebut?

Jawab:

Diketahui: variabel independen (x) adalah tinggi badan
variabel dependen (y) adalah berat badan

Perhatikan tabel di samping.

Nilai Variabel		Siswa 1			Siswa 2		
x	y	\hat{y}_1	$y - \hat{y}_1$	$(y - \hat{y}_1)^2$	\hat{y}_2	$y - \hat{y}_2$	$(y - \hat{y}_2)^2$
154	40	41,8	-1,8	3,24	40,8	-0,8	0,64
150	45	37	8	64	36	9	81
142	36	27,4	8,6	73,96	26,4	9,6	92,16
160	49	49	0	0	48	1	1
159	50	47,8	2,2	4,84	46,8	3,2	10,24
150	34	37	-3	9	36	-2	4
145	42	31	11	121	30	12	144
160	54	49	5	25	48	6	36
159	64	47,8	16,2	262,44	46,8	17,2	295,84
165	68	55	13	169	54	14	196
		$\sum(y - \hat{y}_1)^2$		732,48	$\sum(y - \hat{y}_2)^2$		860,88

Persamaan garis regresi yang mempunyai jumlah kuadrat residu lebih kecil adalah persamaan garis yang lebih tepat. Jumlah kuadrat residu untuk persamaan garis regresi siswa 1 lebih kecil daripada jumlah kuadrat residu untuk persamaan garis regresi siswa 2. Jadi, siswa 1 mendapatkan persamaan garis regresi yang lebih tepat untuk menunjukkan hubungan tinggi badan dan berat badan siswa di kelasnya.

2. Perhatikan himpunan pasangan data (x, y) berikut.

$$\{(1, 2), (2, 3), (3, 5), (4, 6)\}$$

Tentukan persamaan garis regresi yang menunjukkan hubungan antara x dan y .

Jawab:

Perhatikan tabel berikut.

x	y	xy	x^2
1	2	2	1
2	3	6	4
3	5	15	9
4	6	24	16
$\sum x = 10$	$\sum y = 16$	$\sum xy = 47$	$\sum x^2 = 30$

Soal
Pendalaman

Perhatikan nilai tryout (X) dan nilai akhir (Y) dari 8 siswa dalam Ujian Sekolah mata Pelajaran Matematika berikut.

X	77	50	71	72	81	94	99	87
Y	82	66	78	34	47	85	99	68

Persamaan garis regresi yang tepat adalah

- A. $Y = 6,500 + 0,633X$
- B. $Y = 21,550 + 0,633X$
- C. $Y = 69,875 + 0,633X$
- D. $Y = 76,375 + 0,633X$

$$\bar{x} = \frac{\sum x}{n} = \frac{10}{4} = 2,5$$

$$\bar{y} = \frac{\sum y}{n} = \frac{16}{4} = 4$$

$$s_{xy} = \frac{\sum xy}{n} - \bar{x}\bar{y}$$
$$= \frac{47}{4} - (2,5)(4)$$
$$= 11,75 - 10$$
$$= 1,75$$

$$s_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$
$$\Leftrightarrow s_x^2 = \frac{\sum x^2}{n} - \bar{x}^2$$
$$= \frac{30}{4} - (2,5)^2$$
$$= 7,5 - 6,25$$
$$= 1,25$$

Persamaan garis regresi:

$$\hat{y} - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x})$$
$$\Leftrightarrow \hat{y} - 4 = \frac{1,75}{1,25}(x - 2,5)$$
$$\Leftrightarrow \hat{y} - 4 = 1,4(x - 2,5)$$
$$\Leftrightarrow \hat{y} - 4 = 1,4x - 3,5$$
$$\Leftrightarrow \hat{y} = 1,4x - 3,5 + 4$$
$$\Leftrightarrow \hat{y} = 1,4x + 0,5$$

Jadi, persamaan garis regresi yang dapat menunjukkan hubungan antara x dan y adalah $\hat{y} = 1,4x + 0,5$.

$$\epsilon_1^2 = (y_1 - \hat{y}_1)^2$$

$$\epsilon_2^2 = (y_2 - \hat{y}_2)^2$$

$$\epsilon_3^2 = (y_3 - \hat{y}_3)^2$$

$$\epsilon_4^2 = (y_4 - \hat{y}_4)^2$$

$$\epsilon_n^2 = (y_n - \hat{y}_n)^2$$

Jumlah kuadrat dari residu adalah sebagai berikut.

$$SS = \epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + \epsilon_4^2 + \dots + \epsilon_n^2$$

$$= \sum_{i=1}^n \epsilon_i^2$$

$$= \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^n (y_i - mx_i - c)^2$$

Meminimalkan jumlah kuadrat residu tersebut dapat dilakukan dengan diferensiasi. Akan tetapi, proses diferensiasi ini tidak akan dijabarkan karena memerlukan ilmu Kalkulus Lanjutan yang akan dipelajari di tingkat universitas. Proses diferensiasi ini menghasilkan estimasi untuk nilai m dan c pada persamaan garis regresi, sehingga diperoleh rumus persamaan garis regresi sebagai berikut.

$$\hat{y} - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x})$$

di mana:

\bar{x} = rata-rata dari nilai x

\bar{y} = rata-rata dari nilai y

s_{xy} = kovariansi dari x dan y

s_x = standar deviasi dari x

$$s_{xy} = \frac{\sum(x - \bar{x})(y - \bar{y})}{n} \text{ atau } s_{xy} = \frac{\sum xy}{n} - \bar{x}\bar{y}$$

$$s_x = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} \text{ atau } s_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

Dengan menggunakan rumus jumlah kuadrat selisih, diperoleh:

$$\hat{y} - \bar{y} = \frac{SS_{xy}}{SS_{xx}}(x - \bar{x})$$

di mana:

SS_{xy} = Jumlah perkalian antara selisih variabel independen x terhadap rata-ratanya dan variabel dependen y terhadap rata-ratanya

SS_{xx} = Jumlah kuadrat selisih variabel independen x terhadap rata-ratanya

$$SS_{xy} = \sum(x - \bar{x})(y - \bar{y}) = \sum xy - \frac{(\sum x)(\sum y)}{n} \text{ atau } SS_{xy} = \sum xy - n\bar{x}\bar{y}$$

$$SS_{xx} = \sum(x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n} \text{ atau } SS_{xx} = \sum x^2 - n\bar{x}^2$$

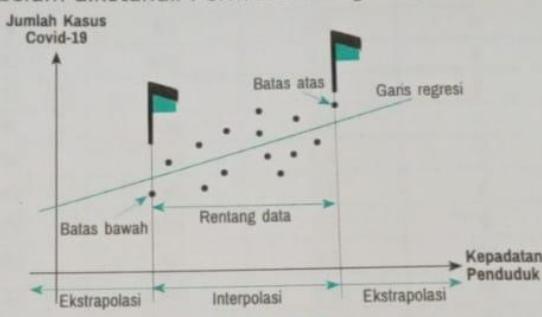
Persamaan garis regresi:

$$\begin{aligned}\hat{y} - \bar{y} &= \frac{SS_{xy}}{SS_{xx}}(x - \bar{x}) \\ \Leftrightarrow \hat{y} - 7,76 &= \frac{73,92}{143}(x - 6,5) \\ \Leftrightarrow \hat{y} - 7,76 &= 0,52(x - 6,5) \\ \Leftrightarrow \hat{y} - 7,76 &= 0,52x - 3,38 \\ \Leftrightarrow \hat{y} &= 0,52x - 3,38 + 7,76 \\ \Leftrightarrow \hat{y} &= 0,52x + 4,38\end{aligned}$$

Jadi, persamaan garis regresi yang menunjukkan hubungan antara umur dan rata-rata berat badan adalah $\hat{y} = 0,52x + 4,38$.

4.2.3 Interpolasi dan Ekstrapolasi

Garis regresi dapat digunakan untuk memprediksi data yang belum diketahui. Perhatikan diagram pencar berikut.



Gambar 4.13

Garis regresi pada diagram pencar tersebut menunjukkan hubungan antara kepadatan penduduk dan jumlah kasus Covid-19. Kita dapat menggunakan persamaan garis regresi tersebut untuk memprediksi jumlah kasus Covid-19 di suatu daerah berdasarkan kepadatan penduduk di daerah tersebut. Penggunaan garis regresi untuk memprediksi nilai yang berada di dalam rentang data disebut **interpolasi**, sedangkan penggunaan garis regresi untuk memprediksi nilai yang berada di luar rentang data disebut **ekstrapolasi**.

Keakuratan interpolasi bergantung pada seberapa linear data aslinya. Hal tersebut dapat diukur dengan menentukan koefisien korelasi dan memastikan data tersebar secara acak di sekitar garis regresi. Keakuratan ekstrapolasi tidak hanya bergantung pada seberapa linear data aslinya, tetapi juga diperlukan asumsi bahwa hubungan linear tersebut berlaku meskipun di luar rentang data. Validitas asumsi tersebut sangat tergantung pada situasi yang sedang diselidiki.

Contoh Soal dan Pembahasan

1. Tabel berikut menunjukkan data harga eceran beras di suatu daerah.
- | Tahun | 2017 | 2018 | 2019 | 2020 | 2021 |
|--|--------|--------|--------|--------|--------|
| Rata-Rata Harga Eceran Beras per Kilogram (rupiah) | 11.200 | 12.000 | 12.800 | 13.200 | 13.500 |
- Gambarlah diagram pencar dari data tersebut.
 - Tentukan persamaan garis regresi dari data tersebut.
 - Hitunglah prediksi rata-rata harga eceran beras per kilogram pada tahun 2022.

Jawab:

- a. Misalkan $t = 1, 2, 3, \dots$ menunjukkan waktu mulai dari tahun 2017

P = harga eceran beras per kilogram dalam ribu rupiah

Data pada soal dapat ditulis dalam bentuk seperti berikut.

t	1	2	3	4	5
P	11,2	12	12,8	13,2	13,5

Diagram pencar dari data tersebut adalah sebagai berikut.

- b. Diketahui:

Variabel independen (x) adalah waktu (t).

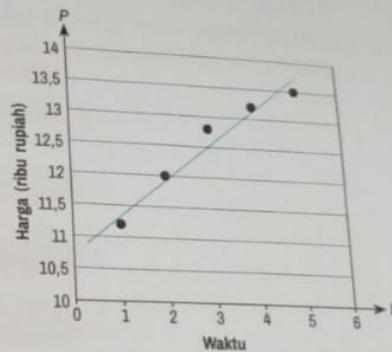
Variabel dependen (y) adalah harga eceran beras per kilogram dalam ribuan rupiah (P).

Perhatikan tabel berikut.

x	y	xy	x^2
1	11,2	11,2	1
2	12	24	4
3	12,8	38,4	9
4	13,2	52,8	16
5	13,5	67,5	25
$\sum x = 15$	$\sum y = 62,7$	$\sum xy = 193,9$	$\sum x^2 = 55$

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{\sum y}{n} = \frac{62,7}{5} = 12,54$$



3. Tabel berikut menunjukkan data rata-rata berat badan bayi umur 1-12 bulan.

Umur (bulan)	1	2	3	4	5	6	7	8	9	10	11	12
Rata-Rata Berat Badan (kg)	4,3	5,2	6,0	6,7	7,3	7,8	8,3	8,8	9,2	9,5	9,9	10,1

Tentukan persamaan garis regresi yang menunjukkan hubungan antara umur dan rata-rata berat badan.

Jawab:

Diketahui: variabel independen (x) adalah umur
variabel dependen (y) adalah rata-rata berat badan

Perhatikan tabel berikut.

x	y	xy	x^2
1	4,3	4,3	1
2	5,2	10,4	4
3	6,0	18	9
4	6,7	26,8	16
5	7,3	36,5	25
6	7,8	46,8	36
7	8,3	58,1	49
8	8,8	70,4	64
9	9,2	82,8	81
10	9,5	95	100
11	9,9	108,9	121
12	10,1	121,2	144
$\sum x = 78$	$\sum y = 93,1$	$\sum xy = 679,2$	$\sum x^2 = 650$

$$\bar{x} = \frac{\sum x}{n} = \frac{78}{12} = 6,5$$

$$\bar{y} = \frac{\sum y}{n} = \frac{93,1}{12} = 7,76$$

Menggunakan rumus jumlah kuadrat selisih, diperoleh:

$$\begin{aligned} SS_{xy} &= \sum xy - n\bar{x}\bar{y} \\ &= 679,2 - 12(6,5)(7,76) \\ &= 679,2 - 605,28 \\ &= 73,92 \end{aligned}$$

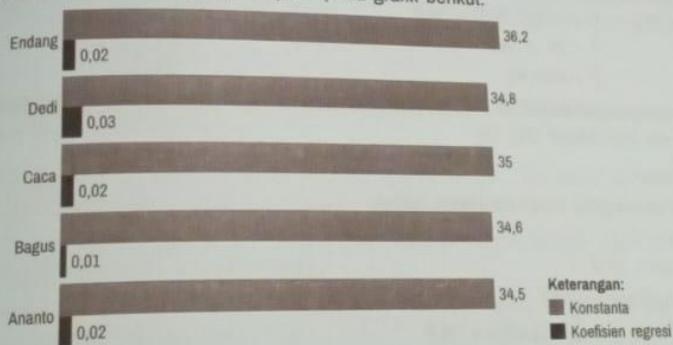
$$\begin{aligned} SS_{xx} &= \sum x^2 - n\bar{x}^2 \\ &= 650 - 12(6,5)^2 \\ &= 650 - 507 \\ &= 143 \end{aligned}$$



Sumber: www.shutterstock.com

Ketika kalori dalam tubuh seseorang lebih sedikit daripada kalori yang dibutuhkan untuk tiga komponen tersebut, maka tubuh orang tersebut dalam kondisi defisit kalori. Kondisi defisit kalori yang dipertahankan dalam jangka waktu panjang dapat menghasilkan penurunan berat badan.

Untuk mengetahui pengaruh kalori terhadap berat badan, beberapa mahasiswa mengadakan penelitian terhadap 30 orang dengan menggunakan regresi linear sederhana. Hasil dari penelitian mereka berbeda-beda seperti yang disajikan pada grafik berikut.



Sumber: <https://food.detik.com/info-sehat/d-5665440/diet-defisit-kalori-untuk-kurangi-berat-badan-ini-aturan-dan-menunya>

2. Berdasarkan stimulus tersebut, pasangkan pernyataan berikut dengan jawaban yang benar.

Pernyataan	Jawaban	
Perkiraan berat badan seseorang yang mengonsumsi 1.300 kalori per hari berdasarkan penelitian Ananto.	<input type="radio"/>	<input type="radio"/> 73,8 kg
Perkiraan berat badan seseorang yang mengonsumsi 1.400 kalori per hari berdasarkan penelitian Bagus.	<input type="radio"/>	<input type="radio"/> 66,2 kg
Perkiraan berat badan seseorang yang mengonsumsi 1.400 kalori per hari berdasarkan penelitian Caca.	<input type="radio"/>	<input type="radio"/> 66 kg
Perkiraan berat badan seseorang yang mengonsumsi 1.300 kalori per hari berdasarkan penelitian Dedi.	<input type="radio"/>	<input type="radio"/> 63 kg
Perkiraan berat badan seseorang yang mengonsumsi 1.500 kalori per hari berdasarkan penelitian Endang.	<input type="radio"/>	<input type="radio"/> 60,5 kg
	<input type="radio"/>	<input type="radio"/> 48,6 kg

Pembahasan:

Misalkan bentuk persamaan regresi linear adalah $\hat{y} = mx + c$, dengan m adalah koefisien regresi dan c adalah konstanta.

Berdasarkan grafik pada stimulus, diperoleh nilai m dan c berikut.

Nama Mahasiswa	Koefisien Regresi (m)	Konstanta (c)
Endang	0,02	36,2
Dedi	0,03	34,8
Caca	0,02	35
Bagus	0,01	34,6
Ananto	0,02	34,5

Pernyataan 1

Persamaan regresi linear penelitian Ananto

$$\hat{y} = mx + c$$

$$\hat{y} = 0,02x + 34,5$$

Banyaknya kalori 1.300

$$x = 1.300 \Rightarrow \hat{y} = 0,02(1.300) + 34,5$$

$$\hat{y} = 26 + 34,5$$

$$\hat{y} = 60,5 \text{ kg}$$

Jadi, berdasarkan penelitian Ananto, perkiraan berat badan seseorang yang mengonsumsi 1.300 kalori per hari adalah 60,5 kg.

Pernyataan 2

Persamaan regresi linear penelitian Bagus

$$\hat{y} = mx + c$$

$$\hat{y} = 0,01x + 34,6$$

Banyaknya kalori 1.400

$$x = 1.400 \Rightarrow \hat{y} = 0,01(1.400) + 34,6$$

$$\hat{y} = 14 + 34,6$$

$$\hat{y} = 48,6 \text{ kg}$$

Jadi, berdasarkan penelitian Bagus, perkiraan berat badan seseorang jika mengonsumsi 1.400 kalori per hari adalah 48,6 kg.

Pernyataan 3

Persamaan regresi linear penelitian Caca

$$\hat{y} = mx + c$$

$$\hat{y} = 0,02x + 35$$

Banyaknya kalori 1.400

$$x = 1.400 \Rightarrow \hat{y} = 0,02(1.400) + 35$$

$$\hat{y} = 28 + 35$$

$$\hat{y} = 63 \text{ kg}$$

Jadi, berdasarkan penelitian Caca, perkiraan berat badan seseorang yang mengonsumsi 1.400 kalori per hari adalah 63 kg.

$$s_{xy} = \frac{\sum xy}{n} - \bar{x}\bar{y}$$

$$= \frac{193,9}{5} - (3)(12,54)$$

$$= 38,78 - 37,62 = 1,16$$

$$s_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

$$\Leftrightarrow s_x^2 = \frac{\sum x^2}{n} - \bar{x}^2$$

$$= \frac{55}{5} - 3^2$$

$$= 11 - 9 = 2$$

Persamaan garis regresi:

$$\hat{y} - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x})$$

$$\Leftrightarrow \hat{y} - 12,54 = \frac{1,16}{2}(x - 3)$$

$$\Leftrightarrow \hat{y} - 12,54 = 0,58(x - 3)$$

$$\Leftrightarrow \hat{y} - 12,54 = 0,58x - 1,74$$

$$\Leftrightarrow \hat{y} = 0,58x - 1,74 + 12,54$$

$$\Leftrightarrow \hat{y} = 0,58x + 10,8$$

$$\Leftrightarrow P = 0,58t + 10,8$$

Jadi, persamaan garis regresi dari data tersebut adalah $P = 0,58t + 10,8$.

- c. Oleh karena $t = 1$ menunjukkan tahun 2017, maka nilai yang menunjukkan tahun 2022 adalah $t = 6$.

$$P = 0,58t + 10,8$$

$$\Leftrightarrow P = 0,58(6) + 10,8$$

$$\Leftrightarrow P = 3,48 + 10,8$$

$$\Leftrightarrow P = 14,28$$

Jadi, prediksi rata-rata harga eceran beras per kilogram pada tahun 2022 adalah Rp14.280,00.



Perhatikan stimulus berikut untuk menjawab soal nomor 2–4.

Kalori Makanan dan Berat Badan

Kalori adalah jumlah energi yang didapat dari konsumsi makanan dan minuman. Kalori yang dibakar setiap hari disebut sebagai "pengeluaran kalori" mencakup 3 komponen berikut.

- Resting energy expenditure (REE)*, mengacu pada kalori yang digunakan tubuh saat istirahat untuk menjaga fungsi-fungsi tubuh tetap normal.
- Efek termis dari makanan, melibatkan kalori yang dikeluarkan tubuh untuk mencerna, menyerap, dan memetabolisme makanan.
- Pengeluaran energi aktivitas, menunjukkan kalori yang kamu keluarkan selama olahraga dan aktivitas lain.

- Menggunakan persamaan regresi linear penelitian Bagus
$$\hat{y} = 0,01x + 34,6$$
$$\Leftrightarrow \hat{y} = 0,01(1.250) + 34,6$$
$$\Leftrightarrow \hat{y} = 12,5 + 34,6$$
$$\Leftrightarrow \hat{y} = 47,1 \text{ kg}$$
- Menggunakan persamaan regresi linear penelitian Caca
$$\hat{y} = 0,02x + 35$$
$$\Leftrightarrow \hat{y} = 0,02(1.250) + 35$$
$$\Leftrightarrow \hat{y} = 25 + 35$$
$$\Leftrightarrow \hat{y} = 60 \text{ kg}$$
- Menggunakan persamaan regresi linear penelitian Dedi
$$\hat{y} = 0,03x + 34,8$$
$$\Leftrightarrow \hat{y} = 0,03(1.250) + 34,8$$
$$\Leftrightarrow \hat{y} = 37,5 + 34,8$$
$$\Leftrightarrow \hat{y} = 72,3 \text{ kg}$$
- Menggunakan persamaan regresi linear penelitian Endang
$$\hat{y} = 0,02x + 36,2$$
$$\Leftrightarrow \hat{y} = 0,02(1.250) + 36,2$$
$$\Leftrightarrow \hat{y} = 25 + 36,2$$
$$\Leftrightarrow \hat{y} = 61,2 \text{ kg}$$

Pernyataan 1 (Benar)

Perkiraan berat badan Ardi menggunakan hasil penelitian Ananto (59,5 kg) lebih besar daripada menggunakan hasil penelitian Bagus (47,1 kg).

Pernyataan 2 (Benar)

Perkiraan berat badan Ardi menggunakan hasil penelitian Bagus (47,1 kg) lebih kecil daripada menggunakan hasil penelitian Dedi (72,3 kg).

Pernyataan 3 (Salah)

Perkiraan berat badan Ardi menggunakan hasil penelitian Caca (60 kg) lebih kecil daripada menggunakan hasil penelitian Dedi (72,3 kg).

Pernyataan 4 (Benar)

Perkiraan berat badan Ardi menggunakan hasil penelitian Bagus (47,1 kg) lebih kecil daripada menggunakan hasil penelitian Endang (61,2 kg).

Pernyataan 5 (Salah)

Perkiraan berat badan Ardi menggunakan hasil penelitian Ananto (59,5 kg) lebih kecil daripada menggunakan hasil penelitian Endang (61,2 kg).

4. Perhatikan kembali stimulus tersebut. Jika seseorang mengonsumsi 1.450 kalori per hari, selisih perkiraan berat badan jika menggunakan hasil penelitian Dedi dan Endang adalah
- A. 13 kg D. 14,1 kg
B. 13,1 kg E. 14,2 kg
C. 13,2 kg

Pernyataan 4
Persamaan regresi linear penelitian Dedi

$$\hat{y} = mx + c$$

$$\hat{y} = 0,03x + 34,8$$

Banyaknya kalori 1.300

$$x = 1.300 \Rightarrow \hat{y} = 0,03(1.300) + 34,8$$

$$\hat{y} = 39 + 34,8$$

$$\hat{y} = 73,8 \text{ kg}$$

Jadi, berdasarkan penelitian Dedi, perkiraan berat badan seseorang yang mengonsumsi 1.300 kalori per hari adalah 73,8 kg.

Pernyataan 5

Persamaan regresi linear penelitian Endang

$$\hat{y} = mx + c$$

$$\hat{y} = 0,02x + 36,2$$

Banyaknya kalori 1.500

$$x = 1.500 \Rightarrow \hat{y} = 0,02(1.500) + 36,2$$

$$\hat{y} = 30 + 36,2$$

$$\hat{y} = 66,2 \text{ kg}$$

Jadi, berdasarkan penelitian Endang, perkiraan berat badan seseorang yang mengonsumsi 1.500 kalori per hari adalah 66,2 kg.

3. Perhatikan kembali stimulus tersebut. Jika Ardi mengonsumsi 1.250 kalori per hari, beri tanda centang (✓) pada kotak di depan pernyataan untuk jawaban-jawaban yang benar.
- Perkiraan berat badan Ardi menggunakan hasil penelitian Ananto lebih besar daripada menggunakan hasil penelitian Bagus.
 - Perkiraan berat badan Ardi menggunakan hasil penelitian Bagus lebih kecil daripada menggunakan hasil penelitian Dedi.
 - Perkiraan berat badan Ardi menggunakan hasil penelitian Caca lebih besar daripada menggunakan hasil penelitian Dedi.
 - Perkiraan berat badan Ardi menggunakan hasil penelitian Bagus lebih kecil daripada menggunakan hasil penelitian Endang.
 - Perkiraan berat badan Ardi menggunakan hasil penelitian Ananto lebih besar daripada menggunakan hasil penelitian Endang.

Pembahasan:

Perkiraan berat badan Ardi jika mengonsumsi 1.250 kalori per hari:

Perkiraan berat badan Ardi jika mengonsumsi 1.250 kalori per hari:

Menggunakan persamaan regresi linear penelitian Ananto

$$\hat{y} = 0,02x + 34,5$$

$$\Leftrightarrow \hat{y} = 0,02(1.250) + 34,5$$

$$\Leftrightarrow \hat{y} = 25 + 34,5$$

$$\Leftrightarrow \hat{y} = 59,5 \text{ kg}$$

5. Seorang siswa menyelidiki hubungan antara usia sebuah mobil bekas dengan harganya. Data yang ia dapatkan adalah sebagai berikut.

Umur Mobil (tahun)	1	2	3	4	5
Harga Mobil (juta rupiah)	210	180	170	140	110

B. Jawablah dengan jelas dan benar.

1. Tentukan persamaan garis regresi jika diketahui sebagai berikut.

- $\bar{x} = 3,6$; $\bar{y} = 2,95$; $s_{xy} = -2,14$; dan $s_x^2 = 1,16$
- $\bar{x} = 43,2$; $\bar{y} = 91,8$; $SS_{xy} = 24,56$; dan $SS_{xx} = 17,54$

2. Tentukan persamaan garis regresi yang menunjukkan hubungan antara x dan y berikut ini.

a.	<table border="1"> <tr> <td>x</td><td>1</td><td>2</td><td>4</td><td>6</td><td>7</td></tr> <tr> <td>y</td><td>7</td><td>8</td><td>13</td><td>17</td><td>18</td></tr> </table>	x	1	2	4	6	7	y	7	8	13	17	18
x	1	2	4	6	7								
y	7	8	13	17	18								

b.	<table border="1"> <tr> <td>x</td><td>3</td><td>5</td><td>6</td><td>8</td><td>11</td></tr> <tr> <td>y</td><td>10</td><td>12</td><td>15</td><td>21</td><td>28</td></tr> </table>	x	3	5	6	8	11	y	10	12	15	21	28
x	3	5	6	8	11								
y	10	12	15	21	28								

3. Seorang siswa ingin menyelidiki hubungan antara kepadatan penduduk dengan jumlah kasus Covid-19. Data yang ia peroleh disajikan dalam tabel berikut.

Perkiraan harga mobil bekas tersebut jika sudah berusia 6 tahun adalah . . .

- Rp100.000.000,00
- Rp97.000.000,00
- Rp94.000.000,00
- Rp90.000.000,00
- Rp89.000.000,00

Kepadatan Penduduk (jiwa per km^2)	Jumlah Kasus Covid-19
2.200	4.300
7.500	7.900
7.600	9.200
2.100	3.900
900	1.200
2.600	2.300
4.800	6.500
8.900	11.500
2.200	2.000
1.000	1.200

- Gambarlah diagram pencar dari data tersebut, kemudian jelaskan hubungan antara kepadatan penduduk dengan jumlah kasus Covid-19 berdasarkan sebaran titik-titik data.
- Tentukan persamaan garis regresi dari data tersebut.
- Tentukan prediksi jumlah kasus Covid-19 pada daerah dengan kepadatan penduduk 1.500 jiwa per km^2 .
- Tentukan prediksi jumlah kasus Covid-19 pada daerah dengan kepadatan penduduk 9.000 jiwa per km^2 .

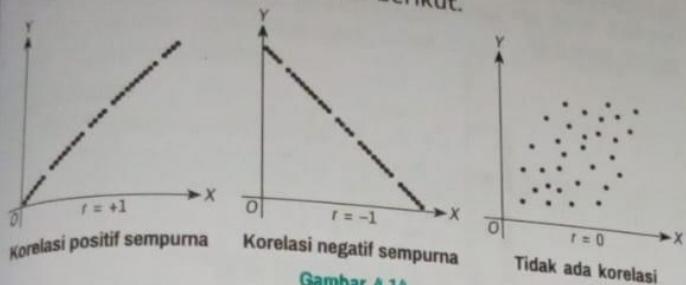
4.3 Analisis Korelasi

4.3.1 Koefisien Korelasi

Koefisien korelasi merupakan nilai yang menunjukkan kekuatan hubungan antara dua variabel. Koefisien korelasi juga memberikan informasi mengenai arah hubungan antara dua variabel tersebut. Nilai koefisien korelasi (r) berkisar antara -1 dan 1 , ditulis $-1 \leq r \leq +1$. Kekuatan hubungan diketahui dari nilai angka, sedangkan arah diketahui dari tanda positif (+) dan negatif (-).

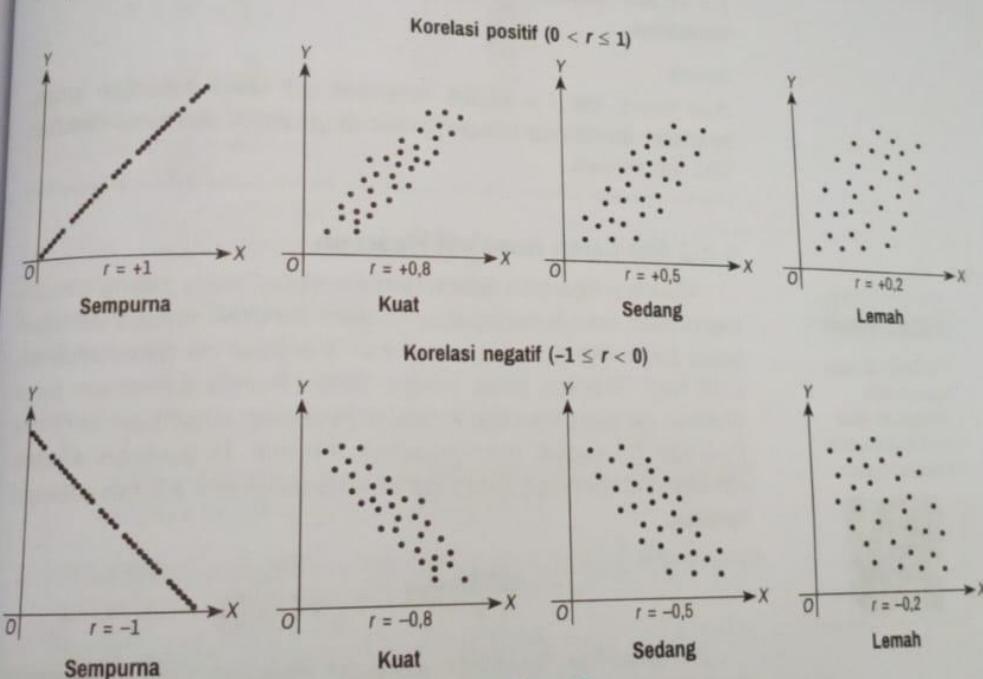
Nilai koefisien korelasi $r = \pm 1$ menunjukkan terjadi hubungan sempurna antara dua variabel. Tanda positif (+) menunjukkan korelasi positif dan tanda negatif (-) menunjukkan korelasi negatif. Nilai $r = 0$ menunjukkan tidak ada hubungan antara dua variabel, artinya tidak ada korelasi.

Perhatikan diagram pencar berikut.



Gambar 4.14

Nilai koefisien korelasi $-1 \leq r < 0$ dan $0 < r \leq 1$ menunjukkan berbagai tingkat hubungan antara dua variabel. Perhatikan diagram pencar berikut.



Gambar 4.15

Korelasi akan semakin kuat jika nilai r mendekati $+1$ atau -1 , sedangkan korelasi akan semakin lemah jika nilai r mendekati 0 .

Pembahasan:

Perkiraan berat badan seseorang jika mengonsumsi 1.450 kalori per hari:

- Menggunakan persamaan regresi linear penelitian Dedi

$$\hat{y} = 0,03x + 34,8$$

$$\Leftrightarrow \hat{y} = 0,03(1.450) + 34,8$$

$$\Leftrightarrow \hat{y} = 43,5 + 34,8$$

$$\Leftrightarrow \hat{y} = 78,3 \text{ kg}$$

- Menggunakan persamaan regresi linear penelitian Endang

$$\hat{y} = 0,02x + 36,2$$

$$\Leftrightarrow \hat{y} = 0,02(1.450) + 36,2$$

$$\Leftrightarrow \hat{y} = 29 + 36,2$$

$$\Leftrightarrow \hat{y} = 65,2 \text{ kg}$$

Selisih = $78,3 - 65,2 = 13,1 \text{ kg}$

Jadi, selisih perkiraan berat badan jika menggunakan hasil penelitian Dedi dan Endang adalah 13,1 kg.

Jawaban: B

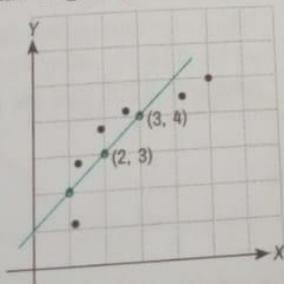


Uji Pemahaman

A. Pilihlah satu jawaban yang benar.

1. Diketahui sebuah persamaan garis regresi memiliki koefisien regresi 34,5 dan konstanta 0,25. Persamaan garis regresi tersebut adalah
- A. $\hat{y} = 34,5x + 0,25$ D. $\hat{y} = 0,25x + 34,5$
 B. $\hat{y} = 34,5x + 0,5$ E. $\hat{y} = 0,25x + 69$
 C. $\hat{y} = 34,5x + 0,75$

2. Perhatikan diagram pencar berikut.



- Garis pada diagram tersebut adalah garis regresi.
 Persamaan garis regresi tersebut adalah
- A. $\hat{y} = x + 1$ D. $\hat{y} = 2x + 2$
 B. $\hat{y} = x + 2$ E. $\hat{y} = x + 0,1$
 C. $\hat{y} = 2x + 1$

3. Perhatikan tabel data berikut.

x	1	2	3	4
y	5	8	10	13

Persamaan garis regresi yang menunjukkan hubungan antara x dan y adalah

- A. $\hat{y} = 2,6x + 5$ D. $\hat{y} = 2,5x + 2,6$
 B. $\hat{y} = 2,6x + 2,6$ E. $\hat{y} = 2,5x + 2,5$
 C. $\hat{y} = 2,6x + 2,5$

4. Tabel berikut menunjukkan data rata-rata tinggi badan siswa umur 14–18 tahun.

Umur (tahun)	14	15	16	17	18
Rata-rata Tinggi Badan (cm)	154	156	157	158	159

Persamaan garis regresi dari data tersebut adalah

- A. $\hat{y} = 1,2x + 136,7$ D. $\hat{y} = 1,2x + 139$
 B. $\hat{y} = 1,2x + 137,6$ E. $\hat{y} = 1,3x + 139$
 C. $\hat{y} = 1,3x + 137,6$

Contoh Soal dan Pembahasan

1. Perhatikan himpunan pasangan berurutan (x, y) berikut.
 $\{(2, 1), (3, 5), (4, 6)\}$
- Hitunglah koefisien korelasi Pearson antara x dan y .
 - Jelaskan tingkat hubungan antara x dan y berdasarkan nilai koefisien korelasinya.

Jawab:

- a. Perhatikan tabel berikut.

x	y	xy	x^2	y^2
2	1	2	4	1
3	5	15	9	25
4	6	24	16	36
$\sum x = 9$	$\sum y = 12$	$\sum xy = 41$	$\sum x^2 = 29$	$\sum y^2 = 62$

$$n = 3 \text{ (jumlah data)}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{9}{3} = 3$$

$$\bar{y} = \frac{\sum y}{n} = \frac{12}{3} = 4$$

$$\begin{aligned} r &= \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2} \sqrt{\sum y^2 - n\bar{y}^2}} \\ &= \frac{41 - (3)(3)(4)}{\sqrt{29 - (3)(3)^2} \sqrt{62 - (3)(4)^2}} \\ &= \frac{41 - 36}{\sqrt{29 - 27} \sqrt{62 - 48}} \\ &= \frac{5}{\sqrt{2} \sqrt{14}} \approx 0,945 \end{aligned}$$

Jadi, koefisien korelasi antara x dan y adalah 0,945.

- b. Oleh karena nilai koefisien korelasi mendekati 1, maka terdapat korelasi positif yang kuat antara x dan y .

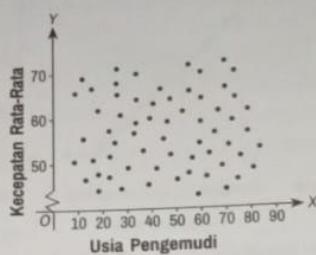
2. Tabel berikut menunjukkan data berat badan dan kecepatan lari jarak 100 m dari 4 orang siswa kelas XI SMA "Tiara".

Berat Badan (kg)	46	42	45	48
Kecepatan Lari (m/s)	14	12	12,8	13,2

Tentukan tingkat korelasi antara berat badan dan kecepatan lari sesuai dengan data tersebut.

Contoh Soal dan Pembahasan

Seseorang melakukan penelitian untuk mengetahui hubungan antara kecepatan rata-rata mobil di suatu daerah dengan usia pengemudi mobil tersebut. Hasil dari penelitian tersebut disajikan dalam bentuk diagram pencar berikut ini.



Berdasarkan hasil penelitian tersebut, diperoleh nilai koefisien korelasi $r = +0,023$. Jelaskan hubungan tersebut berdasarkan nilai koefisien korelasinya.

Jawab:

Oleh karena nilai $r = +0,023$ mendekati nol, maka hubungan antara kecepatan rata-rata dan usia pengemudi sangat lemah atau dapat dikatakan tidak ada korelasi.

4.3.2 Koefisien Korelasi Pearson

Korelasi Pearson adalah jenis korelasi yang paling umum digunakan untuk mengukur tingkat korelasi antara variabel yang berhubungan secara linear. Korelasi ini dikemukakan oleh Karl Pearson pada tahun 1900. Korelasi Pearson juga disebut dengan korelasi *Product Moment*. Koefisien korelasi Pearson (r) untuk menunjukkan tingkat hubungan antara variabel independen X dan variabel dependen Y adalah sebagai berikut.

$$r = \frac{s_{xy}}{s_x s_y} \text{ atau } r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$

Berdasarkan koefisien korelasi Pearson tersebut, maka diperoleh rumus sebagai berikut.

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}} \text{ atau } r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2} \sqrt{\sum y^2 - n\bar{y}^2}}$$

Bahas Materi

Pindailah QR code berikut untuk mengakses video pembahasan materi Korelasi Pearson.



X	Y	XY	X²	Y²
4	4	16	16	16
5	3	15	25	9
6	2	12	36	4
7	1	7	49	1
$\sum x = 22$	$\sum y = 10$	$\sum xy = 50$	$\sum x^2 = 128$	$\sum y^2 = 30$

$n = 4$ (jumlah data)

$$\bar{x} = \frac{\sum x}{n} = \frac{22}{4} = 5,5$$

$$\bar{y} = \frac{\sum y}{n} = \frac{10}{4} = 2,5$$

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2} \sqrt{\sum y^2 - n\bar{y}^2}}$$

$$= \frac{50 - (4)(5,5)(2,5)}{\sqrt{126 - (4)(5,5)^2} \sqrt{30 - (4)(2,5)^2}}$$

$$= \frac{50 - 55}{\sqrt{126 - 121} \sqrt{30 - 25}}$$

$$= \frac{-5}{\sqrt{5} \sqrt{5}} = -1$$

Oleh karena nilai korelasi koefisien $r = -1$, maka harga suatu produk dan jumlah produk yang terjual memiliki korelasi negatif sempurna. Artinya, makin bertambah harga produk, makin sedikit jumlah produk yang terjual.

4.3.3 Koefisien Determinasi

Koefisien determinasi digunakan untuk memprediksi dan melihat seberapa besar pengaruh variabel independen terhadap variabel dependen. Koefisien determinasi (r^2) adalah kuadrat dari koefisien korelasi (r). Tingkat korelasi berdasarkan nilai koefisien determinasi adalah sebagai berikut.

Tabel 4.1 Tingkat korelasi berdasarkan nilai koefisien determinasi

Nilai Koefisien Determinasi	Tingkat Korelasi
$r^2 = 0$	tidak ada korelasi
$0 < r^2 < 0,25$	sangat lemah
$0,25 \leq r^2 < 0,50$	lemah
$0,50 \leq r^2 < 0,75$	sedang
$0,75 \leq r^2 < 0,90$	kuat
$0,90 \leq r^2 < 1$	sangat kuat
$r^2 = 1$	sempurna

Jawab:

Misalkan: x = berat badan (kg)
 y = kecepatan lari (m/s)

Perhatikan tabel berikut.

x	y	xy	x^2	y^2
46	14	644	2.116	
42	12	504	1.764	196
45	12,8	576	2.025	144
48	13,2	633,6	2.304	183,84
$\sum x = 181$	$\sum y = 52$	$\sum xy = 2.357,6$	$\sum x^2 = 8.209$	$\sum y^2 = 678,08$

$n = 4$ (jumlah data)

$$\bar{x} = \frac{\sum x}{n} = \frac{181}{4} = 45,25$$

$$\bar{y} = \frac{\sum y}{n} = \frac{52}{4} = 13$$

$$\begin{aligned}
 r &= \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2} \sqrt{\sum y^2 - n\bar{y}^2}} \\
 &= \frac{2.357,6 - (4)(45,25)(13)}{\sqrt{8.209 - (4)(45,25)^2} \sqrt{678,08 - (4)(13)^2}} \\
 &= \frac{2.357,6 - 2.353}{\sqrt{8.209 - 8.190,25} \sqrt{678,08 - 676}} \\
 &= \frac{4,6}{\sqrt{18,75} \sqrt{2,08}} \approx 0,737
 \end{aligned}$$

Oleh karena nilai koefisien korelasi mendekati 1, maka terdapat korelasi positif yang kuat antara berat badan dan kecepatan lari. Semakin besar berat badan siswa, semakin bertambah kecepatan larinya.

3. Tabel berikut menunjukkan data harga suatu produk dan jumlah produk yang terjual.

Harga Produk (rupiah)	4.000	5.000	6.000	7.000
Jumlah Produk Terjual (unit)	4.000	3.000	2.000	1.000

Tentukan tingkat korelasi antara harga suatu produk dan jumlah produk yang terjual.

Pembahasan:

Misalkan: x = harga produk (dalam ribu rupiah)

y = jumlah produk yang terjual (dalam ribu unit)

Perhatikan tabel berikut.

Seorang siswa melakukan penelitian tentang pengaruh jarak tempuh ke sekolah dengan nilai ujian akhir siswa. Data yang ia dapatkan adalah sebagai berikut.

Jarak Tempuh (km)	2	2	3	1	4
Nilai Ujian	70	85	90	75	80

Bagaimana hasil penelitian siswa tersebut? Jelaskan berdasarkan nilai koefisien determinasi dari data tersebut.

Jawab:

Misalkan: x = jarak tempuh (km)

y = nilai ujian

Perhatikan tabel berikut.

x	y	xy	x^2	y^2
2	70	140	4	4.900
2	85	170	4	7.225
3	90	270	9	8.100
1	75	75	1	5.625
4	80	320	16	6.400
$\sum x = 12$	$\sum y = 400$	$\sum xy = 975$	$\sum x^2 = 34$	$\sum y^2 = 32.250$

$n = 5$ (jumlah data)

$$\bar{x} = \frac{\sum x}{n} = \frac{12}{5} = 2,4$$

$$\bar{y} = \frac{\sum y}{n} = \frac{400}{5} = 80$$

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2} \sqrt{\sum y^2 - n\bar{y}^2}}$$

$$= \frac{975 - (5)(2,4)(80)}{\sqrt{34 - (5)(2,4)^2} \sqrt{32.250 - (5)(80^2)}}$$

$$= \frac{975 - 960}{\sqrt{34 - 28,8} \sqrt{32.250 - 32.000}}$$

$$= \frac{15}{\sqrt{5,2} \sqrt{250}} \approx 0,416$$

$$r^2 = (0,416)^2 = 0,173$$

$$\text{Koefisien determinasi} = r^2 \times 100\% \\ = 0,173 \times 100\% \\ = 17,3\%$$

Nilai koefisien determinasi 17,3% menunjukkan bahwa jarak tempuh ke sekolah tidak berpengaruh besar terhadap nilai ujian akhir siswa.

Nilai koefisien determinasi berkisar antara $0 \leq r \leq 1$ dan biasa diubah ke dalam bentuk persentase (%) untuk menyatakan besar pengaruh variabel independen terhadap variabel dependen.

$$\text{Koefisien determinasi} = r^2 \times 100\%$$



Contoh Soal dan Pembahasan

- Perhatikan data pada tabel berikut.

x	1	2	3
y	5	8	10

Hitunglah koefisien determinasi dari data tersebut dalam bentuk persentase.

Jawab:

Perhatikan tabel berikut.

x	y	xy	x^2	y^2
1	5	5	1	25
2	8	16	4	64
3	10	30	9	100
$\sum x = 6$	$\sum y = 23$	$\sum xy = 51$	$\sum x^2 = 14$	$\sum y^2 = 189$

$$n = 3 \text{ (jumlah data)}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{6}{3} = 2$$

$$\bar{y} = \frac{\sum y}{n} = \frac{23}{3}$$

$$\begin{aligned} r &= \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2} \sqrt{\sum y^2 - n\bar{y}^2}} \\ &= \frac{51 - (3)(2)\left(\frac{23}{3}\right)}{\sqrt{14 - (3)(2)^2} \sqrt{189 - (3)\left(\frac{23}{3}\right)^2}} \\ &= \frac{51 - 46}{\sqrt{14 - 12} \sqrt{189 - \frac{529}{3}}} \\ &= \frac{5}{\sqrt{2} \sqrt{\frac{38}{3}}} \approx 0,993 \end{aligned}$$

$$r^2 = (0,993)^2 = 0,986$$

$$\begin{aligned} \text{Koefisien determinasi} &= r^2 \times 100\% \\ &= 0,986 \times 100\% = 98,6\% \end{aligned}$$

Jadi, koefisien determinasi dari data tersebut adalah 98,6%.

Tingkat korelasi antara berat badan dan tinggi badan lima orang siswa tersebut adalah

B. Jawablah dengan jelas dan benar.

1. Tentukan nilai r dan r^2 untuk setiap kumpulan data berikut, kemudian tentukan tingkat korelasinya.

x	-2	-1	3
y	-1	1	2

x	-1	0	5	9
y	7	5	2	-3

x	2	3	4	5
y	5	8	10	5

x	10	10	25	20	30
y	50	30	40	10	15

2. Tabel berikut menunjukkan data biaya promosi suatu produk dan hasil penjualan produk tersebut.

Biaya Promosi (ratus ribu rupiah)	7	8	9	10	15
Hasil Penjualan (juta rupiah)	8	10	12	15	20

- a. Hitunglah koefisien korelasi dan koefisien determinasi dari data tersebut.
b. Jelaskan hubungan antara biaya promosi dan hasil penjualan produk tersebut.

- A. sangat lemah
B. lemah
C. sedang
D. kuat
E. sangat kuat

3. Seseorang melakukan penelitian tentang pengaruh suhu udara terhadap jumlah pengunjung suatu taman wisata. Selama sepekan, ia mencatat suhu udara pada siang hari dan jumlah pengunjung di taman wisata tersebut dalam bentuk tabel berikut ini.

Hari ke-	1	2	3	4	5	6	7
Suhu Udara (°C)	25	28	30	30	27	25	31
Jumlah Pengunjung	640	620	540	480	580	610	360

- a. Gambarlah diagram pencar dari data tersebut.
b. Hitunglah koefisien korelasi dan koefisien determinasi dari data tersebut.
c. Apakah suhu udara berpengaruh terhadap jumlah pengunjung taman wisata tersebut? Jika ya, seberapa besar pengaruhnya? Jelaskan jawaban Anda.
d. Ada empat orang mahasiswa menyelidiki korelasi antara tinggi badan dan tekanan darah. Mereka mengumpulkan data dari responden yang sama, tapi mahasiswa 1 dan 2 menggunakan satuan cm dan mahasiswa 3 dan 4 menggunakan satuan m untuk tinggi badan. Mahasiswa 1 dan 3 menggunakan variabel x untuk tinggi badan dan y untuk tekanan darah, sedangkan mahasiswa 2 dan 4 sebaliknya. Apakah keempat mahasiswa akan memperoleh nilai koefisien korelasi yang sama? Jika ya, jelaskan alasan Anda. Jika tidak, siapa yang mendapatkan hasil yang berbeda?

4.4 Menggunakan Aplikasi dalam Statistika

Penggunaan aplikasi dalam statistika dapat memudahkan kita menganalisis data dalam jumlah yang banyak. Contoh aplikasi statistika yang dapat digunakan adalah Microsoft Excel, SPSS, GeoGebra, SAS, dan lain-lain. Pada subbab ini, kita akan menggunakan Microsoft Excel untuk menggambar diagram pencar, menentukan persamaan garis regresi linear, serta menghitung koefisien korelasi.

4.4.1 Menggambar Diagram Pencar

Kita akan menggambar diagram pencar untuk mengetahui hubungan antara kepadatan penduduk dan jumlah kasus Covid-19. Langkah-langkah menggambar diagram pencar dengan Microsoft Excel adalah sebagai berikut.

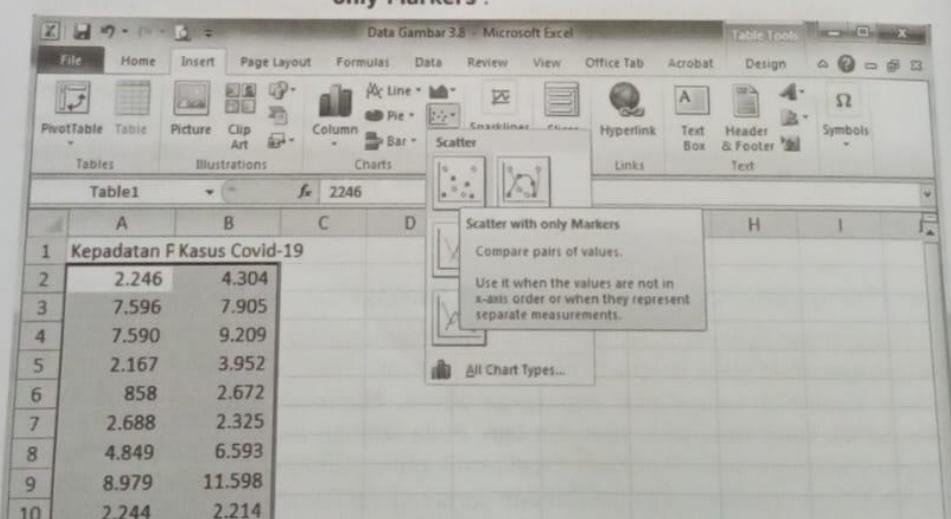
1. Input data kepadatan penduduk dan jumlah kasus Covid-19.

The screenshot shows a Microsoft Excel spreadsheet titled "Data Gambar 3.8 - Microsoft Excel". The table has two columns: A (Kepadatan) and B (Kasus Covid-19). The data is as follows:

	A	B
1	Kepadatan F	Kasus Covid-19
2	2.246	4.304
3	7.596	7.905
4	7.590	9.209
5	2.167	3.952
6	858	2.672
7	2.688	2.325
8	4.849	6.593
9	8.979	11.598
10	2.244	2.214

Gambar 4.16

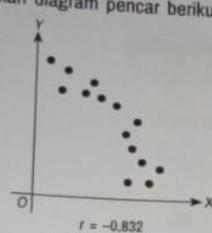
2. Data terdiri atas 27 amatan. Blok sel yang akan dibuat diagram pencar, yaitu sel A2–A28 dan B2–B28. Kemudian, klik menu "Insert" → klik icon "Charts" → klik icon "Scatter" → klik icon "Scatter with only Markers".



Gambar 4.17

Uji Pemahaman

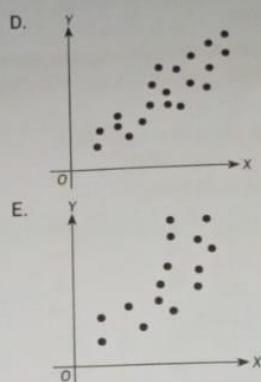
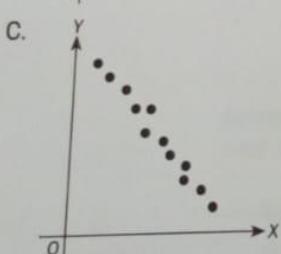
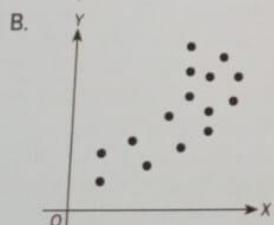
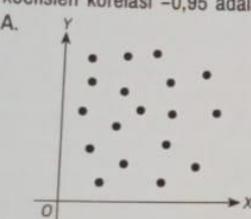
- A. Pilihlah satu jawaban yang benar.
1. Perhatikan diagram pencar berikut.



Tingkat korelasi dari data pada diagram pencar tersebut adalah

- A. sangat lemah D. kuat
B. lemah E. sangat kuat
C. sedang

2. Diagram pencar yang mungkin untuk nilai koefisien korelasi $-0,95$ adalah



3. Perhatikan data berikut.

x	3	5	7	10
y	1	4	5	8

Nilai koefisien korelasi Pearson dari data tersebut adalah

- A. 0,996 D. 0,886
B. 0,986 E. 0,872
C. 0,972

4. Perhatikan data berikut.

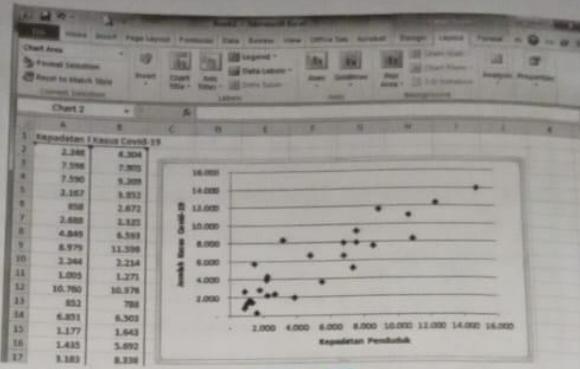
x	5	3	2	1
y	4	6	7	5

Koefisien determinasi dari data tersebut adalah

- A. 25% D. 38%
B. 28% E. 98%
C. 32%

5. Tabel berikut menunjukkan data berat badan dan tinggi badan lima orang siswa kelas XI SMA "Mandiri".

Berat Badan (kg)	42	38	45	48	40
Tinggi Badan (cm)	140	142	148	150	150



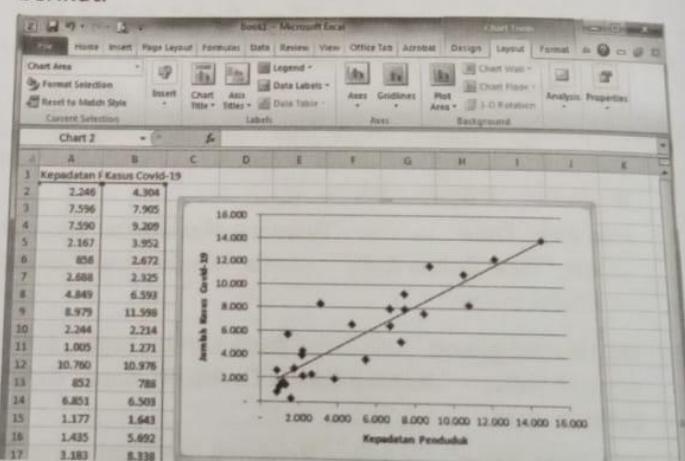
Gambar 4.19

4.4.2 Menampilkan Gambar dan Persamaan Garis Regresi

Kita akan menggambar garis regresi dan menentukan persamaannya menggunakan diagram pencar kepadatan penduduk dan jumlah kasus Covid-19. Langkah-langkah menggambar garis regresi dengan Microsoft Excel adalah sebagai berikut.

1. Tampilkan dahulu menu "Chart Tools" dengan cara: klik diagramnya → klik "Chart Tools" → klik "Layout" → klik "Trendline" → klik "Linear Trendline".

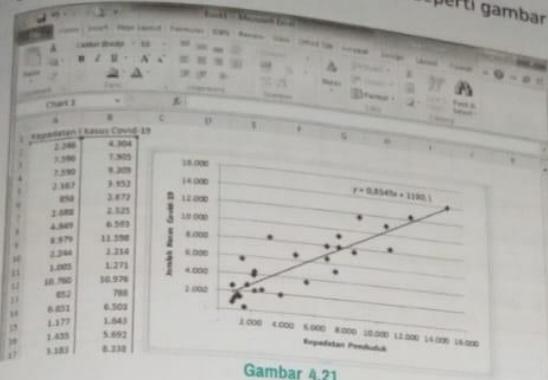
Hasil yang akan tampil di layar komputer seperti gambar berikut.



Gambar 4.20

2. Tampilkan persamaan garis regresi dengan cara: klik diagramnya → klik "Chart Tools" → klik "Layout" → klik "Trendline" → klik "More Trendline Options" dan pilih "Linear" pada "Trend/Regression Type" → centang "Display Equation on Chart".

Hasil yang akan tampil di layar komputer seperti gambar di bawah ini.

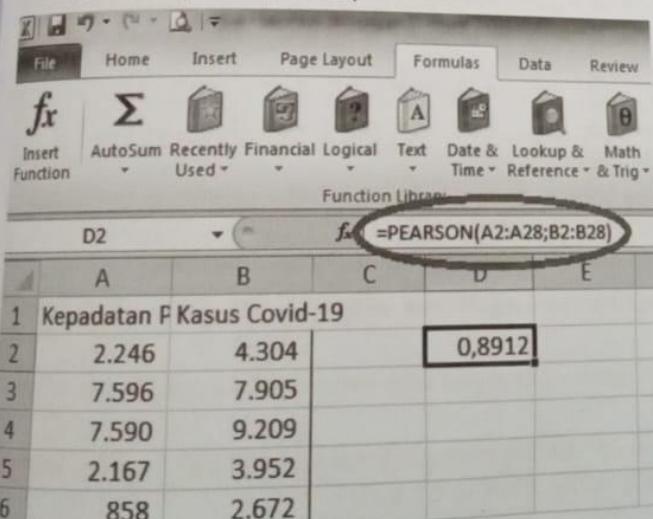


Gambar 4.21

4.4.3 Menghitung Koefisien Korelasi

Kita akan menghitung koefisien korelasi Pearson dari data kepadatan penduduk dan jumlah kasus Covid-19. Langkah-langkah menghitung koefisien korelasi dan koefisien determinasi dengan Microsoft Excel adalah sebagai berikut.

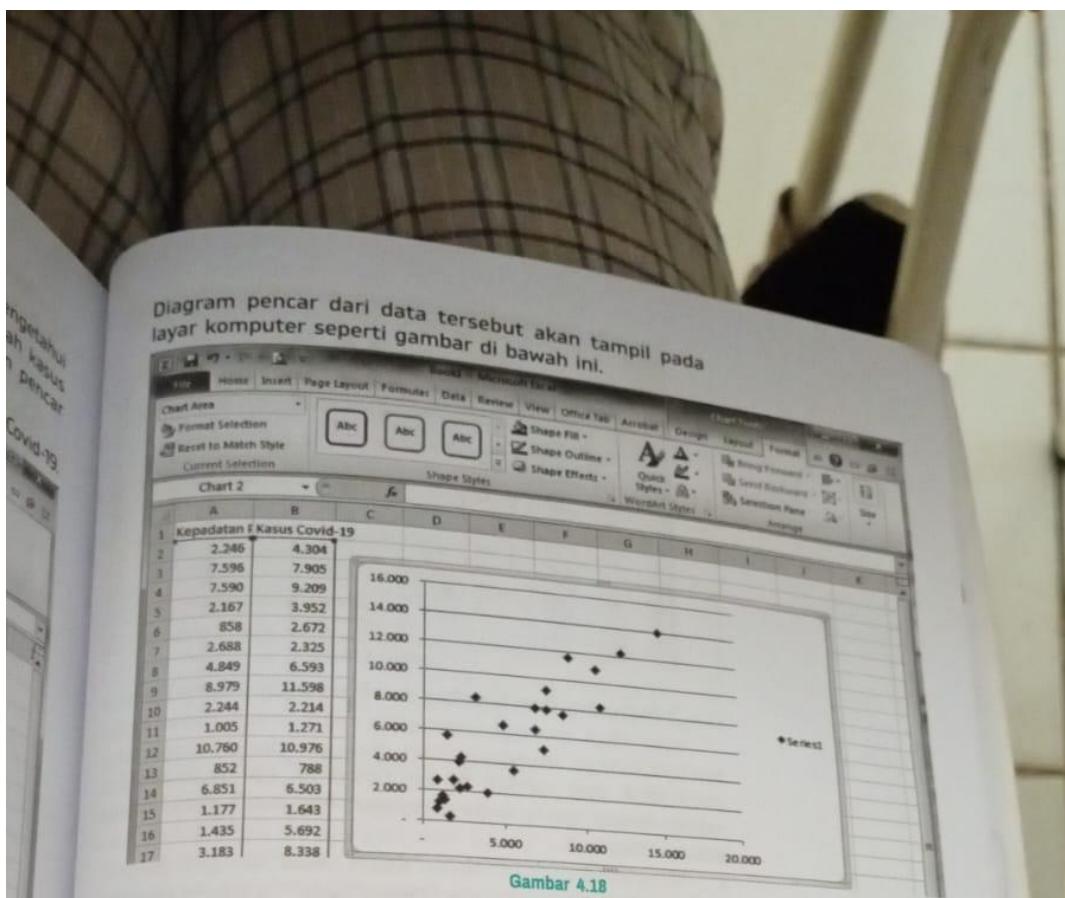
1. Blok sel yang akan ditentukan nilai koefisien korelasinya, yaitu sel A2–A28 dan B2–B28.
2. Pada sel yang diinginkan, ketik formula fungsi seperti berikut ini:
 $=PEARSON(A2:A28,B2:B28)$



Gambar 4.22

BAB 4 Statistika Regresi

201



Gambar 4.18

Kita dapat menambahkan judul "**Kepadatan Penduduk**" untuk sumbu mendatar dan judul "**Jumlah Kasus Covid-19**" untuk sumbu tegak dengan cara berikut.

1. Tampilkan dahulu menu "**Chart Tools**" dengan cara:
klik diagramnya → klik "**Chart Tools**" → klik "**Layout**" → klik "**Axis Title**".
2. Untuk mengganti "Axis Title" pada sumbu mendatar (horizontal) dengan tulisan "**Kepadatan Penduduk**", dapat dilakukan dengan cara:
klik "**Axis Title**" → klik "**Primary Horizontal Axis Title**" → klik "**Title Below Axis**" → ketik "**Kepadatan Penduduk**" → tekan enter.
3. Untuk mengganti "Axis Title" pada sumbu tegak (vertikal) dengan tulisan "**Jumlah Kasus Covid-19**", dapat dilakukan dengan cara:
klik "**Axis Title**" → klik "**Primary Vertical Axis Title**" → klik "**Rotated Axis**" → ketik "**Jumlah Kasus Covid-19**" → tekan enter.

Hasil yang akan tampil di layar komputer seperti gambar berikut.

13. Rumus koefisien korelasi Pearson:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}} \text{ atau } r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2} \sqrt{\sum y^2 - n\bar{y}^2}}$$

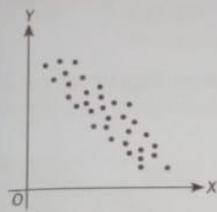
14. Rumus koefisien determinasi:

$$\text{Koefisien determinasi} = r^2 \times 100\%$$

Latihan Soal Akhir Bab 4

A. Pilihlah satu jawaban yang benar.

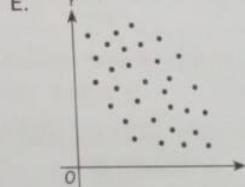
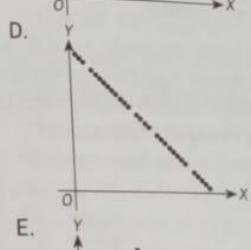
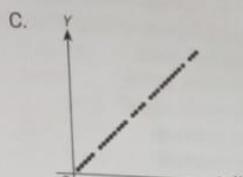
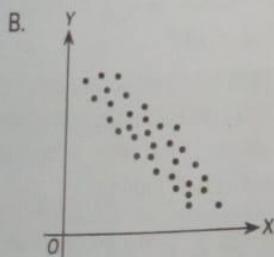
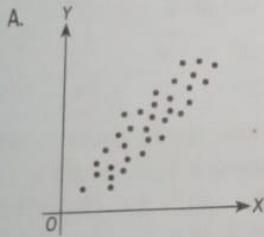
1. Perhatikan diagram pencar berikut.



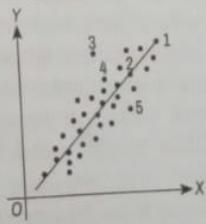
Jenis korelasi antara variabel x dan y pada diagram tersebut adalah

- A. korelasi negatif
- B. korelasi positif
- C. korelasi kuadratik
- D. korelasi nonlinear
- E. tidak ada korelasi

2. Diagram pencar yang menunjukkan adanya korelasi negatif linear sempurna antara dua variabel adalah



3. Perhatikan diagram pencar berikut.



Data outlier pada diagram pencar tersebut ditunjukkan oleh nomor

- A. 1
- B. 2
- C. 3
- D. 4
- E. 5

Rangkuman

- Diagram pencar atau *scatter plot* dapat digunakan untuk menunjukkan hubungan antara dua variabel yang digambarkan dalam bentuk titik-titik (points) pada bidang koordinat Cartesius.
- Variabel independen adalah variabel yang mempengaruhi atau menyebabkan berubahnya variabel dependen.
- Variabel dependen adalah variabel yang nilainya dipengaruhi oleh variabel independen.
- Hubungan linear antara dua variabel pada diagram pencar disebut korelasi.
- Outlier atau pencilan adalah data yang menyimpang terlalu jauh dari data lain dari sekumpulan data.
- Korelasi antara dua variabel tidak selalu menunjukkan hubungan sebab-akibat pada keduanya.
- Garis regresi atau garis *best-fit* adalah garis yang paling tepat untuk menggambarkan hubungan antara dua variabel. Bentuk persamaan garis regresi:

$$\hat{y} = mx + c$$

di mana:

\hat{y} = nilai variabel dependen yang diprediksi

x = nilai variabel independen

m = koefisien regresi (gradien garis regresi)

c = konstanta (nilai \hat{y} jika $x = 0$)

- Residu adalah jarak vertikal antara titik-titik data dan garis regresi. Nilai residu:

$$\text{Residu } (\varepsilon) = y - \hat{y}$$

di mana: y = nilai variabel dependen yang diamati

\hat{y} = nilai variabel dependen yang diprediksi.

- Rumus persamaan garis regresi sebagai berikut.

$$\hat{y} - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \text{ atau } \hat{y} - \bar{y} = \frac{SS_{xy}}{SS_{xx}}(x - \bar{x})$$

di mana:

\bar{x} = rata-rata dari nilai x

\bar{y} = rata-rata dari nilai y

s_{xy} = kovariansi dari X dan Y

s_x = standar deviasi dari X

SS_{xy} = Jumlah perkalian selisih variabel independen x terhadap rata-ratanya dan variabel dependen y terhadap rata-ratanya.

SS_{xx} = Jumlah kuadrat selisih variabel independen x terhadap rata-ratanya.

- Interpolasi adalah penggunaan garis regresi untuk memprediksi nilai yang berada di dalam rentang data.

- Ekstrapolasi adalah penggunaan garis regresi untuk memprediksi nilai yang berada di luar rentang data.

- Koefisien korelasi merupakan nilai yang menunjukkan kekuatan dan arah hubungan antara dua variabel dengan nilai berkisar antara $-1 \leq r \leq +1$.

Persamaan regresi yang menunjukkan data tersebut adalah . . .

- $\hat{y} = -0,21x + 18,4$
- $\hat{y} = -0,21x + 17,4$
- $\hat{y} = -0,23x + 15,4$
- $\hat{y} = -0,23x + 16,4$
- $\hat{y} = -0,23x + 17,4$

11. Perhatikan data pada tabel berikut.

x	1	3	7	9
y	3	4	5	6

Perkiraan nilai y untuk $x = 6$ adalah . . .

- 2,1
- 2,75
- 4,75
- 4,80
- 4,85

Perhatikan tabel berikut untuk menjawab soal nomor 12–15.

Usia Bayi (bulan)	1	2	3	4	5
Berat Badan (kg)	4	4,5	5,2	5,6	6,8

12. Persamaan garis regresi yang menunjukkan hubungan antara usia dan rata-rata berat badan bayi adalah . . .

- $\hat{y} = 0,67x + 3,21$
- $\hat{y} = 0,76x + 3,21$
- $\hat{y} = 0,67x + 3,31$
- $\hat{y} = 0,76x + 3,51$
- $\hat{y} = 0,67x + 3,51$

13. Perkiraan rata-rata berat badan bayi usia 3,5 bulan adalah . . .

- 6,07 kg
- 5,87 kg
- 5,86 kg
- 5,66 kg
- 5,56 kg

14. Perkiraan rata-rata berat badan bayi usia 7 bulan adalah . . .

- 7,5 kg
- 7,9 kg
- 8,1 kg
- 8,3 kg
- 8,4 kg

15. Tingkat korelasi antara usia dan berat badan bayi adalah . . .

- sangat lemah
- lemah
- sedang
- kuat
- sangat kuat

16. Perhatikan data pada tabel berikut.

x	10	30	50	70	90
y	30	50	70	90	

Nilai koefisien korelasi Pearson dari data tersebut adalah . . .

- 0,96
- 0,97
- 0,98
- 0,99
- 1

17. Perhatikan data pada tabel berikut.

x	10	5	6	15
y	1	6	7	10

Koefisien determinasi dari data tersebut adalah . . .

- 8,9%
- 9,8%
- 9,9%
- 89%
- 98%

18. Perhatikan data pada tabel berikut.

Tinggi Badan Ibu (cm)	154	158	160	162	165
Tinggi Badan Anak (cm)	120	110	140	120	135

Persentase besar pengaruh tinggi badan ibu terhadap tinggi badan anak adalah . . .

- 20,2%
- 23,2%
- 24,2%
- 49,2%
- 50,2%

Perhatikan tabel berikut untuk menjawab soal nomor 19 dan 20.

Lama Menonton TV per Hari (jam)	2	3	4	5
Nilai Ujian Matematika	80	85	82	75

19. Nilai koefisien korelasi antara lama menonton TV dalam sehari terhadap nilai ujian Matematika adalah . . .

- 0,853
- 0,753
- 0,653
- 0,553
- 0,453

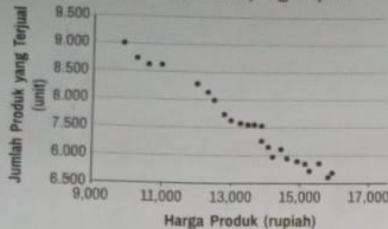
20. Persentase besar pengaruh lama menonton TV dalam sehari terhadap nilai ujian Matematika adalah . . .

- 60,6%
- 50,6%
- 55,3%
- 30,6%
- 25,3%

BAB 4 Statistika Regresi

205

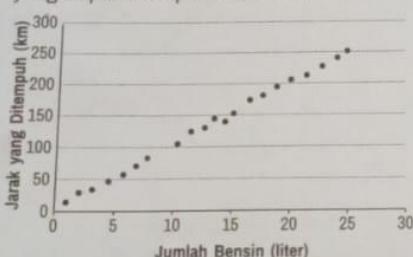
4. Berikut adalah diagram pencar yang menunjukkan hubungan antara harga sebuah produk dan jumlah produk yang terjual.



Jenis korelasi yang sesuai dengan data pada diagram tersebut adalah . . .

- A. korelasi linear positif
- B. korelasi linear negatif
- C. korelasi nonlinear
- D. korelasi kuadratik
- E. tidak ada korelasi

5. Berikut adalah diagram pencar yang menunjukkan hubungan antara jumlah bensin yang dibutuhkan sebuah mobil dengan jarak yang dapat ditempuh mobil tersebut.



Pernyataan yang tepat sesuai dengan diagram tersebut adalah . . .

- A. Makin sedikit jumlah bensin, makin jauh jarak yang dapat ditempuh.
- B. Makin banyak jumlah bensin, makin jauh jarak yang dapat ditempuh.
- C. Makin banyak jumlah bensin, makin dekat jarak yang dapat ditempuh.
- D. Jumlah bensin tidak mempengaruhi jarak yang dapat ditempuh.
- E. Tidak ada hubungan antara jumlah bensin dan jarak yang dapat ditempuh.

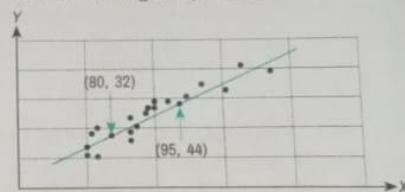
6. Diketahui $\bar{x} = 5$; $\bar{y} = 3$; $SS_{xy} = -0,8$; dan $SS_{xx} = 2$. Persamaan garis regresi yang menunjukkan hubungan variabel x dan y adalah . . .

- A. $\hat{y} = -0,2x + 0,2$ D. $\hat{y} = 0,4x + 5$
- B. $\hat{y} = -0,2x + 1$ E. $\hat{y} = 4x + 3$
- C. $\hat{y} = -0,4x + 2$

7. Diketahui $n = 5$, $\sum x = 20$; $\sum y = 35$; $\sum xy = 150$; $\sum(x - \bar{x})^2 = 10$ dan $\sum(y - \bar{y})^2 = 10$. Persamaan regresi yang menunjukkan hubungan variabel x dan y adalah . . .

- A. $\hat{y} = 0,1x - 7$ D. $\hat{y} = x + 3$
- B. $\hat{y} = 0,1x + 3$ E. $\hat{y} = x + 11$
- C. $\hat{y} = x - 7$

8. Perhatikan diagram pencar berikut.



Persamaan garis regresi pada diagram pencar tersebut adalah . . .

- A. $\hat{y} = 0,8x - 64$ D. $\hat{y} = 0,8x + 32$
- B. $\hat{y} = 0,8x + 64$ E. $\hat{y} = 8x + 32$
- C. $\hat{y} = 0,8x - 32$

9. Perhatikan data pada tabel berikut.

x	12	14	18	20
y	10	8	6	4

Persamaan garis regresi yang menunjukkan hubungan antara x dan y adalah . . .

- A. $\hat{y} = 0,7x - 18,2$ D. $\hat{y} = -0,7x + 18,2$
- B. $\hat{y} = 0,7x + 18,2$ E. $\hat{y} = -0,7x + 9,1$
- C. $\hat{y} = -0,7x - 18,2$

10. Seseorang ingin menyelidiki pengaruh berat badan terhadap kecepatan lari 100 meter dari lima orang siswa. Data yang ia kumpulkan disajikan dalam tabel berikut.

Berat Badan (kg)	48	54	56	58	60
Kecepatan Lari (m/s)	5	4	4	3	2

- a. Gambarlah diagram pencar yang menunjukkan hubungan antara durasi belajar siswa di rumah dengan rata-rata nilai UNBK.
- b. Tentukan persamaan garis regresi untuk menunjukkan hubungan antara durasi belajar siswa di rumah dengan rata-rata nilai UNBK.
- c. Hitunglah koefisien korelasi dan koefisien determinasi untuk menentukan kekuatan hubungan antara durasi belajar siswa di rumah dengan rata-rata nilai UNBK.
- d. Jelaskan hubungan antara durasi belajar siswa di rumah dengan rata-rata nilai UNBK berdasarkan jawaban a-c.

Soal Model AKM

Perhatikan stimulus berikut.

Uang Saku Anak Sekolah



Sumber: www.shutterstock.com

Bagi para orang tua memberikan uang saku kepada anak sudah menjadi hal yang lumrah. Namun, sebenarnya berapa uang saku yang tepat untuk diberikan kepada anak-anak? Jawabannya tentu tidak ada jumlah yang pasti. Karena tiap usia anak membutuhkan jumlah uang saku yang berbeda-beda. Begitu pula tingkat penghasilan orang tua, gaya hidup dan aktivitas anak-anak yang berbeda-beda.

Lalu, berapa sebenarnya jumlah yang tepat untuk anak? Anak-anak di bawah umur 9 tahun mungkin hanya membutuhkan uang untuk cemilan/minuman di sekolah, serta ongkos kendaraan umum bagi mereka yang tidak diantar/dijemput.

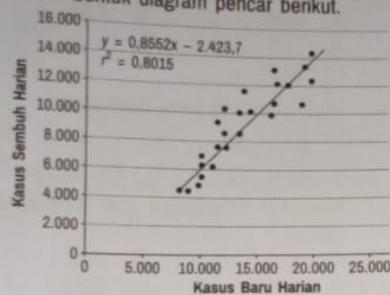
Berbeda dengan anak-anak yang lebih besar. Kebutuhan mereka saat ini, selain untuk makanan dan ongkos kendaraan umum, juga ditambah dengan biaya pulsa. Bahkan saat ini, mereka terbiasa membelanjakan uang untuk membeli aplikasi dan berlangganan musik di gadget mereka. Pada umumnya, remaja juga membutuhkan uang untuk bersosialisasi dengan teman-temannya dan membeli buku ataupun berlangganan musik.

Sumber: <https://finance.detik.com/portfolio/d-2428782/tips-memberikan-uang-saku-ideal-untuk-anak>

Lima orang mahasiswa melakukan penelitian terhadap 30 siswa SMA Negeri "Mutia Bangsa" untuk menyelidiki pengaruh jarak sekolah terhadap besar uang saku yang diberikan orang tua. Hasil penelitian tersebut disajikan pada grafik berikut.

B. Jawablah dengan jelas dan benar.

1. Seseorang melakukan penelitian untuk menyelidiki hubungan antara kasus baru harian dan kasus sembuh harian Covid-19 di suatu daerah. Hasil penelitian tersebut ditampilkan dalam bentuk diagram pencar berikut.



Sumber: <https://link.springer.com/article/10.1007/s10389-021-01512-2>

Jelaskan hubungan antara kasus baru harian dan kasus sembuh harian Covid-19 berdasarkan hasil penelitian tersebut.

2. Tabel berikut menunjukkan data jumlah pelanggaran dan jumlah kecelakaan lalu lintas di suatu provinsi pada tahun 2018–2021.

Tahun	2018	2019	2020	2021
Jumlah Pelanggaran	180.000	90.000	80.000	30.000
Jumlah Kecelakaan	5.000	8.000	4.500	5.400

- a. Tentukan persamaan garis regresi di mana jumlah pelanggaran adalah variabel independen dan jumlah kecelakaan adalah variabel dependen.
 b. Hitunglah perkiraan jumlah kecelakaan pada tahun 2022 jika terdapat jumlah pelanggaran sebanyak 20.000 kasus.
 c. Seseorang ingin menyelidiki pengaruh jam kerja pegawai di suatu perusahaan terhadap besar gaji yang diterima. Data yang ia peroleh disajikan dalam tabel berikut.

Jam Kerja per Hari (jam)	6	4	8	7	7	8	8	8
Besar Gaji per Bulan (juta rupiah)	3	2	5	4	5	6	3	4

- a. Gambarlah diagram pencar dari data tersebut.
 b. Jelaskan pengaruh jam kerja pegawai terhadap gaji yang diterima berdasarkan nilai r dan r^2 .

4. Anto ingin menyelidiki hubungan antara durasi belajar siswa di rumah dengan rata-rata nilai UNBK. UNBK terdiri atas mata pelajaran Bahasa Indonesia, Matematika, Bahasa Inggris, dan 1 mata pelajaran jurusan. Anto mendapatkan data yang disajikan dalam tabel berikut.

Siswa	Durasi Belajar	Nilai			
		Bahasa Indonesia	Matematika	Bahasa Inggris	Mata Pelajaran Jurusan
1	4	80	70	80	90
2	6	90	80	80	70
3	2	80	60	90	80
4	6	90	80	90	90
5	2	70	60	70	80
6	3	80	70	80	70
7	2	60	60	70	60
8	1	50	60	70	50
9	2	70	60	70	70
10	6	80	80	90	80
11	4	60	60	70	60
12	3	80	90	90	80
13	4	90	90	90	80
14	2	70	80	70	70
15	2	60	50	70	60
16	2	70	80	80	70
17	1	50	60	50	70
18	2	60	60	70	50
19	4	70	80	70	60
20	5	80	90	60	90