MODUL 1

Studi Deskriptif Data Bivariat

Prof. Dr. Zanzawi Soejoeti

PENDAHULUAN

Modul pertama mata kuliah *Metode Statistika 2* ini akan mengantarkan kita untuk mempelajari hubungan suatu variabel dengan variabel lainnya (bivariat). Pembahasan diawali dengan bagaimana meringkas data kategorik bivariat atau tabel silang dengan menghitung banyaknya individu atau frekuensi, frekuensi relatif, dan frekuensi marginal. Pembahasan dilanjutkan dengan hubungan bivariat data kontinu yang berupa ukuran numerik koefisien korelasi, memperkirakan garis lurus, dan diagram pencar (*scatter plot*). Pembahasan diakhiri dengan inferensi distribusi normal bivariat yang menggambarkan hubungan dua variabel numerik atau kontinu. Inferensi ini mencakup pengujian hipotesis tentang koefisien korelasi dengan ukuran sampel yang besar dan ukuran sampel yang kecil, serta pendugaan interval atau selang kepercayaan untuk koefisien korelasi.

Setelah mempelajari Modul 1 ini, Anda diharapkan dapat memperoleh gambaran tentang deskripsi data bivariat, baik untuk data kategorik maupun data numerik beserta aplikasinya. Secara khusus, setelah mempelajari Modul 1 ini Anda diharapkan dapat:

a. mendeskripsikan data kategorik bivariat;
b. menganalisis data bivariat yang kontinu dengan koefisien korelasi;
c. melakukan pengujian hipotesis tentang koefisien korelasi;
d. melakukan pendugaan interval kepercayaan tentang koefisien korelasi.

KEGIATAN BELAJAR 1

Data Kategorik dan Pengukuran Bivariat

Observasi dua variabel atau lebih kerap kali diperoleh dari unit-unit sampel. Dengan mempelajari data bivariat atau multivariate seperti itu, kita ingin mengungkap apakah ada hubungan antara variabel-variabel itu, seberapa kuat hubungan itu, dan apakah satu variabel yang menjadi perhatian kita dapat diperkirakan secara efektif dari informasi nilai-nilai variabel yang lain. Guna melukiskan konsep ini, kita membatasi perhatian kita pada kasus yang paling sederhana, yakni hanya 2 karakteristik yang diamati pada tiap unit sampel yang ada. Beberapa contoh misalnya berikut ini.

1. Jenis kelamin dan jenis pekerjaan para sarjana.
2. Kebiasaan merokok dan penyakit jantung pada orang laki-laki dewasa.
3. Rata-rata karbohidrat dan protein yang diserap setiap hari oleh anak-anak umur 10 tahun.
4. Banyak pupuk yang digunakan dan hasil panen per hektar.

Dua sifat yang diamati dapat kedua-duanya berupa variabel kualitatif, atau variabel numerik, atau salah satu kualitatif dan yang lain variabel numerik. Di sini kita hanya akan mempelajari keadaan di mana karakteristik-karakteristik yang diamati keduanya kategorik atau keduanya numerik.

A. MERINGKAS DATA KATEGORIK BIVARIAT

Jika bagi setiap unit sampel diamati dua sifatnya maka data hasil pengamatan itu dapat diringkaskan dan disajikan dalam bentuk tabel frekuensi 2 arah. Kategori-kategori bagi sifat yang pertama dituliskan pada tepi sisi kiri, dan bagi sifat yang kedua pada tepi sisi atas, serta cacah frekuensinya dituangkan dalam tiap sel. Data dalam bentuk ringkasan ini biasanya dinamakan data *klasifikasi silang* atau data *tabulasi silang*. Dalam terminologi statistik dinamakan juga *tabel kategorik*.

Contoh 1.1

Suatu survei dilakukan terhadap 400 orang mahasiswa di suatu kota. Mereka dimintai pendapat tentang rencana pembangunan gedung olah raga di kota itu, diperoleh data sebagai berikut.

Tabel 1.1

Mahasiswa	Mendukung	Menentang	Tidak Berpendapat	Jumlah
Putri	71	20	33	124
Putra	112	84	80	276
Jumlah	183	104	113	400

Angka-angka dalam tabel itu cukup jelas artinya. Misalnya, dari 400 orang mahasiswa, terdapat 124 orang mahasiswa putri. Di antara 124 orang mahasiswa putri ini 71 orang menyatakan mendukung, 20 orang menolak dan 33 orang tidak menyatakan pendapat. Pemahaman kita selanjutnya tentang bagaimana jawaban-jawaban itu didistribusikan dapat diperoleh dengan menghitung frekuensi relatif bagi tiap-tiap selnya. Untuk ini, frekuensi tiap sel kita bagi dengan ukuran sampel 400. Misalnya, $71/400 = 0{,}1775$ adalah frekuensi relatif mahasiswa putri yang mendukung. Tabel 1.2 menunjukkan frekuensi relatif tiap-tiap sel.

Tabel 1.2

Mahasiswa	Mendukung	Menentang	Tidak Berpendapat	Jumlah
Putri	0,1775	0,0500	0,0825	0,3100
Putra	0,2800	0,2100	0,2000	0,6900
Jumlah	0,4575	0,2600	0,2825	1,0000

Bergantung pada konteks suatu tabulasi silang, mungkin kita juga ingin mempelajari frekuensi relatif sel terhadap jumlah tepi (kanan atau bawah). Dalam contoh di atas, mungkin kita ingin membandingkan pola setiap mahasiswa putri dan mahasiswa putra. Ini diperoleh dengan menghitung frekuensi relatif secara terpisah untuk dua kelompok itu, seperti ditunjukkan dalam Tabel 1.3 (misalnya $71/124 = 0{,}5726$).

Tabel 1.3

Mahasiswa	Mendukung	Menentang	Tidak Berpendapat	Jumlah
Putri	0,5726	0,1613	0,2661	1,0000
Putra	0,4058	0,3043	0,2899	1,0000
Jumlah	0,9784	0,4656	0,5560	2,0000

Dari Tabel 1.3 tampak bahwa pola sikap kedua kelompok mahasiswa itu berbeda, yang menentang kelihatannya kelompok mahasiswa putra lebih kuat dari kelompok mahasiswa putri.

Pertanyaan yang penting selanjutnya apakah perbedaan yang kita amati itu karena kebetulan atau memang benar-benar oleh perbedaan sikap yang nyata antara populasi mahasiswa putra dan mahasiswa putri? Jawaban pertanyaan ini akan kita jumpai nanti dalam Modul 5.

B. DIAGRAM TITIK DATA BIVARIAT

Sekarang kita pelajari deskripsi himpunan data tentang dua variabel, masing-masing diukur pada skala numerik. Guna memudahkan menyebutnya, kedua variabel itu akan kita beri nama variabel x dan variabel y. Jadi, observasi numerik dua variabel itu (x, y) dicatat untuk semua unit sampel. Observasi-observasi ini berpasangan, dalam arti bahwa suatu pasangan (x, y) diperoleh dari unit sampel yang sama. Satu observasi x dari suatu pasangan tidak mempunyai hubungan dengan nilai x atau y dari pasangan yang lain. Untuk n unit sampel, kita dapat menuliskan pengukuran (observasi) berpasangan itu sebagai (x_1, y_1), (x_2, y_2), ..., (x_n, y_n).

Himpunan observasi x sendiri, dengan mengabaikan observasi-observasi y, merupakan himpunan data satu variabel. Data seperti ini pernah kita pelajari dalam kuliah Metode Statistik I. Hal serupa akan terjadi jika kita hanya memandang observasi y saja, dan mengabaikan observasi x. Tetapi dalam modul ini kita akan mempelajari data bivariat (berpasangan) untuk menjawab pertanyaan-pertanyaan, seperti berikut.

1. Apakah variabel-variabel itu berhubungan?
2. Seperti apakah bentuk hubungan yang ditunjukkan oleh data?
3. Dapatkah kita ukur eratnya hubungan itu?
4. Dapatkah kita perkirakan nilai satu variabel jika nilai variabel yang lain diketahui?

Jika kita hanya mempelajari observasi x saja, atau observasi y saja maka pertanyaan-pertanyaan, seperti di atas tidak akan dapat dijawab.

Langkah pertama yang penting dalam mempelajari hubungan antara dua variabel adalah menggambar diagram titik data berpasangan itu. Untuk ini, variabel x dituangkan pada sumbu mendatar dan variabel y pada sumbu tegak pada kertas grafik. Selanjutnya observasi berpasangan (x, y) dituangkan dalam kertas grafik itu dan kita peroleh satu titik. Maka hasil seluruhnya adalah titik-titik pada kertas itu dan dinamakan *diagram titik*. Dengan memperhatikan diagram titik, kita akan memperoleh kesan tentang pola kecenderungan titik-titik itu. Misalnya, kita memperoleh kesan bahwa titik-titik itu cenderung berbentuk lurus, atau melengkung, atau tidak berpola sama sekali.

Contoh 1.2

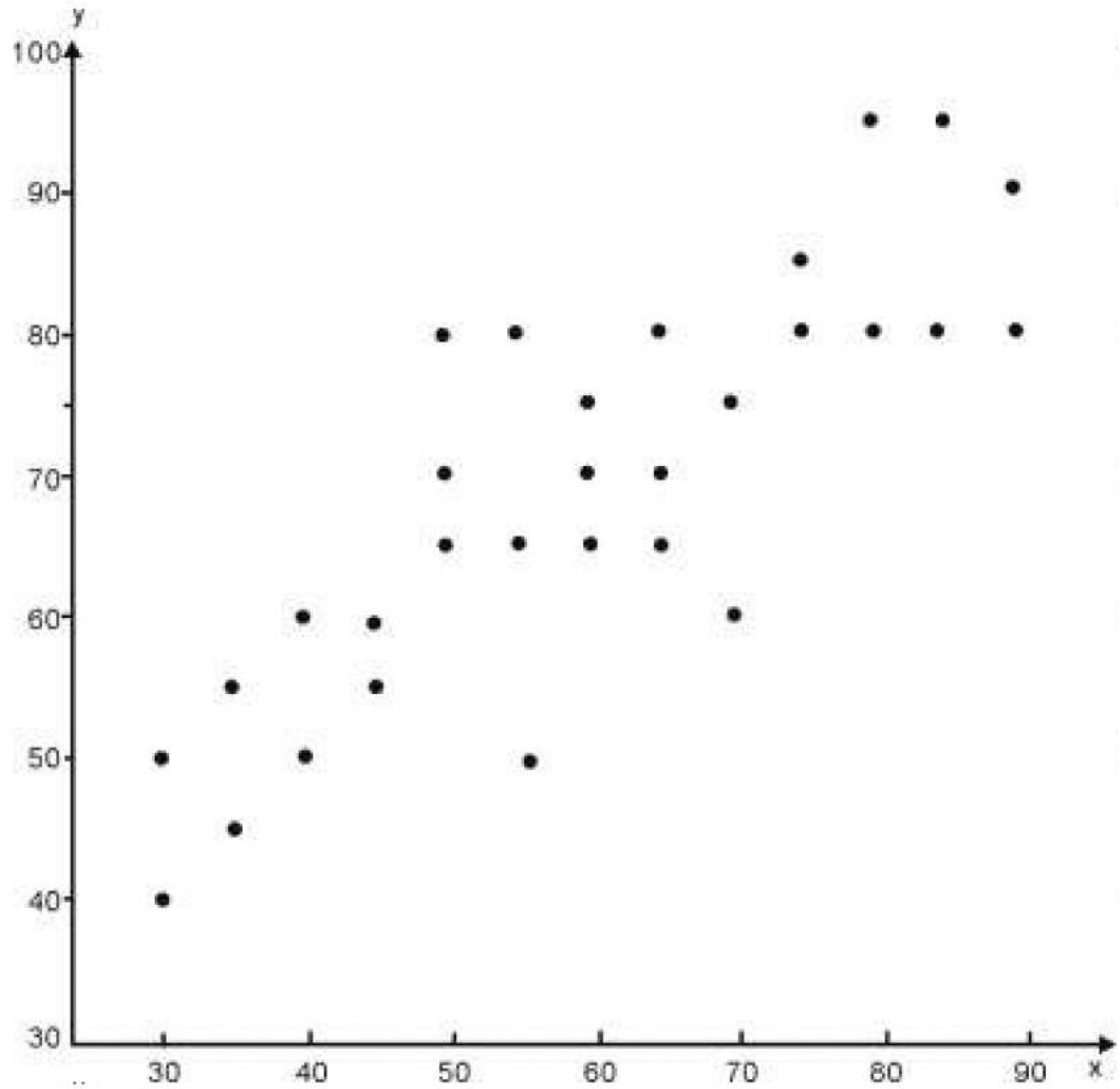
Dalam Tabel 1.4 kita punya data tentang nilai ujian saringan pelamar pekerjaan dan nilai pekerjaan mereka setelah bekerja tiga bulan pertama pada perusahaan HANOR.

x = nilai ujian saringan

y = nilai pekerjaan tiga bulan pertama

Tabel 1.4
Data Nilai Ujian Saringan (x) dan Nilai Pekerjaan (y)

x	y	x	y	x	y
70	60	45	55	45	60
50	65	50	70	30	50
80	95	65	80	35	45
30	40	70	75	55	80
90	90	85	80	60	65
60	70	60	75	75	80
90	80	35	55	75	85
40	60	40	50	80	80
65	65	50	80	85	95
55	50	65	70	55	65



Gambar 1.1
Diagram Titik Data Ujian dalam Tabel 1.4

LATIHAN

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

1) Sakit mual sering kali menyerang orang-orang yang bepergian menggunakan pesawat udara. Perusahaan obat-obatan ingin menentukan efektivitas tablet obat mual buatannya. Untuk ini ia memberikan tablet itu dan tabel serupa yang hanya memuat gula masing-masing kepada 100 orang yang dipilih secara random diperoleh tabel sebagai berikut.

	Tingkat Kemualan				Jumlah
	Tidak	Ringan	Sedang	Berat	
Tablet obat	43	36	18	3	
Bukan obat	19	33	36	12	
Jumlah					

a. Lengkapilah jumlah marginalnya, (tepinya)
b. Hitunglah frekuensi relatif secara terpisah untuk masing-masing baris
c. Berilah komentar tentang perbedaan yang tampak dalam respons antara obat dan bukan obat.

2) Catatan tentang pengemudi dengan kondisi kesehatan utama (gula, jantung, dan epilepsi) dan juga kelompok pengemudi yang tidak diketahui kondisi kesehatannya diperoleh dari kantor polisi lalu lintas. Pengemudi dalam tiap kelompok itu diklasifikasi menurut catatan pengemudi tahun lalu diperoleh tabel sebagai berikut.

Kondisi kesehatan	Pelanggaran Lalu-Lintas		Jumlah
	Tidak Pernah	Satu atau Lebih	
Gula	119	41	160
Jantung	121	39	160
Epilepsi	72	78	150
Sehat (control)	157	43	200

Bandingkan tiap-tiap kondisi kesehatan dengan kelompok kontrol dengan menghitung frekuensi relatif yang sesuai.

3) *Interview* terhadap 185 orang yang memangku jabatan dengan urusan-urusan yang sangat pelik dan dapat membuat stress mengungkapkan fakta bahwa 76 orang adalah alkoholik (peminum), 81 orang mental tertekan, dan 54 orang yang alkoholik dan tertekan.
a. Berdasarkan catatan itu, lengkapilah tabel frekuensi dua arah di bawah.
b. Hitunglah frekuensi relatifnya.

	Alkoholik	Bukan alkoholik	Jumlah
Tertekan			
Tidak tertekan			

4) Suatu survei dilakukan untuk mempelajari sikap staf pengajar, karyawan, dan mahasiswa terhadap suatu usulan proyek pembangunan di kampus. Diperoleh fakta sebagai berikut.

	Setuju	Tidak berpendapat	Menentang	Jumlah
Staf pengajar	36	42	122	
Karyawan	44	77	129	
Mahasiswa	106	178	116	

a. Hitunglah jumlah tepi!
b. Ubahlah frekuensi-frekuensi itu menjadi frekuensi relatif!
c. Hitunglah frekuensi relatif secara terpisah untuk tiap-tiap baris!

5) Dipunyai skor dalam mata kuliah matematika (X) dan mata kuliah ekonomi (Y) 36 orang mahasiswa sebagai berikut.

x	41	39	53	67	61	46	50	55	72	63	59	67
y	29	19	30	27	28	27	22	29	24	33	25	29
x	53	62	65	48	32	64	59	54	52	64	51	62
y	28	22	27	22	27	28	30	29	21	36	20	29
x	56	38	52	40	65	61	64	64	53	51	58	65
y	34	21	25	24	32	29	27	26	24	25	34	28

Gambarkan diagram titik data itu!

6) Dipunyai data tentang kandungan magnesium (x) air sumber dan nilai rasa (y) air itu yang diambil dari delapan lokasi sumber (x dalam mg per liter).

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

Gambarkan diagram titik data itu!

RANGKUMAN

Data klasifikasi silang dapat digambarkan dengan menghitung frekuensi relatif.

Diagram titik memberikan impresi visual tentang hubungan antara dua variabel, apakah gerombolan titik-titik memberi kesan garis lurus, atau garis lengkung, atau tidak berbentuk garis apa pun.

TES FORMATIF 1

Pilihlah satu jawaban yang paling tepat!

1) Satu kelompok 1083 sukarelawan yang berisiko tinggi diikutkan dalam percobaan klinis guna menguji vaksin baru untuk hepatitis B. Vaksin itu diberikan kepada 549 orang yang dipilih secara random dari kelompok di atas, sedangkan sisanya diberi injeksi bahan netral yang pada dasarnya tidak berbahaya. Sebelas orang yang divaksinasi di kemudian hari terserang penyakit itu. Fakta ini dituangkan dalam tabel sebagai berikut.

	Terserang hepatitis	Tidak terserang hepatitis	Jumlah
Divaksinasi	x_1	x_2	x_3
Tidak divaksinasi	x_4	x_5	x_6
Jumlah	x_7	x_8	x_9

a. Maka, x_1 sama dengan ….
   A. 11
   B. 70
   C. 549
   D. 1083

b. Maka, x_6 sama dengan ….
   A. 464
   B. 534
   C. 549
   D. 1083

c. Maka, x_8 sama dengan ….
   A. 534
   B. 549
   C. 1002
   D. 1083

2) Lihat kembali data soal nomor 1.
   Kita hitung frekuensi relatif data dalam tabel di atas. Kita peroleh

	Terserang Hepatitis	**Tidak terserang Hepatitis**	**Jumlah**
Divaksinasi	y_1	y_2	y_3
Tidak divaksinasi	y_4	y_5	y_6
Jumlah	y_7	y_8	y_9

a. Maka, y_2 sama dengan ….
   E. 0,40
   F. 0,45
   G. 0,50
   H. 0,55
b. Maka, y_4 sama dengan ….
   A. 0,06
   B. 0,36
   C. 0,67
   D. 0,92

c. Maka, y_6 sama dengan ….
   A. 0,29
   B. 0,36

   C. 0,41
   D. 0,49

3) Pandang kembali data soal nomor 1.
   Kita hitung frekuensi relatif secara terpisah bagi tiap-tiap baris. Kita peroleh

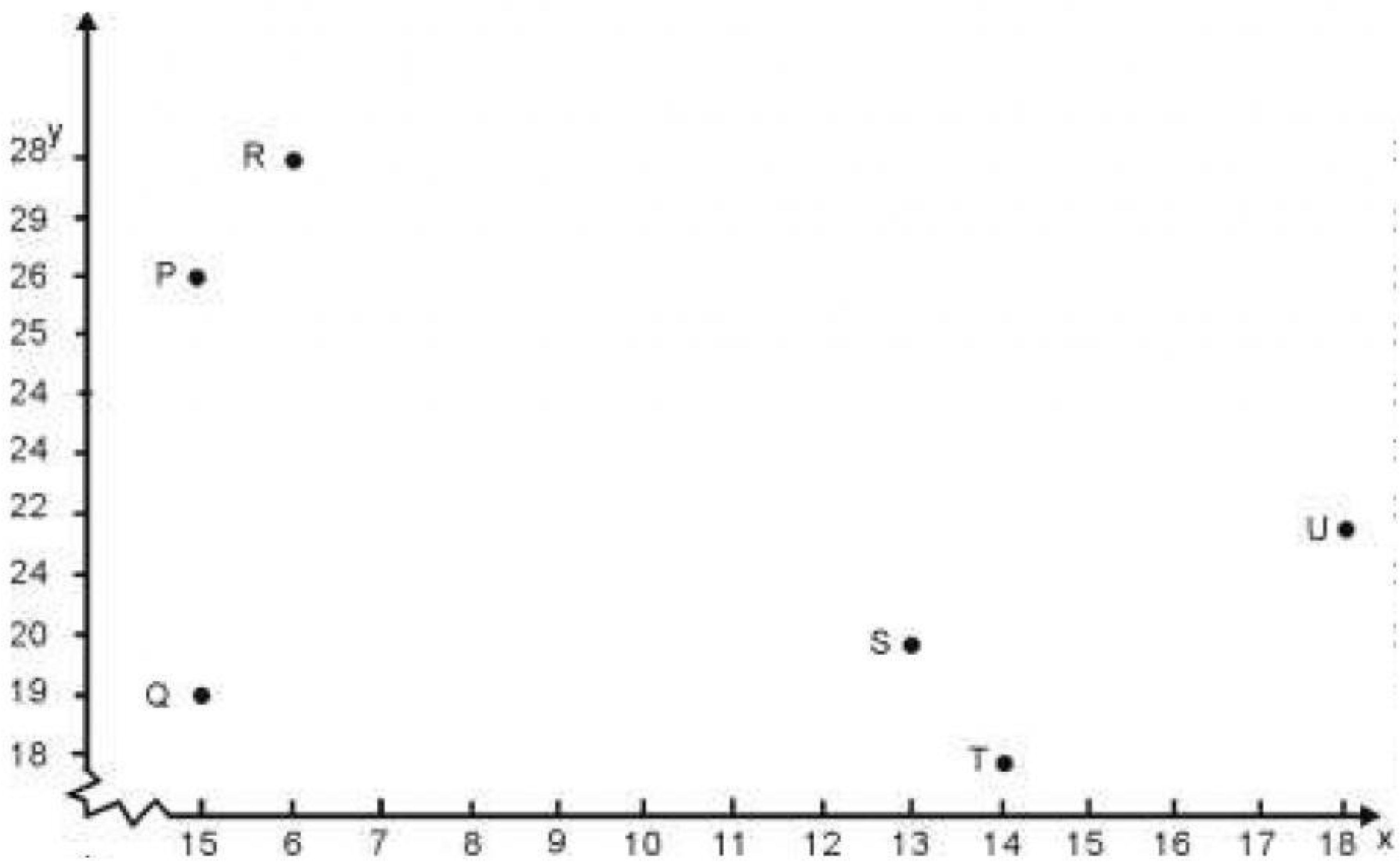
	Terserang Hepatitis	Tidak Terserang Hepatitis	Jumlah
Divaksinasi Tidak divaksinasi	z_1 z_4	z_2 z_5	z_3 z_6
Jumlah	z_7	z_8	z_9

a) Maka, z_1 sama dengan ….
   A. 0,02
   B. 0,22
   C. 0,42
   D. 0,62

b) Maka, z_5 sama dengan ….
   A. 0,14
   B. 0,35
   C. 0,56
   D. 0,87

4) Dipunyai data berpasangan (x; y) sebagai berikut.

x	6	7	5	18	13	5	13	14
y	28	23	26	22	20	19	28	18



b. Maka, titik P menggambarkan titik (x, y) dengan ….
   A. $x = 6$

$y = 28$
B. $x = 5$
$y = 26$
C. $x = 4$
$y = 22$
D. $x = 4$
$y = 19$

c. Maka, titik S menggambarkan titik (x, y) dengan ….
A. $x = 5$
$y = 18$
B. $x = 5$
$y = 19$
C. $x = 13$
$y = 20$
D. $x = 20$
$y = 15$

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.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{10} \times 100\%$$

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

Data Bivariat Kontinu

A. KOEFISIEN KORELASI

Diagram titik memberikan kesan visual tentang hubungan antara nilai-nilai x dan y dalam himpunan data bivariat. Sangat sering titik-titik dalam diagram itu tampak menggerombol di sekeliling garis lurus, tetapi dalam banyak hal kemungkinan berpencar-pencarnya titik-titik itu menyediakan gambaran hubungan itu bukan hubungan linear yang tegas. Kesan visual kita tentang dekatnya titik-titik ke hubungan linear dapat dikuantifikasikan dengan menghitung ukuran numerik yang dinamakan *koefisien korelasi*.

Koefisien korelasi, ditulis dengan lambang r, adalah ukuran kuatnya hubungan linear antara variabel x dan y. Sebelum kita kenalkan rumusnya, terlebih dahulu kita jelaskan secara garis besar beberapa sifat penting koefisien korelasi itu. Kita bicarakan juga bagaimana koefisien korelasi itu mengukur kuatnya hubungan linear.

1. Nilai r selalu antara -1 dan +1
2. Besarnya r mutlak menunjukkan kuatnya hubungan linear, sedangkan tanda aljabarnya menunjukkan arah hubungan itu. Artinya,

   $r > 0$ jika pola nilai-nilai (x, y) berkecenderungan dari kiri ke kanan naik.

   $r < 0$ jika pola nilai-nilai (x, y) berkecenderungan dari kiri ke kanan turun.

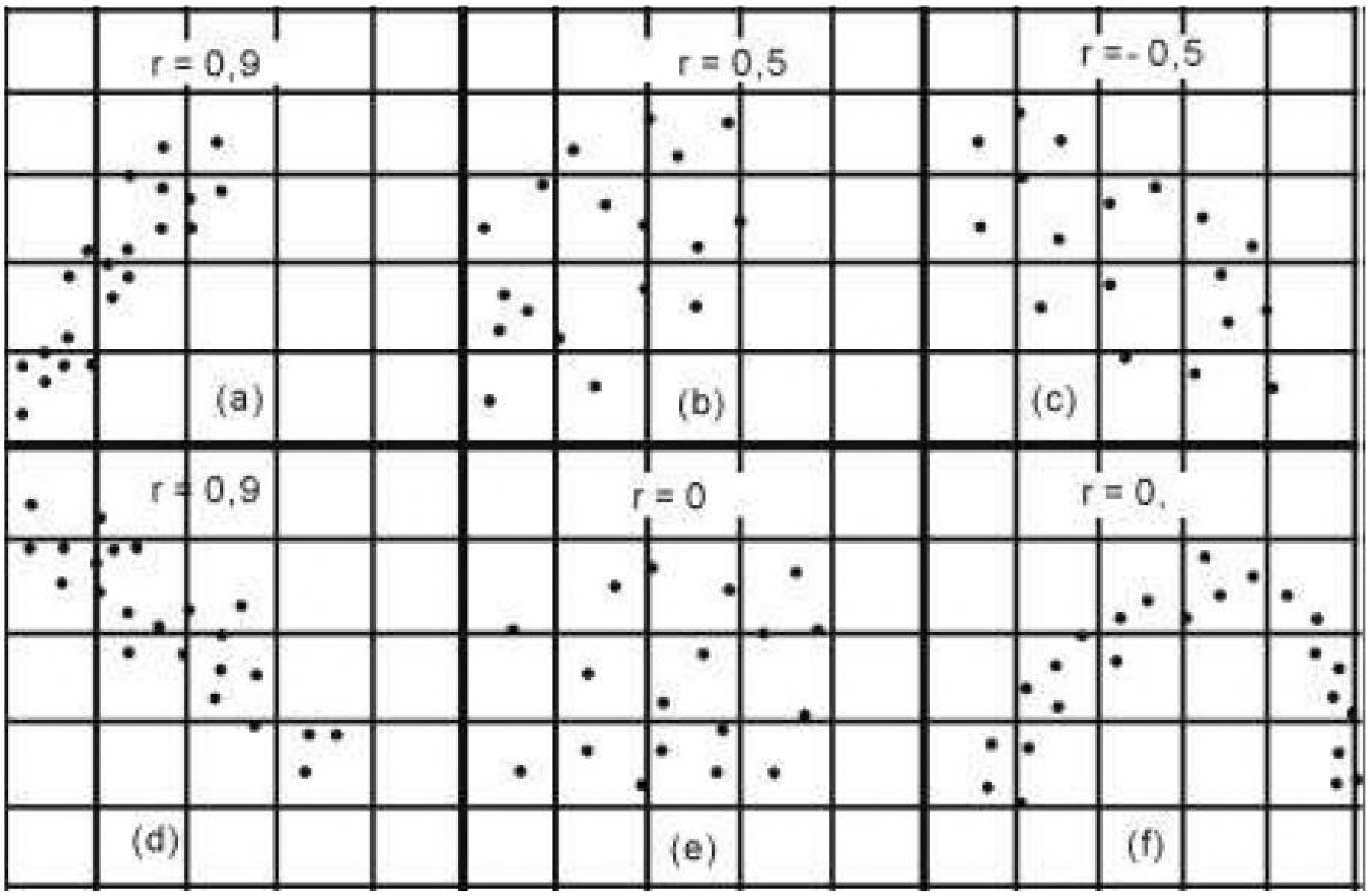
   $r = +1$ jika semua nilai-nilai (x, y) terletak tepat pada garis lurus dengan lerengan positif (hubungan linear positif sempurna).

   $r = -1$ jika semua nilai-nilai (x, y) terletak tepat pada garis lurus dengan lerengan negatif (hubungan linear negatif sempurna).

   Nilai r numerik mutlak yang tinggi, yakni nilai yang dekat dengan +1 atau -1 menunjukkan adanya hubungan linear yang kuat.
3. Nilai r yang dekat dengan nol berarti hubungan linear itu sangat lemah

Koefisien korelasi akan dekat dengan nol jika diagram titik tidak menampakkan adanya pola hubungan yang jelas; yakni, nilai-nilai y tidak menunjukkan perubahan ke sesuatu arah dengan berubahnya nilai-nilai x. Nilai r dekat dengan nol dapat juga terjadi karena titik-titik pada diagram cenderung bergerak melengkung, yakni tidak linear. Memang, r mengukur hubungan linear, dan kurva yang jelas melengkung berarti jauh dari linear.

Gambar 1.2 e dan Gambar 1.2 f berkaitan dengan situasi di mana $r = 0$. Korelasi nol dalam Gambar 1.2 e disebabkan karena tidak adanya hubungan antara x dan y, sedangkan dalam Gambar 1.2 f ini disebabkan hubungannya mengikuti kurva melengkung yang jauh dari linear.



Gambar 1.2.
Diagram Pencar Menunjukkan Berbagai Koefisien Korelasi, r

Menghitung *r*.

Nilai r dihitung dari n pasang observasi (x, y) menggunakan rumus sebagai berikut.

Koefisien korelasi:

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

dengan $S_{xy} = \sum (x - \overline{x})(y - \overline{y})$

$$S_{xx} = \sum (x - \overline{x})^2$$

$$S_{xy} = \sum (y - \overline{y})^2$$

Kuantitas S_{xx} dan S_{yy} adalah masing-masing jumlah kuadrat deviasi observasi x terhadap *mean*-nya, dan deviasi observasi y terhadap mean-nya. Sedang S_{xy} adalah jumlah hasil kali deviasi observasi x terhadap mean-nya dan deviasi observasi y terhadap mean-nya.

Contoh 1.3

Hitunglah r untuk $n = 4$ pasang observasi:
(2 ; 5), (1 ; 3), (5 ; 6), (0 ; 2)

Pertama-tama kita hitung mean $\overline{x}$ dan deviasi $(x - \overline{x})$, selanjutnya mean $\overline{y}$ dan deviasi $(y - \overline{y})$, dan seterusnya. Lihat Tabel 1.5.

Tabel 1.5
Menghitung r

x	**y**	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
2	5	0	1	0	1	0
1	3	-1	-1	1	1	1
5	6	3	2	9	4	6
0	2	-2	-2	4	4	4
Jumlah 8 $\overline{x} = 2$	16 $\overline{y} = 4$	0	0	14 S_{xx}	10 S_{yy}	11 S_{xy}

Maka,

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{11}{\sqrt{14}\sqrt{10}} = 0{,}93$$

Untuk menghitung r ini sering kali lebih baik jika kita gunakan rumus alternatif untuk S_{xy}, S_{xy}, dan S_{yy} sebagai berikut.

$$S_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

$$S_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$

$$S_{xy} = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}$$

Menghitung r dengan rumus di atas ini kita tunjukkan dalam Tabel 1.6.

Tabel 1.6
Menghitung *r* (cara lain)

x	***y***	x^2	y^2	xy
2	5	4	25	10
1	3	1	9	3
5	6	25	36	30
0	2	0	4	0
Jumlah 8 $\sum x$	16 $\sum y$	30 $\sum x^2$	74 $\sum y^2$	43 $\sum xy$

Maka,

$$r = \frac{43 - \frac{(8)(16)}{4}}{\sqrt{30 - \frac{8^2}{4}}\sqrt{74 - \frac{16^2}{4}}} = 0{,}93$$

Contoh 1.4

Kita hitung r untuk data dalam tabel berikut ini.

Tabel 1.7
Menghitung r

x	y	x^2	y^2	xy
1	6	1	36	6
3	4	9	16	12
2	3	4	9	6
4	1	16	1	4
5	2	25	4	10
Jumlah 15 $\sum x$	16 $\sum y$	55 $\sum x^2$	66 $\sum y^2$	38 $\sum xy$

Maka,

$$r = \frac{38 - \frac{(15)(16)}{5}}{\sqrt{55 - \frac{15^2}{5}}\sqrt{66 - \frac{16^2}{5}}} = \frac{-10}{(3{,}162)(3{,}847)} = -0{,}822$$

Sampai di sini kita ingin mengingatkan bahwa r mengukur dekatnya pola titik-titik terhadap suatu garis lurus. Gambar 1.2 f menunjukkan hubungan yang kuat antara x dan y, tetapi hubungan itu tidak linear. Nilai r yang kecil untuk data ini tidak mencerminkan secara wajar kuatnya hubungan. Jelas bahwa r bukan ukuran yang cocok untuk pola kurva. Contoh lain yang menggambarkan bahwa koefisien korelasi sampel r bukan ukuran yang sesuai terjadi jika diagram titik terbagi menjadi dua kelompok titik-titik. Dihadapkan dengan dua kelompok titik-titik yang terpisah seperti dilukiskan dalam Gambar 1.3, hal terbaik yang harus kita lakukan adalah mencoba dan menentukan sebab yang menyebabkan keadaan seperti itu. Mungkin sekali bahwa satu bagian dari titik-titik sampel itu diambil dari satu populasi dan bagian lain dari titik-titik itu dari populasi yang lain.

Gambar 1.3
Koefisien *r* yang Tidak Sesuai Sampel dari Dua Populasi

Korelasi dan Sebab Akibat

Penganalisis data sering kali terjebak pada kesimpulan-kesimpulan yang tidak benar karena salah mengamati korelasi untuk hubungan sebab akibat. Nilai koefisien korelasi sampel yang tinggi tidak harus berarti adanya hubungan sebab akibat antara 2 variabel. Suatu contoh klasik tentang adanya korelasi positif yang tinggi yang diamati antara banyak bangau yang tampak dan banyak kelahiran di suatu kota di Eropa. Tentu saja kita berharap bahwa tidak seorang pun akan menggunakan kenyataan ini untuk menyimpulkan bahwa bangau membawa bayi atau bahkan, lebih gawat lagi bahwa membunuh bangau akan mengendalikan pertumbuhan penduduk.

Mengamati bahwa dua variabel bersama-sama cenderung berubah-ubah ke arah tertentu tidak berarti ada hubungan langsung di antara mereka. Jika kita catat banyak pembunuhan bulanan x dan banyak pertemuan keagamaan bulanan y untuk beberapa kota yang berbeda-beda besarnya, data itu mungkin menunjukkan korelasi positif yang tinggi. Sebenarnya, berubah-ubahnya variabel ketiga (yakni, banyak penduduk kota itu) yang menyebabkan x dan y berubah-ubah dalam arah yang sama meskipun kenyataannya x dan y mungkin tidak berhubungan atau bahkan berhubungan secara negatif. Gambarnya, variabel ketiga yang dalam contoh ini menyebabkan tampak ada hubungan antara banyak kriminal dan pertemuan keagamaan dinamakan variabel tersembunyi. Korelasi yang salah yang dihasilkannya dinamakan korelasi palsu. Lebih bersifat akal sehat daripada alasan statistik untuk menentukan apakah korelasi yang tampaknya ada itu benar-benar praktis ada atau hanya palsu saja.

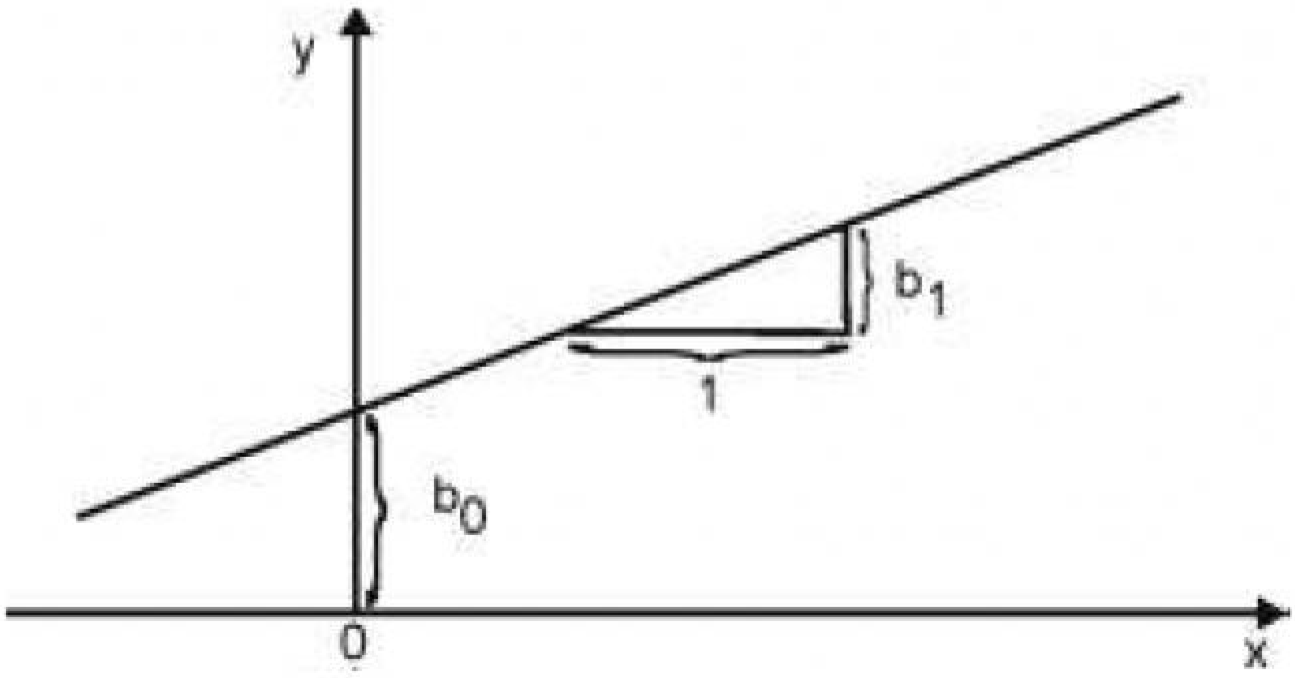
> Peringatan:
> Korelasi yang tampak ada antara dua variabel mungkin palsu. Yakni, mungkin disebabkan karena pengaruh variabel ketiga.

Jika kita menggunakan koefisien korelasi sebagai ukuran hubungan, kita harus berhati-hati untuk menghindarkan kemungkinan adanya variabel tersembunyi yang mempengaruhi salah satu atau kedua variabel yang kita pelajari.

B. PRAKIRAAN GARIS LURUS

Studi eksperimental hubungan antara dua variabel sering kali didorong oleh kebutuhan untuk memperkirakan satu variabel dari variabel yang lain. Pengelola suatu program pelatihan kerja mungkin ingin mempelajari hubungan antara lama pelatihan dan skor peserta pada ujian keterampilan setelah pelatihan selesai. Seorang pengelola hutan mungkin ingin menaksir volume kayu suatu pohon dari pengukuran diameter batang pohon itu beberapa meter di atas tanah. Seorang teknolog kesehatan mungkin tertarik untuk memperkirakan pengukuran alkohol darah dari membaca skala alat baru penganalisis nafas.

Dalam konteks seperti itu, variabel prediktor atau masukan ditulis sebagai x, dan variabel respons atau luaran dilambangkan dengan y. Tujuannya adalah untuk menentukan sifat hubungan antara x dan y dari data eksperimental, dan menggunakan hubungan itu untuk memperkirakan variabel respons y dari variabel prediktor x. Tentu saja, langkah pertama dalam studi semacam ini adalah menggambar diagram titik dan memeriksanya. Jika tampak (terkesan) ada hubungan linear, menghitung nilai numerik r akan menegaskan kekuatan hubungan linear itu. Nilai menunjukkan seberapa efektif nilai y dapat diperkirakan dari nilai x dengan menggunakan garis lurus yang dihitung dari data. Sebuah persamaan garis ditentukan oleh dua konstan. Tingginya di atas titik nol (*intercept*), dan besarnya kenaikan dalam y jika nilai x naik satu satuan (lerengan). Lihat Gambar 1.4 sebagai ilustrasi. Dalam modul-modul mendatang akan kita pelajari secara lebih dalam metode kuadrat terkecil yang menghasilkan rumus-rumus untuk menghitung persamaan garis. Setelah persamaan garis kita peroleh, selanjutnya kita gunakan untuk memperkirakan nilai y berdasarkan nilai x.



Gambar 1.4
Garis $\hat{Y} = b_0 + b_1 x$

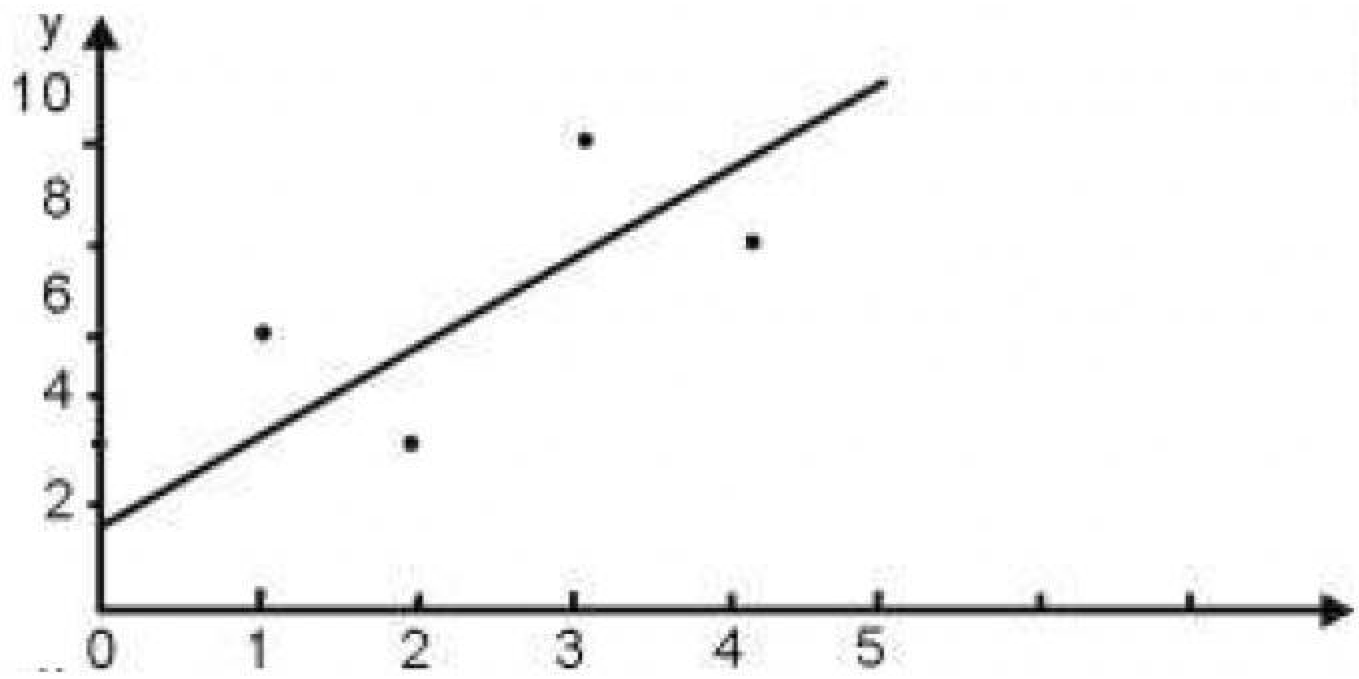
> Persamaan garis:
>
> $$\hat{Y} = b_0 + b_1 x$$
>
> dengan
>
> $$\textit{lerengan}:\ b_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x-\overline{x})(y-\overline{y})}{\sum (x-\overline{x})^2}$$
>
> $$\textit{intercept}:\ b_0 = \overline{y} - b_1 \overline{x}$$

Contoh 1.5

Seorang ahli kimia ingin mempelajari hubungan antara waktu keringnya suatu cat dan konsentrasi pelarut dasar yang memberikan aplikasi yang halus. Data ukuran konsentrasi (x) dan waktu pengeringan pengamatan (y) dicatat dalam dua kolom pertama Tabel 1.8.

Tabel 1.8
Data Konsentrasi *x* dan Waktu Pengeringan *y*
(dalam menit) dan hitungan-hitungan Dasar

x	y	x^2	y^2	xy
0	1	0	1	0
1	5	1	25	5
2	3	4	9	6
3	9	9	81	27
4	7	16	49	28
Jumlah 10	15	30	165	66



Gambar 1.5
Diagram Titik Data Konsentrasi *x* dan Waktu Pengeringan *y*

Diagram titik dalam Gambar 1.5 memberikan kesan adanya hubungan linear. Untuk menghitung r dan menentukan persamaan garis, pertama-tama kita hitung kuantitas dasar $\overline{x}, \overline{y}, S_{xx}, S_{yy}$, dan S_{xy} dalam Tabel 1.8.

$$\overline{x} = \frac{10}{5} = 2 \quad ; \qquad \overline{y} = \frac{25}{5} = 5$$

$$S_{xx} = 30 - \frac{10^2}{5} = 10$$

$$S_{yy} = 165 - \frac{25^2}{5} = 40$$

$$S_{xy} = 66 - \frac{(10)(25)}{5} = 16$$

$$r = \frac{16}{\sqrt{(10)(40)}} = \frac{16}{20} = 0,8$$

$$b_1 = \frac{16}{10} = 1,6$$

$$b_0 = 5 - (1,6)2 = 1,8$$

Maka, persamaan garisnya adalah

$$\hat{y} = 1,8 + 1,6x$$

Garis ini juga digambarkan pada diagram titik dalam Gambar 1.5.

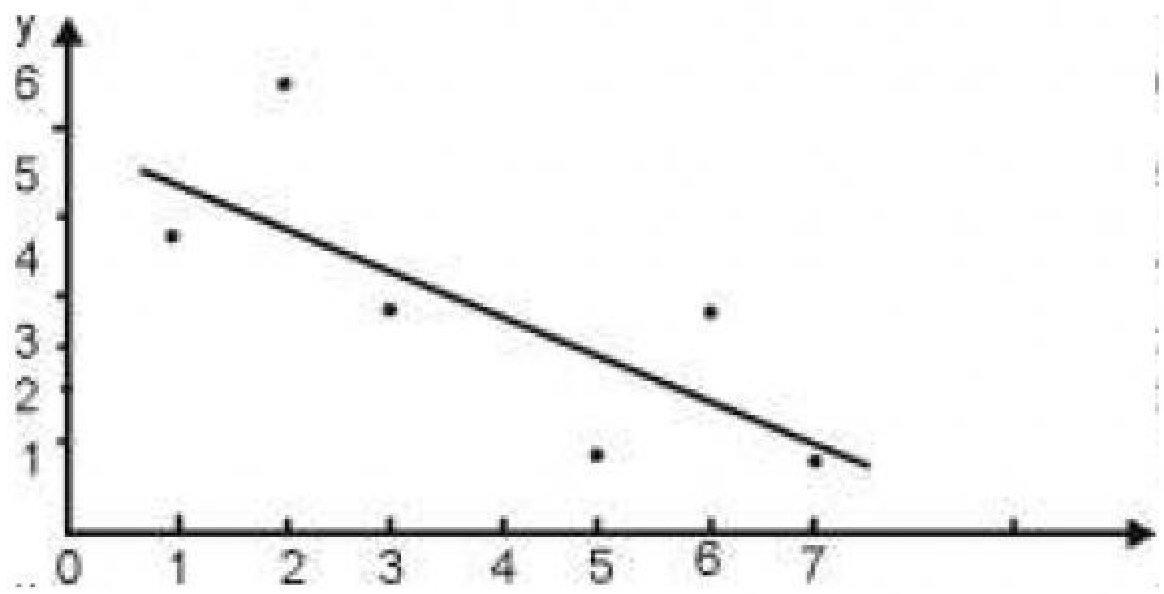
Jika kita harus memperkirakan waktu pengeringan y yang berkaitan dengan konsentrasi 2,5 maka kita substitusikan nilai x = 2,5 ini dalam persamaan garis perkiraan maka kita peroleh

$$\hat{y} = 1,8 + (1,6)(2,5) = 5,8 \text{ menit}$$

Dengan grafik, kuantitas ini kita peroleh dengan membaca ordinat titik pada garis vertikal di atas x = 2,5

Contoh 1.6

Dipunyai enam pasang observasi (x, y), seperti dalam kolom pertama dan kedua Tabel di bawah. Akan kita hitung garis perkiraannya serta nilai perkiraannya untuk x = 5,5



x	y	x^2	y^2	xy
1	4	1	16	4
2	6	4	36	12
3	3	9	9	9
5	1	25	1	5
6	3	36	9	18
7	1	49	1	7
Jumlah 24 $\sum x$	18 $\sum y$	124 $\sum x^2$	72 $\sum y^2$	55 $\sum xy$

Maka,

$$b_1 = \frac{55 - \frac{(24)(18)}{6}}{124 - \frac{(24)^2}{6}} = \frac{-17}{28} = -0,61$$

$$b_0 = \frac{18}{6} - (-0,61)\left(\frac{24}{6}\right) = 5,44$$

$$r = \frac{55 - \frac{(24)(18)}{6}}{\sqrt{\left(124 - \frac{(24)^2}{6}\right)\left(72 - \frac{(18)^2}{6}\right)}} = \frac{-17}{22,45} = -0,76$$

Sehingga

$$\hat{y} = 5,44 - 0,61x$$

Untuk x = 5,5 ; $\hat{y} = 5{,}44 - (0{,}61)(5{,}5) = 2{,}085$

LATIHAN

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

1) Gambarkan garis lurus $y = 10 - 3x$ dengan pertama-tama menentukan titik-titik untuk x = 0 dan x = 3. Berapakah *intercept* dan berapa pula lerengannya?
2) Gambarkan diagram titik untuk dua himpunan data di bawah ini.

(i)

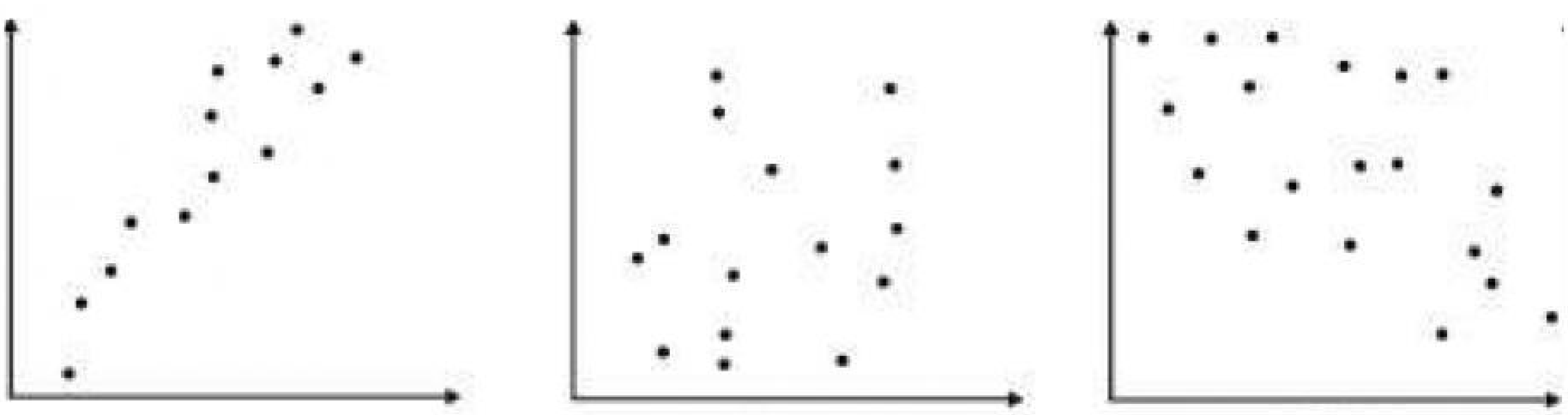
x	-1	3	1	5	2
$f(x)$	2	4	0	6	3

(ii)

x	-1	3	1	5	2
$f(x)$	6	0	3	2	4

a) Hitunglah r untuk himpunan data (i)
b) Tebaklah nilai r untuk himpunan data (ii), kemudian hitunglah r. (Catatan: Untuk kedua himpunan itu nilai-nilai x dan y sama, tetapi nilai-nilai itu berpasangannya berbeda-beda)

3) Pasangkan nilai-nilai r ini dengan grafik titik-titik di bawah secara benar.
   a. r = -0,3
   b. r = 0,1 ; dan
   c. r = 0,9



(a) (b) (c)

4) Penghitungan dari himpunan data dengan n = 48 pasang nilai-nilai (x, y) memberikan hasil-hasil sebagai berikut:

$$\sum\left(x-\overline{x}\right)^2 = 260,2\,;\ \sum\left(y-\overline{y}\right)^2 = 403,7\,;\ \sum\left(x-x\right)\left(y-\overline{y}\right) = 298,8$$

Hitunglah koefisien korelasinya!

5) Untuk himpunan data berpasangan (x, y) kita peroleh hasil hitungan sebagai berikut: n = 26; $\sum x = 1287$; $\sum x^2 = 66831$; $\sum y = 1207$; $\sum y^2 = 59059$; $\sum xy = 62262$.

Hitunglah koefisien korelasinya!

6) Data tinggi (dalam inci) dan berat (dalam *pound*) tujuh belas Miss Amerika dicatat sebagai berikut:

Tinggi	65	67	66	65,5	65	66,5	66	67	66
Berat	114	120	116	118	115	124	124	115	116

Tinggi	69	67	65,5	68	67	68	69	68
Berat	135	125	110	121	118	120	125	119

a. Gambarkan diagram titiknya

b. Hitung koefisien korelasinya

7) Sifat r yang lain.

Misalkan, semua pengukuran x diubah menjadi $x' = ax + b$, dan semua pengukuran y menjadi $y' = cy + d$, di mana a, b, c, dan d bilangan-bilangan tetap. Maka koefisien korelasinya tidak berubah jika a dan c mempunyai tanda aljabar yang sama; koefisien korelasi akan berubah tandanya, jika a dan c mempunyai tanda aljabar yang berlainan.

Sifat r ini dapat dijelaskan sebagai berikut. Jika x berbuah menjadi ax + b maka deviasi $\left(x-\overline{x}\right)$ berubah menjadi $a\left(x-\overline{x}\right)$; demikian juga jika y berubah menjadi $cy + b$ maka deviasi $\left(y-\overline{y}\right)$ berubah menjadi $c\left(y-\overline{y}\right)$. Akibatnya, $\sqrt{S_{xx}}$, $\sqrt{S_{yy}}$, dan $\sqrt{S_{xy}}$ masing-masing berubah menjadi $|a|\sqrt{S_{xx}}$, $|c|\sqrt{S_{yy}}$, dan acS_{xy}; jadi, r berubah menjadi:

$$\frac{ac}{|a||c|}r = \begin{cases} r & \text{jika } a \text{ dan } c \text{ bertanda aljabar sama} \\ -r & \text{jika } a \text{ dan } c \text{ bertanda aljabar berbeda} \end{cases}$$

a. Untuk melihat sifat r itu dengan contoh angka, pandanglah pasangan nilai (x, y):

x	1	3	2	4	5	6
y	1	1	3	4	6	5

Hitunglah koefisien korelasi r.

b. Hitunglah nilai-nilai x' dan y' dari pasangan nilai-nilai x dan y di atas dengan rumus berikut.

$$x' = 3x - 2 \quad \text{dan} \quad y' = -y + 10$$

Dari pasangan nilai-nilai (x', y') ini hitunglah koefisien korelasinya r'. Bandingkan r dari soal a) dan r' ini.

c. Pandang kembali soal nomor 6. Ubahlah data itu menjadi tinggi dalam satuan sentimeter dan berat dalam kilogram. Selanjutnya, hitunglah koefisien korelasi untuk himpunan, data dalam satuan baru ini.

8) Seorang manajer toko alat-alat mobil menentukan bahwa keuntungan bulanan (y) yang diperoleh dari menjual baterai mobil merek HANOR diberikan dengan rumus:

$$y = 10x - 145$$

dengan x menunjukkan banyak baterai yang terjual dalam satu bulan.

a. Jika ada 45 baterai yang terjual dalam satu bulan, berapakah keuntungan yang diperolehnya?

b. Berapa baterai paling sedikit harus dapat terjual dalam satu bulan supaya toko itu memperoleh keuntungan?

9) Identifikasi variabel prediktor x dan variabel respons y dalam tiap-tiap keadaan berikut ini.

a. Seorang direktur pelatihan ingin mempelajari hubungan antara lama pelatihan karyawan baru dengan kinerja mereka dalam menjalankan pekerjaan

b. Tujuan suatu studi adalah untuk menentukan hubungan antara tingkat *carbon monoxide* dalam sampel daerah para pecandu rokok dengan banyak rata-rata rokok yang mereka isap setiap harinya.

c. Seorang ahli pertanian ingin mempelajari hubungan antara tingkat pertumbuhan cendawan dengan tingkat kelembaban sekelilingnya.

d. Seorang analisis pasar ingin mempelajari hubungan antara biaya untuk promosi suatu produk dalam uji pemasaran dengan besar penjualan produk itu selanjutnya.

10) Dipunyai 5 pasang nilai (x, y) sebagai berikut.

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

a. Gambarkan grafik titik
b. Dari pemeriksaan visual, gambarkan garis lurus yang tampaknya cukup cocok dengan datanya.
c. Hitunglah nilai taksiran b_0 dan b_1, dan gambarkan garis lurus sesuai dengan persamaan taksiran $\hat{Y} = b_0 + b_1 x$

11) Dalam suatu eksperimen yang dirancang untuk mempelajari hubungan antara hasil (y dalam gram) suatu proses kimia dan temperatur (x dalam derajat F) untuk suatu fase reaksi yang penting proses itu, dicatat beberapa statistik sebagai berikut.

$n = 8$; $\sum x = 1278$; $\sum y = 396$

$S_{xx} = 480$; $S_{yy} = 1628$; $S_{xy} = 935$

a. Hitunglah persamaan regresi taksirannya
b. Dengan menggunakan persamaan garis taksiran itu perkirakan hasil y_0 jika temperatur diatur pada $x_0 = 170^0 F$
c. Hitung koefisien korelasinya.

RANGKUMAN

Koefisien korelasi r mengukur seberapa dekat diagram titik mendekati pola garis lurus.

Nilai koefisien korelasi yang positif menunjukkan kecenderungan nilai-nilai x yang besar terjadi dengan nilai-nilai y yang besar, dan untuk nilai-nilai kecil kedua variabel itu terjadi secara bersama-sama juga.

Nilai koefisien korelasi yang negatif menunjukkan kecenderungan nilai-nilai x yang besar terjadi dengan nilai-nilai y yang kecil, dan sebaliknya.

Suatu nilai koefisien korelasi yang tinggi tidak harus berarti adanya hubungan sebab akibat.

Taksiran garis lurus kuadrat terkecil menolong menggambarkan hubungan antara variabel respons *y* dengan variabel prediktor *x*.

Nilai y dapat diperkirakan untuk nilai *x* yang diketahui dengan membaca dari persamaan regresi estimasi

$$\hat{Y} = b_0 + b_1 x$$

Untuk pasangan observasi (x, y):

Koefisien korelasi sampel $r = \dfrac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$

dengan $S_{xx} = \sum (x - \bar{x})^2$; $S_{yy} = \sum (y - \bar{y})^2$; dan $S_{xy} = \sum (x - \bar{x})(y - \bar{y})$.

Persamaan taksiran garis: $\hat{Y} = b_0 + b_1 x$

dengan $b_1 = \dfrac{S_{xy}}{S_{xx}}$ dan $b_0 = \bar{y} - b_1 \bar{x}$.

TES FORMATIF 2

Pilihlah satu jawaban yang paling tepat!

1) Dipunyai delapan pasang nilai (x, y) sebagai berikut.

x	6	7	5	21	13	5	13	14
y	28	23	29	22	20	19	28	19

a. Maka koefisien korelasi *r* sama dengan ….
   A. -0,212
   B. -0,313
   C. 0,212
   D. 0,313

b. Maka, persamaan garis taksiran kuadrat terkecil adalah ….
   A. $\hat{y} = 2,44 - 0,23x$

B. $\hat{y} = 4,24 + 0,32x$

C. $\hat{y} = 4,42 - 2,03x$

D. $\hat{y} = 24,24 - 3,20x$

c. Perkiraan nilai y untuk $x = 8$ adalah ….

A. 0,4

B. 0,5

C. 0,6

D. 0,7

d. Jika x diubah menjadi $x' = 6x - 7$ dan y menjadi $y' = 2x + 7$ maka r berubah menjadi r' sama dengan

A. -0,212

B. -0,313

C. 0,414

D. 0,515

2) Dari himpunan data dengan 20 pasang nilai-nilai (x, y) kita peroleh statistik:

$\sum x = 156$; $\sum y = 1178$; $\sum x^2 = 1262$; $\sum y^2 = 69390$

dan $\sum xy = 9203$

a. Maka, koefisien korelasi r sama dengan ….

A. 0,6

B. 0,7

C. 0,8

D. 0,9

b. Maka, persamaan garis taksiran kuadrat terkecil adalah ….

A. $\hat{y} = 3,126 - 0,124x$

B. $\hat{y} = 2,776 + 0,221x$

C. $\hat{y} = 2,519 + 0,323x$

D. $\hat{y} = 2,124 - 0,427x$

c. Perkiraan nilai y untuk $x = 8$ adalah ….

A. 2,371

B. 5,103

C. 7,214
D. 9,331

d. Jika x diubah menjadi $x' = 2x + 6$ dan y menjadi $y' = -4x - 3$ maka r berubah menjadi r' sama dengan ….
A. 0,6
B. -0,7
C. 0,8
D. -0,9

3) Dipunyai enam pasang nilai (x, y) sebagai berikut.

x	1	2	3	4	5	6
y	8	9	7	10	12	11

a. Maka, sama dengan ….
A. $S_{xx} = 13,5$
B. $S_{xy} = 14,5$
C. $S_{yy} = 16,5$
D. $\hat{y} = 2,312 - 0,212x$

b. Maka, r sama dengan ….
a. 0,55
b. 0,66
c. 0,77
d. 0,88

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.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{10} \times 100\%$$

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

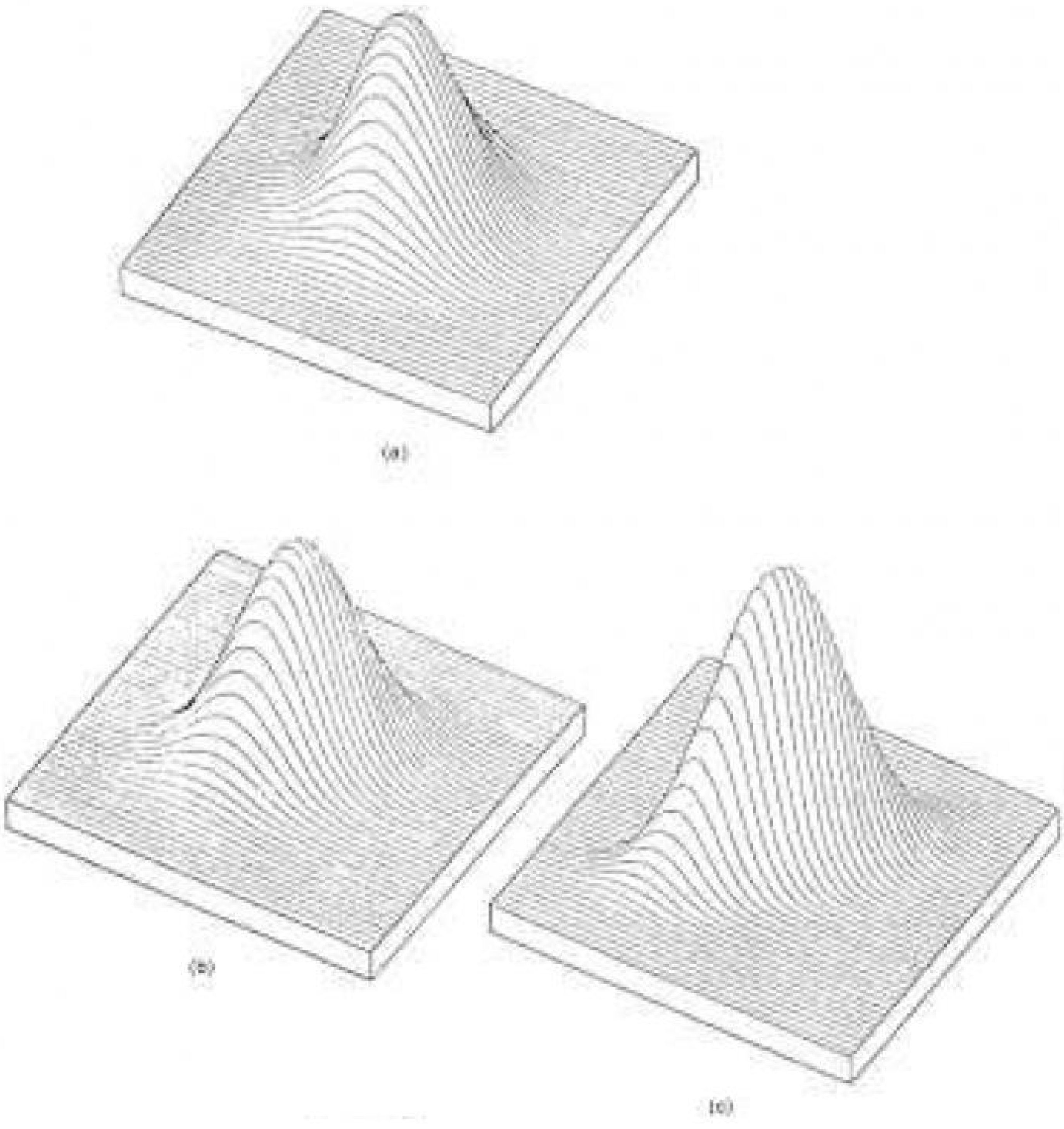
Apabila mencapai tingkat penguasaan 80% atau lebih, Anda dapat meneruskan dengan Kegiatan Belajar 3. **Bagus!** Jika masih di bawah 80%, Anda harus mengulangi materi Kegiatan Belajar 2, terutama bagian yang belum dikuasai.

KEGIATAN BELAJAR 3

Distribusi Normal Bivariat

A. MODEL POPULASI UNTUK KORELASI

Distribusi normal bivariat adalah model populasi yang sangat banyak digunakan dalam mempelajari observasi pada dua variabel random kontinu X dan Y. Dalam model ini masing-masing variabelnya berdistribusi normal. Koefisien korelasi populasinya ρ adalah satu-satunya parameter tambahan dalam distribusi bersama pada dua mean dari dua deviasi standar yang ada pada distribusi marginal X dan Y.



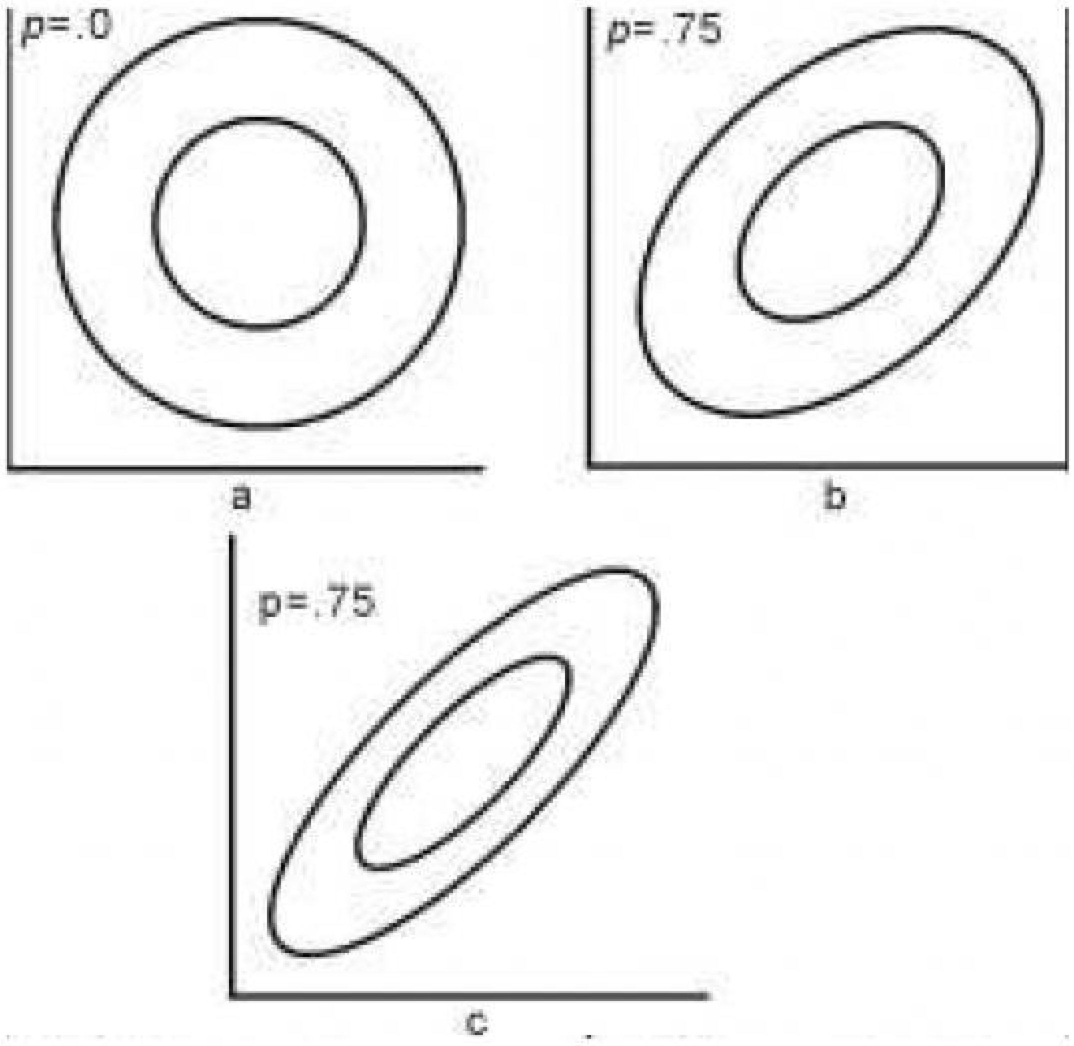
Gambar 1.5
(a) Permukaan Normal Bivariat dengan $\rho = 0$
(b) Permukaan Normal Bivariat dengan $\rho = 0{,}4$
(c) Permukaan Normal Bivariat dengan $\rho = 0{,}75$

Dengan dua variabel random, grafik fungsi peluang bersamanya berbentuk gunungan dalam 3 dimensi yang menggambarkan bagaimana peluang disebarkan pada bidang nilai-nilai (x, y). Rumus matematik untuk fungsi peluang normal bivariat cukup rumit dalam taraf pembicaraan kita di sini, tetapi beberapa ilustrasi akan menarik di sini. Tiga distribusi normal bivariat ditunjukkan dalam Gambar 1.5. Semua distribusi ini mempunyai mean yang sama, dan kedua komponen, X dan Y, mempunyai deviasi standar yang sama. Perbedaan penampilan semata-mata disebabkan karena perbedaan dalam koefisien korelasi populasinya ρ. Dengan nilai ρ yang tinggi (positif atau negatif), permukaan peluang cenderung berbentuk gunungan yang tajam. Tanda aljabar ρ menentukan orientasi gunungan itu; ρ yang positif (negatif) berkaitan dengan lerengan yang positif (negatif) dalam bidang (x, y).

Pandangan lain distribusi normal bivariat diturunkan dengan menentukan letak semua pasangan nilai (x, y) yang mendefinisikan tinggi tertentu fungsi peluang. Dengan perkataan lain, kita cari garis edar di dalam bidang di mana gunungan peluang mempunyai tinggi yang konstan. Garis edar atau kontur berbentuk ellips dengan pusat pada pasangan mean. Gambar 1.6 menunjukkan garis edar yang bersesuaian dengan fungsi peluang dalam Gambar 1.5. Dalam tiap kasus garis edar dalam memuat 50% peluang dan garis edar luar memuat 90% peluang.

Jika suatu distribusi normal bivariat sebagai model populasi maka suatu sampel random diharapkan menyerupai populasi itu. Misalnya, kira-kira 50% dan 90% observasi diharapkan terletak di dalam garis edar yang ditunjukkan dalam Gambar 1.6.

Dalam sampel bivariat satu pertanyaan yang penting ditanyakan adalah apakah kedua variabel random itu berkorelasi atau tidak. Jika populasinya dimodelkan sebagai normal bivariat, tersedia uji untuk $H_0 : \rho = 0$ yang cukup sederhana. Dalam model ini $\rho = 0$ adalah ekuivalen dengan dua variabel itu independen. Statistik uji yang cocok guna menguji independensi dalam model normal bivariat adalah $t = \dfrac{\sqrt{n-2}\, r}{\sqrt{1-r^2}}$.



Gambar 1.6

Kontur 50% dan 90% distribusi Normal Bivariat Bervariansi sama yang berdistribusi t dengan derajat bebas db = n – 2. Dengan alternatif dua sisi, hipotesis nol ini ditolak jika nilai hitungan statistik uji ini lebih besar dari $t_{\alpha/2}$ atau lebih kecil dari $-t_{\alpha/2}$.

Untuk menguji $H_0 : \rho = 0$ *vs* $H_1 : \rho \neq 0$ berdasarkan n pasang observasi dari populasi normal bivariat

Menolak H_0 jika $\left|\dfrac{\sqrt{n-2}\, r}{\sqrt{1-r^2}}\right| > t_{\alpha/2}$ di mana t mempunyai db = n – 2.

Contoh 1.7

Dipunyai data tentang status jabatan ayah (x) dan anak laki-laki (y) dalam skala Duncan sebagai berikut.

x	22	14	14	14	68	12	32	22	19	14	44	18	61
y	13	49	72	44	44	19	17	13	22	14	21	15	66

x	82	14	18	44	32	72	86	26	65	53	14	25	37
y	67	44	13	16	15	40	17	31	65	65	14	25	31

x	53	19	14	15	49	36	21	14	18	53	44	24	87
y	64	17	18	18	47	18	41	15	44	72	37	44	45

x	61	19	44	19
y	19	15	50	41

Kita hitung koefisien korelasi sampel r data di atas, kita peroleh r = 0,412. Jika kita anggap sampel itu diambil dari distribusi normal bivariat maka kita dapat menguji $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$. Dengan mengambil α = 0,01 maka H_0 ditolak jika $t > 2,70$ atau $t < -2,70$. Selanjutnya, kita hitung statistik uji t, kita peroleh

$$t = \frac{\sqrt{n-2}\ r}{\sqrt{1-r^2}} = \frac{\sqrt{41}\,.\,0,412}{\sqrt{0,830}} = 2,90 \quad \text{dengan db} = 41$$

Oleh karena t = 2,90 > 2,70 maka H_0 ditolak.

Guna menguji hipotesis yang lebih umum $H_0 : \rho = \rho_0$ dalam populasi normal bivariat, suatu uji dengan sampel besar didasarkan atas kenyataan bahwa distribusi

$$Z = \sqrt{n-3}\left(\frac{1}{2}\ln\frac{1+r}{1-r} - \frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0}\right)$$

mendekati distribusi normal standar. Dengan alternatif dua sisi dan $\alpha = 0,05$ maka H_0 ditolak jika $|Z| > 1,96$.

LATIHAN

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

1) Dipunyai 15 pasang observasi (x, y) sebagai berikut.

x	58	53	55	44	38	69	40	38	53	54	56	57	58	44	66
y	59	58	47	48	40	61	45	41	47	44	51	58	47	56	62

   a. Hitunglah koefisien korelasi r.
   b. Ujilah $H_0 : \rho = 0$ *vs* $H_1 : \rho \neq 0$ dengan $\alpha = 5\%$.

2) Dipunyai data berpasangan (x, y) dengan x = nilai matematika dan y = nilai ilmu pengetahuan sosial dalam UMPTN beberapa orang calon mahasiswa:

x	56	38	52	40	65	61	64	64	53	51	58	65	53	62	65	48	32	64	59	54
y	34	21	25	24	32	29	27	26	24	25	34	28	28	22	27	22	27	28	30	29

   a. Hitunglah koefisien korelasi r.
   b. Ujilah $H_0 : \rho = 0$ *vs* $H_1 : \rho \neq 0$ dengan $\alpha = 1\%$.

3) Seorang anthropologist mengukur panjang dan lebar tengkorak, dan memperoleh r = 0,82 berdasarkan 27 pasang observasi. Dengan menganggap data diambil dari populasi normal bivariat, ujilah $H_0 : \rho = 0$ *vs* $H_1 : \rho \neq 0$ dengan $\alpha = 5\%$.

4) Seorang ahli teknologi pangan mempelajari pelayanan makan di rumah sakit mendapatkan koefisien korelasi r = 0,68 antara waktu penyiapan dan waktu penyajian berdasarkan n = 11 observasi. Dapatkah waktu penyiapan dan waktu penyajian dipandang independen? Ambil $\alpha = 5\%$.

5) Kita dapat juga menggunakan $Z = \frac{1}{2}\ln\left[(1+r)/(1-r)\right]$ untuk memperoleh interval kepercayaan pendekatan bagi ρ dalam populasi normal. Interval kepercayaan 95% untuk $\frac{1}{2}\ln\left[(1+\rho)/(1-\rho)\right]$ diberikan oleh $\frac{1}{2}\ln\frac{(1+r)}{(1-r)} \pm \frac{1,96}{\sqrt{n-3}}$
dan batas-batas interval ini dapat ditransformasi untuk batas-batas ρ, diperoleh $tgh\left(Z - \frac{1,96}{\sqrt{n-3}}\right) < \rho < tgh\left(Z + \frac{1,96}{\sqrt{n-3}}\right)$
di mana *tgh* adalah fungsi tangen hiperbolik.
Hitunglah interval kepercayaan 95% ρ jika r = 0,70, berdasarkan n = 39 observasi dari populasi normal.

6) Hitunglah interval kepercayaan 90% untuk ρ menggunakan data latihan nomor 2.

RANGKUMAN

Guna menguji $H_0 : \rho = 0$ dalam suatu populasi normal bivariat dapat kita gunakan statistik uji $t = \frac{\sqrt{n-2}\, r}{\sqrt{1-r^2}}$ yang berdistribusi t dengan derajat bebas $n - 2$.

Guna menguji $H_0 : \rho = \rho_0$ dalam suatu populasi normal bivariat dapat kita gunakan statistik uji (untuk sampel besar) $Z = \sqrt{n-3}\left(\frac{1}{2}\ln\frac{1+r}{1-r} - \frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0}\right)$ yang mendekati distribusi normal standar.

TES FORMATIF 3

Pilihlah satu jawaban yang paling tepat!

1) Dipunyai 5 pasang nilai (x, y) sebagai berikut.

x	5	1	4	3	2
y	0	4	2	0	-1

a. Maka, r sama dengan ….
   A. -0,395
   B. -0,921
   C. 0,412
   D. 0,783

b. Guna menguji $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$ kita hitung statistik penguji t, sama dengan ….
   A. -0,745
   B. -0,235
   C. 0,124
   D. 0,437

c) Dengan $\alpha = 5\%$, daerah kritik uji itu adalah ….
   A. $|t| > 2{,}115$
   B. $|t| > 3{,}182$
   C. $|t| > 4{,}621$
   D. $|t| > 5{,}116$

2) Dari 36 pasang observasi (x, y) diperoleh koefisien korelasi $r = 0{,}437$.

a. Untuk menguji $H_0 : \rho = 0$ kita hitung statistik uji t, sama dengan ….
   A. 1,236
   B. 2,833
   C. 3,416
   D. 4,112

b. Untuk menguji $H_0 : \rho = 0,6$ kita hitung statistik uji sampel besar Z, sama dengan ….
   A. 3,16
   B. 2,74
   C. -1,29
   D. -4,27
c. Dengan $\alpha = 5\%$ daerah kritik uji b) di atas adalah ….
   A. $|Z| > 1,64$
   B. $|Z| > 1,96$
   C. $|Z| > 2,12$
   D. $|Z| > 2,33$
d. Interval kepercayaan 95% pendekatan untuk $\frac{1}{2}\ln\left[(1+\rho)/(1-\rho)\right]$ adalah ….
   A. $0,25 \pm 0,21$
   B. $0,29 \pm 0,24$
   C. $0,36 \pm 0,31$
   D. $0,47 \pm 0,34$
e. Interval kepercayaan 95% pendekatan untuk ρ adalah ….
   A. $tgh(0,02) < \rho < tgh(0,47)$
   B. $tgh(0,09) < \rho < tgh(0,66)$
   C. $tgh(0,13) < \rho < tgh(0,81)$
   D. $tgh(0,51) < \rho < tgh(0,92)$

3) Dari n=26 pasang observasi (x, y) diperoleh koefisien korelasi r =0,818.
   a. Untuk $H_0 : \rho \le 0,8$ *vs* $H_1 : \rho > 0,8$ kita hitung statistik uji sampel besar Z, sama dengan ….
      A. 0,14
      B. 0,18
      C. 0,20
      D. 0,25

b. Untuk $\alpha = 1\%$, daerah kritik uji di atas adalah ….
   A. $t > 1,64$
   B. $t > 1,96$
   C. $t > 2,33$
   D. $t > 2,56$

Cocokkanlah jawaban Anda dengan Kunci Jawaban Tes Formatif 3 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar. Kemudian, gunakan rumus berikut untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 3.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{10} \times 100\%$$

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 modul selanjutnya. **Bagus!** Jika masih di bawah 80%, Anda harus mengulangi materi Kegiatan Belajar 3, terutama bagian yang belum dikuasai.

Kunci Jawaban Tes Formatif

Tes Formatif 1

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

Tes Formatif 2

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

Tes Formatif 3

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

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.