

BAB II

MATERI PENUNJANG

2.1 Probabilitas

Berikut dibawah ini akan diberikan beberapa definisi mengenai probabilitas yang ada kaitannya dengan Statistik Nonparametrik:

Definisi 2.1.1

Ruang sampel adalah himpunan dari semua hasil kejadian dalam suatu percobaan

Definisi 2.1.2

Titik sampel adalah himpunan kejadian yang mungkin dalam suatu percobaan.

Contoh 2.1.1

Jika eksperimen atau percobaan adalah melempar dua kali koin maka ruang sampel terdiri dari empat titik yaitu $\{G,G\}$, $\{G,A\}$, $\{A,G\}$, dan $\{A,A\}$.

Definisi 2.1.3

Event (peristiwa) adalah satu atau beberapa himpunan bagian dari ruang sampel

Dalam contoh 2.1.1 kejadian 2 gambar adalah satu titik yaitu $\{G,G\}$ dan kejadian satu angka terdiri dari dua titik yaitu $\{A,G\}$ dan $\{G,A\}$.

Definisi 2.1.4

Jika himpunan kejadian A adalah himpunan kejadian yang berhubungan dengan percobaan dan jika n_A adalah banyaknya himpunan kejadian A yang terjadi dalam suatu percobaan yang diulang sebanyak n kali maka probabilitas kejadian dinyatakan dengan $P(A)$ yang dinotasikan dengan

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n} \quad (2.1.1)$$

Contoh 2.1.2

Dalam percobaan melempar koin homogen sebanyak 3 kali akan menghasilkan delapan titik sampel yaitu $\{G,G,G\}$, $\{G,G,A\}$, $\{G,A,G\}$, $\{G,A,A\}$, $\{A,G,G\}$, $\{A,G,A\}$, $\{A,A,G\}$, $\{A,A,A\}$ yang mempunyai kemungkinan sama, maka probabilitas tiap kejadian tersebut adalah $1/8$, sedang probabilitas munculnya dua gambar = $3/8$.

Definisi 2.1.5

Dua himpunan kejadian A dan B adalah saling bebas jika $P(AB) = P(A) P(B)$

Contoh 2.1.3

Dalam percobaan melempar dua kali koin yang seimbang ada empat titik dalam ruang sampel yang mempunyai peluang sama. Misal A himpunan kejadian G terjadi pada pada lemparan pertama dan B himpunan kejadian G terjadi pada lemparan kedua, maka himpunan kejadian A mempunyai titik-titik $\{G,G\}$ dan $\{G,A\}$, himpunan kejadian B mempunyai titik-titik $\{G,G\}$ dan $\{A,G\}$, dan $\{AB\}$ mempunyai titik $\{G,G\}$, karena $P(A) = 2/4$, $P(B) = 2/4$ dan $P(AB) = 1/4$, maka himpunan kejadian A dan B adalah bebas.

2.2 Variabel Random

Definisi 2.2.1

Jika E sebuah percobaan yang memiliki ruang sampel K dan X adalah sebuah fungsi yang mengaitkan himpunan kejadian pada ruang sampel dengan bilangan riil $X(e)$ untuk setiap hasil $e \in K$, maka X disebut variabel random.

Definisi 2.2.2

Jika E sebuah percobaan yang memiliki ruang sampel K dan X, Y adalah sebuah fungsi yang mengaitkan himpunan kejadian pada ruang sampel dengan bilangan riil $X(e)$ dan $Y(e)$ untuk setiap hasil $e \in K$, maka (X, Y) disebut variabel random berdimensi dua atau variabel random bivariat.

Definisi 2.2.3

Jika nilai yang mungkin dari variabel random X , R_x , adalah himpunan terbatas atau tak terbatas yang dapat dihitung, maka variabel random X disebut sebuah variabel random diskrit.

Definisi 2.2.4

Jika nilai yang mungkin dari variabel random X , R_x , adalah sebuah interval atau kumpulan dari interval-interval, maka variabel random X disebut variabel random kontinu.

Fungsi Probabilitas dan Fungsi Distribusi

Berikut akan diberikan definisi-definisi dari fungsi distribusi dan fungsi probabilitas:

Definisi 2.2.5

Fungsi distribusi variabel random yang biasa dinotasikan dengan $F(x)$ adalah fungsi yang memberikan probabilitas X lebih kecil atau sama dengan bilangan riil x dengan kata lain dapat ditulis

$$F(X) = P(X \leq x) \quad (2.2.1)$$

Jika X diskrit, maka

$$F(X) = \sum_{x_i \leq x} P(X_i) \quad (2.2.2)$$

Jika X kontinu, maka

$$F(X) = \int_{-\infty}^x f(t) dt \quad (2.2.3)$$

Definisi 2.2.6

Fungsi probabilitas dari variabel random X biasa dinyatakan dengan $p(x)$ adalah fungsi yang memberikan probabilitas X sama dengan x , x adalah bilangan riil dengan kata lain :

$$p(x) = P(X=x) \quad (2.2.4)$$

2.3 Populasi, Sampel, dan Statistik

Pada eksperimen-eksperimen selalu didapatkan hasil-hasil yang terdiri dari data-data, sehingga berdasarkan data-data tersebut dapat ditarik suatu kesimpulan yang menyatakan hasil-dari eksperimen tersebut. Guna mencapai hasil yang cukup akurat, maka eksperimen biasanya dilakukan dengan mengambil sampel-sampel dari suatu populasi. Untuk itu diperlukan suatu populasi-sasaran yang akan diuji. Berikut akan didefinisikan populasi sasaran yaitu:

Definisi 2.3.1

Himpunan seluruh anggota yang dibuat dan yang akan diambil kesimpulan disebut populasi sasaran.

Agar mendapatkan informasi atau hasil yang akurat seluruh populasi diperiksa tetapi hal ini tidak praktis, maka digunakan sampel.

Definisi 2.3.2

Sampel adalah himpunan bagian dari populasi

Biasanya diasumsikan sampel adalah sampel random, sampel random dapat diperoleh dengan memberi nomor anggota populasi dari 1 sampai N dan kemudian mengambil n nomor secara acak, maka definisi dari sampel random adalah

Definisi 2.3.4

Sampel adalah sampel random jika tiap sampel mempunyai kemungkinan sama untuk diperoleh.

Istilah random dari definisi 2.3.4 diatas berkaitan dengan metode bagaimana sampel itu diperoleh. Pada dasarnya untuk melihat apakah sampel itu random atau tidak diperhatikan cara memperoleh sampel tersebut. Jika populasi terhingga mempunyai N anggota dan sampel diperoleh tanpa pengembalian maka ada $\binom{N}{n}$ kemungkinan, dengan ukuran sampel n. Jika sampel dengan pengembalian maka ada N^n kemungkinan.

Definisi 2.3.5

Sampel random ukuran n adalah kumpulan n variabel random $X_1, X_2, X_3, \dots, X_n$ yang bebas dan berdistribusi identik.

Tiap-tiap variabel random X_i dalam definisi 2.3.5 dapat berupa variabel random yang terdiri dari satu atau lebih dari satu variabel random.

Sekarang dipunyai dua definisi untuk menggambarkan sampel random. Definisi berhubungan dengan ruang sampel, jika tiap titik sampel disajikan dengan satu titik dalam ruang sampel dan tiap titik mempunyai peluang yang sama sebagai sampel maka metode sampling adalah random dan menghasilkan sampel random.

Definisi yang kedua sampel random berkaitan dengan variabel random, dari definisi 2.2.1 variabel random adalah fungsi yang menghubungkan bilangan-bilangan riil pada titik-titik dalam ruang sampel, sehingga secara tidak langsung variabel random berkaitan dengan sampel random.

Statistik

Pada mulanya kata statistik hanyalah merupakan kumpulan dari angka-angka yang disajikan dalam bentuk data yang dilakukan oleh pemerintah dalam suatu negara. Kemudian orang mulai berpikir bahwa statistik adalah angka yang berdasar dari beberapa angka seperti rata-rata dari angka-angka dalam suatu sampel, proporsi dari suatu populasi dan lain sebagainya. Jika diperhatikan angka-angka yang dirata-rata dapat berubah dari sampel satu ke sampel berikutnya atau populasi dapat berubah dari tahun ketahun berikutnya. Dibawah ini adalah definisi dari statistik.

Definisi 2.3.6

Statistik adalah adalah fungsi dari beberapa variabel random.

Contoh 2.2.1

Misal X_1, X_2, \dots, X_n adalah nilai-nilai dari n murid, maka masing-masing X_i adalah variabel random. Misal W adalah rata-rata nilai murid $W =$

$$\frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \text{ maka } W \text{ adalah statistik. Jika } X_1 = 76, X_2 =$$

84 dan $X_3 = 85$ adalah nilai ketiga siswa tersebut maka W samadengan

$$1/3(76+84+85) = 81 \frac{2}{3}$$

2.4 Pengukuran

Dalam pengukuran suatu variabel kita kenal ada empat jenis pengukuran, berikut akan diberikan jenis-jenis skala pengukuran dari yang terlemah hingga yang terkuat:

1. Skala nominal

Pengukuran pada tingkatan paling lemah adalah ketika angka-angka atau lambang-lambang lain digunakan semata-mata untuk mengklasifikasikan suatu objek, orang, atau sifat. Jika angka-angka atau lambang-lambang lain itu digunakan untuk mengidentifikasi kelompok-kelompok sebagai induk objek-objek yang berlainan angka atau lambang tadi merupakan skala nominal.

2. Skala Ordinal

Mungkin terjadi, bahwa objek-objek dalam satu kategori dalam suatu skala tidak saja berbeda dari objek-objek lain dalam kategori lain dalam skala itu, tetapi juga objek-objek itu berada dalam suatu jenis hubungan dengan kategori-kategori tersebut. Hubungan yang biasa terdapat antara kelas-kelas adalah lebih tinggi, lebih disukai, lebih sulit, dan sebagainya. Hubungan semacam itu diberi tanda dengan tanda $>$, yang pada umumnya berarti lebih besar daripada. Sehubungan dengan skala tertentu tanda itu dapat dipakai untuk menunjukkan bahwa sesuatu lebih disukai daripada, lebih tinggi daripada dan seterusnya. Maknanya yang khusus bergantung pada sifat hubungan yang menentukan skala itu.

Jika terdapat suatu kelompok yang terdiri dari kelas-kelas ekuivalensi (yakni merupakan suatu skala nominal), kalau hubungan $>$ berlaku untuk hubungan berapa pasangan kelas, tetapi tidak untuk semua pasangan yang ada, maka yang dihadapi adalah skala berurutan sebagian. Jika hubungan itu berlaku untuk semua pasangan kelas yang ada, sehingga terjadi susunan kelas-kelas secara lengkap, maka yang kita punyai adalah suatu skala ordinal.

3. Skala Interval

Kalau suatu skala mempunyai segala sifat skala ordinal, dan kalau disamping itu jarak antara dua angka pada skala itu diketahui ukurannya, maka telah dicapai ukuran yang lebih kuat daripada yang memperlihatkan urutan semata. Dalam kasus ini pengukuran dalam arti skala interval telah dicapai. Artinya jika pemetaan atas beberapa kelas obyek sebegitu tepatnya sehingga tahu berapa besar interval (jarak) antara obyek satu dengan lainnya, maka telah dicapai suatu pengukuran interval. Skala interval ditandai dengan sebuah unit pengukuran yang umum dan konstan, yang melekatkan suatu angka riil kepada semua pasangan obyek dalam himpunan berurut. Dalam pengukuran jenis ini, rasio dua interval yang manapun, tidak bergantung pada unit pengukuran dan titik nol. Dalam skala interval, titik nol dan unit pengukurannya adalah sembarang.

4. Skala Rasio

Kalau suatu skala memiliki semua ciri suatu skala interval, dan disamping itu memiliki suatu titik nol sejati sebagai titik asalnya maka skala itu dinamakan skala rasio.

Dalam pengujian metode parametrik skala pengukuran yang digunakan sekurang kurangnya adalah skala interval sedang pada metode nonparametrik skala pengukuran bisa menggunakan skala nominal atau ordinal.

2.5 Distribusi sampling

Distribusi sampling adalah suatu distribusi teoritis. Ini adalah suatu distribusi yang akan didapatkan jika diambil semua sampel yang mungkin dengan ukuran sama, dari populasi yang sama, dan masing masing ditarik secara random. Hal ini juga dapat dikatakan dengan cara lain distribusi sampling adalah distribusi dibawah H_0 dari semua harga yang dijalani oleh suatu statistik tertentu (misalnya mean sampel) kalau statistik itu dihitung dari sampel-sampel yang sama besarnya dan yang didapatkan secara random.

Distribusi sampling suatu statistik menunjukkan kemungkinan dibawah H_0 , yang berkaitan dengan berbagai harga angka yang mungkin dari statistik itu. Kemungkinan yang berkaitan dengan terjadinya harga tertentu dari statistik itu, dibawah H_0 , bukanlah kemungkinan yang eksak dari tentang harga itu saja. Tetapi kemungkinan yang terjadi dibawah H_0 disini di gunakan untuk menunjukkan harga tertentu, di tambah dengan kemungkinan harga yang lebih ekstrim. Dengan demikian kemungkinan yang bersangkutan atau kemungkinan yang berkaitan dengan kejadian dibawah H_0 adalah kemungkinan akan terjadinya suatu harga yang sama ekstrim atau lebih ekstrim dari harga tertentu statistik uji dibawah H_0 .

2.6 Pengujian Hipotesis

Pengujian hipotesis adalah termasuk bagian dalam statistika inferensia, pengujian hipotesis adalah proses membuat kesimpulan berdasarkan data sampel yang diperoleh dari populasi yang bertujuan untuk menolak atau tak dapat menolak tentang pernyataan suatu populasi. Pernyataan tersebut adalah yang disebut dengan hipotesis atau dugaan sementara. Sebagai salah satu contoh hipotesis adalah laki-laki pada umumnya lebih kuat dari perempuan.

Contoh hipotesis diatas dapat diuji dengan jalan membandingkan hasil terhitung dengan hasil sampel yang bersifat empiris. Jika hipotesis tersebut ternyata tak sesuai dengan data empirisnya, maka harus diperbaiki atau ditolak keabsahannya. Jika cara pengumpulan data sampelnya memang baik sekali, maka penolakan atau penerimaan hipotesis diatas dapat menggunakan prosedur pengujian secara statistik. Prosedur tersebut dapat dibuat langkah-langkah sebagai berikut:

1. Menyusun data
2. Membuat asumsi-asumsi
3. Membuat hipotesis
4. Menentukan taraf signifikan
5. Menentukan harga statistik uji
6. Menentukan daerah penolakan
7. Pengambilan keputusan

Hipotesis yang akan diuji disebut dengan hipotesis nol dan dinyatakan dengan H_0 . Sedang hipotesis alternatifnya dinyatakan dengan H_1 adalah kebalikan dari hipotesis nol yang berisi pernyataan H_0 tak dapat diterima.

Kemudian untuk membuat keputusan mengenai suatu hipotesis diperlukan suatu statistik uji.

Definisi 2.6.1

Statistik uji adalah statistik yang digunakan untuk membantu membuat keputusan mengenai pengujian hipotesis

Sifat-sifat yang diinginkan dari pengujian hipotesis adalah menghubungkan titik-titik dalam ruang sampel, titik-titik disusun dengan tujuan untuk membedakan antara H_0 yang dapat diterima dan H_0 yang ditolak. Letak daerah kritis (daerah penolakan H_0) berkorespondensi dengan bentuk dari perumusan H_1 . Nilai-nilai besar atau nilai-nilai kecil dari statistik uji dari suatu kejadian dalam percobaan dapat menunjukkan daerah penolakan. Uji dimana daerah penolakan berhubungan dengan nilai-nilai besar atau nilai-nilai kecil dari statistik uji disebut uji satu sisi. Jika daerah penolakan merupakan kombinasi nilai-nilai besar dan nilai-nilai kecil maka uji tersebut merupakan uji dua sisi.

Definisi 2.6.2

Daerah kritis adalah himpunan seluruh titik dalam ruang sampel yang menghasilkan keputusan menolak H_0

Kadang-kadang daerah kritis disebut juga daerah penolakan dan himpunan titik dalam ruang sampel yang tidak berada dalam daerah kritis tersebut disebut daerah penerimaan.

Ada dua kemungkinan membuat kesalahan mengambil keputusan dalam pengujian hipotesis. Jika hipotesis nol seharusnya adalah dapat diterima tetapi ditolak maka telah melakukan kesalahan, dan kesalahan tersebut disebut kesalahan tipe I.

Definisi 2.6.3

Kesalahan tipe I adalah kesalahan menolak H_0 yang seharusnya dapat diterima.

Kemungkinan kedua melakukan kesalahan dalam pengujian hipotesis dengan menerima H_0 yang seharusnya ditolak. Kesalahan seperti ini disebut dengan kesalahan tipe II.

Definisi 2.6.4

Kesalahan tipe II adalah kesalahan menerima H_0 yang seharusnya ditolak

Definisi 2.6.5

Tingkat signifikan atau α adalah probabilitas kesalahan menolak H_0 yang seharusnya diterima

Jika probabilitas kesalahan menolak H_0 adalah α maka probabilitas menerima H_0 atau membuat keputusan yang bisa diterima dari uji hipotesis adalah $1 - \alpha$.

Probabilitas melakukan kesalahan tipe kedua dinyatakan dengan β . Jika H_0 tidak dapat diterima maka keputusan dapat menerima H_0 mempunyai probabilitas sebesar β atau untuk menolak H_0 dengan probabilitas $1 - \beta$.

Jelaslah dalam statistika inferensia ada kemungkinan untuk berbuat satu diantara dua tipe kesalahan itu. Dengan demikian pembuat eksperimen harus mencapai harga kompromi yang merupakan keseimbangan yang optimal antara

kemungkinan-kemungkinan berbuat kedua tipe kesalahan itu. Uji-uji statistik yang bermacam-macam menawarkan kemungkinan untuk terciptanya keseimbangan yang berbeda. Dalam mencapai keseimbangan inilah maka pengertian mengenai fungsi kekuatan (power function) suatu uji statistik menjadi relevan.

Definisi 2.6.6

Kekuatan (power) suatu uji didefinisikan sebagai kemungkinan untuk menolak H_0 jika H_0 memang pada kenyataannya harus ditolak

3.7 Permutasi

Berikut akan diberikan definisi dan teorema yang berkaitan dengan permutasi.

Definisi 2.7.1

Setiap susunan yang berbeda dari sehimpunan objek dinamakan permutasi

Secara umum jika kita memiliki n objek dan disusun r unsur dari mereka (dimana $r \leq n$) dalam susunan terhingga, maka susunan yang demikian ini dinamakan permutasi dari n objek yang disusun. Permutasi biasa dinyatakan dengan lambang $P(n,r)$.

Teorema 2.7.1

Jumlah keseluruhan permutasi dari n objek yang diambil r , $P(n,r)$, diberikan dengan pernyataan :

$$P(n,r) = n(n-1)(n-2)\dots\dots(n-r+1) \quad (2.7.1)$$

Bukti

Jumlah yang mungkin dari susunan n objek yang diambil r ekuivalen dengan cara memilih dari n objek berbeda untuk mengisi r posisi. Ada n pilihan untuk posisi pertama, $n-1$ untuk pilihan kedua, dan $n-2$ untuk posisi ketiga, dan seterusnya. Posisi ke r dapat diisi dengan $n-(r-1) = n-r+1$, sehingga r posisi ketiga dan seterusnya dapat diisi dengan banyak cara $n(n-1)(n-2)\dots(n-r+1)$.

Jika akan disusun suatu permutasi dari n objek yang diambil dari n unsur (semua unsur), maka dicatat bahwa $r = n$, berdasarkan persamaan 2.7.1 maka diperoleh :

$$P(n,n) = n(n-1)(n-2)\dots 3.2.1. = n! \quad (2.7.2)$$

Dengan mengalikan penyebut dan pembilang, ruas kanan persamaan 2.7.1 dengan $(n-r)!$, kita peroleh bentuk lain rumus $P(n,r)$:

$$P(n,r) = \frac{n(n-1)(n-2)\dots(n-r+1)(n-1)!}{(n-r)!} \quad (2.7.3)$$

Atau

$$P(n,r) = \frac{n!}{(n-r)!} \quad (2.7.4)$$

2.8 Korelasi

Masalah korelasi akan timbul apabila dihadapkan pada pertanyaan apakah ada hubungan antara variabel-variabel dari sekumpulan data yang sedang diselidiki. Misalnya ingin diselidiki apakah ada hubungan antara jumlah penduduk dengan banyaknya pemukiman baru yang dibangun dimasing-masing kota terpilih tadi.

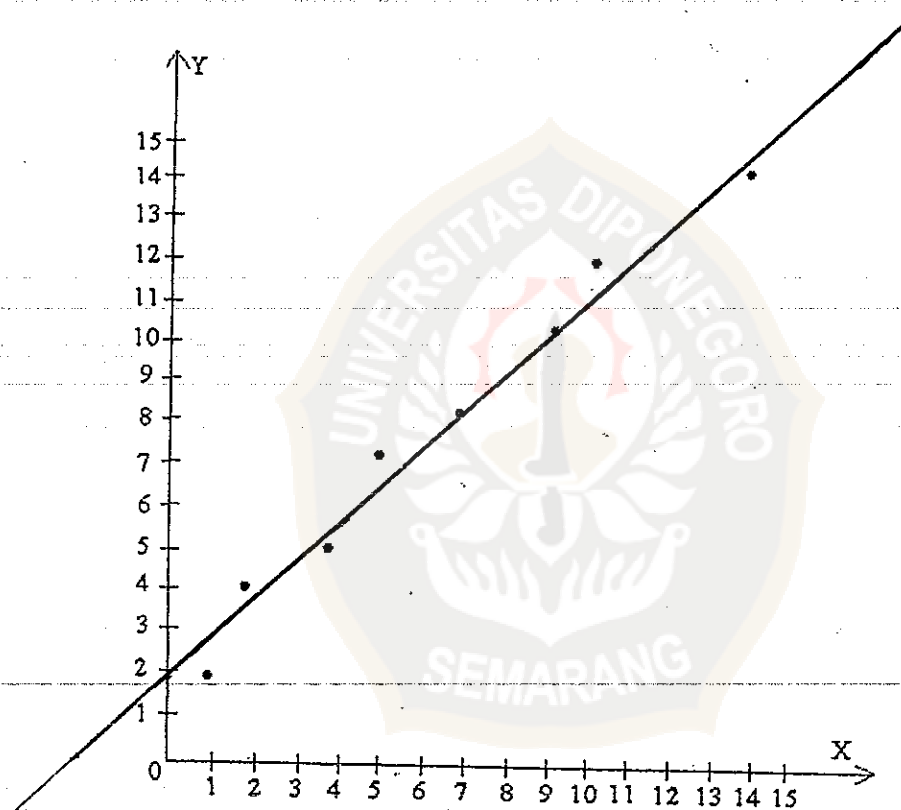
Penyelidikan untuk mengetahui hubungan antara kedua variabel tersebut biasanya dimulai dengan suatu usaha untuk menentukan hubungan terdekat daripada hubungan itu dengan disajikan dalam sebuah grafik yang disebut dengan diagram pencar (scatter diagram). Diagram ini melukiskan titik-titik yang diletakkan pada sumbu koordinat, dari dua variabel X dan Y.

Salah satu syarat penggunaan metode korelasi adalah bahwa hubungan antara dua variabel X dan Y adalah hubungan yang linier. Hubungan yang linier salah satu cara dibuktikan dengan apakah titik-titik dalam diagram pencar tersebut bisa ditarik suatu garis lurus yang mewakili yang terpencar tadi. Apabila bisa ditarik garis paling sesuai berarti variabel-variabel tersebut mempunyai hubungan linier (gambar 2.1 dan gambar 2.2). Apabila tidak bisa dibuat garis lurus yang sesuai maka variabel-variabel tersebut mempunyai hubungan nonlinier. Pembuatan garis yang sesuai tadi digunakan metode kuadrat terkecil.

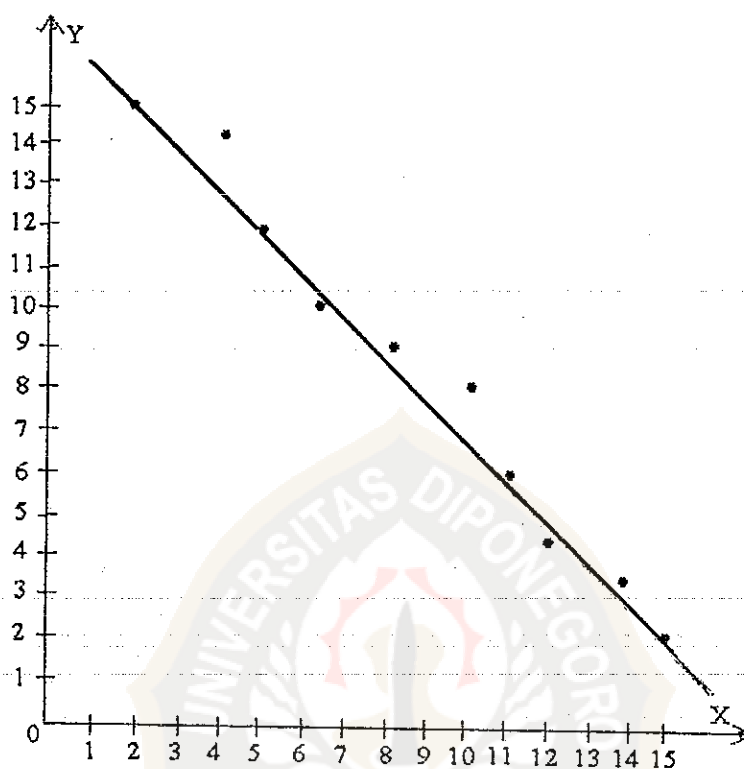
Ukuran yang menentukan terpencarnya titik-titik diagram pencar di sekitar garis lurus yang sesuai dengan titik-titik tersebut, jika variabel-variabel tersebut mempunyai hubungan linier, dinamakan koefisien korelasi. Dengan perkataan lain koefisien korelasi merupakan ukuran besar kecilnya atau kuat tidaknya hubungan antara variabel-variabel tersebut apabila bentuk hubungan tersebut linier. Koefisien korelasi hanyalah mengukur besarnya hubungan linier variabel-variabel dan tidak selalu menunjukkan hubungan sebab akibat.

Hubungan dua variabel ada yang positif dan ada yang negatif, hubungan X dan Y dikatakan positif apabila kenaikan nilai-nilai X pada umumnya diikuti dengan kenaikan nilai-nilai Y atau penurunan nilai-nilai X diikuti penurunan

nilai-nilai Y (gambar 2.1). Sebaliknya hubungan dikatakan negatif apabila kenaikan nilai-nilai X pada umumnya diikuti oleh penurunan nilai-nilai Y atau penurunan nilai-nilai X diikuti oleh kenaikan nilai-nilai Y (gambar 2.2).



Gambar 2.1.



Gambar 2.2.