

Ringkasan Materi Kuliah

Statistika Dasar

Irmatul Hasanah, M.Si.



Fakultas Dakwah
UIN Sultan Maulana Hasanuddin Banten
2018

Daftar Isi

| | |
|---|----|
| Materi 1 Teori Pendahuluan Statistika; Data dan Variabel..... | 3 |
| Materi 2 Skala Pengukuran; Pengumpulan dan Pengolahan Data..... | 8 |
| Materi 3 Penyajian Data | 13 |
| Materi 4 Penyajian Data | 17 |
| Materi 5 Penyajian Data | 21 |
| Materi 6 Distribusi Frekuensi..... | 23 |
| Materi 7 Distribusi Frekuensi..... | 25 |
| Materi 8 Ukuran Pemusatan..... | 28 |
| Materi 9 Ukuran Pemusatan..... | 31 |
| Materi 10 Ukuran Pemusatan..... | 34 |
| Materi 11 Ukuran Pemusatan..... | 38 |
| Materi 12 Ukuran Variasi atau Dispersi | 43 |
| Materi 13 Analisis Korelasi..... | 49 |
| Materi 14 Analisis Regresi | 57 |
| Referensi..... | 60 |

Materi 1

Pendahuluan Teori Statistika; Data dan Variabel

Pengertian Statistik dan Statistika

1. Statistik

Statistik merupakan:

- Sekumpulan data bilangan maupun non bilangan yang disusun pada sebuah tabel/diagram.
- Ukuran yang diperoleh dari pengukuran dan menggambarkan data.

2. Statistika

Statistika merupakan ilmu yang mempelajari:

- Cara pengumpulan data
- Pengolahan/pengelompokkan data
- Penyajian data
- Analisis data
- Pengujian hipotesis
- Cara pengambilan kesimpulan

Jenis Statistika

1. Statistika Deskriptif

Statistika deskriptif adalah tata cara untuk mengatur dan meringkas informasi. Statistika deskriptif berupa grafik, diagram, dan tabel serta perhitungan beberapa macam pengukuran, seperti: rata-rata, variasi, persentil, dan lain sebagainya.

Contoh 1-1

Pesta Olahraga Asia Tenggara (Southeast Asian Games disingkat SEA Games) adalah ajang olahraga yang diadakan setiap 2 tahun dan melibatkan 11 negara Asia Tenggara, yaitu Kamboja, Laos, Myanmar, Thailand, Vietnam, Brunei, Filipina, Indonesia, Malaysia, Singapura, dan Timor Leste.

Pada tahun 2015, Indonesia berhasil meraih sebanyak 182 medali yang terdiri dari medali emas sebanyak 47, medali perak sebanyak 61, dan medali perunggu sebanyak 74.

Tabel 1-1

| Negara | Emas | Perak | Perunggu | Jumlah |
|---------------|------------|------------|------------|-------------|
| Thailand | 95 | 83 | 69 | 247 |
| Singapura | 84 | 73 | 102 | 259 |
| Vietnam | 73 | 53 | 60 | 186 |
| Malaysia | 62 | 58 | 66 | 186 |
| Indonesia | 47 | 61 | 74 | 182 |
| Filipina | 29 | 36 | 66 | 131 |
| Myanmar | 12 | 26 | 31 | 69 |
| Kamboja | 1 | 5 | 9 | 15 |
| Laos | 0 | 4 | 25 | 29 |
| Brunei | 0 | 1 | 6 | 7 |
| Timor Leste | 0 | 1 | 1 | 2 |
| Jumlah | 403 | 401 | 509 | 1313 |

[http://id.wikipedia.org/wiki/Pesta\Olahraga\Asia\Tenggara\2015](http://id.wikipedia.org/wiki/Pesta%20Olahraga%20Asia%20Tenggara%202015)

2. Statistika Inferensial

Statistika inferensial adalah tata cara untuk menggambarkan dan menarik kesimpulan mengenai populasi berdasarkan informasi yang diperoleh dari sampel.

Contoh 1-2

Quick Count atau penghitungan cepat digunakan untuk memprediksi hasil pemungutan suara.

Lembaga Penelitian, Pendidikan, dan Penerangan Ekonomi dan Sosial (LP3ES) melakukan perhitungan cepat pemilu presiden tahun 2009. Berdasarkan perhitungan statistika diperoleh pasangan Megawati-Prabowo 27,40%, SBY-Boediono 60,28%, dan JK-Wiranto 12,7%. **Sehingga pilpres 2009 akan dimenangkan oleh SBY-Boediono dengan perolehan suara sebesar 60,28%. Pada kenyataannya, hasil pilpres dimenangkan oleh SBY-Boediono sebesar 60,80%.**

Catatan:

Statistika deskriptif dan statistika inferensial sangat berkaitan. Karena sebelum kita melakukan analisis inferensial, terlebih dahulu kita harus melakukan teknik statistika deskriptif untuk mengumpulkan dan meringkas informasi dari sampel.

Kegunaan Statistika

1. Menjabarkan dan memahami suatu hubungan

Contoh 1-3

Seorang wiraswasta, dengan mengumpulkan data pendapat dan biaya, dapat membandingkan hasil pengembalian atas investasi dalam satu periode dengan data dari periode-periode sebelumnya.

2. Mengambil keputusan yang baik.

Contoh 1-4

Misalkan manajer Perusahaan Kosmetika “Selalu Cakep”, Suryono Dasamuko, mengiklankan bahwa 90 persen konsumen puas dengan produk perusahaannya. Jika Ibu Aminah Sari, seorang aktivis politik, merasa bahwa pernyataan ini berlebihan. Dalam kasus ini, ibu Aminah dapat menguji kebenaran dari pernyataan Suryono.

3. menangani perubahan

Contoh 1-5

Misalkan seorang manajer personalia telah mencatat bahwa pelamar yang mempunyai nilai tinggi untuk tes ketangkasan manual cenderung berprestasi baik dalam perakitan suatu produk, sedangkan mereka yang lebih rendah nilainya cenderung kurang produktif. Dengan menerapkan teknik statistic yang dikenal sebagai analisis regresi, manajer itu dapat memperkirakan atau meramalkan bagaimana produktivitas seorang pelamar baru dalam pekerjaannya berdasarkan hasil tes.

Data dan Variabel

Definisi

Data merupakan sesuatu yang diketahui atau dianggap. Dengan demikian, data dapat memberikan gambaran tentang suatu keadaan atau persoalan. Data tentang sesuatu pada umumnya dikaitkan dengan tempat dan waktu.

Contoh 1-6

Harga beras yang bermutu sedang, di pasar Senen, Jakarta, pada tanggal 2 Januari 2014 adalah Rp8500,- per kg.

- Data : Beras yang bermutu sedang
- Tempat : pasar Senen Jakarta
- Waktu : 2 Januari 2014

Sesuatu yang dianggap juga merupakan data walaupun data seperti itu belum tentu benar, sebab hasilnya masih merupakan suatu hipotesis yang perlu diuji terlebih dahulu. Anggapan atau asumsi yang dipergunakan sebagai dasar pembuatan keputusan.

Contoh 1-7

Karena pemerintah menganggap persediaan beras cukup, maka diputuskan untuk tidak mengimpor beras.

Syarat Data yang Baik dan Pembagian Data

1. Objektif
2. Representatif (mewakili)
3. Kesalahan sampling kecil
4. Tepat waktu
5. Relevan

Data menurut sifatnya:

1. Data kualitatif

Data kualitatif adalah data yang tidak berbentuk angka. Misalnya produksi daging sapi meningkat, harga daging ayam mahal, dan sebagainya.

2. Data kuantitatif

Data kuantitatif adalah data yang dinyatakan dalam bentuk angka. Misalnya produksi padi meningkat 10 persen, harga daging sapi per kilogram rata-rata adalah Rp 15.000, dan sebagainya.

Data menurut sumbernya:

1. Data Internal

Data internal adalah data yang bersumber dari keadaan atau kegiatan suatu organisasi atau kelompok. Misalnya, data penjualan dan data produksi suatu perusahaan.

2. Data Eksternal

Data eksternal adalah data yang bersumber dari luar organisasi atau kelompok. Misalnya, suatu perusahaan mencari data mengenai daya beli konsumen dari kantor Badan Pusat Statistik setempat.

Data menurut cara memperolehnya:

1. Data Primer

Data primer adalah data yang dikumpulkan dan diolah sendiri oleh suatu organisasi atau perorangan langsung dari objeknya. Misalnya, suatu perusahaan ingin mengetahui konsumsi susu rata-rata penduduk di suatu daerah dengan cara melakukan wawancara langsung kepada penduduk setempat.

2. Data Sekunder

Data sekunder adalah data yang diperoleh dalam bentuk jadi dan telah diolah oleh pihak lain, yang biasanya dalam bentuk publikasi (BPS, LIPI, KPU, BI).

Data menurut waktu pengumpulannya:

1. Data *Cross Section*

Data *cross section* adalah data yang dikumpulkan dalam suatu periode tertentu, biasanya menggambarkan keadaan atau kegiatan dalam periode tertentu. Misalnya, hasil sensus penduduk tahun 2010 menggambarkan keadaan Indonesia pada tahun 2010 menurut umur, jenis kelamin, agama, tingkat pendidikan, dan sebagainya.

2. Data Berkala

Data berkala adalah data yang dikumpulkan dari waktu ke waktu untuk melihat pembahasan yang terjadi. Tujuannya adalah untuk menggambarkan perkembangan suatu kegiatan dari waktu ke waktu atau untuk melihat perubahan yang terjadi. Misalnya perkembangan produksi padi selama lima tahun terakhir.

Variabel

Variabel atau peubah ialah sesuatu yang nilainya dapat berubah atau berbeda. Nilai karakteristik suatu elemen merupakan nilai variabel, misalknya harga (karakteristik harga suatu barang akan berubah-ubah menurut waktu atau berbeda-beda menurut tempat), produksi, hasil penjualan, ekspor, pendapatan, nasional, umur, tinggi badan, berat badan, tekanan darah, temperature/suhu, modal perusahaan, dan sebagainya. Huruf X,Y,Z biasa digunakan untuk menunjukkan suatu variabel.

Contoh 1-8

Diketahui 3 perusahaan dengan X=modal perusahaan dalam jutaan rupiah, dimana $X_1=5$, $X_2=7$, $X_3=4$. Artinya perusahaan pertama mempunyai modal Rp5 juta, kedua Rp7 juta, dan ketiga Rp4 juta.

Materi 2

Skala Pengukuran; Pengumpulan dan Pengolahan Data

Ada 4 tingkatan variabel yang disebut 4 skala utama, yaitu:

1. Skala Nominal

Skala nominal ialah angka yang berfungsi hanya untuk membedakan, sebagai lambing/symbol. Disebut data kategori/nonmetrik/kualitatif.

Contoh 2-1

Jenis kelamin : Perempuan=0, Laki-laki=1
 Agama : Islam=1, Kristen=2, Hindu=3
 Suku Bangsa : Jawa=1, Sunda=2, Batak=3, Minang=4

2. Skala Ordinal

Skala ordinal adalah angka selain berfungsi sebagai nominal juga menunjukkan urutan, dengan jarak tidak sama. Tidak sampai menunjukkan berapa kali.

Contoh 2-2

| Peringkat | | | | | | | | |
|-----------|-------------|---------|---|---|---|----------|---|----------|
| Nama | Nilai Ujian | Ordinal | | | | Jarak | | |
| Abas | 70 | 3 | | | | 1-2 : 5 | | |
| Budi | 90 | 1 | 1 | 2 | 3 | 4 | 5 | 2-3 : 15 |
| Charles | 85 | 2 | | | | 3-4 : 15 | | |
| Dadang | 65 | 4 | | | | 4-5 : 5 | | |
| Eko | 60 | 5 | | | | | | |

Pendidikan Johny: Bukan Sarjana=1, Sarjana Muda=2, Sarjana=3

Pangkat Paulus: Mayor=1, Let.Kol=2, Kol.=3

Pendapatan Pak Karto: rendah=1, menengah=2, tinggi=3 (<Rp5jt, Rp5-10jt, >Rp10jt)

3. Skala Interval

Skala interval ialah angka yang selain berfungsi sebagai nominal dan ordinal juga menunjukkan jarak yang sama tetapi tidak sampai menunjukkan berapa kali, tidak mempunyai titik asal nol.

Contoh 2-3

Interval mencakup rating (penilaian), misalnya sangat puas=5, puas=4, netral=3, tidak puas=2, dan sangat tidak puas=1. Puas=4 tidak berarti dua kalinya tidak puas=2. Tetapi jarak dari puas dan sangat puas dianggap sama jarak dari netral ke puas (dari 3 ke 4) atau jarak dari sangat tidak puas ke tidak puas (dari 1 ke 2), yaitu sebesar 1(=satu).

4. Skala Rasio

Skala rasio ialah angka yang selain berfungsi sebagai nominal, ordinal, dan interval juga menunjukkan berapa kali, sebab mempunyai titik asal nol. Misalnya berat badan Abas 90 kg, Budi 60 kg. Berat badan Abas 1,5 kali berat badan Budi.

Contoh 2-4

Penelitian terhadap nasabah Bank Mandiri, mereka ditanya mengenai

- Jumlah tabungannya (jutaan Rp) : Rasio (R)
- Jumlah penghasilannya (jutaan Rp) : R
- Jumlah anggota keluarganya yang ditanggung : R
- Pendidikan : BS, SM, S : Ordinal (O)
- Tingkat kepuasan terhadap mutu pelayanan : Interval (I)
- Suku bangsa : Jawa=1, Sunda=2, Batak=3 : Nominal (N)
- Agama : I=1, K=2, H=3 : N

Nominal dan Ordinal disebut nonmetrik (kualitatif)

Interval dan Rasio disebut metrik (kuantitatif)

Pengumpulan Data

Data statistik harus merupakan data yang dapat dipercaya dan tepat waktu. Sebelum mengumpulkan data, perlu diketahui tujuan dari pengumpulan data. Apapun tujuan pengumpulan data, perlu diketahui jenis elemen atau objek yang akan diselidiki. Elemen adalah unit terkecil dari objek penelitian, disebut juga "unit analysis" atau "unit sampling". Elemen atau unit terkecil dapat berupa orang (pegawai negeri, mahasiswa, pedagang, konsumen, karyawan, nasabah bank, dan sebagainya), organisasi, atau badan usaha (perusahaan, sekolah/universitas, departemen, provinsi, kabupaten, kecamatan, desa, rumah tangga, pasar, dan lain sebagainya), atau barang (kendaraan, mesin, gedung, senjata, dan sebagainya).

Tujuan pengumpulan data:

1. Mengetahui jumlah elemen.
2. Mengetahui karakteristik dari elemen-elemen.

Karakteristik adalah sifat-sifat, ciri-ciri atau hal-hal yang dimiliki oleh elemen, yaitu semua keterangan mengenai elemen.

Contoh 2-5

- Elemen: pegawai pemerintah/swasta
Karakteristik: jenis kelamin, pendidikan, agama, umur, masa kerja, golongan dan gaji.
- Elemen: perusahaan
Karakteristik: jumlah karyawan, jumlah kekayaan, hasil produksi, dan hasil penjualan.
- Elemen: universitas
Karakteristik: jumlah mahasiswa, jumlah dosen, dan banyaknya fakultas.

Populasi dan Sampel

1. Populasi

Populasi adalah kumpulan dari seluruh elemen sejenis tetapi dapat dibedakan satu sama lain karena karakteristiknya.

Contoh 2-6

Populasi: seluruh karyawan perusahaan
Elemen: Orang, yaitu karyawan perusahaan

Meskipun jenisnya sama, namun karakteristiknya secara keseluruhan akan berlainan, misalnya umur, pendidikan, masa kerja, jumlah anak, gaji pokok, dan lain sebagainya.

Karena pengumpulan data akan menghasilkan nilai observasi sebagai nilai karakteristik dari masing-masing elemen, maka kumpulan seluruh nilai (data) observasi disebut populasi. Kumpulan seluruh kemungkinan hasil eksperimen juga disebut populasi.

2. Sampel

Sampel adalah sebagian dari populasi. Jika n adalah banyaknya elemen sampel dan N adalah banyaknya elemen populasi, maka $n < N$.

Populasi : $X_1, X_2, \dots, X_i, \dots, X_N$

Sampel : $X_1, X_2, \dots, X_i, \dots, X_n$

Metode Pengumpulan Data

1. Sensus

Sensus adalah cara pengumpulan data apabila seluruh elemen populasi diselidiki satu per satu. Data yang diperoleh sebagai hasil pengolahan sensus disebut data yang sebenarnya (true value), atau disebut parameter.

Contoh 2-7

Hasil sensus penduduk tahun 1980 memberikan data sebenarnya mengenai penduduk Indonesia. Jumlahnya menurut umur, menurut jenis kelamin, lapangan kerja, agama, dan pendidikan.

2. Sampling

Sampling adalah cara pengumpulan data apabila diselidiki hanya elemen sampel dari suatu populasi. Data yang diperoleh dari hasil sampling merupakan data perkiraan (*estimated value*). Jadi, jika dari 1000 perusahaan hanya akan diselidiki 100 saja, maka hasil penyelidikannya merupakan suatu perkiraan. Data yang dihitung berdasarkan sampel disebut statistik.

Cara pengambilan Sampel

1. Cara acak

Cara acak adalah suatu cara pemilihan elemen dari populasi untuk menjadi anggota sampel, dimana pemilihannya dilakukan sedemikian rupa sehingga setiap elemen populasi mendapat kesempatan yang sama (*equal chance*) untuk dipilih menjadi anggota sampel.

Jenis Sampling Acak

Sampling acak (*random or probability sampling*) ialah sampling yang pemilihan elemen-elemen populasinya dilakukan secara acak (*random*). Pemilihan dilakukan dengan lotere, undian, atau tabel bilangan acak.

- (i) Simple random sampling
- (ii) Stratified random sampling
- (iii) Multistage random sampling
- (iv) Cluster random sampling
- (v) Systematic random sampling

2. Cara bukan acak

Cara bukan acak adalah suatu cara pemilihan elemen-elemen dari populasi untuk menjadi anggota sampel dimana setiap elemen tidak mendapat kesempatan yang sama untuk dipilih.

Alat Pengumpulan Data

1. Daftar pertanyaan
2. Wawancara
3. Observasi atau pengamatan langsung
4. Melalui pos, telepon, atau alat komunikasi lainnya
5. Alat ukuran seperti meteran, timbangan, thermometer, dan lain sebagainya

Metode Pengolahan Data

1. Pengolahan data secara manual

Contoh 2-8

Diasumsikan bahwa para pemilih pada Pemilu tahun 1999 di suatu tempat hanya akan memilih 4 partai besar, yaitu PDI-P (No.11), PAN (No.15), Golkar (No.33), dan PKB (No.35). Jika hasil suara yang masuk sebanyak 50, maka dapat dituliskan sebagai berikut (angka menunjukkan nomor partai):

15 15 11 11 11 33 15 35 33 11 15 33 35 11 35
 11 15 33 35 11 15 15 35 35 33 35 11 35 11 11
 11 11 33 15 15 35 33 35 11 11 15 15 35 15 11

Dari data tersebut, buatlah tally mark.

Contoh 2-9

Suatu survey dilakukan terhadap 10 perusahaan di daerah A yang diambil secara acak, untuk mengetahui jumlah dan rata-rata modal serta persentase perusahaan dengan modal kurang dari suatu nilai tertentu. Dari hasil penelitian diperoleh data modal 10 perusahaan sebagai berikut (nilai dalam jutaan rupiah):

50 40 70 60 30 100 120 80 110 90

Tentukan jumlah dan rata-rata modal serta persentase perusahaan dengan modal kurang dari 100 juta rupiah.

2. Pengolahan data secara elektronik

Materi 3

Penyajian Data

Data statistik tidak hanya cukup dikumpulkan dan diolah, tetapi juga perlu disajikan dalam bentuk yang mudah dibaca dan dimengerti oleh pengambil keputusan. Penyajian data ini bias berupa tabel atau grafik.

1. Tabel

Tabel merupakan kumpulan angka-angka yang disusun menurut kategori-kategori. Misalnya, jumlah pegawai menurut pendidikan dan masa kerja.

2. Grafik

Grafik merupakan gambar-gambar yang menunjukkan secara visual data berupa angka (mungkin juga dengan simbol-simbol) yang biasanya juga berasal dari tabel-tabel yang telah dibuat.

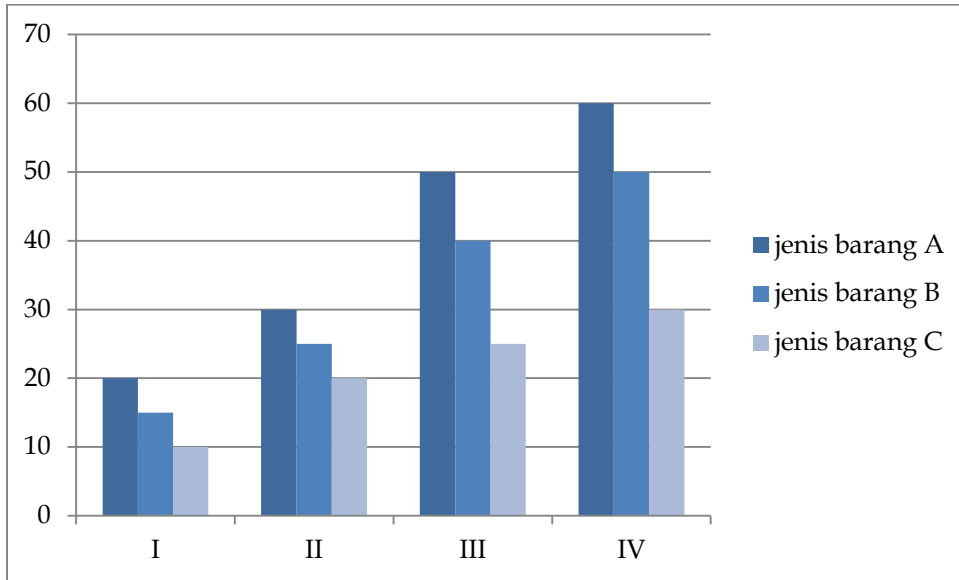
Cross Section Data

Penyajian dengan Tabel

Tabel 3-1 Penjualan PT.SINAR SAKTI menurut Jenis Barang dan Daerah Penjualan pada Tahun 2013 (dalam satuan)

| Jenis Barang | Daerah Penjualan | | | | Total |
|-----------------|------------------|-----|-----|-----|-------|
| | I | II | III | IV | |
| (1) | (2) | (3) | (4) | (5) | (6) |
| A | 20 | 30 | 50 | 60 | 160 |
| B | 15 | 25 | 40 | 50 | 130 |
| C | 10 | 20 | 25 | 30 | 85 |
| Total | 45 | 75 | 115 | 140 | 375 |

Penyajian dengan Grafik



Data Berkala

Tabel 3-2 Perkembangan Seluruh Hasil Penjualan PT. SINAR SAKTI menurut Jenis Barang dari 2007 sampai dengan 2013 (dalam satuan)

| Tahun | Jenis Barang A | Jenis Barang B | Jenis Barang C | Jumlah |
|-------|----------------|----------------|----------------|--------|
| (1) | (2) | (3) | (4) | (5) |
| 2007 | 90 | 85 | 50 | 225 |
| 2008 | 110 | 90 | 55 | 255 |
| 2009 | 115 | 105 | 60 | 280 |
| 2010 | 130 | 110 | 65 | 305 |
| 2011 | 140 | 120 | 75 | 335 |
| 2012 | 155 | 125 | 80 | 360 |
| 2013 | 160 | 130 | 85 | 375 |

Bentuk Tabel

1. Tabel satu arah (*one way table*)

Tabel satu arah adalah tabel yang memuat keterangan mengenai satu hal atau satu karakteristik saja, misalnya

Data peralatan: jumlah kendaraan bermotor menurut: a) merek, b) jenis, c) umur, d) harga, dan lain sebagainya.

Tabel 3-3 Produksi Kayu Hutan menurut Jenis Produksi 1996/1997 (000 M3)

| Jenis | Banyaknya |
|----------------|---------------|
| (1) | (2) |
| Kayu Bulat | 26.069 |
| Kayu Gergajian | 3.427 |
| Kayu Lapis | 10.498 |
| Jumlah | 39.994 |

2. Tabel dua arah

Tabel dua arah adalah tabel yang menunjukkan hubungan dua hal atau dua karakteristik, misalnya

Data peralatan, menurut umur dan merek, umur dan jenis, dan lain sebagainya.

Tabel 3-4 Jumlah Mahasiswa Universitas Persada Jakarta, menurut Fakultas dan Agama, 1999

| Fakultas | Katolik | Bukan Katolik | Jumlah |
|---------------------------------------|------------|---------------|-------------|
| Fakultas Ekonomi | 266 | 292 | 558 |
| Fakultas Ilmu Pendidikan dan Keguruan | 72 | 68 | 140 |
| Fakultas Ilmu Pasti | 108 | 88 | 196 |
| Fakultas Teknik | 150 | 162 | 312 |
| Fakultas Hukum | 55 | 65 | 120 |
| Fakultas Kedokteran | 273 | 186 | 459 |
| Jumlah | 924 | 861 | 1785 |

3. Tabel tiga arah

Tabel tiga arah adalah tabel yang menunjukkan tiga hal atau tiga karakteristik, misalnya

Data peralatan, menurut umur, merek, dan jenis; jenis, merek, dan unit kerja, dan lain sebagainya.

Tabel 3-5 Penjualan Perusahaan X menurut Jenis Barang, Tempat Penjualan, dan Tempat Asal Pembeli 2013 (dalam juta rupiah)

| Jenis Barang | I | | II | | III | | IV | | Jumlah | | Jumlah |
|-----------------|------------|------|------------|------|------------|------|------------|------|-------------|------|--------|
| | Kota | Desa | Kota | Desa | Kota | Desa | Kota | Desa | Kota | Desa | |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| A | 20 | 10 | 75 | 30 | 60 | 20 | 90 | 60 | 245 | 120 | 365 |
| B | 50 | 45 | 80 | 40 | 65 | 15 | 80 | 40 | 275 | 140 | 415 |
| C | 30 | 20 | 90 | 50 | 40 | 25 | 85 | 45 | 245 | 140 | 385 |
| D | 35 | 15 | 65 | 45 | 35 | 10 | 60 | 30 | 195 | 100 | 295 |
| E | 25 | 15 | 85 | 55 | 45 | 30 | 55 | 20 | 210 | 120 | 330 |
| | 160 | 105 | 395 | 220 | 245 | 100 | 370 | 195 | 1170 | 620 | 1790 |
| Jumlah | 265 | | 615 | | 345 | | 565 | | 1790 | | |

Materi 4

Penyajian Data (Lanjutan)

Bentuk Grafik

1. Grafik garis tunggal (*single line chart*)

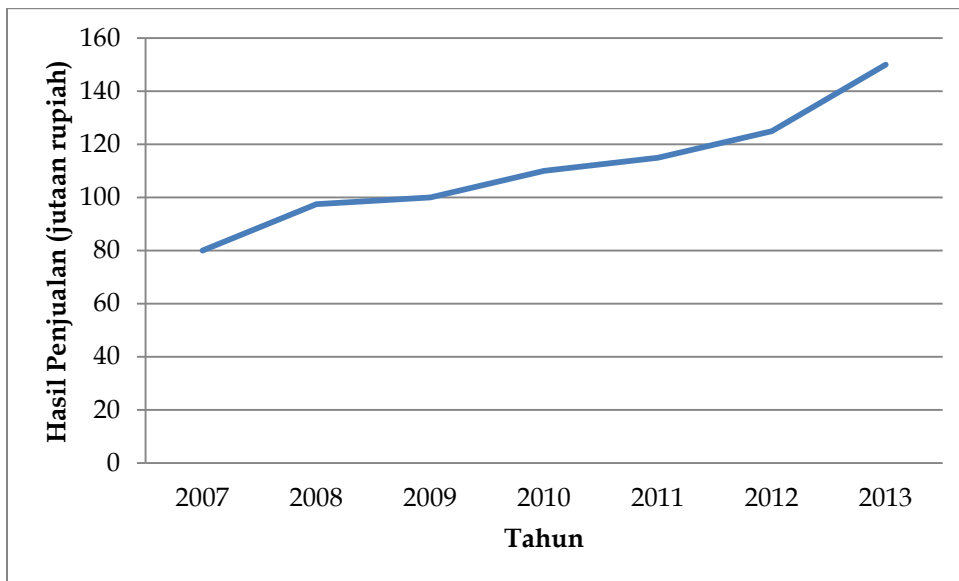
Grafik garis tunggal (*single line chart*) adalah grafik yang terdiri dari satu garis untuk menggambarkan perkembangan (trend) dari suatu karakteristik.

Contoh 4-1

Tabel 4-1 Penjualan Perusahaan "Harapan Kita", 2007-2013 (jutaan rupiah)

| Tahun | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|-----------------|------|------|------|------|------|------|------|
| Hasil Penjualan | 80 | 97.5 | 100 | 110 | 115 | 125 | 150 |

Berdasarkan data pada Tabel 4-1, buatlah grafiknya



2. Grafik garis berganda (*multiple line chart*)

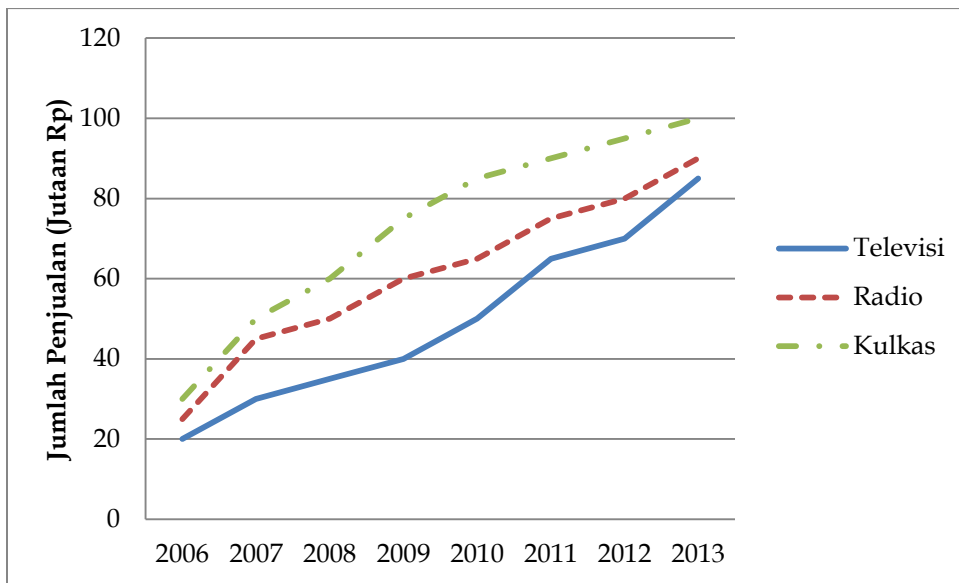
Grafik garis berganda (*multiple line chart*) adalah grafik yang terdiri dari beberapa garis untuk menggambarkan perkembangan beberapa hal/kejadian sekaligus.

Contoh 4-2

Tabel 4-2 Penjualan Toko "Terang" menurut Jenis barang dan Waktu (dalam juta rupiah)

| Jenis Barang | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|--------------|------|------|------|------|------|------|------|------|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Televisi | 20 | 30 | 35 | 40 | 50 | 65 | 70 | 85 |
| Radio | 25 | 45 | 50 | 60 | 65 | 75 | 80 | 90 |
| Kulkas | 30 | 50 | 60 | 75 | 85 | 90 | 95 | 100 |
| Jumlah | 75 | 125 | 145 | 175 | 200 | 230 | 245 | 275 |

Grafik garis berganda berdasarkan tabel di atas yaitu:



3. Grafik Batangan Tunggal (*Single Bar Chart*)

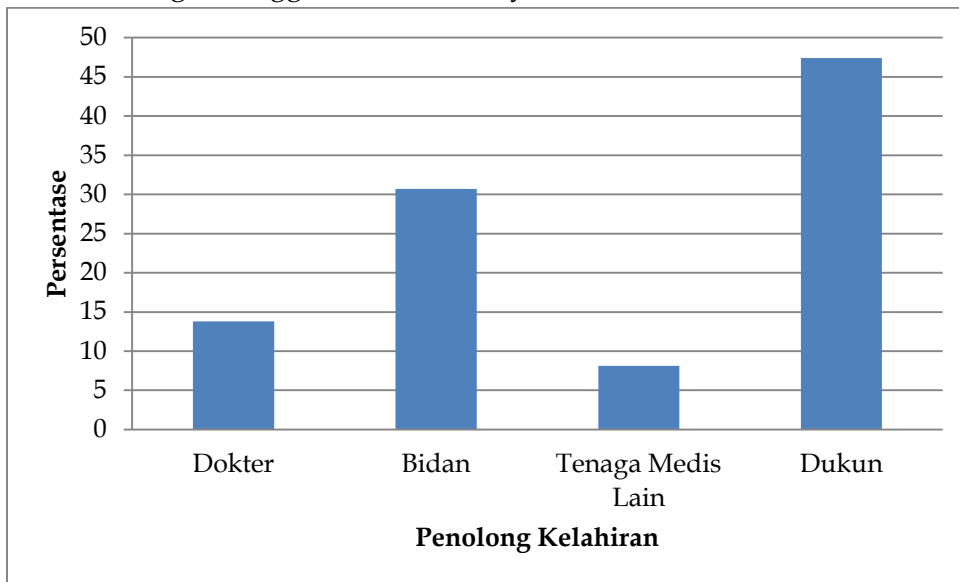
Grafik batangan tunggal (*single bar chart*) adalah grafik yang terdiri dari satu batang untuk menggambarkan perkembangan (trend) dari suatu karakteristik.

Contoh 4-3

Tabel 4-3 Persentase Balita menurut Tenaga Penolong Kelahiran di Indonesia, 1999

| Penolong Kelahiran | Persentase Balita |
|--------------------|-------------------|
| Dokter | 13.8 |
| Bidan | 30.7 |
| Tenaga Medis Lain | 8.1 |
| Dukun | 47.4 |
| Total | 100 |

Grafik batangan tunggal dari tabel 4-3 yaitu



4. Grafik Batangan Berganda (*Multiple Bar Chart*)

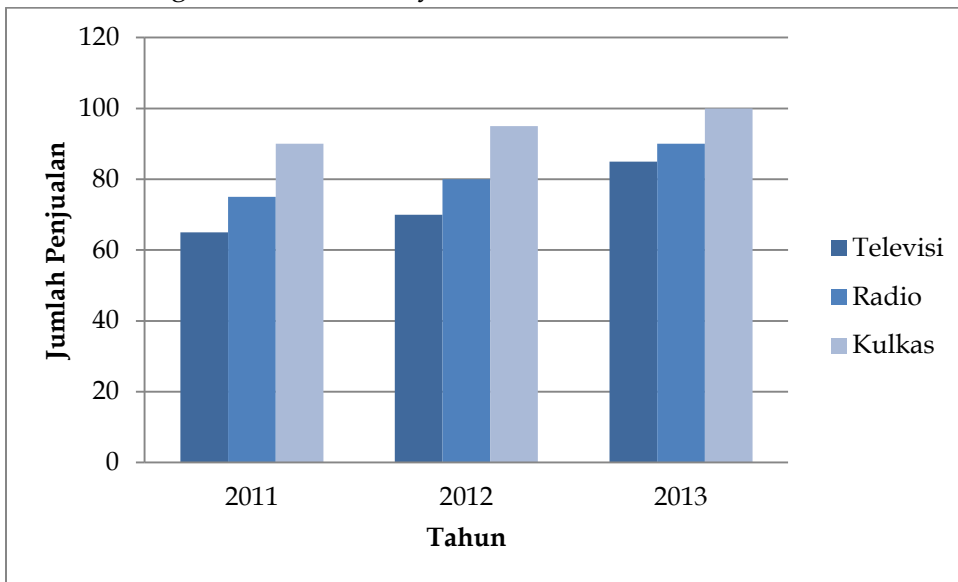
Grafik batangan berganda (*multiple bar chart*) adalah grafik yang terdiri dari beberapa garis untuk menggambarkan perkembangan beberapa hal/kejadian sekaligus.

Contoh 4-4

Tabel 4-4 Penjualan Toko "Terang" menurut Jenis barang dan Waktu
 (dalam juta rupiah)

| Jenis Barang | 2011 | 2012 | 2013 |
|--------------|------|------|------|
| (1) | (7) | (8) | (9) |
| Televisi | 65 | 70 | 85 |
| Radio | 75 | 80 | 90 |
| Kulkas | 90 | 95 | 100 |
| Jumlah | 230 | 245 | 275 |

Grafik Batang dari tabel di atas yaitu



Materi 5

Penyajian Data (Lanjutan)

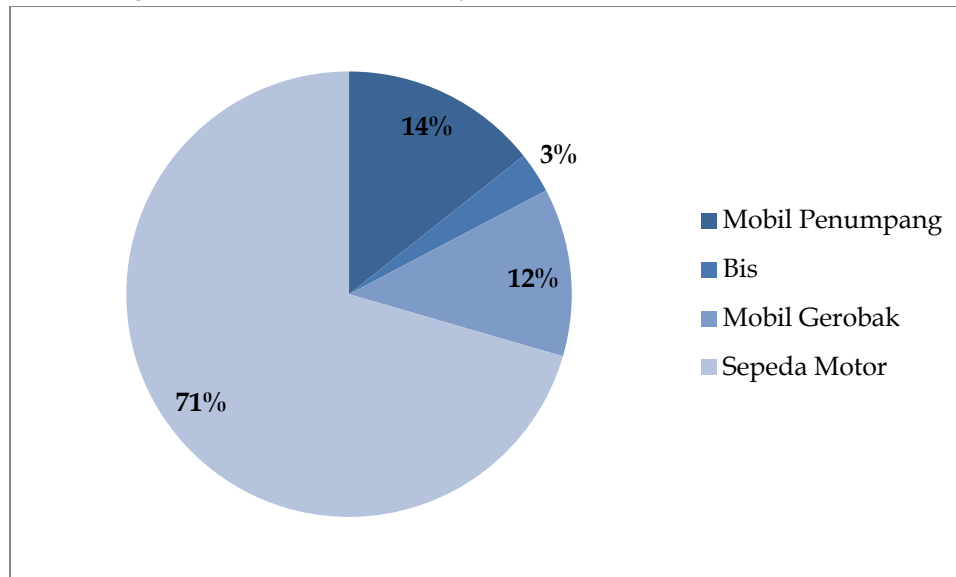
5. Grafik Lingkaran Tunggal (*Single Pie Chart*)

Contoh 5-1

Tabel 5-1 Banyaknya Kendaraan Bermotor di Indonesia menurut jenisnya, Tahun 1999 (dalam ribuan)

| Jenis Kendaraan | Mobil Penumpang | Bis | Mobil Gerobak | Sepeda Motor | Jumlah |
|-----------------|-----------------|-----|---------------|--------------|--------|
| (1) | (2) | (3) | (4) | (5) | (6) |
| Jumlah | 925 | 191 | 788 | 4551 | 6455 |

Grafik lingkaran dari tabel di atas, yaitu



6. Grafik Lingkaran Berganda (*Multiple Pie Chart*)

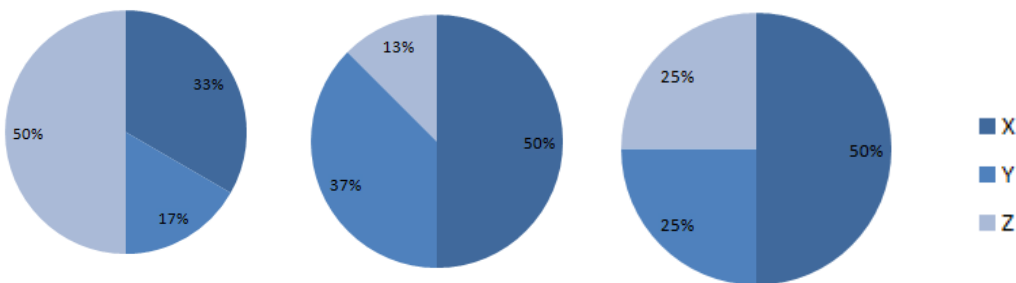
Contoh 5-2

Tabel 5-2 Hasil Tambang (Tambang X,Y, dan Z) dari Negara-negara A, B, dan C Tahun 1999 (dalam jutaan ton)

| Negara | Hasil Tambang | | | Jumlah |
|--------|---------------|-----|-----|--------|
| | X | Y | Z | |
| (1) | (2) | (3) | (4) | (5) |
| A | 4 | 2 | 6 | 12 |
| B | 8 | 6 | 2 | 16 |
| C | 10 | 5 | 5 | 20 |

Tabel Persentase

| Negara | Hasil Tambang | | | Jumlah |
|--------|---------------|------|------|--------|
| | X | Y | Z | |
| (1) | (2) | (3) | (4) | (5) |
| A | 33.3 | 16.7 | 50.0 | 12 |
| B | 50.0 | 37.5 | 12.5 | 16 |
| C | 50.0 | 25.0 | 25.0 | 20 |



Materi 6

Distribusi Frekuensi

Salah satu cara untuk meringkas data adalah dengan distribusi frekuensi, yaitu pengelompokan data ke dalam beberapa kelompok (kelas) dan kemudian dihitung banyaknya data yang masuk ke dalam tiap kelas.

Distribusi Frekuensi Data Kualitatif

Berikut ini disajikan sebuah tabel mengenai data 50 orang pembeli komputer dari lima jenis perusahaan komputer.

Tabel 6-1 Data mengenai 50 Orang Pembeli Komputer dari beberapa Jenis Perusahaan Komputer

| | | | | |
|--------------|--------|--------------|--------------|--------------|
| IBM | Compaq | Compaq | IBM | IBM |
| Compaq | Compaq | Packard Bell | Gateway 2000 | Packard Bell |
| Apple | Apple | IBM | Apple | Compaq |
| Packard Bell | Compaq | Compaq | IBM | Packard Bell |
| Gateway 2000 | Apple | Apple | Packard Bell | Compaq |
| IBM | Apple | Apple | Packard Bell | Packard Bell |
| Apple | Apple | Compaq | Gateway 2000 | Compaq |
| Packard Bell | IBM | Gateway 2000 | Compaq | Apple |
| Packard Bell | IBM | Packard Bell | Compaq | Packard Bell |
| Gateway 2000 | Apple | IBM | Apple | Apple |

Kita mengalami kesulitan untuk mengetahui jenis komputer mana yang paling banyak diminati. Untuk itu diperlukan distribusi frekuensi yang menunjukkan jumlah atau banyaknya item dari setiap kategori atau kelas.

Tabel 6-2 Distribusi Frekuensi Pembelian Komputer dari 5 Merek

| Perusahaan | Frekuensi |
|--------------|-----------|
| Apple | 13 |
| Compaq | 12 |
| Gateway 2000 | 15 |
| IBM | 9 |
| Packard Bell | 11 |
| Jumlah | 50 |

Dari distribusi frekuensi di atas, diperoleh merek komputer yang paling banyak diminati oleh pembeli adalah Apple dan yang paling sedikit diminati adalah Gateway 2000.

Distribusi Frekuensi Relatif dan Persentase Data Kualitatif

Frekuensi relatif dari suatu kelas adalah proporsi item dalam setiap kelas terhadap jumlah keseluruhan item dalam data tersebut. Jika sekelompok data memiliki *n* observasi, maka frekuensi relatif dari suatu kelas dihitung dengan rumus berikut.

$$\text{Frekuensi relatif suatu kelas} = \frac{\text{Frekuensi kelas}}{n}$$

Sedangkan frekuensi persentase dari suatu kelas adalah frekuensi relatif kelas tersebut dikalikan dengan 100.

$$\text{PERSENTASE} = \text{PROPORSI} \text{ dikalikan } 100$$

Distribusi frekuensi relatif adalah ringkasan dalam bentuk tabel dari sekelompok data yang menunjukkan frekuensi relatif bagi setiap kelas.

Tabel 6-3 Distribusi Frekuensi Relatif dan Persentase Pembelian Komputer

| Perusahaan | Frekuensi | Frekuensi Persentase |
|--------------|-----------|----------------------|
| Apple | 0.26 | 26 |
| Compaq | 0.24 | 24 |
| Gateway 2000 | 0.30 | 30 |
| IBM | 0.18 | 18 |
| Packard Bell | 0.22 | 22 |
| Jumlah | 1.00 | 100 |

Materi 7

Distribusi Frekuensi (Lanjutan)

Distribusi Frekuensi Data Kuantitatif

Ada tiga hal yang perlu diperhatikan dalam menentukan kelas bagi distribusi frekuensi untuk data kuantitatif, yaitu jumlah kelas, lebar kelas, dan batas kelas.

Jumlah Kelas

Banyaknya kelas sebaiknya antara 7 dan 15, atau paling banyak 20. Pada tahun 1926, H.A. Sturges menulis artikel dengan judul: "The Choice of a Class Interval" dalam Journal of the American Statistical Association, yang mengemukakan suatu rumus untuk menentukan banyaknya kelas.

$$k = 1 + 3,322 \log n$$

dengan

k = banyaknya kelas

n = banyaknya nilai observasi

rumus tersebut dinamakan Kriteria Sturges dan merupakan suatu perkiraan tentang banyaknya kelas.

Contoh 7-1

Data dengan $n=100$, maka banyaknya kelas k adalah

$$k = 1 + 3,322 \log 100 = 1 + 3,322(2) = 7,644$$

Jadi, banyaknya kelas sebaiknya 7.

Interval Kelas

Rumus untuk menentukan besarnya kelas (panjang interval)

$$c = \frac{X_n - X_1}{k}$$

Dengan:

c = perkiraan besarnya (*class, width, class size, class length*)

k = banyaknya kelas

X_n = nilai observasi terbesar

X_1 = nilai obeservasi terkecil

Batas Kelas

Batas kelas bawah menunjukkan kemungkinan nilai data terkecil pada suatu kelas. Sedangkan batas kelas atas mengidentifikasi kemungkinan nilai data terbesar dalam suatu kelas.

Jika diketahui kelas-kelas interval adalah 30-39, 40-49, 50-59, dan seterusnya, maka untuk nilai batas bawahnya (*lower limit*) adalah 30, 40, 50, dan seterusnya. Sedangkan nilai batas atasnya (*upper limit*) adalah 39, 49, 59, dan seterusnya.

Contoh 7-2

Suatu penelitian dilakukan oleh pejabat dari Badan Koordinasi Penanaman Modal (BKPM) terhadap 100 perusahaan. Salah satu karakteristik yang diteliti ialah besarnya modal yang dimiliki perusahaan-perusahaan tersebut. Apabila X adalah modal dalam jutaan rupiah, maka nilai X adalah sebagai berikut:

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| 75 | 86 | 66 | 86 | 50 | 78 | 66 | 79 | 68 | 60 |
| 80 | 83 | 87 | 79 | 80 | 77 | 81 | 92 | 57 | 52 |
| 58 | 82 | 73 | 95 | 66 | 60 | 84 | 80 | 79 | 63 |
| 80 | 88 | 58 | 84 | 96 | 87 | 72 | 65 | 79 | 80 |
| 86 | 68 | 76 | 41 | 80 | 40 | 63 | 90 | 83 | 94 |
| 76 | 66 | 74 | 76 | 68 | 82 | 59 | 75 | 35 | 34 |
| 65 | 63 | 85 | 87 | 79 | 77 | 76 | 74 | 76 | 78 |
| 75 | 60 | 96 | 74 | 73 | 87 | 52 | 98 | 88 | 64 |
| 76 | 69 | 60 | 74 | 72 | 76 | 57 | 64 | 67 | 58 |
| 72 | 80 | 72 | 56 | 73 | 82 | 78 | 45 | 75 | 56 |

Tabel 7-1 Frekuensi Modal Perusahaan

| Batas Kelas Modal (Jumlah Rp) | Nilai Tengah/Mean (M) | Frekuensi (f) |
|----------------------------------|--------------------------|------------------|
| (1) | (2) | (3) |
| 30 - 39 | 34.5 | 2 |
| 40 - 49 | 44.5 | 3 |
| 50 - 59 | 54.5 | 11 |
| 60 - 69 | 64.5 | 20 |
| 70 - 79 | 74.5 | 32 |
| 80 - 89 | 84.5 | 25 |
| 90 - 99 | 94.5 | 7 |
| Jumlah | | 100 |

Materi 8

Ukuran Pemusatan

Definisi Ukuran Pemusatan

Rata-rata (average) adalah nilai yang mewakili himpunan atau sekelompok data (a set of data)

Tabel 8-1 Hasil Ujian Hipotetis Toni dan Joni

| Mata pelajaran | Hasil Ujian Toni | Hasil Ujian Joni |
|----------------|------------------|------------------|
| | (X) | (Y) |
| Statistika | 8 | 7 |
| Matematika | 7 | 6 |
| Teori Ekonomi | 6 | 5 |
| Pemasaran | 8 | 6 |
| Metode Riset | 7 | 6 |
| Jumlah | 36 | 30 |
| Rata-rata | 7.2 | 6 |

Dari nilai rata-rata tersebut dapat disimpulkan bahwa Toni lebih pandai dari Joni.

Rata-rata Hitung

Apabila kita mempunyai variabel X , sebagai hasil pengamatan atau observasi sebanyak N kali, yaitu X_1, X_2, \dots, X_n , maka

a) Rata-rata sebenarnya (populasi):

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i = \frac{1}{N} (X_1 + X_2 + \dots + X_N)$$

Simbol rata-rata sebenarnya disebut parameter.

b) Rata-rata perkiraan (sampel):

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

Simbol rata-rata perkiraan disebut statistik.

Contoh 8-1

Berikut disajikan data penjualan suatu perusahaan selama 10 tahun.

X = hasil penjualan selama 10 tahun dalam jutaan rupiah.

X_1 = 50 (hasil penjualan tahun pertama)

X_2 = 60 (hasil penjualan tahun kedua)

X_3 = 40 (hasil penjualan tahun ketiga)

X_4 = 70 (hasil penjualan tahun keempat)

X_5 = 80 (hasil penjualan tahun kelima)

X_6 = 90 (hasil penjualan tahun keenam)

X_7 = 100 (hasil penjualan tahun ketujuh)

X_8 = 65 (hasil penjualan tahun kedelapan)

X_9 = 75 (hasil penjualan tahun kesembilan)

X_{10} = 85 (hasil penjualan tahun kesepuluh)

- a) Hitung rata-rata hasil penjualan sebenarnya.
- b) Ambil sampel sebanyak $n=5$, misalnya setelah diambil sampelnya diperoleh $X_2, X_4, X_5, X_8, X_{10}$.

Penyelesaian:

- a) Rata-ratanya:

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i = \frac{1}{10} (715) = 71,5$$

- b) Rata-rata perkiraan

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{5} (60 + 70 + 80 + 65 + 85) = 72$$

Rata-rata Hitung (Data Berkelompok)

Apabila data sudah disajikan dalam bentuk tabel frekuensi, dimana X_1 terjadi f_1 kali, X_2 terjadi f_2 kali, dan seterusnya sampai X_k terjadi f_k kali, maka rumus rata-rata dari data yang sudah dibuat tabel frekuensinya adalah sebagai berikut:

$$\bar{X} = \frac{\sum_{i=1}^k f_i X_i}{\sum_{i=1}^k f_i} \text{ atau } \bar{X} = \frac{\sum_{i=1}^k M_i f_i}{\sum_{i=1}^k f_i}$$

di mana M_i = nilai tengah kelas interval ke-I (untuk data berkelompok).

Contoh 8-2

| | | | | | | |
|---|---|---|---|---|---|---|
| X | 8 | 6 | 4 | 5 | 7 | 9 |
| F | 2 | 3 | 4 | 5 | 2 | 1 |

$$\bar{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{88}{15} = 5.87$$

Jadi, rata-rata dari data di atas adalah 5.87.

Contoh 8-3

| Berat Badan (kg) | Banyaknya Mahasiswa (f) |
|---------------------|----------------------------|
| 60-62 | 5 |
| 63-65 | 18 |
| 66-68 | 42 |
| 69-71 | 27 |
| 72-74 | 8 |

Penyelesaian:

| Berat Badan (kg) | M | f | Mf |
|------------------|----|----|------|
| 60-62 | 61 | 5 | 305 |
| 63-65 | 64 | 18 | 1152 |
| 66-68 | 67 | 42 | 2814 |
| 69-71 | 70 | 27 | 1890 |
| 72-74 | 73 | 8 | 584 |

$$M_1 = \frac{60+62}{2} = 61, \dots, M_5 = \frac{72+74}{2} = 73$$

$$\bar{X} = \frac{\sum M_i f_i}{\sum f_i} = \frac{6745}{100} = 67.45$$

Jadi, rata-rata perkiraan berat per mahasiswa adalah 67.45 kg.

Materi 9

Ukuran Pemusatan (Lanjutan)

Median (Data Tidak Berkelompok)

Apabila ada sekelompok nilai sebanyak n diurutkan mulai dari yang terkecil X_1 sampai dengan yang terbesar X_n , maka nilai yang ada di tengah disebut Median (Med) Setengah (=50%) dari nilai lebih kecil atau sama dengan median, sedangkan setengah lainnya lebih besar atau sama dengan median.

Untuk n Ganjil

Kalau k adalah suatu bilangan konstan dan n ganjil, maka

$$n = 2k + 1 \text{ atau } k = \frac{n-1}{2}$$

Kelompok nilai $X_1, X_2, \dots, X_{k-1}, X_k, X_{k+1}, \dots, X_n$

Median = X_{k+1} atau nilai yang ke $(k+1)$.

Contoh 9-1

Nilai ujian Linear Programming dari 9 mahasiswa FE-UI, masing-masing adalah sebagai berikut: 90, 70, 60, 75, 65, 80, 40, 45, 50.

Penyelesaian:

$$X_1 = 40, X_2 = 45, X_3 = 50, X_4 = 60, X_5 = 65, X_6 = 70, X_7 = 75, X_8 = 80, X_9 = 90.$$

Karena $n=9$, maka $k = \frac{9-1}{2} = 4$. Jadi Med = $X_{4+1} = X_5 = 65$.

Untuk n Genap

Jika k adalah suatu bilangan konstan dan n genap, maka selalu dapat ditulis $n = 2k$, atau $k = \frac{n}{2}$

. Misalkan $n = 8$, maka $k = 4$.

$$\text{Median} = \frac{1}{2}(X_k + X_{k+1})$$

Contoh 9-2

Nilai ujian Linear Programming dari 10 mahasiswa FE-UI, masing-masing adalah sebagai berikut: 90, 70, 60, 75, 65, 80, 40, 45, 50, 95.

Penyelesaian:

$$X_1 = 40, X_2 = 45, X_3 = 50, X_4 = 60, X_5 = 65, X_6 = 70, X_7 = 75, X_8 = 80, X_9 = 90, X_{10} = 95.$$

Karena $n=10$, maka

$$k = \frac{10}{2} = 5. \text{ Jadi Med} = \frac{1}{2}(X_5 + X_{5+1}) = \frac{1}{2}(X_5 + X_6) = \frac{1}{2}(65 + 70) = 67.5.$$

Median (Data Berkelompok)

Untuk data berkelompok, nilai median dapat dicari dengan

$$med = L_0 + c \left(\frac{\frac{n}{2} - (\sum f_i)_0}{f_m} \right)$$

Di mana:

L_0 = nilai batas bawah sebenarnya dari kelas yang memuat nilai median;

n = banyaknya observasi = jumlah semua frekuensi;

$(\sum f_i)_0$ = jumlah frekuensi dari semua kelas di bawah kelas yang mengandung median (kelas yang mengandung median tidak termasuk);

f_m = frekuensi dari kelas yang mengandung median;

c = besarnya interval kelas, jarak antara kelas yang satu dengan lainnya atau besarnya kelas yang mengandung median.

Contoh 9-3

Hitunglah median berat badan 100 orang mahasiswa.

| Berat Badan (kg) | Banyaknya Mahasiswa (f) |
|---------------------|----------------------------|
| 60-62 | 5 |
| 63-65 | 18 |
| 66-68 | 42 |
| 69-71 | 27 |
| 72-74 | 8 |
| Jumlah | 100 |

Penyelesaian:

Setengah dari observasi = $\frac{100}{2} = 50 \rightarrow f_1 + f_3 = 23$, untuk mencapai 50 masih kurang 27, sehingga perlu ditambah dengan frekuensi kelas ketiga. Jadi, median terletak pada kelas ke-3, yaitu kelas 66-68, setelah dikoreksi menjadi $65,5 - 68,5 \rightarrow c = 68,5 - 65,5 = 3$.

$$L_0 = 65,5, \quad \frac{n}{2} = 50, \quad (\sum f_i)_0 = 23, \quad f_m = 42$$

$$med = L_0 + c \left(\frac{\frac{n}{2} - (\sum f_i)_0}{f_m} \right) = 65,5 + 3 \left(\frac{50 - 23}{42} \right) = 65,64.$$

Materi 10

Ukuran Pemusatan (Lanjutan)

Modus (Data Tidak Berkelompok)

Modus dari suatu kelompok nilai adalah nilai kelompok tersebut yang mempunyai frekuensi tertinggi, atau nilai yang paling banyak terjadi di dalam suatu kelompok nilai.

| X | F |
|-------|-------|
| (1) | (2) |
| X_1 | f_1 |
| X_2 | f_2 |
| . | . |
| . | . |
| X_5 | f_5 |
| . | . |
| . | . |
| X_n | f_n |

$X_i = \text{modus} = \text{mod}$ kalau f_i mempunyai nilai terbesar dibandingkan dengan frekuensi lainnya

$$f_i > f_{i+1}$$

$$f_i > f_{i-1}$$

Contoh 10-1

Dari data berikut, apakah ada mod-nya? Kalau ada, tentukan nilainya.

- a) 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18
- b) 3, 5, 8, 10, 12, 15, 16
- c) 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9

Penyelesaian:

a) Tabel frekuensi:

| X | f |
|----|---|
| 2 | 2 |
| 5 | 1 |
| 7 | 1 |
| 9 | 3 |
| 10 | 2 |
| 11 | 1 |
| 12 | 1 |
| 18 | 1 |

Jadi, mod=9, sebab nilai observasi ini yang paling banyak atau mempunyai frekuensi terbesar.

b) Tabel frekuensi:

| X | f |
|----|---|
| 3 | 1 |
| 5 | 1 |
| 8 | 1 |
| 10 | 1 |
| 12 | 1 |
| 15 | 1 |
| 16 | 1 |

Karena semua nilai mempunyai frekuensi yang sama, maka distribusi ini tidak mempunyai mod.

c) Tabel frekuensi:

| X | f |
|---|---|
| 2 | 1 |
| 3 | 1 |
| 4 | 3 |
| 5 | 2 |
| 7 | 3 |
| 9 | 1 |

Karena terdapat dua nilai observasi yang mempunyai frekuensi terbanyak, maka distribusi memiliki dua mod, yaitu 4 dan 7.

Modus (Data Berkelompok)

Apabila data sudah dikelompokkan dan disajikan dalam tabel frekuensi, maka dalam mencari modulusnya dihitung dengan:

$$\text{mod} = L_0 + c \left(\frac{(f_1)_0}{(f_1)_0 + (f_2)_0} \right)$$

Di mana:

L_0 = nilai batas bawah sebenarnya dari kelas yang memuat nilai median;

f_{m0} = frekuensi kelas yang memuat modus;

$(f_1)_0 = f_{m0} - f_{(m0-1)}$ {selisih frekuensi kelas yang memuat modus dengan frekuensi kelas sebelumnya (bawahnya)};

$(f_2)_0 = f_{m0} - f_{(m0+1)}$ {selisih frekuensi kelas yang memuat modus dengan frekuensi kelas sesudahnya (atasnya)};

c = besarnya jarak antara suatu kelas dengan kelas sebelumnya, misalnya antara kelas kedua dengan kelas kesatu, kelas ketiga dengan kelas kedua atau antara a_2 dengan a_1 , a_3 dengan a_2 .

Contoh 10-2

| Kelas | f |
|---------|----|
| 30 - 39 | 4 |
| 40 - 49 | 6 |
| 50 - 59 | 8 |
| 60 - 69 | 12 |
| 70 - 79 | 9 |
| 80 - 89 | 7 |
| 90 - 99 | 4 |
| Jumlah | 50 |

Penyelesaian:

Dari tabel, $f_{m_0} = 12$ merupakan frekuensi dari kelas yang memuat modus (nilai tertinggi). Kelas interval yang memuat modus, mempunyai nilai batas bawah $\frac{1}{2}(59 + 60) = 59,5$ dan nilai batas atas $\frac{1}{2}(59 + 70) = 69,5$. Jadi, antara $59,5 - 69,5$ terdapat observasi $f_{m_0} = 12$.

$$c = 69,5 - 59,5 = 10$$

$$L_0 = 59,5$$

$$f_{(m_0-1)} = 8$$

$$f_{(m_0+1)} = 9$$

$$(f_1)_0 = 12 - 8 = 4$$

$$(f_2)_0 = 12 - 9 = 3$$

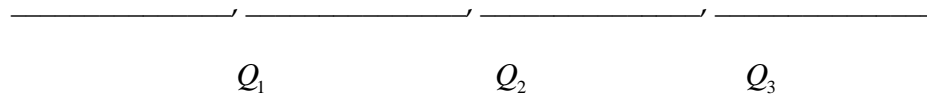
$$\text{mod} = L_0 + c \left(\frac{(f_1)_0}{(f_1)_0 + (f_2)_0} \right) = 59,5 + 10 \left(\frac{4}{4 + 3} \right) = 65,214$$

Materi 11

Ukuran Pemusatan (Lanjutan)

Distribusi yang Dibagi Menjadi 4, 10, 100 Bagian yang sama: Kuartil, Desil, dan Persentil (Data Tidak Berkelompok)

Untuk kelompok data dimana $n \geq 4$, kita tentukan tiga nilai, katakanlah Q_1, Q_2, Q_3 yang membagi kelompok data tersebut menjadi 4 bagian yang sama, yaitu setiap bagian memuat data yang sama. Nilai-nilai tersebut dinamakan kuartil pertama, kedua, dan ketiga. Pembagian itu adalah sedemikian rupa sehingga nilai 25% data/observasi sama atau lebih kecil dari Q_1 , 50% data/observasi sama atau lebih kecil dari Q_2 , 75% data/observasi sama atau lebih kecil dari Q_3 .



Dimana $Q_2 = \text{med}$.

Jika suatu kelompok data atau nilai sudah diurutkan dari yang terkecil (X_1) sampai yang terbesar (X_n), maka menghitung Q_1, Q_2 , dan Q_3

$$Q_i = \text{nilai yang ke-} \frac{i(n+1)}{4}, i = 1, 2, 3$$

Contoh 11-1

Berikut ini adalah data upah bulanan dari 13 karyawan dalam ribuan rupiah, yaitu 40, 30, 50, 65, 45, 55, 70, 60, 80, 35, 85, 95, 100, ($n=13$). Cari nilai Q_1, Q_2 , dan Q_3 .

Penyelesaian:

Pertama, data diurutkan:

$$X_1 = 30, X_2 = 35, X_3 = 40, X_4 = 45, X_5 = 50, X_6 = 55, X_7 = 60, X_8 = 65$$

$$X_9 = 70, X_{10} = 80, X_{11} = 85, X_{12} = 95, X_{13} = 100.$$

$$Q_1 = \text{nilai yang ke-} \frac{i(n+1)}{4} = \frac{1(13+1)}{4} = 3\frac{1}{2}. \text{ (Berarti rata-rata dari } X_3 \text{ dan } X_4 \text{).}$$

Jadi: $Q_1 = \frac{1}{2}(X_3 + X_4) = \frac{1}{2}(40 + 45) = 42,5$.

$Q_2 =$ nilai yang ke- $\frac{i(n+1)}{4} = \frac{2(13+1)}{4} = 7$.

Jadi: $Q_2 = X_7 = 60$.

$Q_2 =$ nilai yang ke- $\frac{i(n+1)}{4} = \frac{3(13+1)}{4} = 10\frac{1}{2}$. (Berarti rata-rata dari X_{10} dan X_{11}).

Jadi: $Q_3 = \frac{1}{2}(X_{10} + X_{11}) = \frac{1}{2}(80 + 85) = 82,5$.

Untuk kelompok data dimana $n \geq 10$, dapat ditentukan 9 nilai yang membagi kelompok data tersebut menjadi 10 bagian yang sama, misalnya D_1, D_2, \dots, D_9 , artinya setiap bagian mempunyai jumlah observasi yang sama sedemikian rupa sehingga nilai 10% observasi sama atau lebih kecil dari D_1 , nilai 20% observasi sama atau lebih kecil dari D_2 , dan seterusnya. Nilai tersebut dinamakan desil pertama, kedua, dan seterusnya sampai desil kesembilan. Kalau nilai kelompok data tersebut sudah diurutkan dari terkecil ($= X_1$) sampai yang terbesar (X_n), maka rumus desil

$$D_i = \text{nilai yang ke-} \frac{i(n+1)}{10}, i = 1, 2, 3, \dots, 9.$$

Terakhir, untuk kelompok data, dimana $n \geq 100$, dapat ditentukan 99 nilai P_1, P_2, \dots, P_{99} , yang disebut persentil pertama, kedua, dan seterusnya sampai persentil ke-99, yang membagi kelompok data tersebut menjadi 100 bagian; masing-masing mempunyai bagian dengan jumlah observasi yang sama, dan sedemikian rupa sehingga 1% dari observasi mempunyai nilai yang sama atau lebih kecil dari P_1 , 2% observasi mempunyai nilai yang sama atau lebih kecil dari P_2 , dan seterusnya. Kalau nilai kelompok data tersebut sudah diurutkan dari terkecil ($= X_1$) sampai yang terbesar (X_n), maka rumus persentil

$$P_i = \text{nilai yang ke-} \frac{i(n+1)}{100}, i = 1, 2, 3, \dots, 99.$$

Kuartil, Desil, dan Persentil (Data Berkelompok)

Untuk data berkelompok, data yang sudah dibuat tabel frekuensinya, maka rumus kuartil, desil, dan persentil adalah

Rumus Kuartil:

$$Q_i = L_0 + c \left(\frac{\frac{in}{4} - (\sum f_i)_0}{f_q} \right), i = 1, 2, 3.$$

Rumus Desil:

$$D_i = L_0 + c \left(\frac{\frac{in}{10} - (\sum f_i)_0}{f_d} \right), i = 1, 2, 3, \dots, 9.$$

Rumus Persentil:

$$P_i = L_0 + c \left(\frac{\frac{in}{100} - (\sum f_i)_0}{f_p} \right), i = 1, 2, 3, \dots, 99.$$

dimana

L_0 = nilai batas bawah dari kelas yang memuat kuartil ke-i;

n = banyaknya observasi = jumlah semua frekuensi;

$(\sum f_i)_0$ = jumlah frekuensi dari semua kelas sebelum kelas yang mengandung kuartil (atau desil atau persentil) ke-i (kelas yang mengandung kuartil (atau desil atau persentil) ke-i tidak termasuk);

f_q = frekuensi dari kelas yang mengandung kuartil ke-i;

f_d = frekuensi dari kelas yang mengandung desil ke-i;

f_p = frekuensi dari kelas yang mengandung persentil ke-i;

c = besarnya kelas yang mengandung kuartil ke-i;

$i = 1, 2, 3.$

in = i kali n.

Contoh 11-2

Berdasarkan data berikut, hitunglah $Q_1, D_6,$ dan P_{50} .

| Nilai Kelas | f |
|-------------|-----|
| (1) | (2) |
| 72,2 - 72,4 | 2 |
| 72,5 - 72,7 | 5 |
| 72,8 - 73,0 | 10 |
| 73,1 - 73,3 | 13 |
| 73,4 - 73,6 | 27 |
| 73,7 - 73,9 | 23 |
| 74,0 - 74,2 | 16 |
| 74,3 - 74,5 | 4 |
| Jumlah | 100 |

Penyelesaian:

Untuk menghitung $Q_1 : f_1 + f_2 + f_3 = 17$ belum mencapai 25% (25). Agar mencapai frekuensi 25, harus ikut dijumlahkan frekuensi kelas yang ke-4. Dengan demikian, diketahui kelas ke-4 memuat Q_1 . Dari data, $(\sum f_i)_0 = 17; n = 100, f_q = 13$. Nilai batas bawah dan batas atas dari kelas yang memuat Q_1 , masing-masing adalah $\frac{1}{2}(73,0 + 73,1) = 73,05$ dan $\frac{1}{2}(73,3 + 73,4) = 73,35$. Jadi, $c = 73,35 - 73,05 = 0,30$.

$$Q_1 = L_0 + c \left(\frac{\frac{in}{4} - (\sum f_i)_0}{f_q} \right) = 73,05 + 0,30 \left(\frac{\frac{1(100)}{4} - 17}{13} \right) = 73,23$$

Untuk menghitung $D_6 : f_1 + f_2 + f_3 + f_4 + f_5 = 57$ belum mencapai 60% (=60). Agar mencapai frekuensi 60, harus ikut dijumlahkan frekuensi kelas yang ke-6. Dengan demikian, diketahui kelas ke-6 memuat D_6 . Dari data, $(\sum f_i)_0 = 57; n = 100, f_d = 23$. Nilai batas bawah dan batas atas dari kelas yang memuat D_6 , masing-masing adalah $\frac{1}{2}(73,6 + 73,7) = 73,65$ dan $\frac{1}{2}(73,9 + 74,0) = 73,95$. Jadi, $c = 73,95 - 73,65 = 0,30$.

$$D_6 = L_0 + c \left(\frac{\frac{6n}{10} - (\sum f_i)_0}{f_d} \right) = 73,65 + 0,30 \left(\frac{\frac{6(100)}{10} - 57}{23} \right) = 73,69$$

Artinya nilai 60% dari observasi sama atau lebih kecil dari 73,69.

Untuk menghitung $P_{50} : f_1 + f_2 + f_3 + f_4 = 30$ belum mencapai 50% (=50). Agar mencapai frekuensi 50, harus ikut dijumlahkan frekuensi kelas yang ke-5. Dengan demikian, diketahui kelas ke-5 memuat P_{50} . Dari data, $(\sum f_i)_0 = 30; n = 100, f_p = 27$. Nilai batas bawah dan batas atas dari kelas yang memuat P_{50} , masing-masing adalah $\frac{1}{2}(73,3 + 73,4) = 73,35$ dan $\frac{1}{2}(73,6 + 73,7) = 73,65$. Jadi, $c = 73,65 - 73,35 = 0,30$.

$$P_{50} = L_0 + c \left(\frac{\frac{50n}{100} - (\sum f_i)_0}{f_p} \right) = 73,35 + 0,30 \left(\frac{\frac{5(100)}{100} - 30}{27} \right) = 73,57$$

Artinya nilai 50% dari observasi sama atau lebih kecil dari 73,57.

Materi 12

Ukuran Variasi atau Dispersi

Kita mengenal tiga kelompok nilai, yaitu kelompok nilai homogen (tidak bervariasi), kelompok nilai heterogen (sangat bervariasi), dan kelompok nilai relatif homogen (tidak begitu bervariasi).

Perhatikan 3 kelompok data berikut:

- (1) 50 50 50 50 50 - Rata-rata hitung = 50
- (2) 50 40 30 60 70 - Rata-rata hitung = 50
- (3) 100 40 80 20 10 - Rata-rata hitung = 50

Walaupun rata-rata hitung dari masing-masing kelompok adalah sama, namun kelompok (1) rata-ratanya dapat mewakili kelompok data dengan baik (sempurna), kelompok (2) cukup baik dan kelompok (3) tidak dapat mewakili dengan baik.

Ada beberapa macam ukuran variasi atau dispersi, misalnya nilai jarak (range), rata-rata simpangan (mean deviation), simpangan baku (standard deviation), dan koefisien variasi (coefficient of variation).

Pengukuran Dispersi Data Tidak Dikelompokkan

Nilai Jarak (*range*)

Apabila suatu kelompok nilai (data) sudah disusun menurut urutan yang terkecil (X_1) sampai dengan yang terbesar (X_n), maka untuk menghitung nilai jarak:

$$\text{Nilai jarak} = X_n - X_1$$

Atau

$$\text{Nilai jarak} = \text{Nilai Maksimum} - \text{Nilai Minimum}$$

Contoh 12-1

Carilah jarak dari data berikut:

50 40 30 60 70

Penyelesaian:

Pertama, urutkan terlebih dahulu:

$$X_1 = 30, X_2 = 40, X_3 = 50, X_4 = 60, X_5 = 70$$

$$\text{Nilai jarak} = X_5 - X_1 = 70 - 30 = 40.$$

Rata-rata simpangan

Apabila tersedia data $X_1, X_2, \dots, X_i, \dots, X_n$, dan rata-rata $\bar{X} = \frac{1}{n} \sum X_p$, maka simpangan terhadap rata-rata hitung diartikan sebagai berikut:

$$(X_1 - \bar{X}), (X_2 - \bar{X}), \dots, (X_i - \bar{X}), \dots, (X_n - \bar{X})$$

Rata-rata simpangan (RS) adalah rata-rata hitung dari nilai mutlak simpangan yang dirumuskan:

$$RS = \frac{1}{n} \sum |X_i - \bar{X}|$$

Contoh 12-2

Cari rata-rata simpangan, baik terhadap rata-rata hitung maupun terhadap median dari Contoh 6.1.

Penyelesaian:

$$\bar{X} = \frac{1}{5}(50 + 40 + 30 + 60 + 70) = 50, \text{ med} = 50.$$

$$RS = \frac{1}{n} \sum |X_i - \bar{X}| = \frac{1}{5}(|0| + |-10| + |-20| + |10| + |20|) = 12$$

$$RS = \frac{1}{n} \sum |X_i - \text{med}| = \frac{1}{5}(|0| + |-10| + |-20| + |10| + |20|) = 12$$

Simpangan Baku

Simpangan baku merupakan salah satu ukuran dispersi yang diperoleh dari akar kuadrat positif varians. Varians adalah rata-rata hitung dari kuadrat simpangan setiap pengamatan terhadap rata-rata hitungnya.

Apabila kita mempunyai suatu populasi dengan jumlah elemen sebanyak N dan sampel dengan n elemen, dan selanjutnya nilai suatu karakteristik tertentu kita kumpulkan (umur, nilai ujian), maka

Populasi: X_1, X_2, \dots, X_N

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i = \text{rata-rata sebenarnya dari } X \text{ (rata-rata populasi)}$$

Sampel: X_1, X_2, \dots, X_n

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \text{rata-rata perkiraan dari } X \text{ (rata-rata sampel)}$$

\bar{X} = adalah perkiraan dari μ .

Varians populasi adalah σ^2 , dihitung dengan

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$$

Varians sampel adalah S^2 , dihitung dengan

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Simpangan baku populasi adalah σ . Sedangkan simpangan baku sampel adalah S .

Pengukuran Dispersi Data Dikelompokkan

Nilai Jarak

Untuk data berkelompok, nilai jarak dapat dihitung dengan dua cara:

- a) Nilai jarak = Nilai tengah kelas terakhir - Nilai tengah kelas pertama
- b) Nilai jarak = Batas atas kelas terakhir - Batas bawah kelas pertama

Contoh 12-3

Hitung nilai jarak dari berat badan 100 mahasiswa.

| Berat Badan (kg) | Banyaknya Mahasiswa (f) |
|---------------------|----------------------------|
| 60-62 | 5 |
| 63-65 | 18 |
| 66-68 | 42 |
| 69-71 | 27 |
| 72-74 | 8 |

Penyelesaian:

Cara I:

$$\text{Nilai tengah kelas terakhir} = \frac{72 + 74}{2} = 73 \text{ kg}$$

$$\text{Nilai tengah kelas pertama} = \frac{60 + 62}{2} = 61 \text{ kg}$$

Maka:

$$\text{Nilai jarak} = \text{Nilai tengah kelas terakhir} - \text{Nilai tengah kelas pertama} = 73 - 61 = 12 \text{ kg.}$$

Cara II:

$$\text{Batas atas kelas terakhir} = 74,5 \text{ kg}$$

$$\text{Batas bawah kelas pertama} = 59,5 \text{ kg}$$

Maka:

$$\text{Nilai jarak} = \text{Batas atas kelas terakhir} - \text{Batas bawah kelas pertama} = 74,5 - 59,5 = 15 \text{ kg.}$$

Simpangan Baku

Untuk data berkelompok dan sudah disajikan dalam tabel frekuensi, rumus simpangan baku populasi adalah

$$\sigma = \sqrt{\frac{\sum_{i=1}^k f_i (M_i - \mu)^2}{N}}$$

M_i = nilai tengah dari kelas ke- i , $i=1,2,\dots,k$

atau

$$\sigma = c \sqrt{\frac{\sum_{i=1}^k f_i d_i^2}{N} \pm \left[\frac{\sum_{i=1}^k f_i d_i}{N} \right]^2}, \text{ untuk interval yang sama.}$$

dimana:

c = besarnya kelas interval

f_i =frekuensi ke- i

d_i =deviasi = simpangan dari kelas ke- i terhadap titik asal asumsi

dan

$$\sigma = \sqrt{\frac{1}{N} \left(\sum_{i=1}^k f_i M_i^2 - \frac{\left(\sum_{i=1}^k f_i M_i \right)^2}{N} \right)}, \text{ untuk interval kelas yang tidak sama}$$

M_i = nilai tengah dari kelas ke- i .

Untuk data sampel diperoleh simpangan baku sampel

$$S = c \sqrt{\frac{\sum_{i=1}^k f_i d_i^2}{n-1} \pm \left[\frac{\sum_{i=1}^k f_i d_i}{n-1} \right]^2}, \text{ untuk interval yang sama.}$$

dan

$$S = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^k f_i M_i^2 - \frac{\left(\sum_{i=1}^k f_i M_i \right)^2}{n-1} \right)}, \text{ untuk interval kelas yang tidak sama}$$

Contoh 12-4

Hitunglah simpangan baku dari data berikut:

(1) 50 50 50 50 50

(2) 50 40 30 60 70

(3) 100 40 80 20 10

Penyelesaian:

$$\sigma_1 = 0$$

$$\sigma_2 = 14,14 (= Rp14.140)$$

$$\sigma_3 = 36,64 (= Rp36.640)$$

Ternyata hasil perhitungan di atas menunjukkan bahwa data yang heterogen mempunyai simpangan baku yang besar.

$$\sigma_1 < \sigma_2 < \sigma_3 \rightarrow 0 < 14,14 < 36,64$$

Materi 13

Analisis Korelasi

Pada semua kejadian, pasti ada faktor yang menyebabkan terjadinya kejadian-kejadian tersebut. Misalnya meningkatnya tekanan darah mungkin karena berat badan bertambah atau meningkatnya harga bahan makanan mungkin karena kenaikan harga minyak.

Hal ini menunjukkan adanya hubungan (korelasi) antara kejadian satu dengan kejadian lainnya. Kejadian itu dapat dinyatakan dengan perubahan nilai variabel. Misalnya, jika X = variabel harga, maka naik turunnya harga dapat dinyatakan dengan perubahan nilai X .

Korelasi Data Kuantitatif

Koefisien Korelasi dan Kegunaannya

Hubungan dua variabel ada yang positif dan negatif. Hubungan X dan Y dikatakan positif apabila kenaikan (penurunan) X pada umumnya diikuti oleh kenaikan (penurunan) Y . Sebaliknya dikatakan negatif bila kenaikan (penurunan) X pada umumnya diikuti oleh penurunan (kenaikan) Y .

Kuat dan tidaknya hubungan antara X dan Y apabila dinyatakan dengan fungsi linear (paling tidak mendekati), diukur dengan suatu nilai yang disebut Koefisien Korelasi, dinotasikan dengan r dengan

$$-1 < r < 1$$

Artinya:

- Jika $r = 1$, hubungan X dan Y sempurna dan positif (mendekati 1, yaitu hubungan sangat kuat dan positif).
- Jika $r = -1$, hubungan X dan Y sempurna dan negatif (mendekati -1, yaitu hubungan sangat kuat dan negatif).
- Jika $r = 0$, hubungan X dan Y lemah sekali atau tidak ada hubungan.
- Kalau $r = 0$, X dan Y tidak berkorelasi.
- $< 0,5$, hubungan X dan Y lemah.
- $0,5 < 0,75$, hubungan X dan Y cukup kuat.
- $0,75 < 0,9$, hubungan X dan Y kuat.
- $0,9 < 1$, hubungan X dan Y sangat kuat.
- 1, hubungan X dan Y sempurna +.
- -1, hubungan X dan Y sempurna -.

X dikatakan mempengaruhi Y jika berubahnya nilai X akan menyebabkan perubahan nilai Y; artinya naik turunnya X akan membuat nilai Y juga naik/turun.

Ketika perubahan nilai Y disebabkan oleh perubahan X, maka besar kontribusi dari X terhadap Y dihitung dengan koefisien penentuan (coefficient of determination)

$$KP = r^2$$

Jika $r = 0,9$, maka $KP = (0,9)^2 = 0,81 (= 81\%)$, yaitu besarnya sumbangan variabel X terhadap variasi atau naik/turunnya Y adalah 81%, sedangkan 19% disebabkan oleh faktor yang lain. Rumus untuk menghitung r :

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2} \sqrt{n \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i\right)^2}}$$

Contoh 13-1

Jika X adalah persentase kenaikan harga, sedangkan Y adalah persentase kenaikan hasil penjualan.

Tabel 13-1

| | | | | | | | | | | |
|---|----|----|----|----|---|----|----|----|----|----|
| X | 2 | 4 | 5 | 6 | 8 | 10 | 11 | 13 | 14 | 15 |
| Y | 15 | 14 | 12 | 10 | 9 | 8 | 6 | 4 | 3 | 2 |

Penyelesaian:

Tabel 13-2

| X | Y | X ² | Y ² | XY |
|----|----|----------------|----------------|----|
| 2 | 15 | 4 | 225 | 30 |
| 4 | 14 | 16 | 196 | 56 |
| 5 | 12 | 25 | 144 | 60 |
| 6 | 10 | 36 | 100 | 60 |
| 8 | 9 | 64 | 81 | 72 |
| 10 | 8 | 100 | 64 | 80 |
| 11 | 6 | 121 | 36 | 66 |
| 13 | 4 | 169 | 16 | 52 |
| 14 | 3 | 196 | 9 | 42 |

| | | | | |
|----|---|-----|---|----|
| 15 | 2 | 225 | 4 | 30 |
| | | | | |

$$r = \frac{10 \sum_{i=1}^{10} X_i Y_i - \sum_{i=1}^{10} X_i \sum_{i=1}^{10} Y_i}{\sqrt{10 \sum_{i=1}^{10} X_i^2 - \left(\sum_{i=1}^{10} X_i \right)^2} \sqrt{10 \sum_{i=1}^{10} Y_i^2 - \left(\sum_{i=1}^{10} Y_i \right)^2}} = \frac{10(548) - (88)(83)}{\sqrt{10(956) - (88)^2} \sqrt{10(875) - (83)^2}} = -0,99$$

Kesimpulan: Hubungan X dan Y kuat dan negatif. Dengan demikian, nilai $KP = (0,99)^2 = 0,9801 = 0,98 = 98\%$.

Dalam hal ini, harga mempunyai pengaruh yang negatif terhadap penjualan; maksudnya, kenaikan harga pada umumnya menyebabkan hasil penjualan menurun.

Koefisien Korelasi Data Berkelompok

Rumus untuk menghitung koefisien korelasi untuk data berkelompok, yaitu

$$r = \frac{n(\sum uvf) - (\sum uf_u)(\sum vf_v)}{\sqrt{n(\sum u^2 f_u) - (\sum uf_u)^2} \sqrt{n(\sum v^2 f_v) - (\sum vf_v)^2}}$$

Contoh 13-2

Misalkan 100 orang mahasiswa Akademi Ilmu Statistik, Jakarta, mengambil ujian, di antaranya mata pelajaran statistik dan matematika. Ada suatu pendapat bahwa pada umumnya mahasiswa yang kemampuan matematikanya lemah akan mengalami kesukaran dalam belajar statistik. Berarti, jika nilai matematika yang diperoleh tinggi, nilai statistik juga tinggi. Sebaliknya, jika nilai matematikanya rendah, nilai statistiknya juga rendah. Dengan kata lain. Ada hubungan positif antara nilai ujian matematika (X) dan nilai ujian (Y).

Tabel 13-3

| Matematika | 40 - 49 | 50 - 59 | 60 - 69 | 70 - 79 | 80 - 89 | 90 - 99 | Jumlah |
|------------|---------|---------|---------|---------|---------|---------|--------|
| Statistika | | | | | | | |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 90 - 99 | | | | 2 | 4 | 4 | 10 |
| 80 - 89 | | | 1 | 4 | 6 | 5 | 16 |
| 70 - 79 | | | 5 | 10 | 8 | 1 | 24 |
| 60 - 69 | 1 | 4 | 9 | 5 | 2 | | 21 |
| 50 - 59 | 3 | 6 | 6 | 2 | | | 17 |

| | | | | | | | |
|---------|---|----|----|----|----|----|-----|
| 40 - 49 | 3 | 5 | 4 | | | | 12 |
| Jumlah | 7 | 15 | 25 | 23 | 20 | 10 | 100 |

Cara membaca tabel:

Jumlah mahasiswa yang mendapat nilai matematika antara 70-79 dan nilai statistik antara 70-79 adalah 10 orang.

Jumlah mahasiswa yang mendapat nilai matematika antara 40-49 dan nilai statistik antara 40-49 adalah 3 orang.

Tabel 13-4

| Kelas Nilai Matematika | Nilai Tengah (X) | u | f_u | Kelas Nilai Statistika | Nilai Tengah (Y) | v | f_v |
|------------------------|------------------|-----|-------|------------------------|------------------|-----|-------|
| (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| 40 - 49 | 44,5 | -2 | 7 | 90 - 99 | 94,5 | 2 | 10 |
| 50 - 59 | 54,5 | -1 | 15 | 80 - 89 | 84,5 | 1 | 16 |
| 60 - 69 | 64,5 | 0 | 25 | 70 - 79 | 74,5 | 0 | 24 |
| 70 - 79 | 74,5 | 1 | 23 | 60 - 69 | 64,5 | -1 | 21 |
| 80 - 89 | 84,5 | 2 | 20 | 50 - 59 | 54,5 | -2 | 17 |
| 90 - 99 | 94,5 | 3 | 10 | 40 - 49 | 44,5 | -3 | 12 |

u dan v adalah skala baru dari X dan Y , dan merupakan simbol untuk menyederhanakan perhitungan. Untuk data genap (contoh 6 kelas), nilai nol dipilih kelas yang paling dekat dengan nilai tengah, misalnya $X > 64,5$ atau $74,5$, disamakan 0, jadi $u = 0$, demikian juga untuk Y . f_u = frekuensi untuk X atau u , dan f_v = frekuensi untuk Y atau v .

Tabel 13-5 Tabel Korelasi

| | | | | | | (I) v | (II) f_v | (III) vf_v | (IV) v^2f_v | (V) uvf |
|-------|----|----|----|---|---|----------|---------------|-----------------|------------------|------------|
| | | | 2 | 4 | 4 | 2 | 10 | 20 | 40 | 44 |
| | | 1 | 4 | 6 | 5 | 1 | 16 | 16 | 16 | 31 |
| | | 5 | 10 | 8 | 1 | 0 | 24 | 0 | 0 | 0 |
| 1 | 4 | 9 | 5 | 2 | | -1 | 21 | -21 | 21 | -3 |
| 3 | 6 | 6 | 2 | | | -2 | 17 | -34 | 68 | 20 |
| 3 | 5 | 4 | | | | -3 | 12 | -36 | 108 | 33 |
| (I) u | -2 | -1 | 0 | 1 | 2 | 3 | | 100 | -55 | 253 |

| | | | | | | | | |
|-------|----------|-----|-----|----|----|----|----|-----|
| (II) | f_u | 7 | 15 | 25 | 23 | 20 | 10 | 100 |
| (III) | uf_u | -14 | -15 | 0 | 23 | 40 | 30 | 64 |
| (IV) | u^2f_u | 28 | 15 | 0 | 23 | 80 | 90 | 0 |
| (V) | uvf | 32 | 31 | 0 | -1 | 24 | 39 | 125 |

Dari tabel korelasi di atas dapat disimpulkan:

$$\begin{array}{lll} n = 100 & \sum uf_u = 64 & \sum u^2 f_u = 236 \\ \sum uvf = 125 & \sum vf_v = -55 & \sum v^2 f_v = 253 \end{array}$$

Jadi,

$$\begin{aligned} r &= \frac{n(\sum uvf) - (\sum uf_u)(\sum vf_v)}{\sqrt{n(\sum u^2 f_u) - (\sum uf_u)^2} \sqrt{n(\sum v^2 f_v) - (\sum vf_v)^2}} \\ &= \frac{100(125) - (64)(-55)}{\sqrt{100(236) - (64)^2} \sqrt{100(253) - (-55)^2}} = 0,7686 \text{ atau } 0,77 \end{aligned}$$

Kesimpulan: Hubungan antara nilai matematika dan statistika cukup kuat dan positif. Artinya, nilai matematika yang diperoleh akan mempengaruhi nilai statistika. (Pada umumnya mahasiswa dengan nilai matematika yang rendah akan memperoleh nilai ujian statistik yang rendah; sebaliknya jika nilai matematikanya tinggi, maka nilai statistika yang akan dicapai juga tinggi).

Korelasi Data Kualitatif

Untuk data kualitatif yang dipergunakan dalam mengukur kuatnya hubungan disebut Contingency Coefficient (Koefisien Bersyarat), dinotasikan C_c :

$$C_c = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Jika nilai C_c sebesar nol, berarti tidak ada hubungan. Nilai tertinggi C_c adalah $\sqrt{(r-1)/r}$ di mana r adalah banyaknya baris atau kolom. Jika banyaknya baris atau kolom tidak sama, pilih nilai yang terkecil.

$$n = \sum_{i=1}^p \sum_{j=1}^q f_{ij} = \sum_{i=1}^p n_{i.} = \sum_{j=1}^q n_{.j} = \sum_{i=1}^p \sum_{j=1}^q n_{ij} = \text{banyaknya observasi}$$

$$\chi^2 = \sum_{i=1}^p \sum_{j=1}^q \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Di mana $f_{ij} = n_{ij}$ = frekuensi atau banyaknya observasi baris i kolom j ;

$i = 1, 2, \dots, p$;

$j = 1, 2, \dots, q$.

Tabel 13-6

| I | II | | 1 | 2 | ... | j | ... | q | |
|-----|--------------------------|--------------------------|-----|-----|-----|--------------------------|-----|--------------------------|----------|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | | |
| 1 | f_{11} (e_{11}) | f_{12} (e_{12}) | | | | f_{1j} (e_{1j}) | | f_{1q} (e_{1q}) | $n_{1.}$ |
| 2 | f_{21} (e_{21}) | f_{22} (e_{22}) | | | | f_{2j} (e_{2j}) | | f_{2q} (e_{2q}) | $n_{2.}$ |
| ... | | | | | | | | | |
| i | f_{i1} (e_{i1}) | f_{i2} (e_{i2}) | | | | f_{ij} (e_{ij}) | | f_{iq} (e_{iq}) | $n_{i.}$ |
| ... | | | | | | | | | |
| p | f_{p1} (e_{p1}) | f_{p2} (e_{p2}) | | | | f_{pj} (e_{pj}) | | f_{pq} (e_{pq}) | $n_{p.}$ |
| | $n_{.1}$ | $n_{.2}$ | | | | $n_{.j}$ | | $n_{.q}$ | n |

f_{ij} = frekuensi kategori i dan j

e_{ij} = frekuensi harapan kategori i dan j

$$e_{ij} = \frac{(n_{i.})(n_{.j})}{n} = \text{frekuensi harapan (expected frequency)}$$

$$n_{i.} = \sum_{j=1}^q f_{ij} \Rightarrow \sum_{i=1}^p n_{i.} = n$$

$$n_{.j} = \sum_{i=1}^p f_{ij} \Rightarrow \sum_{j=1}^q n_{.j} = n$$

- Jika nilai perbandingan C_c dengan batas tertinggi $< 0,50$, maka hubungannya lemah, antara 0,50 dan 0,75 maka hubungannya sedang/cukup,
- antara 0,75 dan 0,90 maka hubungannya kuat,
- antara 0,90 dan 1 maka hubungannya sangat kuat,
- sama dengan 1 maka hubungannya sempurna.

Contoh 13-3

Para mahasiswa pada suatu universitas di Jakarta mengambil kuliah di 4 fakultas, yaitu Ekonomi, Hukum, Teknik, dan Kedokteran. Mereka kemudian digolongkan menjadi 3 kategori: pertama mereka yang sudah membayar uang kuliah. Kedua, yang sudah membayar tetapi belum lunas. Ketiga, mereka yang sama sekali masih menunggak. Perhatikan data berikut:

Tabel 13-7

| Kategori | FE | FH | FT | FK |
|----------------|-----|-----|-----|-----|
| (1) | (2) | (3) | (4) | (5) |
| Sudah membayar | 42 | 31 | 56 | 28 |
| Belum lunas | 16 | 82 | 47 | 21 |
| belum membayar | 13 | 26 | 39 | 19 |

Penyelesaian:

Tabel 13-8

| Kategori | FE | FH | FT | FK | Jumlah |
|----------|-------------|--------------|--------------|-------------|--------------|
| (1) | (2) | (3) | (4) | (5) | (6) |
| 1 | 42 (26,54) | 31 (51,96) | 56 (53,08) | 28 (25,42) | $n_{1.}=157$ |
| 2 | 16 (28,06) | 82 (54,96) | 47 (56,12) | 21 (26,88) | $n_{2.}=166$ |
| 3 | 13 (16,40) | 26 (32,10) | 39 (32,80) | 19 (15,70) | $n_{3.}=97$ |
| Jumlah | $n_{.1}=71$ | $n_{.2}=139$ | $n_{.3}=142$ | $n_{.4}=68$ | $n=420$ |

$$e_{11} = \frac{(n_{1.})(n_{.1})}{n} = \frac{(157)(71)}{420} = 26,54;$$

$$e_{12} = \frac{(n_{1.})(n_{.2})}{n} = \frac{(157)(139)}{420} = 51,96$$

$$e_{13} = \frac{(n_{1.})(n_{.3})}{n} = \frac{(157)(142)}{420} = 53,08$$

$$e_{14} = \frac{(n_{1.})(n_{.4})}{n} = \frac{(157)(68)}{420} = 25,42$$

$$e_{21} = \frac{(n_{2.})(n_{.1})}{n} = \frac{(166)(71)}{420} = 28,06$$

$$e_{22} = \frac{(n_{2.})(n_{.2})}{n} = \frac{(166)(139)}{420} = 54,94$$

$$e_{23} = \frac{(n_{2.})(n_{.3})}{n} = \frac{(166)(142)}{420} = 56,12$$

$$e_{24} = \frac{(n_{2.})(n_{.4})}{n} = \frac{(166)(68)}{420} = 26,88$$

$$e_{31} = \frac{(n_{3.})(n_{.1})}{n} = \frac{(97)(71)}{420} = 16,40$$

$$e_{32} = \frac{(n_{3.})(n_{.2})}{n} = \frac{(97)(139)}{420} = 32,10$$

$$e_{33} = \frac{(n_{3.})(n_{.3})}{n} = \frac{(97)(142)}{420} = 32,80$$

$$e_{34} = \frac{(n_{3.})(n_{.4})}{n} = \frac{(97)(68)}{420} = 15,70$$

$$\chi^2 = \sum_{i=1}^3 \sum_{j=1}^4 \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = 42,86$$

$$C_c = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{42,86}{42,86 + 420}} = 0,30$$

Batas atas = $\sqrt{(3-1)/3} = 0,82$. Perbandingan C_c dengan batas atas adalah $(0,30)/0,82 = 0,36$.

Karena $0,36 < 0,50$, maka hubungannya dikatakan lemah.

Materi 14

Analisis Regresi

Analisis korelasi bertujuan untuk mengetahui kuatnya hubungan antara X dan Y. Sedangkan analisis regresi bertujuan untuk mengetahui pengaruh dari X terhadap Y jika naik 1 unit dan untuk meramalkan nilai Y jika variabel X yang berkorelasi dengan Y nilainya sudah diketahui dengan menggunakan persamaan regresi linear sederhana.

Persamaan Regresi Linear

Bentuk persamaan regresi linear

$$Y' = a + bX$$

dimana

a = Y pintasan, (nilai Y' bila $X=0$)

b = kemiringan dari garis regresi (kenaikan atau penurunan Y' untuk setiap perubahan satu satuan X) atau koefisien regresi, yang mengukur besarnya pengaruh X terhadap Y kalau X naik satu unit.

X = nilai tertentu dari variabel bebas

Y' = dibaca “Y aksen” artinya nilai yang diukur/dihitung pada variabel tidak bebas.

Nilai dari a dan b pada persamaan regresi dapat dihitung dengan rumus berikut:

$$b = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$a = \bar{Y} - b\bar{X}$$

Contoh 14-1

Tabel 14-1 Hasil Produksi dan Skor Tes Kecerdasan dari 8 Karyawan Pabrik Mainan Anak-anak “Trackey”.

| Karyawan | Hasil Produksi (lusin) (Y) | Skor Tes Kecerdasan (X) |
|----------|----------------------------|-------------------------|
| A | 30 | 6 |

| | | |
|---|----|----|
| B | 49 | 9 |
| C | 18 | 3 |
| D | 42 | 8 |
| E | 39 | 7 |
| F | 25 | 5 |
| G | 41 | 8 |
| H | 52 | 10 |

Penyelesaian:

| Y | X | X ² | XY |
|----|----|----------------|-----|
| 30 | 6 | 36 | 180 |
| 49 | 9 | 81 | 441 |
| 18 | 3 | 9 | 54 |
| 42 | 8 | 64 | 336 |
| 39 | 7 | 49 | 273 |
| 25 | 5 | 25 | 125 |
| 41 | 8 | 64 | 328 |
| 52 | 10 | 100 | 520 |

Dari tabel di atas, diperoleh

$$\begin{array}{lll}
 n = 8 & \sum Y_i = 296 & \bar{X} = 7 \\
 \sum X_i = 56 & \sum X_i Y_i = 2257 & \bar{Y} = 37
 \end{array}$$

Sehingga,

$$b = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{(8)(2257) - (56)(296)}{(8)(428) - (56)^2} = 5,24$$

$$a = \bar{Y} - b\bar{X} = 37 - (5,24)(7) = 0,32$$

Persamaan regresi yang memperlihatkan kedua variabel antara hasil produksi dan hasil tes kecerdasan karyawan pada pabrik Mainan Anak-anak "Tackey" adalah:

$$Y' = 0,32 + 5,24X$$

Adapun tujuan utama penggunaan persamaan regresi adalah untuk memperkirakan nilai dari variabel tak bebas pada nilai variabel bebas tertentu. Misalnya seorang manajer akan

mempekerjakan seorang calon karyawan dengan nilai tes kecerdasan 4. Sedangkan supervisor pada departemen tersebut menginginkan seorang karyawan baru yang dapat memproduksi pada tingkat minimal 30 lusin.

$$Y' = 0,32 + 5,24X = 0,32 + 5,25(4) = 21$$

Dari hasil perhitungan, diramalkan bahwa calon karyawan dengan nilai tes kecerdasan 4 tidak dapat memproduksi minimal 30 lusin.

Referensi

1. Sudjana. 2005. *Metoda Statistika*. Bandung: Tarsito.
2. Supranto, J. 2016. *Statistik Teori dan Aplikasi*. Jakarta: Erlangga.
3. Weis, Neil A. 2012. *Elementary Statistics*. United State of America: Addison-Wesley.