

MSIM4310
Edisi 1

MODUL 01

Prinsip Dasar Penyajian Data

Dr. Dewi Juliah Ratnaningsih, S.Si., M.Si.
Dr. R. Bagus Fajriya Hakim, S.Si., M.Si.

Daftar Isi

Modul 01	1.1
Prinsip Dasar Penyajian Data	
Kegiatan Belajar 1	1.5
Ukuran Pemusatan dan Penyebaran Data	
Latihan	1.11
Rangkuman	1.12
Tes Formatif 1	1.13
Kegiatan Belajar 2	1.16
Bentuk-bentuk Penyajian Data	
Latihan	1.26
Rangkuman	1.27
Tes Formatif 2	1.28
Kegiatan Belajar 3	1.31
Membandingkan Kelompok Data	
Latihan	1.34
Rangkuman	1.35
Tes Formatif 3	1.36
Kunci Jawaban Tes Formatif	1.39
Daftar Pustaka	1.40



Pendahuluan

Data merupakan sekumpulan informasi atau keterangan-keterangan mengenai suatu hal yang diperoleh melalui suatu pengamatan atau pencarian dari sumber-sumber tertentu. Data yang diperoleh tersebut harus diolah dan disajikan dengan baik, menarik, dan tepat agar mudah dipahami dan memberikan informasi yang bermakna bagi yang membutuhkan. Dalam hal ini, penyajian data memegang peranan yang sangat penting.

Data memiliki kekuatan (*power*) sebagai sebuah informasi yang dapat mengungkap berbagai kendala atau masalah yang terjadi. Dengan menggunakan data, sebuah perusahaan/*startup* dapat membidik pelanggan (*customer*) dengan tepat sasaran dan tepat guna, sehingga menghasilkan keuntungan yang luar biasa. Kepiawaian dalam membaca dan menyajikan data dalam hal ini sangat diperlukan untuk dapat dijadikan sebagai informasi kunci untuk meningkatkan keuntungan (*profit*) bagi suatu perusahaan. Oleh karena itu, dewasa ini tidak mengherankan jika data memiliki nilai yang tinggi untuk mengungkap berbagai informasi dalam upaya peningkatan layanan terhadap pelanggan dan sebagai sarana untuk meningkatkan keuntungan perusahaan.

Bagaimana data dapat disajikan dengan baik, menarik, dan tepat diperlukan suatu pemahaman mengenai data itu sendiri dan siapa yang akan menggunakan data tersebut. Oleh karena itu, pemahaman terhadap prinsip-prinsip penyajian data sangat diperlukan. Penyajian data merupakan bagian dari statistika deskriptif. Sebagaimana diketahui dan mungkin sudah dipelajari dalam mata kuliah sebelumnya (Metode Statistika I/SATS4121) bahwa statistika diklasifikasikan ke dalam 2 kelompok yaitu: statistika deskriptif dan statistika inferensia. Statistika deskriptif secara umum meliputi dua bagian, yaitu penyajian data dan ukuran-ukuran numerik. Ukuran-ukuran numerik dibagi menjadi dua bagian, yaitu ukuran pemusatan dan ukuran penyebaran data. Melalui statistika deskriptif, kita dapat mengeksplorasi data seoptimal mungkin sehingga dapat diperoleh informasi yang lebih bermakna. Penyajian data yang tepat akan memudahkan pengguna dalam memahami informasi yang terkandung dalam data. Oleh karena itu, penyajian data merupakan salah satu bagian penting dalam eksplorasi data.

Buku Materi Pokok (BMP) Analisis dan Visualisasi Data ini akan memberikan berbagai materi terkait data dan penyajian data. Pada Modul 1 akan dibahas beberapa materi mengenai Prinsip Dasar Penyajian Data. Materi dalam Modul 1 ini akan diberikan dalam tiga Kegiatan Belajar:

1. Kegiatan Belajar 1 : Ukuran Pemusatan dan Penyebaran Data
2. Kegiatan Belajar 2 : Bentuk-bentuk Penyajian Data
3. Kegiatan Belajar 3 : Membandingkan Kelompok Data

Setelah mempelajari modul ini diharapkan Anda dapat menjelaskan:

1. pengertian penyajian data statistik;
2. prinsip-prinsip dasar penyajian data;
3. ukuran pemusatan dan penyebaran data;
4. bentuk-bentuk penyajian data;
5. perbandingan kelompok data.

Kegiatan
Belajar

1

Ukuran Pemusatan dan Penyebaran Data

Mahasiswa yang berbahagia, di bagian pendahuluan telah disinggung bahwa ukuran numerik dalam statistika deskriptif ada dua yakni: ukuran pemusatan data dan ukuran penyebaran data. Ukuran pemusatan atau *central tendency* atau dikenal dengan istilah nilai sentral atau nilai tengah sekumpulan data adalah nilai yang menggambarkan kondisi umum dari sebagian besar amatan dalam data. Statistik yang tergolong dalam ukuran pemusatan adalah rata-rata, median, dan modus.

Bagaimana kita mengetahui selisih atau penyimpangan antara suatu nilai data dengan nilai rata-ratanya? Seberapa jauh amatan-amatan yang ada menyebar dari rata-ratanya? Bagaimana bentuk keragaman datanya? Hal ini dapat dinyatakan dalam sebuah nilai yang disebut sebagai ukuran penyebaran data. Ukuran penyebaran data digunakan untuk menganalisis pemencaran, penyebaran, penyimpangan, atau keragaman suatu kelompok data. Beberapa ukurannya adalah jangkauan (*range*), keragaman (*variance*), simpangan baku (*standard deviation*), dan koefisien keragaman (*coefficient of variation*).

Para mahasiswa yang budiman, untuk memahami apa dan bagaimana kedua ukuran numerik ini penting dalam sekumpulan data dan bagaimana penerapannya dalam kasus atau kehidupan sehari-hari, mari kita bahas secara rinci ukuran numerik dari statistika deskriptif ini. Selamat membaca dan mempelajari materi ini.

A. UKURAN PEMUSATAN DATA

Satu hal penting yang harus kita pahami adalah fungsi data dalam kehidupan sehari-hari. Setiap saat kita selalu berhadapan dengan data, ketika pagi hari saat bersantai menonton TV, sajian pertama yang kita peroleh adalah data berupa informasi misalnya terkait dengan jumlah yang terpapar Pandemi Covid-19 hari ini, persentase kenaikan jumlah yang terpapar Covid-19, kondisi penyediaan vaksin Covid-19 dan lain-lain, itu semua tidak lain merupakan data. Setiap saat kita tak terlepas dari data. Dari mulai bangun tidur sampai malam hari menjelang tidur, tidak terlepas dari sajian data. Oleh karena itu, sebagai awal dari pembahasan ukuran numerik data, kita akan bahas sedikit mengenai fungsi data. Apakah Anda semua sudah paham mengenai fungsi data? Apakah Anda semua menyadari bahwa data itu sangat penting dalam kehidupan sehari-hari? Ada yang bisa menjelaskan apa itu fungsi data? Baiklah, kita

ulas sedikit berkaitan dengan fungsi data. Jika kita perhatikan beberapa hal di sekeliling kita, maka dapat kita katakan bahwa fungsi data di antaranya adalah:

1. untuk membuat keputusan terbaik di dalam memecahkan sebuah masalah;
2. dapat dijadikan sebagai dasar suatu perencanaan atau juga penelitian;
3. dapat dijadikan sebagai acuan dalam tiap-tiap implementasi suatu kegiatan atau aktivitas;
4. dapat dijadikan sebagai bahan evaluasi untuk peningkatan suatu kegiatan.

Kapankah ukuran pemusatan data diperlukan? Untuk apa ukuran pemusatan data tersebut? Ini adalah pertanyaan dasar yang harus dipahami. Ukuran pemusatan data digunakan untuk memperoleh informasi lokasi berkumpulnya data atau titik pusat sekumpulan data. Ukuran ini diperlukan untuk menyatakan keterwakilan data dalam kumpulannya tersebut. Misalnya, ingin diketahui berapa gambaran tinggi badan mahasiswa di kelas A? Atau berapa perolehan nilai Mata Kuliah Analisis dan Visualisasi Data di Program Studi Sistem Informasi (Prodi SI) pada masa registrasi 2021.1? Untuk menjawab ini digunakan ukuran pemusatan data, yaitu rata-rata (*mean*).

Ukuran pemusatan data yang sering digunakan adalah: rata-rata, median, dan modus. Ketiga ukuran ini memiliki kelebihan dan kekurangan. Penggunaan ketiga ukuran pemusatan yang tepat sangat dipengaruhi oleh kondisi atau sebaran data. Kapan sebaiknya kita menggunakan ketiga ukuran tersebut? Hal ini akan kita bahas secara rinci berikut ini.

1. Rata-rata

Rata-rata (*mean*) didefinisikan sebagai jumlah semua data yang dibagi dengan banyaknya data. Apabila terdapat n data yang dinyatakan dengan X_1, X_2, \dots, X_n , maka rata-ratanya adalah:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (1.1)$$

Contoh 1.1:

Perolehan nilai ujian Metode Statistika I dari 10 orang mahasiswa Prodi SI adalah: 84, 75, 80, 69, 76, 87, 85, 90, 75, 80. Nilai rata-rata ujian Metode Statistika I adalah:

$$\begin{aligned} \bar{X} &= \frac{\sum_{i=1}^n X_i}{n} \\ &= \frac{84 + 75 + 80 + 69 + 76 + 87 + 85 + 90 + 75 + 80}{10} \\ &= \frac{801}{10} = 80,1 \end{aligned}$$

Dengan demikian dari data di atas dapat diketahui bahwa rata-rata perolehan nilai Metode Statistika I untuk 10 mahasiswa Prodi SI adalah 80,1.

Lantas yang menjadi pertanyaan selanjutnya adalah kapan sebaiknya kita menggunakan rata-rata sebagai ukuran pemusatan data? Beberapa referensi menyatakan bahwa rata-rata digunakan jika:

- jenis datanya adalah numerik (interval atau rasio);
- sebaran atau distribusi datanya simetris;
- tidak terdapat data pencilan (*outlier*).

Jika dalam sekumpulan data terdapat pencilan, maka nilai rata-rata menjadi tidak memadai. Misalnya, dalam sebuah pemukiman yang berada di desa tertentu, pada umumnya pekerjaan penduduknya adalah petani. Namun, di desa itu ada tiga orang kepala keluarga merupakan orang terkaya di desa itu. Jika ingin diperoleh informasi berapa pendapatan masyarakat di desa tersebut, maka tiga kepala rumah tangga terkaya merupakan pencilan (*outlier*). Penggunaan ukuran pemusatan data rata-rata dalam hal ini kurang memadai, karena dengan mengikutsertakan 3 data pencilan tersebut akan memengaruhi pendapatan masyarakat yang pada umumnya petani. Oleh karena itu, dalam kasus demikian, median akan lebih bermakna dalam menggambarkan ukuran pemusatan data pendapatan desa tersebut.

2. Median

Median biasa dituliskan Q_2 (kuartil kedua) adalah suatu nilai yang membagi data yang telah diurutkan menjadi dua bagian yang sama. Frekuensi data yang di bawah median sama dengan data di atas median. Nilai median dipengaruhi oleh banyaknya pengamatan dan tidak tergantung pada besarnya nilai pengamatan walaupun nilainya tersebut sangat ekstrem. Dengan demikian, median sangat cocok digunakan untuk mewakili data yang sebarannya tidak simetris.

Formulasi median untuk data tunggal adalah sebagai berikut.

a. Untuk n ganjil, median = X_k di mana $k = \frac{n+1}{2}$ (1.2)

b. Untuk n genap, median = $\frac{X_k + X_{k+1}}{2}$ di mana $k = \frac{n}{2}$ (1.3)

Contoh 1.2:

Pendapatan dari 20 orang penduduk di Desa Bakti Mulya adalah sebagai berikut.

No.	Pendapatan (Rp)	No.	Pendapatan (Rp)
1	300.000	11	315.000
2	250.000	12	650.000
3	2.000.000	13	225.000

No.	Pendapatan (Rp)	No.	Pendapatan (Rp)
4	320.000	14	450.000
5	525.000	15	15.000.000
6	200.000	16	500.000
7	10.000.000	17	700.000
8	2.350.000	18	565.000
9	450.000	19	550.000
10	385.000	20	625.000

Apabila ingin diperoleh gambaran mengenai besarnya pendapatan masyarakat di Desa Bakti Mulya akan lebih bermakna jika menggunakan median. Perhitungan median untuk data di atas adalah datanya diurutkan terlebih dahulu dari yang terkecil sampai dengan yang terbesar. Karena datanya ($n=20$) genap, maka gunakan persamaan (1.3). Nilai $k = n/2=20/2=10$, sehingga nilainya mediannya:

$$Q_2 = \frac{X_k + X_{k+1}}{2}$$

$$= \frac{X_{10} + X_{10+1}}{2} = \frac{500.000 + 525.000}{2} = 512.500$$

Besarnya penghasilan masyarakat di Desa Bakti Mulya sebesar Rp512.500. Besaran ini masuk akal, karena pada umumnya mereka adalah petani. Median baik digunakan pada kondisi berikut ini:

- rata-rata tidak memenuhi syarat pada data berjenis interval/rasio yaitu ketika distribusi data tidak simetris;
- skala data numerik ordinal.

3. Modus

Ukuran pemusatan ketiga yang sering digunakan adalah modus. Modus adalah nilai yang sering muncul (frekuensi terbesar). Pada data tunggal, jika terdapat data sebanyak n dengan X_1, X_2, \dots, X_n maka modus adalah nilai X yang sering muncul atau nilai data yang memiliki frekuensi terbesar. Modus tidak dipengaruhi oleh nilai ekstrem atau pencilan (*outlier*). Biasanya modus hanya digunakan untuk tujuan deskriptif karena nilai modus tidak mempertimbangkan sebaran atau distribusi data. Jika nilai-nilai pengamatan sangat bervariasi dari nilai pusatnya, maka modus kurang cocok digunakan sebagai ukuran pemusatan data. Modus sangat berguna dalam menentukan titik pusat data khususnya untuk data nominal dan ordinal.

Misalnya seorang marketing produk Sampo ingin mengetahui Sampo merek apa dan berapa volumenya yang laku di wilayah tertentu? Untuk mengetahui hal ini, maka marketing tersebut tentu menggunakan ukuran pemusatan data. Nilai apakah yang sesuai untuk menyatakan ukuran pemusatan data merek dan ukuran Sampo tersebut?

Dapatkah Anda memberikan penjelasannya? Ya, dalam kasus ini yang digunakan tentunya modus. Artinya berapa banyak masyarakat di wilayah tersebut yang menggunakan Sampo merek tertentu dan berapa ukuran volumenya yang paling banyak dibeli oleh masyarakat di wilayah tersebut. Mengapa demikian? Hal ini karena untuk mencerminkan pengguna Sampo tentu yang paling bermakna adalah modus. Berapa banyak atau berapa frekuensi masyarakat yang menggunakan Sampo jenis tertentu dan yang ukuran volume berapa yang sering dibeli. Oleh karena itu, dalam kasus ini modus akan memberikan informasi lebih valid.

Contoh 1.3:

Lihat kembali data perolehan nilai Metode Statistika I pada Contoh 1.1. Berapa modus dari data tersebut? Silakan kerjakan sebagai latihan.

B. UKURAN PENYEBARAN DATA

Pengamatan terhadap data tidak terlepas dari ukuran pemusatan dan ukuran penyebaran data. Kadang kala kita ingin mengetahui seberapa jauh amatan-amatan yang ada menyebar dari rata-ratanya? Bagaimana bentuk sebaran data tersebut? Bagaimana keragaman datanya? Pertanyaan-pertanyaan tersebut dapat dijawab dengan menentukan ukuran penyebaran data. Dengan demikian, apa yang dimaksud dengan ukuran penyebaran data? Seberapa penting ukuran ini untuk pengamatan data? Silakan pahami penjelasan berikut ini.

Jika kita ingin membandingkan dua kelompok data yang memiliki keragaman yang berbeda dengan ukuran pemusatan data yang sama, maka kelompok mana yang akan Anda ambil? Mengapa? Ukuran penyebaran data digunakan untuk mengetahui seberapa jauh pengamatan-pengamatan yang kita peroleh menyebar dari rata-ratanya. Semakin dekat penyebaran data ke rata-ratanya atau ukuran sebaran datanya semakin kecil, maka rata-rata tersebut benar-benar representatif mewakili sekumpulan data yang diamati. Oleh karena itu, ukuran penyebaran data ini dijadikan indikator untuk menentukan ketepatan dalam menentukan titik pusat data. Selain itu, ukuran penyebaran data digunakan untuk menentukan perbandingan keragaman/variabilitas data.

Beberapa ukuran penyebaran data yang sering digunakan adalah: jangkauan (*range*), keragaman (*variance*), simpangan baku (*standard deviation*), dan koefisien keragaman (*coefficient of variation*). Berikut penjelasan mengenai ukuran penyebaran data.

1. Jangkauan

Jangkauan (*range*) merupakan selisih antara data tertinggi (maksimum) dan data terendah (minimum). Perhitungan ukuran ini mudah didapatkan untuk menunjukkan seberapa jauh data memencar, tetapi tidak menunjukkan keragaman. Ukuran ini jarang

digunakan karena hanya berdasarkan dua amatan, yaitu nilai maksimum dan minimum.

$$\text{Range} = x_{\max} - x_{\min} \quad (1.4)$$

2. Keragaman dan Simpangan Baku

Berbeda dengan jangkauan, keragaman menggunakan keseluruhan data dan menunjukkan deviasi atau keragaman. Ukuran ini didasarkan pada perbedaan antara nilai-nilai suatu amatan dengan rata-ratanya. Dalam perhitungannya dirumuskan sebagai penjumlahan kuadrat selisih antara amatan dengan rata-ratanya, kemudian dibagi dengan total amatan.

Misalkan dari suatu populasi terdiri atas N pengamatan yaitu x_1, x_2, \dots, x_N , maka keragaman populasi (σ^2) didefinisikan sebagai:

$$\sigma^2 = \sum_{i=1}^N \frac{(x_i - \mu)^2}{N} \quad (1.5)$$

Sedangkan jika dari populasi tersebut diambil sampel berukuran n , maka keragaman sampelnya adalah:

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \quad (1.6)$$

dengan n adalah total amatan sampel dan \bar{x} adalah rata-rata sampel. Perbedaan masing-masing amatan (x_i) dengan rata-ratanya dinamakan deviasi dari rata-rata. Misalnya untuk data sampel memiliki deviasi ($x_i - \bar{x}$).

Rumus keragaman sampel pada persamaan 1.6 dapat dijabarkan menjadi rumus hitung sebagai berikut.

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1} \quad (1.7)$$

Nilai keragaman lebih baik digunakan dan diinterpretasikan ketika membandingkan beberapa keragaman dari dua atau lebih variabel. Dari beberapa nilai keragaman tersebut, keragaman atau deviasi yang besar menunjukkan penyimpangan dan keragaman yang besar pula. Dalam hal interpretasi, nilai keragaman yang diperoleh dari pengkuadratan menyebabkan sulit untuk mendapatkan interpretasinya. Oleh karena itu, diperlukan standar deviasi.

Standard deviasi dikenal juga sebagai simpangan baku, yang merupakan akar dari nilai keragaman. Muncul pertanyaan, mengapa perlu mengubah nilai keragaman menjadi standar deviasi?. Hal ini dikarenakan standar deviasi merupakan hasil pengukuran yang memberikan nilai sama dengan data asli.

Untuk populasi, standar deviasi (σ) merupakan akar dari keragaman populasi, sedangkan standar deviasi sampel merupakan akar dari keragaman sampel. Sebagai gambaran, untuk standar deviasi sampel dihitung dengan rumus berikut:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (1.8)$$

3. Koefisien Keragaman

Nilai koefisien keragaman menunjukkan nilai relatif standar deviasi terhadap rata-rata. Secara umum, nilai ini digunakan untuk membandingkan keragaman beberapa variabel yang memiliki standar deviasi dan rata-rata yang berbeda. Dapat juga dinyatakan dengan persentase, seperti berikut:

$$CV = \frac{s}{\bar{x}} \times 100\% \quad (1.9)$$



Latihan

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

- 1) Mengapa penyajian data sangat diperlukan?
- 2) Apa yang dimaksud ukuran pemusatan dan ukuran penyebaran data?
- 3) Kapankah nilai rata-rata diperlukan dalam ukuran pemusatan data?
- 4) Kapankah median diperlukan dalam ukuran pemusatan data?
- 5) Kapankah modus diperlukan dalam ukuran pemusatan data?
- 6) Mengapa ukuran penyebaran data diperlukan?
- 7) Apa kegunaan ukuran penyebaran data?

Petunjuk Jawaban Latihan

- 1) Penyajian data sangat diperlukan untuk memberikan gambaran atau deskripsi yang mudah dipahami oleh pengguna karena data merupakan informasi penting dalam pengambilan keputusan.
- 2) a) Ukuran pemusatan data adalah nilai yang dipandang dapat menggambarkan ukuran pemusatan data yang berkaitan dengan letak

- (lokasi). Statistik yang tergolong dalam ukuran pemusatan adalah rata-rata, median, modus, kuartil, dan persentil.
- b) Ukuran penyebaran data adalah nilai yang menggambarkan ukuran seberapa jauh amatan-amatan yang ada menyebar dari rata-ratanya. Ukuran penyebaran data digunakan untuk menganalisis pemencaran, penyebaran, penyimpangan, atau keragaman suatu kelompok data.
 - 3) Nilai rata-rata digunakan sebagai ukuran pemusatan data jika jenis datanya numerik, distribusi data simetris, dan tidak ada data pencilan (*outlier*).
 - 4) Median digunakan sebagai ukuran pemusatan data jika datanya numerik, distribusi data tidak simetris.
 - 5) Modus digunakan sebagai ukuran pemusatan data jika data pengamatan tidak simetris dan memiliki pencilan (*outlier*). Modus baik digunakan dalam menentukan titik pusat data khususnya untuk data nominal dan ordinal.
 - 6) Ukuran penyebaran data diperlukan karena mengetahui seberapa jauh amatan-amatan tersebut menyebar dari pusat data (nilai rata-rata).
 - 7) Kegunaan ukuran penyebaran data adalah untuk membandingkan kelompok data yang memiliki keragaman berbeda dengan ukuran pemusatan data yang relatif sama.



Rangkuman

1. Statistik adalah karakteristik dari suatu sampel, atau alat untuk menunjukkan pada pencatatan angka-angka dari suatu kejadian atau kasus tertentu. Dengan kata lain statistik adalah kumpulan fakta berbentuk angka yang disusun dalam daftar atau tabel dan atau diagram, yang melukiskan atau menggambarkan suatu persoalan.
2. Statistik berbeda dengan statistika, perbedaannya yakni statistik menunjukkan pencatatan angka-angka suatu kejadian, sedangkan statistika adalah ilmu tentang cara-cara mengumpulkan, mentabulasi dan menggolongkan, menganalisis dan mencari keterangan yang berarti dari data yang berupa angka.
3. Secara menyeluruh, peran dan fungsi statistika digunakan untuk menunjukkan tubuh pengetahuan (*body of knowledge*) tentang cara-cara pengumpulan data, analisis, dan penafsiran data.
4. Pengklasifikasian statistika dilihat dari isi yang dipelajari terbagi menjadi statistika teoritis dan terapan, sedangkan dari aktivitas yang dilakukan yakni statistika deskriptif dan statistika inferensial. Dalam penerapannya, statistika menggunakan istilah-istilah umum yang sering digunakan dalam rumus statistik.
5. Jenis data statistik terbagi dua yakni data kualitatif dan kuantitatif. Jenis data tersebut dianalisis dengan menggunakan skala pengukuran, yakni skala nominal, ordinal, interval, dan rasio. Dengan menggunakan skala tersebut, maka data statistik dapat dilihat ukuran hasilnya.



Tes Formatif 1

Pilihlah satu jawaban yang paling tepat!

- 1) Berikut merupakan ukuran pemusatan data, *kecuali*
 - A. rata-rata
 - B. ragam
 - C. median
 - D. modus

- 2) Ukuran pemusatan data yang tidak dipengaruhi oleh pencilan (*outlier*) adalah
 - A. rata-rata
 - B. ragam
 - C. simpangan baku
 - D. median

- 3) Ukuran-ukuran pemusatan data berikut yang nilainya tidak sama dengan median adalah
 - A. desil D5
 - B. kuartil Q1
 - C. persentil ke 50
 - D. kuartil Q2

- 4) Berikut adalah ukuran-ukuran pemusatan data yang memungkinkan bernilai negatif
 - A. ragam
 - B. simpangan baku
 - C. koefisien keragaman
 - D. kurtosis

- 5) Banyaknya jawaban yang salah pada suatu kuis dengan soal benar salah dari lima belas siswa yang dipilih secara acak adalah 2, 1, 3, 0, 1, 3, 6, 0, 3, 3, 5, 2, 1, 4, dan 2. Nilai rata-rata dan median data tersebut adalah
 - A. 2, 4, dan 3
 - B. 2, 5, dan 2
 - C. 2, 4, dan 2
 - D. 2, 5, dan 3

- 6) Pegawai di sebuah pabrik memberikan sumbangan dalam dolar pada *United Fund* dengan data: 10, 40, 25, 5, 20, 10, 27, 40, 50, 32, dan 35. Dengan menganggap data itu sebagai populasi, maka modusnya adalah
- 10
 - 40
 - 20
 - 10 dan 40
- 7) Data percobaan laboratorium tentang ketidakmurnian senyawa diberikan di bawah ini. Anggap bahwa data tersebut adalah sampel, maka keragaman dari data tersebut adalah

Observasi Ketidakmurnian Senyawa (%)				
0,04	0,14	0,17	0,19	0,22
0,06	0,14	0,17	0,21	0,24
0,12	0,15	0,18	0,21	0,25

- 0,0034
 - 0,0036
 - 0,166
 - 0,058
- 8) Diberikan data grup populasi seperti tertera pada tabel berikut. Nilai simpangan bakunya sama dengan

Kelas (dollar)	Frekuensi (f)
00,00 – 49,95	78
50,00 – 99,95	123
100,00 – 149,95	187
150,00 – 199,95	112

- 48,46
- 48,44
- 48,54
- 48,41

- 9) Diketahui kelas 12A suatu sekolah menengah memiliki rata-rata nilai matematika 7,5 dan simpangan baku sebesar 1,5. Sementara itu, pada pelajaran yang sama, Kelas 12B memiliki rata-rata sebesar 8 dengan simpangan baku sama dengan 3. Koefisien keragaman kelas 12A adalah
- A. 0,2%
 - B. 37,5%
 - C. 20%
 - D. 15%
- 10) Perhatikan kembali soal nomor 9 di atas. Nilai koefisien keragaman kelas 12B adalah
- A. 0,2%
 - B. 37,5%
 - C. 20%
 - D. 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.

$$\text{Tingkat Penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{\text{Jumlah Soal}} \times 100$$

Arti tingkat penguasaan



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.

Bentuk-bentuk Penyajian Data

Para mahasiswa yang budiman, setelah kita mempelajari ukuran numerik statistika deskriptif, berikut kita bahas mengenai bentuk-bentuk penyajian data. Bentuk penyajian data perlu dipahami agar data atau informasi yang diperoleh dapat dipahami dengan mudah oleh pembaca. Secara umum, bentuk-bentuk penyajian data dikelompokkan ke dalam 2 bagian yaitu: tabel dan grafik.

Tabel menyajikan data ke dalam bentuk baris atau kolom sedemikian rupa sehingga memberikan informasi lebih kepada peneliti. Sedangkan grafik menyajikan data dari tabel tersebut menjadi bentuk visual yang lebih informatif lagi. Penyajian data ini tidak hanya sangat membantu peneliti untuk mengetahui gambaran data awal, namun digunakan juga pada analisis inti penelitian atau pelaporan. Mari kita bahas bentuk penyajian data tersebut satu persatu.

A. TABEL

Bentuk penyajian data dengan tabel adalah menyajikan data ke dalam bentuk baris atau kolom. Adapun baris dan kolom tersebut dapat berupa kategori-kategori dan angka frekuensi. Terdapat dua jenis tabel yang akan diterangkan lebih lanjut di bawah ini.

1. Tabel Satu Arah (*One-Way Table*)

Tabel satu arah (*one-way table*) merupakan tabel yang menunjukkan satu variabel atau satu karakteristik saja. Sebagai contoh Tabel 1.1 jumlah responden berdasarkan kategorisasi umur. Dari Tabel 1.1 dapat ditunjukkan hanya terdapat satu 1 karakteristik jumlah responden, yaitu kelompok usia.

2. Tabulasi Silang (Lebih dari Satu Arah/*Two-Way Table*)

Tabulasi silang atau sering disebut tabel dua arah merupakan metode tabulasi untuk merangkum data dengan dua atau lebih variabel secara bersamaan. Pada bentuk tabel, sisi (kolom) sebelah kiri dan baris atas menyatakan kelas untuk kedua variabel yang digunakan. Metode tabulasi silang ini dapat digunakan jika:

- a. salah satu variabel bersifat kualitatif dan lainnya kuantitatif;
- b. kedua variabel berupa variabel kualitatif;
- c. kedua variabel berupa variabel kuantitatif.

Tabel 1.1
Jumlah Responden Berdasarkan Umur

Kategori umur	Jumlah responden
< 20 tahun	10
21 – 30 tahun	15
31 – 40 tahun	25
> 40 tahun	10
Total	60

Contoh tabulasi silang dapat kita amati pada Tabel 1.2 yang terdiri dari 2 variabel, yaitu umur dan jenis kelamin. Umur ada di baris dan jenis kelamin ada di kolom.

Tabel 1.2
Jumlah Responden Berdasarkan Umur dan Jenis Kelamin

Kategori umur	Jenis Kelamin	
	Laki-laki	Perempuan
< 20 tahun	5	5
21 – 30 tahun	10	5
31 – 40 tahun	12	13
> 40 tahun	2	8
Total	29	31

3. Tabel Frekuensi

Tabel frekuensi merupakan tabel ringkasan data yang menunjukkan frekuensi atau banyaknya item atau obyek pada setiap kelas yang ada. Tujuannya adalah untuk mendapatkan informasi lebih dalam tentang data yang ada yang tidak dapat secara cepat diperoleh dengan melihat data aslinya. Langkah-langkah dalam menyusun distribusi frekuensi ini adalah:

- menentukan range (R), yaitu selisih data tertinggi dengan data terendah;
- menentukan banyak kelas (k);
- banyak kelas yang baik antara 5 – 15 kelas. Pendekatan yang cukup baik digunakan adalah $k = 1 + 3,3 \log n$, di mana n menyatakan banyaknya data;
- menentukan lebar interval (i) = (maksimum-minimum)/k;
- menetapkan batas-batas kelas;
- menghitung banyaknya data yang termasuk dalam tiap-tiap kelas;
- menentukan titik tengah kelas;
- menghitung frekuensi kumulatif dan relatifnya.

Tabel 1.3
Data Usia Responden

19	20	23	24
27	28	32	36
37	38	38	40
41	41	45	47
50	51	55	56

Mari kita pelajari langkah pembuatan tabel frekuensi dengan contoh data usia dari 20 responden di Tabel 1.3.

- Range = $56 - 19 = 37$
- Banyak kelas = $1 + 3,3 \log (20) = 5,293 \sim 5$
- Lebar interval = $37/5 = 7,4 \sim 7$
- Batas kelas = 19-26; 27-34; 35 - 42; 43 - 50; 51-58
- Sehingga dihasilkan tabel frekuensi seperti berikut:

Kategori umur	Frekuensi	Frekuensi kumulatif	Frekuensi relatif
19 - 26 tahun	4	4	0,20
27 - 34 tahun	3	7	0,35
35 - 42 tahun	7	14	0,70
43 - 50 tahun	3	17	0,85
51 - 58 tahun	3	20	1,00
Total	20		

Tabel tersebut menunjukkan, dari 20 responden, sejumlah 4 (atau 20%) di antaranya berumur 19-26 tahun. Selain itu, sejumlah responden terbanyak berumur 35-42 tahun.

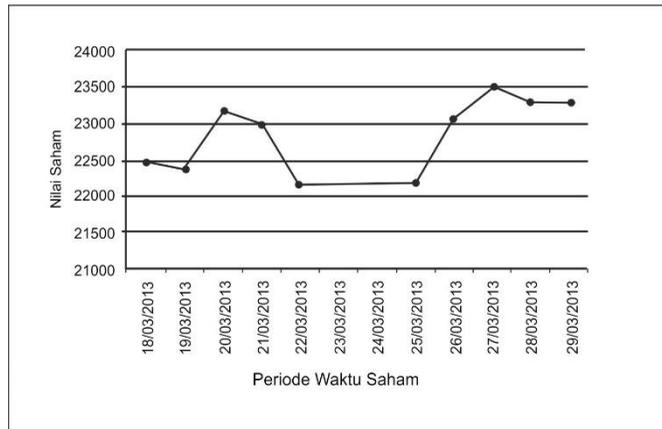
Apabila kita amati contoh tersebut merupakan aplikasi untuk data kuantitatif. Sedangkan untuk data kualitatif tentu akan lebih mudah, di mana perhitungan frekuensi langsung dilakukan berdasarkan kategorinya atau tidak perlu menghitung range hingga batas kelas.

B. GRAFIK

Penyajian data ke dalam tabel yang telah dibahas sebelumnya hanya menunjukkan data berupa angka-angka. Sebagai kelanjutannya, kita dapat membuat penyajian data dari tabel tersebut menjadi grafik-grafik yang lebih menarik dan informatif. Beberapa jenis grafik yang dapat digunakan sebagai berikut.

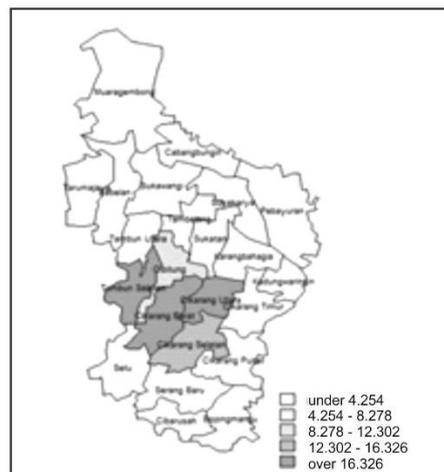
- Grafik Garis (*Line Chart*), untuk melihat pertumbuhan atau perkembangan suatu kejadian. Gambar 1.1 berikut menunjukkan perkembangan saham

perusahaan PT “Y” periode 18 Maret hingga 29 Maret 2013. Dari grafik ini terlihat fluktuasi naik dan turun nilai saham, misalnya tertinggi adalah 23.500 dan terjadi pada 27 Maret 2013.



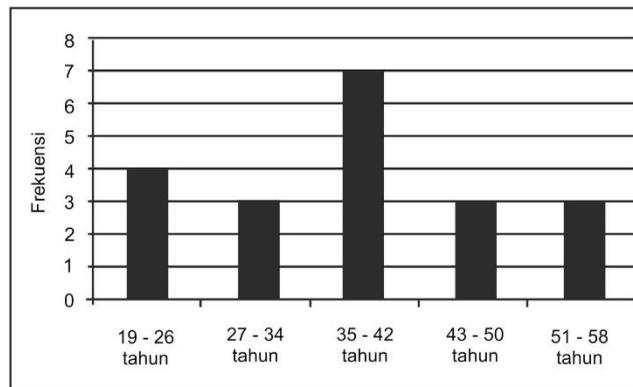
Gambar 1.1
Grafik Line Saham (Close) PT “Y”

- b. Grafik Peta, untuk melihat atau menunjukkan lokasi suatu wilayah beserta atribut atau karakteristiknya. Grafik ini sangat penting apabila Anda tertarik pada analisis data spasial. Seperti contoh pada Gambar 1.2, yang menunjukkan lokasi 23 kecamatan di Kabupaten Bekasi. Sementara itu, degradasi warna menunjukkan karakteristik persentase kontribusi Produk Domestik Regional Bruto setiap kecamatan pada 2010. Detail penjelasan penyajian ini akan dibahas pada kegiatan belajar selanjutnya.



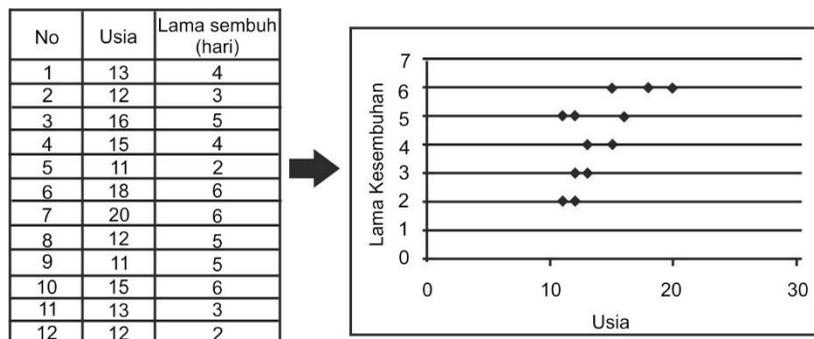
Gambar 1.2
Grafik Peta Persentase Kontribusi Produk Domestik Regional Bruto di Kabupaten Bekasi

- c. Batang (*Bar Graph*), berfungsi untuk melihat distribusi atau perbandingan nilai, frekuensi, atau persentase di setiap kelas (kategori). Sumbu vertikal berupa nilai/frekuensi/persentase, di mana lebar tiap batang sama dengan interval kelas dan tinggi batang sesuai dengan nilai/frekuensi/persentase tiap-tiap kelas. Sedangkan sumbu horizontal menunjukkan nama-nama kelas. Data tabel frekuensi dari data Tabel 1.3 dapat kita lihat pada Gambar 1.3.



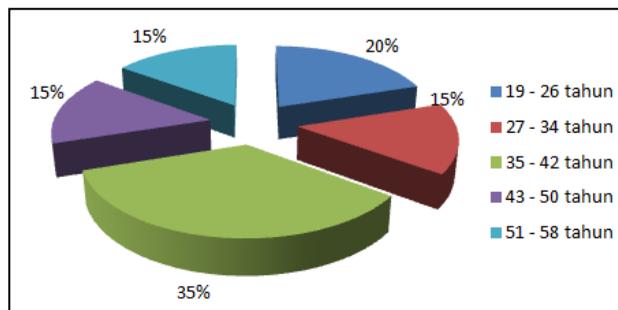
Gambar 1.3
Grafik Batang Frekuensi Responden berdasarkan Umur

- d. Plot Pencaran (*Scatter Plot*), merupakan metode persentase secara grafis untuk menggambarkan hubungan antara dua variabel kuantitatif (Santoso, 2004). Salah satu variabel digambarkan pada sumbu horizontal dan variabel lainnya digambarkan pada sumbu vertikal. Pola yang ditunjukkan oleh titik-titik yang ada menggambarkan hubungan yang terjadi antar variabel. Sebagai contoh kita buat grafik titik yang menunjukkan hubungan antara usia (sumbu horizontal) dan lama kesembuhan penyakit diare (sumbu vertikal). Dapat kita amati secara visual bahwa ada hubungan sebanding antara keduanya, di mana semakin tinggi usia maka waktu kesembuhan juga semakin lama atau panjang.



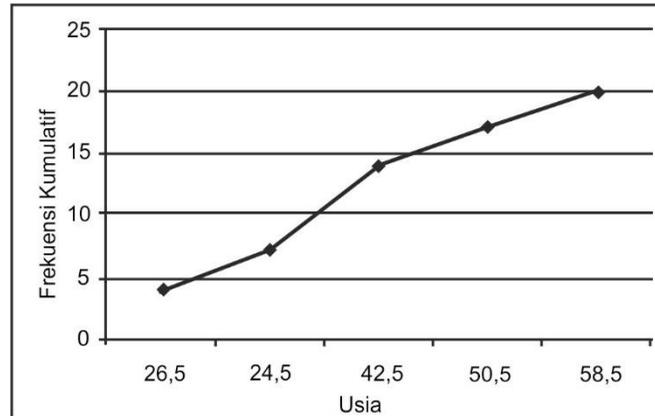
Gambar 1.4
Grafik Titik Usia dan Lama Kesembuhan

- e. Grafik Lingkaran (*Pie Chart*), berfungsi untuk melihat perbandingan (dalam persentase atau proporsi). Grafik ini juga mempresentasikan distribusi frekuensi relatif dari data kualitatif maupun data kuantitatif yang telah dikelompokkan. Caranya adalah dengan menggambar sebuah lingkaran, kemudian menggunakan frekuensi relatif untuk membagi daerah pada lingkaran menjadi sektor-sektor yang luasnya sesuai dengan frekuensi relatif tiap kelas atau kelompok. Contoh: bila total lingkaran adalah 360° maka suatu kelas dengan frekuensi relatif 0,25 akan membutuhkan daerah seluas $(0,25)(360^{\circ})=90^{\circ}$ dari total luas lingkaran. Hasil tabel frekuensi dari data umur Tabel 1.3 telah disajikan ke grafik batang di Gambar 1.3. Data tersebut dapat kita sajikan juga ke *pie chart* seperti Gambar 1.5 berikut.



Gambar 1.5
Pie Chart Umur Responden

- f. Ogive, merupakan grafik yang menyajikan nilai kumulatif. Sumbu horizontal merupakan nilai data, sedangkan sumbu vertikal adalah dapat berupa frekuensi kumulatif, frekuensi relatif kumulatif, atau persen frekuensi kumulatif. Frekuensi yang digunakan (salah satu di atas) masing-masing kelas digambarkan sebagai titik. Setiap titik dihubungkan oleh garis lurus. Hasil tabel frekuensi dari data umur Tabel 1.3 dapat kita sajikan ke Ogive seperti pada Gambar 1.6 berikut. Ogive ini merupakan jenis ogive positif.

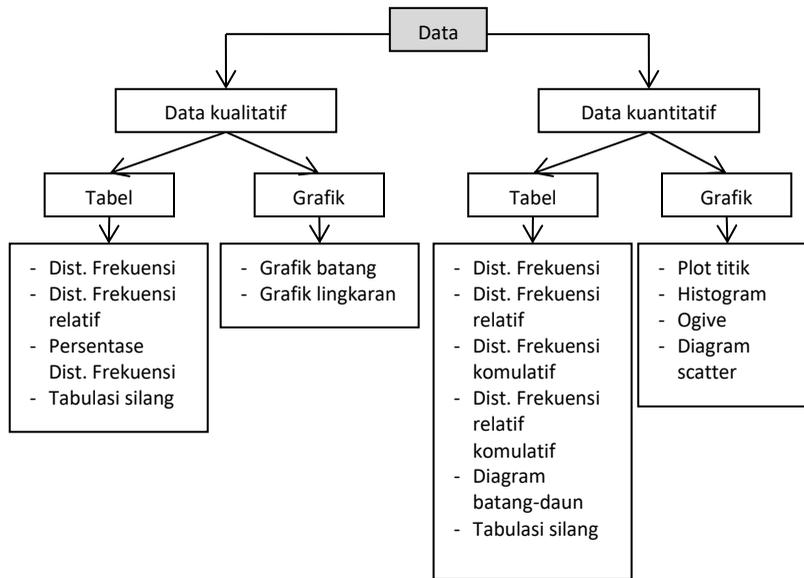


Gambar 1.6
Ogive Data Umur Responden

Dari beberapa penjelasan dan contoh penyajian data yang telah Anda pelajari, dapat kita simpulkan beberapa manfaat tabel dan grafik (Aunuddin, 2005), di antaranya:

- meringkas data, baik data kualitatif maupun kuantitatif;
- data kualitatif berupa distribusi frekuensi, frekuensi relatif, persen distribusi frekuensi, grafik batang, dan grafik lingkaran;
- data kuantitatif berupa distribusi frekuensi, relatif frekuensi dan persen distribusi frekuensi, diagram atau plot titik, distribusi kumulatif, dan ogive;
- dapat digunakan untuk melakukan eksplorasi data;
- membuat tabulasi silang dan diagram sebaran data.

Kita tahu bahwa jenis data yang digunakan berbeda, bisa berupa data kualitatif atau kuantitatif. Dengan jenis data yang berbeda maka bentuk penyajian tabel dan grafik pun berbeda atau langkah-langkahnya pun berbeda. Untuk membedakannya, silakan Anda baca diagram pada Gambar 1.7 berikut ini.



Gambar 1.7
Prosedur Penggunaan Tabel dan Grafik

C. PETA

Penyajian data dalam bentuk peta pada dasarnya dilakukan dengan mengikuti kaidah-kaidah kartografis yang pada intinya menekankan pada kejelasan informasi tanpa mengabaikan unsur estetika dari peta sebagai sebuah karya seni. Kartografis itu sendiri merupakan studi dan praktik membuat peta atau globe.

Para mahasiswa yang budiman, visualisasi data spasial pada prinsipnya adalah bagaimana menampilkan data spasial tersebut. Konsep dasar yang digunakan dalam visualisasi adalah dimensi dari data yang dapat dikelompokkan menjadi tiga yaitu titik, garis, dan area. Data spasial selanjutnya divisualisasikan dalam bentuk simbol dengan memperhatikan beberapa aspek yaitu: sifat dan ukuran data, cara penggambaran simbol, serta variabel visual yang dapat digunakan.

Mari kita bahas satu persatu dari aspek tersebut. Sifat dan ukuran data sama seperti yang telah dibahas pada Kegiatan Belajar 1. Sifat data spasial berupa kuantitatif maupun kualitatif. Bentuk simbol dapat dikelompokkan menjadi simbol titik, garis, ataupun area.

Variabel visual merupakan variabel yang digunakan untuk membedakan unsur yang diwakili pada setiap simbol. Variabel-variabel tersebut dapat berupa (a) bentuk, (b) ukuran, (c) kepadatan, (d) arah, (e) nilai, dan (f) warna, serta (g) posisi. Namun pada perkembangannya dapat juga berupa variabel *transparency*, *shadow*, dan *animation*. Dengan demikian, pertimbangan untuk menentukan simbol pada peta saat

ini dapat menggunakan variabel-variabel visual tersebut. Variabel pada Gambar 1.2 merupakan nilai persentase PDRB (Produk Domestik Regional Bruto).

Variabel visual berkaitan dengan kesan (persepsi) yang akan diperoleh bagi pengguna atau pembaca peta. Tiga tingkatan persepsi tersebut meliputi:

1. asosiatif, apabila pembaca dengan cepat memperoleh kesan yang sama (setingkat) terhadap semua fenomena yang ada pada peta;
2. order, apabila pembaca dengan cepat memperoleh kesan bertingkat terhadap semua fenomena yang dipetakan;
3. kuantitatif, apabila pembaca dengan cepat memperoleh kesan terhadap kuantitas data atau fenomena yang dipetakan.

Lebih mudah lagi Anda dapat melihat simbol persepsi tersebut pada Gambar 1.8 (Cressie, 1991).

1. Pembuatan Peta Tematik

Penyajian data spasial identik dengan peta tematik. Peta tematik itu sendiri adalah peta yang menyajikan tema tertentu dan untuk kepentingan tertentu, sebagai contoh status lahan, kependudukan, distribusi penyebaran penyakit, dan lain-lain. Peta ini menggunakan peta rupa bumi yang telah disederhanakan sebagai dasar untuk meletakkan informasi-informasi tematiknya.

Secara umum, unsur-unsur yang perlu ada di peta tematik di antaranya:

- a. *grid* dan *graticule*,
- b. pola aliran,
- c. relief,
- d. permukiman,
- e. jaringan perhubungan,
- f. batas administrasi,
- g. nama-nama geografi,
- h. detail-detail lain yang erat kaitannya dengan tema yang dipetakan.

	point	line	area		associative	selective	ordered	quantitative
				size		☺	☺	☺
				value		☺	☺	
				texture		☺	☺	
				colour	☺	☺		
				orientation	☺			
				shape	☺			

Gambar 1.8
Variabel Visual dan Persepsi dalam Desain Simbol Grafis

Misalnya, dari Gambar 1.8 tersebut Anda amati jenis poin di baris terakhir. Poin-poin tersebut dapat digambarkan menjadi beberapa bentuk yang berbeda, seperti lingkaran, segitiga, bintang, dan lain-lain. Karena Anda atau orang lain mampu menyimpulkan bentuk poin tersebut secara sama, maka persepsi tersebut merupakan jenis asosiatif.

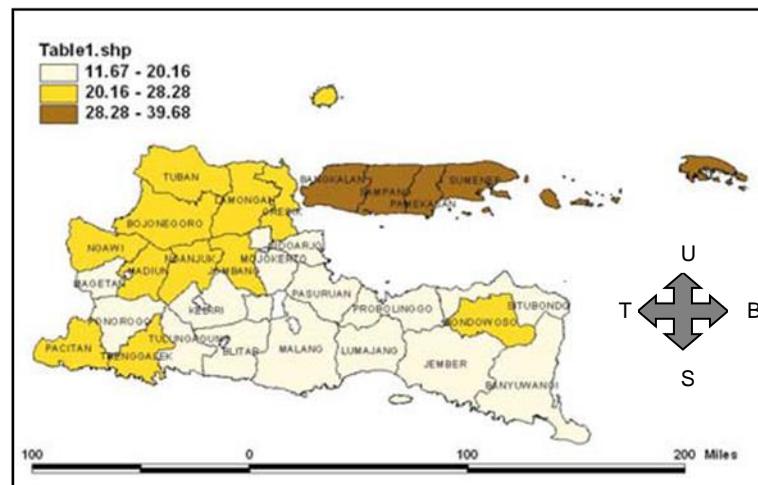
Selanjutnya, perlu ada juga keterangan yang dicantumkan pada tiap lembar peta. Hal ini supaya pembaca dapat dengan mudah memahami isi peta dan arti dari informasi yang disajikan. Informasi tepi setidaknya-tidaknya memuat:

- judul peta,
- skala,
- legenda,
- arah utara,
- angka koordinat geografis,
- diagram lokasi dan petunjuk letak peta,
- sumber data,
- pembuat peta,
- gambaran informasi tepi dan muka peta seperti dalam gambar diagram tata letak peta.

Anda dapat menggunakan beberapa jenis *software* untuk pembuatan peta tematik, yaitu Arcview, Arcgis, atau Open Geo Suit. Misalnya pada *software* Arcview, Anda dapat membuat peta tematik dari awal atau menggunakan file .shp yang kemudian mendemokan variabel yang dibutuhkan.

Sebagai contoh hasil pembuatan peta di Arcview disajikan pada Gambar 1.9. Peta ini menggunakan data interval yang dibentuk oleh ketiga kelompok data dan simbol area. Variabel yang digunakan adalah *Head Count Index* (HCI) atau persentase

penduduk miskin di Provinsi Jawa Timur tahun 2005. Pada peta tersebut menampilkan distribusi HCI di 29 kabupaten. Misalnya, HCI di semua kabupaten di Pulau Madura memiliki HCI yang lebih tinggi (kelompok 1) dibandingkan kabupaten lain di Pulau Jawa, yaitu antara 28.28% - 39.68%. Selanjutnya di beberapa Kabupaten di Pulau Jawa bagian utara memiliki HCI di kelompok kedua. Dengan demikian, melalui visualisasi pada peta tematik ini Anda dapat melihat pengelompokan karakteristik HCI berdasarkan lokasi kabupaten tersebut. Inilah yang menjadi ciri khas data spasial, di mana Anda dapat melihat pengelompokan-pengelompokan lokasi berdasarkan variabel tertentu.



Gambar 1.9
Peta Tematik HCI Provinsi Jawa Timur 2005



Latihan

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

Angka berikut menunjukkan hasil survei curah hujan tahunan pada suatu wilayah (dalam mm) selama 30 tahun antara tahun 1960-1990.

123	117	83	140	97	110	117	86
116	79	130	63	95	103	98	119
84	136	87	91	107	122	74	98
80	82	90	125	105	97		

Buatlah tabel frekuensi dari data di atas!

Petunjuk Jawaban Latihan

Dari data pada soal dapat dibuat tabel frekuensi dengan langkah-langkah sebagai berikut.

- 1) Range : $140-63 = 77$
- 2) $k = 1 + 3,3 \log 30 = 5,8745 \approx 6$
- 3) $i = 77 / 6 = 12,8 \approx 13$
- 4) interval kelasnya: 63-75, 76-88, 89-101, 102-114, 115-127, 128-140.
- 5) Batas kelasnya : 62,5 75,5 88,5 101,5 114,5 127,5 140,5
- 6) Menetapkan turus, yaitu menandai setiap data (angka atau nilai pengamatan) dalam suatu interval kelas tertentu.
- 7) Menentukan frekuensi f, frekuensi kumulatif, dan frekuensi relatif.

Tabel Frekuensi Curah Hujan

Kelas	Interval	Titik Tengah	Turus	Frekuensi (f)	Frek. Kumulatif	Frek. Relatif
1	63 - 75	69		2	2	0,067
2	76 - 88	82		7	9	0,233
3	89 - 101	95		7	16	0,233
4	102 - 114	108		4	20	0,133
5	115 - 127	121		7	27	0,233
6	128 - 140	134		3	30	0,100
Jumlah				30		1,000



Rangkuman

Tahapan pengumpulan data terdiri dari observasi, wawancara, kuesioner, pengukuran fisik, dan percobaan laboratorium. Sedangkan tahapan penyajian data meliputi editing, koding, kemudian membuat penyajian data. Beberapa bentuk penyajian data di antaranya:

1. tabel: tabel satu arah, tabulasi silang, tabel frekuensi;
2. grafik: grafik garis, grafik peta, grafik batang, plot pencaran (*scatter plot*), grafik lingkaran, dan ogive.



Tes Formatif 2

Pilihlah satu jawaban yang paling tepat!

- 1) Salah satu alasan diperlukan tabulasi data dalam penyajian data adalah
 - A. untuk mendapatkan variabel-variabel penelitian
 - B. karena data kasar sulit diinterpretasikan
 - C. untuk memudahkan peneliti mengumpulkan data
 - D. agar data yang telah terkumpul bisa dipahami oleh orang lain

- 2) Yang bukan merupakan penyajian data adalah
 - A. tabel
 - B. grafik
 - C. ogive
 - D. modus

- 3) Tabulasi silang dapat digunakan jika, *kecuali*
 - A. salah satu variabel bersifat kualitatif dan lainnya kuantitatif
 - B. kedua variabel berupa variabel kualitatif
 - C. kedua variabel berupa variabel kuantitatif
 - D. kedua variabel tidak berupa variabel kualitatif atau kuantitatif

- 4) Grafik dari distribusi frekuensi kumulatif, nilai data disajikan pada garis horizontal, pada sumbu vertikal dapat disajikan: frekuensi kumulatif, frekuensi relatif kumulatif, persen frekuensi kumulatif disebut
 - A. tabel
 - B. grafik
 - C. ogive
 - D. modus

- 5) Grafik frekuensi bertangga adalah nama lain dari
 - A. diagram batang
 - B. histogram
 - C. *pie chart*
 - D. diagram pareto

- 6) Penyajian data untuk melihat perbandingan (dalam persentase atau proporsi) dapat berupa
- A. lingkaran
 - B. ogive
 - C. tabel frekuensi
 - D. grafik batang
- 7) Grafik peta untuk menunjukkan
- A. pertumbuhan
 - B. lokasi
 - C. perbandingan
 - D. perkembangan
- 8) Data berbentuk kategori sangat tepat disajikan dalam bentuk
- A. diagram batang daun
 - B. diagram lingkaran
 - C. *dot plot*
 - D. diagram batang
- 9) Diagram yang berfungsi menunjukkan pencilan dan membandingkan dua populasi secara deskriptif adalah
- A. diagram batang daun
 - B. diagram kotak garis
 - C. histogram
 - D. diagram batang
- 10) Diagram yang banyak digunakan untuk membandingkan data maupun menunjukkan hubungan suatu data dengan data keseluruhan adalah
- A. histogram
 - B. diagram batang
 - C. *box plot*
 - D. *dot plot*

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.

$$\text{Tingkat Penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{\text{Jumlah Soal}} \times 100$$

Arti tingkat penguasaan



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.

Membandingkan Kelompok Data

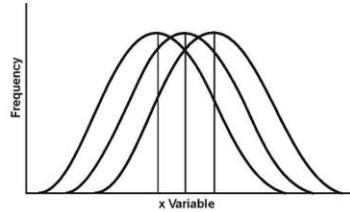
Mahasiswa yang budiman, pada kegiatan belajar sebelumnya Anda telah mempelajari ukuran numerik dan bentuk penyajian data. Pada Kegiatan Belajar 3 ini, kita akan membahas mengenai perbandingan kelompok data. Masih ingat mengenai ukuran pemusatan data dan penyebaran data? Bahasan pada kegiatan belajar ini tidak terlepas dari ukuran numerik yang telah dibahas sebelumnya. Jika Anda belum paham, silakan baca kembali materi yang ada pada Kegiatan Belajar 1.

A. MEMBANDINGKAN KELOMPOK DATA

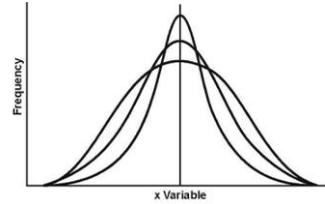
Adakalanya dalam statistika deskriptif kita ingin membandingkan kelompok data. Bagaimana kita bisa membandingkan kelompok data? Apa ukuran yang digunakan untuk membandingkan kelompok data? Apakah Anda masih ingat materi pada kegiatan sebelumnya?

Nilai statistik yang dapat dijadikan indikator untuk membandingkan dua atau lebih kelompok data adalah ukuran penyebaran data. Hal ini karena ukuran penyebaran data digunakan untuk mengetahui seberapa jauh pengamatan-pengamatan yang kita peroleh menyebar dari rata-ratanya. Semakin dekat penyebaran data ke rata-ratanya atau ukuran sebaran datanya semakin kecil, maka rata-rata tersebut benar-benar representatif mewakili sekumpulan data yang diamati.

Jika terdapat dua sekumpulan data yang memiliki ukuran pemusatan data rata-rata yang sama, maka data tersebut mungkin memiliki ukuran penyebaran data yang berbeda. Ilustrasi tersebut disajikan seperti Gambar 1.10 dan 1.11 di bawah ini. Perhatikanlah Gambar 1.10. Dari gambar tersebut terlihat terdapat tiga kurva yang menunjukkan tiga kumpulan data dengan ukuran pemusatan rata-rata berbeda namun ukuran penyebaran datanya sama. Pemusatan rata-rata ditunjukkan oleh garis tengah yang membagi kurva menjadi dua bagian simetris. Sedangkan penyebaran ditunjukkan oleh lebar gelembung dan tinggi kurva. Ketiga kurva memiliki lebar gelembung kurva yang sama, maka dikatakan memiliki penyebaran yang sama.



Gambar 1.10
Tiga Kurva dengan Ukuran
Pemusatan Berbeda dan
Penyebaran Sama



Gambar 1.11
Tiga Kurva dengan Ukuran
Pemusatan Sama dan Ukuran
Penyebaran Berbeda

Perhatikan juga Gambar 1.11. Gambar 1.11 menyajikan perbandingan kurva dari tiga sekumpulan data yang memiliki ukuran pemusatan data rata-rata sama dengan ukuran penyebaran berbeda. Terlihat bagaimana kurva tersebut memiliki sebaran yang berbeda ditunjukkan dengan lebar gelembung kurva berbeda. Misalnya kurva pertama memiliki lebar gelembung kurva lebih besar dibandingkan kurva kedua, maka kurva pertama memiliki keragaman yang lebih besar dibandingkan kurva kedua. Demikian juga halnya dengan kurva kedua dan ketiga. Dari Gambar 1.11 terlihat bahwa kurva ketiga yang paling dekat dengan nilai tengah yang ditunjukkan dengan garis vertikal memiliki ketajaman yang lebih kecil.

Untuk memahami perbandingan kelompok data perhatikan contoh berikut.

Contoh 1.4:

Andaikan terdapat 2 sekumpulan data usia responden sebagai berikut.

1. Kelompok 1 : 2, 2, 6, 6, dan 6. Memiliki rata-rata = 4,4 dan keragaman = 4,8.
 2. Kelompok 2 : 1, 1, 1, 4, dan 15. Memiliki rata-rata = 4,4 dan keragaman = 36,8.
- Jika kita diminta untuk memilih, kelompok data manakah yang memiliki data homogen?

Perhatikan data 2 kelompok di atas. Rata-rata digunakan sebagai ukuran pemusatan dan keragaman sebagai ukuran penyebaran data. Kedua data memiliki rata-rata yang sama. Secara umum dapat dilihat bahwa data pertama hanya bernilai 2 dan 6, sedangkan data kedua lebih beragam yaitu 1, 4, dan 15. Maka dapat dikatakan kelompok kedua memiliki keragaman atau penyebaran lebih besar dibandingkan kelompok pertama. Hal ini juga ditunjukkan oleh nilai keragaman yang lebih besar pada kelompok kedua. Data pada kelompok kedua berkisar antara nilai 1, 4, dan 15 dan memiliki rata-rata 4,4. Apakah angka 4,4 tersebut benar-benar mampu menggambarkan pemusatan data? Sedangkan kita lihat nilai minimumnya adalah 1 dan maksimumnya adalah 15. Tentu saja rata-rata tersebut belum mampu menggambarkan data secara keseluruhan.

Namun jika kita lihat data pada kelompok pertama terlihat rata-ratanya 4,4 dan ukuran penyebarannya (keragaman = 4,8). Nilai keragaman mendekati nilai rata-ratanya, sehingga data yang ada pada kelompok pertama lebih homogen dan ukuran pemusatan data lebih mewakili kondisi data sesungguhnya. Oleh karena itu, dalam hal ini kelompok data pertama lebih mewakili dan dapat kita ambil sebagai data yang baik karena memiliki keragaman yang minimum dan mendekati kepada nilai pemusatan data.

Contoh 1.5:

Perhatikan data pengeluaran untuk konsumsi rumah tangga (dalam ribuan rupiah per hari) dari sampel masing-masing berukuran $n = 10$ rumah tangga di Perkotaan dan di Pedesaan.

Rumah Tangga	Perkotaan	Rumah Tangga	Pedesaan
A	110	P	54
B	102	Q	45
C	98	R	60
D	107	S	55
E	120	T	75
F	125	U	72
G	115	V	62
H	95	W	65
I	122	X	85
J	112	Y	35

Jika dari kasus tersebut ingin diketahui pengeluaran mana yang lebih merata? Apakah di Pedesaan atau Perkotaan? Bagaimana Anda menjelaskan fenomena tersebut?

Untuk menjawab pertanyaan tersebut, hal pertama yang harus Anda ingat adalah ukuran pemusatan dan ukuran penyebaran data. Apa ukuran pemusatan data yang Anda gunakan? Ya, tepat sekali. Ukuran pemusatan data yang digunakan adalah rata-rata.

Untuk dapat menentukan pengeluaran mana yang lebih merata, artinya kita bisa menentukan keragaman dari 2 kelompok tersebut yang lebih kecil. Atau Anda pun bisa menggunakan koefisien keragaman. Masih ingat apa yang dimaksud koefisien keragaman? Koefisien keragaman (KK) adalah perbandingan antara standar deviasi dan nilai rata-rata yang dinyatakan dengan persentase. Nilai koefisien keragaman berguna untuk mengamati keragaman data atau sebaran data dari rata-rata hitungannya. Dalam hal ini jika koefisien keragamannya semakin kecil, datanya semakin seragam (homogen). Sebaliknya, jika koefisien keragamannya semakin besar datanya semakin heterogen.

Kita kembali ke contoh soal pengeluaran untuk konsumsi rumah tangga pada Contoh 1.5. Untuk dapat menunjukkan pengeluaran mana yang lebih merata dapat ditentukan nilai rata-rata, simpangan baku dan koefisien keragaman masing-masing kelompok. Silakan hitung sebagai latihan. Diperoleh hasil sebagai berikut.

Rumah Tangga	Perkotaan	Rumah Tangga	Perdesaan
A	110	P	54
B	102	Q	45
C	98	R	60
D	107	S	55
E	120	T	75
F	125	U	72
G	115	V	62
H	95	W	65
I	122	X	85
J	112	Y	35
Jumlah	1106	Jumlah	608
Rata-rata	110,6	Rata-rata	60,8
Simpangan Baku	10,20	Simpangan Baku	14,63
KK	9,22	KK	24,07

Dari nilai Koefisien Keragaman (KK) Perkotaan = 9,22% < KK Perdesaan = 24,07%, Dengan demikian maka pengeluaran untuk konsumsi rumah tangga di Perkotaan lebih homogen dibandingkan di Perdesaan.



Latihan

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

- 1) Apa ukuran yang digunakan untuk membandingkan dua kelompok data?
- 2) Perhatikan kasus data produktivitas padi IR64 dan Cisadane (dalam ton) di 10 kabupaten di Jawa Barat.

No	Kabupaten	Varietas IR64	Varietas Cisadane
1	Bekasi	5,9	5,4
2	Kerawang	7,0	6,4
3	Purwakarta	6,2	6,0
4	Subang	6,4	5,8
5	Bogor	6,0	5,2
6	Sukabumi	6,1	5,4
7	Cianjur	5,8	5,9
8	Bandung	5,7	5,3
9	Sumedang	6,5	6,2
10	Garut	6,7	6,1

Padi varietas manakah yang produktivitasnya lebih merata di 10 kabupaten tersebut? Jelaskan jawaban Anda!

Petunjuk Jawaban Latihan

- 1) Ukuran yang dapat digunakan untuk membandingkan dua kelompok data adalah ukuran pemusatan dan penyebaran data. Ukuran pemusatan data yang digunakan bergantung pada kondisi sebaran data dan skala pengukuran data.
- 2) Untuk kasus produktivitas padi varietas IR64 dan Cisadane lakukan perhitungan rata-rata, simpangan baku, dan koefisien keragaman (KK) dari masing-masing varietas terbut. Semakin kecil KK, semakin homogen (merata).



Rangkuman

1. Ukuran yang digunakan untuk membandingkan dua kelompok data adalah ukuran pemusatan dan penyebaran data.
2. Salah satu ukuran yang digunakan adalah koefisien keragaman (KK), yaitu perbandingan antara simpangan baku dan nilai rata-rata yang dinyatakan dengan persentase.
3. Nilai koefisien keragaman berguna untuk mengamati variasi data atau sebaran data dari rata-rata hitungannya. Dalam hal ini jika koefisien keragamannya semakin kecil, datanya semakin seragam (homogen). Sebaliknya, jika koefisien keragamannya semakin besar datanya semakin heterogen.



Tes Formatif 3

Pilihlah satu jawaban yang paling tepat!

- 1) Berikut merupakan ukuran pemusatan data
 - A. modus
 - B. keragaman
 - C. jangkauan
 - D. simpangan baku

- 2) Ukuran berikut yang merupakan ukuran penyebaran data adalah
 - A. rata-rata
 - B. median
 - C. jangkauan
 - D. modus

- 3) Terdapat data berikut: 6,7,8,9,10,14. Rata-rata dari data tersebut adalah
 - A. 7,0
 - B. 8,0
 - C. 8,5
 - D. 9,0

- 4) Dari soal no.3 keragaman data tersebut adalah
 - A. 2,8
 - B. 3,0
 - C. 8,0
 - D. 9,0

- 5) Dari soal no.3 simpangan baku data tersebut adalah
 - A. 2,8
 - B. 3,0
 - C. 8,0
 - D. 9,0

- 6) Dari soal no.3 nilai koefisien keragaman dari data tersebut adalah
 - A. 31,4%
 - B. 33,3%
 - C. 35,4%
 - D. 88,9%

- 7) Perhatikan contoh kasus berikut (untuk soal no.7 sampai dengan no. 10).
Lampu di sebuah gedung A rata-rata digunakan selama 800 jam dengan simpangan baku 300 jam. Lampu di gedung B rata-rata digunakan 1.200 jam dengan simpangan baku 400 jam. Lampu di gedung C rata-rata digunakan selama 2.000 jam dengan simpangan baku 1.000 jam. Koefisien keragaman lampu di gedung A adalah
- A. 33,0%
 - B. 37,5%
 - C. 45,0%
 - D. 50,0%
- 8) Koefisien keragaman lampu di gedung B adalah
- A. 33,0%
 - B. 37,5%
 - C. 45,0%
 - D. 50,0%
- 9) Koefisien keragaman lampu di gedung C adalah
- A. 33,0%
 - B. 37,5%
 - C. 45,0%
 - D. 50,0%
- 10) Lampu yang lebih baik terdapat di
- A. gedung A
 - B. gedung B
 - C. gedung C
 - D. gedung A, B, dan C tidak ada yang baik

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.

$$\text{Tingkat Penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{\text{Jumlah Soal}} \times 100$$

Arti tingkat penguasaan



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) B
- 2) D
- 3) B
- 4) D
- 5) C
- 6) D
- 7) B
- 8) D
- 9) C
- 10) B

Tes Formatif 2

- 1) D
- 2) D
- 3) D
- 4) C
- 5) B
- 6) A
- 7) B
- 8) B
- 9) B
- 10) B

Tes Formatif 3

- 1) A
- 2) C
- 3) D
- 4) C
- 5) A
- 6) A
- 7) B
- 8) A
- 9) D
- 10) B

Daftar Pustaka

- Aunuddin. (2005). *Statistika : Rancangan dan analisis data*. Bogor: IPB Press.
- Cressie, N. A. C. (1991). *Statistics for spatial data*. USA : John Wiley & Sons, Inc.
- Hill, R. C., Griffiths, W. E., & Lim, G. C. (2011). *Principles of econometrics*. New Jersey: John Wiley & Sons, Inc.
- Indriatno, I., & Irwinsyah, R. (1998). Aplikasi analisis tabulasi silang (*Crosstab*) dalam perencanaan wilayah dan kota. *Jurnal PWK No.2*.
- Lind, A. D., Marchal, W.G., & Wathen, S.A. (2012). *Statistical techniques in business & economics*. USA: McGraw-Hill.
- Santoso. (2004). *Mengolah data statistik secara profesional*. Jakarta: Elex Media Komputindo.
- Sudjana. (1992). *Metoda statistika*. Bandung: Tarsito.
- Sugiyono. (2003). *Statistika untuk penelitian*. Bandung: Alfabeta.
- Walpole, R. E. (1995). *Pengantar statistika*. Jakarta: Gramedia Pustaka Utama.