Statistika dan Probabilitas untuk Teknik Informatika

Buku Ajar

Penulis:
1. Almed Hamzah
2. Aridhanyati Arifin
3. Elyza Gustri Wahyuni
4. Sri Mulyati

Program Studi Teknik Informatika
Universitas Islam Indonesia
2017

Kata Pengantar

Puji syukur Kami sampaikan kepada Allah SWT, Tuhan Yang Maha Esa atas limpahan rahmatnya sehingga buku ajar "Statistika dan Probabilitas untuk Teknik Informatika" dapat diselesaikan.

Banyak terdapat beragam buku-buku Statistika dan Probabilitas, namun masih sedikit bahkan jarang buku Statistika dan Probabilitas yang dikhususkan untuk bidang Teknik Informatika atau yang serumpun dengannya. Buku ini merupakan buku ajar matakuliah Statistika dan Probabilitas yang ditujukan bagi dosen maupun mahasiswa S-1 khususnya di bidang Teknik Informatika atau yang serumpun dengannya. Materi buku ini, merujuk berbagai buku ajar Statistika dan Probabilitas yang judul-judul bukunya dapat dilihat pada bagian Daftar Pustaka. Materi, narasi tiap materi, contoh soal dan latihan dalam buku ini, disesuaikan dengan kebutuhan kurikulum di Teknik Informatika.

Tiada gading yang tak retak, sebagai karya manusia, tentunya buku ajar ini tidak luput dari kekurangan, karenanya Kami selaku penyusun menerima dengan hati terbuka atas kritik dan saran membangun. Saran tersebut dapat dikirimkan ke email editor: almed.hamzah@uii.ac.id. Terimakasih Kami ucapkan kepada Universitas Islam Indonesia, terutama BPA UII yang telah memberikan dukungan penuh terhadap penyelesaian penulisan buku ajar ini. Ucapan terimakasih juga kami sampaikan kepada berbagai pihak atas dukungannya dalam penyusunan dan penerbitan buku ini. Harapannya, buku ini dapat memberikan sumbangan dan nilai tambah dalam ilmu pengetahuan baik bagi dosen yang mengajar maupun mahasiswa S-1 Teknik Informatika yang sedang mengambil matakuliah Statistika dan Probabilitas.

Yogyakarta, Maret 2017

Tim Penulis

Daftar Isi

Daftar Tabel

Daftar Gambar

1 Peran Statistika dan Probabilitas

Capaian Pembelajaran

Mahasiswa mampu menjelaskan pengertian statistika dan probabilitas serta peran statistika dan probabilitas dan ilmu pengetahuan

1.1 Peran Statistika dan Probabilitas

Statistika dan Probabilitas telah banyak berperan dalam kehidupan kita. Peran yang dimainkan Statistika dan Probabilitas yakni, dalam penyajian informasi pada suatu pemberitaan yang dirilis media ataupun pemerintah, dalam penyediaan metode analisis maupun pemecahan masalah untuk berbagai bidang penelitian dan berperan dalam pembuatan keputusan seorang leader/manager.

Fungsi Statistika dan Probabilitas dalam penyajian data, yaitu merupakan metode atau cara mencatat data yang matematis dan sistematis. Statistika dan Probabilitas telah memberikan cara meringkas data kedalam bentuk yang mudah untuk dianalisis dengan jelas.

Fungsi Statistika dan Probabilitas dalam penelitian, memberikan petunjuk bagi peneliti supaya berpola pikir dan bekerja dengan pasti dan mantap dalam melakukan suatu penelitian. Statistika dan Probabilitas memberi landasan berfikir yang logis untuk digunakan meramal secara ilmiah terhadap suatu gejala atau kejadian dan merupakan alat bagi peneliti untuk menganalisis proses sebab akibat yang kompleks dan rumit.

Teknologi Informasi (TI) dan Statistika dan Probabilitas memiliki hubungan mutualisme. Pengolahan data statistik yang rumit dapat dibantu oleh komputer dan software yang dirancang khusus untuk keperluan perhitungan Matematika dan Statistika seperti minitab, SPSS dan Ms. Excel atau software yang berbasis open source seperti GNU SPSS, GNU Octave dan Gnumeri. Sebaliknya, Statistika dan Probabilitas digunakan oleh bidang TI, ilmu komputer, teknik informatika dan yang serumpun untuk kegiatan penelitian akademis di berbagai konsentrasi bidang serta implementasinya di dunia praktisi.

Beberapa penggunaan Statistika dan Probabilitas dalam bidang TI, Ilmu Komputer, Teknik Informatika dan yang serumpun antara lain:

a. Analisis kebutuhan pengguna (*Requirement Analysis*)

Proses membangun suatu perangkat lunak dimulai dengan tahap analisis kebutuhan pengguna dengan tujuan memperoleh informasi, model dan spesifikasi tentang perangkat lunak yang dibutuhkan. Teknik-teknik dalam Statistika dapat digunakan dalam penelitian analisis kebutuhan tersebut, mulai dari pengumpulan data, pengolahan hingga proses inferensi. Misalnya, dilakukan survei terhadap kebutuhan yang diinginkan oleh pengguna sekaligus mendeskripsikan sistem informasi yang ideal. Perusahan-perusahan raksasa di bidang ITC (Information Technology and Communication) melakukan hal ini sebelum mereka meluncurkan produk barunya.

b. Analisis dan evaluasi sistem (software)

Analisis sistem dapat memanfaatkan metode-metode Statistika untuk mengidentifikasi masalah dan memahami kerja dari sistem yang telah berjalan. Tahapan ini dilakukan dengan cara mempelajari secara terperinci tentang cara kerja sistem yang sudah ada. Untuk menguji apakah kualitas sistem, kualitas informasi dan kualitas pelayanan berpengaruh terhadap penggunaan sistem dan kepuasan pengguna, serta menguji apakah penggunaan sistem, kepuasan pengguna serta struktur organisasi berpengaruh terhadap keuntungan/kerugian.

Sejumlah data mengenai sistem lama dan pengguna perlu dikumpulkan dengan menggunakan bermacam-macam usaha pengumpulan data sehingga dapat dikenali permasalahan-permasalahan dan kelemahan-kelemahannya. Umumnya data diperoleh dengan pengambilan sampel melalui kuesioner yang diisi oleh responden dan diukur dengan skala Likert. Metode analisis data yang digunakan adalah Partial Last Square atau teknik Regresi Linier Berganda. Regresi Linier digunakan untuk melihat hubungan sebab-akibat antar variabel dan mengukur seberapa besar pengaruh antar variabel tersebut. Berdasarkan hasil analisis terhadap sistem lama dan pengguna, analis sistem dapat menentukan apakah sistem lama perlu dipertahankan tapi perlu perbaikan ataukah sistem lama digantikan oleh sistem yang baru.

Penerapan metode statistik dalam analisis sistem dilakukan oleh Monalisa dan Setia (2016,50:53). Penelitian tersebut mengevaluasi penerapan Sistem Informasi Pengolahan Data Statistik Rutin di BKKBN Provinsi Riau. Variabel yang digunakan adalah tingkat penerimaan user terhadap penerapan Sistem Informasi menunjukkan hasil sebesar 41.7 %. Hasil tersebut dipengaruhi oleh faktor manfaat (perceived usefulness) dan faktor kemudahan (perceived ease of use) dengan hubungannya tergolong cukup berarti sebesar 64.6%.

c. Solusi terhadap masalah ketidakpastian

Kita sering berhadapan dengan masalah yang mengandung ketidakpastian contohnya berapa peluang file-file dalam suatu direktori yang terinfeksi virus bisa dipulihkan? Berapa lama suatu software sebesar 250 Kb dapat di-install? Penentuan siswa diterima di suatu sekolah, peramalan cuaca, prediksi stok barang, dan sebagainya. Masalah-masalah yang tidak dapat dipastikan hasilnya, menyangkut masa depan dan memunculkan peluang benar atau salah ini lah yang dinamakan ketidakpastian. Probabilitas merupakan salah satu model yang digunakan untuk memberi dukungan dalam pengambilan keputusan dalam situasi yang mengandung ketidakpastian. Salah satu metodenya adalah Bayes. Selain itu untuk kasus-kasus peramalan atau prediksi dapat digunakan Regresi dan Interpolasi.

1.2 Statistik dan Statistika

Di dunia ini, banyak kejadian yang dicatat baik dalam bentuk bilangan maupun non bilangan yang selanjutnya disajikan sedemikian rupa sehingga lebih mudah untuk dibaca dan diinterpretasikan. Data yang ditampilkan tersebut dinamakan sebagai statistik. Konteks masalah yang akan diwakilkan oleh bilangan maupun non bilangan tersebut dijadikan sebagai nama dari data statistik, seperti statistik pendidikan, statistik pendapatan penduduk dan lain-lain. Data-data tersebut menggambarkan situasi umum, sebagian besar, sebagian kecil, hampir semuanya, kira-kira atau kecenderungan menyangkut suatu permasalahan yang dijadikan obyek penelitian. Misalnya, jika diteliti 10 unit printer yang merknya sama dan dicatat waktu pemrosesan (*processing time*) untuk setiap lembar yang dicetak kemudian dihitung rata-rata waktu pemroresan sebesar 10 detik per lembar, inilah yang dinamakan sebagai statistik, jika 40 % dari printer tersebut memiliki waktu pemroresan kurang dari 6 detik, maka nilai 40 % tersebut dinamakan sebagai statistik.

Statistik dapat didefinisikan sebagai kumpulan data, bilangan atau non bilangan yang disusun dalam bentuk tabel, diagram atau grafik, menggambarkan suatu persoalan atau keadaan, selain itu menyatakan karakteristik sampel seperti rata-rata, standar deviasi dan koefisien korelasi (Sudjana, 1996:2).

Diperlukan kegiatan statistika untuk menghasilkan data statistik. Statistika dan statistik memiliki keterkaitan satu sama lain, lalu apakah statistika itu? Sudjana (1996:3) mendefenisikan statistika sebagai pengetahuan yang berhubungan dengan cara-cara pengumpulan, penyajian, pengolahan data dan teknik-teknik analisis data, penarikan kesimpulan serta pembuatan keputusan yang cukup beralasan berdasarkan data dan analisis yang dilakukan.

Statistika, pada awal perkembangannya digunakan sebagai sarana pemerintah untuk mengumpulkan data-data mengenai kondisi masyarakat. Pemerintah Indonesia melalui Biro Pusat Statistik mengumumkan statistik mengenai ekspor, impor, kelahiran, kematian dan sebagainya dilengkapi dengan tabulasi data, diagram dan grafiknya kemudian menerbitkannya dalam Buku Saku Statistik Indonesia (Purwanto, 2010:4). Kegiatan ini digolongkan sebagai statistika deskriptif karena hanya berisi pengumpulan data, penyajian dan pengolahan data tanpa diakhiri penarikan kesimpulan. Statistika yang tidak hanya berisi pengumpulan, penyajian dan pengolahan data tapi juga melibatkan kegiatan penarikan kesimpulan dinamakan statistika inferensial atau statistika induktif. Aktivitas penarikan kesimpulan itu meliputi pengujian hipotesis dan peramalan dengan regresi.

1.3 Populasi dan sampel

Statistika deskriptif dilakukan atas data yang diambil dari populasi, sementara statistika induktif dilakukan pada sampel. Baron (2014:208) menjelaskan apa yang dimaksud dengan populasi, parameter dan sampel, sebagai berikut:

d. Populasi adalah keseluruhan pengamatan atau semua unit yang dijadikan obyek penelitian.
e. Sejumlah karakteristik dari suatu populasi disebut sebagai parameter.

f. Sampel merupakan unit dari populasi yang menjadi pengamatan dan memiliki karakteristik yang sama dengan karakteristik populasi.

Jika dinyatakan bahwa 30% sistem operasi yang digunakan perkantoran pemerintah di Indonesia adalah Windows 10 maka kesimpulan ini berlaku umum untuk seluruh kantor pemerintah Indonesia dari Sabang hingga Merauke, baik di kota-kota maupun pedesaan, tidak hanya berlaku untuk sekelompok perkantoran saja. Pernyataan tersebut harus dibangun oleh data, populasi atau sampel.

Pengumpulan data populasi tidak bisa selalu dilakukan sebab adakalanya dihadapkan pada jumlah populasi yang berukuran tak hingga. Yakni, populasi yang berisikan banyak obyek mencapai tidak terhingga. Pengumpulan data populasi dengan ukuran terhingga misalnya jumlah pengguna internet di seluruh dunia atau jumlah pengguna akun media sosial di seluruh Indonesia, juga tidak selalu dapat dilakukan, baik karena alasan minimnya biaya atau banyaknya waktu yang dihabiskan, kurangnya sumber daya, tidak praktis, masalah ketelitian dan lain-lain. Contohnya, populasi sebesar 1 milyar obyek, bagaimana cara menganalisis data sebanyak itu lalu teknik pengumpulan data yang digunakan apa? berapa lama waktu yang dibutuhkan untuk mendapatkan semua obyek tersebut?

Berdasarkan pertimbangan tersebut maka ditempuh jalan pengambilan sebagian populasi (sampel). Data sampel dikumpulkan kemudian dianalisis untuk selanjutnya dibuat kesimpulan mengenai karakteristik populasinya, yakni sesuai dengan keadaan yang sebenarnya. Oleh karena itu sampel tersebut harus mewakili populasi, maksudnya segala karakteristik populasi semestinya diwakili oleh sampel yang diambil.

Sampel adakalanya memberikan informasi keliru mengenai populasi meskipun ini terjadi dengan probabilitas yang rendah. Misalnya pada kasus kepuasan layanan internet kampus, jika 80 % mahasiswa merasa puas terhadap layanan internet kampus, itu bukan berarti bahwa benar-benar 8 dari 10 mahasiswa dari sampel yang diambil merasa puas. Apabila misalnya diketahui nilai probabilitasnya 0.0328, hanya setengah dari 10 sampel mahasiswa yang merasa puas. Maksudnya, masih terdapat 3% peluang untuk sampel random yang berlawanan dengan parameter populasi, tidak lebih dari 50% mahasiswa merasa puas terhadap layanan internet kampus.

Oleh karena itu, kualitas sampel dapat mempengaruhi kualitas kesimpulan sebab kesimpulan tersebut akan digeneralisasi kepada populasi, sehingga mempengaruhi pengambilan keputusan. Ketepatan pengambilan sampel merupakan hal yang penting. Misalnya, untuk mengevaluasi suatu Sistem Informasi Manajemen Perpustakaan, survey perlu dilakukan terhadap para pustakawan dan pengunjung perpustakaan, bukan kepada selain mereka.

1.4 Pengumpulan Data

Tahap pengumpulan data merupakan tahapan krusial dalam statistika. Data adalah keterangan mengenai suatu keadaan pada sejumlah obyek (Purwanto, 2011:8). Obyek tersebut bisa bersumber dari pupulasi atau sampel. Ketika melakukan pengumpulan data, terdapat tiga hal yang perlu ditentukan yaitu:

g. Obyek pengumpulan data

Obyek dapat berupa orang atau benda, seperti dokumen presensi, laporan hasil usaha suatu CV, anak balita, karyawan suatu instansi dan seterusnya. Jika obyek berupa orang disebut sebagai responden. Responden adakalanya diambil dari populasi, adakalanya bersumber dari sampel. Menurut Nazir dalam Purwanto (2011:40), data mentah dapat dikumpulkan melalui 2 jalan yaitu:

- Sensus, apabila pengumpulan data dilakukan terhadap tiap unit populasi beserta parameter yang diperlukan. Misalnya seluruh mahasiswa Teknik Informatika dengan parameter indeks prestasi kumulatif di atas 3,00.
- Sampling, apabila pengumpulan data dilakukan pada sampel. Misalnya mahasiswa Teknik Informatika angkatan 2015/2016 saja yang menjadi obyek penelitian.

h. Keadaan atau sifat yang akan dikumpulkan datanya

Keadaan atau sifat ini disebut sebagai variabel. Misalnya, dari dokumen presensi dikumpulkan data kehadiran, ijin, sakit dan absen atau dari karyawan perlu dikumpulkan data kinerja. Varibel ini bergantung pada tujuan penyelesaian masalah.

i. Alat pengumpulan data

Data adakalanya dikumpulkan dengan cara mengukur selain itu dengan cara menghitung. Kedua cara ini memerlukan alat atau disebut instrumen. Data yang dikumpulkan dengan cara mengukur misalnya untuk mendapatkan data lamanya waktu respon suatu halaman web dengan menggunakan alat ukur waktu seperti jam. Data yang dikumpulkan dengan jalan menghitung misalnya pengumpulan data jumlah pengunjung suatu minimarket menggunakan alat hitung yang dipasang di pintu masuk minimarket tersebut. Selain memperhatikan 3 hal di atas, kebenaran data merupakan hal yang sangat penting dalam proses pengumpulan data. Keputusan yang keliru dapat dibuat akibat data-data yang tidak valid. Sebelum data diolah, periksalah data tersebut apakah ada yang meragukan? Apakah ada data yang salah pengukuran atau penghitungannya? Pihak yang mengumpulkan data, harus mengetahui syarat-syarat data yang baik.

Adapun syarat-syarat data yang baik adalah sebagai berikut:

a. Obyektif, menggambarkan masalah apa adanya bukan hasil rekayasa.
b. Representatif, maksudnya data harus mewakili keadaan sebenarnya. Khususnya bila menggunakan data sampel maka harus mewakili populasi.
c. Kesalahan baku kecil. Biasanya digunakan nilai standar toleransi tertentu yang disepakati. Setiap permasalahan biasanya memiliki standar toleransi yang berbeda-beda.
d. Tepat waktu. Data dikumpulkan pada durasi waktu yang ditentukan, di beberapa kasus adakalanya data yang usang, sudah tidak lagi mengandung kebenaran. Misalnya untuk mengalokasikan bantuan RASKIN pada tahun 2017, diperlukan data jumlah KK miskin 1 tahun terakhir. Data penduduk miskin 5 tahun sebelumnya dianggap tidak tepat lagi.

n. Relevan. Data yang diambil, relevan dengan masalah yang dikaji.

Pengelompokan data dilakukan berdasarkan 5 kategori yakni sifat, sumber, cara memperoleh, waktu pengumpulan dan jenis. Beberapa kategori pengelompokkan data, rinciannya adalah berikut ini:

a. Sifat

   Data diklasifikasi menjadi kualitatif dan kuantitatif. Data kualititatif adalah data yang dinyatakan secara non numerik, linguistik dan menunjukkan derajat kualitas suatu hal. Misalnya data yang menyebutkan istilah sangat buruk, buruk, cukup, baik, sangat baik.

b. Sumber

   Berdasarkan sumbernya data, dikelompokkan menjadi data internal dan data eksternal. Data internal adalah data yang diperoleh dari dalam institusi, misalnya data pegawai dan data gaji. Data eksternal adalah data yang diperoleh dari luar institusi, misalnya Keputusan Menteri yang berisi daftar komponen kimia yang digolongan sebagai narkotika atau data trend selera pasar, biasanya didapatkan dari koran atau jurnal atau data fluktuasi harga saham didapatkan dari Bursa Efek.

c. Cara memperoleh

   Berdasarkan perolehannya, data dibagi menjadi data primer dan data sekunder. Data primer adalah data yang dikumpulkan secara langsung oleh individu atau suatu lembaga. Data sekunder adalah data yang tidak dikumpulkan secara langsung tapi mengambil data yang dikumpulkan oleh pihak lain misalnya data yang diambil dari seorang peneliti, suatu lembaga survey atau suatu badan penelitian.

d. Waktu pengumpulan

   Data yang dikelompokkan berdasarkan waktu pengumpulan terdiri atas data cross section dan data berkala (*time series*). Data cross section adalah data yang dikumpulkan pada periode waktu tertentu misalnya data pemilu hanya dikumpulkan pada periode pemilu. Data berkala adalah data yang dikumpulkan dari waktu ke waktu misalnya data kelahiran, data kematian, data kemiskinan dan sebagainya.

e. Jenis

   Jenis data ditentukan berdasarkan skala pengukuran. Pengelompokan data berdasarkan jenisnya akan dibahas lebih rinci pada sub bab selanjutnya.

Kegiatan pengumpulan data, baik melalui sensus maupun sampling dapat menempuh bermacam-macam usaha. Upaya-upaya pengumpulan data antara lain:

a. Wawancara (interview), yaitu data diperoleh melalui proses wawancara, dimana responden akan menjawab sejumlah pertanyaan yang diajukan. Lalu jawaban tersebut direkam atau dicatat.
b. Kuesioner (angket), yaitu dengan menyebarkan kuisioner kepada obyek penelitian. Misalnya jika ingin mengetahui kepuasan pelanggan smartphone merk Samsung di Toko A maka disebarkan kuisioner yang berisi sejumlah pertanyaan yang relevan kepada sejumlah pelanggan smartphone merk Samsung pada durasi waktu yang telah ditentukan. Calon responden hanya tinggal mengisi kuisioner tersebut sesuai petunjuk yang diberikan.

c. Observasi, yakni data diperoleh melalui pengamatan terhadap obyek penelitian. Pengamatan tersebut dapat dilakukan di laboratorium maupun dengan terjun langsung ke lapangan. Misalnya untuk mendapatkan data citra sel kanker perlu dilakukan pengamatan terhadap sampel sel kanker melalui mikroskop di laboratorium, sehingga diketahui panjang dan bentuk sel. Untuk mendapatkan data mengenai kondisi jalan raya di area ringroad utara, kabupaten Sleman, Yogyakarta, diperlukan terjun langsung ke lapangan.
d. Tes dan Skala Obyektif yaitu dengan melakukan tes pada obyek yang diteliti dan memberikan nilai tertentu pada jawaban yang diberikan. Misalnya pengumpulan data capaian hasil pembelajaran dilakukan dengan format quiz atau ujian semester, baik berupa ujian tulis maupun lisan.
e. Metode proyektif mengamati atau menganalisis suatu obyek melalui ekspresi luar dari obyek tersebut dalam bentuk karya (lukisan) atau tulisan. Misalnya meminta obyek penelitian untuk menggambar sesuatu atau membuat karangan tentang suatu, umumnya metode ini digunakan pada bidang ilmu sosial humaniora.

1.5 Jenis Data

Jenis-jenis data menurut skala pengukuran terdiri atas data nominal, ordinal, interval dan rasio. Pengetahuan atas jenis data berhubungan dengan bagaimana sebaiknya suatu data itu disajikan (cara penyajian data).

1.5.1 Data nominal

Data nominal berupa angka, hanya memiliki arti sebagai variasi jenis dari sekumpulan data tanpa menunjukkan tingkatan dari kumpulan data tersebut. Data nominal tidak bisa diurutkan. Termasuk data nominal yaitu data dari variabel jenis kelamin, agama, pekerjaan seperti yang ditunjukkan oleh Tabel 1.1 Misalnya untuk data jenis kelamin, diberikan angka 1 untuk kode laki-laki dan angka 2 untuk kode perempuan, angka 1 dan 2 bukan merupakan tingkatan, maksudnya 2 bukan berarti lebih tinggi daripada 1. Data jenis ini belum bisa dilakukan operasi matematis seperti penambahan,pengurangan, perkalian dan pembagian.

Tabel 1.1. Data nominal

No	Nama	Jenis Kelamin	Agama	Pekerjaan
1	Aisyah	Perempuan	Islam	Dosen
2	Sulaiman	Laki-laki	Islam	Kontraktor
3	Ignatius	Laki-laki	Kristen	Dosen
4	Baharudin	Laki-laki	Islam	Psikolog
5	Muh. Surya	Laki-laki	Islam	Dokter
6	I Made Kurniawan	Laki-laki	Hindu	Dokter
7	Maria Magdalena	Perempuan	Katholik	Guru
8	Fransiskus Aditya	Laki-laki	Kriste	Pengacara
9	Nurul Fatikah	Perempuan	Islam	Psikiater
10	Nur Cahyo	Laki-laki	Islam	Kontraktor

1.5.2 Data ordinal

Data ordinal berupa angka yang sudah memiliki tingkatan, data yang satu memiliki tingkatan yang lebih tinggi atau lebih rendah dari data yang lain. Masing-masing klasifikasi yang berupa tingkatan tersebut tidak memiliki jarak yang sama. Tingkat pengurutan dimulai dari yang paling rendah hingga tingkat yang paling tinggi. Ukuran ini berupa urutan ranking berdasarkan tingkatan tertentu, misalnya tingkat pendidikan dan status kelulusan (lihat Tabel 1.2).

Contoh kasus pada perlombaan balap motor. Kita dapati ada juara 1, juara 2 dan juara 3. Angka 1, 2, 3 tersebut memiliki makna tingkatan, yaitu juara 1 lebih cepat daripada juara 2 dan juara 3. Juara 2 lebih cepat daripada juara 3

Meskipun menunjukkan tingkatan, data ordinal ini tidak memiliki jarak yang sama, antara juara 1 dan 2 selisih waktunya 2 menit, antara juara 2 dan juara 3 selisih waktunya 5 menit. Operasi matematis belum bisa dijalankan pada data ordinal karena angka 1, 2 dan 3 itu hanya berupa ranking saja.

Tabel 1.2. Data ordinal

No	Nama	Jenjang Pendidikan	Status Kelulusan
1	Aisyah	S3	Cumlaude
2	Sulaiman	S1	Cumlaude
3	Ignatius	S3	Sangat memuaskan
4	Baharudin	S2	Sangat memuaskan
5	Muh. Surya	S1	Memuaskan
6	I Made Kurniawan	S2	Sangat memuaskan
7	Maria Magdalena	S1	Cumlaude
8	Fransiskus Aditya	S2	Memuaskan
9	Nurul Fatikah	S2	Sangat memuaskan
10	Nur Cahyo	S1	Cumlaude

1.5.3 Data interval

Data interval merupakan data yang menggunakan ukuran interval berupa angka kuantitatif sehingga dapat dilakukan operasi aritmatik penjumlahan dan pengurangan, namun belum bisa dikenai operasi matematis perkalian dan pembagian karena data interval tidak memiliki angka nol mutlak.

Ukuran tersebut dapat memberikan infomasi tentang interval antara suatu obyek dengan obyek yang lainnya, namun informasi absolut suatu obyek tidak dapat diperoleh. Contoh variabel yang menggunakan data interval yakni temperatur, tahun, nomor sepatu.

Pada contoh berat badan berikut ini: 39, 40, 41, 42, angka berat badan tersebut sudah bermakna tingkatan. Maksudnya, 42 lebih berat daripada 41 dan seterusnya. Skala data interval telah memiliki jarak yang sama masing-masing tingkatan tersebut yaitu sebesar 1 kg. Dengan adanya interval tersebut, bisa dimaknai bahwa berat 42 adalah berat 39 ditambah 3 kg, tapi belum bisa dimaknai bahwa berat 42 adalah berat 39 dikali 3.

1.5.4 Data rasio

Data rasio memiliki ukuran berupa angka kuantitatif yang memiliki nilai nol mutlak. Ukuran rasio mirip dengan ukuran interval hanya saja jaraknya diukur tidak dari angka rata-rata atau angka minimal namun bermula dari titik nol yang sesungguhnya, sementara titik nol pada data interval ditentukan secara sembarang.

data dengan tingkatan yang tertinggi karena telah memiliki angka nol mutlak. Contohnya ukuran berat atau ukuran umur dan tinggi badan (lihat Tabel 1.3).

Tabel 1.3. Data rasio

No	Nama	Umur (thn)	Tinggi Badan (cm)
1	Hendriadi	32	176
2	Gde Mahendra	51	173
3	Kamelia	43	169
4	Darwis	41	175
5	Sila	40	169
6	Irwan	41	175
7	Bayu	35	162
8	Sigit	45	157
9	Nurul	32	176
10	Rahadi	51	173

Pada kasus ukuran berat, berat 0 kg berarti tidak ada massa yang ditimbang. Berat 3 kg lebih besar daripada berat 2 kg, berat 2 kg lebih besar daripada berat 1 kg. Dapat dimaknai bahwa 3 kg adalah 2kg + 1kg atau 3kg adalah 3x1kg. Dengan demikian data rasio sudah bisa dikenai semua operasi matematis.

1.6 Penyajian Data

Tahap selanjutnya dalam proses statistika, setelah pengumpulan data ialah menyusun data-data yang berserakan sehingga tersusun dalam bentuk yang sistematis. Aktivitas ini disebut penyajian data. Ferguson dan Takane (1989:16) menjelaskan tujuan penyajian data yakni agar memudahkan membaca, memahami data dan analisis data. Data dapat disajikan dalam bentuk skor, persentase maupun indeks. Bentuk penyajian manakah yang paling tepat ? itu bergantung pada bentuk mana yang paling memberikan manfaat maksimal dalam memahami data.

Adakalanya data disajikan dalam bentuk skor. Skor langsung diambil dari pengukuran variabel tertentu atas obyek penelitian. Data disajikan dalam bentuk persentase pada kondisi yang lain bilamana ingin diketahui posisi data diantara total keseluruhan. Data dapat diubah ke bentuk indeks jika nilai skor di antara keseluruhan data ingin diketahui. Data indeks ini disajikan dalam bentuk desimal.

Contoh 1.1. Diketahui skor hasil ujian matakuliah Statistika dan Probabilitas kelas A dari 50 orang peserta. Jumlah peserta yang gagal dalam ujian sebanyak 20 orang.

- Persentase kegagalan dapat diperoleh dengan cara membagi jumlah peserta yang gagal dengan total peserta ujian Statistika probabilitas kemudian hasilnya dikalikan 100%, sehingga kita dapati (20/50) x 100% = 40%. Berdasarkan perhitungan tersebut, dapat dipahami bahwa persentase maksimal kelulusan adalah 100%, dimana peserta gagal sebesar 40% dan peserta berhasil (lulus ujian) sebesar 60%.
- Indeks kegagalan dapat diketahui dengan mengubah hasil perhitungan dalam format persen menjadi desimal, sehingga diperoleh indeks peserta gagal sebesar 0,40. Berdasarkan angka tersebut dapat dipahami bahwa indeks maksimal adalah 1,00 dimana indeks yang gagal sebesar 0,40 dan indeks yang berhasil sebesar 0,60.

Data dapat disajikan dalam beberapa cara, antara lain dengan membuat statemen (naskah pernyataan), tabel atau daftar dan grafik atau diagram (Sudjana, 1996:15). Penyajian data menggunakan statemen, kurang begitu populer berhubung sulit untuk dipelajari secara cepat. Lain halnya dengan penyajian menggunakan tabel dan grafik, penggunaannya cenderung lebih banyak. Umumnya hasil penelitian atau suatu laporan mendeskripsikan data menggunakan ketiga cara tersebut.

1.6.1 Tabel

Data mentah disusun ke dalam sebuah tabel. Tabel dapat dibagi menjadi 2 jenis berdasarkan cara penyajiannya, yakni:

a. Tabel baris kolom

   Tabel ini memuat keterangan mengenai baris dan kolom. Berdasarkan jumlah karakteristiknya, tabel dibagi lagi menjadi 3 kategori, yaitu:

   1. Tabel satu arah (one way table)

      Tabel satu arah memuat keterangan mengenai satu hal atau satu karakteristik saja, misalnya tahun angkatan (lihat Tabel 1.4).

Tabel 1.4. Tabel satu arah

No	Angkatan	Jumlah Mahasiswa
1	2013	200
2	2014	215
3	2015	190
4	2016	203
5	2017	170

   2. Tabel dua arah (two way table)

      Tabel dua arah menunjukkan dua hal atau dua karakteristik. Misalnya karakteristik jenis kelamin dan usia, seperti dalam Tabel 1.5.

Tabel 1.5. Tabel dua arah

Jenis Kelamin	Usia				Total
	<15	16-31	32-47	>47	
Laki-laki	10	30	40	25	105
Perempuan	15	40	50	30	135

3. Tabel tiga arah (three way table)

   Tabel yang menampilkan keterangan tentang tiga hal atau tiga karakteristik, seperti karakteristik jenis barang, harga dan tahun pembelian, seperti dalam contoh pada Tabel 1.6 berikut ini.

Tabel 1.6. Tabel tiga arah

Barang	2014		2015		2016	
	Banyak	Harga	Banyak	Harga	Banyak	Harga
Proyektor	10	81	9	80,5	13	92
Komputer	8	234,4	13	307,8	11	290,4
Jumlah	18	315,8	22	388,3	24	382,4

b. Tabel distribusi frekuensi

   Tabel menampilkan keterangan tentang distribusi data dalam frekuensi. Misalnya diketahui data ujian Statistika 60 siswa, setiap nilai memiliki frekuensi kemunculan, maka penyajiannya menggunakan tabel distribusi dapat dlihat pada Tabel 1.7. Tabel yang digunakan untuk menyusun data dalam frekuensi dengan sifat distribusi tunggal disebut tabel distribusi tunggal. Tabel yang digunakan untuk menyusun data dalam frekuensi dengan sifat distribusi bergolong disebut tabel distribusi bergolong. Uraian lengkap mengenai bagaimana menyusun tabel distribusi frekuensi bergolong akan dibahas pada Bab 2.

Tabel 1.7. Tabel distribusi frekuensi

No	Nilai	Frekuensi
1	45	2
2	50	3
3	55	1
4	60	3
5	65	5
6	70	6
7	75	19
8	80	5
9	85	4
10	90	6
11	95	2
12	100	4

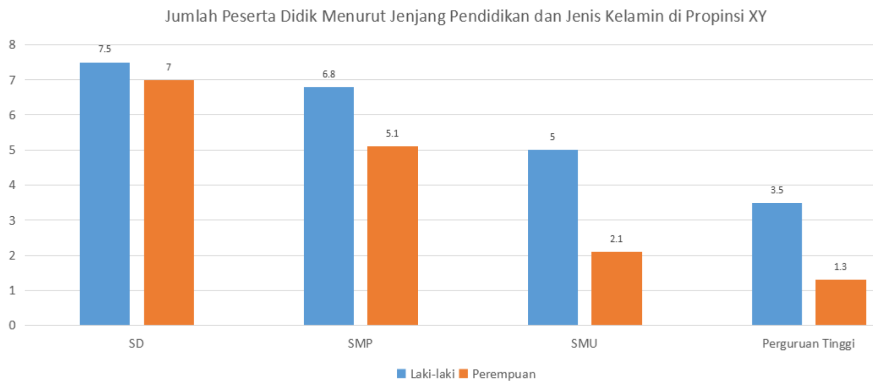
1.6.2 Grafik

Data dapat disajikan secara visual dalam bentuk grafik atau diagram. Penyajian data dalam bentuk ini lebih mudah dipahami. Deskripsi data melalui grafik atau diagram dapat ditampilkan dalam bermacam-macam format, sesuai jenis datanya. Data nominal hendaknya disajikan menggunakan grafik batang, lambang, garis maupun lingkaran (pie chart) Adapun data kontinum yaitu data yang dikumpulkan dengan cara mengukur menggunakan alat ukur dengan skala tertentu, hendaknya disajikan dalam bentuk histogram, poligon atau kurva.

a. Grafik batang

Grafik batang tepat digunakan untuk menyajikan data yang berbentuk kategori atau data dengan masa waktu waktu tertentu misalnya tahun, asalkan masa waktunya tidak terlalu banyak. Cara menyusun datanya yakni sumbu datar untuk menyatakan waktu dan sumbu tegak untuk nilai data, ini berlaku jika grafik dibuat vertikal. Selain itu, grafik batang juga bisa divisualisasikan secara horisontal. Penempatan batang pada kategori satu dengan batang pada kategori berikutnya harus terpisah. Lebar jarak antara batang pertama dengan batang berikutnya digambarkan secara serasi.

Visualisasi data dengan grafik batang dapat diterapkan terdiri atas satu variabel, misalnya jenjang pendidikan, digambarkan dengan diagram batang satu komponen pada Gambar 1.1.



Gambar 1.1. Grafik batang

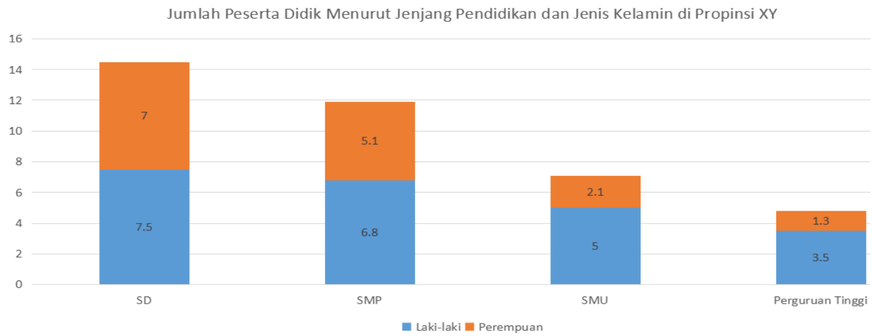
Selain itu, grafik batang dapat digunakan untuk lebih dari satu variabel misalnya jenjang pendidikan dan jenis kelamin, digambarkan dengan grafik batang dua komponen dan tiga komponen.

Contoh 1.3. Diketahui data pada Tabel 1.8 berikut ini. Data tersebut dapat disajikan memakai grafik batang dua komponen dan grafik batang tiga komponen.

Tabel 1.8. Data contoh grafik batang

Jenjang Pendidikan	Banyak Peserta Didik		Jumlah (dalam ribuan)
	Laki-laki	Perempuan	
SD	7.5	7	14.5
SMP	6.8	5.1	11.9
SMU	5	2.1	7.1
Perguruan Tinggi	3.5	1.3	4.8

Gambar 1.1 menunjukkan penyajian Tabel 1.8 dalam grafik batang 2 komponen. Pada grafik batang tiga komponen, yang ditunjukkan oleh Gambar 1.2, terdapat satu komponen tambahan yang melukiskan jumlah peserta didik di tiap kategori.

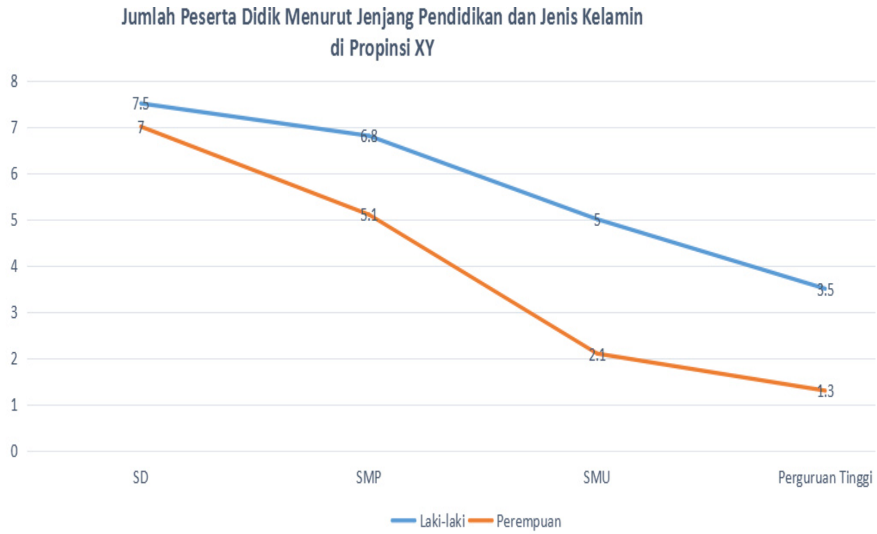


Gambar 1.2. Contoh grafik batang

b. Grafik garis

Grafik garis digunakan untuk melukiskan kondisi yang bersifat kontinu misalnya cadangan beras per tahun, jumlah kelahiran per tahun dan lain-lain. Grafik dapat dilukiskan dengan garis tegas patah-patah maupun garis melengkung.

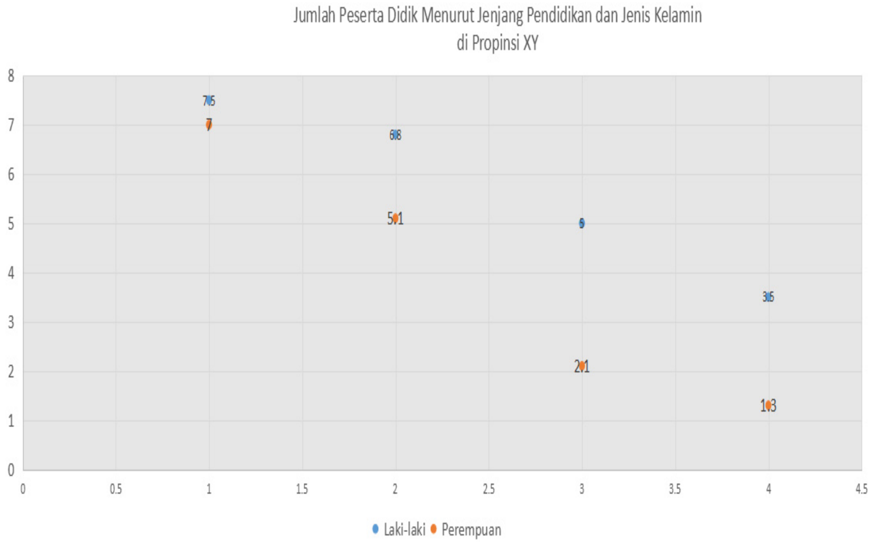
Cara penyusunannya, jenjang pendidikan diletakkan pada sumbu datar dan data tiap jenjang pendidikan diposisikan pada sumbu tegak, sebagaimana yang ditampilkan oleh Gambar 1.3.



Gambar 1.3. Grafik garis

c. Grafik pencar

Grafik pencar digunakan untuk kumpulan data bernilai kuantitatif dengan dua variabel. Grafik diilustrasikan sebagai kumpulan titik-titik yang terpencar. Perhatikan Gambar 1.4.

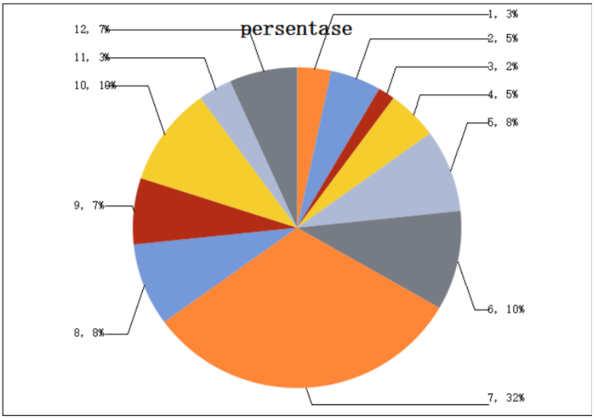


Gambar 1.4. Grafik pencar

d. Grafik lingkaran

Penyajian data menggunakan lingkaran apabila setiap bagian melukiskan kategori data yang telah diubah dalam bentuk derajat. Data disusun ke dalam bagian-bagian lingkaran dengan cara data mentah pada kategori A dibagi 100 kemudian hasilnya dikali 360 derajat, maka akan didapatkan derajat bagian A.

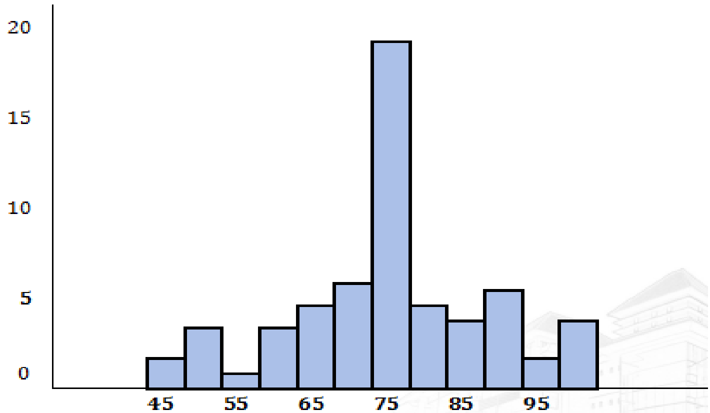
Contoh 1.3 Diketahui data ujian Statistika seperti pada Tabel 1.7. Setiap kelas data menjadi bagian dari lingkaran, berhubung terdapat 12 kelas maka dibuat 12 bagian. Setiap bagian berisi persentase nilai tiap kelas. Besar bagian 7 diukur dengan (32/100) x 360 derajat = 115,2 derajat. Pengukuran ini dilanjutkan ke 11 bagian lainnya sehingga diperoleh hasil seperti pada Gambar 1.5.



Gambar 1.5. Grafik lingkaran

e. Grafik histogram

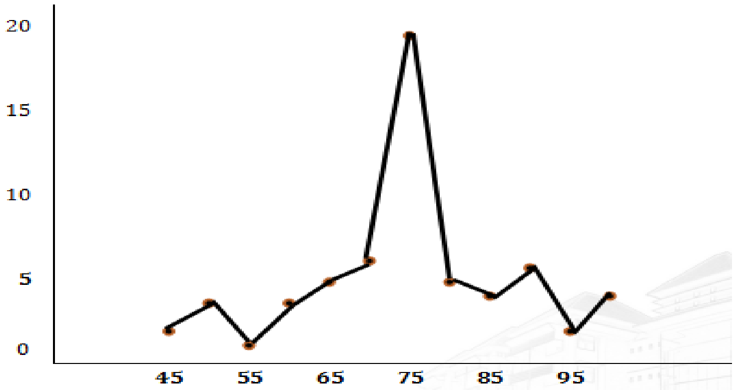
Grafik histogram menyajikan data kontinum menggunakan batang-batang histogram. Batang-batang histogram disusun secara berhimpit. Gambar 1.6 dibentuk dari Tabel 1.7, pada sumbu tegak diletakkan frekuensi dan nilai ujian tiap kelas data diletakkan pada sumbu datar.



Gambar 1.6. Histogram

f. Grafik poligon

Penyusunan grafik poligon dengan cara menghubungkan titik-titik tengah dari batang histogram. Menggunakan data dari Tabel 1.7, diperoleh hasil seperti pada Gambar 1.7.



Gambar 1.7. Poligon

g. Grafik kurva

Kurva digambarkan dengan menghubungkan titik-titik tengah pada batang histogram dengan garis-garis yang dihaluskan.

1.7 Latihan Soal

1. Berikan 5 contoh lain dari masalah ketidakpastian dalam kehidupan sehari-hari yang dapat diselesaikan dengan metode-metode Statistika dan Probabilitas!
2. Catatan: Contoh yang diberikan tidak boleh sama dengan contoh pada materi!
3. Jelaskan peran Statistika dan Probabilitas bagi bidang TI dan yang serumpun!

4. Jelaskan perbedaan dari:
   a. Statistik dan statistika
   b. Populasi dan sampel
   c. Sensus dan Sampling
5. Mengapa diperlukan sampel?
6. Berikan masing-masing contoh dari jenis-jenis data berikut ini:
   a. Data nominal
   b. Data ordinal
   c. Data rasio
   d. Data interval

   Catatan: Contoh yang diberikan tidak boleh sama dengan contoh pada materi!
7. Jelaskan jenis-jenis tabel berdasarkan cara penyajiannya!
8. Abdullah sedang mengerjakan skripsi dengan judul Sistem Pendukung Keputusan (SPK) Manajemen Distribusi Bantuan Sosial. Sistem tersebut akan dipakai oleh suatu lembaga sosial kemasyarakatan. Lembaga tersebut beroperasi hanya di sekitar region DIY-Jawa Tengan saja. Tujuan pengembangan sistem tersebut yaitu untuk membantu pihak lembaga dalam pengambilan keputusan terkait pendistribusian bantuan.
   a. Data-data apa sajakah yang harus dimiliki oleh Abdullah?
   b. Bagaimanakah upaya pengumpulan data yang harus dilakukan Abdullah menurut disiplin Statistika agar dia dapat membangun SPK tersebut?
9. Buatlah grafik batangan (*Bar Chart)* 2 komponen dan 3 komponen berdasarkan data-data yang tercantum dalam Tabel berikut ini ! (Anda bisa menggunakan fasilitas di *Microsoft Excel*).

Tabel 1.9. Data untuk grafik batangan (soal)

Tingkat Sekolah	Banyak Murid		Jumlah
	Laki-Laki		
SD	875	687	1.562
SMP	512	507	1.019
Madrasah Aliyah	347	85	432
SMU	476	342	818
SMK	316	427	743

10. Sajikan data dalam Tabel berikut ini ke dalam diagram lingkaran (*Pie Chart*) ! (Anda bisa menggunakan fasilitas di *Microsoft Excell*)

Tabel 1.10. Data untuk diagram lingkaran (soal)

Keperluan Biaya	
Untuk	%
Pos A	28
Pos B	18
Pos C	14
Pos D	22
Pos E	10
Pos F	8

11. Sajikan data dalam Tabel berikut ini ke dalam grafik garis (*Line Chart*) !

Tabel 1.11. Data untuk grafik garis (soal)

Tahun	1971	1972	1973	1974	1975	1976	1977
Barang yang Digunakan	376	524	412	310	268	476	316

2 Distribusi Frekuensi

Capaian Pembelajaran

Mahasiswa mampu menjelaskan jenis-jenis distribusi frekuensi, tahapan pembuatan tabel distribusi frekuensi dan menggambar grafik poligon dan histogram.

2.1 Pendahuluan

Distribusi frekuensi, penyajiannya dapat dalam bentuk tabel distribusi frekuensi tunggal maupun tabel distribusi frekuensi bergolong. Tabel distribusi frekuensi tunggal telah dibahas di bab 1. Pada bab 2 ini akan dirincikan pembahasan mengenai distribusi frekuensi bergolong. Caranya dengan membuat beberapa data yang nilainya berdekatan dikelompokkan menjadi 1 kelompok agar lebih mudah dibaca.

Distribusi data dengan penggolongan bertujuan untuk memudahkan dalam memahami data, menyajikan data dan mudah dibaca sebagai bahan informasi. Terdapat beberapa komponen penyusun tabel distribusi frekuensi. Perlu dipahami terlebih dahulu beberapa komponen tersebut, sebelum membuat tabel distribusi frekuensi. Mari kita gunakan Tabel 2.1 berikut ini guna membahas komponen penyusunan tabel frekuensi.

Tabel 2.1. Tabel distribusi frekuensi

No	Data	Frekuensi (f)
1	60-62	5
2	63-65	18
3	66-68	42
4	69-71	27
5	72-74	8

Komponen penyusunan tabel frekuensi yaitu:

a. **Kelas interval**

Data dihimpun dalam kelompok-kelompok berbentuk interval a-b. Penggolongan inilah yang dinamakan sebagai kelas interval atau dinotasikan dengan *i*. Kelas interval biasa juga disebut kelas saja, disusun terutur mulai dari data terkecil hingga data terbesar. Kelas yang paling atas dinamakan kelas pertama, kelas kedua hingga kelas terakhir, bergantung pada banyak kelas. Kelas diisikan pada kolom paling kiri dari tabel distribusi frekuensi. Kolom sebelah kanan berisi frekuensi data per kelas.

Berdasarkan data pada Tabel 2.1, diketahui jumlah kelas sebanyak 5 kelas. Kelas pertama berisi data dengan interval 60-62 dengan frekuensi sebesar 5, artinya terdapat 5 buah data dalam rentang 60-62. Kelas kedua berisi data dengan interval 63-65 dengan frekuensi sebanyak 18, maksudnya terdapat 18 data dalam interval tersebut dan seterusnya.

b. **Ujung bawah**

Ujung bawah (Ub) merupakan nilai-nilai yang berada di sebelah kiri kelas, seperti 60,63,66,69 dan 72.

c. **Ujung atas**

Ujung atas (Ua) merupakan nilai-nilai yang berada di sebelah kanan kelas, seperti 62,65,68,71 dan 74.

d. **Rentang**

Rentang (R) adalah selisih dari data terbesar dengan data terkecil. Pada Tabel 2.1 kita telah mengetahui data terbesar adalah 74 dan data terkecil adalah 60.

e. **Panjang kelas**

Panjang kelas (p) adalah selisih dari tiap ujung bawah yang berurutan, yaitu selisih dari ujung bawah kelas setelahnya (ubi+1) dengan ujung bawah kelas sebelumnya (ubi). Perhatikan rumus 2.1 berikut ini:

$$p = Ub_{i+1} - Ub_i \qquad (2.1)$$

Pada Tabel 2.1, panjang kelasnya atau p adalah 3, ini diperoleh dari selisih ujung bawah kelas 2 dengan ujung bawah kelas 1 atau 63-60=3. Jika tiap kelas yang berurutan dicari panjangnya maka semuanya juga bernilai 3. Tabel 2.1 memiliki panjang kelas yang sama.

Panjang kelas juga dapat ditentukan menggunakan cara lain yaitu dengan mencari selisih dari batas atas dengan batas bawah kelas, dinyatakan oleh rumus (2.2) berikut ini:

$$p = La - Lo \qquad (2.2)$$

Untuk kasus dimana tabel distribusi belum disusun, maka panjang kelas dapat dicari dengan membagi rentang (R) dengan banyak kelas (i), sehingga dapat dinyatakan rumus (2.3)

$$p = \frac{R}{k} \tag{2.3}$$

f. Batas kelas.

Batas kelas terdiri atas batas atas (La) dan batas bawah (Lo). Batas kelas merupakan batas nyata yang membatasi kelas yang satu dengan kelas lainnya. Terdapat aturan dalam menentukan batas kelas, yaitu:

1. Jika data dicatat dengan ketelitian hingga satuan, maka:
   i. Batas atas, didapatkan dengan cara ujung atas ditambah 0,5
   ii. Batas bawah, didapatkan dengan cara ujung bawah dikurangi 0,5
2. Jika data dicatat dengan ketelitian hingga satu desimal, maka:
   i. Batas atas, didapatkan dengan cara ujung atas ditambah 0,05
   ii. Batas bawah,didapatkan dengan cara ujung bawah dikurangi 0,05
3. Bagaimana jika data dicatat dengan ketelitian hingga dua desimal, tiga desimal atau n desimal ? Dapat dinyatakan rumus 2.4 dan 2.5 untuk menentukan batas kelas, dimana *n* adalah banyak desimal yang tercantum pada ujung atas dan ujung bawah, yakni:

$$La = ua + \left(5 \times 10^{-(n+1)}\right) \tag{2.4}$$

$$Lo = ub - \left(5 \times 10^{-(n+1)}\right) \tag{2.5}$$

Contoh 2.1. Kelas interval pertama = 30,025-40,025. Ujung atas kelas interval pertama = 40,025, ujung bawahnya = 30,025. Data mengandung ketelitian hingga 3 desimal maka n=3. Batas bawah kelas pertama (*La ke 1*)= 40,025 + (5 x $10^{-(3+1)}$)=40,025-0,0005=40, 0255. Batas bawah kelas pertama (*Lo ke 2*)= 30,025 - (5 x $10^{-(3+1)}$)= 30,0,25-0,0005= 30,0245.

g. Tanda kelas

Tanda kelas dilambangkan dengan (Xi) merupakan nilai tengah dari ujung bawah dan ujung atas, dinyatakan dalam rumus 2.6.

$$x_i = \frac{1}{2}(ua + ub) \qquad (2.6)$$

h. Banyak kelas

Banyak kelas yang dinotasikan sebagai *k*, dihitung menggunakan aturan Sturgess (Dayan, 1986:84), dimana *n* adalah banyak data, yaitu:

$$k = 1 + (3,3)\log n \qquad (2.7)$$

2.2 Pembuatan tabel distribusi frekuensi

Tahap selanjutnya setelah pengenalan komponen tabel distribusi frekuensi adalah membuat tabelnya. Data di bawah ini akan diubah penyajiannya kedalam tabel distribusi frekuensi dengan interval yang sama panjangnya dan nilai tertutup.

Langkah-langkah membuat tabel distribusi yakni:

a. Mengurutkan kumpulan data sehingga terlihat data terkecil hingga terbesar.
b. Menentukan rentang (R) = data terbesar – data terkecil.
c. Menghitung banyak kelas interval (k) menggunakan rumus 2.7 dan memilih mana hasil yang mau dipakai.
d. Menghitung panjang kelas interval dengan menerapkan rumus maka 2.3. Nilai p diambil menurut ketelitian data yang digunakan. Jika data dalam bentuk satuan maka nilai p yang diambil juga dalam bentuk satuan. Jika data dalam bentuk desimal maka nilai p yang diambil juga dalam bentuk desimal.
e. Menetapkan panjang kelas interval yang akan dipakai untuk perhitungan selanjutnya.
f. Menyusun data ke dalam tabel. Langkah awal untuk menyusun data ke dalam tabel distribusi ialah :
   a) Memilih ujung bawah kelas interval pertama dan ujung atas kelas interval terakhir. Bisa diambil data yang lebih kecil dari data yang terkecil tetapi selisihnya harus kurang dari panjang kelas yang telah ditentukan maupun data yang lebih besar dari data terbesar.
   b) Membentuk kelas interval.
      i. Ujung bawah kelas pertama (Ub_1) adalah data dengan nilai terkecil.
      ii. Ujung atas kelas terakhir (Ua_k) adalah data dengan nilai terbesar.
      iii. Untuk kelas *ke-i* sampai kelas ke k–1, dimana k adalah banyak kelas, berlaku rumus 2.8 :

$$Ub_{i+1} = Ub_i + p$$
$$Ua_i = Ub_{i+1} - 1 \qquad (2.8)$$

iv. Kelas interval ke-i adalah: Ub_i s/d Ua_i

c) Kelas interval ini diletakkan di kolom sebelah kiri dan frekuensi tiap kelas diletakkan di kolom sebelah kanan.

d) Frekuensi ditentukan dengan cara menghitung jumlah angka yang berada dalam kelas interval.

e) Hasil tabel distribusi frekuensi bergolong.

Contoh 2.2. Diketahui data nilai peserta ujian masuk di perguruan tinggi ABC, dari 100 orang peserta seperti terlihat pada Gambar 2.1. Buatlah sebaran data dengan menggunakan tabel distribusi frekuensi bergolong.

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	80	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

Gambar 2.1. Sebaran data

Jawaban: membuat tabel distribusi frekuensi menggunakan 6 langkah di atas.

1. Data terbesar 99 dan terkecil 35.
2. Rentang (R)= 99 – 35 = 64
3. Diketahui n = 100, sehingga k=1 + 3,3 log 100 = 7,6.

4. Nilai k =7,6 maka banyaknya kelas yang dapat dibuat adalah 7 atau 8. Misalkan nilai yang dipakai sebagai k= 7
5. Untuk banyak kelas= 7 maka p = 64/7 =9,2. Data dalam bentuk satuan maka nilai p = 9 atau p=10.
6. Menetapkan nilai p yang diambil adalah 10.
7. Menyusun data ke dalam tabel sebagai berikut:
   a) Untuk p=10 maka ujung bawah kelas ke-1 (Ub_1) = 30 dan ujung atas kelas terakhir (Ub_7) = 99. Angka 30 lebih kecil dari data terkecil yang bernilai 34, namun selisihnya < p. Angka 99 lebih besar dari data terbesar yang bernilai 98, namun selesihnya juga < p, sehingga kedua angka tersebut yakni 30 dan 99, dapat dipilih.
   b) Membentuk kelas interval ke-1 hingga kelas ke-7, menggunakan rumus 2.8, sebagai berikut ini:
      i. Kelas interval ke-1 :
         - ✓ Ub_2=Ub_1+p=30+10=40
         - ✓ Ua_1=Ub_2-1=40-1=39
         - ✓ Nilai kelas ke-1: Ub_1 s/d Ua_1 = 30 s/d 39
      ii. Kelas interval ke-2 :
         - ✓ Ub_3=Ub_2+p=40+10=50
         - ✓ Ua_2=Ub_3-1=50-1=49
         - ✓ Nilai kelas ke-2: Ub_2 s/d Ua_2 = 40 s/d 49
      iii. Kelas interval ke-3 :
         - ✓ Ub_4=Ub_3+p=50+10=60
         - ✓ Ua_3=Ub_4-1=60-1=59
         - ✓ Nilai kelas ke-3: Ub_3 s/d Ua_3 = 50 s/d 59
      iv. Kelas interval ke-4 :
         - ✓ Ub_5=Ub_4+p=60+10=70
         - ✓ Ua_4=Ub_5-1=70-1=79
         - ✓ Nilai kelas ke-4: Ub_4 s/d Ua_4 = 60 s/d 69

v. Kelas interval ke-5 :

- ✓ $Ub_6 = Ub_5 + p = 70 + 10 = 80$
- ✓ $Ua_5 = Ub_6 - 1 = 80 - 1 = 89$
- ✓ Nilai kelas ke-5: Ub_5 s/d Ua_5 = 70 s/d 79

vi. Kelas interval ke-6 :

- ✓ $Ub_7 = Ub_6 + p = 80 + 10 = 90$
- ✓ $Ua_6 = Ub_7 - 1 = 90 - 1 = 89$
- ✓ Nilai kelas ke-6: Ub_6 s/d Ua_6 = 80 s/d 89

vii. Kelas interval ke-7 :

- ✓ $Ub_7 = 90$
- ✓ $Ua_7 = 99$
- ✓ Nilai kelas ke-7: Ub_7 s/d Ua_7 = 90 s/d 99

c) Pada kelas ke-1, angka 34 dan 35 terdapat berada diantara rentang 30 s/d 39 maka frekuensinya sejumlah 2. Pada kelas ke-2 terdapat angka 40, 41 dan 45 berada diantara rentang 40 s/d 49 maka frekuensinya sejumlah 3 dan seterusnya hingga kelas ke-7.

d) Tabel distribusi frekuensi selengkapnya ditunjukkan oleh Tabel 2.1. Tabel tersebut menampilkan frekuensi dengan nilai mutlak (tidak dalam bentuk presentase).

Tabel 2.2. Hasil perhitungan tabel distribusi frekuensi

Data	Frekuensi
30-39	2
40-49	3
50-59	11
60-69	20
70-79	32
80-89	25
90-99	17
N	100

2.3 Distribusi Frekuensi Relatif dan Kumulatif

Distribusi frekuensi dapat dibagi menjadi tiga berdasarkan bentuk frekuensinya. Pertama, jika frekuensinya dituliskan dalam bentuk mutlak maka disebut sebagai distribusi frekuensi mutlak (telah selesai dibahas di sub bab 2.2). Kedua, apabila frekuensinya dinyatakan dalam persen maka dinamakan distribusi frekuensi relatif. Ketiga, jika frekuensinya merupakan akumulasi dari frekuensi demi frekuensi maka dinamakan distribusi frekuensi kumulatif.

Penyusunan data ke dalam tabel distribusi frekuensi relatif menggunakan data dari nilai frekuensi mutlak. Misalnya diketahui data nilai ujian Statistika Probabilitas seperti yang tercantum pada Tabel 2.2. Nilai frekuensi mutlak setiap kelas ($f_{abs(i)}$) diubah menjadi frekuensi relatif ($f_{ref(i)}$) menggunakan rumus 2.9, dimana N adalah total frekuensi mutlak.

$$f_{ref(i)} = \frac{f_{abs(i)}}{N} \times 100\% \qquad (2.9)$$

Tabel 2.3. Distribusi frekuensi mutlak

No	Nilai Ujian	Frekuensi Mutlak
1	31-40	2
2	41-50	3
3	51-60	5
4	61-70	14
5	71-80	24
6	81-90	20
7	91-100	12

Semua nilai frekuensi relatif dari kelas ke-1 s/d kelas terakhir diposisikan pada kolom sebelah kanan sesuai dengan masing-masing kelas. Misalnya nilai frekuensi mutlak dari kelas ke-1 (fabs(1))=2 dan N=80, maka nilai frekuensi relatif untuk kelas ke-1 adalah $f_{ref(1)} = \frac{2}{80} \times 100\% = 2{,}5\%$; selengkapnya ditampilkan dalam tabel 2.4.

Tabel 2.4. Distribusi frekuensi relatif

No	Nilai Ujian	Frekuensi	Persentase (%)
1	31-40	2	2,50
2	41-50	3	3,75
3	51-60	5	6,25
4	61-70	14	17,50
5	71-80	24	30,00
6	81-90	20	25,00
7	91-100	12	15,00

Distribusi frekuensi kumulatif mengakomodir data kelas interval yang bersifat terbuka. Distribusinya terbagi menjadi dua, yaitu: distribusi frekuensi kumulatif "KURANG DARI" dan distribusi frekuensi "ATAU LEBIH". Data pada tiap kelas di tabel distribusi frekuensi kumulatif mutlak, diambil dari nilai ujung bawah kelas. Misalnya pada data Tabel 2.3 kelas ke-3, nilai ujung bawahnya adalah 51. Ketika disusun sebagai nilai data bagi distribusi frekuensi kumulatif "KURANG DARI" maka dinyatakan sebagai <51, jika disusun sebagai nilai data bagi distribusi frekuensi kumulatif "LEBIH DARI" maka dinyatakan sebagai >51. Nilai frekuensi kumulatifnya ada yang bersifat mutlak dan bersifat relatif.

Umumnya kelas interval akan bertambah 1 ketika mengubah tabel distribusi frekuensi menjadi tabel distribusi frekuensi kumulatif maupun tabel distribusi frekuensi kumulatif relatif.

2.3.1 Distribusi frekuensi kumulatif mutlak

Perhatikan Tabel 2.3, jika ingin diubah menjadi tabel distribusi frekuensi kumulatif maka akan terbentuk 8 kelas di tabel yang baru karena banyak kelas di Tabel 2.3 adalah 7 kemudian ditambah 1.

Proses untuk mengubah data dari frekuensi mutlak ke frekuensi kumulatif "KURANG DARI" khusus pada kelas ke-1 nilai frekuensi kumulatifnya atau $f_{kum(i)}$ adalah nol karena tidak ada nilai frekuensi mutlak sebelum kelas pertama. Adapun untuk kelas ke-2 dan kelas selanjutnya mengikuti rumus 2.10.

$$f_{kum(i+1)} = f_{abs(i)} + f_{kum(i)} \ \ ; \text{untuk } i = 1 \text{ sampai ke-}n+1 \qquad (2.10)$$

Notasi n adalah banyak kelas pada tabel frekuensi mutlak, notasi $f_{abs(i)}$ adalah nilai frekuensi mutlak pada kelas sebelumnya, $f_{kum(i)}$ adalah nilai frekuensi kumulatif pada kelas sebelumnya, $f_{kum(i+1)}$ adalah nilai frekuensi kumulatif pada kelas yang dicari.

Pembentukan tabel distribusi frekuensi kumulatif "KURANG DARI" untuk kelas ke-1 sampai kelas ke-8 rinciannya sebagai berikut:

a. Kelas ke-1 : $f_{kum(1)} = 0$
b. Kelas ke-2 : $f_{kum(2)} = f_{abs(1)} + f_{kum(1)} = 2 + 0 = 2$
c. Kelas ke-3 : $f_{kum(3)} = f_{abs(2)} + f_{kum(2)} = 3 + 2 = 5$
d. Kelas ke-4 : $f_{kum(4)} = f_{abs(3)} + f_{kum(3)} = 5 + 5 = 10$
e. Kelas ke-5 : $f_{kum(5)} = f_{abs(4)} + f_{kum(4)} = 14 + 10 = 24$
f. Kelas ke-6: $f_{kum(6)} = f_{abs(5)} + f_{kum(5)} = 24 + 24 = 48$
g. Kelas ke-7 : $f_{kum(7)} = f_{abs(6)} + f_{kum(6)} = 20 + 48 = 68$
h. Kelas ke-8 : $f_{kum(8)} = f_{abs(7)} + f_{kum(7)} = 12 + 68 = 80$

Hasil selengkapnya ditampilkan oleh Tabel 2.5.

Tabel 2.5. Distribusi Frekuensi Kumulatif Mutlak "KURANG DARI"

No	Nilai Ujian	Frekuensi Mutlak	Frekuensi Kumulatif
1	<31	2	0
2	<41	3	2
3	<51	5	5
4	<61	14	10
5	<71	24	24
6	<81	20	48
7	<91	12	68
8	<101	0	80

Proses untuk mengubah data dari frekuensi mutlak ke frekuensi kumulatif "LEBIH DARI" khusus pada kelas ke-1 nilai frekuensi kumulatifnya atau $f_{kum(i)}$ adalah total nilai frekuensi mutlak atau dinotasikan sebagai *N*. Adapun untuk kelas ke-2 dan kelas selanjutnya mengikuti rumus 2.11.

$$f_{kum(i+1)} = f_{kum(i)} \text{-} f_{abs(i)} \quad ; \text{untuk } i = 1 \text{ sampai ke-}n+1 \qquad (2.11)$$

Notasi n adalah banyak kelas pada tabel frekuensi mutlak, notasi $f_{abs(i)}$ adalah nilai frekuensi mutlak pada kelas sebelumnya, $f_{kum(i)}$ adalah nilai frekuensi kumulatif pada kelas sebelumnya, $f_{kum(i+1)}$ adalah nilai frekuensi kumulatif pada kelas yang dicari.

Pembentukan tabel distribusi frekuensi kumulatif "LEBIH DARI" untuk kelas ke-1 sampai kelas ke-8 rinciannya sebagai berikut:

a. Kelas ke-1 : $f_{kum(1)} = N = 80$

b. Kelas ke-2 : $f_{kum(2)} = f_{kum(1)} - f_{abs(1)} = 80 - 2 = 78$

c. Kelas ke-3 : $f_{kum(3)} = f_{kum(2)} - f_{abs(2)} = 78 - 3 = 75$

d. Kelas ke-4 : $f_{kum(4)} = f_{kum(3)} - f_{abs(3)} = 75 - 5 = 70$

e. Kelas ke-5 : $f_{kum(5)} = f_{kum(4)} - f_{abs(4)} = 70 - 14 = 56$

f. Kelas ke-6: $f_{kum(6)} = f_{kum(5)} - f_{abs(5)} = 56 - 24 = 32$

g. Kelas ke-7 : $f_{kum(7)} = f_{kum(6)} - f_{abs(6)} = 32 - 20 = 12$

h. Kelas ke-8 : $f_{kum(8)} = f_{kum(7)} - f_{abs(7)} = 12 - 12 = 0$

Hasil selengkapnya ditampilkan oleh Tabel 2.6.

Tabel 2.6. Distribusi Frekuensi Kumulatif Mutlak "LEBIH DARI"

No	Nilai Ujian	Frekuensi Mutlak	Frekuensi Kumulatif
1	<31	2	80
2	<41	3	78
3	<51	5	75
4	<61	14	70
5	<71	24	56
6	<81	20	32
7	<91	12	12
8	<101	0	0

2.3.2 Distribusi frekuensi kumulatif relatif

Misalnya Tabel 2.3 akan diubah menjadi tabel distribusi frekuensi relatif, maka di tabel yang baru, kelas interval bertambah 1 untuk mengakomodir rentang nilai 91 s/d 100.

Nilai frekuensi kumulatif relatif ditentukan dengan mengubah nilai frekuensi relatif menjadi nilai frekuensi kumulatif relatif "KURANG DARI". Khusus nilai pada kelas pertama adalah nol sebab tidak ada nilai frekuensi relatif sebelum kelas pertama. Penentuan nilai kelas ke-2 dan kelas selanjutnya mengikuti rumus 2.12.

$$f_{kum\%(i+1)} = f_{ref(i)} + f_{kum\%(i)} \quad ;\text{ untuk } i = 1 \text{ sampai ke-}n+1 \qquad (2.12)$$

Notasi n adalah banyak kelas pada tabel frekuensi mutlak, notasi $f_{ref(i)}$ adalah nilai frekuensi relatif pada kelas sebelumnya, $f_{kum\%(i)}$ adalah nilai frekuensi kumulatif relatif pada kelas sebelumnya, $f_{kum\%(i+1)}$ adalah nilai frekuensi kumulatif relatif pada kelas yang dicari.

Pembentukan tabel distribusi frekuensi kumulatif relatif "KURANG DARI" untuk kelas ke-1 sampai kelas ke-8 rinciannya sebagai berikut:

a. Kelas ke-1 : $f_{kum\%(1)} = 0$

b. Kelas ke-2 : $f_{kum\%(2)} = f_{ref(1)} + f_{kum\%(1)} = 2,5 + 0 = 2,50$

c. Kelas ke-3 : $f_{kum\%(3)} = f_{ref(2)} + f_{kum\%(2)} = 3,75 + 2,50 = 6,25$

d. Kelas ke-4 : $f_{kum\%(4)} = f_{ref(3)} + f_{kum\%(3)} = 6,25 + 6,25 = 12,50$

e. Kelas ke-5 : $f_{kum\%(5)} = f_{ref(4)} + f_{kum\%(4)} = 17,50 + 12,50 = 30,00$

f. Kelas ke-6: $f_{kum\%(6)} = f_{ref(5)} + f_{kum\%(5)} = 30,00 + 30,00 = 60,00$

g. Kelas ke-7 : $f_{kum\%(7)} = f_{ref(6)} + f_{kum\%(6)} = 25,00 + 60,00 = 65,00$

h. Kelas ke-8 : $f_{kum\%(8)} = f_{ref(7)} + f_{kum\%(7)} = 15,00 + 85,00 = 100,00$

Hasil selengkapnya ditampilkan oleh Tabel 2.7.

Tabel 2.7. Distribusi Frekuensi Kumulatif Relatif "KURANG DARI"

No	Nilai Ujian	Frekuensi Relatif (%)	Frekuensi Kumulatif (%)
1	<31	2,50%	0%
2	<41	3,75%	2,50%
3	<51	6,25%	6,25%
4	<61	17,50%	12,50%
5	<71	30,00%	30,00%
6	<81	25,00%	60,00%
7	<91	15,00%	85,00%
8	<101	0	100,00%

Khusus pada kelas ke-1 nilai frekuensi kumulatif relatif "LEBIH DARI" atau $f_{kum\%(1)}$ adalah total nilai frekuensi relatif atau dinotasikan sebagai *N*. Adapun untuk kelas ke-2 dan kelas selanjutnya mengikuti rumus 2.13.

$$f_{kum\%(i+1)} = f_{kum\%(i)} \text{-} f_{ref(i)} \quad ;\text{untuk } i = 1 \text{ sampai ke-}n+1 \qquad (2.13)$$

Notasi *n* adalah banyak kelas pada tabel frekuensi relatif, notasi $f_{ref(i)}$ adalah nilai frekuensi relatif pada kelas sebelumnya, $f_{kum\%(i)}$ adalah nilai frekuensi kumulatif relatif pada kelas sebelumnya, $f_{kum\%(i+1)}$ adalah nilai frekuensi kumulatif relarif pada kelas yang dicari.

Pembentukan tabel distribusi frekuensi kumulatif "LEBIH DARI" untuk kelas ke-1 sampai kelas ke-8 rinciannya sebagai berikut:

a. Kelas ke-1 : $f_{kum\%(1)} = N = 100$
b. Kelas ke-2 : $f_{kum\%(2)} = f_{kum\%(1)} - f_{ref(1)} = 100 - 2{,}50 = 97{,}50$
c. Kelas ke-3 : $f_{kum\%(3)} = f_{kum\%(2)} - f_{ref(2)} = 97{,}50 - 3{,}75 = 93{,}75$
d. Kelas ke-4 : $f_{kum\%(4)} = f_{kum\%(3)} - f_{ref(3)} = 93{,}75 - 6{,}25 = 87{,}50$
e. Kelas ke-5 : $f_{kum\%(5)} = f_{kum\%(4)} - f_{ref(4)} = 87{,}50 - 17{,}50 = 70{,}00$

f. Kelas ke-6: $f_{kum\%(6)} = f_{kum\%(5)} - f_{ref(5)} = 70,00 - 30,00 = 40,00$

g. Kelas ke-7 : $f_{kum\%(7)} = f_{kum\%(6)} - f_{ref(6)} = 40,00 - 25,00 = 15,00$

h. Kelas ke-8 : $f_{kum\%(8)} = f_{kum\%(7)} - f_{ref(7)} = 15,00 - 15,00 = 0$

Hasil selengkapnya ditampilkan oleh Tabel 2.8.

Tabel 2.8. Distribusi Frekuensi Kumulatif Relatif "LEBIH DARI"

No	Nilai Ujian	Frekuensi Relatif (%)	Frekuensi Kumulatif (%)
1	<31	2,50%	100%
2	<41	3,75%	97,50%
3	<51	6,25%	93,75%
4	<61	17,50%	87,50%
5	<71	30,00%	70,00%
6	<81	25,00%	40,00%
7	<91	15,00%	15,00%
8	<101	0	0%

2.4 Pembuatan histogram, poligon dan ogive

Pada sub bab ini, cara penyusunan tabel distribusi frekuensi relatif dan kumulatif dibahas sebagai bagian dari langkah untuk membuat grafik ogive.

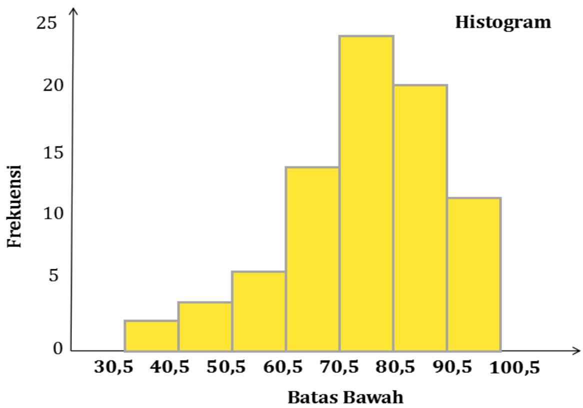
a. Histogram

Grafik histogram dibuat dengan meletakkan batas bawah kelas interval pada sumbu mendatar dan frekuensi mutlak diletakkan pada sumbu tegak. Data contoh yang digunakan yaitu data Tabel 2.2. Pembuatan grafik histogram diawali dengan mencari nilai batas bawah dari setiap kelas interval. Misalnya pada kelas ke-1, diketahui ujung bawahnya adalah 31 kemudian batas bawah dicari dengan menggunakan rumus 2.5 sehingga diperoleh batas bawah kelas ke-1 adalah 30,5. Proses ini dilanjutkan terus hingga batas bawah kelas terakhir sehingga diperoleh Tabel 2.9.

Tabel 2.9. Tabel Distribusi Frekuensi Dengan Batas Bawah

No	Nilai Ujian	Batas Bawah Kelas	Frekuensi
1	31-40	30,5	2
2	41-50	40,5	3
3	51-60	50,5	5
4	61-70	60,5	14
5	71-80	70,5	24
6	81-90	80,5	20
7	91-100	90,5	12

Nilai batas bawah diletakkan pada setiap sisi batang. Batang histogram disusun secara berimpitan dan tinggi batang disesuaikan dengan frekuensi masing-masing kelas interval. Jumlah batang yang terbentuk selaras dengan banyak kelas interval sehingga dihasilkan grafik histogram seperti pada Gambar 2.2.



Gambar 2.2. Grafik Histogram

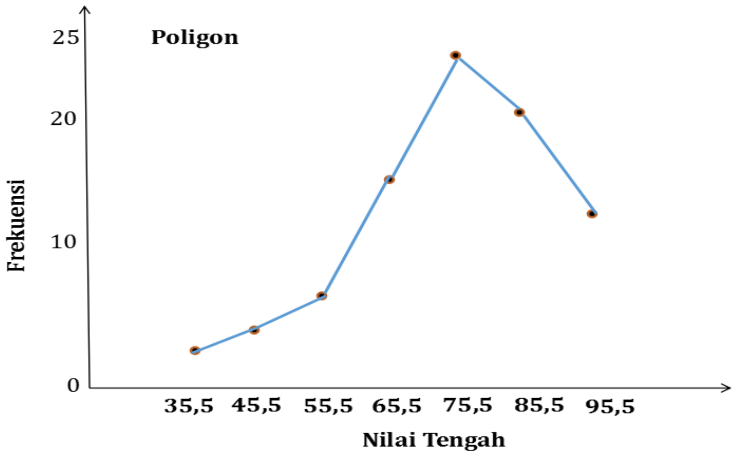
b. Poligon

Grafik poligon dibuat dengan meletakkan tanda kelas atau dinamakan juga nilai tengah kelas interval pada sumbu mendatar dan frekuensi mutlak diletakkan pada sumbu tegak. Data contoh yang digunakan yaitu data Tabel 2.2. Pembuatan poligon diawali dengan mencari nilai tengah dari setiap kelas interval. Misalnya pada kelas ke-5, diketahui ujung bawahnya adalah 71 dan ujung atasnya adalah 80 kemudian diterapkan rumus 2.6 sehingga diperoleh nilai tengah adalah 75,5. Proses ini dilanjutkan terus hingga dihasilkan nilai tengah kelas terakhir.

Tabel 2.10. Tabel Distribusi Frekuensi Dengan Nilai Tengah

No	Nilai Ujian	Nilai Tengah	Frekuensi
1	31-40	35,5	2
2	41-50	45,5	3
3	51-60	55,5	5
4	61-70	65,5	14
5	71-80	75,5	24
6	81-90	85,5	20
7	91-100	95,5	12

Setiap titik poligon diletakkan pada sudut pertemuan antara frekuensi dengan nilai tengah sehingga membetuk koordinat (x,y)=(nilai tengah, frekuensi). Jumlah titik koordinat yang terbentuk selaras dengan jumlah kelas interval. Setiap titik koordinat dihubungkan dengan garis tegas yang patah-patah. Hasilnya dapat dilihat pada Gambar 2.3.



Gambar 2.3. Grafik Poligon

c. Ogive

Grafik ogive dapat dibuat dalam beberapa versi, tergantung pada kebutuhannya. Ogive biasanya digunakan untuk menggambarkan suatu perkembangan, apakah meningkat ataukah menurun, ataukah keduanya sekaligus, misalnya perkembangan penjualan suatu antivirus, perkembangan penggunaan media sosial seperti facebook, instagram dan lain-lain.

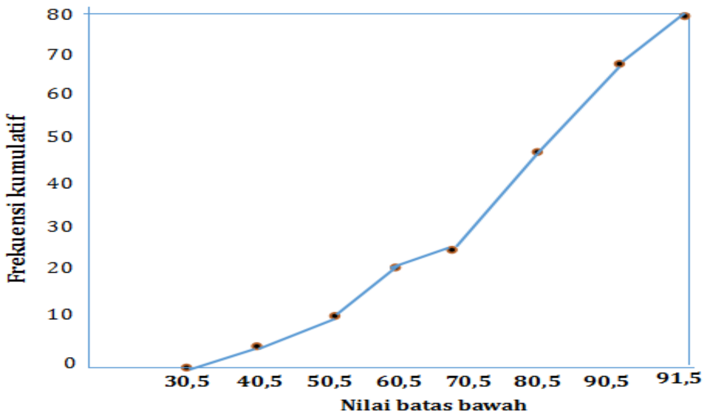
Pembuatan grafik ogive untuk menunjukkan kecenderungan data yang meningkat menggunakan frekuensi kumulatif "KURANG DARI" sedangankan yang menurun menggunakan frekuensi kumulatif "LEBIH DARI". Sumbu mendatar diisi nilai batas bawah dan sumbu tegak diisi nilai frekuensi kumulatif. Menyusun nilai batas bawah dan frekuensi kumulatif sebagai pasangan koordinat (x,y). Titik temu keduanya ditandai dengan titik ogive kemudian setiap titik dihubungkan dengan garis patah-patah.

Contoh 2.3. Buatlah grafik ogive untuk menunjukkan distribusi frekuensi kumulatif meningkat menggunakan data dari Tabel 2.2. Prosesnya dimulai dengan menentukan nilai batas bawah dan frekuensi kumulatif “KURANG DARI”, hasilnya adalah Tabel 2.11.

Tabel 2.11. Tabel Distribusi Frekuensi Kumulatif Meningkat

No	Nilai Ujian	Nilai Tengah	Frekuensi
1	31-40	35,5	2
2	41-50	45,5	3
3	51-60	55,5	5
4	61-70	65,5	14
5	71-80	75,5	24
6	81-90	85,5	20
7	91-100	95,5	12

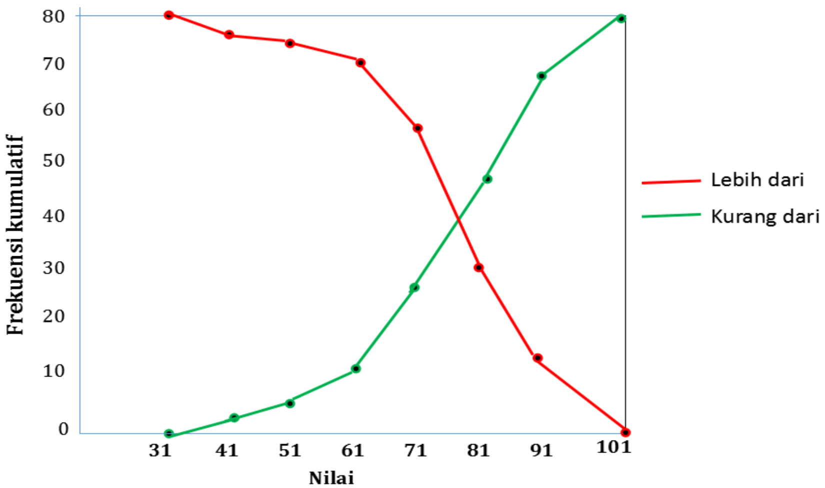
Representasinya di dalam grafik ogive bisa dilihat pada Gambar 2.4.



Gambar 2.4. Grafik Ogive Distribusi Frekuensi Kumulatif Meningkat

d. Grafik ogive dapat dibuat melihat gabungan dari frekuensi kumulatif "KURANG DARI" dengan frekuensi kumulatif "LEBIH DARI". Sumbu mendatar diisi nilai data dan sumbu tegak diisi nilai frekuensi kumulatif. Menyusun nilai data tiap kelas dengan frekuensi kumulatif sebagai pasangan koordinat (x,y). Titik temu keduanya ditandai dengan titik ogive kemudian setiap titik dihubungkan dengan garis patah-patah. Hanya saja, pada grafik ini terhimpun 2 grafik ogive sekaligus, garis grafik yang satu merepresentasikan nilai frekuensi kumulatif "KURANG DARI" dan yang lainnya merepresentasikan nilai frekuensi kumulatif "LEBIH DARI".

Misalnya digunakan data dari Tabel 2.5 dan Tabel 2.6 maka akan dihasilkan grafik ogive seperti pada Gambar 2.5 di bawah ini.



Gambar 2.5. Grafik Ogive Distribusi Frekuensi Kumulatif Gabungan

2.5 Latihan Soal

1. Kapankah kita menggunakan kelas interval yang terbuka, ketika membuat daftar distribusi frekuensi?
2. Jelaskan perbedaan dari ke-3 hal ini:
   a. Distribusi frekuensi mutlak
   b. Distribusi frekuensi relatif
   c. Distribusi frekuensi kumulatif
3. Data di bawah ini merupakan data perkembangan impor beras (dalam ribuan ton) di negara XYZ selama periode 2015-2016.

75	60	85	55	50	70	70	75	80	90
90	45	50	55	60	65	70	75	90	90
70	65	60	60	55	80	85	85	45	40
95	95	70	75	75	70	95	85	60	65
50	50	65	80	70	90	85	40	65	70
70	75	70	75	60	65	95	100	90	80

   a. Buatlah daftar distribusi frekuensi mutlak, gunakan kedua versi banyak kelas yang dihasilkan.
   b. Jelaskan perbandingan hasilnya! Manakah yang lebih tepat?
4. Menggunakan data tabel distribusi frekuensi soal nomor 1, susunlah:
   a. Tabel distribusi frekuensi relatif
   b. Tabel distribusi frekuensi kumulatif
   c. Tabel distribusi frekuensi kumulatif relatif

5. Diketahui sampel data usia perempuan yang mengalami osteoporosis di suatu rumah sakit dalam tabel distribusi frekuensi berikut ini:

Umur	Frekuensi
48-53	2
54-59	2
60-65	6
66-71	5
72-77	5
78-83	2
Jumlah	22

Sajikan data tersebut dengan menggunakan:

a. Histogram,

b. Poligon

6. Sajikan data soal nomor 5 menggunakan Ogive, dengan aturan sebagai berikut:

a. Menunjukkan perkembangan frekuensi kumulatif turun

b. Menunjukkan perkembangan frekuensi kumulatif naik

c. Menunjukkan nilai distribusi frekuensi kumulatif "LEBIH DARI"

d. Menunjukkan nilai distribusi frekuensi kumulatif "KURANG DARI"

e. Menggabungkan soal c dan d

3 Pengukuran Data Statistik

Capaian Pembelajaran

Mahasiswa mampu menjelaskan tendensi sentral berupa mean, median, dan modus serta memahami ukuran variasi data berupa kuartil, desil, dan persentil.

3.1 Ukuran Pemusatan

Ukuran pemusatan merupakan cara mengukur nilai rata-rata dari data yang sudah terdistribusi. Pembagian data dikelompokan menjadi dua yaitu data berkelompok (Group data) dan data tidak berkelompok (Un-group Data) yaitu sejumlah data yang bisa dibuat dalam jumlah kelas tertentu dan interval kelasnya, maka dalam mengukur nilai rata-rata juga harus di bagi menjadi dua, apakah data tersebut termasuk yang berkelompok atau yang tidak berkelompok (Saleh, 1998).

Ukuran Pemusatan juga dikelompokan menjadi tiga jenis skala pengukuran yaitu : Mean, Median dan Modus, dimana setiap pengukuran memiliki fungsi serta sifat yang berbeda-beda, sehingga penggunaannya juga tergantung kebutuhan masing-masing. Untuk setiap jenis skala pengukuran juga dibagi lagi menjadi dua jenis data yaitu data berkelompok dan data berkelompok seperti yang dijelaskan sebelumnya.

Manfaat dari pengukuran pemusatan dan ukuran letak (beberapa penyajian data yang berbentuk tabel, grafis dan diagram) adalah untuk bisa menyaring serta mendapatkan data yang menunjukkan pusat atau nilai tengah dari beberapa data yang tersebar, karena dengan melihat rata-rata berkelompok ini dapat dijadikan data yang mewakili dari keseluruhan data yang ada dalam kelompok tersebut. Menurut Riduwan (2010) pengukuran pemusatan dibagi menjadi rata-rata hitung (mean), rata-rata ukur, rata-rata harmonik, modus, sedangkan ukuran penempatan terdiri dari median, kuartil, desil dan persentil.

3.1.1 Rata-rata hitung (Mean)

Rata-rata hitung atau mean adalah penjumlahan total nilai seluruh data yang dibagi dengan banyaknya data yang ada. Sedangkan menurut Siregar (2010) rata-rata hitung adalah jumlah dari serangkaian data dibagi dengan jumlah data. Sedangkan menurut Rachman (1996) mean adalah jumlah nilai dibagi dengan jumlah/ banyaknya individu.

Mean Data tak berkelompok

Rumus mean untuk data tak berkelompok Un-Group data sebagai berikut: (Siregar, 2010)

$$\bar{x} = \frac{\sum xi}{n} \qquad (3.1)$$

Keterangan :

$\bar{x}$: Rata-rata (mean)

$\sum x_i$: Penjumlahan unsur pada variabel X

n : Jumlah data

Contoh:

Nilai ujian remediasi matematika dari 7 orang mahasiswa yang ikut ujian adalah sbb:

50 65 75 60 70 80 50, tentukan nilai rata-ratanya!

Dengan menggunakan rumus rata-rata diatas maka:

$$\bar{x} = \frac{50+65+75+60+70+80+50}{7} = 64{,}29$$

Mean Data Berkelompok

Perhitungan rata-rata untuk data berkelompok menggunakan rumus sebagai berikut : (Siregar, 2010)

$$\bar{x} = \frac{\sum (t_i.f_i)}{\sum f_i} \qquad (3.2)$$

Keterangan:

t_i : titik tengah kelas ke-i

f_i : Frekuensi kelas ke-i

$\bar{x}$: Rata-rata (mean)

Contoh :

Nilai ujian Matematika dari 60 mahasiswa adalah sebagai berikut:

Tabel 3.1. Distribusi frekuensi nilai ujian Matematika (1)

No	Interval Kelas	Frekuensi
1	38 – 46	3
2	47 – 55	7
3	56 – 64	6

4	65 – 73	17
5	74 – 82	11
6	83 – 91	11
7	92 – 100	5

Berapakah nilai rata-rata ujian matematika dari 60 mahasiswa diatas?

Tabel 3.2. Distribusi frekuensi nilai matematika (2)

No	Interval Kelas	Frekuensi (f_i)	Titik Tengah (t_i)	Perkalian ($t_i.f_i$)
1	38 – 46	3	42	126
2	47 – 55	7	51	357
3	56 – 64	6	60	360
4	65 – 73	17	69	1173
5	74 – 82	11	78	858
6	83 – 91	11	87	957
7	92 – 100	5	96	480
		$\sum f_i = 60$		$\sum t_i.f_i = 4311$

$$\overline{x} = \frac{\sum (t_i.f_i)}{\sum f_i}$$

$$= \frac{4311}{60} = 71.85$$

Jadi rata-rata untuk nilai ujian Matematika diatas adalah 71.85

3.1.2 Modus

Modus adalah nilai data yang paling sering muncul atau dapat juga dikatakan nilai dari data yang memiliki frekuensi terbesar. Modus tidak selalu harus ada dalam data, karena bisa jadi dalam data tidak ada nilai yang sering muncul atau bisa dikatakan frekuensi nilai datanya sama sehingga tidak ada modus didata tersebut. Kemungkinan lainnya yaitu bisa juga lebih dari satu nilai data yang sering muncul karena nilai frekuensinya sama-sama besar. Menurut (Riduwan, 2010) bahwa modus adalah nilai data yang memiliki frekuensi tertinggi dari data yang tidak berkelompok maupun data yang berkelompok atau berbentuk distribusi.

Data tak berkelompok

Mencari modus untuk data tak berkelompok yaitu dengan mencari nilai yang paling sering muncul diantara data yang ada

Contoh : diketahui data nilai UAS 15 orang mahasiswa matakuliah Statistika

45 60 55 75 80 95 70 65 80 70 50 55 70 75 90

Modus data diatas dengan mencari nilai yang paling sering muncul adalah 70 yaitu memiliki kemunculan sebanyak 3 kali.

Modus Data berkelompok

Rumus Modus untuk data berkelompok menurut Siregar (2010):

$$Mo = B_b + P\left(\frac{F_1}{F_1 + F_2}\right) \qquad (3.3)$$

Keterangan:
Mo = Modus
B_b = batas bawah kelas yang mengandung nilai modus
F_1 = Selisih antara nilai frekuensi di kelas modus (*f*) dengan frekuensi sebelum kelas modus (f_{sb})

F_2 = Selisih antara nilai frekuensi di kelas modus (f) dengan frekuensi setelah kelas modus (f_{sd})

Contoh :

Diketahui tabel distribusi frekuensi untuk data UTS nilai Kalkulus

Tabel 3.3. Distribusi frekuensi nilai kalkulus

No	Interval Kelas	Frekuensi
1	38 – 46	3
2	47 – 55	6
3	56 – 64	7
4	**65 – 73**	**17**
5	74 – 82	9
6	83 – 91	13
7	92 – 100	5

Berapakah Modus dari tabel distribusi frekuensi nilai UTS Kalkulus diatas?

Langkah penyelesaian:

a. Mencari nilai frekuensi (f) yang paling tinggi, kalau melihat tabel bisa diketahui nilai frekuensi tertinggi adalah 17. sehingga disimpulkan bahwa modus terletak di interval kelas ke-4

b. Menentukan batas bawah kelas modus (B_b) = 65 - 0.5 = 64.5

c. Menentukan panjang kelas yang mengandung modus P = 65 sampai 73 = 9

d. Menghitung Nilai $f_1 = f - f_{sb} = 17 - 7 = 10$

e. Menghitung Nilai $f_2 = f - f_{sd} = 17 - 9 = 8$

f. Menghitung nilai Modus

$$Mo = B_b + P\left(\frac{F_1}{F_1 + F_2}\right)$$

$$Mo = 64{,}5 + 9\left(\frac{10}{10 + 8}\right)$$

$$Mo = 69{,}79$$

Jadi modus untuk tabel distribusi frekuensi nilai UTS Kalkulus diatas adalah 69,79

3.1.3 Median

Median adalah nilai yang ada ditengah dari data yang sudah diurutkan. Jika jumlah data ganjil maka nilai median didapat dari nilai tengah data yang sudah terurut, sebaliknya jika jumlah data genap maka mediannya didapat dari dua data yang ditengah setelah diurutkan. menurut saleh (1998), median merupakan ukuran rata-rata yang pengukurannya didasarkan atas nilai data yang berada ditengah-tengah distribusi frekuensinya.

Data tak berkelompok

Berikut rumus median data tak berkelompok menurut Siregar (2010).

Letak Median =

$$\frac{n+1}{2} \qquad (3.4)$$

Contoh soal :
Data ganjil sebagai berikut: 60, 53, 62, 49, 40, 45, 51.

Untuk kelompok data diatas, diurutkan terlebih dahulu dari nilai terkecil sampai terbesar, sehingga didapatkan hasil sebagai berikut:

40, 45, 49, 51, 53, 60, 62.

Langkah selanjutnya cari letak median dari data diatas dengan menggunakan rumus $\frac{n+1}{2}$, karena jumlah datanya ganjil jaitu 7, sehingga mediannya didapat dari nilai tengahnya yaitu $\frac{7+1}{2}$ = 4, artinya posisi median berada pada data ke-4, sehingga nilai mediannya = 51 yang persis berada di tengah-tengah.

Data genap sebagai berikut:

38, 28, 34, 42, 25, 30, 46, 23.

Untuk kelompok data diatas, diurutkan terlebih dahulu dari nilai terkecil sampai terbesar, sehingga didapatkan hasil sebagai berikut:

23, 25, 28, 30, 34, 38, 42, 46

Langkah selanjutnya cari letak median dari data diatas dengan menggunakan rumus $\frac{n+1}{2}$, karena jumlah datanya genap jaitu 8, sehingga mediannya didapat dari dua data ditengah setelah diurutkan atau bisa dicari dengan menggunakan letak median yaitu: $\frac{8+1}{2}$ = 4,5, artinya posisi median berada pada data ke-4,5. yaitu bisa disederhakan dengan menggambil data ke-4 = 30 dan ke-5 = 34 sehingga: $\frac{30+34}{2}$ = 32, jadi mediannya adalah 32.

Data berkelompok

Untuk data berkelompok menentukan mediannya dengan persamaan berikut : (Siregar, 2010)

$$Me = B_b + P\left(\frac{\frac{1}{2}n - jf}{f}\right) \qquad (3.5)$$

Keterangan:

Me : Median

B_b : Batas bawah kelas yang mengandung median

P : Panjang kelas

n : Jumlah data

f : banyak frekuensi kelas median

jf : Jumlah dari semua frekuensi kumulatif sebelum kelas median

Contoh :

Berikut data penduduk lansia di DIY berdasarkan usia, berdasarkan tabel distribusi tersebut tentukan nilai mediannya!

Tabel 3.4. Data usia lansia

No	Usia Lansia	Frekuensi
1	51-60	25
2	61-70	50
3	71-80	60
4	81-90	45
5	91-100	20
	Jumlah	200

Langkah penyelesaian:

a. Mencari nilai interval yang mengandung unsur median dengan rumus $\frac{1}{2}n = 1/2\ (200) = 100$

b. Menentukan letak kelas median dengan cara menjumlahkan nilai frekuensi dari kelas awal sampai dengan menunjukan hasil penjumlahan mencapai 100 atau lebih, (25+50+60) = 135, jadi median terletak dikelas ke-3

c. Menentukan batas bawah kelas yang mengandung median (B_b) = 71 - 0,5 = 70,5

d. Menentukan panjang kelas median P = 71 sampai 80 = 10

e. Menentukan jumlah frekuensi dikelas median (f) = 60

f. Hitung jumlah semua frekuensi komulatif dibawah kelas yang mengandung median *jf* = 25+50 = 75

g. Hitung median dengan rumus

$$Me = B_b + P\left(\frac{\frac{1}{2}n - jf}{f}\right)$$

$$Me = 70{,}5 + 10\left(\frac{100 - 75}{60}\right) = 74{,}67$$

Jadi median untuk tabel distribusi frekuensi data lansia DIY adalah 74,67

3.2 Ukuran Letak

Pengertian ukuran letak adalah pengukuran yang dilakukan untuk mengetahui pengelompokan dalam distribusi frekuensi yang nilainya dibagi sama besar, dengan mengelompokannya menjadi tiga yaitu : Kuartil, Desil dan Persentil.

3.2.1 Kuartil

Kuartil adalah nilai atau angka yang membagi data dalam empat bagian yang sama setelah disusun dari yang terkecil sampai data terbesar atau sebaliknya dari data terbesar sampai data terkecil (Riduwan, 2010), atau bisa diartikan secara singkat kuartil adalah membagi data yang sudah terurut menjadi empat bagian nilai yang sama besar. Kuartil dibagi menjadi tiga bentuk yaitu :

a. Kuartil pertama (Q1) yaitu nilai data yang membatasi 25% frekuensi dibagian atas dan 75% frekuensi dibagian bawah
b. Kuartil kedua (Q2) yaitu nilai data yang membatasi 50% frekuensi dibagian atas dan 50% bagian bawah.
c. Kuartil ketiga (Q3) yaitu nilai data yang membatasi 75% frekuensi bagian atas dan 25% bagian bawah.

Data tak berkelompok

Menurut Usman dan Akbar (2008)

$$Letak\ Ki = data\ ke\ \frac{i(n+1)}{4}\text{, dengan i = 1,2,3} \tag{3.6}$$

Contoh :

Contoh data dengan n=12, yaitu : 40, 42, 45, 47, 50, 54, 60, 65, 72, 72, 86, 97

Letak K_1 = data ke $\frac{12+1}{4}$ = data ke 3,25 (yaitu antara data ke-3 dengan data ke-4)

Nilai K_1 = data ke-3 + $\frac{1}{4}$(data ke-4 - data ke-3)

$= 45 + \frac{1}{4}(47 - 45)$

$= 45 + \frac{1}{4}(2)$

$= 45\frac{1}{2}$

Letak K_2 = data ke $\frac{2(12+1)}{4}$ = data ke 6,5 (yaitu antara data ke-6 dengan data ke-7)

Nilai K_2 = data ke-6 + $\frac{1}{2}$ (data ke-7 - data ke-6)

$$= 54 + \frac{1}{2}(60 - 54)$$

$$= 45 + \frac{1}{2}(6)$$

$$= 48$$

Letak K_3 = data ke $\frac{3(12+1)}{4}$ = data ke 9,75 (yaitu antara data ke-9 dengan data ke-10)

Nilai K_3 = data ke-9 + $\frac{3}{4}$(data ke-10 - data ke-9)

$$= 72 + \frac{3}{4}(72 - 72)$$

$$= 72 + \frac{3}{4}(0)$$

$$= 72$$

Data berkelompok

Menurut Usman dan Akbar (2008)

$$Ki = b + p\left(\frac{\frac{in}{4} - F}{f}\right);\ i = 1,2,3 \tag{3.7}$$

Keterangan:

b : Batas bawah kelas K_i yaitu yang memuat K_i

p : panjang kelas interval

F : jumlah frekuensi sebelum kelas K_i

f : frekuensi kelas K_i

in : *i* kali *n* (banyak data)

Contoh :

Diketahui data imunisasi balita dipuskesmas ngaglik yang di kelompokkan berdasarkan umur anak (bulan), tentukan kuartil ke-2 dari data tersebut!

Tabel 3.5. Data imunisasi balita di Puskesmas berdasarkan umur

No	Interval Kelas (bulan)	Frekuensi
1	0 – 6	10
2	7 – 13	12
3	14 – 20	15
4	21 – 27	5
5	28 – 34	3
6	35 – 41	2
7	42 – 48	2
8	49 – 55	1
	Jumlah	50

Letak K_2 = data ke $\frac{2(50+1)}{4}$ = data ke 25,5 (yang artinya K_2 terletak dikelas ke-3) sehingga didapatkan nilai sbb:

b = 14 - 0,5 = 13,5

p = 7

f = 15

F = 22

$$K_2 = 13,5 + 7\left(\frac{\frac{2x50}{4} - 22}{15}\right)$$

$$= 13,5 + 7(0,2)$$

$$= 14,9$$

3.2.2 Desil

Desil adalah ukuran letak yang membagi suatu distribusi frekuensi menjadi 10 bagian yang sama.(saleh, 1998), sehingga nilai-nilai dalam distribusi dapat dibagi menjadi D_1,D_2, $D_{3,\ldots,}D_9$. Desil dapat juga diartikan bahwa membagi data yang sudah diurutkan menjadi sepuluh bagian nilai yang sama besar.

Data tak berkelompok

Menurut Usman dan Akbar (2008)

$$Letak\ Di = data\ ke\ \frac{i(n+1)}{10}, \text{dengan i} = 1,2,3,\ldots,9 \qquad (3.8)$$

Contoh:

Contoh data dengan n=11, yaitu : 30, 32, 35, 37, 40, 44, **50, 55**, 62, 62, 76

Tentukan Desil ke-6 dari data diatas:

Letak D_6 = data ke $\frac{6(11+1)}{10}$ = data ke 7,2 (yaitu antara data ke-7 dengan data ke-8)

$$\text{Nilai } D_6 = \text{data ke-7} + \frac{1}{5}(\text{data ke-8 - data ke-7})$$

$$= 50 + \frac{1}{5}(55 - 50)$$

$$= 50 + \frac{1}{5}(5)$$

$$= 51$$

Data berkelompok

Menurut Usman dan Akbar (2008)

$$Di = b + p\left(\frac{\frac{in}{10} - F}{f}\right);\ i = 1,2,3,\ldots,9 \qquad (3.8)$$

Keterangan:

b : Batas bawah kelas D_i yaitu yang memuat D_i

p : panjang kelas interval

F : jumlah frekuensi sebelum kelas D_i

f : frekuensi kelas D_i

in : *i* kali *n* (banyak data)

Contoh :

Diketahui data anak yang menderita ISPA yang di kelompokkan berdasarkan umur anak, tentukan desil ke-8 dari data tersebut!

Tabel 3.6. Data anak yang menderita ISPA berdasarkan umur

No	Interval Kelas (tahun)	Frekuensi
1	0 – 1	8
2	2 – 3	11
3	4 – 5	7
4	6 – 7	5
5	8 – 9	3
6	10 – 11	4
7	11 – 12	2
	Jumlah	40

Letak D_8 = data ke $\frac{8(40+1)}{10}$ = data ke 32,8 (yang artinya D_8 terletak dikelas ke-5 karena nilai 32,8 ada dijumlah frekuensi kelas ke-5) sehingga didapatkan nilai sbb:

b = 8 - 0,5 = 7,5

p = 2

f = 3

F = 34

$$D_8 = 7{,}5 + 2\left(\frac{\frac{8x40}{10} - 34}{3}\right)$$

$$= 7{,}5 + 2\frac{(-2)}{3}$$

$$= 7{,}5 - \frac{4}{3} = 6{,}17$$

3.2.3 Persentil

Persentil adalah sekumpulan data yang dibagi menjadi 100 bagian yang sama besar, setelah itu disusun mulai dari yang terendah sampai yang tertinggi, sehingga menghasilkan 99 pembagi (usman dan akbar, 2008). atau dalam bisa juga diartikan dengan singkat bahwa persentil adalah ukuran letak yang membagi 100 bagian data yang sama besar dari data yang sudah diurutkan.

Data tak berkelompok

Menurut Usman dan Akbar (2008)

$$Letak\ Pi = data\ ke\ \frac{i(n+1)}{100}, \text{dengan i} = 1,2,3,\ldots,99 \qquad (3.9)$$

Data berkelompok

Menurut Usman dan Akbar (2008)

$$Pi = b + p\left(\frac{\frac{in}{100} - F}{f}\right);\ \ i = 1,2,3,\ldots,99 \qquad (3.10)$$

Keterangan:

b : Batas bawah kelas P_i yaitu yang memuat P_i

p : panjang kelas interval

F : jumlah frekuensi sebelum kelas P_i

f : frekuensi kelas P_i

in : *i* kali *n* (banyak data)

Contoh :

Diketahui data Nilai Ujian Nasional di salah satu SMU di jogja, tentukan desil ke-60 dari data tersebut!

Tabel 3.7. Nilai Ujian Nasional di salah satu SMU di Yogyakarta

No	Interval Kelas (Nilai)	Frekuensi
1	50 – 55	5
2	56 – 61	10
3	62 – 67	8
4	68 – 73	13
5	74 – 79	24
6	80 – 85	15
7	86 – 92	12
8	83 - 99	3
	Jumlah	90

Letak P_{60} = data ke $\frac{60(90+1)}{100}$ = data ke 54,6 (yang artinya D_{60} terletak dikelas ke-5 karena nilai 54,6 ada dijumlah frekuensi kelas ke-5) sehingga didapatkan nilai sbb:

b = 74 - 0,5 = 73,5

p = 6

f = 24

F = 60

$$P_{60} = 73{,}5 + 6\left(\frac{\frac{60x90}{100} - 60}{24}\right)$$

$$= 73{,}5 + 6\frac{(-6)}{24}$$

$$= 73{,}5 - 1{,}5$$

$$= 72$$

3.3 UKURAN VARIAN DAN STANDAR DEVIASI

Pengukuran simpangan menunjukkan seberapa jauh atau dekatnya nilai suatu data dengan nilai rata-rata. Simpangan sering digunakan untuk mengukur luas penyimpangan data/ homogenitas data. Contohnya jika ada tiga data yang memiliki nilai rata-rata (mean) yang sama tapi untuk nilai simpangannya belum tentu memiliki nilai yang sama, seperti dibawah ini:

Tabel 3.8. Mean

No	X1	X2	X3	X4	Mean ($\overline{X}$)
1	40	40	40	40	40
2	20	50	30	60	40
3	10	80	20	50	40

Dari data diatas dapat disimpulkan bahwa nilai simpangan data ke-1 jika dibanding dengan data ke-2 atau data ke-3 memiliki nilai penyebaran yang berbeda walaupun memiliki nilai mean yang sama.

Terdapat beberapa metode pengukuran simpangan data, yaitu rentang atau range, rentangan antar kuartil, rentangan semi antar kuartil, simpangan rata-rata, varians dan koefisien varians, serta angka baku dan simpangan baku.

3.3.1 Rentang (range)

Range adalah jarak penyebaran yang diukur dari nilai terbesar dikurangi dengan nilai terkecil pada sekelompok data. Semakin kecil selisih antara jarak nilai terbesar dan nilai terkecil menunjukkan hasil yang baik atau mengartikan bahwa data mendekati nilai yang terpusat.

Data Tidak Berkelompok

Rumus untuk data tidak berkelompok adalah sebagai berikut :

Jarak (range) = Nilai Terbesar – Nilai Terkecil (3.11)

Contoh :

Data nilai Ujian Metode Numerik sbb: 60 80 75 60 85 95 100 50 55 90, tentukan nilai Range untuk data tersebut!

Pertama data yang ada jika belum terurut maka harus diurutkan dari nilai terkecil sampai nilai terbesar yaitu : 50 55 60 60 75 80 85 90 95 100

Setelah terurut baru dilakukan perhitungan dengan rumus range: 100 - 50 = 50

Sehingga didapatkan nilai range adalah 50

Data Berkelompok

Rumus Range untuk data berkelompok adalah sebagai berikut:

$$\textbf{Range = batas atas kelas tertinggi – batas bawah kelas terendah} \quad (3.12)$$

Contoh:

Contoh data pasien paruhbaya yang menderita penyakit kardiovaskular di sebuah Rumah Sakit di Yogyakarta data bulan Januari 2017. Hitunglah Range dari data tersebut.

Tabel 3.9. Data pasien paruhbaya menderita Kardiovaskular

No	Usia Paruhbaya	Frekuensi
1	51-60	8
2	61-70	3
3	71-80	10
4	81-90	4
5	91-100	2

Range = batas atas kelas tertinggi – batas bawah kelas terendah
= 100 - 51 = 49

3.3.2 Varians Dan Standar Deviasi

Varians adalah salah satu ukuran penyimpangan atau variasi data. Varians dapat menjelaskan berpencarnya suatu data kuantitatif. Varians diberi simbol σ^2 (baca: sigma kuadrat) untuk populasi dan untuk s^2 sampel.

Standar Deviasi adalah akar kuadrat dari varians dan menunjukkan ukuran dispersi yang paling baik dan paling sering digunakan dalam analisis data untuk melihat penyimpangan data terhadap nilai rata-rata karena, standar deviasi mempunyai bentuk linier dari kuadrat selisih antara semua nilai data dengan rata – rata hitungnya, sehingga akan selalu menghasilkan nilai yang positif.

Data Tidak Berkelompok

Rumus Varians untuk data tidak berkelompok adalah sebagai berikut:

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N} \text{ dimana } \mu = \frac{\sum X}{N} \qquad (3.13)$$

σ : Varians Populasi

X : nilai data dalam populasi

μ : Nilai rata-rata hitung dalam populasi

N : jumlah total data/pengamatan dalam populasi

$\sum$: Jumlah data

Rumus Standar Deviasi untuk data tidak berkelompok adalah sebagai berikut:

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}} \qquad (3.14)$$

Data Berkelompok

Rumus Varians untuk data berkelompok adalah sebagai berikut:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \text{ atau } S^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} \qquad (3.15)$$

Ketarangan:

S^2 : Varian sampel

xi : nilai *x* ke-*i*

$\bar{x}$: rata-rata

n : banyak data sampel

$\sum$: Jumlah data

Rumus Standar Deviasi untuk data berkelompok adalah sebagai berikut:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \text{ atau } s = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_1)^2}{n(n-1)}} \qquad (3.16)$$

Keterangan:

s : standar deviasi (simpangan baku)

xi : nilai *x* ke-*i*

$\bar{x}$: rata-rata

n : banyak data sampel

Contoh:

Diketahui data berat badan beberapa siswa yang dijadikan sampel adalah sebagai berikut :47, 57, 78, 50, 56, 80, 46, 40

Dari data tersebut diketahui bahwa jumlah data (*n*) = 8, dan (*n* - 1) = 7. Selanjutnya dapat dihitung komponen untuk rumus varian.

Tabel 3.10. Data berat badan siswa

n	x_i	x_i^2
1	47	2.209
2	57	3.249
3	78	6.084
4	50	2.500
5	56	3.136
6	80	6.400
7	46	2.116
8	40	1.600
	456	27.294

Dari tabel tersebut dapat ketahui:

$$\sum_{i=1}^{n} x_i = 456$$

$$\sum_{i=1}^{n} x_i^2 = 27.294$$

$$\left(\sum_{i=1}^{n} x_i\right)^2 = (456)^2 = 207.936$$

Langkah selanjutnya cari hasil dengan menggunakan rumus varian:

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

$$= \frac{8(27.294) - (207.936)}{(8)(7)}$$

$$= \frac{(218.352) - (207.936)}{(56)}$$

$$= \frac{10.416}{56} = 186$$

Dapat dilihat hasil perhitungan di atas bahwa nilai varian adalah 186. Dari nilai tersebut bisa langsung diperoleh nilai standar deviasi (simpangan baku) dengan cara mengakarkuadratkan nilai varian.

$$S = \sqrt{186} = 13,638$$

3.4 Latihan Soal

1. Data Nilai Tes saringan masuk perusahaan BUMN dari 20 peserta

70	90	85	94
74	87	93	79
74	80	80	74
74	80	92	79
75	73	92	70

   Tentukanlah nilai mean dari data diatas!

2. Dari data No 1, tentukan juga modus serta mediannya!
3. Diketahui tabel distribusi frekuensi untuk data Nilai Ujian Remediasi Algoritma dan Pemrograman dari 20 mahasiswa.

No	Interval Kelas	Frekuensi
1	38 – 46	1
2	47 – 55	2
3	56 – 64	2
4	65 – 73	5
5	74 – 82	7
6	83 – 91	2
7	92 – 100	1

   Berapakah Mean dari tabel distribusi frekuensi Nilai Ujian Remediasi Algoritma dan Pemrograman?

4. Dari soal no.3 tentukan juga nilai modus data berkelompok!
5. Dari soal no.3 tentukan juga nilai median data berkelompok!
6. Diketahui nilai data X sebagai berikut: 3 4 4 5 5 6 7 7 8 9 9 10, tentukanlah nilai kuartil ke-1, kuartil ke-2 dan kuartil ke-3 dari data tersebut!
7. Diketahui nilai pre tes praktikum 12 mahasiswa sbb:
   45, 50, 55, 60, 65, 70, 75, 85, 30, 35, 90, 80
   Tentukan nilai desil ke-3, ke-6 dan ke-8!

8. Diketahui data Nilai Try out di salah satu Bimbingan Belajar, tentukan Kuartil ke-1, dan kuartil ke-3 dari data tersebut!

Tabel 3.11. Nilai Try Out (soal)

No	Interval Kelas (Nilai)	Frekuensi
1	50 – 55	10
2	56 – 61	8
3	62 – 67	15
4	68 – 73	12
5	74 – 79	5
6	80 – 85	8
7	86 – 92	12
8	83 - 99	10
	Jumlah	80

9. Dari tabel Nilai Trayout diatas, tentukan Desil ke-4, ke-5 dan ke-9!
10. Dari tabel Nilai Trayout diatas, tentukan Persentil ke-34, ke-76 dan ke-95!
11. Diketahui nilai data Y sebagai berikut: 3 4 4 5 5 6 7 7 8 9 9 10, Berapakah **rentang/range** dari data tersebut!
12. Diketahui data pasien penderita leukimia berdasarkan umur disebuah Rumah Sakit di Yogyakarta data bulan Februari 2017. Hitunglah Range dari data tersebut.

Tabel 3.12. Data penderita leukimia (soal)

No	Penderita leukimia (tahun)	Frekuensi
1	25-34	6
2	35-44	8
3	45-54	10
4	55-64	2
5	65-74	1

13. Diketahui data tinggi badan anak umur 8 tahun yang dijadikan sampel adalah sebagai berikut :110, 123, 126, 140, 130, 135, 142, 134, 120, 127. tentukan varian dan standar deviasi data sampel tersebut!
14. Diketahui data Nilai X untuk 3 data adalah sbb:

Tabel 3.13. Data nilai X (soal)

No	X1	X2	X3	X4	Mean ($\bar{x}$)
1	40	40	40	40	40
2	20	50	30	60	40
3	10	80	20	50	40

Tentukanlah nilai varian serta nilai standar deviasi untuk data 1, 2, dan ke 3 bandingkan apa perbedaannya?

4 Angka Indeks

Capaian Pembelajaran

Mahasiswa mampu menjelaskan pengertian dan jenis-jenis angka indeks, cara penentuan angka indeks sederhana, agregatif, dan berantai.

Angka indeks adalah adalah sebuah angka yang menunjukkan perubahan yang terjadi terhadap sebuah variabel berdasarkan perbedaan waktu, lokasi atau karakteristik yang lainnya. (Spiegel, Susila, & Gunawan, 1992). Perubahan tersebut dapat berupa kenaikan ataupun penurunan. Angka indeks banyak digunakan dalam bidang bisnis dan perdagangan. Salah satunya adalah angka indeks harga konsumen atau *consumer index price* yang dapat menunjukkan tinggi atau rendahnya biaya hidup di suatu daerah. Jika kita melihat situs bps (bps.go.id) maka akan ditemukan banyak sekali contoh angka indeks lainnya, seperti indeks pembangunan manusia, indeks harga produsen, indeks harga perdagangan dan lain sebagainya.

Tabel 4.1. Daftar indeks harga konsumen (sumber : bps.go.id)

Bulan	2015	2016
Januari	118,71	123,62
Februari	118,28	123,51
Maret	118,48	123,75

Bulan	2015	2016
April	118,91	123,19
Mei	119,50	123,48
Juni	120,14	124,29
Juli	121,26	125,15
Agustus	121,73	125,13
September	121,67	125,41
Oktober	121,57	125,59
November	121,82	126,18
Desember	122,99	126,71

Pada penghitungan angka indeks minimal harus ada dua titik yang digunakan sebagai perbandingan. Dalam konteks perbedaan waktu, maka terdapat dua macam waktu yang bisa dibandingkan, yaitu waktu dasar dan waktu berjalan. Waktu dasar digunakan sebagai dasar perhitungan, sedangkan waktu berjalan adalah nilai yang akan dibandingkan dengan nilai pada waktu dasarnya. Angka indeks dapat digunakan untuk membandingkan dua tipe data, yaitu harga dan jumlah suatu produk. Indeks harga digunakan untuk membandingkan harga sebuah produk, sementara indeks jumlah digunakan untuk membandingkan jumlah sebuah produk pada dua waktu yang berbeda (Lungan, 2006). Perubahan nilai pada harga dan jumlah yang ditunjukkan oleh angka indeks ini sangat bermanfaat untuk mendukung pengambilan keputusan.

4.1 Jenis Angka Indeks

1. Angka Indeks Sederhana

Angka indeks sederhana adalah penghitungan angka indeks untuk satu macam variabel saja (Lind, Marchal, & Waite, 2012). Misalnya, untuk membandingkan harga sebuah laptop pada dua masa yang berbeda maka dapat digunakan penghitungan angka indeks sederhana ini.

Jika p_t adalah nilai suatu variabel pada periode sekarang/berjalan dan p_0 adalah nilai suatu variabel pada periode dasar, maka angka indeks variabel pada periode berjalan terhadap periode dasarnya dapat dihitung dengan rumus sebagai berikut.

$$I_{t,0} = \frac{p_t}{p_0} \; x \; 100 \qquad (4.1)$$

Sebagai contoh sebuah monitor berukuran 15,6 inci dijual seharga Rp. 1.046.000,00 pada tahun 2010. Namun monitor dengan ukuran dan merek yang sama dijual seharga Rp. 756.000,00 pada tahun 2016. Dari data tersebut maka dapat dihitung angka indeks harga monitornya sebagai berikut.

$$I_{2016,2010} = \frac{756000}{1046000} \; x \; 100 = 72{,}27$$

Hasilnya adalah 72,27 yang berarti harga monitor mengalami penurunan dalam waktu 6 tahun sebesar **27,72** persen. Angka tersebut didapatkan dari selisih 100 (persen) dengan 72,27.

Contoh yang lain sebuah ISP (internet service provider) mematok harga untuk pelanggannya sebesar Rp. 150.000,00 per bulan pada tahun 2007. Seiring dengan perkembangan pasar teknologi informasi, harga tersebut dinaikkan setiap tahun hingga mencapai Rp. 325.000,00 pada tahun 2016. Dari data tersebut maka dapat dihitung angka indeks harga berlangganan internet tahun 2016 jika dibandingkan dengan tahun 2007 sebagai berikut.

$$I_{2016,2010} = \frac{325000}{150000} \; x \; 100 = 216{,}67$$

Hasilnya adalah 216,67 yang berarti harga layanan internet berlangganan mengalami kenaikan dalam kurun waktu 11 tahun sebesar 116,67 persen. Angka tersebut didapatkan dari selisih 100 (persen) dengan 216,67.

4.2 Angka Indeks Agregatif

Angka indeks agregatif adalah penghitungan angka indeks yang melibatkan lebih dari satu variabel. Hal ini bertujuan untuk menyederhanakan representasi angka indeks dari berbagai variabel tersebut. Beberapa variabel ini dapat bersifat homogen maupun heterogen. Misalnya, untuk membandingkan harga perangkat komputer yang terdiri dari monitor, keyboard, cpu, dan mouse maka dapat digunakan penghitungan angka indeks agregatif ini. Angka indeks agregatif sangat bermanfaat untuk melihat perubahan harga dari beberapa produk secara keseluruhan.

Indeks agregatif terbagi menjadi dua macam, yaitu indeks agregatif tidak tertimbang dan indeks agregatif tertimbang. Angka indeks agregatif tidak tertimbang adalah angka indeks dari beberapa variabel yang dibandingkan pada dua waktu yang berbeda dengan tidak memperhatikan volume dari masing-masing variabel tersebut. Sementara angka indeks agregatif tertimbang adalah angka indeks dari beberapa variabel yang dibandingkan pada dua waktu yang berbeda dengan memperhatikan volume dari masing-masing variabel tersebut. Perhitungan cacah variabel bertujuan untuk membedakan tingkat prioritas dari masing-masing variabel sehingga angka indeks yang dihasilkan memiliki tingkat representasi yang lebih tinggi terhadap data asalnya.

a. **Angka indeks agregatif tidak tertimbang**

Angka indeks agregatif tidak tertimbang dapat dihitung dengan menjumlahkan harga seluruh variabel pada waktu berjalan kemudian dibagi dengan jumlah harga seluruh variabel pada waktu dasarnya. Rumus dasar untuk menghitung angka indeks agregatif tidak tertimbang adalah sebagai berikut.

$$I_{t,0} = \frac{\sum p_t}{\sum p_0} \, x \, 100\% \qquad (4.2)$$

Tabel 1.x memperlihatkan data harga lima macam perangkat komputer pada Tahun 2011 dan 2017.

Tabel 4.2. Harga perangkat komputer

No	Jenis	Harga	
		2011	2017
1	CPU	2.000.000	3.500.000
2	Monitor	750.000	1.040.000
3	Mouse	40.000	50.000
4	Keyboard	120.000	90.000
	Jumlah	2.910.000	4.680.000

Dari data tersebut maka dapat dihitung angka indeks harga perangkat komputer tahun 2017 jika dibandingkan dengan tahun 2011 sebagai berikut.

$$I_{t,0} = \frac{4680000}{2910000} \, x \, 100\% = 160.82$$

Hasilnya adalah 160.82 yang berarti harga perangkat komputer secara keseluruhan mengalami kenaikan dalam waktu 6 tahun sebesar **60.82** persen. Angka tersebut didapatkan dari selisih 100 (persen) dengan 160.82.

b. Angka indeks agregatif tertimbang

Angka indeks agregatif tertimbang disebut juga dengan angka indeks agregatif terboboti karena menggunakan jumlah variabel sebagai bobot untuk variabel tersebut. Untuk menghitung angka indeks agregatif tertimbang, dapat digunakan rumus Laspeyres maupun Paasche. Pada prinsipnya kedua rumus tersebut memiliki kesamaan, yaitu mengalikan harga keseluruhan dari variabel dengan jumlah keseluruhan variabel tersebut. Perbedaannya terletak pada periode bobot yang digunakan.

Pada rumus Laspeyres, bobot yang digunakan adalah bobot pada waktu dasar, sedangkan pada rumus Paasche, bobot yang digunakan adalah bobot pada waktu berjalan. Perbedaan ini akan menyebabkan angka indeks yang dihasilkan berbeda. Rumus Laspeyres cenderung menghasilkan angka indeks yang lebih besar daripada rumus Paasche untuk data yang sama. Meskipun demikian, angka indeks yang dihasilkan dari rumus Paasche lebih merepresentasikan kondisi pada waktu berjalan dibandingkan rumus Laspeyres (Lind et al., 2012).

Rumus Laspeyres adalah sebagai berikut.

Jika p_t adalah nilai suatu variabel pada periode sekarang/ berjalan dan p_0 adalah nilai suatu variabel pada periode dasar dan q_0 adalah volume dari variabel pada waktu dasar , maka angka indeks variabel pada periode berjalan terhadap periode dasarnya dapat dihitung dengan rumus sebagai berikut.

$$L_{t,0} = \frac{\sum p_t q_0}{\sum p_0 q_0} x\ 100\% \qquad (4.3)$$

Rumus Paasche adalah sebagai berikut.

Jika p_t adalah nilai suatu variabel pada periode sekarang/berjalan dan p_0 adalah nilai suatu variabel pada periode dasar dan q_t adalah volume dari variabel pada waktu berjalan, maka angka indeks variabel pada periode berjalan terhadap periode dasarnya dapat dihitung dengan rumus sebagai berikut.

$$P_{t,0} = \frac{\sum p_t q_t}{\sum p_0 q_t} x\ 100\%$$

Sebagai contoh dengan data yang sama pada Tabel 1.x dan dimodifikasi dengan menambahkan volume penjualan dari masing-masing perangkat sehingga menjadi Tabel 1.y, maka dapat dihitung nilai angka indeks agregatif tertimbang tahun 2017 terhadap waktu dasar tahun 2011 sebagai berikut.

Tabel 4.3. Harga dan volume penjualan perangkat komputer

No	Jenis	2011		2017	
		Harga	Volume	Harga	Volume
1	CPU	2.000.000	30	3.500.000	34
2	Monitor	750.000	12	1.040.000	42
3	Mouse	40.000	31	50.000	33
4	Keyboard	120.000	22	90.000	41
	Jumlah				

Sebelum menggunakan rumus angka indeks, harga dan volume harus dikalikan terlebih dahulu satu per satu pada setiap variabel sehingga Tabel 1.y dimodifikasi menjadi Tabel 1.z (a) dan (b).

Tabel 4.4. Harga dan volume penjualan perangkat komputer dihitung dengan rumus (a) Laspeyres dan (b) Paasche

(a) menggunakan rumus Laspeyres

No	Jenis	2011		2017		p_0q_0 (juta)	p_tq_0 (juta)
		Harga	Volume	Harga	Volume		
1	CPU	2.000.000	30	3.500.000	34	60	105
2	Monitor	750.000	12	1.040.000	42	9	12,480
3	Mouse	40.000	31	50.000	33	1,240	1,550
4	Keyboard	120.000	22	90.000	41	2,640	1,980
	Jumlah					72,880	121,010

(b) menggunakan rumus Paasche

No	Jenis	2011		2017		p_0q_t (juta)	p_tq_t (juta)
		Harga	Volume	Harga	Volume		
1	CPU	2.000.000	30	3.500.000	34	68	119
2	Monitor	750.000	12	1.040.000	42	31,5	43,680
3	Mouse	40.000	31	50.000	33	1,320	1,650
4	Keyboard	120.000	22	90.000	41	4,920	3,690
	Jumlah					105,740	168,020

Dengan rumus Laspeyres

$$L_{t,0} = \frac{121010000}{72880000} \; x \; 100\% = 166$$

Hasilnya adalah 166 yang berarti harga perangkat komputer secara keseluruhan mengalami kenaikan dalam waktu 6 tahun (dengan memperhatikan volume penjualannya) sebesar **66** persen. Angka tersebut didapatkan dari selisih 100 (persen) dengan 166.

Dengan rumus Paasche

$$P_{t,0} = \frac{168020000}{105740000} \, x \, 100\% = 159$$

Hasilnya adalah 159 yang berarti harga perangkat komputer secara keseluruhan mengalami kenaikan dalam waktu 6 tahun (dengan memperhatikan volume penjualannya) sebesar **59** persen. Angka tersebut didapatkan dari selisih 100 (persen) dengan 159.

Seperti telah dijelaskan di awal, terdapat perbedaan angka indeks yang dihasilkan dari rumus Laspeyres dan rumus Paasche untuk data yang sama. Jika kita lihat hasil perhitungan di atas, angka indeks yang dihitung dengan rumus Laspeyres (166) lebih besar dari angka indeks yang dihitung dengan rumus Paasche (159). Angka indeks yang dihasilkan dari rumus Paasche lebih mendekati harga sebenarnya.

4.3 Latihan Soal

1. Jelaskan pengertian dari angka indeks dan berikan contoh penggunaan angka indeks dalam kehidupan sehari-hari.
2. Diketahui harga berlangganan internet pada tahun 2014 sebesar Rp. 350.000,00 per bulan. Hitunglah berapa angka indeks tahun 2016 terhadap tahun 2014 jika diketahui harga berlangganan internet pada tahun 2016 sebesar Rp. 500.000,00 per bulan.
3. Jika diketahui indeks gaji seorang sistem analis pada tahun 2017 terhadap tahun dasar 2013 adalah 1,2 sedangkan besaran gaji pada tahun 2017 adalah Rp. 1.440.000,00 per jam. Hitunglah berapa besaran gaji per jam yang diterima oleh seorang sistem analis pada tahun 2013.

4. Pada Tabel 1.x berikut ini ditampilkan data biaya perbaikan beserta jumlah beberapa perangkat jaringan komputer di suatu instansi pada 3 periode yang berbeda.

Tabel 4.5. Biaya perawatan sejumlah perangkat jaringan komputer

No	Perangkat	2014		2016		2017	
		Biaya (juta)	Jumlah	Biaya (juta)	Jumlah	Biaya (juta)	Jumlah
1	Router	10	5	10	5	10	5
2	Wireless Card	5	10	5	11	5	11
3	LAN Card	4.5	15	4.5	15	4.5	15
4	Modem	3	4	3	4	3	4
5	Switch	1.4	3	1.5	3	1.6	3
6	Bridge	1	5	1	5	1	4
7	Access Point	11	10	10	14	12	16

a. Gunakan rumus Laspeyres untuk menghitung angka indeks biaya perbaikan perangkat jaringan komputer pada tahun 2017 dengan waktu dasar 2014

b. Gunakan rumus Paasche untuk menghitung angka indeks biaya perbaikan perangkat jaringan komputer pada tahun 2017 dengan waktu dasar 2016

c. Gunakan rumus Paasche dan Laspeyres untuk menghitung angka indeks biaya perbaikan perangkat jaringan komputer pada tahun 2016 dengan waktu dasar 2014

5 Analisis Runtun Waktu

Capaian Pembelajaran

Mahasiswa mampu menjelaskan tentang analisis runtun waktu, klasifikasi data berkala, dan penggunaan data berkala

Analisis runtun waktu adalah analisis terhadap variabel atau data yang dikumpulkan secara berurutan dari waktu ke waktu (Chatfield, 2003). Data yang terkumpul disebut juga dengan data berkala. Data berkala memiliki karakteristik interval antara suatu waktu dengan waktu yang lainnya memiliki jarak yang sama (Spiegel et al., 1992). Analisis semacam ini sangat diperlukan terutama untuk meramalkan pola data di masa depan berdasarkan keberadaan data tersebut pada masa berjalan (saat ini).

Analisis runtun waktu dilakukan dengan menarik garis trend pada data berkala yang sudah dipetakan ke dalam diagram pencar.

Menurut Spiegel et al (1992), terdapat empat metode untuk menentukan titik koordinat garis trend tersebut, yaitu metode tangan bebas, metode semi rata-rata, metode rata-rata bergerak, dan metode kuadrat terkecil.

5.1 Metode Tangan Bebas

Dengan menggunakan metode ini, sekumpulan data dipetakan ke dalam diagram pencar. Data tersebut menjadi titik-titik koordinat dalam diagram pencar. Dari sekumpulan titik-titik tersebut kemudian ditarik sebuah garis yang dianggap mewakili seluruh titik-titik yang ada.

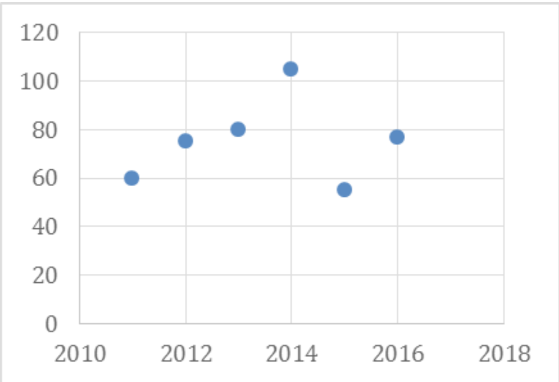
Contoh.

Tabel berikut ini menunjukkan jumlah gangguan jaringan yang dialami oleh sebuah ISP (Internet Service Provider) selama 6 tahun berturut-turut.

Tabel 5.1. Data contoh analisis runtun waktu

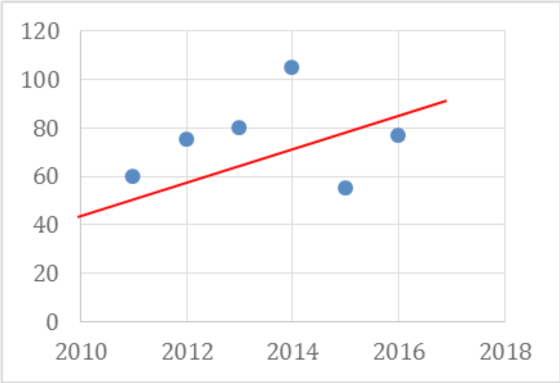
Tahun	2011	2012	2013	2014	2015	2016
Jumlah Gangguan	60	75	80	105	55	77

Jika digambarkan ke dalam diagram pencar maka hasilnya adalah sebagai berikut.



Gambar 5.1. Diagram contoh analisis runtun waktu (1)

Dari titik-titik yang ada di diagram pencar, maka dapat ditarik sebuah garis yang dianggap mewakili arah atau pola keseluruhan titik-titik tersebut.



Gambar 5.2. Diagram contoh analisis runtun waktu (2)

Garis merah yang ada di diagram pencar tersebut tidak menggunakan rumus tertentu. Dalam metode tangan bebas, seseorang dapat dengan bebas menafsirkan arah atau pola data yang dianggap merepresentasikan keseluruhan data.

Garis merah tersebut dapat dijadikan dasar untuk menentukan jumlah gangguan (sumbu Y) pada tahun-tahun yang akan datang.

5.2 Metode Semi Rata-rata

Dengan menggunakan metode ini, sekumpulan data dibagi menjadi dua bagian sama besar. Masing-masing bagian kemudian dihitung rata-ratanya. Setelah itu nilai rata-rata ini, digabungkan dengan nilai tengah (median) dari masing-masing bagian, dijadikan dua titik koordinat untuk menggambar sebuah garis lurus di dalam diagram.

Contoh.

Tahun	2011	2012	2013	2014	2015	2016
Jumlah Gangguan	60	75	80	105	55	77

Langkah penyelesaian:

1. Bagi data menjadi dua bagian sama besar seperti gambar berikut.

2011	2012	2013
60	75	80

(a)

2014	2015	2016
105	55	77

(b)

2. Selanjutnya, dicari rata-rata untuk masing-masing bagian.

   Bagian (a) memiliki nilai semi rata-rata sebesar :

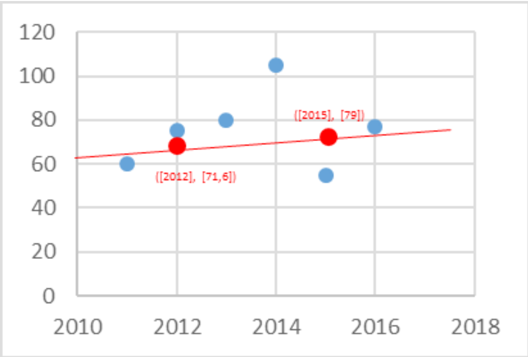
   $$(a) = \frac{60 + 75 + 80}{3} = 71{,}6$$

   Bagian (b) memiliki nilai semi rata-rata sebesar :

   $$(b) = \frac{105 + 55 + 77}{3} = 79$$

3. Didapatkan dua titik koordinat : ([2012], [71,6]) dan (2015,79)
4. Nilai median pada bagian (a) terletak pada titik 2012. Oleh karena itu titik 2012 menjadi titik dasar bagi perhitungan persamaan garis pada langkah 5.

5. Jika digambarkan ke dalam diagram maka hasilnya adalah sebagai berikut.



Gambar 5.3. Diagram contoh analisis runtun waktu dengan metode semi rata-rata

6. Dari dua titik koordinat tersebut dapat dihitung persamaan garisnya sebagai berikut.

$$Y = \mathrm{a} + \mathrm{bX}$$

Nilai a diperoleh dari nilai semi rata-rata pada bagian (a), yaitu sebesar 71,6

Nilai b diperoleh dengan rumus berikut.

$$\mathrm{b} = \frac{Y2 - Y1}{n} \tag{5.1}$$

Y1 = nilai semi rata-rata bagian (a)
Y2 = nilai semi rata-rata bagian (b)
n = jumlah data pada masing-masing bagian

$$\mathrm{b} = \frac{79 - 71{,}6}{3} = 2{,}46$$

Sehingga persamaan garisnya menjadi

$$Y = 71.6 + (2{,}46)X$$

7. Dengan menggunakan persamaan garis tersebut bisa diperkirakan jumlah gangguan yang terjadi pada tahun 2024 sebagai berikut.

Tahun	2011	2012	2013	2014	2015	2016
Nilai X	-1	0	1	2	3	4
Jumlah Gangguan	60	75	80	105	55	77

$Y = 71.6 + (2{,}46)12$
$Y = 101{,}12$

Nilai $x = 12$ karena jika diurutkan dari tahun 2012, tahun 2024 memiliki urutan ke-12.

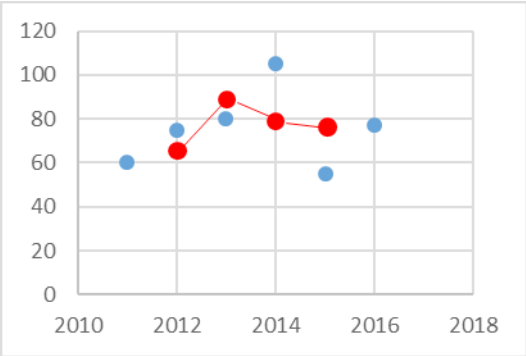
5.3 Metode Rata-rata Bergerak

Dengan menggunakan metode ini, sekumpulan data akan dibagi menjadi kelompok yang terdiri 3,5, atau 7 data kemudian dihitung rata-rata tiap kelompok. Teknik pembagian kelompoknya berurutan mulai dari data ke-1 s.d data ke-3, dilanjutkan data ke-2 s.d data ke-4, dilanjutkan lagi data ke-3 s.d data ke-5 dan seterusnya hingga semua data telah dihitung. Tujuan dari metode ini bukanlah menggambar garis lurus seperti pada metode lainnya melainkan untuk menghilangkan data-data yang sifatnya fluktuatif atau disebut dengan pemulusan data (smoothing) (Spiegel et al., 1992).

Contoh.

Tahun	2011	2012	2013	2014	2015	2016
Jumlah Gangguan	60	75	80	105	55	77
		71.7	86.7	80.0	79.0	

Jika data tersebut digambarkan ke dalam diagram pencar maka hasilnya adalah sebagai berikut.



Gambar 5.4. Diagram contoh analisis runtun waktu dengan metode rata-rata bergerak

Terlihat pada Gambar di atas pergerakan data hasil analisis runtun waktu dengan metode rata-rata bergerak (titik koordinat berwarna merah) tidak terlalu fluktuatif dibandingkan dengan data asalnya (titik koordinat berwarna biru).

5.4 Metode Kuadrat Terkecil

Dengan menggunakan metode ini, sekumpulan data dipetakan ke dalam diagram pencar. Kemudian dengan rumus kuadrat terkecil dicari persamaan garis yang memiliki jarak terkecil dengan tiap-tiap titik koordinat yang berasal dari sekumpulan data tersebut. Metode ini dibahas lebih mendalam dalam Bab 10. Regresi.

5.5 Latihan Soal

1. Apakah yang dimaksud dengan data berkala?
2. Tabel 5.2 berikut ini menunjukkan total penjualan software di suatu daerah dalam 10 tahun terakhir.

Tabel 5.2. Data Penjualan Software

No	Tahun	Jumlah (Juta)
1	2007	13
2	2008	18
3	2009	35
4	2010	35
5	2011	35
6	2012	34
7	2013	28
8	2014	25
9	2015	32
10	2016	24

   Lakukan analisis runtun waktu dengan metode rata-rata bergerak. Gambarkan hasil analisis tersebut ke dalam diagram pencar.
3. Dari hasil analisis pada Soal No. 2, gambarkan garis lurus yang menggambarkan trend data dengan metode semi rata-rata.
4. Carilah persamaan garis pada Soal No. 3
5. Berdasarkan persamaan garis pada Soal No. 4, hitunglah perkiraan total penjualan software pada tahun 2020.

6 Probabilitas

Capaian Pembelajaran

Mahasiswa mampu menjelaskan pengertian probabilitas, operasi dengan kejadian dan menghitung titik sampel, permutasi, dan kombinasi.

Probabilitas atau Peluang merupakan ukuran mengenai seberapa sering suatu peristiwa/kejadian tertentu akan terjadi. Manfaat yang diharapkan untuk menghitung suatu peluang/probabilitas suatu kejadian adalah dapat membantu pengambilan keputusan yang tepat terhadap suatu kejadian yang belum terjadi, karena hampir semua kejadian di kehidupan sehari-hari memiliki peluang yang berbeda-beda.peluang bisa dinyatakan dalam prosentase maupun dalam desimal atau pecahan. Untuk dapat menghitung peluang harus memahami teori ruang sampel.

Beberapa contoh sederhana suatu peluang:

- Berapa peluang munculnya angka 6 dari dadu?
- Peluang Lulus ujian Nasional SMU?
- Peluang Lulus Tes masuk PNS?

6.1 Ruang Sampel, dan kejadian

Ruang Sampel merupakan himpunan semua hasil yang mungkin terjadi dari suatu percobaan dinotasikan dengan lambang S.

Sedangkan yang dinamakan titik sampel adalah tiap-tiap hasil yang mungkin dalam ruang sampel. Definisi dari kejadian atau peristiwa adalah merupakan himpunan bagian dari ruang sampel.

Contoh:

Pada percobaan melempar tiga mata uang, maka diperoleh :

S = {AAA, AAG, AGA, GAA, AGG, GGA, GAG, GGG}

Dimana :
Titik sampel untuk kejadian muncul Angka pada semua dadu : {AAA}
Titik sampel untuk kejadian muncul 2 Angka dan 1 Gambar dalam tiga mata uang : {AAG, AGA, GAA}
Titik sampel kejadian muncul 1 Angka dan 2 Gambar pada tiga mata uang : {AGG, GGA, GAG}
Titik sampel kejadian muncul Gambar pada semua dadu : {GGG}

Kejadian A menyatakan munculnya 2 Gambar pada lemparan 3 mata uang, maka

A = {AGG, GGA, GAG}

Kejadian B menyatakan jumlah kemunculan Angka lebih dari 1, maka
A = {AAA, AAG, AGA, GAA} atau A= {2,3}

6.2 Probabilitas Kejadian

Nilai peluang berada dalam kisaran 0 sampai 1 dan peluang biasanya dinyatakan dalam bentuk pecahan. bila peluang suatu kejadian bernilai 0, maka kejadian tersebut tidak akan terjadi.

Sedangkan bila peluang suatu kejadian makin mendekati 1 maka kejadian tersebut memiliki nilai yang tinggi untuk kejadian tersebut pasti terjadi. Untuk menentukan Peluang suatu kejadian A, semua bobot titik sampel dalam A dijumlahkan. Jumlah ini dinamakan ukuran A atau peluang A dan dinyatakan dengan P(A).

Menurut definisi klasik, bila suatu percobaan dapat menghasilkan N macam hasil yang berkemungkinan sama dan bila tepat sebanyak n dari hasil yang berkaitan dengan kejadian A, maka probabilitas kejadian A adalah:

$$P(A) = \frac{n}{N} \tag{6.1}$$

Keterangan:

P(A) : Probabilitas kejadian A

n : Jumlah hasil yang mungkin untuk satu kejadian

N : Jumlah seluruh kejadian yang mungkin

Contoh :

Percobaan melempar sebuah dadu, Misal A kejadian munculnya angka ganjil dengan diketahui jumlah seluruh kejadian yang mungkin N = 6 sedangkan jumlah hasil yang mungkin untuk kejadian A adalah n = 3, maka tentukan Probabilitas kejadian A, P(A)!

Dengan menggunakan rumus probabilitas maka:

$$P(A) = \frac{3}{6} = \frac{1}{2} \tag{6.2}$$

6.3 Operasi Himpunan

Union atau Gabungan adalah kejadian yang mengandung semua elemen yang termasuk A atau B atau keduanya, dinyatakan dengan $A \cup B$.

Contoh:
A={1,4,5,8} dan B={2,6,8}, maka $A \cup B$={1,2,4,5,6,8}

dan
R={x|2<x<8} dan S={y|6<y<15}, maka
$R \cup S$={z|2<z<15}

Interseksi atau Irisan adalah kejadian yang elemennya termasuk dalam A dan B dua kejadian A dan B, dinyatakan dengan $A \cap B$, Artinya unsur dalam himpunan $A \cap B$ adalah *jika dan hanya jika* yang termasuk dalam himpunan A dan B

Contoh:
A={1,4,5,8} dan B={2,6,8}, maka $A \cap B$={8}

Dan
R={A, B, C, D} dan S={C,F,G}, maka
$R \cap S$={C}

Komplemen suatu kejadian A adalah himpunan semua elemen dalam S yang tidak termasuk dalam A, dinyatakan dengan A'.

Contoh:

S={0,1,2,3,4,5,6,7,8,9,10} A={1,4,5,8} maka A'={0,2,3,6,7,9,10}

Berikut contoh kasus lainnya untuk mengetahui Irisan, Gabungan dan Komplemen:

Misalkan sebuah dadu dilempar maka akan didapatkan nilai angka yang muncul sbb:

Ruang sampel S = {1, 2, 3, 4, 5, 6}

Kejadian munculnya angka ganjil, A = {1, 3, 5}

Kejadian munculnya angka 4 atau lebih, B = {4, 5, 6}

Irisan A dan B, $A \cap B$ = {5}

Gabungan A dan B, $A \cup B$ = {1, 3, 4, 5, 6}

Komplemen dari A'= {2, 4, 6}

6.4 Probabilitas Bersyarat dan Bebas

Definisi Kejadian bebas atau saling pisah adalah dua kejadian A dan B dikatakan saling terpisah (*mutually exclusive*) jika kejadian-kejadian tersebut tidak dapat terjadi secara bersamaan, atau sederhananya antara kejadian A dan B tidak ada yang sama.

Contoh:
Pada percobaan melempar tiga mata uang, maka diperoleh :

Ruang Sampel S = {AAA, AAG, AGA, GAA, AGG, GGA, GAG, GGG}

Kejadian munculnya minimal 2 Angka pada semua dadu
A = {AAA, AAG, AGA, GAA}

Kejadian munculnya minimal 2 Gambar pada semua dadu

B = {AGG, GGA, GAG, GGG}

Sehingga dapat disimpulkan bahwa Kejadian A dan B saling terpisah yaitu :

$A \cap B = \phi$

Definisi kejadian bersyarat adalah dua kejadian A dan B yang menyatakan dua kejadian yang memiliki syarat tertentu dalam koleksi kejadian dalam ruang sampel S, misalkan peluang bersyarat dari B dapat dicari jika peluang kejadian A diketahui, dinyatakan dengan P(B|A), ditentukan dengan rumus :

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \text{, jika P(A)} > 0 \qquad (6.3)$$

Contoh :

Data penerimaan PNS dijogja didapati bahwa laki-laki yang keterima PNS sebanyak 250 sedangkan yang tidak keterima sebanyak 450 orang, untuk data wanita yang keterima PNS sebanyak 230 sedangkan yang tidak keterima sebanyak 370 orang. Dari data tersebut jika dipilih secara acak satu orang, berapa peluang yang terpilih seorang laki-laki dengan status keterima PNS?
L : laki-laki yang keterima PNS
M : semua orang yang keterima PNS

Maka dapat di cari untuk nilai probabilitas semua orang yang keterima PNS tanpa melihat jenis kelamin adalah P(M) = 480/1300 = 0,369

$P(M \cap L) = 250/1300 = 0,192$

Langkah berikutnya dimasukkan kedalam rumus kejadian bersyarat yaitu :

$$P(L|M)=\frac{P(M\cap L)}{P(M)}=\frac{0,192}{0,369}=0,52$$

Maka peluang yang terpilih seorang laki-laki dengan status keterima PNS sebesar 0,52.

6.5 Permutasi

Permutasi adalah kemungkinan pemilihan obyek *r* yang berlainan (*distinguishable*) dari suatu himpunan *n* (Baron, 2014:23). Permutasi menerima susunan atau urutan yang berbeda/berlainan yang dibentuk oleh sebagian atau keseluruhan obyek atau unsur *r*, yang diambil dari sebagian atau keseluruhan himpunan *n*, dimana berlaku r ≤ n. Urutan diperhatikan dalam permutasi, sehingga AB ≠ BA.

Misalnya, diketahui bilangan asli atau A dari 1 s/d 5, A={1,2,3,4,5}. Banyaknya anggota himpunan obyek yang tersedia (n) sebanyak 5 . Susunan permutasinya antara lain: permutasi dengan r=3 yakni {1,2,3} atau {1,3,2} atau r=5 yakni {1,3,2,4,5} dan seterusnya, dengan syarat dalam 1 susunan permutasi tersebut, jumlah maksimal obyek adalah 5 atau r ≤ n.

Banyak permutasi dapat dihitung dengan memperhatikan 2 kondisi dari obyek *r* :

a. Permutasi dengan penukaran/penggantian obyek *r* (*with replacement*)

$$_nP_r=n.n.n.......n=n^r \qquad (6.4)$$

b. Permutasi tanpa penukaran/penggantian obyek r (*without replacement*)

Selama pengambilan sampel tanpa penggantian obyek r, jumlah obyek yang mungkin terpilih berkurang 1 setiap kali sebuah obyek dijadikan sampel, sehingga banyaknya permutasi yang dihasilkan dinotasikan sebagai $P(n,r)$ atau ${}_nP_r$:

$$ {}_nP_r = n(n-1)(n-2)...(n-r+1) = \frac{n!}{(n-r)!} \tag{6.5} $$

Contoh

Membuat *password* acak dimana urutan karakter sangat penting disini dengan kondisi karakter dapat ditukar/diganti-ganti. Jika ketentuannya dari 10 digit huruf alpabet, 26 huruf kecil (*lower case*) dan 26 huruf besar (*capital*), berapa banyak *password* 8 karakter berbeda dapat dibuat?

n= 10 + 26 +26 = 62

r = 8

$$ {}_{62}P_8 = 62^8 = 218{,}340{,}105{,}584{,}896 $$

Maka dapat dibuat 218,340,105,584,896 password 8 karakter yang berbeda.

Contoh

Tentukan banyaknya susunan 2 huruf yang diambil dari himpunan huruf= { V,W,X,Y,Z} !

n= jumlah himpunan huruf=5

r=2

$$_5P_2 = \frac{5!}{(5-2)!} = \frac{5\times4\times3\times2\times1}{3\times2\times1} = 20$$

Contoh

Berapa banyak cara pemilihan tempat duduk, jika 5 orang disediakan 3 kursi, sementara ada 1 orang yang selalu duduk di satu kursi?

n= 4 (karena 1 orang selalu duduk)

r= 2 (karna 1 kursi selalu diduduki)

$$_4P_2 = \frac{4!}{(4-2)!} = \frac{4\times3\times2\times1}{2\times1} = 12 \text{ cara}$$

6.6 Kombinasi

Kombinasi adalah kemungkinan pemilihan obyek *r* yang tidak berlainan (*indistinguishable*) dari suatu himpunan *n* (Baron, 2014:24). Pada kombinasi, susunan-susunan dapat dibentuk dari obyek-obyek dengan mengambil sebagian atau seluruh obyek anggota himpunan tanpa memberi arti pada urutan obyek dari masing-masing susunan tersebut. Kombinasi tidak mementingkan urutan atau susunan dalam pemilihannya sehingga AB=BA.

Banyaknya kombinasi dari r obyek yang diambil dari himpunan n dinotasikan sebagai $C(n,r)$ atau $_nC_r$ atau $\binom{n}{r}$:

$$_nC_r = \frac{n!}{r!(n-r)!} \tag{6.6}$$

Contoh 4 :

Bila ada 4 kimiawan dan 3 fisikawan, hitung banyaknya panitia beranggota 3 orang terdiri dari 2 kimiawan dan 1 fisikawan?

Banyaknya terpilih kimiawan = ${}_4C_2 = \frac{4!}{2!(4-2)!} = 6$

Banyaknya terpilih fisikawan = ${}_3C_1 = \frac{3!}{1!(3-1)!} = 3$

Banyaknya panitia beranggota 3 orang gabungan dari kimiawan dan fisikawan= 6 x3 =18

6.7 Latihan Soal

1. Dari 5 pria dan 3 wanita akan dibentuk panitia beranggota 3 orang. Berapa banyak panitia dapat dibuat jika:
   a. Tanpa pembatasan?
   b. Dengan 2 pria dan 1 wanita?
2. Lima orang sarjana Ekonomi dan tujuh sarjana Hukum, akan dibuat Tim yang beranggota 2 SE dan 3 SH. Berapa banyak cara untuk membuat Tim tsb jika:
   a. Tiap orang dapat dipilih secara bebas
   b. 1 SH harus ikut Tim
   c. 2 SE tidak boleh ikut Tim
3. Berapa banyak kemungkinan 3 dari 10 folder di direktori D dapat terinfeksi virus?
4. Dalam kompetisi tenis meja yang terdiri dari 10 regu akan dipilih juara 1, 2, dan 3. Berapakah banyak cara memilih
5. Dari 10 orang pengurus suatu ekstrakurikuler akan dipilih seorang ketua, wakil ketua, sekretaris, bendahara, dan humas. Berapakah banyak cara pemilihan pengurus?

7 Distribusi Probabilitas Diskrit

Capaian Pembelajaran

Mahasiswa mampu menjelaskan distribusi peluang diskrit, distribusi peluang seragam, dan distribusi peluang binomial

7.1 Data Kuantitatif

Data kuantitatif terdiri dari : Variable Diskrit dan Variabel Kontinyu

a. Variabel Diskrit

Setiap nilai yang ada pada variabel diskrit memiliki nilai peluang yang jika dijumlahkan akan memiliki nilai total sama dengan satu. Variabel diskrit berbentuk bilangan bulat.

b. Variabel Kontinyu

Variabel kontinu merupakan kebalikan dari variable acak diskrit. Jika pada variable acak diskrit nilainya didapat dari atau diperoleh dengan cara menghitung atau membilang, pada Variabel acak kontinu nilainya diperoleh dari atau diperoleh dengan melakukan pengukuran.

7.2 Distribusi Peluang Variabel Acak Diskrit

Ciri-ciri distribusi variabel acak diskret :

a. Jumlah total peluangnya sama dengan 1

b. Peluang dari suatu hasil adalah antara 0 sampai 1

c. Hasilnya tidak terkait satu sama lain

Contoh :

a. Lemparan 2 buah mata uang logam
b. Lemparan sebuah dadu 2 kali

7.3 Macam – Macam Distribusi Diskrit

7.3.1 Distribusi Binomial

Distribusi Probabilitas Binomial adalah ukuran penyebaran data pada sebuah percobaan dimana hasilnya sesuai dengan percobaan Bernoulli. Distribusi Binomial dapat ditandai dari:

1. Setiap percobaan hanya menghasilkan 2 kejadian.
2. Percobaan bersifat independent atau dengan pengembalian.

Persyaratan percobaan binomial:

1. Percobaan terdiri atas *n* usaha yang berulang
2. Tiap usaha memberi hasil yang dapat ditentukan dengan **sukses** atau **gagal**
3. Peluang sukses dinyatakan dengan *p*, tidak berubah dari usaha yang satu ke usaha berikutnya
4. Tiap usaha bebas dengan usaha lainnya

Contoh :

1. Besarnya probabilitas lahirnya 2 orang anak wanita dari keluarga dengan 4 orang anak jika peluang kelahiran wanita adalah sama dengan kelahiran pria.
2. Besarnya probabilitas kemunculan pengambilan 2 bola merah dan biru dalam box.

Distribusi binomial

$$f(x;n;p) = (c_x^n)(p)^x(q)^{n-x}, \text{ untuk } x = 0,1,2,3,4 \qquad (7.1)$$

n: banyaknya ulangan
x: banyak keberhasilan dalam peubah acak X
p: peluang berhasil pada setiap ulangan
q: peluang gagal = 1 - p pada setiap ulangan

Contoh Soal :

Pada sebuah klinik terdapat 4 pasien yang menderita suatu penyakit yang jarang terjadi, ditempatkan pada ruang isolasi yang saling terpisah. Berdasarkan penelitian, diketahui peluang penyakit tersebut dapat disembuhkan adalah 1/3. Berapakah peluang tepat 2 orang yang sembuh?

Penyelesaian :
Diketahui :

$$n = 4, p = \frac{1}{3}, dan\ Q = 1 - \frac{1}{3} = \frac{2}{3}$$

maka dapat dirumuskan distribusi peluangnya sebagai berikut:

$$f\left(x;4;\frac{1}{3}\right) = \binom{4}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{4-x}, \text{ untuk } x = 0,1,2,3,4$$

- tepat 2 orang yang sembuh : $P(X = 2)$

$$P(X = 2) = f\left(2;4;\frac{1}{3}\right) = \binom{4}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{4-2}$$

$$= \frac{4!}{(4-2)!\,2!}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2 = \frac{24}{81}$$

7.3.2 Distribusi Multinom

Perluasan dan distribusi binomial adalah distribusi multinomial. Misalkan sebuah eksperimen menghasilkan peristiwa-peristiwa $E_1, E_2, \ldots\ldots, E_3$ dengan peluang $P(E_1), \pi^2 = P(E_2), \ldots. \pi_K = P(E_K)$.terhadap eksperimen ini kita lakukan percobaan sebanyak 4 kali,maka peluang akan terdapat x_1, peristiwa E_1, x_2 peristiwa $E_2, \ldots.,$ x_k peristiwa E_K diantara N, ditentukan oleh distribusi multinomial berikut:

$$P(x_1, x_2, \ldots\ldots x_k) = \frac{N!}{X_1! X_2! \ldots X_K!} \pi_1^{X_1} . \pi_2^{X_2} \ldots\ldots \pi_K^{X_K} \qquad (7.2)$$

Dengan $X_1 + X_2 + \ldots\ldots + X_K = N$ dan $\pi_1 + \pi_2 + \ldots\ldots + \pi_K = 1$, sedangkan $0 < \pi_1 < 1$

Contoh:

Sebuah kartu diambil dan sekotak kartu Bridge berisi 52 yang dikocok sempurna.Hasilnya dicatat,kemudian kartu dikembalikan. Bila percobaan itu diulangi 3 kali, berapa peluang terambil 1 As,1 Quen, dan 1 King.

Penyelesaian:

Diketahui

$N=3, x_1=1, x_2 = 1, x_3 = 1$

$\pi_1 = P(E_1) = \frac{4}{52} = \frac{1}{13}, \pi_2 = P(E_2) = 4/52 = 1/13, \pi_3 = P(E_3) = \frac{4}{52} = \frac{1}{13}$

$P(1,1,1) = \frac{3!}{1!1!1!} . (\frac{1}{3})^1 . (\frac{1}{3})^1 . . (\frac{1}{3})^1 = \frac{3.2.1}{1!1!1!} = (\frac{1}{2197}) = \frac{6}{2197}$

Bila dua dadu dilempar 6 kali,berapakah peluang mendapat jumlah 7 atau 11 muncul 2 kali,sepasang bilangan yang sama satu kali kombinasi lainnya 3 kali?

Jawab:

Diketahui

E_1=jumlah 7 atau 11 muncul,π_1=P(E_1)=8/36= 2/9

E_2=pasangan bilangan yang sama muncul, π_2=P(E_2)=6/36=1/6

E_3 =baik pasangan yang sama maupun jumlah 7 atau 11 tidak muncul,

π_3=P(E_3)=(36-14)136=22/36=11/18

$$P(2,1,3)=\frac{6!}{2!1!3!}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{8}\right)^3=\frac{6.5.4.3!}{2.1.1.3.2.1}\left(\frac{4}{8}\right)\left(\frac{1}{6}\right)\left(\frac{1331}{5832}\right)=\frac{5324}{283852}=0{,}1127$$

7.3.3 Distribusi Hipergeometrik

Suatu percobaan disebut percobaan hipergeometrik jika:

1. Sampel acak berukuran n diambil dari N benda
2. Sebanyak k benda disebut sukses, dan N-k disebut gagal

Definisi: Banyaknya sukses x dipercobaan hipergeometrik disebut p.a hipergeometrik

Fungsi massa peluangnya (fmp)

$$p(X = x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, x = 0,1, \dots n \qquad (7.3)$$

Karena nilainya bergantung pada banyaknya yang sukses k dalam n barang yang dipilih acak dari sebanyak N

$sifat: x \sim h(x)$

$(i) E(x) = \mu = \frac{nk}{N}$

$(ii) \sigma^2 = Var(x) = \frac{N-n}{N-1} = n.\frac{k}{N}(1 - \frac{k}{N})$

Contoh:

1. Suatu kotak berisi 40 suku cadang yang dapat diterima bila terdapat paling banyak 3 yang cacat. Jika diambil sampel sebanyak 5 kotak, berapa peluang terdapat 1 yang cacat dari sampel?

   Jawab :

   $N = 40; k = 3; n = 5; x = 1; N - k = 40 - 3 = 37;$

   $n - x = 5 - 1 = 4$

   $$p(X = 1) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{58}} = 0{,}3011$$

2. Suatu panitia 5 orang akan dipilih secara acak dari 3 Kimiawan dan 5 Fisikawan. Hitung distribusi peluang banyaknya Kimiawan dalam panitia tersebut?

   Jawab :

   Misal: p.a x menyatakan banyaknya Kimiawan dalam panitia.

$$p(X=0)=h(0;8,5,3)=\frac{\binom{3}{0}\binom{5}{5}}{\binom{8}{5}}=\frac{1}{56}$$
$$p(X=1)=h(1;8,5,3)=\frac{\binom{3}{1}\binom{5}{4}}{\binom{8}{5}}=\frac{15}{56}$$
$$p(X=2)=h(2;8,5,3)=\frac{\binom{3}{2}\binom{5}{3}}{\binom{8}{5}}=\frac{30}{56}$$
$$p(X=3)=h(3;8,5,3)=\frac{\binom{3}{3}\binom{5}{2}}{\binom{8}{5}}=\frac{10}{56}$$

Distribusi Hipergeometrik x dalam bentuk tabel

x	0	1	2	3
h(x;8,5,3)	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{30}{56}$	$\frac{10}{56}$

Distribusi peluangnya dapat dirumuskan:

$$h(0;8,5,3)=\frac{\binom{3}{x}\binom{5}{5-x}}{\binom{8}{5}};x=0,1,2,3$$

Jika n<<<N-> peluang tiap pengambilan hanya berubah sedikit. Jadi pada dasarnya percobaan adalah binomial

Maka Distribusi Hipergeometrik dapat dihampiri dengan menggunakan Distribusi Binomial dengan $p=\frac{k}{N}$

$$(x\sim h(x)\approx b(x;n,p) dengan\ p=\frac{k}{N}$$
$$rataan\ E(x) dapat\ dihampiri\ dengan:\ \mu=np=\frac{nk}{N}$$
$$variansi\ v(x) dapat\ dihampiri\ dengan:\sigma^2=npQ=\ n.\frac{k}{N}(1-\frac{k}{N})$$

Contoh:

1. Sebuah pabrik ban melaporkan bahwa dari pengiriman sebanyak 5000 ban ke suatu toko terdapat 1000 cacat. Jika seseorang membeli 10 ban ini secara acak dari toko tersebut, berapa peluang tepat 3 yang cacat?

Jawab :

$$N = 5000; n = 10; k = 1000 \;\; n \lll N \rightarrow p = \frac{k}{N} = \frac{1000}{5000} \rightarrow dengan\ d.Binom$$

$$x \sim h(x) = b(x; n, p) = b(3; 10{,}0{,}2) = \binom{10}{3}(0{,}2)^3(0{,}8)^7$$

$$= \sum_{x=0}^{3} b(x; 10, 0{,}2) - \sum_{x=0}^{2} b(x; 10, 0{,}2) = 0{,}8791 - 0{,}6778 = 0{,}2013$$

7.3.4 Distribusi Binomial Negatif

Suatu percobaan disebut percobaan Binomial negatif jika memenuhi syarat:

1. Usaha diulangi sampai terjadi sejumlah sukses tertentu
2. Tiap usaha memberikan hasil yang dapat ditentukan saling sukses atau gagal
3. Peluang sukses yang dinyatakan dengan P, tidak berubah dari usaha yang satu ke usaha yang berikutnya
4. Tiap usaha bebas dengan usaha yang lainnya

Definisi:

Banyaknya usaha x untuk menghasilkan k sukses dalam percobaan Binomial Negatif disebut p.a Binomial Negatif

Distribusi peluang Binomial Negatif (fmp):

b* (x; k,p)=p(X=x)= $\binom{x-1}{k-1}p^k(1-p)^{x-k}$

Dimana X=k, k+1, k+2,...
b* (x; k,p) = Banyaknya usaha yang berakhir tepat pada sukses ke k
p =peluang sukses

$$sifat\ x\sim b^*(x;k,p):$$
$$E(x)=\frac{k}{p}$$

$$\sigma^2=\frac{k(1-p)}{p^2} \qquad (7.4)$$

Fungsi Pembangkit Momen (FPM) dari distribusi Binomial Negatif:

$$M(t)=p^k(1-(1-p)e^t)^{-k}\ untukt<-\ln(1-p))$$

Dengan FPM diatas buktikan:

$$E(x)=\frac{k}{p}\ dan\ \sigma^2=\frac{k(1-p)}{p^2}$$

1. Carilah peluang bahwa seorang yang melantunkan 3 uang logam sekaligus akan menghasilkan semuanya muka atau semuanya belakang untuk kedua kalinya pada lantunan ke lima.

   Jawab:
   Distribusi Binomial Negatif dengan $x=5;k=2;p=\frac{1}{4}$

$$b^*\left(5;2,\frac{1}{4}\right)=\binom{5-1}{2-1}(\frac{1}{4})^2(\frac{3}{4})^{5-2}=\binom{4}{1}(\frac{1}{4})^2(\frac{3}{4})^3=\frac{27}{256}$$

7.3.5 Distribusi Poisson

Distribusi Poisson merupakan distribusi probabilitas untuk variabel diskrit acak yang mempunyai nilai bilangan bulat mulai dari 0 dan seterusnya. Suatu bentuk dari distribusi ini adalah rumus pendekatan peluang Poisson untuk peluang Binomial yang dapat digunakan untuk pendekatan probabilitas Binomial dalam situasi tertentu.

Rumus Poisson dapat digunakan untuk menghitung probabilitas dari jumlah kedatangan, misalnya : probabilitas jumlah kedatangan nasabah pada suatu bank pada jam kantor. Distribusi Poisson ini digunakan untuk menghitung probabilitas menurut satuan waktu.

Misalnya banyaknya telepon per jam yang diterima suatu kantor, banyaknya kesalahan ketik perhalaman pada sebuah buku, banyakna kedatangan pesawat tiap unit waktu, banyaknya kendaraan lewat tiap jam dan lain sebagainya. Proses poisson memiliki ciri-ciri sebagai berikut:

1. Banyaknya sukses (mean) terjadi dalam selang waktu/ daerah tertentu, dimana antara selang waktu/ daerah yang satu dengan yang lain independen. Dalam hal ini, proses poison adalah proses yang tidak memiliki ingatan
2. Peluang sukses sebanding dengan panjang selang/ besar daerah dan tidak tergantung pada banyak sukses diluar selang waktu/ daerah tersebut.
3. Peluang terjadinya lebih dari satuukses dalam selang waktu/ daerah yang pendek diabaikan.

7.4 Latihan Soal :

1. Peluang seseorang calon pegawai lulus ujian masuk UII adalah 0,8. Bila 100 pegawai mengikuti seleksi ujiam masuk UII tentukan peluang bahwa ada 8 sampai 16 orang yang lulus ujian ?
2. Suatu ujian terdiri atas 40 pertanyaan pilihan berganda, masing-masing dengan empat kemungkinan jawaban dan hanya ada satu yang benar. Berapa peluang seseorang yang menjawab secara menebak-nebak saja memperoleh 5 sampai 10 jawaban yanag benar ?
3. Peluang seseorang sembuh dari flu karena minum obat herbal adalah 0,7. Bila 30 orang diketahui menderita flu berapa peluang bahwa terdapat 5 yang sembuh.
4. Bila dua dadu dilemparkan 6 kali, berapa peluang mendapatkan jumlah bilangan yang muncul sebesar 7 atau 11 sebanyak dua kali, bilangan yang sama pada kedua dadu sekali, dan kemungkinan lainnya tiga kali
5. Hitunglah peluang mendapatkan dua bilangan 1, satu bilangan 2, satu bilangan 3, dua bilangan 4, tiga bilangan 5, dan satu bilangan 6, bila sebuah dadu setimbang dilemparkan 10 kali.
6. Menurut teori genetika, suatu persilangan kelinci percobaan akan menghasilkan keturunan warna merah, hitam, dan putih dalam perbandingan 8:4:4. Hitunglah peluang bahwa di antara 8 keturunan semacam ini ada 5 yang berwarna merah, 2 hitam, dan 1 putih
7. Seorang mahasiswa mengamati kendaraan yang lewat pada salah satu jalan di kota Bogor. Hasil pengamatan menunjukkan bahwa 75% kendaraan yang lewat berasal dari DKI Jakarta.

Berapa peluang bahwa paling sedikit tiga dari lima mobil yang lewat jalan yang sama pada waktu mendatang berasal dari luar DKI?

8. Seorang penderita penyakit langka mempunyai peluang 0,4 untuk sembuh. Bila diketahui ada 15 orang yang telah mengidap penyakit tersebut, berapakah peluang: (1) paling sedikit 10 akan sembuh; (2) antara 2 sampai 8 yang sembuh; dan (3) tepat 5 yang sembuh.
9. Bila dua dadu dilantunkan 6 kali, berapakah peluang muncul jumlah 7 atau 11 dua kali, sepasang bilangan yang sama satu kali dan pasangan yang lain tiga kali?
10. Menurut teori genetika, persilangan tertentu dari sejenis marmut akan menghasilkan keturunan berwarna merah, hitam dan putih dalam perbandingan 8 : 4 : 4. Hitung peluang bahwa 5 dari 8 turunan akan berwarna merah, 2 hitam dan 1 putih.
11. Dalam suatu penelitian inventori diketahui bahwa permintaan rata-rata dari gudang terhadap suatu bahan tertentu 5 kali sehari. Berapa peluang pada suatu hari tertentu bahan tersebut:
    a. Diminta lebih dari 5 kali?
    b. Tidak diminta sama sekali?

8 Distribusi Probabilitas Kontinyu

Capaian Pembelajaran

Mahasiswa mampu menjelaskan sifat kurva normal, luas daerah, dan hampiran normal terhadap binomial.

8.1 Variabel Acak Kontinu

Variabel acak kontinu adalah variabel acak yang memiliki nilai bersambung dalam sebuah interval data. Olehkarenanya, variabel acak kontinu dapat memiliki nilai berupa bilangan bulat maupun desimal. Varibel acak kontinu jika digambarkan pada sebuah garis interval, akan berupa sederetan titik yang bersambung membantuk suatu garis lurus.

Contoh :

1. Usia penduduk suatu daerah.
2. Panjang beberapa helai kain.
3. Berat badan pasien pada rumah sakit.
4. Masa studi mahasiswa (tahun) di perguruan tinggi.

Sifat distribusi probabilitas variabel acak kontinu dinyatakan dengan fungsi f(x) dan sering disebut sebagai fungsi kepadatan atau fungsi kepadatan probabilitas dan bukan fungsi probabilitas. Nilai f(x) bisa lebih besar dari 1.

Fungsi kepadatan probabilitas harus memenuhi syarat sebagai berikut.

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$ (integral seluruh fungsi kepadatan probabilitas $f(x) = 1$)
3. $P(a < X < b) = \int_{a}^{b} f(x)dx$

   Catatan : $f(x)\, dx = P\{x \leq X \leq (x + dx)\}$, yaitu probabilitas bahwa nilai X terletak pada interval x dan x + dx.

Contoh :

Diketahui *f* (*x*) adalah fungsi distribusi peluang kontinu

a. $f(x)=1/4$, dimana $0 < X < 4$

b. tunjukkan f (x) adalah fungsi peluang!

c. tentukan fungsi peluang kumulatifnya

Penyelesaian:

f(*x*) adalah fungsi peluang tampak bahwa $0 \leq f(x) \leq 1$

$$\int_{x} f(x)dx = 1$$

$$\int_{x} f(x)dx = \int_{0}^{4} f(1/4)dx = 1$$

b . Fungsi distribusi peluang kumulatif

8.2 Fungsi Probabilitas Kumulatif Variabel Acak Kontinu

Kalau pada variabel acak diskrit, fungsi probabilitas kumulatif dihitung dengan cara penjumlahan maka pada variabel acak kontinu, probabilitas kumulatif dicari dengan integral.

Rumusnya adalah sebagai berikut.

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$$

Nilai-nilai dalam rumus ini harus kontinu atau dalam suatu interval.

8.3 Distribusi Normal

Distribusi Normal adalah model distribusi kontinyu yang jika digambarkan dalam sebuah grafik akan berbentuk seperti lonceng terbalik yang simetris. Dua parameter yang menentukan distribusi normal adalah mean (μ) dan standar deviasi (σ).

Fungsi kerapatan probabilitas dari distribusi normal diberikan dalam rumus berikut:

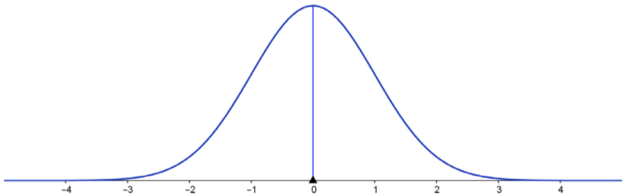
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{8.1}$$

Dimana μ adalah rata-rata, σ adalah standar deviasi dan π = 3,14159…

e = 2,71828..

Sifat- sifat penting distribusi normal adalah sebagai berikut :

1. Grafiknya selalu berada di atas sumbu x
2. Bentuknya simetris pada x = μ
3. Mempunyai satu buah modus, yaitu pada x = μ
4. Luas grafiknya sama dengan satu unit persegi, dengan rincian
    a) Kira-kira 68% luasnya berada di antara daerah $\mu - \sigma$ dan $\mu + \sigma$
    b) Kira-kira 95% luasnya berada di antara daerah $\mu - 2\sigma$ dan $\mu + 2\sigma$
    c) Kira-kira 99% luasnya berada di antara daerah $\mu - 3\sigma$ dan $\mu + 3\sigma$

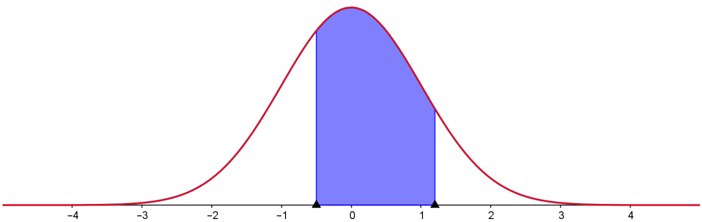


Gambar 8.1. Kurva distribusi normal baku

dengan mean 0 dan Variasi 1 atau ditulis $Z \sim N(0,1)$
bentuk transformasinya adalah :

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1) \qquad (8.2)$$

luas dibawah kurva normal :



Gambar 8.2. Luas kurva normal

Pada X = x, maka z = (x – μ)/σ atau pada X = x_1 , variabel acak Z akan berharga:

$$z_1 = \frac{x_1 - \mu}{\sigma} \tag{8.2}$$

Pada X = x_2 , variabel acak Z akan berharga:

$$z_2 = \frac{x_2 - \mu}{\sigma} \tag{8.3}$$

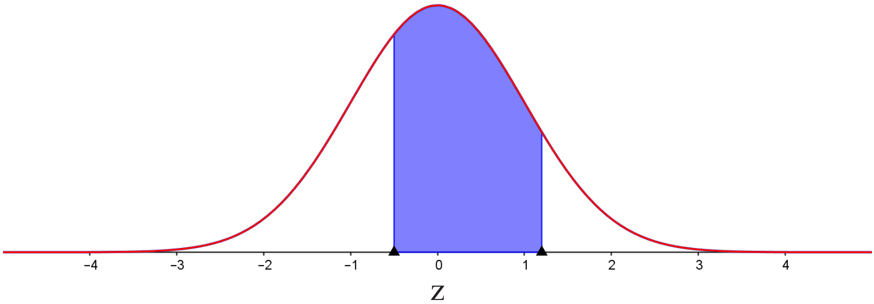
Sehingga fungsi peluang distribusi normal standar dapat ditulis: Selain dapat dihitung dengan rumus, terdapat juga Tabel Distribusi Normal yang akan mempermudah dalam perhitungan kurva normal.. Tabel distribusi normal baku disebut juga dengan Tabel Z dan dapat digunakan untuk mencari peluang di bawah kurva normal secara umum, asal saja nilai μ dan σ diketahui. Sebagai catatan nilai μ dan σ dapat diganti masing-masing dengan nilai $\bar{x}$ dan S

Contoh :

Diketahui suatu distribusi normal dengan $\mu = 50$ dan $\sigma = 10$, hitung peluang bahwa *X* mendapat harga antara 45 dan 62

$$z_1 = \frac{45 - 50}{10} = -0,5$$

$$z_2 = \frac{62 - 50}{10} = 1,2$$



Gambar 8.3. Kurva distribusi normal (contoh soal)

dengan menggunakan tabel z :

$$
\begin{aligned}
P(45 < X < 62) &= P(-0,5 < Z < 1,2) \\
&= P(Z < 1,2) - P(Z < -0,5) \\
&= 0,8849 - 0,3085 \\
&= 0,5764
\end{aligned}
$$

8.4 Distribusi Gamma

Distribusi Gamma digunakan untuk sekumpulan data yang memiliki variasi ukuran kemencengan atau kemiringan yang cukup signifikan. Mesikupun begitu, distribusi Gamma merupakan salah satu alternatif model yang banyak digunakan. Berikut adalah fungsi Gamma.

1) Fungsi gamma r (α) adalah :

$$r(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \text{ , untuk } \alpha > 0 \qquad (8.4)$$

Sifat-sifat penting fungsi gamma adalah :

1. Untuk sebuah bilangan bulat positif n, $\Gamma(n) = (n-1)!$
2. Didefinisikan = $\Gamma(1/2) = \sqrt{\pi}$

2) Distribusi gamma

Peubah acak kontinu x berdistribusi gamma, dengan parameter α dan β, bila padatnya diberikan oleh :

$$f(x : \alpha, \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x \geq 0 \tag{8.4}$$

$$= 0 \text{ untuk x lainnya}$$

Bila $\alpha > 0$ dan $\beta > 0$

3) Distribusi Gamma Standard

Jika parameter skala sebuah distribusi gamma $\beta = 1$ diperoleh suatu distribusi gamma standar.

$$F_G = (x : \alpha) = P\,(X \leq x) = \int_0^x \frac{t^{\alpha-1} e^{-t}}{\Gamma(\alpha)} dt$$

(8.5)

$$P\,(X \leq x) = F_G\,(x ; \alpha, \beta) = F_G \left(\frac{x}{\beta} ; \alpha \right)$$

Contoh :

Variable acak kontinu x yang menyatakan ketahanan suatu bantalan peluru (dalam ribaun jam) yang diberi pembebanan dinamis pada suatu putaran kerja tertentu mengikuti suatu distribusi gamma dengan $\alpha = 8$ dan $\beta = 15$, Tentukan, probabilitas sebuah bantalan peluru dapat digunakan selama 60 ribu-120 ribu jam dengan pembebanan dinamik pada putaran kerja tersebut!

Jawab :

$$\begin{aligned} P(60 \leq x \leq 120) &= P(x \leq 120) - P(x \leq 60) \\ &= F_G(120; 8, 15) - F_G(60; 8, 15) \\ &= F_G(120/15; 8) - F_G(60/15; 8) \\ &= F_G(8; 8) - F_G(4; 8) \\ &= 0{,}5470 - 0{,}0511 = 0{,}4959 \end{aligned}$$

8.5 Distribusi Eksponensial

Distribusi Gamma khususnya dengan $\alpha = 1$ disebut distribusi eksponensial. Peubah acak kontinu x distribusi eksponensial dengan parameter β, bila fungsi padatnya diberikan oleh

1. $f_E(x; \frac{1}{\beta}) = \frac{1}{\beta} e^{-x/\beta}\ x \geq 0$

   $= 0$ untuk x lainnya

   Dengan $\beta > 0$

2. $F_E(x; \frac{1}{\beta}) = P(X \leq x) = \frac{1}{\beta}\int_0^x e^{-x/\beta} dx = 1 - e^{-x/\beta}$

Contoh :

Misalkan x adalah waktu (response time) suatu terminal komputer on-line yang merupakan tenggang waktu antara masuknya suatu permintaan dari pengguna sampai sistem mulai memberikan tanggapan atas permintaan tersebut, memiliki suatu distribusi eksponensial dengan waktu tanggap rata-rata 5 detik. Jika seseorang perintah tersebut akan dijalankan selambat-lambatnya setelah 10 detik.

$$P(X \leq 10) = F(10; \frac{1}{5}) = 1 - e^{-10/0,2} = 1 - e^{-2} = 1 - 0{,}135 = 0{,}865$$

8.6 Latihan Soal :

1. Nilai rata-rata dalam suatu ujian metode statistika adalah 74 dan variasi 49. Jika 12% dari peserta mendapatkan nilai A dan ujian tersebut mengikuti distribusi normal , berapakah nilai terkecil yang mendapat nilai A?
2. Berat bayi yang baru lahir rata-rata 3.750 gram dengan simpangan baku 325 gram. Jika berat bayi berdistribusi normal, mak tentukanlah:
   a. Berapa persen yang beratnya lebih dari 4.500 gram?
   b. Berapa bayi yang beratnya 3.500 gram dan 4.500 gram, jika semuanya ada 10.000 bayi?
   c. Berapa bayi yang beratnya lebih kecil atau sama dengan 4.000 gram jika semuanya ada 10.000 bayi?
   d. Berapa bayi yang beratnya 4.250 gram apabila semuanya ada 5.000 bayi?
   e. Berapa persen bayi yang beratnya 3500 gram?
   f. Berapa persen bayi yang memiliki berat 3.250 dan 4.250
3. Dari penelitian terhadap 150 orang laki-laki yang berumur 40–60 tahun didapatkan rata-rata kadar kolesterol (μ) mereka 215 mg % dan simpangan baku $\sigma = 45$ mg %. Hitunglah peluang kita mendapatkan seorang yang kadar kolesterolnya:
   a. < 200 mg %
   b. > 250 mg %
   c. antara 200 –275 mg %

9 Hipotesis dan Uji Hipotesis

Capaian Pembelajaran

Mahasiswa mampu menjelaskan pengertian hipotesis, membuat hipotesis, dan pengujian hipotesis tentang rata-rata

9.1 Mengenal Uji Hipotesis

Hipotesis merupakan proposisi atau anggapan atau pernyataan yang mungkin benar atau salah tentang suatu populasi atau lebih, dan sering digunakan untuk pembuatan keputusan atau pemecahan permasalahan.Uji Hipotesis adalah suatu prosedur yang digunakan untuk menguji validitas hipotesis statistik suatu populasi, dengan menggunakan data dari sampel populasi. Uji Hipotesis adalah bagian terpenting dalam statistik inferensi, karena berdasarkan uji tersebut, suatu keputusan atau pemecahan permasalahan dari suatu penelitian dapat diselesaikan

Contoh :

a. Metode pembelajaran Student Center Learning lebih efektif dibandingkan metode pembelajaran ceramah dalam meningkatan pemahaman siswa.
b. Rata-rata IPK Mahasiswa Jurusan Teknik Informatika pada tahun Ajaran 2016/2017 adalah 3.00

berdasarkan rumusan hipotesisnya, uji hipotesis dapat dibedakan menjadi tiga jenis, yaitu :

a. Uji Hipotesis dua pihak

   $H_0 =$ dan $H_1 \neq$

b. Uji Hipotesis satu pihak kiri

   $H_0 =$ dan $H_1 <$

c. Uji hipotesis satu pihak kanan

   $H_0 =$ dan $H_1 >$

9.2 Taraf Signifikansi (α):

Data sampel yang diolah dengan statistika sangat dimungkinkan terjadi suatu kesalahan (error). Kesalahan dalam inferensi statistik dibagi menjadi 2 :

Tabel 9.1. Kesalahan dalam inferensi statistik

Keputusan	Keadaaan Alam	
	H_0 Benar	H_0 Salah
Menerima H_0	Keputusan Benar	Kesalahan Tipe II (β)
Menolak H_0	Kesalahan Tipe I (α)	Keputusan Benar

Keterangan :

1. **Kesalahan tipe I**: artinya kesalahan yang terjadi karena menolak H_0 padahal H_0 tersebut benar. Dalam kasus ini perumusan H_0 telah sesuai dengan teori (deduktif) dan kondisi faktual (induktif).
2. **Kesalahan tipe II**: artinya kesalahan yang terjadi karena menerima H_0 padahal H_0 tersebut salah.

Setiap pengujian hipotesis, harus ditentukan terlebih dahulu besarnya α = kesalahan tipe I yang dapat ditolerir, yang disebut **taraf signifikansi/keberartian** (*significant level*). Besarnya nilai α biasanya 1%, 5% dan 10%.

9.3 Pengujian Hipotesis

Untuk uji hipotesis, prosedur pengujianya adalah sebagai berikut:

a. merumuskan hipotesis
b. menentukan tingkat signifikan
c. menentukan nilai kritis
d. menghitung statistik Uji
e. membuat kesimpulan

Tahap 1 Merumuskan hipotesis

Hipotesis yang berupa anggapan atau pendapat dapat didasarkan atas:

1. Teori
2. Pengalaman sendiri atau pengalaman orang lain
3. Ketajaman berpikir, mempunyai pendapat tentang pemecahan suatu persoalan

Pada prosedur uji hipotesis terdapat dua pernyataan hipotesis, yaitu hipotesis nol, H0 dan Hipotesis alternatif , H1. Kedua hipotesis ini bersifat saling asing dan hanya salah satu dari keduanya yang benar atau yang salah.

Contoh :

Hipotesis Nol (Ho) dan Hipotesis alternatif (Ha) untuk hipotesis yang tidak mempunyai arah misalnya adalah sebagai berikut :

$$H_0 = 100 \text{ unit dan } H_1 \neq 100 \text{ unit}$$

Hipotesis Nol (Ho) dan Hipotesis alternatif (Ha) untuk hipotesis yang mempunyai arah misalnya adalah sebagai berikut :

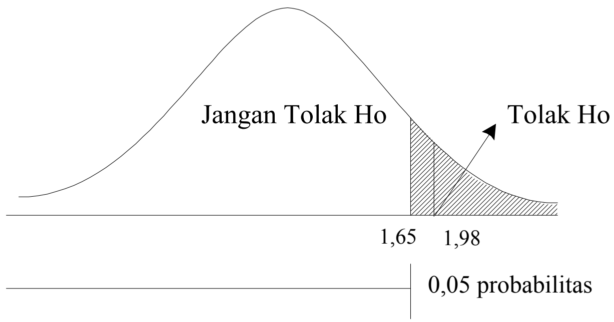
$$H_0 > 100 \text{ unit dan } H_1 < 100 \text{ unit}$$

Tahap 2 Menentukan tingkat signifikan

Tingkat signifikan menunjukkan besarnya interval keyakinan di kurva normal dan menunjukkan probabilitas keyakinan bahwa suatu -ilmu eksakta biasanya menggunakan tingkat signifikansi antara 98% (α 2%) sampai 99% (α 1%).

Tahap 3 Menentukan Nilai Kritis

Daerah Penolakan dan Penerimaan H_0



Gambar 9.1. Daerah penolakan dan penerimaan H_0

Titik Kritis

Titik kritis adalah titik yang membagi daerah di mana hipotesis *null* di terima atau hipotesis *null* di tolak.

Tahap 4 Menentukan uji statistik

Tahap ini dilakukan setelah menentukan tingkat signifikansi. Setelah menentukan tingkat signifikansi langkah selanjutnya adalah menentukan uji statistik yang akan digunakan. Hal ini karena masing-masing uji statistik memerlukan asumsi yang berbeda dalam penerapannya.

Tahap 5 Pengambilan keputusan

Tahap terakhir adalah pengambilan keputusan untuk menolak atau tidak menolak hipotesis *null*. Berdasarkan Gambar 9.1 diatas apabila Z hitung ditemukan sebesar 1,98 maka hipotesis *null* ditolak pada level kepercayaan 95%. Ho ditolak karena Z hitung berada pada daerah penolakan H_0 yaitu disebelah kanan nilai Z sebesar 1,65.

9.4 Uji Hipotesis Tentang Rata-Rata

Kriteria keputusan: jika **distribusi normal**, maka:

1. $H_0 : \mu \leq \mu_0$ $(Z_{hit} < Z_\alpha$, H_0 diterima)
   $Ha : \mu > \mu_0$
2. $H_0 : \mu \geq \mu_0$ $(Z_{hit} > Z_\alpha$, H_0 diterima)
   $Ha : \mu < \mu_0$
3. $H_0 : \mu = \mu_0$ $(-Z_{\alpha/2} < Z_{hit} < Z_{\alpha/2}$, H_0 diterima)
   $Ha : \mu \neq \mu_0$

Contoh :

Data Depsos menunjukkan bahwa rata- rata penghasilan penjual koran di suatu kota tertentu Rp 7000 per hari, dengan alternatif lebih besar dari itu. Diketahui simpangan baku = Rp 1600. Untuk menguji informasi tersebut, dilakukan penyelidikan terhadap

256 orang penjual koran yang dipilih secara acak, ternyata rata2 penghasilan mereka sebesar Rp 7100. dengan menggunakan α = 5%, ujilah data informasi tersebut.

Jawab:

a) H_0 : $\mu \leq 7000$

b) Ha : $\mu > 7000$

c) $\alpha = 0{,}05$

d) Daerah kritis: $Z > Z_{tab}$

Dari tabel normal, untuk $\alpha = 0{,}05$ atau tingkat keyakinan = $1 - \alpha = 95\%$, diperoleh $Z_{tab} = 1{,}64$

$$Z_{hit}=1 \; Z_{tab}=1{,}64$$

Sehingga H_0 diterima, yang berarti rata-rata penghasilan penjual koran adalah Rp 7000 per hari.

9.5 Latihan Soal:

1. Apa yang disebut dengan kesalahan tipe I?
2. Apa yang disebut dengan kesalahan tipe II?
3. Sebutkan Jenis – Jenis Hipotesis?
4. Sebutkan langah-langkah dalam uji hipotesis?
5. Sebuah perusahan mengadakan pelatihan teknik penjualan. Untuk keperluan ini diambil sampel karyawan sebanyak 12 orang dengan metode biasa dan 10 orang dengan terprogram. Pada akhir pelatihan di berikan evaluasi dengan materi yang sama. Kelas pertama mencapai nilai rata-rata 75 dengan simpangan baku 4,5. Ujilah hipotesis kedua metode pelatihan, dengan alternative keduanya tidak sama! Gunakan taraf nyata 10%! Asumsikan kedua populasi menghampiri distribusi normal dengan varians yang sama!

10 Regresi

Capaian Pembelajaran

Mahasiswa mampu menjelaskan pengertian regresi serta menghitung regresi dengan metode kuadrat terkecil dan kesalahan baku taksiran.

10.1 Definisi regresi

Analisis regresi digunakan untuk menunjukkan tingkat hubungan antara dua buah variabel atau lebih berdasarkan data yang tersedia dari kedua variabel tersebut. Data tersebut dapat berasal dari masa lalu untuk memprediksi pola data di masa depan atau data yang ada saat ini untuk memprediksi pola data di masa lalu. Di dalam regresi, hubungan antara kedua variabel bersifat sebab akibat.

Dengan minimal ada dua variabel, maka satu variabel berperan sebagai variabel independen atau bebas dan satu variabel lagi berperan sebagai variabel dependen atau tak bebas. Variabel independen adalah sebab yang mempengaruhi hasil dari variabel dependen (akibat). Pada umumnya analisis regresi digunakan untuk meramalkan nilai variabel di masa yang akan datang berdasarkan pola hubungan yang terjadi pada variabel tersebut di masa sebelumnya.

Analisis regresi menggunakan persamaan garis untuk menentukan pola hubungan suatu data. Persamaan garis ini berdasarkan pada metode kuadrat terkecil. Persamaan garis regresi linier sederhana adalah sebagai berikut.

$$Y = a + bX \tag{10.1}$$

Untuk mencari a dapat menggunakan rumus berikut ini.

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n \sum X^2 - (\sum X)^2} \tag{10.2}$$

Untuk mencari b dapat menggunakan rumus berikut ini.

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2} \tag{10.3}$$

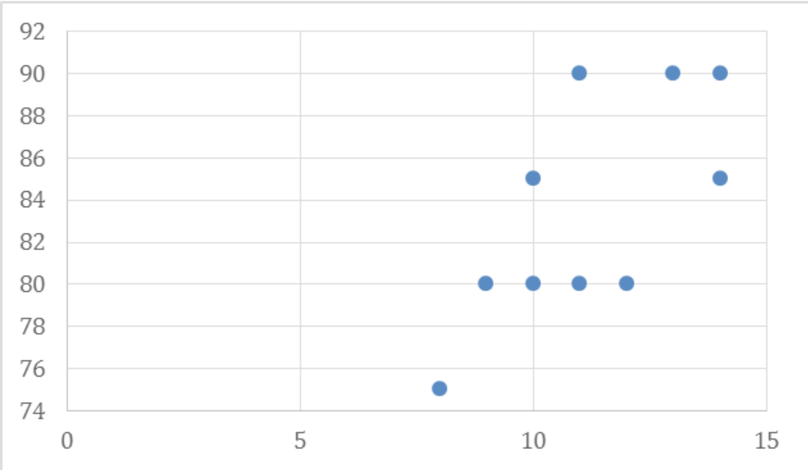
Contoh perhitungan analisis regresi.
Tabel berikut ini memaparkan tentang data jumlah presensi mahasiswa dengan nilai akhir yang didapatkannya.

Tabel 10.1. Contoh regresi

No.	Jumlah Presensi (X)	Nilai Akhir (Y)
1	10	80
2	11	80
3	13	90
4	14	90
5	9	80
6	14	85
7	8	75
8	12	80
9	11	90
10	10	85

Buatlah persamaan regresi dari tabel tersebut untuk mencaritahu pola hubungan antara jumlah presensi mahasiswa dengan nilai akhirnya. Variabel X adalah variabel independen. Variabel Y adalah variabel dependen.

Sebelum menghitung persamaan garis regresi, ada baiknya digambarkan dulu diagram pencar untuk melihat pola data awal.



Gambar 10.1. Contoh diagram pencar

Persamaan regresi: $Y = a + bX$

Mencari nilai a (dengan metode kuadrat terkecil):

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n \sum X^2 - (\sum X)^2}$$

$$a = \frac{(835)(1292) - (112)(9420)}{10(1292) - (112)^2}$$

$$a = 63{,}24$$

Mencari nilai b:

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

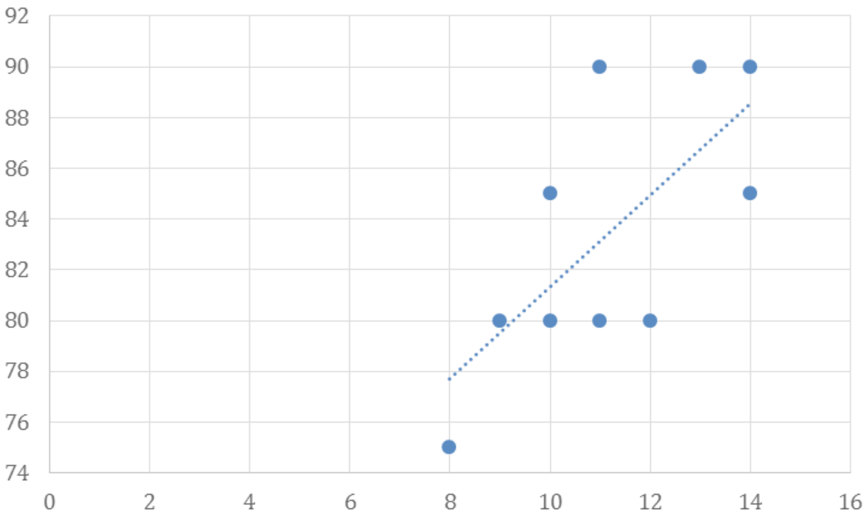
$$b = \frac{(10 \ x \ 9420) - (112 \ x \ 835)}{(10 \ x \ 1292) - \ (112)^2}$$

$$b = 1{,}8$$

Berdasarkan perhitungan tersebut maka persamaan garis regresinya adalah sebagai berikut.

$Y = 63{,}24 + 1{,}8X$

Jika digambarkan ke dalam diagram garis maka hasilnya sebagai berikut.



Gambar 10.2. Diagram pencar dengan garis regresi

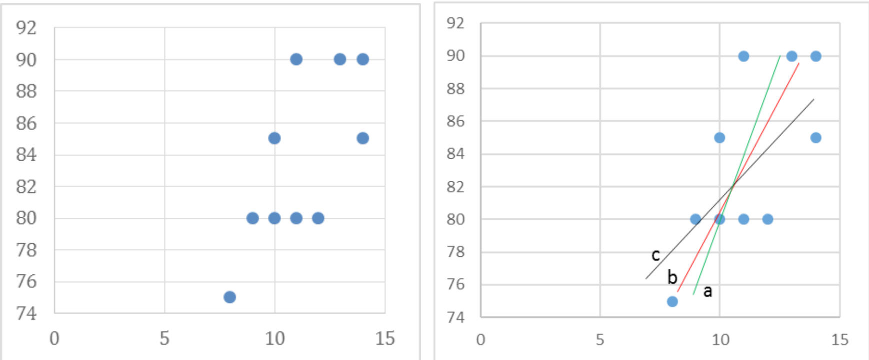
Dengan begitu, dapat diperkirakan jika jumlah presensi (variabel X) nilainya 6 maka nilai akhir yang akan diperoleh dapat dihitung dengan rumus sebagai berikut.

$$Y = 63{,}24 + 1{,}8(X)$$
$$Y = 63{,}24 + 1{,}8(6)$$
$$Y = 74{,}4$$

Perlu dicatat bahwa hasil perhitungan dengan analisis regresi hanya perkiraan dan mendekati kebenaran. Nilai perkiraan ini dapat bernilai kurang dari nilai sebenarnya (*underestimate*) atau melebihi nilai sebenarnya (*overestimate*)

10.2 Metode Kuadrat Terkecil

Garis regresi pada dasarnya adalah garis yang dianggap mewakili sekumpulan data. Garis ini dapat digambar secara manual (bebas) sesuai dengan penafsiran orang yang melihat data. Sebagai contoh terdapat diagram pencar dan garis regresi pada dua gambar berikut ini.



(a) (b)

Gambar 10.3. Diagram pencar dengan garis regresi

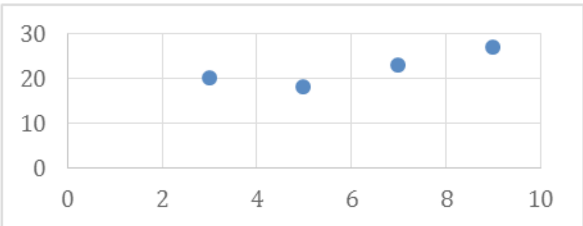
Gambar (a) memperlihatkan pola data terpencar. Berdasarkan data ini dapat ditarik garis yang berbeda-beda sesuai dengan penafsiran orang yang melihat data. Pada gambar (b) terdapat tiga garis yang berbeda. Metode ini memiliki kekurangan, yaitu tidak dapat menghasilkan perhitungan yang tepat dan berimbang untuk data yang tersedia maupun untuk data yang belum tersedia (estimasi).

Metode kuadrat terkecil adalah metode untuk mencari garis regresi dengan menghitung jumlah jarak terkecil antara titik-titik dalam diagram pencar dengan garis regresinya. Tujuannya adalah mendapatkan data estimasi yang akurat dan mendekati keadaan sebenarnya. Sebagai contoh dapat dilihat perbandingan gambar di bawah ini.

Terdapat data di bawah ini beserta diagram pencarnya.

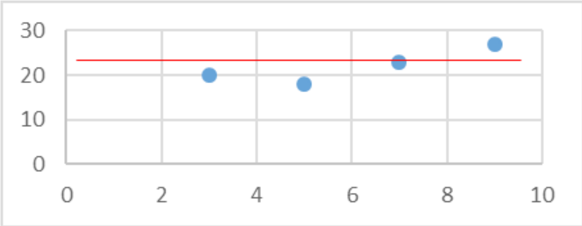
Tabel 10.2. Data sampel untuk metode kuadrat terkecil

y	data
3	20
5	18
7	23
9	27



Gambar 10.4. Diagram pencar untuk Tabel 9.2

Jika digambarkan garis regresi secara manual maka hasil terlihat seperti Gambar 10.5 di bawah ini (misalkan garis lurus dengan persamaan y=23):

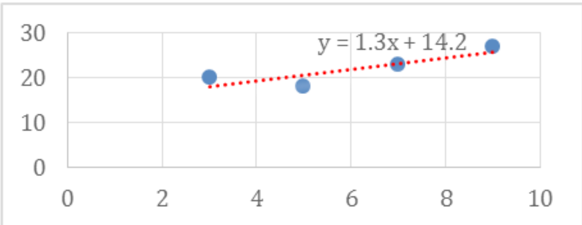


Gambar 10.5. Diagram pencar untuk Tabel 9.3

Tabel 10.3. Data sampel untuk metode kuadrat terkecil (metode bebas)

y	data	regresi	Jarak data dengan garis regresi
3	20	23	3
5	18	23	5
7	23	23	0
9	27	23	4
Jumlah jarak			12

Jika digambarkan garis regresi dengan metode kuadrat terkecil maka hasilnya dapat terlihat pada Gambar 10.6 di bawah ini (beserta dengan persamaan regresi yang dihasilkan):



Gambar 10.6. Diagram pencar untuk Tabel 9.4

Tabel 10.4. Data sampel untuk metode kuadrat terkecil

y	data	regresi	Jarak data dengan garis regresi
3	20	18.1	1.9
5	18	20.7	2.7
7	23	23.3	0.3
9	27	25.9	1.1
Jumlah jarak			6

Dari dua perhitungan di atas maka dapat dilihat bahwa perhitungan dengan metode kuadrat terkecil menghasilkan jumlah jarak yang lebih kecil daripada garis dengan cara manual. Selain itu, metode kuadrat terkecil juga menghasilkan persamaan yang lebih objektif.

Jarak data dengan garis regresi disebut dengan residual, yaitu selisih antara data prediksi dengan data sebenarnya. Jika data prediksi adalah y data sebenarnya dan ŷ adalah data prediksi, maka residual (e_i) dapat dihitung dengan rumus sebagai berikut.

$$e_i = y - \hat{y} \qquad (10.4)$$

10.3 Latihan soal

1. Tabel berikut ini menunjukkan kecepatan processor (variabel 1) dan daya tahan baterai (variabel 2) dari 10 laptop. Carilah persamaan regresi untuk dua variabel tersebut.

No.	Kecepatan Processor	Daya Tahan Baterai
1	2,5 GHz	4 Jam
2	2,3 GHz	3 Jam
3	2,2 GHz	3,5 Jam
4	2,6 GHz	5 Jam

5	3,0 GHz	2,3 Jam
6	2,1 GHz	7 Jam
7	2,4 GHz	4,5 Jam
8	3,1 GHz	3,1 Jam
9	2,0 GHz	6,8 Jam
10	2,9 GHz	3,5 Jam

2. Gambarkan diagram pencar dan garis regresi pada data Soal No. 1.
3. Dengan menggunakan data persamaan regresi yang ditemukan pada Soal No. 1, hitunglah perkiraan daya tahan baterai untuk kecepatan processor 2,8 GHz.

11 Korelasi

Capaian Pembelajaran

Mahasiswa mampu menjelaskan pengertian korelasi serta menghitung koefisien korelasi dan analisis korelasi.

11.1 Definisi Korelasi

Korelasi adalah metode statistik untuk mengetahui keberadaan hubungan antar variabel (Bluman, 1995). Terdapat perbedaan antara korelasi dan regresi, yaitu korelasi menunjukkan tingkat hubungan antar variabel, sedangkan regresi menyatakan hubungan sebab akibat antar variabel. Dengan analisis regresi dapat diketahui (estimasi) pola data di masa yang akan datang. Sementara dengan analisis korelasi hal itu tidak dapat dilakukan (Lungan, 2006). Korelasi menjadi pelengkap informasi bagi persamaan yang dihasilkan dari analisis regresi. Bagian berikut ini akan menjelaskan dan memberikan contoh tentang analisis regresi dan korelasi.

Analisis korelasi digunakan untuk menunjukkan tingkatan dan arah hubungan antara dua variabel (Lind, Marchal, & Wathen, 2006). Sebagai contoh, jika sebuah komputer dengan kapasitas penyimpanan 1 TB dihargai Rp. 10.000.000,00 sedangkan komputer dengan kapasitas penyimpanan 500 GB dihargai Rp. 5.000.000,00 maka terdapat korelasi antara kapasitas penyimpanan dengan harga komputer atau dengan kata lain, kapasitas penyimpanan mempengaruhi harga komputer. Contoh yang lainnya, jarak laptop dengan access point mempengaruhi kekuatan sinyal internet yang didapatkan oleh laptop tersebut.

Terdapat tiga macam kondisi korelasi, yaitu korelasi positif, korelasi negatif, dan tidak ada korelasi. Misalkan ada dua variabel, X dan Y. Keduanya memiliki korelasi atau mempengaruhi satu sama lain. Jika nilai X meningkat dan Y juga meningkat maka dapat dikatakan X dan Y memiliki korelasi positif. Namun jika nilai X meningkat dan Y menurun maka dapat dikatakan X dan Y memiliki korelasi negatif. Jika tidak terdapat pola hubungan nilai di antara keduanya maka dapat dikatakan X dan Y tidak memiliki korelasi sama sekali.

Rumus dasar untuk korelasi (Pearson) adalah sebagai berikut.

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}} \qquad (11.1)$$

Dengan $-1 \leq r \leq 1$

Jika r mendekati -1 maka kedua variabel memiliki korelasi negatif. Sedangkan jika r mendekati +1 maka kedua variabel memiliki korelasi positif.

Contoh:

Sebuah penelitian ingin menyelidiki hubungan antara banyaknya SKS yang diambil oleh seorang mahasiswa dengan IPK yang diraihnya dalam satu semester. Oleh karena itu, diambil 10 sampel mahasiswa dengan data sebagai berikut.

Tabel 11.1. Tabel contoh korelasi

Mahasiswa	Variabel 1 (SKS diambil)	Variabel 2 (IPK Semester)
A	20	3.1
B	18	4.0
C	15	2.8
D	20	4.0
E	10	3.0
F	12	3.6
G	16	4.0
H	14	3.2
I	18	3.5
J	12	4.0

Dengan rumus korelasi, maka dapat dicari tingkat hubungan antara dua variabel tersebut.

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

$$r = \frac{(10\ x\ 549) - (155\ x\ 35{,}2)}{\sqrt{(10\ x\ 2513) - 24025}\sqrt{(10\ x\ 125{,}9) - 1239{,}04}}$$

$$r = 0{,}23$$

Dari hasil perhitungan korelasi didapatkan koefisien korelasi sebesar 0,23. Hal ini berarti terdapat korelasi positif antara jumlah SKS yang diambil dengan besaran IPK yang didapatkan oleh seorang mahasiswa dalam satu semester. Meskipun begitu, korelasi atau hubungan di antara dua variabel ini tidak terlalu kuat karena belum mendekati angka 1.

11.2 Latihan Soal

1. Tabel berikut ini menunjukkan banyaknya jumlah mahasiswa di suatu universitas dengan jumlah penjualan smartphone dalam satu tahun. Dengan analisis korelasi Pearson, tentukan adakah hubungan antara jumlah mahasiswa dengan jumlah penjualan smartphone.

No	Jumlah Mahasiswa	Jumlah Penjualan Smartphone
1	105	96
2	133	84
3	122	96
4	138	83
5	114	84
6	140	87
7	142	94
8	112	89
9	113	85
10	144	93
11	109	87
12	134	82

Referensi

Baron, Michael (2014). *Probabilities and Statistics for Computer Scientist*. Second Edition. USA: CRC Press Taylor & Francis Group.

Bluman, A. G. (1995). Elementary statistics. Melbourne: Brown.

Chatfield, C. (2003). The Analysis of Time Series (Sixth Edit). Chapman & Hall / CRC Press Company.

Hadi, S. (2015). Statistik. Pustaka Pelajar : Yogyakarta

Lind, D., Marchal, W., & Wathen, S. A. (2006). Basic Statistics for Business & Economics (5th ed.). New York, NY, USA: McGraw Hill.

Lind, D., Marchal, W., & Waite, C. A. (2012). Basic Statistics for Business and Economics (4th ed.). McGraw Hill Ryerson.

Lungan, R. (2006). Aplikasi Statistika dan Hitung Peluang. Graha Ilmu, Yogyakarta.

Monalisa, Siti dan Setia, Putri D (2016). Analisis Penerapan Sistem Informasi Pengolahan Data Statistik (SISR) Menggunakan Technology Acceptance Model (Studi Kasus: BKKBN Riau). *Jurnal Rekayasa dan Manajemen Sistem Informasi Vol 2 No 1*.

Purwanto (2011). *Statistika untuk Penelitian*. Yogyakarta: Pustaka Pelajar

Spiegel, M. R., Susila, I. N., & Gunawan, E. (1992). Teori dan Soal-soal Statistik Versi SI (Metrik). Erlangga: Jakarta.

Sudjana (1996). *Metoda Statistika*. Edisi ke-6. Bandung: Tarsito.

Qudratullah, M.F. (2014). Statistika terapan. Penerbit ANDI: Yogyakarta.

Usman, Husaini & Setiady Akbar, Purnomo. (2006). PENGANTAR STATISTIKA. Yogyakarta: BUMI AKSARA.

Saleh, Samsubar. (1998). STATISTIK DESKRIPTIP. Yogyakarta: UPP AMP YKPN.

Riduwan. (2010). Dasar-Dasar Statistika. Bandung : Alfabeta.

Rachman, M., dan Muchsin. (1996). Konsep dan Analisis Statistik. Semarang : CV.IKIP Semarang Press

Siregar, Syofian. (2010). Statistika Deskriptif untuk Penelitian Dilengkapi Perhitungan Manual dan Aplikasi SPSS Versi 17. Jakarta : Rajawali Pres.

Suharyadi, dan S. K. Purwanto. (2009). Statistika: Untuk Ekonomi dan Keuangan Modern, Edisi 2, Buku 1, Penerbit Salemba Empat, Jakarta.

Glosari

Korelasi	:	hubungan antara dua variabel
Korelasi positif	:	hubungan antara dua variabel dengan arah positif
Korelasi negatif	:	hubungan antara dua variabel dengan arah negatif
Variabel	:	obyek perbandingan
Metode kuadrat terkecil	:	perhitungan jarak terpendek antara data dengan garis regresi
Estimasi	:	perkiraan nilai data di masa yang akan datang
Regresi	:	mengukur hubungan antara dua variabel
Laspeyres	:	rumus untuk menghitung angka indeks
Paasche	:	rumus untuk menghitung angka indeks
Angka indeks	:	angka yang mewakili sekumpulan data dalam satu variabel
Angka indeks agregatif	:	angka yang mewakili sekumpulan data dari lebih dari satu variabel
Un-group	:	tidak berkelompok
Group	:	berkelompok
Mean	:	nilai rata-rata
Median	:	Nilai tenggah
Modus	:	nilai yang sering muncul
Interval	:	jarak
Kuartil	:	membagi data dalam 4 bagian
Desil	:	membagi data dalam 10 bagian
Persentil	:	membagi data dalam 100 bagian
Range	:	rentang penyebaran data

Varians : salah satu ukuran dispersi atau ukuran variasi

Standar Deviasi: analisis data untuk melihat penyimpangan data terhadap nilai rata-rata

Probabilitas : Peluang mengenai seberapa sering suatu peristiwa /kejadian tertentu akan terjadi

Union : Gabungan

Interseksi : Irisan

Komplemen : Kejadian di luar A tetapi masih di dalam S

Mutually exclusive : Kejadian saling pisah

Ruang sampel : Seluruh kemungkinan yang dapat terjadi dari suatu percobaan

Titik Sampel : Tiap hasil dalam ruang sampel

Permutasi : Suatu susunan yang dapat dibentuk dari suatu kumpulan benda yang diambil sebagian atau seluruhnya.

Kombinasi : Semua susunan yang mungkin terjadi yang terdiri dari r unsur yang berbeda yang diambil dari n unsur itu, tanpa memperhatikan urutannya.

Indeks