

DRAFT BUKU AJAR

STATISTIKA

Dr. Widodo

Laboratorium Statistika

Program Studi Agribisnis

Universitas Muhammadiyah Yogyakarta

BAB I

KONSEP DASAR STATISTIKA

A. Statistik dan Statistika

Statistik berasal dari kata latin modern *statisticum collegium* ("dewan negara") dan bahasa Italia *statista* ("negarawan" atau "politikus"). Pada awalnya statistik lebih berfungsi untuk melayani keperluan administrasi negara. Pada awalnya, negara (raja) memerlukan data jumlah penduduk untuk memberikan gambaran tentang jumlah tentara yang dapat direkrut, jumlah hewan piaraan dan hasil pertanian lain untuk memberi gambaran tentang upeti yang dapat diambil, jumlah pangan yang dihasilkan, dll.

Statistik (*statistic*), dapat diartikan sebagai sekumpulan data baik berupa bilangan/angka maupun non angka yang melukiskan sesuatu, baik yang telah tersusun dalam daftar yang teratur (tabel) ataupun grafik/diagram maupun yang belum. Misalnya statistik penduduk melukiskan jumlah penduduk berdasar usia penduduk, jenis kelamin, tingkat pendidikan, pekerjaan, dll. Statistik pertandingan sepakbola melukiskan jalannya pertandingan sepakbola, misalnya skor 4-4 menggambarkan terjadinya pertandingan yang sangat terbuka diantara kedua tim sepakbola yang berlaga.

Statistika (*statistics*) membahas mengenai metode yang digunakan untuk pengumpulan dan analisis data sehingga diperoleh informasi yang bermanfaat. Statistika juga merupakan pengetahuan yang berhubungan dengan cara-cara pengumpulan data, penyajian data, pengolahan atau penganalisaannya dan penarikan kesimpulan berdasarkan kumpulan data dan hasil analisis yang dilakukan. Berdasar pengertian tersebut, Statistika dapat dibedakan menjadi 2, yaitu :

1. Statistika Deskriptif, yaitu bagian Statistika yang membicarakan mengenai cara-cara pengumpulan data dan penyederhanaan data pengamatan yang diperoleh, serta melakukan penghitungan nilai pemusatan dan penyebaran data.
2. Statistika Induktif, yaitu bagian dari Statistika yang membicarakan semua aturan dan cara menganalisa data dan menarik kesimpulan. Statistika Induktif pada umumnya berkaitan dengan dua masalah utama, yaitu

pendugaan nilai parameter populasi dan pengujian hipotesis. Dari Statistika Induktif ini dapat diketahui besarnya peluang untuk memperoleh kesimpulan yang sama bila penelitian tersebut diulang, dan dapat memberikan tingkat kepercayaan tertentu terhadap suatu dugaan atas parameter populasi.

Berdasarkan metode yang digunakan, Statistika Induktif dapat dibedakan menjadi Statistika Parametrik dan Statistika non Parametrik. Statistika Parametrik adalah Statistika Inferensia yang mempertimbangkan nilai dari satu atau lebih parameter populasi dan sehubungan dengan kebutuhan penarikan kesimpulan, Statistika Parametrik membutuhkan data dengan skala pengukuran minimal interval. Penurunan prosedur dan penetapan teori pada Statistika Parametrik berpijak pada asumsi bahwa bentuk distribusi populasi yang biasanya diasumsikan berdistribusi normal.

Statistika non Parametrik adalah bagian dari Statistika Induktif yang tidak memperhatikan nilai dari satu atau lebih parameter populasi dan hanya membutuhkan data dengan skala pengukuran nominal atau ordinal. Prosedur dan teori pada Statistika non Parametrik tidak mensyaratkan terpenuhinya asumsi data yang berdistribusi normal.

B. Populasi dan Sampel

Populasi adalah keseluruhan elemen (satuan atau individu) yang menjadi perhatian dalam suatu penelitian yang karakteristiknya hendak dianalisa. Keseluruhan elemen populasi tersebut sering dinamakan sebagai satuan analisa. Elemen populasi dapat merupakan orang, rumah tangga, usahatani, industri rumah tangga, perusahaan. Selain itu elemen populasi dapat berupa tanaman, hasil produksi tanaman, hasil produksi industri, produk yang dihasilkan oleh mesin.

Saat melakukan penelitian, kita memberikan batasan terhadap populasi yang menjadi obyek penelitian. Batasan ini diperlukan agar penelitian kita lebih terfokus kepada obyek yang benar benar menjadi perhatian dalam penelitian. Misalnya, jika kita melakukan penelitian tentang pengeluaran rumah tangga pedesaan maka seluruh rumah tangga yang berada di wilayah pedesaan merupakan elemen yang menjadi perhatian penelitian kita. Kita dapat lebih fokus dengan memberikan batasan yang lebih khusus, misalnya

kita melakukan penelitian tentang pengeluaran rumah tangga petani di pedesaan, maka hanya rumah tangga petani yang menjadi perhatian penelitian kita. Batasan terhadap populasi ini dapat dilakukan lebih khusus lagi untuk mendapatkan data yang khusus terkait dengan penelitian kita.

Dalam penelitian, kita sering dihadapkan pada kekeliruan dalam penentuan populasi yang disebabkan karena kita menemui kesulitan untuk memisahkan elemen populasi. Kekeliruan ini sudah dianggap jamak oleh banyak peneliti, karena seringkali tidak mengakibatkan kesalahan dalam pengambilan data yang diperlukan. Dalam penelitian tentang pendapatan usahatani padi daerah pasang surut, maka elemen populasi penelitian ini adalah usahatani padi di daerah pasang surut yang dikelola oleh petani. Namun, seringkali dalam banyak laporan penelitian disebutkan elemen populasi dalam penelitian tersebut adalah petani padi di daerah pasang surut.

Banyaknya elemen dalam populasi yang diteliti dinamakan ukuran populasi. Berdasarkan ukuran populasi, maka dapat dibedakan menjadi populasi berhingga/terbatas dan populasi tak berhingga/tak terbatas. Populasi berhingga adalah populasi yang jumlah elemennya dapat diketahui dengan pasti, sedangkan populasi tak berhingga adalah populasi yang jumlah elemennya tidak dapat diketahui dengan pasti. “Petani padi di Kabupaten Bantul bulan Januari 2015” merupakan contoh populasi berhingga, karena pada prinsipnya kita dapat memperoleh informasi tentang elemen populasi tersebut, yaitu yang berasal dari pihak lain. Namun “petani padi” merupakan contoh populasi tak berhingga, karena ukuran populasi “petani padi” tidaklah dapat diketahui dengan pasti karena tidak adanya batasan yang khusus sebagaimana halnya pada contoh “Petani padi di Kabupaten Bantul bulan Januari 2015”. Konsumen restoran MasDono merupakan contoh lain populasi tak berhingga. Kita tidak dapat memperoleh informasi tentang jumlah konsumen baik dari restoran MasDono maupun menghitung sendiri, karena konsumen selalu berubah dari waktu ke waktu.

Sampel merupakan sebagian dari populasi yang diambil dengan metode tertentu sehingga dapat mewakili dan menggambarkan karakteristik populasinya. Oleh karenanya kesimpulan tentang populasi sering ditarik dari analisa sampel. Misalnya dari sepetak lahan cabai yang diamati ternyata

terserang penyakit layu, maka sering serta merta disimpulkan bahwa lahan cabai di hamparan lahan itu terserang penyakit layu. Namun untuk mendapatkan kesimpulan yang dapat diterima secara logis memerlukan sejumlah prosedur.

Satuan yang akan diteliti dalam sampel dinamakan dengan satuan sampel. Satuan sampel dapat sama dengan satuan analisa, namun dapat juga berbeda dengan satuan analisa. Contoh satuan sampel yang berbeda dengan satuan analisa adalah misalkan suatu penelitian tentang kesukaan anak terhadap minuman "SARIJO" mengambil keluarga sebagai satuan sampel, dan seluruh anak dalam keluarga terpilih sebagai satuan analisa. Penelitian lain tentang perilaku pembelian minuman "SARIJO" dapat mengambil keluarga sebagai satuan sampel dan keluarga itu pula sebagai satuan analisa, sehingga satuan sampel sama dengan satuan analisa.

Satuan sampel diambil dari kerangka sampel. Keseluruhan satuan sampel membentuk kerangka sampel. Kerangka sampel dapat berupa daftar keluarga dalam suatu wilayah (misalnya RT), daftar pengusaha bakso dalam suatu paguyuban, daftar petani dalam suatu kelompok tani. Dalam survei, kita mengambil data dari satuan sampel ini. Pengambilan satuan sampel dilakukan dengan metode tertentu.

Besaran yang menyatakan karakteristik populasi dinamakan dengan parameter, sedangkan besaran yang menyatakan karakteristik sampel dinamakan dengan statistik. Contohnya, μ adalah simbol bagi rata-rata populasi, sedangkan \bar{x} adalah simbol bagi rata-rata sampel. Kegiatan untuk mengambil data dari populasi dinamakan dengan sensus, sedangkan kegiatan untuk mengambil data dari sampel dinamakan dengan survei.

Survei dilakukan untuk menarik kesimpulan tentang populasi. Informasi tentang karakteristik yang diperoleh dari satuan sampel yang diambil melalui survei diharapkan sama dengan karakteristik populasi. Agar kesimpulan dari satuan sampel berlaku pada populasi, maka pengambilan sampel harus memenuhi syarat tertentu. Syarat utamanya adalah sampel harus menggambarkan populasinya, sampel harus mewakili populasi, sampel harus merupakan populasi dalam jumlah yang kecil. Kalau syarat tersebut tidak terpenuhi, maka kesimpulan mengenai populasi tidak dapat

dipertanggung jawabkan secara ilmiah, karena kesimpulan tersebut akan menyimpang.

Pengambilan data melaalui survei mempunyai keuntungan dan kelemahan. Keuntungan pengambilan data melalui survei bila dibandingkan dengan sensus:

- a. Memerlukan biaya yang lebih kecil. Oleh karena data diambil dari satuan sampel yang jumlahnya lebih kecil dibandingkan keseluruhan elemen populasi, maka biaya pengambilan data, pengolahan data lebih sedikit.
- b. Menghemat waktu dan tenaga. Konsekuensi lain dari jumlah data yang dikumpulkan yang lebih sedikit adalah data dapat segera dikumpulkan, diolah, dan disajikan sehingga dapat segera diketahui kesimpulan mengenai populasi.
- c. Pengambilan data lebih cermat. Beberapa jenis penelitian memerlukan data melalui wawancara yang lama dan padat, sehingga tidak mungkin dilakukan dengan cara sensus tetapi harus dengan survei terhadap sampel. Dengan sampel yang jumlahnya lebih kecil daripada populasi, memungkinkan perhatian peneliti lebih tertuju pada sejumlah hal yang memerlukan pencermatan secara mendalam.
- d. Memiliki ketepatan yang tinggi. Meskipun data diambil dari sampel yang jumlah lebih sedikit, namun kualitas data dan hasil penelitian dapat lebih baik dibandingkan dengan sensus. Pengolahan data yang lebih sedikit menyebabkan pekerjaan pengolahan data dapat diserahkan kepada orang yang benar benar ahli.
- e. Tidak bersifat merusak. Seorang peneliti melakukan penelitian dengan mengambil data tentang pertumbuhan tanaman pada akan mengambil sampel beberapa tanaman saja untuk diukur, karena pengukuran terhadap seluruh tanaman padi akan merusak pertanaman.

Dalam keadaan tertentu, kemanfaatan pengambilan data dengan survei dari satuan sampel menimbulkan keraguan. Keraguan ini timbul dalam beberapa keadaan, yaitu:

- a. Jika data diambil dari populasi yang ukurannya kecil, sehingga ukuran sampel hampir sama dengan ukuran populasi. Dalam hal demikian,

biaya, waktu dan tenaga yang dibutuhkan untuk survei terhadap sampel hampir sama dengan untuk sensus.

- b. Jika dalam suatu survei, pengambilan sampel harus mengeluarkan biaya yang besar yang disebabkan oleh pekerjaan pengambilan sampel, pengawasan dan sebagainya, maka survei dengan pengambilan sampel kurang praktis.

C. Data

Data (bentuk jamak dari datum) merupakan informasi yang diperoleh dari pengamatan. Data merupakan keterangan yang dapat memberikan gambaran tentang suatu keadaan atau suatu permasalahan. Data dapat berupa bilangan atau dapat juga berbentuk katagori, misalnya keluarga menengah, laki-laki, senang, bagus, puas, setuju, berhasil, gagal, efisien, dan sebagainya.

Data dapat diklasifikasikan berdasarkan:

- a. Sifatnya.

Berdasarkan sifatnya, data dapat dibedakan menjadi data kualitatif dan data kuantitatif.

- 1). Data kualitatif adalah data yang bersifat menggolongkan saja, hasilnya berupa katagori. Misalnya laki-perempuan, kelompok bawah- menengah-atas.

- 2). Data kuantitatif adalah data yang berbentuk angka, contohnya data tentang pendapatan petani per musim tanam dapat Rp 2.500.000 atau Rp. 3.000.000, atau bilangan lainnya; data tentang jumlah anggota keluarga petani dapat bernilai 2 orang, 3 orang, 4 orang, atau bilangan lainnya.

Data kuantitatif jika dilihat dari nilainya dapat digolongkan lagi menjadi dua golongan, yaitu data kontinyu dan data diskrit. Data kontinyu adalah data yang nilainya dapat menempati semua nilai bilangan, contohnya data tentang pendapatan petani per musim tanam. Pendapatan petani per musim tanam dapat bernilai 2.500.000 atau 2.500.001 atau 2.500.001,1 atau semua nilai lainnya. Walaupun secara teoritis data kontinyu dapat menempati semua nilai, namun dalam prakteknya diperlukan pengukuran yang rasional. Misalnya, angka di belakang koma pada pendapatan tidak

mempunyai makna yang berarti, maka data pendapatan tidak menggunakan angka di belakang koma.

Sedangkan, data diskrit adalah data yang nilainya hanya mempunyai nilai tertentu saja, contohnya data tentang jumlah anggota keluarga petani. Data tentang jumlah anggota keluarga petani, dapat bernilai 1 atau 2 atau 3 atau bilangan bulat lainnya. Data diskrit ini tidak dapat bernilai lain, selain bilangan bulat

b. Sumbernya.

Berdasarkan sumbernya, data dapat dikelompokkan menjadi data internal dan data eksternal.

1). Data internal adalah data yang menggambarkan keadaan diri sendiri, organisasi sendiri, dan sebagainya. Perusahaan perkebunan kelapa sawit ingin menggambarkan tentang keadaan karyawan buruh panen sawit. Perusahaan ini akan mengambil data tentang jumlah karyawan, produktivitas karyawan, gaji karyawan, umur karyawan, jarak tempuh ke perusahaan, waktu tempuh ke perusahaan, yang berasal dari perusahaan perkebunan tersebut.

2). Data eksternal adalah data yang menggambarkan tentang keadaan di luar diri sendiri. Jika mahasiswa ingin membuat tugas tentang keadaan karyawan suatu perusahaan perkebunan, maka dia akan mengambil data tentang jumlah karyawan, produktivitas karyawan, gaji karyawan, jarak tempuh ke perusahaan, waktu tempuh ke perusahaan, yang berasal dari perusahaan perkebunan (data ini bukan berasal dari mahasiswa tersebut).

Jadi data yang sama sumbernya yaitu dari perusahaan perkebunan, dapat menjadi data internal bagi perusahaan perkebunan itu sendiri, dan dapat menjadi sumber eksternal bagi mahasiswa. Namun jika mahasiswa tersebut merupakan karyawan perusahaan perkebunan yang diteliti, maka data yang sama tersebut merupakan data internal.

c. Cara memperolehnya.

Berdasarkan cara memperolehnya, data dikelompokkan menjadi data primer dan data sekunder.

Data primer yaitu data yang diperoleh dan diolah oleh peneliti langsung dari sumbernya. Sekali lagi contoh mahasiswa di atas. Untuk mendapatkan

data tentang jumlah karyawan, produktivitas karyawan, gaji karyawan, jarak tempuh ke perusahaan, waktu tempuh ke perusahaan, maka mahasiswa tersebut harus mengumpulkan dari setiap karyawan buruh panen kelapa sawit. Data yang dikumpulkan ini merupakan data primer. Pun demikian jika mahasiswa tersebut memperoleh data dengan cara menyalin dari basis data yang dimiliki perusahaan. Data demikian merupakan data primer.

Secara berkala, Biro Pusat Statistik (BPS) melakukan Survei Sosial Ekonomi Nasional (SUSENAS) yang menghasilkan basis data tentang berbagai aspek sosial dan ekonomi pada tingkat rumah tangga terpilih. Peneliti dapat memanfaatkan data yang dikumpulkan oleh BPS melalui SUSENAS tersebut untuk diolah dan disajikan menurut keperluan peneliti tersebut. Basis data tentang berbagai aspek sosial dan ekonomi ini juga merupakan contoh data primer yang telah dikumpulkan oleh BPS.

Data sekunder yaitu data yang bukan diperoleh peneliti secara langsung dari obyek penelitian tetapi dari data telah diolah dan disajikan oleh instansi tertentu. Pada umumnya instansi atau pihak tertentu menyajikan data untuk kepentingan tertentu yang tidak mesti sama dengan maksud peneliti. BPS dengan berbagai tingkatannya telah menerbitkan banyak data sekunder yang menggambarkan keadaan wilayah tertentu, seperti Geografi, Pemerintahan, Penduduk dan Tenaga Kerja, Sosial, Pertanian, Industri, Perhubungan dan Komunikasi, Keuangan dan Harga, Konsumsi, dan Pendapatan Regional. Kementrian dan Dinas juga menerbitkan data yang berkaitan dengan bidang kerjanya, bahkan asosiasi tertentu juga mengeluarkan data sekunder yang berkaitan dengan bidang perhatiannya.

d. Waktu pengumpulan.

Berdasarkan waktu pengumpulan, data dapat dikelompokkan menjadi data *cross section* yaitu data yang dikumpulkan pada rentang waktu tertentu; dan data *time series*.

Data cross section adalah data yang dikumpulkan dari survei atau secara sensus di suatu wilayah tertentu pada satu waktu yang sama. Penerimaan per hektar usahatani padi organik yang diperoleh oleh 3 orang petani di

Kabupaten Sragen adalah petani-1 Rp 48 juta, petani-2 Rp 47,5 juta, petani-3 Rp 49 juta.

Data time series adalah data yang dikumpulkan dari satu obyek penelitian selama beberapa kurun waktu. Penerimaan per hektar usahatani padi organik yang diperoleh seorang petani di Kabupaten Sragen pada 3 musim tanam tahun 2015 adalah musim tanam-1 Rp 46 juta, musim tanam-2 Rp 46,5 juta, musim tanam-3 Rp 48 juta.

D. Variabel

Variabel adalah sifat yang dimiliki oleh satuan analisa yang berbeda antara satu individu (kelompok individu) dengan individu (kelompok individu) lain. Suatu sifat mungkin dapat merupakan variabel bagi suatu populasi tertentu, namun bukan merupakan variabel bagi populasi lainnya. Misalnya jenis kelamin, merupakan variabel pada siswa sekolah umum, namun bukan variabel sekolah pesantren khusus putri.

Variabel yang dapat memiliki sebarang nilai diantara dua nilai yang diberikan dinamakan dengan variabel kontinyu, misalnya variabel pendapatan merupakan variabel kontinyu karena diantara nilai 1 (juta Rupiah) dengan 1,1 (juta Rupiah) masih mempunyai nilai 1,01; 1,02 (juta Rupiah) dslb. Variabel yang tidak dapat memiliki sebarang nilai diantara dua nilai yang diberikan dinamakan variabel diskrit, misalnya variabel ukuran keluarga merupakan variabel diskret karena diantara 1 (orang) dengan 2 (orang) tidak terdapat sebarang nilai.

E. Skala Pengukuran

1. Skala Nominal, digunakan untuk mengklasifikasikan obyek amatan atau kejadian dalam kelompok (katagori) yang terpisah untuk menunjukkan kesamaan atau perbedaan ciri-ciri tertentu dari obyek amatan. Kelompok (katagori) didefinisikan sebelumnya dan dilambangkan dengan kata-kata, huruf, simbol, atau angka yang sekedar sebagai lambang yang menunjukkan suatu obyek amatan termasuk dalam katagori tertentu. Misalnya Ruang F.5.3 menunjukkan ruang tersebut berada di gedung F, blok urutan ke 5, ruang urutan ke 3. Ruang F.5.2 menunjukkan ruang tersebut berada di Gedung F, blok urutan ke, 5 ruang urutan ke 2. Kedua data tentang ruang tersebut menunjukkan kesamaan, yaitu berada di

Gedung F dan blok urutan ke 5. Sedangkan yang membedakan adalah urutan ruangnya, yaitu ruang urutan ke 3 dan urutan ke 2. Kita tidak dapat memperoleh informasi lain, seperti apakah Ruang F.5.3 lebih luas dibandingkan Ruang F.5.2 atau sebaliknya berdasarkan informasi ruang tersebut. Nomor punggung pemain sepakbola menunjukkan perbedaan antar pemain belaka dan tidak memberi informasi lain, misalkan pemain dengan nomor punggung 1 tidak mesti nilai transfERNYA lebih tinggi dibandingkan pemain dengan nomor punggung 2, demikian juga sebaliknya. Jika mahasiswa laki-laki diberi kode 1 dan mahasiswa perempuan diberi kode 2, maka kode tersebut hanyalah sekedar pembeda belaka dan tidak menunjukkan bahwa 1 lebih mahir dibandingkan 2.

2. Skala Ordinal, digunakan untuk mengklasifikan obyek amatan atau kejadian dalam kelompok atau katagori yang terpisah dan menunjukkan adanya urutan peringkat. Data dengan skala ordinal tidak mempunyai angka 0 (nol) yang mutlak. Urutan peringkat dapat menaik atau menurun. Contohnya, kualitas A, B, dan C menunjukkan katagori yang berbeda dan kualitas A lebih baik dibandingkan kualitas B, dan kualitas B lebih kecil dibandingkan kualitas C. Sikap konsumen sering dikelompokkan menjadi sangat suka (5), suka (4), netral (3), tidak suka (2) atau sangat tidak suka (1). Pada contoh ini angka menunjukkan peringkat sikap, 5 lebih baik dibandingkan 4, 4 lebih baik dibandingkan 3, dan seterusnya. Selain dengan cara itu, sikap konsumen juga dapat diperingkatkan menjadi sangat suka (2), suka (1), netral (0), tidak suka (-1) atau sangat tidak suka (-2). Contoh terakhir berbeda dengan contoh sebelumnya, yaitu pada pemberian skor yang berbeda. Kedua pemberian skor boleh digunakan, karena variabel sikap konsumen dapat dikelompokkan sebagai variabel dengan skala pengukuran ordinal. Angka skor tetap menunjukkan peringkat sikap konsumen.
3. Skala Interval, suatu pemberian angka kepada kelompok obyek yang mempunyai sifat skala nominal, ordinal dan ditambah dengan satu sifat, yaitu terdapat jarak yang sama pada skala interval ini. Jarak ini menunjukkan perbedaan jarak yang sama dari sifat obyek yang diukur.

Contoh yang paling populer bagi variabel dengan skala interval adalah suhu. Tiga gelas berisi air dengan suhu, yaitu: gelas pertama diisi air dengan suhu 10°C , gelas kedua diisi air dengan suhu 20°C , sedangkan gelas ketiga diisi air dengan suhu 35°C . Sifat skala nominal ditunjukkan dari adanya perbedaan suhu air pada ketiga gelas tersebut. Sifat skala ordinal ditunjukkan dari suhu air pada gelas ketiga lebih tinggi dibandingkan gelas kedua, dan suhu air pada gelas kedua lebih tinggi dibandingkan gelas pertama. Perbedaan suhu air pada gelas ketiga dengan gelas kedua adalah sebesar 15°C , sedangkan perbedaan suhu air pada gelas kedua dan gelas pertama adalah sebesar 10°C . Sifat skala interval ditunjukkan dari perbedaan suhu air gelas ketiga dengan gelas kedua adalah 1,5 kali dari perbedaan suhu air antara gelas kedua dengan pertama. Hal ini juga berlaku jika suhu air pada ketiga gelas tersebut diukur dengan termometer Reamur atau Fahrenheit.

4. Skala rasio, adalah skala pengukuran yang mempunyai sifat dari ketiga skala pengukuran sebelumnya ditambah dengan satu sifat lagi, yaitu memberikan keterangan tentang nilai absolut obyek yang diukur. Sebagai contoh misalnya upah per jam kerja yang diperoleh pekerja pertama adalah Rp 10.000; sedangkan pekerja kedua adalah Rp 20.000; dan pekerja ketiga adalah Rp 35.000;. Sifat skala nominal ditunjukkan dari adanya perbedaan upah diantara ketiga pekerja tersebut. Sifat skala ordinal ditunjukkan dari upah pekerja ketiga lebih tinggi dibandingkan upah pekerja kedua, dan upah pekerja kedua lebih tinggi dibandingkan upah pekerja pertama. Sifat skala interval ditunjukkan dari perbedaan upah pekerja ketiga dengan pekerja kedua adalah 1,5 kali dibandingkan perbedaan upah pekerja kedua dengan pekerja pertama. Sifat skala rasio ditunjukkan dari upah pekerja ketiga adalah 3,5 kali lebih besar dibandingkan upah pekerja pertama, dan 1,75 kali lebih besar dibandingkan upah pekerja kedua. Apakah sifat skala interval ini juga berlaku pada Suhu pada contoh sebelumnya?

BAB II

DISTRIBUSI FREKUENSI

A. Data Mentah

Informasi yang diperoleh oleh peneliti yang berasal dari sebuah survei atau sensus akan dikumpulkan berdasarkan nomor urut dari obyek penelitian. Informasi yang telah disusun tersebut merupakan data mentah. Data mentah adalah data yang dikumpulkan namun belum diatur secara numerik.

Contoh :

Data berat hasil penimbangan yang dilakukan oleh mahasiswa terhadap 40 ekor ikan lele yang dijual di pasar disajikan pada tabel sebagai berikut:

Tabel 2.1. Berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

No Urut	Berat lele (gr)								
1	138	9	173	17	147	25	150	33	154
2	146	10	145	18	176	26	125	34	140
3	168	11	150	19	147	27	148	35	145
4	146	12	140	20	142	28	119	36	157
5	151	13	138	21	144	29	153	37	144
6	164	14	142	22	136	30	156	38	165
7	158	15	135	23	163	31	149	39	135
8	126	16	132	24	135	32	152	40	128

Dari sajian data pada Tabel 2.1. kita belum memperoleh suatu gambaran apapun tentang permasalahan yang diteliti, karena data mentah hanyalah pengumpulan data yang dituliskan dalam suatu lembaran, yang belum dianalisa.

B. Array

Sebuah Array adalah penyusunan data mentah secara numerik dalam urutan besaran yang menaik atau menurun. Dari tabel 2.1. kita array dengan cara mengurutkan data dari yang terkecil sampai yang terbesar. Kita lihat pada Tabel 2.2 bahwa berat lele telah disusun secara berurutan mulai dari berat yang terendah namun kolom Nomor Urut menjadi tidak beraturan.

Tabel 2.2. Berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

No Urut	Berat lele (gr)								
28	119	22	136	37	144	31	149	36	157
26	125	1	138	10	145	11	150	7	158
8	126	13	138	35	145	25	150	23	163
40	128	12	140	2	146	5	151	6	164
16	132	34	140	4	146	32	152	38	165
15	135	14	142	17	147	29	153	3	168
24	135	20	142	19	147	33	154	9	173
39	135	21	144	27	148	30	156	18	176

Array dapat memberikan informasi bagi peneliti, namun masih sangat terbatas. Dari array yang disajikan pada Tabel 2.2. kita sedikit dapat memperoleh informasi tentang berat ikan lele hasil penimbangan mahasiswa, misalnya: berat lele berkisar antara 119 gram sampai dengan 176 gram.

3. Distribusi Frekuensi

Apabila jumlah data pengamatan cukup banyak maka kita akan mengalami kesulitan untuk mendapatkan informasi berdasarkan data tersebut. Untuk mempermudah penggalan informasi, maka data harus terlebih dahulu melalui proses pengolahan sampai pada bentuk tampilan data yang dibutuhkan sesuai kepentingannya. Sebagian pengolahan dan penyajian data adalah dengan membuat distribusi frekuensi.

Distribusi frekuensi adalah suatu tabel distribusi yang menjelaskan frekuensi anggota populasi menurut beberapa kelompok tertentu. Kelompok kelompok tersebut dapat berupa angka atau dapat juga berupa kategori tertentu. Pada setiap kelompok tersebut selanjutnya ditentukan banyaknya individu amatan. Banyaknya individu amatan ini disebut sebagai frekuensi. Suatu tabel yang berisi susunan data yang terbagi ke dalam frekuensi disebut frekuensi atau tabel frekuensi.

Pembagian kelompok dalam distribusi frekuensi dapat berupa angka ataupun bukan angka. Distribusi frekuensi yang pembagian kelompoknya berupa angka dinamakan Distribusi Frekuensi Kuantitatif. Angka angka

kelompok pada distribusi frekuensi, dapat berupa angka tunggal atau angka dalam interval tertentu. Pengelompokan data dalam distribusi frekuensi dengan kelompok angka tunggal dapat dilakukan apabila banyaknya data yang dianalisa tidak terlalu besar.

Contoh distribusi frekuensi angka tunggal.

Tabel 2.3. Distribusi frekuensi banyaknya anggota keluarga petani jahe

Nomor	Kelompok banyaknya anggota keluarga	Frekuensi petani jahe (orang)
1	0	2
2	1	3
3	2	5
4	3	2
5	4	2
6	5	1

Distribusi frekuensi yang disajikan dalam Tabel 2.3. menunjukkan kelompok angka dinyatakan sebagai “Kelompok banyaknya anggota keluarga” dan frekuensi dinyatakan sebagai “Frekuensi petani jahe (orang)”. Tabel 2.3. dapat kita ketahui bahwa ada sebanyak 2 orang petani jahe yang anggota keluarganya sebanyak 0 orang (baca: tidak mempunyai anggota keluarga), ada sebanyak 3 orang petani jahe yang anggota keluarganya sebanyak 1 orang, dan seterusnya.

Apabila data cukup banyak, maka penyajian data dalam distribusi frekuensi angka tunggal tidak akan mendapatkan banyak manfaat, dan apabila didapatkan keragaman yang tinggi pada data maka kita akan mengalami kesulitan untuk menyajikannya, namun dengan bantuan program komputer akan lebih mudah dilakukan tetapi tetap kurang bermanfaat. Jika data pada Tabel 2.1. kita sajikan dalam distribusi frekuensi angka tunggal, maka kita akan mempunyai 20 buah kelompok berat lele, dan akan diperoleh banyak kelompok berat lele yang frekuensinya adalah 1.

Kelompok yang berupa angka dapat juga dinyatakan dalam interval tertentu. Misalnya Tabel 2.3. di atas kita sajikan dalam angka berkelompok.

Tabel 2.4. Distribusi frekuensi banyaknya anggota keluarga petani jahe

Nomor	Kelompok banyaknya anggota keluarga	Frekuensi petani jahe (orang)
1	0 – 1	5
2	2 – 3	7
3	4 – 5	3

Penentuan kelompok angka dapat dilakukan dengan mendasarkan pada konsep tertentu, agar data dapat disajikan sesuai dengan konsep tersebut. Misalnya kita menggunakan konsep keluarga kecil adalah keluarga yang jumlah anggota keluarganya 3 (yang berarti 4 jika ditambah dengan kepala keluarga), maka kita dapat menyajikan Tabel 2.3. menjadi distribusi frekuensi baru.

Tabel 2.5. Distribusi frekuensi banyaknya anggota keluarga petani jahe

Nomor	Kelompok banyaknya anggota keluarga	Frekuensi petani jahe (orang)
1	0 – 3	12
2	4 – 5	3

Pada Tabel 2.5. kita mendapatkan informasi bahwa didapatkan sebanyak 12 orang petani jahe yang keluarganya termasuk katagori keluarga kecil.

Penyajian data dalam suatu distribusi frekuensi, dapat menggunakan katagori sebagai dasar pembagian kelompoknya. Disitribusi yang pembagian kelompoknya bukan dinyatakan dalam angka, tetapi dalam bentuk katagori tertentu dinamakan Distribusi Frekuensi Kualitatif (Katagorik).

Contoh distribusi frekuensi katagorik adalah sebagai berikut

Tabel 2.6. Distribusi pencari kerja terdaftar menurut tingkat pendidikan di Gunung Kidul, 2014

Nomor	Tingkat pendidikan	Frekuensi pencari kerja (orang)
1	SD	22
2	SLTP	60

3	SMA	321
4	SMK	1.302
5	D1/D2/D3	189
6	S1	324
7	S2	1

Sumber: Gunung Kidul Dalam Angka 2015

4. Pembuatan distribusi frekuensi

Untuk membuat distribusi frekuensi, maka harus memahami beberapa pengertian yang berkaitan dengan distribusi frekuensi. Perhatikan contoh distribusi frekuensi berikut:

Tabel 2.7. Tingkat pendapatan konsumen pada warung kopi XXX

Nomor	Tingkat pendapatan (ribu Rupiah)	Frekuensi konsumen (orang)
1	500 – 999	15
2	1.000 – 1.999	9
3	2.000 – 2.999	9
4	3.000 – 3.999	2
5	4.000 – 4.999	3
6	5.000 – 5.999	2

Pengertian yang harus dipahami dalam distribusi frekuensi adalah sebagai berikut:

1. Lambang yang mendefinisikan kelompok kelas, yaitu 500 – 999; 1.000 – 1.999 dan seterusnya dinamakan dengan interval kelas
2. Bilangan yang terletak di ujung interval kelas dinamakan ujung interval kelas. Bilangan 500 pada interval kelas pertama dinamakan ujung bawah interval kelas pertama, bilangan 999 dinamakan ujung atas interval kelas pertama.
3. Bilangan yang berada tepat antara ujung bawah suatu interval kelas dengan ujung atas selang kelas sebelumnya dinamakan batas bawah interval kelas.

$$\text{Batas bawah interval} = \frac{\text{Ujung bawah interval} + \text{ujung atas interval sebelumnya}}{2}$$

$$\text{Batas bawah interval kedua} = \frac{1.000 + 999}{2} = 999,5$$

Bilangan 999,5 adalah batas bawah interval kelas kedua, bilangan 1.999,5 batas bawah interval kelas ketiga.

Bilangan yang tepat berada antara ujung atas suatu interval kelas dengan ujung bawah interval kelas berikutnya dinamakan batas atas interval kelas.

$$\text{Batas atas interval} = \frac{\text{Ujung atas interval} + \text{ujung bawah interval berikutnya}}{2}$$

$$\text{Batas atas interval kedua} = \frac{1.999 + 2.000}{2} = 1.999,5$$

Bilangan 1.999,5 adalah batas atas interval kelas yang kedua, bilangan 2.999,5 adalah batas atas interval yang ketiga.

Batas bawah suatu interval kelas sama dengan batas atas interval kelas sebelumnya, sedangkan batas atas suatu interval kelas sama dengan batas bawah interval kelas berikutnya.

4. Selisih antara batas atas dengan batas bawah suatu interval kelas dinamakan ukuran kelas.

$$\text{Ukuran interval kelas} = \text{batas atas} - \text{batas bawah}$$

$$\text{Ukuran interval kelas pertama} = 999,5 - 499,5 = 500$$

$$\text{Ukuran interval kelas kedua} = 1.999,5 - 999,5 = 1.000$$

Ukuran interval kelas juga dapat dihitung dari selisih ujung bawah suatu interval kelas dengan ujung bawah interval kelas berikutnya.

$$\text{Ukuran interval kelas pertama} = 500 - 1.000 = 500 \text{ (diambil harga mutlaknya)}$$

$$\text{Ukuran interval kelas kedua} = 1.000 - 2.000 = 1.000 \text{ (diambil harga mutlaknya)}$$

5. Suatu titik yang berada tepat di tengah suatu interval kelas dinamakan dengan markah kelas. Markah kelas diperoleh dengan menambahkan ujung bawah dengan ujung atas dan dibagi 2.

$$\text{Markah kelas} = \frac{\text{Ujung atas} + \text{ujung bawah}}{2}$$

$$\text{Markah kelas interval pertama} = \frac{999 + 500}{2} = 749,5$$

Markah kelas interval pertama adalah 749,9; markah kelas interval kedua adalah 1.499,5; markah kelas interval ketiga adalah 2.499,5.

Tabel 2.8. Perhitungan markah kelas tingkat pendapatan konsumen pada warung kopi XXX

Nomor	Tingkat pendapatan (ribu Rupiah)	Frekuensi konsumen (orang)	Markah kelas (X)
1	500 – 999	15	749,5
2	1.000 – 1.999	9	1.499.5
3	2.000 – 2.999	9	2.499.5
4	3.000 – 3.999	2	3.499.5
5	4.000 – 4.999	3	4.499.5
6	5.000 – 5.999	2	5.499.5

Dari contoh pada Tabel 2.8. dapat diketahui bahwa interval kelas pertama adalah 500 – 999 dengan frekuensi 15. Ujung bawah interval kelas pertama adalah 500 dan ujung atasnya adalah 999. Batas bawah interval kelas pertama tersebut adalah 499,5 dan batas atasnya adalah 999,5. Markah kelas interval pertama adalah 749,5.

Langkah-langkah Pembuatan Distribusi Frekuensi :

Pembuatan distribusi frekuensi dilakukan melalui langkah yang jelas dan berurutan. Langkah dalam pembuatan distribusi frekuensi tersebut adalah:

1. Menentukan data amatan tertinggi (H) dan data amatan terendah (L).
2. Menentukan rentang data amatan (R).
3. Menentukan banyaknya interval kelas.
4. Menentukan ukuran interval kelas.
5. Menentukan ujung bawah dan ujung atas setiap interval kelas.
6. Menghitung frekuensi pada setiap interval kelas.

Untuk menjelaskan langkah langkah pembuatan distribusi frekuensi kita gunakan data yang telah tersusun dalam array yaitu pada Tabel 2.2. yang dituliskan kembali sebagai berikut.

Tabel 2.2. Berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

No Urut	Berat lele (gr)								
28	119	22	136	37	144	31	149	36	157
26	125	1	138	10	145	11	150	7	158
8	126	13	138	35	145	25	150	23	163
40	128	12	140	2	146	5	151	6	164
16	132	34	140	4	146	32	152	38	165
15	135	14	142	17	147	29	153	3	168
24	135	20	142	19	147	33	154	9	173
39	135	21	144	27	148	30	156	18	176

1. Menentukan data amatan tertinggi (H) dan data amatan terrendah (L).

Perhatikan Tabel 2.2.. Dalam tabel tersebut telah disajikan berat lele hasil penimbangan mahasiswa yang telah disajikan secara berurutan, mulai data amatan terrendah sampai dengan data amatan tertinggi. Dengan mudah kita dapat menentukan data tertinggi dari tabel tersebut, yaitu $H = 176$, dan kita dapat juga menentukan data terrendah, yaitu $L = 119$. Jika data yang kita punyai belum tersusun dalam array, maka kita harus mencermati satu per satu data amatan untuk menentukan data amatan tertingginya (H) dan juga data amatan terrendahnya (L). Kesalahan dalam menentukan data amatan tertinggi atau data amatan terrendah atau kedua duanya akan menyebabkan beberapa data tidak dapat dikelompokkan dalam interval kelas yang kita buat atau interval kelas yang kita buat menjadi kurang efisien.

2. Menentukan rentang data amatan (R).

Rentang data amatan disebut juga dengan luas penyebaran data amatan atau *range* dan disimbulkan dengan R. Untuk menentukan rentang data amatan (R) harus diketahui terlebih dahulu data amatan tertinggi dan data amatan terrendah secara tepat agar seluruh data amatan tercakup dalam

distribusi frekuensi yang kita buat. Rentang data amatan (R) dihitung dengan cara mengurangi data amatan tertinggi (H) dengan data amatan terendah (L).

$$R = H - L$$

Keterangan:

R = Rentang data amatan

H = Data amatan tertinggi

L = Data amatan terendah

Untuk data tentang berat lele hasil penimbangan mahasiswa seperti yang disajikan pada Tabel 2.2. dapat diketahui bahwa rentang data amatannya adalah sebesar $176 - 119 = 57$

3. Menentukan banyaknya interval kelas.

Banyaknya interval kelas disimbulkan dengan k, sebaiknya tidak terlalu banyak tetapi juga jangan terlalu sedikit. Banyaknya interval kelas yang terlalu besar dapat memberikan gambaran yang kurang jelas dan sederhana tentang informasi yang tergambar dalam distribusi frekuensi tersebut, namun jika terlalu sedikit akan diperoleh distribusi frekuensi yang kasar. Banyaknya interval kelas disarankan berkisar antara 5 sampai 20, dipilih menurut keperluan. Bahkan jika kita mendasarkan pada suatu konsep yang ada, kita dapat menentukan banyaknya interval kelas berdasarkan konsep tersebut. Misalnya, dengan menggunakan konsep pengelompokan penduduk menurut produktivitas, kita mempunyai 3 kelompok usia yaitu usia belum produktif (0 - 15 tahun) , usia produktif (16 - 60), dan usia tidak produktif (lebih dari 60 tahun). Berdasarkan konsep tersebut kita dapat membuat distribusi frekuensi dengan 3 interval kelas, yaitu 0 - 15, 16 - 60, dan lebih dari 60.

Herbert A. Sturges memberikan suatu pedoman untuk menentukan banyaknya interval kelas yang dikenal dengan aturan *Sturges* sebagai berikut :

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

Keterangan:

k = banyaknya interval kelas

n = banyaknya data amatan

Banyaknya interval kelas hasil perhitungan dengan aturan Sturges tidak mesti merupakan bilangan bulat, namun banyaknya interval kelas yang harus kita buat harus merupakan bilangan bulat. Oleh karena itu kita mesti mengambil

bilangan bulat yang lebih rendah atau yang lebih tinggi dari hasil perhitungan dengan aturan Sturges. Pengambilan bilangan bulat ini bukan merupakan pembulatan ke bilangan bulat terdekat, namun kita bebas memilih bilangan bulat yang lebih rendah atau yang lebih tinggi.

Sebagai contoh data amatan yang disajikan pada Tabel 2.2. kita dapat ketahui bahwa banyaknya data amatan adalah 40 buah. Dengan menggunakan aturan Sturges, kita tentukan banyaknya interval kelas.

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

$$k = 1 + 3,322 \log 40$$

$$= 1 + 5,322 = 6,322 \approx 7 \text{ (pembulatan selalu keatas)}$$

Perlu dicatat bahwa aturan Sturges merupakan suatu yang mutlak harus diikuti. Dalam praktek sering dipakai pertimbangan-pertimbangan lain dalam menentukan jumlah selang kelas. Biasanya banyak selang kelas yang diambil antara 5 - 20 tergantung banyaknya data. Dalam contoh di atas dipilih pembulatan menjadi bilangan bulat yang lebih tinggi dari hasil perhitungan dengan menggunakan aturan Sturges.

4. Menentukan ukuran interval kelas.

Untuk menentukan ukuran selang kelas (i) dibutuhkan nilai rentang data amatan (R) dan banyaknya interval kelas (k) yang akan kita buat. Ukuran interval kelas (i) dihitung dengan menggunakan aturan sebagai berikut:

$$i = \frac{R}{k}$$

Keterangan:

i = ukuran interval kelas

R = rentang data amatan

k = banyaknya interval kelas

Ingat bahwa rentang data amatan adalah data amatan tertinggi (H) dikurangi dengan data amatan terendah (L), atau $R = H - L$.

Seperti halnya hasil perhitungan banyaknya interval kelas dengan menggunakan aturan Sturges, hasil perhitungan ukuran interval kelas juga tidak mesti menghasilkan bilangan bulat. Oleh karena itu, hasil perhitungan ukuran interval kelas diambil sesuai dengan satuan data amatan. Untuk data amatan satuan, sebaiknya pengambilan bilangan untuk ukuran interval kelas juga berupa bilangan bulat. Jika data amatan menggunakan satu desimal,

maka ukuran interval kelas juga satu desimal, dan begitu seterusnya. Untuk menentukan nilai ukuran interval kelas (i), dipilih bilangan yang lebih tinggi. Pemilihan yang lebih tinggi dilakukan agar semua data amatan dapat tercakup dalam distribusi frekuensi yang akan kita buat sesuai aturan yang telah kita buat. Misalnya, data amatan kita berupa satuan dan perhitungan ukuran interval kelas menghasilkan nilai 12,4321 maka kita ambil bilangan 13 sebagai ukuran interval kelas. Contoh lain, data amatan kita berupa satu desimal dan perhitungan ukuran interval kelas menghasilkan nilai 12,4321 maka kita ambil bilangan 12,5 sebagai ukuran interval kelas.

Dari contoh data amatan tentang berat lele hasil penimbangan yang dilakukan oleh mahasiswa yang disajikan pada Tabel 2.2. dapat diketahui bahwa data amatan tertinggi adalah 176, dan data amatan terendah adalah 119, sehingga rentang data amatan adalah sebesar $176 - 119 = 57$. Dari perhitungan dengan menggunakan aturan Sturges kita tentukan banyaknya interval kelas sejumlah 7, sehingga ukuran interval kelas yang kita gunakan adalah:

$$i = \frac{57}{7} = 8,1428 = 9 \text{ (pembulatan selalu ke atas)}$$

Karena data amatan pada Tabel 2.2. sampai pada satuan, maka ukuran interval kelas ditentukan sampai dengan satuan (bilangan bulat). Jadi ukuran interval kelas yang dibuat untuk data amatan yang disajikan pada Tabel 2.2. adalah 9. Sebagai catatan, jika kita ambil angka 8 sebagai ukuran interval kelas, maka dalam penghitungan frekuensi yang dilakukan pada langkah selanjutnya akan menyisakan beberapa data amatan yang tidak dapat dimasukkan dalam distribusi frekuensi.

5. Menentukan ujung bawah dan ujung atas setiap interval kelas.

Ujung bawah interval kelas pertama dapat dipilih dari data amatan yang terkecil atau bilangan lain yang lebih kecil dari data amatan yang terkecil tetapi selisihnya kurang dari ukuran selang kelas yang telah ditentukan. Jika data amatan terkecil kita adalah 50 dan ukuran interval kelasnya adalah 6, maka ujung bawah interval kelas pertama dapat dipilih bilangan 50 atau 49 atau 48 bilangan lain sampai dengan 45.

Untuk menjaga agar data amatan tersusun dalam distribusi frekuensi secara seimbang maka disarankan menentukan ujung bawah interval pertama dengan data lebih kecil dari data amatan terkecil. Adapun penurunan dari data amatan terkecil ini adalah kurang lebih sebesar setengah selisih rentang data amatan dengan rentang distribusi frekuensi. Rentang distribusi frekuensi kita adalah perkalian antara banyaknya interval kelas dengan ukuran interval kelas. Misalnya banyaknya interval kelas adakah sebesar 7 dan ukuran interval kelas sebesar 6 maka rentang distribusi frekuensi kita adalah $7 \times 6 = 42$. Misalkan data amatan terkecil kita adalah 50 dan rentang data amatan kita adalah 39, maka kita sebaiknya menurunkan sebesar $(42 - 39):2 = 1$ dari data amatan terkecil. Sehingga kita gunakan angka 49 sebagai ujung bawah interval pertama.

Pada contoh data amatan tentang berat lele hasil penimbangan yang dilakukan oleh mahasiswa seperti yang disajikan pada Tabel 2.2. kita ketahui bahwa banyaknya interval kelas yang ditentukan adalah sebesar 7 dan ukuran interval kelasnya adalah sebesar 9, sehingga rentang distribusi frekuensinya adalah $7 \times 9 = 63$. Rentang data amatannya adalah sebesar 57, sehingga sebaiknya menurunkan sebesar 2 atau 3 satuan dari data terkecil. Dalam contoh ini kita memilih penurunan sebesar 2 sehingga ujung bawah interval pertama ditentukan 117.

Ujung bawah interval kedua adalah ujung bawah interval pertama ditambah dengan ukuran interval kelas, ujung bawah interval ketiga adalah ujung bawah interval kedua ditambah ukuran interval kelas, dan seterusnya. Dalam contoh data amatan tentang berat lele hasil penimbangan yang dilakukan oleh mahasiswa yang disajikan pada Tabel 2.2. maka ujung bawah interval pertama adalah 117, ujung bawah interval kedua adalah $117 + 9 = 126$, ujung bawah interval ketiga adalah 135 dan seterusnya.

Ujung atas interval pertama dihitung sebesar 1 satuan lebih kecil dari ujung bawah interval kedua, ujung atas interval kedua juga dihitung sebesar 1 satuan lebih kecil dari ujung bawah interval kelas ketiga, dan seterusnya. Pada contoh, ujung bawah interval kedua adalah 126, sehingga ujung atas interval pertama adalah $126 - 1 = 125$, ujung atas interval kedua adalah 134, dan seterusnya.

Setelah penghitungan ujung bawah dan ujung atas seluruh interval telah dilakukan maka disusun interval kelas. Pada contoh amatan tentang berat lele hasil penimbangan yang dilakukan oleh mahasiswa yang disajikan pada Tabel 2.2. maka interval kelas pertama adalah 117 – 125, interval kelas kedua adalah 126 – 134, dan seterusnya. Selanjutnya disusun dalam tabel distribusi frekuensi secara lengkap, sebagai berikut:

Interval kelas berat lele
117 – 125
126 – 134
135 – 143
144 – 152
153 – 161
162 – 170
171 – 179

6. Menghitung frekuensi pada setiap interval kelas.

Untuk menghitung frekuensi diperlukan tabel bantu yang berfungsi untuk menentukan frekuensi pada setiap interval kelas. Penghitungan frekuensi setiap interval kelas dapat menggunakan salah satu dari dua cara berikut ini:

- a. Cara toli (*tally*)
- b. Cara entri (*entry*)

a. Menghitung frekuensi pada interval kelas dengan cara toli

Penghitungan frekuensi dengan cara toli dilakukan dengan membuat satu kolom untuk mencatat setiap data amatan pada interval kelas tertentu. Pencatatan data amatan pada kolom toli, dilakukan secara berurutan mulai data amatan pertama, kedua, ketiga, dan seterusnya sampai data amatan yang terakhir. Kita akan membuat distribusi dari data yang disajikan pada Tabel 2.2. Hasil pencatatan disajikan pada tabel 2.9.

Tabel 2.9. Hasil pencatatan berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Interval kelas berat lele	Toli	Frekuensi
117 – 125	II	2
126 – 134	III	3
135 – 143	IIII IIII	10
144 – 152	IIII IIII III	14

153 – 161	IIII	5
162 – 170	III	4
171 – 179	II	2
J u m l a h		40

b. Menghitung frekuensi pada interval kelas dengan cara entri

Penghitungan distribusi frekuensi dengan cara entri, dilakukan dengan membuat kolom bantu bagi setiap interval, sehingga banyaknya kolom bantu sama dengan banyaknya interval. Setiap data amatan dituliskan kembali pada kolom interval yang bersesuaian. Data pada Tabel 2.2. kita tuliskan kembali dalam tabel entri data, yang hasilnya disajikan pada Tabel 2.10 berikut ini.

Tabel 2.10. Hasil entri data berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

	Interval kelas berat lele						
	117 – 125	126 – 134	135 – 143	144 – 152	153 – 161	162 – 170	171 – 179
D	119	126	135	144	153	163	173
A	125	128	135	144	154	164	176
T		132	135	145	156	165	
A			136	145	157	168	
			138	146	158		
			138	146			
			140	147			
			140	147			
			142	148			
			142	149			
				150			
				150			
				151			
				152			
N	2	3	10	14	5	4	2

Setelah dilakukan penghitungan frekuensi setiap interval baik dengan cara toli ataupun dengan cara entri, selanjutnya tabel tersebut disajikan dengan memberikan judul dan sumber, sehingga didapatkan bentuk

distribusi frekuensi yang lazim digunakan dalam penulisan laporan. Tabel distribusi frekuensi tersebut adalah berikut :

Tabel 2.11. Distribusi frekuensi berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Nomor	Interval kelas berat lele	Frekuensi
1	117 – 125	2
2	126 – 134	3
3	135 – 143	10
4	144 – 152	14
5	153 – 161	5
6	162 – 170	4
7	171 – 179	2
	J u m l a h	40

5. Distribusi Frekuensi Relatif

Frekuensi yang dihitung dan disajikan dalam Tabel 2.11. menunjukkan banyaknya data amatan untuk setiap interval kelas, dan frekuensinya dinyatakan dalam bentuk absolut. Pada interval pertama misalnya, berat ikan lele antara 117 – 125 gram adalah sebanyak 2 ekor. Selain secara absolut, nilai frekuensi juga dapat dinyatakan secara relatif. Distribusi frekuensi yang nilai frekuensinya dinyatakan secara relatif disebut distribusi frekuensi relatif.

Nilai frekuensi relatif dihitung dengan cara mencari proporsi frekuensi suatu interval terhadap frekuensi totalnya, dan biasanya dinyatakan dalam persentase. Frekuensi relatif suatu interval dicari dengan cara membagi frekuensi suatu interval dengan total frekuensi dan dikalikan dengan 100%. Sebagai contoh frekuensi interval pertama Tabel 2.11. adalah 2, sedangkan frekuensi totalnya adalah 40, sehingga frekuensi relatif interval pertama adalah $(2/40) \times 100\% = 5\%$. Dihitung dengan cara yang sama, frekuensi relatif interval kedua adalah 7,5%, frekuensi relatif interval ketiga adalah 25%, dan seterusnya. Distribusi frekuensi relatif berat ikan lele dapat dilihat pada Tabel 2.12.

Tabel 2.12. Distribusi frekuensi relatif berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Nomor	Interval kelas berat lele	Frekuensi	Frekuensi relatif (%)
1	117 – 125	2	5
2	126 – 134	3	7,5

3	135 – 143	10	25
4	144 – 152	14	35
5	153 – 161	5	12,5
6	162 – 170	4	10
7	171 – 179	2	5
	J u m l a h	40	100

6. Distribusi Frekuensi Kumulatif

Penyusunan distribusi frekuensi kumulatif pada dasarnya sama dengan penyusunan frekuensi absolut, yang membedakan adalah frekuensi dinyatakan secara akumulatif. Frekuensi kumulatif menyatakan frekuensi seluruh data amatan yang lebih kecil atau lebih besar dari suatu nilai tertentu. Nilai tertentu ini biasanya digunakan ujung bawah suatu interval. Distribusi frekuensi yang nilai frekuensinya dinyatakan secara akumulatif dinamaka distribusi frekuensi kumulatif.

Distribusi frekuensi yang mengandung frekuensi seluruh data amatan yang lebih kecil dari suatu nilai tertentu dinamakan dengan distribusi frekuensi kumulatif “kurang dari”, karena pada dasarnya menyatakan frekuensi dari data amatan yang nilai kurang dari suatu nilai tertentu. Misalnya dari Tabel 2.11. kita dapat menyatakan frekuensi “kurang dari 117” adalah 0, frekuensi “kurang dari 126” adalah 2, frekuensi “kurang dari 135” adalah 5, dan seterusnya. Secara mudah frekuensi kumulatif dihitung dengan menjumlahkan seluruh frekuensi yang dinyatakan sebelumnya.

Jika diperlukan frekuensi kumulatif dapat dinyatakan sebagai nilai relatifnya dan sering dinamakan frekuensi kumulatif relatif “kurang dari”. Frekuensi kumulatif relatif diperoleh dari membagi frekuensi kumulatif dengan frekuensi total dan mengkalikannya dengan 100%. Misalnya dari Tabel 2.11. kita dapat menyatakan frekuensi “kurang dari 117” adalah $(0/40) \times 100\% = 0\%$, frekuensi “kurang dari 126” adalah $(2/40) \times 100\% = 5\%$, frekuensi “kurang dari 135” adalah 12,5%, dan seterusnya. Jika frekuensi relatif sudah diketahui, frekuensi kumulatif relatif dapat dihitung dengan menjumlahkan seluruh frekuensi relatif yang dinyatakan sebelumnya.

Tabel 2.13. Distribusi frekuensi kumulatif “kurang dari” berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Nomor	Berat lele	Frekuensi kumulatif	Frekuensi kumulatif relatif (%)
1	Kurang dari 117	0	0
2	Kurang dari 126	2	5
3	Kurang dari 135	5	12,5
4	Kurang dari 144	15	37,5
5	Kurang dari 153	29	72,5
6	Kurang dari 162	34	85
7	Kurang dari 171	38	95
8	Kurang dari 180	40	100

Dari distribusi frekuensi kumulatif “kurang dari” kita dapat dengan cepat mengetahui banyaknya data amatan yang nilainya nilai kurang dari suatu nilai tertentu. Misalkan, dari Tabel 2.13. kita dapat dengan cepat mengetahui banyaknya lele yang beratnya kurang dari 117 gram, yaitu sebanyak 0 ekor (dibaca tidak satu ekor pun lele yang beratnya kurang dari 117 gram), banyaknya lele yang beratnya kurang dari 126 gram adalah sebanyak 2 ekor atau 5%.

Distribusi frekuensi yang mengandung frekuensi seluruh data amatan yang lebih besari dari suatu nilai tertentu dinamakan distribusi frekuensi kumulatif “lebih dari”, karena frekuensinya menyatakan banyaknya data amatan yang lebih dari nilai tertentu. Misalnya pada Tabel 2.11 kita dapat menyatakan frekuensi “117 atau lebih” adalah 40, frekuensi “126 atau lebih” adalah 38, frekuensi “135 atau lebih” adalah 35, dan seterusnya. Frekuensi kumulatif dapat dihitung dengan menjumlahkan frekuensi yang bersangkutan dengan seluruh frekuensi sesudahnya.

Jika diperlukan frekuensi kumulatif “lebih dari” juga dapat dinyatakan sebagai nilai relatif dan dinamakan frekuensi relatif kumulatif “lebih dari”. Frekuensi kumulatif relatif diperoleh dari membagi frekuensi kumulatif dengan frekuensi total dan mengkalikannya dengan 100%. Misalnya pada Tabel 2.11 kita dapat menyatakan frekuensi “117 atau lebih” adalah 100%, frekuensi “126 atau lebih” adalah 95, frekuensi “135 atau lebih” adalah 87,5%, dan seterusnya. Frekuensi kumulatif relatif dapat dihitung dengan menjumlahkan frekuensi relatif yang bersangkutan dengan seluruh frekuensi relatif sesudahnya.

Tabel 2.14. Distribusi frekuensi kumulatif “lebih dari” berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Nomor	Berat lele	Frekuensi kumulatif	Frekuensi kumulatif relatif (%)
1	117 atau lebih	40	100
2	126 atau lebih	38	95
3	135 atau lebih	35	87,5
4	144 atau lebih	25	62,5
5	153 atau lebih	11	27,5
6	162 atau lebih	6	15
7	171 atau lebih	2	5
8	180 atau lebih	0	0

Dari distribusi frekuensi kumulatif “lebih dari” kita dapat dengan mudah mengetahui banyaknya data amatan yang nilainya lebih dari suatu nilai tertentu. Misalkan, dari Tabel 2.14 kita dapat mengetahui banyaknya lele yang beratnya 117 gram atau lebih adalah sebanyak 40 ekor atau 100%, banyaknya lele yang beratnya 126 gram atau lebih adalah sebanyak 38 ekor atau 95%.

5. Histogram, Poligon dan Ogive

Histogram, poligon dan ogive merupakan grafik yang menyajikan distribusi data statistik selain distribusi frekuensi. Grafik distribusi frekuensi ini merupakan penggambaran data amatan yang memberi penjelasan kepada pembaca tentang hal yang berkenaan grafik itu. Grafik distribusi frekuensi biasanya menggunakan sumbu mendatar (X) guna menyatakan kelas interval yang diberi nama sumbu pada ujung sumbu atau di bawah sumbu. Sumbu tegak (Y) guna menyatakan frekuensi yang diberi nama sumbu pada ujung sumbu atau di kiri sumbu. Pada sumbu X dituliskan semua batas interval kelas, sedangkan pada sumbu Y dituliskan angka angka frekuensi yang memenuhi interval kelas. Pada umumnya perbandingan antara sumbu X dengan sumbu Y adalah sepuluh berbanding tujuh atau empat berbanding tiga.

Histogram

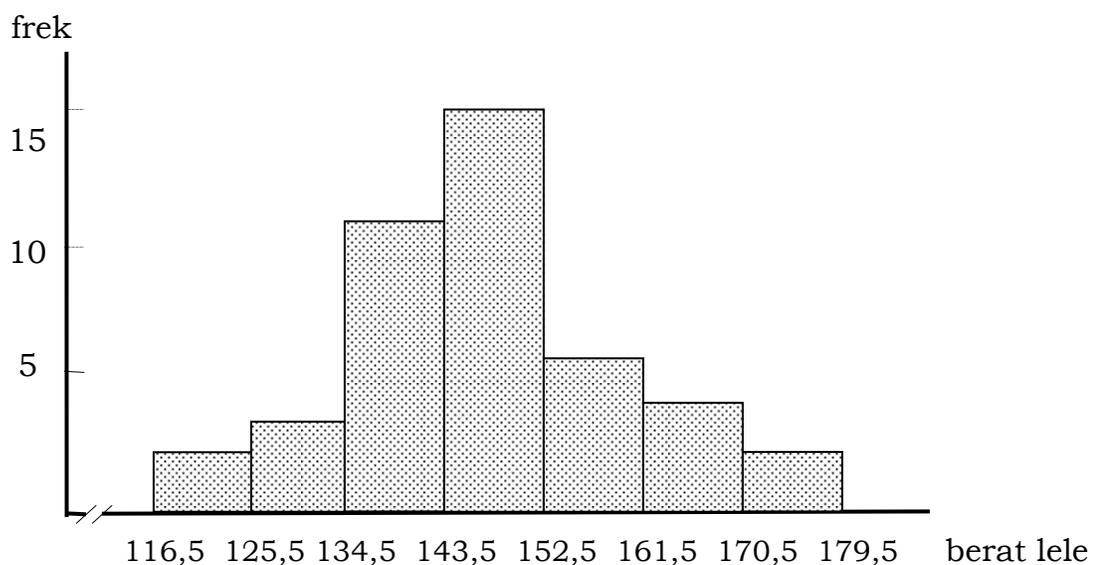
Histogram adalah grafik frekuensi variabel tunggal. Pada grafik ini luas batang sebanding dengan frekuensi. Lebar setiap bagian menggambarkan ukuran interval kelas. Sisi batang pada histogram saling berhimpitan, karena sisi batang pada histogram ini merupakan batas interval kelas. Jumlah batang

pada histogram sesuai dengan banyaknya interval kelas pada distribusi frekuensi.

Untuk membuat histogram diperlukan langkah langkah sebagai berikut:

- Menggambar sumbu X dan sumbu Y dengan perbandingan 10:7 atau 4:3.
- Memberi nama sumbu X dengan nama sesuai dengan data yang akan dibuatkan histogram, dan sumbu Y dengan nama "frekuensi".
- Menuliskan semua batas interval kelas pada sumbu mendatar, dan skala frekuensi pada sumbu Y. Skala sumbu X tidak harus sama dengan sumbu Y, tetapi harus memenuhi sumbu X dan juga sumbu Y.
- Jika jarak antara 0 dengan nilai pertama pada sumbu X ataupun sumbu Y memiliki perbandingan yang tidak sama dengan yang lain, maka antara titik 0 dan nilai pertama dibuat tanda pemotong grafik.
- Membuat batangan pada sumbu mendatar yang tingginya sama dengan frekuensi interval kelas.
- Memberi keterangan selengkap lengkapnya tentang histogram tersebut.

Berdasarkan Tabel 2.11. dapat dibuatkan histogram sebagaimana ditunjukkan pada Gambar 2.1.

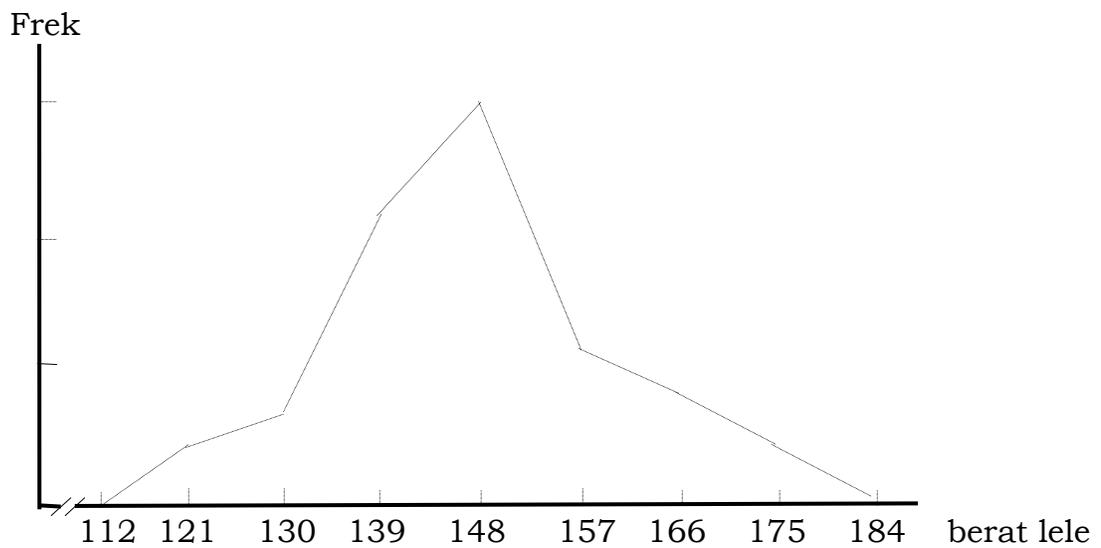


Gambar 2.1. Histogram berat berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Poligon

Poligon merupakan grafik garis yang menggambarkan distribusi frekuensi. Sumbu mendatar (sumbu X) menyatakan kelas interval yang diberi nama sumbu sesuai data yang dibuat distribusi frekuensi. Nama ini dituliskan pada ujung sumbu atau di bawah sumbu. Sumbu tegak (sumbu Y) menyatakan frekuensi interval kelas yang bersesuaian. Untuk data kontinyu, titik ordinat pada sumbu Y digambar pada markah masing-masing interval kelas.

Garis pada poligon dibuat dengan menghubungkan titik titik ordinat pada sumbu Y pada setiap markah interval kelas. Dengan cara yang mudah, poligon dibuat dengan membuat garis lurus yang menghubungkan titik tengah sisi atas batang pada histogram yang berdekatan. Dengan menggunakan histogram seperti pada Gambar 2.1. kita membuat poligon berat berat ikan lele.



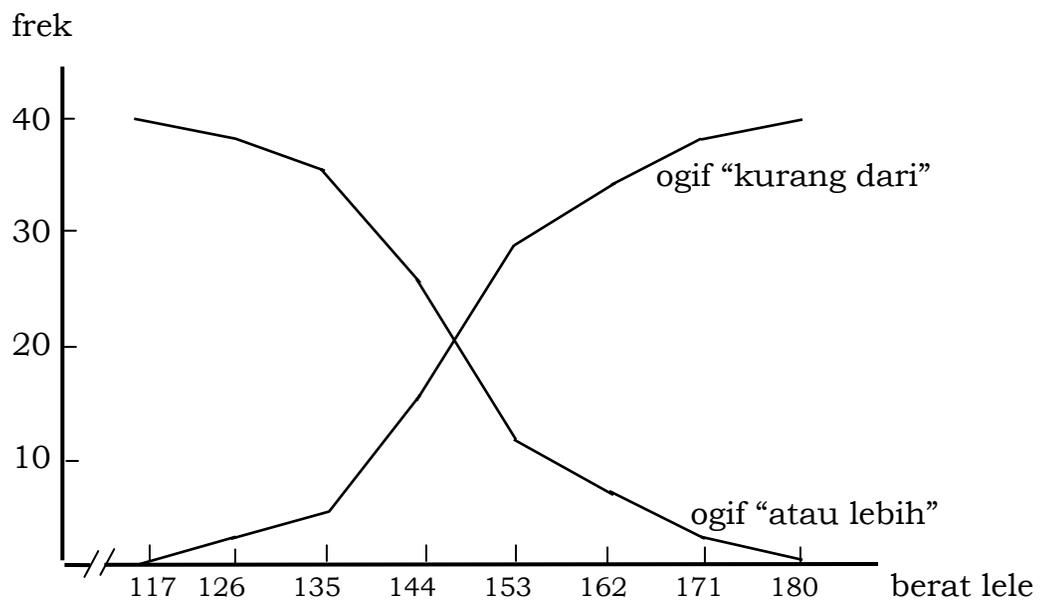
Gambar 2.2. Poligon berat berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Ogive

Ogive atau ogif adalah penyajian grafis dari distribusi frekuensi kumulatif. Ogif merupakan diagram garis dari distribusi frekuensi kumulatif. Oleh karena distribusi frekuensi kumulatif ada dua macam, yaitu distribusi frekuensi kumulatif “lebih dari” dan distribusi frekuensi kumulatif “kurang dari”, maka juga dapat dibuat dua macam ogif, yaitu ogif “kurang dari” dan ogif “lebih dari”.

Titik koordinat pada ogif ditentukan dari ujung bawah interval dan frekuensi kumulatif interval kelas yang bersangkutan. Ogif dibuat dengan menarik garis dari koordinat interval pertama ke koordinat interval kedua, dilanjutkan ke koordinat interval ketiga, sampai koordinat interval terakhir.

Data berat ikan lele yang disajikan pada Tabel 2.13 dan Tabel 2.14 dapat dibuatkan ogif sebagaimana disajikan pada Gambar 2.3.



Gambar 2.3. Ogif berat berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

BAB III UKURAN GEJALA PUSAT

Untuk mendapatkan gambaran yang lebih jelas tentang permasalahan yang digambarkan dari sekumpulan data, baik data dari sampel maupun dari populasi, diperlukan ukuran yang merupakan wakil kumpulan data tersebut. Suatu kumpulan data pada umumnya mempunyai kecenderungan untuk memusat pada suatu nilai tertentu. Nilai tertentu tersebut berupa nilai tunggal atau nilai tendensi pusat yang sering disebut Nilai Pusat, karena pada umumnya berlokasi di bagian tengah dari suatu distribusi. Ukuran gejala pusat berfungsi untuk memperoleh gambaran himpunan data. Ukuran gejala pusat mempunyai nilai yang khas, yaitu rata-rata. Terdapat beberapa konsep rata-rata, yaitu rata-rata hitung (sering ditulis “rata-rata” saja), modus, median, rata-rata geometri.

1. Rata-rata Hitung (*arithmetic mean*)

Rata-rata hitung dilambangkan dengan μ atau \bar{X} . Rata-rata dari suatu himpunan data $X_1, X_2, X_3, \dots, X_n$ dihitung dengan menjumlahkan seluruh data dan dibagi dengan banyaknya data:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

Lebih sederhana dapat ditulis dengan rumus:

$$\bar{X} = \frac{\sum X}{n}$$

Contoh.

Marjin pemasaran dari 8 orang pedagang untuk setiap kilogram beras adalah sebagai berikut (Rp/kg) adalah 700; 500; 500; 700; 640; 550; 660; dan 750;. Tentukan rata-rata hitung marjin pemasaran pedagang tersebut!

Rata-rata hitung pedagang adalah

$$\bar{X} = \frac{700 + 500 + 500 + 700 + 640 + 550 + 660 + 750}{8}$$

= 650

Jika nilai $X_1, X_2, X_3, \dots, X_n$ masing-masing sebanyak $f_1, f_2, f_3, \dots, f_n$, maka rata-rata hitung dihitung dengan rumus:

$$\bar{X} = \frac{\sum fX}{\sum f}$$

Misalnya, marjin pemasaran 4 orang pedagang untuk setiap kilogram beras adalah Rp. 500; 6 orang pedagang lainnya memperoleh marjin Rp. 650; 7 orang lainnya memperoleh marjin Rp. 700; dan 3 orang lainnya memperoleh marjin Rp. 600;. Hitunglah rata-rata hitung!

$$\begin{aligned}\bar{X} &= \frac{(4 \times 500) + (6 \times 650) + (7 \times 700) + (3 \times 650)}{4 + 6 + 7 + 3} \\ &= 637,5\end{aligned}$$

Untuk data yang telah disajikan dalam distribusi frekuensi, nilai X didekati dengan markah selang kelas, sedangkan $f_1, f_2, f_3, \dots, f_n$ menunjukkan frekuensi.

Contoh: Penimbangan yang dilakukan oleh mahasiswa terhadap 40 ekor ikan lele yang dijual pedagang di pasar disajikan pada Tabel 2.11 sebagai berikut.

Hitunglah rata-rata hitung berat 40 ekor ikan lele tersebut!

Tabel 2.11. Distribusi frekuensi berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Nomor	Interval kelas berat lele	Frekuensi
1	117 – 125	2
2	126 – 134	3
3	135 – 143	10
4	144 – 152	14
5	153 – 161	5
6	162 – 170	4
7	171 – 179	2
	J u m l a h	40

Penyelesaian:

Untuk menghitung rata-rata hitung kita harus menghitung markah setiap interval kelas. Markah interval kelas ke-1 adalah $(117+125):2=121$; markah interval kelas ke-2 adalah $(126+134):2=130$; dan seterusnya. Markah kelas

tersebut digunakan sebagai pendekatan atas nilai X pada interval yang bersesuaian. Selanjutnya nilai setiap markah interval kelas dikalikan dengan frekuensi interval yang bersesuaian. Hasilnya disajikan dalam Tabel 3.1.

Tabel 3.1. Perhitungan rata-rata hitung terhadap berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Interval kelas berat lele	Markah (X)	Frekuensi (f)	$f.X$
117 – 125	121	2	242
126 – 134	130	4	520
135 – 143	139	10	1.390
144 – 152	148	12	1.776
153 – 161	157	6	942
162 – 170	166	4	664
171 – 179	175	2	350
J u m l a h		40	5.884

Maka rata-rata hitungnya adalah

$$\begin{aligned}\bar{X} &= \frac{5.884}{40} \\ &= 147,1\end{aligned}$$

Penghitungan rata-rata hitung dapat dilakukan dengan cara yang lebih pendek. Untuk menghitung rata-rata hitung ini, kita dapat mengambil sebarang angka sebagai terkaan atas rata-rata hitung. Angka terkaan rata-rata hitung tersebut yaitu A . Untuk mempermudah perhitungan, nilai A diambil dari salah satu markah interval kelas yang diperkirakan paling mendekati rata-rata hitung. Selanjutnya kita mencari nilai d yang diperoleh dari $d = X - A$. Penghitungan rata-rata hitung dapat dilakukan dengan rumus:

$$\bar{X} = A + \frac{\sum f \cdot d}{\sum f}$$

Keterangan:

A = angka terkaan rata-rata hitung

$d = X - A$

f = frekuensi

Contoh:

Dengan menggunakan data berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa sebagaimana telah disajikan dalam Tabel 2.10. Hitunglah rata-rata hitung dengan menggunakan metode pendek!

Penyelesaian:

Untuk menghitung rata-rata hitung dengan cara pendek, kita ambil 148 sebagai angka terkaan rata-rata hitung (A). Nilai-nilai d kita hitung dengan $d = X - A$. Untuk interval ke-1 nilai d -nya adalah $(121-148)=-27$, interval ke-2 nilai d -nya adalah $(130-148)=-18$, dan seterusnya. Dengan cara mudah kita dapat menentukan $d=0$ pada interval yang markahnya yang kita gunakan sebagai angka terkaan rata-rata hitung (A). Nilai d pada interval sebelumnya secara berurutan adalah sebesar -1 kali ukuran interval kelas, -2 kali ukuran interval kelas, dan seterusnya. Sedangkan nilai d pada interval sesudahnya secara berurutan adalah sebesar 1 kali ukuran interval kelas, 2 kali ukuran interval kelas, dan seterusnya. Perhatikan contoh kita ini, $A=148$ sehingga kita tentukan $d=0$. Ukuran interval kelas kita adalah $125,5-116,5=9$, sehingga nilai d pada interval sebelumnya secara berurutan adalah -9 , -18 , dan -27 , sedangkan nilai d pada interval setelahnya adalah 9 , 18 , dan 27 . Hasil perhitungan selengkapnya disajikan pada Tabel 3.2.

Tabel 3.2. Perhitungan rata-rata hitung dengan menggunakan cara pendek terhadap berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Interval kelas berat lele	Markah (X)	d	Frekuensi (f)	$f.d$
117 – 125	121	-27	2	- 54
126 – 134	130	-18	4	- 72
135 – 143	139	-9	10	- 90
144 – 152	148 = A	0	12	0
153 – 161	157	9	6	54
162 – 170	166	18	4	72
171 – 179	175	27	2	54
J u m l a h			40	-36

Maka rata-rata hitungnya adalah

$$\begin{aligned} \bar{X} &= 148 + \frac{-36}{40} \\ &= 147,1 \end{aligned}$$

Dengan mengambil A sebagai angka terkaan rata-rata hitung, $d = X - A$, dan jika ukuran intervalnya sama yaitu sebesar c , sehingga $u = d/c$ maka penghitungan rata-rata hitung dapat dilakukan dengan menggunakan cara sandi dengan rumus sebagai berikut:

$$\bar{X} = A + \left(\frac{\sum fu}{\sum f} \right) c$$

Keterangan:

A = angka terkaan rata-rata hitung

f = frekuensi

c = ukuran interval kelas

$u = d/c = (X - A)/c$

Contoh:

Dengan menggunakan data berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa sebagaimana telah disajikan dalam Tabel 2.10. Hitunglah rata-rata hitung dengan menggunakan cara sandi.

Penyelesaian:

Untuk menghitung rata-rata hitung dengan cara sandi, kita ambil 148 sebagai angka terkaan rata-rata hitung (A). Nilai u kita hitung dengan $u = (X - A)/c$. Untuk interval ke-1 nilai u -nya adalah $(121 - 148)/9=3$, untuk interval ke-2 nilai u -nya adalah $(130-148)/9=2$, dan seterusnya. Dengan cara mudah kita dapat menentukan $u=0$ pada interval yang markahnya kita gunakan sebagai angka terkaan rata-rata hitung (A), sedangkan nilai u pada interval sebelumnya secara berurutan adalah $-1, -2, -3$, dan seterusnya. Sedangkan nilai u pada interval sesudahnya secara berurutan adalah $1, 2, 3$, dan seterusnya. Perhatikan contoh kita, $A=148$ sehingga kita tentukan $u=0$. Nilai u pada interval sebelumnya adalah $-1, -2$, dan -3 , sedangkan nilai u pada interval sesudahnya adalah $1, 2$, dan 3 . Hasil perhitungan selengkapnya disajikan pada Tabel 3.3.

Tabel 3.3. Perhitungan rata-rata hitung dengan menggunakan cara sandi terhadap berat ikan lele hasil penimbangan yang dilakukan oleh mahasiswa

Interval kelas berat lele	Markah kelas (X)	u	Frekuensi (f)	$f.u$
------------------------------	-------------------------	-----	----------------------	-------

117 - 125	121	-3	2	-6
126 - 134	130	-2	4	-8
135 - 143	139	-1	10	-10
144 - 152	148 = A	0	12	0
153 - 161	157	1	6	6
162 - 170	166	2	4	8
171 - 179	175	3	2	6
J u m l a h			40	-4

Maka nilai tengahnya adalah

$$\begin{aligned}\bar{X} &= 148 + \left(\frac{-4}{40}\right) 9 \\ &= 147,1\end{aligned}$$

2. Nilai tengah berbobot

Merupakan modifikasi dari nilai tengah hitung, yaitu dengan memberikan bobot tertentu pada masing-masing nilai amatan, sehingga akan menghasilkan nilai tengah yang lebih menggambarkan kondisi data yang sebenarnya.

$$\bar{X} = \frac{\sum wX}{\sum w}$$

Contoh

Hasil panen lima petani dengan luas sawah yang berbeda-beda sebagai berikut:

Petani	Luas Sawah (Ha)	Panen per Ha (Ku)
A	3	40
B	4	30
C	5	45
D	2	50
E	3	70

Berapa rata-rata panen per hektarnya ?

$$\begin{aligned}\bar{X} &= \frac{3(40) + 4(30) + 5(45) + 2(50) + 3(70)}{3 + 4 + 5 + 2 + 3} \\ &= 45,59 \text{ kuintal}\end{aligned}$$

3. Nilai tengah geometrik

Jika kuantitas mengalami perubahan pada setiap periode dan kita ingin mengetahui nilai tengah tingkat pertumbuhan setiap periode, maka dicari dengan metode nilai tengah geometrik.

Contoh

Jika Mohd Alfa menabung sebesar Rp 1.000.000 selama 5 tahun, dengan tingkat bunga pada tahun ke 1, 2, 3, 4, dan 5 masing-masing sebesar 7%, 8%, 10%, 12%, dan 18% maka tingkat bunga rata-rata harus dihitung dengan metode nilai tengah geometrik sebagai berikut:

faktor pertumbuhan $(x) = 1 + \text{tingkat bunga}$

$$\bar{x} = \sqrt[5]{1,07(1,08)(1,10)(1,12)(1,18)}$$
$$= 1,10933$$

Sehingga tingkat bunga rata-rata adalah sebesar $1,10933 - 1 = 0,10933$ atau 10,933%

4. Median

Median adalah nilai yang terletak di tengah dari suatu pengamatan yang telah disusun secara teratur mulai dari terkecil sampai terbesar atau sebaliknya. Nilai media dipengaruhi oleh letak urutan dari kumpulan data, sehingga median sering disebut rata-rata letak.

Untuk data yang tidak dikelompok, median dicari nilai data yang tepat berada di urutan tengah, yaitu nilai yang berada pada posisi ke $(n+1):2$ untuk data sebanyak n

Contoh: skor tingkat penerapan teknologi yang diperoleh oleh 7 orang petani adalah 1, 1, 2, 3, 4, 4, 6

Median dari skor penerapan teknologi tersebut adalah 3, yaitu data yang ke 4

Untuk data yang telah disusun dalam distribusi frekuensi, maka letak median adalah pada posisi ke $(n+1):2$, sedangkan nilai median dicari dengan rumus

$$Med = L_1 + \left(\frac{n/2 - \sum f_1}{f_{med}} \right) \cdot c$$

L_1 = batas bawah selang kelas yang mengandung median

n = banyaknya data

$\sum f_1$ = jumlah frekuensi selang kelas yang lebih rendah dari selang kelas median

f_{med} = frekuensi yang mengandung median

c = ukuran selang kelas

Contoh

Panjang daun salam(mm)	Frekuensi
117 – 125	2
126 – 134	3
135 – 143	10
Posisi median → 144 – 152	13
153 – 161	6
162 – 170	4
171 – 179	2
J u m l a h	40

Pada contoh tabel frekuensi di atas, banyaknya data adalah 40 buah, sehingga median ditentukan dengan perhitungan $\frac{40+1}{2} = 20,5$. Sehingga median merupakan data yang terletak antara 20 sampai 21. Data tersebut terletak pada selang kelas 144 – 152.

$$\begin{aligned} Med &= 143,5 + \left(\frac{40/2 - 15}{13} \right) \cdot 9 \\ &= 146,96 \end{aligned}$$

5. Modus

Modus adalah nilai data yang paling sering muncul dalam suatu pengamatan. Oleh karenanya modus digunakan untuk mengetahui tingkat keseringan terjadinya suatu peristiwa. Modus merupakan nilai gejala pusat yang dapat digunakan untuk semua skala pengukuran, namun yang paling cocok adalah data nominal.

Contoh, jika ditanyakan kepada PPL tentang hama/penyakit yang menjadi musuh petani maka akan dijawab dengan hama yang paling sering menyerang di lahan petani, misalnya hama tikus. Hal ini menunjukkan bahwa hama tikus merupakan modus dari hama/penyakit yang menyerang lahan petani.

Pada data yang berupa angka, modus dapat diketahui dari data yang paling sering muncul. Jika data belum disajikan dalam distribusi frekuensi, maka modus dapat diketahui dengan cara memeriksa yang paling sering muncul atau yang mempunyai frekuensi tertinggi.

Suatu data yang belum dikelompokkan boleh jadi:

- tidak mempunyai modus, misalnya 1, 2, 3, 5, 8
- mono modus, misalnya 1, 2, 3, 5, 5, 8 (modusnya adalah 5)
- bimodus, misalnya 1, 2, 2, 5, 5, 8 (modusnya adalah 2 dan 5)

Pada data yang telah dikelompokkan dalam distribusi frekuensi, modus terletak pada selang kelas yang mempunyai frekuensi terbesar. Modus dari data ini didekati dengan rumus:

$$Mo = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \cdot c$$

L_1 = batas selang kelas yang mengandung modus

Δ_1 = selisih frekuensi selang kelas yang mengandung modus dengan selang kelas sebelumnya

Δ_2 = selisih frekuensi selang kelas yang mengandung modus dengan selang kelas sesudahnya

Contoh

Panjang daun salam(mm)	Frekuensi
117 – 125	2
126 – 134	3
135 – 143	10
Interval kelas modus → 144 – 152	13
153 – 161	6
162 – 170	4
171 – 179	2
J u m l a h	40

Selang kelas yang mempunyai frekuensi terbesar pada tabel distribusi frekuensi di atas adalah selang kelas 144 – 152. Sehingga modus berada pada selang kelas 144 – 152.

$$\begin{aligned}
 Mo &= 143,5 + \left(\frac{3}{3 + 7} \right) \cdot 9 \\
 &= 146,2
 \end{aligned}$$

BAB IV

UKURAN VARIASI

Informasi yang diperoleh dari nilai gejala pusat pada suatu gugus data belumlah lengkap, karena kita belum tahu penyebaran data dalam gugus data tersebut dari nilai gejala pusat. Perhatikan contoh berikut:

A : 2, 3, 7, 7, 16

B : 5, 6, 7, 7, 10

Nilai gejala pusat dari gugus data A dan juga adalah sama, yaitu nilai tengah sebesar 7, median adalah 7, modus juga sama yaitu 7. Namun kalau diperhatikan secara seksama, penyebaran data kedua gugus tersebut tidaklah sama, gugus A lebih menyebar dibandingkan gugus B. Untuk memberikan gambaran penyebaran data dari nilai gejala pusatnya, maka diperlukan UKURAN VARIASI/KERAGAMAN atau UKURAN SEBARAN.

Ukuran variasi adalah besaran yang menyatakan jauhnya nilai pengamatan yang sebenarnya menyimpang atau menyebar dari nilai gejala pusatnya.

Macam ukuran variasi yang akan dibahas adalah:

1. Rentang
2. Nilai tengah simpangan
3. Simpangan baku
4. Variasi
5. Koefisien variasi, dan
6. Skor baku

1. Rentang (range)

Rentang (R) adalah selisih antara nilai terbesar (H) dengan nilai terkecil (L) dalam suatu gugus data. Rentang ini memberikan ukuran keragaman yang bersifat kasar atas suatu gugus data. $R = H - L$.

Pada contoh data di atas, ukuran variasi gugus data A adalah $16 - 2 = 14$, sedangkan gugus B adalah $10 - 5 = 5$. Ukuran variasi ini memberikan informasi kepada kita bahwa data pada gugus A lebih menyebar dibandingkan gugus B.

Perhatikan contoh lain pada produksi per hektar kedelai dan kacang tanah di Kabupaten Sragen, mulai tahun 2004 sampai 2008 berikut ini.

Kacang tanah: 1,19; 1,16; 1,13; 1,56; 1,29 (dalam satuan ton/ha)

Kedelai: 1,46; 1,37; 1,44; 1,94; 1,79 (dalam satuan ton/ha)

Rentang produksi/hektar kacang tanah adalah 0,43 sedangkan kedelai 0,57

Untuk data yang telah dikelompokkan dalam distribusi frekuensi nilai rentang (R) didekati dengan

$$R = \text{markah tertinggi} - \text{markah terrendah}$$

Panjang 40 daun salam (mm)

Panjang daun salam(mm)	Markah kelas (X)
117 - 125	121
126 - 134	130
135 - 143	139
144 - 152	148
153 - 161	157
162 - 170	166
171 - 179	175

$$R = 175 - 121$$

$$= 54$$

2. Nilai tengah simpangan (mean deviation - MD)

Nilai tengah simpangan (MD) adalah jumlah harga mutlak penyimpangan setiap data pengamatan terhadap nilai tengahnya, dan dibagi dengan banyaknya pengamatan.

Untuk data yang belum dikelompokkan dalam distribusi frekuensi

$$MD = \frac{\sum |X - \bar{X}|}{n}$$

Contoh produksi per hektar kacang tanah di Kabupaten Sragen tahun 2004 - 2008 adalah 1,19; 1,16; 1,13; 1,56; 1,29

Nilai tengah (\bar{X}) adalah $(1,19 + 1,16 + 1,13 + 1,56 + 1,29) : 5 = 1,266$

$$MD = \frac{|1,19 - 1,266| + |1,16 - 1,266| + |1,13 - 1,266| + |1,56 - 1,266| + |1,29 - 1,266|}{5}$$

$$= \frac{0,076 + 0,106 + 0,136 + 0,294 + 0,024}{5}$$

$$= 0,1272$$

Untuk data yang telah dikelompokkan dalam distribusi frekuensi, nilai tengah simpangan didekati dengan rumus

$$MD = \frac{\sum f|X - \bar{X}|}{n}$$

Contoh

Panjang 40 daun salam (mm)

Panjang daun salam(mm)	Markah kelas (X)	$ X - \bar{X} $	Frekuensi (f)	$f \cdot X - \bar{X} $
117 – 125	121	26,55	2	53,1
126 – 134	130	17,55	3	52,65
135 – 143	139	8,55	10	85,5
144 – 152	148	0,45	13	5,85
153 – 161	157	9,45	6	56,7
162 – 170	166	18,45	4	73,8
171 – 179	175	27,45	2	54,9
J u m l a h			40	382,5

$$MD = \frac{382,5}{40}$$

$$= 9,5625$$

3. Varians dan Simpangan baku

Varians, adalah jumlah kuadrat dari selisih data amatan dengan nilai tengahnya dibagi dengan banyaknya data. Varians dilambangkan dengan s^2 atau σ^2 .

Untuk data yang belum dikelompokkan dalam distribusi frekuensi, varians dihitung dengan rumus

varians populasi :

$$\sigma^2 = \frac{\sum(X - \bar{X})^2}{n} = \frac{\sum x^2}{n}$$

varians sampel :

$$s^2 = \frac{\sum(X - \bar{X})^2}{n - 1} = \frac{\sum x^2}{n - 1}$$

$$x = (X - \bar{X})$$

Contoh untuk produksi per hektar kacang tanah (ton/ha) di Kabupaten Sragen tahun 2004 – 2008 adalah 1,19; 1,16; 1,13; 1,56; 1,29

Nilai tengah (\bar{X}) adalah $(1,19 + 1,16 + 1,13 + 1,56 + 1,29) : 5 = 1,266$

Varians dari produksi kacang tanah pada lima tahun tersebut adalah:

$$\sigma^2 = \frac{(1,19 - 1,266)^2 + (1,16 - 1,266)^2 + (1,13 - 1,266)^2 + (1,56 - 1,266)^2 + (1,29 - 1,266)^2}{5}$$

$$= \frac{(-0,076)^2 + (-0,106)^2 + (-0,136)^2 + (0,294)^2 + (0,024)^2}{5}$$

$$= \frac{0,005776 + 0,011236 + 0,018496 + 0,086436 + 0,000576}{5} = \frac{0,12252}{5} = 0,024504$$

Untuk data yang telah disajikan dalam distribusi frekuensi, maka nilai X diganti dengan markah kelas, varians dihitung dengan rumus

variens populasi :

$$\sigma^2 = \frac{\sum f(X - \bar{X})^2}{n} = \frac{\sum fx^2}{n}$$

variens sampel :

$$s^2 = \frac{\sum f(X - \bar{X})^2}{n - 1} = \frac{\sum fx^2}{n - 1}$$

$$x = (X - \bar{X})$$

Contoh

Panjang 40 daun salam (mm)

Panjang daun salam (mm)	Markah kelas (X)	(X - \bar{X})	(X - \bar{X}) ²	Frek. (f)	f(X - \bar{X}) ²
117 - 125	121	- 26,55	704,9025	2	1409,8050
126 - 134	130	- 17,55	308,0025	3	924,0075
135 - 143	139	- 8,55	73,1025	10	731,0250
144 - 152	148	0,45	0,2025	13	2,6325
153 - 161	157	9,45	89,3025	6	535,8150
162 - 170	166	18,45	340,4025	4	1361,6100
171 - 179	175	27,45	753,5025	2	1507,0050
J u m l a h				40	6471,9000

Varians

$$\sigma^2 = \frac{6471,9}{40} = 161,7975$$

Simpangan Baku

Simpangan baku adalah akar dari varians, dilambangkan dengan s atau σ .

Untuk data yang belum dikelompokkan dalam distribusi frekuensi, simpangan baku dihitung dengan rumus

Simpangan baku populasi :

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{n}} = \sqrt{\frac{\sum x^2}{n}}$$

simpangan baku sampel :

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{\sum x^2}{n - 1}}$$

$$x = (X - \bar{X})$$

Dari contoh produksi per hektar kacang tanah di Kabupaten Sragen tahun 2004 – 2008, maka simpangan bakunya

$$\sigma = \sqrt{\frac{0,12252}{5}} = 0,1565$$

Untuk data yang telah disajikan dalam distribusi frekuensi, nilai X didekati dengan nilai markah kelas, sehingga simpangan bakunya adalah

simpangan baku populasi :

$$\sigma^2 = \sqrt{\frac{\sum f(X - \bar{X})^2}{n}} = \sqrt{\frac{\sum fx^2}{n}}$$

simpangan baku sampel :

$$s^2 = \sqrt{\frac{\sum f(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{\sum fx^2}{n - 1}}$$

$$x = (X - \bar{X})$$

Pada contoh di atas dapat dihitung simpangan bakunya adalah

$$\sigma = \sqrt{\frac{6471,9}{40}} = 12,72$$

Simpangan merupakan ukuran sebaran yang banyak digunakan dalam statistika inferensia.

4. Koefisien Variasi

Ukuran sebaran yang telah dibicarakan di atas tidak dapat digunakan untuk memperbandingkan keragaman dua gugus data atau lebih, karena nilai ukuran sebaran tersebut masih ditentukan juga oleh satuan pengukuran yang digunakan dalam pengamatan. Misalnya pada contoh produksi per hektar kacang tanah (ton/ha) di Kabupaten Sragen tahun 2004 – 2008 adalah 1,19; 1,16; 1,13; 1,56; 1,29 yang ditemukan mempunyai rentang sebesar 0,43. Namun jika menggunakan satuan pengukuran yang berbeda (misalnya ku/ha) maka akan mempunyai rentang sebesar 4,3.

Untuk keperluan melakukan perbandingan, maka digunakan ukuran sebaran yang bebas satuan yaitu koefisien variasi. Koefisien variasi adalah nilai relatif

dari simpangan baku terhadap nilai tengah yang biasanya dinyatakan dalam satuan persen.

$$kv = \frac{\sigma}{\bar{X}} \times 100\%$$

Pada contoh produksi per hektar kacang tanah (ton/ha) di Kabupaten Sragen tahun 2004 – 2008 ditemukan nilai tengah sebesar 1,266 dengan simpangan baku sebesar 0,1565. Pada produksi kedelai ditemukan nilai tengah sebesar 1,6 dengan simpangan baku sebesar 0,2235. Sehingga dapat dihitung koefisien variasi produksi kacang tanah dan kedelai sebagai berikut:

$$kv_{kcg\ th} = \frac{0,1565}{1,266} \times 100\% = 12,36\%$$

$$kv_{kdl} = \frac{0,2235}{1,6} \times 100\% = 13,97\%$$

Jika dilihat dari koefisien variasi maka variasi produksi kedelai sedikit lebih besar dibandingkan produksi kacang tanah.

5. Skor baku

Skor baku merupakan data baru yang diperoleh dari penyimpangan data terhadap nilai tengahnya, dan dinyatakan relatif terhadap simpangan baku.

$$z = \frac{X - \bar{X}}{\sigma}$$

Data baru yang berupa skor baku tersebut, yaitu $z_1, z_2, z_3, \dots, z_n$ merupakan data yang unik, yaitu mempunyai nilai tengah = 0 dan simpangan baku = 1

Dalam penggunaannya, skor baku (z) sering dirubah menjadi distribusi data baru, yang mempunyai nilai tengah \bar{X}_0 dan simpangan baku σ_0

$$z = \bar{X}_0 + \sigma_0 \left(\frac{X - \bar{X}}{\sigma} \right)$$

Skor baku ini sering digunakan untuk membandingkan 2 macam distribusi.

Contoh

Dalam perlombaan tahun 2013, kelompok tani “Suka suka” memperoleh nilai 436. Dari panitia diperoleh informasi bahwa nilai tengah dari seluruh peserta adalah 400 dengan simpangan baku 24. Pada tahun perlombaan tahun sebelumnya, kelompok tani “Suka suka” memperoleh nilai 350. Dari panitia diperoleh informasi bahwa nilai tengah dari seluruh peserta adalah 335 dengan simpangan baku 12. Apakah hasil perlombaan tahun 2013 kelompok tani “Suka suka” lebih baik dari tahun sebelumnya?

Untuk membandingkan 2 distribusi data tersebut digunakan skor baku.

$$z_{2013} = \frac{436 - 412}{24} = 1,00$$

$$z_{2012} = \frac{350 - 335}{12} = 1,25$$

Jadi, peringkat hasil perlombaan tahun 2013 lebih rendah dibandingkan tahun sebelumnya.

Jika kedua distribusi tersebut dinyatakan dalam skor baku, dengan nilai tengah $\bar{X}_0 = 100$ dan simpangan baku $\sigma_0 = 10$, maka skor baku pada dua tahun perlombaan adalah sebagai berikut:

$$z_{2013} = 100 + 10 \left(\frac{436 - 412}{24} \right) = 110$$

$$z_{2012} = 100 + 10 \left(\frac{350 - 335}{12} \right) = 112,5$$

Dari perbandingan skor baku tersebut memberikan informasi bahwa peringkat hasil perlombaan tahun 2013 lebih rendah dibandingkan tahun 2012.

BAB V

DISTRIBUSI PROBABILITAS BINOMIAL, POISSON, DAN NORMAL

Probabilitas/pelemparan merupakan bagian yang tidak dapat dipisahkan dari kehidupan sehari-hari, karena setiap orang selalu berhadapan dengan masalah ketidakpastian (uncertainty). Petani dihadapkan pada masalah berhasil atau gagalnya usahanya yang dikelolanya. Untuk menghasilkan produk komoditas tertentu, petani dihadapkan kepada ketidakpastian yang terjadi pada cuaca, faktor lingkungan biologi, lingkungan sosial, dll. Permasalahan ketidakpastian ini akan dikuantifikasi dengan konsep probabilitas.

Pengertian probabilitas

Andaikan kejadian E dapat terjadi dalam h cara dari seluruh n cara yang mungkin, dan setiap cara ini berkemungkinan sama, maka terjadinya peristiwa tersebut (sukses) adalah

$$Pr(E) = p = \frac{h}{n}$$

Sedangkan peluang tidak terjadinya peristiwa itu (gagal) adalah

$$Pr(\text{bukan } E) = q = \frac{n - h}{n}$$
$$q = 1 - p$$

Dalam pelemparan mata uang logam, kemungkinan didapatkan sisi angka adalah

$$Pr(\text{angka}) = p = \frac{1}{2}$$

Sedangkan peluang didapatkan sisi gambar adalah

$$Pr(\text{gambar}) = 1 - \frac{1}{2} = \frac{1}{2}$$

Peristiwa munculnya sisi angka atau sisi gambar dinamakan dengan titik sampel. Himpunan titik sampel dinamakan dengan ruang sampel.

Dalam pelemparan dadu, titik sampelnya adalah munculnya mata 1, mata 2, mata 3, mata 4, mata 5, atau mata 6.

Pendekatan Peluang

- a. Pendekatan klasik, menyatakan bahwa terjadinya suatu peristiwa, E adalah rasio antara peristiwa E tersebut dengan seluruh peristiwa yang mungkin terjadi apabila setiap peristiwa mempunyai peluang sama.

$$\Pr(E) = p = \frac{h}{n}$$

- b. Pendekatan frekuensi relatif, adalah pendekatan yang menggunakan perhitungan frekuensi relatif yang didasarkan pada terjadinya peristiwa yang lalu sebagai peluang. Contohnya, dari pengamatan 40 daun salam, ternyata banyaknya daun yang mempunyai luas antara 117 – 126 m² adalah sebanyak 2 lembar. Peluang satu lembar daun salam mempunyai luas antara 117 – 126 m² adalah sebesar

$$\begin{aligned} \Pr(117 - 126) &= \frac{2}{40} \\ &= 0,05 \end{aligned}$$

- c. Pendekatan subyektif, adalah pendekatan yang didasarkan pada tingkat kepercayaan individu yang membuat dugaan terhadap suatu peluang. Individu yang melakukan pendekatan pada umumnya telah memiliki sejumlah pengalaman berkaitan dengan peristiwa yang diduganya, sehingga mampu membuat dugaan dengan baik. Misalnya, pedagang sayur mayur berkeyakinan bahwa kemungkinan besar permintaan sayur mayur akan mengalami peningkatan.

Distribusi Binomial

Adalah distribusi probabilitas yang hanya mempunyai 2 peristiwa yang mungkin terjadi. Misalnya dalam pelemparan mata uang, peristiwa yang mungkin terjadi adalah munculnya sisi gambar atau sisi angka. Dalam pelaksanaan inseminasi buatan terhadap ternak sapi, maka peristiwa yang mungkin terjadi adalah berhasil ataukah gagal. Jika petani mempunyai seekor sapi yang sedang bunting, maka jenis kelamin sapi anakan yang akan diperoleh adalah jantan atau betina.

Jika p adalah probabilitas suatu peristiwa yang akan terjadi (dan q adalah probabilitas tidak terjadinya suatu peristiwa) dalam suatu percobaan tunggal, maka probabilitas akan terjadi sebanyak X peristiwa dari percobaan sebanyak N adalah dinyatakan:

$$\Pr(X) = {}_N C_X \cdot p^X (1 - p)^{N-X} = \frac{N!}{X!(N-X)!} p^X (1 - p)^{N-X}$$

$$X = 1, 2, 3, \dots, N$$

$$N! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot N \quad \text{misalnya } 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24 \quad 0! = 1$$

Sifat-sifat distribusi binomial adalah

- Nilai tengah suatu distribusi binomial adalah $\mu = Np$
- Simpangan baku suatu distribusi binomial adalah $\sigma = \sqrt{Np(1-p)}$
- Varians distribusi binomial adalah $\sigma^2 = Np(1-p)$

Contoh 1: Probabilitas memperoleh “sisi angka” (X) sebanyak 2 dalam pelantunan mata uang (N) 6 kali, dapat dicari dengan probabilitas binomial sebagai berikut

$$\begin{aligned} Pr(X = 2) &= {}_6C_2 \cdot 0,5^2(1 - 0,5)^{6-2} \\ &= \frac{6!}{2!(6-2)!} 0,5^2(1 - 0,5)^{6-2} \\ &= \frac{720}{2(24)} (0,25)(0,0625) = 0,234375 \end{aligned}$$

Dalam kasus tersebut peristiwa munculnya sisi angka mempunyai peluang yang sama dengan munculnya sisi gambar, sehingga probabilitas tunggalnya adalah 0,5.

Contoh 2:

Jika 20% dari buah durian yang dihasilkan adalah cacat, hitunglah probabilitas bahwa dari 4 buah durian yang dipilih secara acak, akan didapati buah durian cacat (a) sebanyak 1 buah, (b) paling sedikit 2 buah.

Dari pernyataan dapat dinyatakan bahwa probabilitas buah durian cacat (p) adalah 0,2; banyaknya buah (percobaan/ N) adalah 4.

Jadi peluang buah durian akan cacat sebanyak 1, $Pr(X=1)$, adalah

$$\begin{aligned} Pr(X = 1) &= {}_4C_1 \cdot 0,2^1(1 - 0,2)^{4-1} \\ &= \frac{4!}{1!(4-1)!} 0,2^1(1 - 0,2)^{4-1} \\ &= 0,4096 \end{aligned}$$

Jadi peluang akan didapati buah durian cacat sebanyak 1 buah adalah sebesar 0,4096

Yang dimaksud didapati buah durian cacat paling sedikit 2 buah adalah didapati buah durian cacat paling sedikit 2, 3 atau 4 buah.

Berarti $Pr(X= 2, 3, 4) = 1 - (Pr(X=1) + Pr(X=0))$

$$Pr(X = 0) = {}_4C_0 \cdot 0,2^0(1 - 0,2)^{4-0} = 0,4096$$

$$\begin{aligned} \text{Dengan demikian } Pr(X= 2, 3, 4) &= 1 - (0,4096 + 0,4096) \\ &= 0,1808 \end{aligned}$$

Distribusi Poisson

Distribusi Poisson merupakan turunan dari distribusi binomial, yang secara khusus berlaku untuk kasus yang mempunyai ukuran populasi (N) yang sangat besar dan mempunyai peluang terjadinya suatu peristiwa (p) teramat kecil, sehingga disebut kejadian langka.

Dalam prakteknya kita menganggap suatu kejadian langka jika ukuran sampel (N) sekurang-kurangnya 50 ($N \geq 50$) dan Np tidak lebih dari 5 ($Np \leq 5$)

Distribusi probabilitas Poisson dirumuskan

$$Pr(X) = \frac{\lambda^X e^{-\lambda}}{X!}$$

λ = konstanta Poisson, nilainya $\lambda = \mu = Np$

$e = 2,71828$

Distribusi Poisson mempunyai sifat

Beberapa sifat distribusi Poisson adalah sebagai berikut:

- Nilai tengah suatu distribusi binomial adalah $\lambda = \mu = Np$
- Simpangan baku suatu distribusi binomial adalah $\sigma = \sqrt{\lambda}$
- Varians distribusi binomial adalah $\sigma^2 = \lambda$

Contoh :

Probabilitas ditolaknya pasokan daging ayam pada suatu supermarket adalah 0,001. Tentukan probabilitas bahwa dari 4.000 potong daging ayam yang disetor ke supermarket akan ditolak sebanyak (a) 3 potong ; (b) tidak ada yang ditolak.

Nilai $\lambda = Np$

$$= 4.000 \times 0,001 = 4$$

a. Probabilitas daging ayam akan ditolak sebanyak 3 potong adalah

$$\begin{aligned} Pr(X = 3) &= \frac{4^3 2,71828^{-4}}{4!} \\ &= 0,04884 \end{aligned}$$

b. Probabilitas daging ayam tidak ada yang ditolak adalah

$$\begin{aligned} Pr(X = 0) &= \frac{4^0 2,71828^{-4}}{4!} \\ &= 0,0007632 \end{aligned}$$

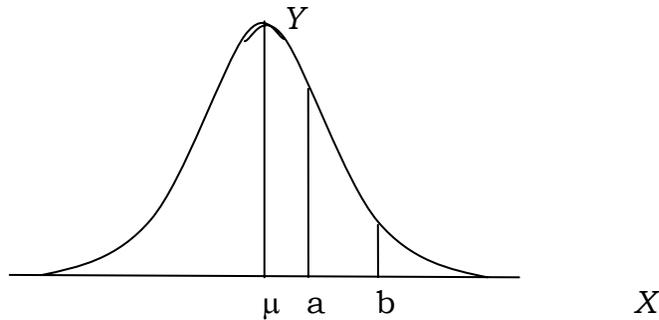
Distribusi Normal

Distribusi normal merupakan distribusi peluang kontinyu, yang didefinisikan dengan persamaan sebagai berikut:

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2/\sigma^2}$$

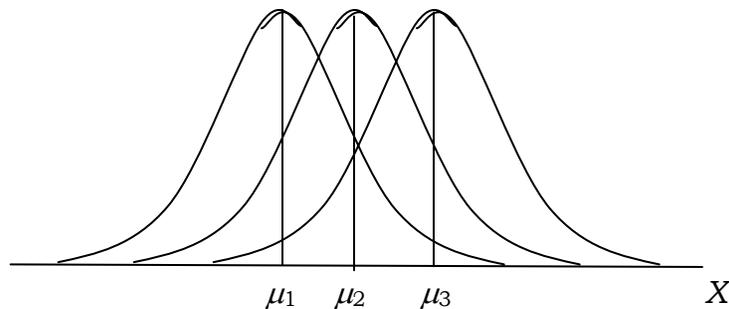
dimana μ = nilai tengah, σ = simpangan baku, $\pi=3,14159$; $e = 2,71828$

Hubungan antara Y dengan X secara grafis dapat digambarkan sebagai berikut:



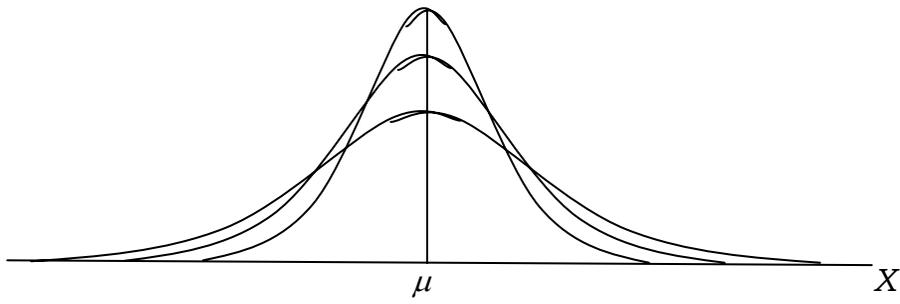
Luas daerah yang dibatasi oleh kurva dengan sumbu X adalah satu, yang menunjukkan peluang. Peluang X mempunyai nilai sebesar a sampai b , yang dinyatakan $Pr \{a < X < b\}$ ditunjukkan oleh luas di bawah kurva antara $X=a$ sampai $X=b$

Bentuk kurva normal tergantung dari nilai tengah (μ) dan simpangan bakunya (σ). Kurva norma yang mempunyai nilai tengah (μ) yang berbeda diilustrasi pada gambar sebagai berikut:



Kurva normal yang mempunyai nilai tengah (μ) berbeda namun simpangan bakunya (σ) yang sama, maka bentuk ketiga kurva normalnya sama namun terjadi pergeseran letak kurva normal.

Kurva normal yang mempunyai nilai tengah (μ) yang sama namun simpangan bakunya (σ) berbeda diilustrasikan pada gambar sebagai berikut:

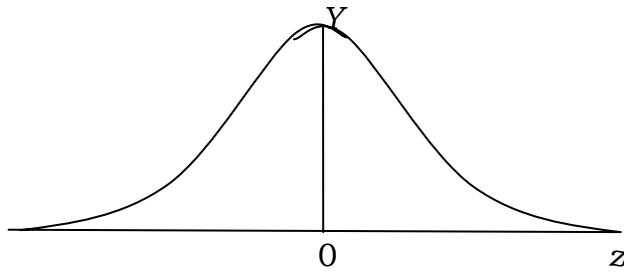


Kurva normal yang mempunyai simpangan baku (σ) yang lebih kecil mempunyai bentuk kurva yang lebih tinggi dan penyebaran data X lebih rendah, sedangkan yang mempunyai simpangan baku (μ) lebih besar, bentuk kurvanya lebih rendah dan penyebaran data X lebih besar.

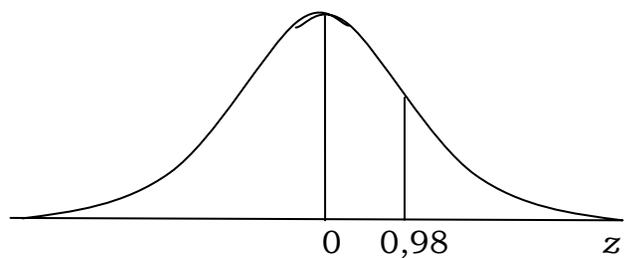
Nilai $(X-\mu)/\sigma$ merupakan skor baku (z) yang mempunyai nilai tengah nol dan simpangan baku satu. Distribusi normal yang nilai X diganti dengan satuan skor baku (z) dinamakan distribusi normal baku, yang dinyatakan dalam fungsi sebagai berikut

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Bentuk kurva normal baku diilustrasikan sebagai berikut:



Luas daerah di bawah kurva normal baku dapat dihitung dengan mengintegalkan fungsi normal baku di atas. Hasil integral fungsi normal baku disajikan dalam tabel “luas kurva normal baku”.



Luas daerah dibawah kurva normal baku antara $z=0$ dan $z=0,98$ dapat dicari dari tabel “luas kurva normal baku” sebagai berikut:

Z	0	1	2		8	9
0,0	0,0000	0,0040	0,0080			
0,1						
0,9					0,3365	

Tabel “luas kurva normal baku” menunjukkan luas daerah dibawah kurva normal baku antara $z=0$ dan $z=0,98$ adalah 0,3365.

Penerapan distribusi normal

Distribusi normal baku digunakan untuk menghitung banyaknya nilai X antara $X=a$ sampai $X=b$ dalam suatu distribusi data yang mempunyai nilai tengah sebesar μ dan simpangan baku sebesar σ .

Misalkan, suatu data tentang hasil panen padi yang dihasilkan 300 orang petani menyebar mengikuti distribusi normal dengan nilai tengah $\mu = 50$ (ku/ha) dan simpangan baku $\sigma = 20$. Berdasarkan informasi ini, carilah:

- Peluang petani menghasilkan panen padi antara 50 sampai 60 (ku/ha)
- Banyaknya petani yang mempunyai hasil panen padi antara 50 sampai 60 (ku/ha)
- Peluang hasil panen padi lebih dari dari 75 (ku/ha)
- Banyaknya petani yang mempunyai hasil panen lebih dari 75 (ku/ha)

Penyelesaian soal a:

- Rubahlah 50 dan 60 dalam bentuk skor baku dengan rumus sebagai berikut

$$z = \frac{X - \mu}{\sigma}$$

$$z_{50} = \frac{50 - 50}{20} = 0$$

$$z_{60} = \frac{60 - 50}{20} = 0,5$$

- Luas daerah di bawah kurva normal antara $z=0$ dan $z=0,5$ diketahui dari tabel luas kurva normal baku, yaitu sebesar 0,1915

Jadi peluang petani menghasilkan panen padi antara 50 sampai 60 (ku/ha) adalah sebesar 0,1915

Penyelesaian soal b:

Banyaknya petani yang mempunyai hasil panen padi antara 50 sampai 60 (ku/ha) merupakan peluang dikalikan banyaknya petani padi, sehingga diperoleh angka $0,1915 \times 300 = 57,45$.

Jadi banyaknya petani yang mempunyai hasil panen padi antara 50 sampai 60 (ku/ha) adalah 57 orang.

Penyelesaian soal c:

Nilai z untuk $X = 75$ adalah

$$z_{75} = \frac{75 - 50}{20} = 1,25$$

Kita perlu memahami bahwa tabel luas kurva norma baku hanya menunjukkan luas kurva antara $z=0$ sampai $z=1,25$. Luas daerah di bawah kurva normal antara $z=0$ sampai $z=1,25$ adalah 0,3944. Sedangkan peluang hasil panen padi lebih dari dari 75 (ku/ha) pada soal nomor "c" menunjukkan luas daerah di bawah kurva normal dari $z=1,25$ sampai tak terbatas. Oleh karena, luas seluruh daerah di bawah kurva normal adalah 1 maka luas separoh daerah di bawah kurva normal adalah 0,5. Sehingga, untuk penyelesaian soal "c" adalah $0,5 - 0,3944 = 0,1056$. Jadi peluang hasil panen padi lebih yang dari dari 75 (ku/ha) adalah 0,1056.

Penyelesaian soal d:

Banyaknya petani yang mempunyai hasil panen lebih dari 75 (ku/ha) adalah $0,1056 \times 300 = 31,68 \approx 31$ orang.

Soal Latihan

1. Hitunglah
 - a. ${}_6C_2$
 - b. ${}_8C_3$
 - c. ${}_7C_5$
 - d. ${}_{15}C_{13}$
 - e. ${}_{300}C_{299}$
2. Carilah probabilitas (peluang) bahwa pada pelantunan 5 keping mata uang logam yang seimbang akan muncul
 - a. 1 sisi angka
 - b. 2 sisi angka
 - c. 3 sisi angka
 - d. paling banyak 3 sisi angka
 - e. paling sedikit 4 sisi angka
3. Peluang diterimanya seorang calon karyawan dalam sebuah wawancara di perusahaan adalah 0,4. Tentukan peluang dari 8 orang pelamar

- a. Tidak ada yang diterima
 - b. Diterima semua
 - c. Diterima sebanyak 6 orang
 - d. Diterima paling sedikit 7 orang
 - e. Diterima paling banyak 2 orang
4. Seorang agen asuransi menjual polis kepada 4 orang wanita dengan usia dan kondisi kesehatan yang sehat semua. Sesuai tabel aktuaria, peluang wanita tersebut tetap hidup selama 30 tahun selanjutnya adalah 0,6. Pada 30 tahun yang akan datang, hitunglah peluang dari 4 orang wanita pembeli polis akan didapati yang masih hidup
- a. Sebanyak 3 orang
 - b. Sebanyak 2 orang
 - c. Tidak ada yang masih hidup
 - d. Paling banyak 3 orang
 - e. Semua masih hidup
5. Peluang masih bertahannya suatu serangga yang disemprot dengan pestisida merek “pestiku” adalah 0,0025. Suatu kelompok serangga diperkirakan sebanyak 4.000 ekor disemprot dengan pestisida “pestiku”, hitunglah peluang
- a. Tidak ditemukan satu pun serangga yang hidup
 - b. Ditemukan 4 ekor serangga yang hidup
 - c. Ditemukan sebanyak-banyaknya 2 ekor serangga yang hidup
6. Dari tabel aktuaria diperoleh peluang orang yang berusia 30 tahun akan meninggal pada usia 31 adalah 0,0014. Jika perusahaan asuransi mempunyai sebanyak 1000 orang pemegang polis yang berusia 30 tahun, tentukan dari pemegang polis tersebut akan ada yang meninggal pada usia 31 tahun, sebanyak
- a. 5 orang
 - b. Kurang dari 3 orang
 - c. Antara 3 dan 5 orang
7. Pada tes akademik pada penyaringan calon karyawan perusahaan didapatkan nilai rata-rata sebesar 72 dengan simpangan baku sebesar 15.

Hitunglah skor baku (z), bagi calon karyawan yang mempunyai nilai sebesar:

- a. 60 b. 67 c. 75 d. 80 e. 85

8. Dengan data seperti nomor 5, hitunglah nilai calon karyawan yang mempunyai skor baku sebesar:

- a. -0,6 b. 0,6 c. -1,2 d. 1,2 e. 0

9. Hitunglah luas daerah di bawah kurva normal baku (gunakan tabel z yang disediakan)

- a. Antara $z = 0$ dan $z = 1,2$
b. Antara $z = 0$ dan $z = -1,86$
c. Antara $z = -1,86$ dan $z = 1,2$
d. Antara $z = -1,86$ dan $z = -1,2$
e. Kurang dari $z = -1,86$
f. Lebih dari $z = -1,86$
g. Lebih dari $z = 1,86$
h. Kurang dari $z = -1,86$ dan lebih dari $z = 1,86$

10. Tentukan nilai z , yang luas daerah di bawah kurva normal bakunya adalah sebagai berikut:

- a. Luas antara 0 dan z adalah 0,4382
b. Luas daerah di sebelah kiri z adalah 0,8508

11. Nilai tengah panjang dari 400 helai daun salam pada suatu ikatan adalah 151 mm dan simpangan bakunya 15 mm. Dengan menganggap bahwa panjang daun salam berdistribusi normal, carilah berapa persen banyaknya daun salam yang mempunyai panjang:

- a. antara 151 dan 175
b. antara 130 dan 151
c. kurang dari 130
d. lebih dari 146
e. antara 160 dan 180
f. antara 135 dan 170

12. Manajemen suatu supermarket menetapkan berat lele minimal yang dapat dipasok untuk supermarket tersebut adalah 175 gram. Seorang petani mempunyai 2000 ekor lele, yang mempunyai nilai tengah berat

sebesar 220 gram dengan simpangan baku 30 gram. Hitunglah banyaknya lele yang dapat dipasok kepada manajemen supermarket !

13. Petani mendapati produksi melon yang dihasilkannya mempunyai berat rata-rata 1,6 kg dengan simpangan baku 0,25 kg.
 - a. Jika ditetapkan grade A adalah melon yang beratnya adalah 1,8 kg, hitunglah berapa persen melon petani tersebut yang masuk grade A
 - b. Jika ditetapkan grade B adalah melon yang beratnya antara 1,35 dan 1,75, hitunglah berapa persen melon petani tersebut yang masuk grade B
 - c. Jika grade C adalah melon yang beratnya antara 1,00 dan 1,30, hitunglah berapa persen melon petani tersebut yang masuk grade C
 - d. Berapa persen banyaknya melon yang tidak lolos grading?

BAB VI

PENDUGAAN

Salah satu aspek penting dalam statistika inferensia adalah melakukan pendugaan terhadap parameter populasi dengan menggunakan statistik yang diperoleh dari sampel. Misalnya suatu perusahaan eceran ingin mengetahui tingkat belanja keluarga masyarakat di sekitar lokasi pendirian perusahaan ritel, maka perusahaan tersebut akan mengambil sampel keluarga yang hasilnya akan digunakan untuk melakukan pendugaan tingkat belanja keluarga masyarakat di sekitar lokasi pendirian secara keseluruhan.

Penduga yang baik adalah penduga yang nilai statistik sampel sama dengan nilai parameter populasi. Misalnya nilai tengah sampel (\bar{X}) sama dengan nilai tengah populasi μ , maka dikatakan nilai tengah sampel merupakan penduga yang baik bagi nilai tengah populasi.

Apabila terdapat dua distribusi data dari dua sampel yang berbeda, yang mempunyai nilai tengah yang sama namun mempunyai varians yang berbeda, maka dikatakan sampel yang mempunyai varians yang lebih kecil disebut penduga yang efisien, sedangkan yang mempunyai varians yang lebih besar disebut penduga yang kurang efisien. Pada kenyataannya sampel yang mempunyai varians yang besar pun digunakan dalam statistika inferensia, karena lebih mudah diperoleh.

Pendugaan terhadap parameter populasi dapat dilakukan pada suatu nilai tertentu yang disebut dengan penduga titik suatu parameter. Misalnya tingkat belanja keluarga masyarakat di sekitar perusahaan eceran diduga sebesar Rp 2,5 juta per bulan. Namun dapat juga dilakukan pada suatu rentang nilai tertentu yang disebut penduga interval suatu parameter, misalnya tingkat belanja keluarga masyarakat di sekitar perusahaan masyarakat di sekitar perusahaan eceran diduga berkisar antara Rp 2,4 juta dan Rp 2,6 juta atau diduga Rp 2,5 juta \pm Rp 0,1 juta.

Dalam membuat pendugaan, dapat ditentukan tingkat ketepatan/kepercayaan terhadap penduga interval suatu parameter tersebut, sehingga lebih baik dibandingkan dengan penduga titik. Misalnya tingkat belanja keluarga masyarakat di sekitar perusahaan masyarakat di sekitar perusahaan

eceran diduga berkisar antara Rp 2,4 juta dan Rp 2,6 juta misalnya dengan tingkat ketepatan/kepercayaan sebesar 90%.

Sampel berukuran besar ($n \geq 30$) yang diambil dari populasi yang berdistribusi normal, menghasilkan nilai tengah (\bar{X}) dan simpangan baku (s), maka dapat diharapkan nilai tengah populasi μ akan mempunyai nilai antara $\bar{X} - s$ dan $\bar{X} + s$ dengan tingkat ketepatan/kepercayaan sebesar 68,27%, secara matematis dinyatakan

$$\Pr\{\bar{X} - s < \mu < \bar{X} + s\} = 68,27\%$$

$\bar{X} - s$ dinamakan batas bawah interval kepercayaan, dan $\bar{X} + s$ dinamakan batas atas interval kepercayaan, sedangkan 68,27% dinamakan tingkat kepercayaan.

Nilai tengah populasi tersebut akan mempunyai nilai antara $\bar{X} - 2s$ dan $\bar{X} + 2s$ dengan tingkat ketepatan/kepercayaan sebesar 95,45%, serta nilai tengah populasi akan mempunyai nilai antara $\bar{X} - 3s$ dan $\bar{X} + 3s$ dengan tingkat ketepatan/kepercayaan sebesar 99,73%.

Batas bawah untuk interval kepercayaan 95% adalah $\bar{X} - 1,96s$ dan batas atasnya adalah $\bar{X} + 1,96s$, sedangkan untuk interval kepercayaan 99% mempunyai batas bawah $\bar{X} - 2,58s$ dan mempunyai batas atas $\bar{X} + 2,58s$. Bilangan 1,96 ; 2,58 masing-masing disebut dengan koefisien kepercayaan 95% dan 99% . Koefisien kepercayaan yang lain dapat dilihat dari tabel distribusi normal baku (distribusi z).

Tingkat kepercayaan (c)	99,73%	99%	98%	96%	95,45%	95%	90%	80%
Koef kepercayaan (z_c)	3,00	2,58	2,33	2,05	2,00	1,96	1,645	1,28

1. Pendugaan rata-rata

Apabila nilai tengah sampel adalah \bar{X} dan simpangan bakunya adalah s , maka simpangan baku populasi (σ) dianggap sama dengan s . Untuk sampel yang diambil dari populasi yang tak hingga ($N \rightarrow \infty$), maka batas bawah interval kepercayaan untuk tingkat kepercayaan sebesar c dinyatakan sebagai

$$\bar{X} - z_c \frac{\sigma}{\sqrt{n}}$$

dan batas atasnya dinyatakan sebagai

$$\bar{X} + z_c \frac{\sigma}{\sqrt{n}}$$

z_c = koefisien kepercayaan pada tingkat kepercayaan c

σ = simpangan baku populasi, yang nilainya sama dengan s

n = ukuran sampel

Contoh:

Suatu sampel acak diambil untuk melakukan pendugaan terhadap berat salak pondoh yang dibudidayakan di Sleman. Dari sampel sebanyak 150 buah salak pondoh, diperoleh nilai tengah berat per buah adalah 60 gr dengan simpangan baku sebesar 15 gr. Hitunglah batas interval kepercayaan untuk tingkat kepercayaan (a) 90% dan (b) 95% bagi berat buah salak pondoh di Sleman.

Jawab:

Diketahui:

$$n = 150$$

$$\bar{X} = 60 \quad \sigma = s = 15$$

a. Interval kepercayaan bagi berat buah salak pondoh di Sleman dengan tingkat kepercayaan 90%.

Dari tabel distribusi normal baku, diperoleh nilai $z_{90\%} = 1,645$

Batas bawah interval kepercayaan

$$= 60 - 1,645 \frac{15}{\sqrt{150}}$$

$$= 60 - 2,01 = 57,99$$

Batas atas interval kepercayaan

$$= 60 + 1,645 \frac{15}{\sqrt{150}}$$

$$= 60 + 2,01 = 62,01$$

Berat salak pondoh per buah di Sleman diduga berkisar antara 57,99 gr dan 62,01 gr dengan tingkat kepercayaan sebesar 90%.

b. Interval kepercayaan bagi berat buah salak pondoh di Sleman dengan tingkat kepercayaan sebesar 95%.

Dari tabel distribusi normal baku, diperoleh nilai $z_{95\%} = 1,96$

Batas bawah interval kepercayaan

$$= 60 - 1,96 \frac{15}{\sqrt{150}}$$

$$= 60 - 2,40 = 57,60$$

Batas atas interval kepercayaan

$$= 60 + 1,96 \frac{15}{\sqrt{150}}$$

$$= 60 + 2,40 = 62,40$$

Berat salak pondoh per buah di Sleman diduga berkisar antara 57,60 gr dan 62,40 gr dengan tingkat kepercayaan sebesar 95%

Untuk sampel yang diambil dari populasi yang berhingga, yaitu sebesar N , maka batas bawah interval kepercayaan untuk tingkat kepercayaan sebesar c dinyatakan:

$$\bar{X} - z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

dan batas atasnya dinyatakan sebagai

$$\bar{X} + z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

\bar{X} = nilai tengah sampel

z_c = koefisien kepercayaan pada tingkat kepercayaan c

σ = simpangan baku populasi, yang nilainya sama dengan s

n = ukuran sampel

N = ukuran populasi

Contoh:

Sebuah sampel acak sebesar 100 orang mahasiswa diambil dari 800 mahasiswa Agribisnis UMY. Dari sampel tersebut didapatkan informasi nilai tengah pengeluaran untuk konsumsi makanan di kampus sebesar Rp 12.500 dengan simpangan baku Rp 5.000. Hitunglah batas interval kepercayaan bagi nilai tengah pengeluaran untuk konsumsi makanan di kampus dengan tingkat kepercayaan (a) 99%; (b) 95%

Jawab:

Diketahui

$$n = 100 \qquad N = 800$$

$$\bar{X} = 12.500 \qquad \sigma = 5.000$$

a. Batas interval kepercayaan dengan tingkat kepercayaan 99%

Dari tabel kurva normal diketahui $z_{99\%} = 2,58$

Batas bawah interval kepercayaan dengan tingkat kepercayaan 99%

$$= 12.500 - 2,58 \frac{5.000}{\sqrt{100}} \sqrt{\frac{800 - 100}{800 - 1}}$$

$$= 12.500 - 1.207,44 = 11.292,56$$

Batas atas interval kepercayaan dengan tingkat kepercayaan 99%

$$= 12.500 + 2,58 \frac{5.000}{\sqrt{100}} \sqrt{\frac{800 - 100}{800 - 1}}$$

$$= 12.500 + 1.207,44 = 13.707,44$$

Jadi dapat dikatakan bahwa nilai tengah pengeluaran 800 orang mahasiswa Agribisnis UMY untuk konsumsi makanan di kampus adalah berkisar antara Rp 11.292,56 dan Rp 13.707,44 dengan tingkat ketepatan/kepercayaan sebesar 99%.

b. Batas interval kepercayaan dengan tingkat kepercayaan 95%

Dari tabel kurva normal dapat diketahui bahwa $z_{95\%} = 1,96$

Batas bawah interval kepercayaan dengan tingkat kepercayaan 95%

$$= 12.500 - 1,96 \frac{5.000}{\sqrt{100}} \sqrt{\frac{800 - 100}{800 - 1}}$$

$$= 12.500 - 917,28 = 11.582,72$$

Batas atas interval kepercayaan dengan tingkat kepercayaan 95%

$$= 12.500 + 1,96 \frac{5.000}{\sqrt{100}} \sqrt{\frac{800 - 100}{800 - 1}}$$

$$= 12.500 + 917,28 = 13.417,28$$

Nilai tengah pengeluaran 800 orang mahasiswa Agribisnis UMY untuk konsumsi makanan di kampus adalah berkisar antara Rp 11.582,72 dan Rp 13.417,28 dengan tingkat ketepatan/ kepercayaan sebesar 95%

2. Pendugaan bagi proporsi

Diambil sampel dengan ukuran n , dan ternyata diperoleh kejadian “sukses” sebanyak X , maka proporsi “sukses” dalam sampel tersebut adalah

$$p = \frac{X}{n}$$

Jika sampel diambil dari populasi tak berhingga, maka batas bawah interval kepercayaan pada tingkat kepercayaan sebesar c adalah

$$p - z_c \sqrt{\frac{p \cdot q}{n}}$$

dan batas atas interval kepercayaan adalah

$$p + z_c \sqrt{\frac{p \cdot q}{n}}$$

p = proporsi sukses

z_c = koefisien kepercayaan pada tingkat kepercayaan c

q = proporsi bukan sukses ($1 - p$)

n = ukuran sampel

Contoh:

Suatu sampel sebanyak 200 buah jeruk diambil secara acak digunakan untuk menduga proporsi buah yang mempunyai kulit cacat/burik. Dari pengamatan diperoleh data bahwa 30 buah diantara sampel tersebut ternyata kulitnya cacat. Hitunglah interval kepercayaan bagi pendugaan proporsi buah jeruk yang berkulit cacat dengan tingkat kepercayaan (a) 99% dan (b) 90%

Jawab:

Diketahui

$$n = 200$$

$$X = 30$$

$$p = 30/200 = 0,15$$

a. Interval kepercayaan bagi proporsi buah jeruk yang berkulit cacat dengan tingkat kepercayaan 99%.

Dari tabel normal diketahui bahwa $z_{99\%} = 2,58$

Batas bawah interval kepercayaan

$$= 0,15 - 2,58 \sqrt{\frac{0,15 \cdot (1 - 0,15)}{200}}$$

$$= 0,15 - 2,58 (0,0252) = 0,0849$$

Batas atas interval kepercayaan

$$= 0,15 + 2,58 \sqrt{\frac{0,15 \cdot (1 - 0,15)}{200}}$$

$$= 0,15 + 2,58 (0,0252) = 0,2151$$

Jadi proporsi buah jeruk yang berkulit rusak diduga sebesar antara 0,0849 dan 0,2151 dengan tingkat kepercayaan 99%.

b. Interval kepercayaan bagi proporsi buah jeruk yang berkulit cacat dengan tingkat kepercayaan 90%

Dari tabel normal baku diketahui bahwa $z_{90\%} = 1,645$

Batas bawah interval kepercayaan

$$= 0,15 - 1,645 \sqrt{\frac{0,15 \cdot (1 - 0,15)}{200}}$$
$$= 0,15 - 1,645 (0,0252) = 0,1085$$

Batas atas interval kepercayaan

$$= 0,15 + 1,645 \sqrt{\frac{0,15 \cdot (1 - 0,15)}{200}}$$
$$= 0,15 + 1,645 (0,0252) = 0,1915$$

Jika sampel diambil dari populasi berhingga dengan ukuran N , maka batas bawah interval kepercayaan pada tingkat kepercayaan sebesar c adalah

$$p - z_c \sqrt{\frac{p \cdot q}{n}} \sqrt{\frac{N - n}{N - 1}}$$

dan batas atas interval kepercayaan adalah

$$p + z_c \sqrt{\frac{p \cdot q}{n}} \sqrt{\frac{N - n}{N - 1}}$$

p = proporsi sukses

q = proporsi bukan sukses ($1 - p$)

n = ukuran sampel

N = ukuran populasi

Contoh:

Pedagang akan membeli 1 keranjang jeruk yang diperkirakan banyaknya adalah 600 buah. Untuk menduga proporsi jeruk yang berkulit rusak, dari jeruk yang akan dibeli tersebut, pedagang mengambil sampel sebanyak 50 buah, dan didapatkan 8 diantaranya berkulit rusak. Hitunglah interval kepercayaan bagi dugaan proporsi jeruk yang berkulit rusak dari sejumlah jeruk yang akan dibeli oleh pedagang tersebut, dengan tingkat kepercayaan 95%.

Jawab:

Diketahui

$N = 600$

$$n = 50$$

$$X = 8$$

$$p = X/n$$

$$= 8/50 = 0,16$$

Dari tabel kurva normal dapat diketahui bahwa $z_{95\%} = 1,96$

Interval kepercayaan bagi dugaan proporsi jeruk yang berkulit rusak

Batas bawah

$$= 0,16 - 1,96 \sqrt{\frac{0,16 \cdot (1 - 0,16)}{50}} \sqrt{\frac{600 - 50}{600 - 1}}$$

$$= 0,16 - 1,96 (0,0518)(0,9582)$$

$$= 0,16 - 0,0974 = 0,0626$$

Batas atas

$$= 0,16 + 1,96 \sqrt{\frac{0,16 \cdot (1 - 0,16)}{50}} \sqrt{\frac{600 - 50}{600 - 1}}$$

$$= 0,16 + 1,96 (0,0518)(0,9582)$$

$$= 0,16 + 0,0974 = 0,2574$$

Jadi, diduga dari 600 butir jeruk yang akan dibeli oleh pedagang, 6,26% sampai 25,74% diantaranya berkulit rusak dengan tingkat kepercayaan/ketepatan 95%.

3. Pendugaan bagi selisih nilai tengah

Diambil sampel ke-1 yang berukuran n_1 dari suatu populasi ke-1 dan ternyata diperoleh nilai tengah sampel \bar{X}_1 dan simpangan bakunya s_1 , maka simpangan baku populasi σ_1 dianggap sama dengan s_1 . Selanjutnya diambil sampel ke-2 dari populasi lain, yaitu populasi ke-2 dan ternyata diperoleh nilai tengah sampel \bar{X}_2 dan simpangan bakunya s_2 , maka simpangan baku populasi σ_2 dianggap sama dengan s_2 .

Jika sampel tersebut diambil dari populasi tak berhingga, maka batas bawah interval kepercayaan selisih nilai tengah kedua populasi tersebut adalah

$$(\bar{X}_1 - \bar{X}_2) - z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

sedangkan batas atasnya adalah

$$(\bar{X}_1 - \bar{X}_2) + z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

\bar{X} = nilai tengah sampel

z_c = koefisien kepercayaan pada tingkat kepercayaan c

σ = simpangan baku populasi, yang nilainya sama dengan s

n = ukuran sampel

N = ukuran populasi

Contoh:

Di pasar lelang agribisnis didapati 2 kelompok salak, yaitu salak kualitas ekspor dan salak kualitas supermarket. Untuk melakukan pendugaan perbedaan berat salak dari kedua kelompok tersebut diambil sampel dari masing-masing kelompok. Dari salak kualitas ekspor diambil sampel sebanyak 40 buah salak, dan didapatkan nilai tengah berat salak sebesar 80 gr/buah dengan simpangan baku 10 gr/buah. Dari salak kualitas supermarket diambil sampel sebanyak 50 buah salak, dan didapatkan nilai tengah berat salak sebesar 70 gr/buah dengan simpangan baku 15 gr/buah. Hitunglah interval kepercayaan dugaan perbedaan berat antara salak kualitas ekspor dan salak kualitas supermarket dengan tingkat kepercayaan 90%.

Jawab:

Diketahui

$$n_1 = 40$$

$$n_2 = 50$$

$$\bar{X}_1 = 80$$

$$\bar{X}_2 = 70$$

$$s_1 = \sigma_1 = 10$$

$$s_2 = \sigma_2 = 15$$

Dari tabel kurva normal dapat diketahui bahwa $z_{90\%} = 1,645$

Interval kepercayaan dugaan perbedaan berat antara salak kualitas ekspor dan salak kualitas supermarket

Batas bawah

$$\begin{aligned} &= (80 - 70) - 1,645 \sqrt{\frac{10^2}{40} + \frac{15^2}{50}} \\ &= 10 - 1,645 (2,6458) \end{aligned}$$

$$= 10 - 4,3523 = 5,6477$$

Batas atas

$$= (80 - 70) + 1,645 \sqrt{\frac{10^2}{40} + \frac{15^2}{50}}$$

$$= 10 + 1,645 (2,6458)$$

$$= 10 + 4,3523 = 14,3523$$

Jadi dugaan selisih berat antara salak kualitas ekspor dan salak kualitas supermarket adalah antara 5,6477 gr/buah dan 14,3523 gr/buah dengan tingkat kepercayaan 95%.

4. Pendugaan bagi selisih proporsi

Diambil sampel ke-1 dengan ukuran n_1 , dan ternyata diperoleh kejadian “sukses” sebesar X_1 , maka proporsi “sukses” dalam sampel ke-1 tersebut adalah

$$p_1 = \frac{X_1}{n_1}$$

Selanjutnya diambil sampel ke-2 dengan ukuran n_2 , dan ternyata diperoleh kejadian “sukses” sebesar X_2 , maka proporsi “sukses” dalam sampel ke-2 tersebut adalah

$$p_2 = \frac{X_2}{n_2}$$

Jika sampel diambil dari populasi tak berhingga, maka batas bawah interval bagi selisih proporsi “sukses” antara kedua populasi tersebut dengan tingkat kepercayaan sebesar c adalah

$$(p_1 - p_2) - z_c \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}$$

Sedangkan batas atasnya adalah

$$(p_1 - p_2) + z_c \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}$$

p = proporsi sukses

z_c = koefisien kepercayaan pada tingkat kepercayaan c

q = proporsi bukan sukses ($1 - p$)

n = ukuran sampel

Contoh:

Jeruk yang dijual di pasar buah Gamping adalah jeruk yang berasal dari Jember dan Banyuwangi. Seorang mahasiswa mendapat tugas untuk melakukan pendugaan perbedaan proporsi jeruk yang berkulit rusak antara jeruk yang berasal dari Jember dan Banyuwangi. Mahasiswa tersebut mengambil sampel 80 buah jeruk Jember, dan setelah diamati didapati 18 diantaranya berkulit rusak. Selanjutnya mahasiswa tersebut mengambil sampel sebanyak 80 buah jeruk Banyuwangi dan didapati 12 diantaranya berkulit rusak. Hitunglah perbedaan proporsi jeruk berkulit rusak antara jeruk yang berasal dari Jember dan Banyuwangi dengan tingkat kepercayaan 90%!

Jawab

Diketahui

Jeruk Jember	Jeruk Banyuwangi
$n_1 = 60$	$n_2 = 80$
$X_1 = 18$	$X_2 = 12$
$p_1 = 0,30$	$p_2 = 0,15$

Dari tabel kurva normal dapat diketahui bahwa $z_{90\%} = 1,645$

Interval kepercayaan selisih proporsi jeruk berkulit rusak antara jeruk yang berasal dari Jember dan Banyuwangi

Batas bawah

$$\begin{aligned} &= (0,30 - 0,15) - 1,645 \sqrt{\frac{0,30 \cdot (1 - 0,70)}{60} + \frac{0,15 \cdot (1 - 0,15)}{80}} \\ &= 0,15 - 1,645 (0,0714) \\ &= 0,15 - 0,11774 = 0,03226 \end{aligned}$$

Batas atas

$$\begin{aligned} &= (0,30 - 0,15) + 1,645 \sqrt{\frac{0,30 \cdot (1 - 0,70)}{60} + \frac{0,15 \cdot (1 - 0,15)}{80}} \\ &= 0,15 + 1,645 (0,0714) \\ &= 0,15 + 0,11774 = 0,26774 \end{aligned}$$

Jadi perbedaan proporsi jeruk berkulit rusak antara jeruk yang berasal dari Jember dan Banyuwangi adalah antara 3,226% sampai 26,774% dengan tingkat kepercayaan/ketepatan 90%.

Soal Latihan

1. Suatu sampel yang terdiri dari 60 keluarga dari dusun Ngrukinam akan digunakan untuk menduga pendapatan per kapita di Kabupaten Bantal. Dari pendataan yang dilakukan terhadap 60 keluarga tersebut diperoleh informasi bahwa nilai tengah pendapatan per kapita adalah sebesar US \$ 4.000 dengan simpangan baku US \$ 100. Hitunglah interval kepercayaan dugaan pendapatan per kapita bagi keluarga di Kabupaten Bantal tersebut dengan tingkat kepercayaan 98%.
2. Di Kecamatan Ngawi diperkirakan terdapat 70 orang petani melon. Seorang mahasiswa melakukan penelitian dengan mengambil sampel sebesar 40 orang petani. Dari 40 orang petani tersebut, diperoleh data nilai tengah keuntungan usahatani melon sebesar Rp 11.000.000 per ha, dengan simpangan baku Rp 2.000.000 per ha. Hitunglah interval kepercayaan dugaan keuntungan usahatani melon 70 orang petani dengan tingkat kepercayaan 95%.
3. Penelitian mahasiswa terhadap 30 orang nasabah petani menunjukkan bahwa 36% nasabah petani mampu mengembalikan kredit usahatani yang diberikan bank secara tepat waktu. Tentukan interval kepercayaan proporsi nasabah petani yang mampu mengembalikan kredit tepat waktu dengan tingkat kepercayaan 95%.
4. BMT "Al Banka" mempunyai nasabah petani sebanyak 160 orang. Untuk menduga persepsi nasabah petani terhadap BMT "Al Banka", bagian RnD mengambil sampel nasabah sebanyak 35 orang, dan diperoleh data bahwa 25 orang diantaranya mempunyai persepsi yang baik terhadap BMT tersebut. Tentukan interval kepercayaan bagi proporsi nasabah petani yang mempunyai persepsi baik dengan tingkat kepercayaan 99%.
5. Suatu sampel random yang terdiri dari 100 keluarga di kota A menunjukkan nilai tengah pendapatan keluarga sebesar Rp. 3.500.000 dengan simpangan baku sebesar Rp 400.000. Sedangkan sampel random

- lain yang terdiri dari 120 keluarga di kota B menunjukkan nilai tengah pendapatan keluarga sebesar Rp 4.000.000 dengan simpangan baku Rp 350.000. Hitunglah interval kepercayaan 95% bagi perbedaan nilai tengah pendapatan keluarga dari semua keluarga yang berada di kedua kota tersebut!
6. Dua sampel masing-masing sebanyak 100 tanaman yang tumbuh di dua tempat yang berbeda. Dari sampel tanaman di tempat pertama diperoleh nilai tengah tinggi tanaman 10 inchi dengan simpangan baku 1 inchi. Dari sampel tanaman di tempat kedua diperoleh rata-rata 12 inchi dengan simpangan baku sebesar 1,1 inchi. Buatlah interval kepercayaan bagi perbedaan nilai tengah tanaman antara kedua tempat tersebut dengan tingkat kepercayaan sebesar (i) 90% dan (ii) 95%
 7. Seorang mahasiswa meneliti dengan mengambil sampel random sebanyak 300 keluarga di perkotaan dan didapatkan penduduk miskin sebanyak 8,6%, sedangkan dari 400 keluarga di pedesaan didapatkan penduduk miskin sebanyak 14,70%. Hitunglah interval kepercayaan bagi selisih proporsi penduduk miskin antara pedesaan dan perkotaan.
 8. Suatu sampel random sebanyak 100 pekerja kebun sawit didapatkan nilai tengah sebesar 12 menit untuk menyelesaikan pemetikan tandan buah segar dengan simpangan baku sebesar 2 menit. Sampel random lain yang terdiri dari 50 pekerja didapatkan nilai tengah sebesar 9 menit untuk menyelesaikan pemetikan tandan buah segar dengan simpangan baku sebesar 3 menit. Hitunglah interval kepercayaan bagi perbedaan nilai tengah waktu pemetikan tandan buah segar diantara kedua kelompok pekerja tersebut.
 9. Di desa pertama, dari 100 sampel petani yang diambil, ternyata 60 orang diantaranya membayar pajak tepat pada waktunya. Sedangkan dari 120 sampel petani di desa kedua, ternyata 90 orang diantaranya membayar pajak tepat pada waktunya. Hitunglah interval kepercayaan bagi perbedaan proporsi petani yang membayar pajak tepat pada waktunya diantara kedua desa tersebut.