Nendra M. S Dwipa, M. Sc.

MODUL BAHAN AJAR

STATISTIKA EKONOMI

PROGRAM STUDI AKUNTANSI
FAKULTAS BISNIS
UNIVERSITAS PGRI YOGYAKARTA
DESEMBER 2022

MODUL KULIAH

STATISTIKA EKONOMI

Oleh:
Nendra M.S Dwipa, M. Sc

DESEMBER 2022

HALAMAN PENGESAHAN

1. Judul Bahan Ajar : Statistika Ekonomi
2. Pelaksana/Penulis
    - a. Nama Lengkap & Gelar : Nendra Mursetya Somasih Dwipa, M.Sc
    - b. Jenis Kelamin : L/P
    - c. Pangkat/Golongan : Penata Muda Tk 1/IIIb
    - d. NIP/NIS : 19831030 201004 1 001
    - e. Program Studi/Fakultas : Akuntansi/Fakultas Bisnis
    - f. Telepon/Faks/Email :085640206090/nendradwipa@upy.ac.id

Mengetahui,	Yogyakarta, 15 Desember 2022
Ketua Program Studi	Penulis,
Ningrum Pramudiati, M. Si., Ak., CA.	Nendra MS Dwipa, M.Sc
NIS. 19930916 201805 2 013	NIS. 19831030 2010 04 1 001

SILABUS		
1. Identitas Mata Kuliah		
a. Nama Mata Kuliah	:	STATISTIKA EKONOMI
b. Nomor Kode	:	EKM 31238
c. Bobot SKS	:	3
d. Semester	:	I
e. Status Mata Kuliah	:	Wajib
2. Tujuan Umum Mata Kuliah		
a. Dengan mengikuti mata kuliah ini Mahasiswa diharapkan mampu memahami konsep-konsep dasar statistika		
b. terampil menjelaskan sifat-sifat peluang		
c. macam-macam data, pengumpulan data, penyajian data dalam tabel baris-kolom, tabel kontingensi, tabel distribusi frekuensi, data dalam bentuk diagram atau grafik,		
d. menafsirkan gejala dengan ukuran pemusatan, mempelajari nilai penyimpangan, ukutan-ukuran yang berkaitan dengan bentuk lengkungan, kurva-kurva normal yang berasal dari distribusi dengan peubah kontinu		
e. kurva-kurva dari distribusi yang tidak normal, populasi beserta sampel dalam penelitian		
3. Materi Kuliah		
a. Tinjauan Umum tentang Statistika		
b. Penyajian Data		
c. Distribusi Frekuensi		
d. Ukuran Pemusatan dan Letak		
e. Pengukuran Dispersi, Skewwnes dan Kurtosis data		
4. Dasar-dasar Probabilitas		
f. Ruang Sampel dan Kejadian		
g. Distribusi Variabel Random		
h. Pengujian Hipotesis		
i. Analisis Regresi dan Korelasi		
5. Pendekatan Pembelajaran		
Ceramah, Tanya Jawab dan Diskusi serta pemberian tugas		
6. Media/Alat Bantu Belajar		
a. Laptop		
b. Lcd		
c. White Board + Boardmaker		
7. Evaluasi Hasil Belajar Mahasiswa		
Ujian Tengah Semester (UTS), Ujian Akhir Semester(UAS), Tugas Terstruktur, dan Tugas Mandiri		
8. Penilaian Hasil Belajar Mahasiswa		
a. Bobot UTS	:	20 %
b. Bobot UAS	:	30 %
c. Bobot Tugas Terstruktur dan partisipasi dalam	:	40 %

pembelajaran		
d. Bobot Kehadiran dan keaktifan	:	10 %
9. Rincian Materi Kuliah Setiap Pertemuan		
a. Pertemuan ke-1	:	Stadium General + Kontrak Perkuliahan
b. Pertemuan ke-2	:	Penyajian data 1,2,3 arah dengan table dan Grafik
c. Pertemuan ke-3	:	Distribusi ferekuensi bagian 1 : Limit kelas,Nilai tengah, & lebar kelas
d. Pertemuan ke-4	:	Distribusi ferekuensi bagian 2 : Cara membuat tabel Distribusi Frekuensi, Distribusi Frekuensi relatif dan Kumulatif serta Histogram, Poligon, Ogif.
e. Pertemuan ke-5	:	Ukuran Pemusatan
f. Pertemuan ke-6	:	Ukuran Letak
g. Pertemuan ke-7	:	Pengukuran Dispersi data
h. Pertemuan ke-8	:	Skewwnes dan Kurtosis data
i. Pertemuan ke-9	:	UTS
j. Pertemuan ke-10	:	Dasar-dasar Probabilitas
k. Pertemuan ke-11	:	Ruang Sampel dan Kejadian
l. Pertemuan ke-12	:	Distribusi Variabel Random
m. Pertemuan ke-13	:	Pengujian Hipotesis
n. Pertemuan ke-14	:	Analisis Regresi dan Korelasi
o. Pertemuan ke-15	:	Review materi
p. Pertemuan ke-16	:	UAS
10. Daftar Pustaka		
1) Walpole, R.E.1995. *Pengantar Statistika*. Edisi ke-3. Terjemahan Bambang Sumantri. Jakarta;Gramedia .		
2) Boediono, Wayan Koster. 2008. *Teori dan Aplikasi Statistika dan Probabilitas.* Bandung: Rosda.		
3) Abadyo dan H. Permadi. 2000. *Metoda Statistika Praktis*. IMSTEP JICA, Bandung		
4) J.D.Edward and Mishra.S. 1988. *Statistika Matematika Modern.* ITB. Bandung.		
5) Walpole, R.E.1995. *Pengantar Statistika*. Edisi ke-3. Terjemahan Bambang Sumantri. Jakarta;Gramedia .		

BAB I
PENDAHULUAN STATISTIKA

A. Pengertian Dasar

Istilah “statistik” dalam kehidupan sehari-hari, bahkan di negara kita terdapat lembaga negara yang bernama Badan Pusat Statistik (BPS). Kita juga sering mendengar istilah “observasi”, “data”, “sensus”, “sample”, “populasi” dan lain-lain. Mirip dengan kata statistik, terdapat kata “statistika” seperti terlihat pada judul bab ini di atas. Kemudian timbul pertanyaan apakah ada perbedaan antara statistik , statistika dan metode statistik . Dari beberapa definisi yang banyak dikemukakan untuk menjelaskan tentang ketiga istilah ini, dapat diambil suatu rangkuman atau kesimpulan istilah-istilah tersebut.Berikut definisi beberapa istilah tersebut:

STATISTIK adalah suatu koleksi metode-metode yang dapat membantu seseorang dalam membuat keputusan-keputusan dari sejumlah informasi yang terbatas atau suatu alat untuk mengumpulkan, mengelola/mengatur dan menganalisa data dari suatu percobaan/survey.
STATISTIKA adalah kumpulan metoda yang digunakan untuk merencanakan eksperimen, mengambil data, dan kemudian menyusun, meringkas, menyajikan, menganalisa, menginterpretasikan dan mengambil kesimpulan yang didasarkan pada data tersebut.
DATA adalah hasil observasi atau pengamatan yang telah dikumpulan. Data dapat berupa hasil pengukuran; misalnya data tinggi dan berat badan, hasil pengelompokan; misalnya jenis kelamin, hasil jawaban responden terhadap suatu quesioner; misalnya tingkat kepuasan.

POPULASI adalah koleksi lengkap semua elemen yang akan diselidiki. Suatu koleksi dikatakan lengkap jika ia memuat semua subjek yang akan diselidiki.

SENSUS adalah koleksi data dari semua anggota dalam populasi.

SAMPEL adalah suatu suatu himpunan bagian dari populasi yang dapat dipilih dengan tehnik tertentu.

SAMPLING adalah proses pemilihan atau penentuan sample dari suatu populasi.

STATISTIKA DESKRIPTIF adalah statistika yang berkaitan dengan analisis dan deskripsi suatu grup sebagai populasinya, tanpa melakukan penarikan kesimpulan apapun untuk komunitas yang lebih luas dari grup tersebut.

STATISTIKA INFERENSIA adalah suatu semua metode yang berhubungan dengan analisis sebagian data untuk kemudian asmpai pada peramalan atau penarikan suatu kesimpulan mengenai keseluruhan gugus data induknya (Pengujian hipotesis).

Sebagai contoh, suatu lembaga survey melakukan wawancara terhadap 2350 penduduk Indonesia untuk mengetahui tingkat kepuasan terhadap kinerja pemerintah. Dalam hal ini sebanyak 2350 penduduk adalah suatu **sampel** dan keseluruhan penduduk Indonesia sekitar 230 juta jiwa adalah **populasinya.** Kalau tidak salah, setiap 5 tahun sekali pemerintah melakukan sensus ekonomi atau sensus pertanian. Pada kegiatan sensus semua kepala keluarga didata dan data yang terkumpul disebut sensus atau **data sensus.**

Pengumpulan data dengan cara sensus membutuhkan biaya, waktu dan tenaga yang banyak. Untuk alasan efisiensi, dalam banyak kasus pola atau kelakukan populasi cukup dipelajari melalui sampelnya. Nantinya, hasil analisis pada sampel ini digunakan untuk memberikan kesimpulan pada populasi asalnya. Agar dapat diharapkan kesimpulan yang valid maka sampel yang diambil haruslah representatif, artinya ia benar-benar mewakili populasinya. Sampel yang tidak valid akan melahirkan kesimpulan yang menyimpang dari keadaan yang sesungguhnya.

Pemilu atau pilkada di Indonesia dilakukan untuk mengatahui aspirasi dari semua pemilih. Jadi pemilu adalah suatu proses sensus untuk populasi pemilih; walaupun kenyataannya tidak semua data populasi dapat diperoleh karena banyaknya "golput". Sedangkan, lembaga survey yang melakukan perhitung cepat atau "quick count" adalah melakukan proses sampling, artinya data hanya diambil dari sebagian TPS yang tersebar dengan cara sedemikian rupa sehingga data yang diperoleh "dipercaya" dapat mewakili para pemilih semuanya. Hasilnya sangat cepat diperoleh dikarenakan data yang diambil hanya sebagian kecil dari data sesungguhnya. Keakuratan kesimpulan yang diambil bergantung pada kualitas sampel yang ambil dan metoda analisis data yang digunakan.

Ingat, dalam sistem sampling terdapat faktor kesalahan yang sudah diperhitungkan sejak awal. Diantara faktor kesalahan ini adalah sampling error yang adalah suatu ukuran peluang ketidakmiripan sampel dengan populasinya. Juga, metoda yang digunakan dalam melakukan analisis data selalu didasarkan pada teori probabilitas. Artinya tidak ada kesimpulan apapun dalam statistik yang bersifat eksak; semuanya memiliki peluang kejadian sebaliknya. Sangat dimungkinkan beberapa lembaga survey perhitungan cepat pilkada memberikan kesimpulan yang berbeda satu sama lainnya; terutama bila kedaan sesungguhnya hanya memberikan selisih yang sangat tipis. Masih ingat dengan kasus pilkada Jawa Timur beberapa waktu yang lalu?

B. Tipe-tipe data

Pada bagian sebelumnya kita telah mendefinisikan sampel dan populasi. Keduanya dibedakan berdasarkan proses melakukan observasi. Untuk membedakan antara data sampel dan data populasi biasanya digunakan istilah **statistik** dan **parameter**.

PARAMETER adalah suatu ukuran numerik yang menggambarkan karakter suatu populasi.

STATISTIK adalah ukuran numerik yang menggambarkan karakter suatu sampel.

CONTOH

1. Berdasarkan sensus ekonomi tahun 2010 terdapat 35% rumah tangga di Indonesia tergolong miskin. Nah, angka 35% ini adalah parameter karena ia diperoleh dari populasi yaitu semua rumah tangga di Indonesia.
2. Berdasarkan hasil survey terhadap 50 orang mahasiswa pendidikan matematika UNMUH Ponorogo angkatan 2008/2009 diperolah bahwa rata-rata NEM matematika mereka adalah 6.75. Angka 6.75 ini adalah statistik karena ia diberikan oleh sampel yang terdiri dari 50 orang mahasiswa tersebut.

Selain data yang berbentuk angka seperti cpntoh di atas, terdapat pula data dalam bentuk kategori. Kedua bentuk data ini didefinisikan secara formal sebagai berikut :

VARIABEL adalah suatu karakteristik yang dapat diukur dari orang, objek atau peristiwa yang sedang diteliti. (Contoh: tinggi badan, berat badan, pendaatan, jenis kelamin, ukuran keluarga, dan lain-lain). Variabel dapat berupa:

a. **Variabel Kuantitatif**
   1. Variabel diskret (variabel diskontinu): variabel yang nilainya hanya terdiri dari bilangan bulat. Contoh: Jumlah penduduk, jumlah anak, jumlah buku dan sebagainya.
   2. Variabel kontinu: variabel yang nilainya dapat berupa pecahan. Contoh: tinggi badan, berat badan, volume, air dan sebagainya.

b. **Variabel Kualitatif**

   Parameter adalah suatu karakteristik dari populasi yang dapat diukur dan dinyatakan dengan bilangan (π, x, σ, S, p, p). **A**dalah data yang dapat dipisahkan dalam beberapa kategori atau kelompok yang dibedakan oleh karakter bukan numerik.

CONTOH

1. **Data kuantitatif**: tinggi badan, nilai NEM, temperatur dalam derajat celsius, besar penghasilan.
2. **Data kualitatif**: jenis kelamin, profesi, temperatur dalam rasa (dingin, panas sejuk).

Selanjutnya, data kuantitatif dibedakan atas data diskrit dan data kontinu.

DATA DISKRIT adalah data yang banyak kemungkinannya berhingga atau terbilang.
DATA KONTINU adalah data yang benyak kemungkinannya takterbilang.
CONTOH

1. **Data diskrit**: jam kerja dalam sehari (kemungkinannya adalah 1, 2, 3, … , 24), banyak telor yang dihasilkan oleh ayam betina, banyak hari libur dalam setiap bulan.
2. **Data kontinu**: temperatur udara di berbagai tempat (kemungkinannya: semua nilai yang ada pada interval, misalnya dari -20 derajat celsius sampai dengan 50 derajat celsius.

C. Skala Pengukuran

Cara umum yang digunakan untuk mengklasifikasikan data adalah ditentukan oleh empat macam skala pengukuruan, yaitu skala **nominal**, **ordinal**, **interval** dan **rasio**. Dalam statistika terapan, level pengukuran data adalah suatu faktor penting dalam menentukan prosedur dan metoda statistika yang digunakan.

SKALA NOMINAL dicirikan oleh data yang terdiri atas nama-nama, label, atau kategori. Data seperti ini tidak dapat diurutkan seperti dari atas ke bawah, atau sebaliknya.

CONTOH: Berikut adalah contoh-contoh yang mengilustrasikan pengukuran skala nominal:

1. **Ya, tidak, tidak tahu**: biasanya diberikan pada lembar kuesioner.

2. **Warna**: warna mobil yang dimiliki oleh dosen UNMUH Ponorogo (hitam, merah, putih, biru, dan lain-lain).

Data-data yang diperoleh pada skala ini tidak dapat diurutkan. Data ini tidak dapat digunakan untuk kalkulasi, misalnya Ya + tidak tahu = ???, merah + hitam = ??? tidak dapat dilakukan.

SKALA ORDINAL data yang diperoleh pada skala ini dapat disusun dalam urutan tertentu, tetapi selisih nilai-nilainya tidak dapat ditentukan atau bahkan tidak bermakna sama sekali.

CONTOH: Berikut adalah contoh data yang diperoleh dari pengukuran skala ordinal

1. Nilai akhir pada KHS mahasiswa yang diberikan oleh pak Julan HERNADI: E, D, C, B-, B, A-, A. Nilai-nilai ini dapat diurutkan, misalnya nilai A lebih baik dari nilai B, tetapi seberapa besar selisih antara A dan B tidak dapat ditentukan. Jelasnya A- B tidak bermakna.
2. *Transparency International* Indonesia (TII) baru-baru ini mengumumkan ranking indeks persepsi korupsi (IPK) untuk 50 kota yang ada di Indonesia. Dari ke 50 kota tersebut, Yogyakarta menduduki kota terbersih pada ranking pertama, disusul Palangkaraya pada rakning kedua, Banda Aceh pada ranking ketiga dan seterusnya sampai Kupang pada ranking ke 50 atau terkorup. Data ranking di sini adalah suatu level pengukuran ordinal. Walaupun ada angka di sini namun selisih antara ranking 2 dan ranking 1, bila ditulis dalam bentuk 2-1 = 1 tidak memiliki makna sama sekali.

SKALA INTERVAL seprerti level ordinal dengan sifat tambahannya adalah selisih antara dua data memiliki makna. Tetapi level ini tidak memiliki **titik nol** alami sebagai titik awal.

CONTOH: Berikut inidata dalam level interval

1. TEMPERATUR: suhu badan 36 derajat celsius dan 37 derajat celsius adalah suatu contoh data dalam level interval. Nilai-nilai ini dapat diurutkan dan selisihnya dapat ditentukan denganjelas, dalam contoh ini selisihnya adalah 1 derajat celsius. Tetapi secara alami tidak ada titik nol dimana suhu atau temperatur ini dimulai. Suhu 0 derajat tidak berarti tidak ada panas. Tidaklah benar mengatakan bahwa suhu badan 40 derajat celsius panasnya 2 kali lipat dari suhu badan 20 derajat celsius.
2. TAHUN: tahun 542, 1000, 2000, 2008 adalah suatu data dalam level interval. Data ini dapat diurutkan dan dapat selisih antara 2 tahun sebarang, namun ia tidak ada titik nol alami. Artinya, waktu tidak dimulai dari tahun 0 dan tahun 0 hanya sebagai titik nol buatan manusia sebagai ganti titik nol alami yang menyatakan "tidak ada waktu".

SKALA RASIO seperti level interval namun ia memiliki titik nol alami sebagai titik awal.
Data dari skala rasio, data data dapat dibandingkan (selisih) dan dirasiokan (pembagian).

CONTOH: Berikut inidata dalam skala rasio

1. HARGA: harga-harga buku teks mahasiswa adalah suatu data level rasio dimana harga 0 rupiah menunjukkan tidak ada harga alias gratis.
2. BOBOT: berat badan manusia adalah suatu data level rasio dimana berat 0 kg

menyatakan tidak ada bobot.

3. Indeks persepsi korupsi (IPK): ketika belum diranking, IPK yang dikeluarkan oleh TII masih dalam bentuk skor skala 10 dengan ketelitian 2 digit dibelakang koma, misalnya Yogyakarta dengan IPK 6.43, Palangkaraya dengan IPK 6.10, Banda Aceh dengan IPK 5.87 dan seterusnya Kendari dengan IPK 3.39, terkecil Kupang dengan IPK 2.97. Di sini nilai 0 menunjukkan kriteria terkorup "di dunia dan akhirat"

D. Notasi Penjumlahan

1. $\sum_{i=1}^{n} x_1 = x_1 + x_2 + \ldots + x_n$

2. $\sum_{i=1}^{n} x_1{}^2 = x_1^2 + x_2^2 = \ldots + x_n^2$

3. $\sum_{i=1}^{n} x_1 y_i = x_1 x_2 = x_2 x_2 = \ldots + x_n y_n$

4. $\sum_{\substack{i=1 \\ (C=konstanta)}}^{n} (x_i + y_i + z_i) = \sum_{i=1}^{n} x_1 + \sum_{i=1}^{n} y_1 + \sum_{z=1}^{n} x_1$

SOAL-SOAL LATIHAN:

1. Jelaskan perbedaan pengertian antara statistik dan statistika !
2. Jelaskan pengertian metode statistik , serta sebutkan dan jelaskan dua jenis metode statistik !
3. Jelaskan perbedaan pengertian antara populasi, sample, sampling, sample acak dan berikan contohnya ! Catatan: Pandanglah bahwa populasinya adalah mahasiswa Semester V Jurusan Ilmu Perpustakan – Universitas Indonesia.
4. Seorang mahasiswa Jurusan Ilmu Perpustakaan – UI ditugaskan untuk mempelajari factor-faktor yang berpengaruh terhadap prestasi kerja seorang pustakawan. Sebutkan variable dan parameter dari masalah tersebut !
5. Apa yang dimaksud dengan hipotesis statistik ! Berikan contohnya yang berkaitan dengan bidang informasi/perpustakaan/dokumentasi minimal 2 macam).
6. [STATISTIK dan PARAMETER] Indentifikasilah apakah nilai (angka) berikut sebagai parameter atau statistik.
   a. Dewan Perwakilan Rakyat (DPRI) saat ini terdiri dari 150 perempuan dan 350 pria.
   b. Sebuah sampel mahasiswa dipilih diperoleh bahwa rata-rata waktu belajar mandiri mereka dalam seminggu adalah 15.2 jam.

c. Dalam mempelajari semua penumpang Titanic yang berjumlah 2223 orang, ditemukan 706 orang selamat pada saat kapal tenggelam.

7. [DATA KONTINU dan DATA DISKRIT] Bedakan apakah nilai (angka berikut) sebagai data kontinu atau data diskrit.
    a. Gaji yang diperoleh oleh pekerja Indonesia di luar negeri mencapai 3.000.000,- rupiah setiap bulannya.
    b. Dalam 1560 orang pria yang disurvey ditemukan 38% dari mereka adalah perokok aktif.
    c. Suatu sampel terdiri dari sejumlah mobil, ditemukan bahwa rata-rata beratnya adalah 1500 kg.

8. [LEVEL PENGUKURAN] Tetapkan level yang paling cocok (nominal, ordinal, interval, rasio) untuk pengukuran berikut.
    a. Tinggi badan pemain sepak bola.
    b. Temperatur saat ini di dalam kelas.
    c. Rating suatu acara televisi: "fantastik, baik, cukup, kurang, tidak diterima".
    d. Nomor punggung pemain basket.
    e. Nomor telepon pada buku telepon.
    f. Majalah konsumen yang memberikan rating: "*best buy, reccomended, not reccomended".*
    g. Kode pos.

9. [SAMPEL – POPULASI] Tentukan yang mana sampel dan yang mana populasinya. Tentukan juga sampel mana yang paling mungkin sebagai representasi dari populasinya.
    a. A reporter for *Newsweek* stands on a street corner and asks 10 adults if they feel that the current president is doing a good job.
    b. Nielsen Media Research surveys 5000 randomly selected households and finds that among the TV sets in use, 19% are tuned to *60 Minutes* (based on data from *USA Today*).
    c. In a Gallup poll of 1059 randomly selected adults, 39% answered "yes" when asked "Do you have a gun in your home?

d. A graduate student at the University of Newport conducts a research project about how adult Americans communicate. She begins with a survey mailed to 500 of the adults that she knows. She asks them to mail back a response to this question: “Do you prefer to use e-mail or snail mail (the U.S. Postal Service)?” She gets back 65 responses,

10. [INTERPRETASI POOLING POLITIK] Andaikan sebuah lembaga survey meminta 200 orang responden tentang preferensi atau pilihan partai politik. Andaikan ada 4 parpol, masing-masing diberi nilai 0 (untuk partai ZERO), 1 (untuk partai ONE), 2 (untuk partai TWO) dan 3 (untuk partai THREE). Berdasarkan hasil survey, diperoleh nilai rata-rata pemilih adalah 0.95. Berikan interpretasi terhadap angka ini ?

11. [SKALA RATING MAKANAN] A group of students develops a scale for rating the quality of the cafeteria food, with 0 representing “neutral: not good and not bad.” Bad meals are given negative numbers and good meals are given positive numbers, with the magnitude of the number corresponding to the severity of badness or goodness. The first three meals are rated as 2, 4, and -5. What is the level of measurement for such ratings? Explain your

BAB II
DISTRIBUSI FREKUENSI

A. Pendahuluan

Data pertama yang diperoleh pada suatu observasi disebut dengan **data mentah** (*raw data*). Data ini belum tersusun secara numerik. Sebagai contoh data mengenai tinggi badan siswa yang penyajiannya masih dalam bentuk presensi kehadiran yang biasanya hanya diurutkan berdasarkan alphabet nama siswa. Terkadang data mentah disajikan berdasarkan urutan naik (*ascending*) atau urutan turun (*descending*). Bentuk penyajian seperti ini disebut array. Selisih antara nilai data terbesar dan terkecil disebut rentang (*range*).

Dalam bekerja dengan jumlah data yang cukup besar, biasanya lebih menguntungkan jika data ini disajikan dalam kelas-kelas atau kategori tertentu bersamaan dengan frekuensi yang bersesuaian. Frekuensi yang dimaksud adalah banyaknya kejadian yang ada pada kelas-kelas tertentu. Suatu tabel yang menyajikan kelas-kelas data beserta frekuensinya disebut **distribusi frekuensi** atau **tabel frekuensi**.

CONTOH: Berikut distibusi frekuensi tinggi badan 100 siswa SMA XYZ

Tabel 2.1 Tinggi 100 siswa SMA XYZ

Tinggi badan (in)	frekuensi
60–62	5
63–65	18
66–68	42
69–71	27
72–74	8
	100

Berdasarkan tabel di atas, banyak siswa yang tingginya berada dalam rentang 66 in dan 68 in adalah 42 orang. Salah satu kelemahan penyajian data dalam tabel frekuensi adalah tidak terlihatnya data asli atau data mentahnya.

B. Beberapa Istilah Pada Tabel Frekuensi

INTERVAL KELAS adalah interval yang diberikan untuk menetapkan kelas-kelas dalam distribusi. Pada tabel 2.1, interval kelasnya adalah 60-62, 63-65, 66-68, 69-71 dan 72-74. Interval kelas 66-68 secara matematis adalah suatu interval tertutup [66, 68], ia memuat semua bilangan dari 66 sampai dengan 68. Bilangan 60 dan 62 pada interval 60-62 disebut **limit kelas**, dimana angka 60 disebut **limit kelas bawah** dan angka 62 disebut **limit kelas atas.**

BATAS KELAS adalah bilangan terkecil dan terbesar sesungguhnya yang masuk dalam kelas interval tertentu. Misalnya jika dalam pengukuran tinggi badan di atas dilakukan dengan ketelitian 0.5 in maka tinggi badan 59.5 in dan 62.5 in dimasukkan ke dalam kelas 60 – 62. Bilangan 59.5 dan 62.5 ini disebut **batas kelas** atau **limit kelas sesungguhnya**, dimana bilangan 59.5 disebut **batas kelas bawah** dan 62.5 disebut **batas kelas atas**. Pada prakteknya batas kelas interval ini ditentukan berdasarkan rata-rata limit kelas atas suatu interval kelas dan limit kelas bawah interval kelas berikutnya. Misalnya batas kelas 62.5 diperoleh dari (62+63)/2. Pemahaman yang sama untuk interval kelas lainnya.

LEBAR INTERVAL KELAS adalah selisih antara batas atas dan batas bawah batas kelas. Misalnya lebar intervl kelas 60-62 adalah 62.5–59.5 = 3.

TANDA KELAS adalah titik tengah interval kelas. Ia diperoleh dengan cara membagi dua jumlah dari limit bawah dan limit atas suatu interval kelas. Contoh tanda kelas untuk kelas interval 66-68 adalah (66+68)/2 = 67.

C. Prosedur umum membuat tabel frekuensi

Berikut langkah-langkah untuk membuat tabel frekuensi:

1. Tetapkan data terbesar dan data terkecil, kemudian tentukan rangenya.
2. Banyaknya kelas tergantung pada tujuan dari penggunaan data. Namun umumnya antara 5 – 20. Dalam menentukan banyak kelas dapat digunakan rumus **Kriterium Sturges:**

   K = 1 + 3.322 log n

   dimana:

   k =jumlah kelas

   n =jumlah data
3. Bagilah range ini ke dalam sejumlah interval kelas yang memiliki ukuran sama. Jika tidak mungkin, gunakan interval kelas dengan ukuran berbeda. Biasanya banyak interval kelas yang digunakan antara 5 dan 20, bergantung pada data mentahnya. Diupayakan agar tanda kelas adalah suatu data observasi sesungguhnya. Hal ini untuk mengurangi apa yang disebut dengan *grouping-error*. Namun batas kelas sebaiknya tidak sama dengan data observasi.
4. Tentukan wilayah datanya

   Wilayah = Nilai data terbesar – Nilai data terkecil
5. Tentukan lebar kelasnya

   Lebar kelas = wilayah/jumlah kelas
6. **Starting point**: mulailah dengan bilangan limit bawah untuk kelas interval pertama. Dapat dipilih sebagai data terkecil dari observasi atau bilangan di

bawahnya.

7. Dengan menggunakan limit bawah interval kelas pertama dan lebar interval kelas, tentukan limit bawah interval kelas lainnya.
8. Susunlah semua limit bawah interval kelas secara vertikal, kemudian tentukan limit atas yang bersesuaian.
9. Tentukan titik tengah kelas bagi masing-masing kelas
10. Kembalilah ke data mentah dan gunakan turus untuk memasukkan data pada interval kelas yang ada.

CONTOH: Berikut nilai 80 siswa pada ujian akhir mata pelajaran matematika:

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

Langkah-langkah untuk membuat tabel distribusi frekuensi dilakukan sebagai berikut:

1. Nilai tertinggi = 97 dan nilai terendah 53. Jadi range = 97-53 = 44.
2. Tetapkan jumlah kelas; dalam hal ini 10.
3. Lebar interval kelas d = 44/10 = 4.4 dibulatkan menjadi 5.
4. Diambil bilangan 50 sebagai limit bawah untuk kelas pertama.
5. Selanjutnya, limit bawah untuk kelas kedua adalah 50+5 = 55, limit bawah kelas ketiga 55+5 = 60 dan seterusnya.
6. Limit atas kelas interval yang bersesuaian adalah 54 untuk kelas pertama, 59 untuk kelas kedua, dan seterusnya.
7. Gunakan turus untuk memasukkan data ke dalam interval kelas.

Hasilnya seperti terlihat pada Tabel 2.3 berikut:

Table 2.3

50–54	53
55–59	59, 57
60–64	62, 60, 61, 62, 63, 60, 61, 60, 62, 62, 63
65–69	68, 68, 65, 66, 69, 68, 67, 65, 65, 67
70–74	73, 73, 71, 74, 72, 74, 71, 71, 73, 74, 73, 72
75–79	75, 76, 79, 75, 75, 78, 78, 75, 77, 78, 75, 79, 79, 78, 76, 75, 78, 76, 76, 75, 77
80–84	84, 82, 82, 83, 80, 81
85–89	88, 88, 85, 87, 89, 85, 88, 86, 85
90–94	90, 93, 93, 94
95–99	95, 96, 95, 97

Akhirnya diperoleh tabel distribusi frekuensi sebagai berikut:

Tabel 2.4 Distribusi nilai matematika 80 siswa SMA XYZ

Rentang nilai	frekuensi
50-54	1
55-59	2
60-64	11
65-69	10
70-74	12
75-79	21
80-84	6
85-89	9
90-94	4
95-99	4
	80

Melalui tabel ini kita dapat mengetahui pola penyebaran nilai siswa. Paling banyak nilai siswa mengumpul pada interval 75-79, paling sedikit data termuat dalam interval 50-54. Sedangkan siswa yang mendapat nilai istimewa atau di atas 90 hanya ada 8 orang.

Pola penyebaran ini akan tampak lebih jelas jika digambarkan dengan menggunakan histogram. Penyajian data dengan menggunakan grafik dan diagram akan dibicarakan minggu depan.

LATIHAN UNTUK MEMANTAPKAN PEMAHAMAN:

Untuk data nilai matematika siswa:

a. Buatlah tabel distibusi frekuensi dengan mengambil banyak kelas 8.

b. Hitung rata-rata nilai siswa dari data mentahnya.
c. Hitung rata-rata nilai siswa dari tabel distribusi frekuensinya dengan menggunakan rumus
d. Lakukan seperti pertanyaan c tetapi untuk tabel distribusi dengan 10 kelas seperti yang diperoleh sebelumnya.
e. Simpulkan, rata-rata mana dari hasil c dan d yang lebih mendekati rata-rata sesungguhnya.

CATATAN : Mahasiswa diharuskan memahami materi ini sampai tuntas karena materi ini tidak diulang lagi. Buku acuan pokok "STATISTICS 4th ed" oleh Murray Spiegel dan Larry Stephens harus sudah dimiliki dan soal-soalnya dikerjakan dan pahami.

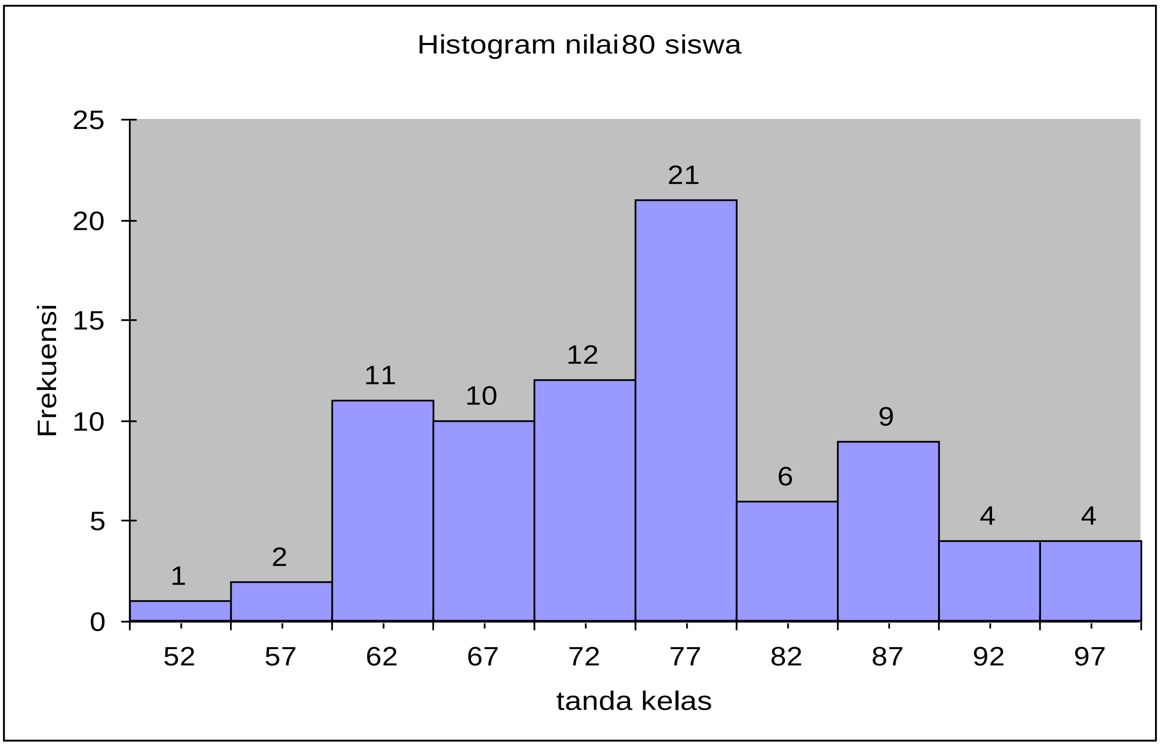
D. Histogram dan poligon frekuensi

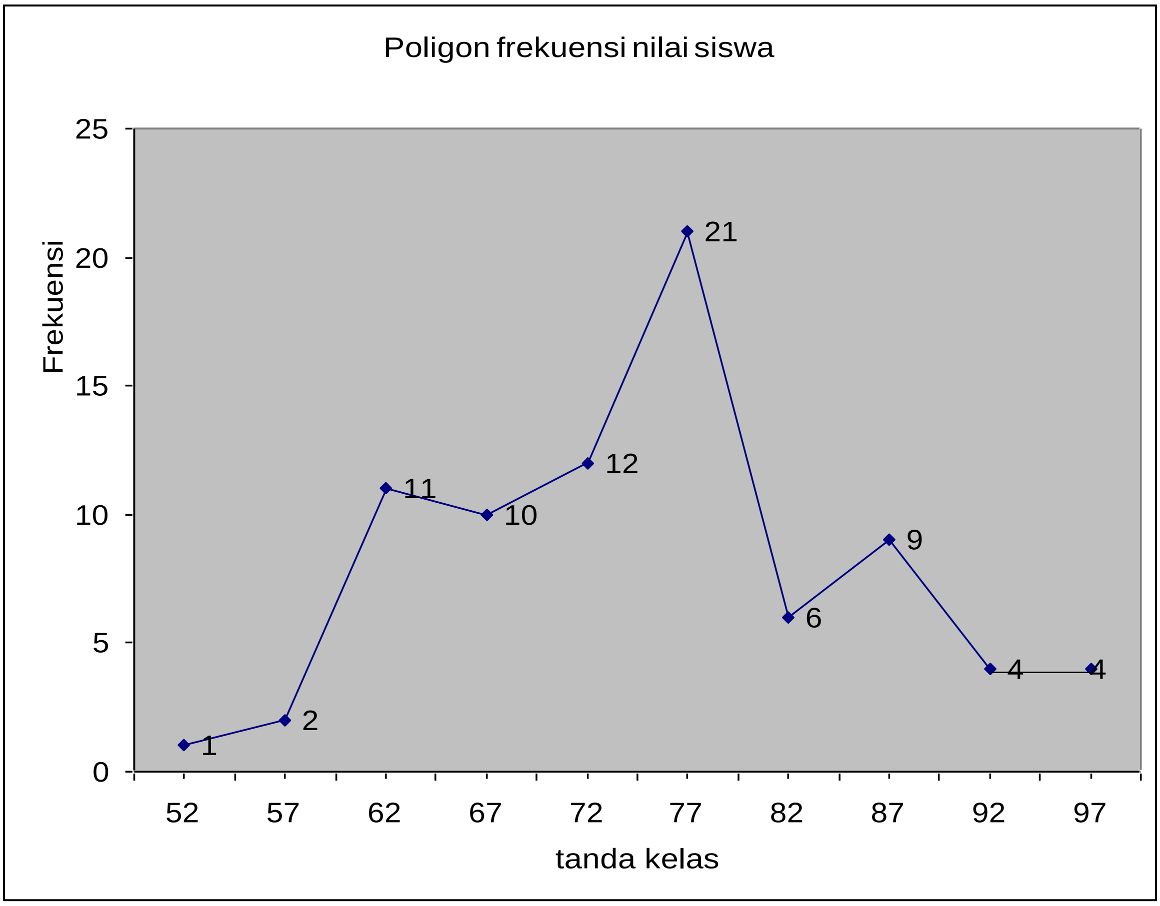
Histogram dan poligon frekuensi adalah suatu representasi grafik untuk distribusi frekuensi.

Histogram berupa sekumpulan persegi panjang dengan

1. Alas pada sumbu X, pusat alasnya adalah tanda kelas dan lebar alasnya adalah lebar kelas interval.
2. Tinggi adalah suatu frekuensi pada kelas yang bersangkutan.

Poligon frekuensi grafik garis yang mengaitkan frekuensi kelas dengan tanda kelas. Ia dapat digambarkan dengan menghubungkan garis lurus yang melalui titik-titik pasangan frekuensi kelas dan titik tengah (tanda) interval kelas.


Histogram nilai 80 siswa
Frekuensi
25
20
15
10
5
0
1
2
11
10
12
21
6
9
4
4
52
57
62
67
72
77
82
87
92
97
tanda kelas


Poligon frekuensi nilai siswa
Frekuensi
25
20
15
10
5
0
1
2
11
10
12
21
6
9
4
4
52
57
62
67
72
77
82
87
92
97
tanda kelas

PENGGUNAAN EXCEL untuk membuat tabel frekuensi, histogram dan poligon frekuensi.

Bila kita bekerja dengan data dalam jumlah besar, katakan lebih dari 1000 data maka menyusun tabel distribusi frekuensi denga metoda klasik yaitu dengan menggunakan turus bukanlah ide yang bagus. Selain waktunya lama juga ketelitiannya pantas diragukan. Oleh karena itu penggunaan teknologi komputer mutlak diperlukan. Salah satu program aplikasi yang dapat digunakan adalah microsoft excel atau disingkat excel saja. Program aplikasi ini sangat sederhana dan ia terpadu satu paket pada microsoft excel. Jadi hampir setiap komputer terinstal program ini. Agar penggunaan excel dalam analisis data dapat maksimal maka harus diinstalkan paket statistiknya yang biasanya sebagai opsional pada CD asalnya. Langkah-langkah mengupdate excel dengan paket analisis data statistik:

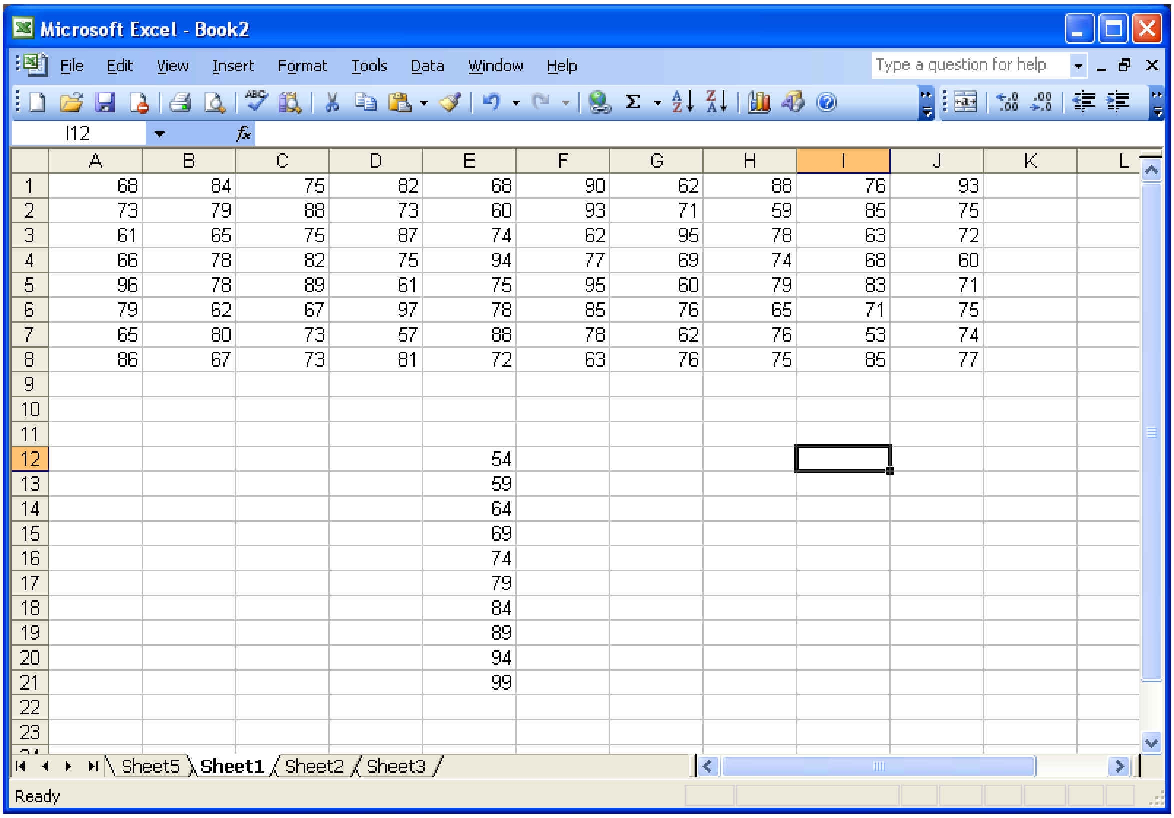
1. Siapkan CD dari mana office diinstall
2. Pada menu pilih **Tools - Add-Ins**.
3. Pada papan dialog Add-Ins, pilih **Analysis ToolPak** dan **Analysis ToolPak-VBA.**
4. Klik OK.

Kalau sekedar membuat tabel distribusi frekuensi, anda dapat menggunakan petunjuk pada modul pelatihan excel oleh Dr. Julan HERNADI dan tidak perlu paket khusus ini. Tetapi bila ingin lebih enak anda disarankan untuk melengkapi excel dengan add-Ins ini.

Dengan asumsi anda sudah bisa bekerja dengan excel paling tidak sudah dapat memasukkan data ke dalam lembar kerja excel, maka berikut ini diberikan langkah-langkah membuat tabel distrubusi frekuensi, histogram dan poligon frekuensi dalam satu paket.

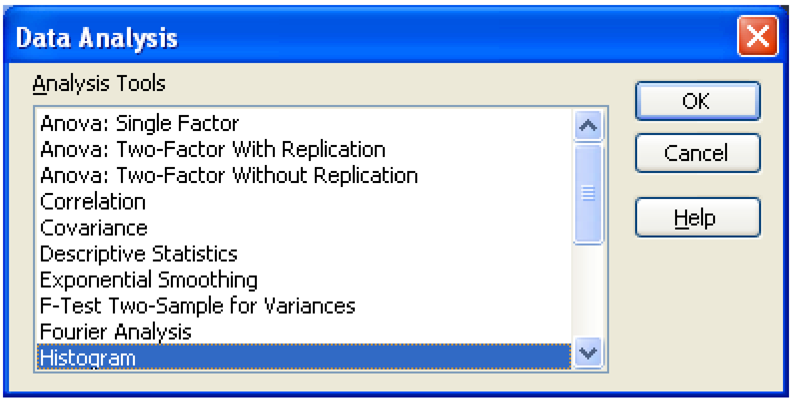
Langkah-langkah:

1. Buka Excel dan masukkan data mentah ke dalam sel-sel yang tersedia, misalnya terlihat pada tampilan berikut.
2. Buat array terpisah untuk memasukkan limit atas masing-masing interval seperti terlihat pada tampilan berikut.

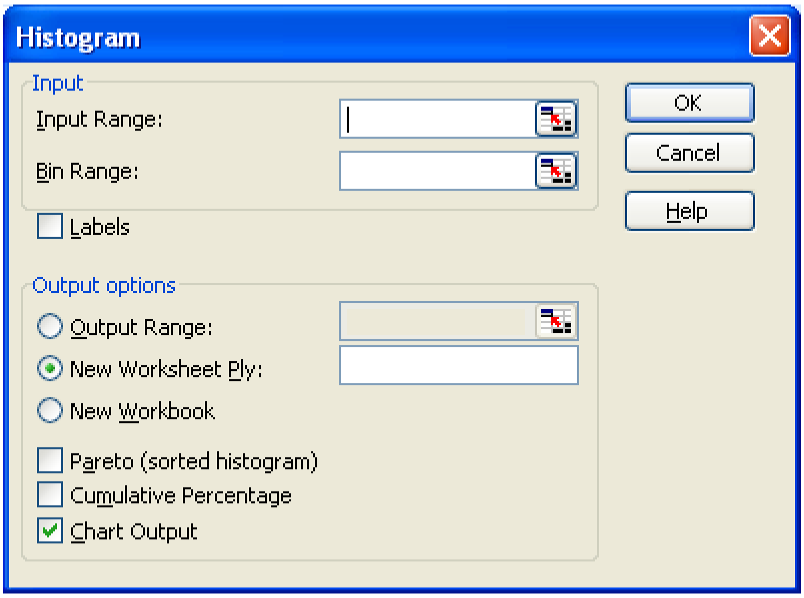


	A	B	C	D	E	F	G	H	I	J
1	68	84	75	82	68	90	62	88	76	93
2	73	79	88	73	60	93	71	59	85	75
3	61	65	75	87	74	62	95	78	63	72
4	66	78	82	75	94	77	69	74	68	60
5	96	78	89	61	75	95	60	79	83	71
6	79	62	67	97	78	85	76	65	71	75
7	65	80	73	57	88	78	62	76	53	74
8	86	67	73	81	72	63	76	75	85	77
12					54					
13					59					
14					64					
15					69					
16					74					
17					79					
18					84					
19					89					
20					94					
21					99					

3. Lakukan langkah-langkah analisis data sebagai berikut:
   a. Melalui menu Tools, pilih Data analysis, kemudian muncul pilihan berikut.
   b. Pilih Histogram, klik Ok.

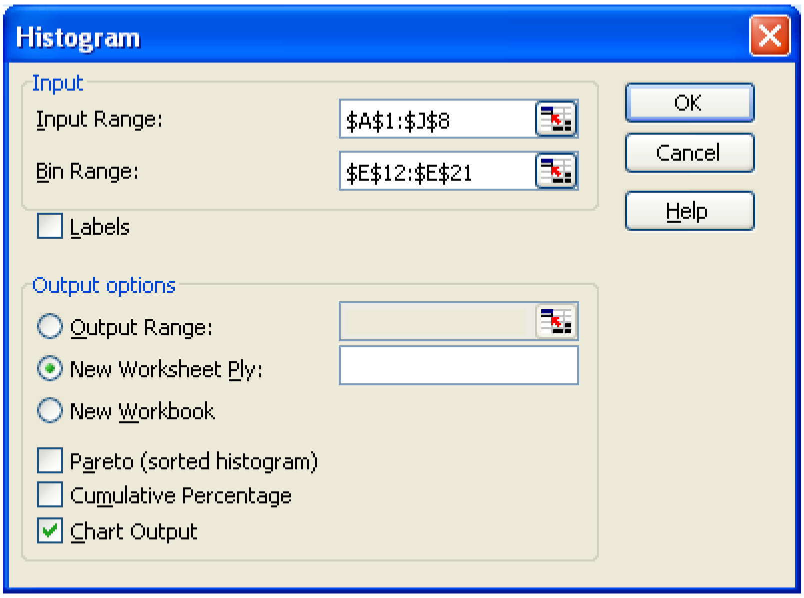


Diperolehlah tampilan berikut:

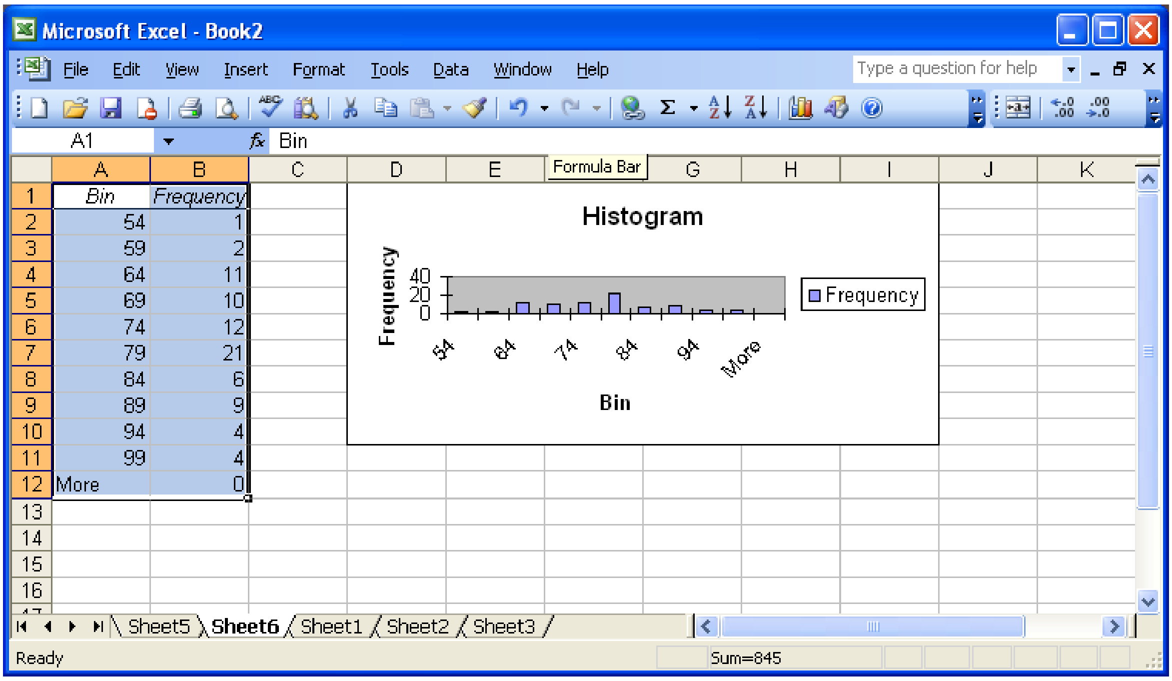


c. Pada baris **Input Range,** isilah dengan semua data dari sel A1 s.d. sel J8. Untuk mudahnya sorot semua sel tersebut.

d. Pada **Bin Range**, sorot semua array limit atas interval kelas. Pada Output option, pilih seperti tampilan berukt.



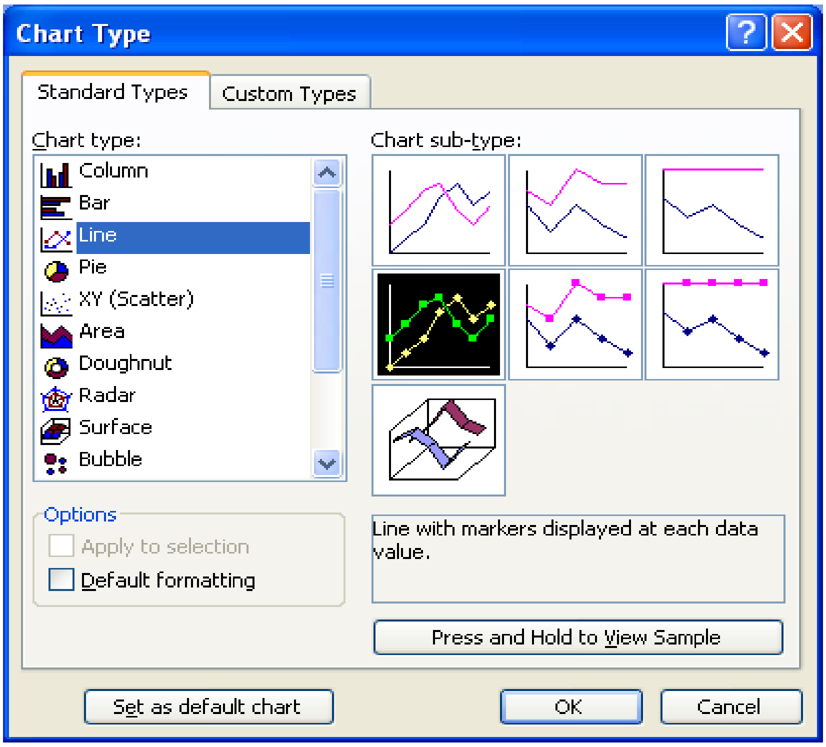
e. Klik Ok. Setelah itu akan muncul tampilan berikut:



f. Dengan melakukan editing pada histogram yang ada seperti memperbesar, menggeser, mengubah label, font, warna dan lain-lain maka akan diperoleh histogram yang diinginkan. Lakukanlah dengan coba-coba sambil mempelajari materi excel lebih lanjut.

g. Untuk menampilkan tanda kelas (titik tengah interval) pada sumbu X seperti pada teorinya maka angka pada kolom Bin diganti dengan tanda kelas, yaitu 52, 57, 62, dan seterusnya.

Coba lakukan langkah-langkah di atas dan berimprovisasilah sesuka anda sehingga diperoleh histogram yang persis gambar histogram pada halaman 4. Untuk membuat poligon frekuensi dilakukan langkah-langkah lanjutan berikut:

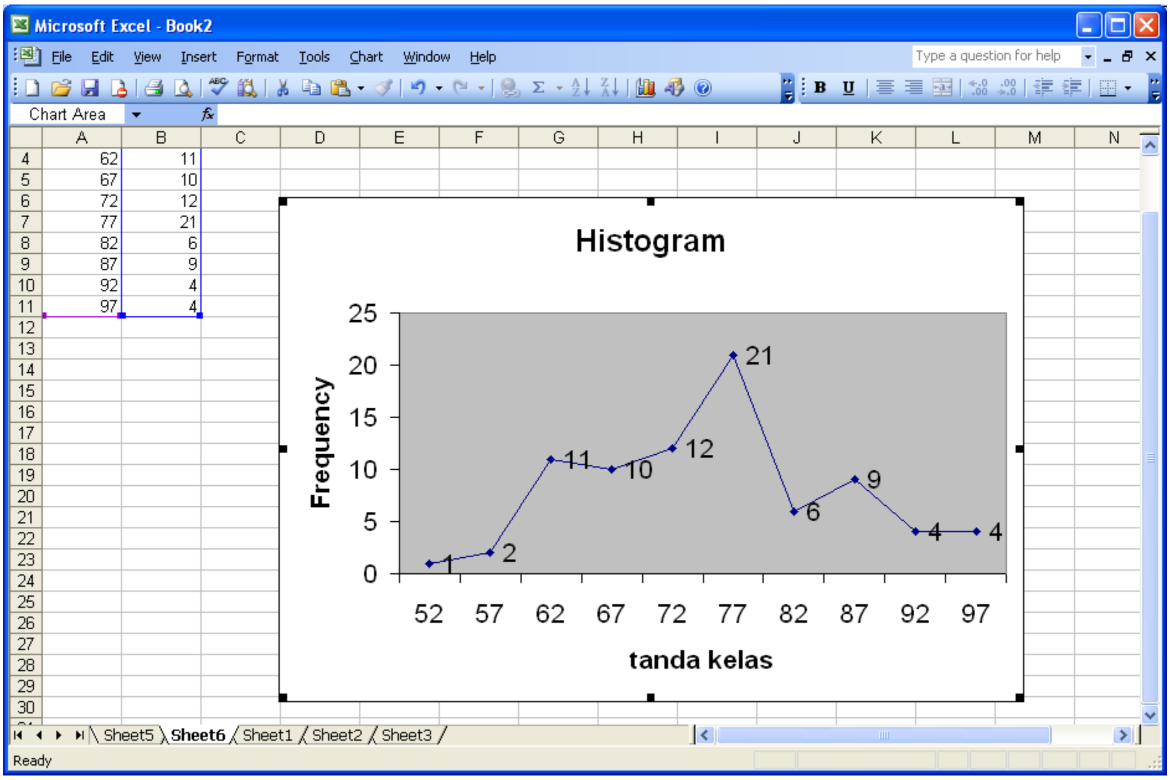


h. Melalui menu **Chart**, pilih **Chart Type**. Diperolehlah tampilan berikut.

i. Pilihlah sub-type sesuai dengan tampilan yang ada.

j. Ok.

Akhirnya, diperoleh poligon yang dimaksud. Selanjutnya lakukan editing, misalnya judul histogram diganti dengan poligon frekuensi, dan lain-lain yang dianggap perlu.



E.Distribusi frekuensi kumulatif, relatif dan ogive

DISTRIBUSI FREKUENSI RELATIF adalah suatu frekuensi kelas interval relatif terhadap total frekuensi. Formula untuk distribusi frekuensi relatif diberikan oleh:

$$frekuensi\ relatif = \frac{\text{frekuensi kelas interval}}{\text{jumlah semua frekuensi}}$$

DISTRIBUSI FREKUENSI KUMULATIF untuk suatu kelas adalah jumlah frekuensi pada kelas tersebut dan semua frekuensi yang terdapat pada kelas sebelumnya. Biasanya digunakan batas atas kelas untuk membuat distribusi frekuensi kumulatif.

CONTOH: Diperhatikan kembali tabel 2.4 sebelumnya.

Tabel 2.5 Distribusi frekuensi relatif nilai matematika 80 siswa SMA XYZ

Rentang nilai	frekuensi	Frekuensi relatif	Frek relatif (%)
50-54	1	1/80	1.25
55-59	2	2/80	2.50
60-64	11	11/80	13.75
65-69	10	10/80	12.50
70-74	12	12/80	15.00
75-79	21	21/80	26.25
80-84	6	6/80	7.50
85-89	9	9/80	11.25
90-94	4	4/80	5.00
95-99	4	4/80	5.00
	80	1.00	100%

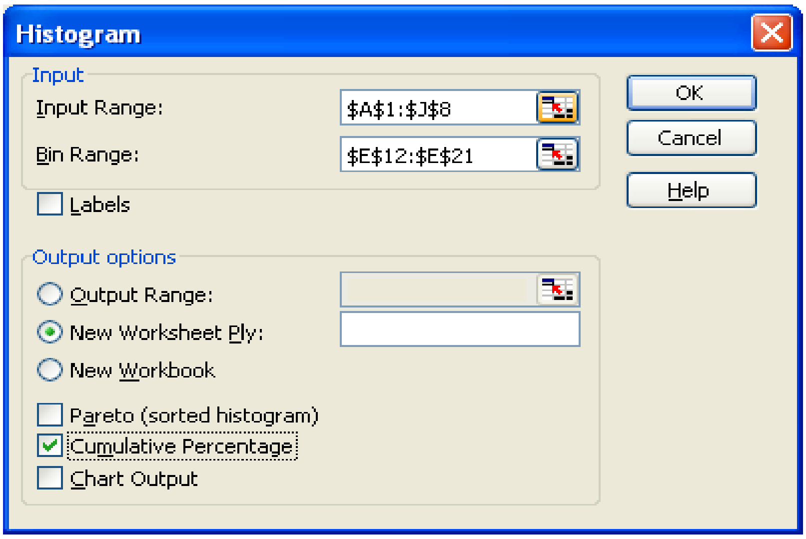
Tabel 2.6 Distribusi frekuensi kumulatif nilai matematika 80 siswa SMA XYZ

Rentang nilai	frekuensi	Frekuensi kumulatif	Frek kum (%)
<54.5	1	1	1.25
<59.5	2	3	3.75
<64.5	11	14	17.50
<69.5	10	24	30.00
<74.5	12	36	45.00

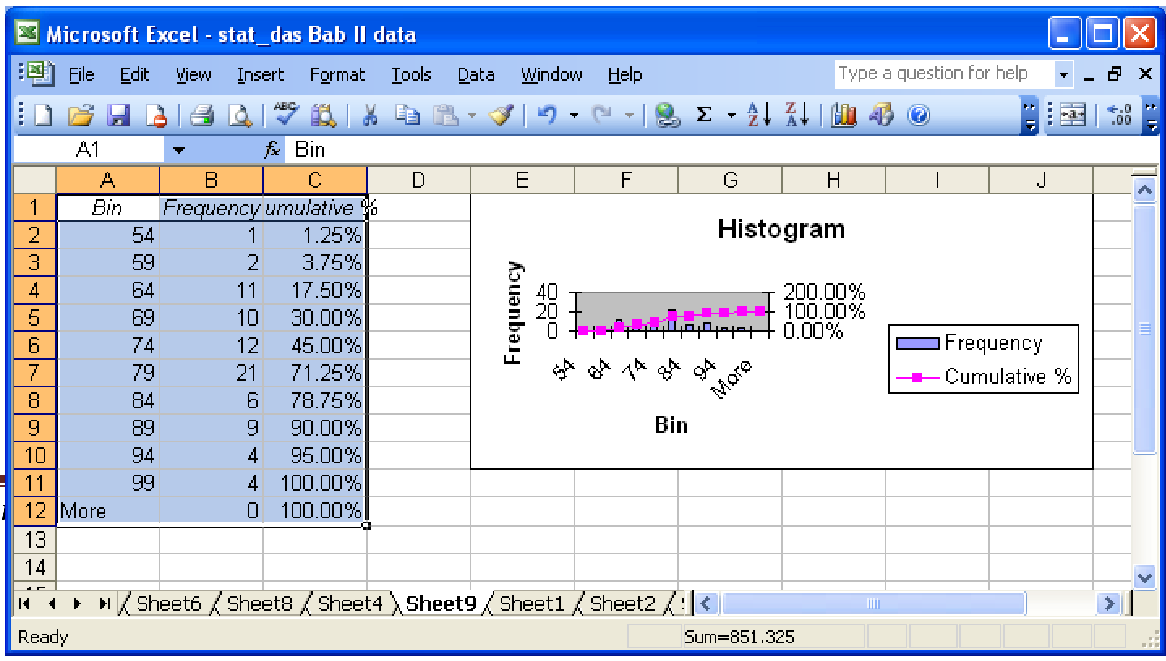
<79.5	21	47	58.75
<84.5	6	53	66.25
<89.5	9	62	77.50
<94.5	4	66	82.50
<99.5	4	80	100.00
	80		

Diperhatikan bahwa frekuensi kumulatif 24 pada kelas 65-69 diperoleh dari 1+2+11+10. Grafik yang menyajikan distribusi kumulatif ini disebut **ogive**. Untuk membuat ogive dengan excel, ikuti langkah-langkah berikut:

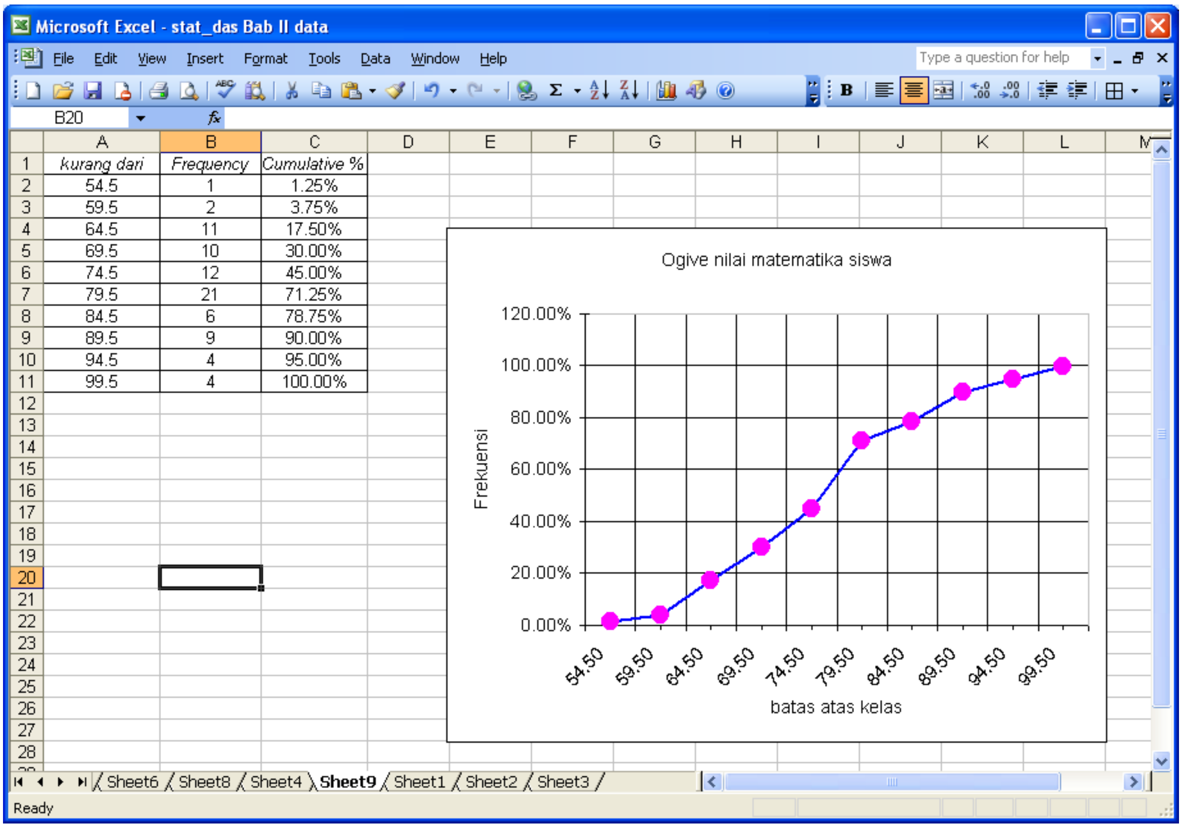
1. Anggaplah data nilai matematika siswa dan limit atas semua kelas sudah dimasukkan ke dalam workshet. Langkah ini sudah dipelajari ketika membuat histogram.
2. Melalui menu **Tools**, pilih **Data Analysis**, pilih **histogram**.
3. Pada **Output option**, pilih **Cumulative Percentage** dan **Chart Output**.
4. Ok



Diperolehlah output sebagai berikut:



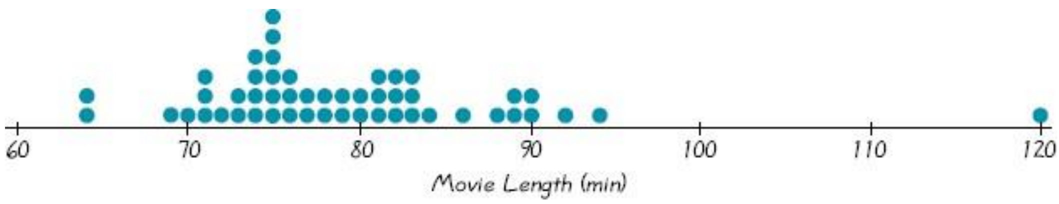
Setelah dilakukan editing, seperti membuang baris More pada Tabel, menghapus histogram frekuensi, menggeser, memperbesar, mengganti judul header, judul sumbu koordinat, dan lain-lain maka diperoleh tampilan yang lebih menarik berikut.



F. Bentuk diagram/kurva lainnya

1. Plot titik (*dotplot*)

Ini adalah grafik dimana setiap data digambarkan sebagai titik (dot) sepanjang garis skala nilai-nilainya.



Pada grafik ini ditampilkan data mengenai lama (durasi) beberapa judul film dengan data mentah sebagai berikut.

83 88 120 64 69 71 76 74 75 75 76 75

75	79	80	73	72	82	74	84	90	89	81	90
89	81	81	90	79	92	82	89	82	74	86	76
81	75	75	77	70	75	64	73	74	71	94	

Berdasarkan grafik ini, terdapat 2 data bernilai 64, terdapat 6 data bernilai 75 dan seterusnya. Data banyak mengumpul di dalam interval 70-90, sedangkan data 120 terpencil jauh dari kelompok data lainnya. Lebih lanjut, data ekstrim seperti ini disebut outlier dan dibutuhkan prosedur khusus untuk menangani data seperti ini.

2. Diagram Pareto

Ini adalah diagram batang untuk data kualitatif dimana batang-batangnya disusun berdasarkan urutan frekuensi. Kelompok dengan frekuensi terbanyak diletakkan paling kiri dan kelompok yang frekuensinya paling sedikit diletakkan paling kanan. Lihat contoh di bawah ini.

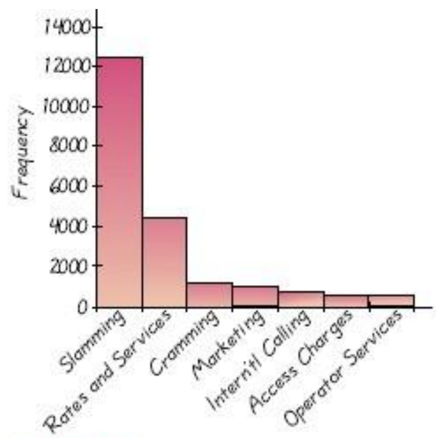


FIGURE 2-6 Pareto Chart of Phone Company Complaints

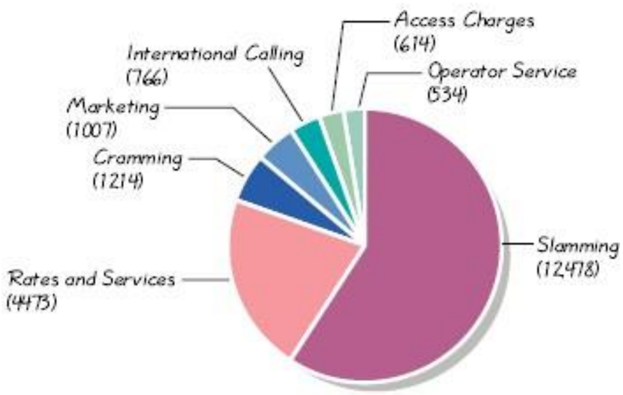


FIGURE 2-7 Pie Chart of Phone Company Complaints

3. Diagram kue (*Pie*)

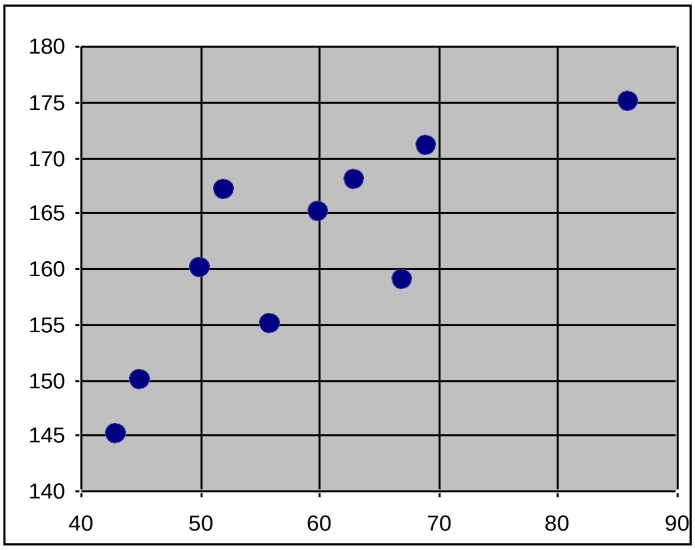
Ini adalah bentuk penyajian data kualitatif dalam bentuk potongan kue. Potongan kue dibuat proposional. Lihat contoh di atas.

4. Diagram pencar (scatter)

Diagram pencar ini digunakan untuk menyajikan pasangan data (x,y). Dengan melihat tampilan pada diagram pencar maka dapat secara umum bentuk hubungan antara dua kelompok data. Andaikan X adalah data tentang berat badan (dalam kg) dan Y adalah data tentang tinggi badan (dalam cm). Kedua data ini berpasangan, artinya setiap pasangan diperoleh dari orang yang sama.

X: 45 56 50 60 67 69 52 43 63 86

Y : 150 155 160 165 159 171 167 145 168 175



Berdasarkan diagram pencar ini terlihat bahwa terdapat hubungan linier antara berat badan dan tinggi badan. Lebih lanjut, konsep ini akan dibahas pada materi regresi dan korelasi.

SOAL-SOAL LATIHAN

1) Diberikan data nilai mahasiswa sebagai berikut:

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

Tentukan:

(a) nilai tertinggi.
(b) nilai terendah.
(c) rentang nilai.
(d) nilai-nilai yang menduduki ranking 5 terbesar.
(e) nilai-nilai yang menduduki rannking 5 terkecil.
(f) banyak siswa yang mendapat nilai tidak kurang dari 75.
(g) banyak siswa yang mendapat nilai kurang dari 85.
(h) prosentase siswa mendapat nilai lebih dari 65 tetapi tidak lebih dari 85.
(i) nilai yang tidak muncul sama sekali.

Selain dengan cara manual, kerjakan soal di atas dengan menggunakan excel. Tuliskan langkah-langkahnya.

2) Tabel berikut menyajikan distribusi frekuensi gaji mingguan pekerja pada PT. AR

Wages	Number of Employees
\$250.00–\$259.99	8
260.00–269.99	10
270.00–279.99	16
280.00–289.99	14
290.00–299.99	10
300.00–309.99	5
310.00–319.99	2
	Total 65

(a) limit bawah kelas ke 4.
(b) limit atas kelas ke 5.
(c) tanda kelas kelas ke 3.
(d) batas-batas kelas ke 6.
(e) lebar kelas ke 5.
(f) frekuensi kelas ke 2.
(g) frekuensi relatif kelas ketiga.
(h) kelas interval yang memiliki frekuensi tertinggi. Kelas ini disebut kelas modal.

3) Berikut data tinggi badan mahasiswa dalam inchi terdekat

67	67	64	64	74	61	68	71	69	61	65	64
62	63	59	70	66	66	63	59	64	67	70	65
66	66	56	65	67	69	64	67	68	67	67	65
74	64	62	68	65	65	65	66	67			

(a) buatlah tabel distribusi frekuensi dengan banyak kelas 5, dilengkapi dengan hsitogramnya.
(b) Buatlah tabel distribusi frekuensi dengan banyak kelas 6, dilengkapi dengan histogramnya.
(c) Buatlah tabel distribusi kumulatif dan ogive untuk hasil (a).

(d) Buatlah tabel distribusi kumulatif dan ogive untuk hasil (b).

4) Andaikan pada soal nomor 2 terdapat 5 pekerja baru dengan gaji sebagai berikut: $285.34, $316.83, $335.78, $356.21, dan $374.50. Buatlah tabel distribusi frekuensi baru untuk total 70 pekerja.

5) Lima koin dilempar sebanyak 1000 kali dan banyak muka (head) yang nampak dari kelima koin tersebut dicatat. Angka 0 menyatakan tidak ada muka yang tampak, angka 1 menyatakan terdapat 1 muka yang tamapak dans seterusnya. Data ke 1000 lemparan tersebut dirangkum pada tabel berikut:

Number of Heads	Number of Tosses (frequency)
0	38
1	144
2	342
3	287
4	164
5	25
	Total 1000

(a) Gambarkan diagram titik (dotplot) untuk data pada tebel di atas.
(b) Buatlah histogramnya.
(c) Buatlah tabel distribusi kumulatif dan ogivenya.

6) The following table shows the weekly-amount of time spent watching on TV by 400 SMA students.

Viewing Time (minutes)	Number of Students
300–399	14
400–499	46
500–599	58
600–699	76
700–799	68
800–899	62
900–999	48
1000–1099	22
1100–1199	6

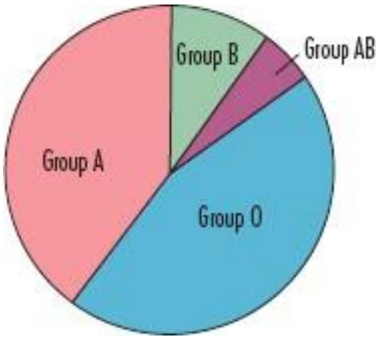
With this refference of table determine:

(*a*) The upper limit of the fifth class
(*b*) The lower limit of the eighth class
(*c*) The class mark of the seventh class
(*d*) The class boundaries of the last class
(*e*) The class-interval size
(*f*) The frequency of the fourth class
(*g*) The relative frequency of the sixth class
(*h*) The percentage of students whose weekly viewing time does not exceed 600 minutes
(*i*) The percentage of students with viewing times greater than or equal to 900 minutes
(*j*) The percentage of students whose viewing times are at least 500 minutes but less than 1000 minutes

7) The following table shows the diameters in centimeters of 60 ball bearings manufactured by a company. Construct a frequency distribution of the diameters, using appropriate intervals.

1.738	1.729	1.743	1.740	1.736	1.741	1.735	1.731	1.726	1.737
1.728	1.737	1.736	1.735	1.724	1.733	1.742	1.736	1.739	1.735
1.745	1.736	1.742	1.740	1.728	1.738	1.725	1.733	1.734	1.732
1.733	1.730	1.732	1.730	1.739	1.734	1.738	1.739	1.727	1.735
1.735	1.732	1.735	1.727	1.734	1.732	1.736	1.741	1.736	1.744
1.732	1.737	1.731	1.746	1.735	1.735	1.729	1.734	1.730	1.740

8) The following pie chart presents the blood groups for large sample of people.



(a) What is appproximate percentage of people with group A blood? If the pie is based on a sample of 500 people, approximately how many of those 500 people have group A blood?
(b) What is the approximate percentage of people with group B blood? Assuming the pie cahrt is based on sample of 500 people, approximately how many of those 500 people have group B blood?

BAB III
UKURAN STATISTIK

A. Pendahuluan

Ukuran Statistik :

a) Ukuran Pemusatan

Bagaimana, di mana data berpusat?

- Rata-Rata Hitung = Arithmetic Mean
- Median
- Modus
- Kuartil, Desil, Persentil

b) Ukuran Penyebaran

Bagaimana penyebaran data?

- Ragam, Varians
- Simpangan Baku

Ukuran Statistik nantinya akan mencakup data :

1. Ungrouped Data
2. Grouped Data

Ungrouped Data : Data yang belum dikelompokkan
Grouped Data : Data yang telah dikelompokkan → Tabel Distribusi Frekuensi

Contoh Ungrouped Data :
Data Nilai Statistika 10 orang mahasiswa FE-GD

78	62	34	57	89
67	55	75	73	56

Contoh Grouped Data

Kelas	Frekuensi
Nilai< 40	15
60 $\leq$ Nilai $\leq$ 80	30
40 $\leq$ Nilai $\leq$ 60	30
Nilai >80	25
Jumlah	100

B. Ukuran Pemusatan

1. Rata-Rata Hitung

Notasi : μ : rata-rata hitung populasi

$\bar{x}$: rata-rata hitung populasi

a) Rata-Rata Hitung untuk Ungrouped Data

$$\boxed{\mu = \frac{\sum_{i=1}^{N} x_i}{N}} \quad \text{dan} \quad \boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

μ : rata-rata hitung populasi

N : ukuran Populasi

$\bar{x}$: rata-rata hitung sampel

n : ukuran Sampel

x_i : data ke-i

Contoh :

Andaikan Di kota A hanya terdapat 6 PTS, masing-masing tercatat memiliki banyak mahasiswa sebagai berikut : 850, 1100, 1150, 1250, 750, 900

Berapakah rata-rata banyak mahasiswa PTS di kota A?

Rata-Rata Populasi atau Sampel ?

Jawab:

$$\mu = \frac{6000}{6} = 1000$$

b) Rata-Rata untuk Grouped Data

Nilainya adalah suatu **pendekatan**

Biasanya berhubungan dengan rata-rata hitung sampel

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}} \quad \text{sehingga :} \quad \boxed{\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{n}}$$

$\bar{x}$: rata-rata hitung sampel

n : ukuran Sampel

f_i : frekuensi di kelas ke-i

x_i : Titik Tengah Kelas ke-i

Kelas	Titik Tengah Kelas (x_i)	Frekuensi (f_i)	f_i x_i
16-23	19.5	10	195
24-31	27.5	17	467.5
32-39	35.5	7	248.5
40-47	43.5	10	435
48-55	51.5	3	154.5
56-63	59.5	3	178.5
Jumlah (Σ)		50	1679

Jawab : $\bar{x}$ = $\frac{1679}{50}$ = 33.58

Selain dengan rumus tersebut, dapat dicari dengan suatu nilai dugaan (M)

$$\boxed{\bar{x} = M + \frac{\sum_{i=1}^{n} f_i d_i}{n}}$$

d_i : TTKi (x_i) - M

Kelas	Titik Tengah Kelas (x_i)	M	d_i	Frekuensi(f_i)	f_i d_i
16-23	19.5	39.5	-20	10	-200
24-31	27.5	39.5	-12	17	-204
32-39	35.5	39.5	- 4	7	-28
40-47	43.5	39.5	4	10	40
48-55	51.5	39.5	12	3	36
56-63	59.5	39.5	20	3	60
Jumlah (Σ)			0	50	-296

Jawab :

$$\bar{x} = M + \frac{\sum_{i=1}^{n} f_i d_i}{n} = 39.5 + \frac{-295}{50} = 39.5 - 5.92 = 33.58$$

Catt : Bagaimana menentukan M?

Tidak ada cara khusus! M dapat ditentukan sembarang !

atau

M dapat ditentukan dengan Titik Tengah Kelas (x_i)

- jika banyak kelas (k) ganjil maka ambil (x_i) pada kelas ke $\left\lceil \frac{k}{2} \right\rceil$(kelas yang di tengah-tengah)
- jika banyak kelas (k) genap maka gunakan (x_i) pada kelas ke $\frac{k}{2}$dan kelas ke $\frac{k}{2}$+1 selanjutnya kedua nilai (x_i) tersebut dibagi dua

2. Modus

Nilai yang paling sering muncul

Nilai yang frekuensinya paling tinggi

a) Modus untuk Ungrouped Data

Contoh : Sumbangan PMI warga Depok

Rp. 7500 8000 9000 8000 3000 5000 8000

Modus : Rp. 8000

Bisa terjadi data dengan beberapa modus (multi-modus)

Bisa terjadi data tanpa modus

b) Modus untuk Grouped Data

Kelas Modus : Kelas di mana Modus berada

Kelas dengan frekuensi tertinggi

$$\text{Modus} = \text{TBB Kelas Modus} + i\left(\frac{d_1}{d_1 + d_2}\right)$$

di mana :

TBB : Tepi Batas Bawah

d_1 : Beda Frekuensi Kelas Modus dengan Frekuensi Kelas ***<u>sebelumnya</u>***

d_2 : Beda Frekuensi Kelas Modus dengan FrekuensiKelas ***<u>sesudahnya</u>***

i : interval kelas

Kelas	Frekuensi (fi)
16-23	10
24-31	17
32-39	7
40-47	10
48-55	3
56-63	3
Jumlah (Σ)	50

Kelas Modus = 24 - 31

TBB Kelas Modus = 23.5

i = 8

frek. kelas Modus = 17

frek, kelas sebelum kelas Modus = 10

frek. kelas sesudah kelas Modus = 7

d_1 = 17 - 10 = 7

d_2 = 17 - 7 = 10

Modus = 23.5 + $8\left(\frac{7}{7+10}\right)$ = 23.5 + 8 $\left(\frac{7}{17}\right)$ =

23.5 + 8 (0.41176...) = 23.5 + 3.2941...=

26.7941... $\approx$ 27

3. Median, Kuartil, Desil dan Persentil

a) Median untuk Ungrouped Data

Median $\rightarrow$ Nilai yang membagi gugus data yang telah tersortir (*ascending*) menjadi 2 bagian yang sama besar

Letak Median $\rightarrow$ Letak Median dalam gugus data yang telah tersortir

Letak Median = $\frac{n+1}{2}$ n : banyak data

- Jika banyak data (n) ganjil dan tersortir, maka:

Median = Data ke $\frac{n+1}{2}$

- Jika banyak data (n) genap dan tersortir, maka:

  Median = [Data ke-$\frac{n}{2}$ + Data ke-($\frac{n}{2}$ +1)] : 2

  Contoh 1 :

  Tinggi Badan 5 mahasiswa :

  1.75 1.78 1.60 1.73 1.78 meter

  Sorted :1.60 1.73 1.75 1.78 1.78 meter

n = 5 Letak Median = $\frac{5+1}{2}$ = $\frac{6}{2}$ = 3

Median = Data ke-3 = 1.75

Contoh 2 :

Tinggi 6 mahasiswa : 1.60 1.73 1.75 1.78 1.78 1.80 meter (Sorted)

n = 6

Letak Median $\rightarrow$ $\frac{6+1}{2}$ = $\frac{7}{2}$ = 3.5

Median = (Data ke 3 + Data ke 4) : 2 = (1.75 + 1.78) : 2 = 3.53 : 2 = 1.765

b) Median, Kuartil, Desil dan Persentil untuk Grouped Data

- Nilainya adalah suatu pendekatan

c) Median

Median $\rightarrow$ Nilai yang membagi gugus data yang telah tersortir (*ascending*) menjadi 2 bagian yang sama besar

Letak Median = $\frac{n}{2}$ n : banyak data

Kelas Median : Kelas di mana Median berada

Kelas Median didapatkan dengan membandingkan Letak Median dengan Frekuensi Kumulatif

$$\boxed{\text{Median} = \text{TBB Kelas Median} + i\left(\frac{s}{f_M}\right)}$$

atau

$$\boxed{\text{Median} = \text{TBA Kelas Median} - i\left(\frac{s'}{f_M}\right)}$$

di mana : TBB : Tepi Batas Bawah

s : selisih antara Letak Median dengan ***Frekuensi Kumulatif*** *sebelum* kelas Median

TBA : Tepi Batas Atas

s' : selisih antara Letak Median dengan ***Frekuensi Kumulatif sampai*** kelas Median

i : interval kelas

f_M : Frekuensi kelas Median

Contoh 3 : Kelas Median

Kelas	Frekuensi	Frek. Kumulatif
16 - 23	10	10
24 - 31	17	27
32 - 39	7	34
40 - 47	10	44
48 - 55	3	47
56 - 63	3	50
Σ	50	----

interval = i = 8

Letak Median = $\frac{n}{2} = \frac{50}{2} = 25$

Median = Data ke-25 terletak di kelas 24-31

$\therefore$ Kelas Median = 24 - 31

TBB Kelas Median = 23.5 dan TBA Kelas Median = 31.5

$f_M = 17$

Frek. Kumulatif sebelum Kelas Median = 10 $\rightarrow$ s = 25 - 10 = 15

Frek. Kumulatif sampai Kelas Median = 27 $\rightarrow$ s' = 27 - 25 = 2

Median= TBB Kelas Median + i $\left(\frac{s}{f_M}\right)$

$= 23.5 + 8\left(\frac{15}{17}\right) = 23.5 + 8\,(0.8823...)$

$= 23.5 + 7.0588... = 30.5588... \approx 30.6$

Median= TBA Kelas Median - i $\left(\dfrac{s'}{f_M}\right)$

= 31.5 - 8$\left(\dfrac{2}{17}\right)$ = 31.5 - 8 (0.1176...)

= 31.5 - 0.9411.. = 30.5588... $\approx$ 30.6

d) Kuartil

Kuartil $\rightarrow$ Nilai yang membagi gugus data yang telah tersortir (*ascending*) menjadi 4 bagian yang sama besar

Letak Kuartil ke-1 = $\dfrac{n}{4}$

Letak Kuartil ke-2 = $\dfrac{2n}{4}$ = $\dfrac{n}{2}$ $\rightarrow$ Letak Median

Letak Kuartil ke-3 = $\dfrac{3n}{4}$ n : banyak data

Kelas Kuartil ke-q : Kelas di mana Kuartil ke-q berada
Kelas Kuartil ke-q didapatkan dengan membandingkan Letak Kuartil ke-q dengan Frekuensi Kumulatif

Kuartil ke-q =TBB Kelas Kuartil ke-q + i $\left(\dfrac{s}{f_Q}\right)$

atau

Kuartil ke-q =TBA Kelas Kuartil ke-q - i $\left(\dfrac{s'}{f_Q}\right)$

q : 1,2 dan 3

di mana : TBB : Tepi Batas Bawah

s : selisih antara Letak Kuartil ke-q dengan ***Frekuensi Kumulatif sebelum*** kelas Kuartil ke-q

TBA : Tepi Batas Atas

s' : selisih antara Letak Kuartil ke-q dengan ***Frekuensi Kumulatif sampai*** kelas Kuartil ke-q

i : interval kelas

f_Q : Frekuensi kelas Kuartil ke-q

Contoh 4 : Tentukan Kuartil ke-3

Kelas	Frekuensi	Frek. Kumulatif
16 - 23	10	10
24 - 31	17	27
32 - 39	7	34
40 - 47	10	44
48 - 55	3	47
56 - 63	3	50
Σ	50	----

Kelas Kuartil ke-3

interval = i = 8

Letak Kuartil ke-3 = $\frac{3n}{4}$ = $\frac{3 \times 50}{4}$ = 37.5

Kuartil ke-3 = Data ke-37.5 terletak di kelas 40 - 47

$\therefore$ Kelas Kuartil ke-3 = 40 - 47

TBB Kelas Kuartil ke-3 = 39.5 dan TBA Kelas Kuartil ke-3 = 47.5

f_Q = 10

Frek. Kumulatif sebelum Kelas Kuartil ke-3 = 34 $\rightarrow$ s = 37.5 - 34 = 3.5

Frek. Kumulatif sampai Kelas Kuartil ke-3 = 44 $\rightarrow$ s' = 44 - 37.5 = 6.5

Kuartil ke-3 =TBB Kelas Kuartil ke-3 + i $\left(\frac{s}{f_Q}\right)$

$$= 39.5 + 8\left(\frac{3.5}{10}\right) = 39.5 + 8\ (0.35)$$

$$= 39.5 + 2.8 = 42.3$$

Kuartil ke-3 =TBA Kelas Kuartil ke-3 - i $\left(\dfrac{s'}{f_Q}\right)$

$= 47.5 - 8\left(\dfrac{6.5}{10}\right)$ $= 47.5 - 8\ (0.65)$

$= 47.5 - 5.2 = 42.3$

e) Desil

Desil $\rightarrow$ Nilai yang membagi gugus data yang telah tersortir (*ascending*) menjadi 10 bagian yang sama besar

Letak Desil ke-1 = $\dfrac{n}{10}$

Letak Desil ke-5 = $\dfrac{5n}{10} = \dfrac{n}{2}$ $\rightarrow$ Letak Median

Letak Desil ke-9 = $\dfrac{9n}{10}$ n : banyak data

Kelas Desil ke-d : Kelas di mana Desil ke-d berada

Kelas Desil ke-d didapatkan dengan membandingkan Letak Desil ke-d dengan Frekuensi Kumulatif

Desil ke-d =TBB Kelas Desil ke-d + i $\left(\dfrac{s}{f_D}\right)$

atau

Desil ke-d =TBA Kelas Desil ke-q - i $\left(\dfrac{s'}{f_D}\right)$

d : 1,2,3...9

di mana : TBB : Tepi Batas Bawah

s : selisih antara Letak Desil ke-d dengan ***Frekuensi Kumulatif sebelum*** kelas Desil ke-d

TBA : Tepi Batas Atas

s' : selisih antara Letak Desil ke-d dengan ***Frekuensi Kumulatif sampai*** kelas Desil ke-d i : interval kelas

f_D : Frekuensi kelas Desil ke-d

Contoh 5: Tentukan Desil ke-9

Kelas	Frekuensi	Frek. Kumulatif
16 - 23	10	10
24 - 31	17	27
32 - 39	7	34
40 - 47	10	44
48 - 55	3	47
56 - 63	3	50
Σ	50	----

Kelas Desil ke-9

interval = i = 8

Letak Desil ke-9 = $\frac{9n}{10}$ = $\frac{9 \times 50}{10}$ = 45

Desil ke-9 = Data ke-45 terletak di kelas 48 - 55

$\therefore$ Kelas Desil ke-9 = 48 - 55

TBB Kelas Desil ke-9 = 47.5 dan TBA Kelas Desil ke-9 = 55.5

f_D = 3

Frek. Kumulatif sebelum Kelas Desil ke-9 = 44 $\rightarrow$ s = 45 - 44 = 1

Frek. Kumulatif sampai Kelas Desil ke-9 = 47 $\rightarrow$ s' = 47 - 45 = 2

Desil ke-9 =TBB Kelas Desil ke-9 + i $\left(\frac{s}{f_D}\right)$

= 47.5 + 8 $\left(\frac{1}{3}\right)$ = 47.5 + 8 (0.333...)

= 47.5 + 2.66... = 50.166...

Desil ke-9 = TBA Kelas Desil ke-9 - i $\left(\frac{s'}{f_D}\right)$

$= 55.5 - 8\left(\frac{2}{3}\right)$ = 47.5 - 8 (0.666...)

= 55.5 -5.33... = 50.166...

f) Persentil

Persentil $\rightarrow$ Nilai yang membagi gugus data yang telah tersortir (*ascending*) menjadi 100 bagian yang sama besar

Letak Persentil ke-1 = $\frac{n}{100}$

Letak Persentil ke-50 = $\frac{50n}{100}$ = $\frac{n}{2}$ $\rightarrow$ Letak Median

Letak Persentil ke-99 = $\frac{99n}{10}$ n : banyak data

Kelas Persentil ke-p : Kelas di mana Persentil ke-p berada
Kelas Persentil ke-p didapatkan dengan membandingkan Letak Persentil ke-p dengan Frekuensi Kumulatif

Persentil ke-p =TBB Kelas Persentil ke-p + i $\left(\frac{s}{f_P}\right)$

atau

Persentil ke-p =TBA Kelas Persentil ke-p - i $\left(\frac{s'}{f_P}\right)$

p : 1,2,3...99

di mana : TBB : Tepi Batas Bawah
s : selisih antara Letak Persentil ke-p dengan ***Frekuensi Kumulatif sebelum*** kelas Persentil ke-p

TBA : Tepi Batas Atas
s' : selisih antara Letak Persentil ke-p dengan ***Frekuensi Kumulatif sampai*** kelas Persentil ke-p

i : interval kelas
f_P: Frekuensi kelas Persentil ke-p

Contoh 6: Tentukan Persentil ke-56

Kelas	Frekuensi	Frek. Kumulatif
16 - 23	10	10
24 - 31	17	27
32 - 39	7	34
40 - 47	10	44
48 - 55	3	47
56 - 63	3	50
Σ	50	----

Kelas Persentil ke-56

interval = i = 8

Letak Persentil ke-56 = $\frac{56n}{100}$ = $\frac{56 \times 50}{100}$ = 28

Persentil ke-56 = Data ke-28 terletak di kelas 32 - 39

$\therefore$ Kelas Persentil ke-56 = 32 - 39

TBB Kelas Persentil ke-56 = 31.5 dan TBA Kelas Persentil ke-56 = 39.5

$f_P = 7$

Frek. Kumulatif sebelum Kelas Persentil ke-56 = 27 $\rightarrow$ s = 28 - 27 = 1

Frek. Kumulatif sampai Kelas Persentil ke-56 = 34 $\rightarrow$ s' = 34 - 28 = 6

Persentil ke-26 =TBB Kelas Persentil ke-56 + i $\left(\frac{s}{f_P}\right)$

$= 31.5 + 8\left(\frac{1}{7}\right) = 31.5 + 8\ (0.142...)$

$= 31.5 + 1.142.. = 32.642...$

Persentil ke-26 = TBA Kelas Persentil ke-56 - i $\left(\frac{s'}{f_P}\right)$

$= 39.5 - 8\left(\frac{6}{7}\right) = 39.5 - 8\ (0.857...)$

= 39.5 - 6.857... = 32.642...

4. Rata-Rata Tertimbang (*Weighted Mean*)

Dalam beberapa kasus setiap nilai diberi beban, misalnya pada kasus perhitungan Indeks Prestasi, Nilai Penjualan Barang, dll

$$\boxed{\bar{x}_B = \frac{\sum_{i=1}^{n} B_i x_i}{\sum_{i=1}^{n} B_i}}$$

Di mana $\bar{x}_B$:rata-rata tertimbang

B_i : beban ke-i

x_i : data ke-i

n : banyak data

Contoh 1 :

Berikut adalah Transkrip Akademik seorang mahasiswa

Mata Kuliah	Nilai Mut u	Angka Mutu (x_i)	SK S (B_i)	$B_i x_i$
Pancasila	B	3	2	6
Teori Ekonomi	A	4	4	16
Bahasa Inggris	C	2	3	6
Manajemen	A	4	3	12
Σ		14	12	40

Indeks Prestasi = $\boxed{\bar{x}_B = \frac{\sum_{i=1}^{n} B_i x_i}{\sum_{i=1}^{n} B_i}}$ = $\frac{40}{12}$ = 3.33

5. Rata-Rata Geometrik (*Geometric Mean*)

Rata-rata geometrik digunakan untuk menghitung rata-rata laju pertumbuhan (*growth rate*), misalnya : pertumbuhan penduduk, penjualan, tingkat bunga dll.

$$\boxed{G = \sqrt[n]{x_1 \times x_2 \times x_3 \times \cdots \times x_n}}$$

atau

$$\boxed{\log\ G = \frac{\log\ x_1 + \log x_2 + \log\ x_3 + \cdots + \log x_n}{n}}$$

ingat G = antilog (log G)

Di mana G : rata-rata geometrik

x_i : data ke-i

n : banyak data

Contoh 2 :

Data pertumbuhan suku bunga dalam 5 hari kerja :

1.5 2.3 3.4 1.2 2.5 %

$G = \sqrt[n]{x_1 \times x_2 \times x_3 \times \cdots \times x_n}$ =

$$\log\ G = \frac{\log\ x_1 + \log x_2 + \log\ x_3 + \log\ x_4 + \log x_5}{5}$$

$$= \frac{\log\ 1.5 + \log 2.3 + \log\ 3.4 + \log\ 1.2 + \log 2.5}{5}$$

$$= \frac{0.176\ldots + 0.361\ldots + 0.531\ldots + 0.079\ldots + 0.397\ldots}{5}$$

$$= \frac{1.5464\ldots}{5} = 0.30928\ldots.$$

G = antilog 0.30928... = 2.03837....

Bandingkan dengan rata-rata hitung

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1.5 + 2.3 + 3.4 + 1.2 + 2.5}{5} = \frac{10.9}{5} = 2.18$$

C. UKURAN PENYEBARAN

1. Ragam/Varians (Variance) dan Simpangan Baku/Standar Deviasi (Standard Deviation)

a) Ragam dan Simpangan Baku untuk Ungrouped Data

POPULASI :

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{\mathrm{N}}$$ atau $$\sigma^2 = \frac{N\sum_{i=1}^{N} x_i^{\,2} - (\sum_{i=1}^{N} x_i)^2}{N^2}$$

dan $$\sigma = \sqrt{\sigma^2}$$

SAMPEL :

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

dan $$s = \sqrt{s^2}$$

x_i : data ke-i

μ : rata-rata populasi $\bar{x}$: rata-rata sampel

σ^2: ragam populasi s^2: ragam sampel

σ : simpangan baku populasi s : simpangan baku sampel

N : ukuran populasi n : ukuran sampel

Contoh 3 :

Data Usia 5 mahasiswa : 18 19 20 21 22 tahun

a. Hitunglah μ, σ^2 dan σ (anggap data sebagai data populasi)

b. Hitunglah $\bar{x}$, s^2 dan s (data adalah data sampel)

Jawab :

x_i	μ atau $\bar{x}$	(x_i -μ) atau (x_i - $\bar{x}$)	(x_i -μ)² atau (x_i - $\bar{x}$)²	$x_i^{\,2}$
18	20	-2	4	324
19	20	-1	1	361
20	20	0	0	400
21	20	1	1	441

	22	20	2	4	484
Σ	100	------	-------	10	2010

POPULASI :

N = 5 $\mu = \frac{100}{5}$ = 20

$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\text{N}}$ = $\frac{10}{5}$ = 2

$\sigma^2 = \frac{N\sum_{i=1}^{N} x_i^{\,2} - (\sum_{i=1}^{N} x_i)^2}{N^2}$ = $\frac{(5\times 2010) - 100^2}{5^2} = \frac{10050 - 10000}{25} = \frac{50}{25}$ =2

$\sigma = \sqrt{\sigma^2}$ = $\sqrt{2}$ = 1.414...

SAMPEL :

n = 5 $\bar{x} = \frac{100}{5}$ = 2 $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ = $\frac{10}{4}$ = 2.5

$s^2 = \frac{n\sum_{i=1}^{n} x_i^{\,2} - (\sum_{i=1}^{n} x_i)^2}{n(n-1)}$ = $\frac{(5\times 2010) - 100^2}{5\times 4} = \frac{10050 - 10000}{20} = \frac{50}{20}$ = 2.5

$s = \sqrt{s^2}$ = $\sqrt{2.5}$ =1.581...

b) Ragam dan Simpangan Baku untuk Grouped Data

POPULASI :

$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{k} f_i \times (x_i - \mu)^2}{\text{N}}} \quad \text{dan} \quad \boxed{\sigma = \sqrt{\sigma^2}}$$

SAMPEL :

$$\boxed{s^2 = \frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{n-1}} \quad \text{dan} \quad \boxed{s = \sqrt{s^2}}$$

x_i : Titik Tengah Kelas ke-i

f_i : frekuensi kelas ke-i

k : banyak kelas

$\bar{x}$: rata-rata sampel

μ : rata-rata populasi

σ²: ragam populasi

s²: ragam sampel

σ : simpangan baku populasi

s :simpangan baku sampel

N : ukuran populasi

n : ukuran sampel

Contoh 4 :

Rata -Rata (μ atau $\bar{x}$) = $\frac{1679}{50}$ = 33.58

Kelas	TTK x_i	Frek. f_i	$f_i\ x_i$	μ atau $\bar{x}$	(x_i -μ) atau (x_i - $\bar{x}$)	(x_i -μ)² atau (x_i - $\bar{x}$)²	f_i (x_i -μ)² atau f_i (x_i - $\bar{x}$)²
16 - 23	19.5	10	195	33.58	-14.08	198.2464	1982.4640
24 - 31	27.5	17	467.5	33.58	-6.08	36.9664	628.4288
32 - 39	35.5	7	248.5	33.58	1.92	3.6864	25.8048
40 - 47	43.5	10	435	33.58	9.92	98.4064	984.0640
48 - 55	51.5	3	154.5	33.58	17.92	321.1264	963.3792
56 - 63	59.5	3	178.5	33.58	25.92	671.8464	2015.5392
Σ	-----	50	1679	----	----------	-----------	6599.68

POPULASI : N = 50

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i \times (x_i - \mu)^2}{\mathrm{N}} = \frac{6599.68}{50} = 131.9936$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{131.9936} = 11.4888....$$

SAMPEL :

$$s^2 = \frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{n-1} = \frac{6599.68}{49} = 134.6873....$$

$$s = \sqrt{s^2} = \sqrt{134.6873\ldots} = 11.6054....$$

2. Koefisien Ragam

Koefisien Ragam = Koefisien Varians

Semakin besar nilai Koefisien Ragam maka data semakin bervariasi, keragamannya data makin tinggi.

Untuk Populasi $\rightarrow$Koefisien Ragam = $\frac{\sigma}{\mu} \times 100\%$

Untuk Sampel $\rightarrow$Koefisien Ragam = $\frac{s}{\bar{x}} \times 100\%$

Contoh :

$\bar{x}$ = 33.58 s = 11.6054

Koefisien Ragam =

$$\frac{s}{\bar{x}} \times 100\% = \frac{11.6054}{33.58} \times 100\% = 34.56\ \%$$

BAB IV

KONSEP PROBABILITAS

Banyak kejadian dalam kehidupan sehari-hari yang sulit diketahui dengan pasti, apalagi kejadian dimasa yang akan datang. Andaikan, apakah nanti malam akan turun hujan? Apakah penerbangan dengan maskapai Garuda pada pagi hari ini akan berangkat tepat waktu? Apakah besok akan terjadi demonstrasi? Begitu juga dalam percobaan statistika, kita tidak bias mengetahui dengan pasti hasil-hasil yang akan muncul. Meskipun kejadian-kejadian tersebut tidak pasti, kita bisa melihat fakta-fakta yang ada untuk menuju derajat kepastian atau derajat keyakinan bahwa sesuatu akan terjadi.

Pemikiran mengenai probabilitas diawali dari pertanyaan seorang bangsawan Prancis bernama Chevalier de Mere kepada Pascal (1623 – 1662). Ia ingin mengetahui bagaimana pola pembagian uang taruhan pada suatu perjudian jika permainannya terpaksa dihentikan sebelum selesai. Pertanyaan ini kemudian menjadi bahan diskusi antara Pascal dan Fermat (1601 – 1665), berdasarkan diskusi tersebut munculah teori-teori probabilitas. Walaupun dasar-dasar probabilitas awalnya muncul untuk menjelaskan masalah-masalah dalam perjudian, dalam perkembangannya, konsep probabilitas dapat diterapkan pada berbagai masalah seperti masalah social, teknik, kesehatan, biologi, industry, transportasi, manajemen, akutansi, pendidikan dll (Algifari, 2010)

Probabilitas adalah suatu besarnya kesempatan (kemungkinan) suatu peristiwa akan terjadi. Besarnya kesempatan dapat ditulis dalam bentuk bilangan decimal, pecahan atau persen.

Dengan demikian, kita dapat menentukan probabilitas terjadinya hujan, munculnya muka 1 pada percobaan pelemparan dadu, probabilitas munculnya kartu AS pada penarikan kartu dari sekelompok kartu Bridge dan seterusnya.

A. Perumusan Probabilitas

Perumusan konsep dasar probabilitas dilakukan dengan tiga cara, yaitu perumusan klasik, cara frekuensi relatif dan pendekatan subjektif. Bila kejadian-kejadian pada contoh di atas kita lambangkan dengan huruf besar E, kita dapat merumuskan probabilitas kejadian E, yaitu P(E).

1. Perumusan Klasik

Bila kejadian E terjadi dalam m cara dari seluruh n cara yang mungkin terjadi dan masing-masing n cara itu memiliki kesempatan atau kemungkinan yang sama untuk muncul, prrobabilitas kejadian E yang ditulis P(E) dirumuskan sebagai berikut :

Rumus 1.1

P(E) = m / n

Contoh :

1. Sebuah uang logam dilemparkan. Andaikan sisi pertama kita sebut muka (m) dan sisi kedua kita sebut belakang (b), maka ada dua kejadian yang mungkin, yaitu kejadian munculnya muka m yang kita sebut E={m} atau kejadian munculnya belakang yang kita sebut {b}. karena uang logam terdiri atas 2 sisi (n=2) dan kedua sisi itu memiliki kesempatan yang sama untuk muncul, probabilitas munculnya kejadian E={m} atau E{b} adalah

   P(E) = P(m) = m/n   P(E) = P(b) = m/n
   = 1/2   = 1/2

   Pada pelemparan uang logam tersebut yang akan muncul adalah salah satu dari E = {m} atau E = {b}.

2. Sebuah dadu dilemparkan. Muka dadu ada 6. Semua muka dadu memiliki kesempatan yang sama untuk muncul. Salah satu muka yang akan muncul dari muka-muka dadu itu (m=1) adalah muka dadu 1, muka dadu 2, muka dadu 3, muka dadu 4, muka dadu 5 atau muka dadu 6. Maka probabilitas kejadian E adalah :

   P(E) = P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = m/n = 1/6

2. Frekuensi Relatif

Perumusan konsep probabilitas dengan cara klasik memiliki kelemahan karena menuntut syarat semua hasil memiliki kesempatan yang sama untuk muncul. Pengertian ini mengaburkan adanya probabilitas yang sama. Sehubungan dengan itu dikembangkan konsep probabilitas berdasarkan statistic, yaitu dengan pendekatan empiris. Probabilitas empiris dari suatu kejadian dirumuskan dengan memakai frekuensi relatif dari terjadinya suatu kejadian dengan syarat banyaknya pengamatan atau banyaknya sampel n adalah sangat besar. Bila n bertambah besar sampai tak terhingga ($n \to \infty$), probabilitas kejadian E sama dengan nilai limit dari frekuensi relatif kejadian E tersebut. Dengan demikian, jika kejadian E berlangsung sebanyak f kali dari keseluruhan pengamatan sebanyak n, dimana n mendekati tak berhingga, probabilitas kejadian E dirumuskan sebagai berikut :

Rumus 1.2

$$\mathbf{P(E) = \lim_{n \to \infty} f / n}$$

Walaupun mudah dan berguna dalam praktek, secara matematis perumusan konsep probabilitas dengan frekuensi relative ini juga memiliki kelemahan karena suatu nilai limit yang benar-benar mungkin sebenarnya tidak ada. Oleh karena itu, konsep probabilitas modern dikembangkan dengan memakai pendekatan aksiomatis, yaitu suatu kebenaran yang diterima secara apa adanya tanpa memerlukan bukti matematis, dimana konsep probabilitas tidak didefinisikan, seperti konsep titik dan konsep garis yang tidak didefinisikan dalam ilmu geometri (Boediono, 2006).

Contoh :

1.Pada suatu percobaan statistic, yaitu pelemparan sebuah dadu yang diulang sebanyak 1000 kali (n=1000), frekuensi munculnya muka dadu X adalah seperti pada tabel berikut ini :

Muka dadu (X)	1	2	3	4	5	6
Frekuensi (f)	164	165	166	167	168	169

Bila E menyatakan kejadian munculnya muka-muka dadu tersebut, maka probabilitas kejadian E untuk masing-masing kemungkinan munculnya muka dadu tersebut adalah

P(E) = P(1) = 164/1000   P(E) = P(2) = 165/1000 P(E) = P(3) = 166/1000, dst

2.Dari 100 mahasiswa yang mengikuti ujian statistika, distribusi frekuensi nilai mahasiswa adalah seperti tabel berikut

Nilai (X)	45	55	65	75	85	95
Frekuensi (f)	10	15	30	25	15	5

Maka probabilitas kejadian E mahasiswa memperoleh nilai tersebut adalah

P(E) = P(45) = 10/100 P(E) = P(55) = 15/1001 P(E) = P(65) = 30/100, dst

3. Pendekatan Subjektif

Pendekatan subjektif yang digunakan untuk menentukan probabilitas suatu peristiwa didasarkan pada selera dan keyakinan individu seseorang. Misalnya, saya ingin menentukan bahwa besok probabilitas naiknya harga dolar Amerika adalah 0.75 atau 75%. Atas dasar apa saya menentukan probabilitas naiknya harga dolar itu 75%? Pengetahuan ini hanya didasarkan pada pengetahuan, pengalaman, dan keahlian yang dimiliki. Dengan demikian, probabilitas suatu peristiwa yang ditentukan dengan pendekatan subjektif menyebabkan penentuan probabilitas suatu peristiwa antara orang yang satu dengan yang lain dapat berbeda. Hal ini disebabkan oleh tingkat pengetahuan, penguasaan informasi, naluri dan faktor-faktor lain yang berkaitan dengan peristiwa itu.

B. Ruang Sampel dan Kejadian

Pada pelemparan sebuah uang logam, ada dua hasil yang mungkin muncul, yaitu muka (m) atau belakang (b). dua hasil yang mungkin muncul ini dapat dihimpun menjadi S = {m,b}. dengan demikian dapat dikatakan bahwa kumpulan himpunan dari semua hasil yang mungkin muncul atau terjadi pada suatu percobaan statistic disebut ruang sampel, yang dilambangkan dengan himpunan S, sedangkan anggota-anggota dari S disebut titik sampel.

Perhatikan bahwa pada pelemparan sebuah uang logam tersebut S = {m,b} dan A = {m}, sehingga A c A, A adalah suatu himpunan bagian dari S. berdasarkan kejadian A dan ruang sampel S tersebut, perumusan konsep probabilitas didefinisikan sebagai berikut. Bila kejadian A berlangsung dalam m cara pada ruang sampel S yang terjadi dalam n cara, probabilias kejadian A adalah :

Rumus 1.3

$$\mathbf{P(A) = n(A) / n(S) = m / n}$$

Dimana n(A) = banyaknya anggota A dan n(S) = banyaknya anggota S
Perhatikan bahwa definisi probabilitas tersebut tidak menuntut syarat bahwa semua titik sampel memiliki kesempatan yang sama untuk muncul. Definisi probabilitas kejadian ini terlepas dari definisi probabilitas yang dirumuskan secara klasik maupun memakai frekuensi relative. Dengan menggunakan rumus 1.3, kita dapat menentukan probabilitas dari sembarang kejadian A yang didefinisikan pada S.

Contoh :

1. Pada pelemparan sebuah dadu, andaikan kejadian A menyatakan munculnya muka dadu genap pada S, A = {2,4,6} sehingga probabilitas kejadiab A adalah P(A) = 3/6 = ½

2. Pada pelemparan dua buah uang logam :
   a. Tentukanlah ruang sampel S
   Hasil-hasil yang mungkin muncul adalah sebagai berikut :

Uang Logam 1	Uang Logam 2	
	M	B
M	(M,M)	(M,B)
B	(B,M)	(B,B)

   Jadi ruang sampel S adalah = {(m,m), (m,b), (b,m), (b,b)}
   Titik sampel (m,m) menyatakan munculnya sisi muka dari uang logam pertama dan kedua, titik sampel (m,b) menyatakan munculnya muka dari uang logam pertama dan belakang dari uang logam kedua, begitu seterusnya.
   b. Bila A menyatakan kejadian munculnya sisi yang sama dari dua uang logam tersebut, tentukanlah probabilitas kejadian A
   A adalah kejadian munculnya sisi-sisi yang sama dari dua uang logam, maka A = {(m,m), (b,b)}. Dengan demikian, n(A) = 2 dan n(S) = 4, sehingga probabilitas kejadian A adalah
   P(A) = n(A) / n(S) = 2/4 = 1/2

TUGAS 1 :

1. Sebuah kotak berisi 8 bola merah, 7 bola putih, dan 5 bola biru. Jika diambil 1 bola secara acak, tentukanlah probabilitas terpilihnya :
   - Bola merah
   - Bola biru
   - Bola putih
   - Bola merah atau biru

2. Peluang seorang pria akan hidup selama 25 tahun adalah 3/5 dan peluang istrinya akan hidup selama 25 tahun adalah 2/3. Tentukanlah peluang :
   - Keduanya akan hidup selama 25 tahun
   - Hanya pria yang hidup selama 25 tahun
   - Hanya istri yang hidup selama 25 tahun
   - Paling sedikit salah satu dari mereka (suami/istri) hisup selama 25 tahun

3. Tiga wanita dipilih secara acak untuk ditanya apakah mereka mencuci pakaian dengan detergen. Tentukanlah :
   - Anggota ruang sampel S dengan memakai huruf Y = ya dan T = tidak

- Tulislah anggota kejadian E dalam S yang menyatakan bahwa paling sedikit dua wanita memakai detergen
- Hitunglah P(E)

4. Dalam pengumpulan nilai probabilitas dan statistika mahasiswa jurusan SI STT NIIT I-Tech, diperoleh daftar nilai sebagai berikut :

ai	40	50	60	70	80	90	100
ekuensi	3	4	5	8	2	2	1

Jika kita mengambil 1 nilai secara random, berapa probabilitas dari :

- Mahasiswa yang memperoleh nilai diatas 60
- Mahasiswa yang memperoleh nilai antara 60 dan 80 (60 < nilai < 80)

BAB V
DISTRIBUSI VARIABEL RANDOM

A. DISTRIBUSI PROBABILITAS DISKRET

1.Pendahuluan

Distribusi teoretis adalah suatu alat bagi kita untuk menentukan apa yang dapat kita harapkan apabila asumsi-asumsi yang kita buat benar. Distribusi frekuensi dapat digunakan sebagai dasar pembanding, dari suatu hasil observasi atau eksperimen, dan sering juga digunakan sebagai pengganti distribusi sebenarnya. Hal ini penting sekali karena, selain sangat mahal, distribusi sebenarnya yang harus diperoleh melalui eksperimen sering kali tidak dapat dilakukan. Distribusi teoretis memungkinkan para pembuat keputusan untuk memperoleh dasar logika yang kuat di dalam keputusan, dan sangat berguna sebagai dasar pembuatan ramalan berdasarkan informasi yang terbatas atau pertimbangan-pertimbangan teoretis dan berguna pula untuk menghitung probabilitas terjadinya suatu peristiwa.

Pengertian mengenai beberapa distribusi yang utama akan meningkatkan kemampuan seseorang untuk membaca atau mengartikan hasil karya ilmiah hampir di semua bidang ilmu pengetahuan. Setiap kejadian yang dapat dinyatakan sebagai perubahan nilai suatu variabel umumnya mengikuti distribusi teoretis tertentu dan apabila sudah ketahuan jenis distribusinya, kita dengan mudah dapat mengetahui besarnya nilai probabilitas terjadinya peristiwa tersebut. Beberapa distribusi teoretis yang akan dibahas antara lain distribusi seragam, distribusi binomial, distribusi multinomial, distribusi hipergeometrik, dan distribusi poisson yang adalah suatu distribusi peubah acak yang bersifat diskret. Sedangkan distribusi kontinu terdiri atas distribusi normal, distribusi student dan khi-kuadrat.

2.Distribusi Seragam

Distribusi seragam (*uninformly distribution*) adalah suatu distribusi probabilitas yang paling sederhana diantara distribusi-distribusi probabilitas yang lain. Dalam distribusi ini setiap nilai peubah acak memiliki probabilitas terjadi yang sama. Distribusi seragam dapat pula didefinisikan seperti berikut. Bila peubah acak X memiliki nilai-nilai X_1, X_2, …X_k, dengan probabilitas yang sama, distribusi seragam diskret dinyatakan sebagai

Rumus 1.1

$$\mathbf{P(x : k) = 1/k \text{ untuk } x = x_1, x_2, \ldots, x_k}$$

Kita menggunakan notasi p(x, k), alih-alih p(x) untuk menunjukkan bahwa distribusi seragam bergantung pada parameter k.

Contoh :

1) Sebuah dadu setimbang dilemparkan sekali. Bila x menyatakan mata dadu yang muncul, buatlah distribusi probabilitas x!

   Jawab

   Ruang contoh S = {1,2,3,4,5,6} dan setiap mata dadu memiliki probabilitas yang sama untuk muncul, yaitu 1/6. Dengan demikian distribusi seragamnya adalah

   p(x : 6) = 1/6 untuk x = 1,2,3,4,5,6

2) Tim bulu tangkis terdiri atas 8 orang. Bila dari tim tersebut dipilih 2 orang secara acak untuk melakukan pertandingan, tentukan distribusi seragam yang diambil secara acak!

   Jawab

   Jumlah dalam satu tim 8 orang, maka kita mengambil 2 orang secara acak dalam (8 2) = 28 orang. Bila cara masing-masing diberi nomor 1 sampai 28, distribusi probabilitasnya adalah

   p(x : 28) = 1/28 untuk x = 1, 2,…, 28

3.Distribusi Binomial

Beberapa percobaan sering kali terdiri atas ulangan-ulangan yang memiliki dua kejadian, yaitu berhasil atau gagal. Percobaan ini adalah suatu percobaan dengan pemulihan (*with replacement*), yaitu setiap cuplikan yang telah diamati dimasukkan kembali dalam populasi semula. Populasi setelah pencuplikan tetap sama, artinya susunan anggota populasi dan nisbah setelah pencuplikan tidak pernah berubah. Seorang petugas pengendali mutu ingin menghitung probabilitas untuk mendapatkan 4 bola lampu yang rusak dari suatu sampel acak sebanyak 20 bola lampu apabila bahwa 10% dari bola lampu tersebut rusak. Nilai probabilitas ini dapat diperoleh dari tabel binomial yang dibuat berdasarkan distribusi binomial (Supranto, 2006).

Percobaan-percobaan pada distribusi binomial bersifat bebas dan probabilitas keberhasilan setiap ulangan tetap sama. Distribusi binomial adalah suatu suatu distribusi probabilitas peubah acak yang bersifat diskret. Distribusi ini sering disebut proses Bernoulli (*Bernoulli Trials*). Nama ini diambil dari seorang ahli matematika berkebangsaan Swiss, yaitu James Bernoulli (1654 – 1705). Pada umumnya, suatu eksperimen atau percobaan dapat dikatakan eksperimen atau percobaan binomial apabila memiliki beberapa syarat berikut :

a. Setiap percobaan selalu dibedakan menjadi dua macam kejadian yang bersifat saling meniadakan (*mutually exclusive*) b.Dalam setiap percobaan hasilnya dapat dibedakan, yaitu berhasil atau gagal

c. Probabilitas kejadian berhasil dinyatakan dengan huruf p, sedangkan probabilitas gagal dinyatakan dengan huruf q, dimana $p + q = 1$ atau $q = 1 - p$

d.Masing-masing percobaan adalah suatu peristiwa yang bersifat bebas, yaitu peristiwa yang satu tidak dapat mempengaruhi peristiwa yang lain

Misalnya, keluarga Markus merencanakan memiliki 3 anak seperti yang terlihat pada tabel 1.1. Disini setiap kelahiran anak laki-laki dikatakan "berhasil" dan setiap kelahiran anak perempuan dikatakan "gagal". Dengan demikian, banyaknya anak laki-laki dipandang sebagai sebuah peubah acak x yang mengambil bilangan 0 sampai 3. Peubah acak x yang adalah suatu banyaknya keberhasilan dalam setiap percobaan disebut *peubah acak binomial.*

Tabel 1.1 hasil "percobaan" keluarga Markus

Ruang contoh	Peubah X	Probabilitas	
PPP	0	1/8	
LPP	1	**1/8**	
PLP	1	**1/8**	**= 3/8**
PPL	1	**1/8**	
LLP	2	**1/8**	
LPL	2	**1/8**	**= 3/8**
PLL	2	**1/8**	
LLL	3	1/8	

Selanjutnya, ilustrasi keluarga Markus di atas akan kita generalisasi dengan mencari rumusan yang lebih umum dari distribusi binomial. Bila kelahiran anak laki-laki dinyatakan sebagai x, probabilitas kelahiran anak laki-laki memiliki nilai yang tetap, yaitu ½. Probabilitas kelahiran anak laki-laki yang dipandang berhasil adalah x dengan probabilitas p dan sebaliknya, setiap kegagalan yaitu kelahiran anak perempuan, adalah $(n - x)$ dengan probabilitas $q = 1 - p$. Dengan demikian, probabilitas untuk urutan tertentu dinyatakan dengan $p^x \cdot q^{n-x}$

Sekarang tinggal menghitung banyaknya kombinasi yang memiliki keberhasilan x dan kegagalan $(n - x)$. Bilangan ini tidak lain adalah bentuk kombinasi. Selanjutnya, banyaknya kombinasi ini dikalikan dengan $p^x \cdot q^{n-x}$ untuk mendapatkan rumus distribusi binomial. Dengan kata lain, jika suatu percobaan binomial memiliki probabilitas keberhasilan p dan probabilitas kegagalan q, distribusi probabilitas peubah acak x adalah banyaknya keberhasilan dalam n percobaan yang bebas dan dinyatakan oleh

Rumus 1.2

$$b(x : n : p) = (n\ x)\ p^{x} \cdot q^{n-x} \text{ dengan } x = 0, 1, 2, \dots n$$

Tabel 1.2 Koefisien probabilitas distribusi binomial

Peubah X	Koefisien Distribusi binomial	Polinomial
0	1	$(p+q)^{0}$
1	$p + q$	$(p+q)^{1}$
2	$p^2 + 2pq + q^2$	$(p+q)^{2}$
3	$p^3 + 3p^2q^1 + 3p^1q^2 + q^3$	$(p+q)^{3}$
4	$p^4 + 4p^3q^1 + 6p^2q^2 + 4p^1q^3 + q^4$	$(p+q)^{4}$
5	$p^5 + 5p^4q^1 + 10p^3q^2 + 10p^2q^3 + 5p^1q^4 + q^5$	$(p+q)^{5}$
…	…………………………………………………………………..	………
n	$p^n + np^{n-1}q^1 + \dots + np^1q^{n-1} + q^n$	$(p+q)^{n}$

Misalnya, besarnya probabilitas keluarga Markus dengan 2 anak laki-laki dari 3 anak yang dimiliki adalah

$b(2 : 3 : ½) = (3\ 2)\ (1/2)^2\ (1-1/2)^{3-2} = 3!\ /\ 2!\ (3-2)!\ (1/2)^2(1/2)^1 = 3/8$

Perumusan 1.2 dapat dirangkum dalam bentuk tabel probabilitas binomial bagi peubah acak x yang memuat kombinasi yang mungkin terjadi.

Nilai rata-rata dan varian distribusi binomial pada dasarnya ditentukan oleh berbagai macam peristiwa yang dihasilkan dari percobaan binomial, terutama probabilitas keberhasilan atau kegagalannya. Andaikan hasil percobaan ke n dinyatakan peubah acak L_n dengan probabilitas p keberhasilan $L_n = 1$ dan probabilitas q kegagalan $L_n = 0$. Suatu percobaan binomial banyaknya keberhasilan dituliskan sebagai jumlah n peubah acak bebas :

$$x = L_1 + L_2 + \dots + L_n$$

Nilai harapan setiap L_n adalah $E(L_n) = 1\ (p) + 0\ (q) = p$ sehingga rata-rata suatu populasi distribusi binomial dapat dinyatakan sebagai perkalian n percobaan dengan probabilitas percobaan.

Rumus 1.3

$$\mu = E(x) = E(L_1) + E(L_2) + \dots + E(L_n)$$
$$= p + p + \dots + p = n.p$$

Sementara besarnya ragam distribusi binomial dapat dicari dari hubungan berikut. Ragam populasi untuk setiap Li adalah

$$\delta^2_{Li} = E\ [(L_i - p)^2] = E(L_i^2) - p^2$$

$= (1)^{2p} + (0)^2 q - p^2 = p.q$

Dengan demikian, total ragam populasi distribusi binomial dirumuskan sebagai berikut :

Rumus 1.4

$$\delta^2 = \delta^2_{L1} + \delta^2_{L2} + \ldots + \delta^2_{Ln} = p.q + p.q + \ldots = npq$$

dan simpangan bakunya adalah

Rumus 1.5

$$\delta = \sqrt{n\,p\,q}$$

Contoh :

1) Keluarga Markus berencana memiliki 3 anak. Bila X menyatakan banyaknya kelahiran anak laki-laki, hitunglah
   a. Probabilitas kelahiran 2 anak laki-laki
   b. Probabilitas memiliki tidak lebih dari 2 anak laki-laki
   c. Rata-rata dan simpangan baku peubah acak X

   Jawab

   Probabilitas kelahiran anak laki-laki sama dengan anak perempuan, p,q = ½ dan n = 3

   a. Probabilitas lahir 2 anak laki-laki

   $p(x = 2) = b(x : n : p) = (n\ x)\ p^x \cdot q^{n-x}$
   $= b(2 : 3 : ½) = (3\ 2)\ (1/2)^2 \cdot (1/2)^{3-2}$
   $= 3! / 2!\ (3 - 2)! \cdot (½)^{2+1}$
   $= 3! / 2!\ 1! \cdot (1/2)^3$
   $= 3 \cdot (½)^3$
   $= 3 \cdot 0.125 = 0.375$

   b. Tidak lebih dari 2 anak laki-laki

   $p(x \le 2)$ dimana x = 0, 1 dan 2

   $b(0 : 3 : ½) = (3\ 0)\ (1/2)^0 \cdot (1/2)^{3-0}$
   $= 3! / 0!\ (3 - 0)! \cdot (½)^{0+3}$
   $= 3! / 3! \cdot (1/2)^3$
   $= 0.125$

   $b(1 : 3 : ½) = (3\ 1)\ (1/2)^1 \cdot (1/2)^{3-1}$
   $= 3! / 1!\ (3 - 1)! \cdot (½)^{1+2}$
   $= 3! / 1!\ 2! \cdot (1/2)^3$
   $= 0.375$

   $b(2 : 3 : ½) = (3\ 2)\ (1/2)^2 \cdot (1/2)^{3-2}$
   $= 3! / 2!\ (3 - 2)! \cdot (½)^{2+1}$
   $= 3! / 2!\ 1! \cdot (1/2)^3$
   $= 3 \cdot (½)^3$

$= 3 \cdot 0.125 = 0.375$

Sehingga $p(x \leq 2) = 0.125 + 0.375 + 0.375$

$= 0.875$

Dapat juga diselesaikan dengan bantuan tabel distribusi binomial

$p(x \leq 2) = \Sigma$ n= 0..2 b (x : 3 : 0.5)

= b (0 : 3 : ½) + b (1 : 3 : ½) + b (2: 3 : ½)

= 0.1250 + 0.375 + 0.375

= 0.875

c. Rata-rata, ragam dan simpangan baku kelahiran anak laki-laki

Rata-rata, $\mu = n \cdot p = 3 \cdot ½ = 1.5$, dengan n = 3 dan p = ½

Simpangan baku, $\delta = \sqrt{n \cdot p \cdot q} = \sqrt{3 \cdot ½ \cdot ½} = 0.866$

Jadi, dalam kelahiran 3 anak, rata-rata anak laki-laki yang dilahirkan adalah 1.5 dengan simpangan baku sebesar 0.866

2) Menurut penelitian, probabilitas seseorang untuk sembuh dari penyakit antraks dengan pemberian obat tertentu adalah sebesar 60%. Jika diambil 10 orang yang terjangkit secara acak, hitunglah :
   a. Probabilitas tidak lebih dari 3 orang sembuh
   b. Sedikitnya 5 orang sembuh
   c. Rata-rata dan simpangan baku pasien sembuh

Jawab

n = 10, p = 60% = 0.6, q = 1 – p = 40% = 0.4

a. Tidak lebih dari 3 orang dapat sembuh

$p(x \leq 3) = \Sigma$ n= 0..3 b (x : 10 : 0.6)

= b (0 : 10 : 0.6) + b (1 : 10 : 0.6) + b (2 : 10 : 0.6) + b (3 : 10 : 0.6)

= 0.0001 + 0.0016 + 0.0106 + 0.0425

= 0.548

b. Sedikitnya 5 orang dapat sembuh

$p(x \geq 5) = 1 - (\Sigma$ n= 0..3 b (x : 10 : 0.6) + b (4 : 10 : 0.6))

= 1 – (0.548 + 0.1114)

= 0.3406

c. Rata-rata, ragam dan simpangan baku pasien dapat sembuh

Rata-rata $\mu = 10 (0.6) = 6$

Simpangan baku, $\delta = \sqrt{10 \cdot 0.6 \cdot 0.4} = 1.55$

4.Distribusi Multinomial

Bila suatu percobaan binomial terhadap ulangannya menghasilkan lebih dari dua kemungkinan berhasil, “nyaris berhasil” atau gagal, percobaan itu menjadi percobaan multinomial. Dengan kata lain, bila pada distribusi binomial hasil sebuah percobaan hanya dikategorikan dua macam, yaitu berhasil atau gagal, dalam distribusi multinomial sebuah percobaan akan menghasilkan beberapa kejadian (lebih dari dua) yang saling meniadakan atau saling lepas (*mutually exclusive*).

Sebagai contoh, keadaan cuaca dapat digolongkan menjadi cerah, hujan, atau mendung. Pilihan kendaraan untuk ke kantor adalah mobil sendiri, bus, kereta api, angkot bahkan ojek. Seluruhnya adalah suatu ulangan-ulangan yang menghasilkan lebih dari dua kemungkinan. Secara umum, bila setiap ulangan dapat menghasilkan satu diantara k kemungkinan hasil percobaan E_1, E_2, …, E_k kali kejadian dalam n ulangan yang bebas dengan $x_1 + x_2 + \ldots + x_k = n$. sedangkan banyaknya sekatan n elemen ke dalam k kelompok dengan x_1 dalam kelompok pertama, x_2 dalam kelompok kedua, … dan x_k dalam kelompok ke k adalah suatu suatu permutasi dari n elemen yang seluruhnya tidak dapat dibedakan. Dengan demikian, probabilitas distribusi multinomial dapat dirumuskan secara matematik dengan persamaan berikut

Rumus 1.6

$$\mathbf{b(x_1, x_2, \ldots x_n : n : p_1, p_2, \ldots, p_k) = (n\ x_1, x_2, \ldots x_k)\ p_1^{x1}\ p_2^{x2} \ldots p_k^{xt}}$$

dengan probabilitas suku-suku pengurai multinomial $p_1 + p_2 + \ldots + p_k = 1$

Contoh :

1.Dalam pemilu legislatif, para konstituen memiliki pilihan mencoblos 3 partai politik dengan probabilitas pilihan : PAN 0.5, Partai Demokrat 0.3, GOLKAR 0.2. berapa probabilitas bahwa di antara 10 konstituen sebanyak 4 konstituen memilih PAN, 3 konstituen memilih PD dan 3 konstituen memilih GOLKAR

Jawab

Kita daftar kejadian yang mungkin

E_1 = 4 konstituen memilih PAN

E_2 = 3 konstituen memilih PD

E_3 = 3 konstituen memilih GOLKAR

Setiap ulangan dengan probabilitas masing-masing, $p_1 = 0.5$, $p_2 = 0.3$ dan $p_3 = 0.2$ oleh karena $x_1=4$, $x_2=3$ dan $x_3=3$, distribusi multinomial adalah

$b(4, 3, 3 : 10 : 0.5, 0.3, 0.2) = (10\ 4,3,3)\ (0.5)^4\ (0.3)^3\ (0,2)^2$

$= 10! / 4!\ 3!\ 3!\ (0.0625)\ (0.027)\ (0.008)$

= 0.057

B. DISTRIBUSI NORMAL

1. Pendahuluan

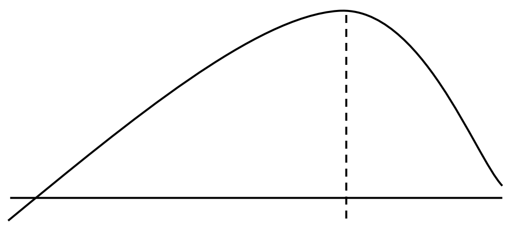
Pada uraian sebelumnya telah dibahas bagaimana cara menentukan probabilitas peristiwa dari suatu percobaan yang menggunakan variabel acak diskret dan dengan nilai yang terbatas. Apabila suatu percobaan menggunakan variabel acak yang kontinu dan nilai yang tidak terbatas, diperlukan distribusi probabilitas kontinu untuk menentukan probabilitas suatu peristiwa yang akan dihasilkan dari percobaan tersebut. Satu hal yang sangat penting dalam distribusi probabilitas kontinu adalah distribusi normal. Sekumpulan nilai data akan terdistribusi secara normal (membentuk kurva yang simetris) apabila rata-rata nilai variabel sama dengan median dan sama dengan modus nilai data tersebut. Distribusi probabilitas normal membentuk suatu kurva normal yang juga sering disebut kurva genta (*bell-shaped curve*) karena bentuknya yang menyerupai sebuah genta (Haryono, 2008).

Ada dua alasan mengapa distribusi normal sering digunakan dalam analisa statistik, yaitu :

1. Distribusi normal memiliki kemampuan yang dapat diterapkan pada banyak situasi, terutama untuk membuat kesimpulan dari sampel yang digunakan.
2. Distribusi normal sangat baik digunakan dalam analisis tentang fenomena yang menggunakan data kontinu, seperti ukuran berat, tinggi rendahnya skor IQ, panjang, jumlah curah hujan, banyaknya botol dalam satu kerat dan lain sebagainya.

Kita telah mengetahui bahwa ada 3 jenis kemiringan, yaitu :

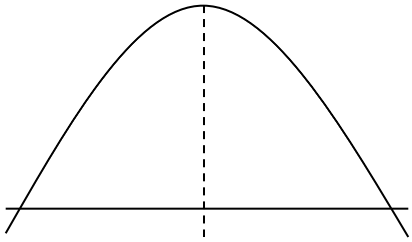
1. Distribusi miring ke kiri (x < med < mod)



Gambar 8.1

Gambar 8.1 menunjukkan distribusi data miring ke kiri, dimana nilai rata-rata hitung lebih kecil dari median dan median lebih kecil dari modus. Kurva tidak simetris sebab puncaknya ada di bagian kanan, tetapi ada sedikit data yang menyebar ke kiri.

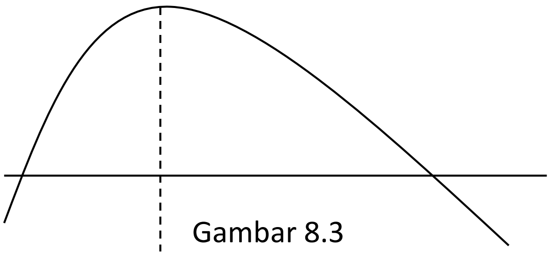
2. Distribusi simetri (med = mod = x)



Gambar 8.2

Gambar 8.2, dimana nilai rata-rata sama atau mendekati median dan modus. Kurvanya simetris dengan puncak distribusi ada dibagian tengah. Distribusi ini disebut dengan distribusi normal.

3.Distribusi miring ke kanan (med < mod < x)



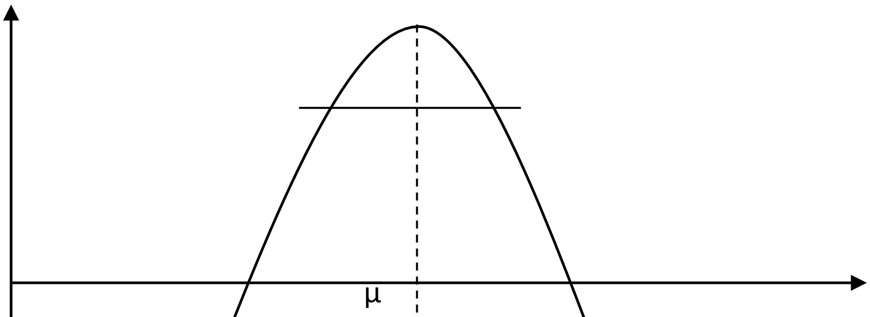
Gambar 8.3

Gambar 8.3 menunjukkan keadaan terbalik dari gambar 8.1, yaitu distribusi data miring ke kanan, dimana nilai modus lebih kecil dari median dan median lebih kecil dari nilai rata-rata hitung. Kurva juga tidak simetris sebab puncaknya ada dibagian kiri, sementara ada sedikit data yang menyebar ke kanan.

2. Distribusi Normal Selengkapnya

Pada pembahasan sebelumnya telah diuraikan empat bentuk distribusi probabilitas variabel diskret, yaitu distribusi binomial, poisson, multinomial, dan hipergeometrik. Variabel diskret adalah variabel yang nilainya bilangan bulat (0, 1, 2, 3,...). Nilai variabel diskret tidak dapat minus dan juga tidak berupa pecahan. Pembahasan berikutnya adalah probabilitas peristiwa untuk variabel acak kontinu yang berbentuk simetris dan memiliki poros di tengah-tengah distribusi. Probabilitas suatu peristiwa yang berdistribusi normal dari variabel acak kontinu ditunjukkan oleh daerah di bawah kurva normal. Pada suatu observasi, berapapun nilai rata-rata dan nilai standar deviasinya, luas seluruh daerah di bawah kurva normal adalah 1. Hal ini sesuai dengan ketentuan nilai probabilitas semua kemungkinan peristiwa yang akan terjadi dalam suatu percobaan adalah 1. Poros kurva normal terdapat pada rata-rata dan populasi (μ).

Distribusi normal disebut juga distribusi Gauss, lihat kembali kurva distribusi normal berikut :



Gambar 8.4 kurva distribusi normal

Rata-rata populasi membagi dua data sama banyaknya sehingga luas daerah di sebelah kiri rata-rata adalah 0.5 dan di sebelah kanan rata-rata juga 0.5. karakteristik kurva normal yang dihubungkan dengan nilai rata-rata dan nilai standar deviasi data adalah sebagai berikut :

1.Sekitar 68% nilai data observasi yang terdistribusi secara normal berada di dalam interval $\mu \pm 1$ standar deviasi

2.Sekitar 95% nilai data observasi yang terdistribusi secara normal berada di dalam interval $\mu \pm 2$ standar deviasi

Distribusi probabilitas normal untuk setiap nilai x yang membentuk kurva normal memiliki persamaan umum :

Rumus 8.1

$$f(x) = \frac{1}{\delta\sqrt{2\pi}}\; e^{-1/2\left(\frac{x-\mu}{\delta}\right)^2}$$

Keterangan :

μ = rata-rata populasi

δ = simpangan baku populasi

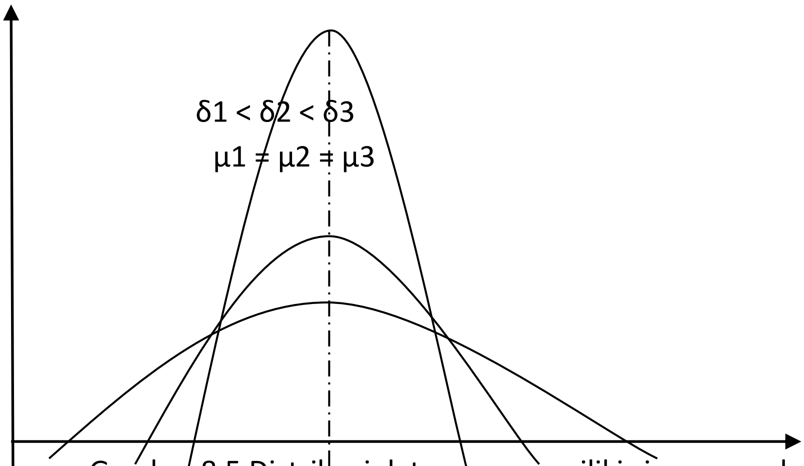
π = konstanta yang nilainya mendekati 3.14159

e = konstanta yang nilainya mendekati 2,7182

x = setiap nilai variabel acak kontinu yang besarnya $-\infty$ sampai dengan $+\infty$

Distribusi normal f(x) didefinisikan pada internal terbuka $-\infty < x < +\infty$. Distribusi normal dengan parameter μ dan δ^2 biasanya ditulis N (μ,δ^2). Dengan memperhatikan persamaan umum dan grafik distribusi normal f(x), tampak bahwa bentuk kurva normal ditentukan oleh dua parameter, yaitu rata-rata (μ) dan simpangan baku (δ). Bila nilai δ mengecil, bentuk kurva akan lebih rapat dan semakin runcing, dan sebagian besar nilai x akan berkumpul atau mendekati nilai rata-rata μ. Sebaliknya, jika nilai δ

semakin besar, bentuk kurva akan lebih renggang dan tumpul, dimana sebagian besar nilai x akan menjauhi nilai rata-rata μ. Perhatikan gambar 8.5 yang menunjukkan uraian tiga distribusi data yang memiliki simpangan baku serta rata-rata.



Gambar 8.5 Distribusi data yang memiliki simpangan baku berbeda

3. Sifat-sifat Distribusi Normal

Ada beberapa sifat penting dari distribusi normal, diantaranya :

1.Grafik simetri terhadap garis tegak x = μ

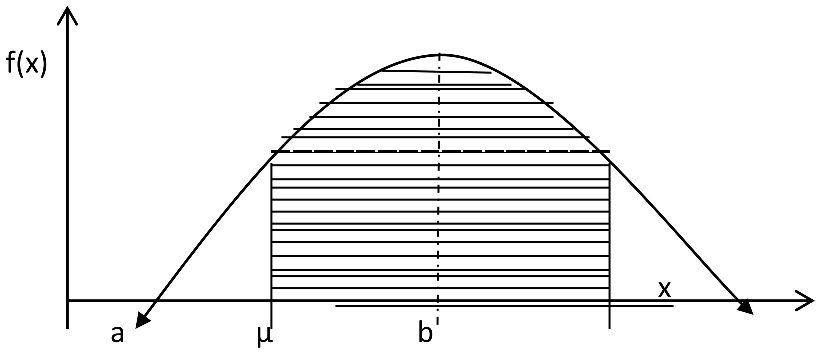
2.Grafik selalu berada diatas sumbu x atau f(x) > 0

3.Memiliki satu nilai modus

4.Luas daerah di bawah kurva f(x) dan di atas sumbu x = 1, yaitu P(-∞< x < +∞) = 1

4. Probabilitas P(A < x < B)

Probabilitas distribusi normal f(x) pada interval a < x < b ditentukan dengan memakai luas daerah di bawah kurva f(x), sebagaimana ditunjukkan oleh gambar 8.6



Gambar 8.6 Luas daerah di bawah kurva

Pada gambar 8.6 probabilitas $P(a<x<b)$ ditunjukkan oleh luas daerah yang diarsir, yang dibatasi oleh kurva f(x), sumbu x, garis tegak x=a dan x=b. Oleh karena f(x) adalah suatu fungsi kontinu, probabilitas $P(a<x<b)$ dihitung dengan menggunakan integral dari fungsi f(x) yang dibatasi oleh x=a dan x=b, yaitu :

Rumus 8.2

$$P(a<x<b) = \int f(x)\,dx = \int \frac{1}{\delta\sqrt{2\pi}}\, e^{-1/2 \frac{(x-\mu)^2}{\delta}}\, dx$$

Rumus integral tersebut sangat berguna untuk menghitung daerah di bawah kurva distribusi normal standar. Akan tetapi, secara matematis bentuk integral dari fungsi f(x) tersebut sulit dipecahkan secara langsung dengan teknik integral. Oleh karena itu, penyelesaiannya dilakukan dengan memakai transformasi nilai-nilai X menjadi nilai-nilai baku Z, yaitu :

Rumus 8.3

$$Z = \frac{x - \mu}{\delta}$$

Dengan transformasi tersebut, kita memperoleh distribusi normal Z yang memiliki rata-rata $\mu = 0$ dan simpangan baku $\delta = 1$, atau ditulis N(0,1). Distribusi normal Z seperti ini disebut distribusi normal standar. Dengan demikian, fungsi distribusi f(x) berubah menjadi fungsi distribusi f(z).

Rumus 8.4

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-½ Z2}$$

Selanjutnya probabilitas $P(z1 < Z < z2)$ dihitung dengan rumus berikut :

Rumus 8.5

$$P(z1 < Z < z2) = \int f(z)\,dz = \int \frac{1}{\sqrt{2\pi}}\, e^{-½ Z2}\, dz$$

Berdasarkan integral dari distribusi normal standar tersebut, probabilitas $P(z1 < Z < z2)$ dihitung dengan memakai tabel distribusi normal standar. Perhatikan bahwa nilai-nilai yang ada dalam tabel tersebut menunjukkan probabilitas dari nilai-nilai Z mulai dari z = 0 sampai dengan $z = Z_0$ (positif) yaitu $P(0 < Z < Z_0)$.

Contoh :

Tentukan probabilitas dari nilai Z berikut (gunakan tabel distribusi normal standar)

1) $P(0 < Z \leq 1.54) = 0.4382$
2) $P(-2.53 \leq Z < 0)$

   Karena fungsi distribusi normal standar simetri terhadap Z = 0, maka probabilitas $P(-2.53 \leq Z < 0) = P(0 < Z < 2.53) = 0.4943$

3) $P(-1.62 \leq Z \leq 1.62)$

   Karena $P(-1.62 \leq Z \leq 0) = P(0 \leq Z \leq 1.62)$, $P(-1.62 \leq Z \leq 1.62)$

   $= 2P(0 \leq Z \leq 1.62)$

   $= 2(0.4474)$

   $= 0.8948$

4) $P(-2.75 < Z < -1.52)$

   Perhatikan bahwa $P(-2.75 < Z < -1.52) = P(1.52 < Z < 2.75)$

   $= P(0 < Z < 2.75) – P(0 < Z < 1.52)$

   $= 0.4970 – 0.4357$

   $= 0.0613$

5) $P(1.42 < Z < 2.54)$

   Perhatikan bahwa $P(1.42 < Z < 2.54) = P(0 < Z < 2.54) – P(0 < Z < 1.42)$

   $= 0.495 – 0.4222$

   $= 0.0723$

6) Bila X adalah variabel acak berdistribusi normal dengan rata-rata $\mu = 25$ dan simpangan baku $\delta = 10$, tentukanlah probabilitas $P(20 < X < 38)$

   Jawab :

   $$Z = \frac{x - \mu}{\delta} = \frac{X - 25}{10}$$

   maka diperoleh

   $Z1 = (20 – 25) / 10 = -0.5$ dan $Z2 = (38 – 25)/10 = 1.3$

   Dengan demikian probabilitas $P(20 < X < 38) = P(-0.5 < Z < 1.3)$

   $= P(-0.5 < Z < 0) + P(0 < Z < 1.3)$

   $= P(0 < Z < 0.5) + P(0 < Z < 1.3)$

   $= 0.1915 + 0.4032$

   $= 0.5947$

5. Fungsi Distribusi Kumulatif

Sering kali perhitungan probabilitas variabel acak Z yang berdistribusi normal standar lebih mudah dilakukan dengan memakai fungsi distribusi kumulatif. Bila variabel acak Z berdistribusi normal standar dengan fungsi padat probabilitas f(z), fungsi distribusi kumulatif dari Z yang ditulis F(z) dirumuskan sebagai berikut (Gunawan, 2007)

Rumus 8.6

$$F(z) = P(Z < z) = \int f(z)\, dz = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z2}\, dz$$

Sifat-sifat fungsi distribusi kumulatif F(z) adalah sebagai berikut ;

1.F(z) monoton naik

2.$0 \le F(z) \le 1$

3.$F(-\infty) = \lim_{X \to \infty} F(x) = 0$ dan $F(+\infty) = \lim_{X \to \infty} F(x) = 1$

Perhatikan bahwa grafik F(z) tidak memotong sumbu Z dan juga tidak memotong garis F9z) = 1. Oleh karena itu, sumbu Z dan garis F9z) adalah suatu garis batas dari grafik F(z). Dengan memakai fungsi distribusi kumulatif F(z), probabilitas P(z1 < Z < z2) dihitung dengan menggunakan rumus berikut :

Rumus 8.7

$$P(z1 < Z < z2) = P(Z < z2) - P(Z < z1) = F(z2) - F(z1)$$

Nilai-nilai probabilitas fungsi distribusi kumulatif dari distribusi normal standar terdapat dalam tabel distribusi kumulatif normal standar.

Contoh :

a.P(-1.43 < Z < 2.53) = F(2.53) – F(-1.43)
= F(2.50) – F(-1.40)
= 0.9938 – 0.0808
= 0.9130

b.P(-0.5 < Z < 1.3) = F(1.3) – F(-0.5)
= 0.9032 – 0.3085
= 0.5947

BAB V
PENGUJIAN HIPOTESIS

A. PENGUJIAN HIPOTESIS STATISTIK

Hipotesis Statistik : Suatu anggapan/pernyataan yang mungkin benar/salah mengenai satu atau lebih populasil.

Hipotesis dengan harapan untuk ditolak → Hipotesis nol (H_0)
Hipotesis tandingannya → Hipoteis satu/Hipotesis alternative (H_1)

Untuk mengetahui kebenaran suatu hipoteis: Pengambilan sample acak dari populasi → berdasarkan informasi yang diperoleh dalam sample terseut, diputuskan apakah suatu hipotesis benar/salah.

- Petunjuk dari sample yang tidak sesuai dengan hipotesis → penolakan hipotesis.
- Petunjuk yang mendukung hipotesis → penerimaan hipotesis.

Dalam pengujian hipotesis dapat membuat dua jenis kesalahan, yaitu:
(1) Kesalahan jenis kesatu (α)
σ = P (Kesalahan jenis I) = P (menolak H_0/H_0 benar)
(2) Kesalahan jenis kedua (β)
β = P (Kesalahan jenis II) = P (menerima H_0/H_0 benar)
Pengujian yang baik → α dan β sekecil-kecilnya.

Pada umumnya α ditentukan terlebih dahulu
<u>Contoh</u> : $\alpha = 0.05$; $\alpha = 0.01\alpha = 0.001$.

α = taraf keberartian dari suatu pengujian

$\alpha = 0.05$ berarti dalam 100 kali tolak H_0, ada 5 kali tola H_0 padahal H_0 benar atau memiliki tingkat kepercayaan 95%.

B. PENGUJIAN HIPOTESIS MENGENAI NILAI TENGAH

Jika peubah acak X menyebar Normal dengan nilai tengah (μ) dan ragam (δ^2), maka hipotesis yang perlu diuji adalah:

(1) H_0 : $\mu = \mu_0$ VS H_1 : $\mu > \mu_0$
(2) H_0 : $\mu = \mu_0$ VS H_1 : $\mu < \mu_0$
(3) H_0 : $\mu = \mu_0$ VS H_1 : $\mu \neq \mu_0$

(1) dan (2) : Pengujian eka-arah
(3) : Pengujian dwiarah

Apabila suatu statistik dengan parameter μ digunakan untuk menguji hipotesis bahwa parameter tersebut sama dengan suatu nilai μ_0 tertentu lawan suatu tandingan yang cocok, maka langkah pengujian tersebut adalah sebagai berikut:

(1) H_0 : $\mu = \mu_0$ adalah sebagai berikut:
(2) H_1 : tandingannya ($\mu > \mu_0$; $\mu < \mu_0$; atau $\mu \neq \mu_0$)
(3) Pilih taraf keberartian α
(4) Pilih uji statistik yang sesuai dan cari daerah kritisnya
(5) Hitunglah nilai statistik dari sample acak berukuran n
(6) Kesimpulan: Tolak H_0 bila nilai statistik uji tersebut berada dalam daerah kritis, sedangkan bila nilai tersebut jatuh di luar wilayah kritiknya terima H_0.

Latihan:

1. In Economics abstracts, July 1972, the mean length of abstract is 79.56 words with a standard deviation of 24.80. A random sample of thirty-two of the abstracts in German language has a mean length of 67.47 words. Is there any significant difference between the random sample of abstracts in German and the whole population.
2. Over a whole year of 52 weeks, the number of issues from a library was 30.000. In 10 weeks during the winter, the number of issues per week were found to be:
   650 693 750 726 804 735 751 751 687 762
3. A random sample of 30 shelves of geography books had amean number of 27.3 books per shelf and a standar deviation of 2.16. A random sample of 40 shelves of books on production had amean number of 32.0 books per shelf and a standard deviation of 6.04. Use a z-test to decide if the number of geography books per shelf is significantly less than from the number of books per shelf on production.

4. The number of issues of junior non-fiction on a random sample of days in May and November were:

May : 66 58 62 69 57 94 67 97 67 63 87 75

Nov : 64 89 52 74 79 63 78 89 64 52 63 47

Does there appear to be significant difference in demand the two months

BAB VI

REGRESI LINIER DAN KORELASI

A. Pengertian

Regresi Linier: hubungan antara variabel terikat/*dependent variable* (y) dengan variabel bebas/*independent variable* (x) yang dinyatakan dalam bentuk persamaan matematik yang bentuknya linier.

Kegunaan Regresi Linier: untuk membuat estimasi atau pendugaan nilai atau harga suatu variabel (variabel terikat) berdasarkan nilai atau harga variabel yang lain (variabel bebas).

B. Persamaan Regresi Linier

Model matemastik untuk persamaan regresi linier:

$Yi = \alpha + \beta \cdot x_i = e_i$

Dengan membuat $\Sigma\ e_i^2$ sekecil-kecilnya atau menggunakan metode kuadrat terkecil *(method of least square)*, diperoleh:

Y = a + b.x

Dimana:

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

$$a = \bar{y} - b.\bar{x}$$

Contoh:

Tabel 8. Nilai rata-rata tes masuk dan nilai statistik mahasiswa APP Bogor.

Mahasiswa	Nilai Rata-rata Tes Masuk (X)	Nilai Statistik (Y)
1	65	85
2	50	74
3	55	76
4	65	90
5	55	85
6	70	87
7	65	94

8	70	98
9	55	81
10	70	91
11	50	76
12	55	74

Tentukan:

a. Persamaan garis regresinya
b. Dugalah nilai statistik seorang mahasiswa apabila nilai tes rata-rata masuknya adalah 70

Jawab: ΣX_i = 725; ΣY_i = 1011; ΣYX = 61685

ΣX_i = 44475; X = 60,417; Y = 84,250

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

b = $\dfrac{(12)(61685) - (725)(1011)}{(12)(44475) - (725)^2} = 0.897$

a = $\bar{y} - b.\bar{x}$

= 84,250 – (0.897)(60.417) = 30.056

a. Persamaan garis regresi → y = 30.056 + 0.897x

b. x = 70 → y = 30.056 + 0.897 (70) = 92.846

jadi nilai statistiknya = 92.846

C. KOEFISIEN KORELASI (R)

Koefisien korelasi adalah ukuran hubungan linier antara dua variabel X dan Y. Koefisien korelasi diduga dengan koefisien korelasi contoh r, yaitu:

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

Nilai koefisien korelasi: $-1 \leq r \leq 1$

a. = 0 artinya antara variabel X dan variabel Y tidak ada hubungan/korelasi.
b. = -1 artinya ada hubungan/korelasi sempurna antara variabel X dan Y dengan koefisien arah negatip
c. = 1 artinya ada hubungan/korelasi sempurna antara variabel X dan Y dengan koefisien arah positip.

D. KOEFISIEN DETERMINASI (r^2)

Koefisien determinasi adalah bilangan yang menyatakan proporsi keragaman total nilai-nilai variabel Y yang dapat dijelaskan oleh nilai-nilai variabel X melalui hubungan linier tersebut. Nilai koefisien determinasi : $0 \leq r^2 \leq 1$.
Contoh: r = 0.80, maka r^2 = 0.64 persen, artinya bahwa 64 persen diantara keragaman dalam nilai-nilai variabel Y dapat dijelskan oleh hubungan liniernya dengan x.

Contoh: Tentukan nilai koefisien korelasi dan koefisien determinasnya, serta interpretasikan berdasarkan data pada Tabel. 8.

KEGIATAN PRAKTIKUM

1. Data luas panen (X) dan produksi ubikayu pada 10 kabupaten di Propinsi jawa timur tahun 1987 – 1988 adalah sebagai berikut:

Kabupaten	Luas Panen (Ha)	Produksi (ton)
Jombang	1.929	22.765
Mojokerto	2.269	27.977
Nganjuk	5.364	59.778
Tulungagung	6.597	74.136
Blitar	8.190	96.994
Pajuruan	11.600	133.909
Probolinggo	14.096	181.561
Trenggalek	16.038	206.213
Malang	20.545	262.837

Ponorogo	31.165	372.215

Pertanyaan:

a. Tentukan persamaan garis regresinya
b. Dugalah produksi ubikayu Kabupaten Bojonegoro, apabila luas panennya 20.410 ha!
c. Tentukan besarnya koefisien determinasi dan koefisien korelasinya! Jelaskan arti koefisien determinasi yang diperoleh!

LATIHAN REGRESI LINIER

1. Nilai laporan (x) dan nilai akhir (y) dari 9 mahasiswa adalah sebagai berikut:

X	77	50	71	72	81	94	96	99	67
X	82	66	78	34	47	85	99	99	68

a. Tentukan persamaan garis regresi liniernya dan gambarkan
b. Dugalah nilai akhir seorang mahasiswa yagn tidak ikut ujian, tetapi nilai laporannya 85

2. Data dibawah ini adalah suatu adalah suatu keterpakaian koleksi dengan jumlah koleksi yang dimiliki oleh sebuah perpustakaan. X adalah suatu jumlah koleksi perpustakaan dalam nilai ribuan, Y adalah suatu keterpakaiannya.

No.	X	Y
1.	1.09	24
2.	7.42	92
3.	4.20	67
4.	8.25	158
5.	8.81	81
6.	1.62	59
7.	3.84	54
8.	9.40	171
9.	3.63	100
10.	14.10	276
11.	2.50	122
12.	11.47	200

a. Tentukan persamaan regresi liniernya

b. Tentukan pula bagaimana korelasinya keduanya

LATIHAN PENGUJIAN HIPOTESIS

1. Sebuah perusahaan mengadakan penelitian tentang jumlah pengunjungnya. Hasil penelitian menunjukkan bahwa jumlah pengunjung menghampiri sebaran normal dengan nilai tengah 800 orang dan simpangan baku 40 orang. Ujilah hipotesis bahwa μ = 800 lawan alternatifnya μ = 800 orang, bila suatu contoh acak 30 hari hanya menghasilkan jumlah pengunjung rata-rata per hari 788 orang. Gunakan taraf nyata 0.04.
2. Suatu contoh acak berukuran n1 = 25, yang diambil dari suatu populasi normal dengan simpangan baku $\sigma 2$ = 3.4, memiliki nilai tengah x2 = 76. Ujilah
3. Dari sembilan contoh tanah yang diambil dari beberapa tempat ditentukan kandungan fosfor anorganiknya dalam ppm (X) dan juga kandungan fosfor yang tersedia bagi tanaman dalam ppm (Y). Data yang diperoleh adalah sebagai berikut:

Contoh Tanah	1	2	3	4	5	6	7	8	9
X (ppm)	1	4	5	9	13	13	23	27	28
Y (ppm)	64	70	71	81	93	96	97	115	119

Pertanyaan:

a. Berdasarkan data yang tersedia, berikan persamaan regresi antar Y sebagai peubah tak bebas dengan X sebagai peubah bebas !

b. Tentukan banyaknya rata-rata kandungan fosfor yang tersedia bagi tanaman pada suatu tanah yang kandungan fosfor anorganiknya sebesar 7,5 ppm !

c. Gambarkan diagram pencar dan garis regresinya !

LATIHAN:

1. In Economics abstracts, July 1972, the mean length of abstract is 79.56 words with a standard deviation of 24.80. A random sample of thirty-two of the abstracts in German language has a mean length of 67.47 words. Is there any significant

difference between the random sample of abstracts in German and the whole population !

2. Over a whole year of 52 weeks, the number of issues from a library was 30.000. In 10 weeks during the winter, the number of issues per week were found to be:
   650 693 750 726 804 735 751 751 687 762
   Perform at-test to determine whether the demand for books is significantly greater or less during the winter.
3. A random sample of 30 shelves of geography books had a mean number of 27.3 books per shelf and a standard deviation of 2.16. A random sample of 40 shelves of books on production had a mean number of 32.0 books per shelf and a standard deviation of 6.04. Use a z-test to decide if the number of geography books per shelf is significantly less than from the number of books per shelf on production.
4. The number of issues of junior non-fiction on a random sample of days in May and November were:
   May : 66 58 62 69 57 94 67 97 67 63 87 75
   Nov : 64 89 52 74 79 63 78 89 64 52 63 47
   Does there appear to be significant difference in demand the two months